# DIVERSE MECHANISMS OF TRANSCRIPTION REGULATION BY RNA POLYMERASE II IN SACCHAROMYCES CEREVISIAE

by

Maria Aristizabal

B.A., The University of British Columbia, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2015

## Abstract

Transcription is the process by which information encoded in the DNA is read by the cell. RNAPII is a key enzyme in this process and it contains a unique and functionally conserved C-terminal domain (CTD) composed of heptapeptide repeats. The CTD plays essential roles in the coordination of transcription events and this is mediated by its ability to be differentially modified throughout the transcription cycle. As such, a number of CTD modifying enzymes act at different stages on the CTD, including the phosphatase Fcp1, which associates with RNAPII along the length of transcribed genes and plays key roles in transcription elongation and recycling. In addition to the modification status, both the sequence and length of the CTD are also important for normal function. This thesis focuses on understanding the role of the RNAPII-CTD and its modifying enzyme, Fcp1, in gene expression and in understanding how the CTD and the process of transcription contribute to genome instability. We found clear evidence that despite their general role in transcription both CTD length and Fcp1 function have gene-specific effects on transcription. In particular, we show that truncation of the CTD results in an increasing number of transcriptional defects. These defects resulted primarily from changes in initiation and were suppressed by loss of a previously reported CTD suppressor, *CDK8*. Additionally, our work on Fcp1 mutants provides a framework for exploring a general role for Fcp1 in the regulation of transcription factors. Furthermore, we find that truncating the CTD also resulted in altered expression of mobile genetic elements, a previously unreported phenotype. Here, we show that loss of *CDK8* was able to suppress the increased expression of retrotransposons, expanding our understanding of retrotransposon regulation *in vivo*. We also find that this effect is mediated by the transcription factors Ste12 and Tec1 and that loss of either of these can suppress growth and

retrotransposon expression defects of CTD truncation mutants. Finally, we report the genome-wide distribution of DNA:RNA hybrids in *S. cerevisiae*. By comparing this profile to that of mutants in genes encoding RNA or DNA:RNA hybrid processing factors, we show differential effects on DNA:RNA hybrid abundance and localization.

## Preface

The work presented in chapter 2 was published in "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 Genetics". As the sole first author, I was involved in the experimental design, data interpretation and manuscript writing. Experimentally, I was involved in the generation of all the strains and plasmids, all of the biochemical and genetic assays, the ChIP-on-chip experiments and some of the data analysis. Gian Luca Negri was involved with the data analysis of the ChIP-on-chip and gene expression results and was involved in writing most scripts necessary for the data analysis. Joris Benschop performed and analyzed the gene expression micro arrays. For chapter 3, I performed and conceived all of the experiments and performed most of the data analysis. Gian Luca helped with data analysis and sequence alignments for the RT-qPCR assay. This work will constitute a future publication. Chapter 4 corresponds to work for a future publication, and here, the gene expression arrays were performed and analyzed by Jorish Benschop. Gian Luca Negri helped with the transcription factor enrichment analysis. I performed and conceived the rest of the experiments for this chapter. The last chapter was work done in collaboration with Dr. Phillip Hieter's laboratory at the University of British Columbia. This work is published in "Chan Y. A., Aristizabal M. J., Lu P. Y. T., Luo Z., Hamza A., Kobor M. S., Stirling P. C., Hieter P. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. PLoS Genetics". As a co-first author of this manuscript I was primarily involved in the data analysis of the DRIP-chip data, which makes up a significant portion of this project. I was also involved in both the figure generation and editing of the manuscript. Experimentally, I

helped with the DRIP-chip protocol and the DRIP-qPCR. Alina Chan performed and conceived all of the remaining experiments and wrote most of the manuscript.

# Table of Contents

## List of Tables

## List of Figures

## List of Abbreviations

| | |
|---|---|
| RNAP | RNA Polymerase |
| RNAPII | RNA Polymerase II |
| CTD | C-terminal domain |
| TSS | Transcription start site |
| ORF | Open reading frame |
| mRNA | Messenger RNA |
| rRNA | Ribosomal RNA |
| snRNA | small nucleolar RNA |
| ncRNA | Non-coding RNA |
| tRNA | Transfer RNA |
| TERRA | Telomere repeat-containing RNA |
| H3K36me3 | Histone H3 lysine 36 tri-methylation |
| GO | Gene ontology |
| E-MAP | Epistasis MiniArray profiling |
| DAmP | Decreased abundance by mRNA perturbation |
| ChIP | Chromatin immunoprecipitation |
| ChIP-on-chip | Chromatin immunoprecipiation microarray |
| qPCR | Quantitative polymerase chain reaction |
| RT | Reverse transcriptase |
| MMS | Methyl methanesulfonate |
| HU | Hydroxyurea |

| | |
|---|---|
| NAT | Gene that confers CloNat resistance |
| OD | Optical density |
| YPD | yeast extract- peptone-dextrose media |
| SC | Synthetic complete media |
| -URA | Growth media lacking uracil |
| -TRP | Growth media lacking tryptophan |
| WT | Wild type |
| BPS | Bathophenanthroline disulfonate |

## Acknowledgements

This thesis would not have been possible without my supervisor Dr. Michael Kobor, whom not only started some of these projects but whom most importantly taught me to work hard. His guidance in science and his encouragement to pursue learning opportunities are most appreciated. Of course, my time in the lab would not have been as enjoyable had it not been for my lab mates with whom not only did I share ideas but most importantly life. Thank you to all the current and previous Kobor lab members for their support, friendship, and fun memories. In particular, I want to thank Grace Leung, Phoebe Lu and Alice Wang who really were sisters throughout my entire PhD and to whom I owe a great deal. I want to thank Dr. Nevan Krogan and Dr. Gerard Cagney for welcoming me into their labs for a temporary period and giving me the opportunity to expand my scientific knowledge. Dr. Ivan Sadowski, thank you for the scientific discussions, reagents, and meaningful collaborations. To my committee members Dr. James Kronstad and Dr. LeAnn Howe, thanks for all the critical advice throughout the years. I also want to thank Alina Chan and Peter Stirling for sharing your research projects with me and for letting me contribute to your work. Finally, to the funding agencies, CIHR, NSERC, CFRI, and ICUF thank you for making it possible for me to train in this and other great environments and to contribute a small piece to our understanding of transcription regulation.

I have been extremely privileged to have a very large and supportive family. Although most of them have been far away during this journey, they have always made sure to share in my accomplishments and celebrate my successes. To La Tia Amparo, who often prayed for the success of my experiments, gracias por mostrarme lo que es el amor incondicional. To my

parents, who have taught me the importance of education and who always told me to be the best at what I do, thanks for all the opportunities you gave me and the sacrifices you made. I am extremely lucky to have you both as my parents. You have both taught me very much and I look forward to continuing to grow and follow this journey together. To my sister, who has promised to fund my lab once she makes millions, thanks for being the best and only sister one could ever hope for and for enriching my life by challenging me and showing me life from many different angles. Gian Luca Negri, thank you for everything you do, specially laughing with me. You have bettered my life in so many ways.

## Dedication

To my family

# Chapter 1: Introduction

## 1.1 Transcription as a mechanism of accessing DNA encoded information

RNA is synthesized during a process called transcription by an enzyme called RNA Polymerase (RNAP) (Werner and Grohmann 2011). Transcription begins with the recruitment of RNAP to the upstream promoter region of a gene, a step that is tightly regulated in part by sequence-specific and general transcription factors. Following, DNA is unwound allowing RNAP to select the template strand and use it as a guide for the production of a complementary RNA molecule. RNA synthesis proceeds, one nucleotide at a time until the end of the gene. Here sequence information and termination factors help RNAP dissociate from the DNA, thus culminating in one round of transcription and the production of an RNA molecule.

Transcription regulation is an important step in maintaining normal cellular function, cell growth, and differentiation. This is in part because RNAs serves many essential roles in the cell. For instance, messenger RNA (mRNA) is the key intermediate in synthesis of proteins by the ribosome, and ribosomal RNA (rRNA) comprises the catalytic center and performs the enzymatic activity of the ribosome (Simonovic and Steitz 2009). Therefore, given its basic importance, transcription is a process common to all domains of life, and as such many of the key enzymes involved in this process are highly conserved and homologs are clearly evident (Allison et al. 1985, Chapman et al. 2008, Guo and Stiller 2005, Werner and Grohmann 2011, Yang and Stiller 2014).

## 1.2 Eukaryotes contain three RNA polymerases

While prokaryotes have only one RNAP responsible for the transcription of all their genes, eukaryotes have evolved specialized complexes, such that transcription is performed by one of three different RNA polymerases (RNAPI, RNAPII or RNAPIII) (Roeder and Rutter 1969). Although RNAPI, RNAPII, and RNAPIII are highly homologous and share a number of subunits, each is responsible for the transcription of a different set of genes and interacts with a different set of general transcription factors (Sakurai and Ishihama 2002, Vannini and Cramer 2012). More specifically, RNAPI is responsible for the production of rRNA, RNAPII for mRNA, many non-coding RNAs (ncRNA), and small nucleolar RNAs (snRNA), and RNAPIII for transfer RNAs (tRNA). Although all three RNAPs perform essential functions in the cell, this dissertation will focus on the biology of RNAPII, in part because it transcribes a large proportion of all encoded genes in the genome and because it synthesizes mRNAs, a key precursor for the synthesis of all proteins.

## 1.3 Composition and structural transitions of RNAPII during transcription progression

RNAPII is composed of 12 subunits, five of which are shared with RNAPI and RNAPIII (Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12) (Cramer et al. 2000, Cramer 2002). Rpb1, the largest and catalytic subunit, along with Rpb2, forms a deep cleft harboring the catalytic center. Rpb4 and Rpb7 associate transiently with the rest of the RNAPII, and play key roles in promoter-dependent initiation and survival during stress conditions (Armache et al. 2005, Maillet et al. 1999, Sheffer et al. 1999). The structure of RNAPII changes during transcription. It begins with

a closed complex, that transitions into an open complex, before becoming a transcription elongation complex (Cramer 2002, Kettenberger et al. 2004, Martinez-Rucobo and Cramer 2013, Shandilya and Roberts 2012, Treutlein et al. 2012, Vannini and Cramer 2012). The closed complex is formed upon RNAPII's association with double-stranded DNA. In this conformation, the catalytic center is closed and RNA synthesis does not occur. DNA unwinding characterizes the open RNAPII complex and is key for establishing the transcription bubble, a 8-9 base pair DNA:RNA hybrid. The open complex is capable of RNA synthesis, but its activity if generally abortive, resulting in the production and release of short RNAs. In contrast, the elongation complex is processive and this is accomplished by stabilization of the transcription bubble and release of RNAPII from promoter elements.

### 1.4 A set of factors help RNAPII initiate transcription and transition to the elongation complex

RNAPII promoter-dependent transcription initiation is mediated by a set of general transcription factors (TFIIs) including TFIIA, TFIIB, TFIID, TFIIF, TFIIH and the Mediator complex, which together with RNAPII form the promoter initiation complex (PIC) (Grunberg and Hahn 2013, Thomas and Chiang 2006). These components can be recruited to the promoter in a stepwise manner or as a pre-assembled complex. Together, they function to recognize the promoter, recruit, anchor and orient RNAPII, unwind DNA, and release RNAPII from promoter elements and into transcription elongation. These roles are mediated by physical interactions between PIC components and DNA motifs or RNAPII subunits, and by catalytic activities that are harbored on a subset of PIC subunits. Focusing on the latter, the TFIIH subunit, Ssl2, is a double stranded

DNA translocase involved in DNA unwinding, thus promoting the transition of RNAPII from a closed to an open complex (Grunberg et al. 2012). The TFIIH subunit, Kin28/Cdk7, is a kinase that phosphorylates the RNAPII C-terminal domain (CTD), a domain that will be introduced in more detail later. CTD phosphorylation by Kin28 promotes release of RNAPII from promoter elements, thus stimulating the transition from an open to an elongation complex (Liu et al. 2004). The Mediator complex kinase, Cdk8 can also target the CTD, resulting in transcription activation or repression (Poss et al. 2013). *In vitro,* phosphorylation of the CTD by Cdk8, prior to PIC formation, inhibits transcription, while phosphorylation following PIC formation stimulates transcription in a manner similar to Kin28 (Liu et al. 2004). Finally, the role of the PIC is not limited to transcription initiation, as some components also have roles in transcription elongation and re-initiation (Poss et al. 2013, Shandilya and Roberts 2012). Briefly, PIC components have been detected at the promoter following RNAPII release into transcription elongation. Here, they have been shown to facilitate transcription re-initiation (Yudkovsky et al. 2000).

### 1.5 The C-Terminal domain of RNAPII coordinates co-transcriptional activities

Rpb1 contains a C-terminal domain (CTD) composed of tandem repeats of a core seven amino acid motif. The CTD is essential for viability and functionally conserved across species. However, the sequence and the number of repeats vary between species, a topic that will be discussed in more detail below (Allison et al. 1985, Corden et al. 1985, Corden 1990). This domain is found exclusively on RNAPII of eukaryotes, and it forms an unstructured tail that extends from the RNAPII complex. Functionally, the CTD plays key auxiliary roles during transcription, including orchestrating co-transcriptional processes such as mRNA capping,

splicing, and polyadenylation (Buratowski and Sharp 1990, Buratowski 2009, Heidemann et al. 2012, Hsin and Manley 2012, Kim and Dahmus 1989, Zehring et al. 1988, Zhang et al. 2012b). This role is mediated by its ability to act as a recruiting platform for many regulatory and RNA-processing factors, a function linked to the CTD's ability to be differentially modified throughout the transcription cycle. Although the CTD can be modified in many different ways, including proline isomerization, methylation, and glycosylation, arguably phosphorylation has emerged as the most important modification studied to date, and will be discussed in detail later (Heidemann et al. 2012).

### 1.6 Properties that influence CTD function

The function of the CTD is highly conserved and this is highlighted by the ability of sequences from other organisms to functionally replace the *S. cerevisiae* CTD. For example, the mammalian CTD is sufficient to support normal growth in *S. cerevisiae,* although it is twice as long and has a more variable core motif (Allison et al. 1988). Similarly, replacing the *S. cerevisiae* CTD with that of a primitive protist, *Mastigamoeba invertens,* results in minimal effects on growth (Stiller et al. 2000). However, and indicative of some degree of species-specific function, replacing the *S. cerevisiae* CTD with that of *D. melanogaster* or a red algae, is not tolerated and results in lethality (Allison et al. 1988, Stiller and Cook 2004). Thus, although generally conserved, the CTD sequence and function likely varies across species. Here we discuss how characteristics such as length, sequence motif, and repeat spacing affect normal CTD function.

### 1.6.1 The role of RNAPII-CTD length in normal function

CTD length is a feature that varies greatly across species and broadly speaking, longer CTDs are correlated with increasing genomic complexity (Chapman et al. 2008, Yang and Stiller 2014). For example, the CTD of *S. cerevisiae* has 25-26 repeats, *Schizosaccharomyces pombe* has 29 repeats, *Caenorhabditis elegans* has 32 repeats, and chordates (rodents, human, and chicken) have 52 repeats. Although it remains unclear why CTD length varies across species, previous work has shown the importance of normal length for CTD function. Furthermore, given that the CTD is essential for viability, organisms vary in the minimal length required to support life and display growth defects and deficiencies in transcription regulation and processing when carrying viable yet shortened version of the CTDs (Zehring et al. 1988). For example, in *S. cerevisiae* a minimum of 8 repeats are required for viability and strains carrying 9-13 repeats display conditional growth phenotypes (Allison et al. 1988, Nonet et al. 1987). Similarly, *S. pombe* requires 10 repeats for viability and strains carrying 10-13 repeats also show conditional growth phenotypes (Schneider et al. 2010). Rodent and chicken cells require 25-26 repeats for viability and introduction of shortened CTDs to these cells results in reduced cell size (Bartolomei et al. 1988, Hsin et al. 2014). Mice with shortened CTDs display increased neonatal lethality and reduced body size compared to wild type littermates (Litingtung et al. 1999). At the molecular level, shortened CTDs have been associated with mRNA capping, splicing and termination defects (de la Mata and Kornblihtt 2006, McCracken et al. 1997a, Rosonina and Blencowe 2004, Ryan et al. 2002, Suh et al. 2010). Furthermore, and indicative of gene-specific roles for the RNAPII-CTD in *S. cerevisiae* and mammals, CTD truncations are specifically associated with defects in activated transcription (Allison and Ingles 1989, Liao et al. 1991, Pan et al. 2006, Scafe et al. 1990). Thus, CTD length is intimately linked with CTD function. This topic is further

investigated in Chapter 2 and 3 of this thesis wherein I investigate how changes in CTD length result in alterations to steady state gene expression in *S. cerevisiae*.

### 1.6.2 The core CTD motif is variable across species

The CTD core heptapeptide motif generally follows an $Y_1S_2P_3T_4S_5P_6S_7$ consensus sequence; however, deviations are common across species (Allison et al. 1985, Corden et al. 1985). Generally speaking, deviations from the consensus motif are correlated with developmental complexity, multicellularity and parasitic lifestyles (Yang and Stiller 2014). For instance, most *S. cerevisiae* CTD repeats follow tightly the consensus sequence, most of the *D. melanogaster*'s repeats diverge from the consensus sequence, and the mammalian CTD is composed of an N-terminal half that follows closely the consensus sequence and a C-terminal half that is more variable. Regardless of the observed species-specific variability a few amino acids making up the CTD sequence are less variant. Specifically, $Y_1$ and $S_5$ are the most conserved residues and mutating either results in lethality in most species tested (Egloff et al. 2007, Hintermair et al. 2012, Schwer and Shuman 2011, West and Corden 1995, Zhang et al. 2012a). Nonetheless, canonical and non-canonical repeats are important for normal CTD function. For instance, in chicken cells, CTDs engineered to contain no canonical repeats are inviable, and in mammals, CTDs made up of only consensus repeats are unable to support life (Bartolomei et al. 1988, Chapman et al. 2005, Hsin et al. 2014). More specifically, it was observed that non-consensus repeats at positions 1-3 and 52 of the mammalian CTD are required for stability. This differs from the *S. cerevisiae* CTD, wherein loss of the most C-terminal repeats have no effect on protein stability (Aristizabal et al. 2013). Overall, the CTD is highly conserved despite sequence

variation being observed frequently, and in some instances deviations from the consensus motif has conferred specific functions to individual CTD repeats.

### 1.6.3 Repeat spacing and positioning are important for normal CTD function

Repeat spacing and CTD location relative to the RNAPII complex are also important for normal function. In yeast separating individual CTD repeats by the addition of alanine residues results in lethality, while separating heptapeptide pairs, results in normal function (Schwer et al. 2012, Stiller and Cook 2004). Therefore, it is broadly accepted that the functional unit of the CTD is a heptapeptide pair, which is consistent with the *Candida albicans* capping enzyme contacting residues spanning multiple CTD repeats in a Cgt1-CTD peptide co-crystal structure (Fabrega et al. 2003).

The CTD extends in an unstructured manner from RNAPII. Recently, CTD positioning relative to RNAPII was shown to be important for function, although in some instances transplanting the CTD to the C-terminus of other RNAPII subunits could confer normal function (Suh et al. 2013). In particular, fusing the CTD to subunits with C-termini near the Rpb1 C-terminus in the RNAPII 3D structure resulted in normal growth, while fusion to subunits far away from the normal location were insufficient to support viability. Therefore, the extended nature of the CTD allows it to function independently of Rpb1, however there are constraints regarding its positioning relative to the RNAPII enzyme, the nature of which requires further investigation.

## 1.7 Suppressors of RNAPII-CTD truncation phenotypes

Seminal work aimed at investigating the role of the RNAPII-CTD utilized genetic screens, resulting in the identification of Suppressor of RNA Polymerase B genes (SRB genes), many of which were later identified as members of the Mediator complex (Hengartner et al. 1995, Koleske et al. 1992, Liao et al. 1995, Nonet and Young 1989, Thompson et al. 1993). *SRB2-SRB11* encode Mediator subunits, and their mutation suppressed the cold sensitive phenotype of strains carrying shortened CTDs. Of these, *SRB8/MED12, SRB9/MED13*, *SRB10/CDK8* and *SRB11/CYCC* all belong to the Mediator's Cdk8 module, which associates transiently with the core Mediator complex and is implicated in transcriptional activation and repression (Galbraith et al. 2010, Tsai et al. 2013). Furthermore, Cdk8 has kinase activity that can modify the CTD and many other transcription factors (Poss et al. 2013). Investigating the genetic relationship between *CDK8* and the RNAPII-CTD revealed that loss of *CDK8/SRB10* could suppress many CTD-dependent phenotypes including the slow growth of CTD truncation mutants when exposed to high temperatures, in the absence of inositol, or upon presence of the chemicals hydroxyurea or formamide (Aristizabal et al. 2013, Hengartner et al. 1998, Wong and Ingles 2001). However, despite its initial discovery in 1989, the molecular underpinning of the genetic suppression of CTD truncation phenotypes by the loss of *CDK8/SRB10*, have remained unclear and are further investigated in Chapter 2 and 3 of this thesis.

In addition to Mediator subunits, other genes were identified as suppressors of RNAPII-CTD truncation phenotypes, namely loss of *SPT2* and mutations of histone H3 (Kruger et al. 1995, Peterson et al. 1991). Spt2 is a histone chaperone involved in histone deposition following

transcription (Thebault et al. 2011, Winston et al. 1984). Loss of *SPT2* or mutations on the histone H3 gene can suppress the growth defects and decreased *INO1* mRNA levels associated with strains carrying shorter CTDs (Kruger et al. 1995, Peterson et al. 1991). Furthermore, loss of *SPT2* can suppress the lethality of *S. cerevisiae* strains carrying CTDs with only seven heptapeptide repeats, although this seems to be yeast strain specific (Peterson et al. 1991, Yuryev and Corden 1996). While the molecular mechanisms of these genetic interactions remain unclear, they are consistent with the well-documented role of the RNAPII-CTD in chromatin remodeling during transcription, which is described in brief later (Zhang et al. 2012b). In Chapter 3, I describe the identification of two new suppressors of CTD truncation phenotypes. These highlight novel roles for the RNAPII-CTD in the transcriptional regulation of genome-encoded retrotransposons.

## 1.8 The CTD phosphorylation cycle marks transcription progression

The transcription of protein-coding genes is roughly divided into four steps: initiation, promoter escape, elongation, and termination, and each step is characterized by a particular CTD phosphorylation status (**Figure 1.1**) (Heidemann et al. 2012). In particular, $Y_1$, $S_2$, $T_4$, $S_5$, and $S_7$ residues of the heptapeptide repeat are phosphorylated during the transcription of protein coding genes and these play key roles during transcription progression. Testifying to the CTD's potential for information complexity, it has been reported that a single mammalian CTD can contain up to 50 phosphates (Dahmus 1994). However, it remains unclear where along the CTD length modifications occur and how this changes as RNAPII moves along the transcription cycle. Much of what we know about the CTD phosphorylation cycle has been elucidated in the model

organism, *S. cerevisiae*. However, the distribution of CTD modifications and the enzymes involved are conserved in other organisms (Bataille et al. 2012, Coudreuse et al. 2010, Garrido-Lecca and Blumenthal 2010, Kim et al. 2010, Mayer et al. 2010, Odawara et al. 2011, Rahl et al. 2010, Tietjen et al. 2010). This thesis focuses on the *S. cerevisiae* enzymes unless otherwise stated. **Table 1.1** lists the majority of CTD modifying enzymes, their mammalian homologs and their specificities.

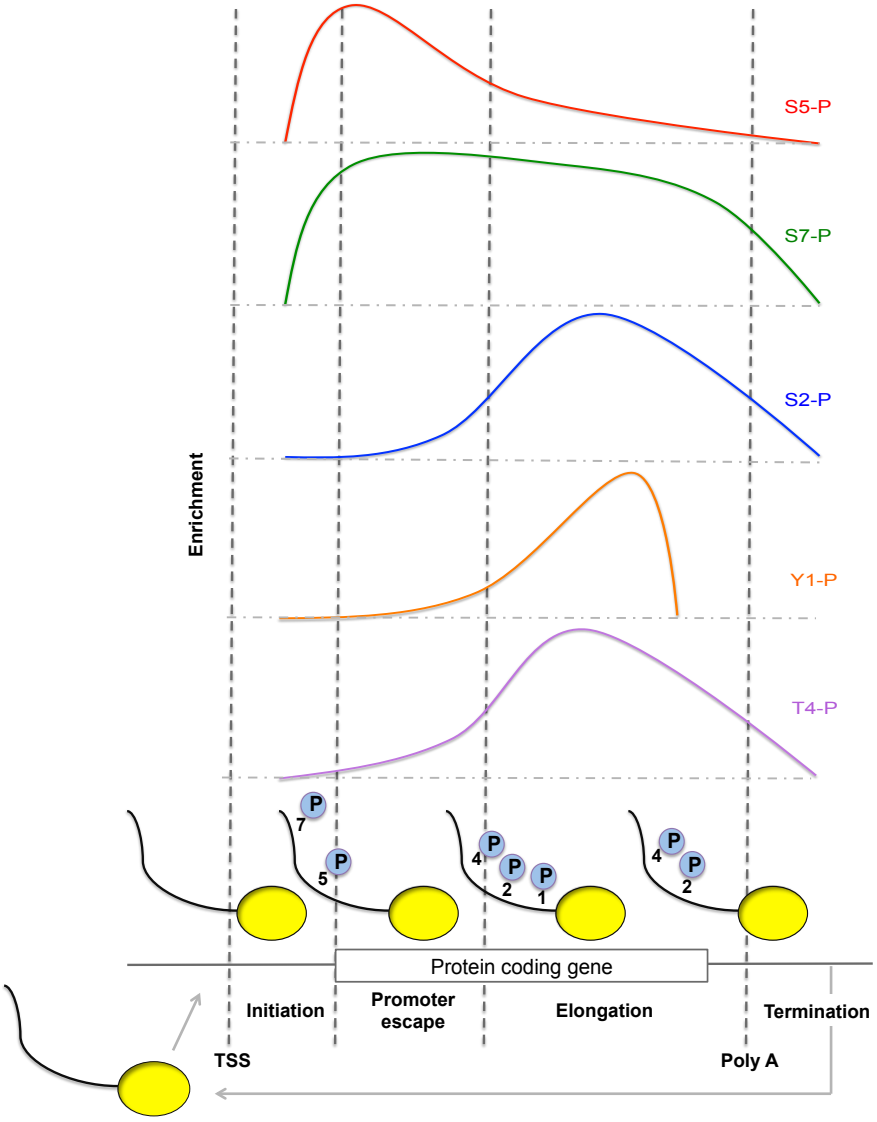



**Figure 1.1 Differential phosphorylation of the RNAPII-CTD marks transcription progression.**
Transcription begins with the recruitment of an unmodified CTD to promoter regions. $S_5$ and $S_7$ phosphorylation are high at the 5' end of genes and decrease towards the 3' end. $Y_1$, $S_2$ and $T_4$ phosphorylation levels are low at the 5' end of genes and increased toward to end of the gene. $Y_1$ is dephosphorylated prior to RNAPII reaching the polyA site, differentiating transcription elongation from termination. Each line graph represents the relative enrichment of each CTD phosphorylation mark along the length of the gene. Transcriptional stages are approximate.

**Table 1.1 Enzymes involved in the CTD modification cycle**

| Enzymatic activity | *S. cerevisiae* Protein | Mammalian homolog | Substrate specificity | Location | Differences *S. cerevisiae* vs mammals | Protein Complex |
|---|---|---|---|---|---|---|
| Kinase | Kin28 | Cdk7 | $S_5$ and $S_7$ | 5' end of genes | | TFIIH |
| Kinase | Bur1 | Cdk9 | $S_2$, $S_5$* and $S_7$* | $S_2$ at 5' end of genes, $S_5$ and $S_7$ middle to 3' end of genes | | pTEFb in mammals |
| Kinase | Cdk8/Srb10 | Cdk8 | $S_5$* | before PIC assembly or after PIC assembly | | Mediator complex |
| Kinase | Ctk1 | Cdk12 | $S_2$ | middle to 3' end of genes | | CTDK-1 |
| Kinase | | Polo-like Kinase 3 | $T_4$ | unclear | *S. cerevisiae* homolog not tested | |
| Phosphatase | Rtr1 | Rpap2 | $S_5$ and $Y_1$* | $S_5$ at 5' end of genes, $Y_1$ unknown | No evidence for $Y_1$ activity in mammals | |
| Phosphatase | Ssu72 | Ssu72 | $S_5$ and $S_7$ | $S_5$ and $S_7$ at 3' end of genes | | |
| Phosphatase | Fcp1 | Fcp1 | $S_2$, $S_5$* and $T_4$ | $S_2$ at 3' end of genes. $S_5$ unknown | | |
| Phosphatase | Glc7 | | $Y_1$ | 3' end of genes | Mammalian homolog not tested | |
| Phosphatase | Cdc14 | HCDC14A | $S_2$ and $S_5$ | at specific loci | Unknown activity mammals | |
| Proline Isomerase | Ess1 | Pin1 | $S_2$-$P_3$ bonds* and $S_5$-$P_6$ | along the length of genes | Different functional consequences | |

| Enzymatic activity | *S. cerevisiae* Protein | Mammalian homolog | Substrate specificity | Location | Differences *S. cerevisiae* vs mammals | Protein Complex |
|---|---|---|---|---|---|---|
| | | | bonds | | in mammals vs *S. cerevisiae* | |

*Indicates substrate specificities that are minor or unclear

### 1.8.1 RNAPII with an unmodified CTD begins transcription

Transcription begins with the recruitment of RNAPII with an unmodified CTD to promoters (Lu et al. 1991). This is mediated by gene-specific and general transcription factors (Grunberg and Hahn 2013, Thomas and Chiang 2006). Gene-specific factors contain DNA binding domains and activation domains. The former determine the sequence specificity of the transcription factor, while the latter enhances transcription. The molecular underpinnings of how gene-specific transcription factors function have remained somewhat of a mystery. Nonetheless, they likely stimulate transcription by interacting with transcription co-activators such as the Mediator complex (Dyson and Wright 2005). Furthermore, *in vitro* they can form polymers that specifically bind components of the transcription machinery, including an unmodified RNAPII-CTD (Kato et al. 2012, Kwon et al. 2013).

### 1.8.2 $S_5$ phosphorylation stimulates promoter escape

Promoter escape is characterized by the release of RNAPII from the PIC. This is achieved in part via phosphorylation of $S_5$ residues by the TFIIH-associated kinase, Kin28, although it can be mediated by the Bur1 or Cdk8 kinases to a lesser extent (**Figure 1.2**) (Akoulitchev et al. 1995, Heidemann et al. 2012, Hengartner et al. 1998, Liu et al. 2004). TFIIH-dependent $S_5$ phosphorylation disrupts key contacts between the CTD and Mediator subunits, and interestingly

this is stimulated by the Mediator complex itself (Esnault et al. 2008, Jeronimo and Robert 2014, Max et al. 2007, Wong et al. 2014). Additionally, $S_5$ phosphorylation also facilitates the recruitment of chromatin remodelers and capping factors to the 5' end of protein coding genes (Buratowski 2009, Kobor et al. 2002, Komarnitsky et al. 2000, Komarnitsky et al. 2000, McCracken et al. 1997a, Zhang et al. 2012b). The latter is thought to underlie the essential role of $S_5$ phosphorylation, as in *S. pombe* fusing a component of the capping machinery to RNAPII can overcome the lethality of a $S_5$ to alanine CTD mutant (Schwer and Shuman 2011). Following $S_5$ phosphorylation, Kin28 also phosphorylates the CTD at $S_7$ residues in a Mediator-dependent fashion (Akhtar et al. 2009, Boeing et al. 2010, Glover-Cutter et al. 2009, Kim et al. 2009). $S_7$ phosphorylation has been detected at most protein-coding genes in a number of species (Bataille et al. 2012, Kim et al. 2010, Mayer et al. 2010, Tietjen et al. 2010). In mammals and *S. pombe* $S_7$ phosphorylation enhances P-TEFb recruitment, thus promoting a switch from initiation to elongation, however its exact role in *S. cerevisiae* remains unclear (Czudnochowski et al. 2012, St Amour et al. 2012).

### 1.8.3 $S_2$ and $Y_1$ phosphorylation mark transcription elongation

Transcription elongation is characterized by low $S_5$ phosphorylation and high of $S_2$ and $Y_1$ phosphorylation levels. This is accomplished in part by the phosphatases Rtr1 and Fcp1, and the kinases Ctk1 and Bur1 (Bartkowiak et al. 2010, Bowman and Kelly 2014, Cho et al. 2001, Egloff et al. 2012, Mosley et al. 2009, Qiu et al. 2009). More specifically, Bur1 is thought to first recognizes $S_5$ phosphorylated CTDs at the 5' end of genes, resulting in the phosphorylation of a small number of $S_2$ residues. This primes the CTD for robust phosphorylation by the primary $S_2$ kinase, Ctk1, along the length of genes (Bartkowiak et al. 2010, Bowman and Kelly 2014, Qiu et

al. 2009). The $S_5$ phosphatase, Rtr1 also functions at the 5' end of genes and mediates the transition from initiation to elongation. Rtr1 recruitment to the CTD is stimulated by $S_2$ phosphorylation and proline isomerization of the CTD by Ess1 (Ma et al. 2012, Singh et al. 2009, Smith-Kinnaman et al. 2014). Therefore, $S_2$ phosphorylation stimulates $S_5$ dephosphorylation, resulting in differential $S_2$ and $S_5$ phosphorylation levels along the length of genes. (Egloff et al. 2012, Kim et al. 2009, Mosley et al. 2009). $S_2$ phosphorylation levels along the length of genes are further fine-tuned by Ctk1 and Fcp1, both of which associate with RNAPII along the length of genes and play roles in stimulating transcription elongation (Bataille et al. 2012, Cho et al. 2001). Although Fcp1 can dephosphorylate the RNAPII-CTD during elongation, how Fcp1 stimulates transcription elongation remains unclear, as there is evidence that this can occur via catalytic-independent mechanisms (Cho et al. 2001, Cho et al. 1999, Kong et al. 2005). During elongation, $Y_1$ is also phosphorylated by a yet unknown kinase (Mayer et al. 2012). High $S_2$ and $Y_1$ phosphorylation levels along the length of genes mediates the recruitment of elongation factors, chromatin-remodelers, and splicing factors, which collectively function to stimulate transcriptional activity (Komarnitsky et al. 2000, Mayer et al. 2012, Zhang et al. 2012b). Finally, $T_4$ residues are also phosphorylated during elongation and in mammals this is mediated by the Polo-like kinase 3 (Hintermair et al. 2012, Mayer et al. 2012). Although $T_4$ phosphorylation is observed at protein coding genes in many species, thus far it is primarily associated with gene and species-specific activities, which are described in more detail below. In mammals and yeast, Fcp1 can dephosphorylate $T_4$ residues, although it remains to be seen where along the length of genes this activity is performed (Allepuz-Fuster et al. 2014, Hsin et al. 2014).

### 1.8.4 Transcription termination involves removal of all CTD phosphorylation marks

Transcription termination is marked by dephosphorylation of $Y_1$ residues by Glc7 and/or Rtr1 prior to RNAPII reaching the poly-adenylation (pA) site (Mayer et al. 2012, Hsu et al. 2014, Schreieck et al. 2014). The differential levels of $S_2$ and $Y_1$ phosphorylation (high $S_2$ and low $Y_1$ phosphorylation) at the 3' end of genes result in the preferential recruitment of termination factors (Mayer et al. 2012). Following mRNA termination, the CTD is completely dephosphorylated to regenerate an unphosphorylated RNAPII complex capable of initiating another transcription cycle (Kobor et al. 1999). At this stage, Ssu72 is responsible for dephosphorylating $S_5$ and $S_7$ residues and Fcp1 for $S_2$ residues, although it is likely that other phosphatases also play a role (Bataille et al. 2012, Cho et al. 1999, Krishnamurthy et al. 2004, Zhang et al. 2012a). Upon termination, the unmodified RNAPII dissociates from the template DNA, freeing the enzyme so it can start another round of transcription. Chapter 4 of this thesis focuses on further investigating the role of Fcp1 during RNAPII-dependent transcription.

### 1.8.5 Deviations from the phosphorylation cycle

The CTD phosphorylation cycle described above is generally observed in most species at most protein-coding genes, however limited reports have described gene-specific and species-specific deviations from this pattern. Focusing on *S. cerevisiae*, short genes accumulate very low levels of $S_2$ phosphorylation and are characterized by high $S_5$ phosphorylation levels along their length (Bataille et al. 2012, Kim et al. 2010, Mayer et al. 2010, Tietjen et al. 2010). However, given the small compact *S. cerevisiae* genome, deviations from the canonical CTD phosphorylation cycle can be partly attributed to the close proximity of transcribed units, such that RNAPII-CTD phosphorylation enrichment of adjacent genes can overlap giving the impression of non-

canonical profiles (Bataille et al. 2012). Similarly, in mammalian cells, genes vary in the levels of $S_2$ and $S_5$ phosphorylation along their length, and these correlate with transcriptional activity (Odawara et al. 2011). Furthermore, comparison of CTD phosphorylation profiles of *S. cerevisiae* and *S. pombe* reveal subtle differences regarding $S_2$ phosphorylation levels with levels being generally lower in the latter (Drogat and Hermand 2012). Therefore the extent to which the CTD modification cycle varies across genes and species remains unclear. However, given that RNAPII transcribes a wide range of different types of transcripts, many of which require different regulatory and processing programs, suggests possible functional implications for non-canonical CTD modification profiles (Chen and Wagner 2010). Thus, if deviations in the CTD modification cycle occur, they may underlie the CTD's ability to elicit gene-specific and species-specific transcription regulatory and processing requirements.

### 1.8.6 CTD phosphorylation and transcription inhibition

In addition to their roles during the transcription cycle, modification of the RNAPII-CTD can also play a role in transcription repression. For example, phosphorylation of the RNAPII-CTD by Cdk8 prior to PIC formation can repress transcription by preventing RNAPII from engaging with the promoter (Hengartner et al. 1998, Liu et al. 2004). Furthermore and indicative of loci-specific roles for RNAPII-CTD modifications, Cdc14 can dephosphorylate $S_5$ and $S_2$ residues at Y element-containing telomeres and rDNA intergenic spacer regions (Clemente-Blanco et al. 2011). Here, CTD dephosphorylation results in transcription repression, an important aspect of ensuring correct chromosomal segregation of these loci during mitosis.

## 1.9 A growing repertoire of CTD modifications and functions

As mentioned previously, the RNAPII-CTD can be modified in a number of different ways, some of which have clearly defined roles during the transcription of protein-coding genes. However, some CTD modifications play gene-specific functions. This section focuses on describing additional CTD modification with known roles in RNAPII biology and transcription.

### 1.9.1 $S_7$ phosphorylation is important for snRNA processing in mammalian cells

snRNAs are transcribed by RNAPII and their regulation and co-transcriptional processing varies significantly from mRNAs (Chen and Wagner 2010). snRNAs contain specialized promoter and termination sequences, and are not spliced, nor polyadenylated. snRNA 3'end formation is dependent on the Integrator complex, which is recruited via direct contacts with $S_2/S_7$ phosphorylated CTD heptapeptides (Baillat et al. 2005, Egloff et al. 2010). $S_7$ phosphorylation is critical for snRNA-specific 3' processing and mutating $S_7$ to alanine results in decreased levels of processed snRNA (Egloff et al. 2007). Additionally, $S_7$ phosphorylation also stimulates the recruitment of RPAP2 (mammalian Rtr1 homolog) to snRNAs, resulting in $S_5$ dephosphorylation and the generation of a predominantly $S_2/S_7$ phosphorylated CTD capable of recruiting the integrator complex (Egloff et al. 2010, Egloff et al. 2012).

### 1.9.2 CTD methylation is important for the regulation of snRNA expression

The mammalian CTD is methylated on R1810 (repeat 31 position 7) by the methyltransferase CARM1 (Sims et al. 2011). R1810 methylation plays a role in transcription regulation of snRNAs, and mutating R1810 to alanine leads to their up-regulation. This effect is likely

mediated by the ability of R1810 methylation to stimulate the recruitment of transcriptional co-activators like TDRD3 (Sikorsky et al. 2012). CARM1 is recruited to transcribed genes during initiation and can only methylate non-phosphorylated CTD peptides *in vitro,* suggesting that CTD methylation precedes CTD phosphorylation (Sims et al. 2011). However, CTD methylation and phosphorylation co-occur *in vivo,* indicating that CTD methylation is likely found along the length of genes and might have roles beyond transcription initiation.

### 1.9.3 $T_4$ phosphorylation plays species-specific functions

Work in a variety of species revealed diverse roles for $T_4$ phosphorylation. In *S. cerevisie*, CTDs containing $T_4$ to valine mutations are viable but display disregulation of genes for survival in media lacking phosphate or containing galactose. This is in part due to defects in INO80-mediated eviction of the histone variant H2A.Z at the promoter of these genes (Rosonina et al. 2014). In mouse cells, $T_4$ to alanine mutations generally result in increased RNAPII levels at the 3' end of genes, suggesting a role in transcription elongation (Hintermair et al. 2012). In contrast, in chicken cells $T_4$ mutants do not display elongation defects, and instead show defects in histone RNA 3' end processing (Hsin et al. 2011). The latter is likely due to $T_4$ phosphorylation's ability to regulate the recruitment of the histone specific 3' processing factor, SLBP.

### 1.9.4 Proline isomerization is important for expression of a subset of genes

Ess1 is a proline isomerase that targets the proline residues of the CTD. Proline isomerization is reversible and likely occurs along the length of all RNAPII transcribed genes (Wu et al. 2000). However, conditional mutations of Ess1 only affect the expression of a subset of mRNAs, and lead to termination defects on a subset of small non-coding transcripts (Singh et al. 2009).

Similarly, it was found that Ess1 is required for Nrd1-dependent termination, a pathway principally involved in early termination and degradation of short non-coding transcripts.

### 1.9.5 $Y_1$ phosphorylation has roles in CTD stability

$Y_1$ phosphorylation is detected in the cytoplasm of chicken cells, where it plays a role in Rpb1 stability (Hsin et al. 2014). Mutating all CTD $Y_1$ residues to phenylalanine results in the degradation of Rpb1 by the proteasome, leaving a RNAPII that is missing the CTD and unable to support life. However, a single $Y_1$ residue is sufficient to restore RNAPII-CTD stability but not long-term viability, an observation consistent with additional roles in transcription elongation.

## 1.10 Eukaryotic transcription takes place on a chromatin template

In eukaryotes, DNA is packaged into nucleosomes that are composed of 147 bp of DNA wrapped around an octamer of histone proteins: two copies of H2A, H2B, H3 and H4 (Luger et al. 1997). Nucleosomes are the basic units of chromatin, a DNA packing structure which is thought to be restrictive to processes requiring access to the underlying DNA, including transcription (Kulaeva et al. 2013, Li et al. 2007, Petesch and Lis 2012). Cells have evolved a variety of mechanisms to remodel chromatin, making it a dynamic structure. These include the alteration of histone properties by the addition of posttranslational modifications, the moving, sliding or eviction of histone proteins using ATP-dependent chromatin remodelers, and the loci-specific incorporation of histone variants, which bear distinct functional properties (Rando and Winston 2012).

Underscoring the tight link between transcription and chromatin are a number of histone modifications that specifically associate with transcribed regions and whose deposition is dependent on RNAPII (Smolle and Workman 2013). For example, histone acetylation is positively correlated with transcriptional activity and histone H2B ubiquitination (H2Bub), histone H3 lysine 4 methylation (H3K4me), histone H3 lysine 36 methylation (H3K36me), and histone H3 lysine 79 methylation (H3K79me) preferentially associate with actively transcribed loci. Briefly, H2Bub is found along the length of active genes and its deposition is dependent on a $S_5$ phosphorylated RNAPII-CTD. H2Bub limits higher order chromatin compaction and is important for nucleosome reassembly following transcription. H2Bub is a prerequisite for H3K4me and H3K79me. Set1 is the enzyme responsible for H3K4me and can mono, di, or tri methylate (me1, me2, and me3 respectively) histone H3 in a manner correlated with transcriptional activity. The H3K4me state plays roles in the recruitment of chromatin modifiers whose functions are important during transcription. Similarly, H3K79me is deposited by the enzyme Dot1, and is found along the length of genes. H3K79me has well-established roles in DNA damage response, cell cycle regulation, and maintenance of silent loci, however its role in transcription is less clear. H3K36me is deposited by Set2, which is recruited to transcribed regions via direct contacts with $S_2$ phosphorylated RNAPII-CTDs. Set2 deposits H3K36me3 towards the 3' end of genes and H3K36me levels are tightly associated with transcribed genes. H3K36me3 is important for recruitment of the histone deacetylase Rpd3 to transcribed regions, where it functions to remove activating acetyl marks following RNAPII transcription. As such, H3K36me and Rpd3 prevent transcription from cryptic promoters found along the length of genes.

## 1.11 Transcription plays a role in maintaining genome integrity

DNA damage can lead to changes or losses in genome-encoded information, and most severely it can lead to cell death. DNA damage can occur from insults by exogenous factors or through normal biological processes (Barnes et al. 1993). Given the importance of maintaining the integrity of coding sequences, cells have evolved specialized pathways to detect and repair damaged DNA on transcribed templates. Transcription-coupled repair (TCR) is tasked with maintaining genome integrity at biologically relevant loci, resulting in decreased levels of damaged DNA at highly transcribed genes compared to weakly transcribed or untranscribed regions of the genome (Svejstrup 2003, Szilard et al. 2010). In this process, damaged DNA result in prolonged RNAPII pausing, which signals the recruitment of nucleotide excision repair factors. If the damage cannot be repaired, RNAPII is removed via ubiquitination- and proteasome-dependent degradation, allowing other repair pathways to attempt to repair the damaged site. In addition, RNAPII and its C-terminus have been further implicated in the maintenance of DNA integrity. First, RNAPII-CTD truncation mutants show increased sensitivity to drugs like the DNA replication inhibitor, hydroxyurea (Wong and Ingles 2001). Second, critically short RNAPII-CTD truncation mutants spontaneously revert to RNAPIIs with longer CTDs, indicating higher rates of genome rearrangements (Nonet and Young 1989). Third, RNAPII-CTD truncation mutants show decreased stability of minichromosomes *in vivo* (Gauthier et al. 2002). Fourth, RNAPII is the enzyme responsible for transcribing genome-encoded retrotransposons, the mobility of which can lead to DNA alterations (Lesage and Todeschini 2005). Fifth, RNAPII's role in synthesizing RNA makes it an important contributor

to the formation of DNA:RNA hybrids, structures that contribute to increased levels of DNA damage (Aguilera and Garcia-Muse 2012).

### 1.11.1 Retrotransposons are transcribed by RNAPII

Retrotransposons are transcribed by RNAPII, a key intermediate step in their multiplication within the host genome (Boeke et al. 1985). The retrotransposon life cycle culminates when integrase inserts retrotransposon cDNA into a new genomic location, an event that can have grave consequences to genome integrity. For example, insertion within a coding sequence can mutate or interrupt the underlying gene, while insertion near a coding sequence can alter the transcriptional regulation of adjacent genes (Lesage and Todeschini 2005). Therefore, cells have evolved mechanisms to maintain low retrotransposition rates in part by limiting their transcription. Nonetheless, retrotransposons take advantage of a number of host transcription factors to drive their own expression including Mcm1, Tea1, Rap1, Gcn4, Tec1, and Ste12 (Errede 1993, Gray and Fassler 1993, Gray and Fassler 1996, Laloux et al. 1990, Madison et al. 1998, Morillon et al. 2000, Morillon et al. 2002). Chapter 3 of this thesis describes a previously unreported role for the RNAPII-CTD in limiting retrotransposon mobility by inhibiting their gene expression.

### 1.11.2 DNA:RNA hybrids contribute to genome instability

Given their complementarity, DNA and RNA can base pair forming a DNA:RNA hybrid. DNA:RNA hybrids form naturally. During transcription, DNA:RNA hybrids form at the active center of RNAP and during DNA replication DNA:RNA hybrids are observed during RNA priming of the lagging strand (Chargaff 1976, Westover et al. 2004). In addition DNA:RNA

hybrids function in maintaining telomere length, regulating antisense transcription, and transcription termination (Balk et al. 2013, Skourti-Stathaki et al. 2011, Sun et al. 2013). However, formation of DNA:RNA hybrids can sometimes result in a displaced DNA strand that is more prone to DNA damage (Aguilera and Garcia-Muse 2012, Thomas et al. 1976). Thus, prolonged or increased levels of DNA:RNA hybrids have been implicated in genome instability. This is supported by the observation that some chromosome instability phenotypes can be suppressed by overexpression of RNAse H, the enzyme responsible for removing DNA:RNA hybrids (Stirling et al. 2012). Nonetheless, very little is known about the formation, regulation, and distribution of DNA:RNA hybrids. Chapter 5 of this thesis describes the first ever genome-wide map of DNA:RNA hybrids in *S. cerevisiae* and reveals how factors with prominent roles in RNA processing affect their occupancy.

### 1.12 Research scope

Since its discovery, the CTD has emerged as a central player in transcriptional regulation and co-transcriptional processing. Recent efforts have expanded our understanding of the CTD modification status and have suggested that some modifications and their distributions vary in a gene-specific manner (Bataille et al. 2012, Kim et al. 2010, Mayer et al. 2010, Tietjen et al. 2010). Furthermore, given that RNAPII synthesizes many types of RNA, it remains to be determined if and how the RNAPII-CTD coordinates co-transcriptional activities in a gene-specific manner. This thesis focuses on understanding how altering CTD length and CTD phosphorylation influences transcriptional activity. Furthermore, this thesis investigates the link between RNAPII function and DNA damage. More specifically, the work described in Chapter

2, focuses on the effect of CTD-length on the expression of protein-coding genes and constitutes the most comprehensive evidence to-date that individual protein-coding genes have different minimal CTD length requirements for normal expression. Work in this chapter also reveals unexpected connections to *CDK8/SRB10,* a previously characterized RNAPII-CTD truncation mutant suppressor. Chapter 3 focuses on the role of the RNAPII-CTD in the regulation of retrotransposon gene expression. This function has significant implications for the CTD in maintaining genome stability. Furthermore, this role was also dependent on CTD-length and *CDK8*. Chapter 4 encompasses work aimed at understanding the role and regulation of the RNAPII phosphatase, Fcp1. *FCP1* and RNAPII-CTD truncation mutants resulted in similar gene expression and genetic interaction profiles. In addition, *FCP1* and RNAPII-CTD truncation mutant phenotypes were suppressed by loss of *CDK8*. However, differences were also observed and these suggested roles additional roles for Fcp1. Chapter 5 focuses on the link between transcription and DNA:RNA hybrid formation. This chapter describes the first genome-wide map of DNA:RNA hybrids in *S. cerevisiae*, and provides evidence that not all hybrids are formed equally. In particular, mutants involved in DNA:RNA hybrid biology resulted in distinct changes to the hybrid landscape. Taken together, this thesis explores the biology of the RNAPII-CTD, expands the role of the CTD to include a direct effect on genome integrity, and explores mechanisms of how transcription can lead to genome instability.

# Chapter 2: High-throughput Genetic and Gene Expression Analysis of the RNAPII-CTD Reveals Unexpected Connections to *SRB10/CDK8*

## 2.1 Introduction[1]

The largest subunit of RNA polymerase II, Rpb1, has a unique C-terminal domain (CTD) composed of the repeated sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser ($Y_1\,S_2\,P_3\,T_4\,S_5\,P_6\,S_7$) (Allison et al. 1985, Corden et al. 1985). Although the CTD is highly conserved across species, the number of repeats varies in a manner resembling genomic complexity, with 25/26 repeats in *Saccharomyces cerevisiae* and 52 in humans (Heidemann et al. 2012). Deletion of the entire CTD is lethal in budding yeast, while strains carrying 9-13 repeats are viable but display conditional phenotypes (Allison et al. 1988, Nonet et al. 1987). While not required to support basal transcription *in vitro*, the CTD is critical for the response to activator signals *in vivo* (Scafe et al. 1990, Zehring et al. 1988). For example, CTD truncation mutants exhibit reduced activation of *INO1* and *GAL10* upon switching to inducing conditions (Scafe et al. 1990).

The CTD is a scaffold for the recruitment of RNA processing and chromatin-remodeling factors, a function linked to its differential phosphorylation at specific residues of the heptapeptide repeat

---

[1] A version of this chapter is published in the Public Library of Science Genetics. 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. doi:10.1371/journal.pgen.1003758.

