Non-canonical kinases expand the RNA polymerase II CTD code

By

Corey M. Nemec

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: August 5, 2016

Thesis committee:
Aseem Z. Ansari (advisor), Professor of Biochemistry and The Genome Center
Mark E. Burkard, Professor of Medicine and the UW Carbone Cancer Center
Catherine A. Fox, Professor of Biomolecular Chemistry
Audrey P. Gasch, Professor of Genetics and The Genome Center
Peter W. Lewis, Professor of Biomolecular Chemistry and Epigenetics

Table of Contents:

Table of Figures:

Table of Tables:

Acknowledgements

First and foremost, I would like to thank Aseem for the privilege of working his lab. While we are all unified in our research as a transcription lab, the diversity of all our projects is immense. Being exposed to all these projects throughout the years has given me a much broader depth of knowledge than I would have received in other labs. I further learned from Aseem that science isn't just about doing experiments…it's about communication, both written and oral. Working with Aseem on papers and grants has shown me how to tell a compelling story. I will be forever thankful for his support and the well-rounded education he provided.

I am also thankful to Dr. Stephen Munroe at Marquette University, my undergraduate mentor. He taught me so much and encouraged me to follow my dreams. Working in his lab gave me a great appreciation for transcription…something that carried over into graduate school. Without his kindness, advice, and support, I doubt that I would have decided to pursue my Ph.D.

I would also like to thank my graduate committee for thoughtful discussion over the years. Their unique insight and perspective has greatly improved the quality of my work. I am especially grateful to Audrey for keeping me on track during our long-standing collaboration. Because our lab is so diverse, I was always excited to try new techniques other lab members have used…and sometimes deviated too far on tangents. From Audrey, I learned to keep the end goal in sight and always focus on getting there.

Next, I must thank current and former lab members. Seeing their faces every day always kept me excited to come to work. I am forever grateful to Juanito for teaching me almost everything I know about yeast and for being the best bay-mate I could have asked for. Sandy, Josefat, Dave, and Josh, the other "yeasties" also played an extensive role in getting me to where I am today. Whenever science wasn't working and I was in a sour mood, AnnaK was always there to

cheer me up. Graham, Asuka, Devesh, José, and I all joined the lab the same year. We struggled together, and celebrated together…and “synergized” really well. I am also thankful to Kristy Moeller, Kennedy Ringelberg, and Tanner Byer, undergraduate students who have assisted me over the last 6 years. Finally, I must also thank honorary lab member Laura Vanderploeg for teaching me how to use Adobe Illustrator to make scientific figures and posters.

I am also lucky to have such a loving and caring family support me in my endeavors over the past six years. As a child, instead of reading me a bedtime story, my dad would ask if I had any questions about science, engineering, etc. I think this is where my curiosity came from, and is likely part of the reason I became interested in science. My mom is one of the most caring and selfless people I know and has always supported me in every decision I made. I could not have asked for better parents. My sister is also a scientist, so she could easily relate to the stress I’ve experienced as a graduate student. Our talks always kept me grounded and helped me realize that everyone gets stressed out, but pushing through to the end is worth it. I’d also like to thank my extended family including grandparents, aunts, uncles, and cousins for all their support.

Regularly hanging out with all the friends I’ve made in Madison has maintained my sanity throughout the years. Whether it is was crazy night downtown, a day trip to New Glarus, a weekend trip to Milwaukee, or a trek in the Himalayas, it was always fun time. Many of them have already graduated and moved on to bigger and better things. I hope to join them soon!

Finally, I am thankful for my funding sources. Being a trainee in the Chemistry-Biology Interface Training program broadened my knowledge and provided me with unique feedback and suggestions from students with diverse backgrounds. I am grateful to Laura Kiessling, Helen Blackwell, and all my CBI colleagues. I am also thankful to the National Science Foundation for supporting me with a Graduate Research Fellowship.

Abstract

The carboxyl-terminal domain (CTD) of RNA polymerase II is a largely unstructured and highly repetitive domain consisting of the heptapeptides (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7) repeated in tandem 26 times in *S. cerevisiae*. A unique aspect of the CTD is that each residue can be post-translationally modified. Tyrosine, threonine, and the three serines can be phosphorylated, and the prolines can undergo cis/trans isomerization. The CTD serves as a scaffold whereby the placement and removal of phosphates spatiotemporally recruits machinery required for transcription initiation, elongation, termination, pre-mRNA splicing, and chromatin remodeling. Previously, only cyclin-dependent kinases (CDKs) were known to phosphorylate the CTD. This thesis describes the characterization of 21 novel kinases that can phosphorylate the CTD, many of which are in a group other than the CDKs (Chapter 2). We extend our studies on these non-canonical CTD kinases and identify a role for the CTD in directly responding to environmental stresses (Chapter 3). Using unique peptide arrays, we are able to characterize the specificity and the targets of any of each of these kinases (Appendix A). Finally, we develop a unique strategy to determine at which CTD repeats each kinase functions (Appendix B).

Chapter 2 describes the discovery of 21 new yeast and human Thr4 CTD kinases, many of which are part of kinases groups other than the “canonical” cyclin dependent CTD kinases. Chemical-genetic inhibition of select kinases in vivo or Thr-to-Ala substitutions in the CTD substrate revealed an essential role for phospho-Thr4 in transcription termination at specific gene classes. Unexpectedly, we identify the phospho-Ser2 binding termination factor, Rtt103, as a direct reader of phospho-Thr4 marks. Our data suggest that non-canonical CTD kinases place phospho-Thr4 marks to engage the termination machinery at select genes whereas phospho-Ser2, placed by canonical CTD kinases, functions as a general mark to terminate Pol II-dependent transcription.

Further examination of the kinases identified in Chapter 2 suggest a role of the CTD in responding to environmental stresses. Previous work suggests that signaling cascades of kinases modify transcription factors and chromatin remodelers to fine tune gene expression; however, little has been done to explore the connectivity between these stimuli and RNA Pol II itself, in part, because these inferred links were assumed to be indirect. In Chapter 3, we show that in response to diverse signals, activated kinases directly phosphorylate the carboxyl-terminal domain (CTD) of Pol II. Further, CTD phosphorylation is critical for proper response to osmotic stress. Thus, we propose a mechanism in which multiple signals all converge at the CTD to quickly modulate the transcriptome, yielding a novel branch in transcriptional regulation.

To identify the specificity and targets of each of these kinases, in Appendix A, we utilize high-density peptide arrays. Tiled peptides from yeast and human transcriptional machinery, as well as combinatorial substitutions to the CTD substrate were synthesized on arrays via maskless array synthesis. Phosphorylation of peptides with Kin28 confirms its target is the canonical CDK motif, even though yeast CTD lack this motif in nature. This platform will allow for an in depth survey of kinase specificity, and can determine how the modification of specific residues alters subsequent kinase activity.

In Appendix B, we utilize a method to covalently crosslink CTD “writers” and “readers” directly to their targets. Covalent crosslinking not only allows for direct evidence of interaction, but spatial evidence as well. Further, by engineering the covalent crosslinker into the active sites of CTD “writers” (kinases), we are also able to identify novel targets of the kinase, allowing for a more complete understanding how individual kinases regulate biological processes in the cell.

In Appendix C, we developed an experimental and computational approach to integrate available protein interaction data with gene fitness contributions, mutant transcriptome profiles, and

phospho-proteome changes in cells responding to salt stress, to infer the salt-responsive signaling network in yeast. The inferred subnetwork presented many novel predictions by implicating new regulators, uncovering unrecognized crosstalk between known pathways, and pointing to previously unknown ‘hubs’ of signal integration. We exploited these predictions to show that Cdc14 phosphatase is a central hub in the network and that modification of RNA Polymerase II coordinates induction of stress-defense genes with reduction of growth-related transcripts.

Finally, Chapter 4 outlines future directions in the CTD field. Single molecule sequencing of CTD marks, as well single cell observation of transcription dynamics in response to stress are discussed.

Chapter 1: Introduction

This chapter has been adapted from "Emerging views on the CTD code." by David W. Zhang, Juan B. Rodríguez-Molina, Joshua R. Tietjen, Corey M. Nemec, and Aseem Z. Ansari. (Zhang et al., 2012b).

1.1 Introduction

The transcription of DNA to RNA in eukaryotes is catalyzed by three structurally related RNA polymerases, with each acting on a different class of genes (Cramer et al., 2008). RNA Polymerase I synthesizes most of the ribosomal RNA (rRNA) subunits while RNA Polymerase III synthesizes tRNAs, 5S rRNA, and other small RNAs (Dieci et al., 2007; Grummt, 2003; Russell and Zomerdijk, 2005). These two polymerases account for 75% and 15% of transcription in the cell, respectively (Werner et al., 2009). However, the most studied polymerase is RNA Polymerase II (Pol II), which is responsible for the transcription of protein-coding genes, small nuclear RNA (snRNA), and small nucleolar RNA (snoRNA) (Davis and Ares, 2006; Wyers et al., 2005). In higher eukaryotes, Pol II generates long non-coding RNA (lncRNA) and microRNA (miRNA) (Faller and Guo, 2008; Lykke-Andersen and Jensen, 2007; Mitchell Guttman et al., 2009). Pol II also transcribes cryptic unstable transcripts (CUTs) and stable unannotated transcripts (SUTs), which are degraded after synthesis (Arigo et al., 2006; Thiebaut et al., 2006). The suppression of CUTs is important to prevent inappropriate transcription within ORFs, to enhance processivity during transcription elongation, and to prevent gene silencing via histone deacetylation (Camblong et al., 2007; De Santa et al., 2010; Kaplan et al., 2003; Ørom et al., 2010; Sato et al., 2011; Ying and Lin, 2009).

Of the twelve Pol II subunits, five are common between the three polymerases (Cramer et al., 2008; Shpakovski et al., 1995; Woychik and Young, 1994; Young, 1991). It is believed that the specific functions attributed to each polymerase arise from the combined action of remaining non-identical subunits and other factors that associate with them. An especially unique feature of Pol II is the carboxyl-terminal domain (CTD) of its large subunit Rpb1 (Figure 1.1A). The CTD serves as the primary point of contact for a wide variety of molecular machines involved in RNA biogenesis during the transcription cycle (reviewed in (Buratowski, 2009; Egloff and Murphy, 2008; Hirose and Manley, 2000; Hirose and Ohkuma, 2007; Kornberg, 2007; Lykke-Andersen

and Jensen, 2007; Nechaev and Adelman, 2011; Perales and Bentley, 2009; Phatnani and Greenleaf, 2006; Richard and Manley, 2009; Sims III et al., 2004; Venters and Pugh, 2009b)). This domain consists of a highly conserved heptapeptide repeat: $Y_1S_2P_3T_4S_5P_6S_7$ (Chapman et al., 2008; Corden, 1990; Corden et al., 1985; Liu et al., 2010). The number of times this sequence is repeated varies among eukaryotic organisms, ranging from 15 repeats in amoeba, to 26 repeats in the budding yeast *Saccharomyces cerevisiae* to 52 repeats in humans. When fully extended, the yeast CTD can span a distance of up to 650 Å, over 4 times the diameter of the core polymerase (Figure 1.1B) (Chapman et al., 2008; Corden, 1990; Phatnani and Greenleaf, 2006).

The ability of this repetitive sequence to interact with a wide range of nuclear factors stems from the dynamic plasticity of its structure and the diversity of binding surfaces generated by the multitude of post-translational modifications it can accommodate. Tyrosine, threonine, and three serines can all be phosphorylated, the threonine and serine can be glycosylated, and the prolines can undergo isomerization (Figure 1.2B) (Egloff and Murphy, 2008; Fuchs et al., 2009; Zeidan and Hart). In humans, CTD repeats further away from core Pol II bear non-canonical repeats that can be methylated (Figure 1.2A) (Sims et al., 2011) (described further in Chapter 4). Taken together, at least 10^{59} unique modification patterns can occur on the CTD. The combinatorial nature of these modifications, which is reminiscent of the histone code, led to the hypothesis of a CTD code, where the patterns of modifications are read by the transcriptional machinery and these patterns dictate the association or disassociation of complexes (Buratowski, 2003; Jenuwein and Allis, 2001). To date, much effort has been made towards characterizing these modifications and understanding the interactions between the CTD and components of various protein machines that play a role in RNA biogenesis. Our current knowledge of the integration of these events by Pol II CTD is summarized in Figure 1.3, and the known yeast CTD-interacting factors are displayed in Table 1.1.

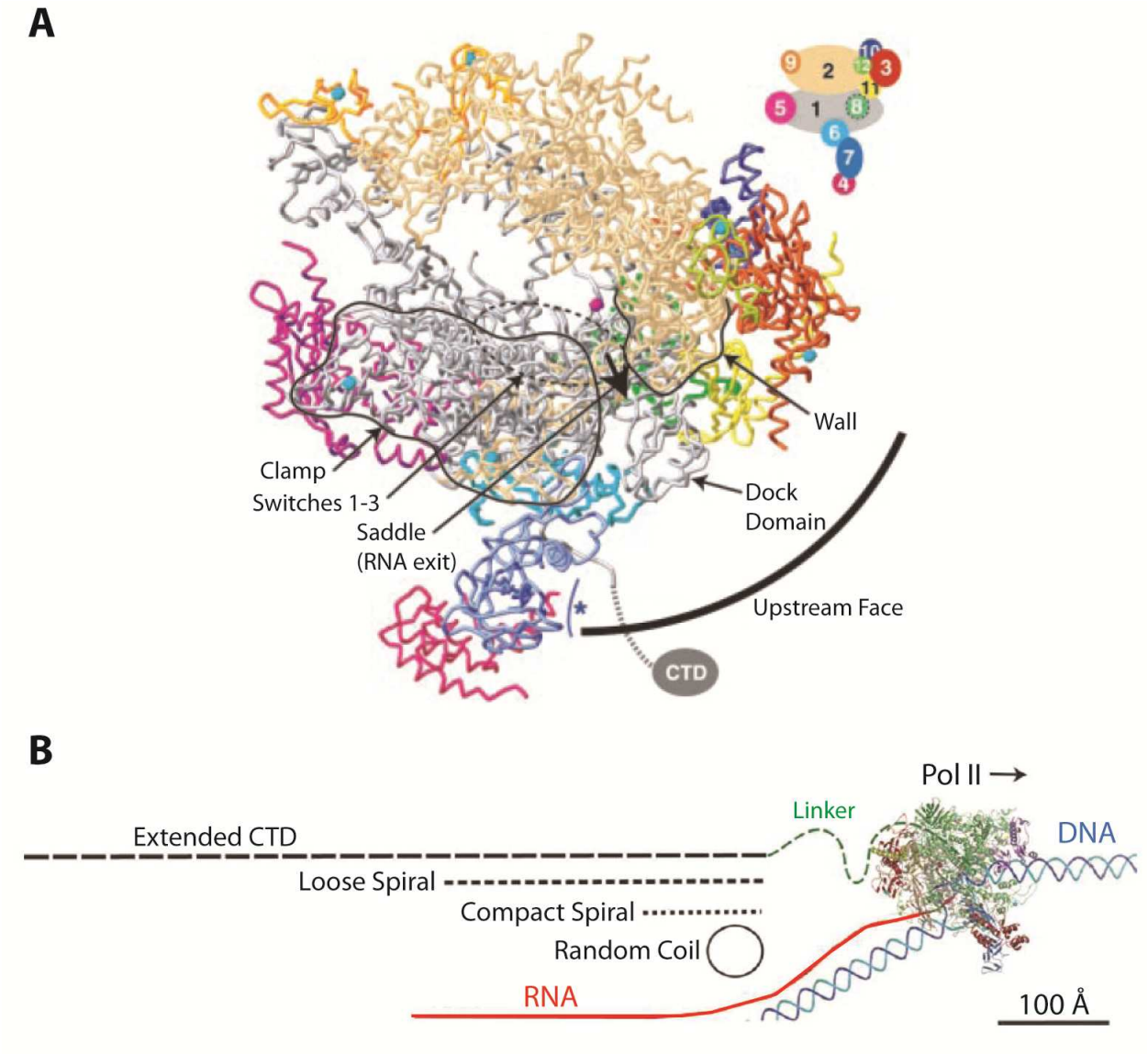

Figure 1.1 RNA polymerase II structure

A) Side view of the core Pol II crystal structure containing all twelve subunits and displaying the RNA exit channel (bold arrow) and the positioning of the CTD adapted from (Armache et al., 2003). Cartoon in the upper right displays the color coding for the Pol II subunits used in the crystal structure.

B) Illustration of the relative length(s) between the CTD in various conformations and the core Pol II adapted from (Meinhart et al., 2005). RNA positioning (red) upon exit of the Pol II and the positioning of the DNA template (blue) upstream and downstream of the core Pol II are also displayed.

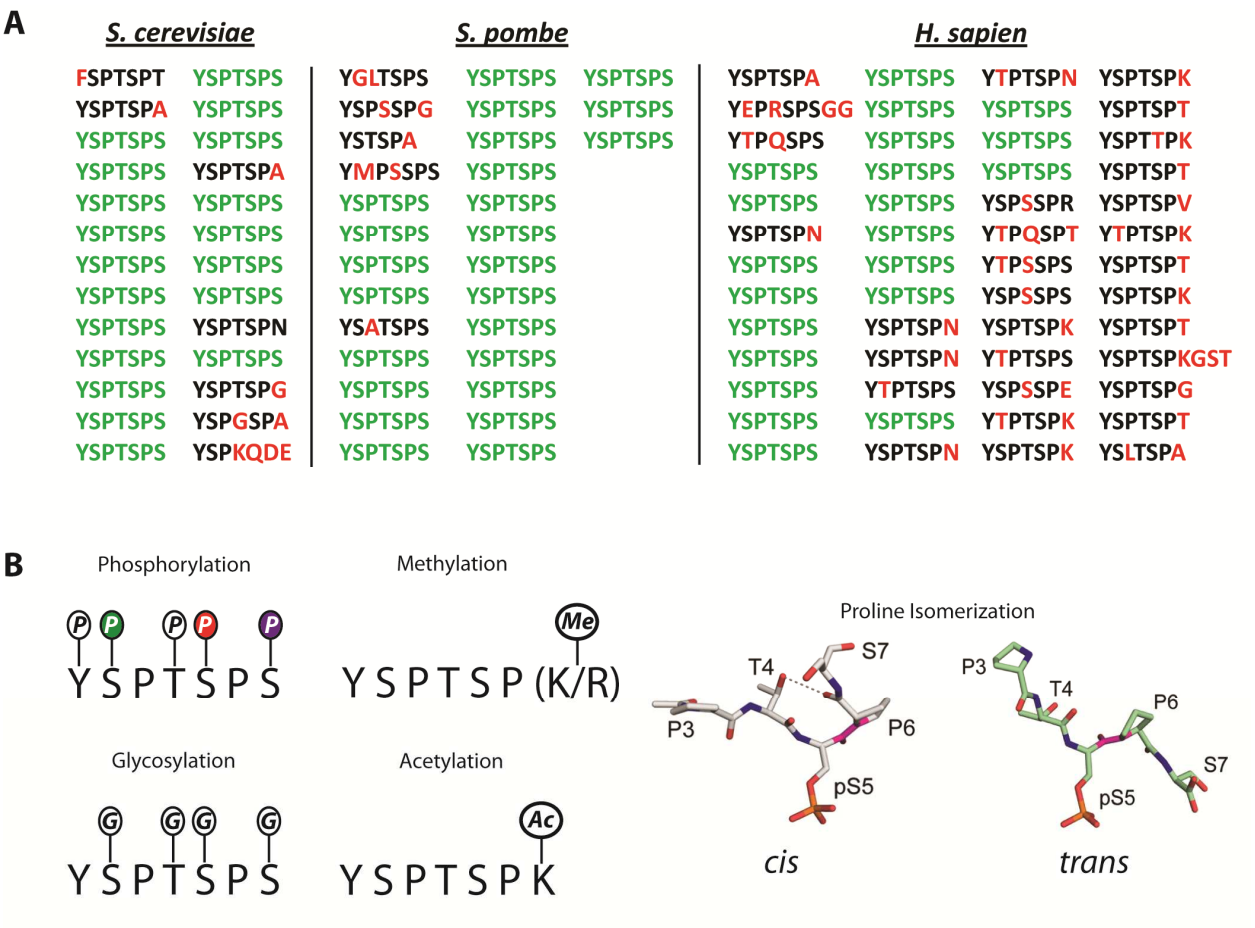

Figure 1.2 RNA polymerase II CTD sequence and known modifications

A) Amino acid sequence of the RNA polymerase II CTD from *S. cerevisiae, S. pombe,* and *H. sapiens*. Repeats highlighted in green represent the consensus CTD sequence (YSPTSPS) while amino acids in red represent deviations from the consensus.

B) Known modifications possible on the RNA polymerase II CTD. Glycosylation and phosphorylation are mutually exclusive modifications. Methylation and acetylation occur exclusively on non-canonical residues at position 7. Structural images of a heptad repeat in the *cis* and *trans* conformation are also shown (Meinhart and Cramer, 2004; Werner-Allen et al., 2011; Xiang et al., 2010). G = β-O-linked N-acetylglucosamine (Kelly et al., 1993); P = O-linked Phosphate.

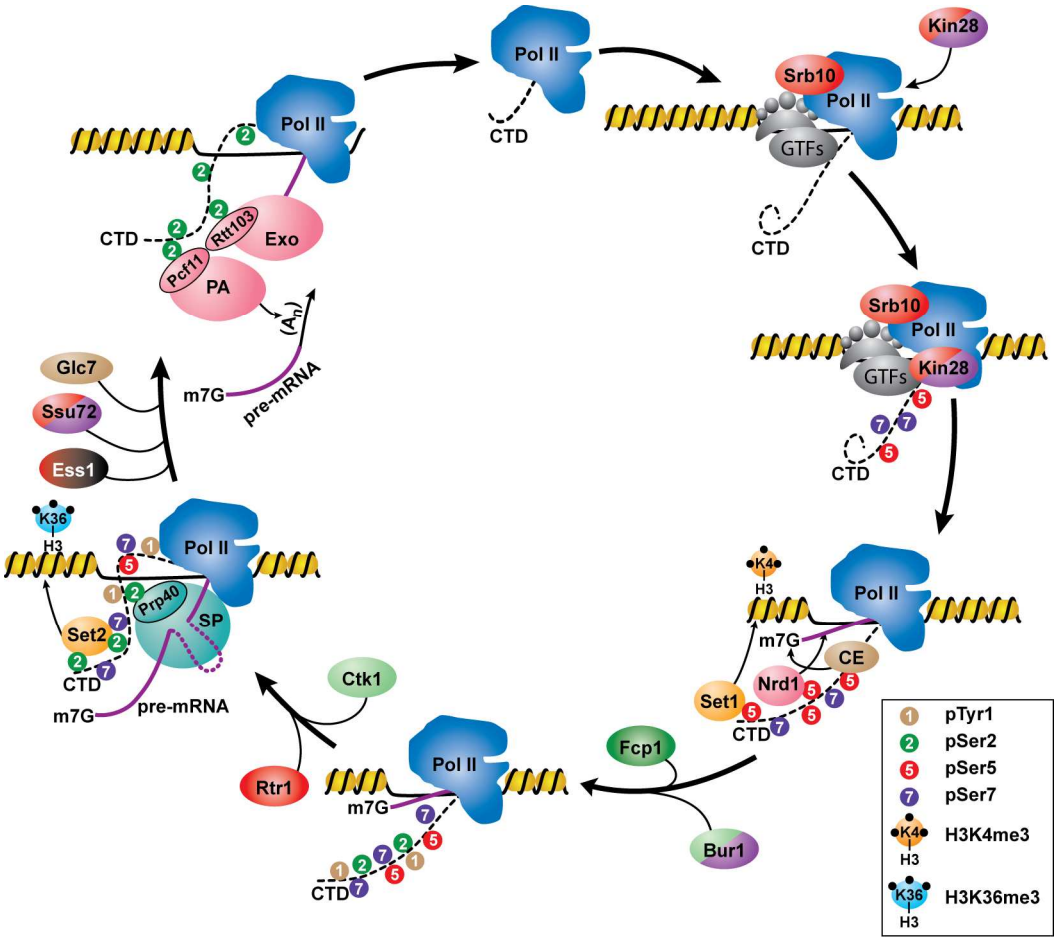

Figure 1.3 The primary components of the RNA biogenesis machinery and their interactions with the RNA polymerase II C-terminal domain (CTD)
Hypophosphorylated Pol II assembles at the pre-initiation complex (PIC) with the Mediator and general transcription factors (GTFs), with TFIIH associating last. The TFIIH-associated kinase Kin28 phosphorylates Ser5 (red) and Ser7 (purple) on the CTD. Mediator-associated kinase Srb10 also contributes to the phosphorylation of pSer5. This mark enables promoter release and mediates interactions with the capping enzyme (CE) complex, Nrd1 component of termination machinery, and Set1 histone methyltransferase, which places trimethyl marks on histone H3K4. The pSer5 mark also facilitates recruitment of Bur1 kinase. Bur1 places initial pSer2 marks, which facilitate recruitment of Ctk1 kinase, and continues to replenish pSer7 marks during elongation. Ctk1 is the primary Ser2 kinase, and its phosphorylation recruits splicing machinery (SP) through Prp40, as well as Set2 histone methyltransferase, which places di- and trimethyl marks on histone H3K36. Cleavage and polyadenylation (PA) machinery are recruited through many factors associating with the CTD. One of the factors, Pcf11, binds cooperatively to pSer2 with Rtt103 and requires dephosphorylation of Tyr1 via Glc7. The exonuclease complex (Exo) is also recruited through interaction between CTD and Rtt103 and through cooperative interaction between Rtt103 and Pcf11. Finally, the hypophosphorylated CTD is regenerated through four CTD phosphatases. pSer2 is removed by the phosphatase Fcp1, Rtr1 and Ssu72 dephosphorylate Ser5 during elongation and termination, respectively. Ssu72 dephosphorylates Ser7, and Glc7 dephosphorylates Tyr1 during termination. Upon dephosphorylation, Pol II is released with the assistance of a mechanism involving Pcf11 and can begin another cycle of transcription.

Table 1.1 Proteins known to bind RNA polymerase II C-terminal domain in *S. cerevisiae*

Protein/ Complex	Role in RNA Biogenesis	Phospho-CTD Bound	References
TFIIE	Pre-initiation Complex	Hypophosphorylated CTD	(Kang and Dahmus, 1995; Maxon et al., 1994)
TFIIF	Pre-initiation Complex	Hypophosphorylated CTD	(Kang and Dahmus, 1995)
TBP	Pre-initiation Complex (TFIID)	Hypophosphorylated CTD	(Usheva et al., 1992)
Mediator Complex	Transcription Activation/Repression	Hypophosphorylated CTD	(Myers et al., 1998),(Svejstrup et al., 1997)
Ceg1	Capping	pSer5	(Cho et al., 1997; Fabrega et al., 2003; Ho and Shuman, 1999; Komarnitsky et al., 2000; McCracken et al., 1997; Schroeder et al., 2000)
Abd1	Capping	PCTD	(Komarnitsky et al., 2000)
Set1	Histone Methylation	pSer5	(Ng et al., 2003)
Rpd3C (Rco1)	Histone Deacetylation	pSer2+pSer5	(Govind et al., 2010), (Drouin et al., 2010)
Spt6	Histone Chaperone	pSer2	(Yoh et al., 2007)
Nrd1	Transcription Termination/Processing	pSer5	(Vasiljeva et al., 2008)
Sen1	Transcription Termination/Processing	Unknown	(Finkel et al., 2010)
Asr1	Pol II Ubiquitylation	pSer5	(Daulny et al., 2008)
Ess1	Proline Isomerase	pSer2	(Morris et al., 1999; Wu et al., 2000)
Set2	Histone Methylation	pSer2+pSer5	(Kizer et al., 2005; Vojnic et al., 2006)
Prp40	Splicing	PCTD	(Morris and Greenleaf, 2000)
Npl3	Promotes Elongation/Prevents Polyadenylation	pSer2	(Dermody et al., 2008)
Pcf11	Cleavage/Polyadenylation (CF1A)	pSer2	(Hollingworth et al., 2006; Noble et al., 2005)
Rna14	Cleavage/Polyadenylation (CF1A)	PCTD	(Barillà et al., 2001)
Rna15	Cleavage/Polyadenylation (CF1A)	PCTD	(Barillà et al., 2001)
Ydh1	Cleavage/Polyadenylation (CPF)	PCTD	(Kyburz et al., 2003)
Yhh1	Cleavage/Polyadenylation (CPF)	PCTD	(Dichtl et al., 2002b)
Pta1	Cleavage/Polyadenylation (CPF)	pSer5	(Rodriguez et al., 2000)
Rtt103	5'-3' Exonuclease (Rat1)	pSer2	(Kim et al., 2004b)
Sus1	mRNA export	pSer5	(Pascual-García et al., 2008)
Yra1	mRNA export	Hyperphosphorylated CTD	(MacKellar and Greenleaf, 2011)
Rsp5	Pol II Ubiquitylation (DNA Damage Response)	pSer2	(Chang et al., 2000) (Somesh et al., 2005)
Hrr25	DNA Damage Repair	PCTD	(Phatnani and Greenleaf, 2006; Phatnani et al., 2004)

CTD interacting proteins, the processes they are involved in, the phosphorylation state of the CTD with which they associate, and where in the literature the interaction is documented. pSer2 refers to phosphorylated Serine 2, pSer5 refers to phosphorylated Serine 5, and PCTD refers to a mixed phosphorylation state generated by *in vitro* phosphorylation of a CTD peptide with cell extracts. Additional protein-CTD interactions are described (Phatnani et al., 2004), but have not been directly tested.

1.2 Transcription initiation

Initiation of transcription begins with the recruitment of gene-specific transcription factors (TFs), general transcription factors (GTFs), the Mediator complex, and Pol II. These factors self-assemble into a pre-initiation complex (PIC) at the promoters of Pol II-transcribed genes (Buratowski, 2009; Nechaev and Adelman, 2011). Recognition of the promoter is only partially understood, but it is believed to occur via the recognition of the various *cis*-elements in the promoter region, such as the TATA box. Binding generally occurs within upstream nucleosome-free regions – the DNA centered over promoters flanked by well-positioned nucleosomes (Basehoar et al., 2004; Pedersen et al., 1999; Venters and Pugh, 2009a; Yuan et al., 2005). There are two main models for how these factors assemble at this region: the sequential model and the holoenzyme model (Figure 1.4). In both models, TFs first bind at the upstream activating/repressing sequences (UAS/URS) and recruit the transcriptional machinery. In the sequential model, TBP/TFIID/SAGA assembly at the promoter is accompanied by TFIIA, followed by TFIIB (Orphanides et al., 1996; Zanton and Pugh, 2006). Then, the Mediator complex arrives, connecting the PIC to transcription factors assembled at the UAS/URS (Asturias et al., 1999; Kornberg, 2005; Malik and Roeder, 2005; Myers and Kornberg, 2000; Svejstrup et al., 1997). This massive complex consists of three large modules known as the head, middle, and tail, and an additional kinase module containing a cyclin-dependent kinase (Srb10 in yeast, Cdk8 in metazoans) (Borggrefe et al., 2002; Casamassimi and Napoli, 2007; Chadick and Asturias, 2005; Guglielmi et al., 2004; Kang et al., 2001; Toth-Petroczy et al., 2008). The Mediator complex is important for basal transcription and plays a central role in facilitating communication between transcription factors bound to regulatory elements and the PIC (Biddick and Young, 2005; Bjorklund and Gustafsson, 2005; Casamassimi and Napoli, 2007; Conaway et al., 2005; Kornberg, 2005; Malik and Roeder, 2005; Myers and Kornberg, 2000; Toth-Petroczy et al., 2008). However, there are studies that suggest the Mediator is not present at most genes, and it only associates with a few UAS/URS in an activator- and stress-

specific manner (Fan et al., 2006; Fan and Struhl, 2009). Pol II is then recruited, followed by the last GTF, TFIIH, which is brought to the PIC by TFIIE (Maxon et al., 1994). It is possible that several pathways of ordered recruitment exist for GTFs. Other components, including Pol II, TFIIE, and TFIIH, may be recruited via interactions with the Mediator (Esnault et al., 2008). The holoenyzme model originated from the observation that Srb proteins, which are components of the Mediator, are tightly associated with core Pol II in the absence of DNA (Koleske and Young, 1994). In this model, Pol II is associated with the Mediator and other general transcription factors as a massive holoenzyme supercomplex that is recruited immediately after TBP binds (Barberis et al., 1995; Hengartner et al., 1995; Liao et al., 1995). These complexes have been identified in yeast and mammalian systems (Chao et al., 1996). Importantly, Pol II is fully able to activate transcription upon arrival in this state (Barberis et al., 1995; Koleske and Young, 1995).

Two complexes of the PIC, TFIIH and the Mediator, contain important kinases that phosphorylate the CTD. TFIIH is a ten-subunit complex containing two helicases, an ATPase, a ubiquitin ligase, a neddylation regulator, and a cyclin-dependent kinase (Kin28 in yeast, Cdk7 in metazoans) (Chang and Kornberg, 2000; Hengartner et al., 1998; Myer and Young, 1998; Rabut et al., 2011; Svejstrup et al., 1994; Svejstrup et al., 1995; Takagi et al., 2005; Tirode et al., 1999). Both Kin28/Cdk7 and Srb10/Cdk8 have been shown to phosphorylate Ser5 (pSer5) *in vivo*, with Kin28/Cdk7 being the dominant kinase (Bensaude et al., 1999; Dahmus, 1996; Feaver et al., 1994; Gebara et al., 1997; Hengartner et al., 1998; Palancade and Bensaude, 2003; Phatnani and Greenleaf, 2006; Rickert et al., 1999).

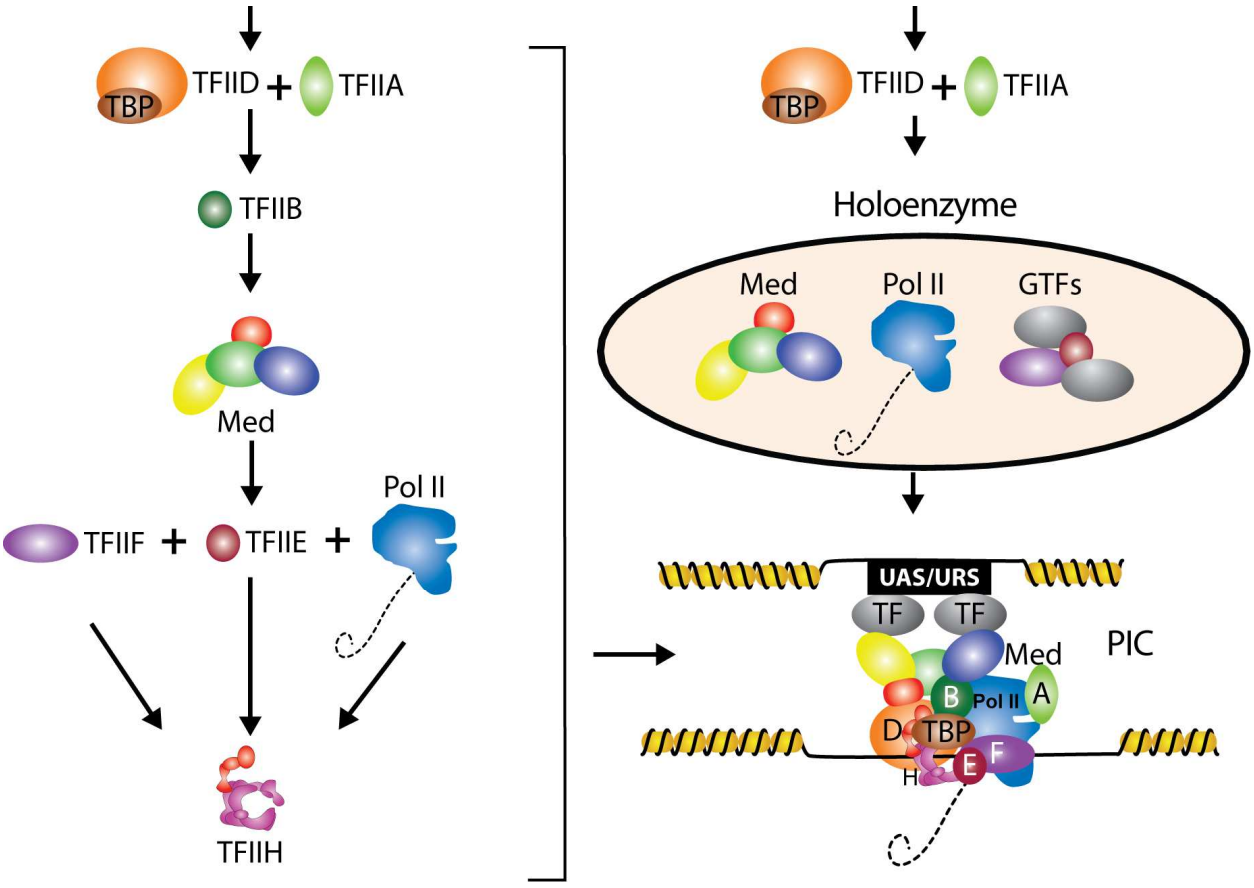

Figure 1.4 Recruitment and composition of PIC components
Sequential recruitment of the Mediator complex, GTFs, and Pol II (left) or the recruitment of the Pol II holoenzyme (top right), which assembles the pre-initiation complex (PIC) at promoters (bottom right).

1.3 Transcription elongation

Phosphorylation of Ser5 is involved in coordinating the placement of several key post-translational modifications on chromatin that constitute the histone code (Jenuwein and Allis, 2001) (reviewed in (Lee et al., 2010; Misri et al., 2008; Munshi et al., 2009)). The structural properties of chromatin, such as the +1 nucleosome that resides immediately after gene promoters, is thought to provide a significant physical barrier to transcription. This barrier is weakened or removed through the combined action of post-translational modifications on the flexible histone tails and chromatin remodeling complexes (Workman, 2006). In this context, the pSer5 mark recruits the yeast histone methyltransferase Set1. Trimethylation of histone H3K4 by Set1 and subsequent trimethylation of H3K79 by Dot1 are frequently associated with active transcription and have a reciprocal effect on H3K14 acetylation by SAGA and NuA3 (Nakanishi et al., 2008; Ng et al., 2003; Venters and Pugh, 2009b; Wood et al., 2003). pSer5 also recruits the histone deacetylase complexes Set3 and Rpd3C(S) (Govind et al., 2010), which are important in suppressing CUT initiation at promoters (Drouin et al.; Govind et al., 2010).

An especially important role of pSer5 is the recruitment of the capping enzyme complex. The capping complex places the m^7G cap on the nascent transcript as it exits the core polymerase, stabilizing the mRNA by preventing its degradation by 5'-3' exonucleases. The CTD repeats proximal to the core Pol II are ideally placed near the RNA exit tunnel to facilitate this capping reaction (Cramer et al., 2001; Ghosh et al., 2011). The guanylyltransferase (Ceg1 in *S. cerevisiae*) and possibly the methyltransferase (Abd1 in *cerevisiae*) directly interact with both the pSer5 and the core polymerase (Cho et al., 1997; Fabrega et al., 2003; Fong and Bentley, 2001; Ho and Shuman, 1999; Komarnitsky et al., 2000; McCracken et al., 1997; Moteki and Price, 2002; Schroeder et al., 2000). Although the recognition of the CTD is structurally different between yeast and mammalian capping enzymes, both complexes require pSer5 for binding (Fabrega et al., 2003; Ghosh et al., 2011). A parallel line of experiments showed that inhibition

of Kin28 kinase activity using a small molecule inhibitor leads to a severe reduction in pSer5 and 5'-capping of transcripts at gene promoters (Kanin et al., 2007; Liu et al., 2004). In agreement with this, tethering the mammalian capping enzyme to the CTD rescues the null Ser5 to alanine mutants in the fission yeast *Schizosaccharomyces pombe* (Schwer, 2011). Interestingly, inactivation of Kin28 does not eliminate transcription: neither steady state mRNA levels nor the ability to initiate transcription at the inducible *GAL1* gene are significantly compromised by the inhibition (Kanin et al., 2007). A subsequent study using the same chemical inhibition system confirmed the earlier observations but incorrectly attributed small differences in transcript levels to inappropriate normalization of earlier microarray data (Hong et al., 2009). No such global normalization was performed by Kanin et al. (Kanin et al., 2007), and it is unclear why the subsequent study (Hong et al., 2009) made the unsubstantiated and erroneous claim that the data was treated incorrectly. Kanin et al. were quite cognizant of the consequences of inhibiting an enzyme that could have a role in global transcription. Moreover, quantitative PCR and Northern blot assays, experiments that were not reliant on microarray normalization, showed little difference in expression (Pei and Ansari, 2007 unpublished data) (Kanin et al., 2007). These results strongly support the conclusion that inactivating Kin28 does not significantly impact global transcription. It is important to note that these studies only focused on chemical inhibition of Kin28 and that the inhibition is not an "all or none" phenomenon due to equilibrium binding of the small molecule to the kinase; it is possible that extremely low levels of Ser5 phosphorylation, by either Srb10 or residual Kin28, suffice for transcription initiation.

We and others have recently demonstrated that Kin28/Cdk7 is also the primary kinase that phosphorylates Ser7 (pSer7) (Akhtar et al., 2009; Boeing et al., 2010). The phosphorylation occurs at protein-coding and non-coding genes and seems to be Mediator-dependent (Boeing et al., 2010). Cyclin-dependent kinases are thought to prefer a substrate bearing Ser-Pro rather

than Ser-Tyr dipeptides (Songyang et al., 1996). Additionally, while Kin28 has been localized to promoters (Komarnitsky et al., 2000), pSer7 marks were thought to be found only at non-coding genes and at the 3' end of protein coding genes(Chapman et al., 2007; Egloff et al., 2007). While the role of pSer7 at promoters remain an active area of investigation, the 5' enriched pSer5 mark has been linked to a variety of chromatin-modifying and RNA processing events.

Chemical inhibition of Kin28 and Srb10 shows a drop in Pol II across the ORF, supporting the model where pSer5/pSer7 may help in promoter clearance. Following promoter clearance, transcription initiation factors are exchanged for transcription elongation factors required for RNA processing, passage through chromatin, and suppressing cryptic transcripts. In budding yeast, this exchange occurs immediately after the +1 nucleosome (Mayer et al., 2010). The association of these elongation factors, which include Paf1, Spt16, Spt4, Spt5, Spt6, Spn1, and Elf1, occur concurrently on all Pol II genes and is independent of gene length, type, or expression (Mayer et al., 2010). The recruitment of these factors is essential for transcription processivity (Spt4/5) (Grohmann et al., 2011; Hartzog et al., 1998; Martinez-Rucobo et al., 2011), histone regulation (Spt6/16, Spn1, Elf1) (Adkins and Tyler, 2006; Jamai et al., 2009; McDonald et al., 2010; Orphanides et al., 1999; Prather et al., 2005; Youdell et al., 2008; Zhang et al., 2008), and gene activation/3' processing (Paf1) (Jaehning, 2010). Similarly, mammalian P-TEFb complex is recruited to Pol II at this stage of transcription (Brès et al., 2008; Peterlin and Price, 2006; Price, 2000; Viladevall et al., 2009). This complex contains a cyclin-dependent kinase (Cdk9) that phosphorylates the DRB-sensitivity inducing factor (DSIF), which allows Pol II to overcome the promoter-proximal pausing induced by the negative elongation factor (NELF) complex (Peterlin and Price, 2006; Sims III et al., 2004). It is unclear if promoter-proximal pausing occurs in yeast, but it is known that Bur1 (the yeast homolog of Cdk9) promotes elongation through post-translational modification of Spt5 (DSIF) (Figure 1.5A) (Zhou et al., 2009). Bur1 also improves transcription elongation through the recruitment of histone modifying

enzymes and the phosphorylation of CTD. Bur1 activity promotes the ubiquitylation of H2BK123 by the ubiquitin conjugating enzyme Rad6 and Bre1 (Wood et al., 2003, 2005). H2BK123Ub promotes Set1 trimethylation of histone H3K4 and subsequent trimethylation of H3K79, both of which are important for transcription activation (Nakanishi et al., 2008; Ng et al., 2003; Venters and Pugh, 2009b; Wood et al., 2003). Bur1 also promotes transcription elongation by coupling promoter-proximal CTD modifications with promoter-distal marks. Bur1 is recruited to the transcription complex by the pSer5 marks placed at the promoter. It then phosphorylates Ser2 (pSer2), priming the CTD for the recruitment of Ctk1 (Cdk12), the major Ser2 kinase (Jones et al., 2004). Initial CTD phosphorylation also increases the activity of Ctk1, thereby coupling sequential CTD modifications (Figure 1.5B) (Kim and Sharp, 2001; Peterlin and Price, 2006; Qiu et al., 2009; Sims III et al., 2004). Interestingly, Bur1 travels with Pol II and phosphorylates Ser7. Although the exact role of this modification is unclear, it is likely a mark that promotes elongation, as genes with uniformly high levels of pSer7 are transcribed at significantly higher levels (Tietjen et al., 2010).

Most pSer5 marks are removed near the +1 nucleosome through the action of the newly characterized CTD phosphatase Rtr1 (Mosley et al., 2009). This phosphatase has been shown to specifically remove pSer5 marks immediately after promoter clearance. The pSer2 phosphatase Fcp1 also associates during elongation, but pSer2 levels remain high across the transcript due to the opposing action of the pSer2 kinase Ctk1 (Cho et al., 2001; Kobor et al., 1999). It is thought that the Ubp8 component of SAGA travels with Pol II and promotes deubiquitylation of H2BK123Ub (Henry et al., 2003), which allows the association of Ctk1 and subsequent phosphorylation of Ser2 on the CTD (Wyce et al., 2007).

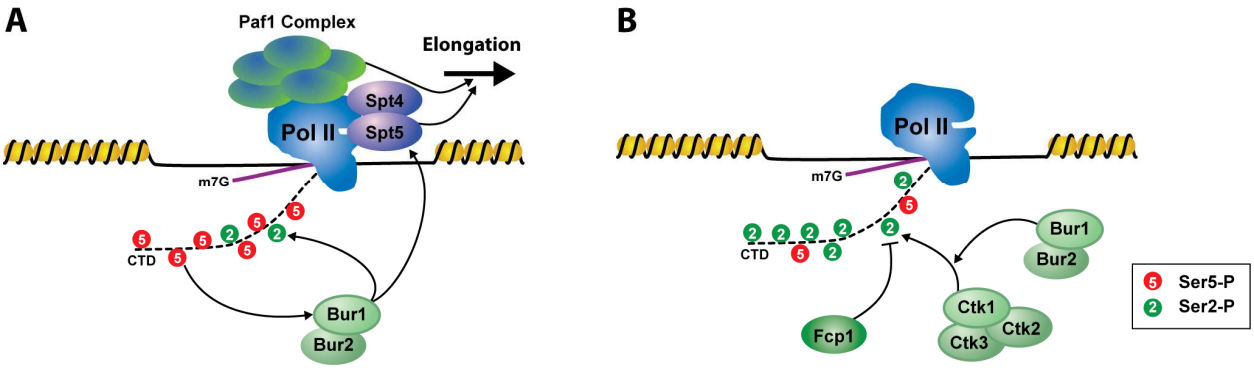

Figure 1.5 Bur1 phosphorylation of the CTD facilitates the transition from initiation to elongation

A) pSer5 enhances recruitment and subsequent phosphorylation of Ser2 by Bur1. Bur1 also phosphorylates Spt5, which acts with the Paf1 complex to promote elongation.

B) CTD phosphorylation by Bur1 enhances the activity of Ctk1 on Ser2. The majority of the pSer2 is maintained by competition between phosphorylation by Ctk1 and dephosphorylation by Fcp1. This increase in pSer2 facilitates recruitment of many pSer2-binding proteins, such as Npl3.

pSer2 is critically important for the interaction between the CTD and many histone modifying and RNA processing machines (Ahn et al., 2004; Bentley, 1999, 2002; Bentley, 2005; Buratowski, 2005; Fong and Bentley, 2001; Komarnitsky et al., 2000; Licatalosi et al., 2002; Meinhart and Cramer, 2004; Proudfoot et al., 2002). Increasing levels of pSer2, in combination with the residual pSer5, promote the recruitment of the Set2 methyltransferase, which catalyzes the formation of H3K36me2 and H3K36me3 (Kizer et al., 2005; Krogan et al., 2003; Li et al., 2003; Li et al., 2002; Vojnic et al., 2006). This leads to the recruitment of the histone deacetylase complex Rpd3C(S) and the removal of acetylation from histones H3 and H4, thereby resetting the transcription state of the nucleosomes and repressing cryptic transcription within ORFs (Carrozza et al., 2005; Govind et al., 2010; Keogh et al., 2005). pSer2 is involved in the co-transcriptional and post-transcriptional processing of RNA. Co-transcriptional processing of introns via splicing involves the yeast protein Prp40, which preferentially associates with pSer2/pSer5 marked CTD (Morris and Greenleaf, 2000). Recent studies have also implicated pThr4 as a critical mark for proper post-transcriptional splicing (Harlen et al., 2016).pSer2 is also bound by the SR-like (Serine/Arginine rich) protein Npl3, which functions in elongation, 3'-end processing, hnRNP formation, and mRNA export (Bucheli and Buratowski, 2005; Bucheli et al., 2007; Gilbert et al., 2001; Lei et al., 2001). Finally, increasing levels of pSer2, coupled with depletion of pSer5, leads to the recruitment of the termination and polyadenylation machinery (discussed below).

1.4 Transcription termination

The role of CTD modifications in orchestrating transcription termination is better described in recent reviews (Proudfoot, 2011; Richard and Manley, 2009). In essence, two models have been proposed to explain how Pol II termination occurs, with the emerging view being that it is likely a combination of the two models that best describes the mechanism (Mischo et al., 2011;

Richard and Manley, 2009). The first model, known as the “allosteric” or “anti-terminator” model, proposes that transcription through the polyadenylation site leads to an exchange of elongation factors for termination factors, resulting in a conformational change of the elongation complex. Indeed, this model is supported by chromatin immunoprecipitation (ChIP) data of elongation factor exchange at the 3’ end of genes (Kim et al., 2004a; Mayer et al., 2010). The second model, known as the “torpedo” model, postulates that cleavage of the transcript at the cleavage and polyadenylation site (CPS) creates an entry site for the 5’-3’ exonuclease Rat1 (Xrn2 in mammals), which degrades the 3’ RNA and promotes Pol II release by “torpedoing” the complex (Connelly and Manley, 1988; Kim et al., 2004a; West et al., 2004). In this model, recruitment of Rat1 is likely to be indirect, possibly though its partner Rtt103. Rtt103 has been shown to bind pSer2 in a cooperative manner with Pcf11 (Lunde et al., 2010), an essential component of the cleavage factor IA (CFIA) complex that also promotes Pol II release (Zhang et al., 2005). pTyr1 serves as an “anti-terminator” mark, which needs to be removed by the Glc7 phosphatase in order for these factors to associate with pSer2 (Mayer et al., 2012; Schreieck et al., 2014). Interestingly, ChIP data shows Pcf11 at both protein-coding and non-coding genes, and mutating Pcf11 results in terminator read-through due to inefficient cleavage at both gene classes (Grzechnik et al., 2015; Kim et al., 2010; Kim et al., 2006; Licatalosi et al., 2002; Meinhart and Cramer, 2004; Sadowski et al., 2003; Zhang et al., 2005). Pcf11 may play an important role in both the termination and processing of protein-coding and non-coding genes.

Processing of Pol II transcripts occurs via one of two distinct, gene class-specific pathways in yeast. Many small mRNAs (<550 bp), CUTs, snRNA, and snoRNAs (non-coding genes) are processed via the Nrd1-Nab3 pathway (Figure 1.6), while longer mRNAs (protein-coding genes) are processed in a polyadenylation-dependent process (Figure 1.7) (Arigo et al., 2006; Birse et al., 1998; Buratowski, 2005; Egloff and Murphy, 2008; Gudipati et al., 2008; Kim et al., 2006; Lykke-Andersen and Jensen, 2007; Richard and Manley, 2009; Steinmetz et al., 2001; Thiebaut

et al., 2006). The decision to proceed down a certain processing path is modulated by the phosphorylation state of the CTD.

Nrd1 preferentially associates with pSer5 and its recruitment is also enhanced via histone H3K4 trimethylation by Set1 (Terzi et al., 2011; Vasiljeva et al., 2008). Nrd1 and Nab3 scan the nascent RNA for specific sequence elements (GUAA/G or UGGA for Nrd1, and UCUU or CUUG for Nab3) as it exits the core polymerase (Carroll et al., 2007; Carroll et al., 2004; Conrad et al., 2000; Hobor et al., 2011; Steinmetz et al., 2001; Steinmetz and Brow, 1996; Steinmetz and Brow, 1998; Vasiljeva et al., 2008; Vojnic et al., 2006; Wlotzka et al., 2011). Upon detecting its consensus sequence elements, the Nrd1 complex and the Rnt1 endonuclease cleave these short transcripts (Allmang et al., 1999; Chanfreau et al., 1998a; Chanfreau et al., 1998b; Kim et al., 2006), which are then trimmed at the 3' end by the TRAMP complex and the exosome (Egecioglu et al., 2006; LaCava et al., 2005; Vasiljeva and Buratowski, 2006; Wyers et al., 2005).

Nrd1 stimulates the association of Pcf11, which allows Nrd1 to disengage from the transcription complex, with help from antagonizing pSer2 marks (Grzechnik et al., 2015; Gudipati et al., 2008). Pcf11 further promotes pSer2 marks, possibly by blocking phosphatase recruitment (Grzechnik et al., 2015). The Sen1 helicase (Senataxin in humans) associates with these pSer2 marks and resolves the DNA:RNA hybrids known as R-loops that form between the template DNA and the nascent RNA, keeping the specific sequence elements exposed and preserving genomic stability (Mischo et al., 2011; Skourti-Stathaki et al., 2011; Ursic et al., 1997). The involvement of Sen1 is dependent on the phosphatase Glc7, which dephosphorylates Sen1 and is essential for the proper termination of snRNA and snoRNA transcripts (Nedea et al., 2008).

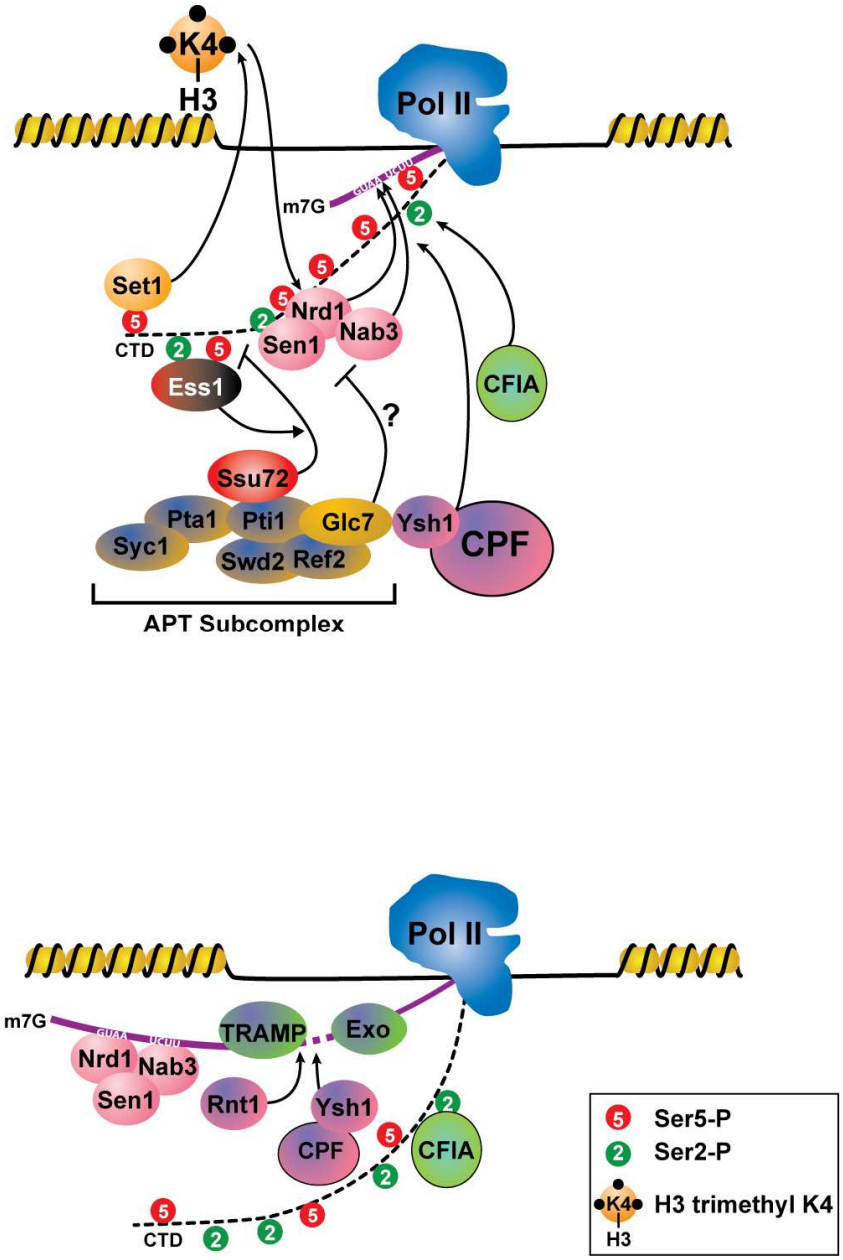

Figure 1.6 Nrd1-dependent termination pathway
The Nrd1-Nab3-Sen1 complex is recruited via interaction between Nrd1 and pSer5. This recruitment is facilitated by H3K4me3, which is placed by the Set1 histone methyltransferase. The mechanisms by which the Ssu72 and Glc7 phosphatases promote termination are still unclear, but it may be that the dephosphorylation of Sen1 by Glc7 and of the CTD by Ssu72 causes the polymerase to pause, and allowing the termination machinery to associate. During elongation, both Nrd1 and Nab3 scan the nascent RNA for their preferred sequences (see text for details). Upon finding their consensus sequences, Nrd1-Nab3-Sen1 complex is able to associate with the RNA. The endonucleases Rnt1 and Ysh1 may contribute to the cleavage of the RNA, which is followed by 3'-5' trimming the transcript by the TRAMP complex and by the degradation of the remaining RNA exiting Pol II by the 5'-3' exonuclease Rat1 (Exo).

Unlike snRNA/snoRNAs, which have protective structural elements in the RNA, Nrd1-terminated CUTs have no protective elements at their 3’ ends and are thus fully degraded by TRAMP after cleavage (Arigo et al., 2006; Lykke-Andersen and Jensen, 2007; Thiebaut et al., 2006). Nrd1 has been mapped to the 5' end of transcribed regions, but a recent study has demonstrated that Nrd1 occupancy is maintained across the open reading frame of genes (Kim et al., 2010).. Although no homolog of Nrd1 has been found in mammalian cells, the Integrator complex that is involved in 3' processing of snRNA transcripts is recruited by pSer7 (Baillat et al., 2005). The association of this complex with pSer7 CTD was demonstrated by the abolishment of this interaction upon mutation of Ser7 to alanine (Egloff et al., 2007). Subsequent analysis using a panel of CTD peptides determined the Integrator prefers to bind a di-phosphorylated CTD substrate spanning two heptad repeats in the pSer7-pSer2 conformation (Egloff et al., 2010). It is possible pSer7 may serve as a similar scaffold for snRNA and snoRNA processing machinery in yeast.

The second pathway, used for the processing of most mRNA transcripts, involves the cleavage and polyadenylation factor (CPF) complex, cleavage factor IA and IB (CFIA and CFIB) complexes, and the exosome (Figure 1.7) (Birse et al., 1998; Kim et al., 2006; Richard and Manley, 2009). Many of the termination and 3’ processing factors involved in this process are known to preferentially associate with pSer2 or pSer2/pSer5 enriched CTD including: Npl3, Rtt103, Rna14, Rna15, Ydh1, Yhh1, Pta1, and Pcf11. In this pathway, Rna15 competes with Npl3 for recognition of a UA-rich site in the nascent RNA (Bucheli et al., 2007; Dermody et al., 2008). This competition is removed upon phosphorylation of Npl3 by casein kinase 2 (CK2) (Dermody et al., 2008). Rna15 can then bind the nascent RNA and promote endonucleolytic cleavage followed by polyadenylation by the polyadenylate polymerase (Pap1) (Birse et al., 1998; Minvielle-Sebastia et al., 1994). Polyadenylation-binding proteins (PAB) then protect the mature transcript from exonucleolytic degradation (Figure 1.7) (Moore, 2005).

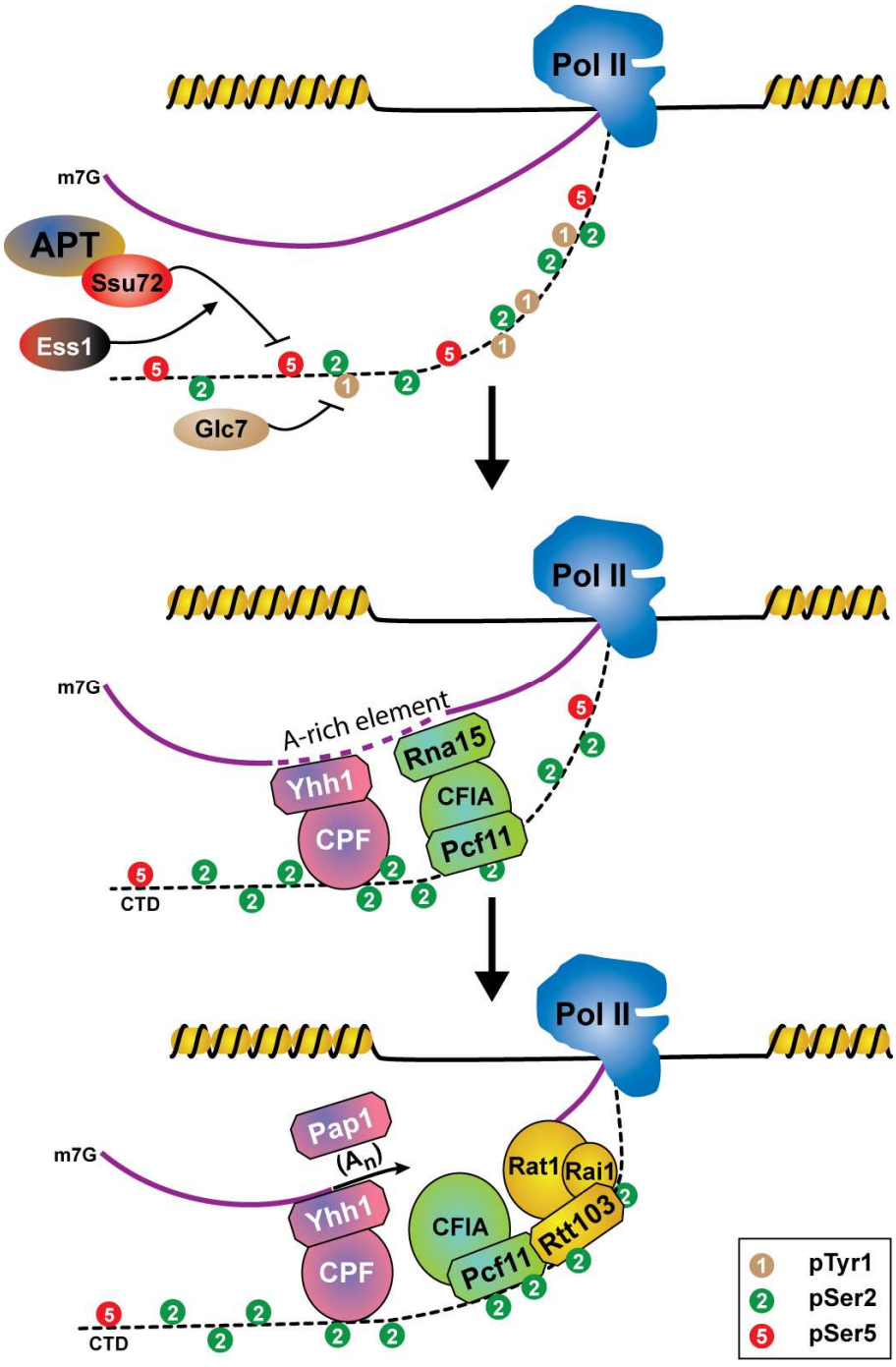

Figure 1.7 The mRNA termination pathway
Rna15 competes with Npl3 for binding to the nascent RNA. CK2 phosphorylates Npl3, allowing Rna15 to find its preferred binding site (an A/U-rich region) in the RNA. The CPF and CFIA components assemble through interactions with the CTD and the Yth1 component of CPF cleaves the nascent RNA at the polyadenylation site, followed by polyadenylation by Pap1. Then the Rat1 exonuclease complex associates via cooperative interaction between Pcf11 and Rtt103 and leads to termination and dissociation of Pol II.

In both pathways, the CTD is hypophosphorylated by the combined action of two essential phosphatases at the end of transcription: Ssu72 and Fcp1. Ssu72 is a member of the Associated with Pta1 (APT) complex, which is present at both gene classes and is involved in 3' processing of non-coding RNAs (Nedea et al., 2003). As such, Ssu72 is primarily localized at the 3' end of transcripts (Nedea et al., 2003), although there is one instance in which it has been found at promoters (Ansari and Hampsey, 2005). Temperature sensitive mutants of Ssu72 exhibit read-through at both protein-coding and non-coding transcripts (Steinmetz and Brow, 2003). Ssu72 is the primary pSer5 and pSer7 phosphatase (Krishnamurthy et al., 2004; Zhang et al., 2012a), and its phosphatase activity is enhanced by the prolyl isomerase Ess1/Pin1 and by interacting with Pta1/Symplekin (Ghazy et al., 2009; Krishnamurthy et al., 2009; Singh et al., 2009b). Recently, crystal structures have shed light on the mechanism of Ssu72: the phosphatase binds to pSer5 only when the adjacent Pro6 is in the *cis* conformation (Werner-Allen et al., 2011; Xiang et al., 2010). In contrast to Ssu72, Fcp1 associates with TFIIF during transcription and is found across the entire transcribed region (Archambault et al., 1997; Cho et al., 2001; Kobor et al., 1999; Kong et al., 2005). Although it has pSer5 and pSer2 phosphatase activity *in vitro*, Fcp1 is considered a pSer2 specific phosphatase *in vivo* (Ghosh et al., 2008; Hausmann et al., 2004). Fcp1 activity is enhanced upon phosphorylation of Fcp1 by CK2 (Abbott et al., 2005). Defects in Fcp1 also result in transcription read-through at Nrd1 dependent transcripts (Gudipati et al., 2008).

Global dephosphorylation of the CTD facilitates the release of Pol II from DNA, which can then recycle to promoters for the next cycle of transcription (Cho et al., 1999; Dichtl et al., 2002a; Steinmetz and Brow, 2003). It has been proposed that transcription termination and subsequent dephosphorylation of the CTD is coupled to transcription reinitiation through gene looping, by which the promoter and terminator regions are brought together, allowing Pol II to associate more rapidly with the PIC (O'Sullivan et al., 2004; Singh et al., 2009a). Intriguingly, Ssu72 and

the GTF TFIIB have been shown to be essential in gene looping (Ansari and Hampsey, 2005; Singh and Hampsey, 2007). Taken together, the phosphorylation and dephosphorylation of the CTD is intimately involved in every phase of transcription, from initiation, to elongation, to termination, and possibly reinitiation.

1.5 mRNA export

In addition to its many roles in transcription initiation, elongation, and termination, the CTD has been implicated in a variety of transcription-extrinsic processes, such as mRNA export. mRNA export (reviewed in (Köhler and Hurt, 2007; Stewart, 2010; Vinciguerra and Stutz, 2004)) requires the packaging of the mRNA into export-competent messenger ribonucleoprotein (mRNP) via association with the Mex67:Mtr2 heterodimer (Segref et al., 1997). This heterodimer is brought to the mRNA by Yra1 and Sub2, components of the THO subunit of the TREX1 complex (Stewart, 2010). The process of mRNP export is coordinated by the protein Sus1. This central protein directly interacts with pSer5 and pSer2/pSer5 of the CTD, Ub8 subunit of the SAGA complex, Yra1 subunit of the TREX1 complex, and Sac3 subunit of the TREX2 complex at the nuclear pore (Figure 1.8) (Jani et al., 2009; Pascual-García et al., 2008).

1.6 The CTD code controversy

The concept of the CTD code was first proposed due to the enormous amount of information that can be encoded via post-translational modification of the CTD repeats (Buratowski, 2003). The code would coordinate the assembly of complexes that “read, write, and erase” the code during transcription. Historically, the pSer5 and pSer2 marks have been the best characterized, with the canonical distribution of pSer5 being enriched at the 5’ end of genes and pSer2 enriched towards the 3’ end. Further, gene-specific phosphorylation profiles, with pSer2 levels being significantly lower at non-coding genes and pSer7 profiles diverging from pSer5 profiles

only at protein-coding genes (Tietjen et al., 2010). The distinct patterns of CTD marks at these two gene classes reflect the different mechanisms of transcription termination and 3' end processing machinery that act on these two classes of RNA. Importantly, an unexpected degree of co-occurrence of CTD marks was observed, suggesting a bivalent or even multivalent mode of recognition by docking partners (Kim et al., 2010; Mayer et al., 2010; Tietjen et al., 2010). In support of this idea, the Set2 histone methyltransferase and the Integrator complex have been shown to prefer a bivalent mark rather than a single phosphorylated residue (Egloff et al., 2010; Kizer et al., 2005; Vojnic et al., 2006). In addition to the various phosphorylation marks, the isomerization state of the CTD also contributes to the complexity of the code. For example, Pcf11 binds the CTD in the *trans* conformation while Ssu72 prefers a *cis* CTD as substrate (Meinhart and Cramer, 2004; Werner-Allen et al., 2011; Xiang et al., 2010).

Many in the transcription field have made the argument that the CTD code is not a true code because it does not convey biological information via a rigorous decoding key. However, research in the last several years has demonstrated that specific phosphorylation marks and proline isomerization are important for conveying information from cis-elements encountered by Pol II to the protein complexes necessary for successful progression through the transcription cycle. This thesis further delves into the mechanisms of this information transfer and seeks to expand the RNA Pol II CTD code.

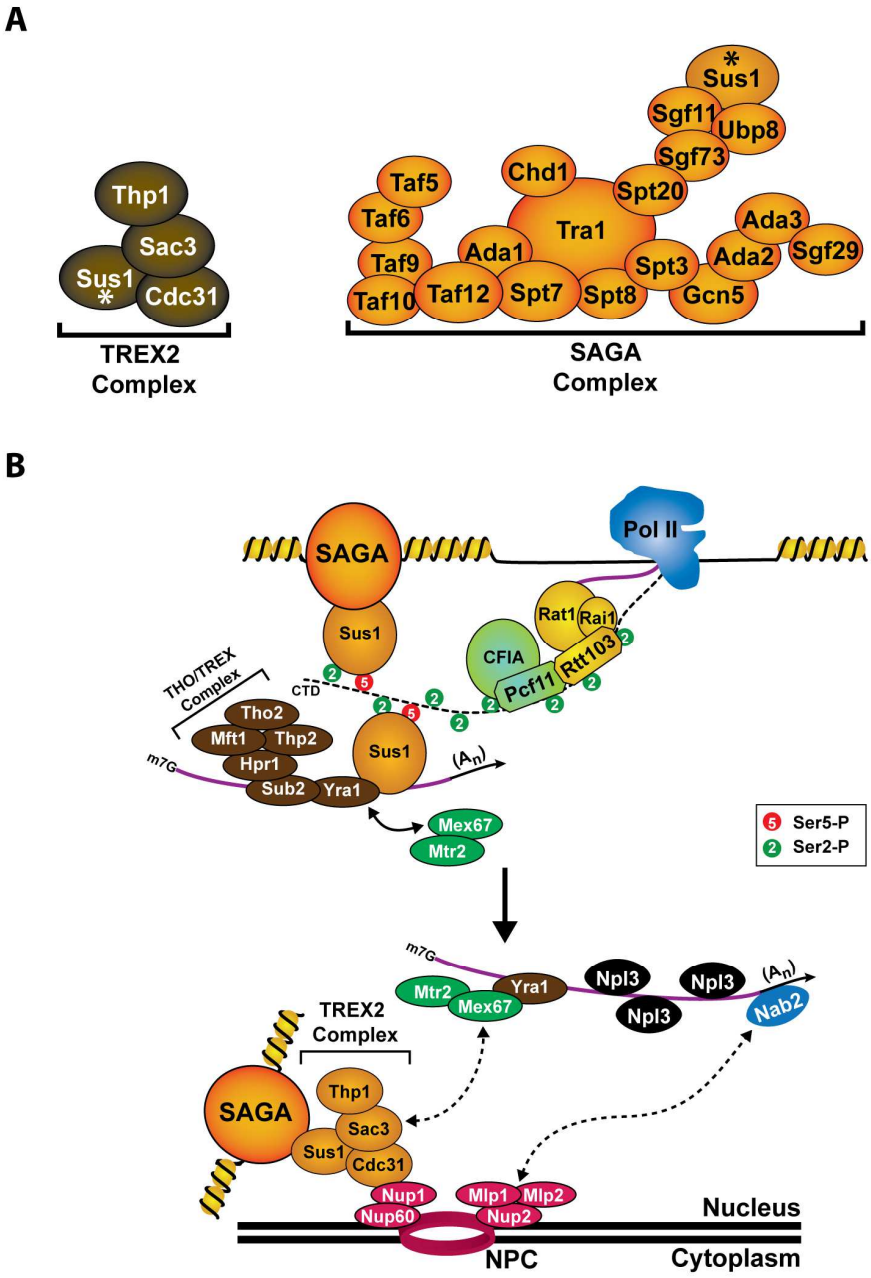

Figure 1.8 Sus1 in TREX2 and SAGA complexes coordinates mRNA export

A) Subunit compositions of TREX2 and SAGA complexes are shown, highlighting Sus1 (asterisk).

