Investigation into the recruitment and functions of Paf1C

by

Mitchell Alden Ellison II

Bachelor of Arts in Biology, Hood College, 2014

Master of Science in Biotechnology, Johns Hopkins University, 2015

Submitted to the Graduate Faculty of

the Dietrich School of Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Mitchell Alden Ellison II

It was defended on

July 16, 2020

and approved by

Committee Members:

Jeffrey L. Brodsky PhD, Professor, University of Pittsburgh, Biological Sciences

B. Franklin Pugh PhD, Professor, Cornell University, Molecular Biology and Genetics

Andrew P. VanDemark PhD, Associate Professor, University of Pittsburgh, Biological Sciences

Miler T. Lee PhD, Assistant Professor, University of Pittsburgh, Biological Sciences

Thesis Advisor and Committee Chair:

Karen M. Arndt PhD, Professor, University of Pittsburgh, Biological Sciences

Investigation into the recruitment and functions of Paf1C

Mitchell Alden Ellison II, PhD

University of Pittsburgh, 2020

My work focuses on how transcription, the first step in gene expression, is affected by accessory factors and epigenetic modifications. Because gene expression defects are the cause of many diseases it is critical that we understand this fundamental process. Regulation of RNA polymerase II (Pol II) transcription requires accessory factors such as the Polymerase Associated Factor 1 complex or Paf1C. Paf1C facilitates a suite of co-transcriptional histone modifications, regulates transcription elongation efficiency, and performs many other functions. Paf1C must be tethered to the active Pol II elongation complex to function correctly. The Cdc73 and Rtf1 subunits of Paf1C are critical to this interaction. The Rtf1-Spt5 interaction is understood, but interactions governing Cdc73 attachment are unclear. Using an *in vivo* cross-linking strategy, interactions between the C-domain of Cdc73 and both Spt6, an essential elongation factor, and the Pol II C-terminal domain were identified. Cross-linking and mass spectrometry analysis data supported by mutational analysis, revealed that the C-domain of Cdc73 interacts with multiple domains of Spt6. *In vitro* protein binding assays support a direct interaction between recombinant Spt6 and Cdc73. Acute depletion of Spt6 followed by ChIP-seq demonstrates that genome-wide Paf1 occupancy requires Spt6, thus revealing an entirely new mechanism for Paf1C recruitment. Once properly recruited to Pol II, Paf1C regulates the transcriptome; however, its effects on non-coding RNAs (ncRNAs) and pervasive transcription are poorly understood. Using a genetic background that enriches for unstable transcripts, I demonstrated that deletion of *PAF1* affects all classes of Pol II transcripts including antisense transcripts and other classes of ncRNAs. Further analysis revealed that H3K36me3 defects in a *paf1*Δ strain are likely causing much of the de-repression we observe

for both coding and ncRNAs. Taken together these results identify Spt6 as a key Paf1C recruitment factor and demonstrate that Paf1C function is critical to maintain proper transcript levels.

Table of contents

List of tables

List of figures

Preface

First, I want to thank my family. I could not have accomplished any of this work without my wife Christy Ellison. She has taught me how to be a better man and motivated me to do better in life since the day we met. She works a full-time job and takes care of nearly everything at home when I am working late or writing papers. She is amazing! My son Lucas keeps me motivated and, on my toes, reminding me what I was like as a kid and teaching me something new every day. My daughter Alexis always cheers me up when I am feeling down with her great sense of humor and encourages me to be more active and playful when I am working too hard.

My mother and father passed away years ago, but before they died, they taught me many things. My mother Bonnie Ellison used to let my brother and me help her in the garden which really got me interested in plants and biology in general from a young age. My father Mitchell Ellison taught me work ethic and how to not overthink a problem, a skill many scientists lack. My brother Myles Ellison has always been there for me and we have been in some tough spots together, but we always come out on top. I want to thank him for all his advice and encouragement over the years and for showing me, by example, that if I work hard, I can do anything I put my mind to.

Before I move on to give much due thanks to my thesis advisor, I want to mention my in-laws who have supported me so much during graduate school. Thank you to Jeannie Wilbern for always being there for my family and teaching me so much about hard work, commitment, and family values over the years. Thank you to John Wilbern Sr. and Jr., Cassie Wilbern and Dawn Wilbern for always being there to lend me an ear when I needed to vent and being so supportive of my work.

Words cannot describe how grateful I am to Dr. Karen Arndt for being such an outstanding mentor, both in the lab and with regards to navigating life as a scientist. Karen has taught me so much while offering me tons of freedom to develop as an independent scientist. During the course of my thesis work I have developed immense respect for Karen as a person and as a leader and she is truly someone that I look up to and am happy to call a friend and mentor. Karen expects a high standard of work from her students, and in return she provides a high standard of mentorship. She truly puts the lab and her students first, often at the expense of her personal time. I can walk into Karen's office at any time no matter how busy she is and receive advice on my project, which is not something all students can say about their thesis advisor.

At the bench, Karen is one of the most rigorous and honest scientists that I have ever met. She will not believe a result until she has seen it multiple times and is highly critical of data. Karen always says: "your only job as a scientist is to report the truth," I love this! It means that the truth is in the data, and we as scientists can only design experiments to test hypotheses, and report on the data generated from those experiments, honestly, even if it does not agree with our favorite hypothesis. These are both important lessons to learn as a young investigator and Karen teaches them to all of her students.

The last thing I will say here about Karen is that she is an outstanding science writer and has literally taught me everything I know about writing. When I came to the MCDB program my writing skills were weak and I had major problems with grammar and usage. I still have some of those problems (you might even find one in this thesis), but through a lot of hard work and careful mentorship from Karen I have become a much stronger and more confident writer.

I would also like to thank Dr. Margaret Shirra (aka Peggy or Peg). No Arndt lab student gets a PhD without lots of help from Peggy. Ninety percent of the time Peg will drop everything

to help you out if you ask her question. She has helped me so much over the years with all types of questions at the bench, and with some good life advice too. She has also helped keep me honest by pointing out when I need to clean something up or do lab chores.

During my time in the Arndt lab I had the pleasure of working with tons of really smart graduate students. When I joined the lab Dr. Branden Van Oss, Dr. Christine Cucinotta, and Dr. Elizabeth Hildreth were working towards their PhD degrees. I really learned a lot from Christine during the short time we worked together. Elizabeth and I spent many years together in the lab and became great friends. She and I shared many high and low points during graduate school, and I learned a lot during my time working with her. In recent years, Karen hired Brendan McShane as a lab technician, and he has been a great colleague. I will miss discussing biochemistry with him and wish him the best of luck in graduate school. Lastly, I have had the pleasure of working with some great new students since I joined the lab. Alex Francette, Sarah Tripplehorn, and Sanchirmaa Namjilsuren were all a pleasure to mentor and work alongside and I am excited to see how their projects develop over the years.

I have also worked with many undergraduate students. When I started in the lab, I mentored Alex Lederer and worked alongside Rachel Schusteff and Chelsea Guan. Alex was incredibly smart and motivated and was a joy to mentor. Later, Rachel Kocik, Matthew Blacksmith, and Lauren Lotka, joined the lab followed by Julia Seraly, and Nicole Horan. I mentored Matt directly for multiple years after Alex left for graduate school. I always appreciated Matt's work ethic, willingness to ask questions, and his interest in the science we were doing. Once Matt left for graduate school, I began working with Julia Seraly who is an extremely fast learner and has become a skilled bench scientist in short time.

I would like to thank my thesis committee composed of Dr. Jeffrey Brodsky, Dr. Franklin Pugh, Dr. Andrew VanDemark, and Dr. Miler Lee for teaching me how to do better science and offering many helpful suggestions over the years. I would also like to thank previous mentors. Dr. Susan Carney mentored me on my undergraduate honors thesis and showed me that I was capable of doing scientific research and even fairly good at it. Without the opportunity to do undergraduate research that she provided I would have never gone on to obtain a PhD. I would also like to thank Dr. Oney Smith, Dr. Douglas Luster, Dr. Reid Frederick, Dr. Nina Shishkoff, Susan Nester, and Mike McMahon at the USDA ARS FWDSRU for mentoring me during my time as a lab technician.

Finally, I would like to dedicate this thesis to the memory of my grandmother Leora Howe, who showed our family that graduate school was a possibility, and to my fallen brothers from the 1-506 infantry: Corey Garver, Jason (Doc) Pardun, and Jacob Nickell who died either as a direct result of combat (Corey) or as the result of post-traumatic stress syndrome (Doc and Jake) after the war. They sacrificed their lives so that we may all live safely in peace.

1.0 Introduction

The need for a better understanding of gene expression and the epigenetic factors promoting diseases like cancer has never been greater, with over half a million deaths per year resulting from cancer in the United States (Siegel *et al.* 2016). Misregulation of gene expression can occur in many ways and there are a suite of accessory factors involved in promoting proper transcription. In order to understand the work presented here and the relevance of the questions being asked it is necessary to cover extensive background information to ensure understanding of all the factors a play.

1.1 Structure and function of chromatin

The eukaryotic genome is organized into chromatin, which is a combination of DNA and protein (Rando and Chang 2009; Rando and Winston 2012). The repeating unit of chromatin is the nucleosome, but chromatin can achieve higher levels of compaction, allowing it to fit into the nucleus and enabling discrete separation of chromosomes during mitosis (Tremethick 2007). Chromatin structure and nuclear architecture are areas of high scientific interest along with the role of chromatin looping in gene regulation (Rowley and Corces 2018). Here the focus is on chromatin at the nucleosome level because that is the degree of compaction observed at actively transcribed regions (Klemm *et al.* 2019).

1.1.1 The nucleosome

A nucleosome is composed of ~147bp of DNA wrapped around a globular core of histone proteins containing two copies each of histones 2A (H2A), H2B, H3, and H4 (McGinty and Tan 2015). Nucleosomes protect the DNA and pose a barrier to DNA-templated processes such as transcription. Nucleosomes form a stereotypical pattern at actively transcribed regions of the genome (Figure 1; Li *et al.* 2007; Smolle and Workman 2013). There is generally an ordered array of nucleosomes over the coding region, and well-defined nucleosome depleted regions (NDR) at the 5' and 3' ends (Venkatesh and Workman 2015). Nucleosomes impose regulation on transcription by both impeding DNA-templated processes and recruiting protein factors via post-translational modifications to the histone proteins (Berger 2007; Li *et al.* 2007; Smolle and Workman 2013; Venkatesh and Workman 2015).

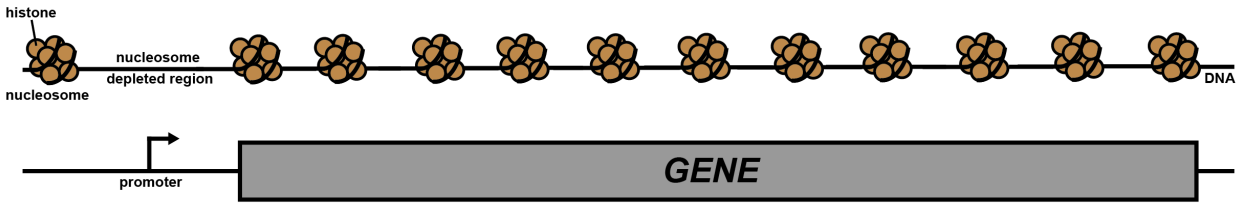


Figure 1. Nucleosomes are composed of histone proteins and form stereotypical patterns at protein-coding loci.

The diagram illustrates the nucleosome occupancy pattern found around an average gene.

1.1.2 Post-translational histone modification

Histones are modified with a suite of chemical moieties of varying sizes including, but not limited to, acetylation, methylation, phosphorylation, ubiquitylation and sumoylation. Most modifications occur on the N-terminal unstructured tail regions of the histones, but modified

residues are also present in the globular core and C-terminal tails of these proteins. Many histone modifications are associated with regions of active transcription. The diagrams below (Figure 2) describe a subset of the post-translational modifications that have been observed (Berger 2007; Li *et al.* 2007; Smolle and Workman 2013; Huang *et al.* 2014; Venkatesh and Workman 2015).

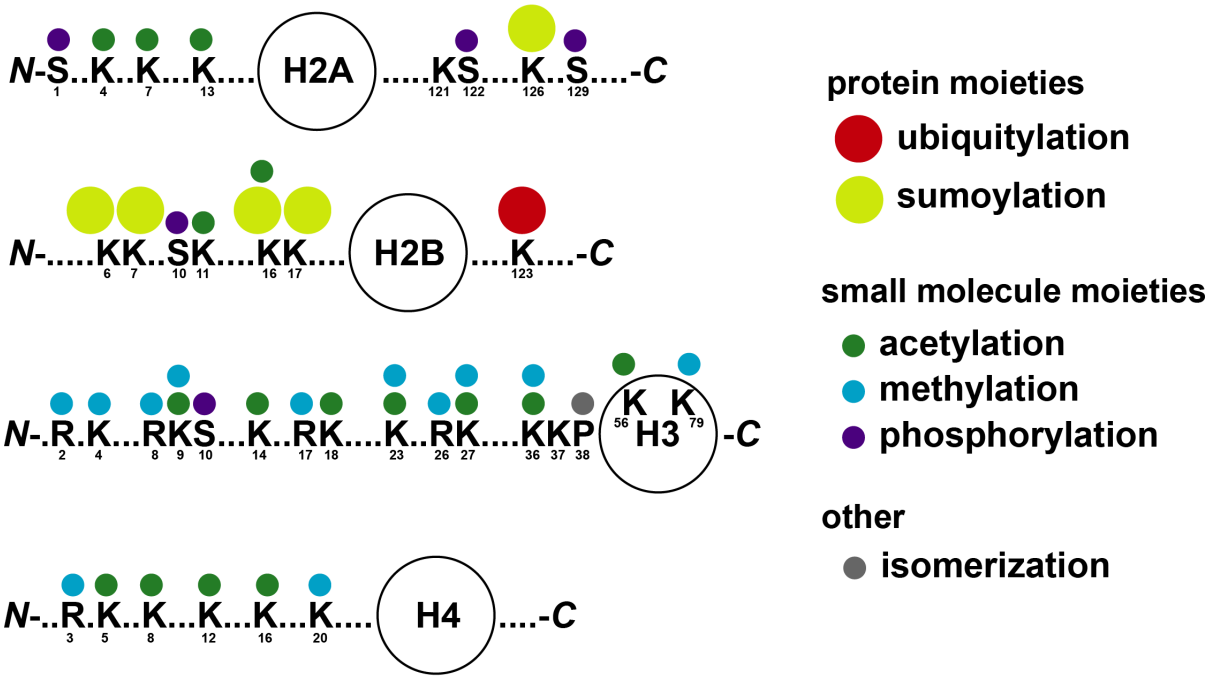


Figure 2. Histone post-translational modifications

The diagram illustrates the four core histones and a subset of the residues that are known to undergo post-translational modification. The residue numbers represent positions in *S. cerevisiae* histone protein sequences. This figure is modified directly from Smolle and Workman, 2012.

1.1.3 Chromatin remodeling factors

Chromatin remodeling factors are molecular machines capable of repositioning nucleosomes using the power of ATP hydrolysis. The chromatin remodeling factors found in yeast are categorized into four subfamilies (listed in Table 1 below) that have overlapping and

independent functions (Bao and Shen 2007; Clapier *et al.* 2017). These molecular machines are critical for maintaining the proper positioning of nucleosomes genome-wide and regulating DNA accessibility at promoters and origins of replication (Lai and Pugh 2017).

Table 1. Chromatin remodeling factors

Subfamily	Chromatin Remodeling Factor	Subunit(s)	Function(s)
INO80	INO80	Rvb1, Rvb2, Arp4, Arp5, Arp8, Act1, Taf14, Ies1 Ies2, Ies3, Ies4, Ies5, Ies6, Nhp10	Regulates transcription, cell cycle, and DNA repair
	SWR1	Swr1, Rvb1, Rvb2, Arp4, Arp6, Act1, Yaf9, Swc4/Eaf2, Swc2/Vsp72, Bdf1, H2AZ, H2B, Swc6/Vps71, Swc3, Swc5, Swc7	Replaces H2A/H2B dimers with H2AZ/H2B dimers, regulates transcription, chromatin stability and cell cycle
ISWI	ISW1a	Isw1, Ioc3	Regulate transcription, promotes proper chromatin assembly and DNA replication through heterochromatin
	ISW1b	Isw1, Itc1, Ioc2, Ioc4	
	ISW2	Isw2, Itc1	
CHD	CHD1	Chd1	Restores regularly spaced nucleosome arrays during transcription elongation.
SWI/SNF	SWI/SNF	Swi2/Snf2, Swi1/Arp6, Swi3, Swp73, Arp7, Arp9, Snf5, Swp82, Taf14, Snf6, Snf11	Directional DNA translocase that regulates transcription and exit from mitosis helps to activate weak origins of replication, binds hyperacetylated nucleosomes.
	RSC	Sth1, Rsc1 or Rsc2, Rsc4, Rsc8, Rsc6, Arp7, Arp9, Sfh1, Rsc3, Rsc5, Rsc7, Rsc9, Rsc10, Rsc30, Htl1, Ldb7, Rtt102	Regulates transcription and cell cycle, involved in sister chromatid cohesion, chromosome segregation and DNA repair

Modified from Bao and Shen (2007) and Clapier et al. (2017)

1.1.4 Histone chaperones

Histone chaperones facilitate the replacement of histone proteins back onto the DNA template in the wake of Pol II transcription (discussed in section 1.2). Two well-studied histone chaperones that play a role in transcription elongation are the FACT complex (Facilitates Chromatin And Transcription complex) made up of Spt16 and Pob3 subunits, and the protein Spt6, which chaperone H2A-H2B and H3-H4 dimers, respectively (Duina 2011). Additional proteins are also categorized as histone chaperones (see Table 2 below) that support or regulate the roles of Spt6 and FACT (Spn1 and Nhp6, respectively) or are involved in DNA replication and repair (Nap1, Asf1, Rtt106, Chz1) (Ransom *et al.* 2010; Duina 2011; Hammond *et al.* 2017; Warren and Shechter 2017).

Table 2. Histone chaperones

Protein/Protein Complex	Function(s)
FACT	Interacts with disrupted nucleosomes produced by Pol II passage, evicts and reassembles nucleosomes, chaperones H2A/H2B dimers during Pol II passage, prevents cryptic initiation and regulates H2BK123ub
Spt6	Interacts directly with Pol II, chaperones H3/H4 dimers helping to reassemble nucleosomes in the wake of Pol II passage, prevents cryptic transcription, required for proper methylation of H3 at lysine 36
Spn1	Inhibits Spt6 binding to nucleosomes by interacting with Spt6 in its N-terminal histone binding region
Nhp6	Facilitates binding of FACT and Spt6 to nucleosomes
Nap1	Promotes nucleosome assembly and interacts directly with H2A/H2B dimers
Asf1	Controls the exchange of histone H3 during transcription
Rtt106	Promotes co-transcriptional H3 deposition via an interaction with H3/H4, Prevents cryptic transcription
Chz1	Exchanges H2A/H2B dimers for H2AZ/H2B and promotes DNA resection

Duina, 2011; McDonald, 2010; Stillman, 2010; Ransom et al., 2010; Cucinotta and Arndt, 2016

1.1.4.1 FACT

Much is known about the structure of the FACT complex and its role in DNA-templated processes such as transcription and DNA replication (VanDemark *et al.* 2006; Vandemark *et al.* 2008; Duina 2011; Liu *et al.* 2020). The FACT complex is composed of the proteins Spt16 and Pob3 in yeast and interacts with disrupted nucleosomes produced by Pol II (with the help of Nhp6) in order to chaperone H2A/H2B dimers, enabling eviction and reassembly of nucleosomes during Pol II passage. FACT stands for FAcilitates Chromatin Transcription because the complex aids in Pol II transcription of chromatin and restoration of chromatin in the wake of the polymerase (Duina 2011). FACT also plays a role in the maintenance of proper histone modification patterns at transcribed regions, ensuring that modified nucleosomes remain properly positioned (Jeronimo *et al.* 2019).

1.1.4.2 Spt6

Suppressor of Ty 6 or Spt6 is an essential elongation factor in *S. cerevisiae*, that was identified in yeast genetic screens for mutants that could suppress the effects of retrotransposon insertions at the 5' end of the *LYS2* gene (*lys2-128∂* allele) and the promoter of the *HIS4* gene (*his4-912∂* allele) (Rando and Winston 2012). Spt6 is conserved in humans where it is referred to as SUPT6H. The protein encoded by the *SPT6* gene was later identified to be a histone chaperone that promotes Pol II CTD serine 2 phosphorylation and the tri-methylation of histone three at lysine 36 (H3K36me3) (Fuchs *et al.* 2008; Dronamraju and Strahl 2014). Spt6 also affects 3' end formation of the *GAL10* transcript and recruits the protein Ctr9, a Paf1C subunit, to *GAL10* and *GAL7* (Kaplan *et al.* 2005). Spt6 is primarily involved in chaperoning H3-H4 histone dimers (Bortvin and Winston 1996) during transcription, where it facilitates their placement back onto the

DNA after Pol II passage (Cheung *et al.* 2008; Jensen *et al.* 2008). Spt6 interacts with Spn1 (Iws1), which competes with histone proteins for binding to its N-terminal domain (McDonald *et al.* 2010). Studies in *Drosophila melanogaster* confirmed that Spt6 co-localizes with another essential elongation factor, Spt4-Spt5, on transcriptionally active genes on polytene chromosomes solidifying its role in transcription elongation (Kaplan *et al.* 2000). Furthermore, Spt6 promotes transcription (Keegan *et al.* 2002), and in its absence, the rate of Pol II elongation decreases (Endoh *et al.* 2004) and the use of cryptic promoters within coding regions (cryptic initiation) increases (Kaplan 2003). Spt6 is structurally homologous to the bacterial transcription factor Tex, but contains two additional domains (see Table 3 below for domain information) (Close *et al.* 2011). One of the two additional domains consists of two Src homology 2 (SH2) domains (Schlessinger 1994) in tandem (Dengl *et al.* 2009; Sun *et al.* 2010; Close *et al.* 2011), and the other is a small globular domain termed the DLD or Death-Like-Domain. Interestingly, the SH2 domains in Spt6 appear to be the common ancestral protein from which all SH2 domains in eukaryotes arose (Dengl *et al.* 2009). It has been suggested that Spt6 is indirectly involved in mRNA turnover by recruiting the Ccr4-Not de-adenylation complex to regions of active transcription, where it is required to degrade specific mRNAs (Dronamraju *et al.* 2018a). Spt6 also prevents cryptic initiation of Pol II transcription; *spt6* mutants exhibit both changes in chromatin architecture and loss of H3K36me3 (Kaplan 2003; Dronamraju and Strahl 2014; Doris *et al.* 2018). Recently, Spt6 was shown to play a role in the maintenance of proper histone modification patterns at transcribed regions by ensuring that modified nucleosomes remain properly positioned (Jeronimo *et al.* 2019).

Table 3. Spt6 protein domains and regions of interest

Domain or Region	*S.c.* Protein Residues	Known Function or Predicted Function	Relevant Mutants in *S.c.*	Reference(s)
Disordered N-terminus	1-300	Interacts with histone proteins, promotes proper chromatin occupancy of Spt6, phosphorylated by CKII, contains Spn1 binding domain		Bortvin and Winston 1996; Dronamraju et al., 2018; Gouot et al., 2018
Spn1 binding	239-268	Binds to Spn1, recruits Spt6 and Swi/Snf to promoters, interacts with nucleosomes		McDonald, 2010; Zhang, 2008;
HtH	336-442	Predicted to be involved in DNA or protein interaction		Close et al., 2011; Vos et al., 2018
YqgF	735-887	Associates with nascent RNA with S1 domain		Close et al., 2011; Vos et al., 2018
HhH	933-1002	Represses cryptic transcription, promotes 3' end formation, promotes H3K36 methylation, predicted to be involved in protein interaction	*spt6-1004*	Kaplan et al., 2003; Kaplan et al., 2005; Youdell et al., 2008; Close et al., 2011; Vos et al., 2018
DLD	1019-1104	Predicted to be involved in protein interaction based on conservation		Close et al., 2011; Vos et al., 2018
S1	1129-1219	associates with nascent RNA with YqgF domain		Close et al., 2011; Vos et al., 2018

tSH2	1250-1440	Binds to the phosphorylated Pol II (Rpb1) CTD linker region, promotes proper Spt6 association with chromatin	*spt6-50*	Dengl et al, 2009; Sun et al., 2010; Close et al, 2011; Sdano et al 2017; Vos et al., 2018, Chun et al., 2019

1.2 The transcription cycle

Transcription, the first step in gene expression, can be broken up into three phases (Figure 3): initiation (Sainsbury et al. 2015), elongation (Arndt and Kane 2003; Saunders et al. 2006; Selth et al. 2010), and termination (Kuehner et al. 2011; Porrua and Libri 2015). The completion of these three phases constitutes one round of the transcription cycle (Svejstrup 2004). Transcription is carried out by RNA polymerase enzymes, which are multi-subunit molecular machines that synthesize RNA using DNA as a template (Cramer et al. 2008). The most well studied RNA polymerase in eukaryotes is RNA polymerase II or Pol II. Transcription by Pol II is the focus of this thesis. Pol II requires numerous accessory factors to help it perform its functions at each stage of the transcription cycle, some of which are essential for life and others that are dispensable (Figure 3). Pol II is responsible for the synthesis of messenger RNA (mRNA), but also is known to synthesize non-coding RNAs, some functional, and others that appear to be the result of pervasive transcription (Xu et al. 2009; Yassour et al. 2010; van Dijk et al. 2011; Schulz et al. 2013).

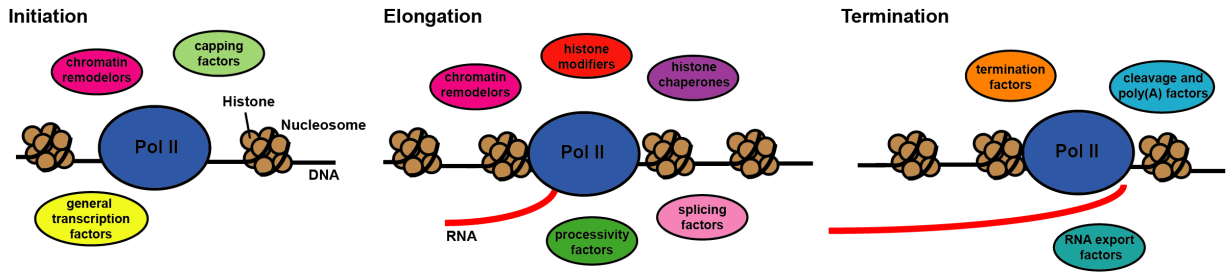


Figure 3. The transcription cycle

A simplified diagram illustrating the transcription cycle and highlighting some of the many accessory factors required to enable Pol II to transcribe a gene.

It is important to note that there are two other RNA polymerases in eukaryotic cells termed Pol I and Pol III. Pol I is responsible for the synthesis of ribosomal RNA, and Pol III is responsible for transcribing tRNA and snRNAs (Cramer *et al.* 2008; Vannini and Cramer 2012). All three polymerases have similar structures and share some subunits, but they vary in protein subunit composition (subunits for Pol I, Pol II, and Pol III listed in Table 4).

Table 4. Three RNA polymerases

Pol I	Pol II	Pol III	Region	Function
A190	Rpb1	C160	Core	Active Center
A135	Rpb2	C128	Core	Active Center
AC40	Rpb3	AC40	Core	
A14	Rpb4	AC17	Stalk	Initiation complex formation
Rpb5	Rpb5	Rpb5	Core	
Rpb6	Rpb6	Rpb6	Core	
A43	Rpb7	C25	Stalk	Initiation complex formation
Rpb8	Rpb8	Rpb8	Core	
A12.2 N-ribbon	Rpb9	C11 N-ribbon	Core	RNA cleavage
Rpb10	Rpb10	Rpb10	Core	
AC19	Rpb11	AC19	Core	
Rpb12	Rpb12	Rpb12	Core	
		C82		
		C34		Pol III specific
		C31		

Modified from Vannini and Cramer (2012) and Cramer et al., (2008)

1.2.1 Initiation

Transcription initiation by Pol II requires the assembly of the Pre-Initiation Complex or PIC at the promoter of a gene, which incorporates the transcription start site (TSS). PIC components (see Table 5) consist of Pol II, transcription factor IIA (TFIIA), TFIIB, TFIID, TFIIE, TFIIF, and TFIIH, which are referred to as the general transcription factors or GTFs (Sainsbury *et al.* 2015). The Mediator coactivator complex is also required for normal levels of transcription and functions at the initiation stage (Plaschka *et al.* 2016).

1.2.1.1 The Pre-Initiation Complex

The PIC (Table 5) begins its assembly with TATA-binding protein (TBP), a member of TFIID, binding to the promoter and creating a bend in the DNA. After TFIID subunit TBP (Spt15 in yeast) has bound and bent the DNA template, TFIIA (Toa1 and Toa2 in yeast) binds and stabilizes the interaction between TFIID and the DNA, which is followed by TFIIB. TFIIF and Pol II join the complex at this point followed by TFIIE (Tfa1 and Tfa2 in yeast) and TFIIH to the polymerase (Sainsbury *et al.* 2015). The TFIIH kinase module member protein Kin28 then phosphorylates the C-terminal domain (CTD; described in detail in 1.2.2.3) of Pol II subunit Rpb1 (Buratowski 2003; Corden 2013; Eick and Geyer 2013). The helicase activity components of the TFIIH core, Rad3 and Ssl2, open the promoter forming the transcription bubble and eventually leading to promoter escape (Sainsbury *et al.* 2015). Thus, the combined function of all these factors allows for promoter recognition, DNA melting and Pol II transcription initiation.

Table 5. Transcription initiation factors.

Initiation Factor	Subunit(s) (S.c.)	Function(s)
TFIIA	Toa1, Toa2	Counteracts negative co-factors and stabilizes TBP
TFIIB	Sua7	TSS selection, TBP binding and Pol II recruitment
TFIID	Tbp, Taf1, Taf2, Taf3, Taf4, Taf5, Taf6, Taf7, Taf8, Taf9, Taf10, Taf11, Taf12, Taf13, Taf14	Promoter recognition and recruitment of Pol II
TFIIE	Tfa1, Taf2	Stabilizes open DNA and recruits TFIIH
TFIIF	Tfg1, Tfg2, Taf14	TSS selection and TFIIB stabilization
TFIIH core	Tfb1, Ssl1, Tfb4, Tfb2, Tfb5, Rad3, Ssl2	Promoter opening and DNA repair
TFIIH kinase module (sometimes referred to as TFIIK)	Ccl1, Kin28, Tfb3	Phosphorylates Pol II CTD

Modified from Sainsbury, Bernecky and Cramer (2015)

1.2.1.2 Transcriptional coactivator complexes

In addition to being a member of the PIC, TFIID is an important coactivator. This coactivator function depends on the conserved TAF proteins within TFIID (TBP-associated factors). In a recent study, acute loss of certain TAFs revealed that ~87% of genes in yeast are TFIID-dependent and the other ~13% depend upon both TFIID and SAGA (described below) (Donczew *et al.* 2020). TFIID functions at both TATA-box containing and TATA-less promoters (Rhee and Pugh 2012; Donczew *et al.* 2020). Thus, TFIID, and not simply its TBP subunit, is extremely important for transcription initiation at nearly all genes.

The Spt-Ada-Gcn5-Acetyltransferase (SAGA) and Mediator complexes are transcriptional coactivators that promote transcription of many genes (Baker and Grant 2007; Samara and Wolberger 2011; Plaschka *et al.* 2016). The Mediator complex is capable of connecting cis-regulatory regions to promoters by forming connections between the transcription factors binding the regulatory regions and the PIC. The Mediator is a very large multi-subunit protein complex (see Table 6 for subunit composition). It is required for normal levels of transcription *in vivo* and is highly conserved in eukaryotes (Plaschka *et al.* 2016).

SAGA is composed of four modules and numerous proteins (described in Table 6) possessing the means to read, write, and erase histone modifications (Baker and Grant 2007; Samara and Wolberger 2011). SAGA has been shown to be as important for transcription initiation as TFIID at a small number of genes and is also important for proper transcription elongation (Baptista *et al.* 2017). The complex has been shown to be involved in activation of genes during the heat shock response in yeast (Vinayachandran *et al.* 2018). In a recent study acute depletion of SAGA subunits was shown to affect ~13% of genes, some of which are also affected by TFIID loss (Donczew *et al.* 2020).

Table 6. The Mediator and SAGA coactivator complexs.

Co-activator Complex	Module	Subunit(s) (S.c.)
Mediator	Head	Med16, Med8, Med11, Srb4, Srb5, Srb2, Srb6
	Middle	Med1, Med7, Med9/Cse2, Med10/Nut2, Rox3, Srb7, Soh1
	Tail	Med2, Med3/Pgd1/Hrs1, Nut1, Gal11, Sin4
	Backbone	Rgr1
	Kinase	Srb8, Srb9, Srb10/Ssn3/Ume5, Srb11/Ssn8/Ume3
SAGA (Spt-Ada-Gcn5-Acetyltransferase)	HAT (histone acetyl-transferase)	Sgf29, Ada3, Gcn5, Ada2
	TAF	Taf5, Taf6, Taf9, Taf10, Taf12
	SPT	Spt3, Spt7, Spt8, Spt20, Tra1, Ada1
	DUB	Sgf73, Sgf11, Ubp8, Sus1

Modified from Plaschka, Kayo and Cramer (2016) and Samara and Wolberger (2011)

1.2.2 Elongation

Transcription elongation begins as initiation ends with promoter escape. In this phase, Pol II begins the productive elongation of the nascent RNA and processive translocation along the DNA. This is a highly dynamic process involving the recruitment of many important protein factors to the elongation complex. Post-translational modifications to histones, Pol II, and elongation factors allow for the selective recruitment of additional factors at various positions relative to the TSS. The following sections will break down the elongation phase of transcription into smaller parts in order to discuss the key components of this complicated process.