(Heidemann et al. 2012). Transcription begins with the recruitment of RNAPII with an unphosphorylated CTD to promoters, where it interacts with components of the transcription pre-initiation complex (PIC) (Kobor et al. 2002, Lu et al. 1991). Following, it is phosphorylated at $S_5$ and $S_7$ by the general transcription factor TFIIH, facilitating recruitment of capping enzymes and release of RNAPII from promoter-bound elements (Akhtar et al. 2009, Kim et al. 2009, Max et al. 2007, McCracken et al. 1997a). Elongation is characterized by phosphorylation of $S_2$ by Ctk1 and $Y_1$ and $T_4$ by yet unidentified kinases (Cho et al. 2001, Mayer et al. 2012). $S_2$ and $Y_1$ phosphorylation play a role in the temporal recruitment of elongation and termination factors (Mayer et al. 2012). Subsequently, termination entails removal of all phosphorylation marks by Fcp1 and Ssu72 to regenerate an initiation competent RNAPII molecule (Cho et al. 1999, Kobor et al. 1999, Zhang et al. 2012a).

While early work aimed at understanding CTD function uncovered a set of *SRB* (Suppressor of RNA Polymerase B) genes, a comprehensive genetic network governing CTD function has yet to be fully elucidated (Nonet and Young 1989). Of the identified *SRB* genes many encode members of a large multisubunit complex known as Mediator (Thompson et al. 1993). Mediator was first identified *in vitro* as a cellular fraction that stimulates RNAPII transcription, and is now known to not only physically interact with the CTD, but also to be important for the response to up-stream regulatory signals (Robinson et al. 2012). Although primarily associated with RNAPII gene promoters, Mediator also resides at open reading frames (ORFs) (Andrau et al. 2006, Zhu et al. 2006). Furthermore, Mediator is organized into four functionally distinct submodules: head, middle, tail and Cdk8 modules (Guglielmi et al. 2004). The head module interacts with the CTD while the tail and middle modules interact with gene-specific and general transcription factors

(Bhoite et al. 2001, Han et al. 1999). The Cdk8 kinase module likely associates transiently with the core Mediator complex and has roles in both transcriptional activation and repression (Galbraith et al. 2010, Tsai et al. 2013). This dual activity is in part mediated by Cdk8's ability to phosphorylate multiple regulatory components of the transcription machinery. These include several transcription factors as well as factors more generally required for transcription such as the CTD itself (Galbraith et al. 2010, Hengartner et al. 1998, Hirst et al. 1999, Liu et al. 2004). While the mechanistic role of some of these phosphorylation events is unclear, CTD phosphorylation by Cdk8 prior to promoter association inhibits RNAPII recruitment and transcription initiation *in vitro* (Hengartner et al. 1998). In contrast, CTD phosphorylation by Cdk8 and Kin28 following promoter association promotes RNAPII release from the PIC and thus stimulates transcription activation (Liu et al. 2004).

The work here highlighted the functional circuitry between the RNAPII-CTD and Mediator in the regulation of cellular homeostasis, gene expression, and the transcription factor Rpn4. Our data uncovered a length-dependent requirement of the CTD for genetic interactions and mRNA levels of genes expressed under normal growth conditions. Truncating the CTD primarily resulted in increased expression and RNAPII association at a subset of genes, in part mediated by changes to transcription initiation. These genes had preferential association of Cdk8 at their promoters and were regulated by the transcription factor Rpn4. The expression and RNAPII binding defects of the majority of this subset of genes were suppressed by deleting *SRB10/CDK8*, suggesting that in CTD truncation mutants, Cdk8 functions to enhance transcription and RNAPII association at a subset of genes. Conversely, our data also revealed that deletion of *CDK8* suppressed the activation defects of CTD truncation mutants at the *INO1*

locus thus indicating that Cdk8 also functions to repress transcription and RNAPII association in CTD truncation mutants.

## 2.2 Materials and methods

### 2.2.1 Yeast strains, plasmids, and growth conditions

Strains and plasmids are listed in **Table 2.1** and **Table 2.2** respectively. Partial, complete gene deletions or integration of a 3XFLAG tag was achieved via the one-step gene replacement method (Longtine et al. 1998). CTD truncations were created at the *RPB1* locus by addition of a TAG stop codon followed by a NAT resistance marker and confirmed by sequencing. As a control for E-MAP and gene expression analysis we used *RPB1-CTDWT*. This strain contained a NAT resistance marker following the endogenous stop codon. pRS314 [RPN4] and pRS314 [rpn4 S214/220A] were obtained from Dr. Youming Xie (Wayne State University School of Medicine). Reporter plasmids were generated by cloning 450bp of the desired promoter into the Sal1 BamH1 sites of pLG669-Z (Guarente and Ptashne 1981).

**Table 2.1 Strains used in this study**

| Name | Genotype | Background |
|---|---|---|
| MKY654 | *Matα rpb1-CTD11-nat his3Δ1 leu2Δ0 LYS2+ met15Δ0 ura3Δ0 Δcan1::MATaPr-HIS3 Δlyp1::MATαPr-LEU2* | *BY4742* |
| MKY655 | *Matα rpb1-CTD12-nat his3Δ1 leu2Δ0 LYS2+ met15Δ0 ura3Δ0 Δcan1::MATaPr-HIS3 Δlyp1::MATαPr-LEU2* | *BY4742* |
| MKY656 | *Matα rpb1-CTD13-nat his3Δ1 leu2Δ0 LYS2+ met15Δ0 ura3Δ0 Δcan1::MATaPr-HIS3 Δlyp1::MATαPr-LEU2* | *BY4742* |
| MKY657 | *Matα rpb1-CTD20-nat his3Δ1 leu2Δ0 LYS2+ met15Δ0 ura3Δ0 Δcan1::MATaPr-HIS3 Δlyp1::MATαPr-LEU2* | *BY4742* |
| MKY658 | *Matα rpb1-CTDWT-nat his3Δ1 leu2Δ0 LYS2+ met15Δ0 ura3Δ0 Δcan1::MATaPr-HIS3 Δlyp1::MATαPr-LEU2* | *BY4742* |
| MKY1507 | *Matα his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1508 | *Matα rpb1-CTD11-nat his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |

| Name | Genotype | Background |
|---|---|---|
| MKY1509 | *Matα rpb1-CTD12-nat his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1510 | *Matα rpb1-CTD13-nat his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1511 | *Matα rpb1-CTD20-nat his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1512 | *Matα rpb1-CTDWT-nat his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1513 | *Matα rpb1-CTD11-nat cdk8::kan his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1514 | *Matα rpb1-CTD12-nat cdk8::kan his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1515 | *Matα cdk8::kan his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1516 | *Matα CDK8-Flag::kan his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1524 | *MatA RPN4-HA-kan ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* | *W303* |
| MKY1525 | *MatA RPN4-HA-kan rpb1-CTD11::hyg ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* | *W303* |
| MKY1526 | *Matα RPN4-HA-kan cdk8::his ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* | *W303* |
| MKY1527 | *Matα RPN4-HA-kan rpb1-CTD11::hyg cdk8::his ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* | *W303* |
| MKY1561 | *Matα CDK8-Flag-kan CTD11-nat his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1562 | *Matα SUA7-Flag-kan his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1563 | *Matα SUA7-Flag-kan CTD11-nat his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1564 | *Matα CET1-Flag-kan his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1565 | *Matα CET1-Flag-kan CTD11-nat his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1566 | *Matα ELF1-Flag-kan his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1567 | *Matα ELF1-Flag-kan CTD11-nat his3Δ1 leu2Δ0 ura3Δ0* | *BY4742* |
| MKY1568 | *Matα rpn4::nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314*[RPN4]* | *W303* |
| MKY1569 | *Matα rpb1-CTD11-hyg rpn4::nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314*[RPN4]* | *W303* |
| MKY1571 | *Matα cdk8::his rpn4::nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314*[RPN4]* | *W303* |
| MKY1572 | *Matα rpb1-CTD11-hyg cdk8::his rpn4::aat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314*[RPN4]* | *W303* |
| MKY1573 | *Matα rpn4::nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314*[rpn4 S214/220A]* | *W303* |
| MKY1574 | *Matα rpb1-CTD11-hyg rpn4::nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314*[rpn4 S214/220A]* | *W303* |
| MKY1575 | *Matα cdk8::his rpn4::nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314*[rpn4 S214/220A]* | *W303* |
| MKY1576 | *Matα rpb1-CTD11-hyg cdk8::his rpn4::Nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314*[rpn4 S214/220A]* | *W303* |
| MKY1577 | *Matα ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314 | *W303* |
| MKY1578 | *Matα rpb1-CTD11-hyg ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314 | *W303* |
| MKY1579 | *Matα cdk8::his ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314 | *W303* |
| MKY1580 | *Matα rpn4::nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314 | *W303* |
| MKY1581 | *Matα rpb1-CTD11-hyg cdk8::his ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314 | *W303* |
| MKY1582 | *Matα rpb1-CTD11-hyg rpn4::nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314 | *W303* |
| MKY1583 | *Matα cdk8::his rpn4::nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314 | *W303* |
| MKY1584 | *Matα rpb1-CTD11-hyg cdk8::his rpn4::nat ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* pRS314 | *W303* |
| MKY1585 | *Matα gcn4::kan ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* | *W303* |

| Name | Genotype | Background |
|---|---|---|
| MKY1586 | *Matα rpb1-CTD11-hyg gcn4::kan ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* | *W303* |
| MKY1587 | *MatA cdk8::his gcn4::kan ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* | *W303* |
| MKY1588 | *Matα rpb1-CTD11-hyg cdk8::his gcn4::kan ade2-1 can1-100 his3-11 leu2-3,112 trp1-1 ura3-1 LYS2* | *W303* |

**Table 2.2 Plasmids used in this study**

| Plasmid | Relevant Genotype | Backbone | Source or Reference |
|---|---|---|---|
| pMK540 | YOR052C 450bp promotor LacZ fusion | PGL669-z | this study |
| pMK541 | *PDR5* (YOR153W) 450bp promotor LacZ fusion | PGL669-z | this study |
| pMK542 | YKL145W 450bp promotor LacZ fusion | PGL669-z | this study |
| pMK543 | YKL096W 450bp promotor LacZ fusion | PGL669-z | this study |
| pMK544 | YDR033W promoter LacZ fusion | PGL669-z | this study |
| pMK545 | CUP1p RPN4 | pRS314 | Donghong *et al* 2007 |
| pMK546 | CUP1p RPN4(S214/220A) | pRS314 | Donghong *et al* 2007 |

### 2.2.2 Epistasis miniarray profiling (E-MAP)

E-MAP screens were performed and normalized as described previously (Collins et al. 2010). RNAPII-CTD truncation mutants were crossed, using a Singer robot, to a library of 1536 mutants (Collins et al. 2010) covering a number of categories, including RNA processing, kinases/phosphatases and chromatin biology. Mutants contain either complete deletions or decrease abundance by mRNA perturbations (DAmP) alleles of the indicated genes. Diploid selection, sporulation, haploid selection, and double mutant selection steps were performed by replicate plating on the appropriate selective media. All strains were screened in triplicate and for each replicate double mutant colony sizes were determined from three technical replicates. Colony size was used to determine a quantitative S-score, which is a modified T-test that compares the observed double mutant growth rate to an expected growth rate based on the average colony size across an entire plate.

### 2.2.3 Microarrays experiments and analysis

Microarrays were performed in duplicate as previously described (Lenstra et al. 2011, van Wageningen et al. 2010). Cultures were grown with a 24-well plate incubator/reader. Spiked-in controls were used to determine global changes in mRNA levels. As no such changes were detected, the expression profiles were normalized to total mRNA levels, a more reproducible measure. Differentially expressed genes were determined by p value < 0.01 and fold change > 1.7 compared to wild type. Suppressed genes were determined as those having fold changes < 1.1 in the *rpb1-CTD11 cdk8Δ* mutant. The Yeast Promoter Atlas database was used for transcription factor enrichment by performing a Hypergeometric test with Bonferroni correction (p value 0.05) (Chang et al. 2011). "Biological Process" ontology annotated in the Bioconductor package org.Sc.sgd.db was used for Gene Ontology enrichment using the conditional Hypergeometric test (adjusted p value <0.05) described (Carlson et al, Falcon and Gentleman 2007).

### 2.2.4 Chromatin immunoprecipitation (ChIP)

Yeast cultures were grown in media containing 200 μM of inositol (uninduced) and switched to media lacking inositol for 4 hrs (induced) (Brickner et al. 2007). Cross-linking was done with 1% formaldehyde for 20 min. Chromatin was prepared as described previously (Schulze et al. 2009). Five μl of anti-Rpb3 antibody (Neoclone) was used. Crosslinking reversal and DNA purification were followed by qPCR analysis of the immunoprecipitated and input DNA. cDNA was analyzed using a Rotor-Gene 600 (Corbett Research) and PerfeCTa SYBR Green FastMix (Quanta Biosciences). Samples were analyzed from three independent DNA purifications and

normalized to an intragenic region of Chromosome V (Keogh and Buratowski 2004). Primers are listed **Table 2.3**.

**Table 2.3 Primers used in this study**

| Primer name | Forward Sequence | Reverse Sequence | Source or Reference |
|---|---|---|---|
| *INO1* Promoter | TTTGGCTTGTTCTGTTGTCG | CCTTGTACGTGCACTTGTCG | This study |
| *INO1* 5' | CGAGCTGCTCACCAAGTACA | AGCCATTGTTGCCACCTAAC | This study |
| *INO1* mid | GATCTGCAACAACGCTTGAA | TCTGGATATCGCGTCTGATG | This study |
| *INO1* 3' | ATGCTGGCAAATTCGAGAAC | TCGTTTTGAGAAGGCAATCC | This study |
| Chr V | GGCTGTCAGAATATGGGGCCGTAGTA | CACCCCGAAGCTGCTTTCACAATAC | Keogh and Buratowski 2004 |
| YKL145W | GGTGAAGGTGCTCGTATGGT | GGGTCAAACCCGTCTAACTG | This study |
| YMR276W | AACAACGCAGGGACTAATGC | AGACGCTAGCAATGCAGGAT | This study |
| YNL241C | CGAACTGGTCATCAGAGTGC | CACCTCGTAAGCCTCTGGAA | This study |
| YIR034C | CTGCCGGGCCTAAATTATCT | ACGAGCGCAATGTCTATCG | This study |
| YML116W | CACCGGTTCACGAGACATAC | ACCCATACCGAGACACAAGG | This study |

### 2.2.5 Genome-wide ChIP-on-chip

ChIP-on-chip cultures were grown overnight in YPD, diluted to 0.15 OD600 and grown to 0.5-0.6 OD600 units. Cross-linking and chromatin isolation were performed as above. Five µl of anti-Rpb3 (Neoclone), 4.2 µl of anti-FLAG (Sigma), or 4 µl of anti-H3K36me3 antibody (Abcam ab9050) were coupled to 60 µl of protein A magnetic beads (Invitrogen). DNA was amplified using a double T7 RNA polymerase method, labeled, hybridized as previously described (Schulze et al. 2009). Samples were normalized as described previously using the rMAT software (Droit et al. 2010). Relative occupancy scores were calculated for all probes using a 300 bp sliding window. Rpb3 and H3K36me3 experiments were normalized to input while Flag-tagged factors were normalized to untagged controls. Samples were carried out in duplicate, quantile normalized and averaged data was used for calculating average enrichment

scores (MAT scores). For ORFs, we averaged probes whose start sites fell within the ORF start and end positions, and for promoters we averaged probes mapping to 500bp upstream of the ORFs. Enriched features had at least 50% of the probes contained in the feature above the threshold of 1.5. Enriched features were identified for each replicate and the overlap was reported as the significantly enriched set.

### 2.2.6 ChIP-on-chip visualization

CHROMATRA plots were generated as described previously (Hentrich et al. 2012). In detail, transcripts were aligned by their TSSs and relative occupancy scores for each transcript were binned into segments of 150 bp and averaged. Transcripts were grouped into five classes according to their transcriptional frequency as per Holstege *et al* 1998 and sorted by their length. Average gene profiles were generated by averaging all probes whose start sites map to a gene of interest. For averaging, ORFs and corresponding probes were split into 40 bins while UTRs and their probes were split into 20 bins.

### 2.2.7 Reporter assays

Reporter plasmids were generated as described above, transformed into wild type and *rpb1-CTD11* mutants, and assayed as previously described (Guarente 1983). Briefly, overnight cultures were diluted to OD600 0.2 and grown to OD600 0.5-1. Cells were collected and resuspended in Z buffer. Twenty one µl of chloroform and 14µl of 0.1%SDS were added and samples were incubated for 5 min at 28 °C. One hundred and sixty µl of 4mg/ml 2-Nitrophenyl-β-D-galactopyranoside (Sigma N1127) was added, mixed by inversion and placed back at 28 °C. Reactions were allowed to proceed until the solution turned light yellow and stopped with 400µl

of 1M $Na_2CO_3$ solution. OD420 of supernatant was measured and along with time of reaction and OD600 units of cells collected were used to determine Miller Units. Measurements were obtained from three independent cultures and error bars represent standard deviations.

### 2.2.8 Growth assays

Overnight cultures grown on YPD or –TRP media were diluted to 0.5 OD600, 10-fold serially diluted and spotted onto YPD or –TRP plates with or without the indicated amounts of hydroxyurea (Sigma), formamide (Sigma), or on plates lacking inositol. Plates were incubated at the indicated temperatures for 2-4 days.

### 2.2.9 Protein blotting

Whole cell extracts were prepared from logarithmic growing cells by glass bead lysis in the presence of trichloroacetic acid. Immunoblotting was carried out with 3E10, 3E8, 4E12, 8WG16 (Millipore), YN-18 (Santa Cruz), Rpb3 (Neoclone), HA-Peroxidase (Roche) and Pgk1 (Molecular Probes) antibodies (Chapman et al. 2007). Immunoblots were scanned with the Odyssey Infrared Imaging System (Licor) or visualized with SuperSignal enhanced chemiluminescence (Pierce Chemical).

### 2.2.10 Reverse transcriptase PCR (RT-PCR)

RNA was extracted and purified using the Qiagen RNeasy Mini Kit. cDNA was generated using the Qiagen QuantiTect Reverse Transcription Kit. cDNA was analyzed by qPCR as described above. *INO1* mRNA levels were normalized to *ACT1* mRNA (Scafe et al. 1990). Samples were analyzed in triplicate from three independent RNA preparations. Primers are listed in **Table 2.3**

### 2.2.11 Protein stability assay

Overnight cultures were diluted to 0.3 OD600 and grown to 1.0 OD600. Ten OD600 units were collected to constitute time 0 and a final concentration of 100ug/ml of cycloheximide (Sigma) was added to the remaining culture. Ten OD600 units were collected at the indicated time points. Proteins were extracted using trichloroacetic acid.

## 2.3 Results

### 2.3.1 The RNAPII-CTD was linked to an extensive genetic interaction network

To broadly determine the requirement of CTD length for cellular function, we used Epistasis Mini Array Profiling (E-MAP) to generate genetic interaction profiles of CTD truncation mutants containing 11, 12, 13 or 20 heptapeptide repeats (*rpb1-CTD11*, *rpb1-CTD12*, *rpb1-CTD13* and *rpb1-CTD20* respectively) against a library of 1532 different mutants involved principally in aspects of chromatin biology and RNA processing (Collins et al. 2010). CTD truncations were created at the *RPB1* locus by addition of a TAG stop codon followed by a NAT resistance marker. As a control for the genetic integration strategy we also generated *RPB1-CTDWT,* which contained a NAT resistance marker following the endogenous stop codon. While the minimal CTD length for viability is 8 repeats, we focused on strains starting at 11 repeats as mutants bearing shorter CTDs were significantly unstable in our hands, consistent with previous findings (West and Corden 1995). Overall our data revealed a greater number of significant genetic interactions as the CTD was progressively shortened, an effect consistent with increasingly disrupted function (**Figure 2.1A**). Furthermore, hierarchical clustering based on

Spearman's rho correlation delineated two major clusters, the first including *rpb1-CTD11*, *rpb1-CTD12* and *rpb1-CTD13* and the second consisting of *rpb1-CTD20* and *RPB1-CTDWT* (**Figure 2.1B**), however, individual genetic interactions revealed more nuanced CTD length-dependent genetic interaction patterns (**Figure 2.2**). For example, aggravating interactions were observed with strains lacking *ASF1, RTT109* and *DST1* when the CTD was truncated to 13 repeats or shorter, while truncation to 11 repeats was required for aggravating interactions with *SET2, RTR1* and *SUB1*. Collectively, this data revealed significant and specific functional alterations to the CTD as a result of shortening its length and suggested that individual pathways required different CTD lengths for normal function. Finally, given that we identified significant genetic interactions with genes involved in a variety of processes, we compared the E-MAP profile of our shortest CTD truncation with all previously generated profiles to determine which pathways were principally affected by truncating the CTD. This analysis revealed that four of the ten most correlated profiles belonged to loss of function alleles of genes encoding subunits of TFIIH and Mediator (*RAD3, MED8, MED31* and *MED20*) suggesting that shortening the CTD results in genetic interaction patterns most similar to mutants affecting transcription initiation (**Figure 2.1C**).

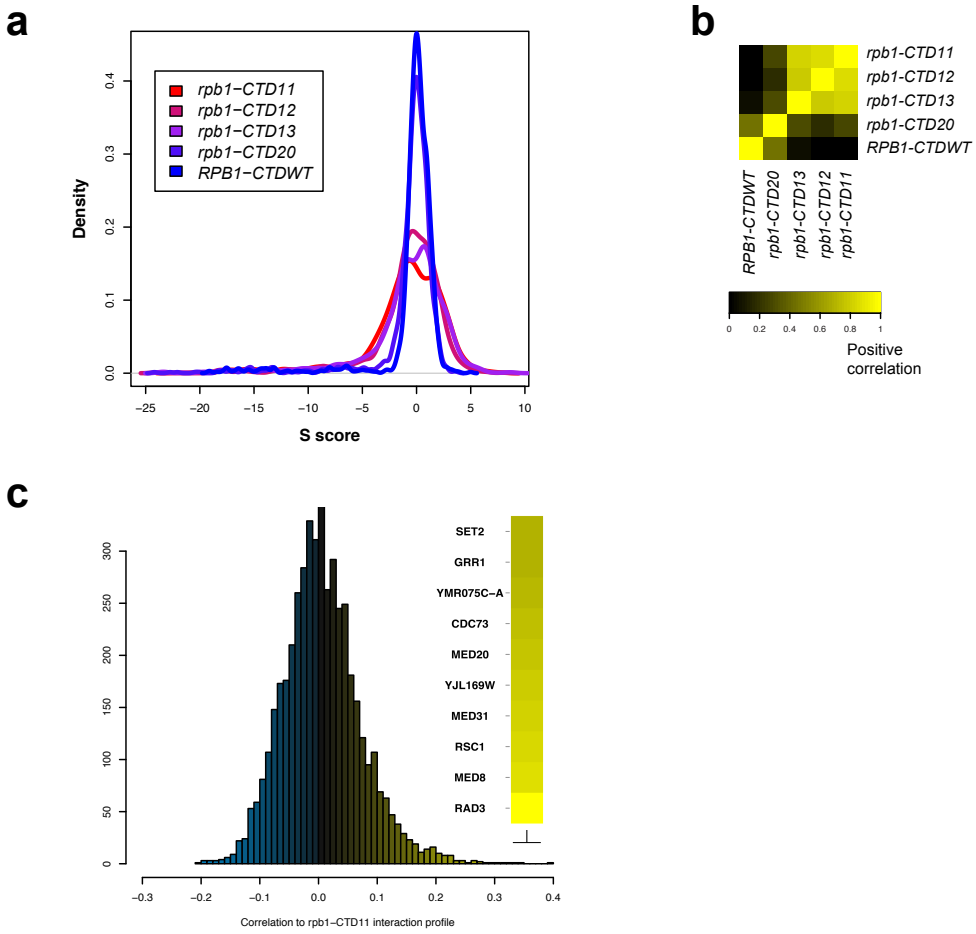



**Figure 2.1 E-MAP uncovered CTD length-dependent genetic interactions with genes involved in transcription.**
The genetic interaction profile of strains containing 11, 12, 13 or 20 heptapeptide repeats (*rpb1-CTD11*, *rpb1-CTD12*, *rpb1-CTD13* and *rpb1-CTD20*) against a library of 1532 different mutants involved principally in aspects of chromatin biology and RNA processing. CTD truncations were created at the endogenous *RPB1* locus by addition of a TAG stop codon followed by a NAT resistance marker. *RPB1-CTDWT* served as a control and contained a NAT resistance marker following the endogenous stop codon (A) Distribution of S scores for CTD truncation mutants revealed an increase in the number of significant genetic interactions as a result of truncating the CTD. The S score is a modified T-statistic measure, which captures both the confidence and strength of the genetic interaction. Scores greater than 2.0 or less than -2.5 are considered significant. (B) Spearman rho correlation of CTD truncated mutants identified two distinct groups (C) Distribution of Pearson's correlation scores derived from comparing the *rpb1-CTD11* interaction profile to all previously assayed strains.

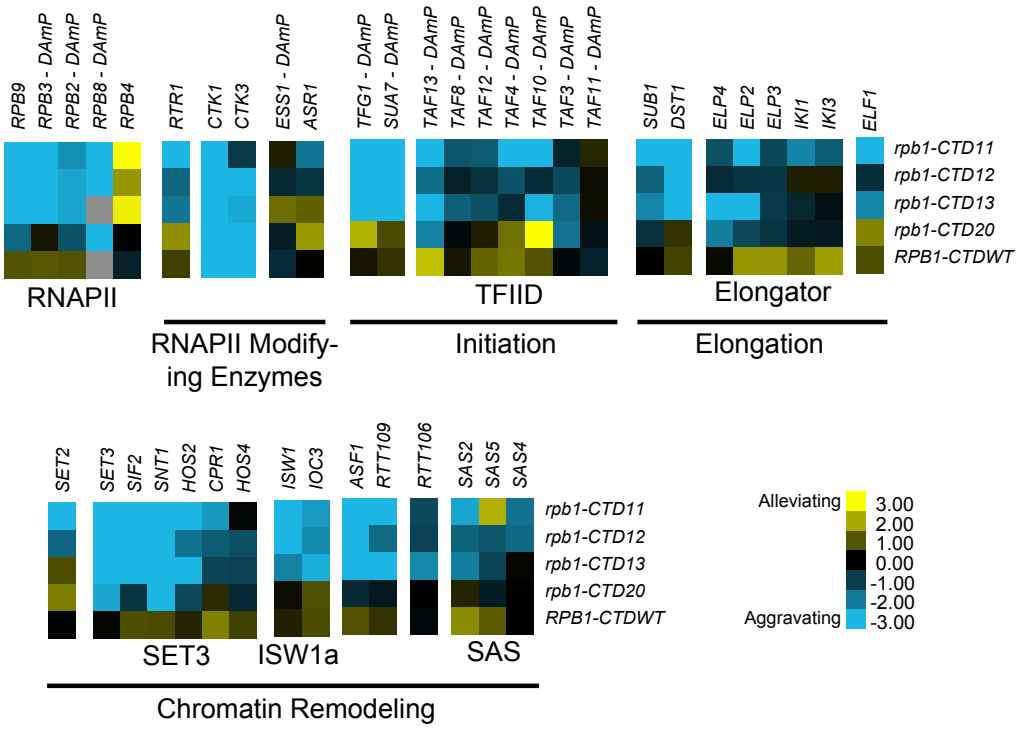
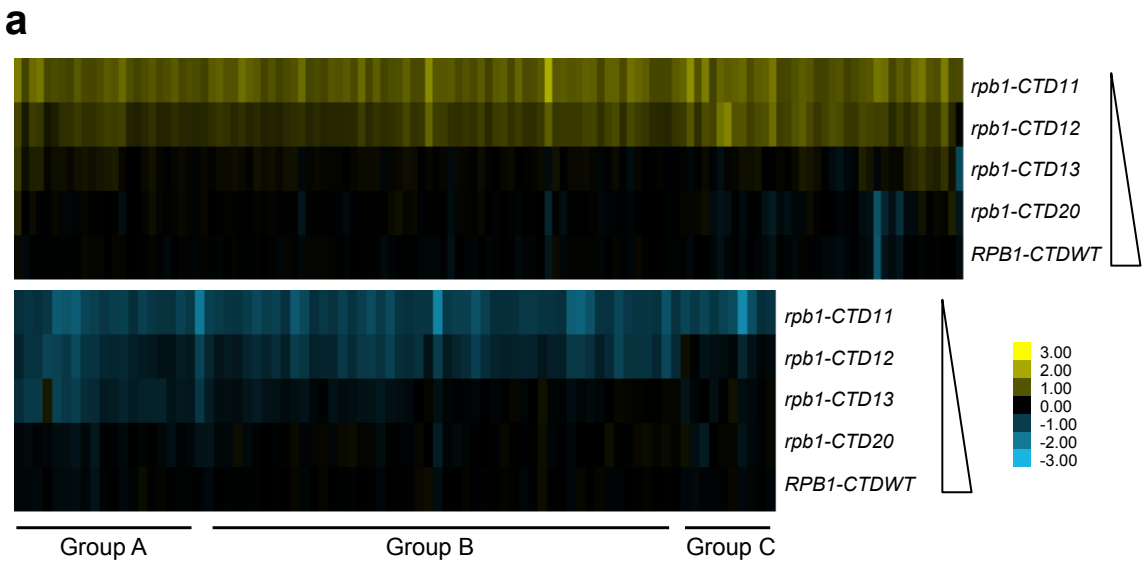


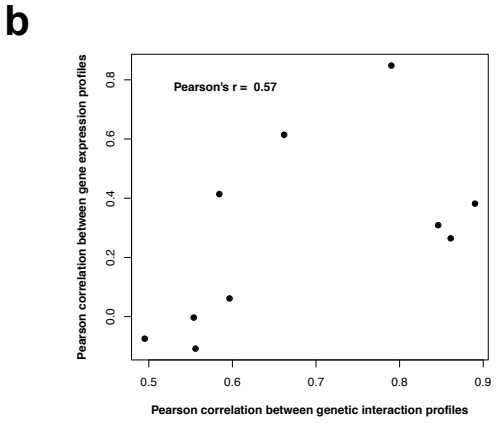

**Figure 2.2 Sample genetic interaction network of RNAPII-CTD truncation mutants revealed CTD length-dependent genetic interactions.**
Subsets of genetic interaction profiles depicting genes involved in transcription and how they interacted with the CTD as it was progressively shortened. Blue and yellow represent aggravating and alleviating genetic interactions respectively. Gray boxes represent missing values.

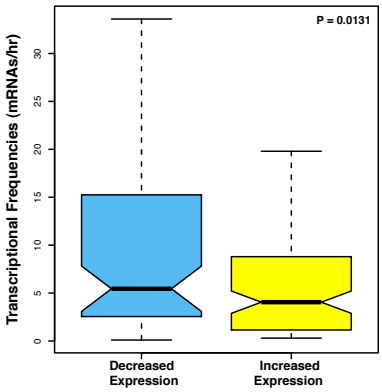
### 2.3.2 CTD serial truncations led to progressive changes in transcription

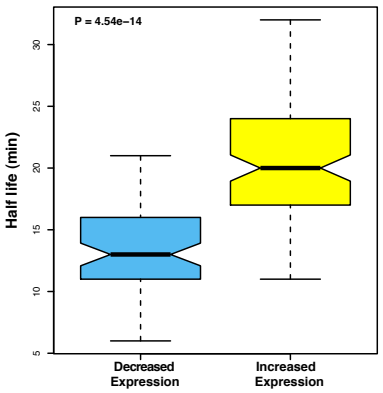
Although the CTD plays a major role in the response to activator signals *in vivo*, its general involvement in transcription is less well defined. To investigate this important aspect, we generated gene expression profiles of CTD truncation mutants in normal growth conditions (Complete dataset can be found in array-express, code E-MTAB-1431 https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1431). Similar to the E-MAP data, the expression data revealed a length-dependent requirement for CTD function, with the severity

and number of transcriptional changes increasing as the CTD was progressively shortened (comparison of E-MAP vs expression profiles Pearson's rho 0.57) (**Figure 2.3A and B).** This gradient effect was clearly visible in the group of genes whose transcript levels decreased upon truncation of the CTD (**Figure 2.3A** groups A, B and C constitute genes requiring greater than 13, 12, and 11 repeats for normal transcription respectively), and thus provided strong evidence of a gene-specific CTD length requirement for normal transcription. Surprisingly, given the central role of the CTD in RNAPII function, our microarray data identified only 127 genes with significant increases in mRNA levels and 80 genes with significant decreases (p value < 0.01 and fold change > 1.7 compared to wild type), in strains carrying the shortest CTD allele, *rpb1-CTD11*. Functional characterization of the set of genes with increased and decreased mRNA levels suggested that the transcriptional alterations were not affecting a random group of genes. Instead, using previously published transcription frequency data, we found that the genes with decreased mRNA levels tended to be highly transcribed with short mRNA half-lives, while the genes with increased mRNA levels were mostly lowly transcribed with long mRNA half-lives (**Figure 2.3C and D**) (Holstege et al. 1998). In addition, these genes belonged to different functional gene ontology (GO) categories. The genes with increased mRNA levels were enriched for proteasome and proteasome-associated catabolism processes while the genes with decreased levels were enriched for iron homeostasis, purine metabolism and pheromone response. Finally, these genes were differentially regulated by transcription factors (**Figure 2.3E**). The genes whose expression levels decreased were principally bound by Ste12, while those with increased expression were bound by Ume6, Met31, Gcn4 and most significantly by Rpn4 which bound 46% of these genes (p value 1.46E-41).


a
rpb1-CTD11
rpb1-CTD12
rpb1-CTD13
rpb1-CTD20
RPB1-CTDWT
rpb1-CTD11
rpb1-CTD12
rpb1-CTD13
rpb1-CTD20
RPB1-CTDWT
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
Group A
Group B
Group C


b
Pearson's r = 0.57
Pearson correlation between gene expression profiles
Pearson correlation between genetic interaction profiles


c
P = 0.0131
Transcriptional Frequencies (mRNAs/hr)
Decreased Expression
Increased Expression


d
P = 4.54e−14
Half life (min)
Decreased Expression
Increased Expression

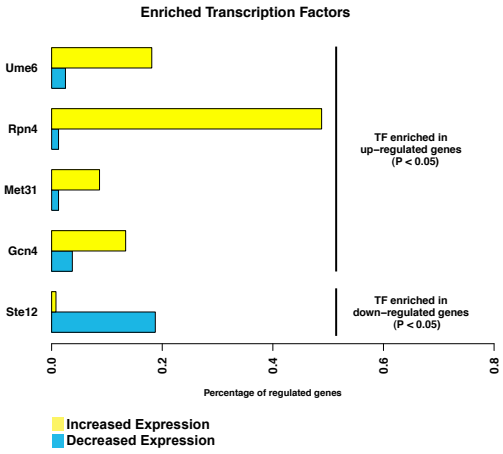
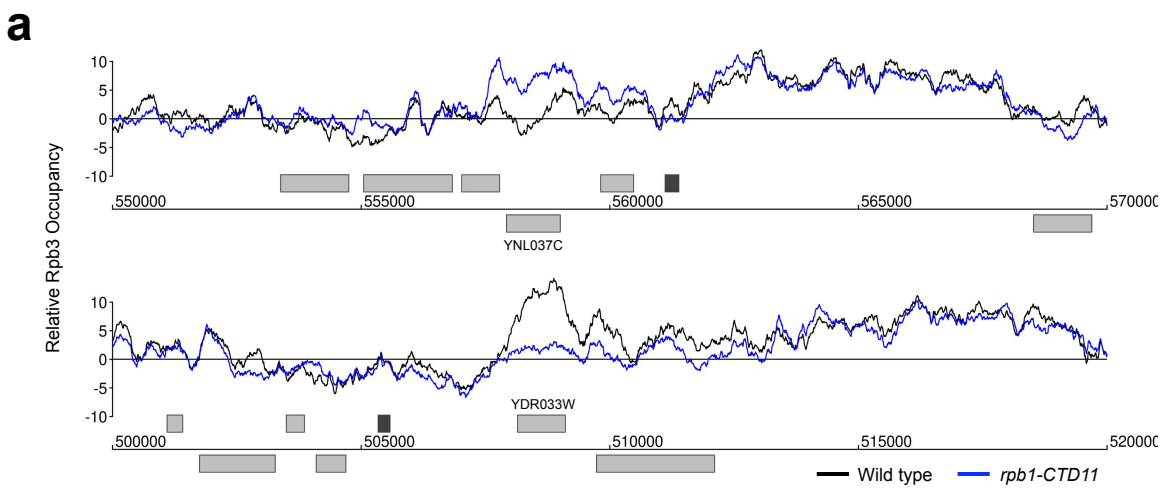

e
Enriched Transcription Factors
Ume6
Rpn4
Met31
Gcn4
Ste12
TF enriched in up−regulated genes (P < 0.05)
TF enriched in down−regulated genes (P < 0.05)
Percentage of regulated genes
Increased Expression
Decreased Expression

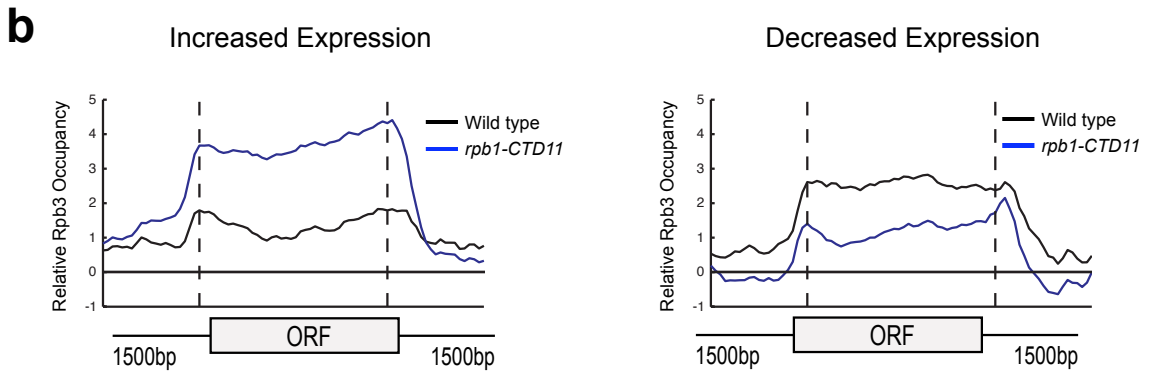
**Figure 2.3 Serial CTD truncations led to progressive steady state transcriptional defects.** (A) Heatmap of genes with significantly increased (top) or decreased (bottom) mRNA levels in the *rpb1-CTD11* mutant. Groups A, B and C approximately outline subsets of genes whose expression were decreased when the CTD was truncated to 13, 12 or 11 repeats respectively. Yellow indicates genes with increased mRNA levels and blue indicates genes with decreased levels. (B) Scatterplot of profile paired correlations in gene expression and genetic interaction. These revealed an overall positive correlation indicating similar function information obtained by the gene expression and genetic interaction profiles. Differences were most apparent for *rpb1-CTD13*, which behaved most similar to the *rpb1-CTD12* and *rpb1-CTD11* mutants when genetic interactions are considered, but grouped with the *rpb1-CTD20* and *rpb1-CTDWT* mutants when gene expression profiles were considered. Boxplot of transcriptional frequency (C) and mRNA half-life (D) showing significant differences in half-life (p value 4.54e-14) and transcriptional frequency (p value 0.0131) between genes with increased or decreased expression in the *rpb1-CTD11* mutant. Outliers are not shown. (E) Differences in enriched transcription factors between genes with increased or decreased mRNA levels.

### 2.3.3 The RNAPII-CTD had varying effects on the genome-wide occupancy profile of transcription related factors

The measured gene expression changes in CTD truncation mutants could result from either effects on the synthesis or stability of the mRNA. To differentiate between these two possibilities, we measured RNAPII occupancy genome-wide and determined if the changes in gene expression correlated with alterations in RNAPII occupancy (Complete dataset can be found in array-express, code E-MTAB-1341 https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1341). Specifically, we measured RNAPII in *rpb1-CTD11* and wild type cells by chromatin immunoprecipitation followed by hybridization on a whole genome tiled microarray (ChIP-on-chip) using an antibody specific to the RNAPII subunit Rpb3. Despite the use of different platforms, antibodies and normalization methods, the genome-wide Rpb3 occupancy profiles obtained in wild type cells were highly correlated with those previously published by several groups (Bataille et al. 2012, Kim et al. 2010, Mayer et al. 2010, Schulze et al. 2011, Tietjen et al. 2010). Furthermore, the occupancy maps revealed highly correlated profiles

between *rpb1-CTD11* and wild type cells (Spearman's rho 0.85), agreeing with the limited transcriptional differences detected by the expression analysis. Nonetheless, our Rpb3 occupancy plots showed clear RNAPII occupancy differences along genes that were identified as either having increased or decreased mRNA levels in the *rpb1-CTD11* mutant (**Figure 2.4A and B**). Accordingly, plotting the average Rpb3 occupancy scores of the differentially regulated genes in *rpb1-CTD11* versus wild type cells revealed that the genes with increased mRNA levels had a significant increase in Rpb3 binding levels along their coding regions while the genes with decreased mRNA levels had a significant decrease (one-tailed t-test p value 2.98e-22 and 3.36e-7, respectively), thus suggesting a direct effect of truncating the CTD on RNAPII levels and mRNA synthesis at specific loci (**Figure 2.4C**).


a
Relative Rpb3 Occupancy
10
5
0
-5
-10
550000
555000
560000
565000
570000
YNL037C
500000
505000
510000
515000
520000
YDR033W
Wild type
rpb1-CTD11


b
Increased Expression
Decreased Expression
Relative Rpb3 Occupancy
Wild type
rpb1-CTD11
ORF
1500bp
1500bp

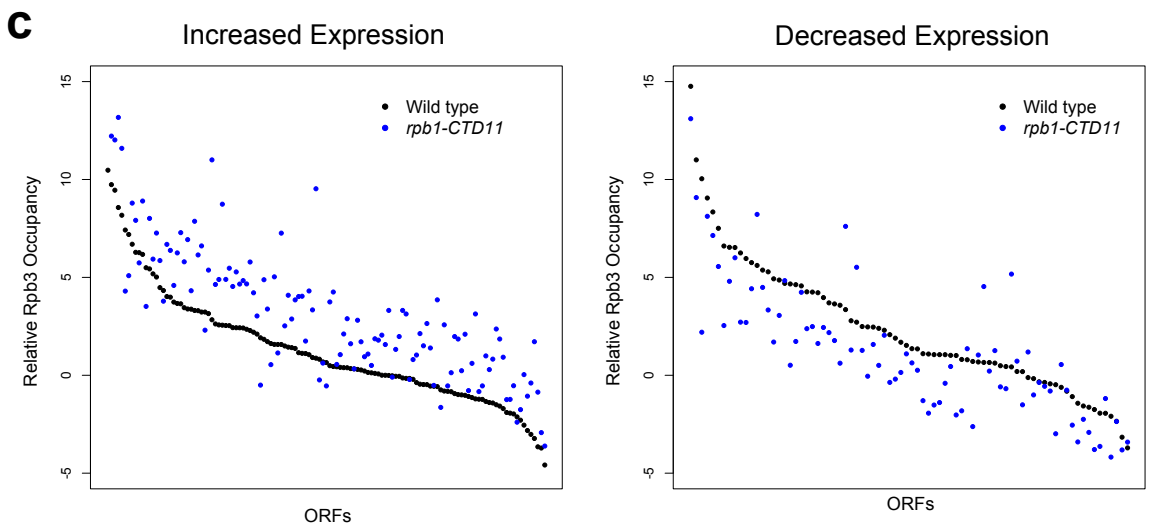

c
Increased Expression
Decreased Expression
Relative Rpb3 Occupancy
Wild type
rpb1-CTD11
ORFs

**Figure 2.4 Genome-wide occupancy profiles of RNAPII identified a direct role for the CTD in transcription regulation.**
(A) Chromosome plots of relative Rpb3 occupancy revealed similar profiles between wild type and *rpb1-CTD11* mutants. Rpb3 occupancy differences were observed in the *rpb1-CTD11* mutant at genes identified to have significantly increased (YNL037C - top) or decreased (YDR033W - bottom) mRNA levels. Light gray boxes depict ORFs and dark gray boxes depict autonomous replicating sequences (ARS). (B) Average gene profile of Rpb3 in genes with increased (left) or decreased (right) mRNA levels upon truncation of the CTD. (C) Average Rpb3 occupancy scores at coding regions with increased (left) (p value 3.36e-7) or decreased (right) (p value 2.98e-22) mRNA levels revealed an intimate link between Rpb3 binding and expression levels.

To better understand the effect of truncating the CTD on transcription, we generated genome-wide association profiles of representative transcription associated factors. These factors included the initiation factor, TFIIB, which is encoded by the *SUA7* gene, the capping enzyme Cet1, the elongation factor Elf1, and the Set2-dependent elongation associated chromatin mark histone H3 lysine 36 trimethylation (H3K36me3) (Complete dataset can be found in array-express, code E-MTAB-1379 https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1379). We note that with the exception of *CET1* (which was not present on our E-MAP array), the genes encoding these factors had negative genetic interactions with our shortest CTD truncation allele. Our genome-wide occupancy profiles under wild type conditions were highly correlated to those previously reported (**Figure 2.5**) (Mayer et al. 2010, Pokholok et al. 2005). Overall, genome-wide occupancy was independent of CTD length for TFIIB, Elf1 and H3K36me3, despite the latter having decreased bulk levels in CTD truncation mutants (**Figure 2.5 and Figure 2.6**) (Xiao et al. 2003). In contrast, Cet1 chromatin association decreased primarily in genes with lower transcriptional frequencies, perhaps reflective of its decreased binding to RNAPII with a shortened CTD (**Figure 2.6B**) (Suh et al. 2010). Focusing on only the genes whose expression levels were altered in the CTD truncation mutants, we observed several interesting patterns.

First, the levels of H3K36me3 correlated well with the transcription changes as its occupancy was decreased in genes whose expression decreased and increased in genes whose expression increased in the *rpb1-CTD11* mutant (paired t-test p value 8.68e-6 and 9.34e-23 respectively) (**Figure 2.6A**). Second, the levels of Cet1 were greatly reduced at the promoters of genes whose expression increased in *rpb1-CTD11* while only slightly reduced at those whose expression decreased (**Figure 2.6B**) (paired t-test p value 7.82e-25 and 2.72e-7 respectively). Lastly, both TFIIB and Elf1 had statistically significant CTD-length dependent occupancy changes, although the overall magnitude of change was minor compared to that of H3K36me3 and Cet1 (**Figure 2.6C and D**).

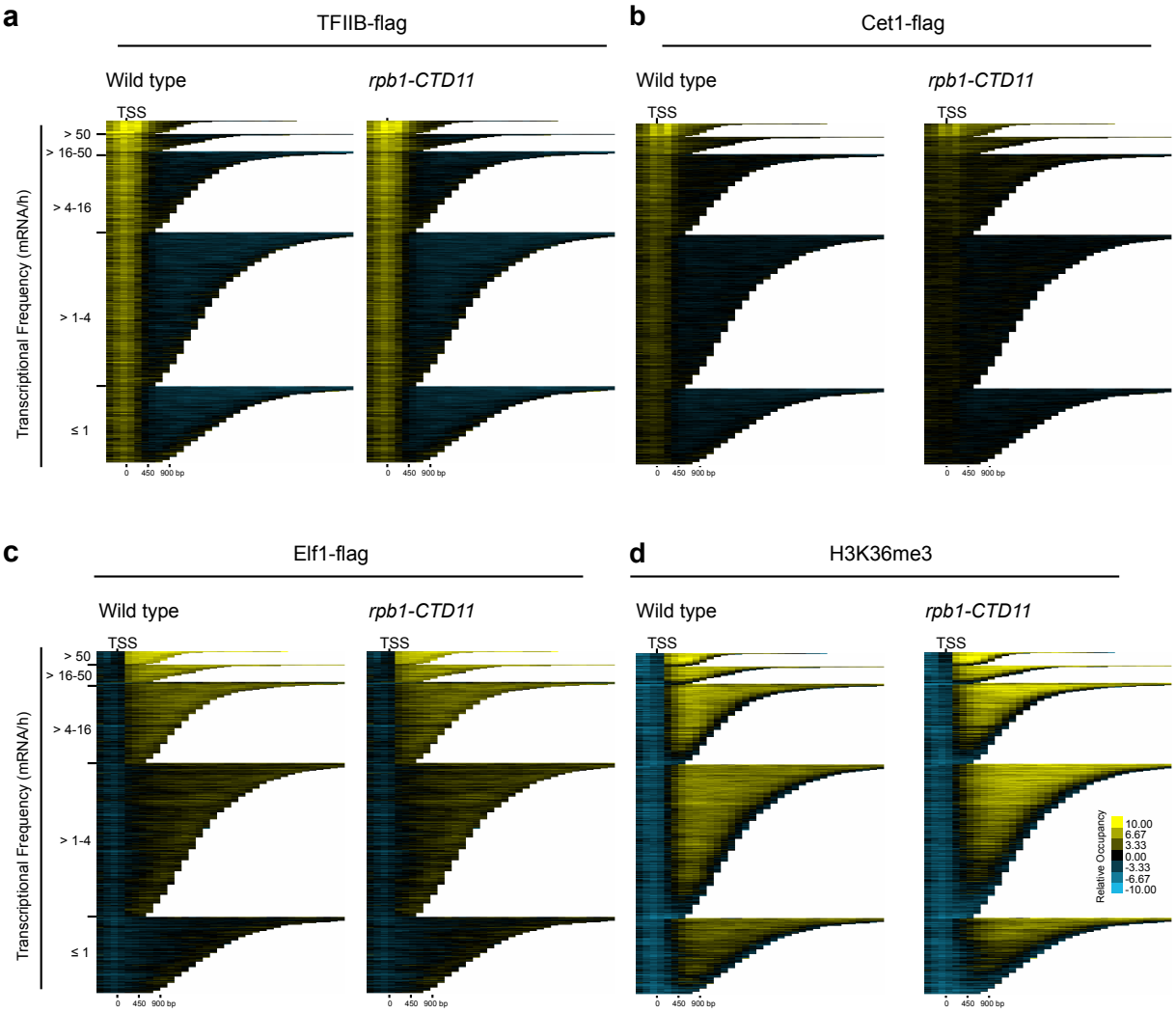



**Figure 2.5 Truncation of the RNAPII-CTD led to changes in the genome-wide association of transcription association factors.**
(A, B, C and D) CHROMATRA plots (Hentrich et al. 2012) of TFIIB, Cet1, Elf1 and H3K36me3 occupancy under wild type and *rpb1-CTD11* mutant conditions. These depict the relative occupancy of TFIIB, Cet1, Elf1 and H3K36me3 along the length of all protein coding genes. Genes are aligned by their transcriptional start side (TSS), grouped into five classes according to their transcriptional frequency as per Holstege *et al* 1998, and sorted by their length. Overall, with the exception of Cet1, truncation of the RNAPII-CTD resulted in few changes to the occupancy of transcription related factors. For Cet1, most protein coding genes showed a reduction in Cet1 occupancy with the exception of highly transcribed genes, which showed unaltered Cet1 occupancy in the *rpb1-CTD11* mutant.