B) mRNA export coordinated by Sus1. Sus1 binds pSer2 and pSer2/pSer5 CTD, connecting the CTD to the SAGA histone acetyltransferase complex. Sus1 also interacts with Yra1 component of the THO/TREX complex on the RNA. Mex67-Mtr2 are recruited by interaction with Yra1 and help form the export-competent mRNP. At the Nuclear Pore Complex (NPC), Mlp1-Mlp2 interact with the polyA mRNA-binding protein Nab2, and Mex67 interacts with Sac3 of the TREX2 complex. This interaction brings the export-competent mRNP to the NPC in preparation for export to the cytoplasm. Sus1 is a component of both TREX2 and SAGA and serves to tether actively transcribed gene promoters to the NPC.

1.6 References

Abbott, K.L., Renfrow, M.B., Chalmers, M.J., Nguyen, B.D., Marshall, A.G., Legault, P., and Omichinski, J.G. (2005). Enhanced binding of RNAP II CTD phosphatase FCP1 to RAP74 following CK2 phosphorylation. Biochemistry *44*, 2732-2745.

Adkins, M.W., and Tyler, J.K. (2006). Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. Mol Cell *21*, 405-416.

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3' End Processing. Mol Cell *13*, 67-76.

Akhtar, M., Heidemann, M., Tietjen, J., Zhang, D., Chapman, R., Eick, D., and Ansari, A. (2009). TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. Mol Cell *34*, 387-393.

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J *18*, 5399-5410.

Ansari, A., and Hampsey, M. (2005). A role for the CPF 3 -end processing machinery in RNAP II-dependent gene looping. Genes Dev *19*, 2969.

Archambault, J., Chambers, R.S., Kobor, M.S., Ho, Y., Cartier, M., Bolotin, D., Andrews, B., Kane, C.M., and Greenblatt, J. (1997). An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae. Proc Natl Acad Sci USA *94*, 14300.

Arigo, J., Eyler, D., Carroll, K., and Corden, J. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Mol Cell *23*, 841-851.

Armache, K.J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America *100*, 6964-6968.

Asturias, F.J., Jiang, Y.W., Myers, L.C., Gustafsson, C.M., and Kornberg, R.D. (1999). Conserved Structures of Mediator and RNA Polymerase II Holoenzyme. Science *283*, 985-987.

Baillat, D., Hakimi, M., Näär, A., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell *123*, 265-276.

Barberis, A., Pearlberg, J., Simkovich, N., Farrell, S., Reinagel, P., Bamdad, C., Sigal, G., and Ptashne, M. (1995). Contact with a component of the polymerase II holoenzyme suffices for gene activation. Cell *81*, 359-368.

Barillà, D., Lee, B.A., and Proudfoot, N.J. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America *98*, 445-450.

Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. Cell *116*, 699-709.

Bensaude, O., Bonnet, F., Casse, C., Dubois, M.F., Nguyen, V.T., and Palancade, B. (1999). Regulated phosphorylation of the RNA polymerase II C-terminal domain (CTD). Biochem Cell Biol *77*, 249-255.

Bentley, D. (1999). Coupling RNA polymerase II transcription with pre-mRNA processing. Curr Opin Cell Biol *11*, 347-351.

Bentley, D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. Curr Opin Cell Biol *14*, 336-342.

Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol *17*, 251-256.

Biddick, R., and Young, E.T. (2005). Yeast Mediator and its role in transcriptional regulation. C R Biol *328*, 773-782.

Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. (1998). Coupling Termination of Transcription to Messenger RNA Maturation in Yeast. Science *280*, 298-301.

Bjorklund, S., and Gustafsson, C.M. (2005). The yeast Mediator complex and its regulation. Trends Biochem Sci *30*, 240-244.

Boeing, S., Rigault, C., Heidemann, M., Eick, D., and Meisterernst, M. (2010). RNA Polymerase II C-terminal Heptarepeat Domain Ser-7 Phosphorylation Is Established in a Mediator-dependent Fashion. J Biol Chem *285*, 188-196.

Borggrefe, T., Davis, R., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (2002). A Complex of the Srb8,-9,-10, and-11 Transcriptional Regulatory Proteins from Yeast. J Biol Chem *277*, 44202-44207.

Brès, V., Yoh, S., and Jones, K. (2008). The multi-tasking P-TEFb complex. Curr Opin Cell Biol *20*, 334-340.

Bucheli, M.E., and Buratowski, S. (2005). Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. EMBO J *24*, 2150-2160.

Bucheli, M.E., He, X., Kaplan, C.D., Moore, C.L., and Buratowski, S. (2007). Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI. RNA *13*, 1756-1764.

Buratowski, S. (2003). The CTD code. Nat Struct Mol Biol *10*, 679-680.

Buratowski, S. (2005). Connections between mRNA 3' end processing and transcription termination. Curr Opin Cell Biol *17*, 257-261.

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

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007). Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell *131*, 706-717.

Carroll, K.L., Ghirlando, R., Ames, J.M., and Corden, J.L. (2007). Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. RNA *13*, 361-373.

Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D., and Corden, J.L. (2004). Identification of cis Elements Directing Termination of Yeast Nonpolyadenylated snoRNA Transcripts. Mol Cell Biol *24*, 6241-6252.

Carrozza, M., Li, B., Florens, L., Suganuma, T., Swanson, S., Lee, K., Shia, W., Anderson, S., Yates, J., Washburn, M.*, et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Casamassimi, A., and Napoli, C. (2007). Mediator complexes and eukaryotic transcription regulation: an overview. Biochimie *89*, 1439-1446.

Chadick, J.Z., and Asturias, F.J. (2005). Structure of eukaryotic Mediator complexes. Trends Biochem Sci *30*, 264-271.

Chanfreau, G., Legrain, P., and Jacquier, A. (1998a). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism1. J Mol Biol *284*, 975-988.

Chanfreau, G., Rotondo, G., Legrain, P., and Jacquier, A. (1998b). Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. EMBO J *17*, 3726-3737.

Chang, A., Cheang, S., Espanel, X., and Sudol, M. (2000). Rsp5 WW domains interact directly with the carboxyl-terminal domain of RNA polymerase II. J Biol Chem *275*, 20562-20571.

Chang, W.H., and Kornberg, R.D. (2000). Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. Cell *102*, 609-613.

Chao, D.M., Gadbois, E.L., Murray, P.J., Anderson, S.F., Sonu, M.S., Parvin, J.D., and Young, R.A. (1996). A mammalian SRB protein associated with an RNA polymerase II holoenzyme.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA Polymerase II Is Phosphorylated at CTD Residue Serine-7. Science *318*, 1780-1782.

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

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

Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev *11*, 3319-3326.

Cho, H., Kim, T.K., Mancebo, H., Lane, W.S., Flores, O., and Reinberg, D. (1999). A protein phosphatase functions to recycle RNA polymerase II. Genes Dev *13*, 1540-1552.

Conaway, R.C., Sato, S., Tomomori-Sato, C., Yao, T., and Conaway, J.W. (2005). The mammalian Mediator complex and its role in transcriptional regulation. Trends Biochem Sci *30*, 250-255.

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

Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., and Corden, J.L. (2000). A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. Genetics *154*, 557-571.

Corden, J.L. (1990). Tails of RNA polymerase II. Trends Biochem Sci *15*, 383-387.

Corden, J.L., Cadena, D.L., Ahearn, J.M., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci USA *82*, 7934-7938.

Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G., Dengl, S., Geiger, S., and Jasiak, A. (2008). Structure of eukaryotic RNA polymerases. Annual Reviews Biophysics *37*, 337-352.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. Science *292*, 1863-1876.

Dahmus, M.E. (1996). Reversible phosphorylation of the C-terminal domain of RNA polymerase II. J Biol Chem *271*, 19009-19012.

Daulny, A., Geng, F., Muratani, M., Geisinger, J.M., Salghetti, S.E., and Tansey, W.P. (2008). Modulation of RNA polymerase II subunit composition by ubiquitylation. Proc Natl Acad Sci USA *105*, 19649-19654.

Davis, C.A., and Ares, M. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomycescerevisiae. Proceedings of the National Academy of Sciences of the United States of America *103*, 3262.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol *8*, e1000384.

Dermody, J.L., Dreyfuss, J.M., Villén, J., Ogundipe, B., Gygi, S.P., Park, P.J., Ponticelli, A.S., Moore, C.L., Buratowski, S., and Bucheli, M.E. (2008). Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation. PLoS One *3*, e3273.

Dichtl, B., Blank, D., Ohnacker, M., Friedlein, A., Roeder, D., Langen, H., and Keller, W. (2002a). A Role for SSU72 in Balancing RNA Polymerase II Transcription Elongation and Termination. Mol Cell *10*, 1139-1150.

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

Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., and Pagano, A. (2007). The expanding RNA polymerase III transcriptome. Trends Genet *23*, 614-622.

Drouin, S., Laramée, L., Jacques, P.É., Forest, A., Bergeron, M., and Robert, F. DSIF and RNA Polymerase II CTD Phosphorylation Coordinate the Recruitment of Rpd3S to Actively Transcribed Genes. PLoS Genet *6*, e1001173.

Drouin, S., Laramée, L., Jacques, P.É., Forest, A., Bergeron, M., Robert, F., and Lieb, J.D. (2010). DSIF and RNA Polymerase II CTD Phosphorylation Coordinate the Recruitment of Rpd3S to Actively Transcribed Genes. PLoS Genet *6*, e1001173.

Egecioglu, D., Henras, A., and Chanfreau, G. (2006). Contributions of Trf4p-and Trf5p-dependent polyadenylation to the processing and degradative functions of the yeast nuclear exosome. RNA *12*, 26-32.

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

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA Polymerase II CTD Is Specifically Required for snRNA Gene Expression. Science *318*, 1777-1779.

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

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., and Werner, M. (2008). Mediator-dependent recruitment of TFIIH modules in preinitiation complex. Mol Cell *31*, 337-346.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol Cell *11*, 1549-1561.

Faller, M., and Guo, F. (2008). MicroRNA biogenesis: there's more than one way to skin a cat. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms *1779*, 663-667.

Fan, X., Chou, D.M., and Struhl, K. (2006). Activator-specific recruitment of Mediator in vivo. Nat Struct Mol Biol *13*, 117-120.

Fan, X., and Struhl, K. (2009). Where does mediator bind in vivo? PLoS One *4*, e5029.

Feaver, W.J., Svejstrup, J.Q., Henry, N.L., and Kornberg, R.D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell *79*, 1103-1109.

Finkel, J.S., Chinchilla, K., Ursic, D., and Culbertson, M.R. (2010). Sen1p performs two genetically separable functions in transcription and processing of U5 small nuclear RNA in Saccharomyces cerevisiae. Genetics *184*, 107.

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

Fuchs, S.M., Laribee, R.N., and Strahl, B.D. (2009). Protein modifications in transcription elongation. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms *1789*, 26-36.

Gebara, M.M., Sayre, M.H., and Corden, J.L. (1997). Phosphorylation of the carboxy terminal repeat domain in RNA polymerase II by cyclin dependent kinases is sufficient to inhibit transcription. J Cell Biochem *64*, 390-402.

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

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

Ghosh, A., Shuman, S., and Lima, Christopher D. (2011). Structural Insights to How Mammalian Capping Enzyme Reads the CTD Code. Mol Cell.

Gilbert, W., Siebel, C.W., and Guthrie, C. (2001). Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. RNA *7*, 302-313.

Govind, C.K., Qiu, H., Ginsburg, D.S., Ruan, C., Hofmeyer, K., Hu, C., Swaminathan, V., Workman, J.L., Li, B., and Hinnebusch, A.G. (2010). Phosphorylated Pol II CTD Recruits Multiple HDACs, Including Rpd3C(S), for Methylation-Dependent Deacetylation of ORF Nucleosomes. Mol Cell *39*, 234-246.

Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R.H., Michaelis, J., and Werner, F. (2011). The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation. Mol Cell *43*, 263-274.

Grummt, I. (2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. Genes Dev *17*, 1691.

Grzechnik, P., Gdula, M.R., and Proudfoot, N.J. (2015). Pcf11 orchestrates transcription termination pathways in yeast. Genes Dev *29*, 849-861.

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

Guglielmi, B., Van Berkum, N.L., Klapholz, B., Bijma, T., Boube, M., Boschiero, C., Bourbon, H.M., Holstege, F.C.P., and Werner, M. (2004). A high resolution protein interaction map of the yeast Mediator complex. Nucleic Acids Res *32*, 5379-5391.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Rep *15*, 2147-2158.

Hartzog, G.A., Wada, T., Handa, H., and Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II inSaccharomyces cerevisiae. Genes Dev *12*, 357.

Hausmann, S., Erdjument-Bromage, H., and Shuman, S. (2004). Schizosaccharomyces pombe Carboxyl-terminal Domain (CTD) Phosphatase Fcp1. J Biol Chem *279*, 10892.

Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S., and Young, R.A. (1998). Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. Mol Cell *2*, 43-53.

Hengartner, C.J., Thompson, C.M., Zhang, J., Chao, D.M., Liao, S.M., Koleske, A.J., Okamura, S., and Young, R.A. (1995). Association of an activator with an RNA polymerase II holoenzyme. Genes Dev *9*, 897-910.

Henry, K.W., Wyce, A., Lo, W.S., Duggan, L.J., Emre, N., Kao, C.F., Pillus, L., Shilatifard, A., Osley, M.A., and Berger, S.L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev *17*, 2648-2663.

Hirose, Y., and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. Genes Dev *14*, 1415-1429.

Hirose, Y., and Ohkuma, Y. (2007). Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. Journal of Biochemistry *141*, 601-608.

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

Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S., and Stefl, R. (2011). Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (Nab) 3 protein. J Biol Chem *286*, 3645-3657.

Hollingworth, D., Noble, C.G., Taylor, I., and Ramos, A. (2006). RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11. RNA *12*, 555-560.

Hong, S.W., Hong, S.M., Yoo, J.W., Lee, Y.C., Kim, S., Lis, J.T., and Lee, D. (2009). Phosphorylation of the RNA polymerase II C-terminal domain by TFIIH kinase is not essential for transcription of Saccharomyces cerevisiae genome. Proc Natl Acad Sci USA *106*, 14276.

Jaehning, J.A. (2010). The Paf1 complex: Platform or player in RNA polymerase II transcription? Biochimica et Biophysica Acta *1799*, 379-388.

Jamai, A., Puglisi, A., and Strubin, M. (2009). Histone chaperone spt16 promotes redeposition of the original h3-h4 histones evicted by elongating RNA polymerase. Mol Cell *35*, 377-383.

Jani, D., Lutz, S., Marshall, N.J., Fischer, T., Köhler, A., Ellisdon, A.M., Hurt, E., and Stewart, M. (2009). Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export. Mol Cell *33*, 727-737.

Jenuwein, T., and Allis, C. (2001). Translating the histone code. Science *293*, 1074-1080.

Jones, J.C., Phatnani, H.P., Haystead, T.A., MacDonald, J.A., Alam, S.M., and Greenleaf, A.L. (2004). C-terminal Repeat Domain Kinase I Phosphorylates Ser2 and Ser5 of RNA Polymerase II C-terminal Domain Repeats. J Biol Chem *279*, 24957-24964.

Kang, J.S., Kim, S.H., Hwang, M.S., Han, S.J., Lee, Y.C., and Kim, Y.J. (2001). The structural and functional organization of the yeast mediator complex. J Biol Chem *276*, 42003-42010.

Kang, M.E., and Dahmus, M.E. (1995). The photoactivated cross-linking of recombinant C-terminal domain to proteins in a HeLa cell transcription extract that comigrate with transcription factors IIE and IIF. J Biol Chem *270*, 23390-23397.

Kanin, E.I., Kipp, R.T., Kung, C., Slattery, M., Viale, A., Hahn, S., Shokat, K.M., and Ansari, A.Z. (2007). Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis. Proc Natl Acad Sci USA *104*, 5812-5817.

Kaplan, C., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096-1099.

Kelly, W.G., Dahmus, M., and Hart, G. (1993). RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. J Biol Chem *268*, 10416-10424.

Keogh, M., Kurdistani, S., Morris, S., Ahn, S., Podolny, V., Collins, S., Schuldiner, M., Chin, K., Punna, T., and Thompson, N. (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.

Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat Struct Mol Biol *17*, 1279-1286.

Kim, J.B., and Sharp, P.A. (2001). Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. J Biol Chem *276*, 12317-12323.

Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004a). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. EMBO J *23*, 354-364.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004b). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517-522.

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

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A Novel Domain in Set2 Mediates RNA Polymerase II Interaction and Couples Histone H3 K36 Methylation with Transcript Elongation. Mol Cell Biol *25*, 3305-3316.

Kobor, M.S., Archambault, J., Lester, W., Holstege, F.C.P., Gileadi, O., Jansma, D.B., Jennings, E.G., Kouyoumdjian, F., Davidson, A.R., and Young, R.A. (1999). An Unusual Eukaryotic Protein Phosphatase Required for Transcription by RNA Polymerase II and CTD Dephosphorylation in S. cerevisiae. Mol Cell *4*, 55-62.

Köhler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nature Reviews Molecular Cell Biology *8*, 761-773.

Koleske, A.J., and Young, R.A. (1994). An RNA polymerase II holoenzyme responsive to activators. Nature *368*, 466-469.

Koleske, A.J., and Young, R.A. (1995). The RNA polymerase II holoenzyme and its implications for gene regulation. Trends Biochem Sci *20*, 113-116.

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

Kong, S.E., Kobor, M.S., Krogan, N.J., Somesh, B.P., Sogaard, T.M.M., Greenblatt, J.F., and Svejstrup, J.Q. (2005). Interaction of Fcp1 Phosphatase with Elongating RNA Polymerase II Holoenzyme, Enzymatic Mechanism of Action, and Genetic Interaction with Elongator. J Biol Chem *280*, 4299-4306.

Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. Trends Biochem Sci *30*, 235-239.

Kornberg, R.D. (2007). The molecular basis of eukaryotic transcription. Proc Natl Acad Sci USA *104*, 12955.

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

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004). Ssu72 Is an RNA Polymerase II CTD Phosphatase. Mol Cell *14*, 387-394.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., and Boone, C. (2003). Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Mol Cell Biol *23*, 4207-4218.

Kyburz, A., Sadowski, M., Dichtl, B., and Keller, W. (2003). The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre mRNA 3 end formation. Nucleic Acids Res *31*, 3936-3945.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell *121*, 713-724.

Lee, J.S., Smith, E., and Shilatifard, A. (2010). The language of histone crosstalk. Cell *142*, 682-685.

Lei, E.P., Krebber, H., and Silver, P.A. (2001). Messenger RNAs are recruited for nuclear export during transcription. Genes Dev *15*, 1771-1782.

Li, B., Howe, L.A., Anderson, S., Yates, J.R., and Workman, J.L. (2003). The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. J Biol Chem *278*, 8897-8903.

Li, J., Moazed, D., and Gygi, S.P. (2002). Association of the Histone Methyltransferase Set2 with RNA Polymerase II Plays a Role in Transcription Elongation. J Biol Chem *277*, 49383-49388.

Liao, S.M., Zhang, J., Jeffery, D.A., Koleske, A.J., Thompson, C.M., Chao, D.M., Viljoen, M., van Vuuren, H.J.J., and Young, R.A. (1995). A kinase-cyclin pair in the RNA polymerase II holoenzyme. Nature *374*, 193-196.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional Interaction of Yeast Pre-mRNA 3' End Processing Factors with RNA Polymerase II. Mol Cell *9*, 1101-1111.

Liu, P., Kenney, J.M., Stiller, J.W., and Greenleaf, A.L. (2010). Genetic Organization, Length Conservation, and Evolution of RNA Polymerase II Carboxyl-Terminal Domain. Mol Biol Evol *27*, 2628-2641.

Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S. (2004). Two Cyclin-Dependent Kinases Promote RNA Polymerase II Transcription and Formation of the Scaffold Complex. Mol Cell Biol *24*, 1721-1735.

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

Lykke-Andersen, S., and Jensen, T.H. (2007). Overlapping pathways dictate termination of RNA polymerase II transcription. Biochimie *89*, 1177-1182.

MacKellar, A.L., and Greenleaf, A.L. (2011). Cotranscriptional association of mRNA export factor Yra1 with the C-terminal domain of RNA polymerase II: a mechanism for cotranscriptional recruitment. J Biol Chem.

Malik, S., and Roeder, R. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci *30*, 256-263.

Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C.M., and Cramer, P. (2011). Architecture of the RNA polymerase–Spt4/5 complex and basis of universal transcription processivity. EMBO J.

Maxon, M., Goodrich, J., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev *8*, 515-524.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. Science *336*, 1723-1725.

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

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., and Foster, S. (1997). 5 -Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev *11*, 3306-3318.

McDonald, S.M., Close, D., Xin, H., Formosa, T., and Hill, C.P. (2010). Structure and Biological Importance of the Spn1-Spt6 Interaction, and Its Regulatory Role in Nucleosome Binding. Mol Cell *40*, 725-735.

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

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

Minvielle-Sebastia, L., Preker, P.J., and Keller, W. (1994). RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. Science *266*, 1702-1705.

Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 Helicase Protects the Genome from Transcription-Associated Instability. Mol Cell *41*, 21-32.

Misri, S., Pandita, S., Kumar, R., and Pandita, T. (2008). Telomeres, histone code, and DNA damage response. Cytogenet Genome Res *122*, 297-307.

Mitchell Guttman, I.A., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., and Cabili, M.N. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223.

Moore, M.J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. Science *309*, 1514-1518.

Morris, D.P., and Greenleaf, A.L. (2000). The Splicing Factor, Prp40, Binds the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. J Biol Chem *275*, 39935-39943.

Morris, D.P., Phatnani, H.P., and Greenleaf, A.L. (1999). Phospho-Carboxyl-Terminal Domain Binding and the Role of a Prolyl Isomerase in Pre-mRNA 3'-End Formation. J Biol Chem *274*, 31583-31587.

Mosley, A., Pattenden, S., Carey, M., Venkatesh, S., Gilmore, J., Florens, L., Workman, J., and Washburn, M. (2009). Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation. Mol Cell *34*, 168-178.

Moteki, S., and Price, D. (2002). Functional coupling of capping and transcription of mRNA. Mol Cell *10*, 599-609.

Munshi, A., Shafi, G., Aliya, N., and Jyothy, A. (2009). Histone modifications dictate specific biological readouts. Journal of Genetics and Genomics *36*, 75-88.

Myer, V.E., and Young, R.A. (1998). RNA polymerase II holoenzymes and subcomplexes. J Biol Chem *273*, 27757-27760.

Myers, L.C., Gustafsson, C.M., Bushnell, D.A., Lui, M., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (1998). The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. Genes Dev *12*, 45-54.

Myers, L.C., and Kornberg, R.D. (2000). Mediator of transcriptional regulation. Annu Rev Biochem *69*, 729-749.

Nakanishi, S., Sanderson, B.W., Delventhal, K.M., Bradford, W.D., Staehling-Hampton, K., and Shilatifard, A. (2008). A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. Nat Struct Mol Biol *15*, 881-888.

Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochimica et Biophysica Acta *1809*, 34-45.

Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C.L., and Greenblatt, J. (2003). Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3 -ends. J Biol Chem *278*, 33000-33010.

Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. (2008). The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. Mol Cell *29*, 577-587.

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

Noble, C.G., Hollingworth, D., Martin, S.R., Ennis-Adeniran, V., Smerdon, S.J., Kelly, G., Taylor, I.A., and Ramos, A. (2005). Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nat Struct Mol Biol *12*, 144-151.

O'Sullivan, J., Tan-Wong, S., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N. (2004). Gene loops juxtapose promoters and terminators in yeast. Nat Genet *36*, 1014-1018.

Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q.*, et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev *10*, 2657-2683.

Orphanides, G., Wu, W.H., Lane, W.S., Hampsey, M., and Reinberg, D. (1999). The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. Nature *400*, 284-288.

Palancade, B., and Bensaude, O. (2003). Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. FEBS J *270*, 3859-3870.

Pascual-García, P., Govind, C.K., Queralt, E., Cuenca-Bono, B., Llopis, A., Chavez, S., Hinnebusch, A.G., and Rodríguez-Navarro, S. (2008). Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2. Genes Dev *22*, 2811-2822.

Pedersen, A.G., Baldi, P., Chauvin, Y., and Brunak, S. (1999). The biology of eukaryotic promoter prediction - a review. Computers & Chemistry *23*, 191-207.

Perales, R., and Bentley, D. (2009). “Cotranscriptionality”: The Transcription Elongation Complex as a Nexus for Nuclear Transactions. Mol Cell *36*, 178-191.

Peterlin, B.M., and Price, D.H. (2006). Controlling the Elongation Phase of Transcription with P-TEFb. Mol Cell *23*, 297-305.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev *20*, 2922-2936.

Phatnani, H.P., Jones, J.C., and Greenleaf, A.L. (2004). Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. Biochemistry *43*, 15702-15719.

Prather, D., Krogan, N.J., Emili, A., Greenblatt, J.F., and Winston, F. (2005). Identification and characterization of Elf1, a conserved transcription elongation factor in Saccharomyces cerevisiae. Mol Cell Biol *25*, 10122.

Price, D.H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. Mol Cell Biol *20*, 2629-2634.

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev *25*, 1770-1782.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA Processing with Transcription. Cell *108*, 501-512.

Qiu, H., Hu, C., and Hinnebusch, A. (2009). Phosphorylation of the Pol II CTD by KIN28 Enhances BUR1/BUR2 Recruitment and Ser2 CTD Phosphorylation Near Promoters. Mol Cell *33*, 752-762.

Rabut, G., Le Dez, G., Verma, R., Makhnevych, T., Knebel, A., Kurz, T., Boone, C., Deshaies, R.J., and Peter, M. (2011). The TFIIH Subunit Tfb3 Regulates Cullin Neddylation. Mol Cell *43*, 488-495.

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

Rickert, P., Corden, J.L., and Lees, E. (1999). Cyclin C/CDK8 and cyclin H/CDK7/p36 are biochemically distinct CTD kinases. Oncogene *18*, 1093-1102.

Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L., and Buratowski, S. (2000). Kin28, the TFIIH-Associated Carboxy-Terminal Domain Kinase, Facilitates the Recruitment of mRNA Processing Machinery to RNA Polymerase II. Mol Cell Biol *20*, 104-112.

Russell, J., and Zomerdijk, J.C.B.M. (2005). RNA-polymerase-I-directed rDNA transcription, life and works. Trends Biochem Sci *30*, 87-96.

Sadowski, M., Dichtl, B., Hübner, W., and Keller, W. (2003). Independent functions of yeast Pcf11p in pre-mRNA 3 end processing and in transcription termination. EMBO J *22*, 2167-2177.

Sato, F., Tsuchiya, S., Meltzer, S.J., and Shimizu, K. (2011). MicroRNAs and Epigenetics. FEBS J *278*, 1598-1609.

Schreieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L.A. (2014). RNA polymerase II termination involves CTD tyrosine dephosphorylation by CPF subunit Glc7. Nat Struct Mol Biol *21*, 175-179.

Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev *14*, 2435-2440.

Schwer, B.S., S. (2011). Deciphering the RNA Polymerase II CTD Code in Fission Yeast. Mol Cell *43*, 9.

Segref, A., Sharma, K., Doye, V., Hellwig, A., Huber, J., and Lührmann, R. (1997). Mex67p, a novel factor for nuclear mRNA export, binds to both poly (A)+ RNA and nuclear pores. EMBO J *16*, 3256-3271.

Shpakovski, G.V., Acker, J., Wintzerith, M., Lacroix, J.F., Thuriaux, P., and Vigneron, M. (1995). Four subunits that are shared by the three classes of RNA polymerase are functionally interchangeable between Homo sapiens and Saccharomyces cerevisiae. Mol Cell Biol *15*, 4702.

Sims III, R.J., Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. Genes Dev *18*, 2437-2468.

Sims, R.J., Rojas, L.A., Beck, D., Bonasio, R., Schüller, R., Drury, W.J., Eick, D., and Reinberg, D. (2011). The C-Terminal Domain of RNA Polymerase II Is Modified by Site-Specific Methylation. Science *332*, 99-103.

Singh, B.N., Ansari, A., and Hampsey, M. (2009a). Detection of gene loops by 3C in yeast. Methods *48*, 361-367.

Singh, B.N., and Hampsey, M. (2007). A Transcription-Independent Role for TFIIB in Gene Looping. Mol Cell *27*, 806-816.

Singh, N., Ma, Z., Gemmill, T., Wu, X., DeFiglio, H., Rossettini, A., Rabeler, C., Beane, O., Morse, R.H., and Palumbo, M.J. (2009b). The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. Mol Cell *36*, 255-266.

Skourti-Stathaki, K., Proudfoot, Nicholas J., and Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Mol Cell *42*, 794-805.

Somesh, B.P., Reid, J., Liu, W.F., Søgaard, T.M.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2005). Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. Cell *121*, 913-923.

Songyang, Z., Lu, K., Kwon, Y.T., Tsai, L., Filhol, O., Cochet, C., Brickey, D.A., Soderling, T.R., Bartleson, C., and Graves, D.J. (1996). A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. Mol Cell Biol *16*, 6486-6493.

Steinmetz, E., Conrad, N., Brow, D., and Corden, J. (2001). RNA-binding protein Nrd1 directs poly (A)-independent 3'-end formation of RNA polymerase II transcripts. Nature *413*, 327-331.

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

Steinmetz, E.J., and Brow, D.A. (1998). Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. Proceedings of the National Academy of Sciences of the United States of America *95*, 6699-6704.

Steinmetz, E.J., and Brow, D.A. (2003). Ssu72 protein mediates both poly (A)-coupled and poly (A)-independent termination of RNA polymerase II transcription. Mol Cell Biol *23*, 6339-6349.

Stewart, M. (2010). Nuclear export of mRNA. Trends Biochem Sci *35*, 609-617.

Svejstrup, J.Q., Feaver, W.J., LaPointe, J., and Kornberg, R.D. (1994). RNA polymerase transcription factor IIH holoenzyme from yeast. J Biol Chem *269*, 28044-28048.

Svejstrup, J.Q., Li, Y., Fellows, J., Gnatt, A., Bjorklund, S., and Kornberg, R.D. (1997). Evidence for a mediator cycle at the initiation of transcription. Proceedings of the National Academy of Sciences USA *94*, 6075-6078.

Svejstrup, J.Q., Wang, Z., Feave, W.J., Wu, X., Bushnell, D.A., Donahue, T.F., Friedberg, E.C., and Kornberg, R.D. (1995). Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. Cell *80*, 21-28.

Takagi, Y., Masuda, C.A., Chang, W.H., Komori, H., Wang, D., Hunter, T., Joazeiro, C.A.P., and Kornberg, R.D. (2005). Ubiquitin ligase activity of TFIIH and the transcriptional response to DNA damage. Mol Cell *18*, 237-243.

Terzi, N., Churchman, L.S., Vasiljeva, L., Weissman, J., and Buratowski, S. (2011). H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. Mol Cell Biol, MCB. 05590-05511v05591.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol Cell *23*, 853-864.

Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S.*, et al.* (2010). Chemical-genomic dissection of the CTD code. Nat Struct Mol Biol *17*, 1154-1161.

Tirode, F., Busso, D., Coin, F., and Egly, J.M. (1999). Reconstitution of the Transcription Factor TFIIH Assignment of Functions for the Three Enzymatic Subunits, XPB, XPD, and cdk7. Mol Cell *3*, 87-95.

Toth-Petroczy, A., Oldfield, C.J., Simon, I., Takagi, Y., Dunker, A.K., Uversky, V.N., and Fuxreiter, M. (2008). Malleable Machines in Transcription Regulation: The Mediator Complex. PLoS Comp Biol *4*, e1000243.

Ursic, D., Himmel, K.L., Gurley, K.A., Webb, F., and Culbertson, M.R. (1997). The yeast SEN1 gene is required for the processing of diverse RNA classes. Nucleic Acids Res *25*, 4778.

Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D., and Aloni, Y. (1992). Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. Cell *69*, 871-881.

Vasiljeva, L., and Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. Mol Cell *21*, 239-248.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1–Nab3–Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat Struct Mol Biol *15*, 795-804.

Venters, B., and Pugh, B. (2009a). A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. Genome Res *19*, 360-371.

Venters, B.J., and Pugh, B.F. (2009b). How eukaryotic genes are transcribed. Crit Rev Biochem Mol Biol *44*, 117-141.

Viladevall, L., St. Amour, C., Rosebrock, A., Schneider, S., Zhang, C., Allen, J., Shokat, K., Schwer, B., Leatherwood, J., and Fisher, R. (2009). TFIIH and P-TEFb Coordinate Transcription with Capping Enzyme Recruitment at Specific Genes in Fission Yeast. Mol Cell *33*, 738-751.

Vinciguerra, P., and Stutz, F. (2004). mRNA export: an assembly line from genes to nuclear pores. Curr Opin Cell Biol *16*, 285-292.

Vojnic, E., Simon, B., Strahl, B.D., Sattler, M., and Cramer, P. (2006). Structure and Carboxyl-terminal Domain (CTD) Binding of the Set2 SRI Domain That Couples Histone H3 Lys36 Methylation to Transcription. J Biol Chem *281*, 13-15.

Werner-Allen, J.W., Lee, C.-J., Liu, P., Nicely, N.I., Wang, S., Greenleaf, A.L., and Zhou, P. (2011). cis-Proline-mediated Ser(P)5 Dephosphorylation by the RNA Polymerase II C-terminal Domain Phosphatase Ssu72. J Biol Chem *286*, 5717-5726.

Werner, M., Thuriaux, P., and Soutourina, J. (2009). Structure-function analysis of RNA polymerases I and III. Curr Opin Struct Biol *19*, 740-745.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5 3 exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522-525.

Wlotzka, W., Kudla, G., Granneman, S., and Tollervey, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J *30*, 1790-1803.

Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2003). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. J Biol Chem *278*, 34739-34742.

Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2005). The Bur1/Bur2 complex is required for histone H2B monoubiquitination by Rad6/Bre1 and histone methylation by COMPASS. Mol Cell *20*, 589-599.

Workman, J. (2006). Nucleosome displacement in transcription. Genes Dev *20*, 2009-2017.

Woychik, N., and Young, R. (1994). Exploring RNA polymerase II structure and function. Transcription: mechanisms and regulation Raven Press, New York, NY, 227-242.

Wu, X., Wilcox, C.B., Devasahayam, G., Hackett, R.L., Arévalo-Rodríguez, M., Cardenas, M.E., Heitman, J., and Hanes, S.D. (2000). The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. EMBO J *19*, 3727-3738.

Wyce, A., Xiao, T., Whelan, K.A., Kosman, C., Walter, W., Eick, D., Hughes, T.R., Krogan, N.J., Strahl, B.D., and Berger, S.L. (2007). H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex. Mol Cell *27*, 275-288.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Régnault, B., Devaux, F., Namane, A., and Séraphin, B. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly (A) polymerase. Cell *121*, 725-737.

Xiang, K., Nagaike, T., Xiang, S., Kilic, T., Beh, M.M., Manley, J.L., and Tong, L. (2010). Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. Nature *467*, 729-733.

Ying, S.Y., and Lin, S.L. (2009). Intron-mediated RNA interference and microRNA biogenesis. Methods in Molecular Biol *487*, 387-413.

Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. Genes Dev *21*, 160.

Youdell, M.L., Kizer, K.O., Kisseleva-Romanova, E., Fuchs, S.M., Duro, E., Strahl, B.D., and Mellor, J. (2008). Roles for Ctk1 and Spt6 in Regulating the Different Methylation States of Histone H3 Lysine 36. Mol Cell Biol *28*, 4915-4926.

Young, R.A. (1991). Rna polymerase ii. Annu Rev Biochem *60*, 689-715.

Yuan, G., Liu, Y., Dion, M., Slack, M., Wu, L., Altschuler, S., and Rando, O. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626-630.

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

Zeidan, Q., and Hart, G.W. The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. J Cell Sci *123*, 13.

Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodríguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. (2012a). Ssu72 Phosphatase-dependent Erasure of Phospho-Ser7 Marks on the RNA Polymerase II C-terminal Domain Is Essential for Viability and Transcription Termination. J Biol Chem *287*, 8541-8551.

Zhang, D.W., Rodriguez-Molina, J.B., Tietjen, J.R., Nemec, C.M., and Ansari, A.Z. (2012b). Emerging Views on the CTD Code. Genetics research international *2012*, 347214.

Zhang, L., Fletcher, A.G.L., Cheung, V., Winston, F., and Stargell, L.A. (2008). Spn1 regulates the recruitment of Spt6 and the Swi/Snf complex during transcriptional activation by RNA polymerase II. Mol Cell Biol *28*, 1393.

Zhang, Z., Fu, J., and Gilmour, D.S. (2005). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. Genes Dev *19*, 1572-1580.

Zhou, K., Kuo, W., Fillingham, J., and Greenblatt, J. (2009). Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. Proceedings of the National Academy of Sciences USA *106*, 6956-6961.

Chapter 2: Non-canonical CTD kinases guide gene-class targeted functions of RNA polymerase II

This chapter has been adapted from "Non-canonical CTD kinases guide gene-class targeted functions of RNA polymerase II" by Corey M. Nemec, Joshua M. Gilmore, Yi-Hsuan Ho, Corinna Hintermair, Amit K. Singh, Kennedy J. Ringelberg, Sandra C. Tseng, Martin Heidemann, Ying Zhang, Laurence Florens, Md. Sohail Akhtar, Dirk Eick, Audrey P. Gasch, Michael P. Washburn, and Aseem Z. Ansari.

2.1 Introduction

Each stage of transcription requires ordered recruitment and exchange of specific protein complexes that act on RNA polymerase II, nascent transcripts and underlying chromatin. This dynamic process is orchestrated via patterned post-translational modifications of the CTD. This unusual and essential domain of Rpb1, the largest component of the 12-subunit polymerase, consists of different number of repeating $Y_1S_2P_3T_4S_5P_6S_7$ heptapeptides (26 repeats in budding yeast and 52 in human) (Corden et al., 1985). The mechanistic consequences of phosphorylating Ser5 and Ser2 have been well documented (Bentley, 2014; Buratowski, 2009; Egloff and Murphy, 2008; Eick and Geyer, 2013; Hsin and Manley, 2012; Phatnani and Greenleaf, 2006; Zhang et al., 2012b). However, the role of Thr4 phosphorylation (pThr4) and even the necessity of Thr4 for cellular survival appears to differ amongst closely related species and between growth conditions within a given species (Hintermair et al., 2012; Hsin et al., 2011; Rosonina et al., 2014; Schwer et al., 2014). Recent mass spectrometric analysis of a highly engineered CTD revealed a paucity of pThr4, raising questions about the importance of this mark (Suh et al., 2016). In contrast, similar studies with two versions of the CTD with fewer sequence alterations found pThr4 marks at levels comparable to or greater than the ubiquitous pSer2 mark in both yeast and human cells (Schüller et al., 2016) (Figure 2.1A). It remains to be determined if the low pThr4 levels observed in the study that relied on a single highly altered CTD is reflective of the abundance and functional relevance of this mark in the unaltered, evolutionarily conserved, endogenous CTD.

The identification of CDK9, a promiscuous kinase that can phosphorylate four of the five hydroxyl-bearing residues of the CTD heptad (Glover-Cutter et al., 2009; Hsin et al., 2011; Hsin et al., 2014; Keogh et al., 2003; Murray et al., 2001; Tietjen et al., 2010) and is active at nearly all Pol II transcribed genes, has done little to clarify the role of pThr4 in regulating distinct signal-responsive genes (Rosonina et al., 2014; Schwer et al., 2014). In humans, PLK3 (Polo-like

kinase 3) places the pThr4 mark to facilitate transcription elongation of protein-coding genes, while related PLK1 and Cdc5, the human and yeast homologs of PLK3, do not do so efficiently (Hintermair et al., 2012) (Figure 2.1). Taken together these observations suggest that signal-responsive kinases may be recruited to specific genes to phosphorylate Thr4 and confer some of the diverse functions of pThr4 that are observed among different eukaryotes.

We surveyed over half of the yeast kinome, representing multiple kinases from all structural groups, and identified 11 that phosphorylate Thr4. Further survey of human kinases revealed additional 10 kinases that also phosphorylate Thr4 on human Rpb1 (e.g. yeast Hrr25 and its human orthologs CK1A and CK1E). Many of the newly identified Thr4 kinases belong to structural groups not known to phosphorylate the CTD. Focusing on yeast Thr4 kinases, we utilized a chemical-genetic approach to conditionally attenuate activity of specific Thr4 kinases in vivo. The results revealed that inhibition of Hrr25, Pho85, Cdc28, and Bur1 reduces bulk pThr4 levels. Selectively inhibiting Hrr25 or substituting all Thr4 in the CTD with an alanine (T4A) results in defective transcription termination of a subset of snoRNA genes. Using quantitative mass spectrometry, we identify discrete complexes (Rtt103 and associated termination machinery, the PAF1 complex, Protein Kinase A, and transcription elongation factors) that show diminished binding to the T4A-bearing Pol II. Surprisingly, Rtt103, a key termination factor, directly binds to pThr4 marks placed by Hrr25. Further, the wild type human Rpb1, phosphorylated by a human Thr4 kinase (PLK3) also enhances Rtt103 binding. Because Rtt103 was previously assumed to bind pSer2, we examined both pThr4 and pSer2 levels across the genome. We observe a reciprocal relationship, with high pThr4 and low pSer2 levels at snoRNA genes that display defective termination. The converse pattern is observed at genes that terminate appropriately despite the attenuation or ablation of pThr4 marks. Examining the NMR derived structural ensembles of Rtt103 in complex with pSer2-CTD reveals that a critical arginine in Rtt103 is well positioned to interact with pThr4 or Ser2. Taken together, our data are

consistent with a model in which non-canonical CTD kinases place pThr4 marks to terminate transcription at distinct sets of genes whereas canonical CTD kinases to place the ubiquitous pSer2 marks and engage the termination machinery at most Pol-transcribed genes.

2.2 Search for Thr4 kinases

Previous studies reported that Thr4 is phosphorylated in humans by PLK3 and CDK9 (Hintermair et al., 2012; Hsin et al., 2014). The ubiquitous association of these Thr4 kinases with Pol II at most genes does not readily explain why phosphorylation of Thr4 is essential for regulation of specific set of signal-responsive genes (Rosonina et al., 2014; Schwer et al., 2014). We reasoned that other, as of yet undiscovered, signal- or condition-responsive kinases phosphorylate Thr4 leading to differential use of this mark at different genes. To test this possibility, 70 of 131 yeast kinases with diverse cellular functions from each of the identified structural groups were purified from cells grown in rich medium (Akhtar et al., 2009). From this diverse set of 70 kinases, we identified 11 that phosphorylate Thr4 at least four-fold above background (Figure 2.1B, Figure 2.2). Hrr25, Yck2, Yck3, Cmk1, Cmk2, Pho85, Rim11, Ssk2, and Sln1 are novel CTD kinases, whereas Cdc28 is known to phosphorylate Ser5 at specific set of cell cycle genes, and Bur1 is well known to phosphorylate all three Ser residues (Ser2, Ser5 and Ser7) of the CTD (Breitkreutz et al., 2010; Chymkowitch et al., 2012; Tietjen et al., 2010). We further tested 39 human kinases and identified 10 new kinases that phosphorylate Thr4. The new human Thr4 kinases include CK1A, CK1E, CDK2/CyclinE, CDK7/CyclinH, DAPK1, GSK3B, IKKE, JNK1, MAPKAPK3 and NEK2. We also validated that PLK3 and CDK9/CyclinT can phosphorylate Thr4 (Hintermair et al., 2012; Hsin et al., 2011; Hsin et al., 2014) (Figure 2.1C, Figure 2.2). Remarkably, five of the human Thr4 kinases are phylogenetically related to the yeast Thr4 kinases: CK1A and CK1E are orthologs of Hrr25, CDK2 is the ortholog of Cdc28, CDK9 of Bur1, and GSK3B of Rim11 (indicated by paired symbols on histograms in Figure 1).

This degree of functional conservation, with the retention of Thr4 in the CTD heptad repeat as well as the conservation of kinases that phosphorylate this residue, underscores the importance of the pThr4 mark from yeast to humans.

Surprisingly, three of the most active yeast Thr4 kinases (Hrr25, Yck2, and Yck3) are casein kinases, whereas canonical CTD kinases are cyclin-dependent kinases in the CMGC group. While Hrr25 is known to interact with the CTD phosphorylated at the Ser2 and Ser5 positions (Phatnani et al., 2004), it was not previously identified as a CTD kinase that controls gene expression. Instead, this kinase has been implicated in diverse processes including DNA-damage repair (Hoekstra et al., 1991), ribosome biogenesis (Schäfer et al., 2006), meiotic chromosome segregation (Petronczki et al., 2006), and autophagy (Pfaffenwimmer et al., 2014). Another unexpected Thr4 kinase, Pho85, has roles in nutrient and phosphate sensing and cell cycle control (Measday et al., 1997; Schneider et al., 1994), which may explain the previously implicated role of pThr4 in phosphate metabolism (Rosonina et al., 2014; Schwer et al., 2014). Pho85 binds and phosphorylates sequence specific transcription factor Pho4, however, the ability to directly phosphorylate the CTD was previously unknown. Rim11 is required for meiotic entry via activation of the meiosis regulatory transcription factor, Ime1 (Mitchell and Bowdish, 1992) and it regulates ribosome biogenesis (Yabuki et al., 2014). Finally, Ssk2, Cmk1, Cmk2, and Sln1 are involved in response to multiple environmental stresses (Chasman et al., 2014; Ding et al., 2014; Posas and Saito, 1998; Posas et al., 1996; Sánchez-Piris et al., 2002). We note that the kinases we tested were purified from cells grown in rich media and it is quite likely that kinases purified from cells stimulated by different signals and growth conditions may reveal additional signal-responsive Thr4 kinases.

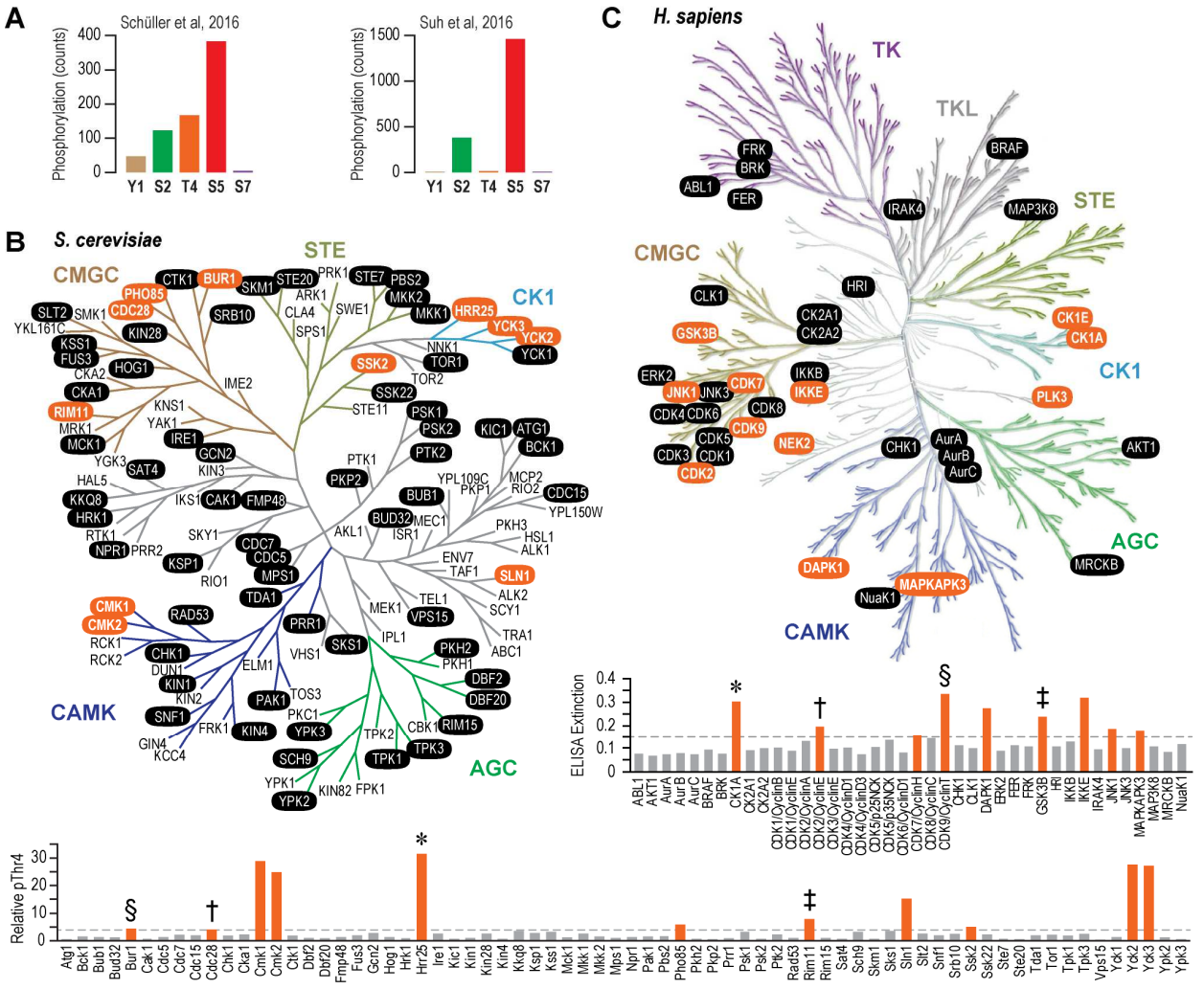

Figure 2.1 Yeast and human kinases phosphorylate Thr4

A) Abundance of phosphorylation at each residue determined by mass spectrometry. Plots are adapted from publicly available data (Schüller et al., 2016; Suh et al., 2016).

B) 131 *S. cerevisiae* kinases organized by maximum likelihood. Colored branches indicate kinase groups. In vitro-tested kinases are denoted in black ovals, and significant pThr4 activity is represented in orange ovals. Quantitation of kinase activity is displayed as histograms below the yeast kinome, with orange indicating >4-fold activity over mock.

C) Human kinome (Cell Signaling) with tested kinases as described in (B). Quantitation of ELISAs is shown below the human kinome, with orange ovals indicating significant phosphorylation over background. Symbols, such as the asterisk, dagger, double dagger, and section symbols indicate yeast and human orthologs that both phosphorylate Thr4.

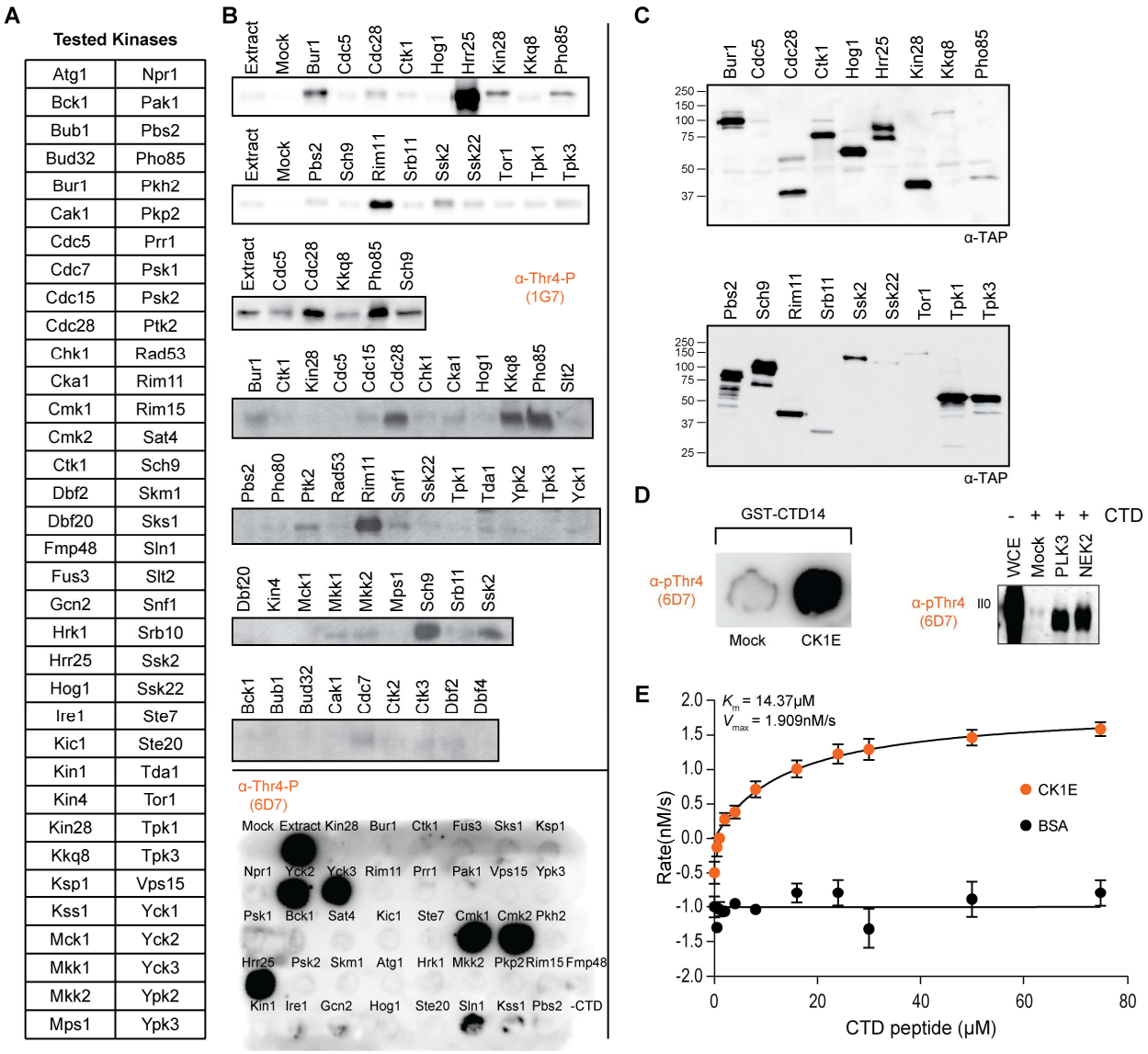

Figure 2.2 Multiple kinases phosphorylate Thr4

A) List of yeast kinases tested in this study. Several kinases (i.e. Cdc7, Ctk1, and Srb10) were purified by co-immunoprecipitating their cyclins or regulatory subunits.

B) Replicates of in vitro kinase assays as performed in Figure 2.1B. Immunoblots were probed with α-pThr4 (1G7 or 6D7 as noted).

C) Kinase purification was validated by probing immunoblots for the TAP-tag.

D) In vitro kinase activity of CK1E, PLK3, and NEK2 were assayed. Recombinant CTD was used as a substrate for CK1E, and unphosphorylated human CTD from HeLa cells (immunoprecipitated with 1C7) was a substrate for PLK3 and NEK2. Blots were probed with the α-pThr4 antibody, 6D7.

E) Kinetic analysis of pThr4 of a three-repeat CTD substrate by CK1E using the ADP biosensor MDCC-ParM. The parameters K_m and V_{max} were obtained from a curve fit according to the Michaelis-Menten equation using GraphPad Prism 5 software. Bovine serum albumin (BSA) was used as a negative control.

2.3 In vivo validation of biochemically identified Thr4 kinases

To further validate the ability of these kinases to place the pThr4 mark in living cells, we utilized a chemical-genetic approach that permits in vivo inhibition of a targeted kinase without impacting the activity of the other kinases in the cell (Akhtar et al., 2009; Bishop et al., 2000; Kanin et al., 2007; Liu et al., 2004; Lopez et al., 2014; Tietjen et al., 2010). As expected, upon inhibition, we saw no reduction in pThr4 in the WT strain; however, unambiguous reduction in pThr4 levels was observed when the analog-sensitive versions of Hrr25, Pho85, Cdc28, or Bur1 were inhibited with small molecule inhibitors that specifically target engineered kinase active sites. The results compellingly demonstrate that Thr4 is a target of these kinases in vivo (Figure 2.3A).

Targeted chemical inhibition of Hrr25 in vivo had the greatest effect on the reduction of the pThr4 mark. To further confirm that Hrr25 is a Thr4 kinase in vivo, we examined the bulk abundance of the pThr4 mark in kinase-inactive *hrr25-3* cells (Figure 2.3B). pThr4 was reduced to comparable levels in chemically inhibited *hrr25as* and genetically inactivated *hrr25-3*, confirming that Hrr25 phosphorylates Thr4 in vivo. Moreover, pThr4 marks were substantially reduced at protein-coding and non-coding genes that were examined via ChIP-qPCR (Figure 2.5B), further supporting the role of Hrr25 as a Thr4 kinase. The presence of residual pThr4 on the CTD in *hrr25-3* cells suggests that while Hrr25 may be the predominant Thr4 kinase, other kinases phosphorylate Thr4 to some extent (Figure 2.3B).

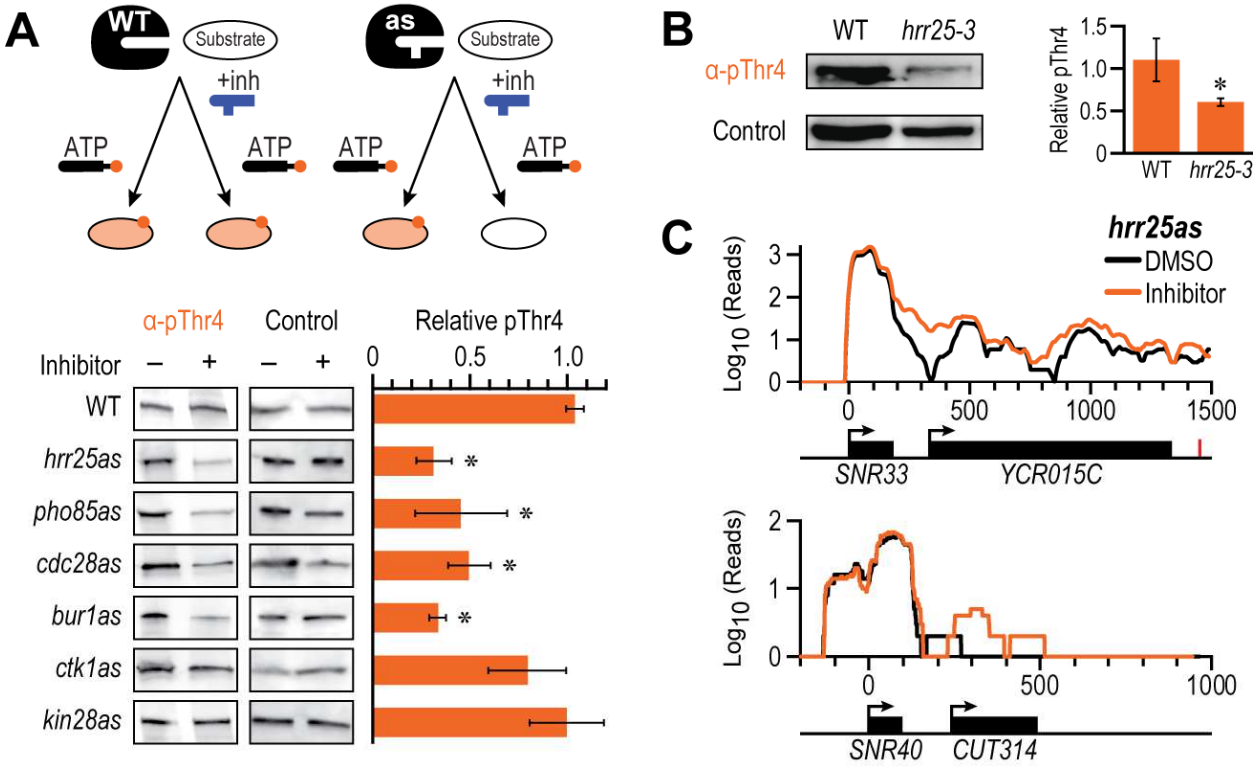

Figure 2.3 Hrr25 phosphorylates Thr4 in vivo

A) Analog-sensitive (-as) kinases were treated in vivo with 20μM of specific chemical inhibitors or an equal volume of vehicle (DMSO). Extract was probed for pThr4 and normalized to Cdc28 as a loading control (*, $p<0.05$. T-test, $n=3$).

B) Extract from WT or kinase-dead *hrr25-3* was probed for pThr4 and normalized to tubulin as a loading control.

C) Log_{10}(RNA-seq reads) of *hrr25as* treated with inhibitor (orange) or DMSO (black) at the *SNR33* or *SNR40* loci. Arrows denote the transcription start site, black bars indicate gene bodies, and red lines indicate cleavage and polyadenylation sites.

While traditional ATP analogs allow for selective and rapid inhibition, the process is dynamic/reversible. Further, while kinase-dead mutants completely inactivate a kinase, other kinases are able to compensate. Therefore, we used an irreversible inhibitor to rapidly, selectively, and permanently inhibit Hrr25 (Figure 2.4A) (Rodríguez-Molina et al., 2016). This method utilizes a similar gatekeeper substitution as with traditional analog sensitive alleles, but also utilizes a strategically placed cysteine substitution to allow for covalent crosslinking to an ATP analog bearing a chloromethylketone (CMK) moiety. Though we observed a clear reduction in pThr4 upon inhibition of *hrr25is*, the Rpb3 subunit was also reduced, suggesting that rapid and irreversible inhibition of Hrr25 may lead to broader defects (Figure 2.4B).

To identify the role of Hrr25-mediated Thr4 phosphorylation, we investigated its impact on the transcriptome by RNA-seq. Upon chemical inhibition of *hrr25as*, a defect in termination at 34 of 77 snoRNAs was observed. For example, significant readthrough transcription was observed at *SNR33* (Figure 2.3C, top and Figure 2.5A), but not at *SNR40* (Figure 2.3C, bottom).

While the effects were apparent at snoRNA, we also observed readthrough defects at some protein-coding genes. Funspec was used to determine if protein-coding genes with readthrough defects had any functional similarities. Indeed, ribosome protein genes and rRNA processing genes displayed significant readthrough defects upon Hrr25 inhibition (p<1E-14). These results corroborate previous work showing that Hrr25 is important for ribosome biogenesis (Ghalei et al., 2015; Ray et al., 2008; Schäfer et al., 2006), and suggest a potential two-pronged mechanism whereby 1) Hrr25-mediated phosphorylation of the CTD modulates ribosome protein and ribosome biogenesis gene synthesis, and 2) Hrr25-mediated phosphorylation of ribogenesis factors cooperate to assemble a complete ribosome.

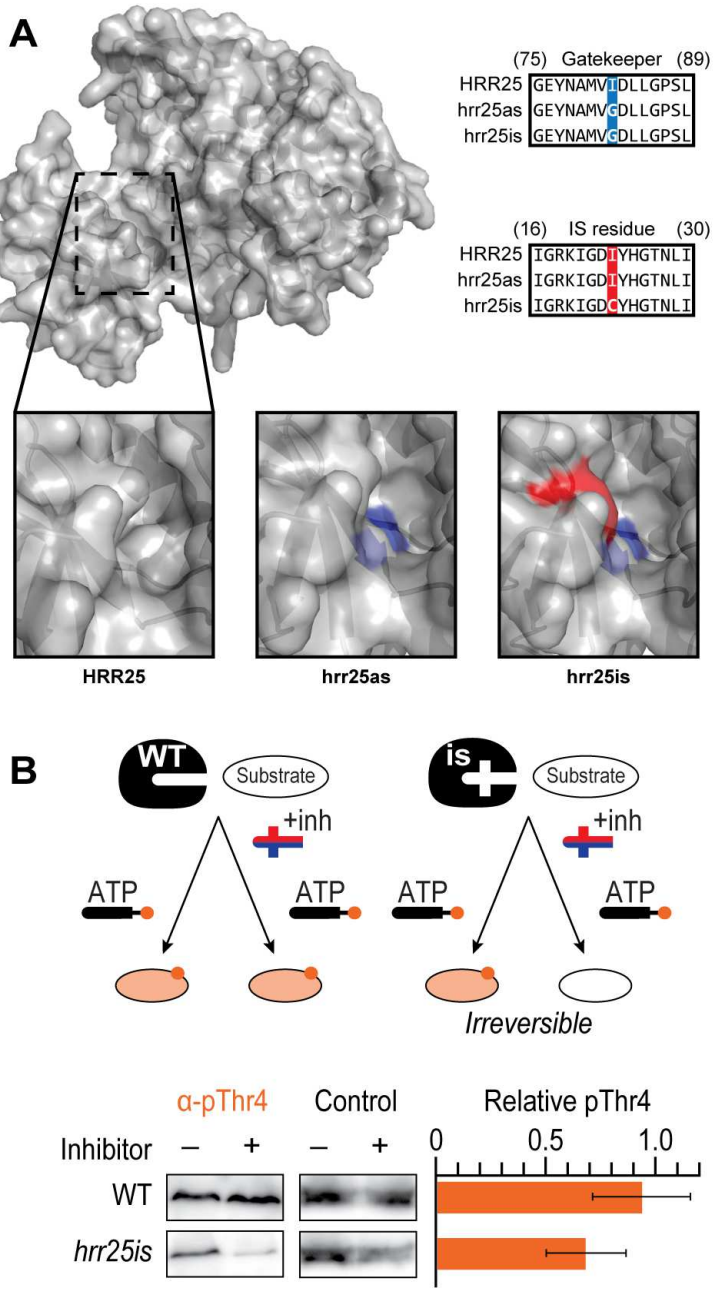

Figure 2.4 Irreversible inhibition of Hrr25

A) Crystal structure of Hrr25 (PDB 4XHL) (Ye et al., unpublished) with ATP binding pocket in the dotted box. Structure was aligned to kin28is (Rodríguez-Molina et al., 2016) to identify residues to substitute. Gatekeeper residue (I82) was substituted with a glycine (blue) in hrr25as and hrr25is (bottom). I23 was substituted with cysteine (red) in hrr25is to allow for covalent inhibition by the inhibitor (bottom right).

B) Schematic for irreversibly inhibiting Hrr25. Irreversibly sensitive (-is) Hrr25 was treated in vivo with 20μM inhibitor (CMK) or an equal volume of vehicle (DMSO). Extract was probed for pThr4 and normalized to Rpb3 as a loading control (p=0.08. T-test, n=5).

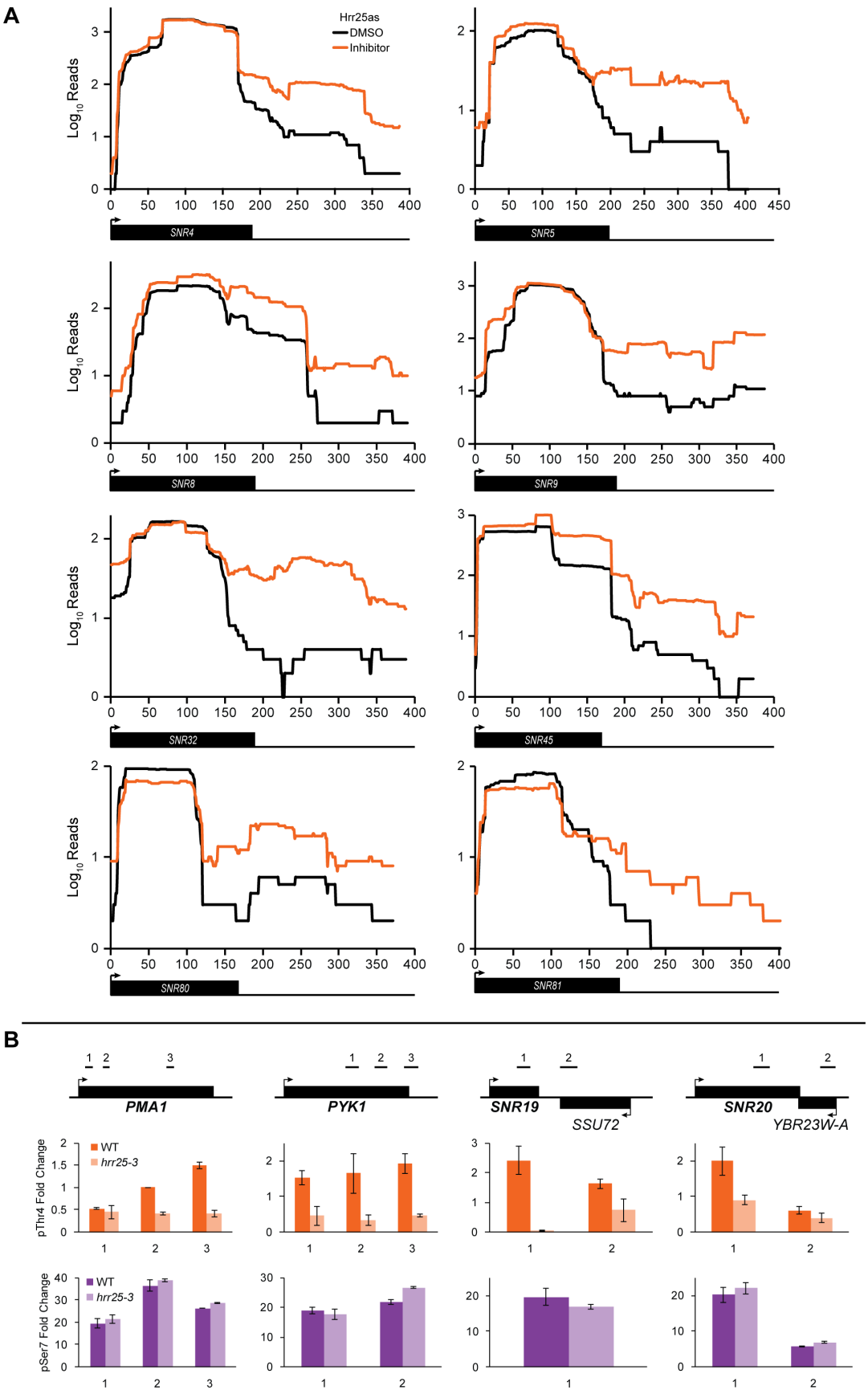
A
Hrr25as
DMSO
Inhibitor
Log10 Reads
SNR4
SNR5
SNR8
SNR9
SNR32
SNR45
SNR80
SNR81
B
PMA1
PYK1
SNR19
SSU72
SNR20
YBR23W-A
WT
hrr25-3
pThr4 Fold Change
pSer7 Fold Change

Figure 2.5 Hrr25 inhibition leads to readthrough at snoRNAs

A) Additional examples of readthrough at snoRNAs in *hrr25as*. Arrows denote transcription start sites and black boxes indicate gene bodies.

B) *hrr25-3* reduces pThr4 across the genome. pThr4 or pSer7 ChIP in WT or *hrr25-3* cells. qPCR was performed at *PMA1*, *PYK1*, *SNR19*, and *SNR20*. Mock IP values were subtracted from the IP values to generate the fold increase in signals using $2^{-\Delta\Delta CT}$ method.

2.4 Thr4 is required for transcription termination

Kinases often act on multiple substrates to evoke complex cellular responses to received signals. To ensure that other substrates of Hrr25 do not indirectly contribute to defective termination, we examined the effect of Rpb1 alleles in which Thr4 was substituted with an alanine (T4A), a residue that cannot be phosphorylated and does not replace a polar hydroxyl moiety of threonine with a hydrophobic methyl group (as is the case with a valine substitution). Alanine substitution is commonly used to map binding interfaces because it is considered a benign perturbation and this residue occurs in structured and unstructured protein folds (Cunningham and Wells, 1989). Alanine substitutions were made across all heptad repeats or in the proximal or distal half of the CTD. We also created a strain in which Thr4 was substituted with the phospho-mimic glutamate (T4E) in each repeat (Figure 2.6A). While the T4E strain was inviable, the T4A strain displayed a slow growth phenotype, as did the kinase-dead *hrr25-3* mutant (Figure 2.6A).

Under exponential growth conditions in rich medium, the transcriptomes of WT and T4A cells displayed few global differences. Avian histone transcripts require pThr4 to process 3'-ends and are not polyadenylated (Hsin et al., 2011); however, in budding yeast, histone genes are polyadenylated (Fahrner et al., 1980). Consistent with this species-specific difference, no defects in 3'-end processing of histone genes were observed in T4A-bearing yeast strains (Figure 2.7A). We also did not observe expression defects similar to those observed in strains containing shortened CTDs (Aristizabal et al., 2013) (Figure 2.7B).