1.2.2.1 Elongation factors

Many protein factors are implicated in transcription elongation including Pol II processivity factors (DSIF), histone chaperones (FACT and Spt6), chromatin remodelers (Chd1), pausing factors (P-TEFb, DSIF, NELF, SEC, PARP1, GDOWN1, and Paf1C), CTD kinases (Kin28 and P-TEFb), general transcription factors (TFIIF and TFIIS), other elongation specific factors (ELL, Elongin, LEC), factors involved in RNA surveillance and transcript export from the nucleus (CCR4-NOT and THO/TREX respectively) and factors that recruit histone-modifying enzymes (Paf1C) (a comprehensive list of elongation factors is given in Table 7).

Table 7. Factors involved in transcription elongation by Pol II

Elongation Factor	Subunit(s)	Function(s)
DSIF	Spt4 and Spt5	Stimulates elongation and increases Pol II processivity, suppresses early termination, recruits capping factors, recruits NELF
Spt6	Spt6	Interacts directly with Pol II, chaperones H3/H4 dimers helping to reassemble nucleosomes in the wake of Pol II passage, prevents cryptic transcription, required for proper methylation of H3 K36
FACT	Spt16 and Pob3	Interacts with disrupted nucleosomes produced by Pol II passage, evicts and reassembles nucleosomes, chaperones H2A/H2B dimers during Pol II passage, prevents cryptic initiation and regulates H2B K123ub
Chd1	Chd1	Chromatin remodeler, associates with the elongation complex via Paf1C (in S.c.), promotes even spacing of nucleosomes over coding regions
P-TEFb	CDK9 and CCNT1 or CCNT2 (H.s.) Bur1 and Bur2 (S.c.)	Phosphorylates Pol II CTD, Spt5 CTR and NELF, promotes release from pausing
BRD4-P-TEFb	BRD4 and P-TEFb	Promotes pause release via BRD4 stimulation of P-TEFb
7SK-P-TEFb	7SK snRNP, MEPCE, LARP7, HEXIM1 or HEXIM2, and P-TEFb	Prevents pause release via 7SK snRNP acting as a sink for PTEFb
SEC	AFF1 or AFF4, ELL2, AF9 or ELN, EAF1 or EAF2, P-TEFb	Contains P-TEFb, promotes rapid release of paused Pol II, most active P-TEFb containing complex

Pol II CTD kinases	P-TEFb (H.s.)/Bur1 and Bur2 in (S.c.), Kin28, and Cdk12 (H.s.)/Ctk1 (S.c.)	Promote transcription elongation factor association and disassociation by affecting Pol II CTD modification state via phosphorylation
Pol II CTD phosphatases	Rtr1, Fcp1, Ssu72	Promote transcription elongation factor association and disassociation by affecting Pol II CTD modification state via de-phosphorylation
NELF	NELF-A, NELF-B, NELF-C or NELF-D, NELF-E	Prevents premature termination during pausing and stabilizes the paused state
PARP1	PARP1	Inhibits NELF during pausing via ADP-ribosylation
TFIIS	TFIIS	Allows Pol II to resume elongation after backtrack arrest by stimulating nascent transcript cleavage
TFIIF	Tgf1, Tfg2, Taf14 (S.c.) GTF2F1 and GTF2F2 (H.s.)	General transcription factor essential for preinitiation complex formation, but during elongation it functions to prevent transient pausing after pause release
ELL	EAP20, EAP30, EAP45	Stimulates Pol II elongation rate and alleviates pausing
Elongin	Elongin A, Elongin B, Elongin C	Stimulates Pol II elongation rate and alleviates pausing
LEC	ELL, Eaf, Ice1 and Ice2	Metazoans require for snRNA expression
Paf1C	Paf1, Ctr9, Leo1, Cdc73 (Parafibromin in H.s.) and Rtf1 (contains WDR61/Ski8 in H.s.)	Associates with Pol II and DSIF, capable of promoting pause release and stabilizing paused Pol II, promotes proper co-transcriptional histone modification placement, recruits chromatin remodeler Chd1 (in S.c.), interacts with histone chaperones, promotes proper termination and 3' end formation at snoRNAs, promotes mRNA export
GDOWN1	GDOWN1	Associates with Pol II to stabilize pause
CCR4-NOT	Not1/Cdc39, Not2/Cdc36, Not5, Caf1/Pop2, Ccr4, Caf40,	Influences mRNA production, export and decay, functions both in the nucleus and in the cytoplasm

	Not4/Mot2/Sig1, Caf130, Not3 (in S.c.) and CNOT1, CNOT2, CNOT3, CNOT7, CNOT8, CCR4a, CCR4b, CNOT9, CNOT4, CNOT10, CNOT11 (in H.s.)	
THO	Hpr1, Tho2, Mft1, Thp2	Supports transcription of long and/or high GC content transcripts
TREX	THO + Sub2 and Yra1	Involved in pre-mRNA splicing and mRNA surveillance and export, interacts with Spt6

Modified from: Sims, Belotserkovskaya and Reinberg (2004); Sainsbury, Bernecky and Cramer (2015); Cucinotta and Arndt (2016); Van Oss, Cucinotta and Arndt (2017); Collart (2016); Chen, Smith and Shilatifard (2018)

1.2.2.2 RNA polymerase II pausing

Promoter proximal pausing is a phenomenon observed in most eukaryotes, but not in *S. cerevisiae*, whereby Pol II initiates transcription and then pauses before entering into productive elongation (Adelman and Lis 2012; Liu *et al.* 2015; Chen *et al.* 2018). This is separate from the pausing that occurs later during transcription elongation (Churchman and Weissman 2011). Promoter proximal pausing allows for genes to be transcribed rapidly upon pause-release. Many elongation factors are involved in the process of Pol II promoter proximal pausing (hereafter referred to as pausing) and pause release (Chen *et al.* 2018). When Pol II is in this paused state, it is bound by NELF (negative elongation factor) and DSIF (Spt4-Spt5) (Vos *et al.* 2018a; b). Some studies suggest that this state is stabilized by Paf1C and with the help of GDOWN1 (Chen *et al.* 2015, 2018; Van Oss *et al.* 2017); however others support a role for Paf1C in pause release (Yu *et al.* 2015b; Lu *et al.* 2016; Vos *et al.* 2018a; b). The Pol II CTD (described in detail below) appears to be phosphorylated on serine residues at positions five and seven (S5-S7) in the heptad repeat

during the pause (Chen *et al.* 2018). Pause-release is achieved as multiple changes take place in the elongation complex. The major factor in pause-release is the protein kinase P-TEFb (Bur1/Bur2 in *S.c.*), which phosphorylates several proteins in the elongation complex, including DSIF (Spt5 CTR), NELF, Paf1C, and the largest subunit of Pol II, Rpb1, both on its CTD and its linker region (Vos *et al.* 2018a; b, 2020; Chen *et al.* 2018; Chun *et al.* 2019). These phosphorylation events result in the release of NELF and exchange for Paf1C, which then occupies the NELF binding surface of Pol II, and entry into productive elongation (Vos *et al.* 2018a; b).

P-TEFb can exist in various forms when combined with additional factors. P-TEFb interacts with the 7SK snRNP in its inactive form. It can interact with Brd4 forming BRD4-P-TEFb, which is capable of promoting pause release. P-TEFb interacts with other proteins in the absence of Brd4 to form what is referred to as super elongation complex (SEC). SEC is the most active form of P-TEFb capable of inducing rapid release from pausing (Chen *et al.* 2018). The mechanism described above is general, and Pol II pausing is still an active area of investigation.

1.2.2.3 RNA polymerase II C-terminal domain modification

The transcription cycle has been carefully characterized at mRNA loci over many decades of scientific study. One fundamental characteristic of the transcription cycle, present at every phase, is the association of additional protein factors with Pol II (Figure 3). One way these interactions are spatially and temporally regulated is by modification of the C-terminal domain (CTD) of the Rpb1 subunit of Pol II, which serves as a binding platform for protein factors (Jeronimo *et al.* 2013; Srivastava and Ahn 2015; Harlen and Churchman 2017a).

The CTD of Pol II consists of a repeated series of seven amino acids. The consensus sequence for these repeats is YSPTSPS, which is repeated 26 times in *Saccharomyces cerevisiae,* and 52 times in *Homo sapiens*. The residues that make up this sequence can be post-translationally

modified. Three of the four residue types in the consensus sequence (Y, S, and T) are competent for phosphorylation and the fourth (P) can undergo cis-trans isomerization (Buratowski 2003; Bataille *et al.* 2012; Jeronimo *et al.* 2013; Corden 2013; Eick and Geyer 2013). These residues are modified in the context of the transcription cycle and have been shown in many cases to mediate the binding of protein factors (Jeronimo *et al.* 2013; Srivastava and Ahn 2015; Harlen and Churchman 2017a).

CTD residues are phosphorylated by specific kinases at different times in the transcription cycle (Table 8 contains a list of the known CTD kinases and phosphatases) (Jeronimo *et al.* 2013; Eick and Geyer 2013; Harlen and Churchman 2017a). During the initiation phase, when Pol II is recruited by the general transcription factors, the CTD is in its un-phosphorylated state, but during the initiation process, it is phosphorylated at S5 and S7 (S5P-S7P) by the essential protein kinase Kin28, which facilitates promoter escape (Sainsbury *et al.* 2015). During the elongation phase of transcription, levels of phosphorylated S5 (S5P) decrease as Y1P and S2P levels increase. The kinase responsible for Y1P is not known in *S. cerevisiae*, but the kinases responsible for modifying S2 are known. Both Ctk1, a non-essential S2 kinase, and Bur1, an essential cyclin-dependent kinase, have been shown to phosphorylate S2 during the transcription cycle. These modifications decrease sharply near the termination site where T4P peaks. Each of these modification states generates a unique binding platform for protein factors that can be used to coordinate the addition of factors involved in transcription regulation (Jeronimo *et al.* 2013; Srivastava and Ahn 2015).

Post-translational modifications on the CTD must be removed to restart the CTD cycle, and there are specific phosphatases that are responsible for the removal of phosphate moieties in *S. cerevisiae*. The phosphatases responsible for removing Y1P are Rtr1 and Glc7 (Hsu *et al.* 2014; Schreieck *et al.* 2014). Phosphatases for T4P have not yet been elucidated. The Ssu72 phosphatase

removes phosphates from both S5P and S7P residues of the CTD. The Fcp1 phosphatase primarily removes phosphates placed at S2P but can also remove phosphates placed at S5P. Interestingly Fcp1 activity is promoted by the presence of Ssu72 in *in vitro* reactions. Additionally, Rtr1 and Cdc14 phosphatases have also been shown to dephosphorylate the CTD at S5 (Hsu *et al.* 2014). Finally, in addition to the phosphatases, there is a CTD proline isomerase, Ess1, that has been shown to affect the activity of Ssu72 (Bataille *et al.* 2012; Jeronimo *et al.* 2013). Ssu72 interacts directly with CTD peptides when P6, is in a *cis* conformation and promotes rapid dephosphorylation of S5P by Ssu72 (Werner-Allen *et al.* 2011). Together the suite of enzymes described here allows for extensive modification of the Pol II CTD to facilitate the selective recruitment of transcription factors.

Table 8. Phosphorylated residues of the Pol II C-terminal domain heptad repeat sequence

	Kinase		Phosphatase		
CTD Residue	**Yeast**	**Human Homologs**	**Yeast**	**Human Homologs**	**Associated Processes**
Tyrosine 1		?	Rtr1, Glc7	Rpap2	CTD stability, antisense and enhancer transcription and inhibition of recruitment of transcription termination factors
Serine 2	Bur1 and Ctk1	Cdk9 and Cdk12	Fcp1	Ctdp1	Promoter-proximal pausing, pause release, mRNA splicing, transcription elongation, transcription termination and DNA topology
Threonine 4	?	Plk3 and Cdk9	?	?	Chromatin remodeling, processing of histone mRNA, transcription elongation, transcription termination and post-transcriptional mRNA splicing

Serine 5	Kin28	Cdk7	Rtr1, Cdc14 and Ssu72	Rpap2 and Ssu72	Chromatin modification, transcription initiation, capping, splicing, ncRNA transcription and transcription termination
Serine 7	Kin28 and Bur1	Cdk7 and Cdk9	Ssu72	Ssu72	Interaction with integrator complex, interaction with Mediator complex, snRNA expression

Harlen and Churchmen, 2017; Cucinotta and Arndt, 2016; Schreieck et al., 2014

1.2.2.4 Spt5 C-terminal repeat modification

The Spt4-Spt5 complex is universally conserved, meaning that it is present in all three domains of life (Werner and Grohmann 2011). Spt5 and its binding partner Spt4 were identified as transcription factors in yeast through a genetic screen (Winston *et al.* 1984) assessing suppression of the effects of a Ty δ element insertion at the 5' end of the *HIS4* locus in *S. cerevisiae*. They were later found to confer sensitivity to the drug DRB in HeLa cell nuclear extract and were together termed the DRB sensitivity inducing factor (DSIF) complex (Wada *et al.* 1998). DSIF, referred to as the Spt4-Spt5 complex in *S. cerevisiae*, is recruited to the Pol II elongation complex very early in transcription elongation. It binds to Pol II at the 5' end of genes and remains associated all the way to the termination site (Mayer *et al.* 2010; Van Oss *et al.* 2016). In humans, there are data supporting the hypothesis that Spt4-Spt5 is handed off from Myc to the polymerase early in the transcription cycle (Baluapuri *et al.* 2019). Once bound, Spt5 has been shown to decrease stalling, thus increasing the processivity of the polymerase (Bourgeois *et al.* 2002; Zhu *et al.* 2007). Interestingly, like Rpb1, Spt5 also has a repeated polypeptide sequence that is competent for phosphorylation.

In *S. cerevisiae*, Bur1 (CDK9 in humans), with its cyclin partner Bur2 (CCNT1 or CCNT2 in humans), phosphorylates Spt5 in addition to phosphorylating the CTD (Zhou *et al.* 2009; Liu *et al.* 2009; Qiu *et al.* 2012). Although promoter-proximal pausing does not appear to occur in *S. cerevisiae,* these proteins are still necessary to promote transcription elongation. The region of Spt5 phosphorylated by Bur1 is referred to as the C-terminal repeat region or CTR, which is made up of 15 repeats of a six amino acid consensus sequence. The consensus sequence (SAWGGQ) has one phospho-competent residue, the serine at position one in yeast (Hartzog and Fu 2013). The serine residues in the CTR of Spt5 are phosphorylated by Bur1 around the same time that serine two on the CTD repeats is being phosphorylated by Bur1 (Qiu *et al.* 2012; Cortazar *et al.* 2019). Mass spectrometry analysis detects phosphorylated serine residues only at repeats 10-13 and 15 suggesting not every phospho-competent residue is phosphorylated *in vivo* (Liu *et al.* 2009).

Phosphorylation of the CTR of Spt5 also controls Pol II elongation speed in humans, where a detailed genomics analysis of the role of Spt5 in the control of Pol II elongation rate and termination has been conducted (Cortazar *et al.* 2019). In this study the authors found that hyper-phosphorylated Spt5 results in faster polymerase speeds and dephosphorylation leads to a slowing of the polymerase (Cortazar *et al.* 2019). Phosphorylated Spt5 is observed beginning at the transcription start site and is rapidly lost following the cleavage and polyadenylation site marking the transition between the elongation and termination phases of transcription (Cortazar *et al.* 2019).

1.2.2.5 Co-transcriptional histone modifications

Many post-translational histone modifications are placed co-transcriptionally as Pol II translocates DNA occupied by nucleosomes (Table 9, Figure 4), a process that requires the help of histone chaperones, elongation factors, and histone-modifying enzymes (Li *et al.* 2007; Smolle and Workman 2013). This section will only cover the modifications most relevant for this thesis,

which are the modifications commonly associated with transcription elongation and the Paf1C. Many of these modifications are part of a cascade of modifications starting with the ubiquitylation of H2B at residue K123 (H2BK123ub).

H2B K123ub is a prerequisite for many of the methyl modifications associated with active transcription (Sun and Allis 2002; Dover *et al.* 2002; Xiao *et al.* 2005; Nakanishi *et al.* 2009). The ubiquitin moiety is placed on H2B K123 by the enzymes Rad6 and Bre1 after they are recruited by the Rtf1 subunit of Paf1C (Van Oss *et al.* 2016). This modification is required for H3 K4me2 and H3 K4me3 placed by Set1 histone methyltransferase and H3 K79me2 and H3 K79me3 placed by Dot1 histone methyltransferase (Sun and Allis 2002; Dover *et al.* 2002; Nakanishi *et al.* 2009). H2B K123ub, H3 K4me2/3 and H3 K79me2/3 are active marks meaning that they are thought to promote transcription (Li *et al.* 2007; Smolle and Workman 2013; Wood *et al.* 2018).

Di- and tri-methylation of H3 at lysine 36 (H3K36me2/3) does not require H2BK123ub and is involved in repressing transcription. H3 K36me2/3 is placed by Set2 methyltransferase and found in the middle (H3K36me2) and 3' ends (H3K36me3) of actively transcribed genes (Li *et al.* 2007; Smolle and Workman 2013). This histone modification is important because it helps prevent transcription from initiating within gene bodies. When this mark is placed, it leads to the recruitment of histone deacetylase (HDAC) complexes: H3K36me2 recruits Set3C, whereas H3K36me3 recruits Rpd3S (Carrozza *et al.* 2005; Joshi and Struhl 2005; Keogh *et al.* 2005; Govind *et al.* 2010; Venkatesh *et al.* 2012). These two complexes remove acetyl groups from the lysine residues in the histones. Histone acetylation is generally known to be an activating mark, so by removing these modifications, these HDAC complexes ensure that transcription initiation from within gene bodies is limited by maintaining a repressive chromatin environment (Kim *et al.* 2016).

Table 9. Co-transcriptional histone modifications relevant to this work

Modification	Modifying Enzyme(s) in S.c	Enzyme(s) Responsible for Modification Removal	Recognized by	Function in Transcription
H2BK123ub (S.c.) H2BK120ub (H.s)	Rad6 and Bre1	Ubp8 and Ubp10	Csp35	Active, promotes H3K4me and H3K79me
H3K4me	Set1	Jhd2	PHD, Chromo, ADD, Tudor, MBT, Zf-CW	Active
H3K36me	Set2	Jdh1, Rph1, Gis1	Chromo, PHD, PWWP	Repressive (recruits Rpd3S histone deacetylase complex)
H3K79me	Dot1		Tudor	Active

Modified from: Smolle and Workman; 2012; Li, Carey and Workman, 2007; Wood, Tellier, and Murphy, 2018

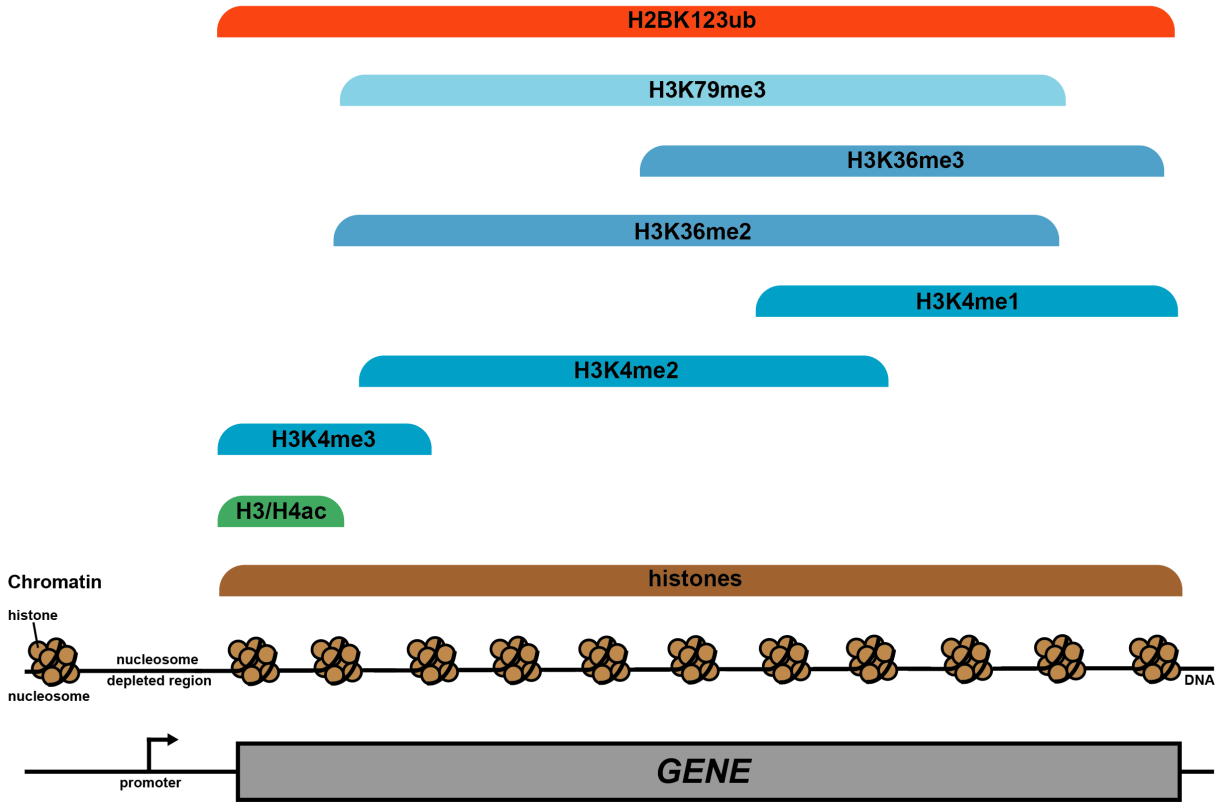


Figure 4. Co-transcriptional histone modification patterns found at transcribed genes.

1.2.3 Termination

Termination of Pol II transcripts occurs by two primary mechanisms in *S. cerevisiae,* and each is responsible for terminating the majority of transcripts of a given class (see Table 10 for a detailed breakdown). The first termination mechanism is cleavage and polyadenylation dependent termination that occurs at mRNA genes and some long ncRNA genes such as stable unannotated transcripts (SUTs) and Xrn1-sensitive unstable transcripts (XUTs). The second is Nrd1-Nab3-Sen1 (NNS) dependent termination that operates at shorter ncRNAs such as cryptic unstable transcripts (CUTs), Nrd1 unterminated transcripts (NUTs), small nuclear RNAs (snRNAs), and

small nucleolar RNAs (snoRNAs). Both of these pathways are essential in *S. cerevisiae* (Kuehner *et al.* 2011; Porrua and Libri 2015).

Table 10. Termination factors

Termination complex	Subunits	Protein interactions and/or enzymatic activity	Suspected role in CPA-dependent termination	Suspected role in NNS-dependent termination	Transcript classes terminated (S.c.)
Cleavage Factor 1A (CF1A)	Pcf11	S2P CTD binding	Disrupts Pol II hybrid, promotes RNA cleavage and Rat1 recruitment	Interacts with Pol II CTD and RNA to disrupt the Pol II hybrid	mRNA, SUT, XUT
Cleavage and Polyadenylation Factor (CPF)	Cft1 (S.c.) CPSF160 (H.s.)	S5P, S2P CTD binding, poly(A) RNA binding	Promotes pausing and RNA cleavage to allow for Rat1 entry.	Unknown	mRNA, SUT, XUT
CPF	Ysh1 (S.c.) CPSF73 (H.s.)	Cleavage of poly(A) RNA	Creates entry point for Rat1	Creates entry-point for exoribonuclease	mRNA, SUT, XUT
CPF	Yth1 (S.c.) CPSF30 (H.s.)	Binds poly(A) RNA and Pol II	Promotes pausing	Unknown	mRNA, SUT, XUT
CPF-APT (associated with Pta1, a subcomplex of CPF)	Glc7 (S.c.) PP1 (H.s.)	S/T phosphatase, dephosphorylates Sen1	Unknown	Promotes Sen1 activity and/or recruitment	mRNA, SUT, XUT
CPF-APT	Pta1 (S.c.) Symplekin (H.s.)	Binds both CFIA and CPF	Unknown	Holds together the APT complex	mRNA, SUT, XUT
CPF-APT	Ssu72	S2P CTD phosphatase	Promotes Pcf11 recruitment to Pol II	Promotes Pcf11 recruitment to Pol II	mRNA, SUT, XUT

Rat1-Rai1-Rtt103	Rat1 (S.c.) XRN2 (H.s.)	5'-3' exoribonuclease that degrades the cleavage product generated by Ysh1/CPSF73	Runs into Pol II near the RNA exit channel and promotes the recruitment of Pcf11	No effect	mRNA, SUT, XUT
Rat1-Rai1-Rtt103	Rai1 (S.c.) DOM3Z (H.s.)	De-capping endoribonuclease and pyrophosphohydrolase	Promotes activity and stability of Rat1	Unknown	mRNA, SUT, XUT
Rat1-Rai1-Rtt103	Rtt103	Binds Rat1 and S2P Pol II CTD	Recruits Rat1 and Pcf11 to Pol II	Unknown	mRNA, SUT, XUT
Nrd1-Nab3-Sen1 (NNS)	Nrd1 (S.c.) SCAF8 and SCAF 4 (H.s.)	S5P CTD binding and RNA binding	No effect	Binds to Pol II CTD and RNA, recruits Sen1 and disrupts Pol II hybrid	snRNA, snoRNA, CUT, NUT
NNS	Nab3	Binds to RNA and both Nrd1 and Sen1	Unknown	Recruits Sen1 to Pol II	snRNA, snoRNA, CUT, NUT
NNS	Sen1 (S.c.) Senataxin (H.s.)	Pol II CTD binding and 5'-3' RNA-DNA helicase	Exposes RNA promoting Rat1 activity	Breaks up the DNA-RNA hybrid in Pol II	snRNA, snoRNA, CUT, NUT

Modified from Kuehner, Pearson and Moore (2011) and Porrua and Libri (2015)

1.2.3.1 CPA-dependent termination

Cleavage and polyadenylation dependent termination is carried out in response to a polyadenylation signal (AAUAAA) in the transcript. The phosphorylation state of the Spt5-CTR has also been shown to play a role in cleavage and polyadenylation dependent termination with hypo-phosphorylation resulting in a slowing of the polymerase (Cortazar *et al.* 2019). This slowdown coincides with the recruitment of the termination factor Xrn2 (Rat1 in yeast) and a slowing of the polymerase. Recruitment of the CF1A and CPF machinery happens in response to

both the polyadenylation signal and Pol II CTD serine two and serine five phosphorylation. Once recruited, CPF cleaves the nascent RNA allowing it to be polyadenylated. The polymerase continues to transcribe even after cleavage and requires Rat1 of the Rat1-Rai1-Rtt103 complex to rapidly translocate the RNA and force it out of the Pol II active site (Kuehner *et al.* 2011; Porrua and Libri 2015).

1.2.3.2 NNS-dependent termination

NNS dependent termination acts in a slightly different manner. This termination mechanism begins with the recruitment of Nrd1 and Nab3 via interactions with the nascent RNA and the Pol II CTD phosphorylated at S5. These RNA binding proteins recruit Sen1, which is a helicase that can disrupt the DNA-RNA hybrid resulting in termination (Kuehner *et al.* 2011; Porrua and Libri 2015).

1.3 Non-coding transcription

In addition to messenger RNAs (mRNA), Pol II transcribes several non-coding RNAs (ncRNA), which are not translated into protein products in the cytoplasm. Some Pol II-transcribed ncRNAs perform well-defined functions (Matera *et al.* 2007), while others appear to be the result of pervasive transcription (Xu *et al.* 2009; Yassour *et al.* 2010; van Dijk *et al.* 2011; Schulz *et al.* 2013). The best-studied classes of ncRNAs are ribosomal RNAs (rRNA), which are transcribed by RNA polymerase I (Pol I), and transfer RNAs (tRNA), which are transcribed by RNA polymerase III (Pol III).

Two other well-studied groups of ncRNAs are small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs). Some snRNAs and snoRNAs are transcribed by Pol III, but the majority, at least in the case of snoRNAs, are transcribed by Pol II. . Two more classes of Pol II transcripts, CUTs, SUTs, were found in experiments performed on strains deleted for *RRP6,* which encodes the catalytic subunit of the nuclear exosome, or grown in varying media conditions, respectively (Xu *et al.* 2009). Additionally, XUTs were identified in strains deleted for the *XRN1* gene, a protein associated with cytosolic RNA degradation. XUTs are often transcribed from regions antisense to coding regions (van Dijk *et al.* 2011).

In more recent years, two more classes of transcripts were identified through the analysis of yeast strains depleted or lacking a key factor, followed by RNA sequencing and *de novo* transcript calling. The first is NUTs, which are enriched upon depletion of the essential Nrd1 protein (Schulz *et al.* 2013). The second is SRATs or Set2-repressed antisense transcripts (Venkatesh *et al.* 2016), which are antisense transcripts that are produced upon deletion of the Set2 methyltransferase due to loss of H3K36me2 and H3K36me3 and thus relief of suppression of intergenic transcription. Antisense transcripts are a major class of ncRNA that often overlap with other non-coding transcripts. These transcripts are produced from the strand of DNA antisense to the mRNA template strand (Faghihi and Wahlestedt 2009; Pelechano and Steinmetz 2013) and can arise from both nucleosome-free regions at the 3' ends of mRNA (Xu *et al.* 2009) genes or from cryptic initiation sites within the body of mRNA genes (Kim *et al.* 2016).

Non-coding RNA transcription has been shown to regulate the transcription of various mRNA genes (Martens *et al.* 2004; Vera and Dowell 2016; Raupach *et al.* 2016). The transcription of CUTs can promote and interfere with mRNA transcription depending upon the location of the CUT in relation to the mRNA gene (Vera and Dowell 2016). Antisense transcription has been

shown to decrease transcript levels at mRNA loci with which they overlap (Crisucci and Arndt 2011b; Castelnuovo *et al.* 2013); however, no global correlation exists between antisense and sense (mRNA) transcription (Murray *et al.* 2015).

1.4 Polymerase associated factor 1 complex

In yeast, Paf1C is composed of 5 protein subunits (Table 11): Paf1, Ctr9, Leo1, Cdc73, and Rtf1 (Wade *et al.* 1996; Shi *et al.* 1997; Mueller and Jaehning 2002; Squazzo *et al.* 2002) and plays numerous critical roles during the transcription cycle.

Table 11. Polymerase associated factor 1 complex subunits and functional domains

Protein	Domain	Residues in S.c	Function
Cdc73	C-domain	230-393	Required for proper Paf1C association with chromatin at coding regions
Rtf1	Chd1 interacting domain	1-30	Required for Rtf1-Chd1 interaction
Rtf1	HMD	74-139	Required for H2BK123ub and downstream methylation marks
Rtf1	OAR/Plus3 domain	238-370	Required for proper Paf1C association with chromatin at coding regions
Rtf1	Paf1C interaction	442-558	Required for Rtf1 to associate with the rest of the Paf1C

Paf1	Leo1 Binding	150-232	Required for Leo1 interaction with Paf1
Leo1	Paf1 binding	164-259	Required for Paf1 interaction with Leo1
Ctr9	TPR1-16	56-89 138-174 183-216 218-251 298-332 338-371 373-405 421-455 462-495 501-534 540-572 664-697 699-731 732-764 768-801 830-863	Predicted to mediate complex assembly and/or other protein-protein interactions

Warner et al., 2007; Amrich et al., 2012; Chu et al., 2013; Mayekar et al., 2013; Wier et al., 2013; Van Oss et al., 2016; Xu et al., 2017

Immunoaffinity chromatography using an antibody targeting the unphosphorylated C-terminal domain (CTD) of Rpb1, the largest subunit of Pol II, identified Paf1 as Polymerase Associated Factor 1, thus providing a name for the complex (Wade *et al.* 1996). Subsequent studies in *S. cerevisiae* revealed Paf1C as a five-member protein complex by discovering interactions with Cdc73, Ctr9, Rtf1, and Leo1 (Shi *et al.* 1997; Mueller and Jaehning 2002; Squazzo *et al.* 2002). In humans, the Paf1C has an additional subunit, Ski8, that is not present in the *S. cerevisiae* Paf1C (Zhu *et al.* 2005). Conservation between yeast and humans for the five shared Paf1C members is high (Tomson and Arndt 2013), like much of the machinery required for transcription (Werner and Grohmann 2011; Vannini and Cramer 2012), meaning that what we learn about the Paf1C and transcription regulation in yeast often applies in the human context.

1.4.1 Functions attributed to Paf1C

Many functions have been attributed to Paf1C (Figure 5). Paf1C is present at every stage of the transcription process and has even been shown to play a role in downstream processes. It can be found at low levels at promoters and the 5' end of genes, but is found at much higher levels in the middle of the gene and near the 3' end (Krogan *et al.* 2002; Pokholok *et al.* 2002) with occupancy peaking near the +3 nucleosome (Van Oss *et al.* 2016). Paf1C is best understood as an elongation factor that promotes histone modifications, such as H2BK123ub (Piro *et al.* 2012; Van Oss *et al.* 2016) and H3K36me3 (Chu *et al.* 2007). H2BK123ub also promotes H3K4me2/3 and H3K79me2/3, so Paf1C is indirectly involved in promoting these marks as well (Sun and Allis 2002; Dover *et al.* 2002). In addition, Paf1C associates with 3' end termination factors, and loss of its subunits leads to transcriptional read-through of small nucleolar RNA (snoRNA) terminators (Sheldon *et al.* 2005; Nordick *et al.* 2008; Tomson *et al.* 2011, 2013a; Ellison *et al.* 2019). Finally, Paf1C is implicated in post-transcriptional processes such as poly(A) tail formation and nuclear export of certain mRNAs (Penheiter *et al.* 2005; Nordick *et al.* 2008; Yang *et al.* 2016; Fischl *et al.* 2017).