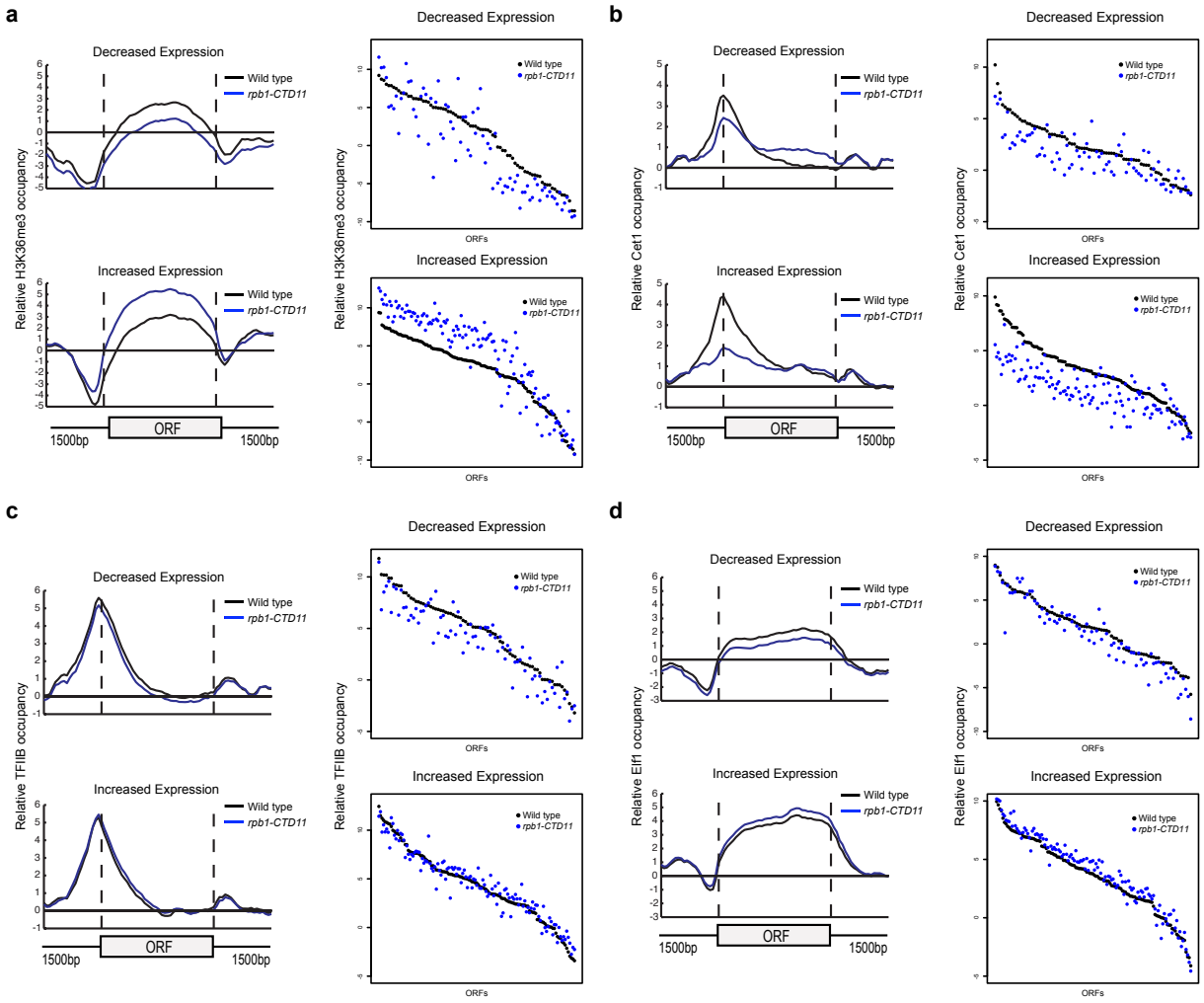



**Figure 2.6 The RNAPII-CTD was critical for the association of transcription related factors.**

(A, B, C and D) Left. Average gene profiles of H3K36me3, Cet1, TFIIB and Elf1 at genes with decreased (top) or increased (bottom) mRNA levels upon truncation of the CTD. Right. Average occupancy scores of H3K36me3, Cet1, TFIIB and Elf1 at genes with decreased (top) (paired t-test p value 8.68e-6, 2.72e-7, 8.66e-8 and 9.17e-6 respectively) or increased (bottom) (paired t-test p value 9.34e-23, 7.82e-25, 0.136 and 4e-15 respectively) mRNA levels upon truncation of the CTD. For H3K36me3 and Efl1, the average occupancy scores were calculated for the coding region. For Cet1 and TFIIB, the average occupancy scores were calculated for the promoter, which consisted of 500 bp upsteam of the start codon.

### 2.3.4 Increases in mRNA levels in CTD truncation mutants were in part a result of increased transcription initiation

The genetic similarity of CTD truncation mutants with mutants encoding initiation factors along with the ChIP-on-chip profiles of RNAPII and transcription associated factors suggested that possible changes to transcription initiation in the CTD truncation mutants might mediate some of the effects on gene expression. Using a LacZ reporter gene strategy we tested if the promoter elements of a set of exemplary genes sufficed to recapitulate the observed changes in expression. These assays revealed significant increases in β-galactosidase activity when the promoter regions of a subset of genes with increased mRNA levels were tested in *rpb1-CTD11* mutants compared to wild type (**Figure 2.7**). These data confirmed that alterations to promoter-directed initiation events were in part responsible for the increased expression observed for these genes at their native loci. In contrast, the promoters of the genes with decreased mRNA levels in *rpb1-CTD11* mutants showed no significant differences in β-galactosidase as compared to wild type cells (The YKL096W and YDR033W promoter were tested) (data not shown).

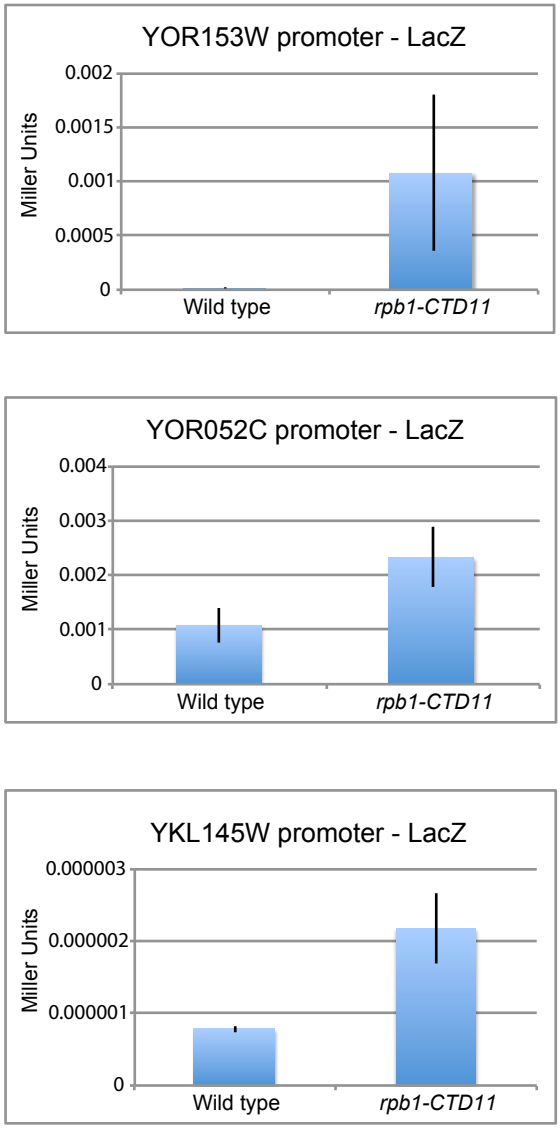



**Figure 2.7 Increases in mRNA levels in CTD truncation mutants were in part a result of increased transcription initiation.**
Plasmid based reporter assays showed that 450 bp of promoter sequence were sufficient to recapitulate the expression levels of three genes with increased mRNA levels in the *rpb1-CTD11* mutant. Assays were performed in triplicate and error bars represent standard deviations.

### 2.3.5 Deletion of *CDK8* normalized mRNA and RNAPII levels at a subset of *rpb1-CTD11* mis-regulated genes

We next expanded our characterization of the CTD to explore the well-established connection to Cdk8 in more detail. First, we showed that in addition to suppressing the cold sensitive phenotype of CTD truncation mutants, loss of *CDK8* could also suppress other known CTD growth defects (**Figure 2.8A**) (Hengartner et al. 1998). Second, despite Cdk8 being able to phosphorylate the CTD, its loss had only very minor effects on the bulk CTD phosphorylation defects seen in CTD truncation mutants (**Figure 2.8B**) (Chapman et al. 2007, Fuchs et al. 2011). Third, we found that loss of *CDK8* had striking effects on the mRNA levels of genes whose expression was dependent on the CTD. Specifically, comparison of mRNA expression profiles for *rpb1-CTD11 cdk8Δ* and *rpb1-CTD12 cdk8Δ* double mutants to the single mutants revealed wide-spread and robust restoration of most of the genes with increased mRNA levels in *rpb1-CTD11*, while only a subset of the genes with decreased mRNA levels appeared to be suppressed (**Figure 2.9A**). The restoration of mRNA levels in the genes with incrreased expression in the *rpb1-CTD11* mutant was mediated by regulation of RNAPII levels, as Rpb3 occupancy changed from an elevated state in the *rpb1-CTD11* mutant to close to wild type levels in the *rpb1-CTD11 cdk8Δ* mutant (**Figure 2.9B**). Accordingly, the average Rpb3 binding scores at these genes in the *rpb1-CTD11 cdk8Δ* mutant were significantly lower than the scores of the *rpb1-CTD11* mutant and were not statistically different from the scores of wild type cells (one-tailed t-test p value 7.17e-18 and 0.159 respectively) (**Figure 2.9C**). Consistent with fewer genes being suppressed in the set of genes with decreased mRNA levels in the *rpb1-CTD11* mutant, a restoring effect on RNAPII levels was not observed at these genes (**Figure 2.9C**).

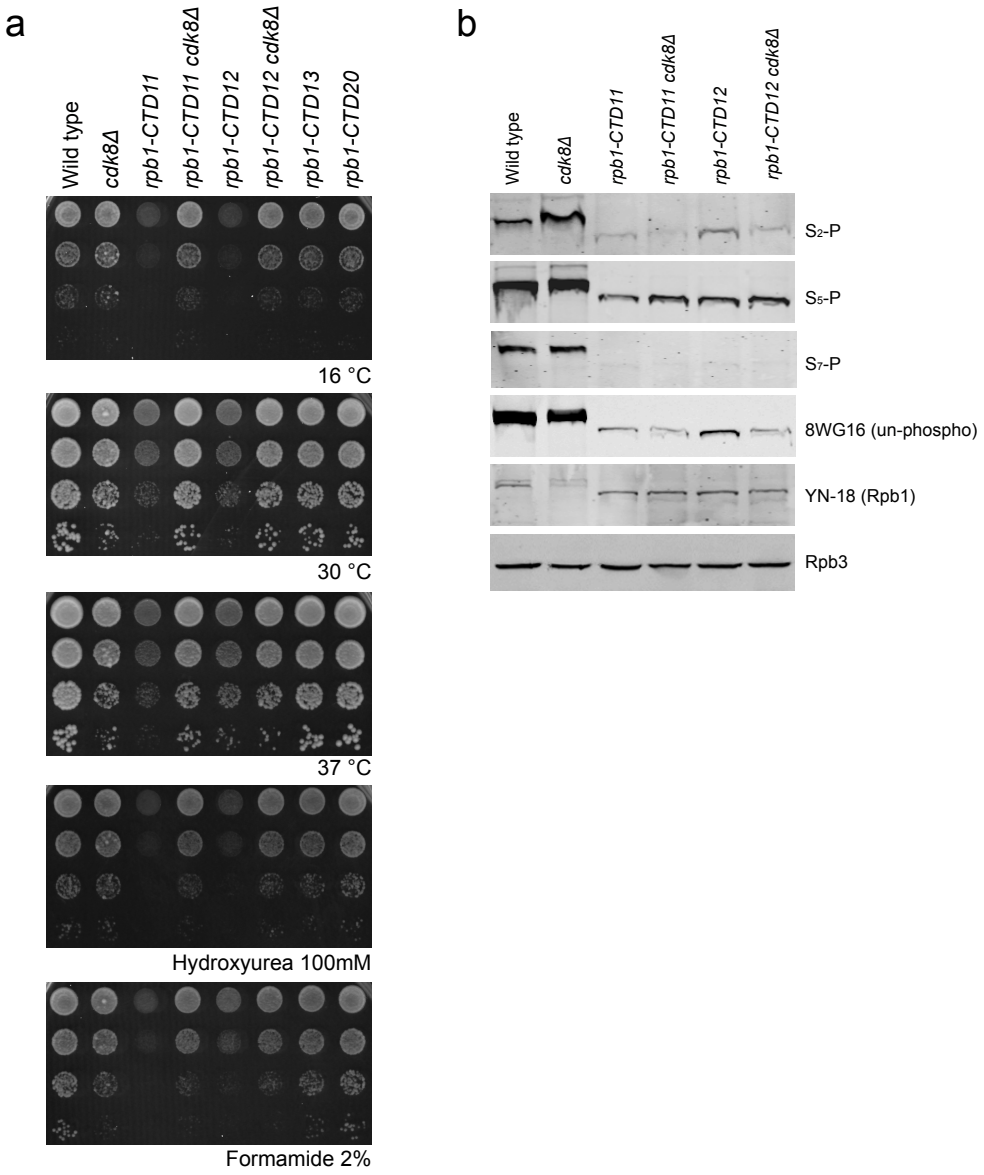



**Figure 2.8 Deletion of *CDK8* suppressed CTD-associated growth phenotypes.**
(A) The sensitivity of CTD truncation mutants containing 11 or 12 repeats to known and novel growth conditions was suppressed by deleting *CDK8*. Ten-fold serial dilutions of strains containing the indicated CTD truncations with and without deletion of *CDK8* were plated and incubated on YPD media at 16, 30 and 37° C and YPD media containing the indicated concentrations of hydroxyurea or formamide. (B) Immunoblots of whole cell extracts with CTD phosphorylation specific antibodies. YN-18 detects the N-terminus of Rpb1 and was used as a control for Rpb1 protein levels. Rpb3 was used as a loading control.

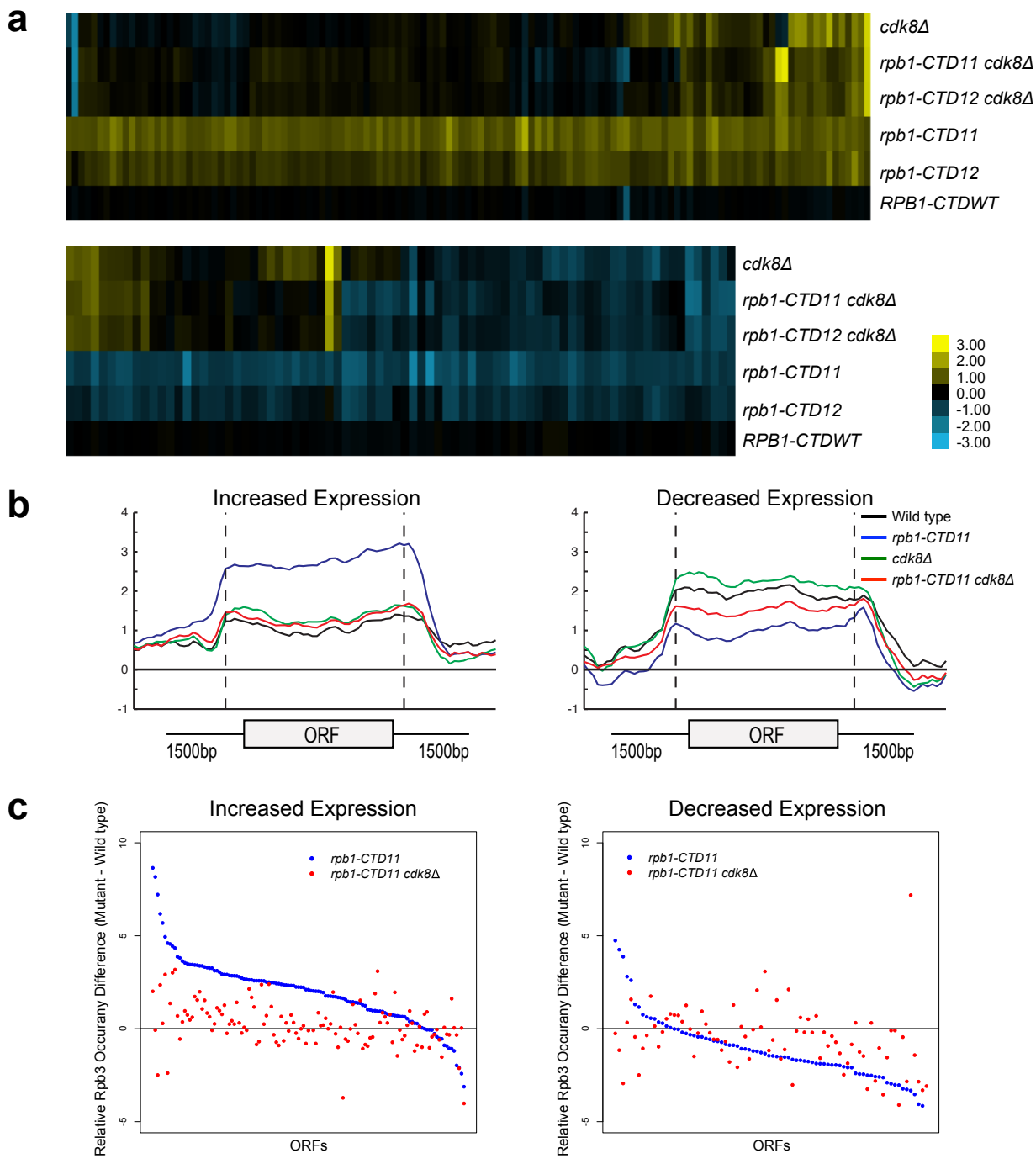



**Figure 2.9 Loss of *CDK8* normalized *rbp1-CTD11* transcriptional defects by altering RNAPII recruitment.**

(A) Heatmap of genes with increased (top) or decreased (bottom) mRNA levels in the *rpb1-CTD11* mutant. Deletion of *CDK8* restored the mRNA levels of genes with increased levels in the *rpb1-CTD11* mutant. (B) Average gene profile of Rpb3 in genes with increased (left) or

decreased (right) mRNA levels upon truncation of the CTD. (C) Average difference from wild type in Rpb3 occupancy for coding regions determined to have significantly increased or decreased mRNA levels in the *rpb1-CTD11* mutant.

A previously characterized phenotype of CTD truncation mutants is reduced activation of *INO1* and *GAL10* upon switching to inducing conditions. Therefore, we investigated if loss of *CDK8* could also suppress these expression defects of CTD truncation mutants (Scafe et al. 1990). Focusing on *INO1*, a gene important for the synthesis of inositol and survival in response to inositol starvation, we measured *INO1* mRNA levels in wild type, *rpb1-CTD11*, *cdk8Δ* and *rpb1-CTD11 cdk8Δ* mutants before and after induction. In agreement with previous work, *rpb1-CTD11* mutants had an impaired ability to activate *INO1* expression upon induction (**Figure 2.10A**) (Brickner et al. 2007, Scafe et al. 1990). Upon deletion of *CDK8, INO1* mRNA levels were robustly and reproducibly restored. This effect was corroborated with the suppression of the growth defect of CTD truncation mutants in media lacking inositol upon removal of *CDK8* (**Figure 2.10B**). Consistent with this being a direct effect on mRNA synthesis, Rpb3 levels throughout the *INO1* gene in *rpb1-CTD11* mutants were significantly lower as compared to wild type. Furthermore, upon deletion of *CDK8*, the levels of RNAPII associated with the *INO1* gene were restored (**Figure 2.10C**). While not statistically significant, we nevertheless observed a tendency for increased Rpb3 occupancy at the 3' end of the gene in *cdk8Δ* and *rpb1-CTD11 cdk8Δ* mutants.

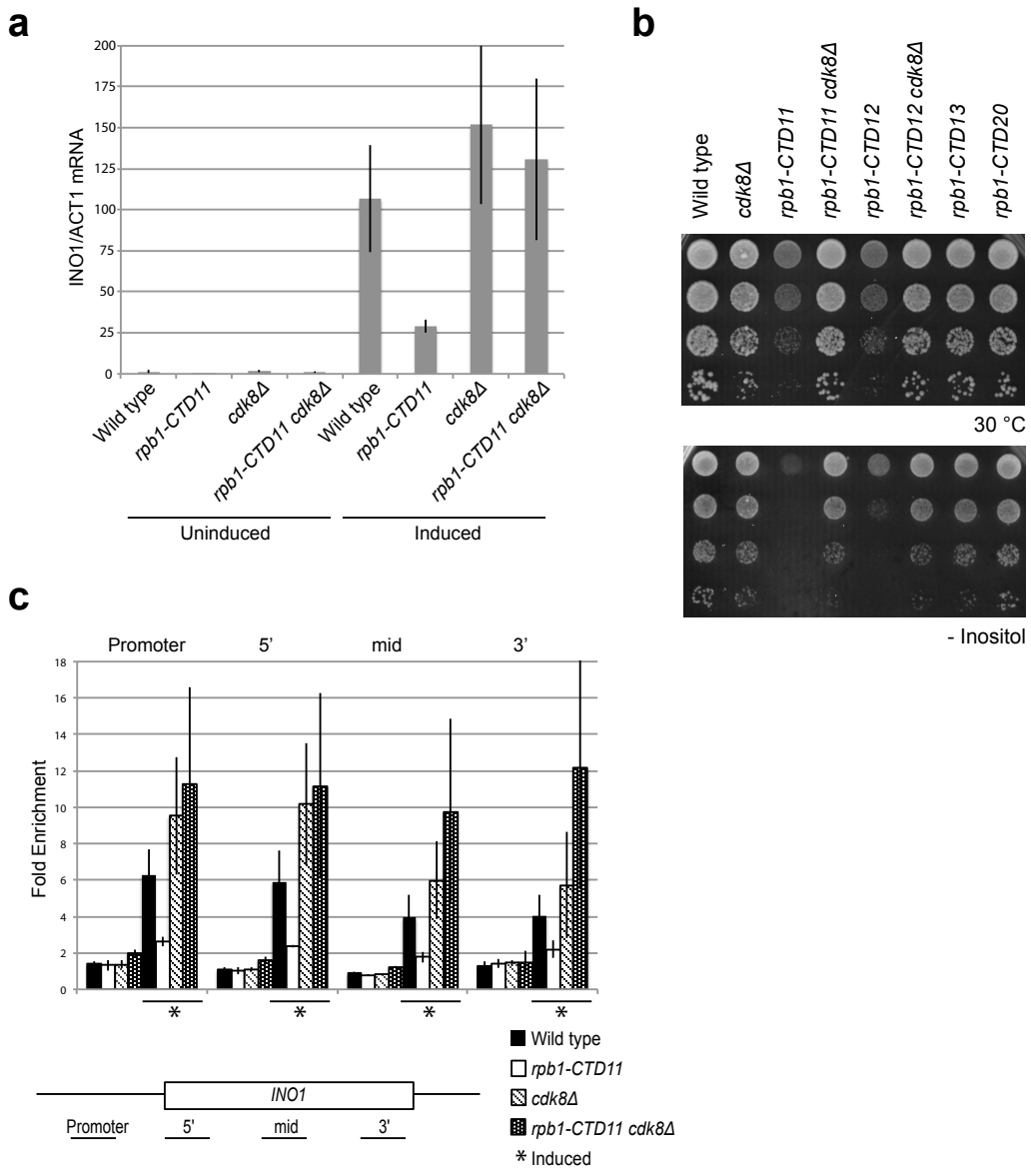



**Figure 2.10 *INO1* expression and RNAPII association defects of *rpb1-CTD11* mutants were suppressed by deleting *CDK8*.**
Cells were grown in inositol containing media (200 μM) to constitute the uninduced sample, and shifted to inositol deplete media for 4hrs to constitute the induced sample. (A) RT-qPCR analysis of *INO1* expression revealed a restoration of expression upon loss of *CDK8*. *INO1* mRNA levels were normalized to *ACT1* levels. (B) The sensitivity of CTD truncation mutants containing 11 or 12 repeats to growth in media lacking inositol was suppressed by deleting *CDK8*. (C) ChIP analysis of Rpb3 binding along the *INO1* gene. Asterisks indicate induced conditions. Rpb3 enrichment along the *INO1* gene was normalized to an intergenic region of chromosome V. Error bars represent standard deviations of values from three replicates.

### 2.3.6 Genes with increased mRNA levels in the *rpb1-CTD11* mutant were directly regulated by Cdk8

To understand the mechanism underlying the restoration of the transcription and RNAPII recruitment changes in the *rpb1-CTD11* mutants upon loss of *CDK8*, we first tried to understand the role of Cdk8 in regulating these genes. To determine if Cdk8 played a direct regulatory role at these genes, we generated a genome-wide map of Cdk8 occupancy under wild type conditions (Complete dataset can be found in array-express, code E-MTAB-1379 https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1379). The average gene occupancy of Cdk8 showed clear enrichment at promoters, although we did identify Cdk8 binding to a small number of ORFs (Andrau et al. 2006, Fan and Struhl 2009, Zhu et al. 2006). Focusing on CTD-length dependent genes, we observed Cdk8 occupancy at the promoters of the genes with increased mRNA levels in the *rpb1-CTD11* mutant (**Figure 2.11A**), while very little Cdk8 was observed at the set of genes with decreased levels (data not shown). Importantly, Cdk8 occupancy was not significantly altered in strains with a truncated CTD (**Figure 2.11A**). In both situations, the preferential association of Cdk8 with the genes having increased expression was significant even when compared to all genes in the genome (one-tailed, unpaired t-test p-value 0.0001079 for wild-type and 0.005898 for *rpb1-CTD11*, respectively), thus supporting a direct regulatory role for Cdk8 at these loci (**Figure 2.11B**). However, despite its significant association and robust effect on normalizing the expression levels of this set of genes, our gene expression analysis clearly showed that Cdk8 was not the sole regulator of these genes as these were generally normal in *cdk8Δ* mutants (**Figure 2.9A**) (van de Peppel et al. 2005).

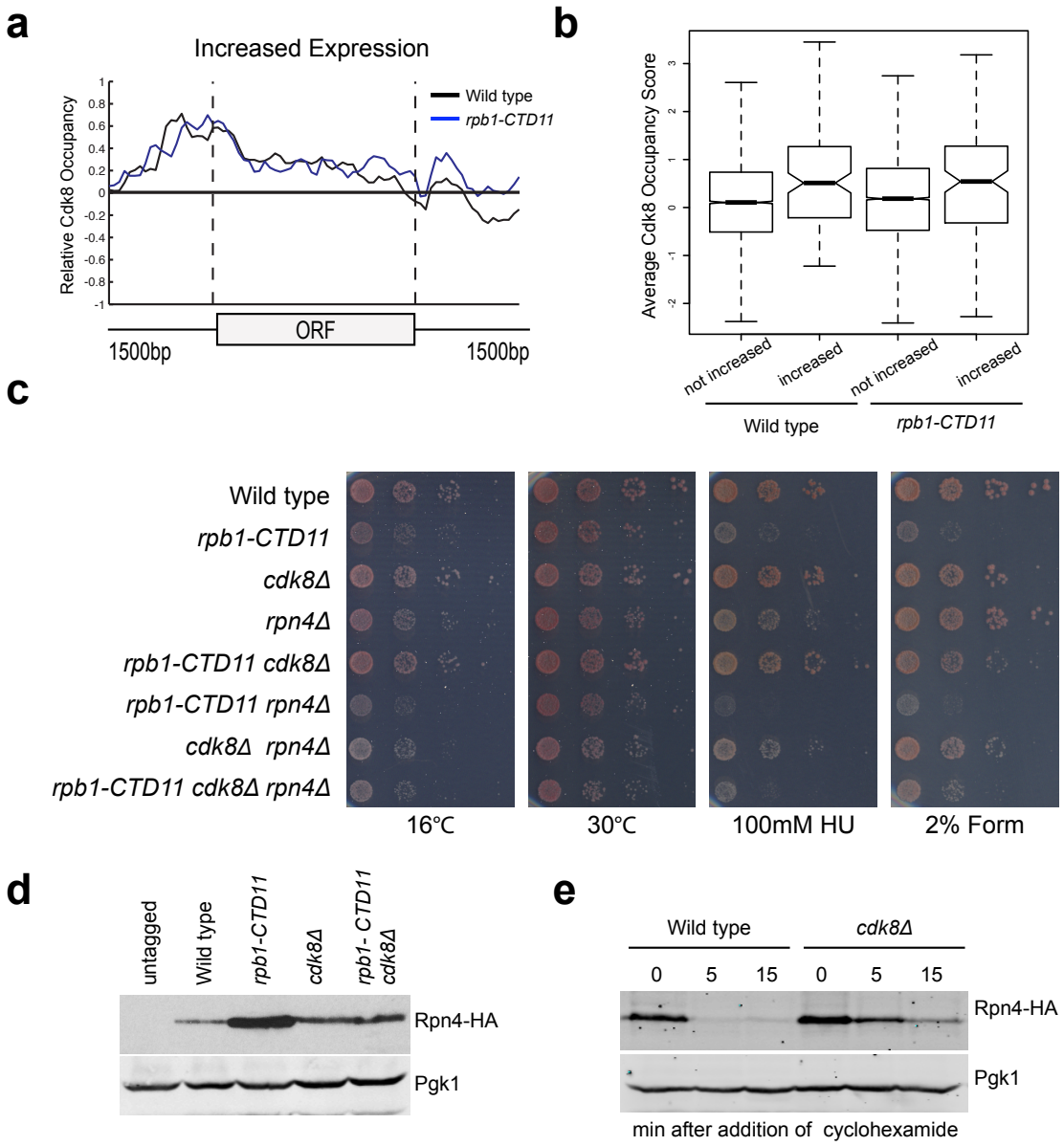



**Figure 2.11 Regulation of Rpn4 levels partly mediated the suppression of *rpb1-CTD11* defects by loss of *CDK8*.**
(A) Cdk8 occupied the promoters of genes whose expression increased in the *rpb1-CTD11* mutant regardless of CTD length. (B) Boxplot comparing average Cdk8 occupancy scores at the promoters of genes whose expression increased in the *rpb1-CTD11* mutant (increased) to all other genes in the genome (not increased). Significantly higher Cdk8 occupancy occurred at the promoters of genes with increased expression levels in both the wild type and the *rpb1-CTD11* mutant. (C) The sensitivity of *rpb1-CTD11*, *cdk8Δ, rpn4Δ* single, double and triple mutants in the W303 background was tested by plating ten-fold serial dilutions on YPD media at 16 or 30° C or in YPD media containing the indicated concentrations of hydroxyurea or formamide. (D)

Immunoblot of Rpn4 protein levels identified an increase of Rpn4 in *rpb1-CTD11* mutants that was reduced upon deletion of *CDK8*. Pgk1 was used as a loading control. (E) Cdk8 regulated the stability of Rpn4 *in vivo.* Rpn4 protein stability was measured at the indicated time points under wild type and *cdk8Δ* conditions. Pgk1 was used as a loading control.

### 2.3.7 The suppression of genes with increased levels in the *rpb1-CTD11* mutant by loss of *CDK8* was through regulation of the protein levels of the transcription factor Rpn4.

Using strict criteria, our profiles of *rpb1-CTD11* and *rpb1-CTD11 cdk8Δ* mutants revealed robust restoration of mRNA levels at 45% of the genes with increased expression levels in the *rpb1-CTD11* mutant and 24% of the genes with decreased levels, when *CDK8* was deleted (**Figure 2.9A**). Among the genes with increased expression, those suppressed were involved in proteasome assembly and proteasome catabolic processes. Consistently, these genes were primarily regulated by Rpn4 (Bonferroni corrected p value of hypergeometric test 1.06E-26). Of the genes with decreased expression, the suppressed set were mainly involved in iron transport, assimilation and homeostasis, however, no significantly associated transcription factors were identified.

Given that our data thus far suggested that the restoring effect was at the level of initiation and mediated by Cdk8, we concentrated our efforts in determining if Rpn4, the only transcription factor found to be significantly involved in regulating the expression of the suppressed set of genes, contributed to the suppression. First, we determined if *RPN4* was genetically required for the suppression of CTD truncation phenotypes by loss of *CDK8* by generating *rpb1-CTD11*, *cdk8Δ* and *rpn4Δ* single, double and triple mutants and testing their growth on different conditions. To test for specificity we also investigated whether the suppression was affected by *GCN4*, which encodes for a transcription factor involved in the regulation of the genes whose

expression increased in the *rpb1-CTD11* mutant but not on those suppressed by deletion of *CDK8*. Deletion of *RPN4* in the *rpb1-CTD11 cdk8Δ* background abolished the suppression, indicating that *RPN4* was genetically required (**Figure 2.11C**; compare *rpb1-CTD11 cdk8Δ* to *rpb1-CTD11 cdk8Δ rpn4Δ*). In contrast, deletion of *GCN4* in the *rpb1-CTD11 cdk8Δ* background had no effect on the suppression, suggesting that the genetic interactions with *RPN4* were specific (**Figure 2.12**).

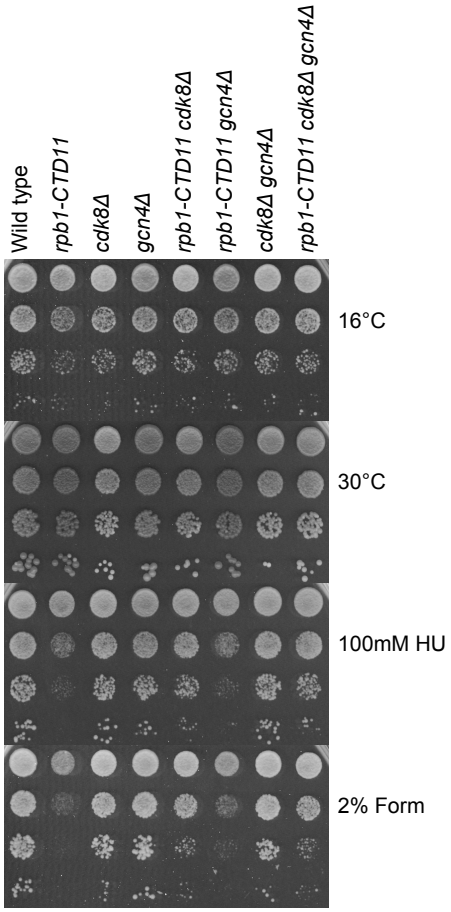



**Figure 2.12 *GCN4* was not involved in the suppression of *rpb1-CTD11* phenotypes by loss of *CDK8*.**
The sensitivity of *rpb1-CTD11*, *cdk8Δ* and *gcn4Δ* single, double and triple mutants in the W303 background was tested by plating ten-fold serial dilutions on YPD media at 16 or 30° C or on YPD media containing the indicated concentrations of hydroxyurea or formamide.

Considering that Rpn4 is a phosphorylated protein, we also tested the involvement of two previously identified phosphorylation sites that are important for its ubiquitin-dependent degradation (Ju et al. 2007). Introduction of the *RPN4* S214/220A mutant restored the suppression in a *rpb1-CTD11 cdk8Δ rpn4Δ* strain in most of the conditions tested, thus demonstrating a general lack of involvement of these phosphorylation sites in the suppression (**Figure 2.13** right panel: compare *rpb1-CTD11 cdk8Δ* and *rpb1-CTD11 cdk8Δ rpn4Δ*) (Ju et al. 2007). Despite our inability to link Rpn4 phosphorylation to the suppression mechanism, the genetic analysis showed that the growth of *rpb1-CTD11 rpn4Δ* double mutants was more compromised than that of *rpb1-CTD11* mutants alone, indicating a clear dependence on Rpn4 function for maintaining *rpb1-CTD11* cell fitness (**Figure 2.11C** compare *rpb1-CTD11* and *rpb1-CTD11 rpn4Δ* mutants). This phenotypic pattern contrasted the apparent increase in Rpn4 function in a *rpb1-CTD11* mutant as suggested by our gene expression analysis, and indicated that mutating *CDK8* normalized, rather than abolished Rpn4 activity in *rpb1-CTD11* mutants. To test this hypothesis, we measured the levels of Rpn4 fused to a hemagglutinin (HA) tag in *rpb1-CTD11* and *cdk8Δ* single and double mutants. Consistent with an increase in Rpn4 function, Rpn4 protein levels were increased in *rpb1-CTD11* mutants compared to wild type cells (**Figure 2.11D**). Surprisingly, Rpn4 protein levels were reduced upon deletion of *CDK8* in the *rpb1-CTD11* mutant, consistent with the observed restoration in gene expression of Rpn4 target genes. In addition, the initial gene expression analysis as well as detailed RT-qPCR analysis of the *RPN4* locus did not detect significant alterations in *RPN4* mRNA levels in *rpb1-CTD11* and *CDK8* single and double mutants, suggesting that the effect of the CTD and Cdk8 on Rpn4 was most likely at the protein level (data not shown). In support of this and consistent with the slightly elevated level of Rpn4 in the *cdk8Δ* strain (**Figure 2.11D**), loss of *CDK8* increased the

half-life of Rpn4 (**Figure 2.11E**) thus indicating a role for Cdk8 was a regulator of Rpn4 stability *in vivo*.

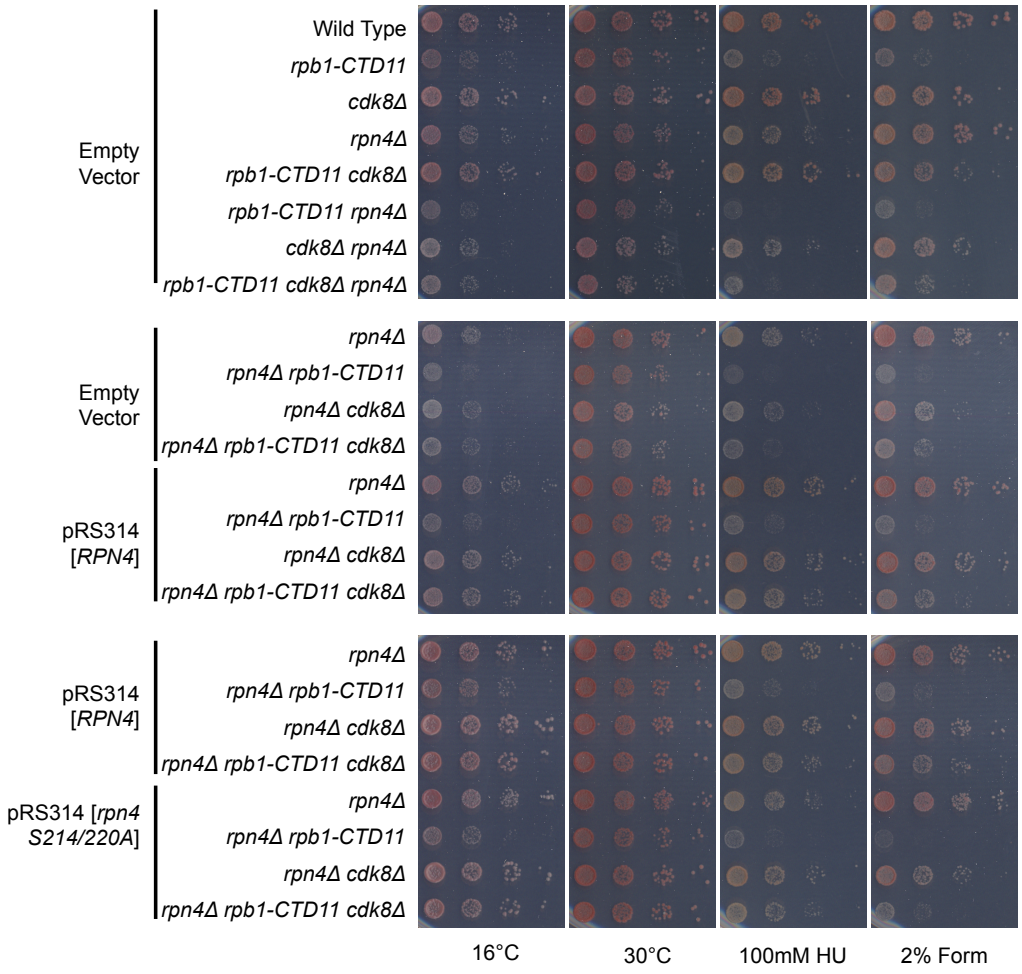



**Figure 2.13 Phosphorylation of Rpn4 at S214/220 was not involved in the suppression of *rpb1-CTD11* defects by loss of *CDK8*.**
The sensitivity of *rpb1-CTD11*, *cdk8Δ, rpn4Δ* single, double and triple mutants carrying an empty vector, or a plasmid containing either *RPN4* or *RPN4 S214/220A* was tested by plating ten-fold serial dilutions on YPD media at 16 or 30° C and YPD media containing the indicated concentrations of hydroxyurea or formamide.

## 2.4 Discussion

Our genetic interaction, mRNA profiling, and RNAPII binding studies illuminated key linkages between CTD function, gene expression, mediator function, and the transcription factor Rpn4. We found distinct CTD length-dependent genetic interactions and gene expression alterations during steady state growth. The majority of the expression changes in the CTD mutants were in genes whose mRNA levels increased and these were accompanied by increased RNAPII binding across their coding regions. CTD truncation mutants primarily had alterations in transcription initiation as suggested by our E-MAP profile of the *rpb1-CTD11* mutant and further supported by reporter assays. Removal of the mediator subunit, Cdk8, in cells with shortened CTD restored the original mRNA levels and RNAPII occupancy profiles at a subset of genes whose expression was increased in the CTD truncation mutant, highlighting an activating role for Cdk8 in gene expression regulation. In contrast, loss of *CDK8* also restored the reduced activation of the *INO1* gene exemplifying the more established repressive role for Cdk8. Finally and highly consistent with the expression results, shortening the CTD resulted in increased cellular amounts of the transcription factor Rpn4, which was normalized upon concomitant removal of *CDK8*. Underscoring its role, we found that *RPN4* was genetically required for the suppression of CTD truncation phenotypes by loss of *CDK8*.

The mRNA analysis identified genes whose expression levels during normal growth were dependent on CTD length, thus expanding the existing knowledge of CTD function *in vivo*, which has been derived from a primary focus on genes activated in response to specific conditions including *INO1* and *GAL10* (Scafe et al. 1990). Despite the CTD being essential for

viability *in vivo*, we detected a seemingly low number of genes with altered expression levels in *rpb1-CTD11* mutants. We reconcile this with the fact that our shortest allele was four repeats above the minimum required for viability in *S. cerevisiae*, suggesting that we were predominantly assaying those genes most sensitive to changes in CTD length rather than the essential function of the CTD. Nonetheless, using stringent criteria our data identified a set of over 200 genes whose transcription was CTD length-dependent. As expected from the well-documented role of the CTD in transcription activation, about 40% of CTD-dependent genes had decreased expression. Surprisingly, we found that about 60% of CTD-dependent genes had increased expression. Functional analysis of the genes with increased or decreased expression upon CTD truncation revealed key differences in mRNA stability, transcriptional frequency, GO categories, and associated transcription factors, suggesting differential effects on groups of genes with distinct properties. In addition, for both groups there was a high correlation between mRNA levels and RNAPII occupancy suggesting a direct effect on RNAPII function rather than changes in posttranscriptional RNA processing. Furthermore, truncating the CTD also caused changes in the association of Cet1 and H3K36me3 at genes whose expression was altered in the *rpb1-CTD11* mutant. Finally, our data linked the alterations observed at the genes with increased mRNA levels to changes in transcription initiation using promoter-fusion experiments. How this latter finding can be reconciled with the minor changes in TFIIB association at the promoters of these genes remains to be determined.

The increased mRNA levels and concurrent increase in occupancy of RNAPII in *rpb1-CTD11* mutants presents an interesting conundrum. Seemingly, these results pointed to a previously unreported inhibitory function of the CTD, as shortening it relieved the inhibition and resulted in

higher RNAPII occupancy. However, we favor a model in which these relationships are reflective of a cellular stress response elicited by impairing CTD function. Consistent with this hypothesis, CTD truncation mutants displayed heightened sensitivity to a variety of stressors, as shown by others and us (Wong and Ingles 2001, Nonet et al. 1987, Nonet and Young 1989). Furthermore, CTD truncation mutants had increased levels of Rpn4 protein and the genes that had increased mRNA levels tended to be regulated by Rpn4, consistent with their important contributions to the cellular stress response (Kruegel et al. 2011, Owsianik et al. 2002, Wang et al. 2010).

We also investigated the molecular underpinnings of the well-established connection between Cdk8 and the RNAPII-CTD. To this end, we found that deletion of *CDK8* normalized the expression of genes with increased mRNA levels in the CTD truncation alleles. This observation is consistent with the less-understood role for *CDK8* as an activator of transcription, likely acting by enhancing recruitment of RNAPII with a shortened CTD to its target genes. Given that Cdk8 was found to be preferentially associated with the promoters of these genes regardless of CTD length, it is likely that this represents a direct mechanism. Importantly, our data clearly showed that Cdk8 was not the sole regulator of this subset of genes as a single deletion of *CDK8* does not alter their expression. Thus, in wild type cells Cdk8 associated at these genes' promoters but it only enhanced transcription when CTD function was disrupted. These observations are in agreement with Cdk8's well-established role in the response to environmental signals (Hirst et al. 1999, Nelson et al. 2003, Raithatha et al. 2011). Furthermore, we showed that Cdk8's role in activating CTD-dependent genes with increased mRNA levels was in part mediated by increasing the protein levels the transcription factor Rpn4, which we found to be genetically

required for the suppression. Accordingly, the levels of Rpn4 protein correlated with the mRNA levels of Rpn4 targets genes in *rpb1-CTD11* and *cdk8Δ* single and double mutants. This is consistent with the known role of Cdk8 in regulating protein levels of transcription regulatory proteins and the established function of Rpn4 in activating gene expression as a result of stress (Hahn et al. 2006). Reminiscent of recent work by several groups showing that loss of Cdk8 stabilizes Gcn4 protein levels, our data on Rpn4 protein stability provided further support of a close linkage between Cdk8 and Rpn4, although the mechanistic details remain to be determined (Chi et al. 2001, Lipford et al. 2005, Rosonina et al. 2012). In addition, we note that not all suppressed genes are known targets of Rpn4, suggesting that it is likely not the only factor linking the RNAPII-CTD and Cdk8 function.

The fact that removal of Cdk8 also suppressed defects in activated transcription suggested an entirely different relationship between the RNAPII-CTD and Cdk8 from the one described above, this time involving a negative role for Cdk8. This is exemplified by the *INO1* locus, where *rpb1-CTD11* mutants have decreased mRNA expression and RNAPII association when grown in inducing conditions, a defect that was restored upon deletion of *CDK8.* While reminiscent of the model postulating that Cdk8-catalyzed phosphorylation of the CTD prevents promoter binding of RNAPII and thus results in transcriptional repression, we do not think this is the mechanism of suppression described here (Hengartner et al. 1998). First, deletion of *CDK8* had no alleviating effects on the bulk phosphorylation status of either full-length or truncated CTD. Second, deletion of *CDK8* alone under non-inducing conditions did not result in de-repression of *INO1*, in contrast to well-characterized Cdk8 target genes (van de Peppel et al. 2005). Lastly, despite our genome-wide Cdk8 occupancy data showing a reproducible, albeit

slight, enrichment of Cdk8 at the *INO1* promoter, it does not meet our enrichment criteria, making it unclear if Cdk8 directly associates and functions at this locus (data not shown). In conclusion, our data revealed a tight link between Cdk8 and the RNAPII-CTD in transcription regulation, where Cdk8 can both enhance and repress transcription, the former in part mediated by regulating the levels of the transcription factor, Rpn4.

# Chapter 3: RNAPII-CTD Maintains Genome Integrity Through Inhibition of Retrotransposon Gene Expression and Transposition

## 3.1 Introduction

RNA polymerase II (RNAPII) is the enzyme responsible for the transcription of a diverse set of genomic loci, including most protein coding genes, many non-coding genes, and retrotransposons. Rpb1, the largest subunit of RNAPII, contains a unique C-terminal domain (CTD) that is composed of heptapeptide repeats ($Y_1$ $S_2$ $P_3$ $T_4$ $S_5$ $P_6$ $S_7$), the number of which increases with genomic complexity (Allison et al. 1985, Corden et al. 1985). The CTD plays key roles in the regulation and coordination of co-transcriptional processes *in-vivo* (Heidemann et al. 2012, Hsin and Manley 2012, Zhang et al. 2012b). In *S. cerevisiae*, deletion of the CTD is lethal, while strains carrying shortened CTDs are viable but display a range of conditional phenotypes, including reduced growth when exposed to high or low temperatures, inositol-deplete conditions, or to the chemicals formamide or hydroxyurea. (Aristizabal et al. 2013, Wong and Ingles 2001, Nonet et al. 1987, Nonet and Young 1989, Scafe et al. 1990). CTD truncation mutants also have alterations in gene expression under normal and inducing conditions as evidenced by increased mRNA and RNAPII levels at a subset of Rpn4-dependent genes and in decreased induction of the *INO1* and *GAL4* genes (Aristizabal et al. 2013, Scafe et al. 1990).

Retrotransposons constitute a major group of genetic elements transcribed by RNAPII. They comprise over 3% of the genome, including 50 full-length elements, which collectively account for 5-10% of the total mRNA in haploid yeast (Elder et al. 1981, Kim et al. 1998). In *S. cerevisiae*, retrotransposons are flanked by long terminal repeats (LTR), which contain promoter and termination sequences for the transcription of the retrotransposon's genes (Havecker et al. 2004). Retrotransposons contain a *gag* gene, which encodes a structural coat protein, and a *pol* gene, which encodes a polypeptide that is processed into the enzymes reverse transcriptase, protease and integrase. Generally, retrotransposons cannot be transmitted across cells, thus they primarily multiply within the host genome. The retrotransposon replication cycle begins with a RNA intermediate that is transcribed from the 5' to the 3' LTR by RNAPII (Boeke et al. 1985). The RNA produced is used for the synthesis of the retrotransposons' proteins and as a template for reverse transcriptase. Transposition involves the integration of retrotransposon cDNA into a new genomic location, where transcription can begin again giving rise to a new replication cycle.

The 50 full-length retrotransposon elements in the *S. cerevisiae* genome belong to five different families called Ty1 to Ty5 (Bleykasten-Grosshans et al. 2013, Kim et al. 1998). These differ primarily in the order and sequence of their encoded genes with Ty1 and Ty2 elements being further divided into subfamilies. Of these, only members of Ty1, Ty2 and Ty3 families are thought to be capable of transposition, whereas Ty4 and Ty5 elements are likely inactive due to the accumulation of deleterious mutations (Voytas and Boeke 1992). In addition to full-length retrotransposons, the yeast genome also contains LTR fragments and lone LTRs, known as delta, sigma, tau, and omega elements (Kim et al. 1998). These are LTR sequences originating from intact retrotransposons that underwent homologous recombination between the almost identical

LTRs flanking the element, and as such they are not associated with retrotransposon coding sequences. Lone LTR sequence and location provide a record of previous retrotransposon integration events.

Transposition can have grave consequences for genome structure and function, making retrotransposons important causes of genome instability (Lesage and Todeschini 2005). Specifically, integration within host genes, although rare, can result in disruption of genetic information, while insertion within a transcription regulatory region can alter the expression of the adjacent genes (Roeder and Fink 1980, Williamson et al. 1981). To restrict genome instability caused by transposition, all stages in the retrotransposon's replication cycle are kept under tight control by the host cell. As such Ty1 gene expression is tightly regulated by cell ploidy and environmental signals, including nutritional status and exposure to DNA damaging agents, and failure to limit Ty1 mRNA levels often results in increased Ty1 mobility (Bradshaw and McEntee 1989, Errede et al. 1980, Sacerdot et al. 2005, Todeschini et al. 2005). Furthermore, several host transcription factors regulate Ty1 transcription (Errede 1993, Gray and Fassler 1993, Gray and Fassler 1996, Laloux et al. 1990, Madison et al. 1998, Morillon et al. 2000, Morillon et al. 2002). Specifically, Ste12 and Tec1 are required for basal Ty1 transcription in haploid yeast as loss of these genes significantly reduces Ty1 mRNA levels (Laloux et al. 1990, Morillon et al. 2002). Furthermore, loss of *TEC1* decreases transposition rates while overexpression of *TEC1* or *GCN4* results in increased Ty1 transposition (Conte and Curcio 2000).

Building on previous work from our laboratory, which illuminated the role of the RNAPII-CTD in the expression of a subset of protein coding genes (Aristizabal et al. 2013), we focused here on examining the role of the RNAPII-CTD in retrotransposon biology. We found that the RNAPII-CTD plays an important role in regulating RNAPII and mRNA levels of Ty1 retrotransposons. Importantly, this effect extended to increased transposition rates, suggesting that the structural integrity of the RNAPII-CTD was required for maintaining genome integrity. Several lines of evidence point to an important role for transcription initiation in mediating the enhanced expression of retrotransposons in cells with shortened RNAPII-CTD, including recapitulation of the effects by promoter-based reporter assays, and requirement for the transcription factors Ste12 and Tec1 and the mediator subunit Cdk8. Lastly, suggesting a broader role for these factors in RNAPII-CTD function, we found that loss of *STE12* or *TEC1* suppressed RNAPII-CTD truncation mutant growth phenotypes, likely in conjunction with *CDK8*.