Plotting the fold change in expression between T4A and WT, ~50 striking peaks were observed (Figure 2.6B). Using a stringent z-score cutoff ≥ 5, we found significant enrichment at 3'-ends of snoRNAs. In the T4A strain, 37 of 77 snoRNA transcripts extended well past the mature 3'-end (Figure 2.6C, top and Figure 2.8A). While many formed chimeric transcripts that utilized the

polyA signals of downstream protein-coding genes to terminate transcription (e.g. *SNR33*) (Figure 2.6C) (Kim et al., 2006; Sheldon et al., 2005; Tomson et al., 2013), others were read through and terminated even in the absence of a downstream gene (e.g. *SNR48*) (Figure 2.8A). In contrast, 40 snoRNAs displayed no 3' readthrough defect (e.g. *SNR40*) (Figure 2.6D, top), and substituting Ser2 with an alanine on the first half of the CTD (S2A/WT) did not result in significant readthrough at any snoRNA gene (Figure 2.6C, D). We quantified readthrough at each snoRNA as previously described (Tomson et al., 2013), and identified 37 snoRNAs with a 3' extension (Figure 2.6E).

Northern blots were performed to confirm that these 3' extensions are bona fide readthrough events and not spurious initiation of new transcripts downstream. RNA from WT, T4A, or S2A/WT cells was analyzed with a probe targeting *SNR33* or *SNR40* (Figure 2.6C, 2.6D right). A strong band at approximately 1.5 kb corresponding to the chimeric *SNR33-YCR015C* transcript was observed in T4A but not in WT. The abundance of the unprocessed *SNR33* transcript was reduced in T4A, consistent with a termination defect. As evident, no chimeric *SNR40-CAF40* transcript was observed in any strain tested.

In addition to Northern blot analysis of specific transcripts, genome-wide chromatin immunoprecipitation of the Rpb3 subunit of Pol II was performed to further distinguish 3' processing defects from bona fide transcription termination defects (Figure 2.6C and 2.6D, bottom). The Pol II readthrough index in the region downstream of the 3'-end of processed snoRNAs was significantly correlated (Pearson R = 0.61) with the 3' extension index (from RNA expression) (Figure 2.6F). However, no correlation was observed between the readthrough defect and the length or the abundance of the snoRNA transcripts (Figure 2.6G,H).

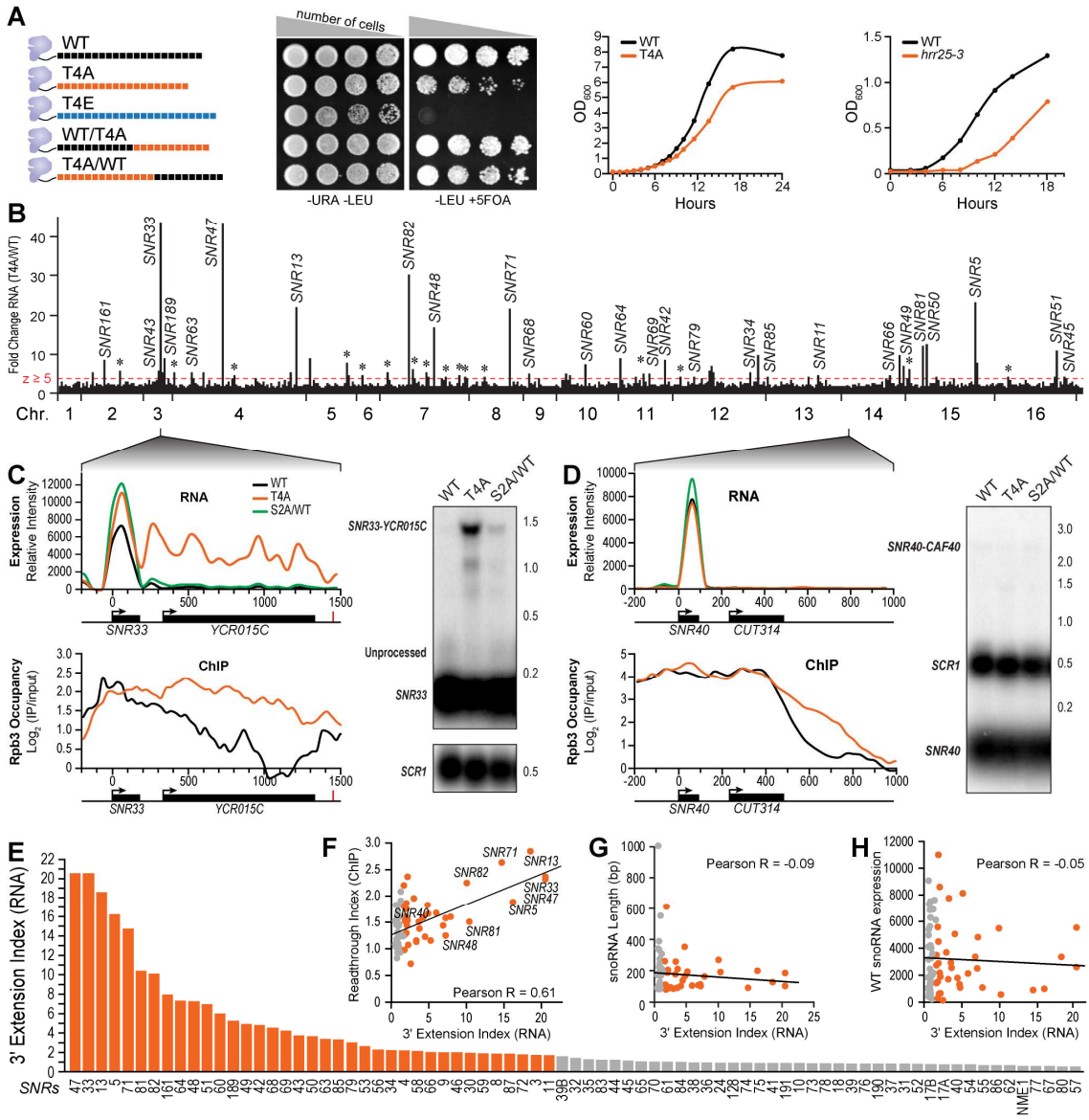

Figure 2.6 pThr4 regulates snoRNA termination

A) Rpb1 with mutant CTDs are shown with black (WT), orange (T4A), and blue (T4E) CTD repeats. Serial dilutions of cells in the presence of 5-FOA are shown. Growth curves of T4A or *hrr25-3* cells are also shown.

B) T4A expression normalized to WT is plotted across the whole yeast genome with chromosomes (Chr.) listed below. Transcripts above the red line denote regions with z ≥ 5 (*, protein-coding genes).

C) RNA expression (top) or Rpb3 ChIP analysis (bottom) from WT (black), T4A (orange), or S2A/WT (green) are shown at *SNR33*. Gene boundaries are illustrated as described in Figure 2.3C. Northern blot probing *SNR33*. *SCR1* is a loading control.

D) Same as panel C at the SNR40 locus.

E) The fold change in 3' extension index (EI) for each snoRNA in T4A is plotted from highest to lowest. Orange bars indicate an EI > 1.7.

F) Comparisons of EI to Rpb3 readthrough, (G) snoRNA length, or (H) snoRNA abundance.

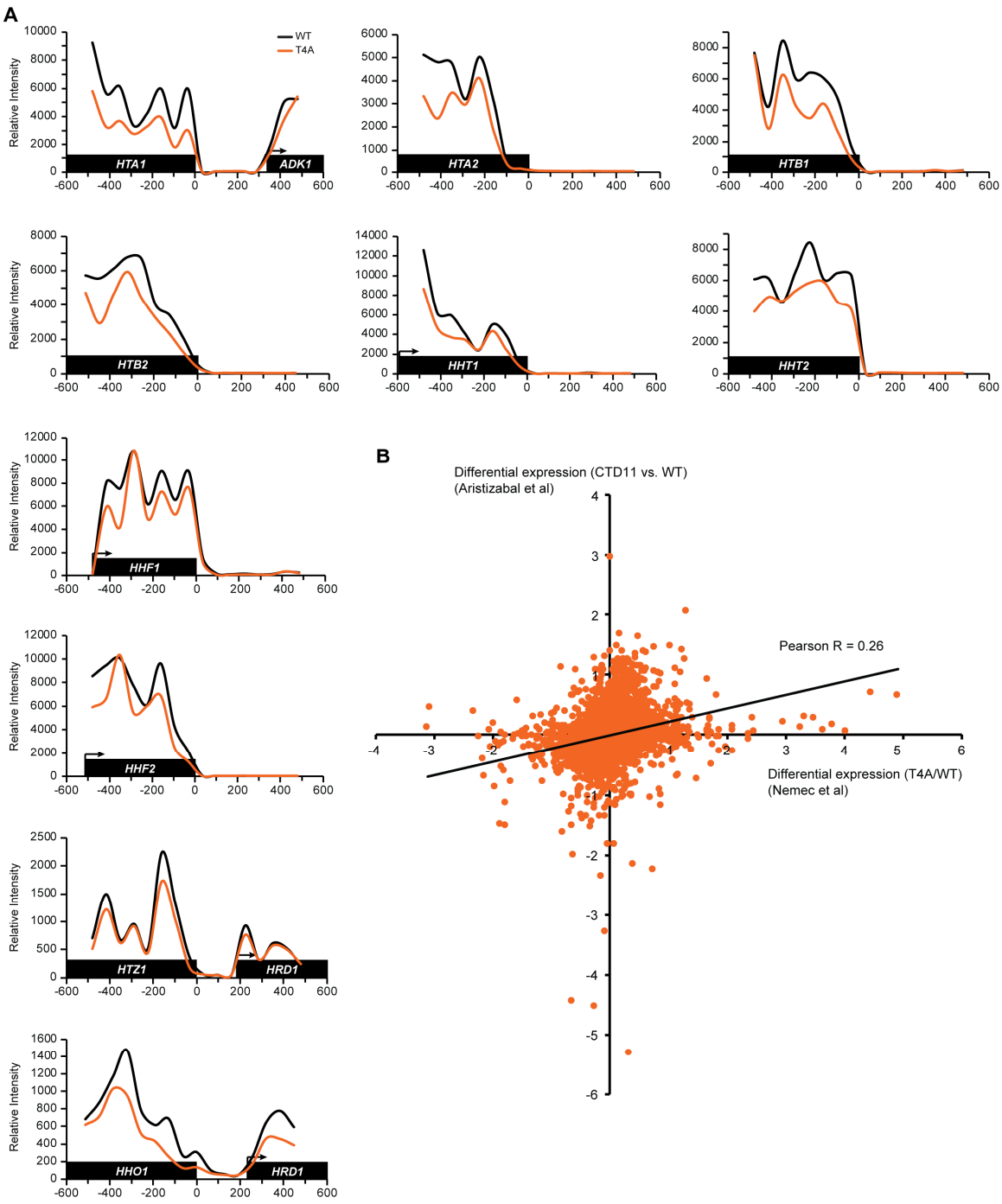

Figure 2.7 T4A does not display defects at histone genes or at genes affected by a short CTD

A) Expression analysis of WT or T4A centered at the 3'-end of histone genes. Black boxes refer to gene bodies. Arrows refer to transcription start sites. No evidence of readthrough is observed at any histone gene.

B) Correlation between T4A and short CTD. The average expression fold change (T4A/WT) was calculated and compared to published data (Aristizabal et al., 2013).

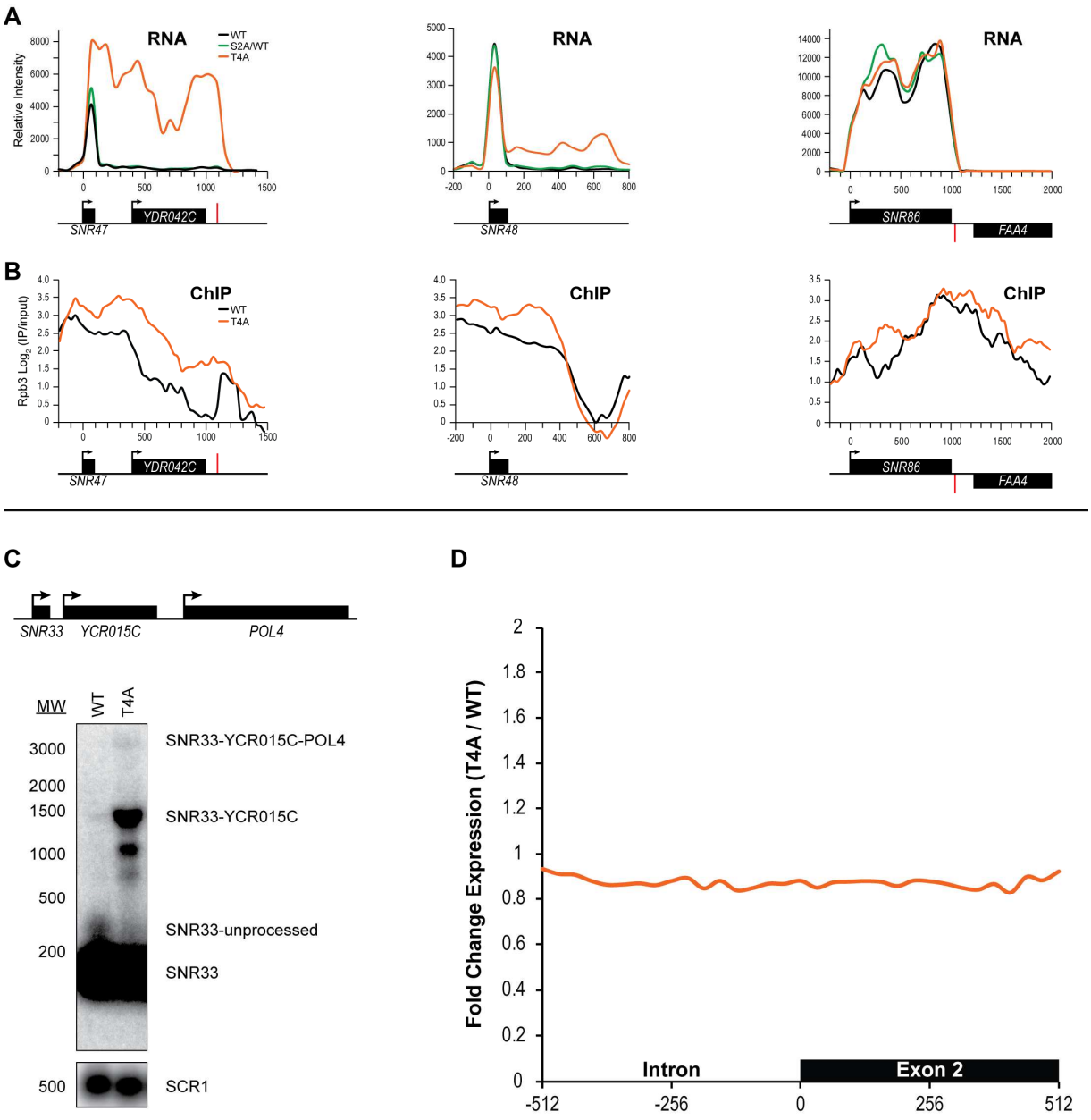

Figure 2.8 Additional evidence of readthrough in T4A

A) Examples of RNA traces or (B) Rpb3 ChIP traces at *SNR47*, *SNR48*, and *SNR86*. Readthrough was observed for *SNR47* and *SNR48* by both RNA expression analysis and ChIP. No evidence of readthrough using either method was observed for *SNR86*. Arrows denote transcription start sites, black boxes indicate gene bodies, and red bars indicate termination sites of protein-coding genes.

C) Northern blot analysis. Longer exposure of the blot from Figure 2.6C. An additional chimeric *SNR33-YCR015C-POL4* transcript is also observed in T4A at low levels.

D) T4A displays no splicing defects. The expression across the genome in T4A was normalized to WT. Metagene analysis of the average fold change in expression across all intron-containing transcripts (except tRNAs) centered at the 3' splice site. Intron retention is not observed in T4A.

2.5 Thr4 is required for association of elongation and termination factors

To identify the cellular factors whose association with RNA polymerase II is dependent on Thr4, we affinity enriched HA-tagged Pol II from strains bearing WT, T4A, and T4E CTDs. A control strain (Z26) lacking the HA-tagged Rpb1 was subjected to identical affinity enrichment procedure. Label-free quantitative proteomics (Liang et al., 2015; Mosley et al., 2013; Zhang et al., 2012a; Zhang et al., 2010) was used to identify and quantify the relative abundance of affinity-enriched complexes (Figure 2.9A). For Pol II enriched from T4A extracts, dramatic reduction was observed for multiple members of the Pol II termination complex (Rtt103, Rat1, Rai1, and Pcf11), c-AMP dependent PKA kinase holoenzyme and regulatory subunit (Tpk2, Tpk3, and Tpk1, and Bcy1), PAF1c complex that bridges elongation and termination (Paf1, Ctr9, Cdc73, Rtf1, Leo1), elongation factors (Set2, Spt4, Spt6, and Elf1), Pol II nuclear import proteins (Npa3, Iwr1, and Gpn3), as well as Fcp1, a phosphatase that acts on pThr4 (Hsin et al., 2014) (Figure 2.9B). The loss of all members of a given protein complex (e.g. Rtt103, PKA, and PAF1c) strongly supports the role of the Thr4 in association of the entire complex with Pol II, and it also serves as an internal control for label-free quantitative proteomic analysis.

A previous report suggested that pThr4 may be important for pre-mRA splicing (Harlen et al., 2016). Under our conditions, only a few components of the spliceosomal machinery (Snu71, Smd2, Syf1, and Lea1) were enriched with Rpb1, irrespective of whether Thr or Ala was at position 4 (Figure 2.9B). The under-representation of the spliceosomal machinery suggests that its interactions with the CTD may be transient or dependent on the presence of RNA. Nevertheless, even in RNA expression data from T4A or *hrr25as* mutant cells, we did not detect significant splicing defects (Figure 2.8D). Any possible effect on splicing may occur at substantially lower levels than termination defects at snoRNAs and some protein-coding genes.

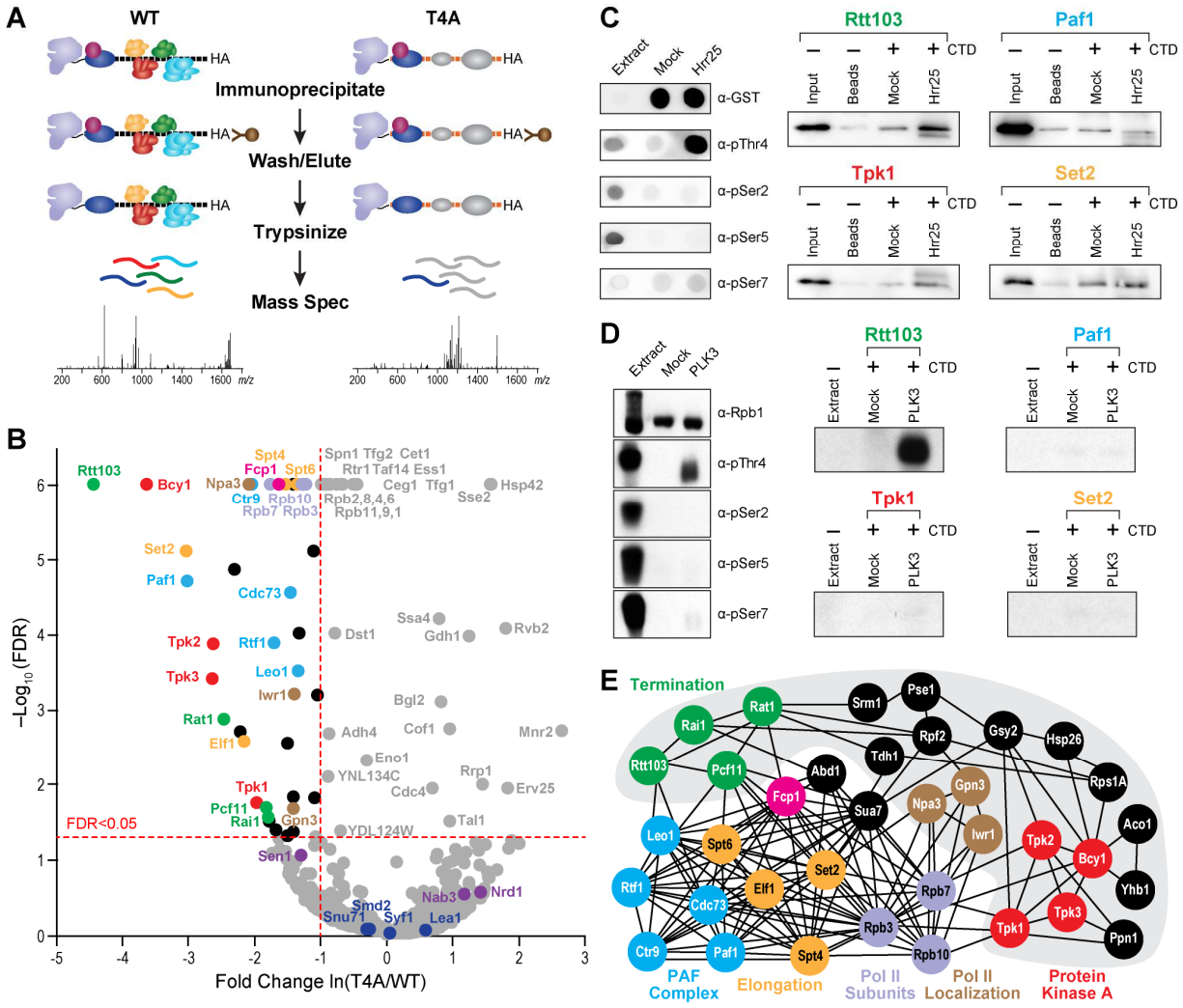

Figure 2.9 Thr4 is required for association of elongation and termination factors

A) Flowchart of the method to identify factors whose binding is compromised in T4A.

B) Volcano plot illustrating the fold change ln(T4A/WT) and the significance $-\text{Log}_{10}$(FDR). Proteins depleted at least one natural log in T4A with an FDR < 0.05 are color-coded and labeled as in (E), and dark blue is used to denote members of the splicing machinery; purple for Nrd1-Nab3-Sen1 (NNS) termination complex.

C) Phosphorylation state of CTD from yeast extract, unphosphorylated GST-CTD, or GST-CTD phosphorylated with Hrr25 (left). Indicated proteins were assayed for binding to pThr4 and probed with α-TAP (right).

D) Phosphorylation state of HeLa extract, unphosphorylated human Rpb1, or Rpb1 phosphorylated with PLK3 (left). Indicated proteins were assayed for binding to pThr4-bearing Rpb1 and probed with α-TAP (right).

E) Physical interconnectivity (string-db.org) between the significantly depleted proteins identified in (B). Grey background indicates low density connectivity.

2.6 Phosphorylation by yeast or human Thr4 kinases facilitates Rtt103 binding

To examine whether factors identified by quantitative mass spectrometry bind pThr4 directly, we first phosphorylated recombinant GST-CTD substrate with Hrr25, or phosphorylated human Rpb1 with PLK3, a human Thr4 kinase (Hintermair et al., 2012). In each case, significant pThr4 was observed with insignificant phosphorylation at other residues (Figure 2.9C and 2.9D left). TAP-tagged Rtt103 from the termination complex, Paf1 from the PAF1 complex, Tpk1, the PKA subunit that associates with Pol II across the transcribed gene (Pokholok et al., 2006), and Set2, an H3K36 methyltransferase, were assayed for direct binding to pThr4-CTD. Rtt103 showed a dramatic increase in binding over the mock treated (unphosphorylated) CTD substrate suggesting direct interaction between the termination factor and the pThr4-bearing CTD. Paf1, Tpk1, and Set2 were subtly enriched in binding to recombinant pThr4-CTD over the unphosphorylated CTD control (Figure 2.9C).

Remarkably, the observation from yeast extends to the human system. The full length human Rpb1, when phosphorylated at Thr4 by PLK3, displayed robust binding to Rtt103, indicative of direct interaction. However, purified Paf1, Tpk1, and Set2 from yeast did not bind appreciably to Thr4-phosphorylated human CTD (Figure 2.9D).

Next, we charted the connectivity between Thr4-dependent partner proteins (Thr4 readers) using publicly available maps of physical contacts. Substantial interconnectivity was evident, as were two groups of high- and low-density networks (Figure 2.9E). Rtt103 termination complex, PKA, and nuclear localization factors form the outer set of interacting nodes, PAF1c, elongation factors, Fcp1, and Pol II subunits form an intricate high-density network. The factors in the high-density network interact with Pol II subunits, suggesting that their interaction with Pol II may be less CTD-dependent than the low-density network containing the Rtt103 termination complex and PKA.

The connectivity map displays remarkable agreement with functional interactions between complexes. Mutually reinforcing functional interactions were reported for PAF1c, Spt6, pSer2 kinase and Set2 (an H3K36 methylase) (Fuchs et al., 2012). Remarkably, PKA, a multifaceted regulator of yeast growth, interacts with Pol II at actively transcribing genes (Pokholok et al., 2006) and phosphorylates the Spt5 subunit of the Spt4/5 elongation complex (Howard et al., 2003). Phosphorylation of both Rpb1-CTD and Spt5 enhance PAF1c recruitment during the early stages of transcription elongation (Liu et al., 2009; Qiu et al., 2012), thus functionally linking many of the Thr4 readers.

2.7 Phospho-Thr4 dependent and independent snoRNAs

To determine why only a subset of snoRNA genes is read through in the T4A mutant, we examined each of the four classes of snoRNA processing sites. Of the 37 snoRNAs with significant readthrough, 36 occurred at terminal processing sites of monocistronic or polycistronic transcripts (p<0.002, Fisher's Exact Test) (Figure 2.10A). We also observed that several snoRNAs even those with minimal 3'-end extension displayed high levels of Pol II downstream of the termination site when assayed by genome-wide ChIP analysis (Figures 2.6D and 2.6F). This result indicates that certain snoRNA transcripts are more defective for transcription termination rather than 3' transcript processing, consistent with a defect in recruiting Rtt103. Remarkably, the termination defects observed in T4A showed significant overlap with the termination defects observed upon chemical inhibition of *hrr25as*, further reinforcing the role of Hrr25 as the Thr4-CTD kinase (Figure 2.10B). Consistent with the reduced association of all subunits of the PAF1 complex (Figure 2.9B) and the role of the PAF1 complex in snoRNA termination (Tomson et al., 2013), the readthrough defects in T4A and *hrr25as* also showed remarkable overlap of the termination defects observed in *paf1Δ* strains (Tomson et al., 2013) (Figure 2.10B, Table 2.1).

Further, we examined the overlap of readthrough defects in T4A and *hrr25as*. While a small fraction of protein coding genes (~2%) displayed readthrough defects in T4A, over 1,000 genes displayed readthrough defects upon Hrr25 inhibition. Nearly half of the read through genes in T4A overlap with *hrr25as*, further supporting the role of Hrr25 as a Thr4 kinase (Figure 2.11). However, readthrough of ribosome protein genes and rRNA processing genes (among others), do not overlap with T4A, suggesting that Hrr25 may have other roles in transcription causing the defects at protein coding genes. Giving credence to this idea, we observed a strong Hrr25 peak over the 3'-ends of protein coding genes (Figure 2.12E).

Previous to this study, Rtt103 and other termination factors for protein-coding genes were only known to interact with pSer2 (Chinchilla et al., 2012; Hollingworth et al., 2006; Kim et al., 2004; Lunde et al., 2010; Noble et al., 2005), we therefore examined the levels of pSer2 normalized to Rpb3 across snoRNAs in a WT strain. Unexpectedly, the snoRNAs with the strongest readthrough defects in T4A displayed low levels of pSer2 (Figure 2.10C, right). To examine the distribution of pThr4 levels across snoRNAs, we performed genome-wide ChIP analysis using the 1G7 α-pThr4 antibody. In contrast to pSer2 profiles, snoRNAs with a readthrough defect in T4A had significantly higher pThr4 levels downstream of the 3'-end (Figure 2.10C, left). As α-pThr4 antibodies yield modest ChIP signals (Mayer et al., 2010; Rosonina et al., 2014), we examined pThr4 signals in other genome-wide studies of this mark. Similar reciprocal patterns of pThr4 and pSer2 profiles were evident in pThr4 ChIP data obtained with a different (6D7) α-pThr4 antibody (Mayer et al., 2010) (Figure 2.12A). Moreover, the respective kinases that place these marks showed similar trends. Genome-wide ChIP analysis with TAP-tagged Hrr25 in the wild type strain revealed enrichment across snoRNA genes that display termination defects in T4A strains. On the other hand, Ctk1, the primary Ser2 kinase, was enriched across snoRNA genes that were Thr4-independent (Mayer et al., 2010). Consistent with low pSer2 levels, Ctk1 was under-represented at snoRNA that were dependent on Thr4 for termination (Figure 2.10D).

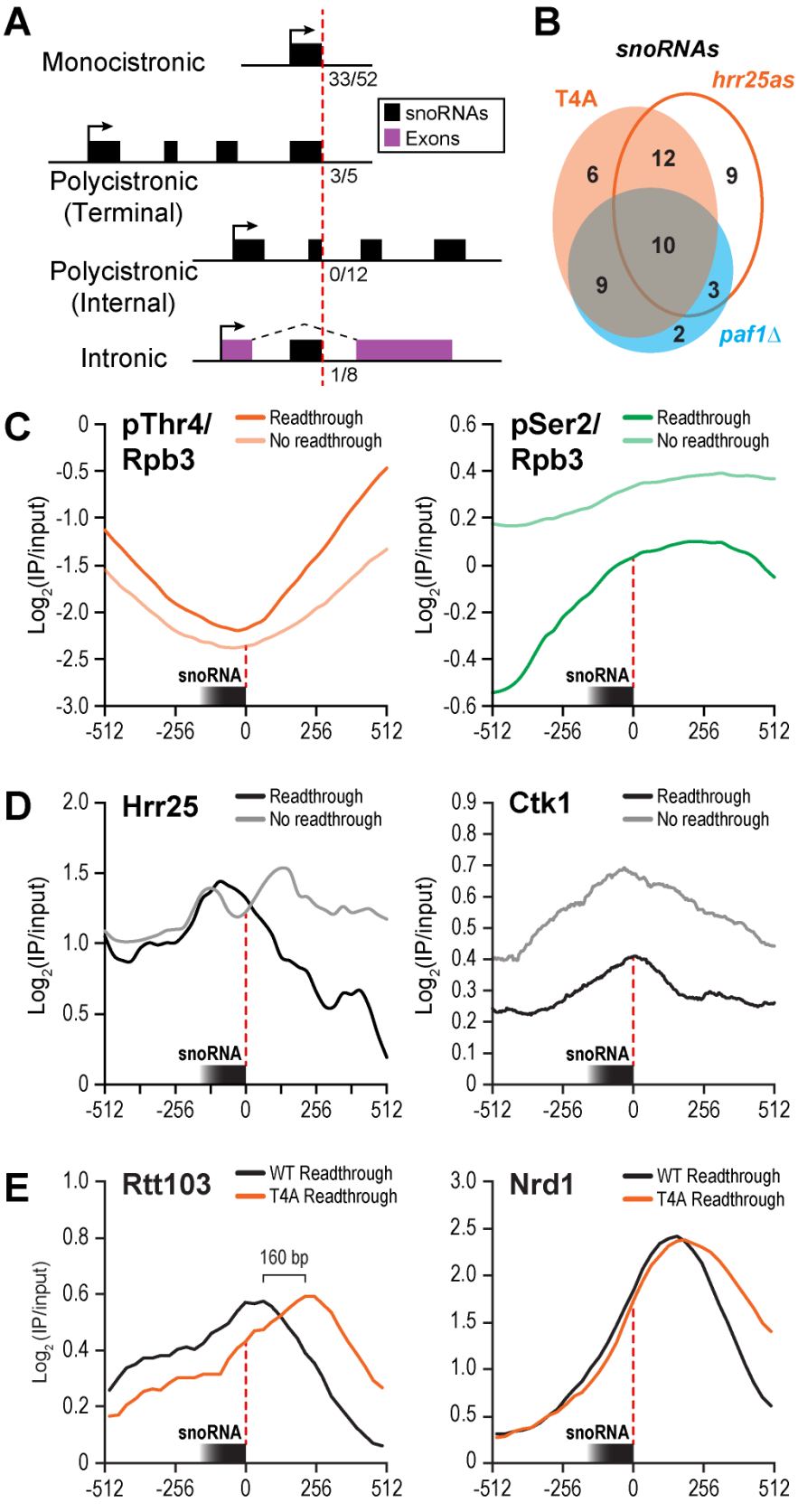
A
Monocistronic
33/52
snoRNAs
Exons
Polycistronic
(Terminal)
3/5
Polycistronic
(Internal)
0/12
Intronic
1/8
B
snoRNAs
T4A
hrr25as
paf1Δ
6
12
9
10
9
3
2
C
pThr4/
Rpb3
Readthrough
No readthrough
Log2(IP/input)
snoRNA
pSer2/
Rpb3
Readthrough
No readthrough
Log2(IP/input)
snoRNA
D
Hrr25
Readthrough
No readthrough
Log2(IP/input)
snoRNA
Ctk1
Readthrough
No readthrough
Log2(IP/input)
snoRNA
E
Rtt103
WT Readthrough
T4A Readthrough
160 bp
Log2 (IP/input)
snoRNA
Nrd1
WT Readthrough
T4A Readthrough
snoRNA

Figure 2.10 Differences between readthrough and non-readthrough snoRNAs

A) Four snoRNA classes are illustrated with black boxes representing snoRNAs and purple representing exons. The fraction of snoRNAs in each class with 3' extensions in T4A are noted next to the red dotted line depicting the mature 3'-end of respective snoRNAs.
B) Venn diagram comparing the number of snoRNAs with readthrough defects from T4A (orange), *hrr25as* (orange outline), and *paf1*Δ (blue) (Tomson et al., 2013).
C) Metagene analysis of pThr4 (orange) or pSer2 (green) ChIP occupancy centered at the 3'-end of snoRNAs either read through in T4A (dark) or not read through (light).
D) Like C), but metagene analysis of kinase (Hrr25 or Ctk1) occupancy.
E) Like C), but metagene analysis of Rtt103 or Nrd1 occupancy in WT (black) or T4A (orange) strains. Profiles are averaged occupancies at snoRNAs read through in T4A.

Table 2.1 Comparison of readthrough indices of snoRNAs

	T4A	*hrr25as*	*paf1*Δ	S2A/WT
SNR47	20.56	1.66	2.33	1.21
SNR33	20.55	2.53	2.18	1.65
SNR13	18.57	1.15	1.12	0.95
SNR5	16.24	4.88	1.90	1.20
SNR71	14.76	0.83	6.28	1.31
SNR81	10.42	5.78	3.75	1.29
SNR82	10.11	5.83	1.73	1.10
SNR161	7.96	1.75	4.77	1.35
SNR64	7.32	1.62	3.06	1.48
SNR48	7.26	1.17	8.60	1.40
SNR51	6.99	1.03	2.37	0.86
SNR60	6.02	1.75	3.86	1.46
SNR189	5.29	9.62	1.56	1.20
SNR49	4.92	2.66	0.96	1.61
SNR42	4.83	1.22	2.29	1.59
SNR68	4.55	1.46	2.35	1.29
SNR69	4.26	1.79	1.86	1.31
SNR43	3.76	2.70	0.65	1.39
SNR50	3.67	1.61	1.60	0.98
SNR63	3.42	2.02	0.70	1.07
SNR85	3.35	0.79	3.52	0.98
SNR79	3.04	3.44	2.83	0.95
SNR53	2.66	2.17	2.30	1.17
SNR56	2.31	1.64	2.13	1.21
SNR34	2.24	8.79	1.72	1.12
SNR4	2.21	4.52	0.98	0.91
SNR58	2.15	0.93	1.56	1.01
SNR66	2.06	1.97	0.66	0.67
SNR9	2.04	8.38	1.63	1.54
SNR46	2.03	4.14	0.86	1.13
SNR30	1.97	9.47	1.55	1.30
SNR59	1.91	0.45	1.03	0.78
SNR8	1.89	2.09	1.16	1.20
SNR87	1.88	1.08	0.25	0.87
SNR72	1.78	0.88	0.83	0.50
SNR3	1.75	10.32	1.47	1.22
SNR11	1.71	2.47	0.95	1.29
SNR39B	1.62	1.11	1.47	1.22
SNR32	1.44	8.88	2.89	1.25
SNR35	1.31	1.63	1.28	1.42
SNR83	1.26	1.57	1.25	1.03
SNR44	1.24	0.82	0.55	0.80

SNR45	1.18	2.01	4.31	1.06
SNR65	1.08	0.29	0.82	1.06
SNR70	1.06	2.48	1.17	0.54
SNR61	1.05	1.64	0.40	0.45
SNR84	1.04	3.04	1.63	0.85
SNR38	0.98	1.12	1.54	1.05
SNR36	0.97	10.79	0.80	1.15
SNR24	0.97	0.52	0.66	0.99
SNR128	0.96	1.86	0.88	1.01
SNR74	0.95	1.12	0.73	1.21
SNR75	0.94	1.59	0.59	1.15
SNR41	0.93	0.86	1.26	1.16
SNR191	0.92	0.79	2.79	0.86
SNR10	0.92	3.72	0.75	1.08
SNR73	0.92	1.05	0.64	1.14
SNR78	0.91	1.08	1.29	0.95
SNR18	0.90	0.41	1.16	0.91
SNR39	0.89	0.22	1.22	0.77
SNR76	0.88	0.83	0.86	1.09
SNR190	0.87	1.63	1.16	1.36
SNR37	0.85	1.24	0.56	1.12
SNR31	0.85	3.45	0.46	1.07
SNR52	0.84	1.15	1.13	1.03
SNR17B	0.82	1.22	0.41	1.12
SNR17A	0.79	1.34	1.64	1.07
SNR40	0.78	0.99	1.47	1.09
SNR54	0.77	1.02	2.12	0.85
SNR55	0.70	1.77	0.99	0.83
SNR86	0.70	3.34	1.79	0.84
SNR62	0.70	1.39	1.04	1.26
NME1	0.65	2.95	1.09	0.68
SNR77	0.64	0.96	0.87	0.94
SNR67	0.64	1.10	0.31	0.89
SNR80	0.56	4.89	0.44	0.76
SNR57	0.55	0.67	0.91	0.78

Readthrough indices were calculated as described in (Tomson et al., 2013), with indices >1.7 (highlighted in red) being deemed significant. T4A and S2A/WT were compared to the consensus WT strain. For *hrr25as*, the inhibited strain was compared to the uninhibited, and *paf1*Δ was compared to WT.

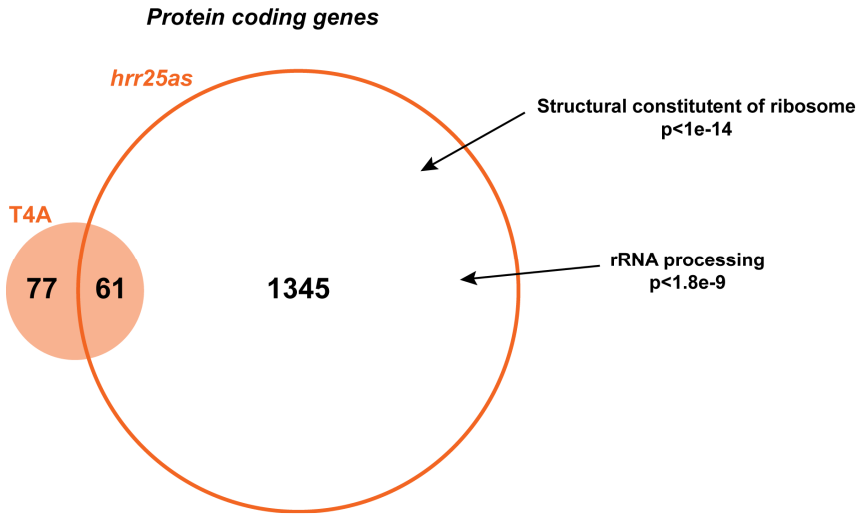

Figure 2.11 Comparison of read through protein coding genes.
Venn diagram comparing the number of protein coding genes with readthrough defects from T4A (orange) or *hrr25as* (orange outline). Funspec was used to identify functional similarities between read through genes. Bonferroni correction was performed.

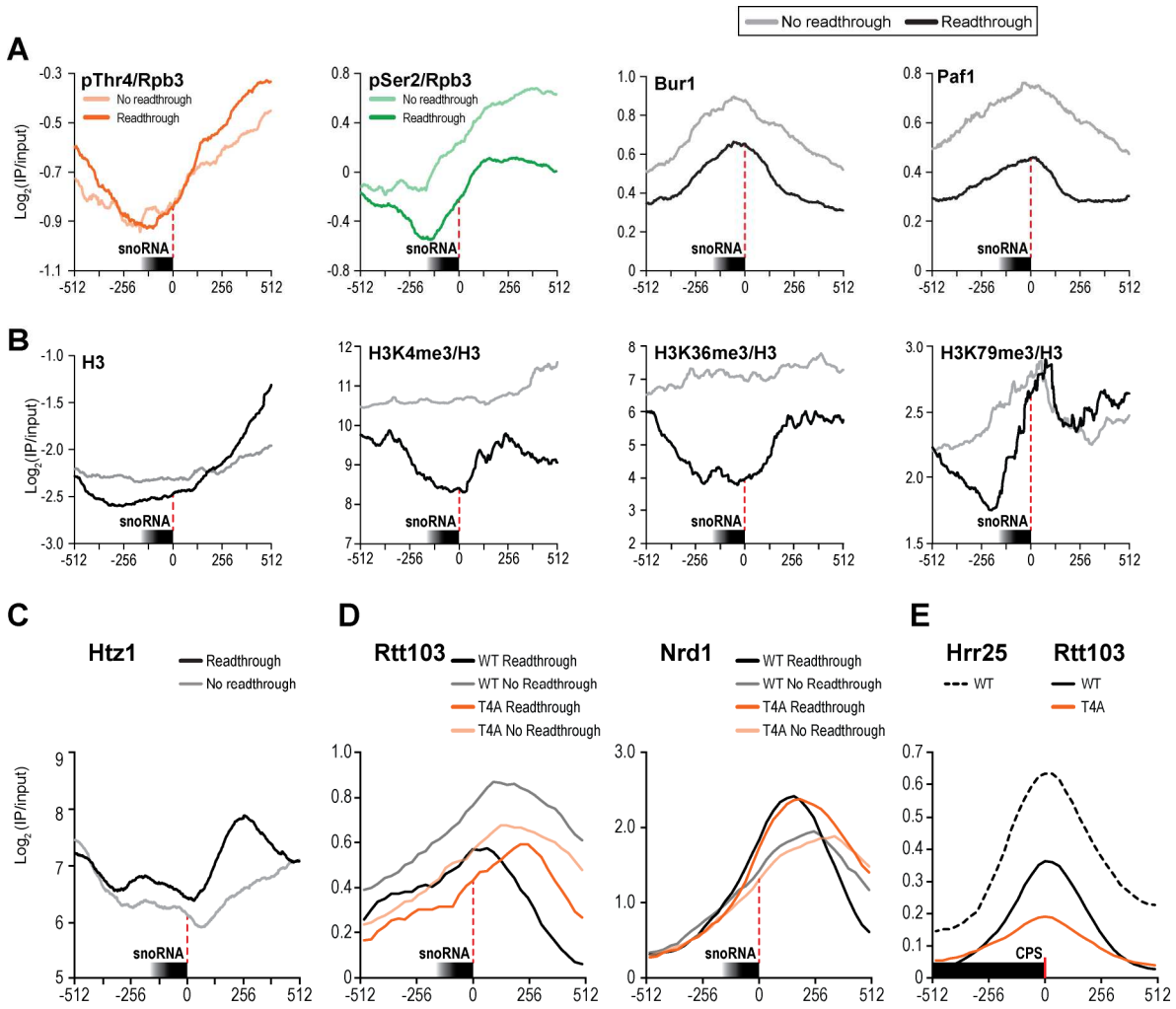

Figure 2.12 Additional ChIP traces

A) Metagene analysis of ChIP occupancy of pThr4 (orange) or pSer2 (green) normalized to Rpb3 centered at the 3'-end of snoRNAs either read through in T4A (dark) or not read through (light). Metagene analysis of ChIP occupancies of Bur1 and Paf1 are also shown at snoRNAs read through in T4A (black) or not read through (grey) (Mayer et al., 2010).

B) Metagene analysis of the ChIP occupancy of histone H3 or its modifications (GSE72802) centered at the 3'-end of snoRNAs either read through in T4A (black) or not read through (grey).

C) Metagene analysis of the ChIP occupancy of Htz1 (Gu et al., 2015) cetered at the 3'-end of snoRNAs either read through in T4A (black) or not read through (grey).

D) Metagene analysis of ChIP occupancy of Rtt103 or Nrd1 centered at the 3'-end of snoRNAs. ChIPs were performed in WT (black) or T4A (orange), and traces are an average of ChIP occupancy across snoRNAs either read through in T4A (dark) or not read through (light).

E) Metagene analysis of the ChIP occupancy of Hrr25 (dotted black) or Rtt103 (solid) centered at the cleavage and polyadenylation signal (CPS) of all protein-coding genes. Rtt103 occupancies were determined in both WT (black) and T4A (orange) strains.

A genetic link between Thr4 and chromatin remodeling through the histone variant Htz1 has been reported (Rosonina et al., 2014). Consistent with a termination defect, Htz1, a histone H2A variant that yields more labile nucleosomes (Li et al., 2005; Santisteban et al., 2011), is highly enriched downstream of snoRNAs that display readthrough in T4A (Gu et al., 2015) (Figure 2.12C). In the context of the chromatin landscape, PAF1c, a complex that is depleted in the T4A interactome, plays a key role in snoRNA termination by recruiting the machinery that modifies histone H3 (Tomson et al., 2013). In agreement with this, significantly lower levels of histone H3 modifications were observed across snoRNA genes that display termination defects in T4A strains (GSE72802) (Figure 2.12B).

Next, we performed genome-wide ChIP analysis of Rtt103 and detected strong reduction in Rtt103 occupancy downstream of snoRNAs with a readthrough defect, confirming the sensitivity of this subset of genes to the absence of pThr4 (Figure 2.12D, left). Further, Rtt103 peaked 160 bp father downstream in T4A than WT (Figure 2.10E, left). Previous work suggested that Rtt103 and Pcf11 cooperatively interact with the Pol II CTD (Lunde et al., 2010) and that Nrd1 is displaced by Pcf11 allowing for Sen1-mediated termination of snoRNA genes (Grzechnik et al., 2015). Therefore, we examined the occupancy of Nrd1 in WT and T4A cells. We observed that in T4A, Nrd1 was retained downstream of the 3'-end of snoRNAs (Figure 2.10E, right) mirroring the readthrough defects in T4A. Similar Nrd1 retention was previously observed in *pcf11-13* mutant cells (Grzechnik et al., 2015).

2.8 Structural basis for pThr4–Rtt103 interaction

The solution-state NMR ensemble of structures of Rtt103 revealed a hydrogen bond between Arg108 of Rtt103 and the phosphate on pSer2 of the CTD (Lunde et al., 2010). The NMR data show that Arg108 side chain can adopt many conformations when bound to the pSer2 CTD peptide (Figure 2.13A). To examine if pThr4 could be accommodated at the same binding

interface of Rtt103, we modeled the CTD with this mark on to the reported NMR structure and performed energy minimization in PyMOL (Schrödinger, 2010) (Figure 2.13B). Strikingly, the resulting models indicate a network of hydrogen bonds between Arg108, the phosphate of pThr4, and hydroxyl of the unmodified Ser2. These interactions include a bifurcated hydrogen bond from the guanidinyl nitrogen of the Arg108 to two oxygens on the phosphate of pThr4 and a bond to hydroxyl of Ser2, this network exceeds that reported for Arg108 and pSer2. These modeling results are consistent with our biochemical experiments that show strong binding between Rtt103 and pThr4 marks placed by yeast and human kinases (Figure 2.9C, 2.9D).

2.9 Discussion

The CTD undergoes sequential waves of post-translational modification during the transcription cycle. These modifications facilitate recruitment of cellular machines that process nascent transcripts, place post-translational marks on histones, and facilitate each stage of the transcription cycle. Unlike other ubiquitous phosphorylation marks on the CTD, pThr4 plays a critical role in regulating transcription at specific gene classes. Due to convergence of results from independent experimental approaches, we have; (i) discovered 21 yeast and human CTD kinases that phosphorylate Thr4, (ii) identified Rtt103, a pSer2 binding termination factor, as a direct reader of the pThr4 mark, (iii) defined a coherent interactome of the pThr4 mark in yeast, (iv) identified two classes of non-coding RNA that depend on either pThr4 or pSer2 to recruit the termination machinery, and (v) functionally linked the diverse set of signaling kinases, CTD marks, and interacting protein complexes to transcriptional termination at distinct sets of non-coding and protein-coding genes.

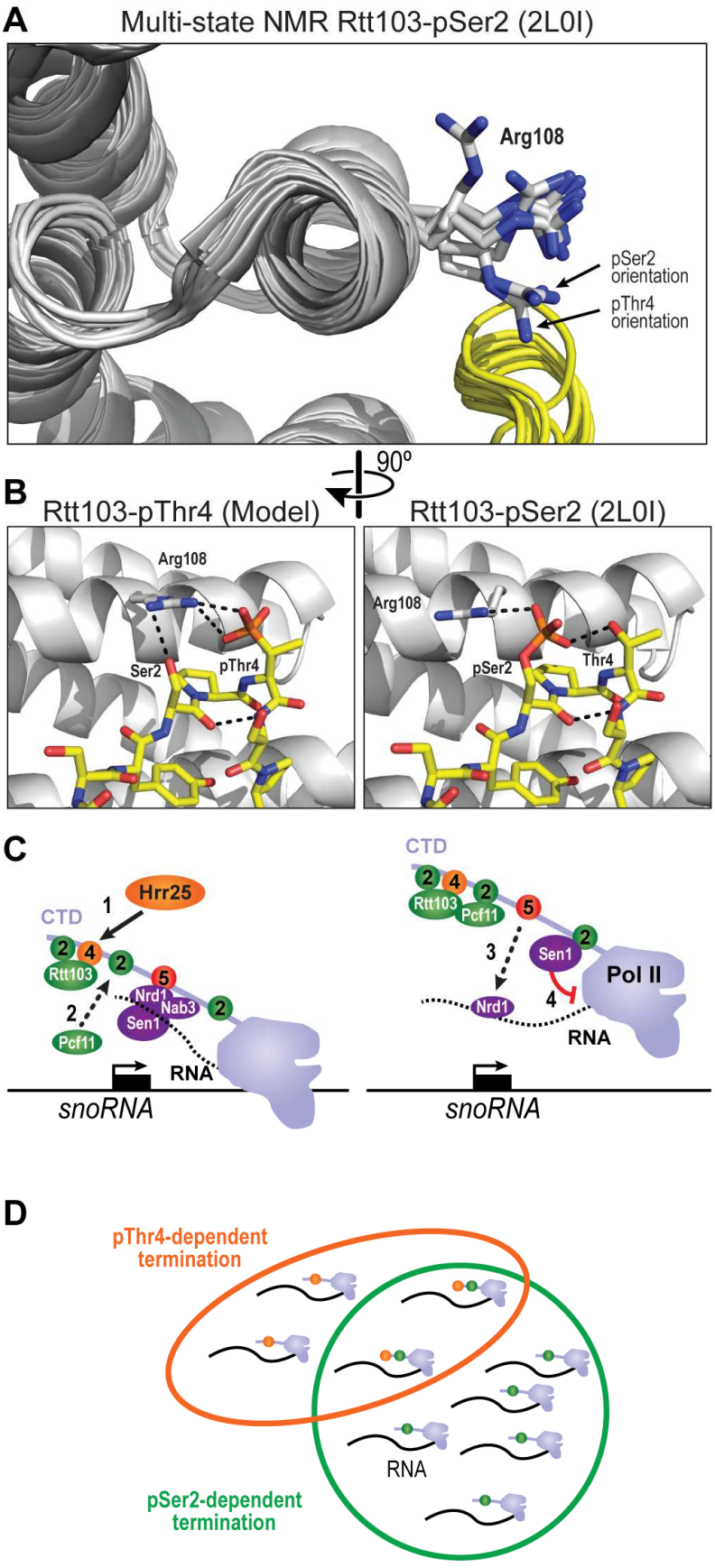
A
Multi-state NMR Rtt103-pSer2 (2L0I)
Arg108
pSer2 orientation
pThr4 orientation
90°
B
Rtt103-pThr4 (Model)
Rtt103-pSer2 (2L0I)
Arg108
Ser2
pThr4
Arg108
pSer2
Thr4
C
CTD
Hrr25
1
2
4
2
Rtt103
5
Nrd1
Nab3
2
2
Pcf11
Sen1
RNA
snoRNA
CTD
2
4
2
Rtt103
Pcf11
5
2
3
Sen1
Pol II
Nrd1
4
RNA
snoRNA
D
pThr4-dependent termination
RNA
pSer2-dependent termination

Figure 2.13 Structural basis for pThr4-Rtt103 interaction

A) Multi-state NMR depiction of Rtt103 in complex with pSer2-CTD (Lunde et al., 2010). Arg108 rotamers in (B) are labeled. All structural figures were generated with PyMOL (Schrödinger, 2010).

B) Representative solution state structure of Rtt103 in complex with pSer2 CTD (right) and a model of Rtt103 in complex with pThr4 CTD (left).

C) Model illustrating how a balance in pThr4 and pSer2 can recruit Rtt103 to different genes to regulate termination in a gene-class specific manner. 1) Hrr25 phosphorylates Thr4 at pThr4-dependent snoRNAs. 2) Pcf11cooperatively interacts with Rtt103 and 3) ejects Nrd1 from the CTD. This allows Sen1 to associate with the CTD and 4) terminate transcription. Solid arrow depicts phosphorylation, dotted arrows illustrate binding to or ejection from the CTD, and the red blunt arrow indicates transcription termination.

D) pThr4 (orange) and pSer2-dependent (green) transcripts, the overlap indicates the potential ability of signal-responsive Thr4 kinases in terminating selective genes under certain conditions.

The model that emerges from our results posits that Hrr25 phosphorylates Thr4 at a subset of snoRNA genes. At this subset, Rtt103 is recruited to the CTD directly by pThr4. Pcf11 cooperatively interacts with Rtt103 and ejects Nrd1 from the CTD, allowing Sen1 to associate. The helicase activity of Sen1 terminates Pol II transcription (Figure 2.13C). We propose that in the absence of pThr4, Rtt103 and Pcf11 cannot cooperatively assemble at pThr4-dependent snoRNAs; thus, Nrd1 is retained, and Sen1-mediated termination does not occur. Reduced pThr4 at pSer2-dependent snoRNAs does not prevent termination because Rtt103 can directly associate with pSer2. This model is consistent with our mass spectrometry results revealing that Sen1 is depleted from T4A CTD, whereas Nrd1 is enriched (Figure 2.9B).

In further support of this model, we find that chemical inhibition of the Thr4 kinase, Hrr25, results in defective transcription termination at a subset of snoRNA genes. This defect is mirrored in a yeast strain with T4A substitution in all heptad repeats of the CTD. Remarkably, in the absence of the pThr4 mark, Pol II transcribes through the *SNR33* termination site and continues to transcribe the entire downstream protein-coding gene, *YCR015C*, using its cleavage and polyadenylation site to terminate transcription of the chimeric *SNR33-YCR015C* transcript. Taken together, our data reveal two classes of snoRNA genes with termination at one class dependent primarily on pThr4 marks and termination at the other class primarily dependent on pSer2. In fact, pThr4 and pSer2 occupancy at two classes of snoRNAs appears inversely correlated (Figure 2.10C). The surprising observation that pThr4 mark is read by Rtt103, a key termination factor, suggests that a balance between pThr4 and pSer2 levels control termination at specific genes (Figure 2.13C, 2.13D). Unlike the global role of pSer2, phosphorylation of Thr4 provides a ready means to regulate a subset of genes to which a given kinase is targeted by specific cellular signals. This model is consistent with variable levels of pThr4 across gene-classes and the different roles that pThr4 has been implicated in.

Recent mass spectrometric analysis of the phosphorylation patterns on the CTD led to two opposing results, with barely detectable levels of pThr4 in one study (Suh et al., 2016) and levels comparable pSer2 in the other (Schüller et al., 2016). It remains to be determined whether this inconsistency is due to the extent of CTD engineering that was required to resolve phosphorylation patterns among the repeating heptads or differences in growth conditions. Nevertheless, even the highly engineered CTD shows evidence of pThr4 marks in the fraction that is bound to Rtt103 (Suh et al., 2016).

Rtt103 was previously only known to function at protein-coding genes (Kim et al., 2004). In human cells, pThr4 levels peak downstream of the 3'-end of protein-coding genes, akin to pSer2 levels (Hintermair et al., 2012). That these two marks occur on the CTD near the 3'-end of protein-coding genes suggests that both may serve a role in termination of protein-coding genes. Indeed, we observe in our ChIP studies that both Hrr25 and Rtt103 peak over the cleavage and polyadenylation signal (CPS) of protein-coding genes in yeast (Figure 2.12E). In the absence of pThr4, a minor but detectable fraction of the Pol II escapes termination at *YCR015C* and continues through the next protein-coding gene, *POL4*, and terminates at its 3' end (Figure 2.8C). Subtle termination defects at other protein-coding genes were also evident in the chemically-inhibited *hrr25as* strain as well as in the T4A strain. The results support the notion that coding or non-coding could be regulated under certain growth conditions by non-canonical kinases that place pThr4 marks and by canonical CTD kinases that place pSer2 marks under other growth conditions (Figure 2.13D).

Beyond Hrr25 and PLK3, many of the Thr4-CTD kinases we identify here are known to play critical roles in cellular signaling pathways and are linked to the etiology of various human diseases (Lei et al., 2011; Malumbres, 2014; Perez et al., 2011). We anticipate that the linkages and regulatory mechanisms we present here will lead to new insights, provoke new hypotheses,

and stimulate the examination of underexplored kinases such as, GSK3B, CDK2, CK1A, NEK2, MAPKAPK3, and DAPK1 in directly targeting the CTD and impacting transcription termination at selective genes in normal and diseased human cells.

2.10 Future directions

To fully validate the proposed model (Figure 2.13C), several experiments must be performed. First, immunoprecipitating Hrr25 (or other kinases) using the TAP tag will co-immunoprecipitate associated machinery. It is possible that an associated kinase phosphorylates Thr4. To validate these in vitro data, we inhibited Hrr25 in vivo using an ATP analog (Figure 2.3). However, many of the kinases identified in Figure 2.1 are involved in signal transduction (see Table 3.1). It is possible that inhibition of our target kinase down-regulates an entire signaling cascade, which ultimately results in decreased pThr4. To confirm that the identified kinases directly phosphorylate Thr4, we propose to purify Hrr25 or Hrr25as and use N^6-benzyl-ATP as a phospho-donor. N^6-benzyl-ATP can be utilized by Hrr25as by docking into the expanded ATP binding pocket, but cannot be used by WT Hrr25 (Figure 2.14). Phosphorylation of Thr4 by Hrr25as using N^6-benzyl-ATP would confirm that Hrr25 can directly phosphorylate Thr4.

Next, if the role of pThr4 is to recruit Rtt103 to the CTD at specific snoRNA genes, *rtt103Δ* cells should display similar readthrough defects as T4A. Therefore, comparative dynamic transcriptome analysis (cDTA) will be performed (Eser et al., 2014). In cDTA, 4-thiouracil is added to media, which is incorporated into nascent transcripts. MTSEA-biotin is conjugated to the thiols, allowing for isolation of nascent transcripts using streptavidin beads. An *S. pombe* spike in control is added to sequencing runs to normalize against global alterations in transcript abundance. This technique will also confirm that our observation of extended snoRNA

transcripts is not merely a result of stabilization of these transcripts, but rather a direct termination defect.

Finally, validation that pThr4-mediated recruitment of Rtt103 has a role in ejecting the Nrd1-Nab3-Sen1 complex from the CTD and onto RNA must be performed. Due to formaldehyde crosslinks, it remains unclear if Nrd1 retention downstream of snoRNAs is due to Nrd1 remaining bound to the CTD, or if the complex has been transferred to RNA but is unable to terminate transcription. Therefore, we propose to ChIP Nrd1 and Sen1 following treatment with RNase. If Rtt103 is important for ejecting the Nrd1-Nab3-Sen1 complex from the CTD to RNA, RNase should have no effect on ChIP signal.

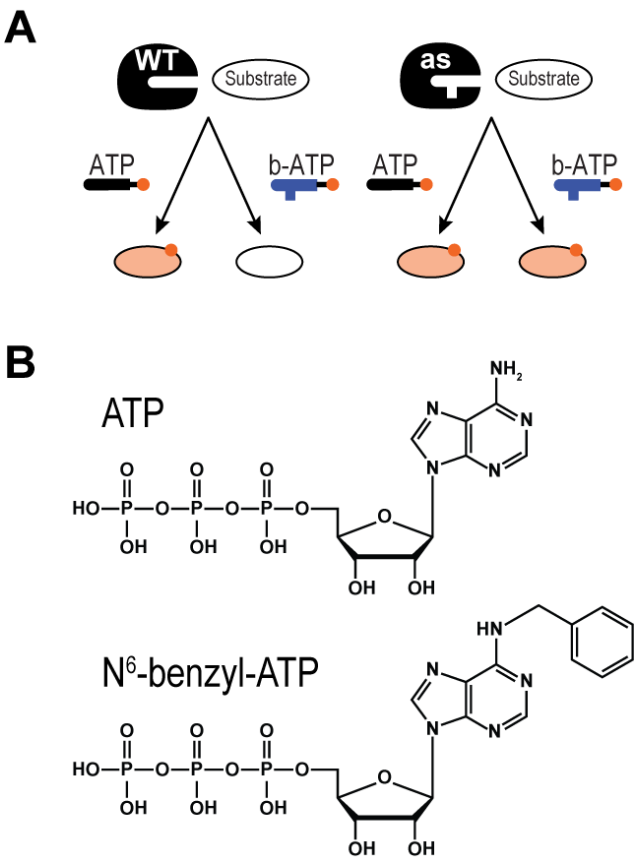

Figure 2.14 Direct phosphorylation utilizing N^6-benzyl-ATP

A) Schematic illustrating the ability of analog sensitive (-as) kinases to use the ATP analog N^6-benzyl-ATP (b-ATP) as a cofactor. WT kinases are unable to use this analog due to the lack of an expanded ATP binding pocket.

B) Structures of ATP and N^6-benzyl-ATP.

2.11 Experimental procedures

Phylogenetic tree

The primary structure of 131 *S. cerevisiae* kinases were aligned with ClustalW. An unrooted tree was developed using a maximum likelihood model with MEGA6, and the tree was plotted with Hypertree.

Kinase purification

Cells expression C-terminally TAP-tagged kinases (Ghaemmaghami et al., 2003) (GE-Healthcare) were grown in 200 mL selective media to an OD_{600} of 1.0, and cells were pelleted and transferred to a 1.5 mL tube. Pellets were resuspended in 500 µL TAP buffer A (20 mM HEPES pH 7.9, 300 mM potassium acetate, 0.5 mM EDTA pH 8.0, 10% glycerol, 0.05% NP40), with 1 mM DTT, 1 mM PMSF, 1 mM benzamidine, 1.45 mM pepstatin, and 1 mM phosphatase inhibitors (NaF, NaN3, and $Na3VO_4$) freshly added. Approximately 200 µL silica beads (OPS Diagnostics) were added, and cells were lysed via bead beating for 15 minutes. A 22 gauge needle was heated with a Bunsen burner and used to puncture a hole in the tube, which was then placed in a 2 mL tube and spun at 7,000 RPM for 30 seconds. 50 µL of a 50/50 slurry of IgG-Sepharose 6 Fast Flow beads (GE-Healthcare) in TAP buffer A was added to the supernatant. After 3 hours nutating at 4°C, beads were pelleted and washed five times with TAP buffer A (500 µL each). Kinases were eluted overnight at 4°C in 25 µL TAP buffer A with 1 mM DTT and 10 U AcTEV (Invitrogen).

GST-CTD purification

The CTD substrate, GST-CTD14 (fourteen repeats of YSPTSPS fused to GST) was expressed in BL21(DE3). Cells were grown to midlog phase at 37°C and induced with 1 mM IPTG overnight at 16°C. Cells were pelleted (5,000 *g* for 20 minutes) and resuspended in 5 mL Buffer B (1X PBS, 10% glycerol, 0.01% NP40) with freshly added protease inhibitors (1 mM PMSF, 1

mM benzamidine, 1.45 mM pepstatin). Cells were lysed with 3 rounds of sonication at power level 5 using a Heat Systems-Ultrasonics Inc. W-220 Sonicator. Cell debris was pelleted at 13,000 rpm for 20 minutes at 4°C and a 200 µL 50/50 slurry of Gutathione-Sepharose 4B beads (GE-Healthcare) in Buffer B was added to the supernatant. After 3 hours nutating at 4°C, beads were pelleted and washed once with Buffer B. Beads were resuspended in 1 mL FastAP buffer (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 100 mM KCl, 0.02% TritonX-100, 100 ug/ mL BSA) and incubated with 100 U FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) for 1 hour at 37°C. Beads were washed four times with Buffer B, then once with Buffer C (5 mM HEPES pH 7.4, 10 mM potassium glutamate, 10 mM magnesium acetate, 0.5 mM EDTA pH 8.0, 10% glycerol, 0.1% NP-40). Reduced glutathione (Fisher Scientific) was added to Buffer C to 50 mM, and the pH was adjusted to 7.0. GST-CTD14 was eluted overnight in 200 µL Buffer C with 50 mM reduced glutathione. Any remaining alkaline phosphatase was heat inactivated at 75°C for 5 minutes.

In vitro kinase assay (yeast)

Each assay was performed with 5 µL of tandem affinity purified (TAP) kinase and 1 µM GST-CTD14 in a 30 µL volume of Buffer D (Ansari et al., 2005) (20 mM HEPES pH 7.5, 10% glycerol, 2.5 mM EGTA, 15 mM magnesium acetate, 100 mM potassium acetate) with 1 mM DTT, 1 mM ATP, freshly added protease inhibitors (1 mM PMSF, 1 mM benzamidine, 1.45 mM pepstatin), and 1 mM phosphatase inhibitors (NaF, NaN3, and $Na3VO_4$). Reactions were performed at 30°C for two hours and resolved via SDS-PAGE. Western blot analysis was performed using standard procedures. Blots were probed with antibodies targeting pThr4 (1G7), or the TAP-tag (Thermo Scientific). Quantitation was performed using ImageJ.

In vitro kinase assay (human)

The CTD-peptide (YSPTSPS YSPTSPS YSPTSPS YSPTSPSC; Peptide Specialty Laboratories GmbH, Heidelberg) was coupled on 96-well maleimide plates for 60 min at 37°C in carbonate buffer at pH 9.5. The kinase assay was performed using 100ng recombinant kinase in 25 µL kinase buffer A (20 mM TrisHCl pH 7.4, 20 mM NaCl, 10 mM $MgCl_2$, 1 µM DTT and 2 µM ATP) at 28°C for 60 min followed by washing and blocking with PBS/milk (1%) for 30 min. Primary antibodies were added and incubated for 30 min. After an additional washing and blocking step, biotin-coupled secondary antibodies were added for 30 min. Following another washing and blocking step, peroxidase attached to avidin was added to the wells. After washing, 50 µL of substrate buffer (o-phenylenediamine and H_2O_2; pH 5.0) was added and after the colour change, samples OD were measured at 405 nm in the ELISA reader.

In vivo inhibition

100 mL cultures were grown to an OD_{600} of 0.3 at 30°C, split to two 50 mL cultures, and inhibitor was added to a final concentration of 20 µM (or an equivalent volume of DMSO). 3-MB-PP1 (*hrr25as* and *bur1as*) was purchased from Toronto Research Chemicals, 1-NA-PP1 (*kin28as*), 1-NM-PP1 (*cdc28as*, *pho85as*, and *ctk1as*), and CMK (*hrr25is)* were from MedChem Express. Cells were grown in the presence of inhibitor for one hour prior to pelleting (4,000 rpm, 5 minutes) and flash freezing. Pellets were resuspended in 200 µL 1X-NET-seq buffer (Churchman and Weissman, 2011) (20 mM HEPES pH 7.4, 110 mM KOAc, 0.5% Triton X-100, 0.1% Tween-20, 10 mM $MnCl_2$) with 1 mM PMSF, 1 mM benzamidine, 1.45 mM pepstatin, 1 mM NaN_3, and 1 mM NaF. Silica beads (OPS Diagnostics) were added, and cells were lysed via bead beating for 15 minutes. A 22 gauge needle was heated with a Bunsen burner and used to puncture a hole in the tube, which was then placed in a 2 mL tube and spun at 7,000 RPM for 10 seconds. Debris was resuspended, and 50U DNase I (Sigma) was added. Samples were mixed on a rotating, rocking mixer for 30 minutes at 4°C, and debris was pelleted. Supernatant

was processed as normal for western blot analysis. Blots were probed with α-pThr4 (1G7 1:1,000), α-Cdc28 (Santa Cruz), or α-Rpb3 (BioLegend, formerly Neoclone).

RNA-seq

Hrr25-as was grown to an optical density (OD_{600}) ~ 0.3 at 30°C in batch culture in YPD (1% yeast extract, 2% peptone and 2% dextrose) medium. Cells were treated with either DMSO or 20 μM 1-MB-PP1 (Toronto Research Chemicals) at final concentration for 1hr. Cells were harvested at 3000 rpm for 2 min followed by flash freezing in liquid nitrogen and maintained at −80°C until RNA extraction. Total RNA was extracted with hot phenol as previously described (Gasch, 2002). DNase-treated at 37°C for 30 min with TURBO DNase (Life Technologies), and then precipitated at −20°C in 2.5 M LiCl for 30 min. rRNA depletion of the DNase-treated total RNA was carried out with ScriptSeq Complete Kit H/M/R (Epicentre, Madison, WI), and samples were purified with a RNeasy MinElute Cleanup Kit (Qiagen). cDNA library preparation of rRNA-depleted RNA was carried out with Index PCR Primers (Epicentre) and FailSafe PCR Enzyme Mix (Epicentre) and purified with Axygen AxyPrep MAG PCR Cleanup beads (Corning). cDNA libraries were sequenced on Illumina's HiSeq 2500 System (UW-Madison DNA Sequencing Facility), generating single-end 100 bp reads. Sequencing reads were processed with Trimmomatic version 0.32 (Bolger et al., 2014) and reads were mapped to the S288c reference 64 using Bowtie2. HTseq version 0.6.1 (Anders et al., 2015) was used to sum read counts per gene, which were then normalized for gene size and the number of reads generated per library via reads per kilobase per million-mapped reads (RPKM). Biological duplicates were performed. Traces are from a single replicate.

Mutant CTD construct and plasmid shuffle

Mutant Rpb1 CTDs were constructed and tested for viability in vivo as previously described with some modifications (West and Corden, 1995). Briefly, the CTD repeats were constructed by

annealing and ligating 5'-phosphorylated oligonucleotides containing WT or mutant codons at position 4 (Table 2.3). WT and mutant Rpb1 CTDs were cut with AvaI and ligated into similarly cut pSB0, which was transformed into DH5α and screened. pSB0 vectors were cut with KpnI and SnaBI and ligated into a similarly cut pY1 vector, a modified pRS315 vector containing full length Rpb1. pY1 constructs were transformed into Z26, which contains a URA3 linked WT Rpb1 gene. Transformants were plated on synthetic complete (SC) media lacking uracil and leucine. Single colonies were grown overnight in YPD, and plated on SC –Leu +5-FOA (1 mg/ mL) (Toronto Research Chemicals) to counterselect against any colonies maintaining the WT Rpb1.

Microarray analysis

Cell collection, RNA preparation, cDNA synthesis and labeling, array hybridization, and normalization were as previously described (Berry and Gasch, 2008), using cyanine dyes (Flownamics) and Superscript III (Life Technologies). Samples were hybridized to whole-genome tiled DNA microarrays (Roche Nimblegen). Biological triplicates were averaged, and significant readthrough was quantified as the fold change (mutant/WT) in the region 200 bp downstream of the 3'-end of processed snoRNAs as described in (Tomson et al., 2013).

Northern blot

25 mL cultures were grown to OD 0.5, pelleted, resuspended in TE lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.5% SDS) and flash frozen. The following day, cells were thawed, and RNA was extracted with 700 µL hot phenol. 2 µL GlycoBlue was added to the aqueous layer, and RNA was extracted again. 700 µL chloroform was added to the aqueous layer and was added to a Heavy Phase Lock Gel Tube and spun for 5 munutes at room temp (16,000 g). RNA was precipitated from the aqueous layer with 10% 3 M NaOAc and 2.5 volumes cold 100% EtOH and frozen at -80°C overnight. The next day, RNA was pelleted, washed with 70% EtOH

and resuspended in 50 µL RNase-free water. 20 ug RNA was run on a 1.5% agarose gel in 1X MOPS, 6% formaldehyde. RNA was transferred to Hybond-N membrane overnight using 20x SSC buffer and was crosslinked to the membrane via UV irradiation. Primers were designed to SNR33 (Kim et al., 2006), SNR40, and SCR1, containing a T7 promoter for in vitro transcription of a radioactive probe (Table 2.3). ^{32}P-UTP was incorporated into probes via in vitro transcription (AmpliScribe T7-Flash transcription kit – Epicentre). G25 columns (GE Healthcare) were used to remove unincorporated UTP. Blots were incubated in ULTRAhyb solution (Ambion) at 68°C for 30 minutes prior to addition of the radioactive probe overnight. Blots were washed twice with 2x SSC, 0.1% SDS and twice with 0.1x SSC, 0.1% SDS. Phosphorimage screens were exposed overnight and scanned the next morning on a Typhoon FLA9000 (GE Healthcare).