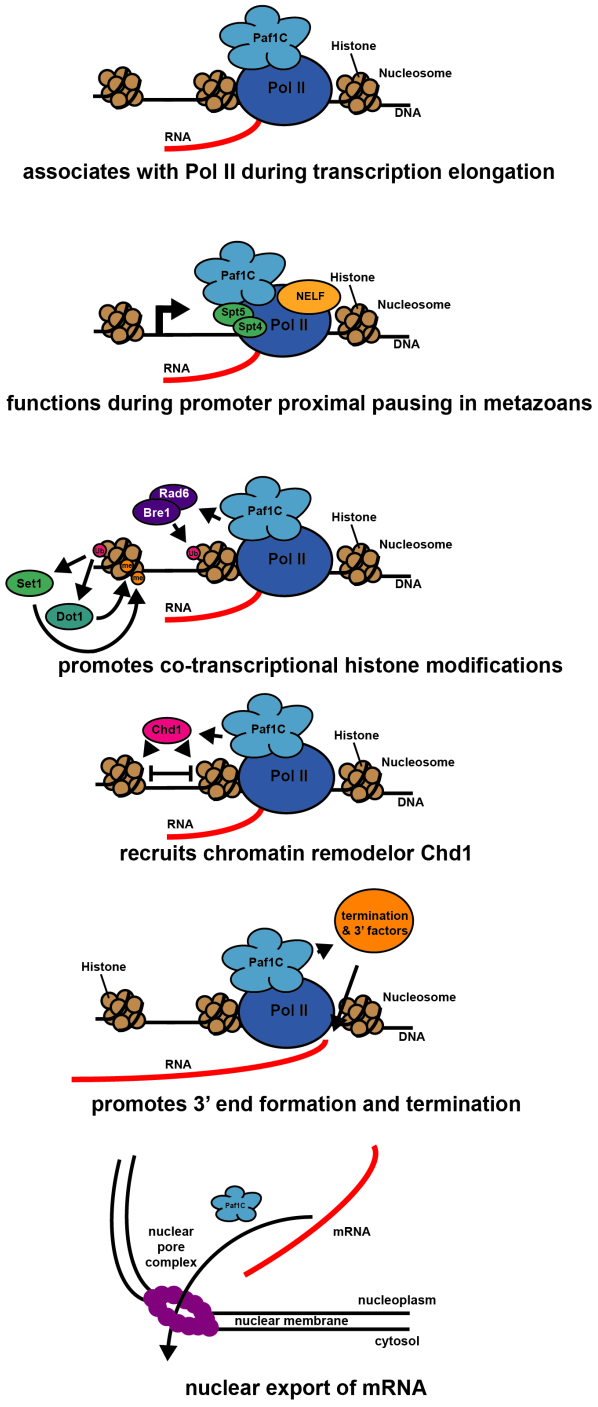


Figure 5. The many functions of the Paf1 complex.

Functions attributed to the Paf1C organized by position in the transcription cycle. Diagrams modified from and/or generated based on information published in Tomson and Arndt (2013), Van Oss, Cucinotta and Arndt (2017), and Fischl *et al.* (2017).

1.4.1.1 Co-transcriptional histone modification

Some examples of histone modifications that are promoted by Paf1C and associated with active transcription are H3K4me2/3, H3K79me2/3, and H2BK123ub (Li *et al.* 2007). Paf1C facilitates the deposition of the H2BK123ub mark in *Saccharomyces cerevisiae* via its Rtf1 subunit (Piro *et al.* 2012; Van Oss *et al.* 2016). Rtf1 directly recruits Rad6 E2 ubiquitin conjugase and Bre1 E3 ubiquitin ligase via its histone modification domain (Van Oss *et al.* 2016). H2BK123ub is a prerequisite to H3K4me2/3 and H3K79me2/3 by the histone methyltransferases Set1 and Dot1 (Sun and Allis 2002; Dover *et al.* 2002); therefore, Paf1C also promotes these modifications through its association with Rad6. Paf1C also mediates the deposition of H3K36me3 by the Set2 histone methyltransferase (Chu *et al.* 2007), although the involvement of Paf1C in the placement of this modification is less well understood. Importantly, all of the histone modifications that Paf1C influences are associated with regions of active transcription (Li *et al.* 2007). These results support a major role for Paf1C in facilitating the proper co-transcriptional modification of chromatin.

1.4.1.2 Transcription termination of snoRNAs

Previous studies in our lab have revealed that when various members of Paf1C are deleted, snoRNAs genes are improperly terminated (Sheldon *et al.* 2005; Tomson *et al.* 2011, 2013a; b). Moreover, Paf1C has been shown to associate with termination and 3' end factors (Tomson and Arndt 2013). This role for Paf1C in controlling proper snoRNA termination suggests the complex works upstream of the Nrd1-Nab3-Sen1 (NNS) transcription termination pathway (Arndt and Reines 2014) responsible for the termination of snoRNAs and other non-coding transcripts (Schulz *et al.* 2013). Many of these same non-coding transcripts are rapidly degraded immediately after their synthesis by the nuclear exosome.

1.4.1.3 Recruitment of the chromatin remodeler Chd1

The chromatin remodeler Chd1 was identified in a yeast 2-hybrid screen for proteins that interact with Rtf1. Rtf1 and Chd1 were both found to associate with chromatin at transcribed regions of the genome (Simic *et al.* 2003), where a deletion of *RTF1* leads to a large reduction in Chd1 occupancy. Further genetic analysis of *chd1*Δ revealed that it could suppress a cold-sensitive phenotype conferred by a specific *spt5 Cs*$^{-}$ mutation, which is also suppressed by mutations in Paf1C members Paf1 (*paf1-49*) and Leo1 (*leo1-43*), but not Rtf1 (Squazzo *et al.* 2002). Further analysis of Rtf1 internal deletion mutants revealed that the N-terminus of Rtf1 is where the Rtf1-Chd1 interaction takes place (Warner *et al.* 2007). These results suggest that the N-terminus of Rtf1 is involved in recruiting Chd1 to active genes during transcription elongation.

1.4.1.4 Nuclear export of mRNA

Paf1 was shown to interact with the Hrp1 protein, a member of the THO/TREX complex involved in the nuclear export of mRNAs (Chang *et al.* 1999; Betz *et al.* 2002). More recent work demonstrated that transcripts from genes with high Paf1 occupancy relative to Pol II subunit Rpb3 (termed Paf1-enriched), measured by transcription elongation factor sequencing (TEF-seq) or native elongating transcript sequencing (NET-seq), show an increase in nuclear retention upon Paf1 deletion. In contrast, for genes at which Paf1 occupancy is equivalent to or less than Rpb3 occupancy, no differences in transcript export were observed (Fischl *et al.* 2017). These results suggest that Paf1C may be helping transcripts to be properly exported from the nucleus at genes with high Paf1/Rpb3 (TEF-seq/NET-seq) occupancy ratios.

1.4.1.5 Pol II pausing in metazoans

When Pol II pauses, it is bound by NELF and DSIF (Spt4-Spt5) and stabilized by GDOWN1 (Chen *et al.* 2018). The major factor in pause-release is P-TEFb (Bur1/Bur2 in S.c.), which phosphorylates elongation complex members (DSIF (Spt5 CTR), NELF and Paf1C) and the Pol II CTD at S2 (Vos *et al.* 2018a; b; Chen *et al.* 2018). These phosphorylation events are thought to be promoted by Paf1C with Ctr9 and Ski8 occupying the NELF binding surface of Pol II (Vos *et al.* 2018a; b). However, in addition to the model of Paf1C promoting in pause release (Yu *et al.* 2015b; Lu *et al.* 2016; Vos *et al.* 2018a; b), there is also research indicating that Paf1C might stabilize the pause (Chen *et al.* 2015). The role of Paf1C in Pol II pausing is not entirely clear and thus is an active area of investigation.

1.4.2 Development

Paf1C has been implicated in development and stem cell biology in humans and mice (Ding *et al.* 2009; Ponnusamy *et al.* 2009). Paf1C is required for organ development in zebrafish and mice. Cdc73 is required for the development of viable mouse embryos (Akanuma *et al.* 2007; Wang *et al.* 2008; Nguyen *et al.* 2010; Langenbacher *et al.* 2011). Decreased protein levels of Paf1C members lead to increased expression of genes responsible for lineage specification and decreased expression of genes responsible for the maintenance of pluripotency (Ding *et al.* 2009; Strikoudis *et al.* 2016). Additionally, depletion of Paf1C members also results in a reduction in expression of genes encoding transcription factors important for maintenance of pluripotency (Ding *et al.* 2009; Ponnusamy *et al.* 2009).

1.4.3 Disease relevance

Paf1C has been demonstrated to both promote and suppress various disease phenotypes in humans. It is implicated in resistance to viruses that infect humans, such as influenza and HIV (Liu *et al.* 2011; Marazzi *et al.* 2012). Through its role in promoting histone modifications and interactions with MLL domain-containing proteins, the Paf1C also plays a role in Mixed Lineage Leukemia (MLL) (Guenther *et al.* 2005; Muntean *et al.* 2010; Smith *et al.* 2011).

In addition to leukemia, mutations in Paf1C may promote numerous other cancers indirectly by affecting its role in promoting histone modifications. H3 lysine 4 di- and tri-methylation (H3K4me2/3) are reduced in lung, liver, prostate, kidney, breast, and pancreatic cancers. These cancers typically result in poor prognosis and low survival rates (Chervona and Costa 2012; Greer and Shi 2012). Over 40% of all new cancer cases and approximately 50% of the cancer deaths in the United States in 2016 came from these six types of cancers (Siegel *et al.* 2016). H3K4me2/3 are both deposited co-transcriptionally, but only if another histone modification is present. The modification of H2B at K120 (K123 in *Saccharomyces cerevisiae*) by the addition of a ubiquitin (H2BK120ub) begins a cascade of co-transcriptional histone modifications resulting in the placement of H3K4me2/3, as well as other important histone modifications (Sun and Allis 2002; Dover *et al.* 2002; Van Oss *et al.* 2016) with links to cancer such as H3K79me2/3 (Audia and Campbell 2017). Thus, if the deposition of H2BK120ub is compromised, there is a reduction in H3K4 and K79 methylation, a hallmark of many cancers (Chervona and Costa 2012; Greer and Shi 2012; Audia and Campbell 2017).

Cdc73, which is termed Parafibromin in humans (Chaudhary *et al.* 2007; Zhao *et al.* 2007; Takahashi *et al.* 2011), is a known tumor suppressor gene. Most of the cancer-related mutations in Cdc73 are truncating mutations that result in loss of its C-domain (Wang *et al.* 2005). Similarly,

truncations and point mutations of Cdc73 result in the development of hyperparathyroidism-jaw tumor syndrome (HPT-JT), a disease of the human endocrine system that results in parathyroid tumor formation (Shattuck *et al.* 2003). These parathyroid tumors are cancerous in about two-thirds of HPT-JT patients (Wang *et al.* 2005).

1.4.4 Recruitment

Paf1C is critical for proper regulation of transcription and associates with Pol II at all stages of the transcription cycle. Thus, it is not surprising that many diseases and developmental defects arise when members of this complex are lost or damaged by mutation. Although the Paf1C associates with Pol II during all phases of the transcription cycle (Krogan *et al.* 2002; Pokholok *et al.* 2002), there is a large increase in its occupancy levels towards the middle and 3' ends of genes (Mayer *et al.* 2010; Van Oss *et al.* 2016). This is most likely because the two known recruitment mechanisms for Paf1C involve phosphorylation events promoted by kinases that modify the CTD and CTR at around the same distance from the TSS that Paf1C occupancy increases (Qiu *et al.* 2009, 2012; Mayer *et al.* 2010; Bataille *et al.* 2012; Cortazar *et al.* 2019). The interactions between Paf1C components and the CTD and CTR have been well characterized (Qiu *et al.* 2012; Mayekar *et al.* 2013; Wier *et al.* 2013), but it is not yet known whether both binding events are equally as important at all genes or if one or the other is utilized more at a subset of genes. It is also possible that there is a universal recruitment mechanism involving Paf1C binding to both the CTD and CTR. If so, there is still the question of how much each of these binding events contributes to Paf1C recruitment.

1.4.4.1 Recruitment via the Pol II CTD – Cdc73 interaction

Paf1C recruitment has been argued to be dependent on the phosphorylation state of the Pol II CTD with S2P and S5P modifications producing the strongest interactions with Paf1C members *in vitro* (Qiu *et al.* 2012). Conversely, it was shown that deletion of the major S2 kinase Ctk1 has no effect on Paf1C subunit occupancy *in vivo* (Ahn *et al.* 2004). It is clear that Bur1 kinase and its cyclin partner Bur2, which promote CTD S2P, play an important role in Paf1C recruitment (Zhou *et al.* 2009; Liu *et al.* 2009; Qiu *et al.* 2012; Mayekar *et al.* 2013; Wier *et al.* 2013); however the Bur1/Bur2 kinase complex is responsible for phosphorylation of the Pol II CTD (Qiu *et al.* 2009; Liu *et al.* 2009; Suh *et al.* 2016; Chun *et al.* 2019), its linker region (Chun *et al.* 2019), and the Spt5 CTR (Liu *et al.* 2009; Qiu *et al.* 2012). Therefore, effects observed in Bur1/Bur2 mutants cannot always be directly attributed to CTD S2P loss. Having described these caveats I will now proceed to describe the data in the literature that suggest an interaction between Cdc73 and the Pol II CTD may promote Paf1C recruitment and the reasoning behind its interpretation.

Inhibition of Kin28 results in a slight reduction of Paf1 occupancy (Qiu *et al.* 2012). Kin28 is a member of the TFIIH kinase module, also known as TFIIK, which is responsible for phosphorylating S5 and S7 of the CTD repeats. The Kin28-phosphorylated CTD is thought to recruit Bur1/2 leading to phosphorylation of S2 of the CTD repeats, which feeds forward to promote the recruitment of the Ctk1 kinase complex, which phosphorylates S2 on additional CTD repeats (Qiu *et al.* 2009, 2012). Thus, it is not surprising that there is an effect on Paf1 occupancy when Kin28 activity is inhibited because it functions at the beginning of this pathway (Buratowski 2003; Qiu *et al.* 2009; Corden 2013; Eick and Geyer 2013; Chun *et al.* 2019).

Detailed analysis of interactions between Pol II CTD peptides in different phosphorylation states and both recombinant and native *S. cerevisiae* Paf1C have been conducted (Qiu *et al.* 2012).

Results from these studies suggest that multiple Paf1C members are competent to interact with the CTD of Pol II in an *in vitro* setting. Cdc73 and Rtf1 purified from yeast bound to phosphorylated peptides S5P, S5P-S7P and S2P-S5P with S2P-S5P peptides resulting in the strongest binding. A phosphomimetic (serine mutated to aspartic acid) mutant S2D-S5D was not competent to bind Cdc73 or Rtf1, indicating that the phosphorylated serine is important for this interaction and not simply the negative charge. Similar results were obtained for Cdc73 and Rtf1 using recombinantly expressed GST-tagged proteins. Ctr9 (594-855) was also competent to bind CTD peptides, but only in the S2P-S5P configuration, suggesting that it might play a role in CTD recognition as well. Upon further analysis of Cdc73, only residues 201-393 at the C-terminus were necessary for S2P, S5P, and S2P-S5P CTD peptide interactions. Mutation of Cdc73 residues W357 and W380 to alanine resulted in loss of Cdc73 occupancy at the *ARG1* locus (Qiu *et al.* 2012).

A study published by our research group in collaboration with the VanDemark lab revealed the structure of the C-domain of Cdc73 (PDB-ID 3V46) (Amrich *et al.* 2012). The C-domain crystal structure revealed that this highly conserved portion of Cdc73 has a Ras-like fold with a flattened out GTP binding pocket that may have been evolutionarily repurposed to participate in a protein-protein interaction. The crystal structure revealed that W357 and W380 are facing inward participating in structurally important interactions, and therefore, it is likely that mutation of those residues to alanine results in unfolding of the C-domain. Loss of the C-domain of Cdc73 did not affect overall Paf1C formation, but did lead to a reduction in Cdc73, Rtf1, and Ctr9 occupancy at coding regions without affecting Rpb3 (Pol II subunit) occupancy. The effects of deleting *CDC73* or its C-domain on Paf1C-promoted histone modifications in yeast are minimal. However, when combined with the deletion of *RTF1*, these mutants completely abolish H3K4me2, H3K4me3, H3K36me3, and H3K79me2/3 marks (Amrich *et al.* 2012).

Together these data support a role for CTD phosphorylation in the recruitment of Paf1C and indicate that Cdc73 is probably the major contact point between the CTD and the Paf1C. Although direct interaction between the CTD and Cdc73 *in vivo* was not observed in these studies, it is inferred from the peptide binding experiments (Qiu *et al.* 2012) and *in vivo* analysis of yeast strains lacking the C-domain of Cdc73 (Amrich *et al.* 2012), that this domain is involved in the Paf1C-CTD interaction.

1.4.4.2 Recruitment via the Spt5 CTR – Rtf1 interaction

In addition to its role in histone modification, the Paf1C subunit Rtf1 interacts directly with the CTR of Spt5 and is extremely important for Paf1C localization. A domain of Rtf1 termed the OAR (Open reading frame Associated Region) in yeast, or Plus3 domain in humans, is responsible for this interaction (Mayekar *et al.* 2013; Wier *et al.* 2013). The Rtf1 OAR binds directly to the phosphorylated CTR of Spt5, and this interaction is critical for proper Paf1C levels at coding regions (Mayekar *et al.* 2013; Wier *et al.* 2013).

Bur1/2 was initially shown to phosphorylate the CTR while other CTD kinases such as Ctk1 and Kin28 do not (Zhou *et al.* 2009). The deletion of the last six repeats of the CTR of Spt5 was shown to affect only H3K36me3, while the deletion of all 15 results in a reduction of H2BK123ub, H3K4me3 and H3K36me3 (Zhou *et al.* 2009). Loss of H3K36me3 had been previously observed in yeast lacking Bur2 (Chu *et al.* 2007). Importantly, Zhou et al. showed that Rtf1 occupancy is reduced *in vivo* when the CTR is deleted.

Further studies revealed that Bur1/2 phosphorylates the Spt5 CTR on repeats 10-13 and 15. This conclusion was reached using an analog sensitive (AS) form of Bur1 and an ATP analog specific to the AS version of the enzyme in conjunction with mass spectrometry analysis (Liu *et al.* 2009). Subsequent experiments further implicated Bur1, and the latter CTR repeats in the

recruitment of the Paf1 subunit of the Paf1C and demonstrated a reduction in H3K4me3 in the context of a full CTR deletion (Liu *et al.* 2009), which was in agreement with previous work (Zhou *et al.* 2009).

An analysis of interactions between phosphorylated Spt5 CTR repeat peptides and both recombinant and native yeast Paf1C revealed interactions with multiple Paf1C members *in vitro* (Qiu *et al.* 2012). Native Rtf1 and Cdc73 were both competent to interact with the phosphorylated CTR peptides or weakly interact with phosphomimetic peptides (S-D). These same proteins did not interact with CTD peptides that were phosphomimetic, demonstrating that CTD peptide binding strictly requires phosphorylation, whereas CTR peptide binding can occur, albeit weakly, if only a similar charge is provided. Experiments using recombinant proteins and CTR peptides revealed that Ctr9 (594-855) was also competent to interact with the phosphorylated CTR peptides. The results of these experiments are very similar to the results of the experiments conducted with Pol II CTD peptides in the same article, suggesting some promiscuity of phosphorylated peptide binding in this *in vitro* context for Cdc73, Rtf1, and Ctr9 (Qiu *et al.* 2012).

In vivo analysis of Spt5 mutants with either all CTR serines mutated to alanine (Liu *et al.* 2009; Qiu *et al.* 2012), aspartic acid, or glutamic acid (Qiu *et al.* 2012) confirmed that the serines in the CTR are necessary for Paf1 occupancy, but that a phosphomimetic residue is also sufficient to promote low levels of Paf1 recruitment. Further investigation by our group revealed that a single domain of Rtf1, termed the OAR, was both necessary and sufficient to promote Rtf1 recruitment to transcribed regions. Analysis of Spt5 CTR, Bur2, and Rtf1 mutants, using co-immunoprecipitation and chromatin immunoprecipitation, support the conclusion that the Spt5 CTR is recruiting Rtf1 via its OAR. In addition, the OAR is important for co-transcriptional histone modifications, such as H3K4me2/3, and H3K79me2/3, because of its involvement in Paf1C

recruitment (Mayekar *et al.* 2013). In a collaborative paper between our group and the VanDemark lab, a crystal structure was reported elucidating the interaction between the Plus-3 domain of human Rtf1 (equivalent to OAR in yeast) and a phosphorylated Spt5-CTR peptide. Guided by the crystal structure, residues R366 and F441 in human Rtf1, corresponding to R251 and Y372 in the yeast protein, were found to be indispensable for this interaction (Wier *et al.* 2013). Structural analysis of Rtf1 bound to the Pol II elongation complex corroborates this finding (Vos *et al.* 2020).

Together, these data support the conclusion that the phosphorylated Spt5-CTR recruits Rtf1 to transcribed regions via a single domain of the Rtf1 protein. We later learned that another domain of Rtf1 termed the HMD (histone modification domain) is sufficient to target the HMD itself to chromatin, likely through its interaction with the acidic patch of the nucleosome, but in this context, the HMD is no longer restricted to coding regions (Van Oss *et al.* 2016; Cucinotta *et al.* 2019). Thus, the Spt5-Rtf1 interaction specifically recruits Rtf1 to the Pol II elongation complex enabling the Paf1C to fulfill its many roles in transcription.

1.4.5 Thesis aims and rationale

The thesis work presented here consists of three Aims, all of which are investigated in *Saccharomyces cerevisiae*. The first is directed at determining the transcriptome-wide effects of deletion of the *PAF1* gene. The second and third are focused on a novel interaction between Spt6 and Cdc73 that I hypothesized to be involved in recruiting Paf1C to Pol II. Aim 2 uses a genomics approach to determine the role of Spt6 in Paf1C recruitment and Aim 3 digs deeper into the interaction in an attempt to identify the residues in each protein that are necessary for the interaction to take place. The work described in Chapters 2 and 3 is the result of these Aims which together inform us about the recruitment and function of Paf1C, a key epigenetic regulator.

1.4.5.1 Biomedical significance

It has become clear in recent years that proteins interacting with our genomic DNA, namely histone proteins, can accept a suite of modifications that are important for the regulation of gene expression (Li *et al.* 2007; Smolle and Workman 2013). There is still very little known about how these modifications are placed and their functions. We do know that misplacement of modifications or loss of protein-protein interactions that normally direct their placement can result in disease states (Chervona and Costa 2012; Greer and Shi 2012; Audia and Campbell 2017). One interaction that is critical for the proper placement of a subset of histone modifications is the interaction between Pol II and Paf1C (Amrich *et al.* 2012; Piro *et al.* 2012; Wier *et al.* 2013). Paf1C is a highly-conserved transcription elongation factor that has been assigned many functions (Tomson and Arndt 2013). Its most well-studied function is promoting co-transcriptional histone modifications (Crisucci and Arndt 2011a; Van Oss *et al.* 2016). Some of the modifications promoted by Paf1C, such as histone H3 lysine 4 (K4) di-methylation (H3K4me2), are associated with cancers where decreases in these marks result in poor patient prognosis (Chervona and Costa 2012; Greer and Shi 2012).

1.4.5.2 Aim 1: Investigation into the *paf1*Δ transcriptome

When I started my dissertation research, the transcriptome-wide effects of *PAF1* deletion were poorly understood, particularly with respect to ncRNA transcription. Trf4 is a factor involved in the termination and degradation of ncRNA transcripts (Schmid and Jensen 2008). By deleting this factor, the degradation of ncRNAs can be inhibited, allowing them to be assayed. *I hypothesized that the loss of Paf1 would affect the entire transcriptome, including ncRNAs.* To test this hypothesis, I performed bioinformatic analysis of tiling array data generated by Travis Mavrich to identify the full impact of Paf1 loss on the transcriptome. Briefly, whole-genome tiling

arrays (work done by Travis) were used to analyze steady-state RNA levels in *paf1*Δ and *paf1*Δ *trf4*Δ strains and isogenic wild-type, or *trf4*Δ controls, respectively. In addition to the analysis I performed follow-up experiments to identify relevant factors that phenocopy the differences observed in *paf1*Δ. *Together these results, which are the subject of Chapter 2, determined the effects of Paf1 loss on the transcriptome and potential mechanisms by which changes in transcription occurred.*

1.4.5.3 Aim 2: Investigate the role of Spt6 in Paf1C recruitment

Rachel Schusteff and Eleanor Kerr (former undergraduates in the lab) identified a previously unrecognized interaction between Cdc73 and Spt6 by site-specific *in vivo* protein cross-linking. Additionally, data published by Kaplan et al. (2004) suggested that Ctr9 recruitment is dependent upon Spt6 at the *GAL10* and *GAL7* genes. *I hypothesized that Paf1C depends upon Spt6 for proper recruitment and localization at transcriptionally active regions.* To test this hypothesis, I rapidly depleted Spt6 from yeast and analyzed the genome-wide occupancy of Paf1 and other relevant elongation complex components. *These experiments, covered in Chapter 3, provided information on how the Cdc73-Spt6 interaction controls Paf1C recruitment.*

1.4.5.4 Aim 3: Elucidate the interacting surfaces on Spt6 and Cdc73

In Aim 3, I determined the location of the binding interface shared between Cdc73 and Spt6 by cross-linking and mass spectrometry (XL-MS). *I hypothesized that mutations resulting in disruption of this interaction would affect Paf1C occupancy*. To test this hypothesis, I substituted residues at this interface to alanine and tested *in vivo* chromatin occupancy to identify residues mediating the interaction and validate the model derived from my XL-MS data. *In Chapter 3, I present results from the chemical crosslinking experiments that identified binding sites on both*

Spt6 and Cdc73 that are mediating their interaction, and further biochemical and mutational analysis validating the Spt6-Cdc73 interaction model.

2.0 Paf1 broadly regulates the transcriptome

The work in this chapter is published in the journal *Genetics* and is re-printed here in altered form.The full citation for the article in print is: Ellison M. A., A. R. Lederer, M. H. Warner, T. N. Mavrich, E. A. Raupach, et al., 2019 The Paf1 Complex Broadly Impacts the Transcriptome of Saccharomyces cerevisiae. Genetics 212: 711–728. https://doi.org/10.1534/genetics.119.302262. This is in accordance with the journal's permission statement which states: "Permission from the GSA is not needed if you will use the material in an article published in GENETICS or if you are reproducing an article (on which you are an author) for your dissertation."

All tiling array data collection was performed by Dr. Travis Mavrich together with Dr. Corey Nislow's lab. Dr. Mavrich performed this work when he was a technician in the Arndt lab, prior to his dissertation research. Data presented for chromatin remodeling factors was from the thesis work of former Arndt lab student Dr. Marcie Warner. The Python script that was used to average tiling array probes over genes in the tiling array data analysis protocol was generated by an undergraduate I mentored, Alex Lederer. Alex and Dr. Beth Raupach, a former Arndt lab student, produced strains used in the northern blot analysis conducted on the *FET4* gene and its upstream CUTs.

2.1 Introduction

In the context of chromatin, accurate and controlled transcription by RNA polymerase II requires the functions of many regulatory factors. One highly conserved regulatory factor is Paf1C,

which in yeast is composed of Paf1, Ctr9, Leo1, Rtf1, and Cdc73 (Jaehning 2010; Crisucci and Arndt 2011a; Tomson and Arndt 2013). Paf1C associates with Pol II during transcription elongation and regulates both transcriptional and post-transcriptional processes, including the co-transcriptional deposition of histone modifications and the nuclear export of RNAs (Tomson and Arndt 2013; Fischl *et al.* 2017; Van Oss *et al.* 2017). Histone modifications dependent on Paf1C include H3 lysine 4 di- and tri-methylation (H3 K4me2/3), H3 K79me2/3, and H2B K123 mono-ubiquitylation (ub) in *S. cerevisiae* (H2B K120 in humans). Paf1C facilitates the deposition of H2B K123ub via its Rtf1 subunit (Piro *et al.* 2012; Van Oss *et al.* 2016), which directly interacts with the ubiquitin-conjugating enzyme Rad6 through its histone modification domain (Van Oss *et al.* 2016). H2BK123ub is a prerequisite to H3 K4me2/3 and H3 K79me2/3, modifications catalyzed by the histone methyltransferases Set1 and Dot1, respectively (Sun and Allis 2002; Dover *et al.* 2002). Paf1C also promotes the deposition of H3 K36me3 by the Set2 histone methyltransferase (Krogan *et al.* 2003; Chu *et al.* 2007). Consistent with the binding of Paf1C to the Pol II elongation machinery (Qiu *et al.* 2006, 2012; Amrich *et al.* 2012; Wier *et al.* 2013; Xu *et al.* 2017; Vos *et al.* 2018b), the histone modifications dependent on Paf1C are found at regions of active transcription (Smolle and Workman 2013).

The absence of specific histone modifications in Paf1C mutants is associated with transcriptional defects. These defects include the transcriptional read-through of terminators found at the 3’ ends of Pol II-transcribed small nucleolar RNA (snoRNA) genes (Sheldon *et al.* 2005; Terzi *et al.* 2011; Tomson *et al.* 2011, 2013). In addition to promoting a histone modification state that facilitates transcription termination, Paf1C physically associates with proteins implicated in transcription termination and RNA 3’-end formation (Nordick *et al.* 2008; Rozenblatt-Rosen *et al.* 2009). The importance of Paf1C in regulating snoRNA termination supports a functional

interaction with the Nrd1-Nab3-Sen1 (NNS) transcription termination pathway (Arndt and Reines 2014; Porrua and Libri 2015), which is responsible for the termination of snoRNAs and other noncoding transcripts including cryptic unstable transcripts (CUTs) in yeast (Schulz *et al.* 2013). Many of these same noncoding transcripts are rapidly degraded by the nuclear exosome through a process mediated by the Trf4/Trf5-Air1/Air2-Mtr4 polyadenylation complex (TRAMP) (Schmid and Jensen 2008). For example, loss of Trf4, the polyA polymerase subunit of the TRAMP complex that adds short polyA tails to transcripts destined for degradation or processing by the nuclear exosome, has been shown to stabilize CUTs and snoRNAs in *S. cerevisiae* (LaCava *et al.* 2005; Vaňáčová *et al.* 2005; Wyers *et al.* 2005; Thiebaut *et al.* 2006; Nishimura *et al.* 2009; Xu *et al.* 2009).

Despite the growing understanding of the molecular functions of Paf1C, few studies have probed how these functions lead to a transcriptional outcome. Moreover, little is known about the roles of Paf1C in controlling noncoding transcription. To begin to address these questions, we sought to comprehensively investigate the importance of Paf1C in modulating the *S. cerevisiae* transcriptome, taking advantage of a genetic background that allows enhanced detection of unstable transcripts. To this end, we used strand-specific whole-genome tiling arrays to measure steady-state RNA levels in *PAF1* and *paf1*Δ strains that contain or lack the TRAMP subunit gene *TRF4*. We found that the deletion of *PAF1* affects all classes of Pol II transcripts, including both stable and unstable noncoding RNAs and antisense transcripts. Comparisons with published NET-seq experiments, which detect Pol II-engaged, nascent transcripts (Harlen and Churchman 2017b), indicate that most, but not all, changes in steady-state transcript abundance in the *paf1*Δ background can be attributed to altered transcription. Analysis of subsets of protein-coding genes suggests that Paf1 represses the transcription of some genes through facilitating H3 K36me3 and

stimulates the transcription of other genes independently of any single Paf1C-dependent histone modification. Finally, we report a regulatory mechanism governing the *FET4* locus, which incorporates both CUT transcription and Paf1. Together these data support a role for Paf1C in multiple regulatory mechanisms that collectively and broadly impact the Pol II transcriptome.

2.2 Materials and methods

2.2.1 Yeast strains and culturing methods

All *S. cerevisiae* strains used in this study are listed in Table 12 and are isogenic to the FY2 strain, which is a *GAL2*$^{+}$ derivative of S288C (Winston *et al.* 1995). The deletion of specific loci was achieved by one-step gene disruption (Lundblad *et al.* 2001) and confirmed by PCR. Genetic crosses were conducted as described in Rose (1991) (Rose *et al.* 1991). Cells were grown to log phase at 30°C in rich media (YPD) supplemented with 400 µM tryptophan and harvested by centrifugation. Cell pellets were washed once with sterile water, flash-frozen, and stored at -80°C prior to RNA isolation for RT-qPCR and Northern blotting experiments.