## 3.2 Materials and methods

### 3.2.1 Yeast strains

Strains are listed in **Table 3.1**. Complete or partial gene deletions were achieved via the one-step gene replacement method (Longtine et al. 1998). CTD truncations were generated previously by addition of a TAG stop codon followed by a NAT, kanamycin or hygromycin resistance marker at the endogenous *RPB1* locus (Aristizabal et al. 2013). All double mutant strains were generated via mating and tetrad dissection. For *STE12* deletion mutants, strains were complemented with pRS316 [STE12] prior to mating. The pRS316 [STE12] plasmid was a gift from Dr. Ivan Sadowski. Plasmids are listed in **Table 3.2**.

**Table 3.1 Strains used in this study**

| Genotype | Background | Reference |
|---|---|---|
| *Matα his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | Aristizabal *et al* 2013 |
| *Matα rpb1-CTD11-nat his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | Aristizabal *et al* 2013 |
| *Matα rpb1-CTD11-nat cdk8::kan his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | Aristizabal *et al* 2013 |
| *Matα cdk8::kan his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | van de Peppel *et al* 2005 |
| *Matα tec1::kan his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | This study |
| *Matα ste12::kan his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | This study |
| *Matα rpb1-CTD11-nat tec1::kan his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | This study |
| *Matα rpb1-CTD11-nat ste12::kan his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | This study |
| *Matα tec1::hygro his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | This study |
| *Matα rpb1-CTD11-nat tec1::hygro his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | This study |
| *Matα cdk8::kan tec1::hygro his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | This study |
| *Matα rpb1-CTD11-nat cdk8::kan tec1::hygro his3Δ1 leu2Δ0 ura3Δ0* | BY4742 | This study |
| *MATa/α his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0* | BY4743 | This study |
| *MATa/α rpb1-CTDWT-nat rpb1-CTDWT-kan his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0* | BY4743 | This study |
| *MATa/α rpb1-CTD11-nat rpb1-CTDWT-kan his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0* | BY4743 | This study |
| *MATa/α rpb1-CTD11-nat rpb1-CTD11-kan his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0* | BY4743 | This study |

**Table 3.2 Plasmids used in this study**

| Relevant Genotype | Backbone | Source |
|---|---|---|
| YJRWTy1-2 promoter | PGL669-z | This study |
| YMLWTy1-2 promoter | PGL669-z | This study |
| STE12 | pRS316 | Dr. Ivan Sadowski |
| YJRWTy1-2 promoter Tec1 binding site deletions | PGL669-z | This study |
| YMLWTy1-2 promoter Tec1 binding site deletions | PGL669-z | This study |
| pJC573 - Ty1*his3AI[Δ1]* | pRS406 | Bryk *et al* 2001 |

### 3.2.2 Genome-wide ChIP-on-chip

All ChIP-on-chip data used were generated previously (Aristizabal et al. 2013). The complete dataset can be found in array-express, code E-MTAB-1341 and E-MTAB-1379 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1341 and

https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1379 respectively). Briefly, overnight cultures were diluted to 0.15 OD600 and grown to 0.5-0.6 OD600 units. Cross-linking was done with 1% formaldehyde for 20 min. Chromatin was prepared as described previously (Schulze et al. 2009). Five µl of anti-Rpb3 (Neoclone) or 4.2 µl of anti-FLAG (Sigma) was used. DNA was amplified using a double T7 RNA polymerase method, biotin labeled, and hybridized to an Affymetrix 1.0R *S. cerevisiae* microarray. Rpb3 samples were normalized to an input and flag tagged samples were normalized to a mock control using the rMAT software (Droit et al. 2010). Relative occupancy scores were calculated for all probes using a 300 bp sliding window. Experiments were carried out in duplicate, quantile normalized and averaged data were used for calculating average enrichment scores. For retrotranposons and LTRs, we averaged probes whose start sites fell within the feature start and end positions. For the box plots, the middle line represents the median and the hinges represent the first and third quartile.

### 3.2.3 Promoter reporter assays

Reporter plasmids were generated by cloning ~1300 bp of the desired promoter region into the Sal1 BamH1 sites of pLG669-Z (Guarente and Ptashne 1981). Specifically, for YJRWTy1-2, 1321 bp were cloned, starting 518 bp upstream of the transposon ORF start and ending 804 bp downstream. For YMLWTy1-2, 1304 bp were cloned, starting 500 bp upstream of the transposons ORF start and ending 804 bp downstream. The cloned sequences were selected such that they included the Tec1 and Ste12 binding sites. Tec1 binding sequences were deleted using nested PCR-based methods and cloned into pLG669-Z using the Sal1 and BamH1 sites. A complete list of plasmids can be found in **Table 3.2**. Reporter plasmids were transformed into

wild type and *rpb1-CTD11* mutants and assayed as previously described (Guarente 1983). Measurements were obtained from three independent cultures and error bars represent standard deviations.

### 3.2.4 Growth assays

Overnight cultures grown on YPD were diluted to 0.5 OD600, 10-fold serially diluted and spotted onto YPD plates with or without the indicated amounts of hydroxyurea (Sigma) or formamide (Sigma). Plates were incubated at the indicated temperatures for 2-4 days.

### 3.2.5 Reverse transcriptase PCR (RT-qPCR)

RNA was extracted and purified using the Qiagen RNeasy Mini Kit. cDNA was generated using the Qiagen QuantiTect Reverse Transcription Kit. cDNA was analyzed using a Rotor-Gene 600 (Corbett Research) and PerfeCTa SYBR Green FastMix (Quanta Biosciences). Samples were analyzed in triplicate from three independent RNA preparations and *TUB1* was used as a control gene (Lu and Kobor 2014). For measuring Ty1 mRNA levels 6 pg/µl of cDNA were used in a 15 µl PCR reaction. For measuring Ty2 mRNA levels 60 pg/µl of cDNA were used. Retrotransposon specific primers were designed, such that the targeted region was unique to all members of a single retrotransposon family. Primer specificity was evaluated by melt curve analysis of the PCR products (data not shown). A complete list of primers used in this study can be found in **Table 3.3**. Error bars represent standard deviations.

**Table 3.3 Primers used in this study**

| Primer name | Forward Sequence | Reverse Sequence | Source |
|---|---|---|---|
| Ty1 | CCAGTTTGGGTGGTATTGGT | TTCTTCGATCTCGGAGGTTC | This study |
| Ty2 | TGCCAACATGGGTAAAACAA | GGCCAATCTGTCGCTAACAT | This study |
| *TUB1* | TCTTGGTGGTGGTACTGGTT | TGGATTTCTTACCGTATTCAGCG | Lu and Kobor 2014 |

### 3.2.6 Ty1 cDNA-mediated mobility assay

This assay tracks the mobility of a genome encoded Ty1 element (Bryk et al. 2001, Scholes et al. 2001). The Ty1 element has a *HIS3* coding sequence inserted in the opposite orientation compared to the Ty1 element. This marker is rendered nonfunctional by the insertion of an artificial intron in the same orientation as the Ty1 element, such that it is only spliced when transcribed from the Ty1 element. During the Ty1 transposition cycle, the Ty1 element is transcribed, the intron is spliced out and the mature RNA is used for the synthesis of cDNA, which is then integrated into a new genomic location. Newly integrated Ty1 elements contain an undisrupted *HIS3* open reading frame and can confer a HIS+ phenotype. Briefly, wild type and *rpb1-CTD11* mutants were transformed with Pac1 digested pJR573 DNA. Transformants were selected in SC-URA media and all subsequent growth procedures were done in this media. Overnight cultures for 12 independent colonies for each strain were started. The next morning, cultures were diluted to OD600 0.3 and incubated at 20 °C for 24 hours. Following, an aliquot was plated onto SC-URA and SC-HIS-URA to count the total number of cells, and cells with retrotransposition events respectively. Plates were grown for 2-3 days until colonies were visible and counted. Results were analyzed using the Fluctuation AnaLysis CalculatOR (FALCOR) web tool using the MSS Maximum Likelihood Method to calculate mutation rates (Hall et al. 2009).

Error bars represent 95% confidence intervals as calculated by the FALCOR web tool. http://www.mitochondria.org/protocols/FALCOR.html.

## 3.3 Results

### 3.3.1 Truncation of the RNAPII-CTD resulted in altered RNAPII occupancy at a subset of retrotransposons

Our previous characterization of genes whose expression is dependent on CTD length focused on protein coding genes (Aristizabal et al. 2013). To test whether the RNAPII-CTD had a role in the regulation of retrotransposons, we explored their occupancy by RNAPII using chromatin immunoprecipitation followed by hybridization to high-density microarrays (ChIP-on-chip). Overall, the *rpb1-CTD11* mutant, which contained only 11 heptapeptide repeats, had significantly increased RNAPII occupancy compared to wild type when all retrotransposons were considered (50 elements) (**Figure 3.1A**)(**Table 3.4**). Given the that high degree of sequence similarity amongst retrotransposon family members limited our ability to uniquely identify single elements in the ChIP-on-chip platform, we focused on retrotransposon families rather than individual retrotransposons. Overall, the Ty1 (31 elements) and Ty2 (13 elements) family of retrotransposons had significant CTD length-dependent increases in RNAPII levels, although the effect at Ty1 elements was more pronounced than at Ty2 elements. The Ty3, Ty4 and Ty5 family of retrotransposons (having 2, 3 and 1 element respectively) were not investigated further because the limited number of members in each family prevented us from doing meaningful statistical analysis. However, individual profiles revealed that truncation of the RNAPII-CTD resulted in elevated RNAPII levels at both Ty3 elements while no effects were observed at Ty4

and Ty5 elements (data not shown). For Ty1 and Ty2 elements, representative examples of individual retrotransposons further manifested the differences apparent from the average occupancy profiles (**Figure 3.1B**). Finally, at Ty1 elements, truncation of the RNAPII-CTD resulted in elevated RNAPII levels along the length of the entire element, whereas at Ty2 elements the increased was focused primarily at the 3' end (**Figure 3.1C and D**).

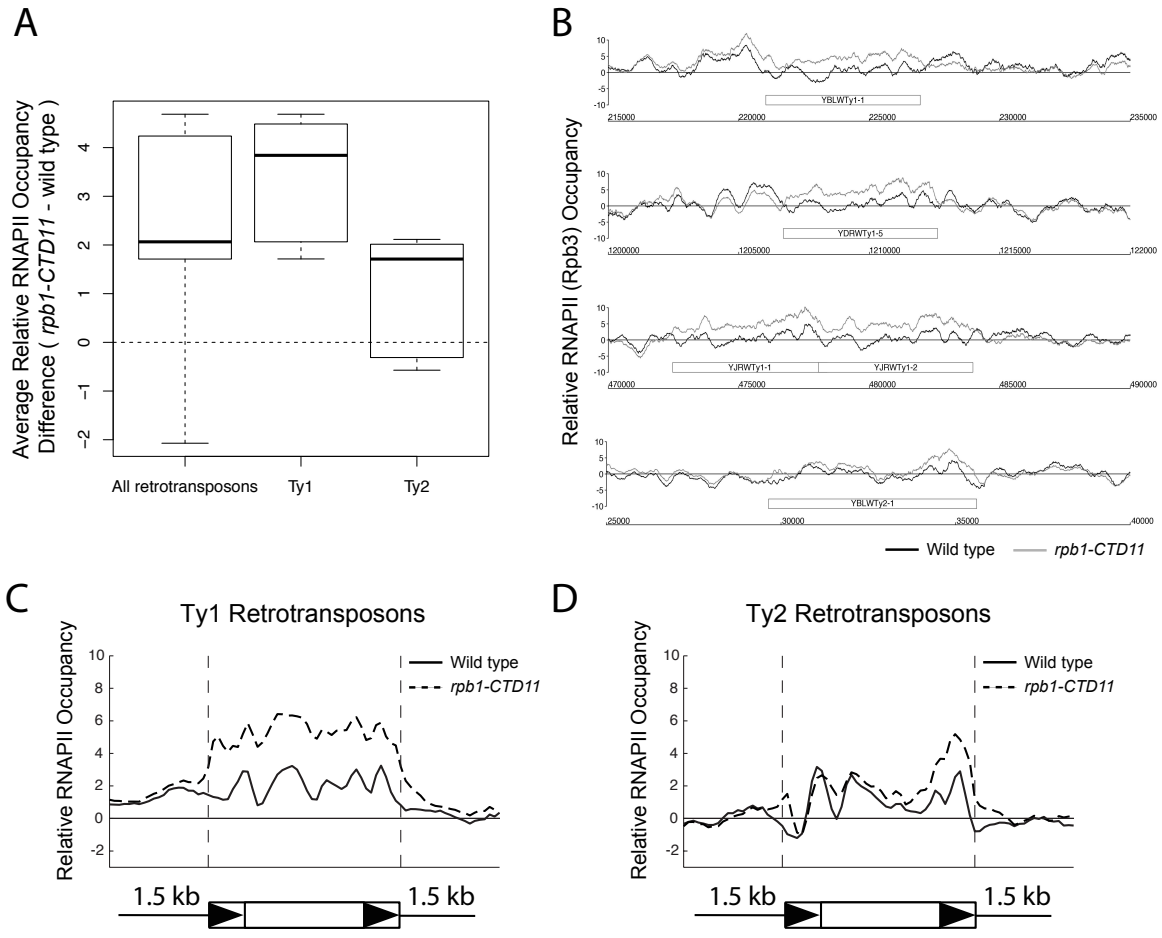



**Figure 3.1 Genome-wide occupancy profiles of RNAPII suggested a role for the RNAPII-CTD in retrotransposon regulation.**
(A) Box plot showing differences in average MAT RNAPII occupancy scores between the wild type and the *rpb1-CTD11* mutant at all, Ty1, or Ty2 retrotransposons. (B) Chromosome plots of relative RNAPII occupancy at representative retrotransposons. Increased RNAPII levels were observed in the *rpb1-CTD11* mutant compared to wild type. Labeled boxes indicate the

retrotransposon. (C) Average gene profile of RNAPII occupancy at Ty1 retrotranspsons showed increased levels along the length of the feature. Below, schematic of an average retrotransposon. Black triangles indicate the LTRs. (D) Average gene profile of RNAPII occupancy at Ty2 retrotransposons revealed increased levels towards the 3' end of the feature.

**Table 3.4 Paired t-test p values comparing RNAPII levels in wild type vs *rpb1-CTD11* at all, Ty1, and Ty2 retrotransposons and derived-LTRs.**

| Element | One tailed paired t-test p value comparing RNAPII levels in wild type vs *rpb1-CTD11* |
|---|---|
| All retrotransposons | 6.42e-12 |
| Ty1 | 1.8e-15 |
| Ty2 | 0.00424 |
| Lone LTRs | 1.85e-16 |
| Ty1-derived LTRs | 1.34e-11 |
| Ty2-derived LTRs | 5.84e-05 |

To test whether the effect of truncating the RNAPII-CTD was specific to full-length retrotransposon elements, we determined RNAPII occupancy at lone LTRs elements and found significantly increased levels in the *rpb1-CTD11* mutant compared to the wild type (**Figure 3.2**) (**Table 3.4**). Focusing on delta elements derived from Ty1 or Ty2 elements revealed that these had significantly increased Rpb3 levels in the *rpb1-CTD11* mutant when compared to wild type, consistent with our findings at intact retrotransposons.

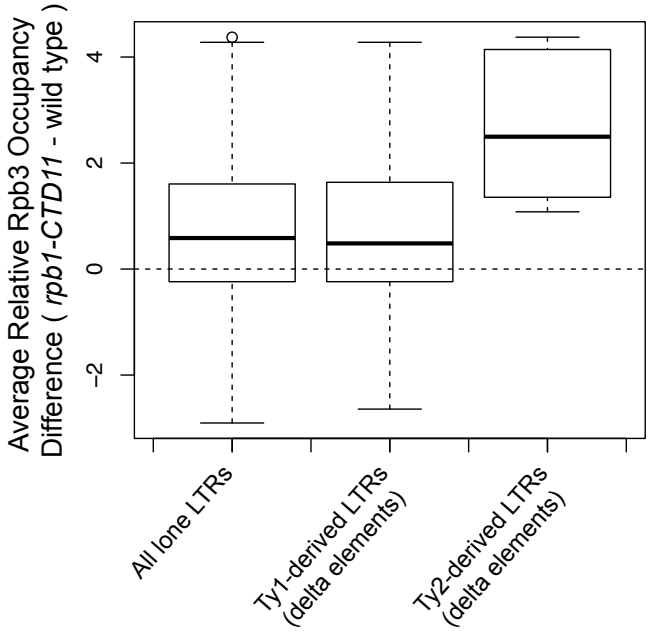



**Figure 3.2 Lone LTRs derived from Ty1 and Ty2 elements showed increased RNAPII levels in the *rpb1-CTD11* mutant when compared to wild type.**
Box plot showing differences in average RNAPII occupancy scores between the wild type and the *rpb1-CTD11* mutant strain at all, Ty1-, or Ty2-derived lone LTRs.

### 3.3.2 Truncation of the RNAPII-CTD resulted in altered occupancy of transcription-associated factors at a subset of retrotransposons

To mechanistically understand the effect of truncating the RNAPII-CTD at retrotransposons, we next determined if the increased binding coincided with occupancy changes of transcription- or chromatin-related factors at these loci. As such, we took advantage of our previously generated genome-wide occupancy maps of the general transcription factor TFIIB, the Mediator subunit Cdk8, the mRNA capping enzyme Cet1, the elongation factor Elf1, and the transcription elongation-associated chromatin mark H3K36me3 (Aristizabal et al. 2013). Truncation of the RNAPII-CTD resulted in significantly increased Cdk8, Cet1 and Elf1 occupancy at Ty1 retrotransposons, albeit with clearly different magnitudes (**Table 3.5**) (**Figure 3.3A**). In contrast,

truncating the RNAPII-CTD had no significant effect on the occupancy of TFIIB and H3K36me3 at Ty1 retrotransposons. Similar effects were observed at Ty2 elements, where Cdk8 and Cet1 occupancy showed significantly increased levels in the *rbp1-CTD11* mutant, and no changes were observed for TFIIB and H3K36me3 occupancy (**Figure 3.3B**). In contrast to the effect of truncating the RNAPII-CTD at Ty1 elements, Ty2 elements did not show any significant changes in Elf1 occupancy.

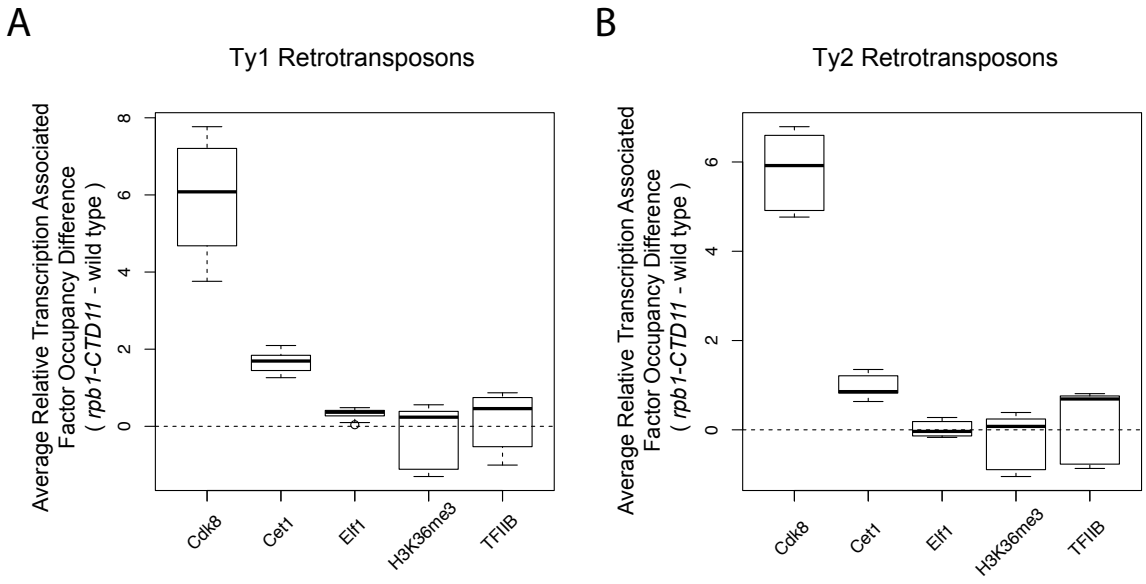



**Figure 3.3 Truncation of the RNAPII-CTD resulted in altered association of a subset of transcription related factors at retrotransposons.**
Comparison of MAT average occupancy scores at Ty1 (A) or Ty2 (B) retrotransposons for the Mediator subunit, Cdk8, the mRNA capping enzyme, Cet1, the elongation factor, Elf1, the transcription elongation-associated chromatin mark, H3K36me3, and the general transcription factor, TFIIB, under wild type and *rpb1-CTD11* conditions.

**Table 3.5 Paired t-test p values comparing the levels of transcription or chromatin-related factors in wild type vs *rpb1-CTD11* at Ty1 and Ty2 retrotransposons.**

| Element | Factor | One tail paired t-test p value comparing wild type vs *rpb1-CTD11* |
|---|---|---|
| Ty1 | **Cdk8** | **3.59e-21** |
| | **Cet1** | **1.42e-27** |
| | **Elf1** | **4.86e-17** |
| | H3K36me3 | 0.0199 |
| | TFIIB | 0.473 |
| Ty2 | **Cdk8** | **1.07e-11** |
| | **Cet1** | **3.93e-9** |
| | Elf1 | 0.672 |
| | H3K36me3 | 0.205 |
| | TFIIB | 0.505 |

### 3.3.3 Increased Ty1 mRNA levels and transposition rates caused by truncation of the RNAPII-CTD

The increased Rpb3 levels at Ty1 and Ty2 retrotransposons suggested concurrent changes in the mRNA levels of these elements. We designed a RT-qPCR based assay to quantitatively measure Ty mRNA levels, focusing on regions that were unique to all members of a single retrotransposon family. Mirroring the RNAPII occupancy data, mRNA levels of Ty1 retrotransposons were significantly increased in the *rpb1-CTD11* mutant compared to wild type (**Figure 3.4A**). Ty2 retrotransposons had a tendency for increased mRNA levels but this did not reach statistical significance (**Figure 3.4B**). The latter, was consistent with the weaker effect of truncating the RNAPII-CTD on RNAPII occupancy levels at Ty2 elements.

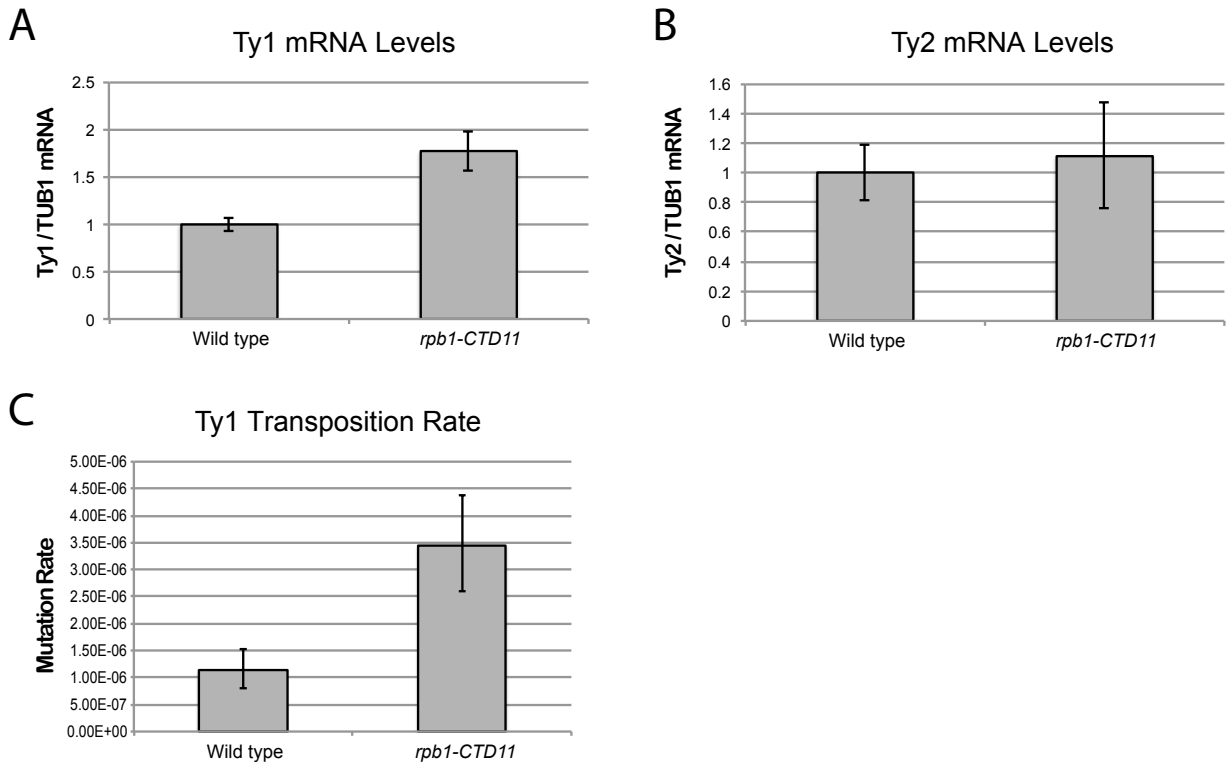



**Figure 3.4 Truncation of the RNAPII-CTD resulted in increased Ty1 mRNA levels and transposition rates.**
(A) The mRNA levels of the Ty1 elements were significantly increased in the *rpb1-CTD11* mutant compared to wild type. (B) RT-qPCR analysis of Ty2 mRNA levels in the *rpb1-CTD11* mutant compared to wild type. (C) Transposition rates for Ty1 were increased in *rpb1-CTD11* mutant compared to wild type. Error bars represent 95% confidence intervals. Transposition rates and confidence intervals were calculated using the Fluctuation AnaLysis CalculatOR (FALCOR) web tool.

Having established that upon truncation of the RNAPII-CTD Ty1 mRNA levels increased, we sought to determine if this had functional consequences on genome stability manifested by increased transposition rates. Using an established Ty1 cDNA-mediated mobility assay in living yeast cells we measured transposition rates in wild type and *rpb1-CTD11* mutants (Bryk et al.

2001, Scholes et al. 2001). Applying fluctuation analysis, we found that truncation of the RNAPII-CTD resulted in a 3-fold increase in Ty1 transposition rates, suggesting that genomic integrity was compromised upon loss of the RNAPII-CTD repeats (**Figure 3.4C**).

### 3.3.4 Loss of *CDK8* normalized the increased RNAPII and mRNA levels of Ty1 retrotransposons

Given that loss of *CDK8* normalizes a number of RNAPII-CTD truncation mutant phenotypes, and Cdk8 occupancy was increased at Ty1 and Ty2 retrotransposons in the *rpb1-CTD11* mutant (Aristizabal et al. 2013, Nonet and Young 1989), we hypothesized that it contributed to the increased RNAPII and mRNA levels at these retrotransposons upon truncation of the CTD. Focusing on Ty1 and Ty2 elements, we found that RNAPII levels were normalized in the *rpb1-CTD11 cdk8*Δ double mutant, as evidenced by average occupancy scores and average gene profiles (**Figure 3.5A and B**) (**Table 3.6**). Occupancy patterns at representative individual retrotransposons were also consistent with the normalizing effect of loss of *CDK8* (**Figure 3.5C**). Average RNAPII binding scores at Ty1 and Ty2 elements in the *rpb1-CTD11 cdk8Δ* mutant were significantly lower than the scores of the *rpb1-CTD11* mutant and were not statistically different from the scores of wild type cells. A similar effect was observed at Ty1- and Ty2-derived LTRs (**Figure 3.5D**). Most importantly, changes in RNAPII occupancy were mirrored by changes in mRNA levels at Ty1 retrotransposons, as loss of *CDK8* also normalized the elevated mRNA levels in the *rpb1-CTD11* mutant (**Figure 3.5E**).

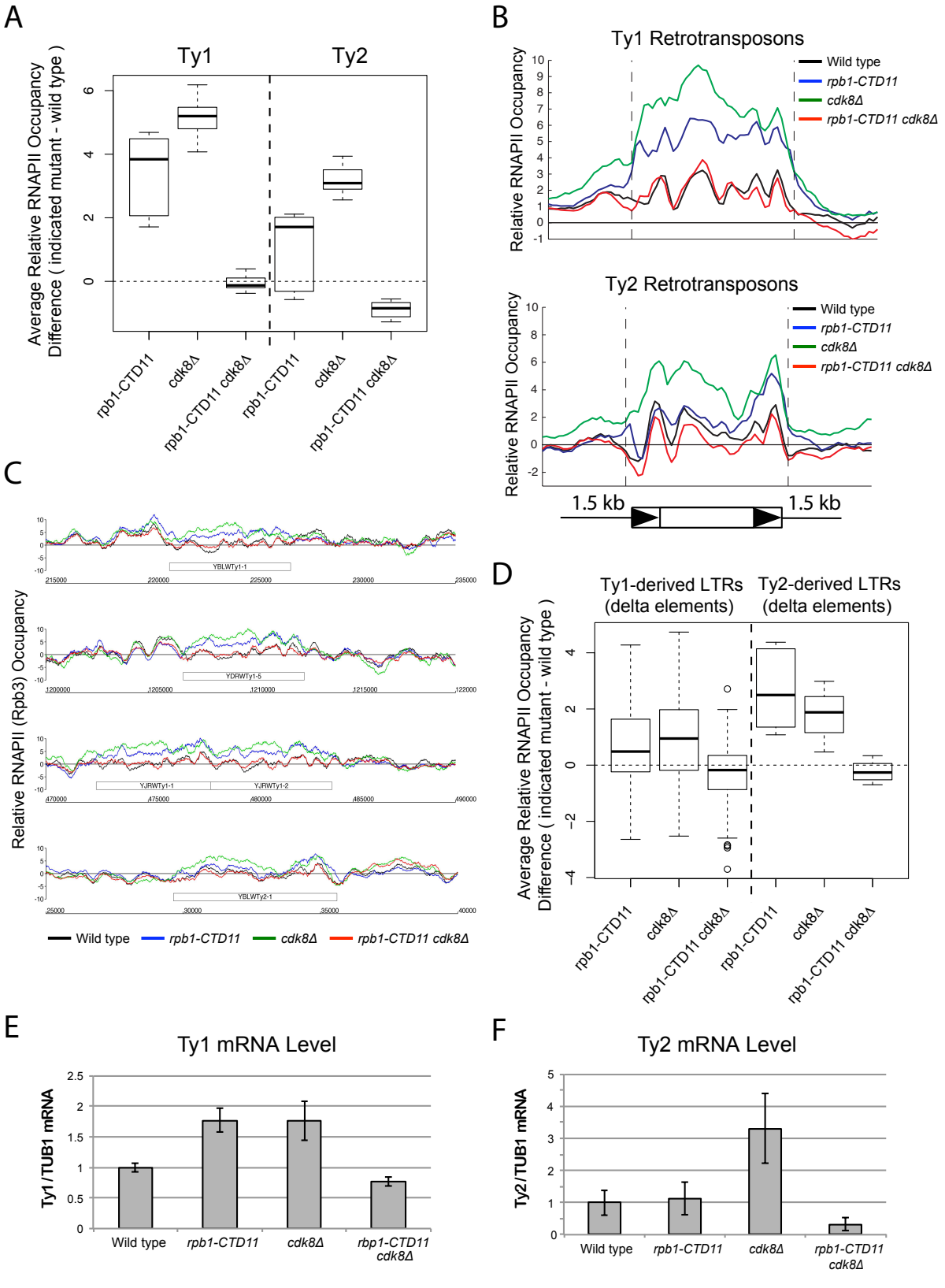

A
Ty1
Ty2
Average Relative RNAPII Occupancy
Difference ( indicated mutant - wild type )
rpb1-CTD11
cdk8Δ
rpb1-CTD11 cdk8Δ
B
Ty1 Retrotransposons
Ty2 Retrotransposons
Relative RNAPII Occupancy
Wild type
rpb1-CTD11
cdk8Δ
rpb1-CTD11 cdk8Δ
1.5 kb
1.5 kb
C
Relative RNAPII (Rpb3) Occupancy
YBLWTy1-1
YDRWTy1-5
YJRWTy1-1
YJRWTy1-2
YBLWTy2-1
Wild type
rpb1-CTD11
cdk8Δ
rpb1-CTD11 cdk8Δ
D
Ty1-derived LTRs
(delta elements)
Ty2-derived LTRs
(delta elements)
Average Relative RNAPII Occupancy
Difference ( indicated mutant - wild type )
E
Ty1 mRNA Level
Ty1/TUB1 mRNA
Wild type
rpb1-CTD11
cdk8Δ
rbp1-CTD11
cdk8Δ
F
Ty2 mRNA Level
Ty2/TUB1 mRNA
Wild type
rpb1-CTD11
cdk8Δ
rpb1-CTD11
cdk8Δ

**Figure 3.5 Loss of *CDK8* normalized the elevated RNAPII and mRNA levels at Ty1 and Ty2 retrotransposons.**
(A) Differential average RNAPII occupancy scores at retrotransposons revealed elevated levels at Ty1 and Ty2 retrotransposons in the single *cdk8Δ* mutant. In the *rpb1-CTD11* background, loss of *CDK8* resulted in normalized RNAPII levels at Ty1 and Ty2 retrotransposons. (B) Average gene profile of RNAPII occupancy at Ty1 (top) or Ty2 (bottom) retrotransposons showed normalized RNAPII levels upon loss of *CDK8*. (C) RNAPII chromosome plots revealed that loss of *CDK8* in the *rpb1-CTD11* background normalized the elevated RNAPII levels at representative retrotransposons. (D) Loss of *CDK8* normalized the elevated RNAPII levels at Ty1- and Ty2-derived LTRs. (E) RT-qPCR analysis of wild type, *rpb1-CTD11, cdk8Δ* and *rpb1-CTD11 cdk8Δ* revealed that loss of *CDK8* significantly normalized the elevated mRNA levels of Ty1 elements in the *rpb1-CTD11* background. (F) Ty2 mRNA levels were significantly elevated in the *cdk8Δ* mutant, an effect that was normalized when combined with an RNAPII-CTD truncation.

**Table 3.6 Paired t-test p values comparing the levels of RNAPII in wild type vs *rpb1-CTD11* at Ty1 and Ty2 retrotransposons and Ty-derived LTRs.**

| Element | Comparison | One tail paired t-test p value |
|---|---|---|
| Ty1 | Wild type vs *rpb1-CTD11* | 1.8e-15 |
| | Wild type vs *cdk8Δ* | 3.68e-32 |
| | Wild type vs *rpb1-CTD11 cdk8Δ* | 0.896 |
| | *rpb1-CTD11* vs *rpb1-CTD11 cdk8Δ* | 2.01e-17 |
| Ty1-derived LTRs | Wild type vs *rpb1-CTD11* | 1.34e-11 |
| | Wild type vs *cdk8Δ* | 9.58e-18 |
| | Wild type vs *rpb1-CTD11 cdk8Δ* | 1 |
| | *rpb1-CTD11* vs *rpb1-CTD11 cdk8Δ* | 3.08e-22 |
| Ty2 | Wild type vs *rpb1-CTD11* | 0.00424 |
| | Wild type vs *cdk8Δ* | 6.24e-12 |
| | Wild type vs *rpb1-CTD11 cdk8Δ* | 1 |
| | *rpb1-CTD11* vs *rpb1-CTD11 cdk8Δ* | 5.68e-06 |
| Ty2-derived LTRs | Wild type vs *rpb1-CTD11* | 5.84e-05 |
| | Wild type vs *cdk8Δ* | 4.63e-05 |
| | Wild type vs *rpb1-CTD11 cdk8Δ* | 0.946 |
| | *rpb1-CTD11* vs *rpb1-CTD11 cdk8Δ* | 2.3e-05 |

### 3.3.5 Suppression of the elevated RNAPII levels at Ty1 and Ty2 elements between the RNAPII-CTD mutant and *CDK8* deletion was reciprocal

Upon closer inspection of our RNAPII binding profiles, we noticed that loss of *CDK8* alone resulted in significantly elevated average RNAPII binding scores at Ty1 and Ty2

retrotransposons when compared to wild type (**Figure 3.5A and B**) (**Table 3.6**). Interestingly, the elevated average RNAPII levels present in the *cdk8Δ* and *rpb1-CTD11* single mutants were significantly reduced in the *rpb1-CTD11 cdk8Δ* double mutant. Having observed a significant and robust effect of loss of *CDK8* on RNAPII occupancy at Ty2 elements, we tested whether this coincided with significant expression changes. Ty2 mRNA levels were significantly increased in the *cdk8Δ* mutant when compared to wild type, but reverted back to baseline levels the *rpb1-CTD11 cdk8Δ* mutant (**Figure 3.5F**). This suggested that *CDK8*-dependent phenotypes could be normalized by functional alteration of the RNAPII-CTD.

### 3.3.6 Increased Ty1 mRNA alterations were in part due to changes to promoter activity mediated by Ste12 and Tec1

Finding that loss of *CDK8* normalized the elevated Ty1 gene expression phenotype of the *rpb1-CTD11* mutant suggested that the regulation occurred at the level of transcription initiation. To formally test this possibility, we focused on Ty1 retrotransposons and employed a plasmid based LacZ reporter strategy wherein we inserted more than 1 kb of promoter sequence for two representative Ty1 elements, YMLWTy1-2 and YJRWTy1-2, into a reporter plasmid. Both of these elements contain features found in many Ty promoters, including putative Ste12 and Tec1 transcription factor binding sites (**Figure 3.6A**) (Servant et al. 2008). The reporter assays showed significantly increased β-galactosidase activity for both elements in the *rpb1-CTD11* mutant compared to wild type, suggesting that their promoter sequences were sufficient to recapitulate the expression changes of the endogenous retrotransposons present in the *rbp1-CTD11* mutant (see left half of **Figure 3.6B and C**). Further expanding the mechanistic details of the RNAPII-CTD-dependent regulation on Ty1 elements, we found that removal of the binding sites

corresponding to the Tec1 consensus sequence affected expression of our reporter constructs (**Figure 3.6A, B and C**). Under wild type conditions, the YMLWTy1-2 promoter showed decreased activity when the Tec1 binding sites were deleted and levels remained low in the *rbp1-CTD11* mutant (**Figure 3.6B**). In contrast, the baseline expression of the YJRWTy1-2 promoter was not dependent on Tec1 binding sites however their removal sufficed to abolish the increased promoter activity caused by truncation of the RNAPII-CTD (**Figure 3.6C**).

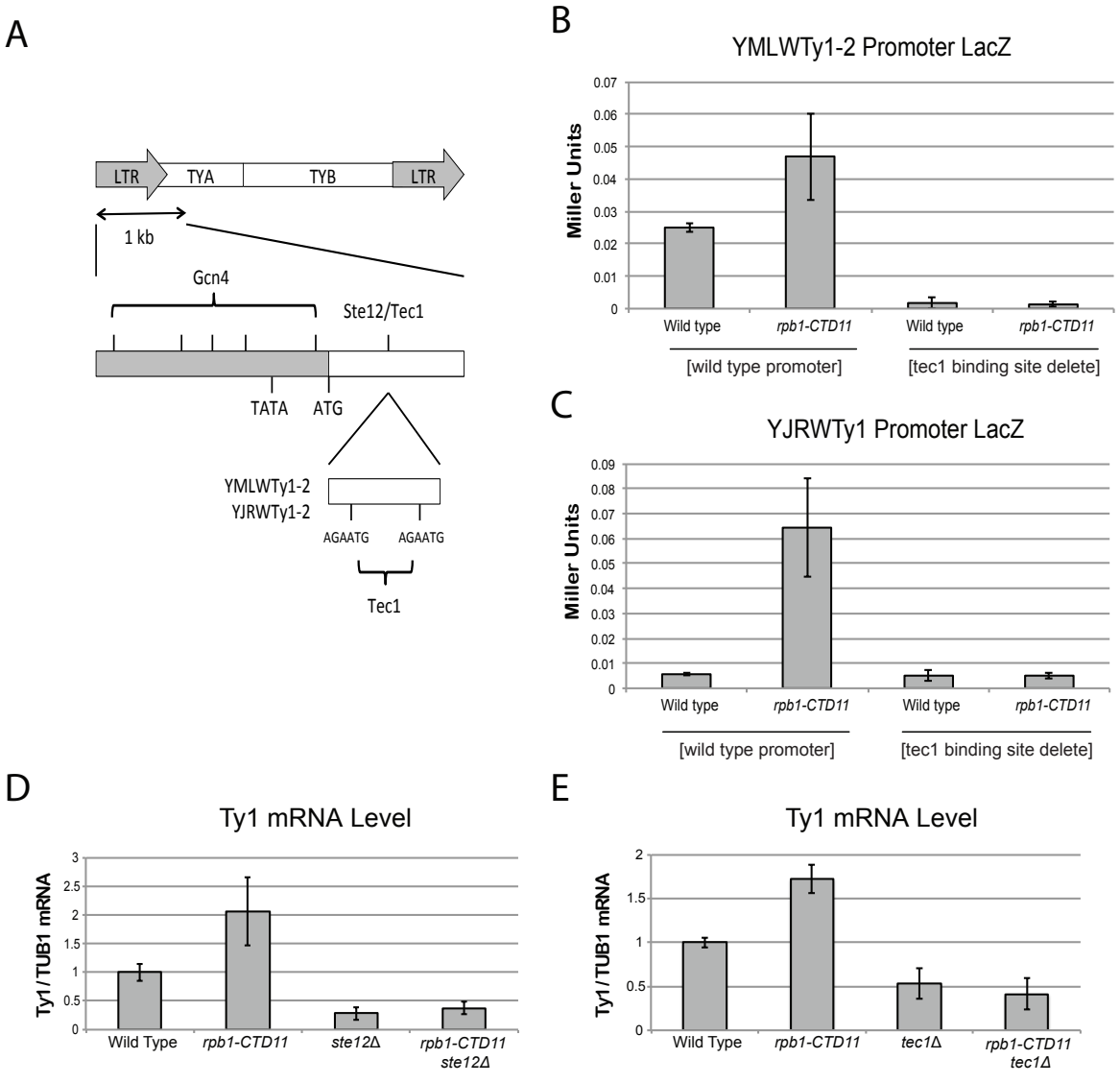


**Figure 3.6 The increased Ty1 gene expression levels observed in the *rpb1-CTD11* mutant were dependent on *TEC1* or *STE12*.**
(A) Adapted from Servant *et al* 2008. Schematic of an average Ty1 promoter with binding sites for Gcn4 and Ste12/Tec1 labeled. TYA and TYB are retrotransposon genes that encode for the coat protein, and reverse transcriptase, protease, integrase, and RNase H respectively. YMLWTy1-2 and YJRWTy1-2 contain two Tec1 binding sites downstream of the ATG start codon, a feature observed on some Ty1 elements. For YJRWTy1-2, 1321 bp of promoter sequence were cloned starting 518 bp upstream of the transposon ORF start and ending 804 bp downstream. For YMLWTy1-2, 1304 bp were cloned, starting 500 bp upstream of the transposons ORF start and ending 804 bp downstream. (B-C) Reporter assay for YMLWTy1-2 or YJRWTy1-2, with or without deletion of Tec1 binding sites. Tec1 binding sites were required for the increased promoter activity of Ty1 reporter constructs upon truncation of the RNAPII-CTD. (D and E) RT-qPCR analysis of wild type, *rpb1-CTD11, ste12Δ* and *rpb1-CTD11 ste12Δ* or wild type, *rpb1-CTD11, tec1Δ* and *rpb1-CTD11 tec1Δ* revealed normalized Ty1 mRNA levels in the *rpb1-CTD11* background upon loss of *STE12* or *TEC1*.

Given that the increased expression of Ty1 elements resulted in part from Tec1 binding site-dependent alterations in transcription initiation, we focused on the effect of loss of *TEC1* or its regulatory partner *STE12* on endogenous Ty1 mRNA levels. The connection to both Ste12 and Tec1 as regulators of Ty1 expression was particularly intriguing given that they are also directly (Ste12) and indirectly (Tec1) regulated by Cdk8 (Nelson et al. 2003, Raithatha et al. 2011). Consistent with their known roles in Ty1 expression, *ste12Δ* or *tec1Δ* single mutants had reduced Ty1 mRNA levels compared to wild type (**Figure 3.6D and E**). More importantly, loss of *TEC1* or *STE12* reduced the elevated Ty1 mRNA levels in the *rpb1-CTD11* mutant. One trivial explanation for the increased Ty1 mRNA levels in the *rpb1-CTD11* mutant could be that the protein levels of Ste12 or Tec1 were increased in the *rpb1-CTD11* mutant, similar to what we observed for Rpn4 (Aristizabal et al. 2013). However this is unlikely given that neither their mRNA nor proteins levels were altered in the *rpb1-CTD11* mutant compared to wild type (data not shown). Additionally, we explored the relationship between the RNAPII-CTD and a different mode of Ty1 regulation, namely its repression by the a1-alpha2 mating repressor pair (Errede et

al. 1980). a1-alpha2 repression is specific to diploid cells, thus we generated diploid strains homozygous or heterozygous for the truncated RNAPII-CTD allele and observed unaltered Ty1 mRNA levels in strains containing *rpb1-CTD11* alleles when compared to a wild type diploid strain (data not shown). Overall, our result revealed that the effect of the a1-alpha2 mating repressor pair on Ty1 mRNA levels could not be overcome by truncation of the RNAPII-CTD, indicating a specific role for Ste12 and Tec1 in the *rpb1-CTD11*-mediated effect on Ty1 gene expression regulation.

### 3.3.7 A broad role for *TEC1* and *STE12* in the regulatory circuitry of the RNAPII-CTD

Having established that loss of *TEC1* and *STE12* normalized the elevated expression levels of retrotransposons caused by loss of the RNAPII-CTD, we tested whether this relationship extended more broadly to CTD-dependent phenotypes. Focusing on four representative protein-coding genes whose expression level is elevated in the *rpb1-CTD11* mutant, we found that further loss of *TEC1* showed a trend towards reduced mRNA levels, although with the exception of YCR061W the effects tended to be small and not statistically significant (**Figure 3.7**) (Aristizabal et al. 2013). Perhaps more significantly, the suppression of phenotypes associated with CTD truncations by loss of *STE12* or *TEC1* extended to restoration of growth defects. Specifically, deletion of *STE12* or *TEC1* in the *rpb1-CTD11* background robustly normalized the slow growth phenotype of *rpb1-CTD11* mutants when grown at 30 and 16 °C (**Figure 3.8A and B**). However, in contrast to *STE12*, loss of *TEC1* also suppressed the growth defects of *rpb1-CTD11* mutants when grown at 37 °C and under hydroxyurea and formamide conditions, suggesting that *TEC1* is a more robust suppressor of *rpb1-CTD11* growth phenotypes than *STE12*. Overall, the suppression pattern observed for loss of *TEC1* was similar to that caused by

loss of *CDK8*. Therefore, we used genetic analysis to test whether this was through independent or overlapping pathways. The strength and condition spectrum of suppression in the *rpb1-CTD11 cdk8Δ tec1Δ* triple mutant was similar to that of the *rpb1-CTD11 cdk8Δ* and *rpb1-CTD11 tec1Δ* double mutants respectively. This suggested that *CDK8* and *TEC1* suppressed RNAPII-CTD truncation phenotypes in part by functioning in the same pathway (**Figure 3.8C**).

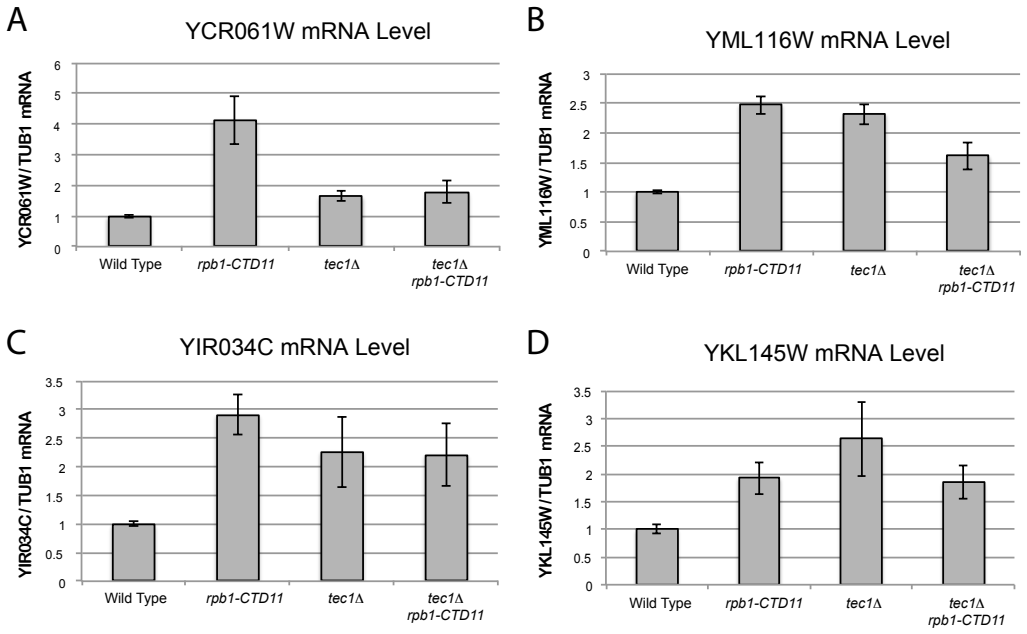



**Figure 3.7 Loss of *TEC1* suppressed additional gene expression alterations observed in the *rpb1-CTD11* mutant.**
(A-D) RT-qPCR analysis of YKL145W, YIR034C, YML116W, and YCR061W mRNA levels in wild type, *rpb1-CTD11, tec1Δ* and *rpb1-CTD11 tec1Δ* mutants.

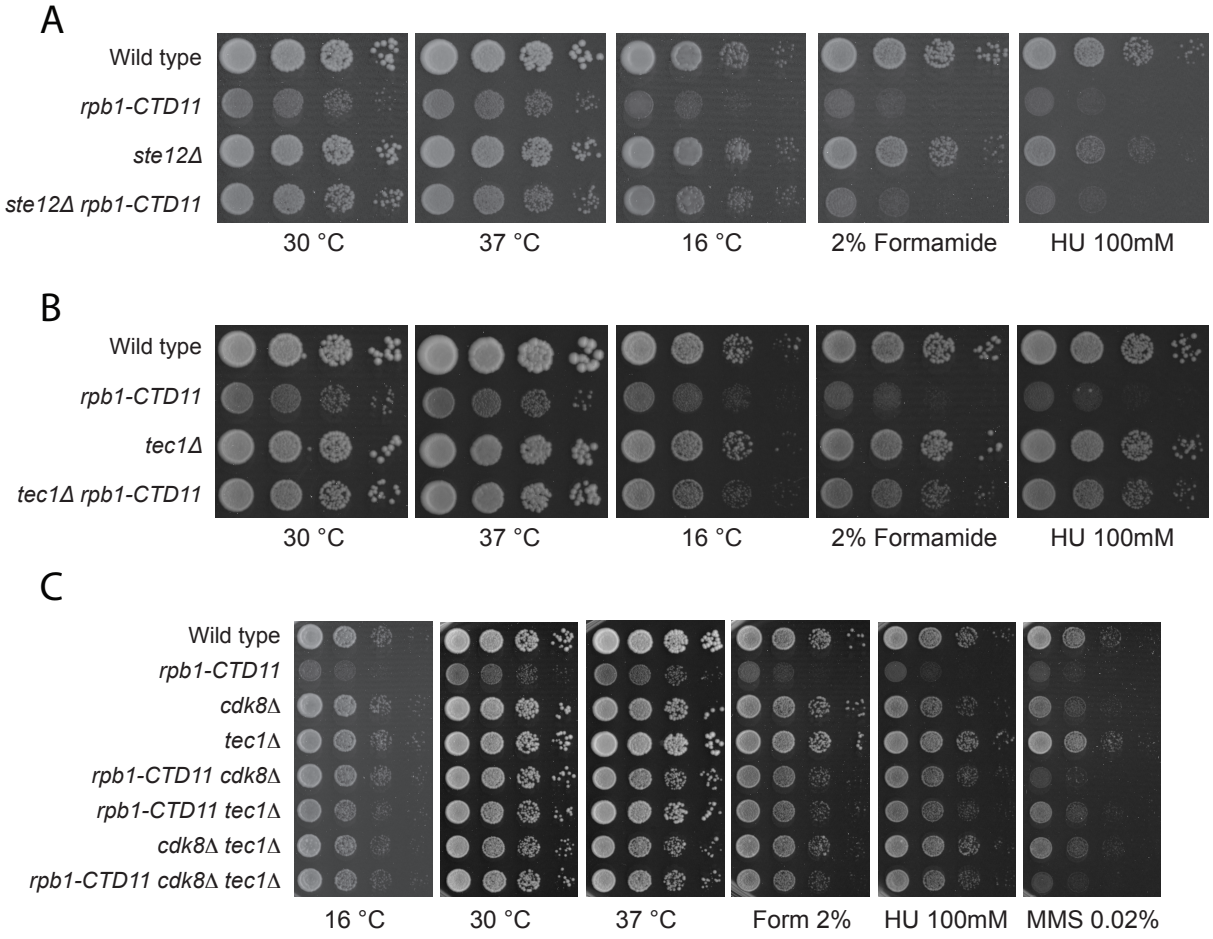



**Figure 3.8 Loss of *STE12* or *TEC1* suppressed growth defects associated with *rpb1-CTD11* and the latter functioned in the same pathway as *CDK8*.**
(A-B) Sensitivity of the *rpb1-CTD11* mutant to growth under normal and low temperature conditions was suppressed by deletion of *STE12* or *TEC1*. Loss of *TEC1* also suppressed the growth defects of the *rpb1-CTD11* mutant upon exposure to high temperatures, formamide and hydroxyurea. Ten-fold serial dilutions of the indicated mutants were plated on YPD media at 16, 30 and 37 °C or media containing the indicated concentrations of hydroxyurea or formamide. (C) Loss of *TEC1* and *CDK8* suppressed the sensitivity of the *rpb1-CTD11* mutant to growth under low and high temperatures and upon exposure to formamide and hydroxyurea. Ten-fold serial dilutions of the indicated mutants were plated and incubated on YPD media at 16, 30 and 37 °C and media containing the indicated concentrations of hydroxyurea or formamide.