Chromatin immunoprecipitation

ChIP was done as previously described with minor modifications (Tietjen et al., 2010). Cells were grown in rich media to OD_{600} ~0.5, crosslinked, pelleted, and flash frozen. Cells were lysed in lysis buffer (5 0 mM HEPES-KOH, 140 mM NaCl 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate), and 1 mM freshly added protease inhibitors (1 mM PMSF, 1 mM benzamidine, 1.45 mM pepstatin). Chromatin was sonicated on a Misonix 4000 sonicator, input was reserved, chromatin was immunoprecipitated overnight (3 µL α-Rpb3 (W0012 from Neoclone, currently BioLegend), 10 µL α-pThr4 (1G7), 25 µL IgG-Sepharose (GE-Healthcare), 5 µL α-FLAG (M2), 5 µL α-Nrd1 (gift from David Brow, (Steinmetz and Brow, 1998)). DNA was reverse crosslinked, amplified, and labeled using ligation-mediated PCR, and was hybridized to full yeast genome tiled microarrays (NimbleGen). Data were normalized as previously described (Tietjen et al., 2010). pSer2 traces were re-analyzed from (Tietjen et al., 2010). Hrr25 and Rtt103 were normalized to a mock TAP-ChIP in the untagged BY4741 strain. Ctk1 ChIP was analyzed from (Mayer et al., 2010), and Htz1 was analyzed from (Gu et al., 2015).

Co-immunoprecipitation and label-free quantitative proteomics

WT, T4A, T4E (unshuffled), and Z26 were grown in two liters of selective media. Cultures were grown to an OD_{600} ~1.0, pelleted, and filtered onto Whatman filter paper. Cells were scraped directly into liquid nitrogen, briefly shaken, and frozen at -80°C. Flash frozen cells were loaded into a 50 mL grinding jar with a grinding ball and immersed into liquid nitrogen. Grinding jars were loaded into a Retsch Mixer Mill MM 400, and cells were pulverized for six cycles (3' at 30 Hz), and grinding jars were re-immersed in liquid nitrogen between each round. Lysed yeast powder was dissolved in 10 mL 1X NET-seq buffer (Churchman and Weissman, 2011) (20 mM HEPES, pH 7.4, 110 mM KOAc, 0.5% Triton X-100, 0.1% Tween-20, 10 mM $MnCl_2$) with 2,000 U DNase (Sigma), and nutated at 4°C for 30 minutes. Debris was pelleted for 10 minutes at 4,000 rpm, and lysate was incubated with 200 µL anti-HA Dynabeads (Life Technologies). Following three washes with NET-seq buffer, samples were eluted with thrice with 100 µL HA peptide (2 mg/ mL in TBS) at room temp for 30 minutes. Fractions were combined, and 100 µL 100 mM Tris pH 8.5 and 100 µL cold TCA was added and nutated overnight. Protein was precipitated and washed twice with cold acetone. Triplicate samples were analyzed by label-free quantitative proteomics analysis as previously described (Zhang et al., 2010). Significant proteins were identified using QSpec (Choi et al., 2008). Proteins identified in all three replicates of the negative control (Z26) were discarded from analysis.

Validation of direct binding to pThr4

GST-CTD was expressed in bacteria, bound to glutathione beads (Amersham Biosciences), and subjected to dephosphorylation using alkaline phosphatase (FastAP, Thermo Fisher). CTD was phosphorylated at position 4 by Hrr25 as described above, and TAP-purified proteins were incubated for 1 hour at 23°C. Beads were washed, and proteins were eluted with Laemmli sample buffer prior to western blot analysis probing for the TAP-tag. GST-CTD was also eluted

from the beads and subjected to dot blot analysis probing for CTD modifications and the GST tag.

Endogenous Pol IIA (unphosphorylated CTD) from HeLa whole cell extract was immunoprecipitated with the antibody 1C7. The immunoprecipitated Pol IIA beads were used as a substrate for three kinases in an in vitro kinase assay as described above. 50 µL of TAP-purified protein was incubated with phosphorylated CTD substrate for 60 minutes at 30°C. Beads were washed six times with 1 mL lysis buffer, and proteins were eluted with Laemmli sample buffer prior to Western blot analysis using antibodies targeting CTD modifications, Pol II, or the TAP-tag.

Molecular Modeling

The multi-state NMR structure of Rtt103 (2L0I) was used in the modeling. The conformation of Arg108 in multiple states could make potential contacts with a pThr4 residue on the CTD. In a representative state, pSer2 was substitute with Ser, and Thr4 was substituted with pThr4 via manual replacement in PyMOL. Energy of the structure was minimized using PyMOL's sculpting tool (Schrödinger, 2010).

Table 2.2 Strains used in this study

Strain	Genotype	Plasmid	Source
TAP library	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0		GE Healthcare
hrr25as	his3Δ leu2Δ ura3Δ Hrr25Δ::kanMX4	pRS315-hrr25-I82G	E. Hurt Lab
pho85as	MATa pho85as1::HPH his3Δ1 leu2Δ0 ura3Δ0 met15Δ0		B. Andrews Lab
cdc28as	MATa cdc28as1::NAT his3Δ1 leu2Δ0 ura3Δ0 met15Δ0		B. Andrews Lab
bur1as	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 bur1 L149G		S. Hahn Lab
ctk1as	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 ctk1-F259G-HA3::kan	RPB1-XTPH LEU2	A. Greenleaf Lab
kin28as	MATα ade2Δ::hisG his3Δ200 leu2Δ0 lys2Δ0 met15Δ0 trp1Δ63 ura3Δ0 kin28 L83G		S. Hahn Lab
hrr25is	his3Δ leu2Δ ura3Δ Hrr25Δ::kanMX4	pRS315-hrr25-I82G-I23C	This study
Z26	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3	pRP112 (Rpb1 URA3)	R. Young Lab
consWT	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3	pY1 (Rpb1 CTD WT21 LEU2)	This study
S2A/WT	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3	pY1 (Rpb1 CTD S2A8-WT7 LEU2)	J. Corden Lab
T4A	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3	pY1 (Rpb1 CTD T4A19 LEU2)	This study
T4E	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3	pY1 (Rpb1 CTD T4E23 LEU2)	This study
WT/T4A	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3	pY1 (Rpb1 CTD WT11-T4A11 LEU2)	This study
T4A/WT	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3	pY1 (Rpb1 CTD T4A4-WT10 LEU2)	This study
consWT Rtt103-Flag	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 *RTT103*-Flagx5::KanMX6	pY1 (Rpb1 CTD WT21 LEU2)	This study
T4A Rtt103-Flag	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 *RTT103*-Flagx5::KanMX6	pY1 (Rpb1 CTD T4A19 LEU2)	This study
LL20 (HRR25 WT)	MATα leu2-3,112 his3-11,15		M. Stark Lab
ARB97 (*hrr25-3*)	MATα leu2-3,112 his3-11,15 hrr25-3/kti4-1		M. Stark Lab

Table 2.3 Oligonucleotides used in this study

Name	Function	Sequence (5'-3')
consWT-F	CTD construct	P-CCGAGCTATAGTCCAACTTCA
consWT-R	CTD construct	P-TCGGTGAAGTTGGACTATAGC
T4A-F	CTD construct	P-CCGAGCTATAGTCCAGCTTCA
T4A-R	CTD construct	P-TCGGTGAAGCTGGACTATAGC
T4E-F	CTD construct	P-CCGAGCTATAGTCCAGAATCA
T4E-R	CTD construct	P-TCGGTGATTCTGGACTATAGC
M3F	Pma1-1 qpCR	GACGACGAAGACAGTGATAACGAT
M3R	Pma1-1 qPCR	GGACCGACGAAAAACATAACGAAC
M5F	Pma1-2 qPCR	CAACTGATGGTCGTATTGTCACTG
M5R	Pma1-2 qPCR	CGAAAGTGTTGTCACCGGTAG
M9F	Pma1-3 qPCR	GGGCATAGTTTTAGCTATAGGTTC
M9R	Pma1-3 qPCR	CACCAGCCAATTGCCAGGAT
Y4F	Pyk1-1 qPCR	GGTAAGCCAGTTATCTGTGCTACC
Y4R	Pyk1-1 qPCR	GCGGTTTCAGCCATAGTGGTAAC
Y5F	Pyk1-2 qPCR	AGATGCCCAAGAGCTGCTAGATT
Y5R	Pyk1-2 qPCR	CCTTGGATGGAAACGTAAGTGTC
Y6F	Pyk1-3 qPCR	CGATGAGGTGTTGCATTTTTGGAA
Y6R	Pyk1-3 qPCR	GTACCCATGTATAACCTTCCAAGT
U1P3	SNR19-1 qPCR	GGGGCCTTTCAAAAGAGAGCT
U1RP3	SNR19-1 qPCR	ATCTTCAAACTACAATCCCGACCA
U1FP4	SNR19-2 qPCR	AGAGAGCCGCAACAAGAAACGTAA
U1RP4	SNR19-2 qPCR	AGCTCACATTCTCAACTACCGT

U2FR3	SNR20-1 qPCR	CCCCCAAGTATCGGCCAAAG
U2RP3	SNR20-1 qPCR	AAGAGCTCCTTCTCCTCAATGAG
U2FP4	SNR20-2 qPCR	CACGTACTCACACATGGCCGA
U2RP4	SNR20-2 qPCR	GGAGAGAACGAGAAAGCGGAT
SNR33-F	Northern Probe	CGGAACGGTACATAAGAATAGAAGAG
SNR33-R	Northern Probe	TAATACGACTCACTATAGGTAAAGAAAACGATAAGAACTAACC
SNR40-F	Northern Probe	AGTACCTTAACACATGACGAAGA
SNR40-R	Northern Probe	TAATACGACTCACTATAGGCTGATCTATTTCACGCCCAGA
SCR1-F	Northern Probe	AGGCTGTAATGGCTTTCTGGTGGGA
SCR1-R	Northern Probe	TAATACGACTCACTATAGGATATGTGCTATCCCGGCCGCCTCCA

2.12 References

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. Mol Cell *34*, 387-393.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Ansari, A.Z., Ogirala, A., and Ptashne, M. (2005). Transcriptional activating regions target attached substrates to a cyclin-dependent kinase. Proc Natl Acad Sci USA *102*, 2346-2349.

Aristizabal, M.J., Negri, G.L., Benschop, J.J., Holstege, F.C.P., Krogan, N.J., and Kobor, M.S. (2013). High-Throughput Genetic and Gene Expression Analysis of the RNAPII-CTD Reveals Unexpected Connections to SRB10/CDK8. PLoS Genet *9*, e1003758.

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

Berry, D.B., and Gasch, A.P. (2008). Stress-activated Genomic Expression Changes Serve a Preparative Role for Impending Stress in Yeast. Mol Biol Cell *19*, 4580-4587.

Bishop, A.C., Ubersax, J.A., Petsch, D.T., Matheos, D.P., Gray, N.S., Blethrow, J., Shimizu, E., Tsien, J.Z., Schultz, P.G., Rose, M.D*., et al.* (2000). A chemical switch for inhibitor-sensitive alleles of any protein kinase. Nature *407*, 395-401.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.Y., Breitkreutz, B.J., Stark, C., Liu, G*., et al.* (2010). A global protein kinase and phosphatase interaction network in yeast. Science *328*, 1043-1046.

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

Chasman, D., Ho, Y.H., Berry, D.B., Nemec, C.M., MacGilvray, M.E., Hose, J., Merrill, A.E., Lee, M.V., Will, J.L., Coon, J.J*., et al.* (2014). Pathway connectivity and signaling coordination in the yeast stress-activated signaling network. Mol Syst Biol *10*, 759.

Chinchilla, K., Rodriguez-Molina, J.B., Ursic, D., Finkel, J.S., Ansari, A.Z., and Culbertson, M.R. (2012). Interactions of Sen1, Nrd1, and Nab3 with Multiple Phosphorylated Forms of the Rpb1 C-Terminal Domain in Saccharomyces cerevisiae. Eukaryot Cell *11*, 417-429.

Choi, H., Fermin, D., and Nesvizhskii, A.I. (2008). Significance Analysis of Spectral Count Data in Label-free Shotgun Proteomics. Molecular & Cellular Proteomics *7*, 2373-2385.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 10.1038/nature09652.

Chymkowitch, P., Eldholm, V., Lorenz, S., Zimmermann, C., Lindvall, J.M., Bjoras, M., Meza-Zepeda, L.A., and Enserink, J.M. (2012). Cdc28 kinase activity regulates the basal transcription machinery at a subset of genes. Proc Natl Acad Sci U S A *109*, 10450-10455.

Corden, J.L., Cadena, D.L., Ahearn, J.M., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci USA *82*, 7934-7938.

Cunningham, B., and Wells, J. (1989). High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science *244*, 1081-1085.

Ding, X., Yu, Q., Zhang, B., Xu, N., Jia, C., Dong, Y., Chen, Y., Xing, L., and Li, M. (2014). The type II Ca2+/calmodulin-dependent protein kinases are involved in the regulation of cell wall integrity and oxidative stress response in Candida albicans. Biochem Biophys Res Commun *446*, 1073-1078.

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

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chemical reviews *113*, 8456-8490.

Eser, P., Demel, C., Maier, K.C., Schwalb, B., Pirkl, N., Martin, D.E., Cramer, P., and Tresch, A. (2014). Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression, Vol 10.

Fahrner, K., Yarger, J., and Hereford, L. (1980). Yeast histone mRNA is polyadenylated. Nucleic Acids Res *8*, 5725-5737.

Fuchs, S.M., Kizer, K.O., Braberg, H., Krogan, N.J., and Strahl, B.D. (2012). RNA Polymerase II Carboxyl-terminal Domain Phosphorylation Regulates Protein Stability of the Set2 Methyltransferase and Histone H3 Di- and Trimethylation at Lysine 36. The Journal of Biological Chemistry *287*, 3249-3256.

Gasch, A.P. (2002). Yeast genomic expression studies using DNA microarrays. In Methods Enzymol, G. Christine, and R.F. Gerald, eds. (Academic Press), pp. 393-414.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Ghalei, H., Schaub, F.X., Doherty, J.R., Noguchi, Y., Roush, W.R., Cleveland, J.L., Stroupe, M.E., and Karbstein, K. (2015). Hrr25/CK1δ-directed release of Ltv1 from pre-40S ribosomes is necessary for ribosome assembly and cell growth. The Journal of Cell Biology *208*, 745-759.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. Mol Cell Biol *29*, 5455-5464.

Grzechnik, P., Gdula, M.R., and Proudfoot, N.J. (2015). Pcf11 orchestrates transcription termination pathways in yeast. Genes Dev *29*, 849-861.

Gu, M., Naiyachit, Y., Wood, T., and Millar, C. (2015). H2A.Z marks antisense promoters and has positive effects on antisense transcript levels in budding yeast. BMC Genomics *16*, 99.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Rep *15*, 2147-2158.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E.*, et al.* (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J *31*, 2784-2797.

Hoekstra, M., Liskay, R., Ou, A., DeMaggio, A., Burbee, D., and Heffron, F. (1991). HRR25, a putative protein kinase from budding yeast: association with repair of damaged DNA. Science *253*, 1031-1034.

Hollingworth, D., Noble, C.G., Taylor, I., and Ramos, A. (2006). RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11. RNA *12*, 555-560.

Howard, S.C., Hester, A., and Herman, P.K. (2003). The Ras/PKA Signaling Pathway May Control RNA Polymerase II Elongation via the Spt4p/Spt5p Complex in Saccharomyces cerevisiae. Genetics *165*, 1059-1070.

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

Hsin, J.-P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD Phosphorylated on Threonine-4 Is Required for Histone mRNA 3′ End Processing. Science *334*, 683-686.

Hsin, J.P., Xiang, K., and Manley, J.L. (2014). Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing. Mol Cell Biol *34*, 2488-2498.

Kanin, E.I., Kipp, R.T., Kung, C., Slattery, M., Viale, A., Hahn, S., Shokat, K.M., and Ansari, A.Z. (2007). Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis. Proc Natl Acad Sci USA *104*, 5812-5817.

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

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517-522.

Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct Pathways for snoRNA and mRNA Termination. Mol Cell *24*, 723-734.

Lei, P., Ayton, S., Bush, A.I., and Adlard, P.A. (2011). GSK-3 in Neurodegenerative Diseases. International Journal of Alzheimer's Disease *2011*, 9.

Li, B., Pattenden, S.G., Lee, D., Gutiérrez, J., Chen, J., Seidel, C., Gerton, J., and Workman, J.L. (2005). Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. Proceedings of the National Academy of Sciences of the United States of America *102*, 18385-18390.

Liang, K., Gao, X., Gilmore, J.M., Florens, L., Washburn, M.P., Smith, E., and Shilatifard, A. (2015). Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. Mol Cell Biol *35*, 928-938.

Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S. (2004). Two Cyclin-Dependent Kinases Promote RNA Polymerase II Transcription and Formation of the Scaffold Complex. Mol Cell Biol *24*, 1721-1735.

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M., and Hahn, S. (2009). Phosphorylation of the Transcription Elongation Factor Spt5 by Yeast Bur1 Kinase Stimulates Recruitment of the PAF Complex. Mol Cell Biol *29*, 4852-4863.

Lopez, M.S., Kliegman, J.I., and Shokat, K.M. (2014). Chapter Eight - The Logic and Design of Analog-Sensitive Kinases and Their Small Molecule Inhibitors. In Methods Enzymol, M.S. Kevan, ed. (Academic Press), pp. 189-213.

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

Malumbres, M. (2014). Cyclin-dependent kinases. Genome Biology *15*, 122-122.

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

Measday, V., Moore, L., Retnakaran, R., Lee, J., Donoviel, M., Neiman, A.M., and Andrews, B. (1997). A family of cyclin-like proteins that interact with the Pho85 cyclin-dependent kinase. Mol Cell Biol *17*, 1212-1223.

Mitchell, A.P., and Bowdish, K.S. (1992). Selection for Early Meiotic Mutants in Yeast. Genetics *131*, 65-72.

Mosley, A.L., Hunter, G.O., Sardiu, M.E., Smolle, M., Workman, J.L., Florens, L., and Washburn, M.P. (2013). Quantitative Proteomics Demonstrates That the RNA Polymerase II Subunits Rpb4 and Rpb7 Dissociate during Transcriptional Elongation. Molecular & Cellular Proteomics : MCP *12*, 1530-1538.

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

Noble, C.G., Hollingworth, D., Martin, S.R., Ennis-Adeniran, V., Smerdon, S.J., Kelly, G., Taylor, I.A., and Ramos, A. (2005). Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nat Struct Mol Biol *12*, 144-151.

Perez, D.I., Gil, C., and Martinez, A. (2011). Protein kinases CK1 and CK2 as new targets for neurodegenerative diseases. Medicinal Research Reviews *31*, 924-954.

Petronczki, M., Matos, J., Mori, S., Gregan, J., Bogdanova, A., Schwickart, M., Mechtler, K., Shirahige, K., Zachariae, W., and Nasmyth, K. (2006). Monopolar Attachment of Sister Kinetochores at Meiosis I Requires Casein Kinase 1. Cell *126*, 1049-1064.

Pfaffenwimmer, T., Reiter, W., Brach, T., Nogellova, V., Papinski, D., Schuschnig, M., Abert, C., Ammerer, G., Martens, S., and Kraft, C. (2014). Hrr25 kinase promotes selective autophagy by phosphorylating the cargo receptor Atg19. EMBO reports *15*, 862-870.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev *20*, 2922-2936.

Phatnani, H.P., Jones, J.C., and Greenleaf, A.L. (2004). Expanding the Functional Repertoire of CTD Kinase I and RNA Polymerase II: Novel PhosphoCTD-Associating Proteins in the Yeast Proteome†. Biochemistry *43*, 15702-15719.

Pokholok, D.K., Zeitlinger, J., Hannett, N.M., Reynolds, D.B., and Young, R.A. (2006). Activated signal transduction kinases frequently occupy target genes. Science *313*, 533-536.

Posas, F., and Saito, H. (1998). Activation of the yeast SSK2 MAP kinase kinase kinase by the SSK1 two-component response regulator. EMBO J *17*, 1385-1394.

Posas, F., Wurgler-Murphy, S.M., Maeda, T., Witten, E.A., Thai, T.C., and Saito, H. (1996). Yeast HOG1 MAP Kinase Cascade Is Regulated by a Multistep Phosphorelay Mechanism in the SLN1–YPD1–SSK1 “Two-Component” Osmosensor. Cell *86*, 865-875.

Qiu, H., Hu, C., Gaur, N.A., and Hinnebusch, A.G. (2012). Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. EMBO J *31*, 3494-3505.

Ray, P., Basu, U., Ray, A., Majumdar, R., Deng, H., and Maitra, U. (2008). The Saccharomyces cerevisiae 60 S Ribosome Biogenesis Factor Tif6p Is Regulated by Hrr25p-mediated Phosphorylation. The Journal of Biological Chemistry *283*, 9681-9691.

Rodríguez-Molina, Juan B., Tseng, Sandra C., Simonett, Shane P., Taunton, J., and Ansari, Aseem Z. (2016). Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization. Mol Cell.

Rosonina, E., Yurko, N., Li, W., Hoque, M., Tian, B., and Manley, J.L. (2014). Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. Proc Natl Acad Sci U S A *111*, 11924-11931.

Sánchez-Piris, M., Posas, F., Alemany, V., Winge, I., Hidalgo, E., Bachs, O., and Aligue, R. (2002). The Serine/Threonine Kinase Cmk2 Is Required for Oxidative Stress Response in Fission Yeast. J Biol Chem *277*, 17722-17727.

Santisteban, M.S., Hang, M., and Smith, M.M. (2011). Histone Variant H2A.Z and RNA Polymerase II Transcription Elongation. Mol Cell Biol *31*, 1848-1860.

Schäfer, T., Maco, B., Petfalski, E., Tollervey, D., Böttcher, B., Aebi, U., and Hurt, E. (2006). Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. Nature *441*, 651-655.

Schneider, K., Smith, R., and O'Shea, E. (1994). Phosphate-regulated inactivation of the kinase PHO80-PHO85 by the CDK inhibitor PHO81. Science *266*, 122-126.

Schrödinger, L. (2010). The PyMOL Molecular Graphics System, Version 1.3.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.-M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol Cell *61*, 305-314.

Schwer, B., Bitton, D.A., Sanchez, A.M., Bahler, J., and Shuman, S. (2014). Individual letters of the RNA polymerase II CTD code govern distinct gene expression programs in fission yeast. Proc Natl Acad Sci U S A *111*, 4185-4190.

Sheldon, K.E., Mauger, D.M., and Arndt, K.M. (2005). A Requirement for the Saccharomyces cerevisiae Paf1 Complex in snoRNA 3′ End Formation. Mol Cell *20*, 225-236.

Steinmetz, E.J., and Brow, D.A. (1998). Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. Proceedings of the National Academy of Sciences of the United States of America *95*, 6699-6704.

Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Mol Cell *61*, 297-304.

Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S.*, et al.* (2010). Chemical-genomic dissection of the CTD code. Nat Struct Mol Biol *17*, 1154-1161.

Tomson, B.N., Crisucci, E.M., Heisler, L.E., Gebbia, M., Nislow, C., and Arndt, K.M. (2013). Effects of the Paf1 Complex and Histone Modifications on snoRNA 3′-End Formation Reveal Broad and Locus-Specific Regulation. Mol Cell Biol *33*, 170-182.

West, M.L., and Corden, J.L. (1995). Construction and Analysis of Yeast RNA Polymerase II CTD Deletion and Substitution Mutations. Genetics *140*, 1223-1233.

Yabuki, Y., Kodama, Y., Katayama, M., Sakamoto, A., Kanemaru, H., Wan, K., and Mizuta, K. (2014). Glycogen synthase kinase-3 is involved in regulation of ribosome biogenesis in yeast. Biosci, Biotechnol, Biochem *78*, 800-805.

Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodríguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. (2012a). Ssu72 Phosphatase-dependent Erasure of Phospho-Ser7 Marks on the RNA Polymerase II C-terminal Domain Is Essential for Viability and Transcription Termination. J Biol Chem *287*, 8541-8551.

Zhang, D.W., Rodriguez-Molina, J.B., Tietjen, J.R., Nemec, C.M., and Ansari, A.Z. (2012b). Emerging Views on the CTD Code. Genetics research international *2012*, 347214.

Zhang, Y., Wen, Z., Washburn, M.P., and Florens, L. (2010). Refinements to Label Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins. Analytical Chemistry *82*, 2272-2281.

Chapter 3: Diverse signals converge at the Pol II CTD to rapidly remodel the transcriptome

This chapter has been adapted from "Diverse signals converge at the Pol II CTD to rapidly remodel the transcriptome" by Corey M. Nemec, Yi-Hsuan Ho, Rajesh K. Kar, Audrey P. Gasch, and Aseem Z. Ansari.

3.1 Introduction

In response to diverse extracellular stimuli, yeast respond by activating the environmental stress response (ESR) network. This response is accomplished through a complex signaling cascade with many kinases playing key roles in signal transduction (Gasch and Werner-Washburne, 2002), yielding a concerted shift in gene expression profiles upon exposure to multiple external stressors (Figure 3.1) (Causton et al., 2001; Gasch, 2007; Gasch et al., 2000). A remarkable aspect of this response is that a diverse array of stresses is funneled by signaling cascades to rapidly remodel an overlapping set of ~900 genes (Gasch and Werner-Washburne, 2002; Gasch et al., 2000). In budding yeast, genes involved in rapid growth (e.g. ribosomal protein genes) are down regulated, while genes required for survival (e.g. cell wall synthesis) are upregulated (Gasch and Werner-Washburne, 2002).

The target of rapamycin complex 1 (TORC1) serves to regulate cell growth and suppress the stress response (Loewith and Hall, 2011). Under optimal growth conditions TORC1 phosphorylates Sch9 (Urban et al., 2007), which subsequently phosphorylates Stb3 and Dot6/Tod6, keeping them inactive (Huber et al., 2011). In response to stress, TORC1 activity is reduced, yielding dephosphorylated (active) Stb3 and Dot6/Tod6, which can bind to polymerase A and C (PAC) and/or RNA processing element (RRPE) elements in promoters of ribosome protein (RP) genes and ribosome biogenesis (RiBi) genes (Badis et al., 2008; Liko et al., 2007; Zhu et al., 2009). These factors recruit the histone deacetylase RPD3L complex, which remodels chromatin and represses transcription of RP and RiBi genes (Alejandro-Osorio et al., 2009; Huber et al., 2011; Lippman and Broach, 2009). Further, pre-mRNA splicing is inhibited when cells encounter stress (Bergkessel et al., 2011; Pleiss et al., 2007). In response to stress, many RP genes remain unspliced and are degraded by the nonsense mediated decay machinery (Garre et al., 2013). Thus, transcriptional repression and active transcript destruction function together to reduce RP transcript abundance.

Conversely, in the absence of stress, the transcription factor Msn2 is localized to the cytoplasm through PKA and TORC1 activity (Beck and Hall, 1999; Görner et al., 1998; Smith et al., 1998). In response to stress, Msn2 and other transcription factors are imported into the nucleus, in part through Hog1 activity (Alepuz et al., 2003; Alepuz et al., 2001; Beck and Hall, 1999; Görner et al., 1998; Rep et al., 1999). These factors bind to the stress-response element (STRE) in promoters of genes induced during ESR (iESR) to modulate their expression. (Amorós and Estruch, 2001; Gasch et al., 2000; Martínez-Pastor et al., 1996; Schmitt and McEntee, 1996).

Thus, current understanding of the ESR suggests that transcription factors and chromatin remodelers are activated in response to stressors, which differentially recruit RNA polymerase II (Pol II) to various genes required to protect cells during stress. The role of Pol II itself in ESR is thought to be indirect. However, Pol II contains a unique carboxyl-terminal domain (CTD) consisting of 26 repeating heptapeptides ($Y_1S_2P_3T_4S_5P_6S_7$) in yeast (Corden et al., 1985; Zhang and Corden, 1991). The “CTD code” hypothesizes that different patterns of post-translational modifications on the CTD encipher a code that is read by different protein complexes involved in Pol II dependent RNA biogenesis (Buratowski, 2003). We hypothesized that ESR-activated kinases might directly target the Pol II CTD giving rise to a stress-specific phosphorylation pattern, offering an elegant mechanism of direct signal-responsive control of transcription.

Here, we show that kinases implicated in ESR can directly phosphorylate the CTD in response to osmostress. Further, phosphorylation of the CTD is critical for cells to properly respond to osmostress. We find that phosphorylation of specific CTD residues controls expression of specific classes of ESR genes, i.e. iESR, RP, and RiBi genes. Specifically, pThr4 is required for repression of RP gene transcripts through a mechanism that may involve defective splicing of RP pre-mRNA. We are also investigating the possibility that CTD phosphorylation may rapidly repress transcription by recruiting termination factors to the CTD.

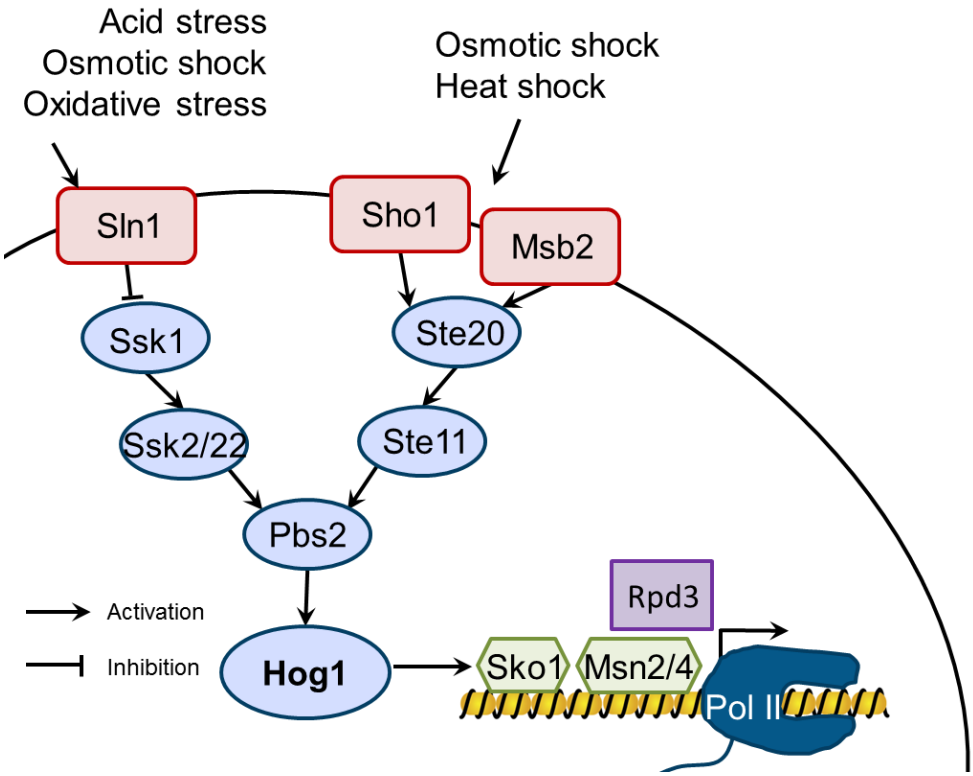

Figure 3.1 Stress response signaling cascade
Membrane-bound proteins (red) sense multiple different stressors and activate a signaling cascade of kinase phosphorylation (blue) centering on the Hog1 kinase. Phosphorylation of Hog1 activates the kinase and localizes is to the nucleus. Activated Hog1 phosphorylates transcription factors (green) and chromatin remodelers (purple), which partially dictate the subset of genes that are activated or repressed in response to stress.

3.2 Stress-responsive kinases phosphorylate the Pol II CTD

We previously identified eleven kinases capable of phosphorylating Thr4 on the RNA Pol II CTD in *S. cerevisiae* (Nemec et al., in preparation). To explore potential functional connectivity between our Thr4 kinases, we performed gene ontology analysis. GO Slim Mapper was used to identify high level, parent GO terms. One of the most abundant GO terms common to the Thr4 kinases was “Response to stress,” which suggested that Thr4 phosphorylation may play a role in the environmental stress response (ESR) (Table 3.1). The ability of ESR kinases to directly phosphorylate the CTD has not been studied, possibly because the transcriptional response to stress was thought to be indirectly regulated by Msn2/4 transcription factors and the chromatin remodeler, Rpd3 (Alejandro-Osorio et al., 2009; Gasch et al., 2000).

To examine the relationship further, we mapped the kinases onto an inferential network of ESR regulators by integrating thousands of calibrated measurements of stress-responsive transcriptome perturbations with the publicly accessible protein–protein interactome information compiled from numerous different platforms (mass spectrometry, synthetic genetic analysis, two-hybrid assays, as well as carefully curated literature reports of traditional genetic and biochemical studies) (Figure 3.2A) (Chasman et al., 2014). We observed a strong connection between Thr4 kinases, Rpb1, stress-activated transcription factors, and the rest of the ESR network.

To determine if these and other known stress-responsive kinases were able to phosphorylate the CTD, kinases were affinity purified as described in Chapter 2, and activity was determined via immunoblot targeting phosphorylated CTD residues. Cdc28 strongly phosphorylated both Ser2 and Ser5 (Figure 3.2B, 3.3). This result confirms a previous report suggesting that Cdc28-mediated phosphorylation of Ser5 is important for recruiting the mRNA capping enzyme to the CTD and for regulating basal transcription (Chymkowitch et al., 2012). The function of Cdc28-

mediated pSer2 remains to be determined. Surprisingly, Sch9 also phosphorylated Ser2 and Ser5 in the absence of osmostress (Figure 3.2B, 3.3). In addition to responding to stress (Pascual-Ahuir and Proft, 2007), Sch9 is involved in regulating ribosome biogenesis (Huber et al., 2011), and nutrient signaling (Smets et al., 2010). Finally, we confirmed that Bur1 can robustly phosphorylate Ser2, Thr4, Ser5, and Ser7 (Glover-Cutter et al., 2009; Hsin et al., 2011; Hsin et al., 2014; Keogh et al., 2003; Murray et al., 2001; Nemec et al., in preparation; Tietjen et al., 2010) (Figure 3.2B, 3.3).

Next, we wondered if stress-activated kinases would phosphorylate the CTD to a larger extent. Yeast expressing TAP-tagged kinases were stressed with 0.7 M NaCl prior to affinity purification, and the activity of stress-activated kinases was compared to an unstressed control. Upon stress-activation, Cdc28 and Sch9 maintained the ability to phosphorylate Ser2 and Ser5, but activity was not increased (Figure 3.2C, 3.3). Conversely, the activity of Hog1, Cmk1, and Tpk1 on the CTD was increased in response to NaCl. To identify significant changes in kinase activity post-osmostress, fold change in kinase activity was compared to fold change in Bur1 activity, a kinase not part of the ESR which is capable of phosphorylating Ser2, Thr4, Ser5, and Ser7. Hog1 activation strongly increased its ability to phosphorylate Ser2, Thr4, Ser5 and Ser7, whereas stress-induced activation of Cmk1 and Tpk1 was only significant for specific CTD residues (T-test, $p<0.05$, $n=3$) (Figure 3.2C, 3.3). Cmk1 is known to play a role in cell wall integrity and redox regulation (Ding et al., 2014), and Tpk1 associates with chromatin in a stress-dependent manner (Baccarini et al., 2015; Pokholok et al., 2006).

Table 3.1 Generic GO slim processes associated with Thr4 kinases

GO-Slim term	Cluster frequency	Genome frequency	*p*-value	Genes annotated to the term
cellular protein modification process	11 out of 11 genes, 100%	660 of 6338 genes, 10.4%	0.0001	CDC28, CMK1, CMK2, HRR25, PHO85, RIM11, SGV1, SLN1, SSK2, YCK2, YCK3
response to stress	6 out of 11 genes, 54.5%	661 of 6338 genes, 10.4%	0.0003	CDC28, HRR25, PHO85, RIM11, SLN1, SSK2
signal transduction	5 out of 11 genes, 45.5%	235 of 6338 genes, 3.7%	0.0001	CMK1, CMK2, PHO85, SLN1, SSK2
cell cycle	4 out of 11 genes, 36.4%	640 of 6338 genes, 10.1%	0.0163	CDC28, HRR25, PHO85, RIM11
cellular nitrogen compound metabolic process	4 out of 11 genes, 36.4%	2478 of 6338 genes, 39.1%	1.0000	CDC28, HRR25, PHO85, SGV1
transport	4 out of 11 genes, 36.4%	1140 of 6338 genes, 18%	0.1043	CDC28, HRR25, YCK2, YCK3
vesicle-mediated transport	4 out of 11 genes, 36.4%	350 of 6338 genes, 5.5%	0.0019	CDC28, HRR25, YCK2, YCK3
biosynthetic process	3 out of 11 genes, 27.3%	2151 of 6338 genes, 33.9%	1.0000	CDC28, PHO85, SGV1
transcription, DNA-templated	3 out of 11 genes, 27.3%	673 of 6338 genes, 10.6%	0.0924	CDC28, PHO85, SGV1
reproduction	3 out of 11 genes, 27.3%	463 of 6338 genes, 7.3%	0.0365	CDC28, HRR25, RIM11
cytoskeleton organization	2 out of 11 genes, 18.2%	230 of 6338 genes, 3.6%	0.0535	CDC28, SSK2
chromosome organization	2 out of 11 genes, 18.2%	549 of 6338 genes, 8.7%	0.2293	CDC28, SGV1
chromosome segregation	2 out of 11 genes, 18.2%	197 of 6338 genes, 3.1%	0.0404	CDC28, HRR25
anatomical structure development	2 out of 11 genes, 18.2%	160 of 6338 genes, 2.5%	0.0275	RIM11, YCK2
DNA metabolic process	2 out of 11 genes, 18.2%	445 of 6338 genes, 7.0%	0.1653	CDC28, HRR25

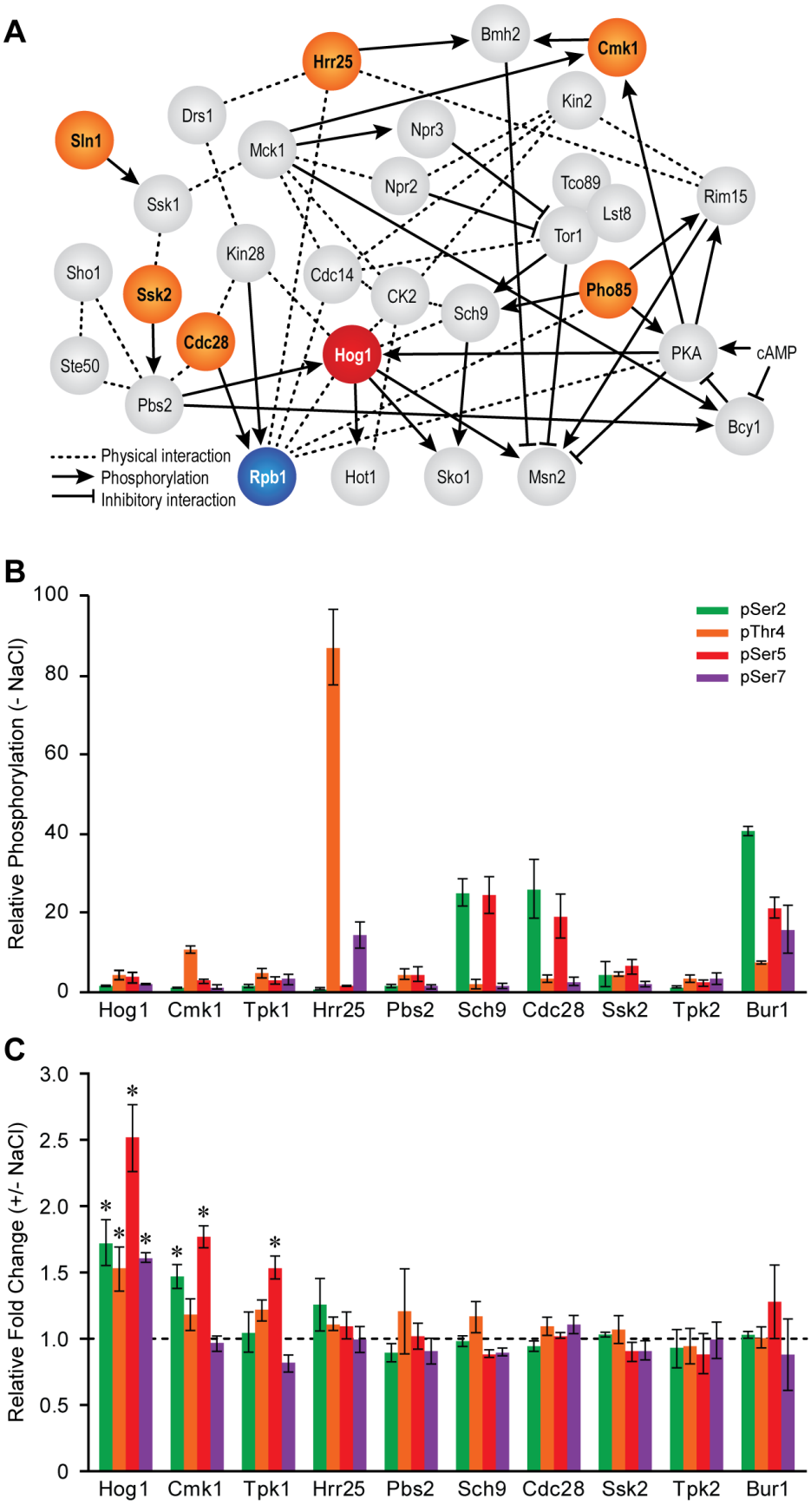
A
Bmh2
Hrr25
Cmk1
Kin2
Drs1
Sln1
Mck1
Npr3
Tco89
Npr2
Rim15
Ssk1
Lst8
Tor1
Kin28
Sho1
Cdc14
Pho85
Ssk2
CK2
Sch9
Cdc28
Hog1
PKA
cAMP
Ste50
Pbs2
Bcy1
Physical interaction
Phosphorylation
Inhibitory interaction
Rpb1
Hot1
Sko1
Msn2
B
Relative Phosphorylation (- NaCl)
pSer2
pThr4
pSer5
pSer7
Hog1
Cmk1
Tpk1
Hrr25
Pbs2
Sch9
Cdc28
Ssk2
Tpk2
Bur1
C
Relative Fold Change (+/- NaCl)

Figure 3.2 Stress responsive kinases phosphorylate the CTD

A) Inferred subnetwork of the ESR (Chasman et al., 2014) illustrating the inclusion of six Thr4 kinases (Nemec et al., in preparation) (orange). The network centers on Hog1 (red), and Rpb1 (blue) is a major node.

B) Fold change in kinase activity with respect to mock. pSer2 (Bethyl), pThr4 (1G7), pSer5 (3E8), and pSer7 (4E12) were probed. Error bars illustrate SEM of three independent experiments.

C) Relative fold change in phosphorylation after stress-activation of kinases. Fold change in kinase activity was compared to the fold change in activity of Bur1, a kinase capable of phosphorylating these four residues, but which is not implicated in the stress response. Error bars indicate SEM. Asterisks indicate a significant increase in phosphorylation after stress ($p<0.05$, $n=3$).

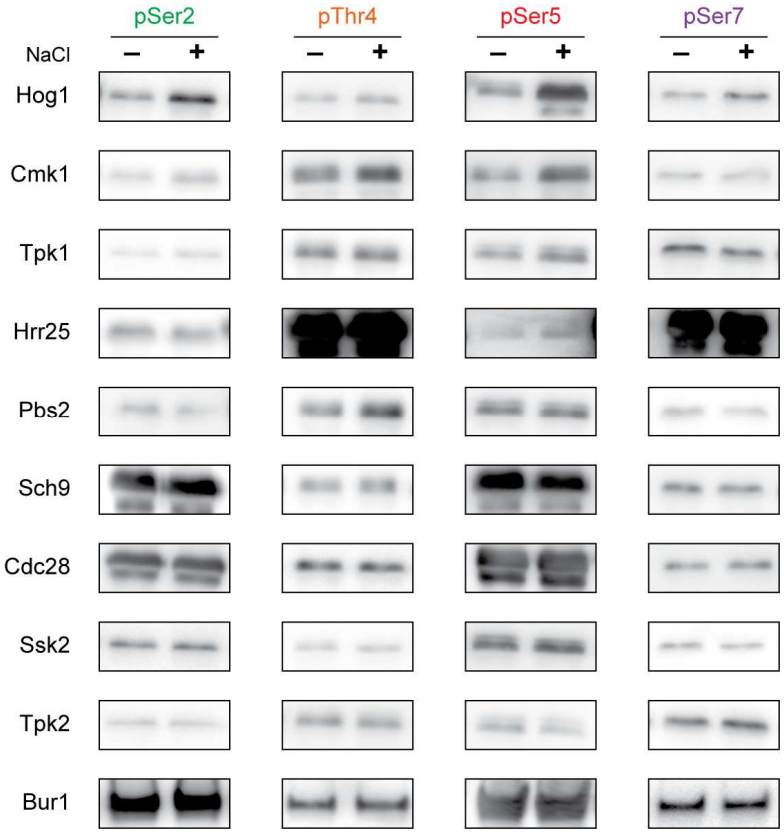

Figure 3.3 Representative stress-responsive kinase activity
Strains expressing the indicated TAP-tagged kinases were grown in YPD to OD 0.5, split, and either stressed with 0.7M NaCl (final) or YPD. Kinases were purified, and activity on the indicated CTD residues were assayed.

3.3 CTD phosphorylation is critical for efficient response to osmostress

To determine the role of each phospho-CTD mark in response to stress, we generated mutant variants of Rpb1 in which Ser2, Thr4, Ser5, or Ser7 were substituted with a non-phosphorylatable alanine. Thr4 and Ser7 were substituted with alanine across all CTD repeats (T4A and S7A) (Nemec et al., in preparation; Zhang et al., 2012), but since substitution of Ser2 or Ser5 with alanine on each repeat is lethal, only repeats in the first half of the CTD bear alanine substitutions (S2A/WT and S5A/WT) (West and Corden, 1995).

Cultures of WT and each CTD mutant strain were stressed with NaCl (0.7M final), and total RNA expression was analyzed. The change in gene expression in response to stress was quantified and plotted in a heatmap (Figure 3.4A, left). Typically, a subset of the transcriptome is induced (iESR genes, red) and an alternate subset is repressed (rESR genes, green) in response to osmostress (Gasch et al., 2000). While WT cells responded to osmostress as expected, CTD mutant strains displayed significant defects (>1.5 fold differential expression) in induction and repression (Figure 3.4A, middle).

To confirm that these differences were due to stress and were not affected under normal growth conditions, the expression of each gene in the CTD mutants were compared to WT in the absence of stress (Figure 3.4A, right). Fewer than 25% of differentially expressed genes were also basally-affected (>1.5 fold) in T4A, S5A/WT, and S7A, confirming that the majority of differential expression was osmostress-dependent. However, in 2A/WT nearly 60% of differentially expressed genes were basally affected, suggesting that these response differences may not be osmostress-dependent. Indeed, nearly one-third of basally-affected genes in S2A/WT were tRNA transcripts, which artificially inflated the induction defect of this gene class. Further, while a previous report observed an upregulation of phosphate-regulated genes in a T4V strain (Rosonina et al., 2014), such a defect was not observed here with the T4A strain.

To compare the expression profiles among CTD mutants, we created a clustergram of pairwise Pearson correlations between each strain in response to NaCl stress (Figure 3.4B). The expression profile of S2A/WT closely clustered with WT, suggesting that the pSer2 may not be as critical for response to stress as phosphorylation of other CTD residues. T4A and S7A expression profiles clustered together, which mirrors the pattern of response differences (Figure 3.4A, middle). Conversely, 5A/WT remained in its own cluster, which was evident based on the intensity and identity of differentially expressed genes in response to NaCl stress (Figure 3.4A, middle). In response to osmostress, the expression differences (mutant vs. WT) were further analyzed via kernel density plot (Figure 3.4C). The lower and broader S5A/WT peak confirmed that more genes were differentially expressed with respect to WT in response to osmostress than the other CTD mutant strains.

Because Hog1 activity on the CTD increased in response to osmostress (Figure 3.2C), we compared the differential expression profiles between CTD mutants and a *hog1Δ* strain (Figure 3.4C) (Chasman et al., 2014). While a substantial overlap in significantly differentially expressed genes was observed between *hog1Δ*, T4A, and S5A/WT, only minor overlap was observed between *hog1Δ*, S2A/WT, and S7A. Further, over half of differentially expressed genes in *hog1Δ* did not overlap with CTD mutants. This is consistent with the role of Hog1 in phosphorylating other targets in response to osmostress (Alepuz et al., 2003; Alepuz et al., 2001; Beck and Hall, 1999; Görner et al., 1998; Rep et al., 1999). GO annotations of genes uniquely differentially expressed in *hog1Δ* included genes involved in oxidoreducatase activity, a class of genes that is upregulated in response to multiple stresses (Figure 3.4B) (Gasch et al., 2000). Interestingly, differential expression of many RP genes were unique to T4A, and differential expression of RiBi genes were unique to S5A/WT (Figure 3.4D), suggesting that specific phospho-CTD marks may play roles in regulating different classes of ESR genes.

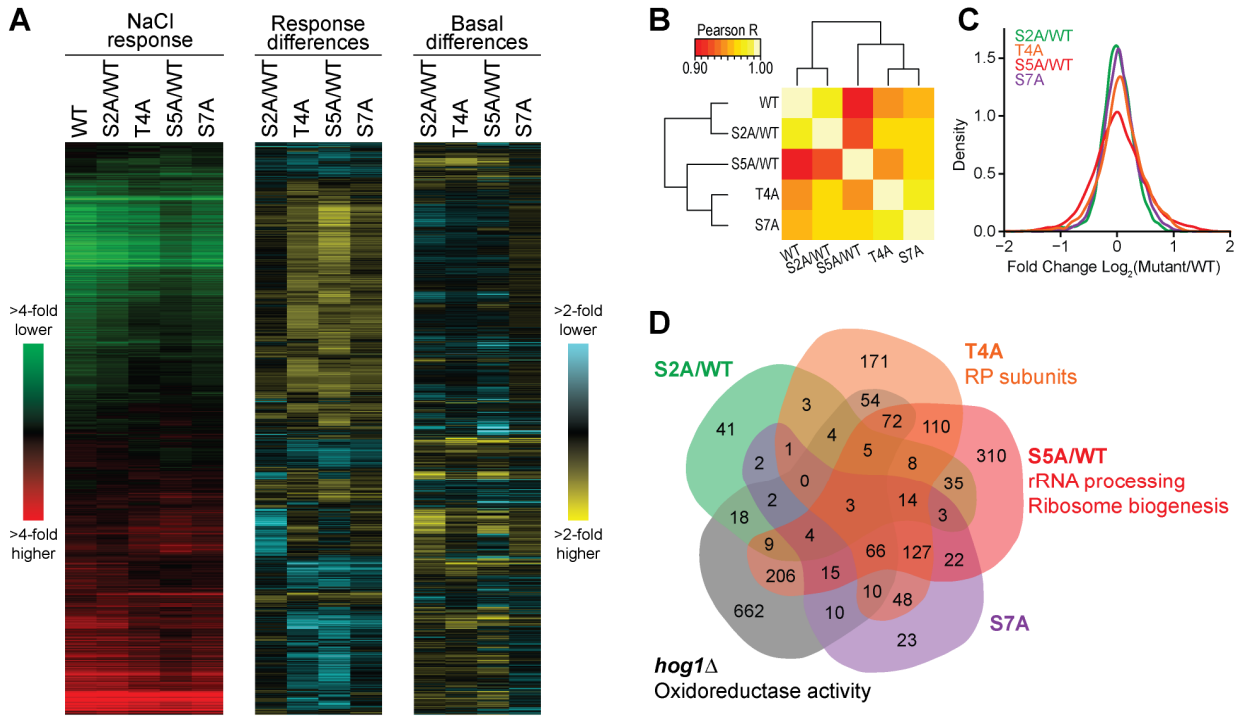

Figure 3.4 CTD phosphorylation is critical for efficient response to osmostress

A) Heatmap of RNA expression changes in WT or CTD mutant strains. The x-axis compares strains' response to NaCl. Induction (red), repression (green). Genes listed on the y-axis are differentially expressed in response to NaCl stress in at least one mutant strain. Fold change between gene expression in mutant strains vs. WT is illustrated with either yellow (increase in expression), or blue (decrease in expression) (middle). Basal differences of mutant strains (expression in mutant vs. WT in the absence of stress) is also shown (right).

B) Clustergrams of Pearson correlations comparing the NaCl response pairwise among all genes in each CTD mutant strain.

C) Kernel density plot of response differences of all genes (CTD mutant vs. WT).

D) Venn diagram comparing the number of genes with significant response differences among CTD mutant strains and *hog1Δ*. GO annotations of genes differentially expressed in only one mutant strain are listed (Funspec, $p<0.0001$).

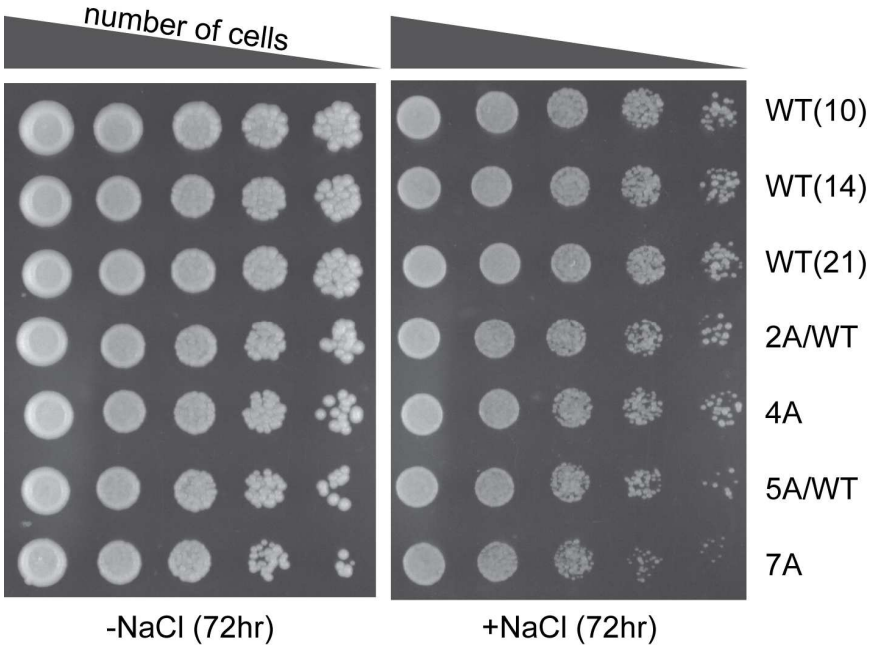

Figure 3.5 Long-term response to stress in CTD mutants
Strains bearing 10, 14, or 21 repeats of WT CTD, as well as S2A/WT, T4A, S5A/WT, and S7A variants were spotted in serial dilution on YPD plates either with or without 0.7M. Plates were imaged after 72 hours.

While the CTD-dependent defects in response to stress were observed within minutes, we wondered how mutant strains would tolerate long term stress. When grown on plates lacking NaCl, all CTD mutant strains displayed minor growth defects with respect to strains bearing WT CTD repeats of various lengths (Figure 3.5, left). Interestingly, not only were all CTD mutant strains able to survive on plates containing 0.7M NaCl, but the reduced growth due to NaCl was no greater than the reduced growth of the WT strains (Figure 3.5, right). While the early response to stress may not be as pronounced in CTD mutant strains, the diminished response is still sufficient to protect cells from death. This suggests that the CTD-dependent response occurs rapidly, whereas long-term adaptation to stress is not mediated by the CTD.

3.4 Residue-specific CTD phosphorylation controls expression of different gene classes

To identify if the differentially expressed genes were similar among each mutant, Venn diagrams illustrating the number of overlapping differentially expressed genes between each mutant strain were created for each class of ESR gene (iESR, RP, and RiBi) (Figure 3.6A). The differentially expressed iESR genes showed only minor overlap between CTD mutants, whereas the differentially expressed RP genes and RiBi genes showed remarkable overlap. Further, among the total 116 differentially expressed RP genes, 113 were differentially expressed in T4A, suggesting a larger role of pThr4 in RP gene repression. Conversely, among the total 246 differentially expressed RiBi genes, 206 were differentially expressed in S5A/WT, suggesting a larger role of pSer5 in RiBi gene repression. These results are consistent with our analysis in Figure 3.4D.

Violin plots illustrating the magnitude of differential expression confirmed that not only were more RP genes differentially expressed in T4A, but the magnitude of the repression defect was largest. Similarly, the large number of RiBi genes with repression defects in S5A/WT mirrored the magnitude of the defects (Figure 3.6B).

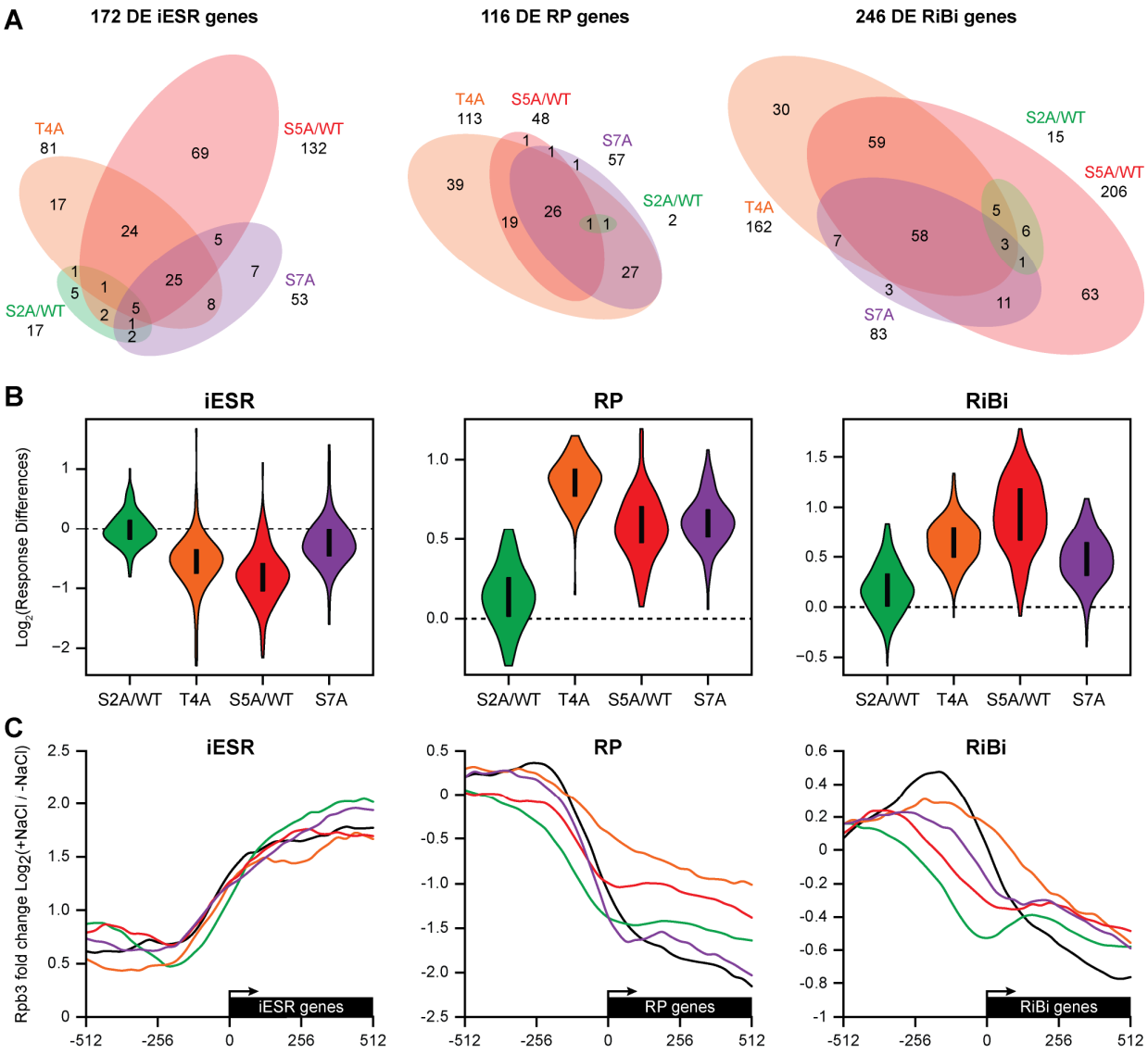

Figure 3.6 Residue-specific CTD phosphorylation controls expression of different gene classes

A) Venn diagrams showing overlap of the number of significantly differentially expressed genes (mutant vs. WT) in response to NaCl in three different ESR gene classes: induced ESR genes (iESR), ribosome protein genes (RP), and ribosome biogenesis genes (RiBi).

B) Violin plots of the magnitude of differential expression (mutant vs.WT) in response to NaCl at iESR, RP, and RiBi genes. Dotted line indicates no expression difference.

C) Average change in Pol II occupancy in response to NaCl in a 1024 bp window centered at the transcription start site of iESR, RP,or RiBi genes.

Since RNA abundance and Pol II occupancy are tightly correlated (Mayer et al., 2010), and ChIP signal is proportional to transcription rate (Miller et al., 2011), we performed chromatin immunoprecipitation of the Rpb3 subunit of Pol II in WT and each of our mutant strains in response to NaCl. The average fold change in Pol II occupancy in response to stress was plotted at the transcription start site of iESR genes, RP genes, or RiBi genes (Figure 3.6C). As expected, WT cells displayed a strong increase in Pol II across the ORF of iESR genes in response to osmostress. Surprisingly, the CTD mutant strains showed a similar increase in Pol II at iESR genes (Figure 3.6C, left). This suggests that the transcripts may be properly produced, but are less stable and suffer from degradation. Within 10-20 minutes post-osmostress, not only are iESR genes rapidly transcribed, but they are also rapidly degraded (Miller et al., 2011). This rapid degradation is thought to be important for rapid removal of iESR transcripts after stress, but the mechanism is poorly understood. Recent work from our lab suggests that the phosphorylation state of the CTD may be important for mRNA stability (Rodríguez-Molina et al., 2016). It remains to be determined if CTD phosphorylation during ESR is involved in iESR transcript stability.

Next, we examined Pol II occupancy across RP genes. While WT cells showed a strong drop in Pol II levels across RP genes, T4A showed only minor reduction in Pol II occupancy (Figure 3.6C, middle). The lack of reduction of Pol II across RP genes suggests that Pol II isn't disengaging in these mutants and potentially explains the strong repression defect observed, in terms of both number of affected genes, and the magnitude of the defect (Figure 3.6A,B, middle). This observation is consistent with our previous work suggesting a role for pThr4 in termination of both noncoding and protein-coding genes (Nemec et al., in preparation). Stronger reduction of Pol II occupancy in S2A/WT, S5A/WT, and S7A mirror the lesser magnitude and number of affected RP genes in these strains.

Finally, we observed an enrichment of Pol II in the promoter region of RiBi genes in the WT strain, consistent with previous reports (Figure 3.6C, right) (Chasman et al., 2014). We hypothesize that Pol II is poised upstream of RiBi genes waiting to be activated upon exiting ESR. We observed a similar, though less defined enrichment in the T4A strain, but no such peak was present in S2A/WT, S5A/WT, or 7A. In fact, a strong reduction in Pol II occupancy in RiBi promoters was observed in these strains.

3.5 Splicing of RP genes is de-repressed in T4A in response to osmostress

Ribosomal protein (RP) encoding genes are significantly enriched with introns (73%) compared to the genome average (5%) (p<0.0001 Fisher's Exact Test) (Garre et al., 2013). As part of a concerted response to repress gene expression, pre-mRNA splicing of RP genes is altered when cells encounter specific stresses (Bergkessel et al., 2011; Garre et al., 2013; Pleiss et al., 2007). Depending on the type and severity of the stressor, RP pre-mRNA transcripts bearing introns can either be stabilized, or rapidly degraded by the exosome or nonsense mediated decay (NMD) machinery (Bergkessel et al., 2011; Garre et al., 2013; Pleiss et al., 2007). Because phosphorylated Pol II CTD is required for splicing (McCracken et al., 1997) and plays a critical role in recruiting essential splicing factors (David et al., 2011), we wondered if stress-dependent CTD phosphorylation plays a role in RP gene regulation in response to osmostress.

To test if CTD phosphorylation mediates the repression of RP genes through enhancement of splicing, we examined the fold change in intron retention of all intron-containing RP genes in response to osmostress using full-genome tiled microarrays (Figure 3.7A). In WT, we observed splicing defects (increased intron retention) of many RP transcripts in response to osmostress. In S2A/WT, S5A/WT, and S7A, we observed an equal or greater osmostress-dependent increase in intron retention. However, in T4A, we observed an increase in splice efficiency (decrease in intron retention) of a subset of RP genes.

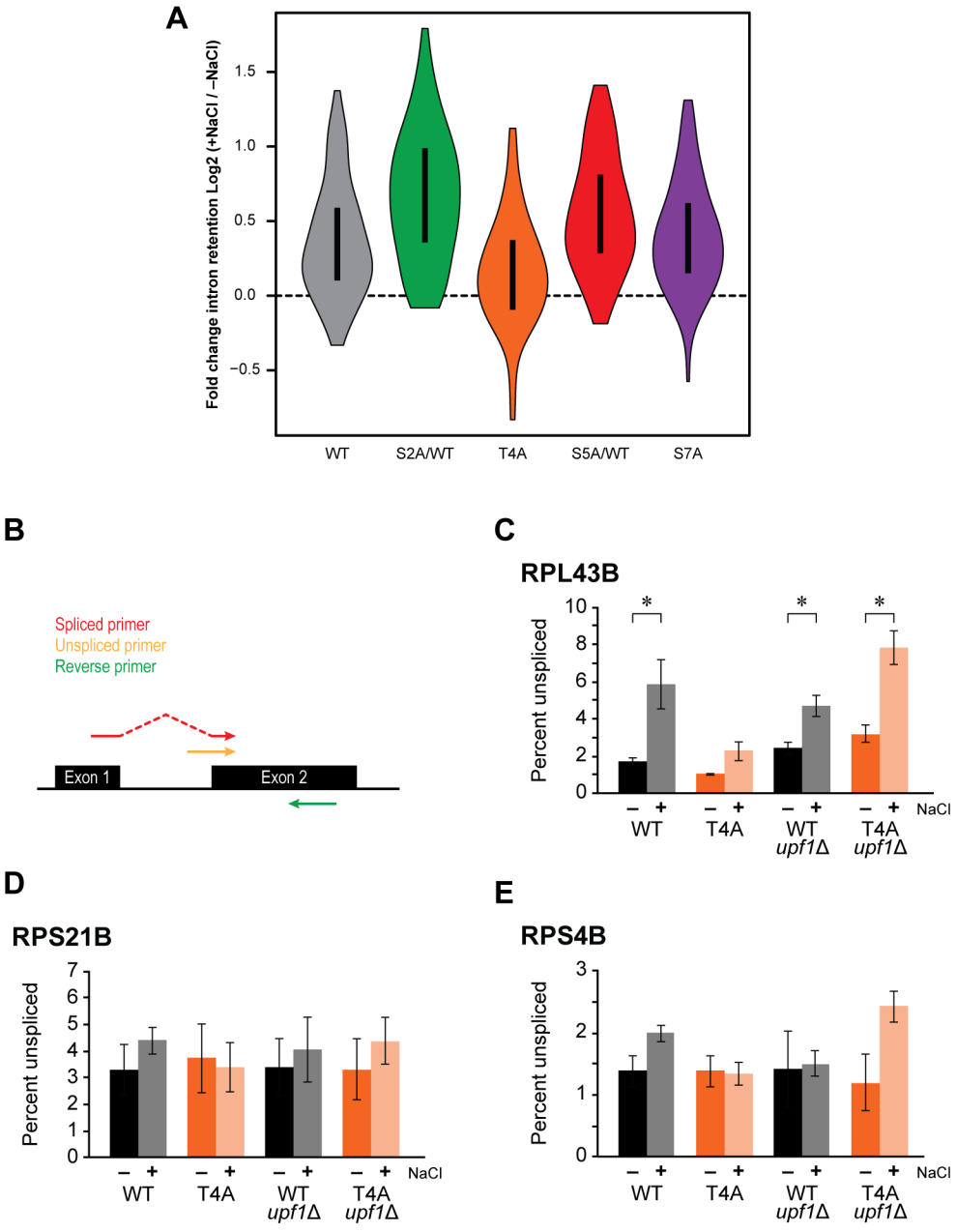

Figure 3.7 Ribosome protein transcripts remain spliced in response to stress

A) Violin plots illustrating the fold change in intron retention of all intron-containing RP genes in response to osmostress.

B) Schematic of spliceform-specific primers used to quantify unspliced transcripts.

C) Histograms of unspliced RPL43B transcripts in WT (black) or T4A (orange strains) in the presence (light) or absence of NaCl treatment (dark). Error bars indicate SEM of three independent experiments. Asterisks indicate $p<0.05$ (T-test, $n=3$).

D) Same as C) at RPS21B.

E) Same as C) at RPS4B.

To validate these results, we performed RT-qPCR using spliceform-specific primers (Figure 3.7B) at three different RP genes: RPL43B, RPS21B, and RPS4B (Figure 3.7C,D,E). Consistent with our microarray data, we observed an increase in intron retention in RPL43B transcripts in WT, but not T4A. Further, in strains lacking Upf1, a component of the NMD machinery, we observed an increase in intron retention in T4A, suggesting that CTD phosphorylation may play a role in stress-dependent degradation of RP transcripts via NMD in response to osmostress.

3.5 Discussion

Previous studies have hinted at a connection between the Pol II CTD and stress. pSer2 was shown to increase upon heat shock and nutrient starvation (Patturajan et al., 1998). Further, in humans, osmotic stress increased pThr4 (Hintermair et al., 2012). And in mammalian fibroblasts, UV-induced DNA damage resulted in hyperphosphorylation of the CTD by the mammalian positive transcription elongation factor b (P-TEFb), which then regulates Pol II ubiquitylation and subsequent degradation (Heine et al., 2008). These stress-dependent changes in CTD phosphorylation were interpreted as indirect effects, and a direct link identifying the role of these marks has never been proposed.

We suggest that the phosphorylation state of the CTD plays an important role in regulating the ESR class of genes in response to specific, physiologically relevant, extracellular signals. We propose that ESR kinases “collaborate” with the canonical CTD kinases to simultaneously or combinatorially phosphorylate the CTD. Intriguingly, Tpk1 may “piggyback” with Pol II, phosphorylating the CTD (Figure 3.2) as it transcribes during osmostress. This would explain previous reports showing Tpk1 occupancy across the ORF of iESR genes (Baccarini et al., 2015; Pokholok et al., 2006). Similarly, the Hog1 peak observed at gene promoters in response to stress may illustrate hyperphosphorylation of the CTD before transcription initiates (Pokholok

et al., 2006). The patterns of marks that are placed by these non-canonical CTD kinases code for differential regulation of various signal-specific genes (Figure 3.6C). Although cells do not require CTD phosphorylation to respond to stress over long time periods (Figure 3.5), rapid regulation is required to efficiently respond to extracellular signals on a short time scale.

Previous results suggest that RP gene repression is not due to an increase decay rate, but rather, is transcriptionally mediated (Li et al., 1999). In addition to a decrease in transcription mediated through transcription factors and chromatin remodelers, we propose that stress-dependent CTD phosphorylation may recruit termination machinery to rapidly repress RP gene transcription. The phosphorylation state of Tyr1, Ser2, Thr4, Ser5, and Ser7 has been implicated in transcription termination. Dephosphorylated Tyr1 and pSer2 are required for recruiting Rtt103 and Pcf11 (Kim et al., 2004; Lunde et al., 2010; Mayer et al., 2012), pThr4 is critical for snoRNA termination (Nemec et al., in preparation), pSer5 recruits the Nrd1-Nab3-Sen1 complex to terminate small non-coding RNAs (Vasiljeva et al., 2008), and pSer7 recruits Integrator, a complex require for snRNA termination in higher eukaryotes (Baillat et al., 2005). Thus, phosphorylation of these residues in response to osmostress (Figure 3.2) may recruit similar machinery to rapidly terminate these transcripts.

Previous studies also suggest a potential mechanism whereby pSer5 recruits the Asr1 ubiquitin ligase, which ubiquitinates lysines beyond the last CTD repeat. Upon stress, Hog1-mediated Ser5 phosphorylation may enhance this ubiquitination, leading to ejection of the Rpb4/7 heterodimer from the core polymerase, thus inactivating Pol II (Daulny et al., 2008). This may provide an alternative mechanism for rapidly aborting Pol II in response to stress.