Table 12. Yeast strains used in Chapter 2

Strain[1]	Genotype
KY292 (FY118[2])	*MAT***a** *his4-912∂ lys2-128∂ leu2Δ1 ura3-52 trp1Δ63*
KY307 (FY838[2])	*MATα his3Δ200 lys2Δ202 leu2Δ1 ura3-52*
KY508 (FY737[2])	*MAT***a** *his3Δ200 lys2-128∂ leu2Δ1 ura3-52 snf2Δ::HIS3*
KY583 (GHY280[3])	*MAT***a** *his3Δ200 lys2-128∂ leu2Δ1 ura3-52 trp1Δ63 chd1Δ::HIS3*
KY632	*MATα his3Δ200 lys2-128∂ leu2Δ1 ura3-52 chd1Δ::URA3*
KY802	*MAT***a** *his3Δ200 lys2-173R2 ura3Δ(0 or 52) paf1Δ::URA3*
KY804	*MATα his3Δ200 leu2Δ2(0 or 1) ura3(Δ0 or -52) paf1Δ::URA3*

KY884	*MAT***a** *his3Δ200 lys2-173R2 leu2Δ1 ura3-52 trp1Δ63 isw2Δ::HIS3*
KY901	*MATα his3Δ200 lys2-128∂ leu2Δ1 ura3-52 trp1Δ63 isw1Δ::HIS3*
KY907	*MAT***a** *his3Δ200 lys2-128∂ leu2Δ1 ura3-52 set1Δ::HIS3*
KY914	*MATα his3Δ200 lys2-173R2 leu2Δ1 ura3-52 set2Δ::HIS3*
KY934	*MATα his3Δ200 leu2Δ1 trp1Δ63 dot1Δ::HIS3*
KY938	*MATα his3Δ200 leu2Δ1 trp1Δ63 set1Δ::HIS3*
KY972	*MATα his3Δ200 lys2-128∂ leu2Δ1 ura3-52 swr1Δ::KANMX*
KY1021	*MAT***a** *his4-912∂ lys2-128∂ leu2Δ1 trp1Δ63*
KY1235	*MAT***a** *his3Δ200 lys2-173R2 ura3-52 rco1Δ::HIS3MX6*
KY1250	*MAT***a** *his3Δ200 leu2Δ1 ura3-52 set2Δ::HIS3 paf1Δ::KANMX*
KY1683	*MATα his3Δ200 lys2-128∂ leu2Δ1 trp1Δ63*
KY1702	*MAT***a** *leu2Δ0 ura3Δ0 paf1Δ::KANMX*
KY1952	*MATα his4-912∂ lys2-128∂ trp1Δ63 leu2Δ1 bre1Δ::KANMX*
KY2012	*MAT***a** *leu2Δ0 ura3Δ0 trf4Δ::CLONAT*
KY2016	*MAT***a** *leu2Δ0 ura3Δ0 trf4Δ::CLONAT paf1Δ::KANMX*
KY2027	*MATα ura3-52 (hta2-htb2)Δ::KANMX*
KY2045	*MATα his3Δ200 leu2Δ1 trp1Δ63 rad6Δ::KANMX*
KY2167	*MATα ura3Δ0 HTA1-htb1-K123R (hta2-htb2)Δ::KANMX*
KY2239	*MATα his4-912∂ lys2-128∂ trp1Δ63 ctr9Δ::KANMX*
KY2241	*MAT***a** *his4-912∂ lys2-128∂ trp1Δ63 cdc73Δ::KANMX*
KY2243	*MATα his4-912∂ lys2-128∂ leu2Δ1 trp1Δ63 rtf1Δ::KANMX*
KY2244	*MAT***a** *his4-912∂ lys2-128∂ ura3-52 trp1Δ63 leo1Δ::URA3*
KY2271	*MATα his4-912∂ lys2-128∂ leu2Δ1 trp1Δ63 paf1Δ::KANMX*
KY2276	*MAT***a** *leu2Δ0 ura3Δ0*
KY2845	*MAT***a** *leu2Δ0 ura3Δ0 trf4Δ::CLONAT paf1Δ::KANMX FET4::HIS3 TTS at -400*
KY2846	*MAT***a** *leu2Δ0 ura3Δ0 paf1Δ::KANMX FET4::HIS3 TTS at -400*
KY2851	*MAT***a** *his3Δ200 leu2Δ0 ura3Δ0 trf4Δ::CLONAT FET4::HIS3 TTS at -400*
KY3460	*MAT***a** *his3Δ200 lys2-128∂ leu2Δ1 ura3-52 trp1Δ63 arp8Δ::HIS3*
KY3461	*MAT***a** *his3Δ200 his4-912∂ leu2Δ1 lys2-173R2 ura3-52 ino80::HIS3*
KY3462	*MAT***a** *his3Δ200 leu2Δ1 ura3-52 trp1Δ63 paf1Δ::URA3*
KY3463	*MAT***a** *his3Δ200 lys2-128∂ leu2Δ1 ura3-52 isw1Δ::HIS3*
KY3464	*MAT***a** *his3Δ200 leu2Δ1 ura3-52 ade8 paf1Δ::URA3*
KY3465	*MAT***a** *his3Δ200 leu2Δ1 ura3-52 ade8 arp8Δ::HIS3*
KY3466	*MAT***a** *his3Δ200 leu2Δ0 ura3Δ0 FET4::HIS3 TTS at -400*

[1] All strains derived from S288C.
[2] FY strains were provided by Fred Winston.
[3] GHY strains were provided by Grant Hartzog.

2.2.2 RNA isolation

RNA was extracted by the hot phenol extraction method (Collart and Oliviero 1993). Briefly, frozen cells were suspended in 400 µL of TES extraction buffer (10 mM Tris-Cl pH 7.5, 10 mM EDTA, 0.5% SDS) and 400 µL of acid phenol, followed by incubation at 65°C for 1 hr. The aqueous phase was collected and re-extracted using acid phenol and then chloroform. Extracted RNA was combined with 40 µL of 3 M sodium acetate and 1 mL of 100% ethanol, mixed, and placed at -80°C for at least 1 hr. Precipitated RNA was collected by centrifugation and suspended in RNase-free water before quantification and quality check by agarose gel electrophoresis.

2.2.3 Northern blot analysis

For Northern blot analyses, 10µg-20µg of total RNA were separated on a gel containing 2% agarose, 6.5% formaldehyde, and 1X MOPS for 500-volt hr and then transferred to a Nytran supercharge nylon transfer membrane (Schleicher & Schuell BioScience, #10416296, Dassel, Germany) prior to hybridization with radiolabeled DNA probes. DNA probes were generated by PCR corresponding to the following genomic regions relative to the +1 nucleotide of the annotated coding sequence for the *FET4* gene: *CUT 793/794* (-479 to -114) and *FET4* (+261 to +651). Detection of *SCR1* (-181 to +284) served as a loading control. Oligonucleotides used to generate these probes are listed in Table 13. Probes were made using [α^{32}P]-dATP (single labeling) or [α^{32}P]-dATP and [α^{32}P]-dTTP (double labeling). Signals were quantified using a Typhoon FLA 7000 phosphorimager (GE, Boston, MA) and normalized to the *SCR1* internal loading control using ImageJ software.

2.2.4 Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

A total of 10 µg of RNA from each sample to be used in RT-qPCR was treated with TURBO DNase (Ambion, AM1907, Thermo Fisher, Waltham, MA) following the manufacturer's instructions. To ensure that there was no DNA contamination after the DNase treatment 1 µL of DNase-treated RNA was subjected to 40 cycles of PCR and analyzed by agarose gel electrophoresis. All samples used in RT-qPCR showed no PCR product after 40 cycles. Reverse transcription reactions were performed on 1 µg of DNase-treated RNA using the RETROScript Reverse Transcription Kit (Ambion, AM1710, Thermo Fisher, Waltham, MA) following the manufacturer's instructions.

RT-qPCR experiments were performed in technical duplicate, and all strains were tested in at least biological triplicate. Reactions were prepared in a volume of 20 µL using Maxima SYBR Green/ROX qPCR Master Mix (2X) (Thermo Fisher # K0221, Waltham, MA) following the manufacturer's instructions. Each 20 µL reaction was then divided into two 10 µL reactions, which were analyzed on a StepOne™ Real-Time PCR System (Thermo Fisher, Waltham, MA) beginning with a hold at 95°C for 10 min followed by 40 cycles of 95°C for 15 sec and 58°C for 1 min and finally terminating with the generation of a melt curve. Efficiencies were determined for all primer sets by measuring C_t values across a series of six ten-fold dilutions starting with 250 ng/µL and ending with 2.5 pg/µL. RT-qPCR data were analyzed using the mathematical formula developed by (Pfaffl 2001) and normalized to *SCR1* levels. RT-qPCR primers and their efficiencies are listed in Table 13.

Table 13. Primers for RT-qPCR and Northern probe generation used in Chapter 2

Target Gene	Sequence	Efficiency
PDA1	5'-ATTTGCCCGTCGTGTTTTGCTGTG-3'	2.02
PDA1	5'-TATGCTGAATCTCGTCTCTAGTTCTGTAGG-3'	
FIT2	5'-CTTTGACAAACGGTTCAGGTTC-3'	1.96
FIT2	5'-AGGAGGATGAGGAGGATGTAG-3'	
FIT3	5'-ACACCTGGTCTCCAAGTAGTA-3'	1.96
FIT3	5'-AGAGGATGTAGCAGAGGAAGA-3'	
SIT1	5'-ACTGTACTAGTCGTTGCAGTTC-3'	1.92
SIT1	5'-CCGAGGATTGTACCAACGATAA-3'	
ARN1	5'-GGATGTAGGTATGTGGGCTTTC-3'	1.95
ARN1	5'-CGTGCCATTCAGGAGTCTTT-3'	
PHO84	5'-CTACTGCCGTCGAATCTCTTG-3'	1.97
PHO84	5'-GAACCAGCAGTACCTAGCAAA-3'	
PHO81	5'-ACTCAACAGGTTTATGCACTCT-3'	2.03
PHO81	5'-GGCGTCCATTTATTAAACCCATC-3'	
PHO5	5'-CAGACTGTCAGTGAAGCTGAAT-3'	1.93
PHO5	5'-TGTCATCATTGGCATCGTAGTC-3'	
SCR1	5'-CTGAAGTGTCCCGGCTATAAT-3'	1.83
SCR1	5'-CTAAGGACCCAGAACTACCTTG-3'	
CUT 793/794	5'-GCGTAAATCACACAGGTGTTG-3'	
CUT 793/794	5'-CAATTAATTCATGCCGTGTGAAG-3'	
FET4	5'-GGATTTCCTGGTACGAGTGG-3'	
FET4	5'-CGTTAGATAAACGGTCGTACC-3'	
SCR1	5'-CAACTTAGCCAGGACATCCA-3'	
SCR1	5'-AGAGAGACGGATTCCTCACG-3'	

2.2.5 Affymetrix tiling array analysis

All RNA samples used in tiling array analysis were prepared using established methods (Juneau *et al.* 2007; Perocchi *et al.* 2007), and quality was assessed by agarose gel electrophoresis. RNA samples (100 µg total) were DNase treated (Fermentas #EN0521, Waltham, MA) and purified using an RNeasy kit (Qiagen #74104, Hilden Germany). RNA was reverse transcribed into cDNA in the presence of 6 ng/µL actinomycin D to prevent antisense artifacts (Perocchi *et al.* 2007), fragmented, and labeled with a 3' biotin tag. Affymetrix custom tiling arrays (A-AFFY-116 - Affymetrix Custom Array - *S. cerevisiae* Tiling Steinmetz, GEO Platform ID: GPL4563) were used to quantify gene expression (David *et al.* 2006; Huber *et al.* 2006; van Bakel *et al.* 2013). Arrays were processed and scanned by published methods (van Bakel *et al.* 2013).

2.2.6 Generation of annotation files for the *tilingArray* R package

Following the guidelines provided with the *davidTiling* Bioconductor package (David *et al.* 2006), tiling array probes were mapped to the *S. cerevisiae* genome (S288C version = R64-2-1) (Cherry *et al.* 2012; Engel *et al.* 2014). Briefly, the probe FASTA file was extracted from the array design file and used as input for MUMmer3.23 (Kurtz *et al.* 2004), along with the chromosome FASTA files for the S288C genome. Both the output of MUMmer3.23 and the S288C genome annotations were read into R (Team 2016), and a slightly modified version of the *makeProbeAnno.R* script was used to generate an up-to-date probe annotation file for use with the *tilingArray* package (Huber *et al.* 2006). All R packages used in this study can be found at www.bioconductor.org (Gentleman et al. 2004) and R scripts can be found at https://github.com/mae92/Paf1C-Transcriptome-Analysis-Code/tree/master/R_Code.

2.2.7 Variance stabilizing normalization

CEL files for 12 tiling arrays were read into R as a single expression set using the *readCel2eSet()* function of the *tilingArray* package. The probe intensities for all 12 arrays were log2 transformed and normalized (variance stabilized normalization R code available at: https://github.com/mae92/Paf1C-Transcriptome-Analysis-Code/tree/master/R_Code) to minimize batch effects using the *vsn* package (Huber *et al.* 2002). This normalization method, as opposed to spike-in methods, assumes that the expression of most genomic regions will be equivalent between experiments, and therefore, the extent of expression changes overall in the *paf1*Δ strains may be underestimated.

2.2.8 Mapping probe intensities to probe positions across the *S. cerevisiae* genome

The expression set containing the normalized log2 transformed probe intensities was used as input for the *segChrom()* function of the *tilingArray* package, and the locations of probes across the genome were extracted for use in downstream analysis using basic R commands (R code can be found at: https://github.com/mae92/Paf1C-Transcriptome-Analysis-Code/tree/master/R_Code). Probe locations were averaged for triplicate samples, and these averaged values were used to generate BedGraph files, which were converted into bigWig files for visualization in the Integrative Genomics Viewer (IGV) from the Broad Institute (Thorvaldsdottir *et al.* 2013).

2.2.9 Annotation guided differential expression analysis

Normalized log2 transformed probe intensity values were extracted from the *tilingArray* output. Using a custom file (available at: https://github.com/mae92/Paf1C-Transcriptome-Analysis-Code/blob/master/Transcript_Annotations/combined.fix.csv) containing transcript annotations (listed in Table 14) from the Saccharomyces Genome Database (SGD) and recent studies of novel noncoding RNA transcripts (Cherry *et al.* 1998; Xu *et al.* 2009; Yassour *et al.* 2010; van Dijk *et al.* 2011; Schulz *et al.* 2013; Venkatesh *et al.* 2016), we calculated the average log2 intensity values for probes spanning each annotation. This process was carried out using an in-house Python script (available at: https://github.com/mae92/Paf1C-Transcriptome-Analysis-Code/tree/master/Python_Code) that calculates the average intensity of all probes occupying a given annotation found in the annotation file.

Average log2 intensity values for all transcripts in each replicate and strain background were loaded into the *limma* package, where a linear model was used to determine statistical significance. Log2 fold change and *p* values for all transcripts tested were extracted from the output for each strain comparison. These *p* values were adjusted using Benjamini and Hochberg's false discovery rate (FDR) (Benjamini and Hochberg 1995) method using *top.table* command in *limma* (R code for *limma* analysis can be found here: https://github.com/mae92/Paf1C-Transcriptome-Analysis-Code/tree/master/R_Code) (Ritchie *et al.* 2015). Significantly differentially expressed genes (adjusted *p* value < 0.05) that were present in both comparisons (*paf1*Δ vs. WT and *paf1*Δ *trf4*Δ vs. *trf4*Δ*)* were loaded into SGD's YeastMine database (Balakrishnan *et al.* 2012), where additional annotation and gene ontology information could be extracted. Gene ontology results are shown in Tables 19 and 20. Plots of differential expression data were produced in R.

Table 14. Transcript counts

Transcript	Count	Source
mRNA	6600	Cherry et al., 1998 (SGD)
snoRNA	77	Cherry et al., 1998 (SGD)
CUT	925	Xu et al., 2009
SUT	847	Xu et al., 2009
NUT	1526	Schulz et al., 2013
XUT	1657	van Dijk et al., 2011
SRAT	532	Venkatesh et al., 2016
Sum	12164	

A **Differentially expressed transcripts identified by the *de novo* differential expression analysis were defined by a six step process:**

1. Average log2(probe intensity) was calculated for three biological replicates.

2. The data were smoothed by averaging across a sliding window of 20 tiling array probes (roughly 160bp).

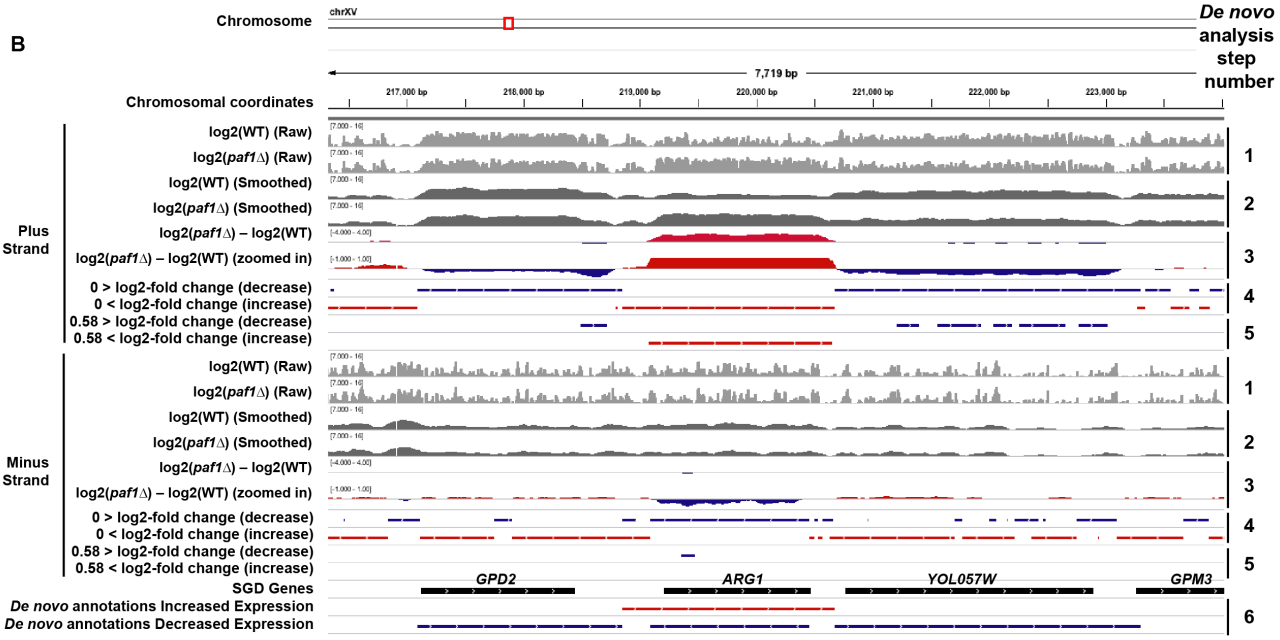
3. The log2 fold change (experimental vs. control) was calculated across the entire genome.

4. All regions with an absolute fold change greater than one (log2 fold change of 0) were identified.

5. Regions of the genome with an absolute change of 1.5-fold (log2 fold change of 0.58) were identified and any of these regions less then 80 bp long (a length comparable to the shortest snoRNA) were excluded.

6. The two lists of regions from steps 4 and 5 were intersected to yield a list of extended differentially expressed regions where some portion of the transcript had an absolute fold change of 1.5-fold or greater.

B



C

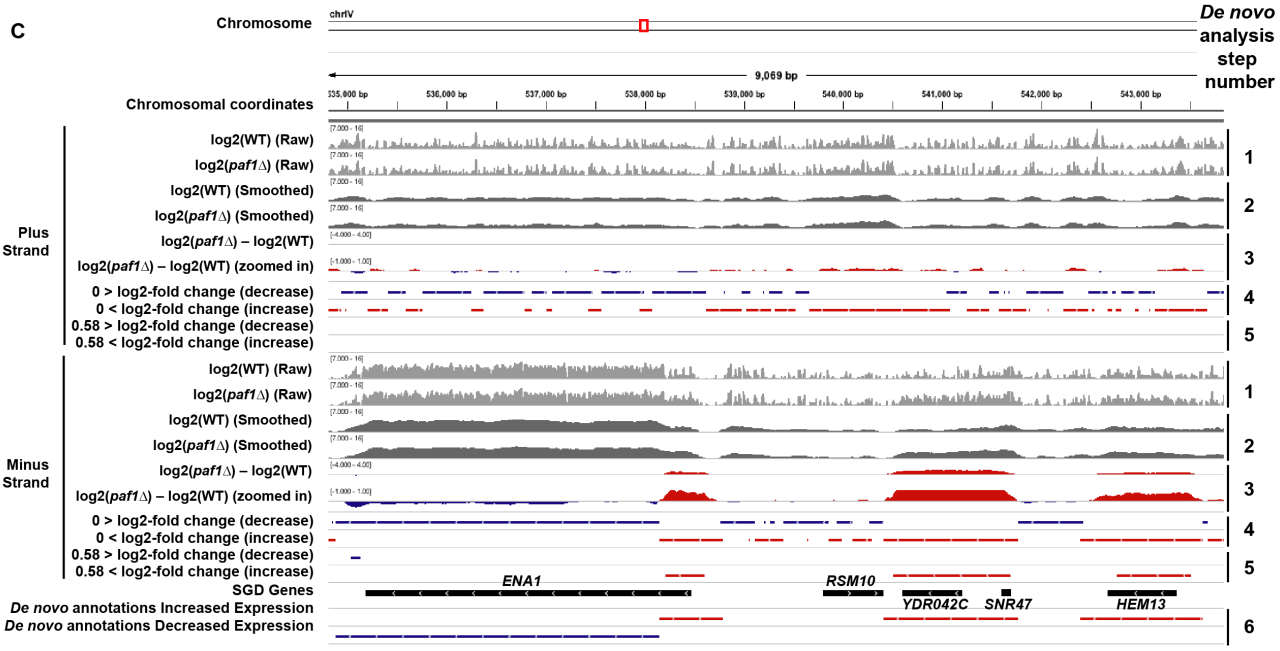


Figure 6. Examples of the steps taken in the *de novo* differential expression analysis.

(A) List of steps taken to identify differentially expressed transcripts using the *de novo* analysis. (B) Genome browser tracks generated using IGV showing step by step how differentially expressed transcripts are identified by the *de novo* analysis. Browser tracks show tiling array data from the *paf1*Δ and WT datasets. Numbers corresponding to steps listed in A are shown on the right. (C) Same as A at a different genomic location.

2.2.10 *De novo* differential expression analysis

BedGraph files were created containing normalized log2 probe intensity values (averaged for the three biological replicate arrays) mapped to the yeast genome. Differentially expressed transcripts identified by the *de novo* differential expression analysis were defined by a six-step process (Figure 6). 1) Average log2(probe intensity) was calculated for three biological replicates. 2) The data were smoothed by averaging across a sliding window of 20 tiling array probes (roughly 160bp). 3) The log2 fold change (experimental vs. control) was calculated across the entire genome. 4) All regions where the absolute value of the observed fold change was greater than one (log2 fold change of 0) were identified. 5) Regions of the genome with an absolute change of 1.5-fold (log2 fold change of 0.58) were identified, and any of these regions less than 80 bp long (a length comparable to the shortest snoRNA) were excluded. 6) The two lists of regions from steps 4 and 5 were intersected to yield a list of extended differentially expressed regions where some portion of the transcript had an absolute fold change of 1.5-fold or greater. The transcripts defined by this method were then treated as their own list of transcript annotations for use in the comparisons described herein. This was done using a combination of AWK (Aho *et al.* 1979) and the BedTools suite (Quinlan and Hall 2010; Quinlan 2014). The shell script used to define differentially expressed transcripts can be found at https://github.com/mae92/Paf1C-

Transcriptome-Analysis-Code/tree/master/Shell_Code and R code used to generate the input BedGraph files can be found at https://github.com/mae92/Paf1C-Transcriptome-Analysis-Code/tree/master/R_Code.

2.2.11 Analysis of published datasets

Next-generation sequencing datasets from previous studies (Churchman and Weissman 2011; Van Oss *et al.* 2016; Harlen and Churchman 2017b) were obtained in BedGraph format directly from the authors or FASTQ format from the NCBI SRA database and converted into BigWig format for use with DeepTools (Ramírez *et al.* 2014, 2016). Files received in BedGraph format were converted using the University of California Santa Cruz (UCSC) Genome Browser (Kent *et al.* 1976) utility *BedGraphToBigWig*. Files downloaded from the SRA database in FASTQ format were mapped to the *S. cerevisiae* genome (S288C version = R64-2-1) (Cherry *et al.* 2012; Engel *et al.* 2014) using HISAT2 (Kim *et al.* 2015) and converted to BAM format using Samtools (Li *et al.* 2009). BAM files were converted to Wig format using the bam2wig utility (found at https://github.com/MikeAxtell/bam2wig) and converted to BigWig format using the UCSC utility *WigToBigWig*. Heatmaps were plotted using *computeMatrix* and *plotHeatmap* tools in the deepTools package by summing the tag counts using 50bp bins.

2.2.12 Statistical analysis

At least three biological replicates were performed for every assay shown in this manuscript, including tiling arrays. Each biological replicate is a pure yeast culture derived from a single colony initiated from a single cell of a given strain. Tiling array data analyzed using the

limma package were subjected to the standard *limma* workflow, which utilizes linear modeling and an empirical Bayes method to determine differentially expressed genes from as little as three biological replicates. The *limma p* values were adjusted for multiple comparisons using Benjamini and Hochberg's false discovery rate (FDR) (Benjamini and Hochberg 1995). All RT-qPCR and Northern blot *p* values were generated using an unpaired, two-sided, students t-test assuming equal variance carried out between the mutant strain and the wild type strain.

2.2.13 Data availability

Tiling array data (raw CEL files, BedGraph files, annotation-guided differential expression results, and files containing annotations for differentially expressed transcripts defined by our *de novo* analysis in BED6 file format) have been deposited in the Gene Expression Omnibus database under accession number GSE122704. The code used for the analysis of tiling array data has been uploaded to the following GitHub repository (https://github.com/mae92/Paf1C-Transcriptome-Analysis-Code).

2.3 Results

2.3.1 Deletion of *PAF1* affects coding and non-coding transcripts genome-wide

To investigate the impact of Paf1C on the *S. cerevisiae* transcriptome, we used high-resolution whole-genome tiling arrays to measure steady-state RNA levels in *S. cerevisiae* strains deleted for the *PAF1* gene, which encodes a core member of Paf1C important for complex integrity (Mueller *et al.* 2004; Deng *et al.* 2018). Additionally, to assess the Paf1-dependency of unstable noncoding RNAs (ncRNAs) in these experiments, we deleted *TRF4* in both *PAF1* and *paf1*Δ strains. When compared to a *trf4Δ* strain, the *paf1Δ trf4Δ* double mutant revealed wide-ranging effects on all Pol II transcript classes examined: mRNAs, small nuclear RNAs (snRNAs), snoRNAs, CUTs, stable unannotated transcripts (SUTs; Xu *et al.* 2009), Xrn1-dependent unstable transcripts (XUTs; van Dijk *et al.* 2011), Nrd1-unterminated transcripts (NUTs; Schulz *et al.* 2013), and Set2-repressed antisense transcripts (SRATs; Venkatesh *et al.* 2016) (Figure 7A-G; Table 15).

Table 15. Summary of differential expression results obtained from the annotation-guided analysis of the tiling array data.

Counts of RNAs with an absolute fold change of 1.5 or greater in the annotation-guided analysis of the tiling array data. The analysis was performed using the limma Bioconductor package in R. These data are graphically presented in Figure 7. Percentages of transcripts within each class that show an expression change greater than 1.5-fold are also shown. The differentially expressed mRNAs detected in the *paf1*Δ *trf4*Δ vs. *trf4*Δ comparison were used in the comparison shown in Figure 6A.

	*paf1*Δ *vs. WT*					
	Increase		Decrease		Sum	
Transcript Class	Count	Class Percent	Count	Class Percent	Count	Class Percent
mRNA	126	1.9	348	5.3	474	7.2
snoRNA	12	15.6	1	1.3	13	16.9
CUT	8	0.9	2	0.2	10	1.1
SUT	4	0.5	3	0.4	7	0.9
NUT	9	0.6	5	0.3	14	0.9
XUT	10	0.6	5	0.3	15	0.9
SRAT	2	0.4	0	0.0	2	0.4
Sum	171		364		535	

	*paf1*Δ *trf4*Δ *vs. trf4*Δ					
	Increase		Decrease		Sum	
Transcript Class	Count	Class Percent	Count	Class Percent	Count	Class Percent
mRNA	281	4.3	320	4.8	601	9.1
snoRNA	21	27.3	2	2.6	23	29.9
CUT	14	1.5	86	9.3	100	10.8
SUT	5	0.6	31	3.7	36	4.3
NUT	16	1.0	61	4.0	77	5
XUT	30	1.8	54	3.3	84	5.1
SRAT	21	3.9	5	0.9	26	4.8
Sum	388		559		947	

Levels of many snRNAs, snoRNAs, and SRATs increased in the *paf1Δ trf4Δ* double mutant relative to the *trf4Δ* single mutant (Figures 7D and 7G; Table 15) indicating that, in wild type cells, Paf1 suppresses their transcription or destabilizes the transcripts. In the case of SRATs, increased transcript levels are consistent with a requirement for Paf1 in facilitating H3 K36me3, a modification that negatively regulates transcription (Churchman and Weissman 2011; Kim *et al.* 2016; Venkatesh *et al.* 2016) by activating the Rpd3S histone deacetylase complex and by inhibiting histone exchange (Carrozza *et al.* 2005; Joshi and Struhl 2005; Keogh *et al.* 2005; Govind *et al.* 2010; Venkatesh *et al.* 2012). Levels of many CUTs, SUTs, XUTs, and NUTs decreased upon deletion of *PAF1* (Figures 7B, C, E, and F; Table 15). For NUTs and CUTs, these changes in transcript abundance suggest that Paf1 impacts NNS-dependent termination beyond the snoRNA genes. At protein-coding genes, Paf1 positively and negatively affects mRNA levels in a locus-specific manner (Figure 7A; Table 15), in agreement with previous studies (Shi *et al.* 1996; Porter *et al.* 2005; Cao *et al.* 2015; Chen *et al.* 2015; Yu *et al.* 2015a; Yang *et al.* 2016; Fischl *et al.* 2017).

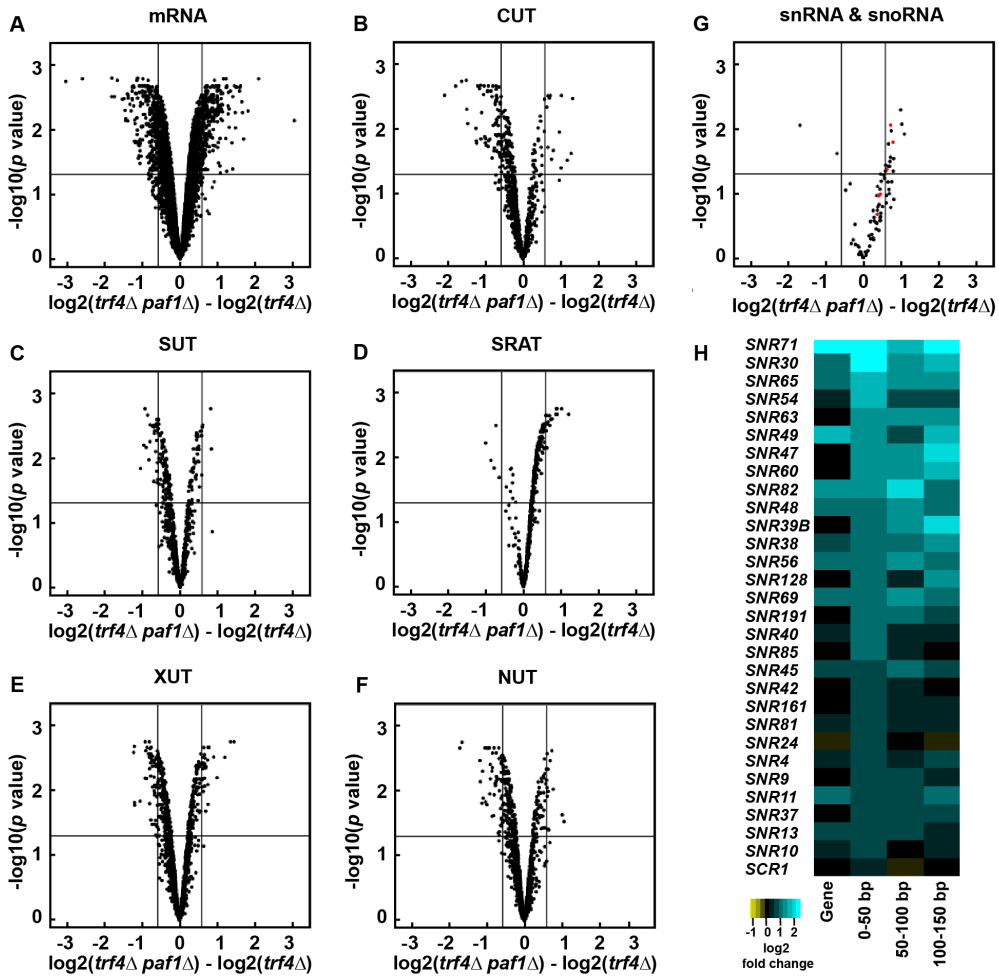


Figure 7. Deletion of PAF1 affects all Pol II transcript classes.

(A-G) Volcano plots graphing statistical significance (y-axis) against expression change (x-axis) between *paf1Δ trf4Δ* and *trf4Δ* strains (KY2016 and KY2012, respectively) for the indicated Pol II transcript classes. In panel G, snRNAs and snoRNAs are shown in red and black, respectively. Each point represents an individual transcript. Tiling array probe intensities were averaged over annotated regions using a custom Python script, and an average log2 fold change and p value were calculated using the limma R package. The horizontal line indicates an FDR adjusted p value of 0.05, and the vertical lines indicate a 1.5-fold change in expression (log2 fold change of 0.58). Counts and percentages of differentially expressed transcripts shown here are listed in Table 15. (H) Heatmap of log2 fold change in expression between *paf1Δ* and WT strains (KY1702 and KY2276, respectively) for the 29 most-affected snoRNA genes. The snoRNA gene bodies and regions 0-50 bp,

50-100 bp, and 100-150 bp downstream of their annotated 3' ends are plotted and sorted by the 0-50 bp region.

For many snoRNA genes, we detected an increase in RNA levels downstream of the annotated gene in the *paf1Δ* and *paf1*Δ *trf4*Δ strains relative to wild type and *trf4*Δ strains, respectively (Figure 7H and Figure 8). This observation is consistent with previous studies showing Paf1 is required for efficient termination at many snoRNA genes (Sheldon *et al.* 2005; Tomson *et al.* 2013a). Note that the log2 fold change values calculated for any particular snoRNA gene and its downstream region do not always agree. In many cases, RNA levels mapping to the gene body do not change expression even when downstream changes are observed, suggesting that read-through transcription is occurring at these loci (Figure 7H). Interestingly, through k-means clustering analysis, we also identified snoRNA genes that are relatively unaffected by the deletion of *PAF1* (Figure 8). Further studies will be needed to elucidate the mechanistic distinctions between the Paf1-dependent and Paf1-independent snoRNA genes.