## 3.4 Discussion

Expanding beyond the traditional role of RNAPII in the transcription of protein coding genes, we characterized its involvement in the biology of retrotransposons. We propose that by virtue of limiting retrotransposon gene expression, the RNAPII-CTD plays an important role in the maintenance of genomic integrity. Several lines of evidence pointed to a direct role for the RNAPII-CTD in restricting retrotransposon mobility and gene expression. First and foremost, deletion of the RNAPII-CTD unmasked this inhibitory role as it caused a significant increase in the rate of transposition of Ty1 elements. Second, higher mRNA and RNAPII occupancy levels underpinned this effect across different families of retrotransposons. Third, Cdk8 and Tec1 regulated the high RNAPII occupancy and mRNA expression caused by shortening the RNAPII-CTD, at least in part through promoter-mediated events. Furthermore, the close regulatory circuitry between the RNAPII-CTD, Tec1 and Cdk8 was not limited to retrotranposon expression, as loss of *TEC1* suppressed additional CTD truncation phenotypes in a manner similar to loss of *CDK8*.

Our key finding of the RNAPII-CTD inhibiting retrotransposition is consistent with an increasing appreciation of a broader involvement of RNAPII and its C-terminus in diverse aspects related to the maintenance of genome integrity. For example, yeast strains with shortened RNAPII-CTDs are sensitive to several DNA damaging drugs, including the DNA replication inhibitor hydroxyurea (Wong and Ingles 2001). Furthermore, strains with critically short CTDs spontaneously revert to RNAPIIs with increased CTD lengths, suggesting enhanced facility for genomic rearrangements (Nonet and Young 1989). A role for the RNAPII-CTD is

also evident in the critical process of transcription coupled repair. This process preferentially monitors the integrity of biologically relevant loci that if damaged result in RNAPII stalling, a signal for DNA repair (Wilson et al. 2013). Repair is attempted first by the nucleotide excision repair pathway, and if unsuccessful, by other repair mechanisms which first require poly-ubiquitination- and proteasome-dependent removal of RNAPII from the template. The latter is dependent on the phosphorylation status of the RNAPII-CTD, which regulates the recruitment and activity of key factors involved in RNAPII ubiquitination such as the E3 ubiquitin ligase, Rsp5 (Somesh et al. 2007).

A 3-fold increase in Ty1 mobility in strains with truncated RNAPII-CTD was comparable to other previously reported Ty1 regulators, although we note that the effect of the *rbp1-CTD11* mutant was in comparison lower on the Ty1 mobility spectrum (Scholes et al. 2001). Nonetheless, the increase in transposition was most likely caused by increased Ty1 mRNA levels due to exacerbated transcriptional initiation. Consistent with this, the increased RNAPII levels at Ty elements and concomitant mRNA increases in these strains were normalized upon loss of transcription factors Tec1 or Ste12, or the mediator subunit Cdk8. Recapitulation of the increased expression and its dependency on Tec1 in a Ty1 promoter reporter assay provided further support of this mechanism. Importantly, the increased levels of RNAPII at lone LTRs not only strongly suggested that the core promoter sequences were sufficient for the initial recruitment of RNAPII with shortened CTDs, but also provided some nuanced insights into the mechanism of activation and the likely role of Tec1. Specifically, lone LTR genomic loci lack functional Tec1 binding sites which tend to be located downstream of the ATG translation start codon. Reconciling this with the requirement for Tec1 in mediating the increased expression

level caused by shortening the RNAPII-CTD suggested that Tec1 acted after the assembly of the transcription complex on the core promoter sequences. Collectively, these data thus revealed multiple layers of control at Ty1 promoters where sequences upstream of the ATG start codon were sufficient for RNAPII recruitment but additional regulatory layers down-stream were important for full transcriptional activation. However, it is likely that this particular model is primarily relevant for Ty1 elements, as Ty2 elements are not dependent on *TEC1* for expression (Laloux et al. 1990).

The effect of the RNAPII-CTD on Ty1 gene expression was reminiscent of previously reported observations on a set of Rpn4-regulated genes (Aristizabal et al. 2013). Specifically, under normal growth conditions both retrotransposons and Rpn4-regulated genes had *CDK8*-dependent increases in RNAPII and mRNA levels in the *rpb1-CTD11* mutant, which were likely mediated by alterations to transcription initiation, the latter in part due to increased protein levels of Rpn4. However, despite the similarities, distinct roles of Cdk8 suggested different transcriptional regulatory processes. Specifically, while Cdk8 was normally present at Rpn4-regulated genes, its loss did not change their expression level. In contrast, at Ty1 elements, Cdk8 alone played a role in their regulation as evidenced by increased RNAPII and mRNA levels in the *CDK8* deletion mutant. Given the well-documented role for Cdk8 in the regulation of Ste12 levels, it is tempting to speculate that in the *cdk8Δ* mutant, stabilization of Ste12 levels indirectly stimulated Ty1 gene expression (Nelson et al. 2003). However, our data suggested that this was unlikely to reflect the full spectrum of Cdk8's role on Ty1 gene expression regulation. Presumably, Ste12 levels would still be increased in the *rbp1-CTD11 cdk8Δ* double mutant, yet RNAPII and mRNA levels at Ty1 elements in this strain were not increased compared to wild type. Instead, the

increased occupancy of Cdk8 at Ty1 elements upon truncation of the RNAPII-CTD suggested that concomitant recruitment of both factors was important for the increased transcriptional activity. One possible model consistent with the data would postulate that Cdk8 stimulated transcription by enhancing RNAPII promoter release via CTD phosphorylation, as reported previously for some protein-coding genes in mammals (Belakavadi and Fondell 2010, Gold and Rice 1998). This model would also be consistent with the reciprocal suppression between Cdk8 and the RNAPII-CTD we observed at Ty1 and Ty2 elements. Focusing on the latter, the RNAPII occupancy results suggested a role for the RNAPII-CTD in their regulation, the details of which remain to be fully characterized.

It is unclear to what extent the various transcriptional regulatory pathways engaged in the cellular response to truncation of the RNAPII-CTD are linked. Retrotransposons can alter the transcriptional regulatory landscape of adjacent genes and thus alterations to their regulation could underlie some of the observed transcriptional defects at other genes (Lesage and Todeschini 2005). However, this effect is limited to genes near retrotransposons and we observed no correlation between being in the vicinity of a retrotransposon and having altered mRNA levels in the *rbp1-CTD11* mutant. Suggestive of a different type of connection, we observed that truncation of the RNAPII-CTD resulted in *STE12*-dependent increases in Ty1 mRNA levels and decreased expression of protein-coding genes primarily regulated by Ste12 (Aristizabal et al. 2013). Given that Ste12 levels were unaltered in the *rbp1-CTD11* mutant, one possibility is that the increased transcriptional output at Ty1 elements reduced the cellular pool of Ste12 protein necessary to drive the expression of other genes. More work beyond the scope of this investigation will be necessary to illuminate the degree of connectivity between the

distinct transcriptional programs found in the *rpb1-CTD11* mutant strains, and their detailed mechanistic underpinning. Our finding that *TEC1*, and to a lesser extent *STE12*, acted in conjunction with *CDK8* as new *SRB* genes by suppressing growth-related phenotypes associated with loss of the RNAPII-CTD offers further evidence for a broad involvement of this network in cellular function.

# Chapter 4: High-throughput Genetic and Gene Expression Analysis of *FCP1* Mutants Supported A Broader Role in Transcription Regulation

## 4.1 Introduction

RNA polymerase II (RNAPII), the enzyme responsible for the transcription of most protein coding genes, contains a unique C-terminal domain (CTD) composed of heptapeptide repeats following a $Y_1S_2P_3T_4S_5P_6S_7$ consensus sequences (Allison et al. 1985, Corden et al. 1985). The CTD functions as a recruiting platform for regulatory, RNA processing and chromatin remodeling factors, a role that is dependent on its differential phosphorylation (Heidemann et al. 2012, Zhang et al. 2012b). Each transcriptional stage is characterized by a unique CTD phosphorylation signature, beginning with an unmodified RNAPII that is recruited to promoters and interacts with components of the pre-initiation complex (Bataille et al. 2012, Kim et al. 2010, Lu et al. 1991, Mayer et al. 2010, Tietjen et al. 2010). Following, the CTD is phosphorylated at $S_5$ residues by the general transcription factor H (TFIIH) kinase, Kin28, a modification that signals the release of RNAPII from the promoter and enhances the recruitment of capping enzymes (Akhtar et al. 2009, Jeronimo and Robert 2014, Kim et al. 2009, Max et al. 2007, McCracken et al. 1997a, Wong et al. 2014). Next, $S_5$ residues are dephosphorylated by the RNAPII-CTD phosphatase Rtr1 and this aids in the transition from initiation to elongation (Mosley et al. 2009). During elongation, the CTD is primarily phosphorylated at $S_2$ and $Y_1$ residues and these enhance the recruitment of elongation factors and splicing factors (Bataille et al. 2012, Kim et al. 2010, Mayer et al. 2010, Mayer et al. 2012, Tietjen et al. 2010). Along the

length of genes, $S_2$ phosphorylation levels are fine-tuned by the opposing action of the CTD kinase, Ctk1, and phosphatase, Fcp1 (Cho et al. 2001). The elongating RNAPII is also phosphorylated at $T_4$ residues and this mark is removed by Fcp1, and deposited in mammals by Polo-like Kinase 3 (Mayer et al. 2012, Allepuz-Fuster et al. 2014, Hintermair et al. 2012, Hsin et al. 2014). Transcription termination is mediated by $Y_1$ dephosphorylation by Glc7 and or Rtr1prior to RNAPII reaching the polyadenylation site (Hsu et al. 2014, Mayer et al. 2012, Schreieck et al. 2014). The differential levels of $S_2$ and $Y_1$ phosphorylation distinguishes the elongation and termination forms of RNAPII and results in the preferential recruitment of termination factors to the 3' end of genes. Finally, all phosphorylation marks are removed to regenerate a RNAPII molecule capable of starting another round of transcription. Fcp1 and Ssu72 play key roles in RNAPII recycling, although it is likely that other RNAPII phosphatases also contribute (Bataille et al. 2012, Cho et al. 1999, Krishnamurthy et al. 2004).

Fcp1 is a highly conserved RNAPII-CTD phosphatase that is essential for viability in a number of species (Archambault et al. 1997, Kimura et al. 2002, Kobor et al. 1999, Son and Osmani 2009, Tombacz et al. 2009). It contains a catalytic FCP homology region (FCPH) and a single BRCA1 C terminus (BRCT) domain (Kobor et al. 1999). The latter likely mediates direct contacts between the RNAPII-CTD and Fcp1 (Ghosh et al. 2008, Yu et al. 2003). Fcp1 is found along the length of most genes, yet is most abundant at the 3' end (Zhang et al. 2012a). Here, it dephosphorylates $S_2$ residues of the RNAPII-CTD, making it important for elongation and recycling (Cho et al. 2001, Cho et al. 1999, Hausmann and Shuman 2002). As such, shifting conditional *FCP1* mutants to non-permissive temperatures ceases transcription of the majority of genes and results in accumulation of hyperphosphorylated RNAPII in coding regions (Bataille et

al. 2012, Cho et al. 2001, Kobor et al. 1999). Fcp1 physically interacts with a number of transcription-related factors, which are thought to aid in the recruitment of Fcp1 to transcribed genes, and might also participate in the regulation of Fcp1 activity. Fcp1 interacts with the RNAPII-CTD, the RNAPII subunit Rpb4, and the general transcription factors TFIIF and TFIIB (Chambers et al. 1995, Kimura et al. 2002, Kobor et al. 2000, Suh et al. 2005). Furthermore, there is evidence that its recruitment to transcribed regions is dependent on the transcription elongation factor Sub1, and that its CTD phosphatase activity is modified by TFIIF, TFIIB, and the CTD proline isomerase, Ess1 (Calvo and Manley 2005, Kobor et al. 2000, Kops et al. 2002). However, despite these interactions and Fcp1 being the first RNAPII-CTD phosphatase identified, the mechanistic details of how it is recruited to transcribed regions and how its activity is regulated remains largely unknown.

In *S. cerevisiae,* the only well characterized substrate of Fcp1 is the RNAPII-CTD. However, given that Fcp1 appears earlier in evolution than the RNAPII-CTD and has been implicated in other biological functions, it is likely that Fcp1 has additional targets in the cell (Fath et al. 2004, Guo and Stiller 2005). In yeast, Fcp1 associates with the RNAPI transcription machinery and is required for efficient rRNA transcription (Fath et al. 2004). Similarly, in mammalian cells Fcp1 dephosphorylates the RNAPI transcription initiation factor TIF-IA, a modification that is important for RNAPI recycling and recruitment to promoters (Bierhoff et al. 2008). Furthermore, in mammals Fcp1 participates in mitotic exit in part by dephosphorylating Wee1, Cdc20, USP44 and Ensa (Hegarat et al. 2014, Visconti et al. 2012).

To understand the extent of Fcp1 function in the cell, a set of *FCP1* C-terminal truncation mutants were examined using a variety of high-throughput methods. Consistent with an intimate relationship between *FCP1* and the RNAPII-CTD, strains carrying *FCP1* or *RPB1* mutant alleles resulted in similar phenotypes as demonstrated by gene expression and genetic interaction profiles. Furthermore, combining a *FCP1* mutant with a *RPB1* C-terminal truncation resulted in lethality. *FCP1* mutants were also similar to *RPB1* C-terminal truncation mutants, in that both were suppressed by loss of *CDK8* although with clearly different magnitudes. However, the mechanism of suppression of *RPB1* C-terminal truncation mutants by loss of *CDK8* likely differed from that of suppression of *FCP1* mutants by loss of *CDK8.* Despite the similarities, clear phenotypic differences were also apparent. Most notably, in contrast to the *rpb1-CTD11* mutant, mutating *FCP1* specifically altered the expression of genes regulated by transcription factors that are known phospho-proteins. Furthermore, differential genetic interactions were observed between *FCP1* and RNAPII-CTD truncations when combined with genes encoding prefoldin complex subunits.

## 4.2 Materials and methods

### 4.2.1 Yeast strains

Strains are listed in **Table 4.1**. Complete or partial gene deletions or integration of a 3XFLAG or VSV tag were achieved via the one-step gene replacement method (Longtine et al. 1998). All double mutant strains were generated via mating and tetrad dissection.

**Table 4.1 Strains used in this study**

| Genotype | Relevant Mutation |
|---|---|
| *Matα his3Δ1 leu2Δ0 ura3Δ0* | *fcp1-594 Flag* |
| *Matα his3Δ1 leu2Δ0 ura3Δ0* | *fcp1-609 Flag* |
| *Matα his3Δ1 leu2Δ0 ura3Δ0* | *fcp1-666 Flag* |
| *Matα his3Δ1 leu2Δ0 ura3Δ0* | *fcp1-713 Flag* |
| *Matα his3Δ1 leu2Δ0 ura3Δ0* | *FCP1-WT Flag* |
| *Matα his3Δ1 leu2Δ0 ura3Δ0* | *fcp1-594 Flag cdk8Δ* |
| *Matα his3Δ1 leu2Δ0 ura3Δ0* | *fcp1-609 Flag cdk8Δ* |
| *Matα his3Δ1 leu2Δ0 ura3Δ0* | *fcp1-666 Flag cdk8Δ* |
| *Matα his3Δ1 leu2Δ0 LYS2+ met15Δ0 ura3Δ0 Δcan1::MATaPr-HIS3 Δlyp1::MATαPr-LEU2* | *fcp1-594 Flag* |
| *Matα his3Δ1 leu2Δ0 LYS2+ met15Δ0 ura3Δ0 Δcan1::MATaPr-HIS3 Δlyp1::MATαPr-LEU2* | *fcp1-609 Flag* |
| *Matα his3Δ1 leu2Δ0 LYS2+ met15Δ0 ura3Δ0 Δcan1::MATaPr-HIS3 Δlyp1::MATαPr-LEU2* | *fcp1-666 Flag* |
| *Matα his3Δ1 leu2Δ0 LYS2+ met15Δ0 ura3Δ0 Δcan1::MATaPr-HIS3 Δlyp1::MATαPr-LEU2* | *fcp1-713 Flag* |
| *Matα his3Δ1 leu2Δ0 LYS2+ met15Δ0 ura3Δ0 Δcan1::MATaPr-HIS3 Δlyp1::MATαPr-LEU2* | *FCP1-WT Flag* |
| *MatA/α ADE2/ade2-1 can1-100/can1-100 his3-11/his3-11 leu2-3,112/leu2-3,112 trp1-1/trp1-1 ura3-1/ura3-1 LYS2/lys2Δ* | *fcp1-594 rpb1-CTD11* |

### 4.2.2 Growth assays

Overnight cultures grown on YPD were diluted to 0.5 OD 600, 10-fold serially diluted and spotted onto YPD plates with or without the indicated amounts of hydroxyurea (HU), methyl methanesulfonate (MMS) (Sigma) or formamide (Sigma). Plates were incubated at the indicated temperatures for 2-4 days.

### 4.2.3 Microarrays experiments and analysis

Microarrays were performed in duplicate as previously described (Lenstra et al. 2011, van Wageningen et al. 2010). Cultures were grown with a 24-well plate incubator/reader. Spiked-in controls were used to determine global changes in mRNA levels. As no such changes were detected, the expression profiles were normalized to total mRNA levels, a more reproducible measure. Differentially expressed genes were determined by p value < 0.01 or an absolute fold

change > 1.7 compared to wild type (Kemmeren et al. 2014). In the *fcp1-594 cdk8Δ* or *fcp1-609 cdk8Δ* double mutant, suppressed genes were determined as those having absolute fold changes < 1.1 and p values > 0.01 compared to wild type. The Yeast Promoter Atlas database was used for transcription factor enrichment by performing a Hypergeometric test with Bonferroni correction (p value 0.05) (Chang et al. 2011). Database for Annotation, Visualization and Integrated Discovery (DAVID) was used for Gene Ontology enrichment. Only biological process and pathway information were considered and significance was determined as Benjamini corrected p value < 0.05 (Carlson et al. , Falcon and Gentleman 2007). The YeastKID database was used do identify phosphorylated transcription factors. These were considered phosphorylated if they had a score greater than 6.4 in the database (Sharifpoor et al. 2011).

### 4.2.4 Epistasis miniarray profiling (E-MAP)

E-MAP screens were performed and normalized as described previously (Collins et al. 2010). RNAPII-CTD truncation mutants were crossed, using a Singer robot, to a library of 1536 mutants (Collins et al. 2010) covering a number of categories, including RNA processing, kinases/phosphatases and chromatin biology. Mutants are either deletions or decrease abundance by mRNA perturbations (DAmP) alleles. Diploid selection, sporulation, haploid selection, and double mutant selection steps were performed by replicate plating on the appropriate selective media. All strains were screened in triplicate and for each replicate double mutant colony sizes were determined from three technical replicates. Colony size was used to determine a quantitative S-score, which is a modified T-test that compares the observed double mutant growth rate to an expected growth rate based on the average colony size across an entire plate.

Significant negative genetic interactions are defined as interactions with scores ≤ −2.5, while significant positive interactions are those with scores ≥ 2.0.

## 4.3 Results

### 4.3.1 *FCP1* C-terminal truncation mutants resulted in few transcriptional alterations

To understand the role of Fcp1 in RNAPII-dependent transcription, a series of previously described and novel *FCP1* C-terminally truncated mutants were subject to gene expression microarrays (Kobor et al. 2000). All *FCP1* alleles were created by incorporation of a 3XFLAG tag after the nucleotides encoding for amino acid number 594, 609, 666 and 713 (*fcp1-594*, *fcp1-609*, *fcp1-666* and *fcp1-713* respectively) at the endogenous *FCP1* locus, and were compared to a strain containing a FLAG tag replacing the natural *FCP1* stop codon (*FCP1-WT*) (**Figure 4.1A**). Of these, only *fcp1-609* and *fcp1-594* showed fitness defects when grown at 30 or 37 °C or when exposed to hydroxyurea (HU), formamide, or methyl methanesulfonate (MMS) (**Figure 4.1B**). In contrast to two previously characterized *FCP1* mutants, *fcp1-1* and *fcp1-2* (encoding Fcp1 R250A/R251A and Fcp1 L117A/L181A/H187A respectively), which result in a general shutdown of transcription when shifted to the non-permissive temperature, the milder *FCP1* C-terminal truncation mutants resulted in relatively few transcriptional alterations, thus revealing the set of genes most dependent on the Fcp1 C-terminus for normal expression (**Figure 4.2A**) (Kobor et al. 1999). Surprisingly, the gene expression profiles showed that the *FCP1* mutant resulting in the most gene expression alterations was *fcp1-609* and not the shortest *fcp1-594* (**Figure 4.2B and C**). Furthermore, although both mutants had similar gene expression profiles

(Pearson's rho 0.90), not all genes whose mRNA levels were significantly affected in the *fcp1-609* mutant were also affected in the *fcp1-594* mutant and vice versa. Thus, to increase our confidence on the set of *FCP1*-dependent genes, only genes that were altered in both mutants were considered. This resulted in 68 genes with decreased mRNA levels and 23 genes with increased mRNA levels. Gene ontology (GO) enrichment analysis revealed that these sets of genes were involved in different biological pathways. The genes whose mRNA levels increased in the *FCP1* mutants were involved in lysine metabolism and biosynthesis while the genes whose mRNA levels decreased were enriched for GO categories related to metabolic and mating pathways. The latter were also regulated by a subset of transcription factors, which included Bas1, Cad1 and Hot1, while no transcription factors met the enrichment criteria when the genes whose mRNA levels increase were considered (**Figure 4.2D**). The enrichment of these transcription factors was particularly interesting, given that a significant proportion of them are known phospho-proteins (Hypergeometric test p value 0.03355).

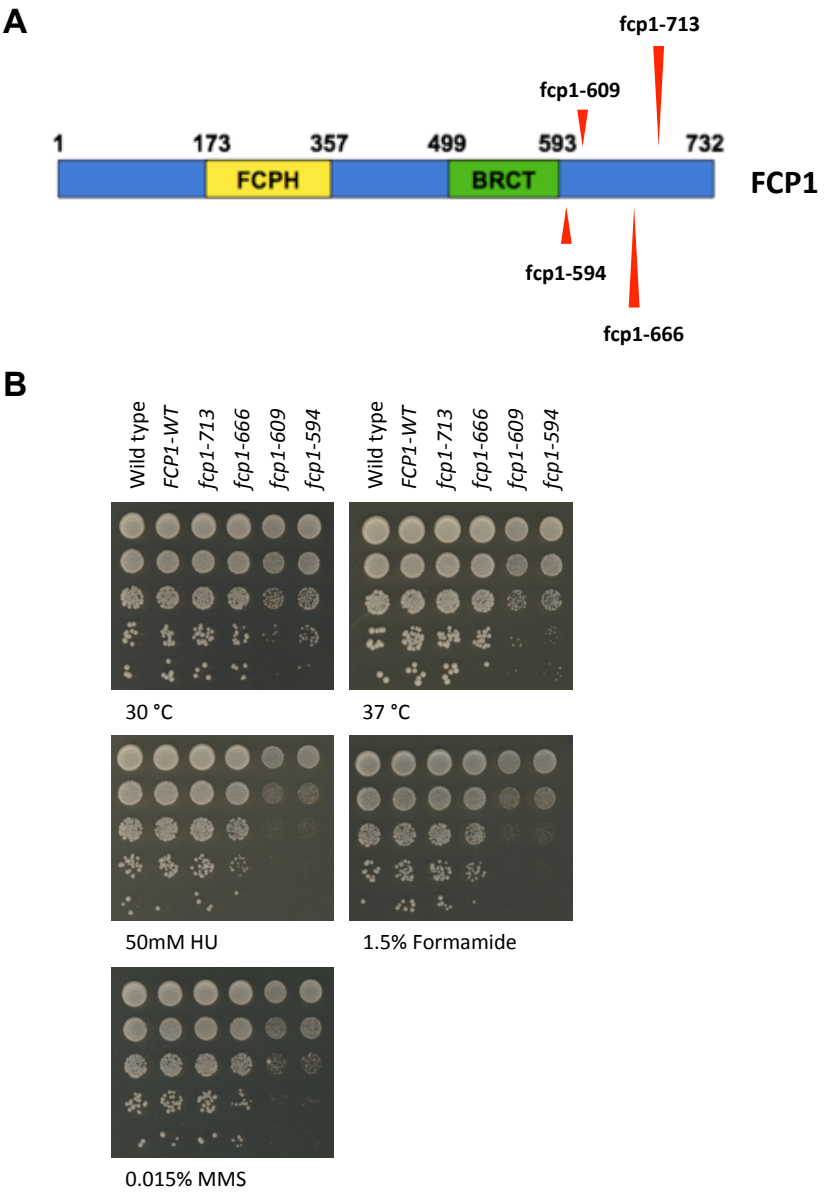
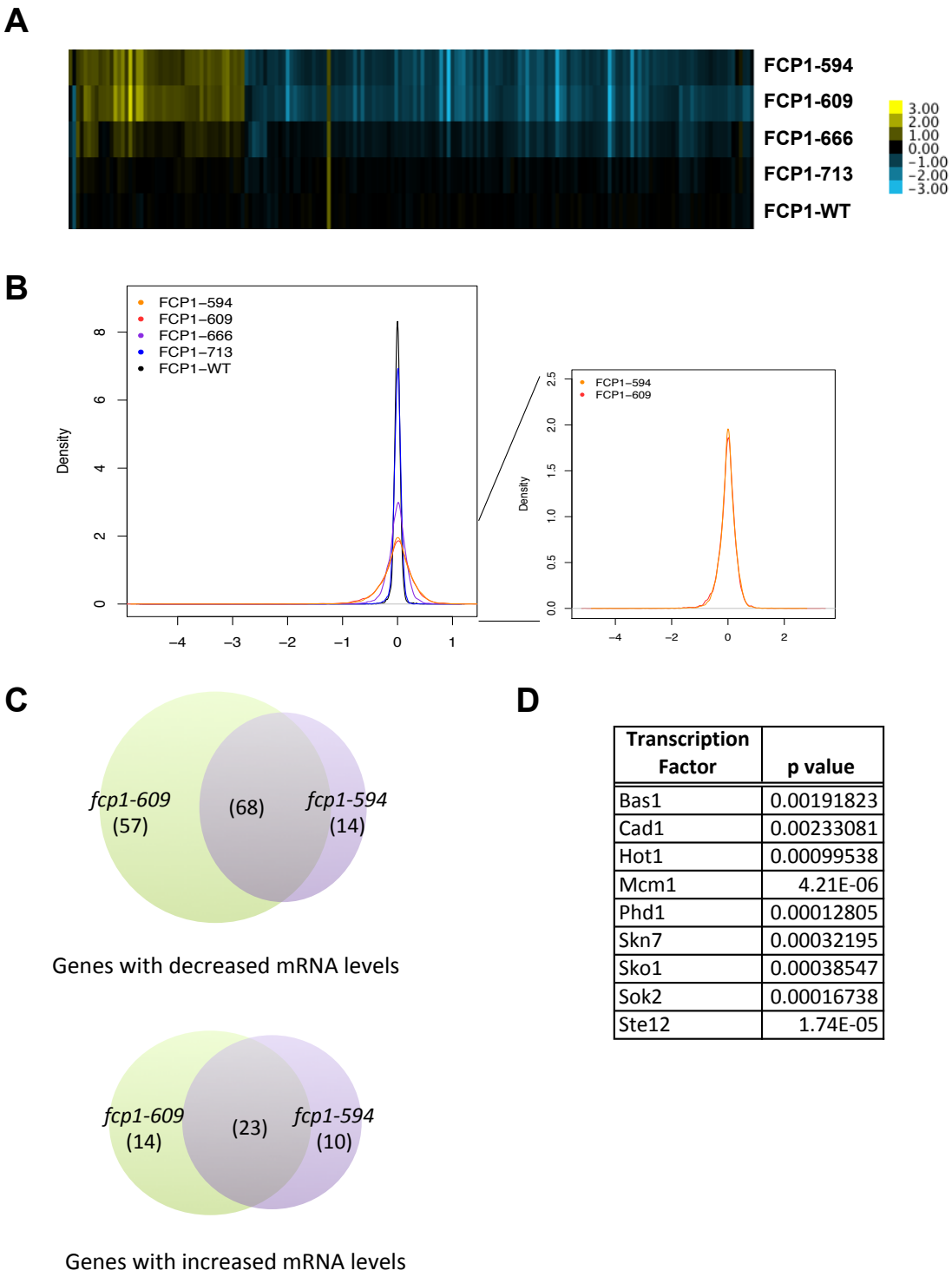



**Figure 4.1 *FCP1* C-terminal truncation mutants displayed growth defects.**
(A) Schematic of *FCP1* highlighting the mutations characterized in this study (B) *fcp1-594* and *fcp1-609* were sensitive to exposure to a variety of growth conditions. Ten-fold serial dilutions of the indicated mutants were plated on YPD media at 30 and 37 °C or media containing the indicated concentrations of formamide, hydroxyurea, or MMS.




| Transcription Factor | p value |
|---|---|
| Bas1 | 0.00191823 |
| Cad1 | 0.00233081 |
| Hot1 | 0.00099538 |
| Mcm1 | 4.21E-06 |
| Phd1 | 0.00012805 |
| Skn7 | 0.00032195 |
| Sko1 | 0.00038547 |
| Sok2 | 0.00016738 |
| Ste12 | 1.74E-05 |

**Figure 4.2 *FCP1* C-terminal truncation mutants resulted in relatively few transcriptional alterations.**
(A) Heatmap of genes with significantly increased or decreased mRNA levels in the *FCP1* truncation mutants. Yellow indicates genes with increased mRNA levels and blue indicates

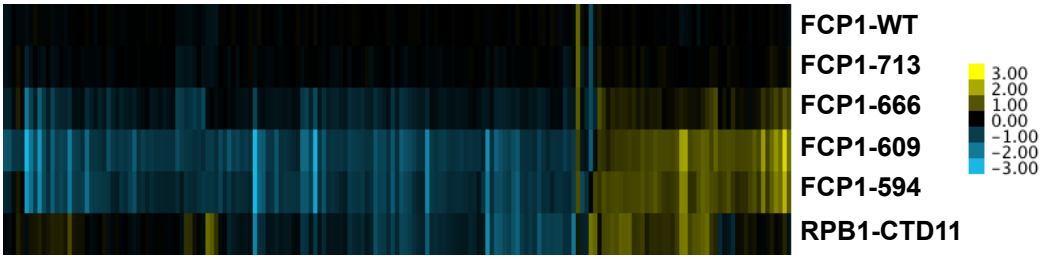
genes with decreased levels. (B) Distribution of S scores for *FCP1* truncation mutants revealed that the *fcp1-609* mutant resulted in the greatest number of transcriptional alterations. (C) Venn diagram showing the overlap of the *fcp1-594* and *fcp1-609* mutant gene expression profile. (D) Table of transcription factors that were enriched for regulation of genes whose mRNA level decrease in the *fcp1-594* and *fcp1-609* mutant.

### 4.3.2 The gene expression profiles of *FCP1* and *RPB1* C-terminal truncation mutants revealed similarities and important differences

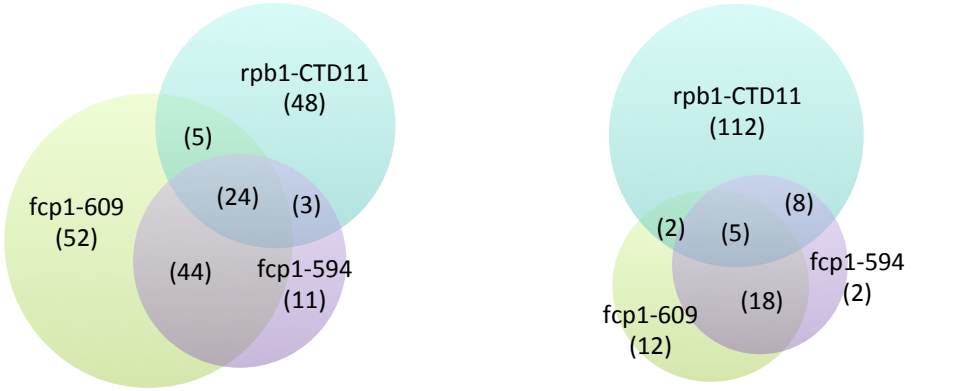
Given that Fcp1's only known target *in vivo* is the RNAPII-CTD, we aimed to understand its contribution to RNAPII biology by comparing the gene expression profiles of the *FCP1* mutants to that of the *rpb1-CTD11* mutant described in chapter 2 and 3 of this thesis, which was generated using the same platform (*RPB1* C-terminal truncation mutant containing only 11 repeats) (Aristizabal et al. 2013). Briefly, the *rpb1-CTD11* mutant resulted in 127 genes with increased mRNA levels that were regulated primarily by Rpn4, and 80 genes with decreased mRNA levels that were primarily regulated by Ste12. Highlighting a shared role, of the genes whose mRNA levels were altered in the *fcp1-594* and *fcp1-609* mutant, a significant proportion was also altered in the *rpb1-CTD11* mutant (Hypergeometric test p value for the genes with decreased mRNA levels 1.33e-29 and for the genes with increased mRNA levels 8.52e-05) (**Figure 4.3A and B**). The shared genes whose mRNA levels decreased in the *FCP1* and *rpb1-CTD11* mutant were primarily involved in pheromone-dependent signal transduction and mating, and were regulated by the Mcm1 transcription factor. In contrast, given that only 5 genes whose mRNA levels increased were shared between the *FCP1* and *rbp1-CTD11* mutant, no significant GO terms or transcription factors were identified. Despite the similarities, differences were also clearly apparent and these suggested differential roles for Fcp1 and the RNAPII-CTD. In particular, a relatively low correlation was obtained when comparing the *fcp1-594* and *rbp1-*

*CTD11* mutant profile or the *fcp1-609* and *rpb1-CTD11* mutant profile (Pearson's rho 0.46 and 0.37 respectively). Focusing on the genes whose mRNA levels were only altered in the *FCP1* mutant revealed that the genes whose mRNA levels increased were involved in lysine degradation, however no transcription factors met the enrichment criteria. In contrast, the genes whose mRNA levels decreased were involved in amine catabolism and serine family amino acid catabolic processes, and were regulated by a subset of transcription factors including, Sko1 and Hot1 which are known phospho-proteins (**Figure 4.3C**). Despite the observed differences, tetrad analysis in the W303 background showed that combining the *fcp1-594* and *rpb1-CTD11* mutant resulted in lethality, thus highlighting a shared role for Fcp1 and the RNAPII-CTD in maintaining cellular viability (**Figure 4.3D**).

A




B



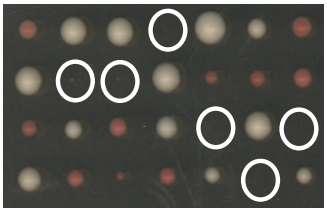

Genes with decreased mRNA levels
Hypergeometric test p value 1.33e-29

Genes with increased mRNA levels
Hypergeometric test p value 8.52e-05

C

| Transcription Factor | p value |
|---|---|
| Bas1 | 0.00149111 |
| Hot1 | 0.00149051 |
| Pdr1 | 0.00121294 |
| Pdr3 | 0.00043621 |
| Sko1 | 2.59E-05 |
| Sok2 | 0.00043613 |

D



*fcp1-594 rpb1-CTD11*

**Figure 4.3 Comparing the gene expression profile of *FCP1* and *rbp1-CTD11* mutants revealed similarities and differences.**
(A) Heatmap of genes with significantly increased or decreased mRNA levels in the *FCP1* truncation mutants revealed a significant overlap with the *rbp1-CTD11* mutant. Yellow indicates genes with increased mRNA levels and blue indicates genes with decreased levels. (B) Venn diagrams showing the overlap of the *fcp1-594*, *fcp1-609*, and *rpb1-CTD11* mutant gene expression profiles. (C) Table of transcription factors that were enriched for regulation of genes whose mRNA level significantly decrease in the *fcp1-594* and *fcp1-609*, but where unaltered in the *rpb1-CTD11* mutant. (D) Tetrad analysis of *fcp1-594 rpb1-CTD11* mutants revealed synthetic lethality. White circles mark the location of double mutants.

### 4.3.3 The genetic interaction network of *FCP1* C-terminal truncation mutants revealed length-dependent genetic interactions

Given that Fcp1 has been implicated in other biological pathways, Epistasis Mini Array Profiling (E-MAP) was performed to more broadly understand the role of Fcp1 in the cell. The same alleles described previously were generated and tested for their genetic requirements against a library of 1532 different mutants whose function primarily related to chromatin biology, transcription and RNA processing (Collins et al. 2010). Overall, the genetic interaction profiles revealed nuanced *FCP1* length-dependent genetic interactions patterns. Nonetheless, testifying to the validity of the profiles, they recapitulated the previously reported genetic relationship between *FCP1* and the translation Elongator complex (Chen et al. 2011), although this was first identified using different *FCP1* mutants (Kong et al. 2005). In detail, the *fcp1-666*, *fcp1-609* and *fcp1-594* mutants had negative genetic interactions when combined with the *ELP2*, *ELP3* and *ELP4* subunits of the Elongator complex (**Figure 4.4A**). The genetic interaction profiles were also consistent with a role for Fcp1 as a $S_2$ CTD phosphatase. Specifically, combining the *FCP1* mutants with deletion of genes encoding subunits of the CTDK-I $S_2$ kinase complex, *CTK1* and *CTK3,* resulted in negative interactions when combined with *fcp1-713* and *fcp1-666*. Further indicative of alterations to CTD phosphorylation upon *FCP1* C-terminal mutation, the *FCP1* mutants were strongly dependent on the *RTR1* gene for viability, and had strong negative genetic interactions when combined with subunits of the COMPASS complex, whose function is dependent on CTD phosphorylation (Shilatifard 2012). *FCP1* mutants also interacted genetically with mutants in genes encoding known *FCP1* regulators. Combining the *fcp1-609* mutant with the *ESS1*- DAmP mutant resulted in a significant positive genetic interaction. Similarly, combining the two shortest *FCP1* mutants with a DAmP mutant in the TFIIF subunit, *TFG1*, or

with the gene encoding TFIIB itself, *SUA7*, resulted in negative genetic interactions. Worth noting, previous work showed that combining the *fcp1-594* mutant and the *tfg1-2* mutant resulted in no genetic interaction, a difference that likely stems from the different *TFG1* mutants used (Kobor et al. 2000). Overall, the genetic interaction profiles were similar to the gene expression profiles, as these also revealed that the *fcp1-609* mutant and not the *fcp1-594* mutant had the most significant genetic alterations (**Figure 4.4B**). Furthermore, not all interactions observed in the *fcp1-609* mutant were also observed in the *fcp1-594* mutant and vice versa (**Figure 4.4C**). More broadly, comparison of the *FCP1* mutant genetic and gene expression profiles revealed that these provided similar functional information as manifested by a relatively high correlation between the two types of profiles (Comparison of E-MAP vs gene expression profiles Person's rho 0.721) (**Figure 4.4D**). Finally, focusing on the genetic interactions that were observed for both *fcp1-594* and *fcp1-609* revealed that the genes which interacted positively with the *FCP1* mutants were enriched for ribosome biogenesis and translation pathways, while no significant GO terms were found for the genes that interacted negatively with the *FCP1* mutants (**Table 4.3**)

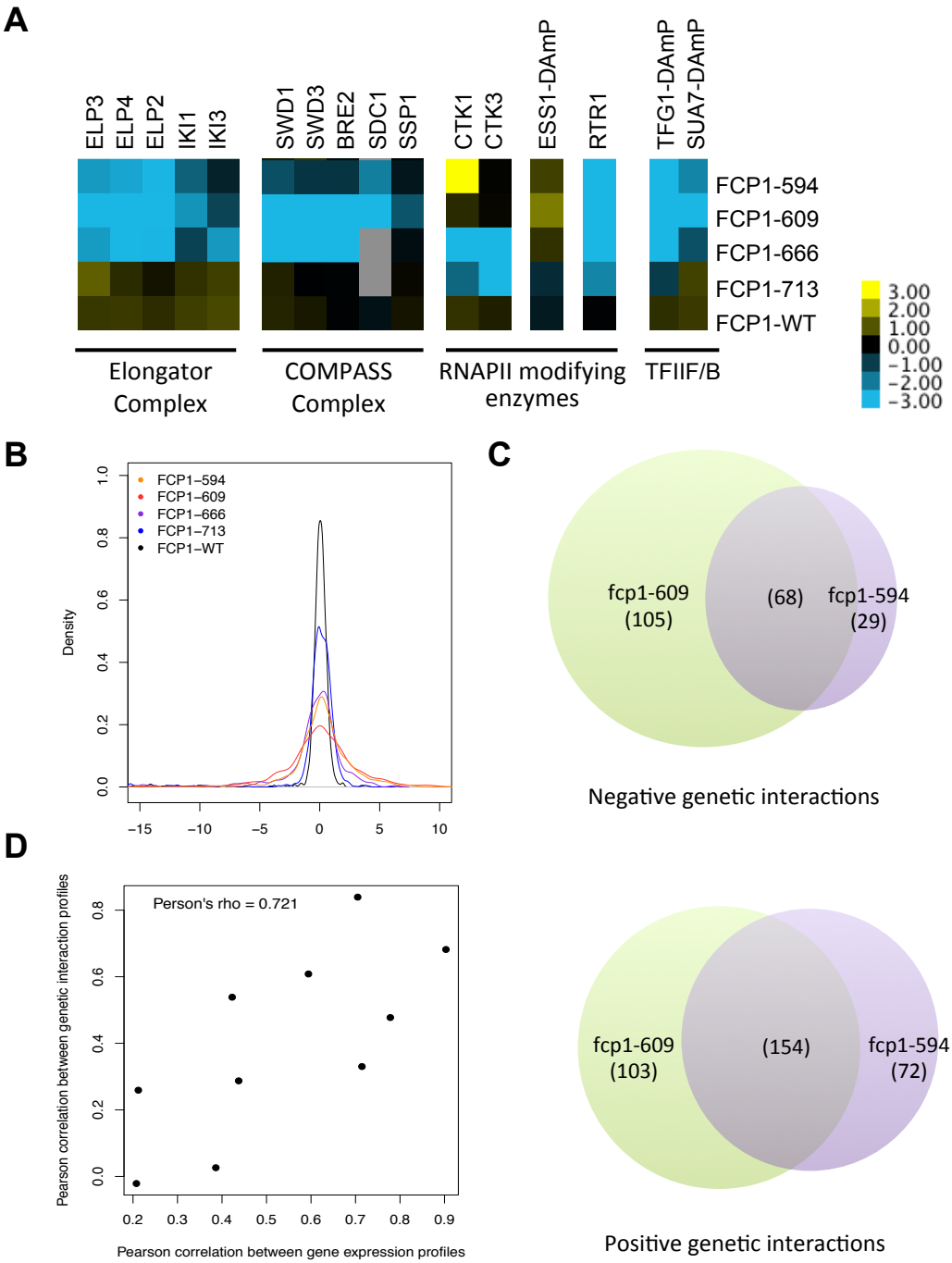



**Figure 4.4 A comprehensive genetic network of *FCP1* C-terminal truncation mutants highlighted roles in regulating RNAPII-CTD phosphorylation levels.**
(A) Subsets of genetic interaction profiles for *FCP1* as it was progressively shortened. Blue and yellow represent aggravating and alleviating genetic interactions respectively. Gray boxes

represent missing values. The S score is a modified T-statistic measure, which captures both the confidence and strength of the genetic interaction. Scores greater than 2.0 or less than -2.5 are considered significant. (B) Distribution of S scores for *FCP1* truncation mutants revealed that the *fcp1-609* mutant had the most significant genetic interactions. (C) Venn diagrams showing the overlap of the *fcp1-594* and *fcp1-609* genetic interaction profiles. (D) Scatterplot of profile of paired correlations in gene expression and genetic interaction.

**Table 4.2 List of GO categories enriched for genes that interact positively with the *FCP1* mutants.**

| GO category | Benjamini corrected p value |
|---|---|
| ribosome biogenesis | 0.000019 |
| translation | 0.000021 |
| ribonucleoprotein complex biogenesis | 0.000039 |
| rRNA processing | 0.00075 |
| rRNA metabolic process | 0.00094 |

### 4.3.4 Similarities and differences of the genetic interaction profiles of *FCP1* and RNAPII-CTD truncation mutants

To investigate the extent of Fcp1's contribution to RNAPII-CTD biology and to identify roles of Fcp1 that differ from that of the RNAPII-CTD, we compared the genetic interaction profiles of the *FCP1* and *rpb1-CTD11* mutants in a manner similar to the way the gene expression profiles were compared previously. While there was a significant overlap between both genetic interaction profiles (hypergeometric test p value 1.07e-11 or 3.01e-30 for negative and positive genetic interaction respectively), differences were also evident as demonstrated by the low correlation between the *fcp1-594* and *rpb1-CTD11* or *fcp1-609* and *rbp1-CTD11* profiles (Pearson's r 0.49 and 0.35 respectively) (**Figure 4.5A**). Highlighting a shared role, the two shortest *FCP1* mutants and the *rpb1-CTD11* mutant were generally similar in their interactions

with *SUA7* and *TFG1* and the RNAPII modifying enzymes, Rtr1, Ess1 and Ctk1 (**Figure 4.5B**). Focusing on genetic interactions that differed between the *FCP1* and *rpb1-CTD11* mutants, revealed differences in interactions with genes encoding subunits of the Elongator, COMPASS and prefoldin complexes. Specifically, only *FCP1* mutants had strong negative genetic interactions when combined with genes encoding subunits of the Elongator or COMPASS complex. Furthermore, *FCP1* mutants had strong positive interactions when combined with a subset of prefoldin subunits, a complex recently found to stimulate RNAPII elongation rate *in vivo* (Millan-Zambrano et al. 2013). This was in contrast to the significant negative genetic interactions observed when the RNAPII-CTD truncation mutants were combined with genes encoding prefoldin subunits. Thus, our genetic interaction profiles suggested differential roles for the RNAPII-CTD and Fcp1 in RNAPII-dependent transcription, which require further investigation.

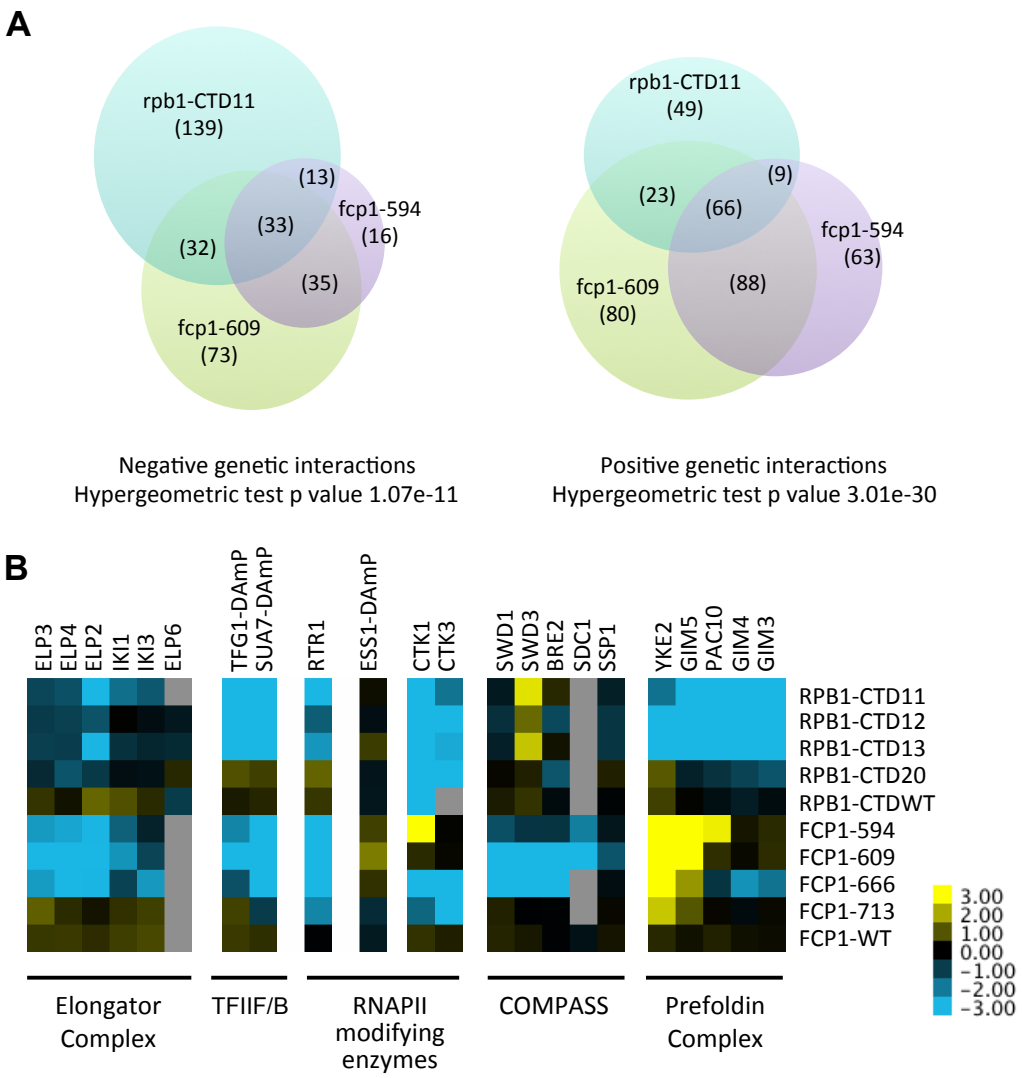



**Figure 4.5 The genetic interaction network of *FCP1* and RNAPII-CTD truncation mutants revealed overlapping and divergent functions.**
(A) Venn diagrams showing the overlap of the *fcp1-594*, *fcp1-609, rpb1-CTD11* mutant genetic interaction profiles. (B) Subsets of genetic interaction profiles for *FCP1* and the RNAPII-CTD as they were progressively shortened. Blue and yellow represent aggravating and alleviating genetic interactions respectively. Gray boxes represent missing values. The S score is a modified T-statistic measure, which captures both the confidence and strength of the genetic interaction. Scores greater than 2.0 or less than -2.5 are considered significant.