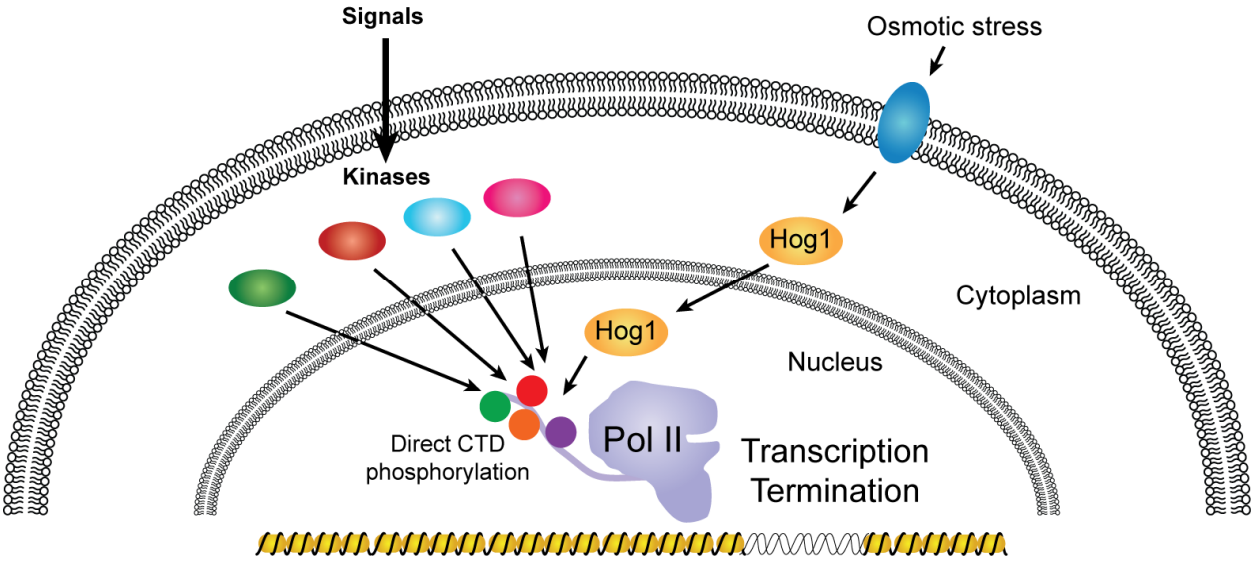

Figure 3.8 Model of stress-responsive, CTD-dependent transcriptional repression
Diverse signals activate a wide range of kinases, which can directly phosphorylate the CTD of Pol II. Along with TF activation and chromatin remodeling, direct phosphorylation of the CTD can remodel the transcriptome, likely through rapid termination of transcription.

Thus, we may uncover a novel branch of the environmental stress response that acts directly on the CTD of RNA polymerase. We propose that signal integration from various stimuli all center on the CTD of RNA Pol II to rapidly respond and remodel the transcriptome (Figure 3.8).

3.6 Future directions

We provide evidence that repression of RP genes in response to osmostress may be a combination of phospho-CTD-dependent transcription termination (Figure 3.6C) and phospho-CTD-dependent splicing inhibition. To decipher these regulatory mechanisms, we will purify WT and mutant Pol II from stressed or unstressed cells. Differential binding of protein factors to the CTDs will be assayed via label-free quantitative proteomics as described in Chapter 2. In T4A, we hypothesize that termination machinery and splicing machinery may be depleted. Obtaining the full CTD interactome in response stress will all us to solidify the mechanisms by which ESR genes are regulated.

Further, analysis of splice isoforms from total RNA (qPCR and microarray analysis) may convolute results. Splicing defects increase intron retention, and these transcripts are subsequently degraded via NMD (Garre et al., 2013). Depending on the length of stress before harvesting, the stability of unspliced transcripts, and the speed of RNA degradation, it can be difficult to discern the effects of stress on splicing. We propose to use comparative dynamic transcriptome analysis (cDTA) (Eser et al., 2014). As described in Chapter 2, 4-thiouracil is added to cultures, which is incorporated into nascent transcripts. MTSEA-biotin is conjugated to thiols, which can isolate nascent RNA using streptavidin beads. Importantly, an *S. pombe* spike in control is added while sequencing to normalize against global changes in transcript abundance in response to stress. More than 50% of splicing occurs within 45 nucleotides downstream of introns (Carrillo Oesterreich et al.). Therefore, cDTA will allow us to definitively determine if the CTD helps control alterations in splicing in response to stress.

3.7 Experimental procedures

Mutant CTD construction and plasmid shuffle

Mutant Rpb1 CTDs were constructed and tested for viability in vivo as previously described with some modifications (West and Corden, 1995). Briefly, the CTD repeats were constructed by annealing and ligating 5'-phosphorylated oligonucleotides containing WT or mutant codons at position 4 (Table 3.2). WT and mutant Rpb1 CTDs were cut with AvaI and ligated into similarly cut pSB0, which was transformed into DH5α and screened. pSB0 vectors were cut with KpnI and SnaBI and ligated into a similarly cut pY1 vector, a modified pRS315 vector containing full length Rpb1 (West and Corden, 1995). pY1 constructs were transformed into Z26 [*MATα his3Δ200 ura3-52 leu2-3,112 rpb1Δ::HIS3 GAL*$^{+}$ *(pRP112)*], which contains a URA3 linked WT Rpb1 gene. Transformants were plated on synthetic complete (SC) media lacking uracil and leucine. Single colonies were grown overnight in YPD, and plated on SC –Leu +5-FOA (1 mg/ mL) (Toronto Research Chemicals) to counterselect against any colonies maintaining the WT Rpb1.

Purification of TAP kinases:

Cells expression C-terminally TAP-tagged kinases (Ghaemmaghami et al., 2003) (GE-Healthcare) were grown in 200 mL selective media to an OD_{600} of 1.0, and cells were pelleted and transferred to a 1.5 mL tube. Pellets were resuspended in 500 µL TAP buffer A (20 mM HEPES pH 7.9, 300 mM potassium acetate, 0.5 mM EDTA pH 8.0, 10% glycerol, 0.05% NP40), with 1 mM DTT, 1 mM PMSF, 1 mM benzamidine, 1.45 mM pepstatin, and 1 mM phosphatase inhibitors (NaF, NaN3, and $Na3VO_4$) freshly added. Approximately 200 µL silica beads (OPS Diagnostics) were added, and cells were lysed via bead beating for 15 minutes. A 22 gauge needle was heated with a Bunsen burner and used to puncture a hole in the tube, which was then placed in a 2 mL tube and spun at 7,000 RPM for 30 seconds. 50 µL of a 50/50 slurry of IgG-Sepharose 6 Fast Flow beads (GE-Healthcare) in TAP buffer A was added to the

supernatant. After 3 hours nutating at 4°C, beads were pelleted and washed five times with TAP buffer A (500 µL each). Kinases were eluted overnight at 4°C in 25 µL TAP buffer A with 1 mM DTT and 10 U AcTEV (Invitrogen).

GST-CTD purification:

The CTD substrate, GST-CTD14 (fourteen repeats of YSPTSPS fused to GST) was expressed in BL21(DE3). Cells were grown to midlog phase at 37°C and induced with 1 mM IPTG overnight at 16°C. Cells were pelleted (5,000 g for 20 minutes) and resuspended in 5 mL Buffer B (1X PBS, 10% glycerol, 0.01% NP40) with freshly added 1 mM PMSF, 1 mM benzamidine, 1.45 mM pepstatin. Cells were lysed with 3 rounds of sonication at power level 5 using a Heat Systems-Ultrasonics Inc. W-220 Sonicator. Cell debris was pelleted at 13,000 rpm for 20 minutes at 4°C and a 200 µL 50/50 slurry of Gutathione-Sepharose 4B beads (GE-Healthcare) in Buffer B was added to the supernatant. After 3 hours nutating at 4°C, beads were pelleted and washed once with Buffer B. Beads were resuspended in 1 mL FastAP buffer (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 100 mM KCl, 0.02% TritonX-100, 100 ug/ mL BSA) and incubated with 100 U FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) for 1 hour at 37°C. Beads were washed four times with Buffer B, then once with Buffer C (5 mM HEPES pH 7.4, 10 mM potassium glutamate, 10 mM magnesium acetate, 0.5 mM EDTA pH 8.0, 10% glycerol, 0.1% NP-40). Reduced glutathione (Fisher Scientific) was added to Buffer C to 50 mM, and the pH was adjusted to 7.0. GST-CTD14 was eluted overnight in 200 µL Buffer C with 50 mM reduced glutathione. Any remaining alkaline phosphatase was heat inactivated at 75°C for 5 minutes.

***In vitro* kinase assay:**

Each assay was performed with 5 µL of tandem affinity purified (TAP) kinase and 1 µM GST-CTD14 in a 30 uL volume of Buffer D (Ansari et al., 2005) (20 mM HEPES pH 7.5, 10% glycerol,

2.5 mM EGTA, 15 mM magnesium acetate, 100 mM potassium acetate) with 1 mM DTT, 1 mM ATP, 1 mM PMSF, 1 mM benzamidine, 1.45 mM pepstatin, and 1 mM phosphatase inhibitors (NaF, NaN3, and $Na3VO_4$). Reactions were performed at 30°C for two hours and resolved via SDS-PAGE. Western blot analysis was performed using standard procedures. Blots were probed with antibodies targeting pSer2 (Bethyl), pThr4 (1G7), pSer5 (3E8), or pSer7 (4E12). Quantitation was performed using ImageJ.

RNA expression analysis

Cell collection, RNA preparation, cDNA synthesis and labeling, array hybridization, and normalization were as previously described (Berry and Gasch, 2008; Lee et al., 2011), using cyanine dyes (Flownamics, Madison WI) and Superscript III (Life Technologies, Carlsbad, CA). Samples were hybridized to whole-genome tiled DNA microarrays (Roche Nimblegen, Madison, WI), comparing cDNA from the salt-treated sample to cDNA generated from the unstressed culture. Comparison of unstressed strains was done as previously described (Lee et al., 2011) by retrieving and comparing single-channel data from mutant and wild type arrays. Genes whose expression was altered in wild-type cells responding to NaCl were identified based on a 1.5 fold cutoff. Heatmaps were generated with MultiExperimentViewer, and GO annotations were identified with FunSpec (Robinson et al., 2002).

Chromatin immunoprecipitation on microarray chips (ChIP-chip)

WT or CTD mutant cells were grown in selective media (100 mL) to an OD_{600} of 0.5 OD. Cultures were split and 9.2 mL YPD +4.5 M NaCl (0.7M final) or YPD were added for 30 minutes. Formaldehyde was added (1%) for 20 minutes to crosslink cells, and excess formaldehyde was quenched with glycine (125 mM) for 5 minutes. Cells were harvested and resuspended in 600 µL lysis buffer (50 mM HEPES-KOH, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate) with 1 mM PMSF, 1 mM benzamidine, and 1.45 mM

pepstatin freshly added. Approximately 200 µL silica beads (OPS Diagnostics) were added, and cells were lysed via bead beating for 15 minutes. A 22 gauge needle was heated with a Bunsen burner and used to puncture a hole in the tube, which was then placed in a 2 mL tube and spun at 7,000 RPM for 30 seconds. Debris was resuspened and samples were sonicated on ice for 32 minutes at 60% amplitude for 10s on and 10s off with a Misonix Sonicator so that DNA fragments were mainly between 200 and 500 bp. Debris was pelleted, and 10 µL extract was reserved as input. To the remaining extract, 3 µL α-Rpb3 antibody (Biolegend, formerly Neoclone) was incubated overnight at 4°C. Protein G Dynabeads (Life Technologies) were incubated, and complexes were pulled down. Immunoprecipitated samples were washed twice with lysis buffer, twice with lysis buffer containing 500 mM NaCl, twice with wash buffer (10 mM Tris-HCl, 250 mM LiCl, 0.5% NP-40, 1% Na-Deoxycholate, 1 mM EDTA), and once with TE. 95 µL elution buffer (50 mM Tris-HCl, 10 mM EDTA, 0.5% SDS, 5 mM $CaCl_2$) was added and beads were incubated at 65°C for 30 minutes. 5 µL of 2.0 mg/ mL Proteinase K (50 mM Tris [pH 8], 10 mM $CaCl_2$) was added and samples were incubated overnight at 65° to reverse crosslinks. 10 µL input, 85 µL elution buffer, and 5 µL Proteinase K were also incubated overnight at 65°C. DNA was purified (Epoch Life Sciences) and blunted with T4 DNA polymerase. Annealed linkers were ligated to DNA overnight with T4 DNA ligase. Ligation-mediated PCR was performed, and Cy3-UTP was incorporated into amplified input DNA, and Cy5-UTP was incorporated into IP DNA. DNA was purified (Epoch Life Sciences) and dye incorporation was quantitated via NanoDrop. 24 pmol Cy5-labeled IP DNA and an equal mass of Cy3-labeled input DNA was mixed and concentrated with a vacuum concentrator. NimbleGen hybridization components were mixed with labeled DNA, boiled, and loaded on to a custom yeast tiled DNA microarray. Samples were incubated overnight at 42°C with a MAUI Hybridization System. Arrays were washed and scanned using an Axon GenePix 4000B scanner. Immunoprecipitated (IP) data were median-scaled against respective “input” sample data, and the ratio of scaled IP to input was Log_2 transformed. A moving average was used to

smooth the microarray data. To compare ChIP profiles between strains, quantile normalization was performed.

RT-qPCR

WT and T4A were grown in 50 mL selective media to OD_{600} 0.5. Cultures were split and 4.6 mL YPD +4.5M NaCl (0.7M final) or YPD were added for 30 minutes. Cells were pelleted, resuspended in 700 µL TE Lysis buffer (10 mM Tris pH 7.5, 10 mM EDTA, 0.5% SDS), and flash frozen. Cells were thawed, 700 µL acid phenol was added, and samples were incubated at 65°C for 45 minutes, vortexing every 15 minutes. Samples were spun for 15 minutes at 13,000 rpm at 4°C. The aqueous layer was added to 750 µL fresh acid phenol, vortexed for 15 seconds, and spun again at 13,000 rpm at 4°C. The aqueous layer was added to 700 µL chloroform and added to pre-spun phase lock gel tubes (5PRIME). The aqueous layer was transferred to a 1.5 mL tube and 50 µL NaOAc and 1.25 mL 100% chilled EtOH was added. Samples were incubated at -80°C overnight. Samples were spun at 13,000 rpm for 15 minutes at 4°C. The pellet was washed with 70% EtOH and spun again. SuperScript III was used to reverse transcribe 1 µg RNA using random hexamers. qPCR was performed using splice specific primers (Table 3.2) using VeriQuest qPCR master mix (Affymetrix) on an ABI7500 Fast thermocycler.

Table 3.2 Strains used in this study

Strain	Genotype	Source
TAP tagged strains	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 kinase-TAG HIS3	GE Healthcare
Z26	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pRP112 (RPB1, URA3)	(Nonet et al., 1987)
WT (21)	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pY1(WT(21))	This study
WT (14)	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pY1(WT(14))	This study
WT (10)	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pY1(WT(10))	This study
S2A/WT	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pY1(S2A(8)WT(7))	(West and Corden, 1995)
T4A	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pY1(T4A(19))	This study
4E	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pY1(4E(23))	This study
S5A/WT	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pY1(S5A(5)WT(7))	(West and Corden, 1995)
S7A	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pY1(S7A(10))	(Zhang et al., 2012)
WT (21) *upf1Δ*	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pY1(WT(21)) *upf1Δ*::KanMX	This study
T4A *upf1Δ*	MATα his3Δ200 leu2-3,112 ura3-52 rpb1Δ187::HIS3 pY1(T4A(19)) *upf1Δ*::KanMX	This study

Table 3.3 Oligonucleotides used in this study

Name	Function	Sequence (5'-3')
consWT-F	CTD construct	P-CCGAGCTATAGTCCAACTTCA
consWT-R	CTD construct	P-TCGGTGAAGTTGGACTATAGC
T4A-F	CTD construct	P-CCGAGCTATAGTCCAGCTTCA
T4A-R	CTD construct	P-TCGGTGAAGCTGGACTATAGC
RPL43B unspliced	qPCR	ATTTTTTTTTTTTGCAGGGCTAAGAG
RPL43B spliced	qPCR	CAAACAAAAAAATGGCTAAGAGAAC
RPL43B reverse	qPCR	CAGTCATATCTGGCATGTTGTTG
RPS21B unspliced	qPCR	TTTACTAGGTTGAATTATATGTTCC
RPS21B spliced	qPCR	GGTCAATTAGTTGAATTATATGTTC
RPS21B reverse	qPCR	GATGTATTCACCTGGAATGGC
RPS4B unspliced	qPCR	TTCATATAGAAAGAAGCATCTAAAG
RPS4B spliced	qPCR	CTAGAGGACCAAAGAAGCATC
RPS4B reverse	qPCR	CAAGGATTCACGCAATTTGTG

3.8 References

Alejandro-Osorio, A., Huebert, D., Porcaro, D., Sonntag, M., Nillasithanukroh, S., Will, J., and Gasch, A. (2009). The histone deacetylase Rpd3p is required for transient changes in genomic expression in response to stress. Genome Biology *10*, R57.

Alepuz, P.M., de Nadal, E., Zapater, M., Ammerer, G., and Posas, F. (2003). Osmostress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II. EMBO J *22*, 2433-2442.

Alepuz, P.M., Jovanovic, A., Reiser, V., and Ammerer, G. (2001). Stress-Induced MAP Kinase Hog1 Is Part of Transcription Activation Complexes. Mol Cell *7*, 767-777.

Amorós, M., and Estruch, F. (2001). Hsf1p and Msn2/4p cooperate in the expression of Saccharomyces cerevisiae genes HSP26 and HSP104 in a gene- and stress type-dependent manner. Mol Microbiol *39*, 1523-1532.

Ansari, A.Z., Ogirala, A., and Ptashne, M. (2005). Transcriptional activating regions target attached substrates to a cyclin-dependent kinase. Proc Natl Acad Sci USA *102*, 2346-2349.

Baccarini, L., Martínez-Montañés, F., Rossi, S., Proft, M., and Portela, P. (2015). PKA-chromatin association at stress responsive target genes from Saccharomyces cerevisiae. Biochim Biophys Acta *1849*, 1329-1339.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L.*, et al.* (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell *32*, 878-887.

Baillat, D., Hakimi, M.-A., Näär, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a Multiprotein Mediator of Small Nuclear RNA Processing, Associates with the C-Terminal Repeat of RNA Polymerase II. Cell *123*, 265-276.

Beck, T., and Hall, M.N. (1999). The TOR signalling pathway controls nuclear localization of nutrient-regulated transcription factors. Nature *402*, 689-692.

Bergkessel, M., Whitworth, G.B., and Guthrie, C. (2011). Diverse environmental stresses elicit distinct responses at the level of pre-mRNA processing in yeast. RNA *17*, 1461-1478.

Berry, D.B., and Gasch, A.P. (2008). Stress-activated Genomic Expression Changes Serve a Preparative Role for Impending Stress in Yeast. Mol Biol Cell *19*, 4580-4587.

Buratowski, S. (2003). The CTD code. Nat Struct Mol Biol *10*, 679-680.

Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, Karla M. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell.

Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S., and Young, R.A. (2001). Remodeling of yeast genome expression in response to environmental changes. Molecular Biology of the Cell *12*, 323-337.

Chasman, D., Ho, Y.H., Berry, D.B., Nemec, C.M., MacGilvray, M.E., Hose, J., Merrill, A.E., Lee, M.V., Will, J.L., Coon, J.J.*, et al.* (2014). Pathway connectivity and signaling coordination in the yeast stress-activated signaling network. Mol Syst Biol *10*, 759.

Chymkowitch, P., Eldholm, V., Lorenz, S., Zimmermann, C., Lindvall, J.M., Bjoras, M., Meza-Zepeda, L.A., and Enserink, J.M. (2012). Cdc28 kinase activity regulates the basal transcription machinery at a subset of genes. Proc Natl Acad Sci U S A *109*, 10450-10455.

Corden, J.L., Cadena, D.L., Ahearn, J.M., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci USA *82*, 7934-7938.

Daulny, A., Geng, F., Muratani, M., Geisinger, J.M., Salghetti, S.E., and Tansey, W.P. (2008). Modulation of RNA polymerase II subunit composition by ubiquitylation. Proc Natl Acad Sci USA *105*, 19649-19654.

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

Ding, X., Yu, Q., Zhang, B., Xu, N., Jia, C., Dong, Y., Chen, Y., Xing, L., and Li, M. (2014). The type II Ca2+/calmodulin-dependent protein kinases are involved in the regulation of cell wall integrity and oxidative stress response in Candida albicans. Biochem Biophys Res Commun *446*, 1073-1078.

Eser, P., Demel, C., Maier, K.C., Schwalb, B., Pirkl, N., Martin, D.E., Cramer, P., and Tresch, A. (2014). Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression, Vol 10.

Garre, E., Romero-Santacreu, L., Barneo-Munoz, M., Miguel, A., Perez-Ortin, J.E., and Alepuz, P. (2013). Nonsense-mediated mRNA decay controls the changes in yeast ribosomal protein pre-mRNAs levels upon osmotic stress. PLoS ONE *8*, e61240.

Gasch, A., and Werner-Washburne, M. (2002). The genomics of yeast responses to environmental stress and starvation. Functional & Integrative Genomics *2*, 181-192.

Gasch, A.P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. Yeast *24*, 961-976.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Molecular Biology of the Cell *11*, 4241-4257.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. Mol Cell Biol *29*, 5455-5464.

Görner, W., Durchschlag, E., Martinez-Pastor, M.T., Estruch, F., Ammerer, G., Hamilton, B., Ruis, H., and Schüller, C. (1998). Nuclear localization of the C(2)H(2) zinc finger protein Msn2p is regulated by stress and protein kinase A activity. Genes Dev *12*, 586-597.

Heine, G.F., Horwitz, A.A., and Parvin, J.D. (2008). Multiple mechanisms contribute to inhibit transcription in response to DNA damage. J Biol Chem *283*, 9555-9561.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E.*, et al.* (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J *31*, 2784-2797.

Hsin, J.-P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD Phosphorylated on Threonine-4 Is Required for Histone mRNA 3′ End Processing. Science *334*, 683-686.

Hsin, J.P., Xiang, K., and Manley, J.L. (2014). Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing. Mol Cell Biol *34*, 2488-2498.

Huber, A., French, S.L., Tekotte, H., Yerlikaya, S., Stahl, M., Perepelkina, M.P., Tyers, M., Rougemont, J., Beyer, A.L., and Loewith, R. (2011). Sch9 regulates ribosome biogenesis via Stb3, Dot6 and Tod6 and the histone deacetylase complex RPD3L. EMBO J *30*, 3052-3064.

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

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517-522.

Lee, M.V., Topper, S.E., Hubler, S.L., Hose, J., Wenger, C.D., Coon, J.J., and Gasch, A.P. (2011). A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. Mol Syst Biol *7*, 514.

Li, B., Nierras, C.R., and Warner, J.R. (1999). Transcriptional Elements Involved in the Repression of Ribosomal Protein Synthesis. Mol Cell Biol *19*, 5393-5404.

Liko, D., Slattery, M.G., and Heideman, W. (2007). Stb3 Binds to Ribosomal RNA Processing Element Motifs That Control Transcriptional Responses to Growth in Saccharomyces cerevisiae. J Biol Chem *282*, 26623-26628.

Lippman, S.I., and Broach, J.R. (2009). Protein kinase A and TORC1 activate genes for ribosomal biogenesis by inactivating repressors encoded by Dot6 and its homolog Tod6. Proc Natl Acad Sci USA *106*, 19928-19933.

Loewith, R., and Hall, M.N. (2011). Target of Rapamycin (TOR) in Nutrient Signaling and Growth Control. Genetics *189*, 1177-1201.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription

termination factors with the RNA polymerase II C-terminal domain. Nat Struct Mol Biol *17*, 1195-1201.

Martínez-Pastor, M.T., Marchler, G., Schüller, C., Marchler-Bauer, A., Ruis, H., and Estruch, F. (1996). The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). EMBO J *15*, 2227-2235.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. Science *336*, 1723-1725.

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

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

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dumcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dolken, L.*, et al.* (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol *7*.

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

Nemec, C.M., Gilmore, J.M., Ho, Y., Hintermair, C., Singh, A.K., Ringelberg, K.J., Tseng, S.C., Heidemann, M., Zhang, Y., Florens, L.*, et al.* (in preparation). Non-canonical CTD-kinases regulate gene-class specific functions of RNA polymerase II.

Nonet, M., Sweetser, D., and Young, R.A. (1987). Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. Cell *50*, 909-915.

Pascual-Ahuir, A., and Proft, M. (2007). The Sch9 kinase is a chromatin-associated transcriptional activator of osmostress-responsive genes. EMBO J *26*, 3098-3108.

Patturajan, M., Schulte, R.J., Sefton, B.M., Berezney, R., Vincent, M., Bensaude, O., Warren, S.L., and Corden, J.L. (1998). Growth-related Changes in Phosphorylation of Yeast RNA Polymerase II. J Biol Chem *273*, 4689-4694.

Pleiss, J.A., Whitworth, G.B., Bergkessel, M., and Guthrie, C. (2007). Rapid, transcript-specific changes in splicing in response to environmental stress. Mol Cell *27*, 928-937.

Pokholok, D.K., Zeitlinger, J., Hannett, N.M., Reynolds, D.B., and Young, R.A. (2006). Activated signal transduction kinases frequently occupy target genes. Science *313*, 533-536.

Rep, M., Reiser, V., Gartner, U., Thevelein, J.M., Hohmann, S., Ammerer, G., and Ruis, H. (1999). Osmotic Stress-Induced Gene Expression in Saccharomyces cerevisiae Requires Msn1p and the Novel Nuclear Factor Hot1p. Mol Cell Biol *19*, 5474-5485.

Robinson, M.D., Grigull, J., Mohammad, N., and Hughes, T.R. (2002). FunSpec: a web-based cluster interpreter for yeast. BMC Bioinformatics *3*, 35-35.

Rodríguez-Molina, Juan B., Tseng, Sandra C., Simonett, Shane P., Taunton, J., and Ansari, Aseem Z. (2016). Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization. Mol Cell.

Rosonina, E., Yurko, N., Li, W., Hoque, M., Tian, B., and Manley, J.L. (2014). Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. Proc Natl Acad Sci U S A *111*, 11924-11931.

Schmitt, A.P., and McEntee, K. (1996). Msn2p, a zinc finger DNA-binding protein, is the transcriptional activator of the multistress response in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America *93*, 5777-5782.

Smets, B., Ghillebert, R., De Snijder, P., Binda, M., Swinnen, E., De Virgilio, C., and Winderickx, J. (2010). Life in the midst of scarcity: adaptations to nutrient availability in Saccharomyces cerevisiae. Curr Genet *56*, 1-32.

Smith, A., Ward, M.P., and Garrett, S. (1998). Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation. EMBO J *17*, 3556-3564.

Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S*., et al.* (2010). Chemical-genomic dissection of the CTD code. Nat Struct Mol Biol *17*, 1154-1161.

Urban, J., Soulard, A., Huber, A., Lippman, S., Mukhopadhyay, D., Deloche, O., Wanke, V., Anrather, D., Ammerer, G., Riezman, H*., et al.* (2007). Sch9 Is a Major Target of TORC1 in Saccharomyces cerevisiae. Mol Cell *26*, 663-674.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat Struct Mol Biol *15*, 795-804.

West, M.L., and Corden, J.L. (1995). Construction and Analysis of Yeast RNA Polymerase II CTD Deletion and Substitution Mutations. Genetics *140*, 1223-1233.

Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodríguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. (2012). Ssu72 Phosphatase-dependent Erasure of Phospho-Ser7 Marks on the RNA Polymerase II C-terminal Domain Is Essential for Viability and Transcription Termination. J Biol Chem *287*, 8541-8551.

Zhang, J., and Corden, J.L. (1991). Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. J Biol Chem *266*, 2290-2296.

Zhu, C., Byers, K.J.R.P., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M*., et al.* (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res *19*, 556-566.

Chapter 4: The future of the CTD

This chapter has been adapted from "Emerging views on the CTD code." by David W. Zhang, Juan B. Rodríguez-Molina, Joshua R. Tietjen, Corey M. Nemec, and Aseem Z. Ansari.

4.1 Expansion of the CTD code

Extraordinarily rapid progress has been made over the last several years in the field of CTD research; however, many important questions remain unanswered. Although the profiles of pSer7 have been mapped and several of its kinases discovered, its function at protein coding genes remains unclear. While Tyr1 can be phosphorylated by c-Abl in mammals (Baskaran et al., 1999), the *S.* cerevisiae Tyr1 kinase remains a specter. Interestingly, pTyr1 and pThr4 were found in both the hyperphosphorylated and hypophosphorylated states of Pol II, opening the possibility of CTD function independent of transcription. Additionally, most of the kinases identified are established members of the transcription initiation or elongation complexes. In Chapter 2, we identify CTD kinases in unexpected groups, and in Chapter 3, we show that these kinases can modulate the CTD in response to signals.

In addition phosphorylation, serines and threonines on the CTD can be glycosylated. Recent studies suggest O-GlcNAc are transfered to Ser5 and Ser7 by O-GlcNac transferase and removed by O-GlcNAc aminidase during PIC assembly. This cycling of O-GlcNAc may be important for preventing aberrant CTD phosphorylation by TFIIH (Kelly et al., 1993; Ranuncolo et al., 2012). The seventh position of the human CTD also contains five asparagines, which can potentially be glycosylated, and six threonines, which may be phosphorylated or glycosylated. Antibodies targeting these marks can be developed, or mass sepctrometry can be used to determine if these post-translational modifications are present.

The role of non-canonical residues and their modification states on mammalian CTD has only begun to be explored (Dias et al., 2015; Simonti et al., 2015; Sims et al., 2011; Voss et al., 2015). In mammals, the Ser7 residue is only weakly conserved in polymerase-distal repeats of the CTD, often changed to lysine or arginine (Chapman et al., 2007). Interestingly, Arg1810 of *rpb1* in the human CTD is methylated by the coactivator-associated methytransferase1

(CARM1/PRMT4) (Sims et al., 2011) (Figure 4.1). This methylation occurs prior to both transcription initiation and phosphorylation of Ser2 or Ser5, and substitution of this residue results in the improper expression of a variety of snRNAs and snoRNAs. PRMT5 can also methylate R1810 and has been implicated in R-loop resolution (Yanling Zhao et al., 2016). Further methylation and acetylation occur on lysines at the seventh position and are marks of early transcription and fine-tune gene expression levels (Dias et al., 2015; Voss et al., 2015). Finally, six of the eight lysines are ubiquitinated by WWP2, an E3 ubiquitin ligase, thus targeting Rpb1 for degradation by the proteasome (Li et al., 2007).

Not only can PRMT4/5 methylate the CTD, but they can also methylate multiple histone residues (Miao et al., 2006). Lys7 is acetylated by p300/CBP, which also acetylates H2K27 (Jin et al., 2011). While the CTD Lys7 methyltransferase is unknown, preliminary work from our lab suggests that PHF8, a known CTD “reader” and H3K9me2 demethylase (Fortschegger et al., 2010), may also demethylate Lys7. Further, COMPASS and SET2 are known to interact with the CTD and post-translationally modify histones (Kim and Buratowski, 2009; Kizer et al., 2005; Lee and Skalnik, 2008; Ng et al., 2003; Vojnic et al., 2006). Clearly, significant crosstalk occurs between the CTD, histone marks, and the enzymes that modify them. Interrogating this crosstalk and integrating the “CTD code” with the “histone code” will be an exciting avenue of future research (Figure 4.1).

Preliminary results from our lab confirm that PRMT4 can methylate the CTD (Figure 4.2), but we were unable to demonstrate methylation by PRMT5. Our attempts to identify the Lys7 methyltransferase were unsuccessful. G9a, SETDB1, SUV39H, and EZH2 were purified and assayed for methyltransferase activity on a mouse CTD substrate. No signal was observed, suggesting that these histone methyltransferases are not involved in crosstalk with the CTD (Figure 4.2). The Lys7 methyltransferase remains elusive.

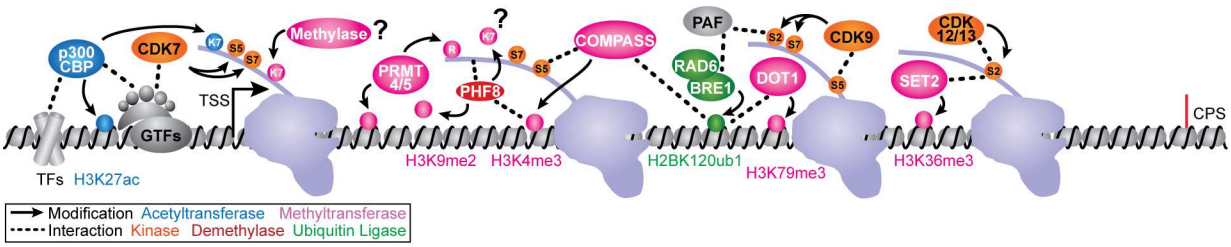

Figure 4.1 Crosstalk between CTD marks, histone marks, and the enzymes that place them in humans

Transcription often begins with the binding of sequence-specific transcription factors (TFs) at gene promoters. Subsequently, acetyltransferases p300/CBP bind these TFs to co-activate transcription through acetylation of H3K27. This permissive chromatin mark allows for the association of general transcription factors (GTFs), which recruit Pol II, forming the pre-initiation complex. Our data suggest that the K7 residues of the CTD are methylated at or before this early stage by a yet to be identified methyltransferase. After Pol II initiates, p300/CBP acetylate CTD K7, and the kinase subunit of TFIIH, CDK7, phosphorylates CTD Ser5 and Ser7. Both pSer5 and TFs recruit COMPASS, which contains Set1-like methyltransferases that trimethylate H3K4. PRMT4 and PRMT5 methylate both histones and R1810 of the CTD. PHF8, a demethylase of the repressive H3K9me1/2 marks, associates with chromatin by binding to H3K4me3. Interestingly, PHF8 physically interacts with the CTD, but the significance of this interaction remains to be determined. Subsequently, transcription elongation is promoted by recruitment of BRD4 through binding to acetylated histones, which in turn recruits CDK9/pTEF-B/Super Elongation Complex to actively transcribed genes. CDK9 phosphorylates the CTD at Ser2 and Ser7, which bind the PAF complex. PAF recruits RAD6/BRE1, which ubiquinates H2BK120. H2BK120ub1 amplifies methylation of H3K4 and H3K79 by binding COMPASS and DOT1 methyltransferases. Elongation is further marked by phosphorylation of Ser2 by CDK12/13, which recruits the SET2 methyltransferase to place H3K36me3 serving to reduce cryptic transcription.

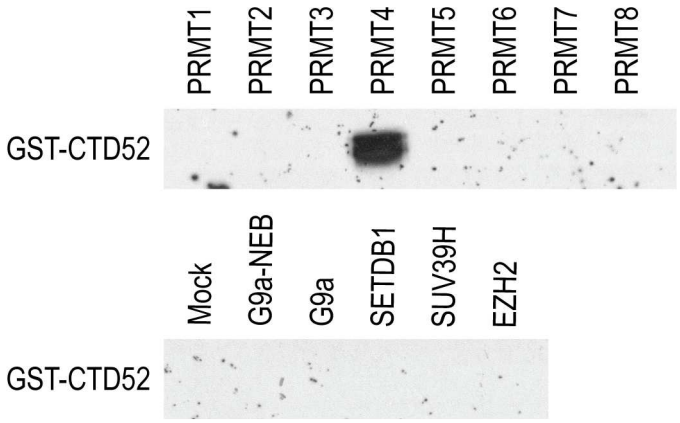

Figure 4.2 Several histone methyltransferases do not methylate Lys7
PRMT1-8 were purified by the Xu lab and tested for in vitro methyltransferase activity on a mouse GST-CTD peptide expressed in bacteria. ^{3}H SAM was used to detect methyltransferase activity. G9a was either purchased from New England Biolabs or expressed in HEK293 cells. SETDB1, SUV39H, and EZH2 were also expressed and purified from HEK293 cells.

4.2 How are phosphates distributed across the length of the CTD?

Besides the characterization of novel marks, significant structural challenges remain for understanding the known phospho-marks. While substitution of Ser2 with glutamate (a phosphomimic) in the core-distal repeats and substitution of Ser5 with glutamate in the core-proximal repeats are lethal (West and Corden, 1995), this does not directly demonstrate whether these CTD marks are generally placed at those repeats. Characterizing the phosphorylation patterns and protein binding at individual repeats will help determine the existence of a "CTD code," and this promises to be one of the most exciting and important challenges in the future of CTD research. However, the extensive use of phospho-CTD specific antibodies presents a challenge. While many phospho-specific CTD antibodies have been developed, the context of nearby residues greatly influences antibody affinity for a specific mark (Figure 4.3) (Chapman et al., 2007; Hintermair et al., 2012; Mayer et al., 2012).

To further test the specificity of phospho-CTD antibodies, immunoblots probing each phospho-CTD mark in strains bearing alanine substitutions were performed (Figure 4.4A). Mutant and WT polymerases were tagged with an HA tag, and immunoprecipitated with α-HA magnetic beads. Samples were immunoblotted and probed with antibodies targeting all five phospho-CTD marks. We observed that T4A is only minimally phosphorylated at Tyr1 (3D12), Ser2 (Bethyl), and Ser7 (4E12). As expected, pThr4 (1G7) was absent in T4A as well, but surprisingly, pSer5 (3E8) was also completely absent. We tested another pSer5 antibody (H14) and found pSer5 at equal levels as the other strains, suggesting that Thr4 may be important for the specificity of many antibodies. To validate this result, we studied chemically-synthesized peptides with two WT or T4A CTD repeats containing either no phosphorylation or pSer2. We observed that pSer2 was recognized by the pSer2 antibody in the context of a WT CTD, but pSer2 on a T4A

peptide was not observed (Figure 4.4B). Thus, we cannot conclude that pY1, pS2, or pS7 are present, but it is likely that their absence by is due to the specificity of the antibodies.

To characterize the pattern of CTD modifications using an antibody-independent approach, the use of mass spectrometry has been proposed. However, the highly repetitive nature of the CTD makes it difficult to distinguish between each of the repeats by mass. Further, the lack of positively charged residues prevents the CTD from being cleaved by trypsin into smaller fragments. To resolve these issues, preliminary experiments using synthetic peptides containing three CTD repeats, a PEG linker, and two C-terminal lysines were performed. The activity of canonical CTD kinases was tested on this substrate and analyzed by mass spectrometry (Figure 4.5). We observed strong phosphorylation on a single repeat using Kin28, Ctk1, and Srb10. As expected, Kin28 phosphorylated Ser5 to a large extent (40 fold over unphosphorylated). Bur1 phosphorylated Ser2 on the second repeat at low levels. Interestingly, neither Kin28 nor Bur1 phosphorylated Ser7 in vitro. As expected, Srb10 (pulled down with Srb11-TAP) phosphorylated Ser5. Finally Ctk1 phosphorylate Tyr1 to a large extent (30 fold over unphosphorylated) in vitro. This was confirmed Western Blot analysis. Unfortunately, inhibition of Ctk1 in vivo did not lead to a reduction of pTyr1, suggesting that Ctk1 activity in vitro may be promiscuous.

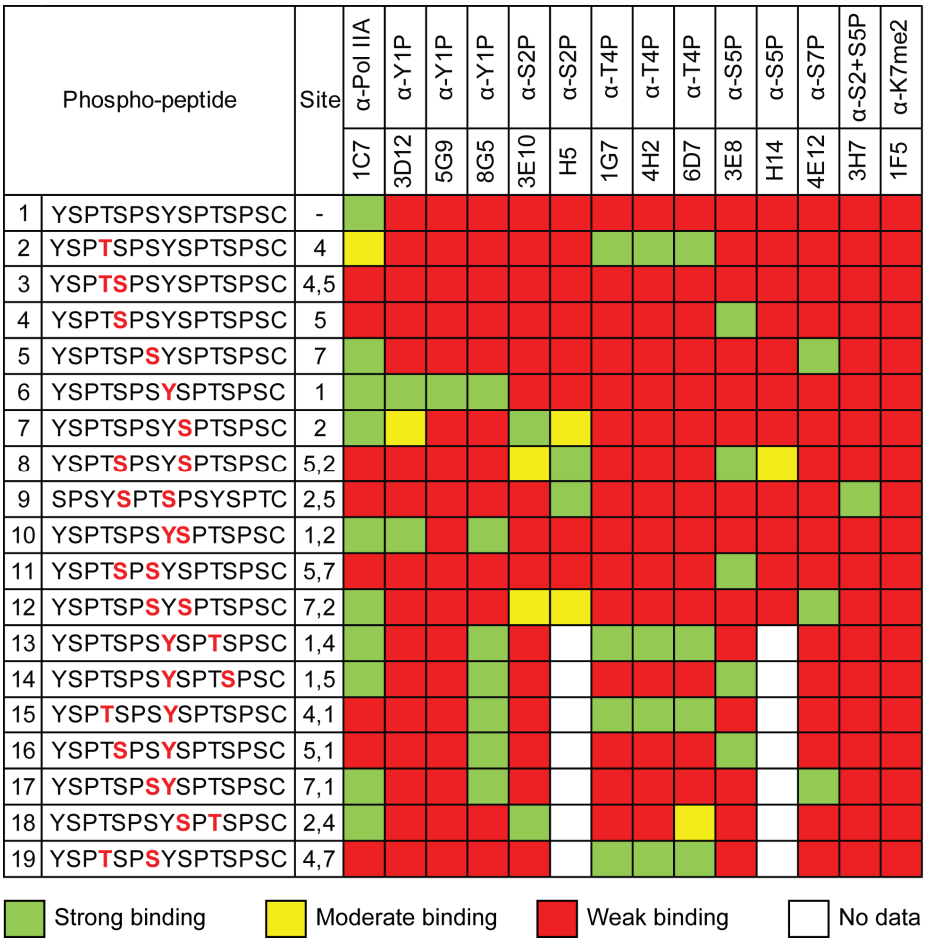

Figure 4.3 Antibody affinities for phospho-CTD peptides
The binding specificity of commercially available α-phospho-CTD antibodies were tested via ELISA. Synthetic phospho-peptides are shown, with phosphorylated residues listed (site) and highlighted in red. Antibodies raised to target specific epitopes are listed along with their clone name. A strong interaction between antibody and peptide is shown in green, moderate in yellow, weak in red, and no data in white. Adapted from (Chapman et al., 2007; Hintermair et al., 2012; Mayer et al., 2012).

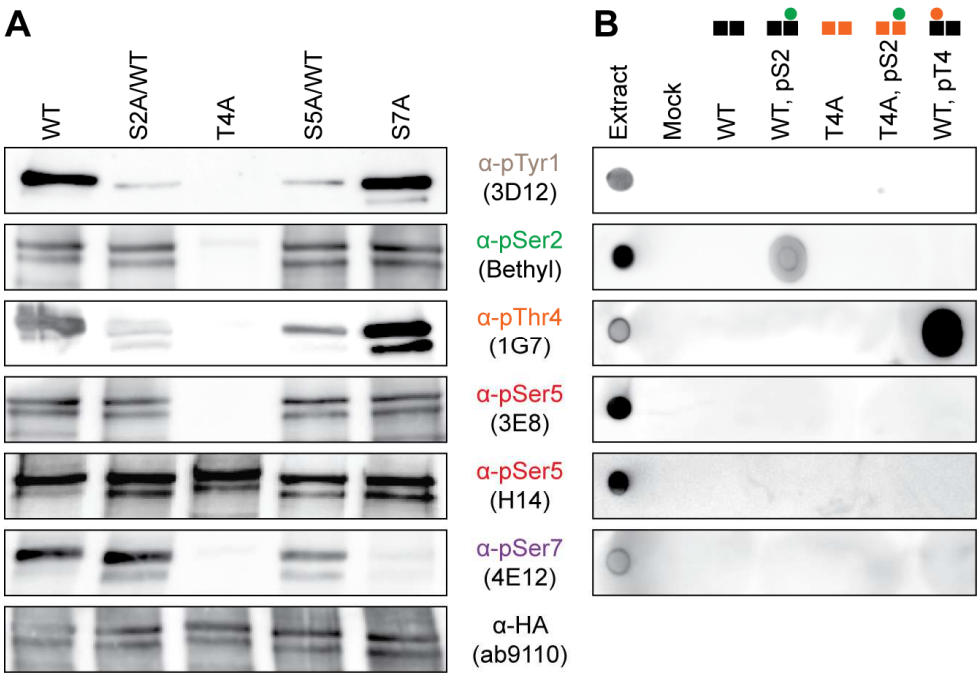

Figure 4.4 Context-dependent specificity of phospho-CTD antibodies

A) HA tagged Pol II were purified from strains bearing WT or mutant CTDs. Samples were probed with antibodies targeting pTyr1 (3D12), pSer2 (Bethyl), pThr4 (1G7), pSer5 (3E8 and H14), and pSer7 (4E12). An antibody targeting the HA tag (Abcam) was used as a loading control.

B) Extract, vehicle, or synthetic peptides were spotted on nitrocellulose and probed with the same antibodies as in A). Black boxes denote WT (YSPTSPS) CTD repeats, orange boxes denote T4A (YSPASPS) repeats, green circles indicate pSer2, and orange circles indicate pThr4.

Kin28

Number of phosphorylations	Relative abundance (compared to unphosphorylated)	Sites of phosphorylation
1	40x	2nd repeat: Ser5
2	0.5x	1st repeat: Tyr1, Ser2, Thr4, or Ser5 2nd repeat: Ser5

Ctk1

Number of phosphorylations	Relative abundance (compared to unphosphorylated)	Sites of phosphorylation
1	30x	2nd repeat: Tyr1
2	0.5x	1st repeat: Tyr1, Ser2, Thr4, or Ser5 2nd repeat: Thr4 or Ser5

Bur1

Number of phosphorylations	Relative abundance (compared to unphosphorylated)	Sites of phosphorylation
1	0.5x	2nd repeat: Ser2

Srb11

Number of phosphorylations	Relative abundance (compared to unphosphorylated)	Sites of phosphorylation
1	3x	2nd repeat: Ser5

Figure 4.5 Analysis of in vitro phosphorylated CTD peptides by mass spectrometry
Synthetic CTD peptides were synthesized. A PEG linker was used to increase peptide flexibility, and lysines were added to aid with mass spectrometry analysis. Peptides were used as a substrate for kinases in an in vitro kinase assay as described in Chapter 2. Peptides were analyzed by mass spectrometry, and phosphorylation sites were determined.

As stated previously, the length of the CTD, its repetitive nature, and the lack of cleavage sites make analysis of phospho-CTD state difficult. Recently, two labs attempted to resolve these issues by interspersing lysines throughout the CTD to allow for tryptic cleavage. Further, the addition or substitution of amino acids throughout creates “barcodes” to uniquely identify each peptide following tryptic digestion (Schüller et al., 2016; Suh et al., 2016). However, with all these substitutions in a single strain, are the identified marks biologically relevant? Though tryptic digestion allows for analysis via mass spectrometry, it prevents characterization of the pattern of marks across the length of a single CTD. While most marks seem uniform across the CTD (Schüller et al., 2016; Suh et al., 2016), is this representative of each CTD molecule, or an average across millions of CTDs? To characterize the pattern of CTD modifications across the entire length of the CTD, a single molecule approach is required.

One approach may be the use of nanopores to directly sequence the CTD. Originally, nanopore technology was used to sequence DNA (Cherf et al., 2012). Briefly, α-hemolysin nanopores are loaded into an electrically resistant synthetic membrane, and an electric current is applied through the nanopore. Phi29 DNA polymerase feeds DNA through the nanopore, and unique changes in the electric current corresponding to each base are measured. Similarly, protein sequence can be determined. Proteins are conjugated to a leader oligo that direct the protein through the nanopore, and ClpX, a translocase, unfolds the protein as it is fed through the α-hemolysin nanopore (Nivala et al., 2013; Rodriguez-Larrea and Bayley, 2013). One difficulty with sequencing proteins is the requirement to distinguish 20 unique electrical signatures corresponding to each amino acid. The complexity is further increased when examining phosphorylation of specific residues (Rosen et al., 2014). Finally, in order to truly analyze single molecules, complex nanofluidics and discrete electrodes are required for every nanopore.

To eliminate these shortcomings, we propose an optical approach in which phosphorylated serine and threonine residues are substituted with fluorophores. Specifically, a variant of Pol II was generated containing a TEV protease site immediately preceding the CTD, and a C-terminal His tag (Figure 4.6B). The CTD lacks lysines and “barcodes” previously required to map phosphorylation marks (Schüller et al., 2016; Suh et al., 2016). Cells are lysed, and Pol II is bound to magnetic Ni-NTA beads via the His-tag. TEV cleavage and harsh washes leave only the CTD bound. Next, well-established β-elimination of phosphoserine and phosphothreonine residues under basic conditions generates dehydroalanine or β-methyldehydroalaline, respectively (Byford, 1991). Subsequently, dehydroalanine and β-methyldehydroalaline act as Michael acceptors for cysteamine, forming aminoethylcysteine or β-methylaminoethylcysteine (Knight et al., 2003) (Figure 4.6A, 4.6B). These residues contain primary amines, which readily react with a Cy5 NHS ester. Thus, each phosphoserine and phosphothreonine along the CTD is substituted with a fluorophore.

To map the locations of the fluorophores, we utilize Förster resonance energy transfer (FRET). ClpX, the translocase used in nanopore technology, is immobilized on a surface, rather than forming a nanopore through the surface. At the exit tunnel of ClpX, a Cy3 donor dye is placed. As the CTD peptides are translocated through, FRET between Cy3 and Cy5 will indicate the presence of a fluorophore (Yao et al., 2015) (Figure 4.6B). This technology allows parallel analysis of thousands of CTD peptides, and can easily be performed by labs capable of single-molecule detection. With this approach, we will 1) precisely identify the phosphorylated serines and threonines along many CTD peptides, 2) determine if marks are stochastically placed, or if there is order, and 3) characterize specific patterns that are represented across many CTDs.

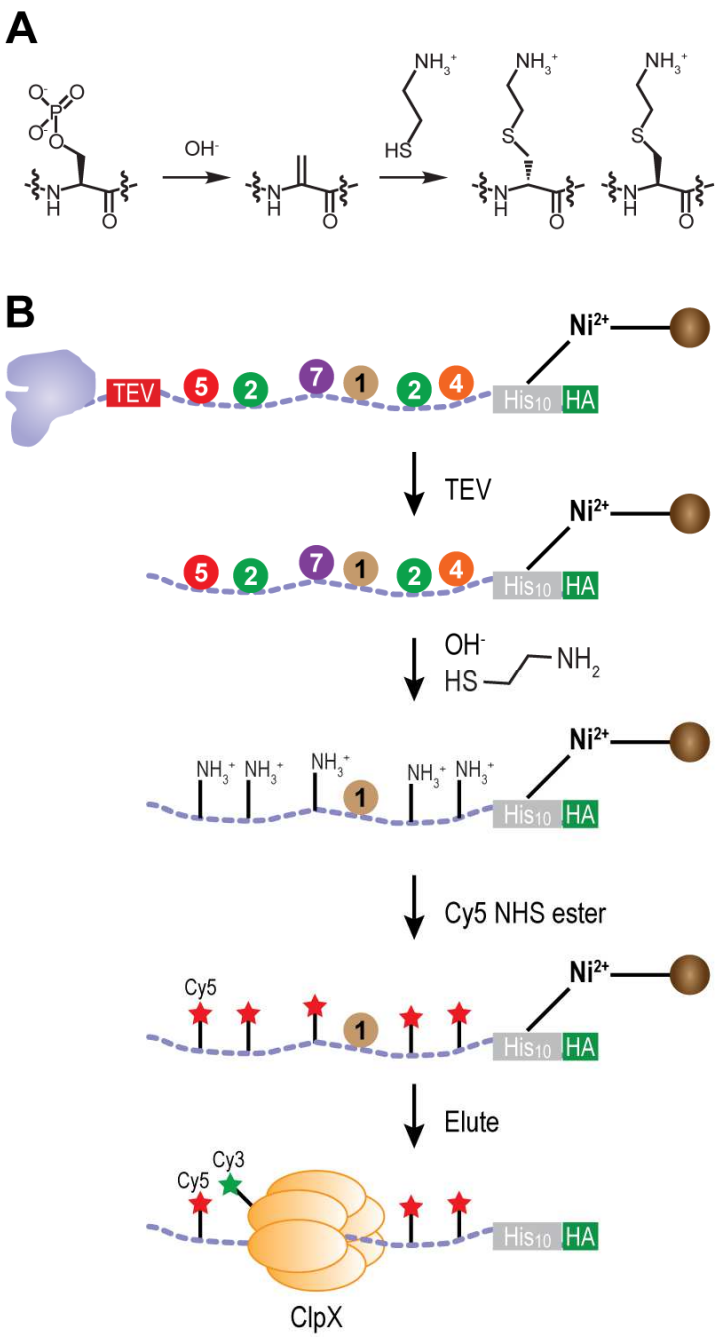

Figure 4.6 Single molecule protein sequencing

A) β-elimination of phosphoserine to dehydroalanine, followed by Michael Addition of cysteamine to form aminoethylcysteine. Figure adapted from (Knight et al., 2003).

B) RNA Pol II containing a TEV protease site immediately preceding the CTD with a C-terminal His tag is bound to Ni-NTA magnetic beads. The core polymerase is removed via TEV cleavage, and phosphoserines and phosphothreonines are converted to aminoethylcysteine or β-methylaminoethylcysteine. Primary amines are labeled with Cy5 and passed through a protein translocase (ClpX) containing a Cy3 donor fluorophore. Originally phosphorylated residues are identified via FRET on individual CTD molecules.

This pipeline can easily be expanded to answer a variety of questions. Inhibition of specific kinases via an analog sensitive approach (Chapter 2) will reveal how the pattern of marks across the CTD is controlled by specific kinases. Further, instead of analyzing the CTD patterns in bulk, the state of CTD phosphorylation throughout the transcription cycle can be discerned by immunoprecipitating factors known to associate with Pol II at various stages (e.g. Mediator during pre-initiation complex formation, Ceg1 at promoter escape, and Rtt103 at termination). Co-immunoprecipitated Pol II can then be processed as described above.

4.3 Spatial and temporal resolution of transcription dynamics during ESR

Many modern techniques including ChIP-seq, RNA-seq, and label free quantitative proteomics require material from millions of cells to obtain a snapshot of the "average" state of a cell. Are data averaged from millions of cells indicative of all cells in the population…or any cell for that matter? In fact, even genetically identical populations can display heterogeneous populations of RNA and protein in each cell (Altschuler and Wu, 2010; Chan et al., 2016; Chang et al., 2008; Klein et al., 2015; Singh et al., 2010). Analogous to single-molecule protein sequencing described above, many labs are moving toward ChIP-seq, RNA-seq, and proteomic analysis using single cells (Klein et al., 2015; Lombard-Banek et al., 2016; Rotem et al., 2015; Treutlein et al., 2016). A surprising amount of heterogeneity between single cells has been found, and therefore, we propose to examine transcription dynamics in single cells.

In Chapter 3, we uncovered a new branch of regulation that cells utilize in response to environmental stresses. Briefly, receptors on the cell surface sense environmental stress and initiate a signaling cascade of MAP kinase phosphorylation (Causton et al., 2001; Gasch, 2007; Gasch et al., 2000). This cascade centers on the Hog1 kinase, which subsequently phosphorylates the Pol II CTD, transcription factors (Msn2/4, Hot1, Sko1), and chromatin

remodelers (Rpd3) to dictate the subset of genes to be up- or down-regulated in order to respond to the stress (Alejandro-Osorio et al., 2009; Beck and Hall, 1999; Gasch and Werner-Washburne, 2002; Gasch et al., 2000; Nemec et al., in preparation) (see Figure 3.1). We showed that kinases in this signaling cascade can directly phosphorylate the Pol II CTD leading to rapid remodeling of the transcriptome.

How does this rapid signaling function at the single-cell level? Within 2 minutes of osmotic stress, Hog1 is phosphorylated by Pbs2, which causes Hog1 to localize to the nucleus (Ferrigno et al., 1998). How rapidly, and in what order does Hog1 phosphorylate transcription factors, chromatin remodelers, and Pol II? How important is the activity of upstream kinases in the signaling cascade? Do alanine substitutions to the CTD alter these dynamics, or is the inability to be phosphorylated the sole mechanism? To answer all these questions, we can use microfluidics to finely control the environment of singe cells, and use fluorescence microscopy to deconvolute the spatial and temporal nuances of the stress response in individual cells.

How critical are upstream kinases on Hog1-dependent stress response? Various factors involved in response to stress have previously been knocked out (Chasman et al., 2014), but cells can compensate for deleted proteins. Further, deleting proteins may disrupt full complexes leading to substantial defects. To characterize the complex interplay between factors in the stress response network, we will utilize the analog sensitive approach described in Chapter 2. Cells of the parent strain harboring Hog1-RFP and Msn2-GFP will be trapped and grown in a microfluidic platform. Individual cells can be stressed with different stressors, of different concentrations, for different lengths of time, and the localization of Hog1 and Msn2 will be monitored by time-lapse microscopy. Next, analog-sensitive variants of each kinase in Figure 3.1 will be created in this parent strain. Individual kinases will be rapidly and selectively inhibited prior to exposure to similar stress conditions as our control set up. We will be able to determine

1) if the stress response is equally affected in all cells, 2) if any kinases are more important than others in properly responding to stress, and 3) if the diminished response to stress is due to defects in temporal or spatial control.

We previously identified that the ability of Hog1 to phosphorylate the CTD of Pol II is critical for proper response to stress (Chapter 3), but what is the mechanism of action? Interestingly, the Pol II in cells harboring a T4A mutant CTD is largely localized outside the nucleus (Figure 4.7A). Further, in WT cells, while the distribution of pSe2 and pSer5 marks is nuclear, pThr4 is distributed throughout the cell (Figure 4.7B). Does Pol II localization itself help control the cell's response to stress? Cells harboring Hog1-RFP and Rpb3-GFP with various CTD mutants will be stressed as above. We will be able to determine if the localization of Pol II mutants is controlled in a Hog1-dependent manner and if Pol II localization is an important regulatory mechanism for proper response to stress.

4.4 Conclusions

Extraordinarily rapid advances in understanding the Pol II CTD code have occurred in recent years. We envision further work on the human CTD, which contains many non-canonical residues capable of being modified. Further, exploring the functional linkage between CTD modification and histone modification will resolve many long withstanding questions.

In Chapters 2 and 3, we identify novel kinases that phosphorylate the repetitive RNA Pol II CTD and characterize the roles these marks play in transcription termination and the environmental stress response. Due to the repetitive nature of the CTD, the pattern of phosphorylation along the CTD remains unclear. Approaches to "sequence" single CTD molecules may decipher this code. These methods do not rely on modifying or fragmenting the CTD and will definitively determine the pattern of modifications across the length of the CTD. Finally, using microfluidics

and single cell microscopy, the role of CTD modifications in the complex dynamics of the environmental stress response will be examined.

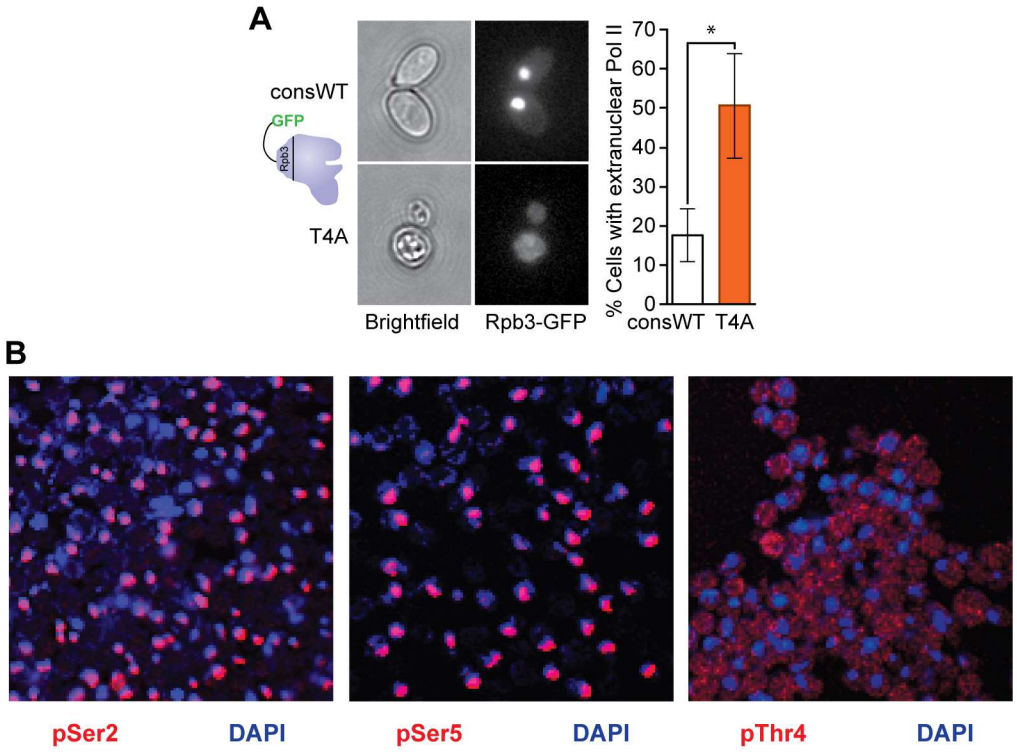

Figure 4.7 RNA Pol II localization in T4A mutants

A) The Rpb3 subunit in strains bearing WT or T4A CTDs was tagged with a C-terminal GFP. Shown are representative cells via brightfield or GFP microscopy. Significantly more cells in the T4A population display Pol II outside the nucleus (p<0.05, Student's T-test, n=3).

B) Immunofluorescence of pSer2 (Bethyl 1:500), pSer5 (3E8 1:200), or pThr4 (6D7 1:20) (red) co-localized with DAPI (1:500) (blue).

4.5 References

Alejandro-Osorio, A., Huebert, D., Porcaro, D., Sonntag, M., Nillasithanukroh, S., Will, J., and Gasch, A. (2009). The histone deacetylase Rpd3p is required for transient changes in genomic expression in response to stress. Genome Biology *10*, R57.

Altschuler, S.J., and Wu, L.F. (2010). Cellular heterogeneity: do differences make a difference? Cell *141*, 559-563.

Baskaran, R., Escobar, S.R., and Wang, J.Y.J. (1999). Nuclear c-Abl is a COOH-terminal repeated domain (CTD)-tyrosine kinase-specific for the mammalian RNA polymerase II: Possible role in transcription elongation. Cell Growth & Differentiation *10*, 387-396.

Beck, T., and Hall, M.N. (1999). The TOR signalling pathway controls nuclear localization of nutrient-regulated transcription factors. Nature *402*, 689-692.

Byford, M.F. (1991). Rapid and selective modification of phosphoserine residues catalysed by Ba2+ ions for their detection during peptide microsequencing. Biochem J *280*, 261-265.

Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S., and Young, R.A. (2001). Remodeling of yeast genome expression in response to environmental changes. Molecular Biology of the Cell *12*, 323-337.

Chan, S.S.-K., Chan, H.H.W., and Kyba, M. (2016). Heterogeneity of Mesp1+ mesoderm revealed by single-cell RNA-seq. Biochem Biophys Res Commun *474*, 469-475.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544-547.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA Polymerase II Is Phosphorylated at CTD Residue Serine-7. Science *318*, 1780-1782.

Chasman, D., Ho, Y.H., Berry, D.B., Nemec, C.M., MacGilvray, M.E., Hose, J., Merrill, A.E., Lee, M.V., Will, J.L., Coon, J.J.*, et al.* (2014). Pathway connectivity and signaling coordination in the yeast stress-activated signaling network. Mol Syst Biol *10*, 759.

Cherf, G.M., Lieberman, K.R., Rashid, H., Lam, C.E., Karplus, K., and Akeson, M. (2012). Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. Nat Biotech *30*, 344-348.

Dias, J.D., Rito, T., Torlai Triglia, E., Kukalev, A., Ferrai, C., Chotalia, M., Brookes, E., Kimura, H., and Pombo, A. (2015). Methylation of RNA polymerase II non-consensus Lysine residues marks early transcription in mammalian cells. eLife.

Ferrigno, P., Posas, F., Koepp, D., Saito, H., and Silver, P.A. (1998). Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin beta homologs NMD5 and XPO1. EMBO J *17*, 5606-5614.

Fortschegger, K., de Graaf, P., Outchkourov, N.S., van Schaik, F.M.A., Timmers, H.T.M., and Shiekhattar, R. (2010). PHF8 Targets Histone Methylation and RNA Polymerase II To Activate Transcription. Mol Cell Biol *30*, 3286-3298.

Gasch, A., and Werner-Washburne, M. (2002). The genomics of yeast responses to environmental stress and starvation. Functional & Integrative Genomics *2*, 181-192.

Gasch, A.P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. Yeast *24*, 961-976.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Molecular Biology of the Cell *11*, 4241-4257.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E.*, et al.* (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J *31*, 2784-2797.

Jin, Q., Yu, L.-R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.-E., Wang, C., Brindle, P.K., Dent, S.Y.R., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J *30*, 249-262.

Kelly, W.G., Dahmus, M., and Hart, G. (1993). RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. J Biol Chem *268*, 10416-10424.

Kim, T., and Buratowski, S. (2009). Dimethylation of H3K4 by Set1 Recruits the Set3 Histone Deacetylase Complex to 5' Transcribed Regions. Cell *137*, 259-272.

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A Novel Domain in Set2 Mediates RNA Polymerase II Interaction and Couples Histone H3 K36 Methylation with Transcript Elongation. Mol Cell Biol *25*, 3305-3316.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187-1201.

Knight, Z.A., Schilling, B., Row, R.H., Kenski, D.M., Gibson, B.W., and Shokat, K.M. (2003). Phosphospecific proteolysis for mapping sites of protein phosphorylation. Nat Biotechnol *21*, 1047-1054.

Lee, J.-H., and Skalnik, D.G. (2008). Wdr82 Is a C-Terminal Domain-Binding Protein That Recruits the Setd1A Histone H3-Lys4 Methyltransferase Complex to Transcription Start Sites of Transcribed Human Genes. Mol Cell Biol *28*, 609-618.

Li, H., Zhang, Z., Wang, B., Zhang, J., Zhao, Y., and Jin, Y. (2007). Wwp2-Mediated Ubiquitination of the RNA Polymerase II Large Subunit in Mouse Embryonic Pluripotent Stem Cells. Mol Cell Biol *27*, 5296-5305.

Lombard-Banek, C., Reddy, S., Moody, S.A., and Nemes, P. (2016). Label-free Quantification of Proteins in Single Embryonic Cells with Neural Fate in the Cleavage-Stage Frog (Xenopus laevis) Embryo using CE-ESI-HRMS. Molecular & Cellular Proteomics.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. Science *336*, 1723-1725.

Miao, F., Li, S., Chavez, V., Lanting, L., and Natarajan, R. (2006). Coactivator-Associated Arginine Methyltransferase-1 Enhances Nuclear Factor-κB-Mediated Gene Transcription through Methylation of Histone H3 at Arginine 17. Mol Endocrinol *20*, 1562-1573.

Nemec, C.M., Gilmore, J.M., Ho, Y., Hintermair, C., Singh, A.K., Ringelberg, K.J., Tseng, S.C., Heidemann, M., Zhang, Y., Florens, L.*, et al.* (in preparation). Non-canonical CTD-kinases regulate gene-class specific functions of RNA polymerase II.

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

Nivala, J., Marks, D.B., and Akeson, M. (2013). Unfoldase-mediated protein translocation through an [alpha]-hemolysin nanopore. Nat Biotech *31*, 247-250.

Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W., and Lewis, B.A. (2012). Evidence of the Involvement of O-GlcNAc-modified Human RNA Polymerase II CTD in Transcription in Vitro and in Vivo. J Biol Chem *287*, 23549-23561.

Rodriguez-Larrea, D., and Bayley, H. (2013). Multistep protein unfolding during nanopore translocation. Nat Nano *8*, 288-295.

Rosen, C.B., Rodriguez-Larrea, D., and Bayley, H. (2014). Single-molecule site-specific detection of protein phosphorylation with a nanopore. Nat Biotech *32*, 179-181.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat Biotech *33*, 1165-1172.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.-M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol Cell *61*, 305-314.

Simonti, C.N., Pollard, K.S., Schroder, S., He, D., Bruneau, B.G., Ott, M., and Capra, J.A. (2015). Evolution of lysine acetylation in the RNA polymerase II C-terminal domain. BMC Evol Biol *15*, 327.

Sims, R.J., Rojas, L.A., Beck, D., Bonasio, R., Schüller, R., Drury, W.J., Eick, D., and Reinberg, D. (2011). The C-Terminal Domain of RNA Polymerase II Is Modified by Site-Specific Methylation. Science *332*, 99-103.

Singh, D.K., Ku, C.-J., Wichaidit, C., Steininger, R.J., Wu, L.F., and Altschuler, S.J. (2010). Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. Mol Syst Biol *6*, 369-369.

Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Mol Cell *61*, 297-304.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M.*, et al.* (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391-395.

Vojnic, E., Simon, B., Strahl, B.D., Sattler, M., and Cramer, P. (2006). Structure and Carboxyl-terminal Domain (CTD) Binding of the Set2 SRI Domain That Couples Histone H3 Lys36 Methylation to Transcription. J Biol Chem *281*, 13-15.

Voss, K., Forné, I., Descostes, N., Hintermair, C., Schüller, R., Maqbool, M.A., Heidemann, M., Flatley, A., Imhof, A., Gut, M.*, et al.* (2015). Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. Transcription, 00-00.

West, M.L., and Corden, J.L. (1995). Construction and Analysis of Yeast RNA Polymerase II CTD Deletion and Substitution Mutations. Genetics *140*, 1223-1233.

Yanling Zhao, D., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F.W., Zhong, G., Liu, K., Li, W., Moffat, J.*, et al.* (2016). SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. Nature *529*, 48-53.

Yao, Y., Margreet, D., Jetty van, G., Dick de, R., and Chirlmin, J. (2015). Single-molecule protein sequencing through fingerprinting: computational assessment. Physical Biology *12*, 055003.

Appendix A: Novel peptide arrays identify targets and binding preferences of CTD "writers" and "readers"

This work was performed in collaboration with Eric Sullivan and Jigar Patel at Roche/NimbleGen in Madison, WI. Anna Kropornicka optimized peptide sequences prior to synthesis.

A.1 Introduction

The carboxyl-terminal domain of RNA polymerase II consists of the repeating heptapeptides ($Y_1S_2P_3T_4S_5P_6S_7$) (Corden et al., 1985). In yeast, within each of the 26 repeats, the tyrosine, serines, and threonine can be reversibly phosphorylated, and the prolines can undergo cis/trans isomerization (Buratowski, 2009; Egloff and Murphy, 2008; Eick and Geyer, 2013; Phatnani and Greenleaf, 2006; Zhang et al., 2012). These dynamic modifications create a slew of unique scaffolds that recruit RNA-processing machinery co-transcriptionally. An estimated 10^{59} unique modification patterns could occur on the CTD, but does the CTD adopt all of these states? Which combinations of marks are critical for proper recruitment of specific factors? Which combination of marks inhibits recruitment of factors? Which marks “prime” specific kinases for subsequent phosphorylation? Do other cellular machines aid in recruiting factors to Pol II? To begin to address these questions, we collaborated with Roche/NimbleGen to create microarrays displaying over 2 million peptides and tested kinase activity on these peptide variants.

A.2 Array design

Peptide arrays allow for high throughput determination of protein:protein interactions. In collaboration with Roche/NimbleGen, we have synthesized nearly 2.1 million peptides (each up to 19 amino acids long) with a maskless array synthesizer (MAS) that uses light-directed activation for spatially resolved chemical coupling. Previous peptide array technology was limited to several hundred hand-spotted features on nitrocellulose (Smith et al., 2011). Our high-density microarrays are blocked off into 12 identical sub-arrays, each with 170,000 unique probes, allowing multiple experiments to be performed in parallel. Five repeats of 6-aminohexanoic acid “linker” are initially synthesized on a synthetic surface to elevate the peptides and minimize any surface-based mass transport effects that may arise. Peptides are

chemically synthesized from these moieties one amino acid at a time using a maskless array synthesizer (Singh-Gasson et al., 1999).

The displayed peptides include protein complexes involved in transcription initiation, elongation, termination, and splicing. In addition, known phosphorylation sites in the proteome (Ficarro et al., 2002) and CTD sequences across species from multiple kingdoms (Liu et al., 2010) were tiled on the array (Table A.1) (tiling method described below).

Furthermore, we designed the microarray to display over 8,000 CTD peptides in which each modifiable site is substituted with a phospho-mimic, a non-phosphorable residue, or a residue found at that position in all recently reported CTD di-heptad repeats (Liu et al., 2010) (Figure A.1). To reduce the combinatorial complexity of sequences, and to adhere to biologically relevant sequences, only CTD peptides containing two or fewer deviations from the WT in each repeat are displayed on the microarray. This is in agreement with recent mass spectrum analyses that observed only mono- or di-phosphorylation on each CTD repeat (Schüller et al., 2016; Suh et al., 2016).