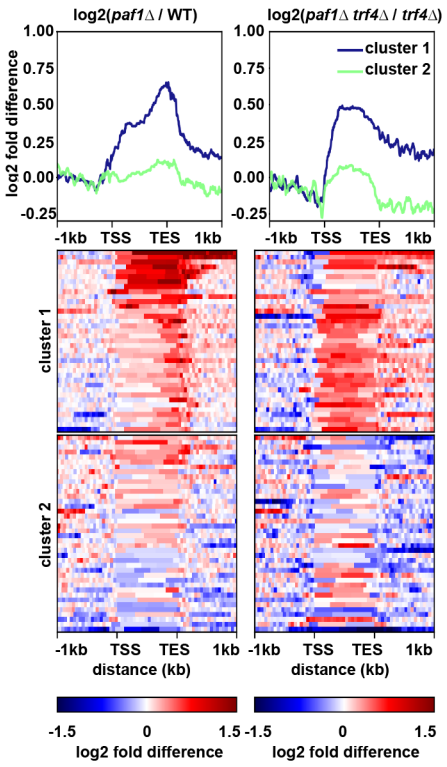


Figure 8. Log2 fold change at snoRNA genes.

Heatmap of log2 fold difference in expression between paf1Δ and WT strains (KY1702 and KY2276, respectively) and *paf1Δ trf4Δ* and *trf4Δ* strains (KY2016 and KY2012, respectively) for the snoRNA genes. The snoRNA transcripts were scaled so that each row in the heatmap represents a single transcript from the transcription start site (TSS) to transcription end site (TES) plus the 1kb region upstream and downstream of the annotated snoRNA. Clusters of snoRNAs were determined by k-means clustering.

2.3.2 *De novo* differential expression analysis reveals effects of Paf1 on antisense transcripts

As an independent analysis and to facilitate the detection of unannotated transcripts, we performed a *de novo* differential expression analysis of our tiling array data. Here, we relied on the data to reveal the boundaries of differentially expressed transcripts instead of using predetermined annotations. Strains lacking *PAF1* were compared to control strains on a probe-by-probe basis. Genomic regions with a 1.5-fold or greater difference in expression between *paf1*Δ and *PAF1* strains were selected as differentially expressed and extended until the expression difference was no longer observed (see Figure 6 and Materials and Methods for a detailed description of this analysis).

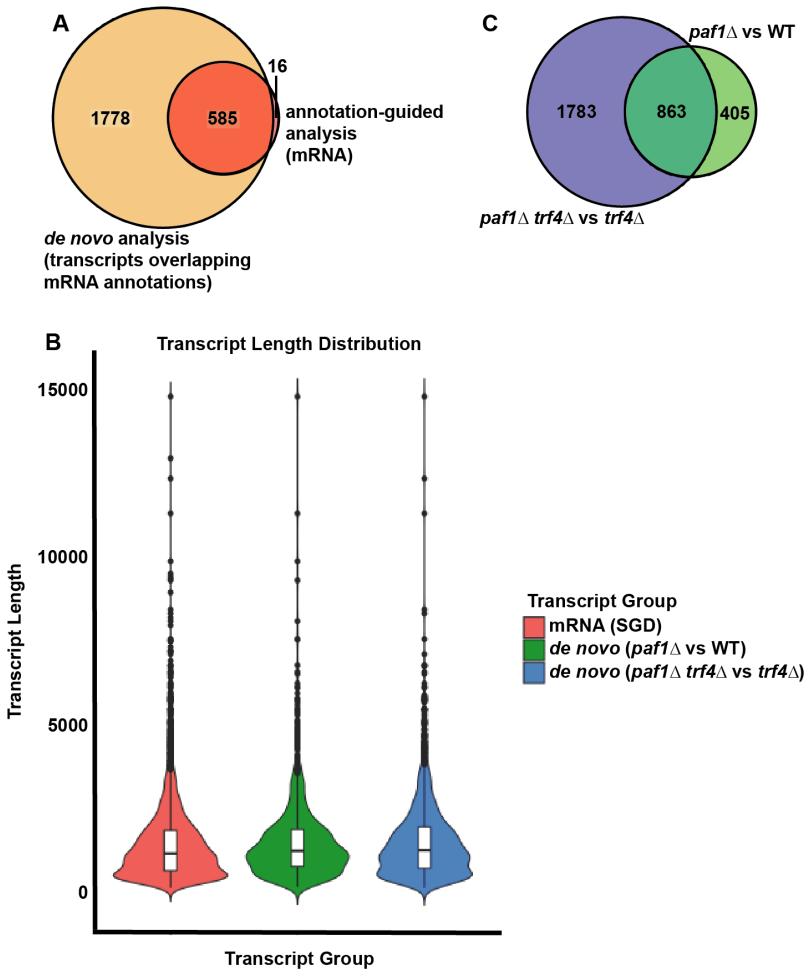


Figure 9. Comparison of the annotation-guided and *de novo* differential expression analyses. **(A) Venn diagram comparing differentially expressed transcripts that overlap with mRNA annotations in the *paf1*Δ *trf4*Δ strain (KY2016) identified by *de novo* analysis (1.5-fold cutoff and length greater than 80bp) and differentially expressed mRNAs identified in our annotation-guided analysis (1.5-fold cutoff). When more than one differentially expressed transcript, as identified by *de novo* analysis, overlapped with the same mRNA, the overlap was only counted once in the intersecting region of the Venn diagram. (B) Violin plots showing the distribution of transcript lengths for mRNA annotations in SGD and the *de novo* annotations from this study. (C) Venn diagram showing overlap between all differentially expressed transcripts identified by *de novo* analysis of *paf1*Δ and *paf1*Δ *trf4*Δ strains (KY1702 and KY2016, respectively).**

Confirming the accuracy of the *de novo* analysis, we found that nearly all the mRNAs identified as differentially expressed in the *paf1*Δ *trf4*Δ strain by our annotation-guided analysis (585 mRNAs;1.5-fold or greater expression change relative to *trf4*Δ strain) were also detected by the *de novo* analysis (Figure 9A). We note that, compared to the annotation-guided analysis, a larger number of differentially expressed transcripts that overlap mRNAs on the sense strand were detected in the *de novo* analysis. This observation is not due to technical differences, but rather a consequence of multiple *de novo* transcripts overlapping with a single annotated mRNA and the increased sensitivity of the analysis. Additionally, we found that the length distribution of all transcripts identified in the *de novo* analysis was similar to that reported in SGD for mRNAs (Figure 9B), confirming that we were not calling exceedingly long or short transcripts. Further, separation of the *de novo* analysis data by transcript class revealed the effects of *paf1*Δ on noncoding and coding RNAs similar to those observed in the annotation-guided analysis (compare Figure 10A to Figure 7A-7G; Table 16).

In the *paf1*Δ *trf4*Δ strain, for example, SRATs and snoRNAs were predominantly up-regulated, and other noncoding RNAs were predominantly down-regulated. When viewed as a whole, the *de novo* analysis detected far more differentially expressed transcripts in the *paf1*Δ *trf4*Δ strain (relative to the *trf4*Δ control strain) than in the *paf1*Δ strain (relative to the *PAF1* control strain) (Figure 9C). Therefore, a functional TRAMP complex, which promotes the processing and degradation of unstable transcripts, obscures many of the transcriptional effects of deleting *PAF1*.

Table 16. Overlap between transcripts identified by *de novo* analysis and annotated transcripts.

Counts and corresponding percentages of transcripts identified by the *de novo* analysis that overlap with existing transcript annotations. Transcripts exhibiting an absolute fold change of 1.5-fold or greater in some portion of the differentially expressed region were counted (see Materials and Methods for a detailed

description). The percentages shown in this table were used to generate Figure 10A, and the counts of mRNA overlaps were used in the comparison shown in Figure 9A. Note that any overlap between a transcript identified in the *de novo* analysis and an annotated RNA was counted. This analysis calculates overlap from the perspective of the previously annotated RNAs. Therefore, if a single *de novo* transcript overlaps with two annotated mRNAs, each mRNA is counted, resulting in a total count of two. This explains the higher number of differentially affected mRNAs in this table compared to Table 17. In Table 17, we calculate from the perspective of the *de novo* transcript, so a *de novo* transcript overlapping with two annotated mRNAs would only be assigned a value of one. It is also important to note that if an overlap is detected between a *de novo* transcript and two annotated transcripts, such as a CUT and a NUT, for example, this will result in the assignment of that *de novo* transcript to both a NUT and a CUT. We chose to allow this because there is overlap between and within some of the annotated RNA classes, and we would consequentially lose most annotated RNAs if we excluded any with overlap within or between classes.

	Counts		Percent of Class	
	paf1Δ vs. WT		*paf1Δ vs. WT*	
	Decreased	Increased	Decreased	Increased
mRNA	1100	359	20.32	6.15
snoRNA	1	28	1.30	27.27
CUT	72	34	2.38	1.41
SUT	62	30	4.14	1.89
XUT	77	57	2.23	1.81
SRAT	11	27	0.19	5.64
NUT	112	101	3.08	5.18

	Counts		Percent of Class	
	paf1Δ trf4Δ vs. trf4Δ		*paf1Δ trf4Δ vs. trf4Δ*	
	Decreased	Increased	Decreased	Increased
mRNA	1519	844	21.47	16.24
snoRNA	17	23	3.90	24.68
CUT	294	70	22.19	4.11
SUT	193	85	19.86	6.86
XUT	372	172	18.53	7.91
SRAT	45	147	5.83	33.46
NUT	426	229	31.72	11.27

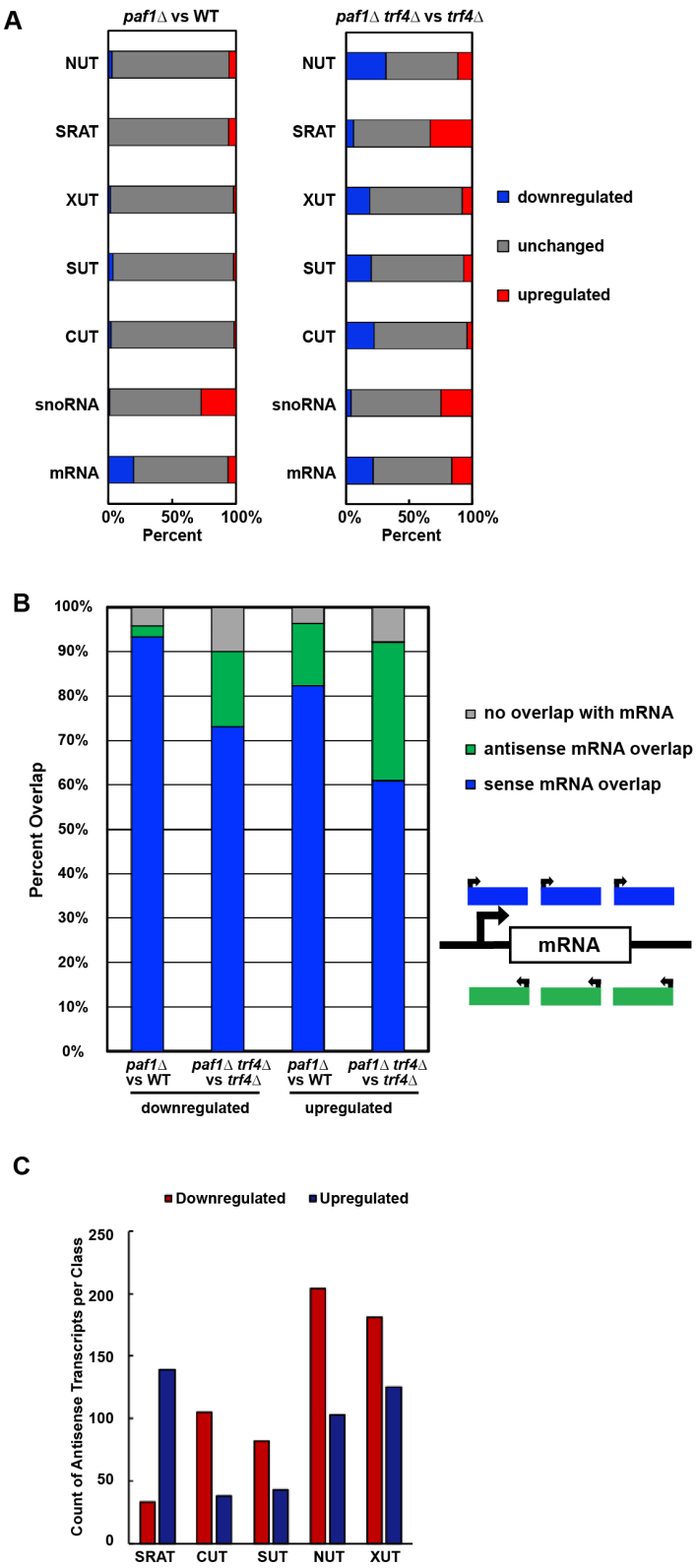


Figure 10. Paf1 positively and negatively regulates antisense transcription.

(A) Horizontally stacked bar graphs showing the percentage of each transcript class (listed in Table 14) found to overlap with a differentially expressed transcript identified in *paf1*Δ or *paf1*Δ *trf4*Δ strains by *de novo* analysis (counts and percentages listed in Table 16). (B) Vertically stacked bar graph plotting percentage of transcripts, identified in the *de novo* analysis, that overlap with mRNA coding regions on the sense or antisense strand. These data are also presented in Table 17. (C) Bar graph summarizing the overlap between differentially expressed antisense transcripts detected by the *de novo* analysis and previously annotated noncoding RNAs (see sums in Table 18 for counts).

Unstable ncRNAs are often found near mRNA loci in tandem or antisense orientations (Neil *et al.* 2009; Xu *et al.* 2009; van Dijk *et al.* 2011; Schulz *et al.* 2013; Castelnuovo *et al.* 2014). Murray *et al.* (2015) demonstrated that regions of high antisense transcription are deficient in H2BK123ub, H3K4me3, H3K79me3, and H3K36me3, while regions of low antisense transcription are enriched for H3K79me2. Levels of all these histone modifications are affected by *PAF1* deletion. Therefore, the deletion of *PAF1* and loss of Paf1C-dependent histone modifications may generate a chromatin landscape that promotes antisense transcription at some loci and represses it at others. Interestingly, in the *de novo* analysis, we observed enrichment of many transcripts oriented antisense to mRNA loci in the *paf1*Δ *trf4*Δ strain, relative to the *trf4*Δ strain, and found that Paf1 both positively and negatively regulates antisense transcript levels in *S. cerevisiae* (Figure 10B; Table 17).

Deeper analysis revealed that many of the antisense transcripts detected by the *de novo* analysis overlapped with previously annotated noncoding transcripts (Figure 10C), consistent with earlier studies showing that many noncoding transcripts are oriented antisense to genes (Xu *et al.* 2009; van Dijk *et al.* 2011; Schulz *et al.* 2013; Venkatesh *et al.* 2016). This suggests that a large portion of the ncRNA differential expression profile observed in the *paf1*Δ *trf4*Δ strain results from antisense transcription.

Table 17. Counts of transcripts identified by the *de novo* analysis that fall into various categories based on position relative to mRNA coding regions.

Counts of transcripts identified by the *de novo* analysis with an absolute fold change of 1.5-fold or more (calculated as described in Table 15) that overlap with mRNAs either in the sense or antisense direction. The counts shown in this table were used in the generation of Figure 10B. The sum of all transcripts identified in each strain background is shown in the bottom row representing the total number of transcripts identified in the *de novo* analysis (used in Figure 9C Venn diagram). Note that the mRNA overlap shown here does not match with Table 16. This analysis calculates overlap from the perspective of the *de novo* transcripts, meaning that if a transcript overlaps with more than one mRNA, it is only counted once in this analysis. This leads to a lower number of mRNA overlaps being counted in Table 17 than in Table 16, because some *de novo* annotations encompass more than one mRNA (see Table 16 legend for more detailed explanation).

	*paf1*Δ *vs. WT*		*paf1*Δ *trf4*Δ *vs. trf4*Δ	
	Decreased	Increased	Decreased	Increased
sense mRNA Overlap	872	275	1111	687
antisense mRNA overlap	23	47	256	353
no overlap with mRNA	39	12	151	88
Sum	1268		2646	

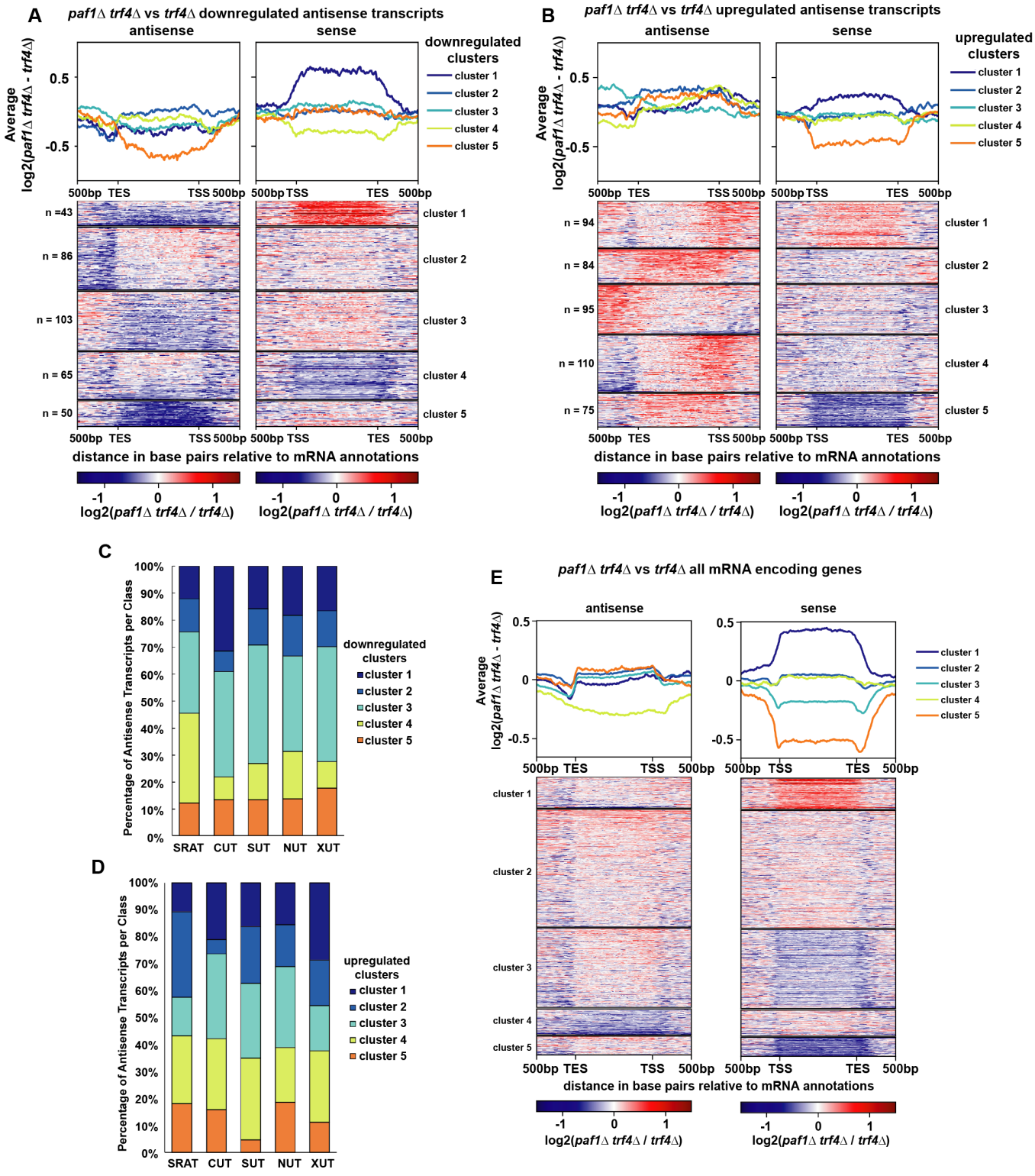


Figure 11. Analysis of antisense transcription in the *paf1*Δ *trf4*Δ mutant.

(A and B) Heatmaps and average gene profiles for clusters generated by k-means clustering of the sense and antisense strands of protein-coding regions that experience changes in antisense transcription in the *paf1*Δ *trf4*Δ vs. *trf4*Δ shown in Figure 10B. Protein-coding genes with decreased and increased antisense transcription are shown in panels A and B, respectively. (C and D) Vertically stacked bar graphs showing the percentage of regions antisense to mRNAs overlapping with various noncoding transcript classes. Clusters were taken from the analyses in A and B. (E) Heatmaps and average gene profiles of tiling array data (log2(*paf1*Δ *trf4*Δ) – log2(*trf4*Δ)) on the sense and antisense strand of all protein-coding genes. These data are scaled over the gene body, and an additional 500 bp upstream and downstream are shown. These data were separated into clusters using k-means clustering.

To investigate the antisense transcriptional landscape further, we plotted sense and antisense transcript levels relative to the transcription start site (TSS) and transcription end site (TES) of protein-coding genes at which we detected an absolute change of 1.5-fold or greater in antisense transcription overlapping the gene (Figure 11A and Figure 11B). For both the antisense-downregulated class and the antisense-upregulated class, k-means clustering analysis revealed five clusters differing in the patterns and levels of antisense transcription relative to sense transcription. (Note that when these clusters are broken down by overlap with various ncRNA classes, no one cluster is dominated by an individual ncRNA class (Figure 11C, 11D, and Table 18). A small number of protein-coding genes show an apparent anti-correlation between sense and antisense transcription in the *paf1*Δ *trf4*Δ mutant (cluster 1 in Figure 11A and cluster 5 in Figure 11B). However, for most genes experiencing a 1.5-fold or greater increase in antisense transcription, a clear relationship between antisense and sense transcript levels was not detected. This result agrees with previous work on antisense transcription (Murray *et al.* 2015). Further, a plot of sense and antisense transcript levels for all protein-coding genes suggests that antisense transcription is not governing the changes we detect in sense transcription for most genes (Figure 11E).

Table 18. Counts of antisense transcripts overlapping with various transcript classes.

Data shown here are graphed in Figure 9C and 9D. Note that the total count of antisense regions captured when calculating the overlap between antisense transcripts and previously published annotations is larger than the number of antisense transcripts used in the analysis. This is due to a single antisense transcript overlapping with more than one previously annotated transcript.

Input: Downregulated Antisense Transcripts (n=256)							
	SRAT	CUT	SUT	NUT	XUT	snoRNA	Count
cluster 1	4	14	11	28	32	0	43
cluster 2	11	9	11	36	18	0	86
cluster 3	10	41	36	72	77	0	103
cluster 4	4	8	11	31	24	0	65
cluster 5	4	33	13	37	30	0	50
sum	33	105	82	204	181	0	347

Input: Upregulated Antisense Transcripts (n=353)							
	SRAT	CUT	SUT	NUT	XUT	snoRNA	Count
cluster 1	25	6	2	19	14	0	94
cluster 2	35	10	13	21	33	0	84
cluster 3	20	12	12	31	21	0	95
cluster 4	44	2	9	16	21	0	110
cluster 5	15	8	7	16	36	0	75
sum	139	38	43	103	125	0	458

2.3.3 Paf1 regulates transcript abundance at the transcriptional level

Paf1C has been shown to regulate both transcriptional and post-transcriptional processes, such as mRNA export, at protein-coding genes (Porter *et al.* 2005; Van Oss *et al.* 2016; Fischl *et al.* 2017). To determine if changes in RNA levels detected by our tiling array analysis occurred at the transcriptional level, we compared our results to published NET-seq data (Harlen and Churchman 2017b). Tiling array data comparing *paf1*Δ and *PAF1* strains or *paf1*Δ *trf4*Δ and *trf4*Δ strains were used to generate heatmaps for comparison to *paf1*Δ NET-seq data (Figure 12A). Overall, we observed similarity between the *paf1*Δ *trf4*Δ tiling array data and the *paf1*Δ NET-seq data, indicating that Paf1 is regulating the abundance of many transcripts, including unstable noncoding RNAs and mRNAs, through an effect on transcription (Figure 12A and 12B). However, our analysis also indicates that at some genes, Paf1 is likely playing a post-transcriptional role. For example, for the majority of Paf1-stimulated protein-coding genes (73%), a decrease in steady-state RNA levels in *paf1*Δ cells was reflected in reduced nascent transcript levels (Figure 12B). In contrast, a smaller fraction of Paf1-repressed genes (52%) showed a corresponding increase in the NET-seq signal in the *paf1*Δ background (Figure 12B). Therefore, both positive and negative effects of Paf1 occur at the transcriptional level at many loci, but for protein-coding genes repressed by Paf1, a larger fraction appears to be post-transcriptionally regulated.

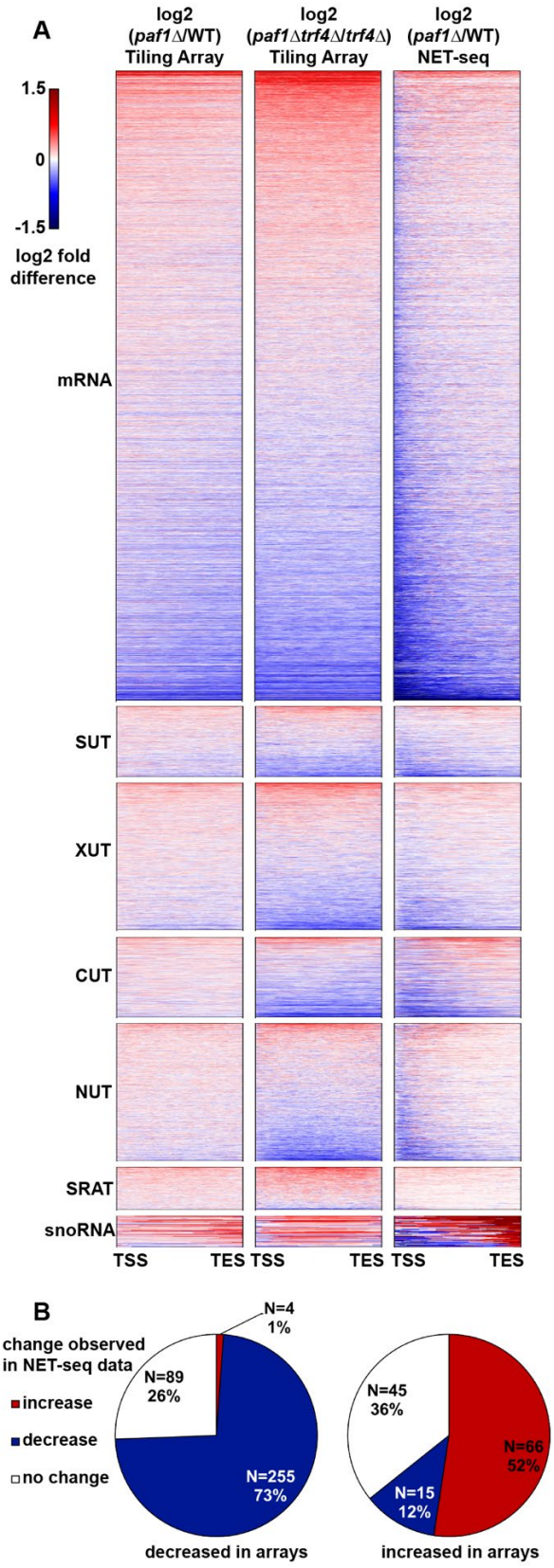

A
log2 (paf1Δ/WT) Tiling Array
log2 (paf1Δtrf4Δ/trf4Δ) Tiling Array
log2 (paf1Δ/WT) NET-seq
1.5
0
-1.5
log2 fold difference
mRNA
SUT
XUT
CUT
NUT
SRAT
snoRNA
TSS
TES
TSS
TES
TSS
TES
B
change observed in NET-seq data
increase
decrease
no change
N=4
1%
N=89
26%
N=255
73%
decreased in arrays
N=45
36%
N=66
52%
N=15
12%
increased in arrays

Figure 12. Paf1 regulates many of its target loci at the transcriptional level

(A) Heatmaps are plotting log2 fold-change in transcript levels detected by tiling array for *paf1*Δ vs. WT (KY1702 vs. KY2276) and *paf1*Δ *trf4*Δ vs. *trf4*Δ (KY2016 vs. KY2012) as well as a *paf1*Δ vs. WT NET-seq comparative analysis (Harlen and Churchman 2017b). Previously annotated coding and non-coding transcripts were scaled so that each row in the heatmap represents a single transcript from transcription start site (TSS) to transcription end site (TES). (B) Pie charts showing the direction of change in NET-seq data (Harlen and Churchman, 2017) for mRNAs that increased or decreased expression by at least 1.5-fold in the *paf1*Δ vs. WT comparison as measured by tiling array. The direction of change in NET-seq was determined by summing the reads in the first 500bp of protein-coding genes in both WT and *paf1*Δ NET-seq datasets and calculating a fold-change (1.5-fold cutoff).

One possible difference between Paf1-stimulated and Paf1-repressed genes is related to their level of expression in wild type cells. To investigate this possibility, we used ChIP-exo data from Van Oss et al. (2016) to analyze Pol II occupancy (Rpb3 subunit) at protein-coding genes with absolute expression changes of 1.5-fold or greater in a *paf1*Δ background as measured by our tiling array analysis. The Rpb3 ChIP-exo data indicate that, in general, Paf1-stimulated genes are more highly transcribed than Paf1-repressed genes (Figure 13A and 13B).

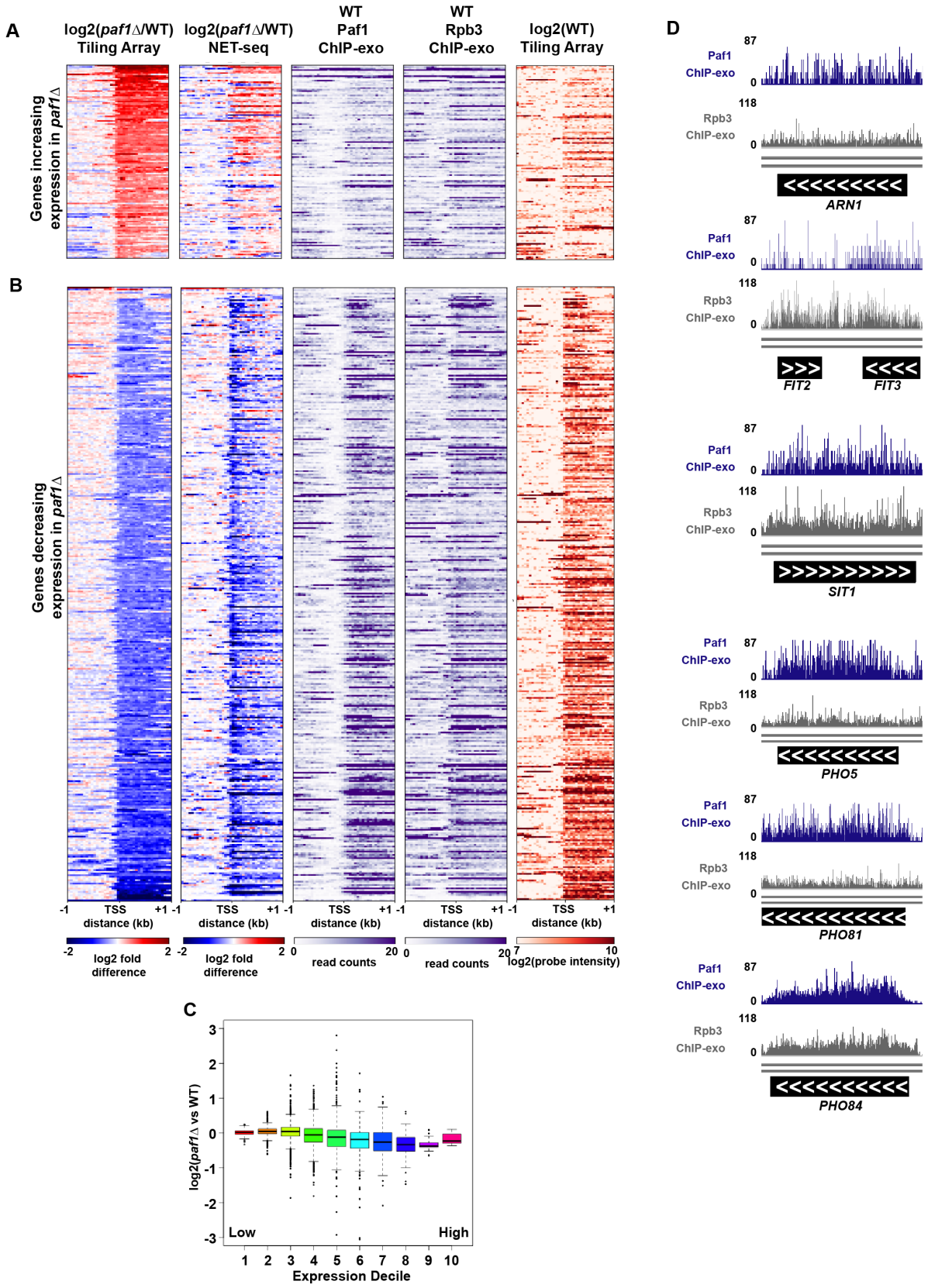

A
log2(paf1Δ/WT) Tiling Array
log2(paf1Δ/WT) NET-seq
WT Paf1 ChIP-exo
WT Rpb3 ChIP-exo
log2(WT) Tiling Array
Genes increasing expression in paf1Δ
B
Genes decreasing expression in paf1Δ
-1 TSS +1 distance (kb)
-2 log2 fold difference 2
0 read counts 20
7 log2(probe intensity) 10
C
log2(paf1Δ vs WT)
Low
High
Expression Decile
D
Paf1 ChIP-exo
Rpb3 ChIP-exo
87
118
0
ARN1
FIT2
FIT3
SIT1
PHO5
PHO81
PHO84

Figure 13. Differentially regulated protein-coding genes shown by tiling array, NET-seq, and ChIP-exo across three different studies.