### 4.3.5 A subset of *FCP1* mutant phenotypes were suppressed by loss of *CDK8*

Growth and gene expression phenotypes associated with *rpb1-CTD11* mutants are robustly suppressed by loss of *CDK8*, thus we determined if loss of *CDK8* could also suppress *FCP1* mutant phenotypes (Aristizabal et al. 2013, Nonet and Young 1989). Focusing on *fcp1-594* and *fcp1-609*, loss of *CDK8* partially suppressed the sensitivity to exposure to HU and MMS to a small degree, an effect that differed from the robust suppression of *rpb1-CTD11* growth phenotypes by loss of *CDK8* when exposed to many different conditions (**Figure 4.6A**). To better understand the biological underpinnings of the suppression phenotype, gene expression profiles were generated for *FCP1* mutants in combination with loss of *CDK8*. These revealed limited normalization of mRNA levels for the *FCP1*-dependent genes (**Figure 4.6B**). Using a stringent threshold (absolute fold changes < 1.1 and p values > 0.01 compared to wild type), only 3 of the 23 genes whose mRNA levels increased in the *FCP1* mutant, were normalized upon loss of *CDK8*. Similarly, of the 68 genes whose mRNA levels decreased in the *FCP1* mutant, only 9 had normalized mRNA levels in the *fcp1-594 cdk8* or *fcp1-609 cdk8* double mutant.

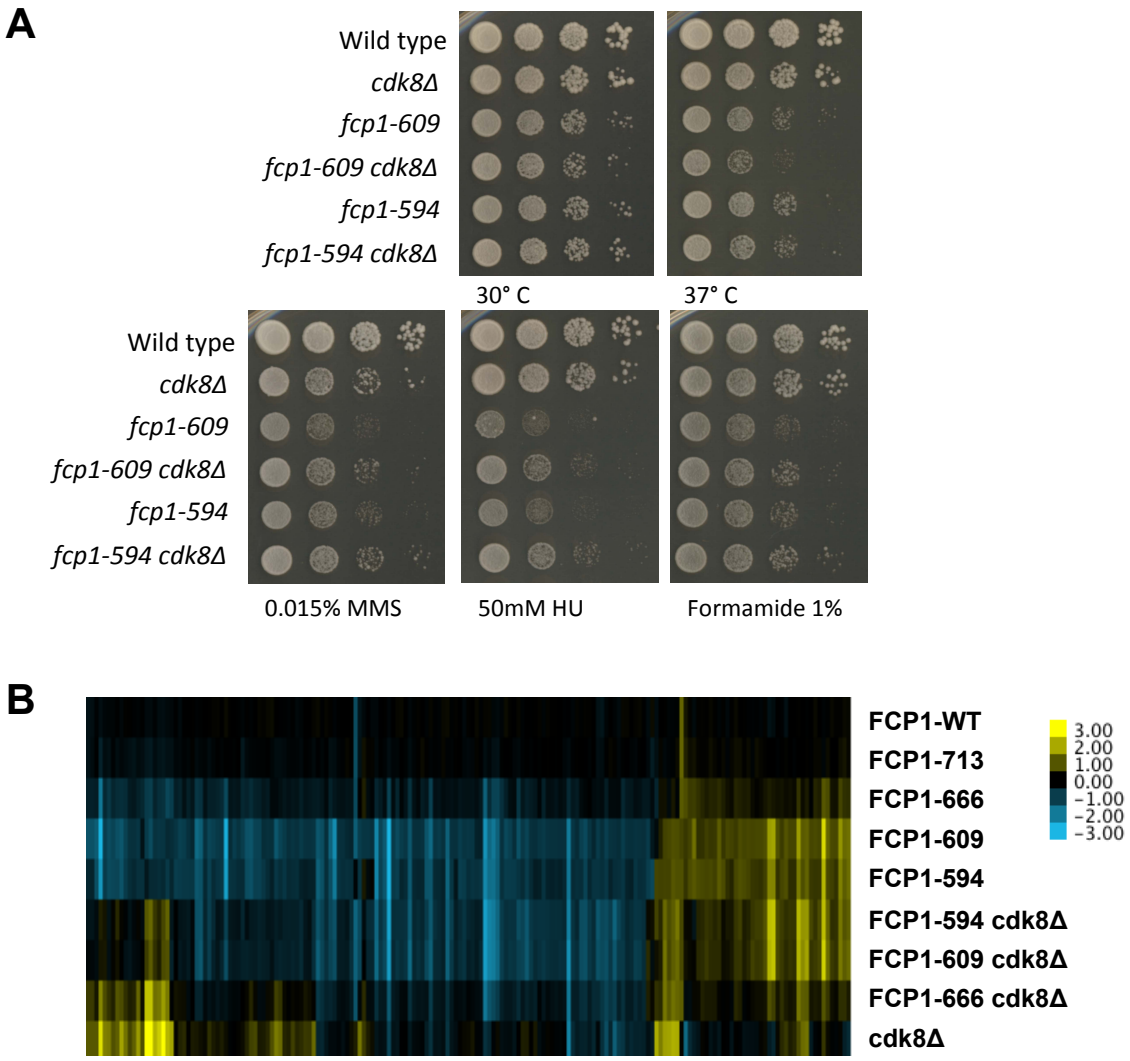



**Figure 4.6 A subset of *FCP1* mutant phenotypes were suppressed slightly by loss of *CDK8*.** (A) Sensitivity of *fcp1-594* and *fcp1-609* to exposure to MMS, HU and formamide were suppressed slightly by loss of *CDK8*. Ten-fold serial dilutions of the indicated mutants were plated on YPD media at 30 and 37 °C or media containing the indicated concentrations of formamide, hydroxyurea, or MMS. (B) Heatmap of genes with significantly increased or decreased mRNA levels in the *FCP1* truncation mutants revealed limited normalization of mRNA levels by loss of *CDK8*. Yellow indicates genes with increased mRNA levels and blue indicates genes with decreased levels

## 4.4 Discussion

To characterize the role and regulation of Fcp1 in the cell, high-throughput gene expression and genetic interaction profiles of *FCP1* C-terminal truncation mutants were generated. These highlighted known roles and regulatory relationships for Fcp1 while suggesting novel functions in transcription initiation and elongation. Overall, our profiles supported the well-established role of Fcp1 as a CTD phosphatase but also suggested differential roles for Fcp1 and the RNAPII-CTD. Importantly, this work provided a framework for identifying additional substrates for Fcp1 as the genes whose mRNA levels were dependent on the Fcp1 C-terminus were primarily regulated by transcription factors that are known phospho-proteins. The significant overlap in gene expression and genetic interaction profiles of the *FCP1* and RNAPII-CTD truncation mutants confirmed a prominent role of Fcp1 in RNAPII biology, however differences were also clearly evident. Further highlighting a shared role, combining the *fcp1-594* and *rpb1-CTD11* mutants resulted in lethality, and similar to *rpb1-CTD11*, loss of *CDK8* could also suppress *FCP1* mutant phenotypes, a connection that was particularly intriguing given that Fcp1 and Cdk8 share a common substrate, the RNAPII-CTD. Highlighting differential effects, the genetic interaction profiles revealed opposing interactions for *FCP1* and RNAPII-CTD truncation mutants when components of the prefoldin complex were considered, suggesting a different regulatory partnership.

The *FCP1* C-terminal truncation mutants described here were comparable to previously characterized alleles although differences were apparent (Kobor et al. 2000). Specifically, a plasmid based *fcp1-594* mutant differed slightly from our genome-encoded version in that it did

not result in reduced growth when exposed to high temperatures, a situation that likely arises from multiple plasmid copies in the cell. Furthermore, the *FCP1* C-terminal truncation mutants differed from previously reported *FCP1* mutants that replaced key catalytic motif amino acids, in that they had milder growth defects and fewer gene expression alterations (Kobor et al. 1999). This is consistent with Fcp1's catalytic activity being important for its role in gene expression. Therefore, this work focused on genes whose expression is most dependent on the Fcp1 C-terminus rather than the catalytic activity of Fcp1. Identifying few transcriptional alterations was reminiscent of recently described gene expression profiles generated from strains carrying mutations in genes with prominent functions in gene expression, including *SET1*, *SET2* and *DOT1* which lay prominent transcription associated histone marks, as these studies also reported limited effects on gene expression (Aristizabal et al. 2013, Kemmeren et al. 2014, Lenstra et al. 2011). Thus, this work contributes to mounting evidence suggesting that gene expression regulation is in part achieved via multiple redundant pathways.

The observation that *FCP1* mutant phenotypes were reminiscent of RNAPII-CTD mutant phenotypes and that both were suppressed by loss *CDK8*, although with clearly different magnitudes, suggested an intimate relationship between the RNAPII-CTD, Fcp1 and Cdk8. Given that Fcp1 and Cdk8 both regulate RNAPII-CTD phosphorylation status, one of the most likely possibilities was that the suppression of *FCP1* mutant phenotypes by loss of *CDK8* was through a normalizing effect on RNAPII-CTD phosphorylation. While this possibility remains to be fully investigated, the relatively few number of *FCP1*-dependent genes whose mRNA levels were normalized upon loss of *CDK8* suggests that a bulk effect on CTD phosphorylation is unlikely. More specifically, loss of *CDK8* in the *FCP1* mutant normalized the altered mRNA

levels of 12 genes, a relatively few number given that loss of *CDK8* in the *rpb1-CTD11* mutant normalized the altered mRNA levels of 67 genes (Aristizabal et al. 2013). Nonetheless, given that Fcp1 might target multiple residues of the RNAPII-CTD, future investigations will need to implement genome-wide mapping of individual RNAPII-CTD phosphorylation marks under wild type, *FCP1* mutant, *cdk8Δ* and *FCP1 cdk8Δ* mutant conditions.

The gene expression and genetic interaction profiles of *FCP1* C-terminal truncation mutants revealed important aspects of Fcp1 function and regulation. In particular, *fcp1-609* and not the shortest *FCP1* truncation mutant resulted in the highest number of gene expression changes and genetic interactions. Most surprisingly, not all genes whose mRNA levels were affected in the *fcp1-609* mutant were also affected in the *fcp1-594* mutant, thus revealing important functional intricacies of the Fcp1 C-terminus. Interestingly, both proteins encoded by the *fcp1-594* and *fcp1-609* alleles are expressed to similar levels (data not shown). Furthermore, both are unable to physically interact with TFIIF and TFIIB *in vitro* suggesting that the gene expression differences between *fcp1-594* and *fcp1-609* were independent of the ability to interact with TFIIF or TFIIB (Kobor et al. 2000). Further work should aim to determine whether these two alleles differ in their ability to interact with other known Fcp1 binding partners, providing a tool to study how these binding partners affect Fcp1 function.

The genetic interaction profiles of *FCP1* mutants highlighted its role as an RNAPII-CTD phosphatase given that *FCP1* mutants were highly dependent on the genes encoding the CTD phosphatase, *RTR1,* for viability. Furthermore, the genetic interaction profiles identified opposing genetic interactions for Fcp1 and the RNAPII-CTD when combined with mutants in

subunits of the prefoldin complex, thus suggesting key roles in its function. A connection to prefoldin is interesting given that Fcp1 and prefoldin function in transcription elongation (Cho et al. 2001, Millan-Zambrano et al. 2013). Furthermore, gene-specific chromatin association of prefoldin is dependent on transcriptional frequency and $S_2$ phosphorylation levels, which are increased in the *FCP1* mutants and decreased in the *rpb1-CTD11* mutant. Thus, our results suggest an important and differential role for Fcp1 and the RNAPII-CTD in prefoldin function.

Finally, our high-throughput characterization of *FCP1* mutants provides a starting point to investigate roles for Fcp1 beyond RNAPII-CTD dephosphorylation. Intriguingly, *FCP1*-dependent genes were primarily regulated by transcription factors that are known phospho-proteins, the majority of which do not have their respective phosphatase identified. Therefore, it is tempting to speculate that Fcp1 might dephosphorylate these transcription factors, and that the unique *FCP1* mutant gene expression profiles resulted from a direct regulatory effect on this set of transcription factors. As such, the transcription factor enrichment analysis of *FCP1*-regulated genes provides an important list of potential Fcp1 substrates for further analysis.

# Chapter 5: Genome-wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with DRIP-chip

## 5.1 Introduction[2]

Elevated DNA:RNA hybrid formation due to defects in RNA processing pathways leads to genome instability and replication stress across species (Chernikova et al. 2012, El Hage et al. 2010, Gan et al. 2011, Gomez-Gonzalez et al. 2011, Mischo et al. 2011, Stirling et al. 2012, Wahba et al. 2011). R loops threaten genome stability and often form under abnormal conditions where nascent mRNA is improperly processed or RNA half-life is increased, resulting in RNA that can hybridize with template DNA, displacing the non-transcribed DNA strand (Aguilera and Garcia-Muse 2012). A recent study also found that hybrid formation can occur in *trans* via Rad51-mediated DNA-RNA strand exchange (Wahba et al. 2013). Persistent R loops pose a major threat to genome stability through two mechanisms. First, the exposed non-transcribed strand is susceptible to endogenous DNA damage due to the increased exposure of chemically reactive groups. The second, more widespread mechanism, identified in *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and human cells, involves the R loops and associated stalled transcription complexes, which block DNA replication fork progression (Aguilera and Garcia-Muse 2012, Castellano-Pozo et al. 2012, Dominguez-Sanchez et al. 2011,

---

[2] A version of this chapter is published in the Public Library of Science Genetics. Chan Y.A.*, Aristizabal M.J.*, Lu P.Y.T., Luo Z., Hamza A., Kobor M.S., Stirling P.C., Hieter P. (2014) Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. PLoS Genet 10:e1004288. doi: 10.1371/journal.pgen.1004288. * These authors contributed equally to the work.

Gan et al. 2011, Gomez-Gonzalez et al. 2011). R loop-mediated instability is an area of great interest primarily because genome instability is considered an enabling characteristic of tumor formation (Hanahan and Weinberg 2011). Moreover, mutations in RNA splicing/processing factors are frequently found in human cancer, heritable diseases like Aicardi-Goutieres syndrome, and a degenerative ataxia associated with Senataxin mutations (Crow et al. 2006, Garraway and Lander 2013, Papaemmanuil et al. 2011, Suraweera et al. 2009, Wang et al. 2011).

To avoid the deleterious effects of R loops, cells express enzymes for the removal of abnormally formed DNA:RNA hybrids. In *S. cerevisiae*, *RNH1* and *RNH201*, each encoding RNase H are responsible for one of the best characterized mechanisms for reducing R loop formation by enzymatically degrading the RNA in DNA:RNA hybrids (Aguilera and Garcia-Muse 2012). Another extensively studied anti-hybrid factor is the THO/TREX complex, which functions to suppress hybrid formation at the level of transcription termination and mRNA packaging (Chavez et al. 2001, Dominguez-Sanchez et al. 2011, Gomez-Gonzalez et al. 2011, Jimeno et al. 2002). In addition, the Senataxin helicase, yeast Sen1, plays an important role in facilitating replication fork progress through transcribed regions and unwinding RNA in hybrids to mitigate R loop formation and RNA polymerase II transcription-associated genome instability (Alzu et al. 2012, Mischo et al. 2011). Several additional anti-hybrid mechanisms have also been identified including topoisomerases and other RNA processing factors (El Hage et al. 2010, Leela et al. 2013, Luna et al. 2005, Sikdar et al. 2008, Stirling et al. 2012, Wahba et al. 2011, Wahba et al. 2013).

To add to the complexity of DNA:RNA hybrid management in the cell, hybrids also occur naturally and have important biological functions (Wahba and Koshland 2013). In human cells, R loop formation facilitates immunoglobulin class switching, protects against DNA methylation at CpG island promoters and plays a key role in pause site-dependent transcription termination (Chaudhuri et al. 2003, Ginno et al. 2012, Ginno et al. 2013, Skourti-Stathaki et al. 2011). Transcription of telomeres by RNA polymerase II also produces telomeric repeat-containing RNAs (TERRA), which associate with telomeres and inhibit telomere elongation in a DNA:RNA hybrid-dependent fashion (Balk et al. 2013, Luke et al. 2008, Pfeiffer et al. 2013). Noncoding RNAs (ncRNAs), such as antisense transcripts, perform regulatory roles in the expression of sense transcripts that may involve R loops (Faghihi and Wahlestedt 2009). The proposed mechanisms of antisense transcription regulation are not clearly understood and involve different modes of action specific to each locus. Current models include chromatin modification resulting from antisense-associated transcription, antisense transcription modulation of transcription regulators, collision of sense and antisense transcription machineries and antisense transcripts expressed in *trans* interacting with the promoter for sense transcription (Camblong et al. 2009, Castelnuovo et al. 2013, Faghihi and Wahlestedt 2009, Hobson et al. 2012, Kanhere et al. 2010, Margaritis et al. 2012, Marinello et al. 2013, van Dijk et al. 2011, Wang et al. 2008). More recently, studies in *Arabidopsis* found an antisense transcript that forms R loops, which can be differentially stabilized to modulate gene regulation (Sun et al. 2013). Similarly, in mouse cells the stabilization of an R loop was shown to inhibit antisense transcription (Powell et al. 2013).

Here we describe, for the first time, a genome-wide profile of DNA:RNA hybrid prone loci in *S. cerevisiae* by DNA:RNA immunoprecipitation followed by hybridization on tiling microarrays

(DRIP-chip). We found that DNA:RNA hybrids occurred at highly transcribed regions in wild type cells, including some identified in previous studies. Remarkably, we observed that DNA:RNA hybrids were significantly associated with genes that have corresponding antisense transcripts, suggesting a role for hybrid formation at these loci in gene regulation. Consistently, we found that genes whose expression was altered by overexpression of RNase H were also significantly associated with antisense transcripts. A small-scale cytological screen found that diverse RNA processing mutants had increased hybrid formation and additional DRIP-chip studies revealed specific hybrid-site biases in the RNase H, Sen1 and THO complex subunit Hpr1 mutants. These genome-wide analyses enhance our understanding of DNA:RNA hybrid-forming regions *in vivo*, highlight the role of cellular RNA processing activities in suppressing hybrid formation, and implicate DNA:RNA hybrids in control of a subset of antisense regulated loci.

## 5.2 Materials and methods

### 5.2.1 Strains and plasmids

All strains are listed in **Table 5.1**. For RNase H overexpression experiments, recombinant human RNase H1 was expressed from plasmid p425-GPD-RNase H1 (2m, *LEU2*, GPDpr-*RNaseH1*) and compared to an empty control plasmid p425-GPD (2m, *LEU2*, GPDpr) (Wahba et al. 2011).

## Table 5.1 Strains used in this study

| YPH number | Relevant Genotype | Source |
|---|---|---|
| BY4741 | MATa *ura3*D*0 leu2*D*0 his3*D*1 met15*D*0* | Jef Boeke |
| YPH2111 | MATa *ura3*D*0 leu2*D*0 his3*D*1 rnh1*D*::KanMX rnh201*D*::KanMX* | Stirling et al. 2012 |
| YPH2233 | MATa *ura3*D*0 leu2*D*0 his3*D*0 sen1-1::KanMX* | Li et al. 2011 |
| YPH2597 | MATa *ura3*D*0 leu2*D*0 his3*D*1 hpr1*D*::KanMX* | Derived from the *hpr1*Δ/*HPR1* heterozygous diploid strain from the *yeast* heterozygous diploid collection (Open Biosystems) |
| YPH2598 | MATa *ura3*D*0 leu2*D*0 his3*D*0 mot1-1033::KanMX* | Li et al. 2011 |
| YPH2599 | MATa *ura3*D*0 leu2*D*0 his3*D*0 taf5-20:KanMX* | Li et al. 2011 |
| YPH2600 | MATa *ura3*D*0 leu2*D*0 his3*D*0 cdc36-16::KanMX* | Li et al. 2011 |
| YPH2601 | MATa *ura3*D*0 leu2*D*0 his3*D*0 pti1-ts7KanMX* | Li et al. 2011 |
| YPH2602 | MATa *ura3*D*0 leu2*D*0 his3*D*0 cet1-2::KanMX* | Li et al. 2011 |
| YPH2603 | MATa *ura3*D*0 leu2*D*0 his3*D*0 hrp1-4::KanMX* | Li et al. 2011 |
| YPH2604 | MATa *ura3*D*0 leu2*D*0 his3*D*0 sub2-1::KanMX* | Li et al. 2011 |
| YPH2605 | MATa *ura3*D*0 leu2*D*0 his3*D*0 rna1-1::KanMX* | Li et al. 2011 |
| YPH2606 | MATa *ura3*D*0 leu2*D*0 his3*D*0 srm1-ts::KanMX* | Li et al. 2011 |
| YPH2607 | MATa *ura3*D*0 leu2*D*0 his3*D*0 brl1-2221:KanMX* | Li et al. 2011 |
| YPH2608 | MATa *ura3*D*0 leu2*D*0 his3*D*0 snu13-L67W::KanMX* | Li et al. 2011 |
| YPH2609 | MATa *ura3*D*0 leu2*D*0 his3*D*0 rpf1-1::KanMX* | Li et al. 2011 |
| YPH2610 | MATa *ura3*D*0 leu2*D*0 his3*D*0 imp4-2::KanMX* | Li et al. 2011 |
| YPH2611 | MATa *ura3*D*0 leu2*D*0 his3*D*0 yhc1-1::KanMX* | Li et al. 2011 |
| YPH2612 | MATa *ura3*D*0 leu2*D*0 his3*D*0 prp31-1::KanMX* | Li et al. 2011 |
| YPH2613 | MATa *ura3*D*0 leu2*D*0 his3*D*0 snu114-60::KanMX* | Li et al. 2011 |
| YPH2614 | MATa *ura3*D*0 leu2*D*0 his3*D*0 prp6-1::KanMX* | Li et al. 2011 |
| YPH2615 | MATa *can1*D::*MFA1pr-HIS3::LEU2 his3*D*1 met15*D*0 rrp4-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2616 | MATa *can1*D::*MFA1pr-HIS3::LEU2 his3*D*1 met15*D*0 dbp6-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2617 | MATa *can1*D::*MFA1pr-HIS3::LEU2 his3*D*1 met15*D*0 snp1-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2618 | MATa *can1*D::*MFA1pr-HIS3::LEU2 his3*D*1 met15*D*0 aar2-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2619 | MATa *can1*D::*MFA1pr-HIS3::LEU2 his3*D*1 met15*D*0 dib1-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2620 | MATa *can1*D::*MFA1pr-HIS3::LEU2 his3*D*1 met15*D*0 spp381-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2621 | MATa *can1*D::*MFA1pr-HIS3::LEU2 his3*D*1 met15*D*0 spp382-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2622 | MATa *can1*D::*MFA1pr-HIS3::LEU2 his3*D*1 met15*D*0 cwc2-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2623 | MATa *can1*D::*MFA1pr-HIS3::LEU2 his3*D*1 met15*D*0 lsm2-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2624 | MATa *can1*D::*MFA1pr-HIS3::LEU2 his3*D*1 met15*D*0* | Ben-aroya et al. 2008 |

| YPH number | Relevant Genotype | Source |
|---|---|---|
| | *hsh155-ts::URA3* | |
| YPH2625 | MATa *can1*D::*MFA1pr-HIS3*::*LEU2 his3*D*1 met15*D*0 msl5-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2626 | MATa *can1*D::*MFA1pr-HIS3*::*LEU2 his3*D*1 met15*D*0 syf1-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2627 | MATa *can1*D::*MFA1pr-HIS3*::*LEU2 his3*D*1 met15*D*0 sts1-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2628 | MATa *can1*D::*MFA1pr-HIS3*::*LEU2 his3*D*1 met15*D*0 kae1-ts::URA3* | Ben-aroya et al. 2008 |
| YPH2629 | MATa *ura3*D*0 leu2*D*0 his3*D*0 lys2*D*0 dis3*D*::KanMX* | Open Biosystems |
| YPH2630 | MATa *ura3*D*0 leu2*D*0 his3*D*0 lys2*D*0 dbp7*D*::KanMX* | Open Biosystems |
| YPH2631 | MATa *ura3*D*0 leu2*D*0 his3*D*0 lys2*D*0 ssf1*D*::KanMX* | Open Biosystems |
| YPH2632 | MATa *ura3*D*0 leu2*D*0 his3*D*0 lys2*D*0 mud2*D*::KanMX* | Open Biosystems |
| YPH2633 | MATa *ura3*D*0 leu2*D*0 his3*D*0 lys2*D*0 snu66*D*::KanMX* | Open Biosystems |
| YPH2634 | MATa *ura3*D*0 leu2*D*0 his3*D*0 lys2*D*0 psh1*D*::KanMX* | Open Biosystems |
| YPH2635 | MATa *ura3*D*0 leu2*D*0 his3*D*0 lys2*D*0 esc2*D*::KanMX* | Open Biosystems |
| YPH2636 | MATa *ura3*D*0 leu2*D*0 his3*D*0 lys2*D*0 rnh1*D*::KanMX* | Open Biosystems |

### 5.2.2 DRIP-chip and qPCR

Briefly, cells were grown overnight, diluted to 0.15 $OD_{600}$ and grown to 0.7 $OD_{600}$. Crosslinking was done with 1% formaldehyde for 20 minutes. Chromatin was purified as described previously (Schulze et al. 2009) and sonicated to yield approximately 500 bp fragments. Forty micrograms of the anti-DNA:RNA hybrid monoclonal mouse antibody S9.6 (gift from Stephen Leppla) was coupled to 60 µL of protein A magnetic beads (Invitrogen). For ChIP-qPCR, crosslinking reversal and DNA purification were followed by qPCR analysis of the immunoprecipitated and input DNA. DNA was analyzed using a Rotor-Gene 600 (Corbett Research) and PerfeCTa SYBR green FastMix (Quanta Biosciences). Samples were analyzed in triplicate on three independent DRIP samples for wild type and *rnh1*D*rnh201*D. Primers are listed in **Table 5.2**.

For DRIP-chip, precipitated DNA was amplified via two rounds of T7 RNA polymerase amplification (van Bakel et al. 2008), biotin labeled and hybridized to Affymetrix 1.0R *S. cerevisiae* microarrays. Samples were normalized to a no antibody control sample (mock) using the rMAT software and relative occupancy scores were calculated for all probes using a 300 bp sliding window. All profiles were generated in duplicate and replicates were quantile normalized and averaged. Spearman correlation scores between replicates are listed in **Table 5.3**. DRIP-chip data is available at ArrayExpress E-MTAB-1656 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1656).

**Table 5.2 Primers used in this study**

| Genomic Target | Forward Primer | Reverse Primer |
|---|---|---|
| SUF2 | TATGATTCTCGCTTAGGGTGCGGGAGG | CATTAACATTGGTCTTCTCCAGCTTACTC |
| tV(UAC)D | GGTCCAATGGTCCAGTGGTTCAAGACGTCGCCTTTACACGGCGAAG | CATCGTTGCTGGGACCC |
| Intergenic region on chromosome V | GGCTGTCAGAATATGGGGCCGTAGTA | CACCCCGAAGCTGCTTTCACAATAC |

**Table 5.3 Spearman correlation between DRIP-chip replicates**

| Profile | Correlation |
|---|---|
| Wild type | 0.7837469 |
| *rnh1*Δ *rnh201*Δ | 0.8454704 |
| *hpr1*Δ | 0.8806051 |
| *sen1-1* | 0.8287079 |

### 5.2.3 DRIP-chip analysis

Enriched features had at least 50% of the probes contained in the feature above the threshold of 1.5. Only features enriched in both replicates were reported. Transcriptional frequency (Holstege et al. 1998), GC content (Kinsella et al. 2011) and gene length were compared using the Wilcoxon rank sum test. Antisense association was analyzed by the Fisher's exact test using R. Statistical analysis of genomic feature enrichment was performed using a Monte Carlo simulation, which randomly generates start positions for the particular set of features and calculates the proportion of that feature that would be enriched in a given DRIP-chip profile if the feature were distributed at random (Schulze et al. 2009). Five hundred simulations were run per feature for each DRIP-chip replicate to obtain mean and standard deviation values. These values were used to calculate the cumulative probability (P) on a normal distribution of seeing a score lower than the observed value by chance.

### 5.2.4 DRIP-chip visualization

CHROMATRA plots were generated as described previously (Hentrich et al. 2012). Relative occupancy scores for each transcript were binned into segments of 150 bp. Transcripts were sorted by their length, transcriptional frequency or GC content and aligned by their Transcription Start Sites (TSS). For transcriptional frequency transcripts were grouped into five classes according to their transcriptional frequency described by Holstege *et al* 1998. For GC content transcripts were grouped into four classes according to their GC content obtained from BioMart (Kinsella et al. 2011). Average protein-coding gene, tRNA or snoRNA profiles were generated by averaging all the probes whose starts sites were encompassed by the feature of interest. For averaging ORFs, corresponding probes were split into 40 bins while 1500 bp of UTRs and their

probes were split into 20 bins. For smaller features like tRNAs and snoRNAs corresponding probes were split into only 3 bins. Average enrichment scores were calculated using in house scripts that average the score of all the probes encompassed by the feature.

### 5.2.5 Gene expression microarray

Gene expression microarray data is available at GEO GSE46652 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46652). Strains harboring the RNase H1 over-expression plasmid or empty vector were grown in SC-Leucine at 30°C. All profiles were generated in duplicate. Total RNA was isolated from 1 $OD_{600}$ using a RiboPure Yeast kit (A&B Applied Biosystems), amplified, labeled, fragmented using a Message-Amp$^{TM}$ III RNA Amplification Kit (A&B Applied Biosystems) and hybridized to a GeneChIP Yeast Genome 2.0 microarray using the GeneChip Hybridization, Wash, and Stain Kit (Affymetrix). Arrays were scanned by the Gene Chip Scanner 3000 7G and expression data was extracted using Expression Console™ Software (Affymetrix) with the MAS5.0 statistical algorithm. All arrays were scaled to a median target intensity of 500. A minimum cut off of p-value of 0.05 and signal strength of 100 across all samples were implemented and only transcripts that had over a 2-fold change in the RNase H over-expression strain compared to wild type were considered significant. The correlation between duplicate biological samples was: control (r=0.9955), RNase H over-expression (r=0.9719). For statistical analysis, GC content, transcription frequencies and antisense association were analyzed as for DRIP-chip analysis.

### 5.2.6 Yeast chromosome spreads

Cells were grown to mid-log phase in YEPD rich media at 30$^{o}$C and washed in spheroplasting solution (1.2 M sorbitol, 0.1 M potassium phosphate, 0.5 M $MgCl_2$, pH 7) and digested in spheroplasting solution with 10 mM DTT and 150 µg/mL Zymolase 20T at 37$^{o}$C for 20 minutes similar as previously described (Michaelis et al. 1997). The digestion was halted by addition of ice-cold stop solution (0.1 M MES, 1 M sorbital, 1 mM EDTA, 0.5 mM $MgCl_2$, pH 6.4) and spheroplasts were lysed with 1% vol/vol Lipsol and fixed on slides using 4% wt/vol paraformaldehyde/3.4% wt/vol sucrose (Klein et al. 1992). Chromosome spread slides were incubated with the mouse monoclonal antibody S9.6 (1 µg/mL in blocking buffer of 5% BSA, 0.2% milk and 1x PBS). The slides were further incubated with a secondary Cy3-conjugated goat anti-mouse antibody (Jackson Laboratories, #115-165-003, diluted 1:1000 in blocking buffer). For each replicate, at least 100 nuclei were visualized and manually counted to obtain the fraction with detectable DNA:RNA hybrids. Each mutant was assayed in triplicate. Mutants were compared to wild type by the Fisher's exact test. To correct for multiple hypothesis testing, we implemented a cut off of $p<0.01$ divided by the total number of mutants compared to wild type, meaning mutants with $p<0.00024$ were considered significantly different from wild type.

### 5.2.7 BPS sensitivity assay

Ten fold serial dilutions of each strain were spotted on 90 µM bethophenanthroline disulfonate (BPS) plates with $FeSO_4$ concentrations of 0, 2.5, 20 or 100 mM and grown at 30$^{o}$C for 3 days (Berthelet et al. 2010).

## 5.3 Results

### 5.3.1 The genomic distribution of DNA:RNA hybrids

DNA:RNA hybrids have been previously immunoprecipitated at specific genomic sites such as rDNA, selected endogenous loci, and reporter constructs (El Hage et al. 2010, Mischo et al. 2011). Subsequently, DRIP coupled with deep sequencing in human cells has demonstrated the prevalence of R loops at CpG island promoters with high GC skew (Ginno et al. 2013). To investigate the global profile of DNA:RNA hybrid prone loci in a tractable model, we performed genome-wide DRIP-chip analysis of wild type *S. cerevisiae* (ArrayExpress E-MTAB-1656 https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1656) using the S9.6 monoclonal antibody which specifically binds DNA:RNA hybrids, as characterized previously (Boguslawski et al. 1986, Hu et al. 2006). DRIP-chip profiles were generated in duplicate (spearman's r=0.78 when comparing each of over 2 million probes after normalization and data smoothing) and normalized to a no antibody control.

Overall, our DRIP-chip profiles identified several previously reported DNA:RNA hybrid prone sites including the rDNA locus and telomeric repeat regions (**Figure 5.1**) (Balk et al. 2013, El Hage et al. 2010, Luke et al. 2008, Pfeiffer et al. 2013). DNA:RNA hybrids were also observed at 1217 open reading frames (ORFs) (containing greater than 50% of probes above the threshold of 1.5 and found in both wild type replicates) (**Figure 5.2A**). These were generally shorter in length than average (p=4.29e-58), highly transcribed (Wilcoxon rank sum test p=2.21e-6), and had higher GC content (p=2.52e-50) (**Figure 5.2B-D**). Importantly, despite the correlation between DNA:RNA hybrid association and transcriptional frequency, the wild type DRIP-chip

profiles compared to the localization profile of the RNA polymerase II subunit Rpb3 revealed very low correlation (r=0.0097; (Aristizabal et al. 2013)). This suggests that the DRIP-chip method was not unduly biased towards the short DNA:RNA hybrids that could theoretically have been captured within active transcription bubbles. Importantly, because genes with high GC content also have high transcriptional frequencies, it is not clear from our findings whether GC content or transcriptional frequency contributed more to DNA:RNA hybrid forming potential. Furthermore, we observe that DNA:RNA hybrid prone loci do not encode for mRNA transcripts with particularly long half-lives (**Figure 5.3**), suggesting that the act of transcription is vital to DNA:RNA hybrid formation and supporting the notion of co-transcriptional hybrid formation as the major source of endogenous DNA:RNA hybrids.

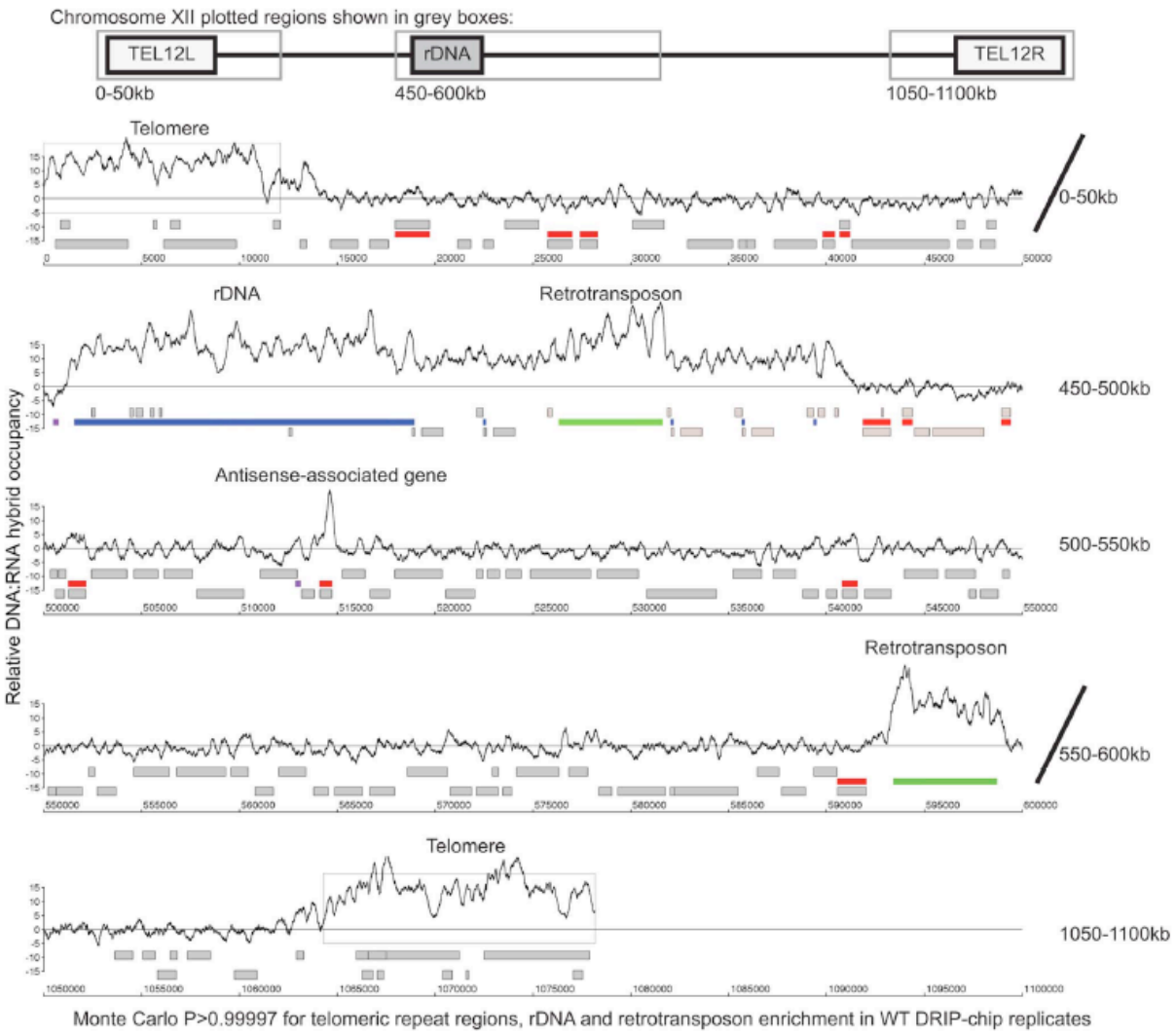



**Figure 5.1 Genome-wide profile of DNA:RNA hybrids in wild type yeast revealed enrichment at rDNA, telomeres, retrotransposons and a subset of genes.**
DRIP-chip chromosome plot of DNA:RNA hybrids in the rDNA region and telomeric ends of chromosome XII. The black line represents the average of two wild type replicates. Bars indicate ORFs (grey), rDNA (blue), retrotransposons (green) or genes associated with an antisense transcript (red) (Xu et al. 2011, Yassour et al. 2010). Grey boxes delineate telomeric repeat regions. Y-axis indicates relative occupancy of DNA:RNA hybrids. X-axis indicates chromosomal coordinates. P indicates probability of observing a number of enriched features by random chance below what was observed (P>0.99997).

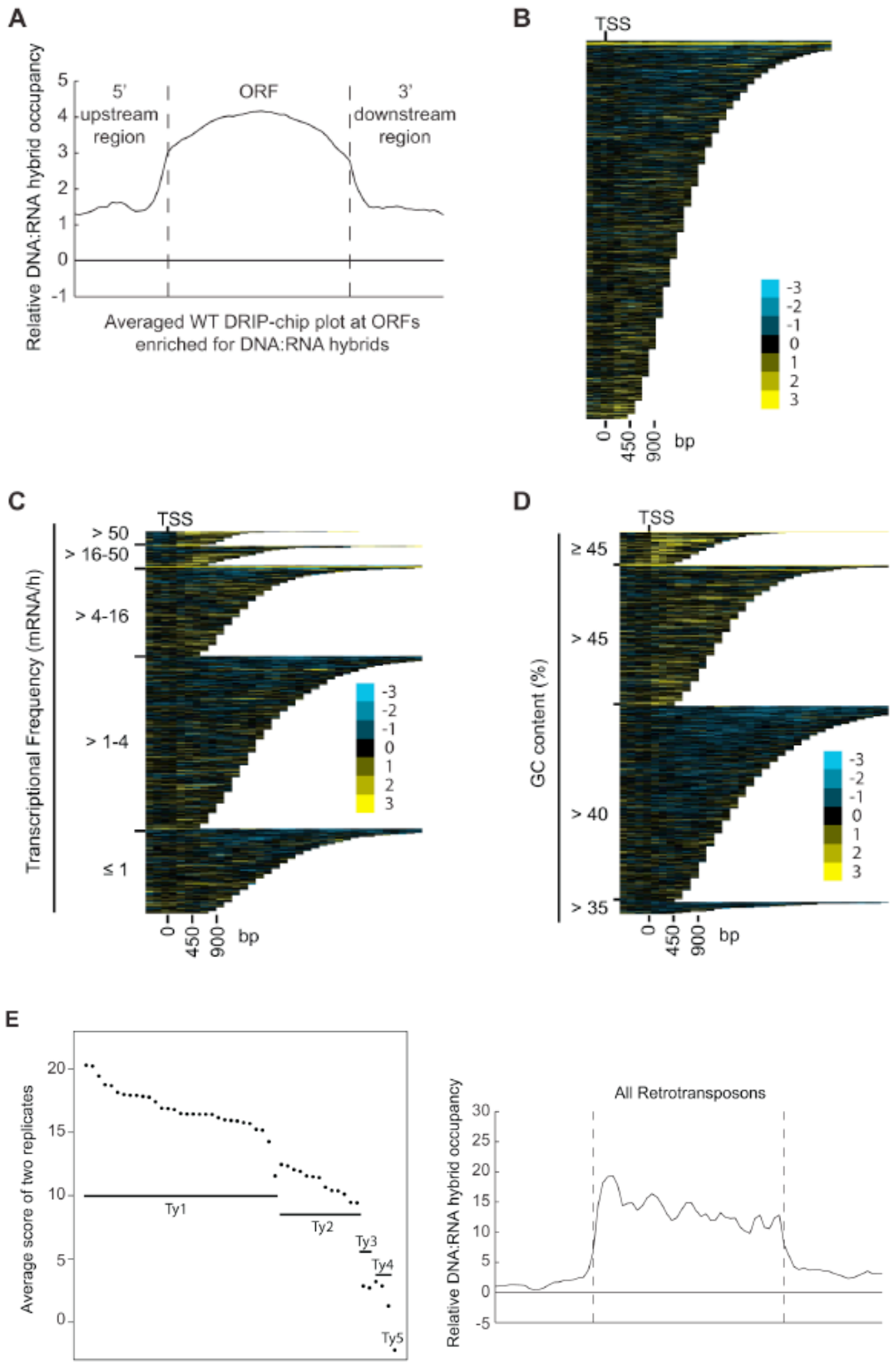

A
Relative DNA:RNA hybrid occupancy
5
4
3
2
1
0
-1
5'
upstream
region
ORF
3'
downstream
region
Averaged WT DRIP-chip plot at ORFs
enriched for DNA:RNA hybrids
B
TSS
-3
-2
-1
0
1
2
3
0
450
900
bp
C
Transcriptional Frequency (mRNA/h)
TSS
> 50
> 16-50
> 4-16
> 1-4
≤ 1
-3
-2
-1
0
1
2
3
0
450
900
bp
D
GC content (%)
TSS
≥ 45
> 45
> 40
> 35
-3
-2
-1
0
1
2
3
0
450
900
bp
E
Average score of two replicates
20
15
10
5
0
Ty1
Ty2
Ty3
Ty4
Ty5
All Retrotransposons
Relative DNA:RNA hybrid occupancy
30
25
20
15
10
5
0
-5

**Figure 5.2 Characteristics of genes enriched for DNA:RNA hybrids.**
Average gene profile of DNA:RNA hybrids at ORFs enriched for DNA:RNA hybrids under wild type conditions. (C-E) CHROMATRA plots of DNA:RNA hybrid distribution along genes sorted by their length (C), grouped into five transcriptional frequency categories as per (Holstege et al. 1998) (D) or grouped into four GC content categories (E). Genes were aligned by their TSSs. (F) The average DNA:RNA hybrid score at Ty1, Ty2, Ty3, Ty4 and Ty5 retrotransposons in the left panel shows higher enrichment at Ty1 and Ty2 retrotransposons. The average profile of DNA:RNA hybrids at all retrotransposons under wild type conditions is shown in the right panel.

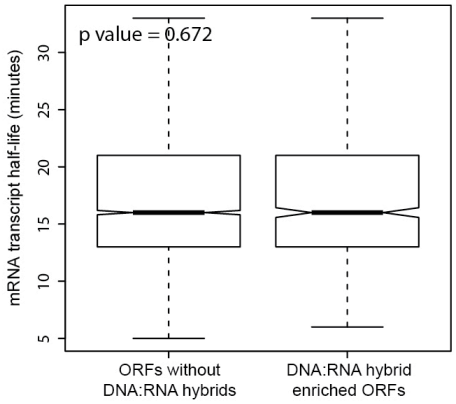



**Figure 5.3 Genes enriched for DNA:RNA hybrids did not produce mRNAs with longer half-lives.**
Box plots comparing the distribution of mRNA half-lives of ORFs enriched for DNA:RNA hybrids versus ORFs not enriched for DNA:RNA hybrids. The p values calculated by the Wilcoxon rank sum test are shown.

Our data also revealed DNA:RNA hybrids highly associated with Ty1 and Ty2 subclasses of retrotransposons (**Figure 5.1F**). Consistent with our findings at ORFs, the levels of DNA:RNA hybrids correspond well with the known levels of expression of these elements. In general, Ty1 which constitutes one of the most abundant transcripts in the cell has the highest levels of DNA:RNA hybrids. Ty3 and Ty4 that are only slightly expressed have much lower levels of hybrids, and the lone Ty5 retrotransposon, which is transcriptionally silent is not enriched for DNA:RNA hybrids (**Figure 5.1F**)(Clark et al. 1988, Hug and Feldmann 1996, Ke et al. 1997). In

contrast to the trends observed with ORFs, GC content in retrotransposons is not highly correlated with the levels of expression, suggesting that expression is the main contributor to DNA:RNA hybrid formation. Specifically, Ty3 retrotransposons have the highest GC content but have only modest levels of expression and DNA:RNA hybrids.

### 5.3.2 DNA:RNA hybrids are significantly correlated with genes associated with antisense transcripts

Certain DNA:RNA hybrid enriched regions identified by our DRIP-chip analysis such as rDNA and retrotransposons are associated with antisense transcripts (Bierhoff et al. 2010, Servant et al. 2012). Therefore, we checked if this was a common feature of DNA:RNA prone sites by comparing our list of DNA:RNA prone loci to a list of antisense-associated genes (Yassour et al. 2010). Because the expression of antisense-associated transcripts may be highly dependent on environmental conditions, we based our analysis on a list of transcripts identified in S288c yeast grown to mid-log phase in rich media, which most closely mirrors the growth conditions of our cultures analyzed by DRIP-chip (Yassour et al. 2010). DNA:RNA hybrid enriched genes significantly overlapped with antisense-associated genes, suggesting that DNA:RNA hybrids may play a role in antisense transcript-mediated regulation of gene expression (Fisher's exact test p=1.03e-12) (**Figure 5.4A-C**).

RNase H overexpression reduces detectable levels of DNA:RNA hybrids in cytological screens and suppresses genomic instability associated with R loop formation presumably through the degradation of DNA:RNA hybrids (Nakama et al. 2012, Stirling et al. 2011, Wahba et al. 2011). To test for a potential role of DNA:RNA hybrids in antisense-mediated gene regulation, we

performed gene expression microarray analysis of an RNase H overexpression strain compared to an empty vector control (GEO GSE46652 http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46652). This identified genes that had increased mRNA levels (upregulated n=212) or decreased mRNA levels (downregulated n=88) as a result of RNase H overexpression. A significant portion of the genes with increased mRNA levels were antisense-associated (Fisher exact test p=2.9e-7) (**Figure 5.4D**) and tended to have high GC content, similar to DNA:RNA hybrid enriched genes in wild type. However, the genes with increased mRNA levels under RNase H overexpression and the antisense-associated genes enriched for DNA:RNA hybrids in our DRIP experiment both tended towards lower transcriptional frequencies (**Figure 5.4E**). These findings suggest that antisense-associated DNA:RNA hybrids moderate the levels of gene expression. Indeed, genes that were both modulated by RNase H overexpression and enriched for DNA:RNA hybrids were all found to be antisense-associated (**Figure 5.4F**).

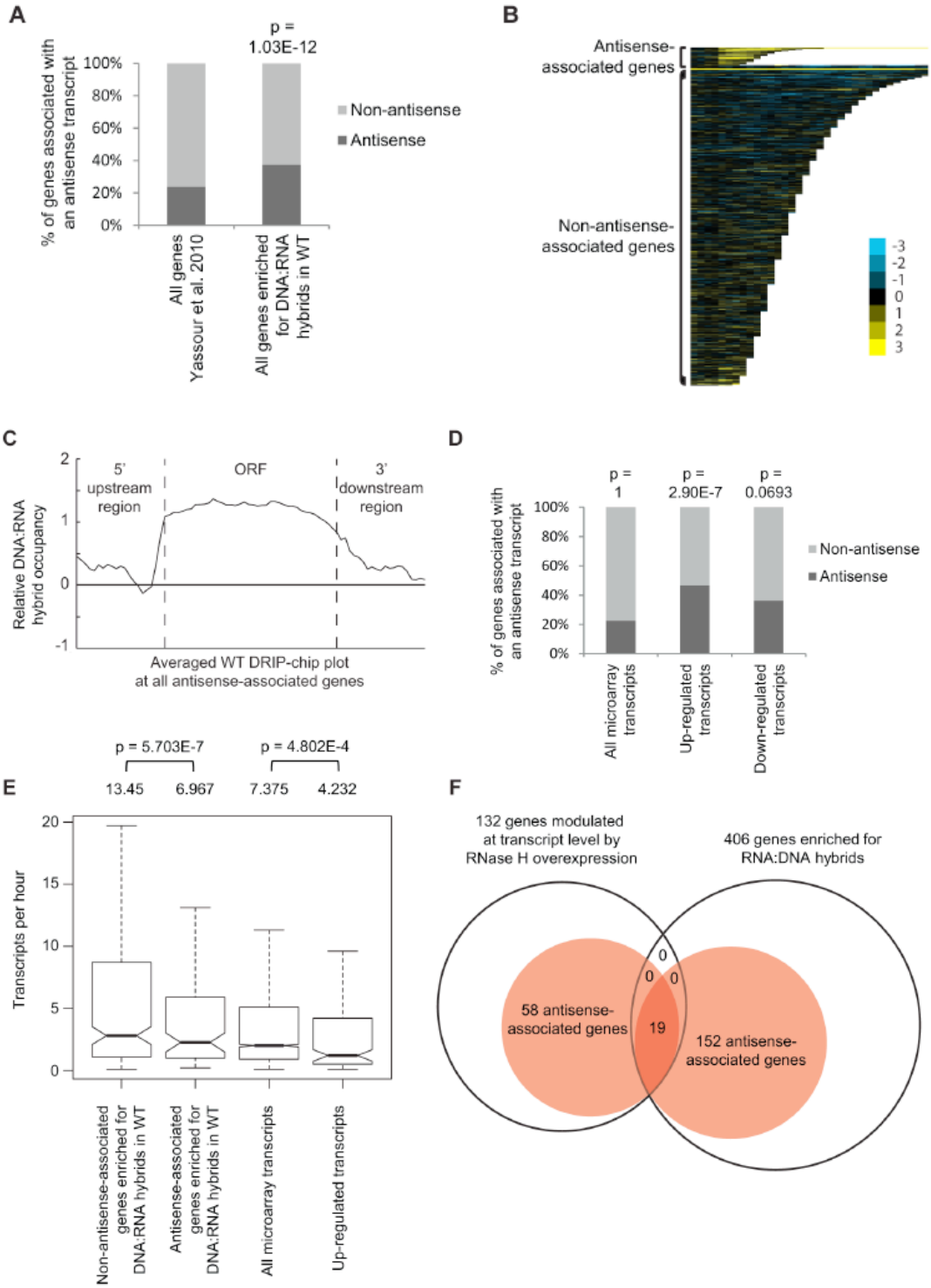

A
p = 1.03E-12
% of genes associated with an antisense transcript
100%
80%
60%
40%
20%
0%
Non-antisense
Antisense
All genes Yassour et al. 2010
All genes enriched for DNA:RNA hybrids in WT
B
Antisense-associated genes
Non-antisense-associated genes
-3
-2
-1
0
1
2
3
C
5' upstream region
ORF
3' downstream region
Relative DNA:RNA hybrid occupancy
2
1
0
-1
Averaged WT DRIP-chip plot at all antisense-associated genes
D
p = 1
p = 2.90E-7
p = 0.0693
% of genes associated with an antisense transcript
100%
80%
60%
40%
20%
0%
Non-antisense
Antisense
All microarray transcripts
Up-regulated transcripts
Down-regulated transcripts
E
p = 5.703E-7
p = 4.802E-4
13.45
6.967
7.375
4.232
Transcripts per hour
20
15
10
5
0
Non-antisense-associated genes enriched for DNA:RNA hybrids in WT
Antisense-associated genes enriched for DNA:RNA hybrids in WT
All microarray transcripts
Up-regulated transcripts
F
132 genes modulated at transcript level by RNase H overexpression
406 genes enriched for RNA:DNA hybrids
0
0
0
58 antisense-associated genes
19
152 antisense-associated genes

**Figure 5.4 Genes associated with DNA:RNA hybrids were significantly associated with antisense transcripts.**
(A) Antisense association of DNA:RNA hybrid-enriched genes in wild type. p-values indicate significant enrichment (Fisher's exact test) of antisense-associated genes among DNA:RNA hybrid-enriched genes compared to the Yassour et al. 2010 antisense-annotated dataset (Yassour et al. 2010). (B) CHROMATRA plots of DNA:RNA hybrid distribution along genes sorted by their length and separated by whether they are antisense associated or not. Genes were aligned by their TSSs. (C) Average gene profile of DNA:RNA hybrids at genes enriched for DNA:RNA hybrids and associated with antisense transcripts. (D) Genes with increased mRNA levels upon RNase H overexpression were significantly associated with antisense transcripts compared to all transcripts represented by the microarray. (E) Antisense-associated DNA:RNA hybrid-enriched genes in wild type have lower transcription frequency compared to non-antisense-associated DNA:RNA hybrid-enriched genes. Genes up-regulated at the transcript level by RNase H overexpression have lower transcription frequency compared to all genes on the expression microarray. Intervals indicate range of the 95% of genes closest to the average in each sample. Averages stated above each bar. P values indicate significant decrease in transcriptional frequency (Wilcoxon rank sum test). (F) Overlap between DNA:RNA hybrid-enriched genes and RNase H-modulated transcripts sorted by antisense association according to the Yassour et al. 2010 database. For genes that are both hybrid-enriched and modulated at the transcript level by RNase H overexpression, the antisense association (100%) is significantly higher (Fisher's exact test $p<2.2e-16$) than those of the parent datasets (37.4% for DNA:RNA hybrid-enriched genes, 43.9% for RNase H-modulated genes).