Table A.1 Proteins displayed on peptide arrays

S. cerevisiae	***H. sapiens***
Histones (including Htz1 and centromeric)	Histones
Mediator subunits	HOX factors
Pol I subunits	bZIPs
Pol II subunits	Nuclear receptors
Pol III subunits	Pol I subunits
THO/TREX	Pol II subunits
TREX2	Pol III subunits
Swi/Snf complex	MLL complex
RSC	PRC1
CFIA	PRC2
CFIB	ESC-specific transcription factors
APT	Miscellaneous transcription factors
CPF	Bromodomains
TRAMP	Chromodomains
PAF1 complex	Methyltransferases
SAGA	Acetyltransferases
TFIIA, B, D, E, F, H, S	TALEs
SKI	
COMPASS	
Sm ribonucleoproteins	
U1, U2, U4/U6, U5 snRNPs	
Prp19C complex, B^{act} Complex, C complex	
Rpd3C	
Set3C	
CTD kinases	
CTD phosphatases	
Elongation factors	
Termination factors	

Peptides tiling the full length proteins listed were synthesized in triplicate on each of 12 sub-arrays.

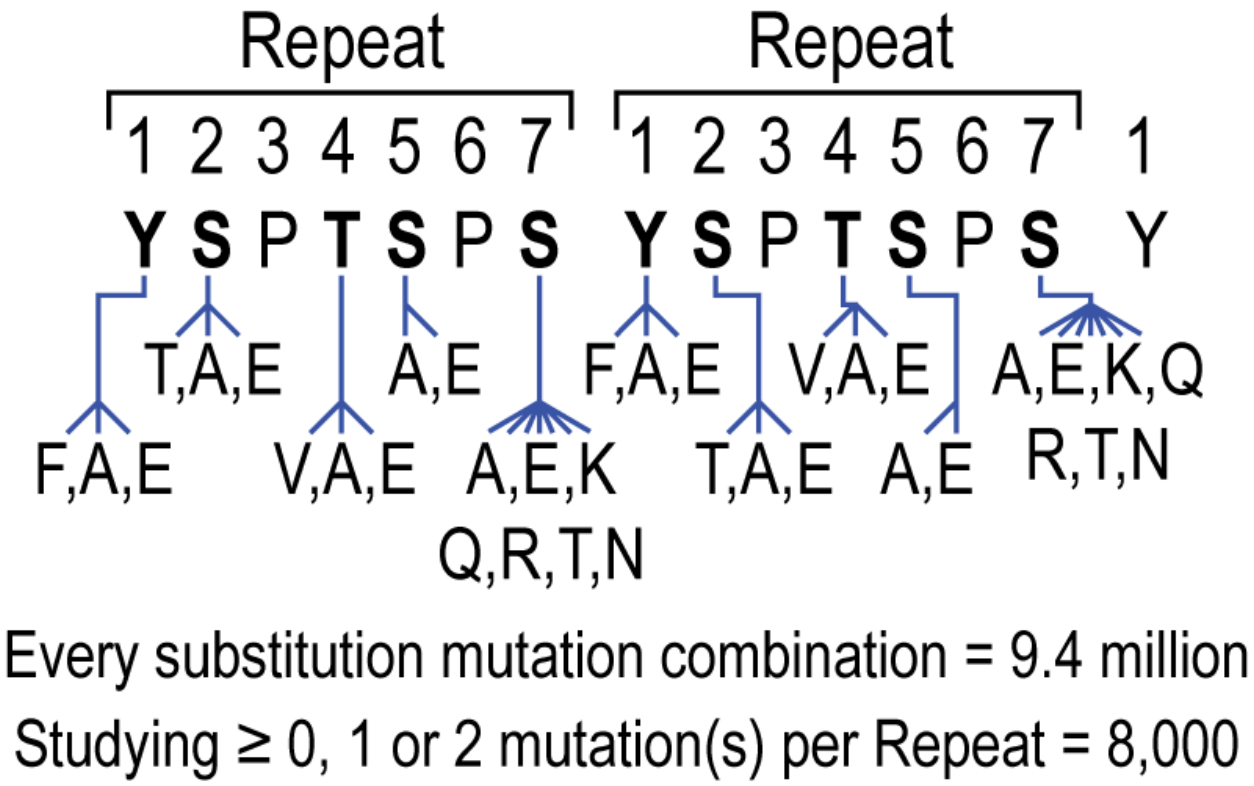

Figure A.1 Combinatorial CTD library
15 amino acid peptides displaying CTD mutants are displayed on the array. WT CTD is listed below each residue number. Substitutions at each residue (phosphomimics, non-phosphorable mimics, and naturally occurring residues) are listed below the WT CTD.

A.3 Results

The combinatorial CTD library portion of the peptide microarray allowed us to determine if the modification of specific residues is required for subsequent modifications to occur, or if specific marks block further modification. As a proof of principle, Kin28 was purified and incubated with the microarray in the presence of ATP (Figure A.2). Initially, we attempted to ascertain phosphorylation state using phospho-CTD specific antibodies. Unfortunately, the fluorescent rat secondary antibody was unable to bind to phospho-CTD specific primary antibodies raised in rats (Figure A.3A). Dye conjugation to these antibodies is out of the question, as the antibodies were not purified from hybridoma medium. However, we will be able to fluorescently label affinity purified antibodies in future experiments (Figure A.3B).

Interestingly, an α-mouse secondary bound well to 8WG16, an antibody that recognizes hypophosphorylated CTD. Kin28 was Flag-tagged, affinity purified, and incubated with the microarray in the presence of ATP. Median 8WG16 intensities among triplicate probes on arrays without Kin28 (Mock) vs. arrays with Kin28 bound (Figure A.4A). A strong correlation (Pearson R=0.97) was observed, but a subset of probes showed substantial deviation from the trend line. We hypothesized that when Kin28 phosphorylated these probes, the 8WG16 antibody could no longer bind (since it recognizes hypophosphorylated CTD), thus reducing the signal.

To identify significant probes, we excluded all probes with intensity less than 10,000, which could be attributed to noise. Using a 2-fold cutoff (Mock/Kin28) with $p<0.05$ among triplicates, we identified the features bound by the kinase and found that the MEME-derived motif (Bailey et al., 2009) was the peptide YSPTSPSYSPTSP**K**Y (Figure A.4B). Interestingly, a non-canonical, but evolutionarily conserved lysine at position 7 appears in the motif. This is consistent with the known motif of other cyclin-dependent kinases (T/S)*PX(R/K) (Morgan, 1997), where the asterisk indicates phosphorylation and X denotes any amino acid.

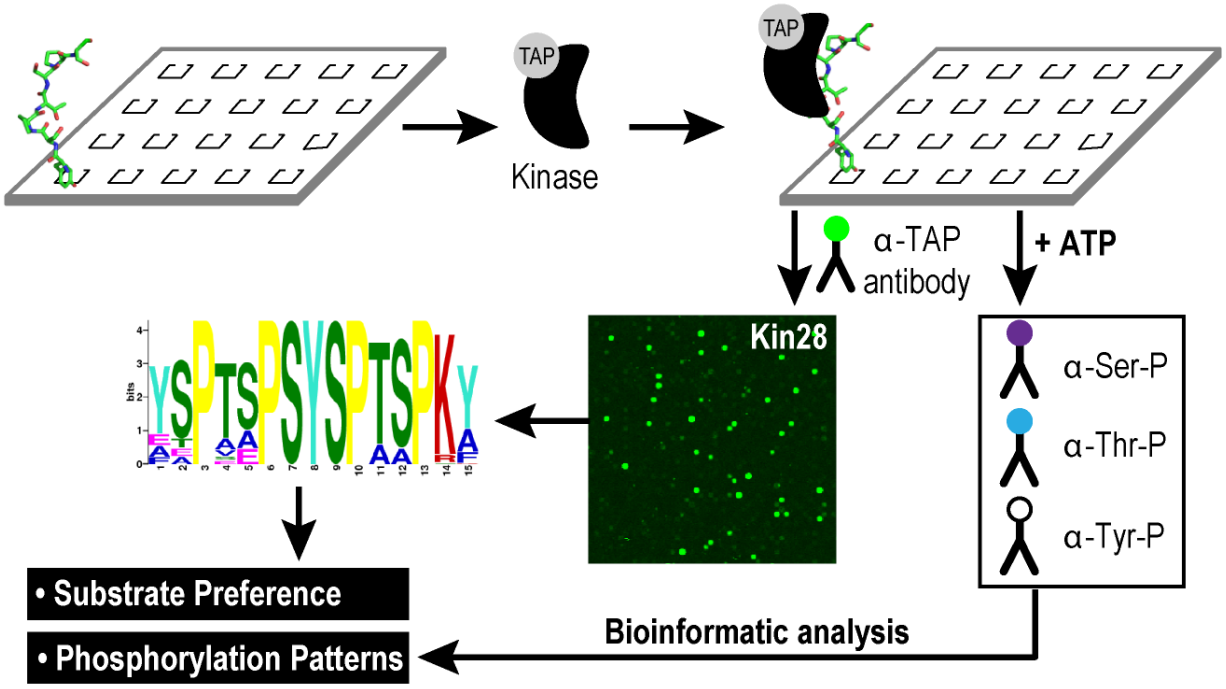

Figure A.2 Peptide array workflow
Peptide arrays displaying transcriptional machinery and CTD peptides were tiled every 7 amino acids. Arrays were incubated with kinase, with or without ATP. In the presence of ATP, phosphorylation is assayed using biotinylated phospho-specific primary antibodies and streptavidin-Cy3. In the absence of ATP, kinase binding is assayed using a Cy3 anti-TAP antibody. Arrays were scanned at 2 micron resolution using a NimbleGen MS 200 scanner. MEME (Bailey et al., 2009) was used to identify the motif that Kin28 binds.

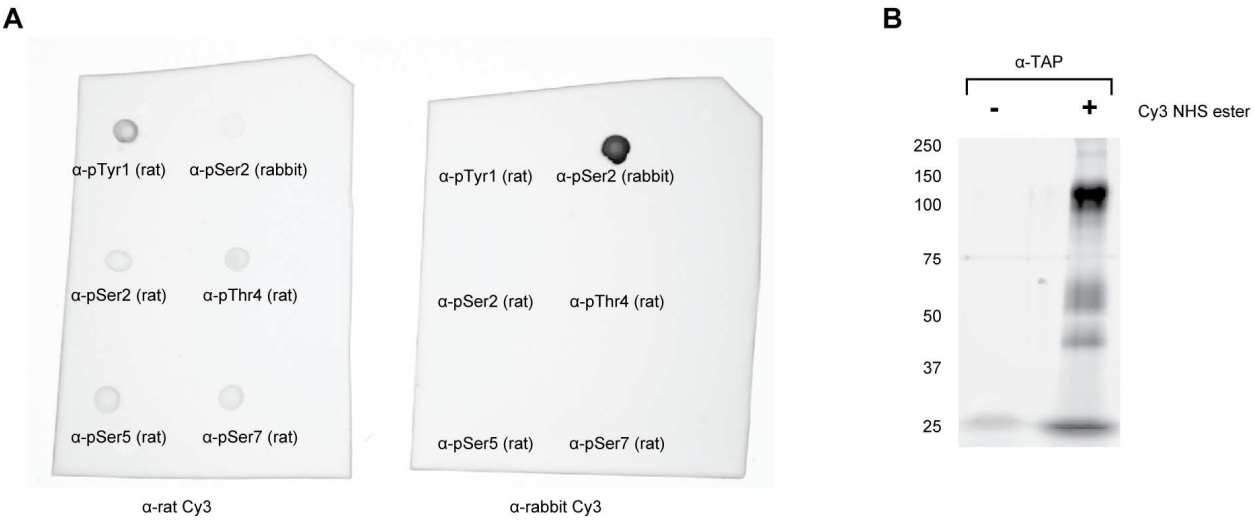

Figure A.3 Fluorescently labeled antibodies

A) Antibodies targeting specific CTD marks, raised in the indicated species, were spotted on nitrocellulose. Cy3 labeled secondary antibodies targeting either rat (left) or rabbit (right) were incubated, and blots were scanned. α-Rat-Cy3 binds poorly to phospho-CTD antibodies.

B) α-TAP antibodies were incubated with Cy3 NHS ester and purified as described below. Antibodies were denatured in sample buffer lacking DTT to prevent the disulfide bonds from reducing. Gels were scanned with an ImageQuant LAS 4000 scanner.

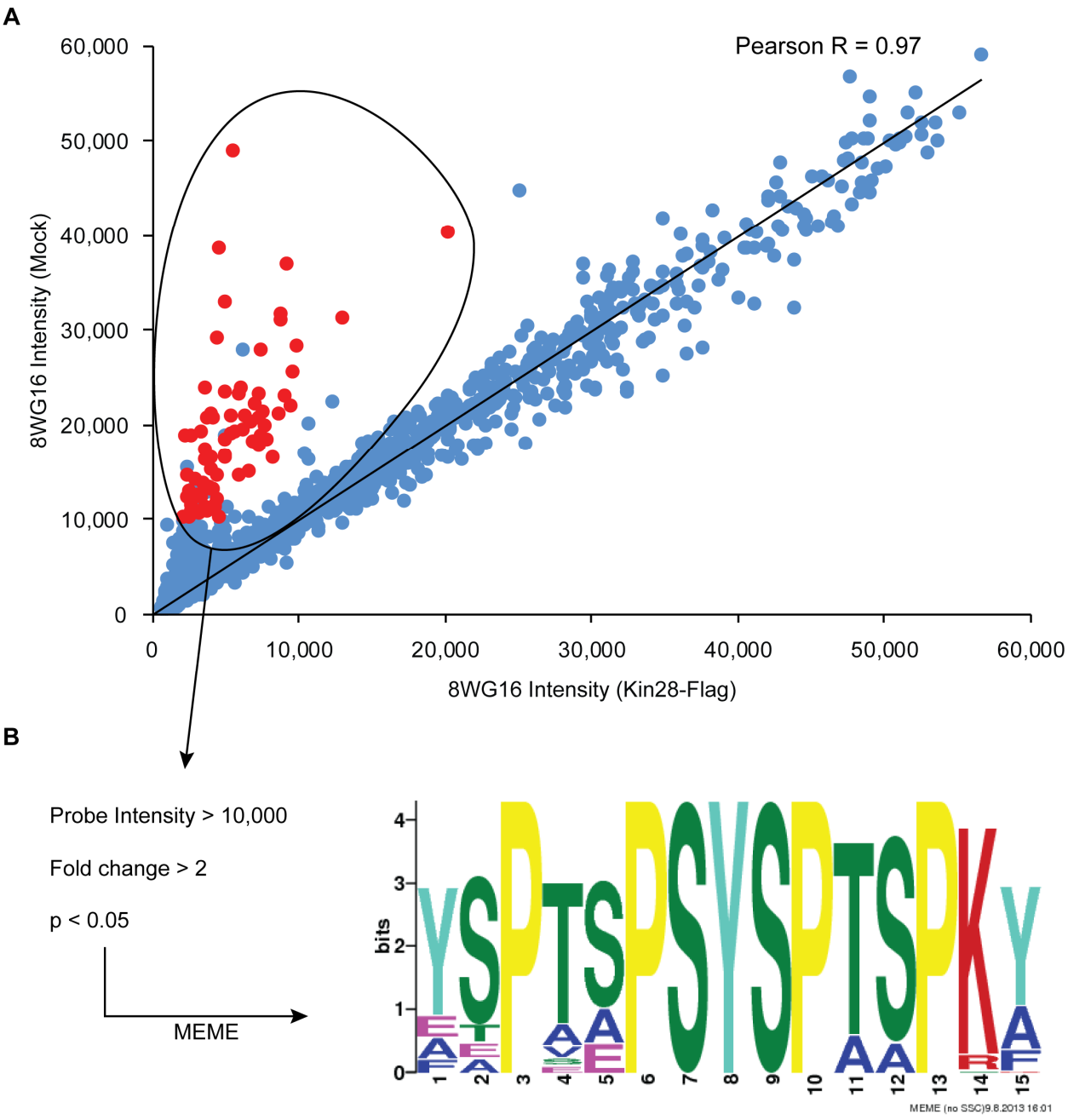

Figure A.4 Kin28 phosphorylates canonical CDK motifs

A) Scatterplot of fluorescence intensities of secondary antibodies bound to 8WG16 in the absence (Mock) or presence of Kin28.

B) Red probes from A) were identified as having >10,000 fluorescence units with at least 2-fold change (Mock vs Kin28-Flag) with a *p*-value among triplicate probes <0.05. A MEME-derived motif of probes is shown.

A.4 Future directions

In future experiments, phosphorylation state will be monitored with fluorescently labeled antibodies targeting either phosphoserine, -threonine, or -tyrosine (Sigma-Aldrich PSR-45, PTR-8, and PT-66, respectively (Figure A.2, A.3B). These antibodies bind their targets with virtually no dependence on modifications of flanking residues. Thus, they overcome the serious limitations and idiosyncratic biases of current α-phospho-CTD antibodies. We will also incubate each antibody with the microarray in the absence of purified kinase in order to control for peptides without a phosphate that may be recognized by these antibodies. Because the antibodies are not specific for a single protein, we will be able to identify all newly placed phosphates by any kinase by subtracting the intensity of the signal from our control arrays lacking kinase. We will be able to determine the prior marks that permit kinase binding and perhaps prime their activity, and which marks are detrimental to the activity of each kinase. Moreover, since much of the proteome is tiled on the array, we will be able to validate the direct targets of each kinase and likely identify new targets as well. Co-immunoprecipitations and yeast two-hybrid asays will be used to validate newly identified targets.

A.5 Scripts to optimize probe synthesis

Probes were synthesized via maskless array synthesis, similar to well-established DNA synthesis on arrays. Probes were optimized to be as long as possible for two reasons. First, increasing the probe length reduces the number of peptides needed to tile a full protein, which allows more unique peptides to be displayed. Second, in our experience, lengthening a substrate dramatically increases the ability of some kinases to phosphorylate their targets.

To synthesize each peptide, each amino acid is flowed 8 times(160 total couplings) in the following order: A, R, N, D, C, E, Q, G, H, I, L, K, M, F, P, S, T, W, Y, V. Probes were optimized

to be as long as possible, while still complying with Roche/NimbleGen's synthesis parameters. For instance, Nrd1, a *S. cerevisiae* protein involved in termination, contains a poly-glutamine region. Thus, one of the shortest peptides displayed on the array is QQQQQQQQ (1 Q for each of the 8 cycles). Conversely, we were able to synthesize a 19 amino acid probe from Rpd3 (MVYEATPFDPITVKPSDKR) while complying with the synthesis paramaters. The script to determine if a peptide can be synthesized within the required 160 couplings is as follows:

```
#!/usr/bin/perl


#Ryan Bannen Note: this next line will need to be modified once you
# get the calculatePcycles code installed:
use lib qw(/home/richmont/lib64/perl5 /home/richmont/lib64/perl5/site_perl);
use strict;
use Getopt::Long;
use File::Basename;
use NimbleGen::CalculatePCycles qw(calculate_pcycles);

my @amino_acids =
("A","R","N","D","C","E","Q","G","H","I","L","K","M","F","P","S","T","W","Y","V");

my $min_length = 9;
my $target_cycles = 160;
my $subX = 0;
my $input_file;
my $output_file;

###############################################################################
#
# usage information
###############################################################################
#
my ($name,$path,undef) = fileparse($0);

if ($#ARGV < 1 ) {
 usage();
 exit(0);
}

sub usage {
 print qq{
usage: $name
 --input input file_name
 --output output file name
 [--min_length minimum peptide length (default = $min_length) ]
```

```
 [--cycles maximum number of synthesis cycles (default = $target_cycles) ]
 [--subX substitute A for X ]

};
}

GetOptions(
 'input=s' => \$input_file,
 'output=s' => \$output_file,
 'min_length=i' => \$min_length,
 'cycles=i' => \$target_cycles,
 'subX!' => \$subX,
);

open(INPUT,$input_file) or die "Can't open intput file - $input_file -$!\n";

my $header_line = <INPUT>;
$header_line =~ s/\n|\r//g;
my @headers = split("\t",$header_line);

my %index = parse_headers(uc $header_line);
if(!exists($index('PEPTIDE_SEQUENCE'))) {
        print STDOUT "No PEPTIDE_SEQUENCE column found. Exiting...\n";
        exit(-1);
}

open(OUTPUT,'>'.$output_file) or die "Can't open output file - $output_file -$!\n";
print OUTPUT join("\t",@headers,"CYCLES"),"\n";
my $logfile = $output_file;
$logfile =~ s/\.txt//g;
$logfile .= '_truncation.log';
open(STDERR,'>' . $logfile ) or die "Can't open logfile $logfile - $!\n";
print STDERR join("\t",@headers,"FILTER_REASON"),"\n";

my $filtered_peptides = 0;
my $total_peptides = 0;
my $truncated_peptides = 0;
my $x_replacements = 0;

while (my $line = <INPUT>) {

        $line =~ s/\n|\r//g;
        my @v = split("\t",$line);

        my $peptide = $v[$index('PEPTIDE_SEQUENCE')];

        if(length($peptide) < $min_length ) {
                print STDERR join("\t",@v,'initial peptide shorter than minimum peptide
length'),"\n";
                $filtered_peptides++;
                next;
```

```
	}

	# filter out peptides with lower-case masked amino acids
	if($peptide =~ m/[acdefghiklmnpqrstvwy]/) {
		print STDERR join("\t",@v,sprintf('masked AAs:%s',$1)),"\n";
		$filtered_peptides++;
		next;
	}

	$peptide = uc $peptide;

	if($subX and $peptide =~ m/X/) {
		my $replacements = $peptide =~ tr/X/A/;
		$x_replacements += $replacements;
		print STDERR join("\t",@v,sprintf('substituted A for X:%d',$replacements)),"\n";
	}

	# Filter out peptides that aren't exclusively known AAs
	# B=aspartic acid or asparagine; Z=glutamine or glutamic acid; X=any amino acid; J=No
meaning; O=No meaning; U=No meaning
	if ($peptide =~ m/[^ACDEF{{GHIKLMNPQRSTVWY]/) {
		print STDERR join("\t",@v,sprintf('unknown AA:%s',$1)),"\n";
		$filtered_peptides++;
		next;
	}
	my ($trunc_peptide,$cycles) = target_cycles($peptide,$target_cycles,$min_length);
	if(length($trunc_peptide) < length($peptide)) {
		$truncated_peptides++;
		print STDERR join("\t",@v,sprintf('truncated:%s',$trunc_peptide)),"\n";
	}
	if($cycles > $target_cycles) {
		print STDERR join("\t",@v,sprintf('exceeded cycles:%d',$cycles)),"\n";
		$filtered_peptides++;
		next;
	}
	$total_peptides++;
	$v[$index('PEPTIDE_SEQUENCE')] = $trunc_peptide;
	print OUTPUT join("\t",@v,$cycles),"\n";

}
close OUTPUT;

print STDOUT "\n";
print STDOUT "Processed a total of $total_peptides peptides\n";
print STDOUT "Truncated a total of $truncated_peptides\n";
print STDOUT "Substituted $x_replacements As for Xs\n" if $subX;
print STDOUT "Filtered out $filtered_peptides peptides\n";

exit();
```

```
sub target_cycles {

	my $base_oligo = shift;
	my $target_cycles = shift;
	my $min_length = shift;

	my $probe_cycles = calculate_pcycles($base_oligo);
	if($probe_cycles <= $target_cycles) {
		return ($base_oligo,$probe_cycles);
	}

	my $temp_seq = $base_oligo;
	my $temp_seq_length = length($temp_seq);

	while($probe_cycles > $target_cycles and $temp_seq_length > $min_length ) {
		$temp_seq_length--;
		$temp_seq = substr($temp_seq,0,$temp_seq_length);
		$probe_cycles = calculate_pcycles($temp_seq);
	}
	return ($temp_seq,$probe_cycles);
}

sub parse_headers {

	my $line = shift;
	$line =~ s/\n|\r//g;

	my @headers = split("\t",$line);

	my %index;

	foreach my $idx (0 .. $#headers) {
		$index{$headers[$idx]} = $idx;
	}

	return %index;


}
```

To optimize probe sequence, Anna Kropornicka developed a python script. Protein sequences from UniProt or the Saccharomyces Genome Database (SGD) were collected in FASTA format, and sequences were reversed, since synthesis occurs C- to N-terminally. Roche/NimbleGen's script was used to determine if a given probe could be successfully synthesized within the required 160 couplings. If so, the next amino acid was added on and re-analyzed. This was repeated until the peptide could no longer be synthesized within the required 160 steps. The peptides are tiled across the protein with 7 amino acid overlap, thus, the next peptide begins 7 amino acids from the end of the previous, and the process is repeated. The code to generate these probes is as follows:

```
import sys, subprocess

f1 = open(sys.argv[1], 'r')
fout = open(sys.argv[2], 'w')
fout.write('PEPTIDE_SEQUENCE\tProtein\tSpecies\tEntry\n')

for i,line in enumerate(f1,start = 1):

 if '>' in line:
 #Grab next line
 searchseq = next(f1)
 searchseq = searchseq.replace('\n','')
 length = len(searchseq)
 line = line.replace('>','')
 line = line.replace('\n','')
 line = line.split(' ')
 oligo = 9
 end = length-1
 limit = end
 beg = end - oligo

 while beg >= 0:
 cont = True
 trouble = False
 while cont == True:
 if beg < 0:
 beg = 0
 end = beg+oligo
 else:
 pass
 probe = searchseq[beg:end]
```

```
 print probe
 ftemp = open('Temp_probe.txt','w')
 ftemp.write('PEPTIDE_SEQUENCE\n')
 ftemp.write(probe+'\n')
 ftemp.close()
 args = ['perl','truncate_peptides/truncate_peptides.pl','--input','Temp_probe.txt','--
output','output_file.txt','>','log.txt']
 #print(' '.join(args))
 perlscrip = subprocess.call(' '.join(args), shell = True)
 f3 = open('log.txt','r')
 seg = f3.readlines()
 #print seg
 if ('Truncated a total of 0\n' in seg[2]) and ('Filtered out 0 peptides\n' in seg[3]):
 if beg == 0:
 print('It\'s less than length',beg)
 cont = False
 elif (oligo == 8):
 cont = False
 else:
 oligo = oligo+1
 if oligo > 18:
 olig = 18
 cont = False
 else:
 pass
 beg = end - oligo
 #print(beg, oligo)
 elif (oligo <= 9) and (('Truncated a total of 1\n' in seg[2]) or ('Filtered out 1 peptides\n' in
seg[3])):
 beg = beg - 1
 end = beg + oligo
 #print(end, end-beg,searchseq[beg:end])
 if (end <= limit) and (oligo == 9):
 oligo = 8
 end = limit
 beg = end - oligo

 elif (beg<=0) and (trouble == False):
 beg = 0
 end = beg + oligo
 trouble = True
 elif (beg <=0) and (trouble == True):
 cont = False
 elif (end <= limit) and (oligo == 8):
 trouble = True
 cont = False
 else:
 continue
 elif (oligo != 9) and (('Truncated a total of 1\n' in seg[2]) or ('Filtered out 1 peptides\n' in seg[3])):
 oligo = oligo - 1
 beg = end - oligo
```

```
 probe = searchseq[beg:end]
 cont = False
 else:
 print('##############################There is a problem.##############################')

 f3.close()

 print('We got out of the loop')
 if (beg <= 0) and (trouble == False):
 print(beg)
 beg = 0
 probe = searchseq[beg:end]
 fout.write(probe+'\t'+'\t'.join(line)+'\t'+str(oligo-1)+'\n')
 oligo = 9
 break
 elif (trouble == True):
 fout.write(probe+'\t'+'\t'.join(line)+'\tWE HAVE AN ERROR HERE:'+ str(limit)+': '+str(beg)+'-
'+str(end)+'\n')
 oligo = 9
 beg = end - oligo
 limit = end
 else:
 fout.write(probe+'\t'+'\t'.join(line)+'\t'+str(oligo)+'\n')
 oligo = 9
 end = beg+7
 limit = end
 beg = end-oligo

 print( 'We got out of the larger loop')
 else:
 pass

fout.close()
f2 = open(sys.argv[2], 'r')
name = sys.argv[2].replace('.txt', '_uniq.txt')
fout = open(name, 'w')
uniqlist = []

for line in f2:
 line = line.split('\t')
 if line[0] in uniqlist:
 continue
 else:
 uniqlist.append(line[0])
 fout.write('\t'.join(line))

fout.close()
```

A.6 Experimental procedures

Kinase purification

Cells expressing C-terminally Flag-tagged Kin28 were grown in 200 mL selective media to OD_{600} of 1.0, and cells were pelleted and transferred to a 1.5 mL tube. Pellets were resuspended in 500 µL TAP buffer A (20 mM HEPES pH 7.9, 300 mM potassium acetate, 0.5 mM EDTA pH 8.0, 10% glycerol, 0.05% NP40), with 1 mM DTT, 1 mM PMSF, 1 mM benzamidine, 1.45 mM pepstatin, and 1 mM phosphatase inhibitors (NaF, NaN3, and $Na3VO_4$) freshly added. Approximately 200 µL silica beads (OPS Diagnostics) were added, and cells were lysed via bead beating for 15 minutes. A 22 gauge needle was heated with a Bunsen burner and used to puncture a hole in the tube, which was then placed in a 2 mL tube and spun at 7,000 RPM for 30 seconds. Extract was incubated with α-Flag antibody (M2 – Sigma) and bound to protein G Dynabeads (Thermo Fisher). Three washes were performed with TAP buffer A prior to elution with Flag peptide.

Optimization of kinase concentration

Arrays were blocked at 4°C overnight in blocking solution (1X TBS, 0.05% Tween-20, 1% alkali soluble casein), were washed twice in TBS for two minutes, and once in nuclease free water for 30 seconds. To validate that purified kinase does not non-specifically coat the microarray, a dilution series of purified Kin28 was incubated with the arrays. Kin28 was diluted 2.5 fold, 5 fold, and 14 fold in kinase buffer (40 mM HEPES pH 7.5, 20% glycerol, 5 mM EGTA, 30 mM MgOAc, 200 mM KOAc) and incubated with the array in triplicates for 3 hours at 30°C. A negative control lacking kinase was also incubated with the array. The array was washed in 1x TBS with 0.05% Tween for 30 minutes prior to a 30 second wash in nuclease free water. Arrays were spun dry. α-Flag antibody was diluted 1:400, 1:1000, or 1:2500 and incubated with the array overnight at 4°C. The array was washed for 30 minutes with 1X TBS with 0.05% Tween-20. α-mouse Cy3 antibody was diluted 1:10,000 in blocking buffer and was incubated for 3 hours at room

temperature. The array was washed in 1X TBS with 0.05% Tween-20 for 30 minutes and 1X TBS for 30s prior to being spun dry. Arrays were scanned at 2 micron resolution. 2.5 fold and 5 fold dilutions of Kin28 revealed background binding, whereas the 14 fold dilution did not, thus subsequent experiments with Kin28 were performed at this dilution.

On-array kinase assay

Each well of the 12 plex array holds 6 µL reaction. 0.43 µL Kin28 (14 fold dilution of stock) was incubated with 3 µL 2x buffer D (Ansari et al., 2005) (40 mM HEPES pH 7.5, 20% glycerol, 5 mM EGTA, 30 mM MgAc, 200 mM KAc) with 1 mM DTT, 1 mM ATP, freshly added protease inhibitors (1 mM PMSF, 1 mM benzamidine, 1.45 mM pepstatin), and 1 mM phosphatase inhibitors (NaF, NaN3, and $Na3VO_4$). 6 of the multiplex arrays were incubated with this reaction, and the additional 6 were negative controls lacking Kin28. Arrays were incubated for 3 hours at 30°C, prior to washing. One reaction and a corresponding negative control were each probed with the following primary antibodies in blocking buffer at 4°C overnight: α-CTD 8WG16 (1:500), α-pTyr1 5G9 (1:100), α-pSer2 Bethyl (1:2500), α-pThr4 6D7 (1:100), α-pSer5 3E8 (1:10,000), α-pSer7 4E12 (1:100). After washes, Cy3 secondary antibodies were diluted 1:10,000 and incubated for 3 hours at room temperature. Arrays were scanned as described above.

Cy3 labeling of primary antibodies

Because the Cy3 α-rat secondary did not bind well to our primary antibodies, we directly labeled primary antibodies. Cy3-NHS ester (GE Healthcare) was resuspended in 50 µL DMSO. 30 µg purified antibody was incubated with 1.5 µL Cy3-NHS ester for 1 hour at room temperature. Tris pH 7.0 was added (5 mM final) to quench excess NHS ester. Reactions were passed through a G-25 PD spin column twice to remove unreacted Cy3-NHS ester.

Data analysis

The median of triplicate probes were analyzed. Probes with intensity >10,000, with >2-fold enrichment in Mock, and with $p<0.05$ among triplicates were deemed significantly phosphorylated by Kin28. MEME was used to identify the substrate preference of Kin28 (Bailey et al., 2009).

A.7 References

Ansari, A.Z., Ogirala, A., and Ptashne, M. (2005). Transcriptional activating regions target attached substrates to a cyclin-dependent kinase. Proc Natl Acad Sci USA *102*, 2346-2349.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res *37*, W202-208.

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

Corden, J.L., Cadena, D.L., Ahearn, J.M., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci USA *82*, 7934-7938.

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

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chemical reviews *113*, 8456-8490.

Ficarro, S.B., McCleland, M.L., Stukenberg, P.T., Burke, D.J., Ross, M.M., Shabanowitz, J., Hunt, D.F., and White, F.M. (2002). Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotech *20*, 301-305.

Liu, P., Kenney, J.M., Stiller, J.W., and Greenleaf, A.L. (2010). Genetic Organization, Length Conservation, and Evolution of RNA Polymerase II Carboxyl-Terminal Domain. Mol Biol Evol *27*, 2628-2641.

Morgan, D.O. (1997). CYCLIN-DEPENDENT KINASES: Engines, Clocks, and Microprocessors. Annu Rev Cell Dev Biol *13*, 261-291.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev *20*, 2922-2936.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.-M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol Cell *61*, 305-314.

Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., and Cerrina, F. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nat Biotech *17*, 974-978.

Smith, B.C., Settles, B., Hallows, W.C., Craven, M.W., and Denu, J.M. (2011). SIRT3 Substrate Specificity Determined by Peptide Arrays and Machine Learning. ACS Chemical Biology *6*, 146-157.

Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Mol Cell *61*, 297-304.

Zhang, D.W., Rodriguez-Molina, J.B., Tietjen, J.R., Nemec, C.M., and Ansari, A.Z. (2012). Emerging Views on the CTD Code. Genetics research international *2012*, 347214.

Appendix B: Covalent crosslinking of CTD “writers” and “readers”

Experiments were performed by undergraduate students Kennedy Ringelberg and Tanner Byer with intellectual input from Corey Nemec. Lysine-bearing strains were created by Dr. Juan B. Rodríguez-Molina.

B.1 Introduction

The carboxyl-terminal domain of Rpb1, the largest subunit of RNA polymerase II, consists of a repeating heptapeptide ($Y_1S_2P_3T_4S_5P_6S_7$) (26 in yeast, 52 in humans) (Corden et al., 1985). Reversible phosphorylation of tyrosine, threonine, and the three serines create a binding scaffold for the association of factors required for RNA processing and chromatin remodeling, termed the “CTD code” (Bentley, 2014; Buratowski, 2003, 2009; Egloff and Murphy, 2008; Eick and Geyer, 2013; Hsin and Manley, 2012; Phatnani and Greenleaf, 2006; Zhang et al., 2012).

Previous work has sought to characterize the CTD interactome via mass spectrometry of CTD-interacting proteins (Harlen et al., 2016; Mosley et al., 2011; Phatnani et al., 2004), but these interactions may be transient or indirect, and give no context as to where along the repetitive CTD the binding occurs. While recent studies suggest a largely uniform distribution of CTD marks along the length of the CTD (Schüller et al., 2016; Suh et al., 2016), a previous study reported that a CTD bearing Ser2 to glutamate (a phosphomimic) in the second half of the CTD is inviable, whereas substitution in the first half is viable (West and Corden, 1995), suggesting that CTD-binding factors may differentially bind across the length of the CTD.

Further, to fully understand the role of CTD “writers” on transcription, the roles and targets of these kinases in a wider cellular context must also be examined. Standard methods for deciphering kinase targets are remarkably indirect and frequently rely on gene knockouts. However, the natural circuitry of a cell is disrupted upon gene knockout, and other kinases can be upregulated to compensate for the loss of a single kinase (Rossi et al., 2015). Kinase inhibition is also used to identify targets, but kinases are often part of larger signaling pathways, and inhibition of one kinase may disrupt the whole pathway, indirectly altering the phosphorylation status of downstream proteins (Kumar et al., 2008).

To circumvent the aforementioned issues, we utilize a method to covalently crosslink CTD "writers" and "readers" directly to their targets. Covalent crosslinking not only enables identification of direct binders, but also spatial information as well. Further, by engineering the covalent crosslinker into the active sites of CTD "writers" (kinases), we can identify novel targets of the kinase, allowing for a more complete understanding how individual kinases regulate biological processes in the cell.

B.2 Method of non-natural amino acid integration

To directly identify targets of CTD "writers" and "readers," amber suppression was used to expand the genetic code to allow for the incorporation of a non-natural amino acid at specific sites. With this technique, a rarely used "amber" stop codon (UAG) is designed into the open reading frame of a protein of interest (Wang et al., 2001; Wang and Schultz, 2005). A tRNA with an anticodon compatible with the amber stop codon was engineered, and a tRNA synthetase capable of charging this tRNA with a *p*-benzoyl-L-phenylalanine (pBpa) was created (Chin et al., 2003; Chin et al., 2002). Cells grown in the presence of pBpa readily incorporate this non-natural amino acid into the nascent polypeptide and suppress the usage of UAG as a stop codon (Figure B.1). Cells lacking pBpa utilize UAG as a stop codon and terminate translation, resulting in a truncated protein. Upon treatment of cells with 360 nm ultraviolet irradiation, pBpa forms a diradical, which reacts with C-H bonds to covalently crosslink to proteins within 3.1 Å (Tanaka et al., 2008). A covalent crosslink forms a substantial molecular weight shift when probing for the protein of interest via Western blot or mass spectrometry. A clear shift in molecular weight can not only validate previously-identified substrates, but also identify new targets.

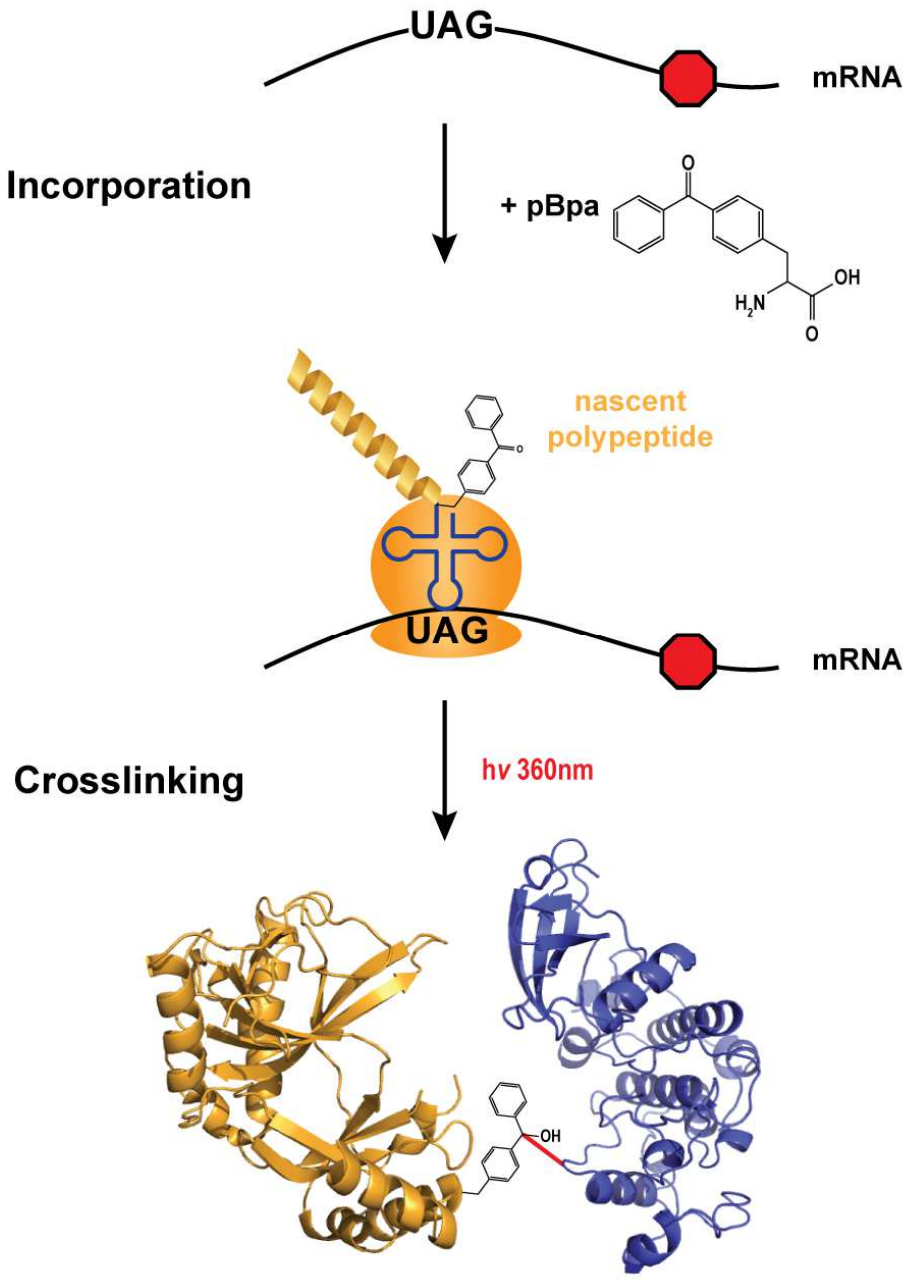

Figure B.1 Schematic of pBpa incorporation into nascent polypeptides
Site directed mutagenesis of a gene replaces a codon with the rarely used amber stop codon (UAG). In the presence of pBpa, an expressed tRNA synthetase mutant charges a modified tRNA with pBpa allowing the ribosome (orange) to incorporate pBpa (not to scale) at the UAG instead of terminating translation. Translation continues until the endogenous stop codon (red octagon) signals translation termination. Protein with incorporated pBpa interacts with binding partners or substrates, and upon UV irradiation, pBpa covalently crosslinks to targets within 3.1 Å (red line).

A plethora of tRNA synthetases have been developed to allow for the incorporation of many different amino acids (Xie and Schultz, 2006), depending on the desired effect. To characterize direct binding of our proteins of interest to their targets, we sought an amino acid capable of photocrosslinking to physically nearby residues. One limitation with pBpa is that crosslinks with methionine are most favorable, and crosslinking to a methionine can occur beyond the 3.1 Å radius (Lancia et al., 2014). Another photocrosslinker, *p*-azido-L-phenylalanine (pAzpa) does not suffer this bias, but if the UV-activated form does not react within 10^{-4}s, a more stable, but less reactive ketenimine will form, which can react with nucleophiles, thus limiting specific crosslinking efficiency (Tanaka et al., 2008). Preliminary results confirm that pAzpa was readily incorporated, but crosslinking to targets was diminished (data not shown). Therefore, experiments were performed with pBpa.

B.3 Identification of sites of non-natural amino acid incorporation

During transcription initiation, Kin28, a subunit of the general transcription factor TFIIH, phosphorylates Ser5 and Ser7 on the RNA Pol II CTD (Akhtar et al., 2009). Phosphorylation of Ser5 recruits Ceg1 (Cho et al., 1997; McCracken et al., 1997), a component of the RNA capping complex, as well as Bur1 (Qiu et al., 2009), a kinase which subsequently phosphorylates Ser2 and Ser7 (Qiu et al., 2009; Tietjen et al., 2010). Therefore, as a proof of concept, Ceg1, Bur1, and Kin28 were engineered to incorporate pBpa via amber suppression.

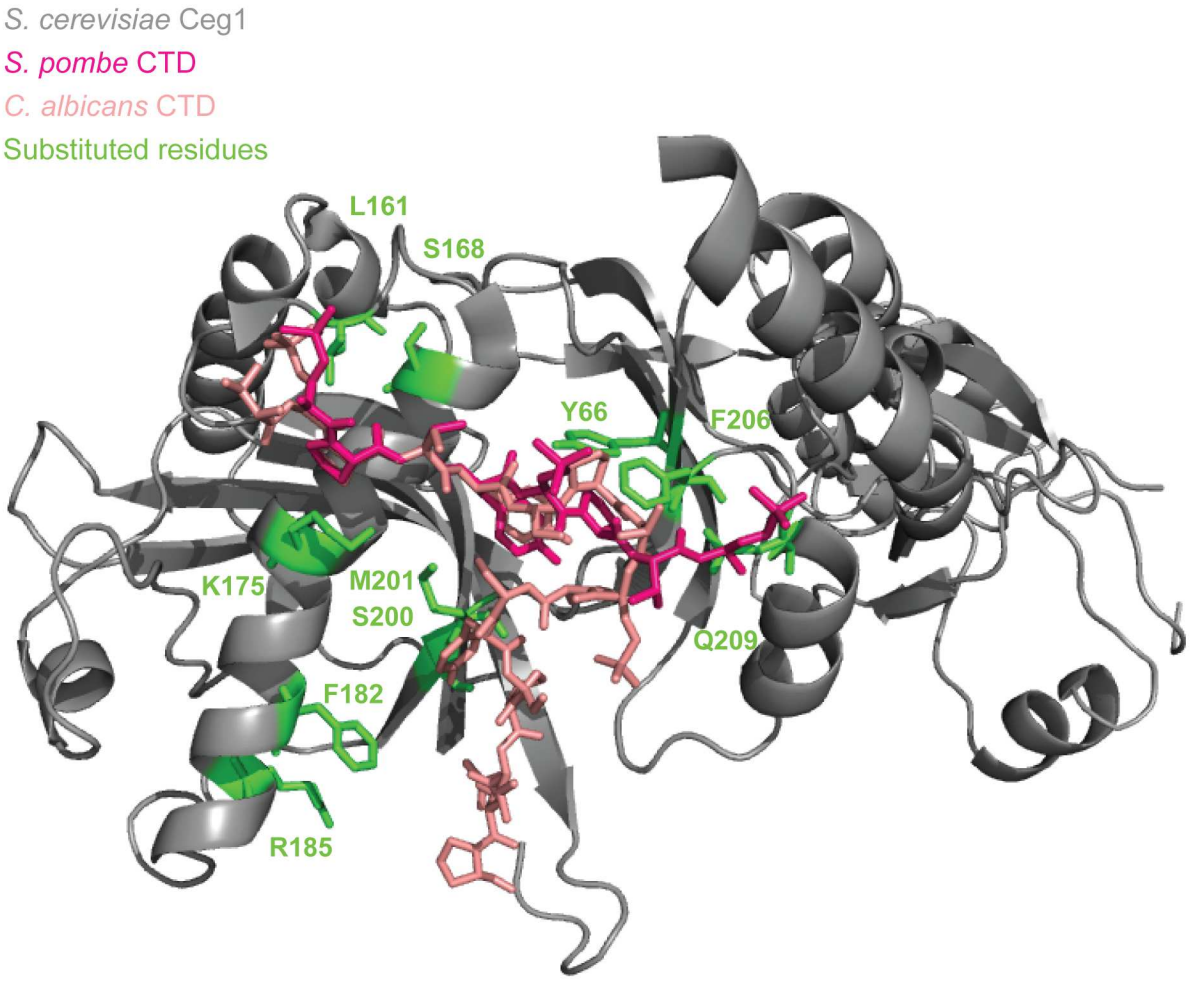

Figure B.2 Non-natural amino acid incorporation sites in Ceg1
Structure of *S. cerevisiae* Ceg1 (grey) aligned with CTDs from *S. pombe* (pink) and *C. albicans* (salmon) in complex with capping enzyme. Green residues in Ceg1 were individually substituted with the amber stop codon to allow for pBpa incorporation.

The crystal structure of *S. cerevisiae* Ceg1 (PDB 3KYH) (Gu et al., 2010) was aligned with the structures of the capping enzyme from *S. pombe* (PDB 4PZ6) (Doamekpor et al., 2014) and *C. albicans* (PDB 1P16) (Fabrega et al., 2003), which were co-crystallized with the CTD. Subsequently, *S. pombe* and *C. albicans* capping enzyme were removed, leaving *S. cerevisiae* Ceg1 in complex with two variants of the CTD (Figure B.2). Surprisingly, while the N-terminal portions of the CTDs overlap quite well, the C-terminal portions of each species' CTD deviate substantially.

Several criteria were used to choose sites of pBpa incorporation. First, residues were within 3.1 Å of the CTD to allow for efficient crosslinking. Amino acids with similar size and charge were favored for substitution to avoid steric clashes due to size differences. And finally, residues with poor evolutionary conservation were favored to preclude residues important for docking. Ten amino acids were selected for substitution with the amber stop codon via site-directed mutagenesis.

While the structures of Bur1 and Kin28 from *S. cerevisiae* have not been resolved, the active sites of many kinases share a high level of sequence similarity. Protein homology/analogy recognition engine (Phyre) was used to model the primary structure of Bur1 and Kin28 based on previously solved kinase structures (Kelley et al., 2015). The generated Bur1 and Kin28 structures were aligned with human CDK2 co-crystallized with its substrate. This substrate approximated the CTD and allowed for the selection of sites to substitute with the amber stop codon. Similar to Ceg1, ten residues within 3.1 Å of the substrate, with similar size and charge, and with poor evolutionary conservation were substituted with UAG for pBpa incorporation (Figure B.3).

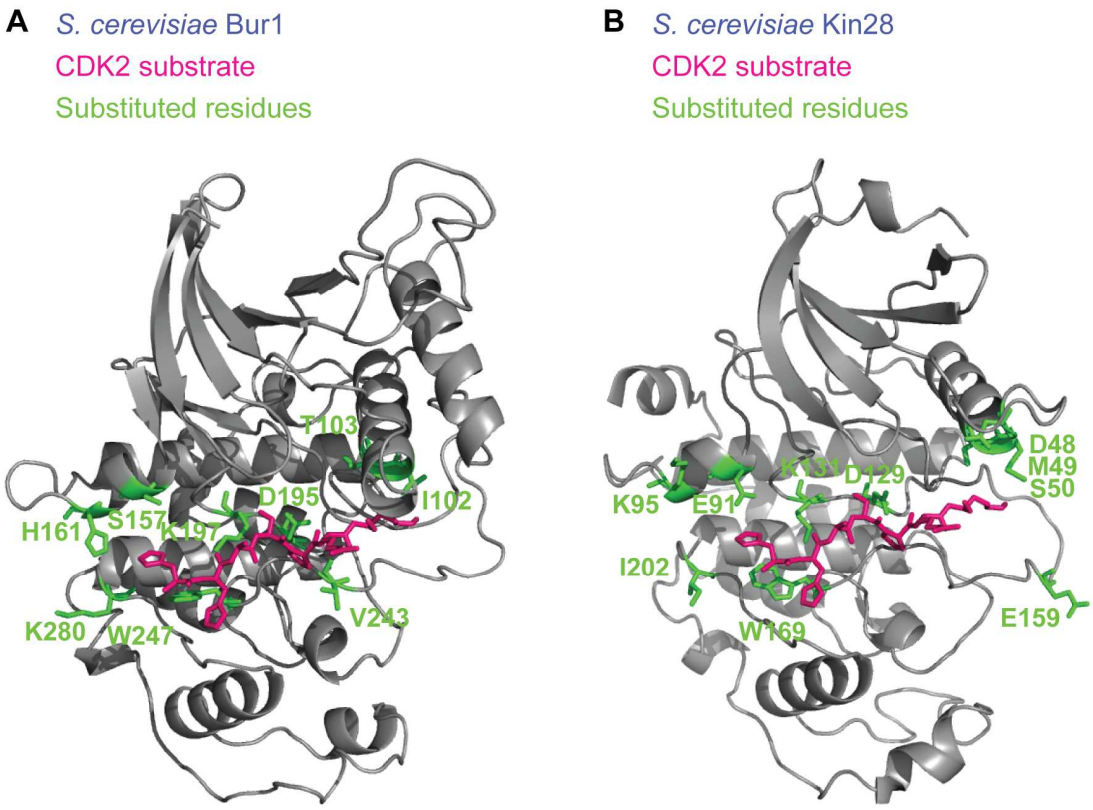

Figure B.3 Non-natural amino acid incorporation sites in Bur1 and Kin28

A) Model of *S. cerevisiae* Bur1 (grey) generated with Phyre and aligned with the CDK2 substrate (pink). Green residues in Bur1 were individually substituted with the amber stop codon to allow for pBpa incorporation.

B) Model of *S. cerevisiae* Kin28 (grey) generated with Phyre and aligned with the CDK2 substrate (pink). Green residues in Kin28 were individually substituted with the amber stop codon to allow for pBpa incorporation.

Additionally, an amber stop codon replaced Kin28 E159, a residue in the T-Loop. The T-Loop contains a threonine that is frequently phosphorylated as a means of kinase activation. Upon activation, the T-Loop undergoes a conformational change and the phosphorylated threonine interacts with a positively charged residue on the substrate. While E159 is further than 3.1 Å from the substrate in the structure, we hypothesized that upon kinase activation, the distance would be reduced and may allow crosslinking. pBpa in place of E159 may allow for the dynamics of Kin28 activity to be more finely discerned.

B.4 Incorporation of pBpa into Ceg1, Bur1, and Kin28

Because Ceg1, Bur1, and Kin28 are all essential, integrating UAG mutations into the genome was impractical. First, in order to express full length protein, pBpa would need to be provided at all times, both on agar plates and in liquid culture. Not only is this cost prohibitive, but cells could easily revert back to WT. Therefore, we decided to express each protein from a plasmid containing a galactose-inducible promoter. Further, using a plasmid allowed for simple site-directed mutagenesis of each mutant. Nine Ceg1 mutants, three Bur1 mutants, and three Kin28 mutants were successfully created and validated by sequencing.

LS41, a yeast strain known for its ability to strongly incorporate non-natural amino acids (Lancia et al., 2014) was transformed with 1) a plasmid encoding the pBpa tRNA and tRNA synthetase, and 2) each mutant variant of Ceg1, Bur1, or Kin28. To prevent premature expression of these proteins, cells were grown overnight in raffinose. In the morning, galactose was added in the presence or absence of pBpa to induce expression of Ceg1, Bur1, or Kin28. In the presence of pBpa, the amino acid is incorporated and translation will continue through a C-terminal HA tag, which can be probed via Western blot. In the absence of pBpa, the UAG codon signals

translation termination, preventing translation through to the HA tag (Figure B.4). WT cells lacking UAG express full length protein even in the absence of pBpa.

Strong expression (and therefore, high incorporation) was observed for Ceg1 mutants incorporating pBpa at position 161, 168, and 201. Significant expression over the negative control was observed for the other mutants, suggesting that pBpa is still incorporated, but less efficiently. Similar results were observed with Bur1 and Kin28. While all mutants displayed higher expression in the presence of pBpa, none were expressed at levels comparable to WT, suggesting poor incorporation.

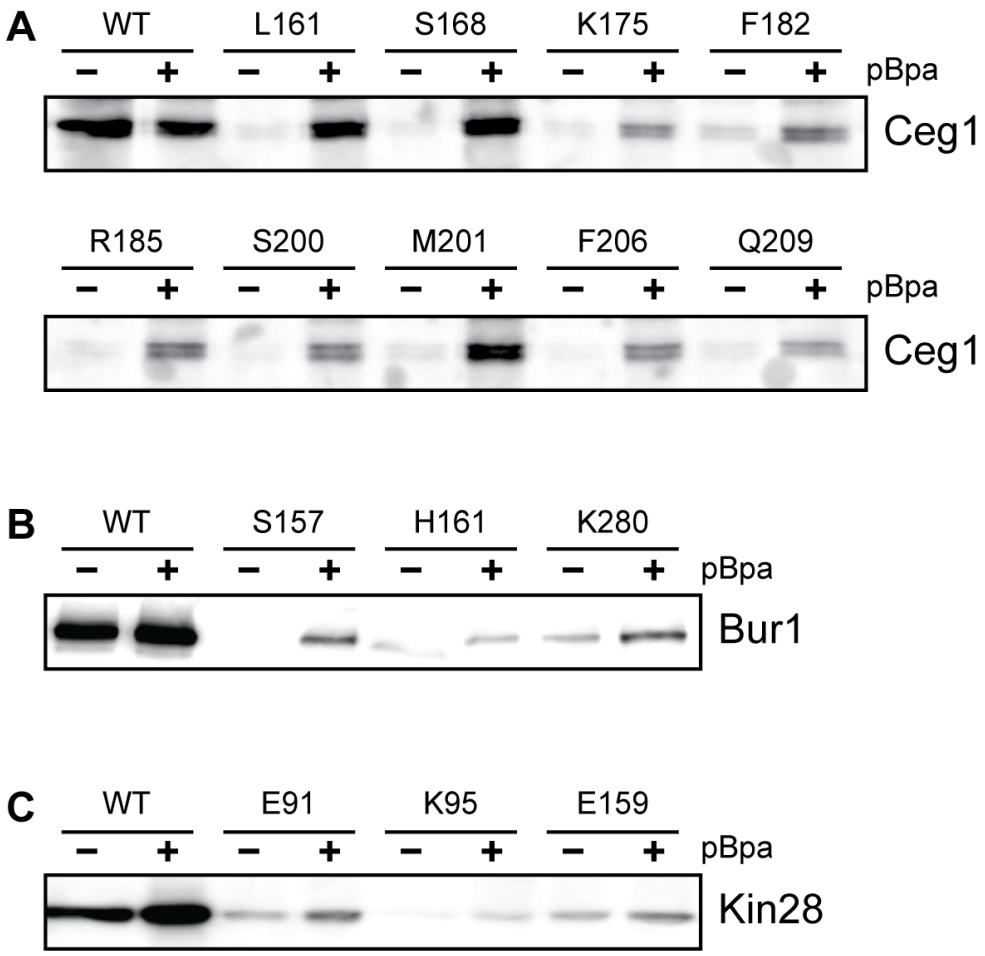

Figure B.4 Incorporation of pBpa into Ceg1, Bur1, and Kin28
Plasmids expressing A) Ceg1, B) Bur1, or C) Kin28 in galactose-dependent manner were mutagenized to allow pBpa incorporation in place of the amino acids listed. Upon addition of galactose, each gene was expressed and translated into protein. Western blots probing a C-terminal HA tag were performed. In the absence of pBpa (–), the UAG codon terminates translation preventing the C-terminal HA tag from being translated. In the presence of pBpa (+), the amino acid is incorporated, and translation continues through to the C-terminus. WT cells lacking a UAG are fully expressed even in the absence of pBpa.

B.5 Crosslinking of pBpa

Next, we tested the ability of Ceg1, Bur1, and Kin28 to crosslink to substrates. WT and mutant variants were expressed in the presence of pBpa. Cells were split and either exposed to 365nm UV light for 30 minutes or not. Cells were lysed, and the lysate was resolved on a polyacrylamide gel, transferred to nitrocellulose, and probed with an antibody targeting the C-terminal HA tag (Figure B.5).

While most of the expressed Ceg1 remained uncrosslinked, a band at approximately 250kd was observed when pBpa was incorporated at position 175, 182, 201, and 209 and crosslinked with UV light (Figure B.5A). This shift corresponds to the mass of Rpb1, the subunit of Pol II containing the CTD. Positions 175, 182, and 201 in Ceg1 are all within physical proximity (Figure B.2), suggesting that this region of Ceg1 may be most amenable to crosslinking, or that the sequence of the CTD bound to this region is preferred by pBpa. Interestingly, *S. pombe* and *C. albicans* structures of Ceg1 show the CTD in different orientations. Only the structure of the CTD in *C. albicans* is capable of crosslinking to pBpa at positions 175, 182, and 201, confirming that this *S. cerevisiae* Ceg1 bound to CTD can adopt this conformation. Surprisingly, pBpa incorporated at position 209 also displayed evidence of crosslinking (Figure B.5A). This residue could only crosslink to CTD in the orientation found in *S. pombe*, suggesting that *S. cerevisiae* Ceg1 can bind to the CTD in two different orientations. Different combinations of phosphorylation marks on the CTD may alter the conformation used, and may serve as a mode of regulation to efficiently cap nascent RNA transcripts.

Unfortunately, no evidence of crosslinking with any Bur1 mutant was observed (data not shown), but a strong shift in Kin28 mobility was observed when pBpa was incorporated at position 91 and exposed to UV light. While the observed shift was not indicative of crosslinking to Rpb1, the identity of the crosslinked species will be elucidated with future work.

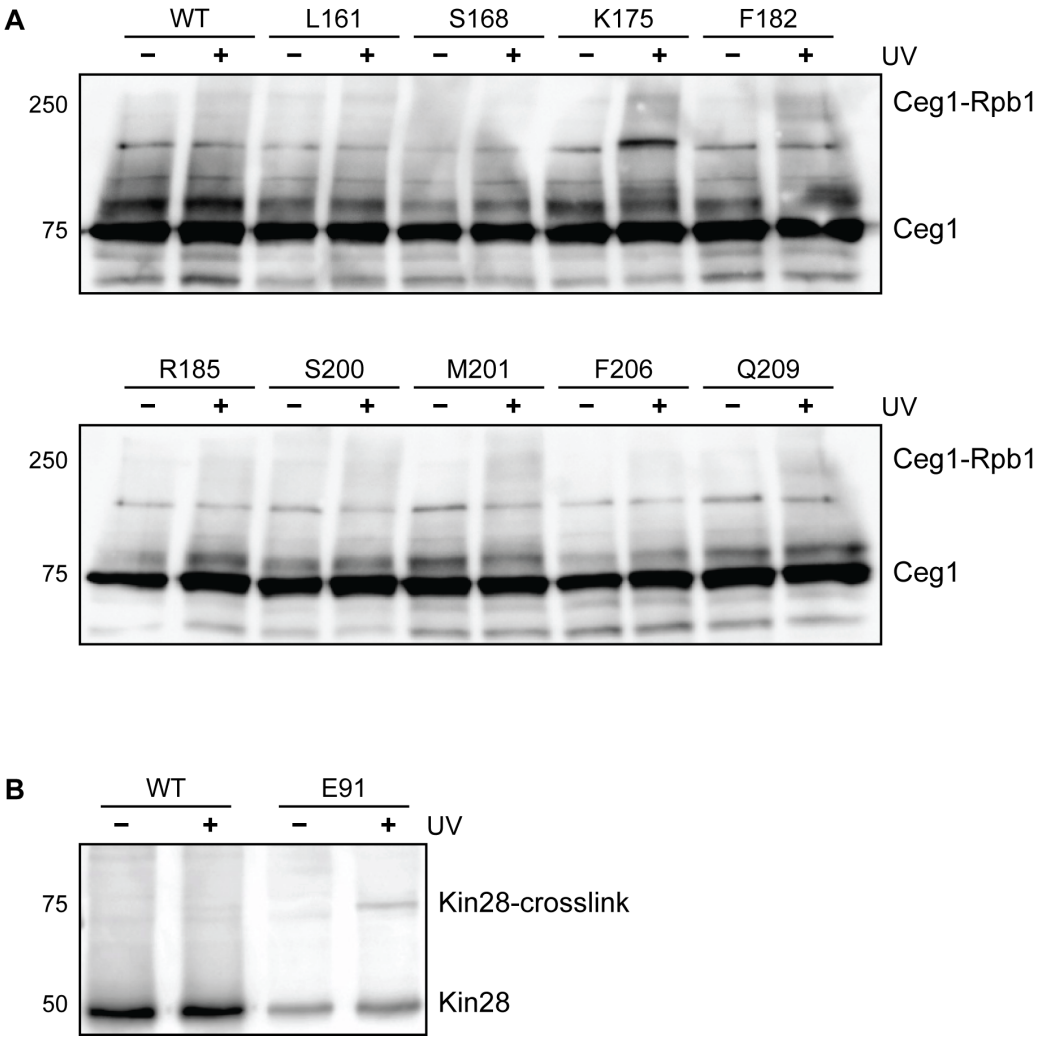

Figure B.5 Crosslinking of pBpa

A) Ceg1 variants were induced in the presence of pBpa. Upon UV treatment (+), Ceg1-pBpa crosslinks to the Pol II CTD (Ceg1-Rpb1).

B) Kin28 variants were induced in the presence of pBpa. Upon UV treatment (+), Kin28-pBpa crosslinks to an unknown target.

B.6 Future directions

Because several Ceg1 mutants displayed prominent crosslinking to the Pol II CTD, we will examine where along the repetitive tail Ceg1 binds. We have engineered a series of strains in which lysine residues are regularly interspersed between diheptad CTD repeats in Rpb1 (Figure B.6A). We reasoned that these insertions would be tolerated by cells because CTD variants with alanine insertions between diheptads do not demonstrate apparent defects in fitness (Liu et al., 2008; Schwer et al., 2012; Stiller and Cook, 2004). Indeed, a plasmid shuffle assay confirms the viability of yeast strains bearing engineered CTDs as their sole source of Rpb1, demonstrating that these mutant Rpb1 alleles associate and function with endogenous Pol II machinery (Figure B.6B). Further, we confirmed that the mutant CTDs are equally capable of being phosphorylated at Tyr1, Ser2, Thr4, Ser5, and Ser7 as the WT strain, which strongly suggests that these mutants will biologically behave similarly to the WT strain (Figure B.6C). All mutant Rpb1 alleles contain a C-terminal 10x His tag and an HA tag for tandem purification. We have also engineered a TEV protease site in the linker between the enzymatic core of Pol II and the CTD to facilitate precise separation of CTD from the rest of Pol II. Immunoblots against the HA epitope demonstrate that the CTD peptide can be cleaved from Rpb1 and resolved by SDS-PAGE (Figure B.6D).

Plasmids expressing these mutants will be transformed into cells capable of incorporating pBpa into Ceg1, and WT Rpb1 will be shuffled out. pBpa will be incorporated and crosslinked, and Rpb1 will be isolated using the 10xHis tag. CTDs will be cleaved from the core polymerase using TEV protease, and will be washed with buffer containing 4M urea to ensure only covalently crosslinked species remain. Following elution with imidazole, cleavage at the lysine residues with trypsin will produce two distinguishable CTD fragments which may then be assayed by immunoblot analysis and mass spectrometry (Figure B.6E).

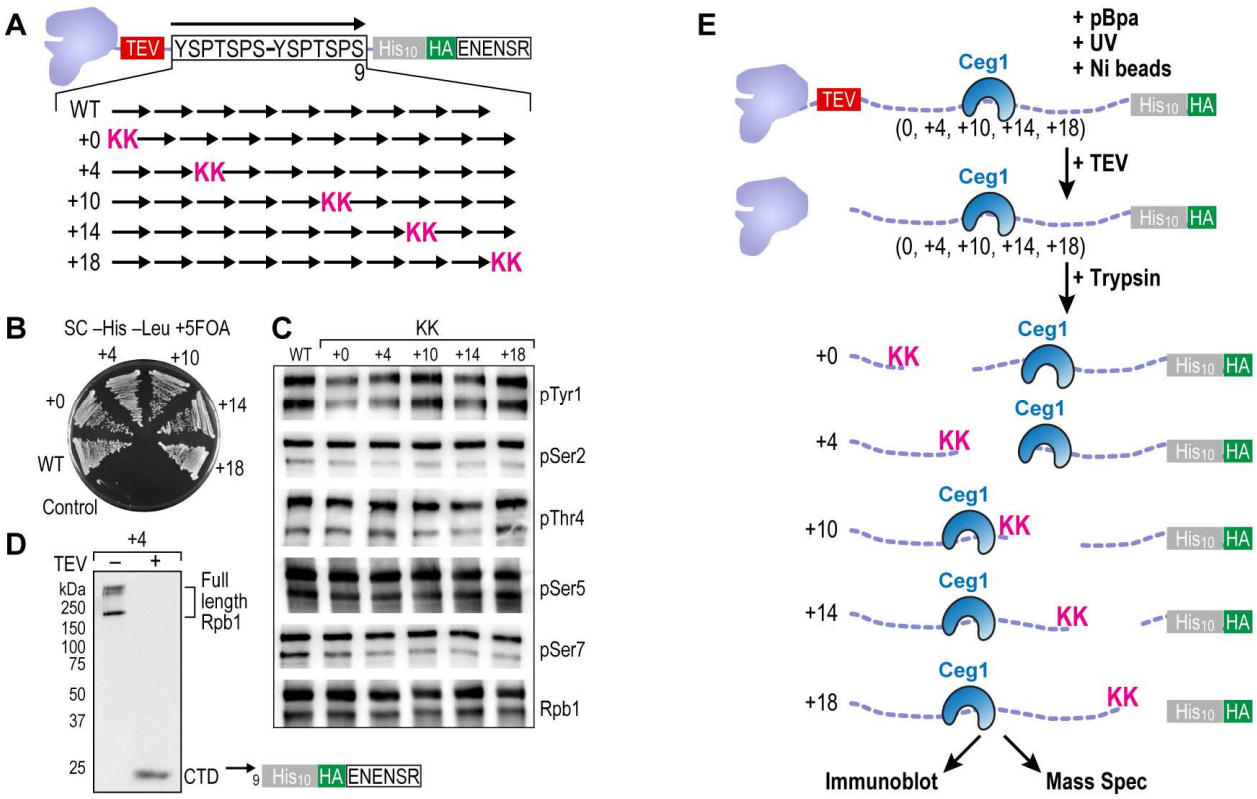

Figure B.6 Deciphering the binding sites along the Pol II CTD

A) CTD variants bearing two lysine residues (pink) throughout the length of an 18 repeat tail, a TEV cleavage site (red) in the linker region, eighteen canonical repeats (black arrows), and a 10xHis (grey) and HA epitope tag (green). The position of each insertion is denoted by the number of repeats between the TEV site and the lysines.

B) Yeast strains expressing Rpb1 18xCTD with lysine substitutions are viable as determined by standard plasmid shuffle assay.

C) Immunoblot analysis against the HA epitope shows CTD can be cleaved from the Rpb1 core.

D) Lysine mutants do not display altered CTD phosphorylation.

E) pBpa is incorporated into Ceg1 mutants and cells are exposed to UV light. Cell lysate is incubated with nickel beads, and the CTD is cleaved with TEV followed by trypsin. CTD fragments are analyzed via immunoblot or mass spectrometry.

B.7 Experimental procedures

Plasmid and strain creation

Plasmid BG1805 containing the GAL1 promoter upstream of Ceg1, Bur1, or Kin28 (GE Healthcare) were site-directed mutagenized to substitute individual amino acids (Figure B.2, B.3) with an amber stop codon (TAG/UAG). Site-directed mutagenesis oligos were designed using PrimerX (bioinformatics.org/primerx) (Table B.2). Standard site-directed mutagenesis was performed, and plasmids were validated by Sanger sequencing (UW Biotechnology Sequencing Facility).

The URA3-bearing pZZ41 was shuffled out of LS41 (Lancia et al., 2014) creating LS42. LS42 was transformed with 1) WT or mutant variants of Ceg1, Bur1, or Kin28 in BG1805, and 2) a plasmid expressing the tRNA-pBpa tRNA synthetase (from *M. jannaschii*).

Incorporation of pBpa

Strains were grown overnight in synthetic complete media lacking uracil and tryptophan containing 2% raffinose. Yeast were subcultured to an OD_{600} ~ 0.5 in 12 mL selective media containing 2% galactose. Cultures were split to 6 mL. pBpa (Chem-Impex) was dissolved in 1 M NaOH to 200 mM and added to the cultures to a final concentration of 2 mM. 1 M NaOH was added to negative control cultures. An equal volume of 1 M HCl was added to normalize the pH of the cultures. Ceg1, Bur1, or Kin28 were induced for 4 hours at 30°C in the absence of light. Incorporation of pBpa was measured via detection of the C-terminal HA tag by western blot (Abcam – ab9110, 1:1000).

UV crosslinking of pBpa

Following pBpa incorporation, cultures were split, and pelleted. Half were immediately flash frozen, while half were resuspended in 1 mL fresh media and exposed to 365 nm UV light

(CalSun Facial Tanning Sun Lamp) at 4°C for 45 minutes in a 6-well plate, with gentle mixing every 10-15 minutes. Cells were pelleted, flash frozen, and processed in parallel to the –UV control samples via western blot. Blots were probed with an α-HA antibody (Abcam – ab9110, 1:1,000) using standard procedures as described above.