(A) Heatmaps of genes that increased expression by 1.5-fold or more in the *paf1*Δ strain relative to WT. (B) Heatmaps of genes that decreased expression by at least 1.5-fold in the *paf1*Δ strain relative to WT. Gene lists were determined by selecting genes that decreased or increased expression by at least 1.5-fold in the tiling array data presented here. Heatmaps were sorted by the tiling array data values for the *paf1*Δ vs. WT comparison. NET-seq data were taken from (Harlen and Churchman 2017b), ChIP-exo data were taken from (Van Oss et al. 2016), and log2 (probe intensity) of RNA levels in WT were from our tiling array data (this work). All heatmaps are plotted relative to the transcription start site (TSS). Regions 1kb upstream and downstream of the TSS are shown. (C) Log2 fold difference distributions for *paf1*Δ vs. WT comparison across ten deciles binned by expression level with decile one containing the most lowly expressed genes and decile ten containing the most highly expressed genes. (D) Genome browser views of Rpb3 and Paf1 occupancy data measured by ChIP-exo (Van Oss et al. 2016) for *ARN1*, *FIT2*, *FIT3*, *SIT1*, *PHO5*, *PHO81*, and *PHO84*.

Similarly, Paf1 occupancy is higher at Paf1-stimulated genes compared to Paf1-repressed genes, consistent with the known association of Paf1C with Pol II. In agreement with the Rpb3 ChIP-exo data, analysis of transcript abundance in wild type strains further demonstrated that Paf1 generally plays a positive role at highly active genes (Figure 13A-C). Since defects in Paf1C cause a disruption in telomeric silencing (Ng *et al.* 2003a; Krogan *et al.* 2003), we also analyzed the chromosomal locations of Paf1-regulated genes. Our analysis revealed broad chromosomal distribution of Paf1-stimulated and Paf1-repressed genes, in both *TRF4* and *trf4*Δ backgrounds, with no obvious bias toward telomeres (Figure 14).

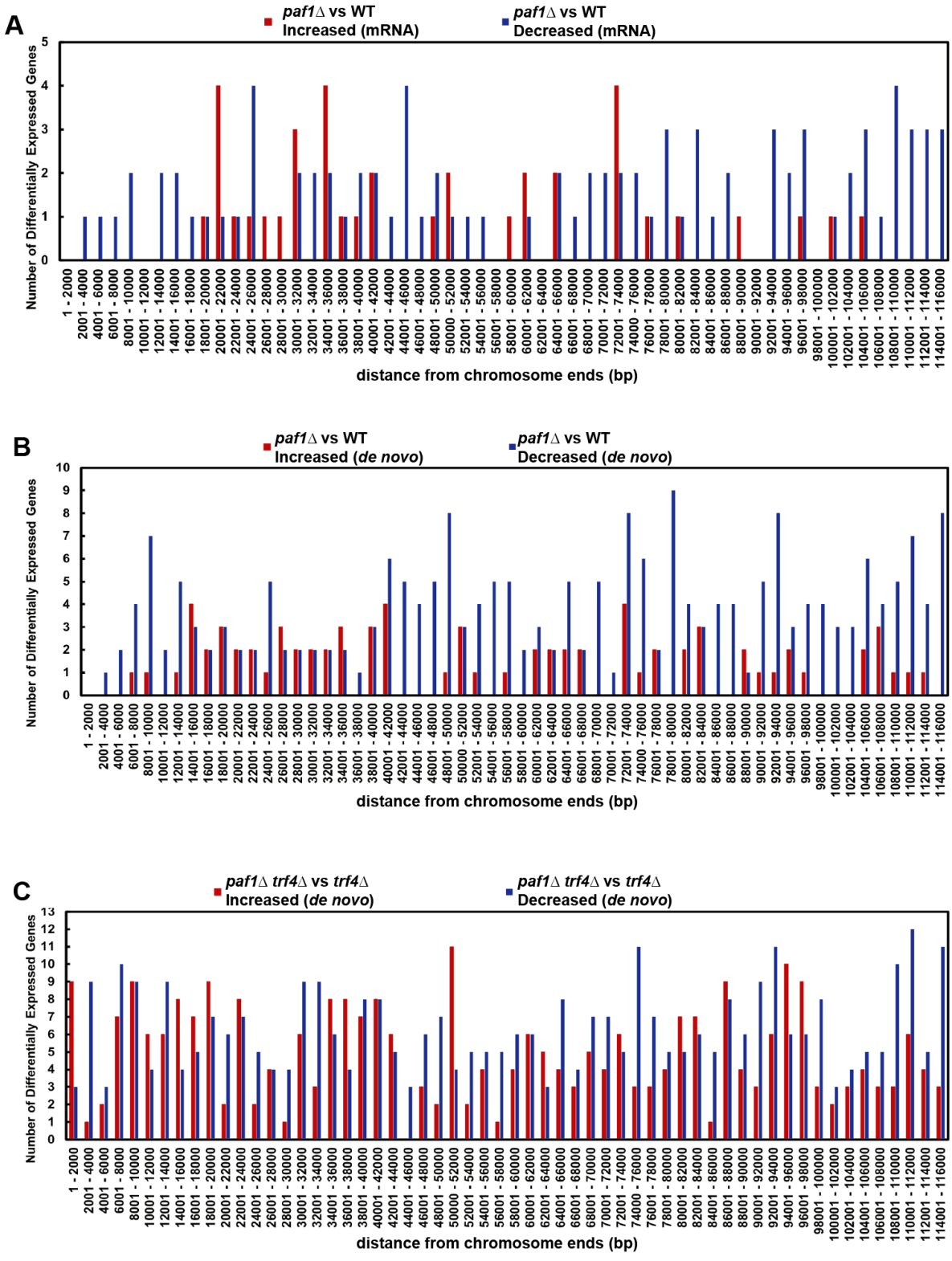

A
paf1Δ vs WT
Increased (mRNA)
paf1Δ vs WT
Decreased (mRNA)
Number of Differentially Expressed Genes
distance from chromosome ends (bp)
B
paf1Δ vs WT
Increased (de novo)
paf1Δ vs WT
Decreased (de novo)
Number of Differentially Expressed Genes
distance from chromosome ends (bp)
C
paf1Δ trf4Δ vs trf4Δ
Increased (de novo)
paf1Δ trf4Δ vs trf4Δ
Decreased (de novo)
Number of Differentially Expressed Genes
distance from chromosome ends (bp)

Figure 14. The positions of Paf1-regulated transcripts are not strongly biased toward the telomere. Non-overlapping bins of 2000bp were generated working in toward the centromere from both ends of each yeast chromosome. Bins were intersected with transcript annotations to generate count tables. Bar graphs show the number of transcripts within each bin that increased (red) or decreased (blue) in the *paf1*Δ background. (A) Counts of differentially expressed mRNAs in a *paf1*Δ strain relative to WT. (B) Counts of differentially expressed transcripts identified in the *de novo* analysis comparing *paf1*Δ to WT. (C) Counts of differentially expressed transcripts identified in the *de novo* analysis comparing *paf1*Δ *trf4*Δ to *trf4*Δ.

2.3.4 Paf1 stimulates the expression of phosphate homeostasis genes through a mechanism independent of its effects on individual histone modifications

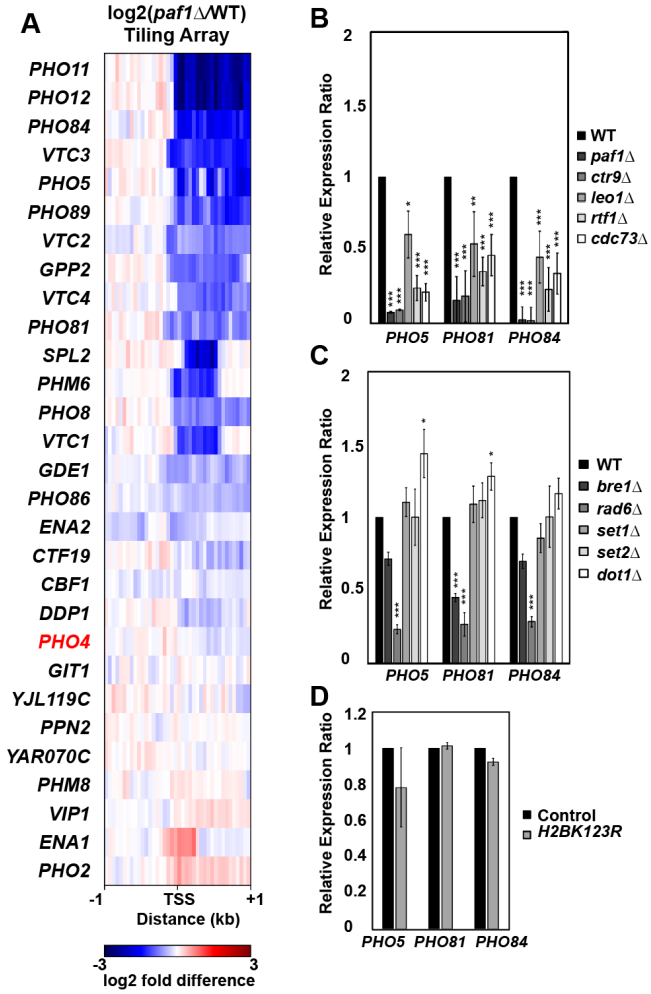
Gene ontology analysis (Ashburner *et al.* 2000) revealed an enrichment in phosphate homeostasis genes among the genes where expression decreased upon the deletion of *PAF1* in both the *TRF4* and *trf4*Δ backgrounds (Table 19). Given the wealth of information on the importance of chromatin structure in regulating genes in the phosphate regulon (Korber and Barbaric 2014), we explored the mechanism of Paf1 involvement at these genes. Our tiling array data show that many but not all genes activated by the Pho4 transcription factor (Zhou and O'Shea 2011) are downregulated in the absence of Paf1 (Figure 15A), indicating that the effects of Paf1 are unlikely to be due to a loss of Pho4 function. Consistent with this, *PHO4* transcript levels are not strongly affected in the *paf1*Δ strain (Figure 15A).

To assess the contribution of individual Paf1C members to the expression of phosphate homeostasis genes, we performed RT-qPCR analysis of RNA isolated from *paf1*Δ, *ctr9*Δ, *cdc73*Δ, *rtf1*Δ, and *leo1*Δ strains. The RT-qPCR results generally agreed with our tiling array results. RNA levels for *PHO5, PHO81,* and *PHO84* were significantly decreased in the absence of any single Paf1C subunit with deletions of *PAF1* and *CTR9,* causing the greatest effects (Figure 15B).

Occupancy of Paf1 on *PHO5*, *PHO81*, and *PHO84*, as demonstrated by ChIP-exo analysis (VanOss *et al.* 2016), argues for a direct role of Paf1 in regulating these genes (Figure 13D).

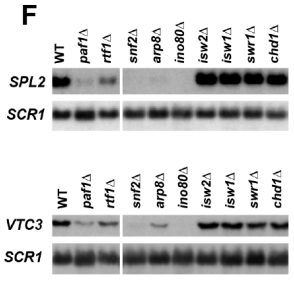
Table 19. Gene ontology results for genes that decreased expression in *paf1*Δ strain

	p value	n
Biological Process		
Small molecule metabolic process [GO:0044281]	1.95E-09	50
Small molecule biosynthetic process [GO:0044283]	6.07E-08	30
Single-organism biosynthetic process [GO:0044711]	8.96E-07	42
Oxoacid metabolic process [GO:0043436]	1.32E-06	33
Organic acid metabolic process [GO:0006082]	1.39E-06	33
Organic acid biosynthetic process [GO:0016053]	9.25E-06	21
Carboxylic acid biosynthetic process [GO:0046394]	9.25E-06	21
Single-organism metabolic process [GO:0044710]	6.55E-05	61
Branched-chain amino acid biosynthetic process [GO:0009082]	9.27E-05	7
Polyphosphate metabolic process [GO:0006797]	7.04E-04	6
Carboxylic acid metabolic process [GO:0019752]	2.44E-03	27
Cellular amino acid biosynthetic process [GO:0008652]	4.44E-03	14
Branched-chain amino acid metabolic process [GO:0009081]	7.73E-03	7
Alpha-amino acid biosynthetic process [GO:1901607]	1.30E-02	13
Organic hydroxy compound metabolic process [GO:1901615]	4.57E-02	13
Cellular Component		
Vacuolar transporter chaperone complex [GO:0033254]	1.08E-03	4



E

Chromatin remodeling complex	Subunit deleted	Growth defect when crossed with *paf1Δ*
CHD1	*chd1Δ*	none
ISW1	*isw1Δ*	severe
ISW2	*isw2Δ*	none
SWI/SNF	*snf2Δ*	severe
INO80	*arp8Δ*	lethal
SWR1	*swr1Δ*	lethal



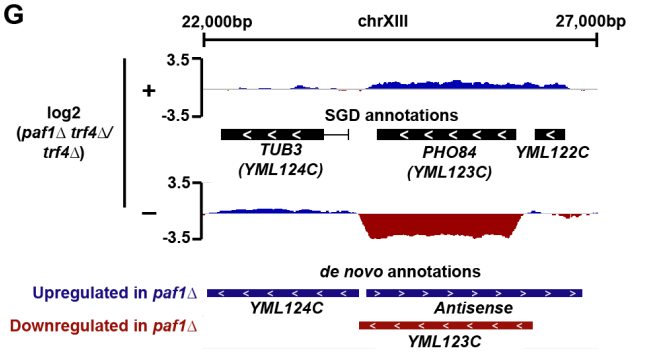


Figure 15. Paf1 positively regulates many phosphate homeostasis genes.

(A) Heatmap of expression differences observed in a *paf1Δ* strain (KY1702) relative to a WT strain (KY2276) at Pho4-responsive genes (Zhou and O'Shea, 2011). (B-D) RT-qPCR analysis of phosphate gene expression in strains lacking (B) individual Paf1C subunits (KY1021, KY2271, KY2239, KY2243, KY2241 and KY2244), (C) histone modification enzymes (KY1683, KY2045, KY1952, KY938, KY914, KY934) or (D) H2B K123. In panel D, RNA levels in the H2B K123R mutant (KY2167) were compared to the appropriate WT control strain (KY2027). Relative expression ratio is calculated using primer efficiency, normalization to the RNA polymerase III transcript *SCR1*, and a comparison to a WT strain (Pfaffl 2001). Error bars represent standard error of the mean and all statistically significant results are reported as asterisks (0.01 < p < 0.05 = *, 0.001 < p < 0.01 = **, 0 < p < 0.001 = *). All p values were derived from a student's t-test between the mutant strain and WT. (E) Cumulative data from crosses between a *paf1*Δ strain and strains deleted for chromatin remodeling factors. Following tetrad analysis of the following crosses, growth defects of double mutants were determined: *chd1*Δ *paf1*Δ = KY583 X KY804; *isw1*Δ *paf1*Δ = KY3464 X KY901; *isw2*Δ *paf1*Δ = KY884 X KY804; *snf2*Δ *paf1*Δ = KY508 X KY804; *arp8*Δ *paf1*Δ = KY3460 X KY804; *swr1*Δ *paf1*Δ = KY3462 X KY972. (F) Northern blot analysis of *SPL2* and *VTC3* RNA levels. Strains used in this analysis were KY292, KY802, KY508, KY3465, KY3461, KY884, KY3463, KY972, and KY632. *SCR1* serves as a loading control. (G) Genome browser view showing antisense transcription at the *PHO84* locus. The browser view shows smoothed differential expression tracks (log2(*paf1*Δ *trf4*Δ / *trf4*Δ), 160bp sliding window) with both SGD and *de novo* transcript annotations. Plus (+) and minus (-) symbols refer to the DNA strand. The *PHO84* gene is oriented right to left.**

Given the prominent role of Paf1C in promoting transcription-coupled histone modifications, we asked if the loss of these modifications could explain the gene expression changes we observed in the Paf1C mutant strains. To this end, we performed RT-qPCR assays on RNA prepared from strains lacking the H2Bub enzymes Rad6 and Bre1, the H3 K4 methyltransferase Set1, the H3 K36 methyltransferase Set2, or the H3 K79 methyltransferase Dot1 (Figure 15C). RNA levels for *PHO5*, *PHO81*, and *PHO84* decreased in the *rad6Δ* and *bre1Δ* strains, which, like an *rtf1*Δ strain (Van Oss *et al.* 2016), are severely deficient in H2Bub. However, the deletion of *PAF1* and *CTR9* had a substantially greater impact on the transcription of these genes than the deletion of either *BRE1*, which encodes the ubiquitin ligase for H2B K123, or *RTF1*, which encodes the primary Paf1C determinant of H2Bub (Figures 15B and 15C). The larger effect of *rad6*Δ compared to *bre1*Δ suggests that, as a ubiquitin-conjugating enzyme, Rad6 may play roles in *PHO* gene regulation beyond catalyzing H2Bub. In agreement with this, we observed only a slight decrease in *PHO5*, *PHO81*, and *PHO84* RNA levels in an H2B K123R mutant compared to the control strain (Figure 15D). Other than a slight, but statistically significant, upregulation of *PHO5* and *PHO81* in the *dot1Δ* strain, loss of individual H3 methyltransferases did not alter transcription of the *PHO* genes. Therefore, the loss of individual Paf1C-mediated histone modifications does not explain the strong reduction in phosphate homeostasis gene expression observed in *paf1*Δ and *ctr9*Δ mutants.

The absence of a clear effect of Paf1C-dependent histone modifications on *PHO* gene regulation prompted us to investigate other connections between Paf1 and chromatin. Previous work by Batta *et al.* (2012) showed reduced nucleosome occupancy within coding regions in a *paf1*Δ strain, and the importance of nucleosome occupancy changes in *PHO* gene expression has

been well documented (Barbaric *et al.* 2007; Korber and Barbaric 2014). Therefore, we investigated genetic interactions between Paf1 and chromatin remodeling factors. Genetic crosses were performed between strains lacking Paf1 and strains mutated for the following chromatin remodeling factors: Chd1, Isw1, Isw2, Swi/Snf, Ino80, and Swr1 (Figure 15E). We observed synthetic lethality or severe synthetic growth defects in double mutants containing *paf1*Δ and a deletion of *SWR1*, *ISW1, SNF2,* or *ARP8*, which encodes a subunit of the Ino80 complex. While the molecular basis for these genetic interactions is unclear, it is likely that Paf1C and chromatin remodeling factors regulate the expression of a shared group of genes. To test this idea for genes in the Pho4 regulon, we focused on two genes, *SPL2* and *VTC3*, known to be stimulated by Ino80 (Ohdate *et al.* 2003). Northern analysis revealed greatly reduced *VTC3* and *SPL2* expression in cells lacking *PAF1, SNF2, ARP8,* or *INO80* (Figure 15F). Although the deletion of *SWR1* or *ISW1* did not affect *SPL2* or *VTC3* mRNA levels, it is possible that other Paf1-dependent genes are regulated by these remodeling factors. Together, these data suggest that Paf1C and chromatin remodeling factors work in parallel to maintain gene expression levels required for cell viability and phosphate homeostasis.

A well-studied example of locus-specific antisense control of transcription occurs at the *PHO84* gene in yeast (Castelnuovo *et al.* 2013). At this locus, the accumulation of an antisense transcript in an *rrp6*Δ strain leads to repression of the sense transcript through a mechanism dependent on particular histone modifications (Castelnuovo *et al.* 2014). In light of the changes in antisense RNAs detected in the *paf1*Δ background (Figure 10B and Figure 11), we examined our *de novo* differential expression data for evidence of Paf1-regulated antisense transcription at *PHO84* (Figure 15G). When comparing the *paf1*Δ *trf4*Δ mutant to the *trf4*Δ control strain, we observed increased antisense and decreased sense transcript levels across the *PHO84* gene.

Interestingly, *PHO84* fell into one of the two clusters of genes for which an anti-correlation between sense and antisense transcription was observed in the *paf1*Δ *trf4*Δ mutant (Figure 11B, cluster 5). These data suggest that, at the *PHO84* gene, the deletion of *PAF1* elevates antisense transcription and coordinately decreases sense transcription.

2.3.5 Paf1 represses iron homeostasis gene expression through its role at H3K36me3

Gene ontology analysis of genes that increased expression in *paf1Δ* strains revealed enrichment for genes in various iron-related processes (Table 20). As with the phosphate genes, an additional motivation for choosing iron homeostasis genes for follow-up experiments was the extent to which they have been characterized in the literature (Yamaguchi-Iwai *et al.* 2002; Rutherford and Bird 2004; Courel *et al.* 2005; Kaplan and Kaplan 2009; Cyert and Philpott 2013). Our tiling array analysis of genes that are normally activated by the Aft1 and Aft2 transcription factors in iron-limiting conditions (Cyert and Philpott 2013) revealed that many but not all of these genes are repressed by Paf1 in iron-replete media (Figure 16A).

To further investigate this subset of genes, we performed RT-qPCR analysis on RNA isolated from strains lacking individual Paf1C members (Figure 16B). A reproducible increase in expression was observed for *ARN1*, *FIT2*, *FIT3,* and *SIT1* in *paf1*Δ and *ctr9*Δ strains. Occupancy of Paf1 on these genes suggests that Paf1 is playing a direct role in mediating their repression (Figure 13D). With the exception of *SIT1*, deletion of *CDC73* also led to derepression of these genes. In contrast, whereas Paf1, Ctr9, and Cdc73 repress the transcription of *ARN1*, *FIT2*, and *FIT3*, Leo1 appears to play a stimulatory role at these genes, while Rtf1 has little effect. Together, these data demonstrate that individual Paf1C subunits differentially regulate iron-responsive genes.

Table 20. Gene ontology results for genes that increased expression in *paf1*Δ strain

Biological Process	*p* value	n
Siderophore transport [GO:0015891]	3.26E-04	5
Iron chelate transport [GO:0015688]	5.91E-04	5
Iron coordination entity transport [GO:1901678]	5.33E-03	5
Glycerol transport [GO:0015793]	2.29E-02	4
Iron ion homeostasis [GO:0055072]	3.25E-02	7
Cellular Component		
Integral component of plasma membrane [GO:0005887]	2.77E-03	10
Intrinsic component of plasma membrane [GO:0031226]	7.38E-03	10
Extracellular region [GO:0005576]	2.00E-02	9
Plasma membrane part [GO:0044459]	2.22E-02	11

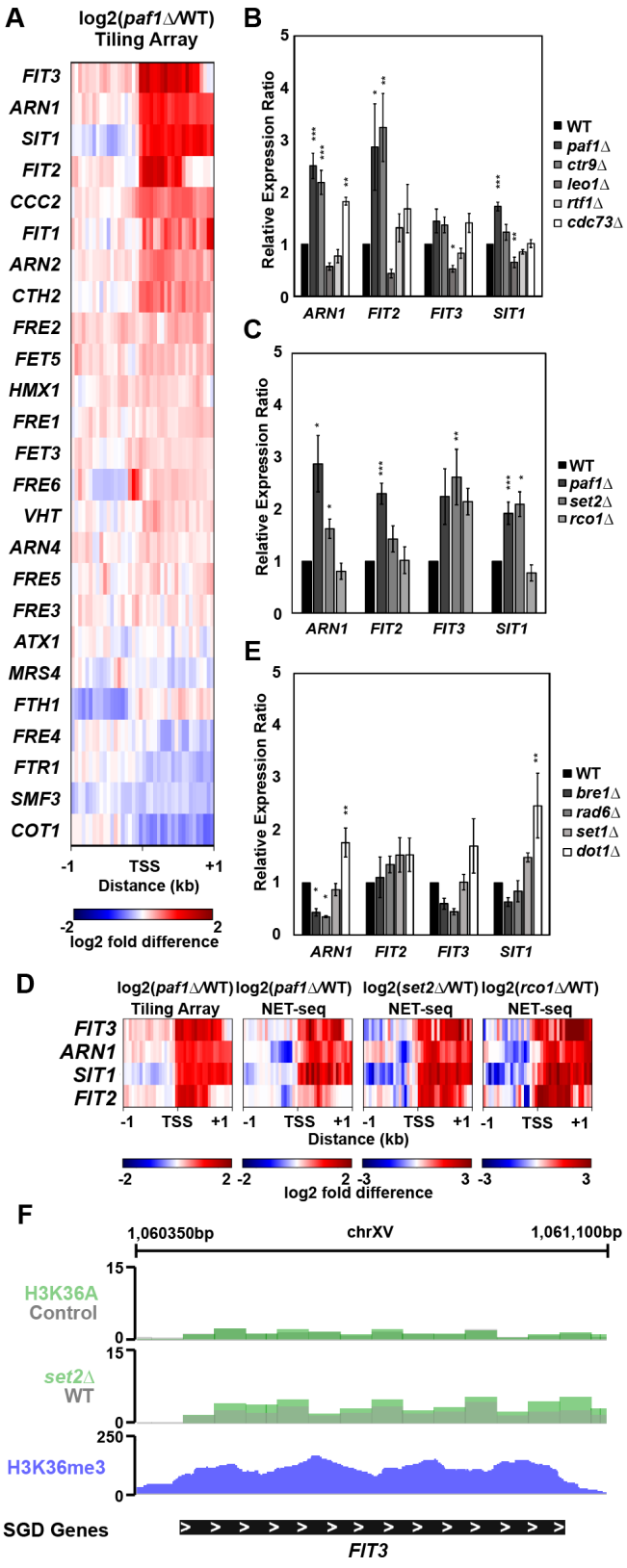

A
log2(paf1ΔWT)
Tiling Array
FIT3
ARN1
SIT1
FIT2
CCC2
FIT1
ARN2
CTH2
FRE2
FET5
HMX1
FRE1
FET3
FRE6
VHT
ARN4
FRE5
FRE3
ATX1
MRS4
FTH1
FRE4
FTR1
SMF3
COT1
-1
TSS
+1
Distance (kb)
-2
2
log2 fold difference
B
Relative Expression Ratio
WT
paf1Δ
ctr9Δ
leo1Δ
rtf1Δ
cdc73Δ
ARN1
FIT2
FIT3
SIT1
C
Relative Expression Ratio
WT
paf1Δ
set2Δ
rco1Δ
ARN1
FIT2
FIT3
SIT1
E
Relative Expression Ratio
WT
bre1Δ
rad6Δ
set1Δ
dot1Δ
ARN1
FIT2
FIT3
SIT1
D
log2(paf1ΔWT) Tiling Array
log2(paf1ΔWT) NET-seq
log2(set2ΔWT) NET-seq
log2(rco1ΔWT) NET-seq
FIT3
ARN1
SIT1
FIT2
-1
TSS
+1
Distance (kb)
-2
2
-3
3
log2 fold difference
F
1,060350bp
chrXV
1,061,100bp
H3K36A
Control
set2Δ
WT
H3K36me3
15
0
250
SGD Genes
FIT3

Figure 16. Paf1 represses iron homeostasis genes.

(A) Heatmap of expression differences observed in a *paf1Δ* strain (KY1702) relative to a WT strain (KY2276) at Aft1 and Aft2 responsive genes involved in the maintenance of iron homeostasis (Cyert and Philpott 2013). (B-C) RT-qPCR analysis of the indicated genes in strains lacking (B) individual Paf1C subunits (KY1021, KY2271, KY2239, KY2243, KY2241 and KY2244) or (C) genes in the Set2/Rpd3S pathway (KY307, KY914, KY1702 and KY1235). Calculation of the relative expression ratio and statistical testing were performed as in Figure 15. (D) Heatmaps of expression differences between mutant yeast strains and their respective WT strains in tiling array (this study) and NET-seq (Churchman and Weissman 2011; Harlen and Churchman 2017b) datasets. (E) RT-qPCR results for iron homeostasis genes in strains lacking enzymes that catalyze Paf1C-associated histone modifications (KY1683, KY2045, KY1952, KY938, and KY934). F) Genome browser view of the *FIT3* locus showing H3 K36A and *set2*Δ RNA-seq data from (Venkatesh *et al.* 2016) and H3 K36me3 occupancy data from (Weiner *et al.* 2015).

The Set2 histone methyltransferase catalyzes H3 K36me3, a modification that is dependent on Paf1 and Ctr9 (Krogan *et al.* 2003; Chu *et al.* 2007). This epigenetic mark leads to the activation of the Rpd3S histone deacetylase complex and inhibition of histone exchange, generating a repressed chromatin state (Carrozza *et al.* 2005; Joshi and Struhl 2005; Keogh *et al.* 2005; Govind *et al.* 2010; Churchman and Weissman 2011; Venkatesh *et al.* 2012, 2016; Kim *et al.* 2016). In the *set2Δ* strain, RNA levels for *FIT3* and *SIT1* increased to match those observed in the *paf1Δ* strain, suggesting that Paf1 represses these genes through stimulating H3 K36me3 (Figure 16C). This conclusion is supported by our observation that a *set2*Δ *paf1*Δ double mutant recapitulates the increased expression of *FIT3* and *SIT1* observed in the *set2*Δ and *paf1*Δ single mutants (Figure 17). Further, NET-seq data (Churchman and Weissman 2011; Harlen and Churchman 2017b) indicate that the increase in steady state mRNA levels for *FIT3* and *SIT1* in *paf1*Δ, *set2*Δ, and *rco1*Δ strains is associated with an increase in transcription (Figure 16D).

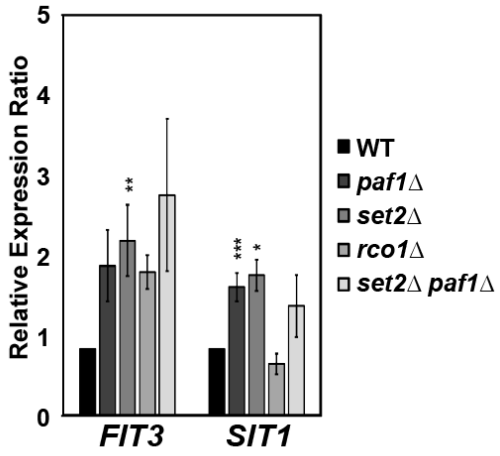


Figure 17. Evidence that Set2 and Paf1 function in the same pathway to repress *FIT3* and *SIT1* expression. **RT-qPCR data collected for *FIT3* and *SIT1* expression in strains deleted for genes in the Paf1-Set2-Rpd3S pathway (KY307, KY914, KY1702, KY1235, and KY1250). Bar graphs show the average of at least 3 biological replicates and standard error of the mean. All statistically significant results are reported as asterisks (0.01 < p < 0.05 = *, 0.001 < p < 0.01 = **, 0 < p < 0.001 = ***). All p values were derived from a student's t-test between the mutant strain and WT.**

For two other genes, *ARN1* and *FIT2*, the level of derepression observed in the *paf1*Δ strain was greater than that observed in the *set2*Δ strain, despite evidence from NET-seq data (Figure 16D; Churchman and Weissman 2011; Harlen and Churchman 2017) that loss of Set2 strongly increases transcription of these genes. Similarly, with the exception of *FIT3*, steady state mRNA levels in a strain lacking the Rpd3S subunit Rco1 did not reflect the increase in transcription detected by NET-seq (Figure 16C and 16D). One likely explanation for the difference between the steady state mRNA measurements (Figure 16C) and the nascent transcript data (Figure 16D) is that mRNA levels for the iron-responsive genes are post-transcriptionally regulated, possibly through a degradation pathway that involves Paf1. This conclusion is in line with observations made through our *de novo* analysis (Figure 12B), which indicated a post-transcriptional role for Paf1 at genes where it negatively regulates mRNA levels, and with previous descriptions of RNA degradation pathways that target mRNAs in the iron regulon (Lee *et al.* 2005; Puig *et al.* 2005).

In addition to the Set2/Rpd3S pathway, we tested other histone modifiers for a role in iron gene repression by examining *ARN1, FIT2, FIT3,* and *SIT1* expression in *bre1Δ, rad6Δ, set1Δ,* and *dot1Δ* strains by RT-qPCR (Figure 16E). With the exception of the *dot1Δ* mutation, which elevated *ARN1* and *SIT1* transcript levels, none of these mutations led to a significant derepression of the iron genes. The prominent role of Set2 in mediating *FIT3* repression is further supported by the enrichment of H3 K36me3 on this gene (Weiner *et al.* 2015) and the derepression of *FIT3* RNA levels in *set2*Δ and H3 K36A mutants (Venkatesh *et al.* 2016) (Figure 16F). Taken together, these results suggest that Paf1 represses expression of genes in the iron regulon by inhibiting transcription, most likely by facilitating histone marks such as H3 K36me3 and by influencing RNA stability.

2.3.6 *FET4* is differentially regulated by Paf1 and upstream CUT transcription

To investigate the interplay between Paf1 and noncoding DNA transcription on a protein-coding gene in the iron regulon, we focused on the *FET4* gene, which encodes a low-affinity iron transporter in *S. cerevisiae* (Dix *et al.* 1994). Two CUTs have been annotated upstream of the *FET4* coding region (Xu *et al.* 2009; Raupach *et al.* 2016) (Figure 18A). We hypothesized that *CUT 794/793* transcription regulates *FET4* transcription, possibly in a *PAF1*-dependent manner. To test this, we generated strains containing a transcription termination sequence (TTS) upstream of the *FET4* gene positioned to stop transcription of both upstream CUTs in wild type, *paf1*Δ, *trf4*Δ, and *paf1*Δ *trf4*Δ backgrounds.