The mechanism underlying altered gene expression in cells overexpressing RNase H remains unclear. While the association with antisense transcription is compelling, alternative models exist. One possibility is that the stress of RNase H overexpression triggers gene expression programs that coincidentally are antisense regulated. We analyzed gene ontology (GO) terms enriched among genes whose expression was changed by RNase H overexpression. Consistent with previous work, genes for iron uptake and incorporation were strongly activated by RNase H overexpression ($p=2.21e-12$) (**Figure 5.5A**) and several of these iron transport genes (i.e. *FRE4*, *FRE2, FRE3*, *FET3*, *FET4*) are antisense-associated (Xu et al. 2011, Yassour et al. 2010) suggesting that loss of RNase H activates transcription of these genes by perturbing antisense-mediated regulation. Alternatively, changes in RNase H levels may increase the cellular iron

requirements since sensitivity to low iron concentration is associated with DNA damage and repair (Berthelet et al. 2010). To test this alternative hypothesis, we tested the RNase H deletion and *sen1-1* mutants for sensitivity to low iron conditions compared to a *fet3*Δ positive control (**Figure 5.5B**). The *sen1-1* mutant, RNase H depletion or overexpression did not induce sensitivity to low iron ruling out the possibility that the transcriptional response in cells overexpressing RNase H was a result of cellular iron requirement. Collectively, our DRIP-chip and microarray analysis suggest that DNA:RNA hybrids may be an important player in antisense-mediated gene regulation.

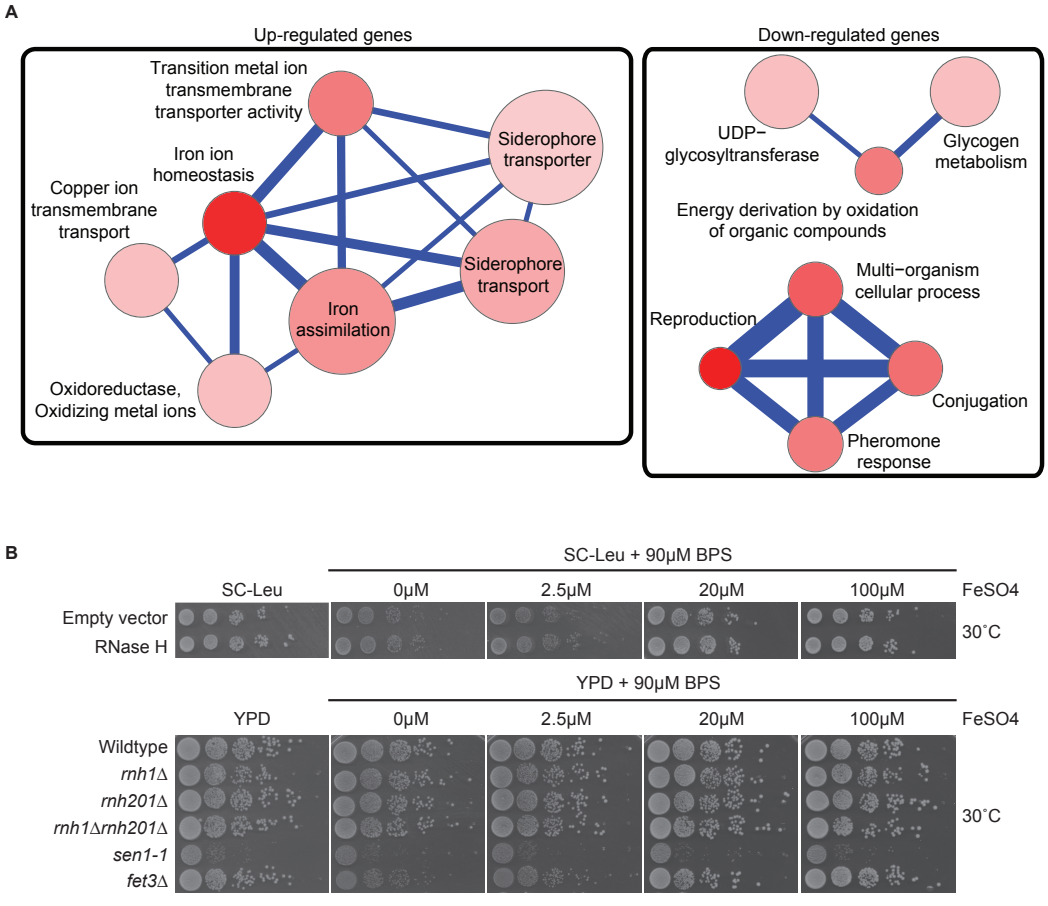


**Figure 5.5 Pathways altered at the transcript level by RNase H overexpression.**
(A) Gene Ontology term network of genes with increased (left) or decreased (right) mRNA levels upon RNase H overexpression. Node size indicates fold enrichment. Node color indicates the number of genes associated with each term (the darkest indicating the greatest number of genes associated). Edge thickness indicates the number of genes shared between terms. (B) Ten fold serial dilutions on plates containing the indicated amounts of FeSO4 and the iron chelator, Bathophenanthroline disulfonate (BPS). The mutant strains showed no difference when grown under normal conditions compared to conditions lacking iron, thus revealing no changes in the cellular iron requirement for these mutant strains.

### 5.3.3 Cytological profiling of RNA processing mutants for R loop formation

Transcription-coupled DNA:RNA hybrids have been shown to accumulate in a diverse set of transcription and RNA processing mutants involved in a wide range of transcription related processes (**Table 5.4**). To gain a broader understanding of factors involved in R loop formation, we performed a cytological screen of RNA processing, transcription and chromatin modification mutants for DNA:RNA hybrids using the S9.6 antibody. Importantly, previous work in our lab has shown that all of the mutants screened exhibit chromosome instability (CIN), which would be consistent with increased hybrid formation (Stirling et al. 2011). Significantly elevated hybrid levels were found in 22 of the 40 mutants tested compared to wild type, including a *SUB2* mutant which has been previously linked to R loop formation (**Figure 5.6**) (Gomez-Gonzalez et al. 2011). We also assayed some of the most well-characterized R-loop forming mutants, RNase H, Sen1 and Hpr1, as positive controls for elevated DNA:RNA hybrid levels (**Figure 5.6**).

## Table 5.4 List of yeast genes that affect DNA:RNA hybrid formation

| Yeast gene linked to DNA:RNA hybrid formation | Reference |
|---|---|
| Exosome and RNA degradation: *DIS3, RRP6, XRN1* | This study, (Luna et al. 2005, Wahba et al. 2011) |
| Helicase: *SEN1, SRS2* | (Mischo et al. 2011, Wahba et al. 2013) |
| mRNA cleavage and polyadenylation: *CLP1, CFT2, FIP1, PCF11, RNA14, RNA15, TRF4* | (Gavalda et al. 2013, Luna et al. 2005, Stirling et al. 2012) |
| mRNA export: *MEX67, MTR2, NAB2, NUP133, RNA1, SAC3, SRM1, SUB2, SUS1, THP1, YRA1* | This study, (Gallardo et al. 2003, Gonzalez-Aguilera et al. 2008, Jimeno et al. 2002, Luna et al. 2005, Stirling et al. 2012) |
| Other processes: *ESC2, KAE1, PSH1, STS1* | This study |
| RNA Polymerase II transcription and chromatin modification: *LEO1, MED12, MED13, MOT1, NPL3, RTT103, SDS3, SIN3, SPT2, TAF5* | This study, (Santos-Pereira et al. 2013, Sikdar et al. 2008, Wahba et al. 2011, Wahba et al. 2013) |
| RNase H: *RNH201, RNH1* | This study, (Stirling et al. 2012, Wahba et al. 2011) |
| rRNA processing factors: *DBP6, DBP7, IMP4, RPF1, SNU13, SNU66* | This study |
| Splicing: *MUD2, SNU114, PRP31, YHC1, SNU13, SNU66* | This study |
| THO transcription elongation: *THO2, HPR1, MFT1, THP2* | (Chavez et al. 2001, Huertas and Aguilera 2003, Stirling et al. 2012) |
| Topoisomerase: *TOP1* | (El Hage et al. 2010) |

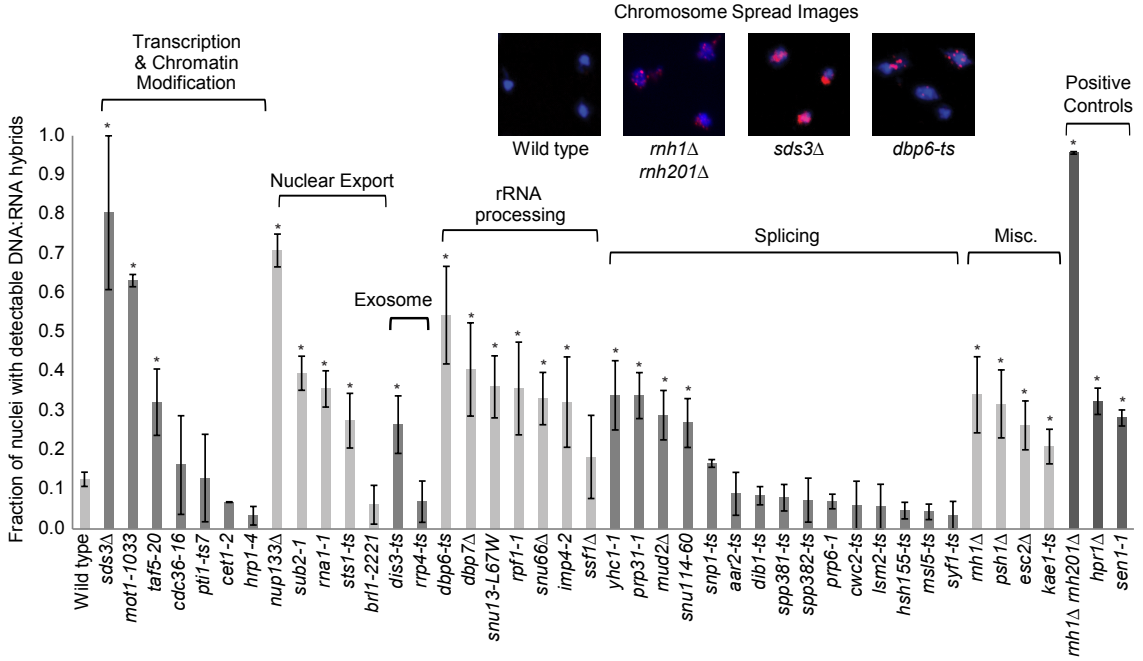


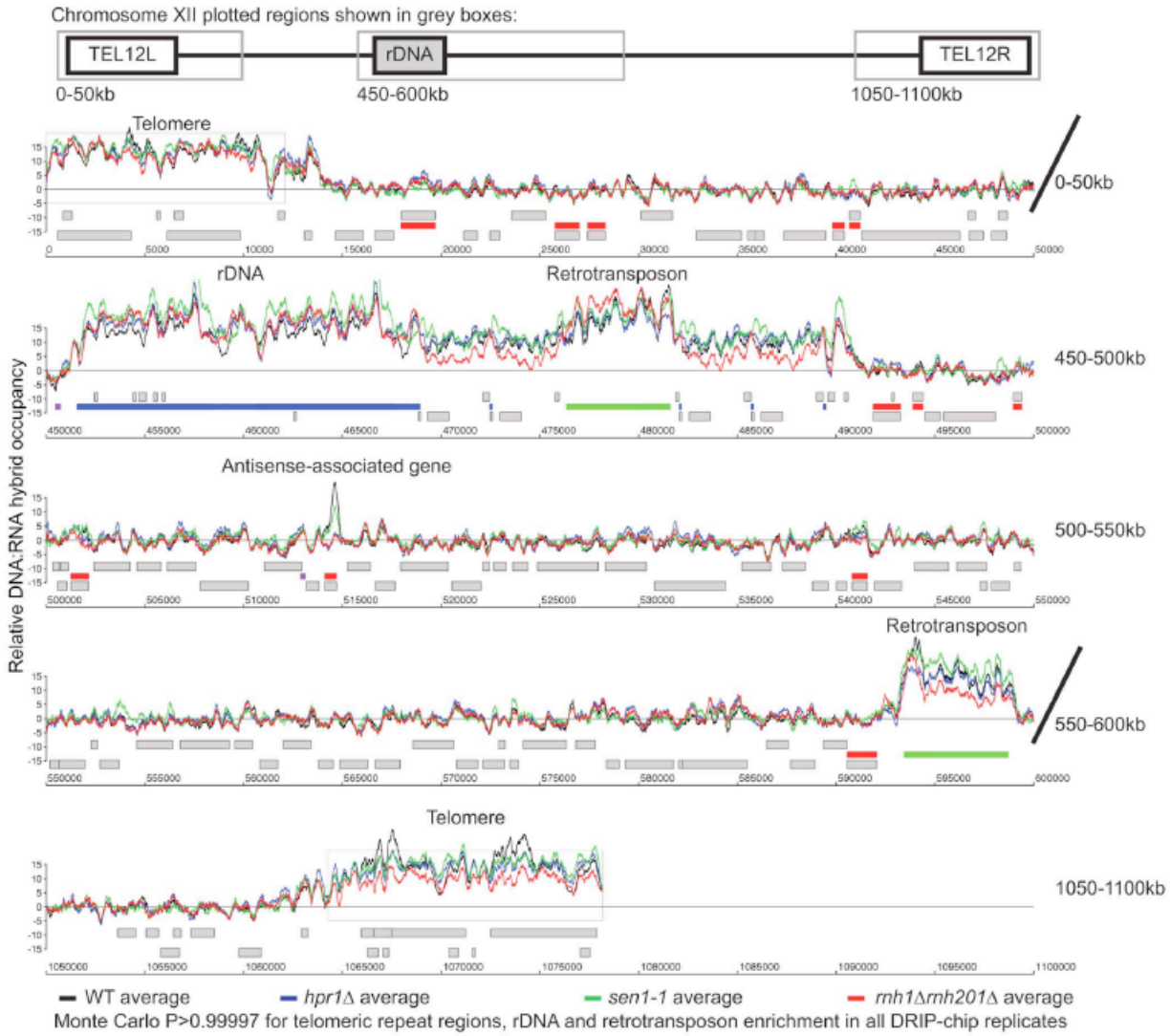
**Figure 5.6 DNA:RNA hybrid cytological screen revealed high DNA:RNA hybrid levels in RNA processing and chromatin modification mutants.**
Asterisks indicate mutants with significantly increased levels of DNA:RNA hybrids compared to wild type ($p<0.00024$). Error bars indicate standard error of the mean. Representative chromosome spreads are shown: blue stain is DNA (DAPI) and the red foci are DNA:RNA hybrids.

In our screen, we detected hybrids in mutants affecting several pathways linked to DNA:RNA hybrid formation such as transcription, nuclear export and the exosome (**Figure 5.6**) (**Table 5.4**). Consistent with findings in metazoan cells, we also observed hybrid formation in some splicing mutants (**Figure 5.6**) (**Table 5.4**) (Li and Manley 2005). Several rRNA processing mutants were enriched for DNA:RNA hybrids (7 out of the 22 positive hits), likely due to DNA:RNA hybrid accumulation at rDNA genes, a sensitized hybrid formation site (**Figure 5.1A**) (El Hage et al. 2010). It is possible that, as seen in mRNA cleavage and polyadenylation mutants, DNA:RNA hybrid formation may contribute to their CIN phenotypes (Stirling et al. 2012). Currently, there are 52 yeast genes whose disruption have been found to lead to DNA:RNA hybrid accumulation, 21 of which were newly identified by our screen (**Table 5.4**). The success of this small-scale screen suggests that most RNA processing pathways suppress hybrid formation to some degree and that many DNA:RNA hybrid forming mutants remain undiscovered.

### 5.3.4 DRIP-chip profiling of R loop forming mutants

To better understand the mechanism by which cells regulate DNA:RNA hybrids, we performed DRIP-chip analysis of *rnh1*Δ*rnh201*Δ, *hpr1*Δ, and *sen1-1* mutants in order to determine if these contribute differentially to the DNA:RNA hybrid genomic profile. The *rnh1*Δ*rnh201*Δ, *hpr1*Δ, and *sen1-1* mutants are particularly interesting because they have well established roles in the regulation of transcription dependent DNA:RNA hybrid formation. Our DRIP-chip profiles

revealed that, similar to wild type profiles, the mutant profiles were enriched for DNA:RNA hybrids at rDNA, telomeres, and retrotransposons (**Figure 5.7A**). The *rnh1*Δ*rnh201*Δ, *hpr1*Δ, and *sen1-1* mutants also exhibited DNA:RNA hybrid enrichment in 1206, 1490 and 1424 ORFs respectively compared to the 1217 DNA:RNA hybrid enriched ORFs identified in wild type. Interestingly, in addition to the similarities described above, our profiles also identified differential effects of the mutants on the levels of DNA:RNA hybrids. In particular, we observed that deletion of *HPR1* resulted in higher levels of DNA:RNA hybrids along the length of most ORFs with a preference for longer genes compared to wild type. This observation is consistent with Hpr1's role in bridging transcription elongation to mRNA export and its localization at actively transcribed genes (Gomez-Gonzalez et al. 2011, Huertas and Aguilera 2003, Strasser et al. 2002, Zenklusen et al. 2002), (**Figure 5.7B-D**). In contrast, mutating *SEN1* resulted in higher levels of DNA:RNA hybrids at shorter genes (**Figure 5.7B-C**), which is consistent with Sen1's role in transcription termination particularly for short protein-coding genes (Mischo et al. 2011, Rondon et al. 2009, Steinmetz et al. 2006). The *rnh1*Δ*rnh201*Δ mutant revealed higher levels of DNA:RNA hybrids at highly transcribed and longer genes (**Figure 5.7B-C**) which is supported by a wealth of evidence of RNase H's role in suppressing R loops in long genes to prevent collisions between transcription and replication machineries (Aguilera and Garcia-Muse 2012, Helmrich et al. 2011).




Monte Carlo P>0.99997 for telomeric repeat regions, rDNA and retrotransposon enrichment in all DRIP-chip replicates

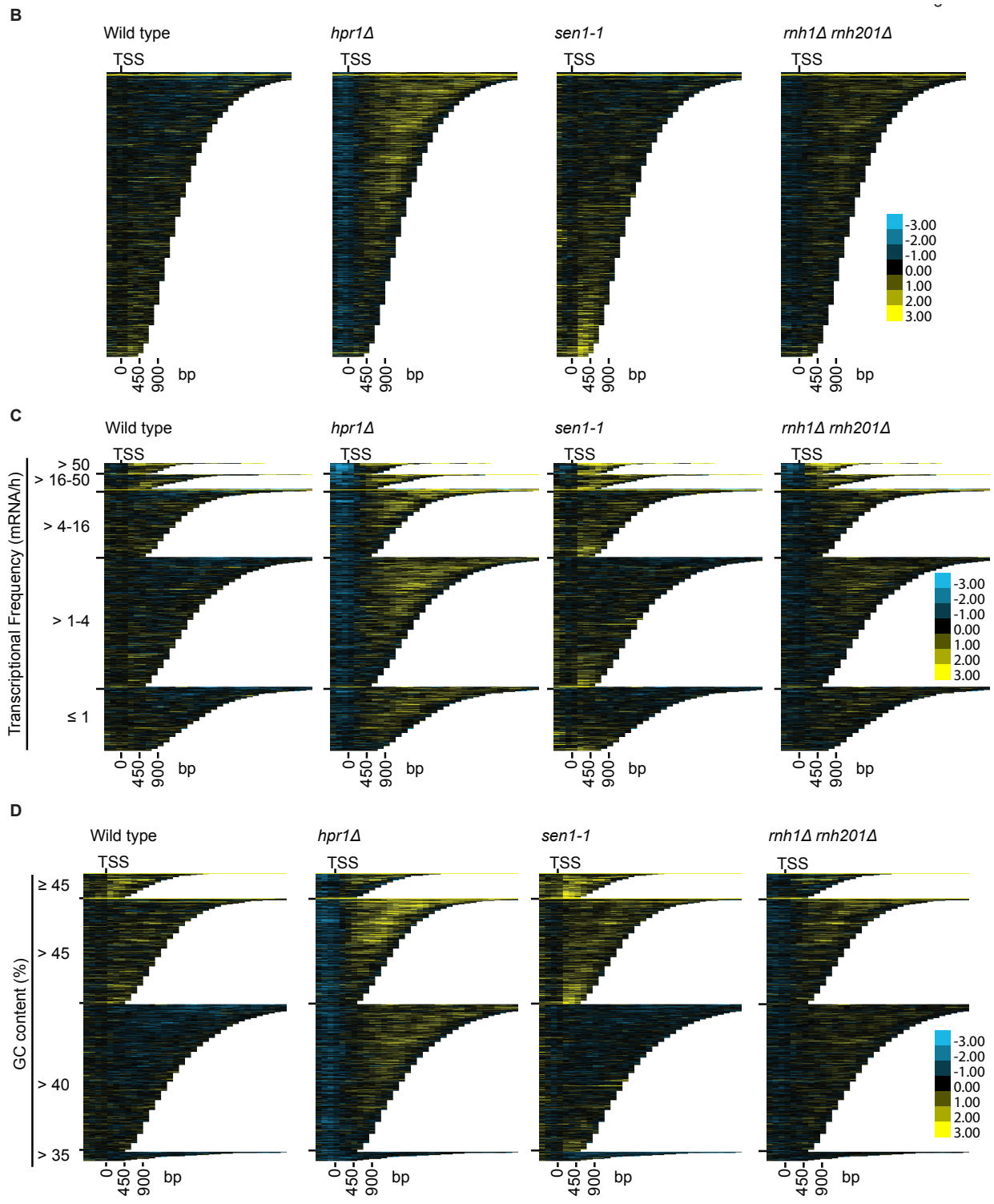

B
Wild type
hpr1Δ
sen1-1
rnh1Δ rnh201Δ
TSS
0
450
900
bp
-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00
C
Transcriptional Frequency (mRNA/h)
> 50
> 16-50
> 4-16
> 1-4
≤ 1
D
GC content (%)
≥ 45
> 45
> 40
> 35

**Figure 5.7 Genome-wide profile of DNA:RNA hybrids in mutants revealed mutant specific DNA:RNA prone loci.**
(A) DRIP-chip chromosome plot of DNA:RNA hybrids in wild type, *rnh1*Δ*rnh201*Δ, *hpr1*Δ and *sen1-1* at chromosome XII. The average of two replicates per strain is shown. Bars indicate ORFs (grey), rDNA (blue), retrotransposons (green) or genes associated with an antisense transcript (red) (Xu et al. 2011, Yassour et al. 2010). Grey boxes delineate telomeric repeat regions. Y-axis indicates relative occupancy of DNA:RNA hybrids. X-axis indicates chromosomal coordinates. P indicates probability of observing a number of enriched features below what was observed (P>0.99997). (B-D) CHROMATRA plots of DNA:RNA hybrid distribution along genes sorted by their length (B) grouped into five transcriptional frequency categories as per (Holstege et al. 1998) (C) or grouped into four GC content categories (D). Genes were aligned by their TSSs.

Further inspection of our profiles also revealed that *rnh1*Δ*rnh201*Δ and *sen1-1* mutants but not the *hpr1*Δ mutant had increased levels of DNA:RNA hybrids at tRNA genes (two tailed unpaired Wilcox test p=1.56e-19 in the *rnh1*Δ*rnh201*Δ mutant and 1.68e-15 in the *sen1-1* mutant) (**Figure 5.8A-C**) and this was confirmed by DRIP-quantitative PCR (qPCR) of two tRNA genes in wild type and *rnh1*Δ*rnh201*Δ (**Figure 5.9**). Because tRNAs are transcribed by RNA polymerase III, this observation indicates that Hpr1 is primarily involved in the regulation of RNA polymerase II specific DNA:RNA hybrids, while RNase H and Sen1 have roles in a wider range of transcripts. Mutation of *SEN1* also led to increased levels of DNA:RNA hybrids at snoRNA (two tailed unpaired Wilcox test p=1.81e-6) (**Figure 5.8D-F**) consistent with its role in 3' end processing of snoRNAs (Ursic et al. 1997).

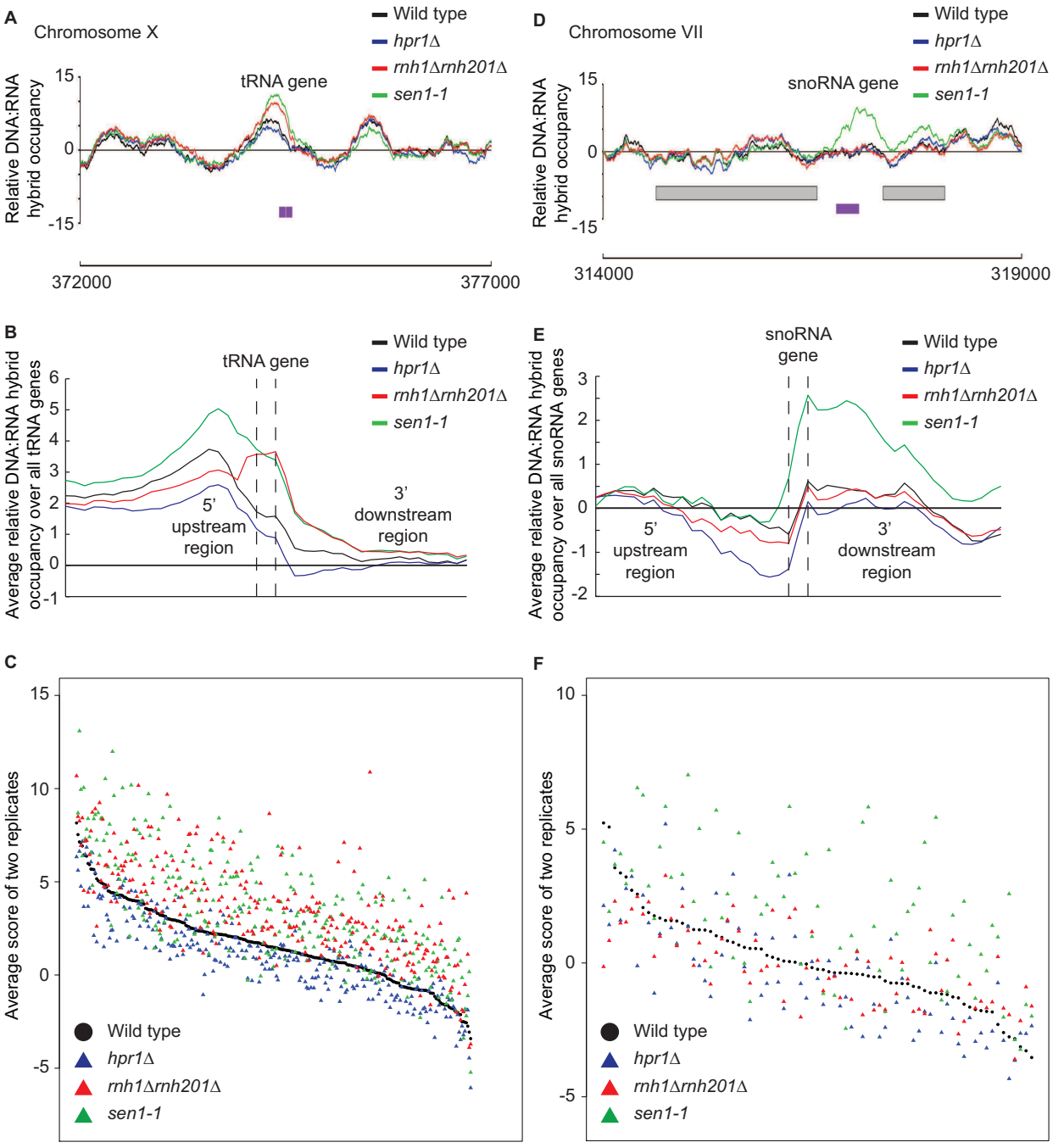



**Figure 5.8 RNase H and Sen1 mutants displayed elevated levels of DNA:RNA hybrids at tRNA and snoRNA genes.**

(A) Sample plot of relative DNA:RNA hybrid occupancy at a tRNA gene on chromosome X. For A and D, Colored lines represent the average enrichment of the indicated strains. Purple bars indicate the tRNA or snoRNA genes respectively and gray boxes represent ORFs. (B) Average profile of DNA:RNA hybrids at all tRNAs. (C) Average DNA:RNA hybrid score at each tRNA. (D) Sample plot of relative DNA:RNA hybrid occupancy at a snoRNA gene on chromosome

VII. (E) Average profile of DNA:RNA hybrids at all snoRNAs. (F) Average DNA:RNA hybrid score at each snoRNA. P indicates probability of observing a number of enriched features below what was observed (P>0.99997).

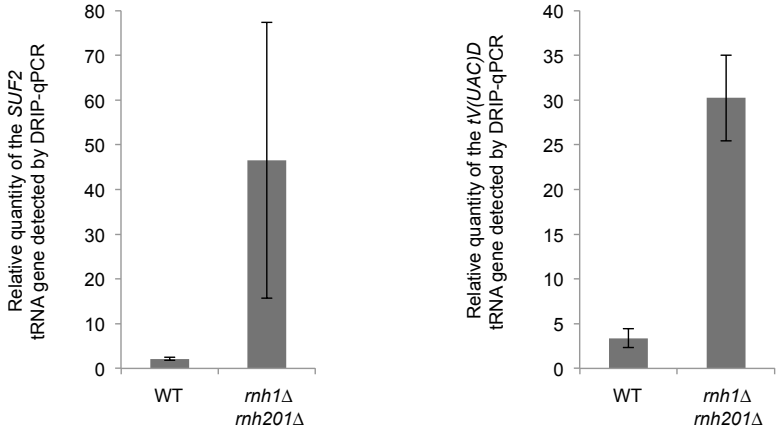



**Figure 5.9 Confirmation of DRIP-chip results by DRIP-qPCR.**
Relative quantities of (A) *SUF2* tRNA gene and (B) *tV(UAC)D* tRNA gene detected in WT or *rnh1*Δ*rnh201*Δ as detected by DRIP-quantitative PCR (qPCR). Error bars indicate standard deviation.

## 5.4 Discussion

### 5.4.1 The genomic profile of DNA:RNA hybrids

Identifying the landscape of genomic loci predisposed to DNA:RNA hybrids is of fundamental importance to delineating mechanisms of hybrid formation and the contributions of various cellular pathways. Although our profiles depend on the specificity of the anti-DNA:RNA hybrid S9.6 monoclonal antibody, this aspect has been well characterized (Boguslawski et al. 1986) and several of our observations are consistent with what has been reported in the literature. Locus specific tests showed that DNA:RNA hybrids occur more frequently at genes with high transcriptional frequency and GC content (Chavez et al. 2001, Gomez-Gonzalez et al. 2011,

Mischo et al. 2011). Moreover, in *rnh201*Δ cells, there is an inverse relationship between GC content and gene expression levels, suggesting that DNA:RNA hybrids accumulate at regions of high GC content and block transcription in the absence of RNase H (Arana et al. 2012). Our work extends the knowledge of DNA:RNA hybrids from a few locus-specific observations to show that, in wild type, there are potentially hundreds of hybrid prone genes that tend to be shorter in length, frequently transcribed and high in GC content (El Hage et al. 2010, Gomez-Gonzalez et al. 2011, Li and Manley 2005). The latter is consistent with recent studies in human cells that demonstrated that genomic regions with high GC skew are prone to R loop formation, which plays a regulatory role in DNA methylation (Ginno et al. 2012, Ginno et al. 2013). However, while we determined the relationship between GC content and DNA:RNA hybrid formation, we were unable to do the same analysis for GC skew, likely due to the low level of GC skew and lack of DNA methylation in *Saccharomyces*. This is unsurprising since the best characterized functional element associated with GC skew, CpG island promoters (Ginno et al. 2012, Ginno et al. 2013), are not found in yeast. Importantly, our findings at retrotransposons support the notion that expression levels and not GC content contribute more to DNA:RNA hybrid forming potential. Additionally, DRIP-chip analysis of wild type cells identified hybrid enrichment at rDNA, retrotransposons, and telomeric regions. Along with previous studies, our DRIP-chip analysis confirms that rDNA is a hybrid-prone genomic site and suggests that many factors of rRNA processing and ribosome assembly suppress potentially damaging rDNA:rRNA hybrid formation (El Hage et al. 2010, Wahba et al. 2011). The presence of TERRA-DNA hybrids at telomeres is supported by our observation of significant hybrid signal at telomeric repeat regions across all DRIP-chip experiments.

### 5.4.2 Antisense association of DNA:RNA hybrids

The DRIP-chip dataset is a resource for future studies seeking to elucidate the localization of DNA:RNA hybrids across antisense-associated regions and the impact of DNA:RNA hybrid removal on genome-wide transcription. We observed that genes associated with antisense transcripts were significantly enriched for DNA:RNA hybrids and modulated at the transcript level by RNase H overexpression. Antisense regulation has been reported at mammalian rDNA and yeast Ty1 retrotransposons, loci that were also enriched for DNA:RNA hybrids in our DRIP-chip (Bierhoff et al. 2010, Servant et al. 2012). The role of DNA:RNA hybrids and RNase H in antisense regulation is currently unclear. However, there are several non-exclusive models of antisense gene regulation. One model proposes that the physical presence of the antisense transcripts is crucial to antisense gene regulation. For instance, *trans*-acting antisense transcripts have been shown to control Ty1 retrotransposon transcription, reverse transcription and retrotransposition (Matsuda and Garfinkel 2009). Another study has further shown that *trans*-acting antisense transcripts that only overlap with the sense strand promoter can block sense transcription, potentially by hybridizing with the non-template DNA strand (Camblong et al. 2009). These suggest that antisense transcription in *cis* is not necessary as long as the antisense transcript is present. It is possible that DNA:RNA hybrids may be formed by the antisense or the sense transcript with genomic DNA. Moreover, DNA:RNA hybrids may play a functional role in antisense transcription regulation as shown by antisense-associated genes both enriched for DNA:RNA hybrids and affected transcriptionally by RNase H overexpression. Experiments comparing the ratio of antisense versus sense transcripts and determining the amount of DNA:RNA hybrid formation by either transcript under conditions known to regulate the

particular gene will further elucidate the role of RNase H and DNA:RNA hybrids in antisense regulation.

### 5.4.3 DRIP-chip analysis of hybrid-resolving mutants

Our investigation of mutant-specific DNA:RNA hybrid formation sites is consistent with the existing literature on Hpr1, Sen1 and RNase H. Significantly, the *hpr1*Δ and *rnh1*Δ*rnh201*Δ mutants exhibited increased DNA:RNA hybrid levels along the length of long genes, while the *sen1-1* mutant exhibited increased DNA:RNA hybrid levels along the length of short genes (**Figure 5.7B**). This coheres with Hpr1's function in transcription elongation and mRNA export, and RNase H's role in preventing transcription apparatus and replication fork collisions, which carry greater consequence for long genes (Gomez-Gonzalez et al. 2011, Helmrich et al. 2011, Huertas and Aguilera 2003, Strasser et al. 2002, Zenklusen et al. 2002). In contrast, Sen1 is particularly important for transcription termination at short genes (Steinmetz et al. 2006).

Deletion of RNase H or mutations in *SEN1* resulted in increased hybrids at tRNA genes, suggesting that they are both required to prevent tRNA:DNA hybrid accumulation. Interestingly, a recent study found that the mRNA levels of genes encoding RNA polymerase III and proteins that modify tRNA are increased in an *rnh1*Δ*rnh201*Δ mutant (Arana et al. 2012), which may be in response to a lack of properly processed tRNA transcripts. The finding that both tRNA and snoRNA genes were enriched for hybrids in *sen1-1* highlights the role of Sen1 in RNA polymerase I, II and III transcription termination and transcript maturation (Kawauchi et al. 2008, Rondon et al. 2009, Ursic et al. 1997). More broadly, our data and the literature support

the notion that transcripts from RNA polymerases I, II and III can be subject to DNA:RNA hybrid formation especially in RNA processing mutant backgrounds.

### 5.4.4 Perspective

Factors regulating ectopic, genome destabilizing DNA:RNA hybrids are best characterized in yeast, although less is known about the functions of native R loop structures. The genome-wide maps of DNA:RNA hybrids presented here recapitulate the known sites of hybrid formation but also add important new insights to potential functions of R loops. Most importantly, we demonstrate the usefulness of DRIP profiling for detecting biologically meaningful differences in mutant strains. Therefore, DRIP profiling of yeast genomes in various mutant backgrounds will be key to understanding the causes and consequences of inappropriate R loop formation and how these are modulated by other cellular pathways.

# Chapter 6: Conclusion

Transcription is an essential process, whose tight control underlies the cell's ability to survive and maintain normal function. In eukaryotes, the RNAPII-CTD functions as a recruiting platform for regulatory and RNA-processing factors making it central to transcription regulation (Allison et al. 1985, Corden et al. 1985, Heidemann et al. 2012). Although much is known about the RNAPII-CTD, the picture is far from complete. For example, it remains unclear how CTD length contributes to normal RNAPII function, and whether the RNAPII-CTD coordinates events in a gene-specific manner. Additionally, while many RNAPII-CTD modifying enzymes have been identified, the full repertoire of CTD modifying enzymes, as well as their specificities, regulation, and functional consequences remain to be fully characterized. This dissertation focused on uncovering how alterations to CTD length affect RNAPII's ability to synthesize different types of transcripts. Furthermore, this thesis explored the role and regulation of the RNAPII-CTD phosphatase, Fcp1, during RNAPII-dependent transcription. Finally, the impact of transcription on genome stability was investigated with a focus on the effect of CTD length on retrotransposon mobility, and of transcription-associated factors on DNA:RNA hybrid occupancy.

The repetitive nature of the RNAPII-CTD has puzzled researchers since its discovery, and despite numerous investigations it remains unclear why its length varies across species. Nonetheless, work in mammalian systems has provided evidence that length is likely the most important contributor to normal CTD function (Chapman et al. 2005). For instance, truncating the RNAPII-CTD results in alterations to growth, transcription regulation, and transcription

processing in a number of species (Allison et al. 1988, Bartolomei et al. 1988, de la Mata and Kornblihtt 2006, Hsin et al. 2014, Litingtung et al. 1999, McCracken et al. 1997a, Nonet et al. 1987, Rosonina and Blencowe 2004, Ryan et al. 2002, Schneider et al. 2010, Suh et al. 2010). Chapter 2 of this thesis built on these observations and described additional CTD length-dependent phenotypes in *Saccharomyces cerevisiae,* which are integral to understanding the full spectrum of CTD function. Specifically, by profiling mRNA levels and RNAPII occupancy on a series of RNAPII-CTD truncation mutants, we identified genes whose expression was dependent on RNAPII-CTD length, under normal growth conditions. This included Ty1 retrotransposons, which are described in chapter 3 and whose mobility was increased when CTD length was reduced.

We note that the phenotypes reported here are most likely due to gross alterations to CTD length, rather than a result of alteration caused by loss of non-consensus repeats. More specifically, the *S. cerevisiae* CTD is primarily composes of consensus sequences and the few non-consensus repeats observed are found at the most C-terminal region of the CTD (repeats 1, 16, 21, 23, 24 and 25) (Allison et al. 1985). This means that most non-consensus repeats are loss in the *rpb1-CTD20* and *rpb1-CTD13* mutants, both of which behave like wild type in most of our assays. The only exception is the non-consensus repeat found at position 16 which is present in the *rpb1-CTD20* mutant but absent in the *rpb1-CTD13* mutant, making it unclear if the phenotypes that first arise when the CTD is truncated to 13 repeats are due to changes in length or by loss of a specific function caused by loss of a non-consensus repeat. More detailed repeat mutation and substitution analysis will be required to differentiate these possibilities and determine the role, if any, of the non-consensus repeat at position 16.

The gene expression strategy described in chapter 2 of this thesis revealed a relatively small number of genes whose mRNA levels were affected by truncating the RNAPII-CTD to 11 repeats, which is 3 heptapeptide repeats above the minimum required for viability. Therefore, this strategy assayed the auxiliary roles of the RNAPII-CTD rather than its essential function. Nonetheless, progressive truncation of the RNAPII-CTD resulted in a greater number and severity of mRNA alterations, which was consistent with progressively impaired function. The gene expression-profiles also demonstrated that individual genes had specific minimal RNAPII-CTD length requirements for normal expression, suggesting gene-specific roles for the RNAPII-CTD. Most of the genes whose expression was dependent on RNAPII-CTD length had increased mRNA levels, although a subset of genes with decreased mRNA levels were also identified. This observation was in contrast to the general view of the CTD as an orchestrator of co-transcriptional processes and instead highlighted important roles in transcription repression. Overall, two sets of genes whose mRNA levels increased upon truncation of the RNAPII-CTD were identified: protein-coding genes regulated by the proteasome-specific transcription factor, Rpn4, and Ty1 retrotransposons.

Chapter 2 revealed that the increased mRNA levels of Rpn4-dependent genes in the *rpb1-CTD11* mutant were likely a result of increased Rpn4 protein levels. How this is mediated remains unclear, although it could be through an indirect effect on the enzymes that regulate Rpn4 protein levels, such as the ubiquitin ligase, Ubr2 (Wang et al. 2004). Nonetheless, despite the increased transcription of genes encoding proteasome subunits, *rbp1-CTD11* mutants accumulated ubiquitinated proteins and had decreased proteasome function, as determined by a reporter substrate (Maria Aristizabal, unpublished data). Therefore, it is likely that truncating the

RNAPII-CTD resulted in proteasome deficiencies, leading to up-regulation of proteasome encoding genes. Connections between RNAPII and the proteasome have been previously reported. The proteasome has been shown to occupy transcribed regions in a transcription-dependent manner and is thought to stimulate transcription elongation by a yet unidentified mechanism that is independent of its degradation activity (Geng and Tansey 2012). The proteasome also removes RNAPIIs that become stalled at sites of damaged DNA to allow for DNA repair (Inukai et al. 2004, Lee et al. 2002). While further work is needed to fully elucidate the connection between RNAPII-CTD length and the proteasome, one hypothesis is that RNAPIIs carrying shortened CTDs may result in a greater load on the proteasome, in part by being more prone to stalling. The genome-wide ChIP-on-chip profiles of Rpb3 under wild type and *rpb1-CTD11* conditions, described in chapter 2 of this dissertation, suggested that this might be the case. In particular, these revealed elevated Rpb3 levels towards the 3' end of genes when the RNAPII-CTD was truncated, an observation that was consistent with termination defects and stalling. However, we observed no changes in RNAPII ubiquitination levels in the *rpb1-CTD11* mutant compared to wild type suggesting that RNAPII itself is not affecting proteasome function or that it might do so though an ubiquitination independent mechanism (data not shown). Further supporting this model was the observation that *rpb1-CTD11* mutants were more dependent on *RPN4*, and likely proteasome function, for survival.

Chapter 3 of this thesis described the second group of *rpb1-CTD11* up-regulated genes, Ty1 retrotransposons. Overall, we found a role for the RNAPII-CTD in minimizing retrotransposon mobility by decreasing their mRNA levels. In particular, the data suggested that the CTD normally functioned to repress Ty1 transcription as truncation of the RNAPII-CTD resulted in

increased Ty1 mRNA levels which were mediated by changes in promoter activity, the transcription factor, Tec1, and the CTD kinase, Cdk8. This observation builds on other well-established roles for RNAPII in DNA repair, and on previously reported genome instability phenotypes for RNAPII-CTD truncation mutants, thus making RNAPII and its CTD prominent players in maintaining genome integrity (Gauthier et al. 2002, Wong and Ingles 2001, Nonet and Young 1989, West and Corden 1995, Wilson et al. 2013).

In *S. cerevisiae* there are 5 classes of retrotransposons. Truncating the RNAPII-CTD had varying effects on their gene expression. This was reminiscent of other reports, which revealed that individual retrotransposons vary in normal gene expression levels, transposition strategies, transposition target sites, and genetic requirements for individual stages of their life cycle (Morillon et al. 2002, Nyswaner et al. 2008). Although chapter 3 of this thesis primarily described a role for the CTD in the regulation of Ty1 elements, it is likely that its effect extends to other Ty elements, namely Ty2 and Ty3. In this regard, the conclusions were primarily limited by technical difficulties, which prevented us from differentiating individual elements. For the Ty2 elements, although Rpb3 occupancy significantly increased upon truncation of the RNAPII-CTD, it did not lead to significant changes in their mRNA levels. Therefore, we were unable to conclude that truncation of the RNAPII-CTD alone resulted in gene expression changes of Ty2 elements. However, deletion of *CDK8* did significantly alter Ty2 mRNA levels, which were normalized upon truncation of the RNAPII-CTD. Thus, there is evidence for the RNAPII-CTD in regulating the expression of Ty2 elements but additional work is required. While it is likely that sequencing techniques might help differentiate individual retrotransposon elements, previous work in *Schizocaccharomyces pombe* revealed that the high degree of sequence similarity could

not be overcome even by sequencing strategies (Mourier and Willerslev 2010). Reporter strategies seem like the most appealing method to differentiate between individual retrotransposon elements, which in future will help address if their differential behavior is rooted in the limited number of sequence differences.

The genes whose expression changed as a result of truncating the RNAPII-CTD had different minimal CTD-length requirements for normal expression, suggesting gene-specific roles for the RNAPII-CTD. Rpn4-regulated genes had a minimum of 12 heptapeptide repeats for normal mRNA levels, and truncation by one repeat resulted in significant increases to their mRNA levels. Similarly, the genes whose mRNA levels decreased upon truncation of the RNAPII-CTD had minimal CTD length requirements that ranged from 11-13 repeats. In this thesis, the detailed characterization of strains carrying progressively truncated RNAPII-CTDs provided an important starting point to understanding the mechanistic details of the observed altered expression. Future work should aim to identify factors whose occupancy profiles are altered in a particular RNAPII-CTD truncation mutant and not its longer counterparts. In this manner, the molecular underpinnings of their differential effect on RNAPII-CTD length-dependent genes can be begin to be determined. Finally, to understand the role of the RNAPII-CTD on retrotransposon gene expression, the minimal CTD length required for normal retrotransposon mRNA levels should be determined.

While the gene expression and RNAPII occupancy profiles identified all genes whose mRNA levels were dependent on CTD-length, they did not provide a complete picture. In particular, information regarding mRNA processing and stability, parameters that have important

consequences to mRNA function, were lacking. Understanding the full effect of reduced CTD length on mRNA processing is important, given that previous work has shown defects in these processes, but have only focused on a subset of representative genes (de la Mata and Kornblihtt 2006, McCracken et al. 1997a, McCracken et al. 1997b, Rosonina and Blencowe 2004, Ryan et al. 2002, Suh et al. 2010). Furthermore, given that strains carrying 11 CTD repeats had relatively few gene expression alterations, yet had strong growth defects, suggests possible defects beyond RNA synthesis. Chapter 2 of this thesis aimed to determine how truncating the RNAPII-CTD changed the genome-wide occupancy of factors involved in various aspects of transcription. Although a limited number of factors were profiled, these provided important insights into RNAPII-CTD biology. Overall, truncating the RNAPII-CTD had minimal effects on the occupancy of the general transcription factor, TFIIB, the Mediator subunit, Cdk8, and the elongation factor, Elf1, suggesting that the transcriptional changes occurred independently of these factors and likely following promoter initiation complex assembly. In contrast, truncating the RNAPII-CTD resulted in decreased histone H3 lysine 36 tri-methylation (H3K36me3) levels at short genes, and an almost complete loss of the capping enzyme, Cet1. The latter was consistent with *in vitro* studies demonstrating that a minimum of 14 CTD repeats are required for normal association of Cet1 with the RNAPII-CTD (Suh et al. 2010, West and Corden 1995). However, the Cet1 genome-wide occupancy profiles also showed that highly transcribed genes had normal Cet1 levels, indicating a gene-specific role for the RNAPII-CTD in Cet1 recruitment. Thus, it remains to be determined how Cet1 recruitment is achieved at these sites and whether capping is generally impaired in the *rpb1-CTD11* mutant. Overall, we conclude that genome-wide occupancy profiles of transcription associated factors using ChIP-on-chip are an important starting point in determining how altering RNAPII-CTD function affects the recruitment and

function of transcription associated factors. In the future, a complete picture of RNAPII-CTD function will require profiling additional factors, and determining how their occupancy is altered when the RNAPII-CTD is progressively truncated. Finally, the genome-wide occupancy profiles also provided insight into the regulatory landscape at retrotransposons, which differed from that of other RNAPII-CTD length-dependent genes. Given that transcription associated factors and chromatin regulators have been implicated in the regulation of retrotransposon gene expression, determining their distribution at these genomic loci is important for understanding the regulatory landscape at these sites and how it changes upon truncation of the RNAPII-CTD (Nyswaner et al. 2008).

Chapter 4 of this thesis aimed to better understand the role of the RNAPII-CTD phosphatase, Fcp1, in RNAPII-dependent transcription. High throughput genetic and gene expression characterization of strains carrying *FCP1* mutant alleles revealed a significant overlap with strains carrying RNAPII-CTD truncations reinforcing an intimate relationship between Fcp1 and the RNAPII-CTD. However, differences were also evident. Focusing on the role of *FCP1* in RNAPII transcription, the characterization of a series of Fcp1 C-terminal truncation mutants showed that the phenotypes of the *fcp1-594* and *fcp1-609* mutants were not fully overlapping. More specifically, strains carrying the *fcp1-609* allele had more gene expression and genetic interactions than strains carrying the shorter *fcp1-594* allele. These data suggested that said mutations resulted in specific disturbances to Fcp1 function and highlighted a nuanced role for the Fcp1 C-terminus. The observation that different alleles resulted in varying phenotypes presented an exciting opportunity to investigate how Fcp1 contributes to different pathways. Future goals are to characterize additional *FCP1* alleles in an effort to determine the scope of

Fcp1 function in the cell. Finally, the *FCP1* mutant gene expression profiles suggested potential substrates for Fcp1. Specifically, we hypothesize that Fcp1 might directly regulate the set of transcription factors involved in the regulation of Fcp1-dependent genes, for which there is evidence of being phosphorylated *in vivo*. In the future, the ability of Fcp1 to directly target these proteins should be investigated *in vitro* and *in vivo*.