Table B.1 Strains used in this study

Strain	Genotype	Plasmid	Source
LS41	JPY9::pZZ41, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4 URA::pZZ41	pZZ41	A. Mapp Lab
Ceg1 WT	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Ceg1 WT, pSNR pBpa aaRS/tRNA	This study
Ceg1 L161TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Ceg1 L161TAG, pSNR pBpa aaRS/tRNA	This study
Ceg1 S178TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Ceg1 S178TAG, pSNR pBpa aaRS/tRNA	This study
Ceg1 K175TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Ceg1 K175TAG, pSNR pBpa aaRS/tRNA	This study
Ceg1 F182TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Ceg1 F182TAG, pSNR pBpa aaRS/tRNA	This study
Ceg1 R185TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Ceg1 R185TAG, pSNR pBpa aaRS/tRNA	This study
Ceg1 S200TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Ceg1 S200TAG, pSNR pBpa aaRS/tRNA	This study
Ceg1 M201TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Ceg1 M201TAG, pSNR pBpa aaRS/tRNA	This study
Ceg1 F206TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Ceg1 F206TAG, pSNR pBpa aaRS/tRNA	This study
Ceg1 Q209TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Ceg1 Q209TAG, pSNR pBpa aaRS/tRNA	This study
Bur1 WT	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Bur1 WT, pSNR pBpa aaRS/tRNA	This study
Bur1 S157TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Bur1 S157TAG, pSNR pBpa aaRS/tRNA	This study
Bur1 H161TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Bur1 H161TAG, pSNR pBpa aaRS/tRNA	This study
Bur1 K280TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Bur1 K280TAG, pSNR pBpa aaRS/tRNA	This study
Kin28 WT	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Kin28 WT, pSNR pBpa aaRS/tRNA	This study
Kin28 E91TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Kin28 E91TAG, pSNR pBpa aaRS/tRNA	This study
Kin28 K95TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Kin28 K95TAG, pSNR pBpa aaRS/tRNA	This study
Kin28 E159TAG	JPY9, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4	BG1805 Kin28 E159TAG, pSNR pBpa aaRS/tRNA	This study

Table B.2 Oligonucleotides used in this study

Ceg1 Y66TAG F	GCTTGCGCATGATTACTAGGTTTGTGAGAAAACAGATG
Ceg1 Y66TAG R	CATCTGTTTTCTCACAAACCTAGTAATCATGCGCAAGC
Ceg1 L161TAG F	CTTGCTATCAATGGTAGATGTTAGACACAATCACCAACAAGTTC
Ceg1 L161TAG R	GAACTTGTTGGTGATTGTGTCTAACATCTACCATTGATAGCAAG
Ceg1 S168TAG F	CACAATCACCAACAAGTTAGAGACTAGCCCACCTTGG
Ceg1 S168TAG R	CCAAGGTGGGCTAGTCTCTAACTTGTTGGTGATTGTG
Ceg1 K175TAG F	CTAGACTAGCCCACCTTGGATAGGAATTTTTTAAACCATACTTC
Ceg1 K175TAG F	GAAGTATGGTTTAAAAAATTCCTATCCAAGGTGGGCTAGTCTAG
Ceg1 F182TAG R	GAATTTTTTAAACCATACTAGGATTTAAGAGCAGCGTACCC
Ceg1 F182TAG F	GGGTACGCTGCTCTTAAATCCTAGTATGGTTTAAAAAATTC
Ceg1 R185TAG R	CCATACTTCGATTTATAGGCAGCGTACCCTAATC
Ceg1 R185TAG F	GATTAGGGTACGCTGCCTATAAATCGAAGTATGG
Ceg1 S200TAG R	CTACTTTTCCGTTCAAAATTTAGATGAAACATATGGATTTCAG
Ceg1 S200TAG F	CTGAAATCCATATGTTTCATCTAAATTTTGAACGGAAAAGTAG
Ceg1 M201TAG R	CTTTTCCGTTCAAAATTTCCTAGAAACATATGGATTTCAG
Ceg1 M201TAG F	CTGAAATCCATATGTTTCTAGGAAATTTTGAACGGAAAAG
Ceg1 F206TAG R	CCATGAAACATATGGATTAGAGTTACCAATTAGTAAAAG
Ceg1 F206TAG F	CTTTTACTAATTGGTAACTCTAATCCATATGTTTCATGG
Ceg1 Q209TAG R	CCATGAAACATATGGATTTCAGTTACTAGTTAGTAAAAGTTGCTAAAAG
Ceg1 Q209TAG F	CTTTTAGCAACTTTTACTAACTAGTAACTGAAATCCATATGTTTCATGG
Kin28 D48TAG F	TTAAAGATGGTTTATAGATGTCAGCTATCCGTGAAGTTAAGTACCTCCAAG
Kin28 D48TAG R	CGGATAGCTGACATCTATAAACCATCTTTAAATTCGGATGTTTTGATCTCC
Kin28 M49TAG F	CATCCGAATTTAAAGATGGTTTAGATTAGTCAGCTATCCG
Kin28 M49TAG R	GGTACTTAACTTCACGGATAGCTGACTAATCTAAACCATC
Kin28 S50TAG F	GTTTAGATATGTCATAGGCTATCCGTGAAGTTAAG
Kin28 S50TAG R	ACTTCACGGATAGCCTATGACATATCTAAACCATC
Kin28 E91TAG F	GGAGTTCCTACCAACTGATCTATAGGTGGTAATAAAAGACAAATC
Kin28 E91TAG R	GATTTGTCTTTTATTACCACCTATAGATCAGTTGGTAGGAACTCC
Kin28 K95TAG F	CCTACCAACTGATCTAGAGGTGGTAATATAGGACAAATCAATACTGTTTACACCAGCAG
Kin28 K95TAG R	CTGCTGGTGTAAACAGTATTGATTTGTCCTATATTACCACCTCTAGATCAGTTGGTAGG
Kin28 D129TAG F	CGTGTATCATTGCCACAGAAATTTCATTTTGCACAGGTAGCTGAAACCAAA
Kin28 D129TAG R	GGCCATCAGGTGAAAATAATAAATTGTTTGGTTTCAGCTACCTGTGCAAAA
Kin28 K131TAG F	GCACAGGGATCTGTAGCCAAACAATTTATTATTTTCACCTGATGGCCAGATAAAAGTAGCAG
Kin28 K131TAG R	ATAAATTGTTTGGCTACAGATCCCTGTGCAAAATGAAATTTCTGTGGCAATGATACACGCCC
Kin28 E159TAG F	GATACCGGCCCCACATTAGATACTGACAAGTAAC
Kin28 E159TAG R	TTACTTGTCAGTATCTAATGTGGGGCCGGTATCG
Kin28 W169TAG F	ACGTCGTAACAAGATAGTATAGAGCGCCAG
Kin28 W169TAG R	TCTGGCGCTCTATACTATCTTGTTACGACG

Kin28 I202TAG F	ATTCGCGGAATTAATGCTAAGGTAGCCTTATTTACCAGGACAG
Kin28 I202TAG R	CATTCTGTCCTGGTAAATAAGGCTACCTTAGCATTAATTCCGC
Bur1 I102TAG F	GAAAAAAATTATAGTTAGCGTTGAAAAAGATCTGTTTCCTTAGACTG CACAACGAGAAATTACTATTTTAAAGCGTTTAAACC
Bur1 I102TAG R	GGTTTAAACGCTTTAAAATAGTAATTTCTCGTTGTGCAGTCTAAGGA AACAGATCTTTTTCAACGCTAACTATAATTTTTTTC
Bur1 T103TAG F	GTTAGCGTTGAAAAAGATCTGTTTCCTATATAGGCACAACGAGAAAT TACTATTTTAAAGCGT
Bur1 T103TAG R	ACGCTTTAAAATAGTAATTTCTCGTTGTGCCTATATAGGAAACAGAT CTTTTTCAACGCTAAC
Bur1 S157TAG F	GTATATGGTAGCAGATCTGTAGGGCGTTTTGCATAATCCGAG
Bur1 S157TAG R	TCGGATTATGCAAAACGCCCTACAGATCTGCTACCATATACG
Bur1 H161TAG F	GCAGATCTGTCGGGCGTTTTGTAGAATCCGAGAATCAACTTAGAG
Bur1 H161TAG R	CTCTAAGTTGATTCTCGGATTCTACAAAACGCCCGACAGATCTGC
Bur1 D195TAG F	GCCTAAATTATATTCATTGTGCAAAATTCATGCATAGGTAGATAAAG ACAGCAAATATTTTGATTGACCACAACGGTG
Bur1 D195TAG R	ACACCGTTGTGGTCAATCAAAATATTTGCTGTCTTTATCTACCTATG CATGAATTTTGCACAATGAATATAATTTAGG
Bur1 K197TAG F	TGTGCAAAATTCATGCATAGGGATATATAGACAGCAAATATTTTGAT TGACCAC
Bur1 K197TAG R	TGTGGTCAATCAAAATATTTGCTGTCTATATATCCCTATGCATGAAT TTTGCAC
Bur1 V243TAG F	CGGCGCCAAATATACATCAGTATAGGTAACGAGATGGTATAGAGCA C
Bur1 V243TAG R	GTGCTCTATACCATCTCGTTACCTATACTGATGTATATTTGGCGCCG
Bur1 W247TAG F	AAATATACATCAGTAGTCGTAACGAGATAGTATAGAGCACCTGAGT TGGTGCTTGGTG
Bur1 W247TAG R	ACCAAGCACCAACTCAGGTGCTCTATACTATCTCGTTACGACTACT GATGTATATTTG
Bur1 K280TAG F	GGATGCGTTTTCGCAGAATTTTTTGAGAAATAGCCTATTTTACAGGG AAAAACGGATATTGATC
Bur1 K280TAG R	ATCAATATCCGTTTTTCCCTGTAAAATAGGCTATTTCTCAAAAAATTC TGCGAAAACGCATCCG

B.8 References

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. Mol Cell *34*, 387-393.

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

Buratowski, S. (2003). The CTD code. Nat Struct Mol Biol *10*, 679-680.

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

Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z., and Schultz, P.G. (2003). An Expanded Eukaryotic Genetic Code. Science *301*, 964-967.

Chin, J.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002). Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. Proc Natl Acad Sci USA *99*, 11020-11024.

Cho, E.-J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev *11*, 3319-3326.

Corden, J.L., Cadena, D.L., Ahearn, J.M., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci USA *82*, 7934-7938.

Doamekpor, S.K., Sanchez, A.M., Schwer, B., Shuman, S., and Lima, C.D. (2014). How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. Genes Dev *28*, 1323-1336.

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

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chemical reviews *113*, 8456-8490.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol Cell *11*, 1549-1561.

Gu, M., Rajashankar, K.R., and Lima, C.D. (2010). Structure of the Saccharomyces cerevisiae Cet1-Ceg1 mRNA capping apparatus. Structure (London, England : 1993) *18*, 216-227.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Rep *15*, 2147-2158.

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

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protocols *10*, 845-858.

Kumar, N., Afeyan, R., Kim, H.-D., and Lauffenburger, D.A. (2008). Multipathway Model Enables Prediction of Kinase Inhibitor Cross-Talk Effects on Migration of Her2-Overexpressing Mammary Epithelial Cells. Mol Pharmacol *73*, 1668-1678.

Lancia, J.K., Nwokoye, A., Dugan, A., Joiner, C., Pricer, R., and Mapp, A.K. (2014). Sequence context and crosslinking mechanism affect the efficiency of in vivo capture of a protein-protein interaction. Biopolymers *101*, 391-397.

Liu, P., Greenleaf, A.L., and Stiller, J.W. (2008). The Essential Sequence Elements Required for RNAP II Carboxyl-terminal Domain Function in Yeast and Their Evolutionary Conservation. Mol Biol Evol *25*, 719-727.

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

Mosley, A.L., Sardiu, M.E., Pattenden, S.G., Workman, J.L., Florens, L., and Washburn, M.P. (2011). Highly Reproducible Label Free Quantitative Proteomic Analysis of RNA Polymerase Complexes. Molecular & Cellular Proteomics *10*.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev *20*, 2922-2936.

Phatnani, H.P., Jones, J.C., and Greenleaf, A.L. (2004). Expanding the Functional Repertoire of CTD Kinase I and RNA Polymerase II: Novel PhosphoCTD-Associating Proteins in the Yeast Proteome†. Biochemistry *43*, 15702-15719.

Qiu, H., Hu, C., and Hinnebusch, A.G. (2009). Phosphorylation of the Pol II CTD by KIN28 Enhances BUR1/BUR2 Recruitment and Ser2 CTD Phosphorylation Near Promoters. Mol Cell *33*, 752-762.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Holper, S., Kruger, M., and Stainier, D.Y.R. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature *524*, 230-233.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.-M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol Cell *61*, 305-314.

Schwer, B., Sanchez, A.M., and Shuman, S. (2012). Punctuation and syntax of the RNA polymerase II CTD code in fission yeast. Proc Natl Acad Sci USA.

Stiller, J.W., and Cook, M.S. (2004). Functional Unit of the RNA Polymerase II C-Terminal Domain Lies within Heptapeptide Pairs. Eukaryot Cell *3*, 735-740.

Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Mol Cell *61*, 297-304.

Tanaka, Y., Bond, M.R., and Kohler, J.J. (2008). Photocrosslinkers illuminate interactions in living cells. Molecular BioSystems *4*, 473-480.

Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S*., et al.* (2010). Chemical-genomic dissection of the CTD code. Nat Struct Mol Biol *17*, 1154-1161.

Wang, L., Brock, A., Herberich, B., and Schultz, P.G. (2001). Expanding the Genetic Code of Escherichia coli. Science *292*, 498-500.

Wang, L., and Schultz, P.G. (2005). Expanding the Genetic Code. Angewandte Chemie International Edition *44*, 34-66.

West, M.L., and Corden, J.L. (1995). Construction and Analysis of Yeast RNA Polymerase II CTD Deletion and Substitution Mutations. Genetics *140*, 1223-1233.

Xie, J., and Schultz, P.G. (2006). A chemical toolkit for proteins - an expanded genetic code. Nat Rev Mol Cell Biol *7*, 775-782.

Zhang, D.W., Rodriguez-Molina, J.B., Tietjen, J.R., Nemec, C.M., and Ansari, A.Z. (2012). Emerging Views on the CTD Code. Genetics research international *2012*, 347214.

Appendix C: Pathway connectivity and signaling coordination in the yeast stress activated signaling network

This appendix has been adapted from "Pathway connectivity and signaling coordination in the yeast stress-activated signaling network" by Deborah Chasman, Yi-Hsuan Ho*, David B. Berry, Corey M. Nemec, Matthew E. MacGilvray, James Hose, Anna E. Merrill, M. Violet Lee, Jessica L. Will, Joshua J. Coon, Aseem Z. Ansari, Mark Craven, and Audrey P. Gasch. (* These authors share first authorship) (Chasman et al., 2014).*

Computational approach was designed by Deborah Chasman, Mark Craven, and Audrey P. Gash. Biological investigation of Cdc14 and CTD analysis was performed by Yi-Hsuan Ho. Additional CTD work, including in vitro kinase assays, were performed by Corey M. Nemec and Aseem Z. Ansari. Proteomic work was conducted by Anna Merrill, M. Violet Lee, and Joshua J. Coon. Expression analysis was conducted by David B. Berry, Matthew E. MacGilvray, James Jose, and Jessica L. Will.

C.1 Introduction

All cells respond to stress by orchestrating complex responses customized for each situation. When grown in optimal conditions, *Saccharomyces cerevisiae* maintains high expression of growth-related genes and low transcription of stress-defense genes, in part via nutrient responsive TOR and RAS-regulated Protein Kinase A (PKA) signaling (Broach, 2012; Smets et al., 2010). Suboptimal conditions suppress these pathways in an unknown manner while activating stress-specific signaling networks that coordinate changes in transcription and translation, protein function, and metabolic fluxes with transient arrest of growth and cell cycle progression. How these disparate physiological processes are coordinated is poorly understood but likely critical for surviving and acclimating to stressful conditions.

At the level of gene expression, stressed yeast activate condition-specific transcript changes that provide specialized stress defenses. These responses are typically regulated by condition-specific transcription factors (TFs) and upstream signaling pathways that are activated under limited circumstances (Hohmann and Mager, 2003). Concurrently, stressed yeast activate the common environmental stress response (ESR) (Causton et al., 2001; Gasch et al., 2000). The ESR includes ~300 induced (iESR) genes that are broadly involved in stress defense and ~600 repressed-ESR (rESR) genes that together encode ribosomal proteins (RPs) and proteins involved in ribosome biogenesis/protein synthesis (RiBi). While the complete set of ESR regulators remains elusive, it is clear that the program is regulated by different upstream signaling factors under different situations (Gasch, 2002; Gasch et al., 2001; Gasch et al., 2000). Activation of the ESR, and of transcript changes more broadly, is in fact not required to survive the initial stressor, but rather is necessary for acquired resistance to subsequent stress (Berry and Gasch, 2008; Berry et al., 2011; Mitchell et al., 2009; Westfall et al., 2008). Therefore, screens for mutants sensitive to a single dose of stress have likely missed many signaling proteins, rendering stress-dependent signaling networks

incomplete. Although several isolated ‘pathways’ are well characterized, how signaling is integrated through a single cellular system is poorly understood.

Here we present an experimental and computational pipeline to infer the complete sodium chloride (NaCl)-activated signaling network from a combination of data types. A key feature of our approach is that we generated several large-scale datasets (including mutant transcriptome profiles, phospho-proteome changes, and gene fitness contributions) under the same culture system in cells responding to acute NaCl stress. Because stress responses are highly context dependent (Berry and Gasch, 2008; O'Rourke and Herskowitz, 2004; Van Wuytswinkel et al., 2000), we restrict our analysis to datasets generated in our own lab, despite many insightful prior studies characterizing the salt response in yeast (e.g. (Capaldi et al., 2008; Causton et al., 2001; Halbeisen and Gerber, 2009; Hirasawa et al., 2006; Martinez-Montanes et al., 2010; Melamed et al., 2008; Miller et al., 2011; Soufi et al., 2009; Warringer et al., 2010; Westfall et al., 2008)).

We wished to develop a computational method to integrate these datasets and infer the stress-activated signaling subnetwork, both to implicate missing regulators and to understand their connections. Prior approaches tackling the challenge of network inference have leveraged large-scale biological datasets, most commonly transcriptome data (see (Friedman, 2004; Schadt et al., 2005)). Extensions focusing on the osmotic response include the work of Gat-Viks *et al.,* whose probabilistic method described regulatory relationships between known regulators of the Hog pathway, assuming a known network topology (Gat-Viks and Shamir, 2007; Gat-Viks et al., 2006). Several approaches leverage protein-protein and protein-nucleic-acid interactions to infer relevant connections between regulators and their downstream gene targets (Huang and Fraenkel, 2009, 2012; Ideker et al., 2000; Liang et al., 1998; Markowetz et al., 2005; Novershtern et al., 2011; Suthram et al., 2008; Tu et al., 2006; Vaske et al., 2009; Yeang et al., 2004a; Yeger-Lotem et al.,

2009; Yeung et al., 2004). The method we present here is most closely related to methods that infer subnetworks by solving an integer linear program (IP) (Gitter et al., 2011; Ourfali et al., 2007; Silverbush et al., 2011). In particular, (Gitter et al., 2013) developed a combined probabilistic/IP method to discern signaling in the potassium chloride-responsive subnetwork from time-series expression data (Gitter et al., 2013). However, their approach incorporated transcriptome data only, whereas we were interested in incorporating other data types. Methods that integrate disparate datasets are emerging, for example the work of Huang *et al* (2013) that considered existing transcriptomic and proteomic data to study oncogene-induced signaling (Huang et al., 2013). In our case, we wanted to design a method that could also take mutant transcriptome profiles generated in our own lab.

We therefore designed an integer linear programming (IP) approach to integrate and interpret our disparate datasets by inferring a signaling subnetwork. The novel facets of our computational approach include a means to integrate these varied data sources, using new types of input paths to the IP, and a multi-part objective function. The resulting subnetwork generated many new insights into stress signaling, by implicating new regulators, unveiling the connections between them, and presenting principles that shed light on stress biology.

C.2 Generation of inferred network coordinating the response to osmostress

We previously identified 225 genes important for acquired stress resistance after NaCl pretreatment (Berry et al., 2011), including a subset of the known signaling proteins activated by NaCl (Figure C.1). Because only a fraction of NaCl-dependent transcript changes are important for acquired stress resistance, the selection misses many of the upstream transcriptome regulators. Therefore, to implicate the complete upstream signaling subnetwork, we began by profiling NaCl-dependent expression changes in 16 mutants implicated in NaCl-induced acquired

stress tolerance (Figure C.1, see Methods). Together, this generated a matrix of regulator-gene target predictions that encompassed 3,300 genes (Figure C.2 and Table C.1). A third of the affected genes were dependent on ≥2 regulators, and there was significant overlap in several target-gene sets (hypergeometric test, Figure C.1). These results hint at the complex upstream signaling that controls the NaCl-responsive transcriptome.

Because much of signal transduction occurs post-translationally, we next measured changes to the phospho-proteome before and at 5 and 15 min after NaCl treatment, using chemical isobaric tags for phospho-peptide quantification (see Methods). Nearly 600 of 1937 identified phosphosites (mapping to 973 proteins) showed a ≥ 2-fold change in phosphorylation, roughly split between sites with increased and decreased modification (Figure C.3). Over 10% of the altered phospho-proteins represented kinases and phosphatases (including regulators of cell-cycle progression, actin organization, and signal transduction) as well as transcriptional regulators (such as activators Hot1, Sko1, and Sub1 and repressors Mot2, Dot6, and Dig1). Proteins affected at the later time point were involved in cytokinesis, bud-site selection, and actin reorganization (Bonferroni-corrected $p < 0.01$, hypergeometric test), implying downstream physiological effects on these processes.

This analysis generated a rich source of datasets (outlined in Figure C.2). To integrate and interpret these disparate datasets, we designed an integer linear programming-based (IP) approach (Figure C.3 and Methods). Using a *background network* of physical or chemical protein interactions, the method infers a *subnetwork* that predicts the *paths* by which each interrogated *source* regulator is connected to its downstream *targets* (identified as dysregulated genes in the *source* mutant responding to NaCl treatment). Each path is a directed, linear chain of interactions between yeast proteins, where the terminal protein node represents a sequence-

specific transcription factor (TF) or RNA-binding protein (RBP) known to bind the downstream promoters or transcripts, respectively. The IP's objective function favors the inclusion of salt-responsive proteins, *i.e.* those with differential phosphorylation or required for acquired-stress fitness after NaCl treatment, and allows the sparing inclusion of additional proteins.

Specifically, we start with a background network of directed and undirected intracellular interactions representing protein-protein, kinase-substrate, and gene regulatory interactions between proteins and genes/mRNAs (Abdulrehman et al., 2011; Breitkreutz et al., 2010; Everett et al., 2009; Fasolo et al., 2011; Guelzim et al., 2002; Heavner et al., 2012; Hogan et al., 2008; Huebert et al., 2012; MacIsaac et al., 2006; Ptacek et al., 2005; Pu et al., 2009; Scherrer et al., 2010; Sharifpoor et al., 2011; Stark et al., 2006; Tsvetanova et al., 2010; Venters et al., 2011). For each interrogated source regulator, we identify candidate TFs and RBPs whose known binding targets significantly overlap with the source's targets (Figure C.3A). We then enumerate all possible directed candidate paths (using an iterative deepening search up to a given length) that connect each of the 16 interrogated *source* regulators to the majority of their *targets*, through candidate TFs or RBPs (Figure C.3B). Other candidate paths connect proteins required for fitness (Figure C.3B, blue nodes), proteins with NaCl-dependent phosphorylation changes (yellow nodes), and two known upstream sensors (pink nodes). The candidate paths serve as input to the IP, which encodes the relevance of each network element as a binary variable and characterizes possible subnetworks using a set of linear constraints over these variables (Figure C.3C). Subnetwork inference is performed by choosing a union of relevant, directed paths that optimize a series of successively applied objective functions that aim to connect experimentally implicated proteins while minimally including proteins not currently supported by experimental evidence. Because many distinct subnetworks may score equally well, we use the IP to identify an ensemble of high-scoring subnetworks. In turn, each protein, interaction, and path is assigned a confidence value based on its frequency across the ensemble.

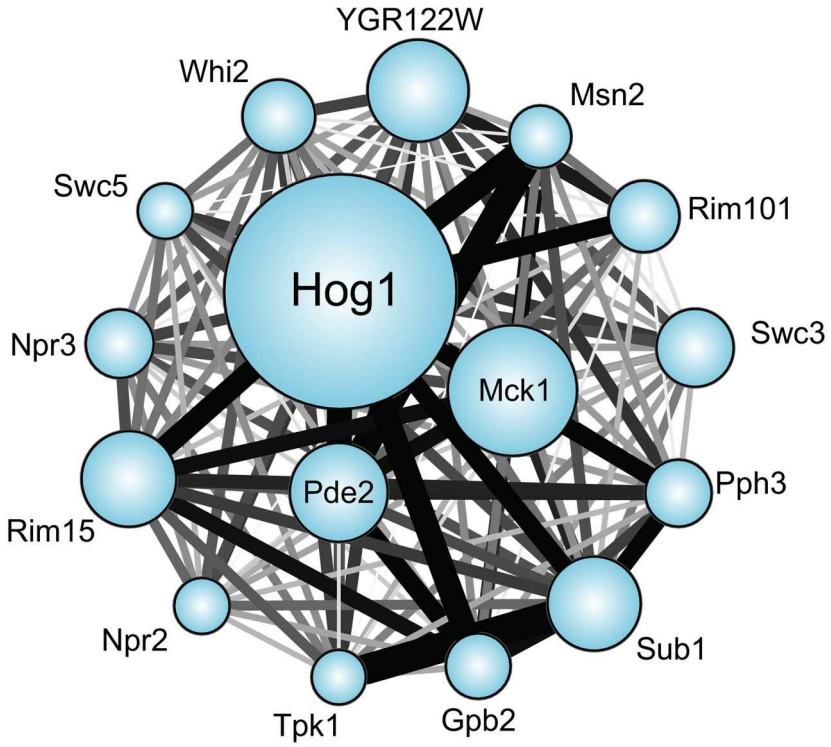

Figure C.1 Overlapping targets of interrogated 'source' regulators
The number of genes whose osmotic response was defective in each of 16 mutants is represented by the size of each circle. Edge thickness represents the fraction of the smaller node's targets that overlap between two nodes. Edge color is proportional to significance of the overlap (hypergeometric test), where black represents a –log(p-value) of 5 or greater.

Table C.1 Gene targets identified in regulator mutants

Mutant[a]	Defective[b]	Amplified[b]
hog1Δ (3)	1378	565
pde2Δ (3)	517	59
mck1Δ (3)	794	101
msn2Δ (3)	184	26
rim101Δ (3)	75	227
gpb2Δ (2)	202	37
rim15Δ (2)	438	106
npr2Δ (2)	75	69
npr3Δ (2)	184	89
swc3Δ (2)	108	257
swc5Δ (2)	84	55
whi2Δ (2)	118	201
pph3Δ (2)	235	21
sub1Δ (2)	431	97
tpk1Δ (2)	35	96
ygr122wΔ(2)	106	502

[a] Mutant and number of replicates in parentheses. [b] Number of genes with smaller ('defective') or larger ('amplified') expression changes compared to the wild type strain. Note this table includes non-coding RNAs that were excluded from the inference.

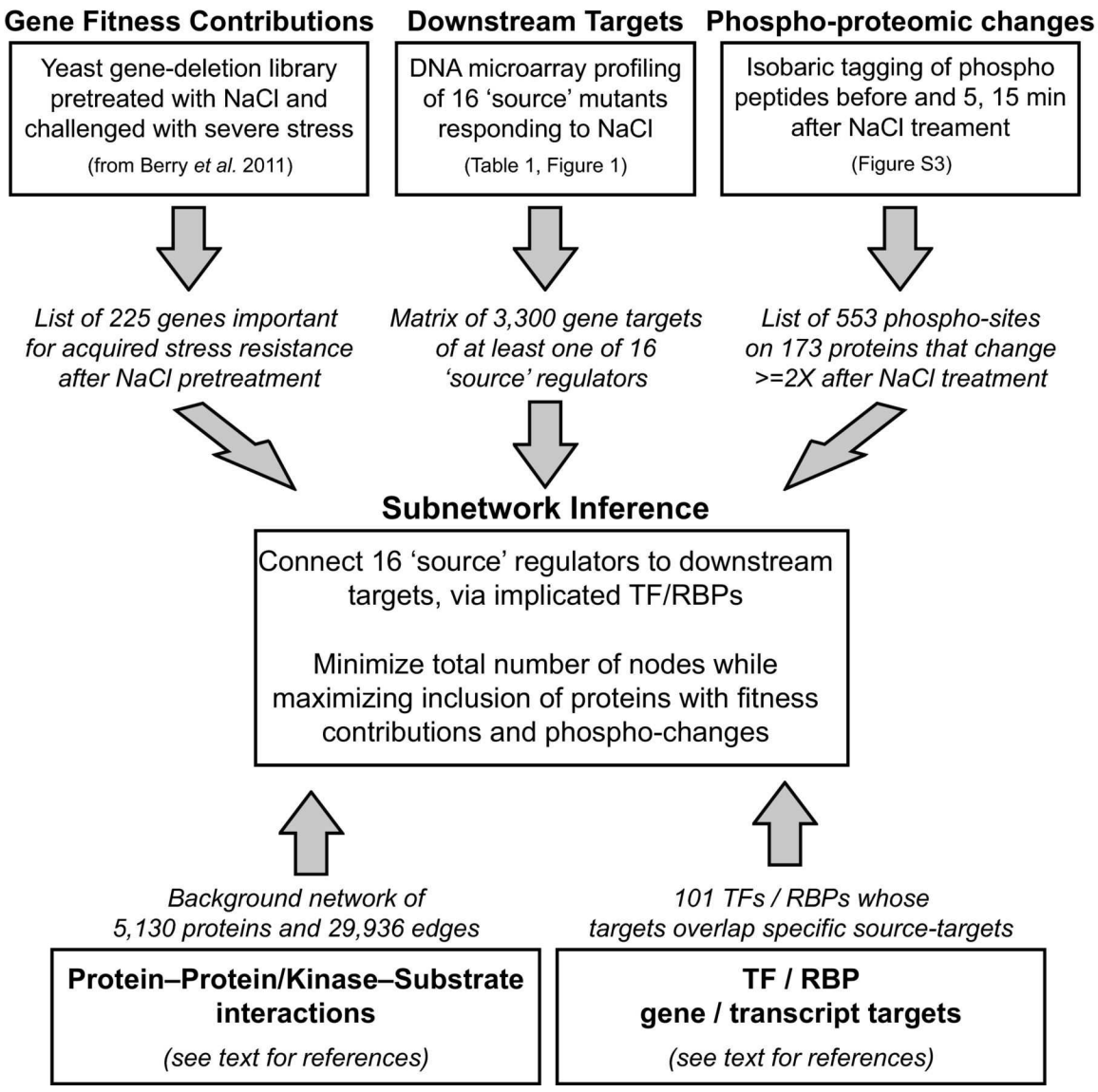

Figure C.2 Overview of the experimental data collection and analysis to generate IP input
See text for details.

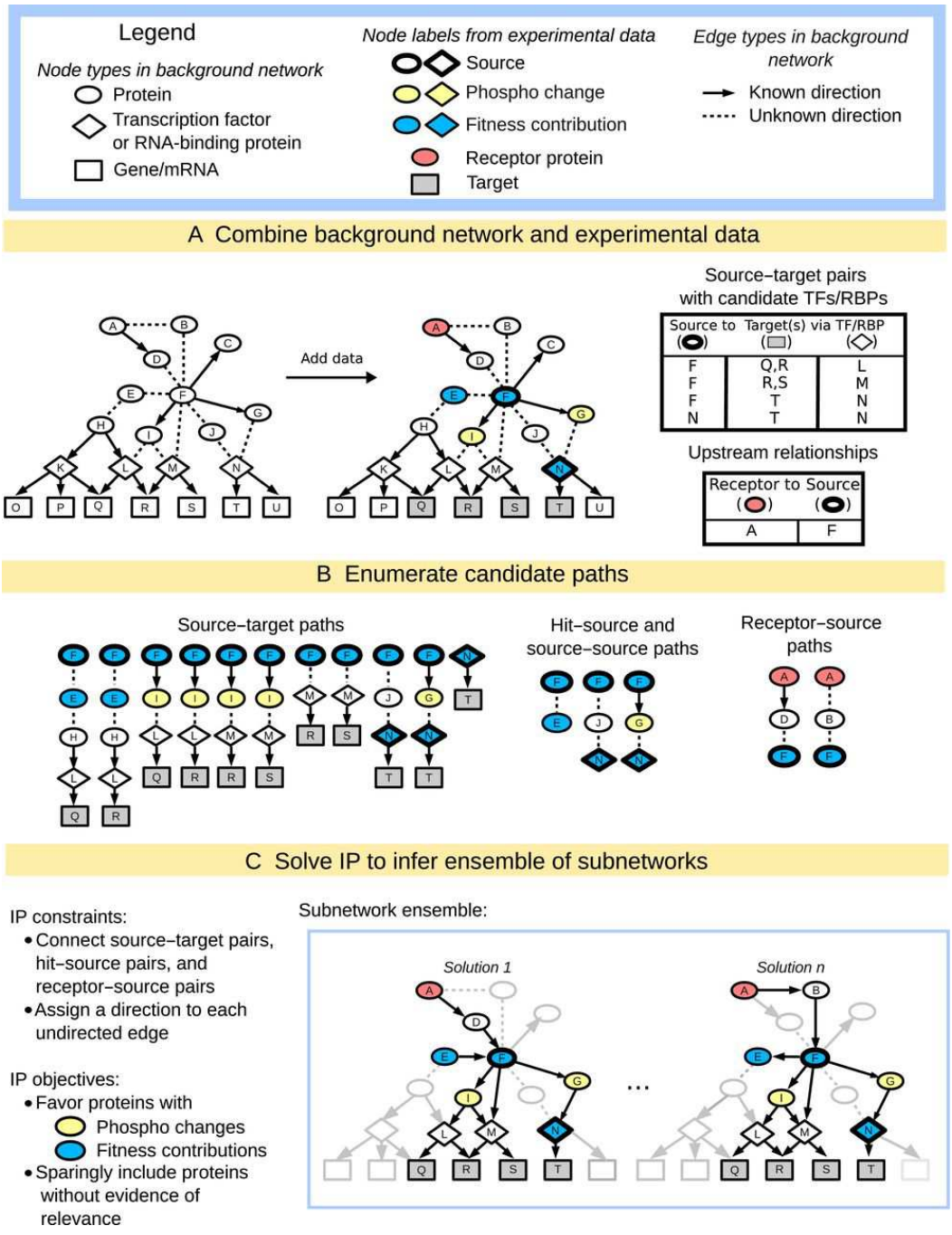

Figure C.3 Overview of the subnetwork inference method

A) The input to the method includes a background network of yeast interactions combined with experimental data that describes the yeast salt stress response, including proteins with phospho-changes (yellow), fitness contribution (blue), or two known upstream regulators (pink), as described in the key.

B) The three different types of paths that we enumerate using the background network and experimental data, where 'hit' refers to proteins identified in the original fitness screen or with significant changes in phosphorylation.

C) The IP for subnetwork inference and the output ensemble of inferred subnetworks.

C.3 Validation analysis provides strong support for the inferred subnetwork

Using the datasets described above, the method identified a consensus subnetwork encompassing 380 nodes (predicted regulators) and 1131 edges (relevant interactions) present at 75% confidence (Figure C.4A). To assess the inferred subnetwork's predictive accuracy, we performed precision-recall analysis using an assembled list of known NaCl regulators and another list of unlikely regulators that included metabolic enzymes and exclusively subcellular proteins. We excluded from consideration proteins with phospho-changes or fitness contributions (since they are preferentially included by the inference) and plotted the precision and recall over varying node-confidence thresholds (Figure C.4B). The inferred ensemble achieved substantially higher accuracy than the enumerated candidate paths provided as input to the IP method, highlighting the power of the inference step (Figure C.4B, green line). To assess the effects of the topological properties in the background network, we ran the method on permuted source-target pairs (maintaining the degree distribution from the real data; see Methods). This permuted baseline achieved high accuracy in the low-recall range, suggesting that some regulators are highly central in the background network. However, our inferred ensemble significantly outperformed the permuted baseline at higher levels of recall; thus, our method's accuracy is not simply due to properties of the background network's topology.

We found additional support for the inferred subnetwork in the non-random inclusion of specific protein functional groups. When compared to the background network, to the enumerated candidate pathways used as input to the IP, and to the permuted subnetworks, the inferred consensus subnetwork was enriched for proteins annotated as 'stress' proteins (background, p = 5e-21; candidates, p = 2e-6; permutations, p = 0.007) and for proteins encoded by genes with genetic interactions (background and candidates, $p \approx 0$; permutations, p = 0.003) (Stark et al., 2006), which suggests functional dependencies. The consensus subnetwork was also slightly enriched for kinases (relative to the candidate paths and background network) and for essential

genes (relative to the background network), but not relative to the permuted subnetworks (suggesting its bias toward kinases and essential genes).

The inferred subnetwork included many regulators not previously linked to the NaCl response. To test some of the novel predictions, we analyzed osmo-dependent transcriptome changes in 14 mutants lacking predicted regulators, with preferences for kinases and phosphatases (see Table C.2). The results provided strong support overall for the inferred subnetwork. All but one of the mutants (93%) displayed a defect in osmo-responsive expression. Furthermore, the predicted targets of 80% of these regulators overlapped significantly (p < 1e-3) with their measured targets, highlighting the accuracy of regulator-target predictions. To garner support for the subnetwork's structure, we investigated the overlap in targets of each interrogated mutant and the known or measured targets of proteins predicted to lie in the interrogated regulator's paths. Using stringent scoring, we found support for 30-100% of nodes in most paths (53% on average, Table C.1). Together these results provide strong support for the validity of the inferred consensus subnetwork.

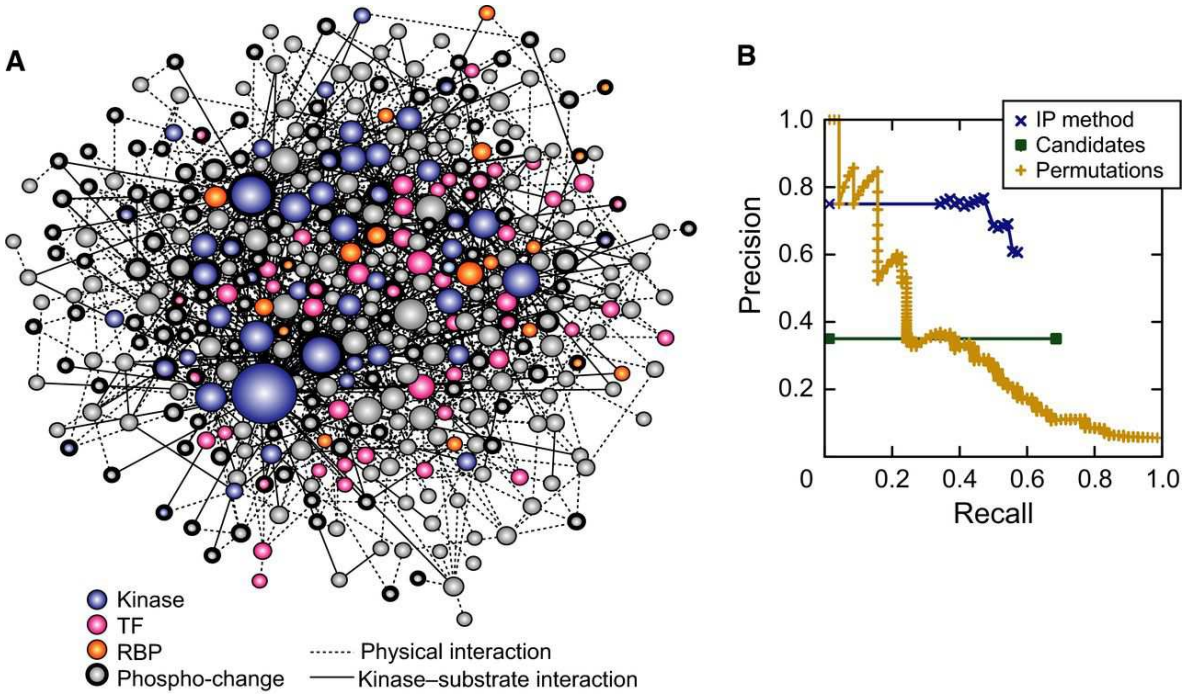

Figure C.4 Inferred NaCl-activated signaling network

A) Inferred consensus subnetwork at 75% confidence, where node size indicates degree (number of connections) and color is according to the key. Nodes representing proteins with phospho-changes are outlined in bold.

B) Precision-recall of the inferred consensus network was calculated using a list of true positives from the literature and a list of likely negatives, after excluding proteins with phospho-changes and those required for fitness (see Methods). *Precision* is the fraction of predicted nodes known to be involved in the osmo response, and *recall* is the fraction of true positives that are above the threshold. The curves represent the performance of the IP method on the real data (blue), of the method on randomized permutations of the input network (yellow), and of the candidate enumerated pathways used as input (green, see Methods).

Table C.2 Gene targets identified for validation mutants

Mutant	Defective	Amplified
cdc14-3 (3)	929	346
nnk1Δ (1)	94	278
bck1Δ (1)	107	169
yak1Δ (1)	226	248
kin2Δ (1)	52	266
pho85Δ (1)	614	342
cka2Δ (2)	155	63
cka1Δ (2)	58	133
ckb1Δcbk12Δ (2)	129	176
arf3Δ (2)	466	331
scd6Δ (2)	0	0

* *cdc14-3* was compared to its isogenic and identically treated wild type.

C.4 Known and new players captured in the NaCl-responsive signaling subnetwork

We therefore explored the consensus subnetwork for new insights into stress signaling. Many expected pathways were captured, including the canonical HOG, PKA, and TOR pathways. The inferred subnetwork included other stress-activated pathways not previously linked to the NaCl response, such as PKC, Pho85, Rim15 pathways and GSK-3 kinase Mck1 (Figure C.5A). We tested the involvement of these pathways by analyzing our phospho-proteomic data and mutant transcriptome profiles: we found that members of all of these pathways showed NaCl-dependent phospho-changes, and cells lacking specific pathway members (including *BCK1*, *YAK1*, *PHO85*, *RIM15*, and *MCK1*) had defects in NaCl-dependent expression changes. The subnetwork also included the 'STE' mating pathway, which shares upstream components with the Hog network and is known to be suppressed by Hog1 signaling (Marles et al., 2004; McClean et al., 2007; Nagiec and Dohlman, 2012; O'Rourke and Herskowitz, 1998; Patterson et al., 2010; Shock et al., 2009; Zarrinpar et al., 2004). The inclusion of the mating pathway indicates that some connections in the consensus subnetwork represent signaling suppression that prevents crosstalk to other pathways. We also validated several newly implicated regulators, including the CK2 kinase complex and the Cdc14 phosphatase (see below).

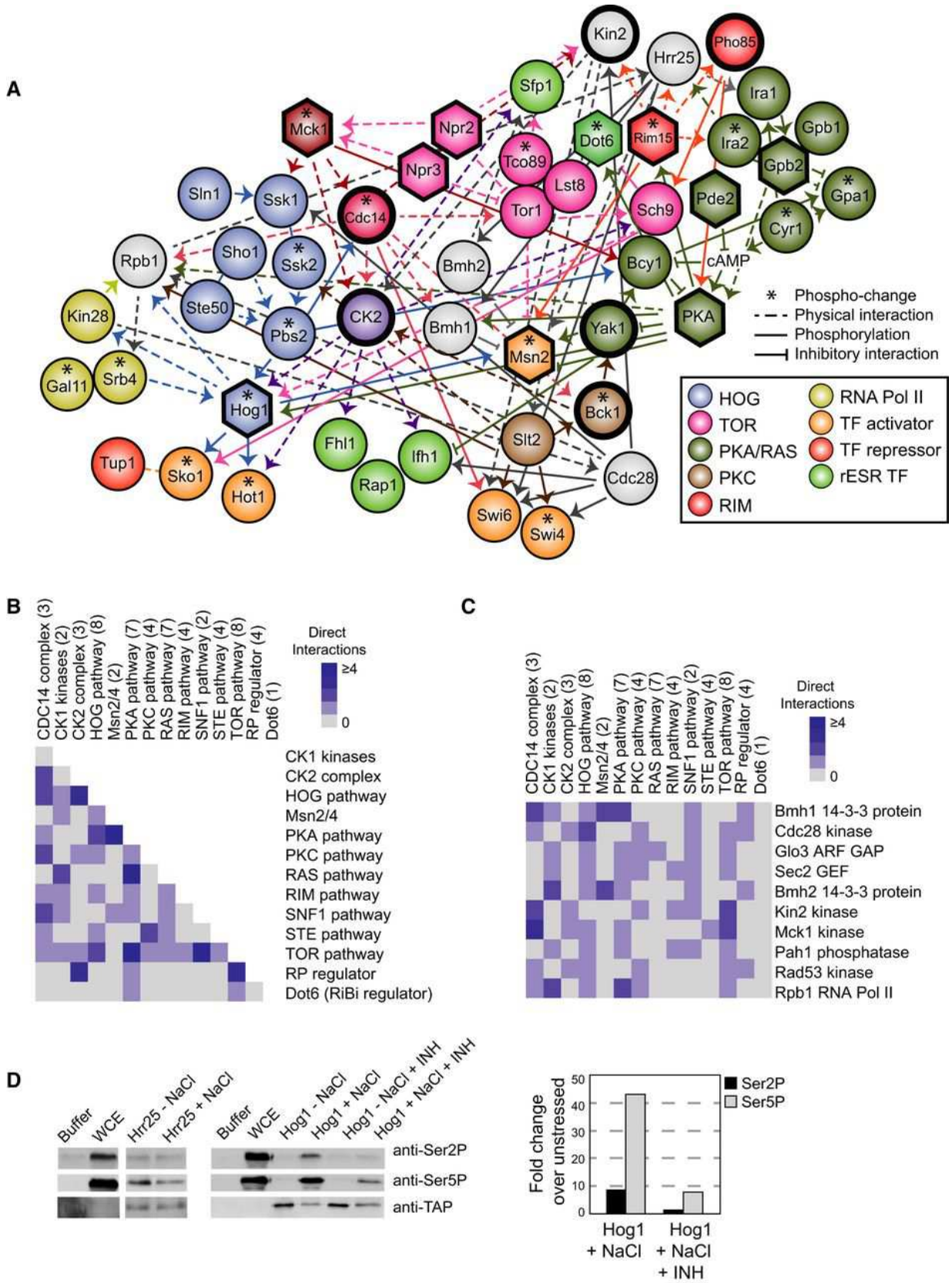
A
Kin2
Hrr25
Pho85
Sfp1
Ira1
Mck1
Npr2
Dot6
Rim15
Gpb1
Ira2
Tco89
Npr3
Lst8
Gpb2
Gpa1
Sln1
Ssk1
Cdc14
Tor1
Sch9
Pde2
Cyr1
Rpb1
Sho1
Ssk2
Bmh2
Bcy1
cAMP
Kin28
Ste50
CK2
Bmh1
Yak1
PKA
Pbs2
Msn2
Gal11
Srb4
Hog1
Bck1
Fhl1
Slt2
Tup1
Sko1
Ifh1
Cdc28
Rap1
Hot1
Swi6
Swi4
Phospho-change
Physical interaction
Phosphorylation
Inhibitory interaction
HOG
TOR
PKA/RAS
PKC
RIM
RNA Pol II
TF activator
TF repressor
rESR TF
B
CDC14 complex (3)
CK1 kinases (2)
CK2 complex (3)
HOG pathway (8)
Msn2/4 (2)
PKA pathway (7)
PKC pathway (4)
RAS pathway (7)
RIM pathway (4)
SNF1 pathway (2)
STE pathway (4)
TOR pathway (8)
RP regulator (4)
Dot6 (1)
Direct Interactions
≥4
0
CK1 kinases
CK2 complex
HOG pathway
Msn2/4
PKA pathway
PKC pathway
RAS pathway
RIM pathway
SNF1 pathway
STE pathway
TOR pathway
RP regulator
Dot6 (RiBi regulator)
C
CDC14 complex (3)
CK1 kinases (2)
CK2 complex (3)
HOG pathway (8)
Msn2/4 (2)
PKA pathway (7)
PKC pathway (4)
RAS pathway (7)
RIM pathway (4)
SNF1 pathway (2)
STE pathway (4)
TOR pathway (8)
RP regulator (4)
Dot6 (1)
Direct Interactions
≥4
0
Bmh1 14-3-3 protein
Cdc28 kinase
Glo3 ARF GAP
Sec2 GEF
Bmh2 14-3-3 protein
Kin2 kinase
Mck1 kinase
Pah1 phosphatase
Rad53 kinase
Rpb1 RNA Pol II
D
Buffer
WCE
Hrr25 - NaCl
Hrr25 + NaCl
Buffer
WCE
Hog1 - NaCl
Hog1 + NaCl
Hog1 - NaCl + INH
Hog1 + NaCl + INH
anti-Ser2P
anti-Ser5P
anti-TAP
Fold change over unstressed
50
40
30
20
10
0
Ser2P
Ser5P
Hog1 + NaCl
Hog1 + NaCl + INH

Figure C.5 Connectivity between known pathways and hubs of signal integration

A) A subregion of the inferred subnetwork, highlighting proteins in known pathways according to the key. Hexagons represent interrogated '*source*' regulators, nodes outlined in bold indicate validated players in the NaCl response, and asterisks represent proteins with phospho-changes upon NaCl treatment. Dashed edges represent physical interactions and solid arrows indicate kinase-substrate relationships. Edge directionality is as predicted by the inference, and edge color is according to the edge's source node. Inhibitory edges were taken from the literature.

B) Connectivity between known pathways, where blue boxes represent the number of interactions between any members of two pathways. Pathway membership is indicated in parentheses.

C) The top 15-ranked 'integrator' nodes with connections to the greatest number of different pathways, as shown in B.

D) A purified CTD peptide was incubated with Hrr25-TAP or Hog1-TAP purified from cells with and without NaCl treatment for 10 min, incubated with and without the reversible p38-specific inhibitor SB203580 (INH) added *in vitro*. Reactions with buffer or yeast whole-cell extract (WCE) served as negative and positive controls, respectively. CTD phosphorylated on serine 2 (Ser2) or Ser5 was detected by immunoblotting (see Methods). TAP-tagged proteins were subsequently quantified on the same blot with the anti-TAP antibody. Quantification of Hog1 phosphorylation, shown to the right, was normalized to Hog1-TAP abundance and then to the corresponding unstressed sample.

C.5 Interconnectivity in the inferred signaling subnetwork

The structure of the subnetwork revealed surprising cross-connectivity between previously defined pathways. We defined stress-activated 'pathways' based on the literature, and then summed the number of direct connections between members of those pathways (Figure C.5B). Many of the pathways were intricately connected, with Tor1 and PKA pathways linked to the greatest number of other pathways. We also identified individual subnetwork nodes as 'integration' points, defined as nodes with the greatest number of connections to distinct pathways (Figure C.5C). Nearly half of the top ten integration nodes were kinases or phosphatases, including Mck1 and cell-cycle regulator Cdc28, which regulates RP genes under optimal conditions (Chymkowitch et al., 2012) but is suppressed during osmotic shock (Adrover et al., 2011; Alexander et al., 2001; Belli et al., 2001). 14-3-3 proteins Bmh1 and Bmh2 were also identified as integration points, confirming their known role as signaling cofactors.

Several of the integration points are also hubs of high connectivity in the consensus subnetwork. While 11 of the top 15 most connected nodes are kinases or phosphatases, the remaining four are known regulatory cofactors – including stress-activated ubiquitin (Ubi4), Sumo (Smt3), and Bmh1 – and the core subunit of RNA polymerase (Pol II), Rpb1. Modification of the Rpb1 carboxyl-terminal domain (CTD) is the basis for the so-called CTD code of transcriptional regulation (Buratowski, 2003; Zhang et al., 2012), making it a logical downstream integration point for complex upstream signaling. Consistent with the predictions of the subnetwork, we found that two of the Rpb1-interacting kinases – Hrr25 and Hog1 – phosphorylate the Rpb1-CTD *in vitro*. TAP-tag-purified Hrr25-TAP phosphorylated Rpb1-CTD serine 5 (Ser5), regardless of prior NaCl treatment (Figure C.5D). In contrast, TAP-purified Hog1-TAP phosphorylated both Ser2 and Ser5, but only after cellular NaCl treatment and in a manner inhibited by a Hog1-specific inhibitor added *in vitro*. Both Hrr25 and Hog1 are known to interact with Pol II and influence transcriptional processes (Alepuz et al., 2003; Cook and O'Shea, 2012; Nadal-Ribelles et al., 2012; Phatnani et

al., 2004; Proft et al., 2006) but neither had been implicated in direct Rpb1-CTD phosphorylation. These results are consistent with the model that the Rpb1-CTD is a direct target of the signaling network and plays a central role in signaling (see more below).

We also dissected the regulatory connections surrounding a second hub, Cdc14. In the process, we found that Cdc14 is critical for coordinating distinct facets of the NaCl response. First, the defect in NaCl transcriptome changes evident in *cdc14-3* cells overlapped significantly with the Hog1 response, raising the possibility that Cdc14 is important for Hog1 regulation (Figure C.4). The subnetwork predicts that Cdc14 is activated in part by the Hog1 regulator Pbs2 (reminiscent of Pbs2 control of Cdc14 localization during the cell cycle (Reiser et al., 2006)) and that Cdc14 affects Hog1 function via the nuclear exporter, Crm1. This prompted us to follow Hog1 localization in the *cdc14-3* mutant. Indeed, Hog1 nuclear localization was defective in the NaCl-treated *cdc14-3* mutant (Figure C.6A), despite Hog1 hyper-phosphorylation under these conditions (Figure C.6B). We found no direct interaction between Cdc14 and Hog1, suggesting that the hyper-phosphorylation of Hog1 is a secondary response to the defect in nuclear localization rather than a deficit of direct Hog1 dephosphorylation by Cdc14. The aberrant Hog1 localization was not a side effect of cell-cycle arrest, since we found no defect in wild-type cells progressing through G2/M phase or in nocodazole-arrested cells (Figure C.6). Instead, these results suggest a direct connection between Cdc14 activity and signaling through the Hog pathway.

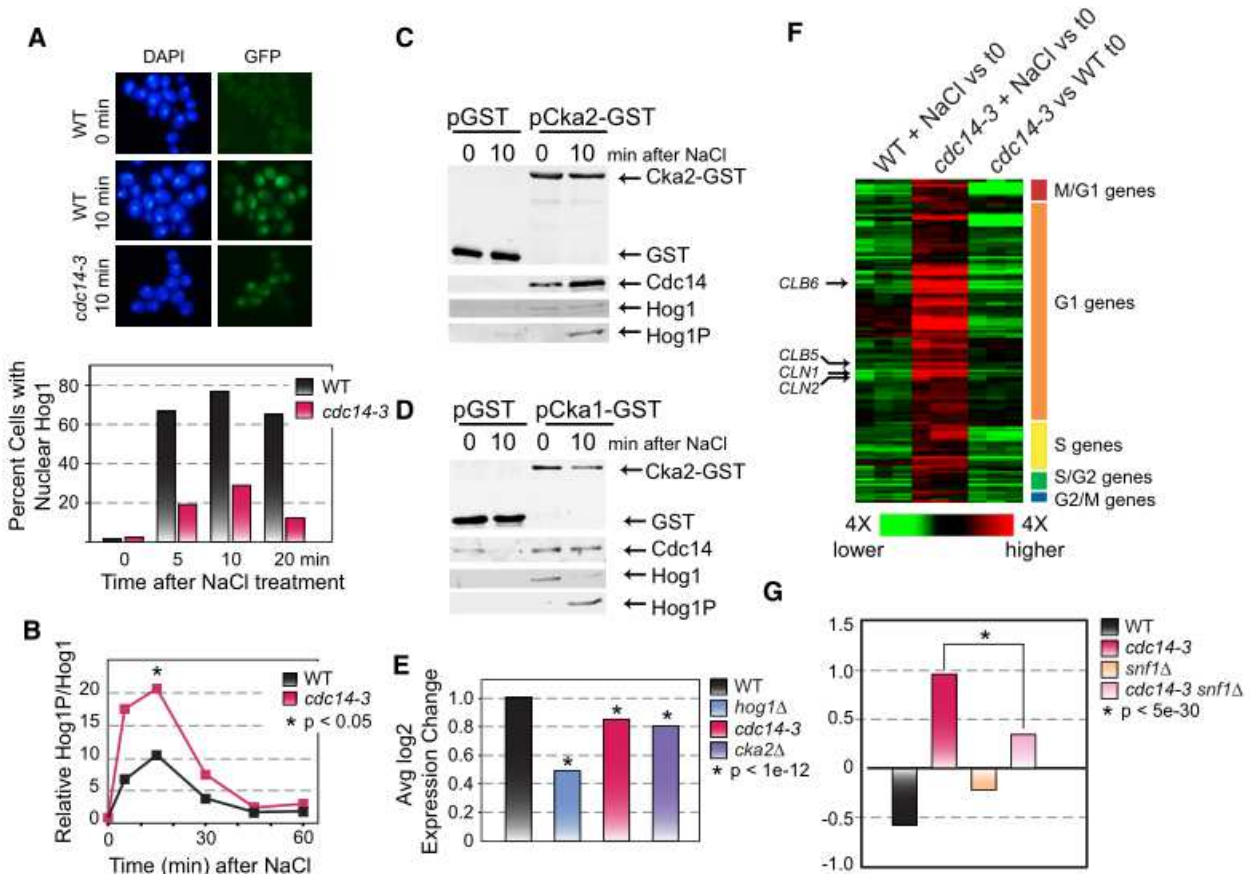

Figure C.6 Cdc14 is a central regulator in the NaCl response

A) Wild-type (WT) and *cdc14-3* cells were shifted to 35°C for 90 min and then exposed to 0.7M NaCl for up to 20 min. Images represent nuclear DNA (DAPI, left) and Hog1-GFP (right) before and at 10 min after NaCl treatment. The plot below quantifies the fraction of cells (n = >75) with nuclear Hog1-GFP signal that overlapped DAPI signal in WT and *cdc14-3* cells.

B) Levels of phospho-Hog1 normalized to total Hog1 in WT and *cdc14-3* cells responding to NaCl at 35°C. Data represent the average of biological duplicates.

C) GST-tagged Cka2 was immunoprecipitated and blotted for Cdc14 and total or phospho-Hog1.

D) Same as C) but for immunoprecipitated Cka1-GST.

E) The average $\log_2$ fold-change of 67 Hot1 targets in replicated WT, *hog1Δ*, *cdc14-3*, and *cka2Δ* strains responding to NaCl. Data for each mutant and its paired WT were scaled to the plotted WT so as to accurately represent the mutant defect. Asterisks represent a significant difference in the mutant versus its paired WT (paired T-test).

F) Expression data in WT or *cdc14-3* cells responding to NaCl at the nonpermissive temperature and in *cdc14-3* cells versus WT at the non-permissive temperature before NaCl addition. Each column represents one of three triplicated expression responses, and each row represents one of 131 cell cycle genes aberrantly induced in *cdc14-3* after NaCl treatment (FDR<0.05). Red represents higher and green represents lower expression in response to NaCl (or in the *cdc14-3* mutant in the case of the last columns), according to the key. Cell-cycle classification of the genes (Spellman et al., 1998) is shown to the right; cyclins are annotated to the left.

G) Average $\log_2$ expression change of genes shown in F, as described in E.

Second, the subnetwork predicts that Cdc14 regulates CK2 subunits to modulate the Hog1-regulated TF, Hot1. We uncovered a salt-enhanced interaction between Cdc14 and CK2 subunit Cka2 (Figure C.6C), and uncovered a constitutive association between CK2 subunits Cka1/Cka2 and Hog1 (Figure C.6C,D). Although the connection between CK2 and Hog1 was not known in yeast, our results are reminiscent of regulation in mammalian systems, in which CK2 is regulated by the human ortholog of Hog1, p38 (De Amicis et al., 2011; Hildesheim et al., 2005; Isaeva and Mitev, 2011; Sayed et al., 2000). As predicted by the subnetwork, we found that Cdc14, Cka2, and Hog1 were all required for normal induction of Hot1 targets (Figure C.6E).

Finally, and surprisingly, we discovered that Cdc14 suppresses NaCl-dependent crosstalk to the cell-cycle network: the *cdc14-3* mutant at the nonpermissive temperature strongly and aberrantly induced G1 and S-phase genes upon NaCl treatment (Figure C.6F), even though cells were completely arrested in M-phase for the duration of the treatment. This included genes encoding G1 and S-phase cyclins *CLN1/2* and *CLB5/6*, respectively. To further understand this effect, we turned to the subnetwork: Cdc14 is predicted to affect these genes via direct interaction with the carbon-responsive kinase Snf1, which is known to be activated by NaCl (Hong and Carlson, 2007; Ye et al., 2008). Snf1 is also required for proper timing of cell-cycle entry in standard conditions (Busnelli et al., 2013; Pessina et al., 2010), raising the possibility that it is responsible for the inappropriate G1/S-gene induction in the absence of Cdc14. We found that deletion of *SNF1* in the *cdc14-3* background largely abrogated the hyper-activation of G1 and S genes in the *cdc14-3* mutant (Figure C.6G). This presents a model for future dissection, in which Cdc14 helps to suppress the cell-cycle effect of Snf1 activation, thereby funneling Snf1 activity toward its stress-specific gene targets. Together, our results demonstrate the remarkable and central role of Cdc14 in coordinating cellular signaling upon osmotic shock, while showcasing the predictive power of our inferred subnetwork.

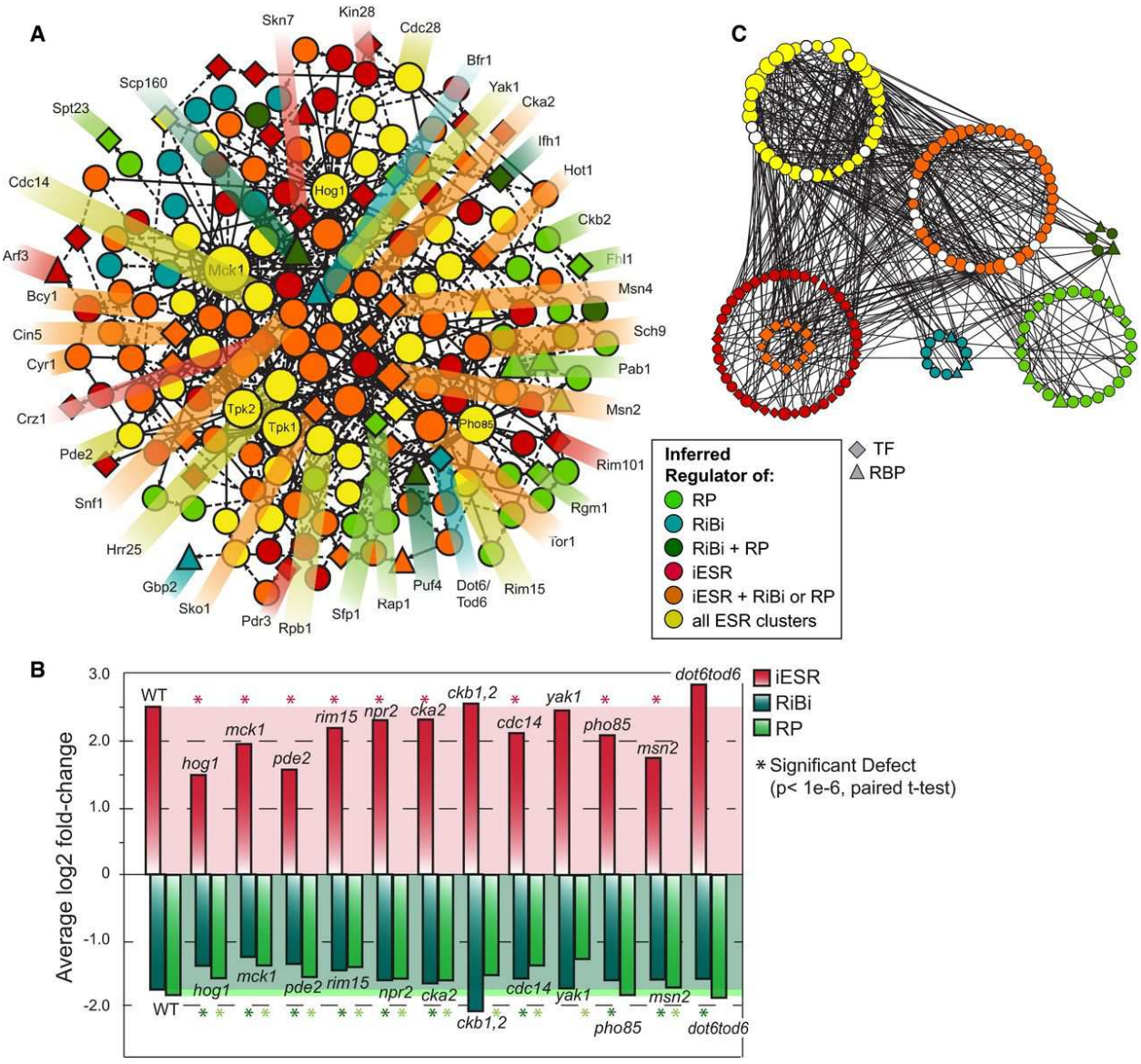

Figure C.7 Inferred ESR regulatory subnetwork

A) Regulators predicted to lie upstream of iESR, RP, and/or RiBi ESR modules are color-coded according to the key and sized according to degree (number of connections). Diamonds and triangles represent TFs and RBPs, respectively.

B) Average $\log_2$ expression changes of iESR, RP, or RiBi genes in mutants responding to salt. Wild-type (WT) levels are highlighted by shaded areas; genes with a significant defect (p<1e-6, paired T-test) are indicated with an asterisk.

C) Same as A) but organized by ESR regulatory potential. Top-ranked bifurcation nodes discussed in the text are colored white. Nested orange nodes represent iESR TFs that are also predicted to lie upstream of RP paths.

C.6 New insights into ESR regulation and coordination

We were especially interested in how distinct modules in the ESR – including iESR genes important for stress defense and RP/RiBi modules required for rapid growth – are regulated and coordinated. Of the 178 nodes implicated in ESR regulation, over half were predicted (Figure C.7A) – and several confirmed (Figure C.7B) – to lie upstream of all three ESR modules. In contrast to common upstream nodes that were enriched for kinases compared to the consensus subnetwork (p = 2.6e-7), nodes exclusive to iESR regulation were enriched for TFs (p = 5e-5), while rESR regulators showed a preponderance of RBPs (p = 1e-5), implicating regulated RNA stability for these genes. Many more regulators and regulatory connections were unique to the iESR versus RP and RiBi modules (the latter being the largest group) (Figure C.7C). This is consistent with the extensive redundancy in iESR control (Gasch, 2002) and hints at a more monolithic regulation of rESR expression during times of adversity.

To better understand how cells coordinate repression of growth-related genes with induction of stress-defense genes in the ESR, we devised a bifurcation score based on information theory, to rank nodes that (a) are upstream of many genes from both modules but (b) have outgoing paths that relatively cleanly divide iESR and rESR genes. A third of top-15 ranked bifurcating nodes are linked to cAMP signaling (including adenylate cyclase Cyr1, cAMP response regulator Bcy1, and phosphodiesterase Pde2). Indeed, we found that the *pde2Δ* mutant has a defect in both iESR induction and rESR repression (Figure C.7B), confirming the role of cAMP in the growth/stress-defense decision (see Discussion). Nearly half of the remaining top-ranked bifurcation proteins associate with RNA Pol II (including Pol II core subunit Rpb3, Pol II associated Sub1 and Ask10, transcription elongation factor Spt5, as well as Sds3 of the Rpd3L chromatin remodeling complex). Together with the identification of Pol II subunit Rpb1 as a hub in the subnetwork, these results implicate RNA Pol II at a key decision point in ESR coordination.

To investigate this, we started by checking *in vivo,* bulk modification of Rpb1-CTD in wild-type and *hog1Δ* cells responding to NaCl. The *hog1Δ* mutant showed an initial drop in Ser5 and Ser2 phosphorylation similar to the wild type, but displayed a reproducible defect in the normal subsequent transient increase in Ser5 and Ser2 phosphorylation (Figure C.8A). The timing of the transient peaks in bulk Rpb1-CTD phosphorylation correlates with the timing of transcription initiation and elongation upon osmotic stress (Berry and Gasch, 2008; Lee et al., 2011; Miller et al., 2011), consistent with the known roles of Hog1 as well as Ser 5 and Ser2 phosphorylation in these processes (Alepuz et al., 2003; Proft et al., 2006; Zhang et al., 2012).

To test our hypothesis that direct modification of Rpb1-CTD is important for ESR regulation, we measured transcriptomic changes upon salt stress in Rpb1-CTD mutant strains that could not be phosphorylated normally on CTD-Ser2 or -Ser5 (S2A and S5A mutants, respectively). Since S2A or S5A substitution in all CTD repeats is lethal, the mutant cells expressed chimeric CTD sequences with half mutant and half wild-type repeat sequences (West and Corden, 1995). Neither mutant showed significant expression differences in the absence of stress, and the S2A mutant showed only a subtle defect in NaCl-dependent expression changes (Figure C.8B). In contrast, the S5A mutant had a significant defect in iESR induction and even more so in rESR repression, comparable to the defect seen in the *hog1Δ* mutant.

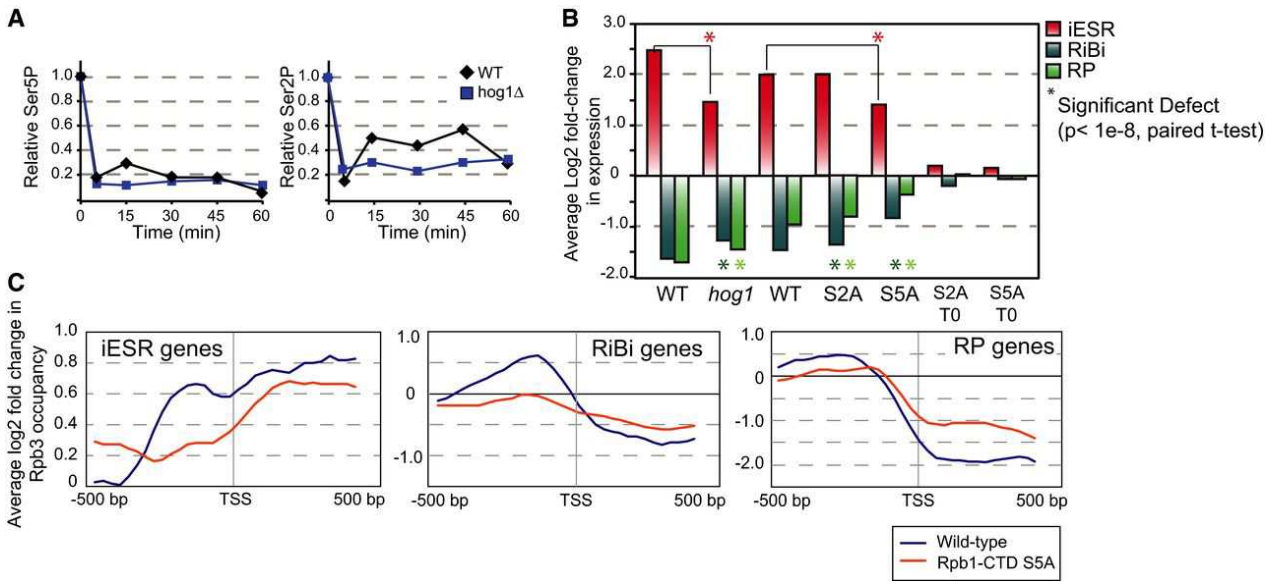

Figure C.8 Pol II CTD modification coordinates ESR regulation

A) Relative abundance of bulk Ser5P (left) and Ser2P (right) normalized to an internal Rpb3 loading control, from yeast cells treated with NaCl for the denoted times. Data are representative of several replicates.

B) The average log2(fold-change) expression of iESR, RP, and RiBi genes is shown for paired wild-type and hog1D strains (as in Figure C.7) and paired wild-type, S2A, and S5 strains. There was little expression difference in the S2A and S5A mutants versus wild type before stress (right bars).

C) Average $\log_2$(fold-change) in Rpb3 occupancy 20 min after NaCl treatment at +/- 500bp around the transcription start site (TSS) of iESR (left), RiBi (middle), and RP (right) genes. Profiles were reproducible across biological replicates.

We reasoned that aberrant ESR coordination may be caused by an inability of polymerase to re-localize from rESR genes, which are highly transcribed before stress, to stress-induced iESR genes. To test this, we measured chromatin occupancy of RNA Pol II subunit Rpb3 in both wild type and S5A mutant strains responding to NaCl stress, using ChIP-chip. In wild-type cells responding to stress, Rpb3 occupancy increased at iESR genes but decreased in the body of RP and RiBi genes, with slight accumulation in the promoter regions of specific repressed genes (Figure C.8C). In contrast, the S5A mutant showed a reproducible defect in Rpb3 recruitment to iESR genes and a concomitant defect in Rpb3 release from rESR genes (Figure C.8C). These results show that direct modification of the Rpb1-CTD is required for normal regulation of iESR and rESR genes (see Discussion).

C.7 The orthologous mammalian networks are enriched for growth-regulating and disease-causing genes

Striking the correct balance between growth rate and stress defense is fundamental for proper cellular function, and improper balance is thought to be a critical driver in diseases such as cancer (Jones and Thompson, 2009). We therefore interrogated the set of human genes orthologous to the yeast NaCl subnetwork. We found that this set is enriched for genes linked to cancer, mostly through somatic mutation, according to the COSMIC database (Forbes et al., 2011): of the 35 human genes in the COSMIC dataset with yeast orthologs, 8 were orthologous to nodes in the consensus-node network, representing a 2.5-fold enrichment above chance (p = 0.0068, Dataset 4A). We also compared the yeast network to Mendelian disease genes in the OMIM database (Hamosh et al., 2005). We identified 25 additional yeast genes whose orthologs are linked to heritable disease (Dataset 4B), with weak enrichment for genes associated with prostate cancer (p = 6e-3, (Woods et al., 2013)). The network was also enriched for yeast proteins whose mouse orthologs are required for pre/perinatal viability, normal growth rate and body size, and male and female fertility (FDR < 5%) (Woods et al., 2013). These results highlight that stress-responsive

signaling is likely important for proper regulation of growth rate, and thus may provide insights into cancer biology (see Discussion).