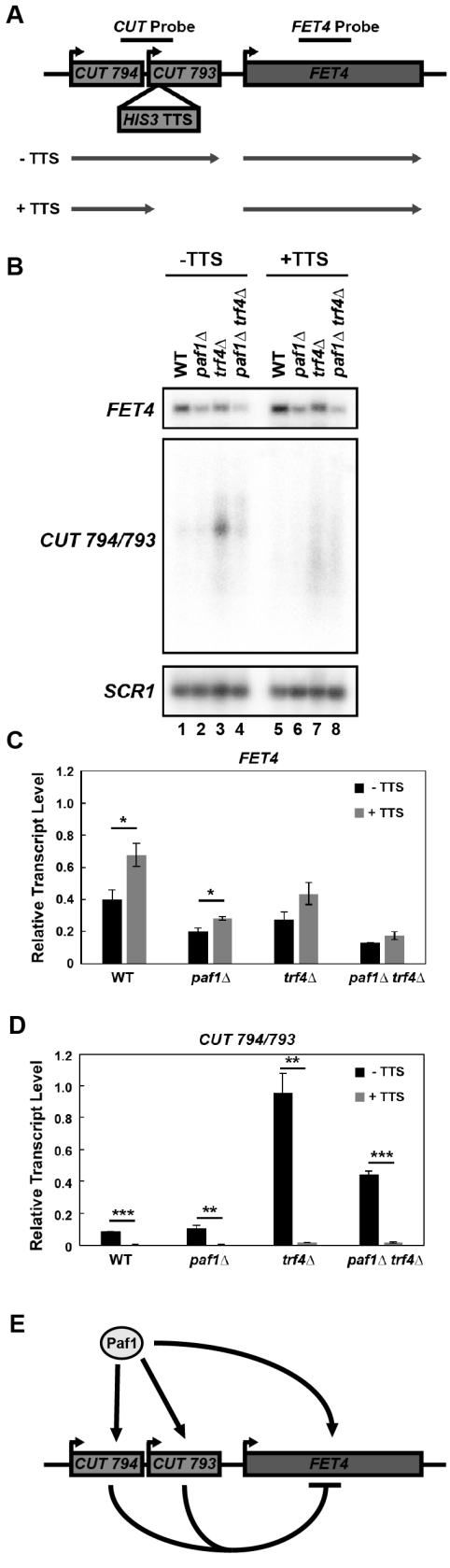

A
CUT Probe
FET4 Probe
CUT 794
CUT 793
FET4
HIS3 TTS
- TTS
+ TTS
B
-TTS
+TTS
WT
paf1Δ
trf4Δ
paf1Δ trf4Δ
FET4
CUT 794/793
SCR1
1 2 3 4 5 6 7 8
C
FET4
Relative Transcript Level
- TTS
+ TTS
WT
paf1Δ
trf4Δ
paf1Δ trf4Δ
D
CUT 794/793
Relative Transcript Level
- TTS
+ TTS
WT
paf1Δ
trf4Δ
paf1Δ trf4Δ
E
Paf1
CUT 794
CUT 793
FET4

Figure 18. The *FET4* locus is regulated by Paf1 and transcription of ncDNA upstream of the coding region. (A) Diagram of the *FET4* locus and the position of a transcription termination sequence (*HIS3* TTS) inserted 400 bp upstream of the *FET4* start codon to block *CUT 794/793* transcription. (B) Northern analysis of *FET4* mRNA, *CUT 794/793* and *SCR1* RNA (loading control) from WT, *paf1Δ*, *trf4Δ*, and *paf1Δ trf4Δ* strains without the inserted TTS (KY2276, KY1702, KY2012, KY2016) or with the TTS (KY3466, KY2846, KY2851, KY2845). (C-D) Quantification of northern blot results for *FET4* and *CUT 794/793* normalized to *SCR1*. Error bars represent standard error of the mean, and all statistically significant results are reported as asterisks that represent p values from students t-test as in Figure 15. (E) Diagram of the observed effects of *PAF1* and *CUT 794/793* at the *FET4* locus.

Northern analysis showed that the deletion of *PAF1* reduced *FET4* transcript levels (Figure 18B top blot, compare lanes 1 and 2 and lanes 3 and 4; Figure 18C) and *CUT794/793* levels (Figure 18B middle blot, compare lanes 3 and 4; Figure 18D). When the TTS was inserted upstream of *FET4* (+TTS), CUT levels decreased, and *FET4* transcript levels increased in all conditions tested, suggesting that transcription of the upstream CUT inhibits expression of the coding region (Figures 18B-D). This is reminiscent of the inhibitory effect of noncoding transcription upstream of the well-studied *SER3* gene (Martens *et al.* 2004). Interestingly, even when CUT transcription was blocked, *FET4* transcript levels were reduced in the *paf1*Δ background. These data suggest that CUT transcription upstream of the *FET4* promoter negatively regulates transcription of the *FET4* gene and that Paf1 has a dual role in regulating *FET4* by stimulating expression of both the ORF and the inhibitory *CUTs 794/793* (Figure 18E).

2.4 Discussion

The many roles of Paf1C and the pleiotropic phenotypes conferred by the deletion of individual Paf1C subunits (Betz *et al.* 2002) suggest that this conserved transcription elongation complex regulates the expression of many genetic loci. While previous studies focused on the regulation of mRNAs (Shi *et al.* 1996; Penheiter *et al.* 2005; Porter *et al.* 2005; Batta *et al.* 2011; van Bakel *et al.* 2013; Cao *et al.* 2015; Chen *et al.* 2015; Yang *et al.* 2016; Harlen and Churchman 2017b; Fischl *et al.* 2017), here we sought to identify the full cohort of Paf1-regulated transcripts, both coding and noncoding, by exploiting a genetic background deficient in the TRAMP/exosome-dependent RNA degradation pathway and by performing *de novo* transcript identification analyses. Our high-resolution tiling array experiments revealed differential expression of transcripts in all Pol II transcribed RNA classes in strains deleted for *PAF1.* A comparison of our *paf1*Δ *trf4*Δ tiling array data with published *paf1*Δ NET-seq (Harlen and Churchman 2017b) data demonstrated that Paf1 regulates many coding and noncoding RNAs at the transcriptional level and that the presence of a functional TRAMP complex obscures many of these transcriptional effects.

Our study revealed both positive and negative roles of Paf1 in regulating noncoding RNA levels. For many transcripts in the CUT, NUT, XUT, and SUT classes, Paf1 stimulates their expression. For two other important classes of noncoding RNAs, snoRNAs, and SRATs, Paf1 functions primarily as a negative regulator. The elevation in SRAT expression was not unexpected given the importance of Paf1 for H3 K36me3 (Krogan *et al.* 2003; Chu *et al.* 2007), a mark important for the maintenance of a repressive chromatin environment. In our previous work, we showed that Paf1 is important for snoRNA termination (Sheldon *et al.* 2005; Tomson *et al.* 2011, 2013a). We recapitulate those results here and identify additional snoRNA loci that exhibit transcription termination defects in the absence of Paf1 (Figure 7H). These results, together with

our finding that Paf1 impacts the transcription of many CUTs and NUTs, extends the functional connections between Paf1 and the machinery that terminates and processes these noncoding RNAs, including the NNS machinery and the nuclear exosome. Similar to a *paf1*Δ strain, snoRNA 3' ends are extended in *rrp6* and *nrd1* mutants, which lack subunits of the nuclear exosome and NNS, respectively (Schulz *et al.* 2013; Fox *et al.* 2015). In contrast, while NUTs and CUTs are elevated in *nrd1* and *rrp6* mutants, levels of many of these unstable noncoding RNAs are decreased in strains deleted for *PAF1*. The reduced levels of these RNAs in *paf1*Δ strains are likely due, at least in part, to the stimulatory effect Paf1 has on their transcription (Figure 12).

By performing a *de novo* differential expression analysis of our tiling array data, we uncovered the effects of Paf1 on antisense transcription (Figure 10B). Interestingly, many of the histone modifications promoted by Paf1C are reduced in regions experiencing higher levels of antisense transcription, but still others are present at high levels in these same regions (Murray *et al.* 2015). The loss of Paf1C-promoted histone modifications may therefore contribute to changes in antisense transcription (Castelnuovo *et al.* 2014; Murray *et al.* 2015). Indeed, we found instances of both increased and decreased antisense transcription in our *paf1*Δ strains. Although some global anticorrelation exists between sense and divergent antisense transcription initiating from nucleosome depleted regions (Xu *et al.* 2009; Churchman and Weissman 2011), antisense transcription does not universally correlate or anticorrelate with sense transcription (Murray *et al.* 2015). Our results agree with this observation (Figure 11E), but also point to a small subset of genes where sense and antisense transcription appear to be anticorrelated when *PAF1* is deleted (Figure 11A cluster 1 and 11B cluster 5).

One gene that fits into this category is *PHO84* (Figure 11B cluster 5; Figure 15G). Our data suggest a role for Paf1 in preventing antisense transcription at *PHO84* independently of its

functional connections with the TRAMP/exosome pathway, as we detect higher antisense and lower sense transcript levels in the *paf1*Δ *trf4*Δ strain relative to the *trf4*Δ strain (Figure 15G). The ability to detect changes in antisense transcription was enhanced by the absence of Trf4. This observation agrees with studies on *PHO84* and other genes, which showed elevated antisense transcription in the absence of Rrp6 (Castelnuovo *et al.* 2013). With respect to *PHO84*, the mechanism by which Paf1 facilitates sense and represses antisense transcription remains undefined. Although previous studies showed that *set1*Δ strongly upregulates *PHO84* sense transcription in the presence of *RRP6* (Castelnuovo *et al.* 2013) and that Paf1C is required for Set1-dependent H3 K4 methylation (Ng *et al.* 2003b; Krogan *et al.* 2003), our strand-specific tiling array data showed that *paf1*Δ strongly downregulates *PHO84* sense transcription (Figure 15A). Similarly, our results do not ascribe the stimulatory effect of Paf1 on *PHO5* and *PHO81* to any single Paf1C-dependent histone modification or an obvious change in antisense transcription; however, it remains possible that the individual modifications function redundantly in promoting the expression of these genes.

At many genes that are normally induced in iron-limiting conditions, Paf1 plays a repressive role under iron-replete conditions. For the four strongly upregulated genes examined, the deletion of *PAF1* increased steady state RNA levels as well as nascent transcript levels, arguing that Paf1 is controlling the transcription of these genes. The increase in transcription in the *paf1*Δ strain correlates with a decrease in Set2 function, as shown by NET-seq data (Churchman and Weissman 2011; Harlen and Churchman 2017b). Indeed, in our tiling array experiments, we detected a global increase in SRAT transcription in a *paf1Δ* strain (Figure 7D and 10A). Interestingly, when comparing steady state RNA levels and nascent transcript levels, we noted an apparent post-transcriptional effect of Paf1 (Figures 12B, 16C, and 16D). With respect to the iron

metabolism genes, we saw a strong overlap between Paf1-repressed mRNAs and Rnt1-repressed mRNAs (Lee *et al.* 2005). Rnt1 is a double-stranded RNA endonuclease that cleaves RNAs with a particular stem-loop structure (Chanfreau *et al.* 2000) and, in iron-replete conditions, executes an RNA degradation pathway for mRNAs that encode iron uptake proteins (Lee *et al.* 2005). While other explanations are possible, the overlap between Paf1- and Rnt1-repressed mRNAs suggests that the post-transcriptional role of Paf1 at iron regulon genes may involve a functional interaction with Rnt1. A recent study showed that the rate of transcription elongation can influence the folding and processing of histone pre-mRNAs (Saldi *et al.* 2018), raising the possibility that the deletion of *PAF1* might alter the rate of elongation in a way that affects the folding of substrates for Rnt1 or another RNA processing factor. Together, our results suggest that through stimulating Set2-mediated H3 K36 methylation, Paf1 represses genes in the iron regulon, but has an additional role in reducing the stability of these mRNAs.

Numerous examples of protein-coding gene regulation by ncDNA transcription or ncRNAs have been observed (Castelnuovo and Stutz 2015). Adding to this body of evidence, we investigated the regulatory mechanisms governing the expression of *FET4*, which encodes a low-affinity iron transporter. Our analysis indicates that *FET4* is regulated by the expression of upstream CUTs and by Paf1. The insertion of a transcription termination sequence upstream of *FET4* decreased *CUT794/793* levels and increased *FET4* transcription. The deletion of *PAF1* reduced both *CUT794/793* and *FET4* transcript levels. Together with these targeted experiments, our tiling array results on the *paf1*Δ strain also revealed a stimulatory effect of Paf1 on *FET4* mRNA levels. However, we note that our tiling array analysis of the *paf1*Δ *trf4*Δ strain indicated that, in some circumstances, Paf1 can repress *FET4* expression. Previous studies have shown that genetic background and growth conditions can influence the levels of *FET4* mRNA and the

noncoding RNAs adjacent to or overlapping *FET4* (*CUT794/793* and the *SUT322* antisense ncRNA) (Xu *et al.* 2009). Since our tiling array and northern blotting experiments used RNA from yeast grown on separate days, it is possible that slight differences in media or growth stage may be responsible for differences in expression dynamics at the *FET4* locus, highlighting the intricacies of the regulatory system operating at this gene. Collectively, our results add to the interesting list of telomere-proximal metal-responsive genes under the control of noncoding transcription (Toesca *et al.* 2011).

The complexity of the transcription process and its regulation by chromatin provides numerous opportunities for multifunctional transcription factors, like Paf1C, to regulate gene expression. Our study reveals the genome-wide effects of Paf1 on both coding and noncoding RNAs and provides mechanistic explanations for its diverse effects on specific classes of protein-coding genes. An understanding of the locus-specific effects of Paf1C will be an important step in elucidating the numerous connections of this complex to gene expression changes that cause human disease (Tomson and Arndt 2013; Karmakar *et al.* 2018).

3.0 Spt6 recruits Paf1C to Pol II through an interaction with Cdc73

The data presented in this chapter will be published in a manuscript describing the role of Spt6 in Paf1C recruitment and the Spt6-Cdc73 interaction. Much of the gel shift assay data presented here was generated by undergraduate student Matthew Blacksmith under my mentorship. Serial dilution growth assays and western blots for Spt6 and Cdc73 binding interface mutants were conducted by undergraduate student Julia Seraly under my mentorship. All BPA crosslinking work was conducted by Rachel Schusteff and Eleanor Kerr under the mentorship of Dr. Margaret Shirra.

3.1 Introduction

The eukaryotic genome is organized into chromatin (Rando and Chang 2009; Rando and Winston 2012), with the nucleosome as its basic repeating unit. Nucleosomes pose a barrier to RNA polymerase enzymes and regulate transcription via post-translational modifications to the histone proteins (Berger 2007; Li *et al.* 2007; Smolle and Workman 2013; Venkatesh and Workman 2015). RNA polymerase II (Pol II) is responsible for the synthesis of messenger RNAs (mRNAs) and a suite of non-coding RNAs, some functional, and others that appear to be the result of pervasive transcription (Xu *et al.* 2009; Yassour *et al.* 2010; van Dijk *et al.* 2011; Schulz *et al.* 2013; Venkatesh *et al.* 2016).

The 12-subunit Pol II holoenzyme requires numerous accessory factors to transcribe the chromatin template faithfully (Kuehner *et al.* 2011; Sainsbury *et al.* 2015; Porrua and Libri 2015;

Harlen and Churchman 2017a; Van Oss *et al.* 2017; Chen *et al.* 2018). Two essential elongation factors, Spt4-Spt5 (reviewed in Hartzog and Fu 2013) and Spt6 (reviewed in Duina 2011), promote elongation by increasing Pol II processivity (Bourgeois *et al.* 2002; Zhu *et al.* 2007) and chaperoning histone proteins displaced by the polymerase (Bortvin and Winston 1996; Keegan *et al.* 2002; Endoh *et al.* 2004; Cheung *et al.* 2008; Jensen *et al.* 2008), respectively. Yeast genetic screens identified Spt4, Spt5, and Spt6 as mutants that could suppress the effects of retrotransposon insertions at the 5' end of the *LYS2* gene (*lys2-128∂* allele) and the promoter of the *HIS4* gene (*his4-912∂* allele) (Simchen *et al.* 1984; Winston *et al.* 1984; Clark-Adams and Winston 1987), and subsequently implicated these genes in the control of transcription elongation (Hartzog *et al.* 1998; Kaplan *et al.* 2000). Now their individual structures (Guo *et al.* 2008; Dengl *et al.* 2009; McDonald *et al.* 2010; Sun *et al.* 2010; Close *et al.* 2011; Meyer *et al.* 2015; Schüller *et al.* 2016), interactions with the elongating polymerase (Bernecky *et al.* 2017; Sdano *et al.* 2017; Vos *et al.* 2018a; b; Ehara *et al.* 2019), and the transcriptional consequences of their loss have been documented (Baejen *et al.* 2017; Shetty *et al.* 2017; Dronamraju *et al.* 2018a; Doris *et al.* 2018; Baluapuri *et al.* 2019; Jeronimo *et al.* 2019).

The polymerase associated factor 1 complex or Paf1C is another critical Pol II accessory factor that is composed of 5 protein subunits: Paf1, Ctr9, Leo1, Cdc73, and Rtf1 (Wade *et al.* 1996; Shi *et al.* 1997; Mueller and Jaehning 2002). Paf1C is an elongation factor that promotes histone modifications, such as H2BK123ub, H3K36me3, H3K79me2/3, and H3K4me2/3 (Krogan *et al.* 2003; Chu *et al.* 2007; Piro *et al.* 2012; Van Oss *et al.* 2016). Paf1C also regulates transcription elongation efficiency (Kim *et al.* 2010; Tous *et al.* 2011; Vos *et al.* 2018b, 2020; Hou *et al.* 2019) and interacts with the chromatin remodeling enzyme Chd1 (Squazzo *et al.* 2002; Simic *et al.* 2003; Warner *et al.* 2007). Additionally, Paf1C promotes proper transcription termination of snoRNA

loci (Tomson *et al.* 2011, 2013a; Ellison *et al.* 2019), poly(A) tail formation (Nordick *et al.* 2008; Yang *et al.* 2016), and mRNA export from the nucleus (Fischl *et al.* 2017). Paf1C has been implicated directly in cancers of the blood, parathyroid, breast, liver, stomach, and colon (Guenther *et al.* 2005; Wang *et al.* 2005; Chaudhary *et al.* 2007; Zhao *et al.* 2007; Muntean *et al.* 2010; Smith *et al.* 2011; Takahashi *et al.* 2011).

The Paf1C must be properly recruited to Pol II to perform its critical functions. Association between Paf1 and Pol II was initially shown by immunoaffinity chromatography (Wade *et al.* 1996). Its recruitment is dependent upon the phosphorylation state of the Pol II CTD (found on the Rpb1 subunit) (Qiu *et al.* 2012) and the Spt5 CTR (Zhou *et al.* 2009; Liu *et al.* 2009; Mayekar *et al.* 2013; Wier *et al.* 2013), both of which are critical for proper Paf1C levels at coding regions. Detailed analysis of interactions between Pol II CTD and Spt5 CTR peptides suggests that multiple Paf1C members are competent to interact with the CTD of Pol II and CTR of Spt5 in an *in vitro* setting (Qiu *et al.* 2012).

There is already a molecular-level understanding of the Paf1C-Spt5-CTR interaction. A domain of Rtf1 termed the OAR (Open reading frame Associated Region) in yeast, or Plus3 domain in humans, is responsible for this interaction (Mayekar *et al.* 2013; Wier *et al.* 2013). The Rtf1 OAR binds directly to the phosphorylated CTR of Spt5, forming an interaction critical to promoting proper Paf1C levels at coding regions (Mayekar *et al.* 2013; Wier *et al.* 2013). A recent cryoEM structure of a Pol II elongation complex containing Rtf1 confirmed the Rtf1-Spt5 CTR interaction (Vos *et al.* 2020). In addition, numerous Paf1C subunit and Paf1C-Pol II interactions are understood (Xu *et al.* 2017; Vos *et al.* 2018b, 2020; Deng *et al.* 2018). However, the interactions between Cdc73 and the elongation complex are still unclear. The C-domain of Cdc73, has a Ras-like fold and flattened out GTP binding pocked and is known to be critical for Paf1C

localization at coding regions (Amrich *et al.* 2012). Loss of the C-domain results in a reduction in Cdc73, Rtf1, and Ctr9 occupancy at coding regions without affecting Pol II occupancy or Paf1C formation (Amrich *et al.* 2012). Although there are no reports of a direct *in vivo* interaction between the Pol II CTD and Cdc73, the interaction occurs in peptide binding experiments (Qiu *et al.* 2012), and *in vivo* analysis of yeast strains lacking the C-domain of Cdc73 show reduced Paf1C recruitment (Amrich *et al.* 2012).

We sought to determine if the interaction between the Pol II CTD and the Cdc73 C-domain occurs *in vivo* and identify additional C-domain interacting proteins by conducting site-specific crosslinking in *Saccharomyces cerevisiae.* Using BPA crosslinking, we demonstrate that the C-domain directly interacts with two core members of the Pol II elongation complex: the Pol II CTD and the histone chaperone Spt6. Here I present my investigation into the role of Spt6 in Paf1C recruitment. I provide functional evidence revealing that loss of Spt6 results in loss of genome-wide Paf1 occupancy and put forth a structural model of the Cdc73-Spt6 interaction. My results suggest that Spt6 is indispensable for recruiting Paf1C to actively transcribed regions and add to our understanding of the Pol II elongation complex and its selective recruitment of factors important for controlling chromatin structure and transcription.

3.2 Materials and methods

3.2.1 Yeast strains, media, and genetic manipulation

S. cerevisiae strains and *K. lactis* spike-in strain used in this study are listed in Table 21. *S. cerevisiae* strains are isogenic to the FY2 strain, which is a *GAL2*$^{+}$ derivative of S288C (Winston *et al.* 1995). Epitope tagging of specific loci was achieved by one-step gene disruption (Lundblad *et al.* 2001) and confirmed by PCR. Genetic crosses were conducted as described in (Rose *et al.* 1991). Yeast transformations were conducted using the lithium acetate method as previously described (Becker and Lundblad 1994). Cells were grown to log phase at 30°C in rich media (YPD) supplemented with 400 μM tryptophan and harvested by centrifugation. In all auxin inducible degron experiments, cells were grown in YPD+W, and either 100% DMSO or 100% DMSO containing 500mM IAA was added at a final concentration of 0.1% during log phase cell growth and incubated at 30 ˚C for the indicated amount of time. In experiments using plasmid-borne, *CDC73* or *SPT6* mutants, selection was maintained by growth in SC-W media. For *SPT6* mutants, the plasmid shuffle technique was used to introduce the mutant copy of the gene. By selection on SC-Trp medium containing 0.1% 5-fluoroorotic acid (FOA), *URA3* plasmids containing *SPT6* were shuffled (Sikorski and Boeke 1991) with *TRP1*-marked *CEN/ARS* plasmids expressing either untagged or 3XV5 tagged alanine-substituted *spt6* or wild-type *SPT6*. Complete plasmid shuffling was ensured by sequential passaging of transformants two times on SC-Trp + 5-FOA media. All plasmids are described in Table 22.

Table 21. Yeast strains used in Chapter 3

Strain ID	Genotype
KY303	*MATα his4-912∂ lys2-128∂ ura3-52 leu2Δ1*
KY600	*MATa his4-912∂ lys2-128∂ leu2Δ1 spt6-14*
KY1203	*MATa his3Δ200 leu2Δ1 ura3-52 trp1Δ63 arg4-12 cdc73Δ::KanMX rpb1Δ187::HIS3[prp112=RPB1 URA3 HA]*
KY1855	*MATα his4-912∂ leu2Δ1 ura3-52 trp1Δ63 cdc73Δ::KANMX*
KY1858	*MATa his4-912∂ trp1Δ63 cdc73Δ::KANMX*
KY3208	MATa his3Δ200::HIS3::osTIR lys2-128∂ leu2Δ1 ura3-52
KY3291	*MATa his3Δ200::HIS3-osTIR lys2-128∂ leu2Δ1 ura3-52 trp1Δ63 3XHSV-PAF1 SPT6-V5-IAA7(AID)-KANMX*
KY3343	*MATa his3Δ200::HIS3-osTIR lys2-128∂ leu2Δ1 ura3-52 ade8 3XHSV-PAF1*
KY3352	*MATa his3Δ200::HIS3-osTIR lys2-128∂ or lys2-173R leu2Δ1 ura3-52 trp1Δ63 ade8 spt6-50 3XHSV-PAF1*
KY2242	*MATα his4-912∂ lys2-128∂ trp1Δ63 cdc73Δ::KANMX*
KY3698	*MATα his4-912∂ lys2-128∂ leu2Δ1 ura3-52 trp1Δ63 spt6Δ::LEU2 PAF1 CDC73 pCC11 (SPT6 and URA3)*
KY3699	*MATα his4-912∂ lys2-128∂ leu2Δ1 ura3-52 trp1Δ63 spt6Δ::LEU2 3XHSV-PAF1 CDC73-HA::kanMX pCC11 (SPT6 and URA3)*
KY3642	*MATa leu2Δ1 trp1Δ63 3XHSV-PAF1 CDC73-HA::kanMX*
KY3638	*MATa lys2-128∂ leu2Δ1 trp1Δ63 ade8 3XHSV-PAF1 spt6-1004 CDC73-HA::kanMX*
KL01	*K. lactis* spike-in

3.2.2 Plasmid construction

All plasmid DNA was purified from *E.* coli DH5α cells grown at 37 ˚C in the presence of antibiotic selection using a GeneJET Plasmid Miniprep kit (Thermo Fisher, Waltham, MA, #K0503) following the manufacturer's instructions. Plasmid backbone was prepared by restriction

digest using NEB standard restriction enzymes (New England Biolabs, Ipswitch, MA) and inserts were prepared by PCR with Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher, Waltham, MA; #F-548S) followed by gel purification using UltraPure™ Low Melting Point Agarose (Thermo Fisher, Waltham, MA; #16520050) and the Promega Wizard™ Plus SV Minipreps DNA Purification System (Promega, Madison, WI; #PRA1465). All plasmids were constructed from purified DNA fragments using Gibson assembly using NEB Gibson Assembly® Master Mix (New England Biolabs, Ipswitch, MA; #E2611S) following the manufacturer's in instructions. Gibson assembly reaction products were transformed into DH5α competent cells. Transformants were purified, and pDNA was extracted and assessed by restriction digest followed by agarose gel electrophoresis. All plasmid DNA building blocks produced by PCR were checked using Sanger sequencing (GENEWIZ, Germany, Leipzig). A list of all plasmids used in this study is provided in Table 22.

Table 22. Plasmids used in Chapter 3

KB Number	Insert Description	*E.coli* Marker(s)	*S. cerevisiae* Marker(s)	*Type*
KB875	Cdc73-3xHA	Amp^R	*TRP1*	*Low copy*
KB890	Cdc73-W321A-3xHA	Amp^R	*TRP1*	*Low copy*
KB1027	*$ADH1_p$-EC Tyr* RS	Amp^R	*TRP1*	*High copy*
KB1273	*$ADH1_p$ cdc73-3XHSV*	Kan^R	*TRP1*	*High copy*
KB1274	*$ADH1_p$ cdc73-3XHSV W321Amber*	Kan^R	*TRP1*	*High copy*
KB1275	*$ADH1_p$ cdc73-3XHSV H319Amber*	Kan^R	*TRP1*	*High copy*
KB1276	*$ADH1_p$ cdc73-3XHSV R268Amber*	Kan^R	*TRP1*	*High copy*
KB1277	*$ADH1_p$ cdc73-3XHSV S272Amber*	Kan^R	*TRP1*	*High copy*
KB1278	*$ADH1_p$ cdc73-3XHSV R300Amber*	Kan^R	*TRP1*	*High copy*
KB1341	*rpb1::1461TEV*	Amp^R	*LEU2*	*Low copy*
KB1347	*$ADH1_p$ cdc73-3XHSV W321A*	Kan^R	*TRP1*	*High copy*
KB1349	*$ADH1_p$ cdc73-3XHSV W321A R300Amber*	Kan^R	*TRP1*	*High copy*
KB1359	*$ADH1_p$ cdc73-3XHSV*	Kan^R	*URA3*	*High copy*
KB1361	*$ADH1_p$ cdc73-3XHSV R268Amber*	Kan^R	*URA3*	*High copy*
KB1435	3XV5-IAA7-KanMX	Amp^R	*KanMX*	*Integrating*
KB1436	GPD1-osTIR-HIS3	Amp^R p	*HIS3*	*Integrating*
KB1453	10xHis-mRuby2-Spt6-239-1451	Kan		*Expression*
KB1482	10xHis-mClover-Cdc73	Kan		*Expression*
KB1504	10xHis-mClover-Cdc73-W321A	Kan		*Expression*
KB1573	Cdc73-10xHis	Amp^R		*Expression*
KB1617	Cdc73-PSVPN271-275AAAAA-3xHA	Amp^R	*TRP1*	*Low copy*
KB1618	Cdc73-VQSIK201-205AAAAA-3xHA	Amp^R	*TRP1*	*Low copy*
KB1619	Cdc73-NPRNL266-270AAAAA-3xHA	Amp^R	*TRP1*	*Low copy*

KB1623	3XV5-Spt6	Amp^R	*TRP1*	*Low copy*
KB1633	3XV5-Spt6ΔYqgF	Amp^R	*TRP1*	*Low copy*
KB1634	3XV5-Spt6ΔDLD	Amp^R	*TRP1*	*Low copy*
KB1635	3XV5-Spt6-DPK744-746AAA	Amp^R	*TRP1*	*Low copy*
KB1636	3XV5-Spt6-NKR771-773AAA	Amp^R	*TRP1*	*Low copy*
KB1637	3XV5-Spt6-DFI775-777AAA	Amp^R	*TRP1*	*Low copy*
KB1638	3XV5-Spt6-NKATD936-940AAAAA	Amp^R	*TRP1*	*Low copy*
KB1639	3XV5-Spt6-NEP970-972AAA	Amp^R	*TRP1*	*Low copy*
KB1640	3XV5-Spt6-SWNE1000-1003AAAA	Amp^R	*TRP1*	*Low copy*
KB1641	3XV5-Spt6-KRQK1004-1007AAAA	Amp^R	*TRP1*	*Low copy*
KB1642	3XV5-Spt6-YEDL1008-1011AAAA	Amp^R	*TRP1*	*Low copy*
KB1643	3XV5-Spt6-PEDY1023-1026AAAA	Amp^R	*TRP1*	*Low copy*
KB1644	3XV5-Spt6-ELLRE1058-1062AAAAA	Amp^R	*TRP1*	*Low copy*
KB1645	3XV5-Spt6-ELLRE1058-1062ALLAA	Amp^R	*TRP1*	*Low copy*

3.2.3 Serial dilution assay

Yeast cells were grown at 30 ˚C to saturation before generating the 10-fold dilution series starting at an OD_{600} = 1. Serial dilutions were pipetted in a 96-well plate to facilitate the use of a pinning tool. Spots were pinned to all experimental and control media using a pinning tool (Sigma-Aldrich, St. Louis, MO, R2383-1EA). All spot test plates were incubated at 30 ˚C for at least three days

3.2.4 Yeast protein extract preparation

All protein extracts used in western analysis were prepared by the TCA method (Cox *et al.* 1997). Before protein extraction, 10 OD_{600} units of log-phase yeast culture were harvested and stored at -80 ˚C. Cells were resuspended in 20% TCA, lysed by vortexing (two pulses on setting 10; 1 min on vortex, 1 min on ice) in the presence of glass beads. The lysate was diluted with 5% TCA and protein was pelleted by centrifugation at 6000 rpm for 10 minutes. The resulting pellet

was washed with 0.5 M Tris-Cl pH 8.0 and resuspended in a 1:1 ratio of 0.5 M Tris-Cl pH 8.0 to 3X protein loading dye (30 % glycerol, 2.15 M BME, 6 % SDS, 187.5 mM Tris pH 6.8, 0.3 % bromophenol blue).

3.2.5 Western blotting

Protein samples obtained by TCA extraction were resolved by SDS PAGE gels before transferring to a nitrocellulose membrane. Blocking was carried out for 2 hr at room temperature in TBST containing 5% milk. Primary antibody incubation with α-G6PDH (Sigma, A-9521, 1:50,000), α-Spt6 (gift from Tim Formosa, 1:1000), α-V5 (Invitrogen, R960-25, 1:1000), α-Spt5 (gift from Grant Hartzog, 1:1000 dilution), α-HSV (Sigma-Aldrich, H6030, 1:1000), α-HA (Roche, 11666606001, 1:1000), α-Rtf1 (1:5,000 dilution), α-Rpb3 (BioLegend, 665004, 1:1000), α-His (Abcam, ab18184, 1:1000) was carried out overnight at 4 ˚C with agitation. Incubation with secondary antibody was carried out at room temperature for 1 hr. Visualization of protein targets was achieved using Pico Plus chemiluminescence substrate (Thermo Fisher, Waltham, MA, #34580) and the ChemiDoc XRS imaging platform (BioRad, Hercules, CA). Quantifications were performed using ImageJ.

3.2.6 *In vivo* BPA crosslinking

All BPA (p-benzoyl-L-phenylalanine) crosslinking experiments were performed as described in (Van Oss *et al.* 2016; Cucinotta *et al.* 2019). Briefly, log-phase yeast cultures were

co-transformed with the tRNA/tRNA synthetase plasmid for BPA incorporation, pLH157/LEU2, and containing a TRP1-marked plasmid harboring *CDC73* with an amber codon at a position of interest to allow for BPA incorporation by nonsense suppression (Chin *et al.* 2002). Yeast cells were grown in SC-L-W media containing 1 mM BPA (Bachem, F-2800) to log phase, and 10 OD units of cells were harvested for protein extraction. Harvested cells were resuspended in 1 mL of ddH_2O and UV irradiated (365 nm) for 10 minutes using a UVG-55 handheld UV lamp (UVP) with the lamp kept at a distance of 2 cm above the cells during crosslinking. Protein extracts were prepared from both UV exposed and unexposed samples using the TCA method described above. Western blot analysis was performed to identify super shifted (crosslinked) bands.

3.2.7 Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation, described in detail here, was performed as described previously (Kuras and Struhl 1999; Komarnitsky *et al.* 2000; Shirra *et al.* 2005) with slight modifications. Cell cultures at a volume of 200 or 250mL were grown to log phase (OD_{600} = 0.8 - 1.0) for all ChIP-seq and ChIP-qPCR experiments. Before harvest, cells were crosslinked for 20 minutes at room temperature by addition of formaldehyde at a concentration of 1% (v/v). Crosslinking was stopped by the addition of a quenching solution (3 M glycine, 20 mM Tris) and incubation for five minutes at room temperature. From the resulting mixture, 250 mL was collected and spun at 3500 rpm for 5 minutes at 4˚C before washing twice with at least 100 mL of cold TBS (20 mM Tris pH 7.5, 150 mM NaCl). Washed cell pellets were resuspended in 5 mL of cold 1 X FA (50 mM HEPES-KOH at pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na

deoxycholate) containing 0.1% SDS and 1 mM PMSF. After washing, all cells were pelleted by centrifugation, flash-frozen, and stored at -80˚C.