Overall, the gene expression profiles of RNAPII-CTD and *FCP1* truncation mutants, described in chapter 2 and 4 of this thesis revealed limited effects on transcription. Although surprising given their central role in transcription, recently published gene expression profiles of mutants in genes with prominent roles in transcription also revealed generally low number of alterations, which are consistent with our observations (Kemmeren et al. 2014, Lenstra et al. 2011). Collectively, these findings suggest that RNAPII-dependent transcription regulation is robust and likely dependent on multiple redundant pathways.

Efforts to understand RNAPII-CTD function took advantage of suppressor screens, which laid the foundation for the discovery of the Mediator complex (Nonet and Young 1989). To complement this approach, chapter 2 described a comprehensive genetic network for the RNAPII-CTD as determined by the E-MAP platform. Using this method, we uncovered the genetic requirements of CTD truncation mutants for viability, as well as potential suppressors currently being validated. Focusing on the original suppressor screen by the Young Laboratory, we investigated the link between the RNAPII-CTD and Cdk8, primarily because of Cdk8's role in regulating RNAPII-dependent transcription and in phosphorylating the RNAPII-CTD (Galbraith et al. 2010, Nonet and Young 1989). Although a genetic connection between *CDK8*

and the RNAPII-CTD has long been established, the molecular underpinnings of this association have remained unexplored. Therefore, we aimed to gain further insight into the biology of the RNAPII-CTD and how it relates to *CDK8*. In this dissertation, in addition to expanding known RNAPII-CTD length-dependent phenotypes, we also showed that loss of *CDK8* could suppress most of these phenotypes. In detail, loss of *CDK8* could normalize the increased expression of Rpn4-dependent genes and Ty1 retrotransposons observed in RNAPII-CTD truncation mutants, and it could overcome the induction defects observed at the *INO1* locus. These results are consistent with the emerging duality of Cdk8 functioning as an activator or repressor of transcription. Furthermore, loss of *CDK8* alone had varying effects on the mRNA levels of these genes, suggesting that it suppressed RNAPII-CTD truncation phenotypes via different roles.

Loss of *CDK8* likely suppressed the increased expression of Rpn4-dependent genes in the *rpb1-CTD11* mutant by playing a role in regulating Rpn4 protein stability. This resulted in normalized expression of Rpn4-dependent genes, proteasome function, and bulk ubiquitinated protein levels (Maria Aristizabal unpublished results). Indicative of a direct role, Cdk8 was enriched at Rpn4-dependent gene promoters and work currently underway is focused on determining if this reflects a role for Cdk8 in phosphorylating Rpn4. Finding a direct role for Cdk8 in targeting Rpn4 would add to the already complex system of regulating Rpn4 protein levels. Briefly, Rpn4 is regulated at the level of transcription, translation and protein turnover by a number of different pathways (Dohmen et al. 2007). We briefly discuss the latter, as this is the most likely step affected by truncation of the RNAPII-CTD and loss of *CDK8*, given that no significant changes in Rpn4 mRNA levels were detected in the *rpb1-CTD11* mutant and that decreased Rpn4 turn over rates were observed in the *cdk8Δ* mutant. At the protein level, Rpn4 is regulated by a feedback loop,

wherein Rpn4 stimulates proteasome gene expression leading to increased levels of functional proteasome complexes in the cell, but itself is also degraded by the proteasome. Proteasome-dependent degradation of Rpn4 can be further achieved via alternative mechanisms including ubiquitin-independent and dependent pathways, the latter regulated by phosphorylation by Casein kinase 2 *in vitro* (Ju et al. 2007). Given that Rpn4 is a known phosphor-protein and that phosphorylation regulates its protein levels, finding a direct role for Cdk8 in its phosphorylation would not be surprising. However, we note that our genetic analysis suggested no relationship between the RNAPII-CTD or *CDK8* and two known Rpn4 phosphorylation sites involved in its ubiquitin-dependent degradation. Thus, if directly involved, Cdk8 likely plays a role by phosphorylating other Rpn4 residues.

Loss of *CDK8* in an RNAPII-CTD truncation mutant background also suppressed the increased Ty1 mRNA and RNAPII levels. Furthermore, chapter 3 identified *TEC1* and *STE12* as novel *rpb1-CTD11* suppressors. Specifically, loss of *TEC1*, and to a lesser extent *STE12*, suppressed the increased expression of Ty1 elements and a subset of RNAPII-CTD truncation mutant growth phenotypes. Using a genetic approach, we showed that *CDK8* and *TEC1* functioned in the same pathway to normalize RNAPII-CTD length-dependent phenotypes, although this was likely independent of *STE12,* as loss of *STE12* led to reduced suppression of RNAPII-CTD truncation phenotypes compared to loss of *TEC1*.

Loss of *CDK8* could also suppress phenotypes associated with mutating the RNAPII-CTD phosphatase, Fcp1, although with clearly different magnitudes. While Cdk8 phosphorylates the RNAPII-CTD and Fcp1 dephosphorylates it, no correlation between the suppression and

RNAPII-CTD phosphorylation levels was observed (Maria Aristizabal unpublished data). Furthermore, the suppression was not mediated by restoring the mRNA levels of *FCP1*-dependent genes. Thus, the molecular underpinning of the suppression of *FCP1* mutant phenotypes by loss of *CDK8* remains to be determined.

Chapter 5 of this thesis described the first ever *S. cerevisiae* genome-wide occupancy map of DNA:RNA hybrids. In particular, we aimed to identify regions most susceptible to DNA:RNA hybrid formation to address how DNA:RNA hybrids are formed, recognized and removed. Overall, the results suggested that not all hybrids are created equally and argued for a focus on genome-wide approaches to study DNA:RNA hybrid biology. This work also revealed that ChIP-on-chip was a suitable method for studying DNA:RNA hybrids because it identified most previously reported prone sites (Balk et al. 2013, El Hage et al. 2010, Luke et al. 2008, Pfeiffer et al. 2013). Furthermore, recently published *S. cerevisiae* DNA:RNA hybrid maps determined using chromatin immunoprecipitation followed by sequencing methods, showed similar results, thus validating the findings of chapter 5 (El Hage et al. 2014). However, these techniques have some limitations. Specifically, the ChIP-on-chip maps of DNA:RNA hybrid occupancy lacked information regarding strand specificity, preventing us from determining which strand was evicted and which strand was paired to the RNA. Modified bisulfite sequencing techniques have been successfully used for this purpose, suggesting that including this into a ChIP-seq platform would overcome the shortcomings of our method (Lin et al. 2010). Determining strand specificity is important because transcription itself is strand-specific and because the displaced DNA is more prone to damage (Aguilera and Garcia-Muse 2012) .

Mutating factors involved in RNA processing led to differential effects on DNA:RNA hybrid occupancy, presenting exciting opportunities to understand how these are created and regulated. Work currently underway includes profiling DNA:RNA hybrids on additional mutants, which paired with gene-expression analysis will be instrumental in determining how changes in the DNA:RNA hybrid landscape ultimately affect gene expression. This is particularly intriguing given evidence of a role for DNA:RNA hybrids in chromatin regulation, antisense transcription, and Ty1 retrotransposition (Castellano-Pozo et al. 2013, El Hage et al. 2014, Powell et al. 2013, Sun et al. 2013). Ultimately, the goal will be to understand what features distinguishes hybrids that contribute to biological functions and hybrids that lead to detrimental effects on genome stability.

Yeast has served as an excellent model organism for studying a diverse range of biological processes. Using this model, we aimed to better understand the role of the RNAPII-CTD, the CTD phosphatase Fcp1 and DNA:RNA hybrids. All of these factors are conserved from yeast to human, making contributions in yeast widely applicable to higher systems (Chan et al. 2014, Guo and Stiller 2004). Work on the CTD is particularly important because the increased complexity of the mammalian CTD makes studies performed in this system more difficult to interpret. Furthermore, the high degree of conservation between the yeast and mammalian CTD, underscored by the observation that the yeast CTD can be replaced by the mammalian version, suggests that phenomena discovered in yeast are likely applicable in higher systems (Allison et al. 1988). Thus, *S. cerevisiae* is a perfect system for discovering basic biological principles of RNAPII-CTD function and for determining which techniques are better suited for studies in higher organisms.

Discovering novel roles for the RNAPII-CTD in genome stability is significant because of its importance in human disease. A role for the CTD in retrotransposons gene expression is relevant as retrotransposons are homologous to mammalian retroviruses. Furthermore, DNA:RNA hybrids have been described in both yeast and mammalian systems, and in the latter they have been implicated in genome instability and cancer (Aguilera and Garcia-Muse 2012). As such, these discoveries could ultimately improve our understanding of endogenous or acquired retrovirus regulation, and of cancer development and progression in mammalian systems.

## Bibliography


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

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

Akoulitchev S, Makela TP, Weinberg RA, Reinberg D. 1995. Requirement for TFIIH kinase activity in transcription by RNA polymerase II. *Nature* 377: 557-560.

Allepuz-Fuster P, Martinez-Fernandez V, Garrido-Godino AI, Alonso-Aguado S, Hanes SD, Navarro F, Calvo O. 2014. Rpb4/7 facilitates RNA polymerase II CTD dephosphorylation. *Nucleic Acids Res*. Pii: gku1227 [Epub ahead of print].

Allison LA, Ingles CJ. 1989. Mutations in RNA polymerase II enhance or suppress mutations in GAL4. *Proc Natl Acad Sci U S A* 86: 2794-2798.

Allison LA, Moyle M, Shales M, Ingles CJ. 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* 42: 599-610.

Allison LA, Wong JK, Fitzpatrick VD, Moyle M, Ingles CJ. 1988. The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function. *Mol Cell Biol* 8: 321-329.

Alzu A, Bermejo R, Begnis M, Lucca C, Piccini D, Carotenuto W, Saponaro M, Brambati A, Cocito A, Foiani M et al. 2012. Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell* 151: 835-846.

Andrau JC, van de Pasch L, Lijnzaad P, Bijma T, Koerkamp MG, van de Peppel J, Werner M, Holstege FC. 2006. Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Mol Cell* 22: 179-192.

Arana ME, Kerns RT, Wharey L, Gerrish KE, Bushel PR, Kunkel TA. 2012. Transcriptional responses to loss of RNase H2 in Saccharomyces cerevisiae. *DNA Repair (Amst)* 11: 933-941.

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

Aristizabal MJ, Negri GL, Benschop JJ, Holstege FC, Krogan NJ, Kobor MS. 2013. High-throughput genetic and gene expression analysis of the RNAPII-CTD reveals unexpected connections to SRB10/CDK8. *PLoS Genet* 9: e1003758.

Armache KJ, Mitterweger S, Meinhart A, Cramer P. 2005. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* 280: 7131-7134.

Baillat D, Hakimi MA, Naar AM, Shilatifard A, Cooch N, Shiekhattar R. 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 123: 265-276.

Balk B, Maicher A, Dees M, Klermund J, Luke-Glaser S, Bender K, Luke B. 2013. Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat Struct Mol Biol* 20: 1199-1205.

Barnes DE, Lindahl T, Sedgwick B. 1993. DNA repair. *Curr Opin Cell Biol* 5: 424-433.

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

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

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

Belakavadi M, Fondell JD. 2010. Cyclin-dependent kinase 8 positively cooperates with Mediator to promote thyroid hormone receptor-dependent transcriptional activation. *Mol Cell Biol* 30: 2437-2448.

Berthelet S, Usher J, Shulist K, Hamza A, Maltez N, Johnston A, Fong Y, Harris LJ, Baetz K. 2010. Functional genomics analysis of the Saccharomyces cerevisiae iron responsive transcription factor Aft1 reveals iron-independent functions. *Genetics* 185: 1111-1128.

Bhoite LT, Yu Y, Stillman DJ. 2001. The Swi5 activator recruits the Mediator complex to the HO promoter without RNA polymerase II. *Genes Dev* 15: 2457-2469.

Bierhoff H, Dundr M, Michels AA, Grummt I. 2008. Phosphorylation by casein kinase 2 facilitates rRNA gene transcription by promoting dissociation of TIF-IA from elongating RNA polymerase I. *Mol Cell Biol* 28: 4988-4998.

Bierhoff H, Schmitz K, Maass F, Ye J, Grummt I. 2010. Noncoding transcripts in sense and antisense orientation regulate the epigenetic state of ribosomal RNA genes. *Cold Spring Harb Symp Quant Biol* 75: 357-364.

Bleykasten-Grosshans C, Friedrich A, Schacherer J. 2013. Genome-wide analysis of intraspecific transposon diversity in yeast. *BMC Genomics* 14: 399-2164-14-399.

Boeing S, Rigault C, Heidemann M, Eick D, 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.

Boeke JD, Garfinkel DJ, Styles CA, Fink GR. 1985. Ty elements transpose through an RNA intermediate. *Cell* 40: 491-500.

Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ. 1986. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods* 89: 123-130.

Bowman EA, Kelly WG. 2014. RNA Polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases. *Nucleus* 5: 224-236.

Bradshaw VA, McEntee K. 1989. DNA damage activates transcription and transposition of yeast Ty retrotransposons. *Mol Gen Genet* 218: 465-474.

Brickner D, Cajigas I, Fondufe-Mittendorf Y, Ahmed S, Lee P, Widom J, Brickner JH. 2007. H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol* 5: e81.

Bryk M, Banerjee M, Conte D,Jr, Curcio MJ. 2001. The Sgs1 helicase of Saccharomyces cerevisiae inhibits retrotransposition of Ty1 multimeric arrays. *Mol Cell Biol* 21: 5374-5388.

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

Buratowski S, Sharp PA. 1990. Transcription initiation complexes and upstream activation with RNA polymerase II lacking the C-terminal domain of the largest subunit. *Mol Cell Biol* 10: 5562-5564.

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

Camblong J, Beyrouthy N, Guffanti E, Schlaepfer G, Steinmetz LM, Stutz F. 2009. Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. *Genes Dev* 23: 1534-1545.

Carlson M, Falcon S, Pages H, Li N. org.Sc.sgd.db: Genome wide annotation for Yeast. R package version 2.7.1.

Castellano-Pozo M, Garcia-Muse T, Aguilera A. 2012. R-loops cause replication impairment and genome instability during meiosis. *EMBO Rep* 13: 923-929.

Castellano-Pozo M, Santos-Pereira JM, Rondon AG, Barroso S, Andujar E, Perez-Alegre M, Garcia-Muse T, Aguilera A. 2013. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol Cell* 52: 583-590.

Castelnuovo M, Rahman S, Guffanti E, Infantino V, Stutz F, Zenklusen D. 2013. Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat Struct Mol Biol* 20: 851-858.

Chambers RS, Wang BQ, Burton ZF, Dahmus ME. 1995. The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II and by the general transcription factors IIF and IIB. *J Biol Chem* 270: 14962-14969.

Chan YA, Hieter P, Stirling PC. 2014. Mechanisms of genome instability induced by RNA-processing defects. *Trends Genet*. 30: 245-253.

Chang DT, Huang CY, Wu CY, Wu WS. 2011. YPA: an integrated repository of promoter features in Saccharomyces cerevisiae. *Nucleic Acids Res* 39: D647-52.

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

Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E, Eick D. 2007. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* 318: 1780-2.

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

Chargaff E. 1976. Initiation of enzymic synthesis of deoxyribonucleic acid by ribonucleic acid primers. *Prog Nucleic Acid Res Mol Biol* 16: 1-24.

Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 422: 726-730.

Chavez S, Garcia-Rubio M, Prado F, Aguilera A. 2001. Hpr1 is preferentially required for transcription of either long or G+C-rich DNA sequences in Saccharomyces cerevisiae. *Mol Cell Biol* 21: 7054-7064.

Chen C, Huang B, Eliasson M, Ryden P, Bystrom AS. 2011. Elongator complex influences telomeric gene silencing and DNA damage response by its role in wobble uridine tRNA modification. *PLoS Genet* 7: e1002258.

Chen J, Wagner EJ. 2010. snRNA 3' end formation: the dawn of the Integrator complex. *Biochem Soc Trans* 38: 1082-1087.

Chernikova SB, Razorenova OV, Higgins JP, Sishc BJ, Nicolau M, Dorth JA, Chernikova DA, Kwok S, Brooks JD, Bailey SM et al. 2012. Deficiency in mammalian histone H2B ubiquitin ligase Bre1 (Rnf20/Rnf40) leads to replication stress and chromosomal instability. *Cancer Res* 72: 2111-2119.

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

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

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

Clark DJ, Bilanchone VW, Haywood LJ, Dildine SL, Sandmeyer SB. 1988. A yeast sigma composite element, TY3, has properties of a retrotransposon. *J Biol Chem* 263: 1413-1423.

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

Collins SR, Roguev A, Krogan NJ. 2010. Quantitative genetic interaction mapping using the E-MAP approach. *Methods Enzymol* 470: 205-231.

Conte D,Jr, Curcio MJ. 2000. Fus3 controls Ty1 transpositional dormancy through the invasive growth MAPK pathway. *Mol Microbiol* 35: 415-427.

Corden JL. 1990. Tails of RNA polymerase II. *Trends Biochem Sci* 15: 383-387.

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

Coudreuse D, van Bakel H, Dewez M, Soutourina J, Parnell T, Vandenhaute J, Cairns B, Werner M, Hermand D. 2010. A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe. *Curr Biol* 20: 1053-1064.

Cramer P. 2002. Multisubunit RNA polymerases. *Curr Opin Struct Biol* 12: 89-97.

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR, Kornberg RD. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288: 640-649.

Crow YJ, Leitch A, Hayward BE, Garner A, Parmar R, Griffith E, Ali M, Semple C, Aicardi J, Babul-Hirji R et al. 2006. Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. *Nat Genet* 38: 910-916.

Czudnochowski N, Bosken CA, Geyer M. 2012. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat Commun* 3: 842.

Dahmus ME. 1994. The role of multisite phosphorylation in the regulation of RNA polymerase II activity. *Prog Nucleic Acid Res Mol Biol* 48: 143-179.

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

Dohmen RJ, Willers I, Marques AJ. 2007. Biting the hand that feeds: Rpn4-dependent feedback regulation of proteasome function. *Biochim Biophys Acta* 1773: 1599-1604.

Dominguez-Sanchez MS, Barroso S, Gomez-Gonzalez B, Luna R, Aguilera A. 2011. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* 7: e1002386.

Drogat J, Hermand D. 2012. Gene-specific requirement of RNA polymerase II CTD phosphorylation. *Mol Microbiol* 84: 995-1004.

Droit A, Cheung C, Gottardo R. 2010. rMAT--an R/Bioconductor package for analyzing ChIP-chip experiments. *Bioinformatics* 26: 678-679.

Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6: 197-208.

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

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

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

El Hage A, French SL, Beyer AL, Tollervey D. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev* 24: 1546-1558.

El Hage A, Webb S, Kerr A, Tollervey D. 2014. Genome-Wide Distribution of RNA-DNA Hybrids Identifies RNase H Targets in tRNA Genes, Retrotransposons and Mitochondria. *PLoS Genet* 10: e1004716.

Elder RT, St John TP, Stinchcomb DT, Davis RW, Scherer S, Davis RW. 1981. Studies on the transposable element Ty1 of yeast. I. RNA homologous to Ty1. II. Recombination and expression of Ty1 and adjacent sequences. *Cold Spring Harb.Symp.Quant.Biol.* 45: 581.

Errede B. 1993. MCM1 binds to a transcriptional control element in Ty1. *Mol Cell Biol* 13: 57-62.

Errede B, Cardillo TS, Sherman F, Dubois E, Deschamps J, Wiame JM. 1980. Mating signals control expression of mutations resulting from insertion of a transposable repetitive element adjacent to diverse yeast genes. *Cell* 22: 427-436.

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

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

Faghihi MA, Wahlestedt C. 2009. Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol* 10: 637-643.

Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* 23: 257-258.

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

Fath S, Kobor MS, Philippi A, Greenblatt J, Tschochner H. 2004. Dephosphorylation of RNA polymerase I by Fcp1p is required for efficient rRNA synthesis. *J Biol Chem* 279: 25251-25259.

Fuchs SM, Kizer KO, Braberg H, Krogan NJ, Strahl BD. 2011. RNA polymerase ii CTD phosphorylation regulates stability of the SET2 methyltransferase and histone H3 DI- and trimethylation at lysine 36. *J Biol Chem*. 287: 3249-3256.

Galbraith MD, Donner AJ, Espinosa JM. 2010. CDK8: a positive regulator of transcription. *Transcription* 1: 4-12.

Gallardo M, Luna R, Erdjument-Bromage H, Tempst P, Aguilera A. 2003. Nab2p and the Thp1p-Sac3p complex functionally interact at the interface between transcription and mRNA metabolism. *J Biol Chem* 278: 24225-24232.

Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X. 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev* 25: 2041-2056.

Garraway LA, Lander ES. 2013. Lessons from the cancer genome. *Cell* 153: 17-37.

Garrido-Lecca A, Blumenthal T. 2010. RNA polymerase II C-terminal domain phosphorylation patterns in Caenorhabditis elegans operons, polycistronic gene clusters with only one promoter. *Mol Cell Biol* 30: 3887-3893.

Gauthier L, Dziak R, Kramer DJ, Leishman D, Song X, Ho J, Radovic M, Bentley D, Yankulov K. 2002. The role of the carboxyterminal domain of RNA polymerase II in regulating origins of DNA replication in Saccharomyces cerevisiae. *Genetics* 162: 1117-1129.

Gavalda S, Gallardo M, Luna R, Aguilera A. 2013. R-loop mediated transcription-associated recombination in trf4Delta mutants reveals new links between RNA surveillance and genome integrity. *PLoS One* 8: e65541.

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

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

Ginno PA, Lim YW, Lott PL, Korf I, Chedin F. 2013. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res* 23: 1590-1600.

Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* 45: 814-825.

Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, Bentley DL. 2009. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol Cell Biol* 29: 5455-5464.

Gold MO, Rice AP. 1998. Targeting of CDK8 to a promoter-proximal RNA element demonstrates catalysis-dependent activation of gene expression. *Nucleic Acids Res* 26: 3784-3788.

Gomez-Gonzalez B, Garcia-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marin A, Foiani M, Aguilera A. 2011. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J* 30: 3106-3119.

Gonzalez-Aguilera C, Tous C, Gomez-Gonzalez B, Huertas P, Luna R, Aguilera A. 2008. The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol Biol Cell* 19: 4310-4318.

Gray WM, Fassler JS. 1996. Isolation and analysis of the yeast TEA1 gene, which encodes a zinc cluster Ty enhancer-binding protein. *Mol Cell Biol* 16: 347-358.

Gray WM, Fassler JS. 1993. Role of Saccharomyces cerevisiae Rap1 protein in Ty1 and Ty1-mediated transcription. *Gene Expr* 3: 237-251.

Grunberg S, Hahn S. 2013. Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci* 38: 603-611.

Grunberg S, Warfield L, Hahn S. 2012. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol* 19: 788-796.

Guarente L. 1983. Yeast promoters and lacZ fusions designed to study expression of cloned genes in yeast. *Methods Enzymol* 101: 181-191.

Guarente L, Ptashne M. 1981. Fusion of Escherichia coli lacZ to the cytochrome c gene of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 78: 2199-2203.

Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, Bourbon HM, Holstege FC, Werner M. 2004. A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res* 32: 5379-5391.

Guo Z, Stiller JW. 2005. Comparative genomics and evolution of proteins associated with RNA polymerase II C-terminal domain. *Mol Biol Evol* 22: 2166-2178.

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

Hahn JS, Neef DW, Thiele DJ. 2006. A stress regulatory network for co-ordinated activation of proteasome expression mediated by yeast heat shock transcription factor. *Mol Microbiol* 60: 240-251.

Hall BM, Ma CX, Liang P, Singh KK. 2009. Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* 25: 1564-1565.

Han SJ, Lee YC, Gim BS, Ryu GH, Park SJ, Lane WS, Kim YJ. 1999. Activator-specific requirement of yeast mediator proteins for RNA polymerase II transcriptional activation. *Mol Cell Biol* 19: 979-988.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* 144: 646-674.

Hausmann S, Shuman S. 2002. Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. *J Biol Chem* 277: 21213-21220.
Havecker ER, Gao X, Voytas DF. 2004. The diversity of LTR retrotransposons. *Genome Biol* 5: 225.

Hegarat N, Vesely C, Vinod PK, Ocasio C, Peter N, Gannon J, Oliver AW, Novak B, Hochegger H. 2014. PP2A/B55 and Fcp1 regulate Greatwall and Ensa dephosphorylation during mitotic exit. *PLoS Genet* 10: e1004004.

Heidemann M, Hintermair C, Voss K, Eick D. 2012. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta*. 1829: 55-62.

Helmrich A, Ballarino M, Tora L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell* 44: 966-977.

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

Hengartner CJ, Thompson CM, Zhang J, Chao DM, Liao SM, Koleske AJ, Okamura S, Young RA. 1995. Association of an activator with an RNA polymerase II holoenzyme. *Genes Dev* 9: 897-910.

Hentrich T, Schulze JM, Emberly E, Kobor MS. 2012. CHROMATRA: a Galaxy tool for visualizing genome-wide chromatin signatures. *Bioinformatics* 28: 717-718.

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.

Hirst M, Kobor MS, Kuriakose N, Greenblatt J, Sadowski I. 1999. GAL4 is regulated by the RNA polymerase II holoenzyme-associated cyclin-dependent protein kinase SRB10/CDK8. *Mol Cell* 3: 673-8.

Hobson DJ, Wei W, Steinmetz LM, Svejstrup JQ. 2012. RNA polymerase II collision interrupts convergent transcription. *Mol Cell* 48: 365-374.

Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95: 717-728.

Hsin JP, Li W, Hoque M, Tian B, Manley JL. 2014. RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *Elife*: e02112.

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

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

Hsin JP, Xiang K, Manley JL. 2014. Function and control of RNA polymerase II CTD phosphorylation in vertebrate transcription and RNA processing. *Mol Cell Biol*. 34: 2488-2498.

Hsu PL, Yang F, Smith-Kinnaman W, Yang W, Song JE, Mosley AL, Varani G. 2014. Rtr1 Is a Dual Specificity Phosphatase That Dephosphorylates Tyr1 and Ser5 on the RNA Polymerase II CTD. *J Mol Biol*. 426: 2970-2981.

Hu Z, Zhang A, Storz G, Gottesman S, Leppla SH. 2006. An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res* 34: e52.

Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* 12: 711-721.

Hug AM, Feldmann H. 1996. Yeast retrotransposon Ty4: the majority of the rare transcripts lack a U3-R sequence. *Nucleic Acids Res* 24: 2338-2346.

Inukai N, Yamaguchi Y, Kuraoka I, Yamada T, Kamijo S, Kato J, Tanaka K, Handa H. 2004. A novel hydrogen peroxide-induced phosphorylation and ubiquitination pathway leading to RNA polymerase II proteolysis. *J Biol Chem* 279: 8190-8195.

Jeronimo C, Robert F. 2014. Kin28 regulates the transient association of Mediator with core promoters. *Nat Struct Mol Biol* 21: 449-455.

Jimeno S, Rondon AG, Luna R, Aguilera A. 2002. The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *EMBO J* 21: 3526-3535.

Ju D, Xu H, Wang X, Xie Y. 2007. Ubiquitin-mediated degradation of Rpn4 is controlled by a phosphorylation-dependent ubiquitylation signal. *Biochim Biophys Acta* 1773: 1672-1680.

Kanhere A, Viiri K, Araujo CC, Rasaiyaah J, Bouwman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW et al. 2010. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol Cell* 38: 675-688.

Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J et al. 2012. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149: 753-767.

Kawauchi J, Mischo H, Braglia P, Rondon A, Proudfoot NJ. 2008. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev* 22: 1082-1092.

Ke N, Irwin PA, Voytas DF. 1997. The pheromone response pathway activates transcription of Ty5 retrotransposons located within silent chromatin of Saccharomyces cerevisiae. *EMBO J* 16: 6272-6280.

Kemmeren P, Sameith K, van de Pasch LA, Benschop JJ, Lenstra TL, Margaritis T, O'Duibhir E, Apweiler E, van Wageningen S, Ko CW et al. 2014. Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* 157: 740-752.

Keogh MC, Buratowski S. 2004. Using chromatin immunoprecipitation to map cotranscriptional mRNA processing in Saccharomyces cerevisiae. *Methods Mol Biol* 257: 1-16.

Kettenberger H, Armache KJ, Cramer P. 2004. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* 16: 955-965.

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

Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF. 1998. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. *Genome Res* 8: 464-478.

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

Kim WY, Dahmus ME. 1989. The major late promoter of adenovirus-2 is accurately transcribed by RNA polymerases IIO, IIA, and IIB. *J Biol Chem* 264: 3169-3176.

Kimura M, Suzuki H, Ishihama A. 2002. Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in Schizosaccharomyces pombe involves direct interaction between Fcp1 and the Rpb4 subunit of pol II. *Mol Cell Biol* 22: 1577-1588.

Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A et al. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* 2011: bar030.

Klein F, Laroche T, Cardenas ME, Hofmann JF, Schweizer D, Gasser SM. 1992. Localization of RAP1 and topoisomerase II in nuclei and meiotic chromosomes of yeast. *J Cell Biol* 117: 935-948.

Kobor MS and Greenblatt J. 2002. Regulation of transcription elongation by phosphorylation. - *Biochim Biophys Acta.* 1577(2):261-275.

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

Kobor MS, Simon LD, Omichinski J, Zhong G, Archambault J, Greenblatt J. 2000. A motif shared by TFIIF and TFIIB mediates their interaction with the RNA polymerase II carboxy-terminal domain phosphatase Fcp1p in Saccharomyces cerevisiae. *Mol Cell Biol* 20: 7438-7449.

Koleske AJ, Buratowski S, Nonet M, Young RA. 1992. A novel transcription factor reveals a functional link between the RNA polymerase II CTD and TFIID. *Cell* 69: 883-894.

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

Kong SE, Kobor MS, Krogan NJ, Somesh BP, Sogaard TM, Greenblatt JF, Svejstrup JQ. 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.

Kops O, Zhou XZ, Lu KP. 2002. Pin1 modulates the dephosphorylation of the RNA polymerase II C-terminal domain by yeast Fcp1. *FEBS Lett* 513: 305-311.

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

Kruegel U, Robison B, Dange T, Kahlert G, Delaney JR, Kotireddy S, Tsuchiya M, Tsuchiyama S, Murakami CJ, Schleit J et al. 2011. Elevated proteasome capacity extends replicative lifespan in Saccharomyces cerevisiae. *PLoS Genet* 7: e1002253.

Kruger W, Peterson CL, Sil A, Coburn C, Arents G, Moudrianakis EN, Herskowitz I. 1995. Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. *Genes Dev* 9: 2770-2779.

Kulaeva OI, Hsieh FK, Chang HW, Luse DS, Studitsky VM. 2013. Mechanism of transcription through a nucleosome by RNA polymerase II. *Biochim Biophys Acta* 1829: 76-83.

Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, Han T, Xie S, Corden JL, McKnight SL. 2013. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* 155: 1049-1060.

Laloux I, Dubois E, Dewerchin M, Jacobs E. 1990. TEC1, a gene involved in the activation of Ty1 and Ty1-mediated gene expression in Saccharomyces cerevisiae: cloning and molecular analysis. *Mol Cell Biol* 10: 3541-3550.

Lee KB, Wang D, Lippard SJ, Sharp PA. 2002. Transcription-coupled and DNA damage-dependent ubiquitination of RNA polymerase II in vitro. *Proc Natl Acad Sci U S A* 99: 4239-4244.

Leela JK, Syeda AH, Anupama K, Gowrishankar J. 2013. Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in Escherichia coli. *Proc Natl Acad Sci U S A* 110: 258-263.

Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NA, Margaritis T, van de Pasch LA, van Heesch SA, Brok MO, Groot Koerkamp MJ et al. 2011. The specificity and topology of chromatin interaction pathways in yeast. *Mol Cell* 42: 536-549.

Lesage P, Todeschini AL. 2005. Happy together: the life and times of Ty retrotransposons and their hosts. *Cytogenet Genome Res* 110: 70-90.

Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell* 128: 707-719.

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

Liao SM, Taylor IC, Kingston RE, Young RA. 1991. RNA polymerase II carboxy-terminal domain contributes to the response to multiple acidic activators in vitro. *Genes Dev* 5: 2431-2440.

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

Lin Y, Dent SY, Wilson JH, Wells RD, Napierala M. 2010. R loops stimulate genetic instability of CTG.CAG repeats. *Proc Natl Acad Sci U S A* 107: 692-697.

Lipford JR, Smith GT, Chi Y, Deshaies RJ. 2005. A putative stimulatory role for activator turnover in gene expression. *Nature* 438: 113-116.

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

Liu Y, Kung C, Fishburn J, Ansari AZ, Shokat KM, Hahn S. 2004. Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Mol Cell Biol* 24: 1721-35.

Longtine MS, McKenzie A,3rd, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* 14: 953-961.

Lu H, Flores O, Weinmann R, Reinberg D. 1991. The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc Natl Acad Sci U S A* 88: 10004-10008.

Lu PY, Kobor MS. 2014. Maintenance of Heterochromatin Boundary and Nucleosome Composition at Promoters by the Asf1 Histone Chaperone and SWR1-C Chromatin Remodeler in Saccharomyces cerevisiae. *Genetics*. 197: 133-145.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389: 251-260.

Luke B, Panza A, Redon S, Iglesias N, Li Z, Lingner J. 2008. The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae. *Mol Cell* 32: 465-477.

Luna R, Jimeno S, Marin M, Huertas P, Garcia-Rubio M, Aguilera A. 2005. Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. *Mol Cell* 18: 711-722.

Ma Z, Atencio D, Barnes C, Defiglio H, Hanes SD. 2012. Multiple Roles for the Ess1 Prolyl Isomerase in the RNA Polymerase II Transcription Cycle. *Mol Cell Biol*. 32: 3594-3607.

Madison JM, Dudley AM, Winston F. 1998. Identification and analysis of Mot3, a zinc finger protein that binds to the retrotransposon Ty long terminal repeat (delta) in Saccharomyces cerevisiae. *Mol Cell Biol* 18: 1879-1890.

Maillet I, Buhler JM, Sentenac A, Labarre J. 1999. Rpb4p is necessary for RNA polymerase II activity at high temperature. *J Biol Chem* 274: 22586-22590.

Margaritis T, Oreal V, Brabers N, Maestroni L, Vitaliano-Prunier A, Benschop JJ, van Hooff S, van Leenen D, Dargemont C, Geli V et al. 2012. Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription. *PLoS Genet* 8: e1002952.

Marinello J, Chillemi G, Bueno S, Manzo SG, Capranico G. 2013. Antisense transcripts enhanced by camptothecin at divergent CpG-island promoters associated with bursts of topoisomerase I-DNA cleavage complex and R-loop formation. *Nucleic Acids Res*. 41: 10110-10123.

Martinez-Rucobo FW, Cramer P. 2013. Structural basis of transcription elongation. *Biochim Biophys Acta* 1829: 9-19.

Matsuda E, Garfinkel DJ. 2009. Posttranslational interference of Ty1 retrotransposition by antisense RNAs. *Proc Natl Acad Sci U S A* 106: 15657-15662.

Max T, Sogaard M, Svejstrup JQ. 2007. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J Biol Chem* 282: 14113-14120.

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

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

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

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

Michaelis C, Ciosk R, Nasmyth K. 1997. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* 91: 35-45.

Millan-Zambrano G, Rodriguez-Gil A, Penate X, de Miguel-Jimenez L, Morillo-Huesca M, Krogan N, Chavez S. 2013. The prefoldin complex regulates chromatin dynamics during transcription elongation. *PLoS Genet* 9: e1003776.

Mischo HE, Gomez-Gonzalez B, Grzechnik P, Rondon AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* 41: 21-32.

Morillon A, Benard L, Springer M, Lesage P. 2002. Differential effects of chromatin and Gcn4 on the 50-fold range of expression among individual yeast Ty1 retrotransposons. *Mol Cell Biol* 22: 2078-2088.

Morillon A, Springer M, Lesage P. 2000. Activation of the Kss1 invasive-filamentous growth pathway induces Ty1 transcription and retrotransposition in Saccharomyces cerevisiae. *Mol Cell Biol* 20: 5766-5776.

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

Mourier T, Willerslev E. 2010. Large-scale transcriptome data reveals transcriptional activity of fission yeast LTR retrotransposons. *BMC Genomics* 11: 167-2164-11-167.

Nakama M, Kawakami K, Kajitani T, Urano T, Murakami Y. 2012. DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes Cells* 17: 218-233.

Nelson C, Goto S, Lund K, Hung W, Sadowski I. 2003. Srb10/Cdk8 regulates yeast filamentous growth by phosphorylating the transcription factor Ste12. *Nature* 421: 187-190.

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

Nonet ML, Young RA. 1989. Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of Saccharomyces cerevisiae RNA polymerase II. *Genetics* 123: 715-24.

Nyswaner KM, Checkley MA, Yi M, Stephens RM, Garfinkel DJ. 2008. Chromatin-associated genes protect the yeast genome from Ty1 insertional mutagenesis. *Genetics* 178: 197-214.

Odawara J, Harada A, Yoshimi T, Maehara K, Tachibana T, Okada S, Akashi K, Ohkawa Y. 2011. The classification of mRNA expression levels by the phosphorylation state of RNAPII CTD based on a combined genome-wide approach. *BMC Genomics* 12: 516.

Owsianik G, Balzi lL, Ghislain M. 2002. Control of 26S proteasome expression by transcription factors regulating multidrug resistance in Saccharomyces cerevisiae. *Mol Microbiol* 43: 1295-1308.

Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD. 2006. A DNA integrity network in the yeast Saccharomyces cerevisiae. *Cell* 124: 1069-1081.

Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D, Pellagatti A, Wainscoat JS, Hellstrom-Lindberg E, Gambacorti-Passerini C et al. 2011. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med* 365: 1384-1395.

Peterson CL, Kruger W, Herskowitz I. 1991. A functional interaction between the C-terminal domain of RNA polymerase II and the negative regulator SIN1. *Cell* 64: 1135-1143.

Petesch SJ, Lis JT. 2012. Overcoming the nucleosome barrier during transcript elongation. *Trends Genet* 28: 285-294.

Pfeiffer V, Crittin J, Grolimund L, Lingner J. 2013. The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *EMBO J*. 32: 2861-2871.

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122: 517-527.

Poss ZC, Ebmeier CC, Taatjes DJ. 2013. The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* 48: 575-608.

Powell WT, Coulson RL, Gonzales ML, Crary FK, Wong SS, Adams S, Ach RA, Tsang P, Yamada NA, Yasui DH et al. 2013. R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proc Natl Acad Sci U S A* 110: 13938-13943.

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

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

Raithatha S, Su TC, Lourenco P, Goto S, Sadowski I. 2011. Cdk8 regulates stability of the transcription factor Phd1 to control pseudohyphal differentiation of Saccharomyces cerevisiae. *Mol Cell Biol*. 32: 664-674.

Rando OJ, Winston F. 2012. Chromatin and transcription in yeast. *Genetics* 190: 351-387.

Robinson PJ, Bushnell DA, Trnka MJ, Burlingame AL, Kornberg RD. 2012. Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc Natl Acad Sci U S A*. 109: 17931-17935.

Roeder GS, Fink GR. 1980. DNA rearrangements associated with a transposable element in yeast. *Cell* 21: 239-249.

Roeder RG, Rutter WJ. 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224: 234-237.

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

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

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

Rosonina E, Yurko N, Li W, Hoque M, Tian B, Manley JL. 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.

Ryan K, Murthy KG, Kaneko S, Manley JL. 2002. Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3' cleavage. *Mol Cell Biol* 22: 1684-1692.

Sacerdot C, Mercier G, Todeschini AL, Dutreix M, Springer M, Lesage P. 2005. Impact of ionizing radiation on the life cycle of Saccharomyces cerevisiae Ty1 retrotransposon. *Yeast* 22: 441-455.

Sakurai H, Ishihama A. 2002. Level of the RNA polymerase II in the fission yeast stays constant but phosphorylation of its carboxyl terminal domain varies depending on the phase and rate of cell growth. *Genes Cells* 7: 273-284.

Santos-Pereira JM, Herrero AB, Garcia-Rubio ML, Marin A, Moreno S, Aguilera A. 2013. The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes Dev* 27: 2445-2458.

Scafe C, Chao D, Lopes J, Hirsch JP, Henry S, Young RA. 1990. RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. *Nature* 347: 491-4.

Schneider S, Pei Y, Shuman S, Schwer B. 2010. Separable functions of the fission yeast Spt5 carboxyl-terminal domain (CTD) in capping enzyme binding and transcription elongation overlap with those of the RNA polymerase II CTD. *Mol Cell Biol* 30: 2353-2364.

Scholes DT, Banerjee M, Bowen B, Curcio MJ. 2001. Multiple regulators of Ty1 transposition in Saccharomyces cerevisiae have conserved roles in genome maintenance. *Genetics* 159: 1449-1465.

Schreieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P, Passmore LA. 2014. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat Struct Mol Biol* 21: 175-179.

Schulze JM, Hentrich T, Nakanishi S, Gupta A, Emberly E, Shilatifard A, Kobor MS. 2011. Splitting the task: Ubp8 and Ubp10 deubiquitinate different cellular pools of H2BK123. *Genes Dev* 25: 2242-2247.

Schulze JM, Jackson J, Nakanishi S, Gardner JM, Hentrich T, Haug J, Johnston M, Jaspersen SL, Kobor MS, Shilatifard A. 2009. Linking Cell Cycle to Histone Modifications: SBF and H2B Monoubiquitination Machinery and Cell-Cycle Regulation of H3K79 Dimethylation. *Mol Cell* 35: 626-641.

Schwer B, Sanchez AM, Shuman S. 2012. Punctuation and syntax of the RNA polymerase II CTD code in fission yeast. *Proc Natl Acad Sci U S A* 109: 18024-18029.

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

Servant G, Pennetier C, Lesage P. 2008. Remodeling yeast gene transcription by activating the Ty1 long terminal repeat retrotransposon under severe adenine deficiency. *Mol Cell Biol* 28: 5543-5554.

Servant G, Pinson B, Tchalikian-Cosson A, Coulpier F, Lemoine S, Pennetier C, Bridier-Nahmias A, Todeschini AL, Fayol H, Daignan-Fornier B et al. 2012. Tye7 regulates yeast Ty1 retrotransposon sense and antisense transcription in response to adenylic nucleotides stress. *Nucleic Acids Res* 40: 5271-5282.

Shandilya J, Roberts SG. 2012. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta* 1819: 391-400.

Sharifpoor S, Nguyen Ba AN, Youn JY, van Dyk D, Friesen H, Douglas AC, Kurat CF, Chong YT, Founk K, Moses AM et al. 2011. A quantitative literature-curated gold standard for kinase-substrate pairs. *Genome Biol* 12: R39-2011-12-4-r39. Epub 2011 Apr 14.

Sheffer A, Varon M, Choder M. 1999. Rpb7 can interact with RNA polymerase II and support transcription during some stresses independently of Rpb4. *Mol Cell Biol* 19: 2672-2680.

Shilatifard A. 2012. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem* 81: 65-95.

Sikdar N, Banerjee S, Zhang H, Smith S, Myung K. 2008. Spt2p defines a new transcription-dependent gross chromosomal rearrangement pathway. *PLoS Genet* 4: e1000290.

Sikorsky T, Hobor F, Krizanova E, Pasulka J, Kubicek K, Stefl R. 2012. Recognition of asymmetrically dimethylated arginine by TDRD3. *Nucleic Acids Res* 40: 11748-11755.

Simonovic M, Steitz TA. 2009. A structural view on the mechanism of the ribosome-catalyzed peptide bond formation. *Biochim Biophys Acta* 1789: 612-623.

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

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

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

Smith-Kinnaman WR, Berna MJ, Hunter GO, True JD, Hsu P, Cabello GI, Fox MJ, Varani G, Mosley AL. 2014. The interactome of the atypical phosphatase Rtr1 in Saccharomyces cerevisiae. *Mol Biosyst*. 10: 1730-1741.

Smolle M, Workman JL. 2013. Transcription-associated histone modifications and cryptic transcription. *Biochim Biophys Acta* 1829: 84-97.

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

Son S, Osmani SA. 2009. Analysis of all protein phosphatase genes in Aspergillus nidulans identifies a new mitotic regulator, fcp1. *Eukaryot Cell* 8: 573-585.

St Amour CV, Sanso M, Bosken CA, Lee KM, Larochelle S, Zhang C, Shokat KM, Geyer M, Fisher RP. 2012. Separate Domains of Fission Yeast Cdk9 (P-TEFb) Are Required for Capping Enzyme Recruitment and Primed (Ser7-Phosphorylated) Rpb1 Carboxyl-Terminal Domain Substrate Recognition. *Mol Cell Biol* 32: 2372-2383.

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

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

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

Stirling PC, Bloom MS, Solanki-Patil T, Smith S, Sipahimalani P, Li Z, Kofoed M, Ben-Aroya S, Myung K, Hieter P. 2011. The complete spectrum of yeast chromosome instability genes identifies candidate CIN cancer genes and functional roles for ASTRA complex components. *PLoS Genet* 7: e1002057.

Stirling PC, Chan YA, Minaker SW, Aristizabal MJ, Barrett I, Sipahimalani P, Kobor MS, Hieter P. 2012. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev* 26: 163-175.

Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Aguilera A, Struhl K, Reed R et al. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* 417: 304-308.

Suh H, Hazelbaker DZ, Soares LM, Buratowski S. 2013. The C-terminal domain of Rpb1 functions on other RNA polymerase II subunits. *Mol Cell* 51: 850-858.

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

Suh MH, Ye P, Zhang M, Hausmann S, Shuman S, Gnatt AL, Fu J. 2005. Fcp1 directly recognizes the C-terminal domain (CTD) and interacts with a site on RNA polymerase II distinct from the CTD. *Proc Natl Acad Sci U S A* 102: 17314-17319.

Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. 2013. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science* 340: 619-621.

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

Svejstrup JQ. 2003. Rescue of arrested RNA polymerase II complexes. *J Cell Sci* 116: 447-451.

Szilard RK, Jacques PE, Laramee L, Cheng B, Galicia S, Bataille AR, Yeung M, Mendez M, Bergeron M, Robert F et al. 2010. Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX. *Nat Struct Mol Biol* 17: 299-305.

Thebault P, Boutin G, Bhat W, Rufiange A, Martens J, Nourani A. 2011. Transcription regulation by the noncoding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. *Mol Cell Biol* 31: 1288-1300.

Thomas M, White RL, Davis RW. 1976. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci U S A* 73: 2294-2298.

Thomas MC, Chiang CM. 2006. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* 41: 105-178.

Thompson CM, Koleske AJ, Chao DM, Young RA. 1993. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* 73: 1361-1375.

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

Todeschini AL, Morillon A, Springer M, Lesage P. 2005. Severe adenine starvation activates Ty1 transcription and retrotransposition in Saccharomyces cerevisiae. *Mol Cell Biol* 25: 7459-7472.

Tombacz I, Schauer T, Juhasz I, Komonyi O, Boros I. 2009. The RNA Pol II CTD phosphatase Fcp1 is essential for normal development in Drosophila melanogaster. *Gene* 446: 58-67.

Treutlein B, Muschielok A, Andrecka J, Jawhari A, Buchen C, Kostrewa D, Hog F, Cramer P, Michaelis J. 2012. Dynamic architecture of a minimal RNA polymerase II open promoter complex. *Mol Cell* 46: 136-146.

Tsai KL, Sato S, Tomomori-Sato C, Conaway RC, Conaway JW, Asturias FJ. 2013. A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. *Nat Struct Mol Biol* 20: 611-619.

Ursic D, Himmel KL, Gurley KA, Webb F, Culbertson MR. 1997. The yeast SEN1 gene is required for the processing of diverse RNA classes. *Nucleic Acids Res* 25: 4778-4785.

van Bakel H, van Werven FJ, Radonjic M, Brok MO, van Leenen D, Holstege FC, Timmers HT. 2008. Improved genome-wide localization by ChIP-chip using double-round T7 RNA polymerase-based amplification. *Nucleic Acids Res* 36: e21.

van de Peppel J, Kettelarij N, van Bakel H, Kockelkorn TT, van Leenen D, Holstege FC. 2005. Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol Cell* 19: 511-522.

van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Ne P, Loeillet S et al. 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475: 114-117.

van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, van Leenen D, Groot Koerkamp MJ, Ko CW, Miles AJ et al. 2010. Functional overlap and regulatory links shape genetic interactions between signaling pathways. *Cell* 143: 991-1004.

Vannini A, Cramer P. 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* 45: 439-446.

Visconti R, Palazzo L, Della Monica R, Grieco D. 2012. Fcp1-dependent dephosphorylation is required for M-phase-promoting factor inactivation at mitosis exit. *Nat Commun* 3: 894.

Voytas DF, Boeke JD. 1992. Yeast retrotransposon revealed. *Nature* 358: 717.

Wahba L, Amon JD, Koshland D, Vuica-Ross M. 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell* 44: 978-988.

Wahba L, Gore SK, Koshland D. 2013. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *Elife* 2: e00505.

Wahba L, Koshland D. 2013. The Rs of biology: R-loops and the regulation of regulators. *Mol Cell* 50: 611-612.

Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L et al. 2011. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med* 365: 2497-2506.

Wang L, Mao X, Ju D, Xie Y. 2004. Rpn4 is a physiological substrate of the Ubr2 ubiquitin ligase. *J Biol Chem* 279: 55218-55223.

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R. 2008. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454: 126-130.

Wang X, Xu H, Ha SW, Ju D, Xie Y. 2010. Proteasomal degradation of Rpn4 in Saccharomyces cerevisiae is critical for cell viability under stressed conditions. *Genetics* 184: 335-342.

Werner F, Grohmann D. 2011. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* 9: 85-98.

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

Westover KD, Bushnell DA, Kornberg RD. 2004. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* 303: 1014-1016.

Williamson VM, Young ET, Ciriacy M. 1981. Transposable elements associated with constitutive expression of yeast alcohol dehydrogenase II. *Cell* 23: 605-614.

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

Winston F, Chaleff DT, Valent B, Fink GR. 1984. Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. *Genetics* 107: 179-197.

Wong JMs and Ingles CJ. 2001. A compromised yeast RNA Polymerase II enhances UV sensitivity in the absence of global genome nucleotide excision repair. *Molecular and General Genetics MGG* 264: 842-851.

Wong KH, Jin Y, Struhl K. 2014. TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Mol Cell*. 54: 601-612.

Wu X, Wilcox CB, Devasahayam G, Hackett RL, Arevalo-Rodriguez M, Cardenas ME, Heitman J, Hanes SD. 2000. The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *EMBO J* 19: 3727-3738.

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

Xu Z, Wei W, Gagneur J, Clauder-Munster S, Smolik M, Huber W, Steinmetz LM. 2011. Antisense expression increases gene expression variability and locus interdependency. *Mol Syst Biol* 7: 468.

Yang C, Stiller JW. 2014. Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain. *Proc Natl Acad Sci U S A* 111: 5920-5925.

Yassour M, Pfiffner J, Levin JZ, Adiconis X, Gnirke A, Nusbaum C, Thompson DA, Friedman N, Regev A. 2010. Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biol* 11: R87-2010-11-8-r87. Epub 2010 Aug 26.

Yu X, Chini CC, He M, Mer G, Chen J. 2003. The BRCT domain is a phospho-protein binding domain. *Science* 302: 639-642.

Yudkovsky N, Ranish JA, Hahn S. 2000. A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408: 225-229.

Yuryev A, Corden JL. 1996. Suppression analysis reveals a functional difference between the serines in positions two and five in the consensus sequence of the C-terminal domain of yeast RNA polymerase II. *Genetics* 143: 661-671.

Zehring WA, Lee JM, Weeks JR, Jokerst RS, Greenleaf AL. 1988. The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. *Proc Natl Acad Sci U S A* 85: 3698-3702.

Zenklusen D, Vinciguerra P, Wyss JC, Stutz F. 2002. Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Mol Cell Biol* 22: 8241-8253.

Zhang DW, Mosley AL, Ramisetty SR, Rodriguez-Molina JB, Washburn MP, Ansari AZ. 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 DW, Rodriguez-Molina JB, Tietjen JR, Nemec CM, Ansari AZ. 2012b. Emerging Views on the CTD Code. *Genet Res Int* 2012: 347214.

Zhu X, Wiren M, Sinha I, Rasmussen NN, Linder T, Holmberg S, Ekwall K, Gustafsson CM. 2006. Genome-wide occupancy profile of mediator and the Srb8-11 module reveals interactions with coding regions. *Mol Cell* 22: 169-178.