C.8 Discussion

A major challenge in network biology remains integrating disparate large-scale datasets in a manner that reveals new insights to biology. The approach we developed here provides a new route to identifying the extensive set of players activated during a response, as well as the connections between them and the flow of information toward the processes they regulate.

The computational approach we developed provides several contributions. First, we provide a means to selectively integrate disparate datasets via four types of paths between proteins in the background network. Each dataset is prioritized separately by a series of objective functions, whereas related approaches for inferring signaling networks use a single objective function. One class of related methods essentially maximizes the number of paths between sources and targets (Gitter et al., 2011; Ourfali et al., 2007; Yeang et al., 2004b; Yeang et al., 2005). In contrast, our method's preference for sparse inferred subnetworks is also employed by approaches based on the prize-collecting Steiner Tree algorithm (Huang et al., 2013; Huang and Fraenkel, 2009, 2012; Yosef et al., 2009) and flow-based algorithms (Lan et al., 2011; Yeger-Lotem et al., 2009). However, those methods require the use of a weight parameter to trade off between subnetwork sparsity and the inclusion of known relevant proteins. Another contribution of our approach is the representation of uncertainty in the underlying network. We assign a confidence value to each protein and interaction according to its frequency in an ensemble of optimal inferred subnetworks. This is similar to the score used by Yeang et al. (2005), who actually enumerate all optimal solutions; doing so is practically intractable for our input and our model. In contrast, Ourfali et al (2007) assign confidence values based on the change in objective value when each protein or interaction is individually excluded, and Gitter et

al. (2011) present several methods for ranking paths based on input experimental data and local topological features of the inferred subnetworks. The confidence values generated by our approach provide useful guidance for subsequent biological examination.

The resulting subnetwork put forward by our approach identified new regulators in the NaCl response and provides a glimpse of their connections in a single cellular signaling system. The extensive physical connectivity between what are traditionally considered ‘distinct’ pathways suggests much greater signaling integration than previously realized. The cross-connectivity between pathways, either direct or through apparent ‘integration’ points, may coordinate the magnitude or timing of signaling through distinct branches, prevent signaling crosstalk, and/or provide important feedback to dampen signaling as cells acclimate to new conditions (Schwartz and Madhani, 2004; Waltermann and Klipp, 2010). Our results implicate Cdc14 as a critical integrator that bridges HOG and CK2 signaling, suppresses inappropriate activation of the cell-cycle network, and connects to several other pathways including Tor1, which is reportedly suppressed by Cdc14 (Breitkreutz et al., 2010). Members of the growth-regulating TOR1 pathway as well as the RAS/PKA pathway show the greatest connectivity to other stress-activated pathways, suggesting that growth regulation is sensitively tuned according to stress conditions.

Our results also shed new light on how growth control and stress defense are related. Optimal growth and maximal stress tolerance are competing interests in the cell: the fastest growing cells are typically the least tolerant of adversity, whereas stress-resistant cells are frequently slow growing or arrested (Elliott and Futcher, 1993; Levy et al., 2012; Lu et al., 2009; Sumner and Avery, 2002; Zakrzewska et al., 2011). Our results here, along with prior studies, suggest that competition for cellular resources – namely those related to transcription and translation – drive the anti-correlated expression of genes involved in stress defense *versus* growth

promotion. Under optimal growth conditions, rESR transcripts are among the most highly transcribed and the most highly translated (Ingolia et al., 2009; Lipson et al., 2009), consuming the bulk of cellular ribosomes (Warner et al., 2001). We previously proposed that the drop in rESR transcripts helps to direct translational capacity to iESR genes by releasing sequestered ribosomes (Lee et al., 2011). Work by You *et al.* suggests that cAMP abundance dictates whether translational capacity is directed to growth versus other processes such as stress defense (You et al., 2013). That our results implicate cAMP in the iESR/rESR regulatory balance is consistent with these models.

In addition to implicating cAMP metabolism, our results show that direct regulation of the RNA Pol II CTD plays a crucial role in the iESR/rESR transcriptional balance, by triggering redistribution of polymerase from highly transcribed growth-related genes to stress-induced defense genes. The ability to fully phosphorylate Ser5 of the Rpb1-CTD is required for normal repression of rESR genes, as indicated by the defect in transcript repression and Pol II redistribution, and is also required for normal recruitment of Rpb3 to iESR promoters for gene induction (Figure C.8). Ser5 phosphorylation has been implicated in both gene repression and induction (Hengartner et al., 1998), in support of our findings. The stress-activated redistribution of Pol II from rESR to iESR genes is at least partly dependent on Hog1 (Cook and O'Shea, 2012; Nadal-Ribelles et al., 2012), which we show phosphorylates the Rpb1-CTD *in vitro* (Figure C.5D) and is required for its normal modification *in vivo* (Figure C.8A). Thus, we propose that direct regulation of RNA Pol II, perhaps in part by the Hog1 kinase, plays a central role in coordinating these opposing transcriptional modules.

Establishing the correct balance between stress tolerance and growth rate is critical for surviving fluctuating environments in nature. But the enrichment for cancer-causing genes in the orthologous human subnetwork highlights the importance of this decision in disease biology,

and it suggests that stress signaling in yeast may serve as a model for cancer signaling in humans. It is notable that orthologs of three key regulators in our network – Hog1, Cdc14, and CK2 – have all been implicated in regulating the mammalian tumor suppressor p53 (Bulavin et al., 1999; Li et al., 2000; Meek et al., 1990), which controls the growth/survival/apoptosis decision in human cells and is mutated in many human cancers (Carvajal and Manfredi, 2013). These results underscore the importance of the growth/survival decision, and hint that the yeast subnetwork could be used to implicate as-yet unidentified human disease genes. An exciting area of future study will be to distinguish signaling dynamics and condition-specific versus common aspects of the signaling, with an eye toward their role in disease biology.

C.9 Experimental Procedures

Growth conditions

All strains were of the BY4741 background, primarily from the deletion collection (Winzeler et al., 1999) (Thermo Scientific, Waltham, MA), except for *cdc14-3* and its isogenic wild type (kindly provided by M. Weinreich (Miller et al., 2009)). BY4741 *ckb1Δckb2Δ* was kindly provided by C. Guthrie (Bergkessel et al., 2011). Knockout strains were verified by diagnostic PCR to ensure correct integration of the drug cassette and to confirm absence of the deleted gene. Unless otherwise noted, cells were grown to log phase in batch YPD cultures at 30°C for at least seven generations before addition of a final concentration of 0.7M NaCl, after which cells were grown for 30 min. *cdc14-3* and isogenic wild-type cells were grown at 25°C, shifted to the non-permissive temperature of 35°C for 90 minutes (or 120 min for experiments from Figure C.6G), and then treated with a final concentration of 0.7M NaCl at 35°C for an additional 30 min before sample collection. Relative physiological changes were compared to the timepoint collected immediately before addition of NaCl (*i.e.* 35°C for 90 min without NaCl). Δ

Microarray analysis

Cell collection, RNA preparation, cDNA synthesis and labeling, array hybridization, and normalization were as previously described (Berry and Gasch, 2008; Lee et al., 2011), using cyanine dyes (Flownamics, Madison WI) and Superscript III (Life Technologies, Carlsbad, CA). Samples were hybridized to whole-genome tiled DNA microarrays (Roche Nimblegen, Madison, WI), comparing cDNA from the salt-treated sample to cDNA generated from the unstressed culture. Dye orientation was performed on select samples to assess dye-specific biases; dye orientation for paired mutant-wild type samples was maintained for statistical analysis to avoid dye-specific effects. Comparison of unstressed strains was done as previously described (Lee et al., 2011) by retrieving and comparing single-channel data from mutant and wild type arrays. Array data are available through the NIH GEO accession # GSE60613.

Genes whose expression was altered in wild-type cells responding to NaCl were identified based on five biological replicates, using the Bioconductor package *limma* (Smyth, 2004) and Q-value (Storey et al., 2005) to assess the false discovery rate (FDR) and taking q< 0.05 as significant. This analysis identified 5,056 genes with a significant change in expression in response to NaCl. Genes with a defect in NaCl-responsive expression in mutants shown in Figure C.1 were assessed in biological triplicate (for *hog1Δ, mck1Δ*, *pde2Δ, msn2Δ,* and *rim101Δ* strains) or duplicate (all other strains). Expression defects were identified using contrast matrices to wild-type expression in *limma* for triplicated samples, with q<= 0.025 taken as significant. For duplicated samples, expression defects were identified if both mutant replicates were outside the wild type mean + 2 standard deviations (95^{th} confidence level), based on 5 replicates of the wild-type samples. Identified targets (summarized in Table 1). Data for the *dot6Δtod6Δ* mutant were taken from (Lee et al., 2011).

Validation experiments were performed on 10 deletion mutants responding to NaCl. For samples done in duplicate, significant expression changes were identified using *limma* with q<0.05 taken as significant. Expression defects from singleton experiments were identified based on a 1.5-fold difference in expression in that mutant versus the paired wild-type sample. Identified expression defects are summarized in Table 1. Expression in unstressed mutant cells was also assessed by comparing the mutant response to unstressed wild-type cells as described above. Unless otherwise noted, we detected few expression differences in unstressed cells. For Figs. 6-8 where average expression values are plotted, data for some paired mutant-wild type experiments (namely *cdc14-3*, *dot6Δtod6Δ*, and *ck2* mutants) were scaled such that their paired wild-type data matched the plotted wild-type data taken from other experiments, in order to accurately represent those mutant defects by accounting for day-to-day variation of paired samples.

Expression analysis for Figure C.8 was done in strains generously provided by JL Corden (West and Corden, 1995). Cells lacking endogenous *RPB1* carried a plasmid expressing *RPB1* with 14 wild-type CTD repeats (YSPTSPS) or a plasmid expressing chimeric *RPB1* genes: the Rpb1-CTD was composed of 5 repeats of S5A (YSPTAPS) followed by 7 wild-type-sequenced repeats in the so-called S5A mutant, or 8 S2A repeats (YAPTSPS) followed by 7 wild-type sequenced repeats in the S2A mutant. There was no difference in salt-responsive gene expression for control plasmids with 14 *versus* 21 wild-type repeats (not shown). Expression was measured as described above, before and at 30 min after treatment with 0.7M NaCl. There were few expression differences in the strains before stress (see Figure C.8B).

Phospho-proteomic analysis

BY4741 was grown as described above, except that samples were taken before and at 5 min and 15 min after NaCl addition. Cells were lysed by three passages through the French press at 4°C in 3 mL of lysis buffer consisting of 50 mM Tris pH 8, 4M urea, 75 mM NaCl, 1 mM DTT, complete mini EDTA-free protease inhibitor (Roche Diagnostics, Indianapolis, IN), and phosSTOP phosphatase inhibitor (Roche Diagnostics). The lysate was centrifuged at 14000 r.p.m. for 10 min and the protein concentration determined by a bicinchoninic acid assay. Cysteine residues were reduced and alkylated by incubating lysate with 5 mM DTT for 45 min at 37°C followed by incubation in 15 mM IAA for 45 min at room temperature in the dark. After adding an additional aliquot of DTT to cap the alkylation reaction, the urea concentration was diluted to a final concentration of 1M with 50 mM Tris and 1 mM $CaCl_2$. Proteins from each time point were digested overnight (37°C, pH 8) with trypsin (Promega, Madison, WI) at an enzyme:substrate ratio of 1:50. TFA was added to a final concentration of 0.5% to quench each digest and the resulting peptides were desalted via solid phase extraction on a 50-mg tC_{18} SepPak cartridge (Waters, Milford, MA) and the eluant lyophilized.

The desalted peptides from each time point were each labeled with a different tandem mass tag (TMT) isobaric label (Thermo-Pierce, Rockford, IL) according to the manufacturer's instructions. The differentially labeled TMT samples were pooled in equal volumes and dried down. Labeled peptides were fractionated by strong cation exchange (SCX) on a polysulfoethyl A column (9.4 mm x 200 mm; PolyLC) with mobile phases A: 5 mM KH_2PO_4 pH 2.65 and 30% acetonitrile; B: 5 mM KH_2PO_4 pH 2.65, 350 mM KCl, and 30% acetonitrile; C: 5 mM KH_2PO_4 pH 6.5 and 500 mM KCl; D: Water. The gradient was generated by a Surveyor LC quaternary pump (Thermo Scientific, Waltham, MA) at 3 mL/min flow rate. Peptides were eluted over the following gradient and detected via a PDA detector (Thermo Scientific): 0-2 min, 100% A; 2-5 min, 0-10% B; 5-41 min 10-100% B; 41-48 min 100% B; followed by washes with C and D prior to re-equilibration with mobile phase A. Fifteen fractions were collected, lyophilized, and desalted. A small portion, 5%, of each was retained for unmodified protein analysis and the remaining material used for phosphopeptide enrichment.

Each fraction was enriched for phosphopeptides using immobilized metal ion affinity chromatography (IMAC). Magnetic beads (Qiagen, Valencia, CA) were washed 3 times with water, incubated with 40 mM EDTA (pH 7.5) for 30 min, and washed with water again. The beads were then incubated with 100 mM $FeCl_3$ for 30 min and washed 4 times with 80% acetonitrile and 0.1% TFA. Peptides from each fraction were resuspended in 1 mL of 80% acetonitrile and 0.1% TFA and incubated with the beads for 30 min. Unbound peptides were removed from the beads by washing 4 times with 80% acetonitrile and 0.1% TFA. Phosphopeptides were eluted using 1:1 acetonitrile:5% NH_4OH in water, immediately acidified with 4% formic acid, and lyophilized.

Phosphopeptide enriched and protein fractions were resuspended in 0.2% formic acid and analyzed by reverse phase liquid chromatography on a nanoAcquity LC (Waters) coupled to an ETD-enabled LTQ Orbitrap Velos (Thermo Scientific). Samples were first loaded onto a 10-cm, 75

µm i.d. pre-column packed with 5 µm C18 particles (Bruker-Michrom, Fremont, CA) in 98% A (0.2% formic acid in water), 2% B (0.2% formic acid in acetonitrile) and then separated across a 25-cm, 50 µm i.d. analytical column packed with 5 µm C18 particles (Bruker-Michrom) using the following gradient: 0-3 min, 2-5% B; 3-123 min, 5-35% B; 123-133 min, 35-70% B; 133-138 min, 70% B; 138-165 min, 2% B. Phosphopeptide and protein fractions were each analyzed in duplicate. Methods to acquire mass spectra started with one MS1 survey scan (R = 30000, 300-1500 Th) followed by data dependent MS2 fragmentation and analysis (R = 15000) of the ten most intense precursors. The exclusion duration was 60 seconds for -0.55 Th to +2.55 Th of the sampled precursor. Ions with an unassigned charge state or a single charge were excluded. The QuantMode instrument control method was employed to reduce reporter ion interference caused by co-isolation of multiple precursors (Wenger et al., 2011a).

Spectral reduction was performed using DTA Generator. Generated text files were searched for fully tryptic peptides with up to three missed cleavages against a UniProt target-decoy database populated with yeast plus isoforms (downloaded 29 July 2011) using the Open Mass Spectrometry Search Algorithm (OMSSA) (Geer et al., 2004). Carbamidomethylation of cysteine (+57.021464), TMT 6-plex on lysine (+229.162932), and TMT 6-plex on peptide N-terminus (+229.162932) were searched as fixed modifications for all samples. Phosphopeptide-enriched fractions were additionally searched for variable phosphorylation modifications. Search results were filtered to 1% FDR at the unique peptide level and identified peptides quantified within the COMPASS software suite (Wenger et al., 2011b). Peptides were grouped into proteins according to previously reported rules and filtered to 1% FDR (Nesvizhskii and Aebersold, 2005). Protein quantification was performed by summing all reporter ion intensities within each channel for each non-phosphorylated peptide mapping uniquely to that protein group.

Phosphorylation events were localized to specific residues using probabilistic methods (Phanstiel et al., 2011). Localized phosphorylated peptides were grouped together by identical modification sites and their reporter ion intensities were summed. For simplicity, phosphorylation isoforms are referred to as phospho-sites. The average of two technical replicates was taken per time point, and phospho-sites with at least two-fold change in recovery were taken as significant for downstream analysis.

Immunoprecipitation analysis

BY4741-*cka2Δ* and *cka1Δ* cells were transformed with empty vector or plasmids encoding GAL-inducible Cka2-GST or Cka1-GST (Sopko et al., 2006; Zhu et al., 2001)(Thermo Scientific), respectively. Cells were grown in YP-2% galactose medium in log phase to 0.6-0.8 OD_{600}, subjected to osmotic stress (0.7N NaCl) for the indicated length of time, and lysed by bead-beating on ice. Cell lysates were incubated with glutathione-sepharose beads (GE Healthcare) at 4°C overnight in 1X PBS buffer with 1 mM DTT, 0.1% NP40, 10% glycerol and protease inhibitors (Millipore, Billerica, MA). Proteins were eluted with 1X SB buffer and resolved by SDS-PAGE and detected by immunoblotting. Antibodies used were goat polyclonal anti-Hog1 (Santa Cruz Biotech, Dallas, TX), rabbit polyclonal anti-phospho-p38 MAPK (Cell Signaling), mouse monoclonal anti-actin (Pierce Biotech), goat polyclonal anti-Cdc14 (Santa Cruz Biotech) and goat-polyclonal anti-GST (Abcam, Cambridge, MA). All blots shown in the manuscript are representatives of at least biological duplicates.

Microscopy

Harvested cells were fixed with 4% final concentration of formaldehyde for 15 min, and GFP was visualized on a Leica DM LB2 microscope with standard GFP filters. DNA was detected via cell staining with 1 µg/ml DAPI for 5 min. Viability of *cdc14-3* cells was measured with Live-Dead staining (Life Technologies), which showed that NaCl-treated *cdc14-3* maintained viability close

to WT cells for over 30 min after treatment with 0.7M NaCl (not shown). Nuclear Hog1 was scored by visual inspection by comparing GFP signal to DAPI signal, in at least 75-100 cells per sample.

In vitro CTD phosphorylation

Cells expressing C-terminally TAP-tagged proteins (Ghaemmaghami et al., 2003)(Thermo) were exposed to NaCl for the denoted times, snap frozen, and then cryo-lysed with a Retsch Mixer Mill MM 400 as described in (Churchman and Weissman, 2011). Ground yeast was added to TAP Buffer A, and TAP-tagged kinase was purified as described in (Liu et al., 2004; Puig et al., 2001), with minor modifications. Kinases were eluted overnight at 4°C in 25 µL TAP Buffer A with 1mM DTT and 10 U AcTEV (Invitrogen).

Peptide substrate GST-CTD14 (fourteen repeats of YSPTSPS fused to GST), was purified essentially as described in (Patturajan et al., 1998). Before elution, glutathione-sepharose beads were resuspended in 1 mL FastAP buffer (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 100 mM KCl, 0.02% TritonX-100, 100 ug/mL BSA) and incubated with 100 U FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) for 1 hour at 37°C to remove any phosphates placed by the bacteria. Beads were washed, and GST-CTD14 was eluted. Any remaining alkaline phosphatase was heat inactivated at 75°C for 5 min. The concentration of GST-CTD14 was determined via Bradford assays.

In vitro kinase assays were performed in at least biological duplicate using 5 µL of tandem affinity purified (TAP) kinase and 3 µM GST-CTD14 in 30 µL Buffer D as described in (Ansari et al., 2005), with minor modifications. For Hog1 inhibition assays, the kinase was pre-incubated with the inhibitor 4-(4-fluorophenyl)-2-(4-methylsulfinylphenyl)-5-(4-pyridyl)imidazole (Cell Signaling Technology) for 10 min prior to the reaction. Reactions were performed at 30°C for

two hours and resolved via SDS-PAGE and Western analysis using antibodies targeting CTD-Ser2P (Bethyl Laboratories), CTD-Ser5P (clone 3E8, gift from Dirk Eick), or the TAP tag (Thermo Scientific). Quantitation was performed using ImageJ. All images are representative of several biological replicates. The plot in Figure C.5D shows background-subtracted levels of Ser2P and Ser5P normalized to Hog1-TAP abundance in each lane, then referred to levels seen in unstressed cells to calculate fold-change in phosphorylation.

Analysis of novel predicted salt-response regulators

Fourteen predicted regulators not previously known to respond to NaCl were chosen for validation analysis. NaCl-responsive gene expression was measured in ten mutants, focusing on kinases and phosphatases not known to respond to NaCl and two RBPs (Scd6 and Arf3), as described above. Data for Rpd3, Bem1, Gal11, and Tpk2 were taken from previous studies probing the osmotic response (Alejandro-Osorio et al., 2009; Gitter et al., 2013), taking $q < 0.05$ from *limma*-q value analysis as significant. Mutants were considered to have a defect in NaCl-dependent expression if there were at least fifty affected genes. Overlap between measured and predicted target genes was based on the hypergeometric test, scoring the probability of getting the number of observations or more compared to random expectation from the 3,330 genes used for IP input. Genes affected in unstressed cells were also identified (see above) and compared to predicted genes in cases where the NaCl-measured targets did not significantly overlap with predicted targets.

We also assessed the connections predicted between the interrogated regulators and other nodes predicted to lie in their paths. From the nodes predicted to lie in each regulator's path (based on the consensus-paths network), we identified those with known downstream targets (*e.g.* TF and RPBs) or targets measured in this study. We then scored the enrichment of each predicted node's known targets within the measured targets of the interrogated regulator, taking $p<1e-6$ from the hypergeometric test as significant. Because the test lacks statistical power for

large gene groups, we scored enrichment against the total list of measured targets as well as induced and repressed targets with defective or amplified expression changes considered separately. The results indicate a lower bound of supported in-path nodes, since the hypergeometric test has lower statistical power for small gene groups (including known targets of several regulators), and targets of several in-path nodes were marginally enriched ($1e\text{-}5 < p < 0.01$) among measured targets of interrogated regulators but did not meet our stringent threshold. It is also possible that regulators that serve redundant roles are difficult to score with our assay, since single-gene knockouts may not identify all of the downstream targets.

Chromatin immunoprecipitation (ChIP)

ChIP was done similarly to as described in (Tietjen et al., 2010), on cells before and 20 min after treatment with 0.7M NaCl. Rpb3 was immunoprecipitated using anti-Rpb3 antibody W0012 (Neoclone, Madison, WI) in strain Z26 carrying “wild-type” or “S5A” *RPB1* gene expressed on a CEN plasmid (West and Corden, 1995), described above. Chromatin was sonicated on a Misonix 4000 machine (Qsonica, Newtown, CT), input and immunoprecipitated material was amplified using ligation-mediated PCR as previously described (Tietjen et al., 2010) and hybridized to tiled Nimblegen arrays designed against the yeast genome (Lee et al., 2011). Data were normalized as in (Tietjen et al., 2010), except without the baseline adjustment procedure. All two-color arrays from two biological replicates were quantile normalized together before further analysis. This procedure did not change any of the trends reported in the manuscript but helped to adjust the baseline across biological replicates done on different days. ChIP-chip data area available in the NIH GEO database under accession # GSE60613

Ortholog analysis

To assess the relationship of the yeast consensus network ensemble to human diseases, we analyzed the orthologous set of human genes. We used the stringent RSD method of ortholog

assignment (Wall et al., 2003), using a BLAST Evalue cutoff of 1e-5 and requiring fewer than 20% gapped positions in the global alignment. The method identified 2381 yeast-human orthologas; we focused on the 1619 of these genes that are reviewed in humans. We compared these genes to those annotated in the COSMIC v67 (Forbes et al., 2011) and OMIM (Hamosh et al., 2005) databases. We also analyzed orthologous mouse proteins using the phenology.org database (Woods et al., 2013).

C.10 Network inference methods

Background network for IP method

To construct the background network, we identified a variety of binary interactions that are relevant to intracellular signaling and gene-expression regulation. The background network, gathered from numerous public databases, represents interactions between pairs of proteins (Fasolo et al., 2011; Heavner et al., 2012; Ptacek et al., 2005; Pu et al., 2009; Sharifpoor et al., 2011), including kinase-substrate interactions, as well as protein-DNA interactions (Abdulrehman et al., 2011; Everett et al., 2009; Guelzim et al., 2002; Huebert et al., 2012; MacIsaac et al., 2006; Ni et al., 2009; Venters et al., 2011) and protein-RNA interactions (Hogan et al., 2008; Scherrer et al., 2010). After manual inspection of the background network neighborhoods of the interrogated mutants, we added a set of 17 missing interactions between the mutants and nearby regulators based on known interactions in the literature.

While the types of biological interactions in the background network are rich and diverse, we use a simplified representation as input to the computational method (illustrated in Figure C.2 and C.3A). The background network is represented as a graph, in which nodes represent genes and gene products, and edges represent interactions. A gene may be represented as two separate nodes in the background network: one representing the protein, and, for targets, one

representing the DNA or mRNA. Each interaction may have a direction: for example, transcriptional regulatory interactions are directed, but most protein-protein interactions are not.

IP method input data and candidate paths

The primary goal of the IP approach is to provide explanations for the salt-specific transcriptomic changes measured for this article. We also use two additional sources of salt-specific experimental data. From these data, we generate directed, acyclic candidate paths that serve as input to the IP (Figure C.3B):

Source-target pairs and paths, source-source paths. From the transcriptomic data measured in each of the original signaling mutants, we identified the set of downstream genes with dysregulated salt-responsive expression. We then extracted what we refer to as *source-target pairs*, each consisting of a single *source* protein and a *target* gene that was dysregulated in the source mutant under salt stress. Next, for each source, we used the hypergeometric test to identify candidate transcription factors (TFs) and RNA-binding proteins (RBPs) whose known binding targets (promoters or transcripts, respectively) are significantly enriched with the genes represented by the source's targets ($p < 0.05$). We also include TFs that are known to bind any number of targets under osmotic stress (Huebert et al., 2012; Ni et al., 2009). Candidate source-target paths were enumerated to connect signaling mutants to their gene targets via candidate TFs/RBPs with up to three intermediate proteins between the source and TF/RBP (for a total of five interactions). Candidate path enumeration for each source was performed in an iterative deepening procedure, which was stopped at the path length at which at least 50% of candidate TFs/RBPs were reached.

Fitness-contribution hits and hit-source paths. Previously, we identified yeast mutants that conferred a defect in acquired stress resistance after salt pretreatment (Berry et al., 2011). We

refer to the gene products represented by these mutants as *fitness-contribution hits* because of the mutation's negative effect on yeast fitness under salt stress. *Candidate hit-source paths* and source-source paths were generated by finding short paths (including at most one intermediate protein) between these hits and the source proteins, and between pairs of source proteins. These paths are useful for interpreting the fitness-contribution hits in terms of connections to known regulators.

Phospho-proteomic hits. We use this name to refer to the proteins that showed differential phosphorylation under salt stress.

Receptor-source paths. Our method can take advantage of domain knowledge about the salt stress response in order to provide a scaffold for the inferred subnetwork. Here, we wanted to capture the most-upstream stress sensors that may otherwise be missed in connecting sources to their downstream targets. We identified well-known indirect relationships between two transmembrane receptors, Sln1 and Sho1, and one of the sources, Hog1 (Saito and Tatebayashi, 2004). We enumerated candidate receptor-source paths (up to four intermediates) from Sln1 to Hog1 and Sho1 to Hog1, and provided them as input to the IP method. To measure the contribution of each input data set, we ran computational experiments in which each component was held aside. We also tested the effect of varying the length of the candidate paths.

IP notation and variables

The salt-specific signaling subnetwork is inferred by solving an integer linear program (IP, for short). We encode the relevance of each node, edge (physical interaction), and candidate path, and the direction of each edge, as binary variables. We characterize possible subnetworks using a set of linear constraints over those binary variables. Subnetwork inference is performed

by choosing a union of relevant, directed paths that together satisfy our constraints and optimize a series of successively applied objective functions.

The values of some variables were determined by data provided as input to the inference process (for example, directions of directed edges), while others are inferred by solving the IP.

Notation. The input to the method is represented as a graph of nodes $\mathcal{N}$, edges $\mathcal{E}$, and candidate paths $\mathcal{P}$. A node represents either a protein or a target gene/mRNA. Protein nodes may belong to one or more of the following subsets: sources $\mathcal{N}^S$, fitness-contribution hits $\mathcal{N}^F$, phospho-proteomic hits $\mathcal{N}^P$, and known membrane receptors $\mathcal{N}^R$. The set $\mathcal{N}^T$ describes targets, and for a given source node n, $\mathcal{N}^T(n)$ is the set of its targets.

The set of edges is $\mathcal{E} = (\mathcal{E}^D \cup \mathcal{E}^U)$, where $\mathcal{E}^D$ is the set of directed edges and $\mathcal{E}^U$ is the set of undirected edges. We denote an edge e between nodes n_i and n_j as $e = (n_i, n_j)$. $\mathcal{N}(e)$ refers to the nodes connected by a particular edge e, and $\mathcal{E}$(n) refers to the edges that touch a particular node n.

We consider four subsets of candidate paths: source-target paths between sources and their targets $\mathcal{P}^{ST}$, hit-source paths between fitness-contribution hits and sources $\mathcal{P}^{FS}$, source-source paths $\mathcal{P}^{SS}$, and receptor-source paths P^{RS} that connect known receptor proteins to sources. (Phosphoproteomic hits and additional fitness-contribution hits may appear in any of these paths.) To refer to the paths between a specific source s and target t, we use the notation $\mathcal{P}^{ST}(s,t)$. We use the same notation to refer to other kinds of paths with specific endpoints: $\mathcal{P}^{FS}(f,s), \mathcal{P}^{SS}(s_i,s_j), \mathcal{P}^{RS}(r,s)$.

Each path specifies a direction for each of its undirected edges e, which is denoted as $dir(p, e)$. $\mathcal{E}(p)$ and $\mathcal{N}(p)$ refer to the edges and nodes in a particular path p.

Variables. The predicted relevance of a path is represented with the variable σ_p which takes the value 1 if the path is included in the inferred subnetwork, and 0 if it is not. As many as two variables describe each edge. The predicted relevance of an edge is represented with the variable x_e, which takes the value 1 if the edge is in at least one relevant path. For undirected edges in the background network, the variable d_e represents the inferred direction of the edge. Each node n has one variable: y_n, representing whether or not the node is present in any relevant paths. Finally, for all pairs of sources (n_i, n_j), and also for all pairs consisting of one source and one fitness-contribution hit, the variable c_{ij} represents whether or not the relevant subnetwork provides a directed path between the two nodes in the pair.

IP constraints

The following linear constraints define a subnetwork that, at minimum, provides consistently directed paths between source-target pairs and receptor-source pairs. Additional constraints are used to count up the number of connected fitness-contribution hit-source pairs and source-source pairs. These counts are optimized during the optimization procedure.

Provide at least one path between each source-target pair. Each source must be connected to each of its targets by at least one relevant path. The following constraint requires that, for each source s, for each of its targets t, at least one source-target path p in $\mathcal{P}^{sT}(s, t)$ from s to t must have $\sigma_p = 1$.

$$\sum_{\text{source-target paths } p \text{ in } \mathcal{P}^{ST}(s,t)} \sigma_p \geq 1 \qquad \text{for all sources } s \text{ in } \mathcal{N}^S, \text{ targets } t \text{ in } \mathcal{N}^T(s) \tag{1}$$

Provide at least one path between each receptor-source pair. We must provide at least one path showing the indirect relationship between an upstream receptor and a source. Similar to the previous constraint, this one requires that for each receptor r and each of its downstream sources s there must be at least one receptor-source path p in $\mathcal{P}^{RS}(r,s)$ for which $\sigma_p = 1$.

$$\sum_{\text{receptor-source paths } p \text{ in } \mathcal{P}^{RS}(r,s)} \sigma_p \geq 1 \qquad \text{for all receptors } r \text{ in } \mathcal{N}^R, \text{ sources } s \text{ in } \mathcal{N}^S(r) \tag{2}$$

Record whether or not there is a path between each fitness-contribution hit-source pair and source-source pair. Rather than require that each of these pairs is connected, we use the optimization procedure to maximize the total count of connected pairs. We use the following constraints to count up the number of connected pairs.

If there is a path between a fitness-contribution hit f and a source s, set the variable $c_{fs} = 1$. Otherwise, set $c_{fs} = 0$:

$$\sum_{\substack{\text{hit-source paths} \\ p \text{ in } \left(P^{FS}(f,s) \cup P^{FS}(s,f)\right)}} \sigma_p - c_{fs} \geq 0 \qquad \begin{array}{l}\text{for all fitness-contribution hits } f \text{ in } \mathcal{N}^F, \\ \text{sources } s \text{ in } N^S\end{array} \tag{3}$$

$$c_{fs} - \sigma_p \geq 0 \qquad \begin{array}{l}\text{for all fitness-contribution hits } f \text{ in } \mathcal{N}^F, \\ \text{sources } s \text{ in } \mathcal{N}^S, \\ \text{hit-source paths } p \text{ in } \mathcal{P}^{FS}(f,s) \cup \mathcal{P}^{FS}(s,f)\end{array} \tag{4}$$

Similarly, if source s_i is connected to source s_j, we set $c_{ij} = 1$. Otherwise, set $c_{ij} = 0$.

$$\sum_{\substack{\text{source-source paths} \\ p \text{ in } \left(\mathcal{P}^{SS}(s_i,s_j) \cup \mathcal{P}^{SS}(s_j,s_i)\right)}} \sigma_p - c_{ij} \geq 0 \qquad \text{for all pairs of sources } (s_i, s_j) \text{ in } \mathcal{N}^S \times \mathcal{N}^S \qquad (5)$$

$$c_{ij} - \sigma_p \geq 0 \qquad \text{for all pairs of sources } (s_i, s_j) \text{ in } \mathcal{N}^S \times \mathcal{N}^S, \text{ source-source paths } p \text{ in } \mathcal{P}^{SS}(s_i, s_j) \cup \mathcal{P}^{SS}(s_j, s_i) \qquad (6)$$

All edges in a relevant path are relevant. For an edge e to be relevant (that is, have $x_e = 1$), there must be at least one relevant path that contains it (that is, a path p for which $\sigma_p = 1$). Similarly, a relevant path p must contain all relevant edges e. The set $\mathcal{P}(e)$ refers to the paths that contain edge e.

$$\sum_{paths\ p\ in\ \mathcal{P}(e)} \sigma_p - x_e \geq 0 \qquad \text{for all edges } e \text{ in } \mathcal{E} \qquad (7)$$

$$x_e - \sigma_p \geq 0 \qquad \text{for all paths } p \text{ in } \mathcal{P}, \text{ edges } e \text{ in } \mathcal{E}(p) \qquad (8)$$

All nodes in a relevant edge are relevant. A node n is relevant if it is connected to a relevant edge e (where $x_e = 1$). Each node n for a relevant edge e must be relevant ($y_n = 1$).

$$\sum_{\text{edges } e \text{ in } \mathcal{E}(n)} x_e - y_n \geq 0 \qquad \text{for all nodes } n \text{ in } \mathcal{N} \qquad (9)$$

$$y_n - x_e \geq 0 \qquad \text{for all edges } e \text{ in } \mathcal{E}, \text{ nodes } n \text{ in } \mathcal{N}(e) \qquad (10)$$

All paths must be uniquely directed. For a relevant path p, all undirected edges e in that path (e in $\mathcal{E}(p) \cap \mathcal{E}^U$) must be uniquely oriented so that the path proceeds only in one direction. This required direction for each edge is determined when the candidate path is generated, and is

given by $dir(p,e)$. (For source-target paths, the required direction allows the path to proceed from the source to the target.) The term including $I(\cdot)$, the indicator function, returns 1 if an edge's inferred direction corresponds to the direction that the path requires for it.

$$I\big(d_e = dir(p,e)\big) - \sigma_p \geq 0 \qquad \text{for all paths } p \text{ in } \mathcal{P}, \text{ undirected edges } e \text{ in } \mathcal{E}(p) \cap \mathcal{E}^U \quad (11)$$

Solving the IP to find an ensemble of subnetworks

An optimal inferred subnetwork satisfies two goals: maximizing the inclusion of salt-response-relevant proteins that are supported by experimental evidence, and minimizing the number of additional nodes that are necessary for connecting each source to each target. To achieve this, we apply four successive objective functions.

To model and solve the IP, we used the GAMS modeling system v. 23.9.3 and the ILOG CPLEX solver v. 12.4.0.1. Both are commercial packages for which an academic license available at a reduced cost.

Step 1: Maximize connections between hits and sources. This involves solving the IP to identify max_connections, the maximum number of connections possible between pairs of sources, and between pairs of fitness-contribution hits and sources. The purpose of this step is to reveal proximal connections between salt-responsive proteins, whether or not they occur between sources and targets. In this constraint, the set $(\mathcal{N}^S \times \mathcal{N}^S) \cup (\mathcal{N}^F \times \mathcal{N}^S)$ gives all source-source pairs and fitness-contribution-hit-source pairs, and the sum counts up the number of pairs that are connected by relevant paths.

$$\text{max\_connections} = \max \sum_{\text{hit-source pairs } (n_i,n_j) \text{ in } (\mathcal{N}^S \times \mathcal{N}^S) \cup (\mathcal{N}^F \times \mathcal{N}^S)} c_{ij} \quad (12)$$

After optimizing this criterion, we add a new constraint to the IP:

$$\sum_{\text{hit-source pairs } (n_i,n_j) \text{ in } (\mathcal{N}^S \times \mathcal{N}^S) \cup (\mathcal{N}^F \times \mathcal{N}^S)} c_{ij} = \text{max\_connections} \quad (13)$$

Step 2: Maximize inclusion of fitness and phospho hits. Next, we solve the IP to identify max_hits, the maximum number of fitness-contribution hits and phosphor-proteomic hits that can be included in the relevant subnetwork. This step prioritizes the use of nodes with experimental evidence of being relevant to the salt stress response.

$$\text{max\_hits} = \max \sum_{\text{nodes } n \text{ in } (\mathcal{N}^F \cup \mathcal{N}^P)} y_n \quad (14)$$

After identifying the maximal number of hits that can be included in the subnetwork, we add a new constraint to the IP:

$$\sum_{\text{nodes } n \text{ in } (\mathcal{N}^F \cup \mathcal{N}^P)} y_n = \text{max\_hits} \quad (15)$$

Step 3: Minimize total nodes and find multiple solutions. Now we solve the IP with a new objective function, which minimizes the number of nodes required to satisfy all of the constraints. The resulting subnetwork will include only those nodes that are required to explain the experimental data.

$$\min \sum_{\text{nodes } n \text{ in } \mathcal{N}} y_n \tag{16}$$

At this point, we find an ensemble of solutions to the IP, where each solution identifies a minimum set of nodes (while still satisfying all other constraints). The CPLEX solver allows for the identification of multiple solutions. First, the CPLEX solver uses a branch-and-cut algorithm to find one optimal solution; this algorithm entails maintaining a tree of linear relaxations of the IP. Next, the solver proceeds down previously rejected branches of the tree to identify additional optimal solutions with different variable settings. For our experiments, we identified 10,000 solutions.

Step 4: Maximize the number of paths in each solution. After predicting the relevant nodes in the previous step, we would like to see all possible relevant connections between them, to aid in their interpretation. For each of the solutions identified in the previous step, we solve the IP again to maximize the number of relevant directed paths between the nodes included in the solution. This step does not change the node content of each solution, but instead reveals all possible directed paths that connect the node set chosen in the previous step.

For each solution:

First, we introduce constraints to fix each value of y_n to its value from the previous solution, $\widehat{y_n}$:

$$y_n = \widehat{y_n} \text{ for all nodes } n \tag{17}$$

Next, we maximize the number of relevant paths:

$$\max \sum_{\text{paths } p \text{ in } \mathcal{P}} \sigma_p \tag{18}$$

At this point, we assemble the solutions into an ensemble of inferred subnetworks. Using the ensemble, we assign a confidence value for a prediction based on the number of solutions in the ensemble that support the prediction. We performed several experiments to assess the effect of each component of our four-part objective function, as well as their ordering.

Precision-recall analysis

To assess the predictive accuracy of the ensemble (as shown in Figure C.4B), we curated a list of true positives and a list of likely negative proteins. True positives were defined as genes previously identified in the Hog network based on literature curation (de Nadal and Posas, 2010; Tiger et al., 2012), genes with ‘osmotic’ or ‘osmolarity’ in their *Saccharomyces* Genome Database (SGD) (Cherry et al., 2012) annotations, and genes with 'stress regulator' in their SGD annotations, if they were also linked to the osmotic response in at least one publication. In all, this identified 112 true positives. Likely negatives were taken as genes with no evidence for nuclear localization and whose GO compartment annotation was ‘mitochondrion’, ‘mitochondrial envelope’, ‘peroxisome’, ‘vacuole’, ‘Golgi’, and/or ‘endoplasmic reticulum’. Proteins annotated in SGD as ‘metabolic enzymes’ were also added to this list of likely negatives. From this list we removed 32 well-known signaling proteins, many of which were already on the true positive list; in all, this left 1,865 likely negative proteins for the network assessment. Among these test cases, the background network contained 108 positives and 1512 likely negatives. In order to separate out the effect of the experimental hits on predictive accuracy, we omitted all hits from the test cases, leaving 70 true positives and 1416 likely negatives. For each test case (true positive or likely negative), we measured the inferred subnetwork ensemble's confidence that it is relevant to the salt response. This is calculated as the fraction of the 10,000 solutions in which the test case appears as a protein node in the subnetwork.

We compared our ensemble's precision-recall curve to two baselines, which we refer to as the *candidate* baseline (Figure C.4B, green) and *permuted* baseline (Figure C.4B, yellow). For the candidate baseline, we computed the precision and recall of the test cases using the complete set of protein nodes present in candidate paths. For the permuted baseline, we compared the inferred ensemble's accuracy to that of a set of 1,000 ensembles inferred using permuted experimental data. For each of 1,000 permutations, we randomly drew a set of sources, proteins with fitness defects, and proteins with phospho-changes from the background network, equal in number and degree distribution to the true experimental data. To generate receptor-source pairs, we randomly drew two proteins from the background network and paired each with a randomly chosen source. To generate permuted source-target pairs, for each source, we randomly drew an equal number of targets from the entire background network. We inferred an ensemble of 1,000 solutions for each permutation, and measured the confidence of each test case as the average confidence over all 1,000 ensembles.

Ranking of putative ESR bifurcation points

We constructed the salt-relevant ESR consensus subnetwork shown in Figure C.7A and C.7C as follows. First, we gathered three clusters of genes defined by (Gasch et al., 2000) based on expression profiles under multiple stress conditions: iESR (induced ESR) and two rESR (repressed ESR) subclusters, RiBi and RP. Using the protein-nucleic acid interactions from the background network, we identified potential transcriptional regulators of the three ESR gene clusters. These were TFs and RBPs whose targets were enriched for a cluster (determined by hypergeometric test, using a threshold of FDR=0.1, calculated by the Benjamini-Hochberg procedure). For iESR targets, we identified 25 total potential TFs/RBPs, of which 22 are TFs and three are RBPs. We found 16 TFs and 10 RBPs for the combined rESR clusters.

Next, we extracted the consensus source-target paths (having confidence ≥ 75%) that end in an interaction between an ESR-relevant TF/RBP and ESR-relevant target gene (of the same cluster). For each protein node in each ESR-relevant consensus path, we assigned a label based on the ESR cluster(s) represented by the downstream ESR-relevant TF/RBPs. These labels were used to perform the coloring in Figure C.7. Finally, we removed the targets that were not a member of any ESR cluster.

Using the ESR consensus paths, we identified candidate bifurcation points, defined as nodes that are upstream of both rESR and iESR targets (yellow and orange nodes in Figure C.7), according to how well their outgoing paths show a distinct division between the induced and repressed clusters. To rank the candidates, we defined a bifurcation score, $B(n)$, that is related to the concept of information gain ratio (Quinlan, 1986). $B(n)$ is calculated as follows.

First, we define the *count* $C(T)$, which counts the number of bits required to represent the cluster membership of all of the targets in a set T. Considering the clusters $c \in$ {iESR, rESR}, let $T^c(n)$ be the set of targets downstream of n that belong to the ESR cluster c.

$$C\big(T(n)\big) = -\sum_{\text{clusters } c \text{ in } \{\text{iESR. rESR}} |T^c(n)| \log_2 \frac{|T^c(n)|}{|T(n)|} \tag{19}$$

An ideal bifurcation point would have a high $C(T(n))$ compared to the paths that emanate from it. To perform this comparison, we next calculate $C(\cdot)$ for each of the paths downstream from n. If the subnetwork were a tree, n's targets would simply be partitioned by n's children. However, since the paths leading out from n's children may converge on the same targets, we instead partition $T(n)$ into disjoint subsets of targets, each of which is reachable via a unique combination of n's children. We refer to n's outgoing partitions as $P_1(n)$ ... $P_m(n)$.

After having calculated $C\big(P_i(n)\big)$ for each partition, we then calculate the *information gain*, $I(n)$,

$$\mathrm{I(n)} = \mathrm{C\big(T(n)\big)} - \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{C\big(P_i(n)\big)} \tag{20}$$

which measures the number of bits that are saved by partitioning the targets downstream of n: Finally, to calculate the bifurcation score $B(n)$, we normalize $I(n)$ by the *split information* $S(n)$, which measures the number of bits required to describe the *partition* assignment of one of n's targets. $I(n)$ is strongly biased toward nodes whose outgoing partitions split each target each into its own partition. The normalized score $B(n)$ prioritizes nodes that have a small number of (relatively) cleanly-split outgoing paths and many downstream targets.

$$S(n) = -\sum_{i=0}^{m} \frac{\left|P_{i(n)}\right|}{|T(n)|} \log_2 \frac{\left|P_{i(n)}\right|}{|T(n)|} \tag{21}$$

$$B(n) = \frac{I(n)}{S(n)} \tag{22}$$

C.11 References

Abdulrehman, D., Monteiro, P.T., Teixeira, M.C., Mira, N.P., Lourenço, A.B., dos Santos, S.C., Cabrito, T.R., Francisco, A.P., Madeira, S.C., Aires, R.S*., et al.* (2011). YEASTRACT: providing a programmatic access to curated transcriptional regulatory associations in Saccharomyces cerevisiae through a web services interface. Nucleic Acids Res *39*, D136-D140.

Adrover, M.A., Zi, Z., Duch, A., Schaber, J., Gonzalez-Novo, A., Jimenez, J., Nadal-Ribelles, M., Clotet, J., Klipp, E., and Posas, F. (2011). Time-dependent quantitative multicomponent control of the G(1)-S network by the stress-activated protein kinase Hog1 upon osmostress. Sci Signal *4*, ra63.

Alejandro-Osorio, A.L., Huebert, D.J., Porcaro, D.T., Sonntag, M.E., Nillasithanukroh, S., Will, J.L., and Gasch, A.P. (2009). The histone deacetylase Rpd3p is required for transient changes in genomic expression in response to stress. Genome Biol *10*, R57.

Alepuz, P.M., de Nadal, E., Zapater, M., Ammerer, G., and Posas, F. (2003). Osmostress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II. EMBO J *22*, 2433-2442.

Alexander, M.R., Tyers, M., Perret, M., Craig, B.M., Fang, K.S., and Gustin, M.C. (2001). Regulation of cell cycle progression by Swe1p and Hog1p following hypertonic stress. Mol Biol Cell *12*, 53-62.

Ansari, A.Z., Ogirala, A., and Ptashne, M. (2005). Transcriptional activating regions target attached substrates to a cyclin-dependent kinase. Proc Natl Acad Sci U S A *102*, 2346-2349.

Belli, G., Gari, E., Aldea, M., and Herrero, E. (2001). Osmotic stress causes a G1 cell cycle delay and downregulation of Cln3/Cdc28 activity in Saccharomyces cerevisiae. Mol Microbiol *39*, 1022-1035.

Bergkessel, M., Whitworth, G.B., and Guthrie, C. (2011). Diverse environmental stresses elicit distinct responses at the level of pre-mRNA processing in yeast. RNA *17*, 1461-1478.

Berry, D.B., and Gasch, A.P. (2008). Stress-activated genomic expression changes serve a preparative role for impending stress in yeast. Molecular biology of the cell *19*, 4580-4587.

Berry, D.B., Guan, Q., Hose, J., Haroon, S., Gebbia, M., Heisler, L.E., Nislow, C., Giaever, G., and Gasch, A.P. (2011). Multiple means to the same end: the genetic basis of acquired stress resistance in yeast. PLoS genetics *7*, e1002353.

Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.Y., Breitkreutz, B.J., Stark, C., Liu, G*., et al.* (2010). A global protein kinase and phosphatase interaction network in yeast. Science *328*, 1043-1046.

Broach, J.R. (2012). Nutritional control of growth and development in yeast. Genetics *192*, 73-105.

Bulavin, D.V., Saito, S., Hollander, M.C., Sakaguchi, K., Anderson, C.W., Appella, E., and Fornace, A.J., Jr. (1999). Phosphorylation of human p53 by p38 kinase coordinates N-terminal phosphorylation and apoptosis in response to UV radiation. EMBO J *18*, 6845-6854.

Buratowski, S. (2003). The CTD code. Nature structural biology *10*, 679-680.

Busnelli, S., Tripodi, F., Nicastro, R., Cirulli, C., Tedeschi, G., Pagliarin, R., Alberghina, L., and Coccetti, P. (2013). Snf1/AMPK promotes SBF and MBF-dependent transcription in budding yeast. Biochim Biophys Acta *1833*, 3254-3264.

Capaldi, A.P., Kaplan, T., Liu, Y., Habib, N., Regev, A., Friedman, N., and O'Shea, E.K. (2008). Structure and function of a transcriptional network activated by the MAPK Hog1. Nat Genet *40*, 1300-1306.

Carvajal, L.A., and Manfredi, J.J. (2013). Another fork in the road--life or death decisions by the tumour suppressor p53. EMBO reports *14*, 414-421.

Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S., and Young, R.A. (2001). Remodeling of yeast genome expression in response to environmental changes. Molecular biology of the cell *12*, 323-337.

Chasman, D., Ho, Y.H., Berry, D.B., Nemec, C.M., MacGilvray, M.E., Hose, J., Merrill, A.E., Lee, M.V., Will, J.L., Coon, J.J.*, et al.* (2014). Pathway connectivity and signaling coordination in the yeast stress-activated signaling network. Mol Syst Biol *10*, 759.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R.*, et al.* (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res *40*, D700-705.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368-373.

Chymkowitch, P., Eldholm, V., Lorenz, S., Zimmermann, C., Lindvall, J.M., Bjoras, M., Meza-Zepeda, L.A., and Enserink, J.M. (2012). Cdc28 kinase activity regulates the basal transcription machinery at a subset of genes. Proc Natl Acad Sci U S A *109*, 10450-10455.

Cook, K.E., and O'Shea, E.K. (2012). Hog1 controls global reallocation of RNA Pol II upon osmotic shock in Saccharomyces cerevisiae. G3 (Bethesda) *2*, 1129-1136.

De Amicis, F., Giordano, F., Vivacqua, A., Pellegrino, M., Panno, M.L., Tramontano, D., Fuqua, S.A., and Ando, S. (2011). Resveratrol, through NF-Y/p53/Sin3/HDAC1 complex phosphorylation, inhibits estrogen receptor alpha gene expression via p38MAPK/CK2 signaling in human breast cancer cells. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *25*, 3695-3707.

de Nadal, E., and Posas, F. (2010). Multilayered control of gene expression by stress-activated protein kinases. The EMBO journal *29*, 4-13.

Elliott, B., and Futcher, B. (1993). Stress resistance of yeast cells is largely independent of cell cycle phase. Yeast (Chichester, England) *9*, 33-42.

Everett, L., Vo, A., and Hannenhalli, S. (2009). PTM-Switchboard--a database of posttranslational modifications of transcription factors, the mediating enzymes and target genes. Nucleic Acids Res *37*, D66-71.

Fasolo, J., Sboner, A., Sun, M.G.F., Yu, H., Chen, R., Sharon, D., Kim, P.M., Gerstein, M., and Snyder, M. (2011). Diverse protein kinase interactions identified by protein microarrays reveal novel connections between cellular processes. Genes \& Development *25*, 767-778.

Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A*., et al.* (2011). COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res *39*, D945-950.

Friedman, N. (2004). Inferring cellular networks using probabilistic graphical models. Science (New York, NY *303*, 799-805.

Gasch, A.P. (2002). The Environmental Stress Response: a common yeast response to environmental stresses. In Yeast Stress Responses. In Yeast Stress Responses, S. Hohmann, and P. Mager, eds. (Heidelberg: Springer-Verlag), pp. 11-70.

Gasch, A.P., Huang, M., Metzner, S., Botstein, D., Elledge, S.J., and Brown, P.O. (2001). Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. Mol Biol Cell *12*, 2987-3003.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Molecular biology of the cell *11*, 4241-4257.

Gat-Viks, I., and Shamir, R. (2007). Refinement and expansion of signaling pathways: the osmotic response network in yeast. Genome Res *17*, 358-367.

Gat-Viks, I., Tanay, A., Raijman, D., and Shamir, R. (2006). A probabilistic methodology for integrating knowledge and experiments on biological networks. Journal of Computational Biology *13*, 165-181.

Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X., Shi, W., and Bryant, S.H. (2004). Open mass spectrometry search algorithm. Journal of proteome research *3*, 958-964.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Gitter, A., Carmi, M., Barkai, N., and Bar-Joseph, Z. (2013). Linking the signaling cascades and dynamic regulatory networks controlling stress responses. Genome Res *23*, 365-376.

Gitter, A., Klein-Seetharaman, J., Gupta, A., and Bar-Joseph, Z. (2011). Discovering pathways by orienting edges in protein interaction networks. Nucleic Acids Res *39*, e22.

Guelzim, N., Bottani, S., Bourgine, P., and Képès, F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. Nat Genet *31*, 60-63.

Halbeisen, R.E., and Gerber, A.P. (2009). Stress-dependent coordination of transcriptome and translatome in yeast. PLoS Biol *7*, e1000105.

Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., and McKusick, V.A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res *33*, D514-517.

Heavner, B.D., Smallbone, K., Barker, B., Mendes, P., and Walker, L.P. (2012). Yeast 5 - an expanded reconstruction of the Saccharomyces cerevisiae metabolic network. BMC Syst Biol *6*, 55.

Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S., and Young, R.A. (1998). Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. Mol Cell *2*, 43-53.

Hildesheim, J., Salvador, J.M., Hollander, M.C., and Fornace, A.J., Jr. (2005). Casein kinase 2- and protein kinase A-regulated adenomatous polyposis coli and beta-catenin cellular localization is dependent on p38 MAPK. J Biol Chem *280*, 17221-17226.

Hirasawa, T., Ashitani, K., Yoshikawa, K., Nagahisa, K., Furusawa, C., Katakura, Y., Shimizu, H., and Shioya, S. (2006). Comparison of transcriptional responses to osmotic stresses induced by NaCl and sorbitol additions in Saccharomyces cerevisiae using DNA microarray. Journal of bioscience and bioengineering *102*, 568-571.

Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D., and Brown, P.O. (2008). Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol *6*, e255.

Hohmann, S., and Mager, P., eds. (2003). Yeast Stress Responses (Heidelberg: Springer-Verlag).

Hong, S.P., and Carlson, M. (2007). Regulation of snf1 protein kinase in response to environmental stress. J Biol Chem *282*, 16838-16845.

Huang, S.S., Clarke, D.C., Gosline, S.J., Labadorf, A., Chouinard, C.R., Gordon, W., Lauffenburger, D.A., and Fraenkel, E. (2013). Linking proteomic and transcriptional data through the interactome and epigenome reveals a map of oncogene-induced signaling. PLoS Comput Biol *9*, e1002887.

Huang, S.S., and Fraenkel, E. (2009). Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. Sci Signal *2*, ra40.

Huang, S.S., and Fraenkel, E. (2012). Swimming upstream: identifying proteomic signals that drive transcriptional changes using the interactome and multiple "-omics" datasets. Methods in cell biology *110*, 57-80.

Huebert, D.J., Kuan, P.F., Keles, S., and Gasch, A.P. (2012). Dynamic changes in nucleosome occupancy are not predictive of gene expression dynamics but are linked to transcription and chromatin regulators. Mol Cell Biol *32*, 1645-1653.

Ideker, T.E., Thorsson, V., and Karp, R.M. (2000). Discovery of regulatory interactions through perturbation: inference and experimental design. Pac Symp Biocomput *5*, 305-316.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Isaeva, A.R., and Mitev, V.I. (2011). CK2 is acting upstream of MEK3/6 as a part of the signal control of ERK1/2 and p38 MAPK during keratinocytes autocrine differentiation. Zeitschrift fur Naturforschung C, Journal of biosciences *66*, 83-86.

Jones, R.G., and Thompson, C.B. (2009). Tumor suppressors and cell metabolism: a recipe for cancer growth. Genes & development *23*, 537-548.

Lan, A., Smoly, I.Y., Rapaport, G., Lindquist, S., Fraenkel, E., and Yeger-Lotem, E. (2011). ResponseNet: revealing signaling and regulatory networks linking genetic and transcriptomic screening data. Nucleic Acids Res *39*, W424-429.

Lee, M.V., Topper, S.E., Huberl, S.L., Hose, J., Wenger, C.D., Coon, J., and Gasch, A.P. (2011). A Dynamic Model of Proteome Changes Reveals New Roles for Transcript Alteration in Yeast Mol Syst Biol.

Levy, S.F., Ziv, N., and Siegal, M.L. (2012). Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. PLoS Biol *10*, e1001325.

Li, L., Ljungman, M., and Dixon, J.E. (2000). The human Cdc14 phosphatases interact with and dephosphorylate the tumor suppressor protein p53. J Biol Chem *275*, 2410-2414.

Liang, S., Fuhrman, S., and Somogyi, R. (1998). Reveal, a general reverse engineering algorithm for inference of genetic network architectures. Pac Symp Biocomput *3*, 18-29.

Lipson, D., Raz, T., Kieu, A., Jones, D.R., Giladi, E., Thayer, E., Thompson, J.F., Letovsky, S., Milos, P., and Causey, M. (2009). Quantification of the yeast transcriptome by single-molecule sequencing. Nature biotechnology *27*, 652-658.

Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S. (2004). Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. Mol Cell Biol *24*, 1721-1735.

Lu, C., Brauer, M.J., and Botstein, D. (2009). Slow growth induces heat-shock resistance in normal and respiratory-deficient yeast. Molecular biology of the cell *20*, 891-903.

MacIsaac, K., Wang, T., Gordon, D.B., Gifford, D., Stormo, G., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC Bioinformatics *7*, 113+.

Markowetz, F., Bloch, J., and Spang, R. (2005). Non-transcriptional pathway features reconstructed from secondary effects of RNA interference. Bioinformatics *21*, 4026-4032.

Marles, J.A., Dahesh, S., Haynes, J., Andrews, B.J., and Davidson, A.R. (2004). Protein-protein interaction affinity plays a crucial role in controlling the Sho1p-mediated signal transduction pathway in yeast. Mol Cell *14*, 813-823.

Martinez-Montanes, F., Pascual-Ahuir, A., and Proft, M. (2010). Toward a genomic view of the gene expression program regulated by osmostress in yeast. Omics : a journal of integrative biology *14*, 619-627.

McClean, M.N., Mody, A., Broach, J.R., and Ramanathan, S. (2007). Cross-talk and decision making in MAP kinase pathways. Nat Genet *39*, 409-414.

Meek, D.W., Simon, S., Kikkawa, U., and Eckhart, W. (1990). The p53 tumour suppressor protein is phosphorylated at serine 389 by casein kinase II. EMBO J *9*, 3253-3260.

Melamed, D., Pnueli, L., and Arava, Y. (2008). Yeast translational response to high salinity: global analysis reveals regulation at multiple levels. RNA (New York, NY *14*, 1337-1351.

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dumcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dolken, L.*, et al.* (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol *7*, 458.

Miller, C.T., Gabrielse, C., Chen, Y.C., and Weinreich, M. (2009). Cdc7p-Dbf4p regulates mitotic exit by inhibiting Polo kinase. PLoS Genet *5*, e1000498.

Mitchell, A., Romano, G.H., Groisman, B., Yona, A., Dekel, E., Kupiec, M., Dahan, O., and Pilpel, Y. (2009). Adaptive prediction of environmental changes by microorganisms. Nature *460*, 220-224.

Nadal-Ribelles, M., Conde, N., Flores, O., Gonzalez-Vallinas, J., Eyras, E., Orozco, M., de Nadal, E., and Posas, F. (2012). Hog1 bypasses stress-mediated down-regulation of transcription by RNA polymerase II redistribution and chromatin remodeling. Genome Biol *13*, R106.

Nagiec, M.J., and Dohlman, H.G. (2012). Checkpoints in a yeast differentiation pathway coordinate signaling during hyperosmotic stress. PLoS Genet *8*, e1002437.

Nesvizhskii, A.I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. Mol Cell Proteomics *4*, 1419-1440.

Ni, L., Bruce, C., Hart, C., Leigh-Bell, J., Gelperin, D., Umansky, L., Gerstein, M.B., and Snyder, M. (2009). Dynamic and complex transcription factor binding during an inducible response in yeast. Genes Dev *23*, 1351-1363.

Novershtern, N., Regev, A., and Friedman, N. (2011). Physical Module Networks: an integrative approach for reconstructing transcription regulation. Bioinformatics *27*, i177-i185.

O'Rourke, S.M., and Herskowitz, I. (1998). The Hog1 MAPK prevents cross talk between the HOG and pheromone response MAPK pathways in Saccharomyces cerevisiae. Genes Dev *12*, 2874-2886.

O'Rourke, S.M., and Herskowitz, I. (2004). Unique and redundant roles for HOG MAPK pathway components as revealed by whole-genome expression analysis. Mol Biol Cell *15*, 532-542.

Ourfali, O., Shlomi, T., Ideker, T., Ruppin, E., and Sharan, R. (2007). SPINE: a framework for signaling-regulatory pathway inference from cause-effect experiments. Bioinformatics *23*, i359-366.

Patterson, J.C., Klimenko, E.S., and Thorner, J. (2010). Single-cell analysis reveals that insulation maintains signaling specificity between two yeast MAPK pathways with common components. Sci Signal *3*, ra75.

Patturajan, M., Schulte, R.J., Sefton, B.M., Berezney, R., Vincent, M., Bensaude, O., Warren, S.L., and Corden, J.L. (1998). Growth-related changes in phosphorylation of yeast RNA polymerase II. J Biol Chem *273*, 4689-4694.

Pessina, S., Tsiarentsyeva, V., Busnelli, S., Vanoni, M., Alberghina, L., and Coccetti, P. (2010). Snf1/AMPK promotes S-phase entrance by controlling CLB5 transcription in budding yeast. Cell Cycle *9*, 2189-2200.

Phanstiel, D.H., Brumbaugh, J., Wenger, C.D., Tian, S., Probasco, M.D., Bailey, D.J., Swaney, D.L., Tervo, M.A., Bolin, J.M., Ruotti, V.*, et al.* (2011). Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nat Methods *8*, 821-827.

Phatnani, H.P., Jones, J.C., and Greenleaf, A.L. (2004). Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. Biochemistry *43*, 15702-15719.

Proft, M., Mas, G., de Nadal, E., Vendrell, A., Noriega, N., Struhl, K., and Posas, F. (2006). The stress-activated Hog1 kinase is a selective transcriptional elongation factor for genes responding to osmotic stress. Molecular cell *23*, 241-250.

Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R.*, et al.* (2005). Global analysis of protein phosphorylation in yeast. Nature *438*, 679-684.

Pu, S., Wong, J., Turner, B., Cho, E., and Wodak, S.J. (2009). Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res *37*, 825-831.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods *24*, 218-229.

Quinlan, J.R. (1986). Induction of decision trees. Machine Learning *1*, 81-106.

Reiser, V., D'Aquino, K.E., Ee, L.S., and Amon, A. (2006). The stress-activated mitogen-activated protein kinase signaling cascade promotes exit from mitosis. Mol Biol Cell *17*, 3136-3146.

Saito, H., and Tatebayashi, K. (2004). Regulation of the osmoregulatory HOG MAPK cascade in yeast. J Biochem *136*, 267-272.

Sayed, M., Kim, S.O., Salh, B.S., Issinger, O.G., and Pelech, S.L. (2000). Stress-induced activation of protein kinase CK2 by direct interaction with p38 mitogen-activated protein kinase. J Biol Chem *275*, 16569-16573.

Schadt, E.E., Lamb, J., Yang, X., Zhu, J., Edwards, S., Guhathakurta, D., Sieberts, S.K., Monks, S., Reitman, M., Zhang, C*., et al.* (2005). An integrative genomics approach to infer causal associations between gene expression and disease. Nature genetics *37*, 710-717.

Scherrer, T., Mittal, N., Janga, S.C., and Gerber, A.P. (2010). A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. PLoS One *5*, e15499.

Schwartz, M.A., and Madhani, H.D. (2004). Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. Annual review of genetics *38*, 725-748.

Sharifpoor, S., Ba, A.N.N., Young, J.-Y., van Dyk, D., Friesen, H., Douglas, A.C., Kurat, C.F., Chong, Y.T., Founk, K., Moses, A.M*., et al.* (2011). A quantitative literature-curated gold standard for kinase-substrate pairs. Genome Biol *12*, R39.

Shock, T.R., Thompson, J., Yates, J.R., 3rd, and Madhani, H.D. (2009). Hog1 mitogen-activated protein kinase (MAPK) interrupts signal transduction between the Kss1 MAPK and the Tec1 transcription factor to maintain pathway specificity. Eukaryot Cell *8*, 606-616.

Silverbush, D., Elberfeld, M., and Sharan, R. (2011). Optimally orienting physical networks. Journal of computational biology : a journal of computational molecular cell biology *18*, 1437-1448.

Smets, B., Ghillebert, R., De Snijder, P., Binda, M., Swinnen, E., De Virgilio, C., and Winderickx, J. (2010). Life in the midst of scarcity: adaptations to nutrient availability in Saccharomyces cerevisiae. Curr Genet *56*, 1-32.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology *3*, Article 3.

Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S.G., Cyert, M., Hughes, T.R*., et al.* (2006). Mapping pathways and phenotypes by systematic gene overexpression. Mol Cell *21*, 319-330.

Soufi, B., Kelstrup, C.D., Stoehr, G., Frohlich, F., Walther, T.C., and Olsen, J.V. (2009). Global analysis of the yeast osmotic stress response by quantitative proteomics. Molecular bioSystems *5*, 1337-1346.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Molecular biology of the cell *9*, 3273-3297.

Stark, C., Breitkreutz, B.-J.J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Res *34*, D535-539.

Storey, J.D., Xiao, W., Leek, J.T., Tompkins, R.G., and Davis, R.W. (2005). Significance analysis of time course microarray experiments. Proc Natl Acad Sci U S A *102*, 12837-12842.

Sumner, E.R., and Avery, S.V. (2002). Phenotypic heterogeneity: differential stress resistance among individual cells of the yeast Saccharomyces cerevisiae. Microbiology *148*, 345-351.

Suthram, S., Beyer, A., Karp, R.M., Eldar, Y., and Ideker, T. (2008). eQED: an efficient method for interpreting eQTL associations using protein networks. Mol Syst Biol *4*, 162.

Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S*., et al.* (2010). Chemical-genomic dissection of the CTD code. Nature structural & molecular biology *17*, 1154-1161.

Tiger, C.-F., Krause, F., Cedersund, G., PalmÈr, R., Klipp, E., Hohmann, S., Kitano, H., and Krantz, M. (2012). A framework for mapping, visualisation and automatic model creation of signal-transduction networks. Mol Syst Biol *8*, 578.

Tsvetanova, N.G., Klass, D.M., Salzman, J., and Brown, P.O. (2010). Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. PLoS One *5*.

Tu, Z., Wang, L., Arbeitman, M.N., Chen, T., and Sun, F. (2006). An integrative approach for causal gene identification and gene regulatory pathway inference. Bioinformatics *22*, e489-e496.

Van Wuytswinkel, O., Reiser, V., Siderius, M., Kelders, M.C., Ammerer, G., Ruis, H., and Mager, W.H. (2000). Response of Saccharomyces cerevisiae to severe osmotic stress: evidence for a novel activation mechanism of the HOG MAP kinase pathway. Mol Microbiol *37*, 382-397.

Vaske, C.J., House, C., Luu, T., Frank, B., Yeang, C.-H.H., Lee, N.H., and Stuart, J.M. (2009). A factor graph nested effects model to identify networks from genetic perturbations. PLoS Comput Biol *5*, e1000274.

Venters, B.J., Wachi, S., Mavrich, T.N., Andersen, B.E., Jena, P., Sinnamon, A.J., Jain, P., Rolleri, N.S., Jiang, C., Hemeryck-Walsh, C*., et al.* (2011). A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. Mol Cell *41*, 480-492.

Wall, D.P., Fraser, H.B., and Hirsh, A.E. (2003). Detecting putative orthologs. Bioinformatics *19*, 1710-1711.

Waltermann, C., and Klipp, E. (2010). Signal integration in budding yeast. Biochemical Society transactions *38*, 1257-1264.

Warner, J.R., Vilardell, J., and Sohn, J.H. (2001). Economics of ribosome biosynthesis. Cold Spring Harbor symposia on quantitative biology *66*, 567-574.

Warringer, J., Hult, M., Regot, S., Posas, F., and Sunnerhagen, P. (2010). The HOG pathway dictates the short-term translational response after hyperosmotic shock. Molecular biology of the cell *21*, 3080-3092.

Wenger, C.D., Lee, M.V., Hebert, A.S., McAlister, G.C., Phanstiel, D.H., Westphall, M.S., and Coon, J.J. (2011a). Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. Nat Methods *8*, 933-935.

Wenger, C.D., Phanstiel, D.H., Lee, M.V., Bailey, D.J., and Coon, J.J. (2011b). COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. Proteomics *11*, 1064-1074.

West, M.L., and Corden, J.L. (1995). Construction and Analysis of Yeast RNA Polymerase II CTD Deletion and Substitution Mutations. Genetics *140*, 1223-1233.

Westfall, P.J., Patterson, J.C., Chen, R.E., and Thorner, J. (2008). Stress resistance and signal fidelity independent of nuclear MAPK function. Proceedings of the National Academy of Sciences of the United States of America *105*, 12212-12217.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H.*, et al.* (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science *285*, 901-906.

Woods, J.O., Singh-Blom, U.M., Laurent, J.M., McGary, K.L., and Marcotte, E.M. (2013). Prediction of gene-phenotype associations in humans, mice, and plants using phenologs. BMC Bioinformatics *14*, 203.

Ye, T., Elbing, K., and Hohmann, S. (2008). The pathway by which the yeast protein kinase Snf1p controls acquisition of sodium tolerance is different from that mediating glucose regulation. Microbiology *154*, 2814-2826.

Yeang, C.-H., Ideker, T., and Jaakkola, T. (2004a). Physical network models. Journal of Computational Biology *11*, 243-262.

Yeang, C.H., Ideker, T., and Jaakkola, T. (2004b). Physical network models. Journal of computational biology : a journal of computational molecular cell biology *11*, 243-262.

Yeang, C.H., Mak, H.C., McCuine, S., Workman, C., Jaakkola, T., and Ideker, T. (2005). Validation and refinement of gene-regulatory pathways on a network of physical interactions. Genome Biol *6*, R62.

Yeger-Lotem, E., Riva, L., Su, L.J., Gitler, A.D., Cashikar, A.G., King, O.D., Auluck, P.K., Geddie, M.L., Valastyan, J.S., Karger, D.R.*, et al.* (2009). Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. Nat Genet *41*, 316-323.

Yeung, K.Y., Medvedovic, M., and Bumgarner, R.E. (2004). From co-expression to co-regulation: how many microarray experiments do we need? Genome Biol *5*, R48.

Yosef, N., Ungar, L., Zalckvar, E., Kimchi, A., Kupiec, M., Ruppin, E., and Sharan, R. (2009). Toward accurate reconstruction of functional protein networks. Mol Syst Biol *5*, 248.

You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C.W., Wang, Y.P., Lenz, P., Yan, D., and Hwa, T. (2013). Coordination of bacterial proteome with metabolism by cyclic AMP signalling. Nature *500*, 301-306.

Zakrzewska, A., van Eikenhorst, G., Burggraaff, J.E., Vis, D.J., Hoefsloot, H., Delneri, D., Oliver, S.G., Brul, S., and Smits, G.J. (2011). Genome-wide analysis of yeast stress survival and tolerance acquisition to analyze the central trade-off between growth rate and cellular robustness. Molecular biology of the cell *22*, 4435-4446.

Zarrinpar, A., Bhattacharyya, R.P., Nittler, M.P., and Lim, W.A. (2004). Sho1 and Pbs2 act as coscaffolds linking components in the yeast high osmolarity MAP kinase pathway. Mol Cell *14*, 825-832.

Zhang, D.W., Rodriguez-Molina, J.B., Tietjen, J.R., Nemec, C.M., and Ansari, A.Z. (2012). Emerging Views on the CTD Code. Genetics research international *2012*, 347214.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T.*, et al.* (2001). Global analysis of protein activities using proteome chips. Science *293*, 2101-2105.