Cell pellets were thawed on ice, resuspended with cold 1xFA + 0.5% SDS + 1 mM PMSF, and mixed with 1.5 mL of glass beads. The lysis was carried out by vortexing the mixture for 15 minutes at max speed (30 seconds on vortex, 30 seconds on ice). Lysate was purified away from beads and spun at 50,000 rpm for 20 minutes at 4˚C. Supernatant was discarded before resuspending the chromatin pellet in 1xFA + 0.1% SDS + 1 mM PMSF and centrifuging at 50,000 rpm for 20 minutes at 4˚C. The supernatant was again discarded, and the resulting pellet was resuspended in 2 mL of 1xFA + 0.1% SDS + 1 mM PMSF before shearing. Chromatin was sheared in Bioruptor® Pico 15 mL tubes with 300 µL beads (Diagenode, Denville, NJ, #C30010017) by 25 cycles of 30 sec on and 30 sec off using a Bioruptor® Pico (Diagenode, Denville, NJ, #B01060010). Sonicated chromatin was brought up to a volume of 6mL and centrifuged for at 50,000 rpm for 20 minutes at 4˚C. Chromatin was aliquoted, flash-frozen, and stored at -80 ˚C. Sonication and chromatin yield was assessed by agarose gel (2%) electrophoresis after pronase treatment (800 µg/mL pronase at 42 ˚C for 1 hr), crosslink reversal (65 ˚C overnight incubation) and phenol-chloroform extraction.

Immunoprecipitation (IP) was carried out after thawing chromatin on ice and bringing the NaCl concentration to 275 mM by the addition of 5 M NaCl. The antibody was added (α-Spt6 (gift from Tim Formosa, 2.5 µL in 350 µL of chromatin), α-V5 (Invitrogen, R960-25, 1.5 µL in 350 µL), α-Spt5 (gift from Grant Hartzog, 0.5 µL in 350 µL), α-HSV (Sigma-Aldrich, H6030, 1.25 µL in 350 µL), α-HA (Santa Cruz, HA-probe Antibody (F-7) AC, #sc-7392 AC, 15 µL (bead-conjugated antibody) in 350 µL), α-Rtf1 (1 µL in 350 µL; (Squazzo *et al.* 2002)), α-Rpb3 (BioLegend, 665004, 1.25 µL in 350 µL), α-Myc (gift from John Woolford, used as non-specific

IgG control, 2.5 µL in 350 µL)) to each IP sample and allowed to incubate overnight at 4 ˚C on an end over end roller. Protein A or G beads were washed in 1xFA + 0.1% SDS + 275 mM NaCl + 1 mM PMSF and added to the chromatin-antibody mixture and placed on the roller for 1 hr at room temperature to affinity purify primary antibodies and bound factors from total chromatin. Beads were washed once for 4 minutes with each of the following buffers 1) 1xFA + 0.1% SDS + 275 mM NaCl + 1 mM PMSF, 2) 1xFA + 0.1% SDS + 500 mM NaCl + PMSF, 3) TLNNE (10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1mM EDTA, 0.5% N-P40, 0.5% sodium deoxycholate) + 1 mM PMSF and 4) TE before elution from beads, pronase treatment, crosslink reversal and DNA purification using a QIAquick PCR Purification Kit (Qiagen, Hilden, Germany, #28106).

3.2.8 Quantitative polymerase chain reaction (qPCR)

All qPCR experiments were performed in biological triplicate and technical duplicate. Reactions were prepared in a volume of 20 µL using qPCRBIO SyGreen Blue 2x reaction mix (Genesee Scientific, El Cajon, CA; #17-505B) following the manufacturer's instructions. Each reaction was split into two reactions of equal volume (technical replicates) and analyzed on a QuantStudio3™ Real-Time PCR System (Thermo Fisher, Waltham, MA) beginning with a hold at 95°C for 10 min followed by 40 cycles of 95°C for 15 sec and 58°C for 1 min and finally terminating with the generation of a melt curve.

Efficiencies were determined for all primer sets by measuring C_t values across a series of six ten-fold dilutions of *S. cerevisiae* genomic DNA. ChIP-qPCR data were analyzed using the following mathematical formula.

$$Relative\ Occupancy = \frac{E_{ROI}^{IP_{ROI} - Input_{ROI}}}{E_{TELVI}^{IP_{TELVI} - Input_{TELVI}}}$$

$$E = Primer\ Efficiency$$

$$IP_{ROI} = IP\ C_t\ Region\ of\ Intrest$$

$$Input_{ROI} = Input\ C_t\ for\ Region\ of\ Intrest$$

$$IP_{TELVI} = IP\ C_t\ for\ a\ region\ of\ Telomere\ VI$$

$$Input_{TELVI} = Input\ C_t\ for\ a\ region\ of\ Telomere\ VI$$

All qPCR primers and their efficiencies are listed in Table 23.

Table 23. Primers used for qPCR in Chapter 3 listed with their efficencies and target regions

Primer	Target	Sequence (5' to 3')	Efficiency
APO95	*TEL VI* FWD	TGCAAGCGTAACAAAGCCATA	2.04
APO96	*TEL VI* REV	TCCGAACGCTATTCCAGAAAG	
ECO234	*PMA1* 5' FWD	GCTAGACCAGTTCCAGAAGAATATTTACA	1.97
ECO235	*PMA1* 5' FWD	CAGCCATTTGATTCAAACCGTA	
ECO236	*PMA1* 3' FWD	GAAATCTTCTTGGGTCTATGGATTG	1.94
ECO237	*PMA1* 3' FWD	CAACATCAGCGAAAATAGCGAT	
ECO238	*PYK1* 5' FWD	ACCAAGGGTCCAGAAATCAGAAC	1.98
ECO239	*PYK1* 5' FWD	TGTCATCGGTGGTGAAGATCAT	
ECO240	*PYK1* 3' FWD	AGAAACTGTACTCCAAAGCCAACCT	1.99
ECO241	*PYK1* 3' FWD	TGGTCTGTACTTGGAAACCAATCTT	
MEO185	*YPS34* 3' FWD	CTTCGAACATACCTGATATCAGAATAG	1.90
MEO186	*YPS34* 3' REV	GCACTGTGGCATCTTCTTCGGAC	
MEO183	*YPS34* 5' FWD	GTGTCTCACAGGATCTGGATGTTCC	1.91
MEO184	*YPS34* 5' REV	GAGATGGCTTCAACAGTGGCTTATG	
MEO189	*RAS1* 3' FWD	CGAGAAGTAAACAGTCTGCTGAG	1.92
MEO190	*RAS1* 3' REV	CAAATTATACAACAACCACCACTAG	
MEO187	*RAS1* 5' FWD	GCAGGGAAATAAATCAACTATAAGAG	2.00
MEO188	*RAS1* 5' REV	GATAGTAGGGTCATATTCGTCCAC	

3.2.9 Next-generation sequencing library preparation

All ChIP-sequencing libraries were prepared using a NEBNext Ultra II kit and indices (New England Biolabs, Ipswitch, MA; E7645, E7335, E7500, E7710, E7730) following manufacturer's instructions. Immunoprecipitated DNA was input into all library build reactions at

0.9 ng/µL along with 0.1 ng/µL of spike-in DNA from *K. lactis*. Please note that unfortunately using a spike-in control in this manner does not allow for the detection of global changes, but rather controls for library build and sequencing effects, which means I can only observe relative changes and not absolute differences. Libraries were quantified by Qubit and assessed by PCR and agarose gel electrophoresis for proper fragment size. All next-generation sequencing was conducted by the Health Sciences Sequencing Core at the University of Pittsburgh Medical Center's Children's Hospital on an Illumina NextSeq 500.

3.2.10 Recombinant protein expression

Proteins were expressed in *Escherichia coli* CodonPlus RIPL cells using the IPTG induction method. Cells were grown at 37 ˚C to an OD_{600} of 0.6-0.8 before the addition of 150 µM isopropylthio-beta-galactosidase (IPTG). After overnight IPTG induction at 25 ˚C, cells were harvested by centrifugation at 6000 rpm for 10 minutes. For Cdc73 expression, 2 L of cells were sufficient, but for Spt6 4 L were required. Harvested cells were stored at -80 ˚C. All plasmids are described in Table 22.

3.2.11 Recombinant protein purification

Cells were thawed and lysed by either homogenization or sonication in the presence of a protease inhibitor cocktail consisting of PMSF, aprotinin, leupeptin, and pepstatin. Lysates were centrifuged at 14000 rpm for 30 minutes at 4 ˚C to isolate soluble protein. A combination of nickel

affinity chromatography and cation exchange chromatography was used to purify 10xHis-mClover-Cdc73 and Cdc73-10xHis full length and mutant fusion proteins. Untagged Cdc73 full length and truncated or mutated proteins were purified starting from 10xHis-mRuby2-Cdc73, 10xHis-mClover-Cdc73 or 6xHis-Cdc73 by nickel exchange chromatography followed by TEV protease digest and cation exchange chromatography. The 10xHis-mClover-Spt6 and 10xHis-mRuby2-Spt6 fusion proteins were purified by a combination of nickel exchange chromatography and heparin affinity chromatography. Untagged Spt6 fusion proteins were purified starting from 10xHis-mRuby2-Spt6 or 10xHis-mClover-Spt6 by nickel exchange chromatography followed by TEV protease digest and heparin affinity chromatography. Spt6-239-1451 used in crosslinking and mass spectrometry experiments was purified in the same way as the tagged Spt6 protein with the exception that a size exclusion chromatography was performed before heparin affinity chromatography. All protein purifications were carried out in a buffer consisting of 25 mM Tris-Cl pH 8.0, 10% glycerol, 1 mM BME, and varying concentrations of NaCl and imidazole at 4˚C. All proteins were concentrated and exchanged into binding buffer consisting of 50 mM HEPES, 100 mM KOAc, 2 mM MgOAc, 10% glycerol, and 1 mM BME.

Protein purity was assessed by SDS PAGE and coomassie blue staining. A Nanodrop was used to assess protein concentration for proteins that were not tagged with a fluorophore. Protein concentration for fluorescently tagged proteins and any protein used for a quantitative binding assay was determined using a lysozyme or BSA standard curve. This quantification was performed by running a dilution series of lysozyme or BSA on a denaturing gel along with proteins to be quantified. Bands were, quantified using ImageJ and the lysozyme or BSA data were fit to a linear equation ($y = mx + b$), which was used to calculate concentrations of the experimental samples.

3.2.12 *In vitro* DSS and EDC crosslinking

DSS and EDC crosslinking was conducted essentially as previously described (Shi *et al.* 2014, 2015; Fernandez-Martinez *et al.* 2016) with some modifications. All reactions were prepared at a final volume of 20 µL. Protein samples were prepared for crosslinking by adding 10 µM 10xHis-mClover-Cdc73 and 2 µM Spt6 239-1451 to either DSS crosslinking buffer/binding buffer (50 mM HEPES, 100 mM KOAc, 2 mM MgOAc, 10% glycerol and 1 mM BME) or EDC crosslinking buffer (100 mM MES pH 6, 100 mM NaCl, 5 mM $MgCl_2$, and 10% Glycerol). For DSS crosslinking, DSS was dissolved in 100% DMSO and added to each reaction at a final concentration of 625 µM (1 µL of 12.5mM stock) and uncrosslinked controls were treated with 1 µL of 100% DMSO. For EDC crosslinking, EDC and NHS were dissolved in sterile water and added to each reaction at a final concentration of 12.5 mM (EDC, 1 µL of 250 mM) and 250 µM (NHS, 1 µL of 5 mM), respectively, to achieve an EDC:NHS ratio of 50:1. Uncrosslinked control reactions for EDC samples were prepared by addition of 1 µL of sterile water and 1 µL of NHS. All reactions were incubated with agitation for 30 minutes at room temperature. Reactions were quenched by addition of 1 µL of 1M Tris-Cl pH 7.5 (DSS) or addition of both 1 µL of 1M Tris-Cl pH 8 and 1 µL of 1M BME (EDC) and incubation at room temperature with agitation for 15 minutes.

Products from DSS or EDC crosslinking reactions prepared for mass spectrometry analysis were combined with 7 µL of 3X SDS PAGE loading dye lacking reducing agent, brought to a final concentration of 20 mM DTT by addition of 0.56 µL of 1 M DTT and incubated at 75 ˚C for 10 minutes. Samples were allowed to cool to room temperature for 5 minutes before addition of 1.43 µL of iodoacetamide (IAA, Sigma, A3221-10VL) dissolved in 80% distilled deionized water (ddH2O)/milliQ H2O and 20% Acetonitrile (ACN, Thermo, 85188) bringing the final

concentration to 50 mM (2.5X that of DTT). Samples were incubated at room temperature for 30 minutes in the dark before running 20 µL of the reaction on an SDS PAGE gel (4%-12% Bis/Tris; Novogene). Crosslinked products excised from the gel were distained and analyzed by mass spectrometry.

For western analysis of crosslinked products 5 µL of SDS PAGE loading dye was added to the 20µL reaction, and 5 µL of the resulting mixture was run on an SDS PAGE gel (4%-12% Bis/Tris; Novogene) and transferred to nitrocellulose before Ponceau staining and Western blotting with the indicated antibodies.

3.2.13 Mass spectrometry analysis

All mass spectrometry methods described in this section were conducted by either Dr. Fei Fang or Yufei Xiang in under the supervision of Dr. Yi Shi at the University of Pittsburgh medical school. In-gel digestion was carried out with trypsin and Lys-C as previously described (Shi *et al.* 2014, 2015). The peptide mixtures were desalted and analyzed with a nano-LC 1200 (Thermo Fisher, Waltham, MA) coupled to a Q Exactive™ HF-X Hybrid Quadrupole-Orbitrap™ mass spectrometer (Thermo Fisher, Waltham, MA). Cross-linked peptides were loaded onto a picochip column (C18, 3 µm particle size, 300 Å pore size, 50 µm × 10.5 cm; New Objective) and eluted using a 60 min LC gradient : 5% B–8% B, 0 – 5 min; 8% B – 32% B, 5 – 45 min; 32% B–100% B, 45 – 49 min; 100% B, 49 - 54 min; 100% B - 5 % B, 54 min - 54 min 10 sec; 5% B, 54 min 10 sec - 60 min 10 sec; mobile phase A consisted of 0.1% formic acid (FA), and mobile phase B consisted of 0.1% FA in 80% acetonitrile. The QE HF-X instrument was operated in the data-dependent mode, where the top 8 most abundant ions (mass range 380–2,000, charge state 3 - 7)

were fragmented by high-energy collisional dissociation (normalized collision energy 27). The target resolution was 120,000 for MS and 15,000 for MS/MS analyses. The quadrupole isolation window was 1.8 Th and the maximum injection time for MS/MS was set at 120 ms. The MS data was searched using pLink2 to identify cross-linked peptides (Chen *et al.* 2019). The mass accuracy was specified as 10 and 20 p.p.m. for MS and MS/MS, respectively. Other search parameters included cysteine carbamidomethylation as a fixed modification and methionine oxidation as a variable modification. A maximum of three trypsin missed-cleavage sites was allowed. Initial search results were obtained using the default 5% false discovery rate. All crosslink spectra were manually inspected to remove potential false-positive identifications essentially as previously described (Shi *et al.* 2014, 2015).

3.2.14 Fluorescent gel-shift assay

In these assays, a standard amount of 150nM 10xHis-mClover-Cdc73 was added to each binding reaction, and 10xHis-mRuby2-Spt6 (239-1451) was titrated into the reaction. The 10xHis-mRuby2-Spt6 (239-1451) titrations began as low as 2 nM and ended as high as 9 μM. 10xHis-mClover-Cdc73 and mRuby2-Spt6 (239-1451) were mixed with binding buffer (50 mM HEPES, 100 mM KOAc, 2 mM MgOAc, 10% glycerol and 1 mM BME) in an Eppendorf tube and given 10 minutes to bind at room temperature before being loaded into the gel. Agarose gels (0.5%) were run in Native PAGE buffer 120V for 1.5 hrs in a 4 ˚C cold room and imaged on an Amersham imager. The gel box and buffer were allowed to equilibrate to 4 ˚C for at least 1 hr before gels were loaded. Native gel-shift data were quantified by measuring the intensity of 10xHis-mClover-Cdc73 (green band). Band intensity values were measured from the raw TIFF files produced by

the imager using LiCor ImagStudioLite. Fraction shifted was calculated using the following equation:

$$fraction\ shifted\ = \frac{shifted\ band\ intensity}{shifted\ band\ intensity\ +\ unshifted\ band\ intensity}$$

Curves were fitted to native gel-shift data using Prism 8 graphing software and the following equation:

$$Y = \frac{(Bmax * X)}{(Kd + X) + baseline}$$

Bmax, Kd, and baseline were optimized for curve fitting.

3.2.15 Fluorescence anisotropy binding assay

For all anisotropy experiments, reactions contained 100nM 10xHis-mClover-Cdc73 and increasing amounts of Spt6 (239-1451). Spt6 (239-1451) titrations began at 750 pM and ended at 7.5 μM. 10xHis-mClover-Cdc73 and Spt6 (239-1451) were mixed in binding buffer (50 mM HEPES, 100 mM KOAc, 2 mM MgOAc, 10% glycerol and 1 mM BME) in black Nunc 384-Well Polystyrene microplates (Thermo Fisher, Waltham, MA, #262360) at a final reaction volume of 80 uL. Data were collected on a Cytation 5 plate reader using a green fluorescent polarization filter (specifications: 5 mP standard deviation at 1 nM sodium fluorescein, excitation 485/20 nm,

emission 528/20 nm, 510 nm mirror, excitation range: 400 to 700 nm, emission range: 400 to 700 nm) and Gen5 software (BioTek, Winooski, VT). Change in anisotropy was calculated as follows:

$$\Delta Anisotropy = \frac{Parallel - Perpendicular}{Parallel + 2 * Perpendicular}$$

Curve fitting was done in Prism 8 using the procedure that was used to analyze the fluorescent gel-shift assay data, above.

3.2.16 Determining a structural model from XL-MS data

Initial crosslinking data visualization in two dimensions was accomplished using xiView and Prism8 to generate a network model and x-y scatterplot, respectively. The generation of a three-dimensional model was accomplished by taking the following steps. A model of full-length Cdc73 was predicted using the I-TASSER threading algorithm (Zhang 2008; Roy *et al.* 2010; Yang *et al.* 2014) and validated by two metrics 1) alignment to various Cdc73 structures (3V46 (Amrich *et al.* 2012), 5YDE (Sun *et al.* 2017) and 6AF0 (Deng *et al.* 2018)) and 2) the number of valid crosslink lengths in our XL-MS data. The best I-TASSER (Zhang 2008; Roy *et al.* 2010; Yang *et al.* 2014) model for full-length Cdc73 was chosen based off comparisons to known crystal structures and a Phyre2 structural model (Figure S6A) and the percentage of crosslinks, from a separate XL-MS experiment using Cdc73-6xHis, that satisfied (Figure 6B) or violated (Figure 6C) a 30 Å cutoff. This model of *Saccharomyces cerevisiae* Cdc73 was then used along with Spt6 structures (3PSF, 3PSI, and 3PSK (Close *et al.* 2011)). Gaps in the Spt6 core structure were filled in before modeling by combining information from 3PSF and 3PSI structures after 3D structural

alignment using the FATCAT server (Ye and Godzik 2003, 2004a; b; Li 2006) to fill in gaps. The tSH2 domains of Spt6 were included in the modeling by including 3PSK (Close *et al.* 2011).

EDC and DSS crosslinking data were incorporated into the model using the Integrative Modelling Platform (IMP) (Russel *et al.* 2012) using a crosslinking distance of 20 Å for EDC and 35 Å for DSS. Only intra-protein crosslinks between the Spt6 core and tSH2 domain and inter-protein crosslinks were considered in the modeling, meaning intra-protein crosslinks within Cdc73 and Spt6 (aside from contacts between the core and tSH2 domain) were excluded (compare tables 26 and 27). During modeling, the position of the Spt6 core was held constant, and Cdc73 and the tSH2 domain structures were allowed to move by specifying them as rigid bodies in the modeling software. In general, the settings used in IMP were the same as shown in the authors' Pol II tutorial (https://integrativemodeling.org/2.4.0/doc/tutorial/rnapolii_stalk.html) (Russel *et al.* 2012).

Combined with our DSS and EDC crosslinking data and the input structures, both empirically determined (from PDB) and predicted (from I- TASSER), a model of the Cdc73-Spt6 interaction was generated. The model presented here was generated by allowing the IMPs Monte-Carlo simulation to run for 100,000 iterations with 10 steps per iteration (1,000,000 steps total). The top 4 models were combined by clustering using IMP software. The top IMP model was assessed using the XlinkAnalyzer (Kosinski *et al.* 2015) plugin in Chimera (Pettersen *et al.* 2004) in order to assess crosslink distance between residues in both proteins. Structures used as input for IMP were aligned to the carbon alpha trace (output by IMP) for the highest scoring to recover side-chain information. This step was performed using the alignment plugin in PyMOL (Schrödinger, LLC 2015). Only data from the 3V46 structure of Cdc73 were included in the final model because the C-domain is the only portion of Cdc73 for which we have rich structural information in *S. cerevisiae* (Amrich *et al.* 2012).

Additionally, our Cdc73-Spt6 interaction model was aligned to the structure of the human Pol II elongation complex structure (6GMH) (Vos *et al.* 2018b) to determine the position of the Cdc73 C-domain in the Pol II elongation complex. This alignment was accomplished by using the alignment plugin in PyMOL to align the human Spt6 core to the yeast Spt6 core. All structural information presented here was visualized in either Chimera (Pettersen *et al.* 2004) or PyMOL (Schrödinger, LLC 2015).

3.2.17 ChIP-seq data analysis

ChIP-seq reads were aligned to the *S. cerevisiae* genome (Ensembl R64-1-1) or *K. lactis* genome (Ensembl ASM251v1), using HISAT2 (Kim *et al.* 2015) (options –no-mixed –no-discordant –no-unal -k 1) before low quality read filtering and SAM to BAM conversion with the SAMtools suite (Li *et al.* 2009) (options: view -bS -q30). The resulting BAM files were used as input to determine read counts for *S. cerevisiae* and *K. lactis* using BAMtools (Barnett *et al.* 2011). The deepTools2 bamCoverage (Ramírez *et al.* 2014, 2016) command (options: --scaleFactor ICPHT --binSize 1 --ignoreDuplicates –extendReads; ICPHT is named for Inverse Counts Per Hundred Thousand) was used to apply the spike-in normalization using the method described in (Orlando *et al.* 2014).

$$ICPHT = \frac{1}{spike\ in\ read\ count/100{,}000}$$

Generation of bigWig files for browser tracks, and count matrices for graphing and statistical analysis was performed in deepTools2 and R Studio (Team 2016). For all ChIPseq samples, biological replicate BigWig files were averaged using the bigwigCompare command (options: --pseudocount 0.1 --operation mean --binSize 1) and log2 fold change BigWigs were

generated from these using the bigwigCompare command (options: --operation log2 --binSize 1). Heatmaps and aggregation plots were plotted either directly in deeptools using a combination of the computeMatrix, and either the plotHeatmap or plotProfile command or in Prism8 using data exported from these deeptools commands. For heatmap and aggregation plot analyses, the computeMatrix command was used to plot data from bigWig files containing spike-in normalized read counts or log2 fold change values over genomic regions specified from a BED file using a bin size of 25 bp and averaging data within each bin (Ramírez *et al.* 2014, 2016). All analysis code can be found at https://github.com/mae92/ChIP-seq_Analysis_Code.

3.2.18 Statistical analysis and reproducibility

A least three biological replicates were performed for all ChIP-qPCR experiments except for the Spt6 and Cdc73 binding interface mutant screen. At least three biological replicates were performed for all spot tests shown and the AID Western blot and viability assay experiments. At least two biological replicates were performed for the Spt6 and Cdc73 binding interface mutant screen, Western blots to confirm protein stability, and ChIP-seq experiments shown in this manuscript. For *in vivo* work, each biological replicate is a pure yeast culture derived from a single colony initiated from a single cell of a given strain. For Spt6 binding interface mutants, where plasmid shuffle was used to introduce alleles, three independent shuffle strains were created. An independent colony from an independent shuffle was assayed for each biological replicate up to three. Where we show more than three replicates, we used additional single colonies from one or more of the independently shuffled strains. All biochemical assays are repeated at least three times. Crosslinking and mass spectrometry were performed in duplicate for DSS and EDC crosslinkers for a total of 4 crosslinking datasets going into the analysis and model generation. All bar graphs

plot mean and standard error of the mean and include all individual data points. For ChIP-qPCR data, all p values were generated using an unpaired, two-sided, Student's t-test, assuming equal variance carried out between the mutant strain and the wild-type strain. For ChIP-seq correlation analysis, Pearson's correlation was used to assess replicate reproducibility and correlation between datasets.

3.2.19 Data availability

Strains and plasmids are available upon request. ChIP-seq data will be deposited in the Gene Expression Omnibus database under upon publication of this work and assigned a GSE accession number. The code used for the analysis of ChIP-seq data has been uploaded to the following GitHub repository (https://github.com/mae92/ChIP-seq_Analysis_Code).

3.3 Results

3.3.1 Cdc73 directly interacts with the Rpb1 CTD and Spt6 *in vivo*

To identify interaction partners for Cdc73 *in vivo*, we employed a site-specific protein crosslinking strategy in which we replaced conserved, surface-exposed amino acids on the Cdc73 C-domain (Figure 19A and 19B) with the photoreactive phenylalanine analog (BPA) through amber codon suppression. Based on the levels of full-length Cdc73 protein produced in cells grown in BPA-containing medium, BPA was incorporated at positions 268, 272, 300 or 321 while incorporation at position 319 was unsuccessful. Upon exposure of cells to UV radiation,

crosslinked products were produced, and these were detected by western blotting using an antibody against the HSV tag on Cdc73 (Figure 19C, lanes 6, 9, 12, and 18). Further resolution of the crosslinked species revealed two slowly migrating, HSV-reactive bands in the R300BPA sample. Western analysis identified the top band as a crosslinked product between Cdc73 and the largest Pol II subunit, Rpb1 (Figure 19D, lanes 4 and 9; blue arrow), and the bottom band as Cdc73 crosslinked to Spt6 (Figure 19D, lane 9; red arrow and Figure 19E, lanes 3 and 7; red arrow). To further investigate the Cdc73-Rpb1 interaction, we performed UV-crosslinking experiments on yeast cells expressing Cdc73 R300BPA and a form of Rpb1 in which a TEV protease cleavage site is inserted between the Rpb1 core domain and the CTD. Upon treatment of the extracts with TEV protease the Cdc73-Rpb1 crosslinked product (Figure 19F, green arrow) shifted to a position on the gel that was inconsistent with crosslinking to the Rpb1 core domain (Figure 19F, lanes 3 and 5). Moreover, the TEV-cleaved product could not be detected with an antibody against the Rpb1 core domain but could be detected with the 8WG16 antibody, which recognizes the Pol II CTD (Figure 19F, lanes 1 and 5). Together, these results demonstrate a direct interaction of Cdc73 with two key components of the Pol II elongation machinery, the Pol II CTD and the essential transcription elongation factor and histone chaperone Spt6 (Duina 2011).

To investigate the functional significance of the Cdc73-Rpb1 and Cdc73-Spt6 interactions, we performed BPA crosslinking experiments with a Cdc73 R300BPA derivative in which the highly conserved tryptophan at position 321 was changed to alanine. We previously showed that the Cdc73 W321A substitution causes phenotypes associated with defects in transcription elongation (Amrich *et al.* 2012). Interestingly, crosslinking of both Rpb1 and Spt6 to Cdc73 R300BPA was greatly diminished by the W321A substitution suggesting that these interactions are strongly dependent on a transcriptionally important residue within Cdc73 (Figure 19E)

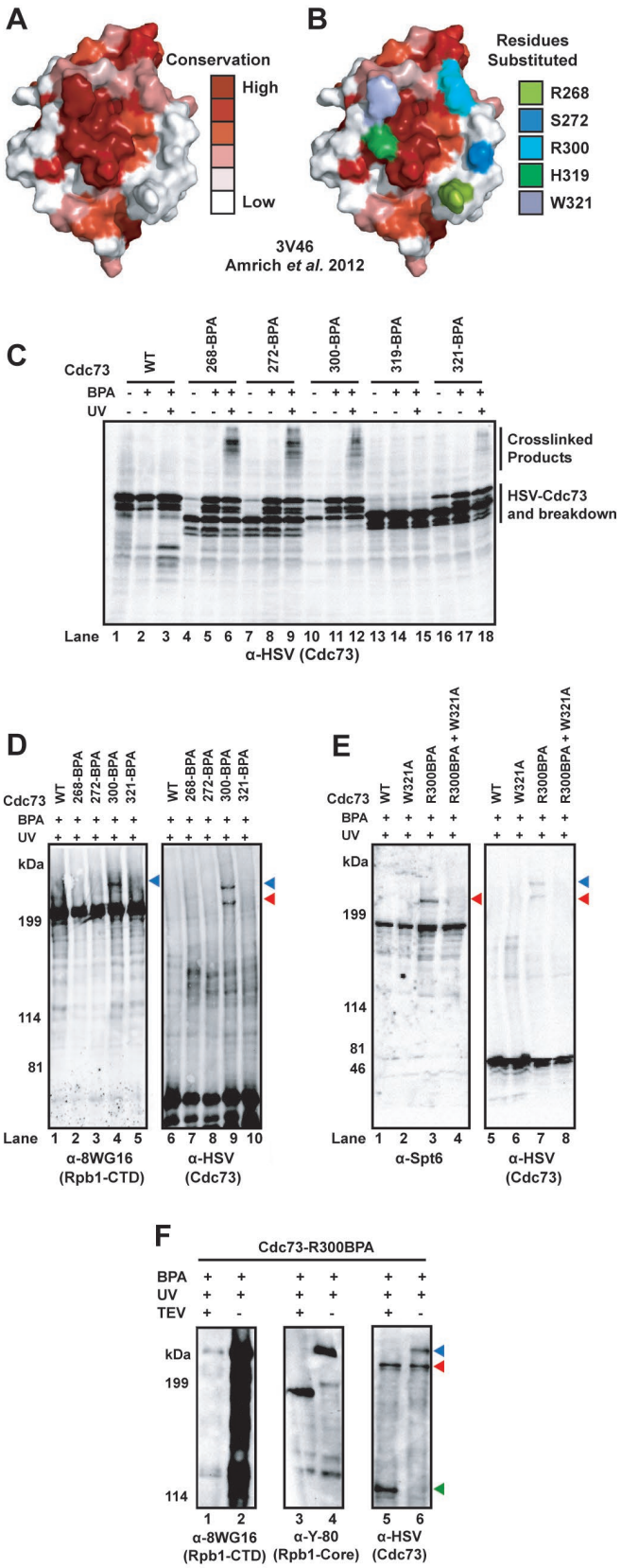

A
Conservation
High
Low
B
Residues Substituted
R268
S272
R300
H319
W321
3V46
Amrich et al. 2012
C
Cdc73
WT
268-BPA
272-BPA
300-BPA
319-BPA
321-BPA
BPA
UV
Crosslinked Products
HSV-Cdc73 and breakdown
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
α-HSV (Cdc73)
D
Cdc73
WT
268-BPA
272-BPA
300-BPA
321-BPA
BPA
UV
kDa
199
114
81
Lane 1 2 3 4 5 6 7 8 9 10
α-8WG16 (Rpb1-CTD)
α-HSV (Cdc73)
E
Cdc73
WT
W321A
R300BPA
R300BPA + W321A
BPA
UV
kDa
199
114
81
46
Lane 1 2 3 4 5 6 7 8
α-Spt6
α-HSV (Cdc73)
F
Cdc73-R300BPA
BPA
UV
TEV
kDa
199
114
1 2 3 4 5 6
α-8WG16 (Rpb1-CTD)
α-Y-80 (Rpb1-Core)
α-HSV (Cdc73)

Figure 19. Cdc73 interacts directly with Spt6 *in vivo*.

A) Cdc73 C-domain crystal structure (3V46) with conservation mapped using the Consurf server. B) As in A except showing the locations of residues substituted with BPA in the experiments in C and D. C-F) Western analysis of protein extracts from BPA crosslinking experiments C) Assessment of optimal location for BPA incorporation. D) Cdc73-Rpb1 interaction captured by BPA crosslinking. E) Cdc73-Spt6 interaction captured by BPA crosslinking. In D and E, adjacent panels are derived from the same gel. F) Cdc73-CTD interaction captured by BPA crosslinking following TEV cleavage of the CTD from the body of Rpb1. Adjacent panels are derived from the same gel. Red arrows indicate the Spt6-Cdc73, blue arrows indicate the Rpb1-Cdc73 crosslinked product, and the green arrow indicates the Cdc73-CTD crosslinked product.

We previously showed that deletion of the Cdc73 C-domain reduces Paf1C occupancy on transcribed genes (Amrich *et al.* 2012).To test the importance of W321 and the position of BPA-crosslinking, R300, in mediating Cdc73 occupancy on chromatin, we performed ChIP-qPCR. These experiments revealed a slight reduction in Cdc73 occupancy (Figure 20A) at the highly transcribed *PMA1* gene in both the Cdc73 R300A and W321A mutants. However, when the data are presented as the ratio of Cdc73 occupancy (Figure 20A) to Spt6 Occupancy (Figure 20B), a decrease in Cdc73 occupancy relative to Spt6 was only observed for the Cdc73 W321A mutant (Figure 20C). The partial reduction in Cdc73 W321A occupancy suggests that Cdc73 remains associated with the elongation machinery either through a partially functional C-domain or through other subunits within Paf1C. Collectively, these results identify physically and functionally important interactions between Cdc73 C-domain and both Rpb1 and Spt6.

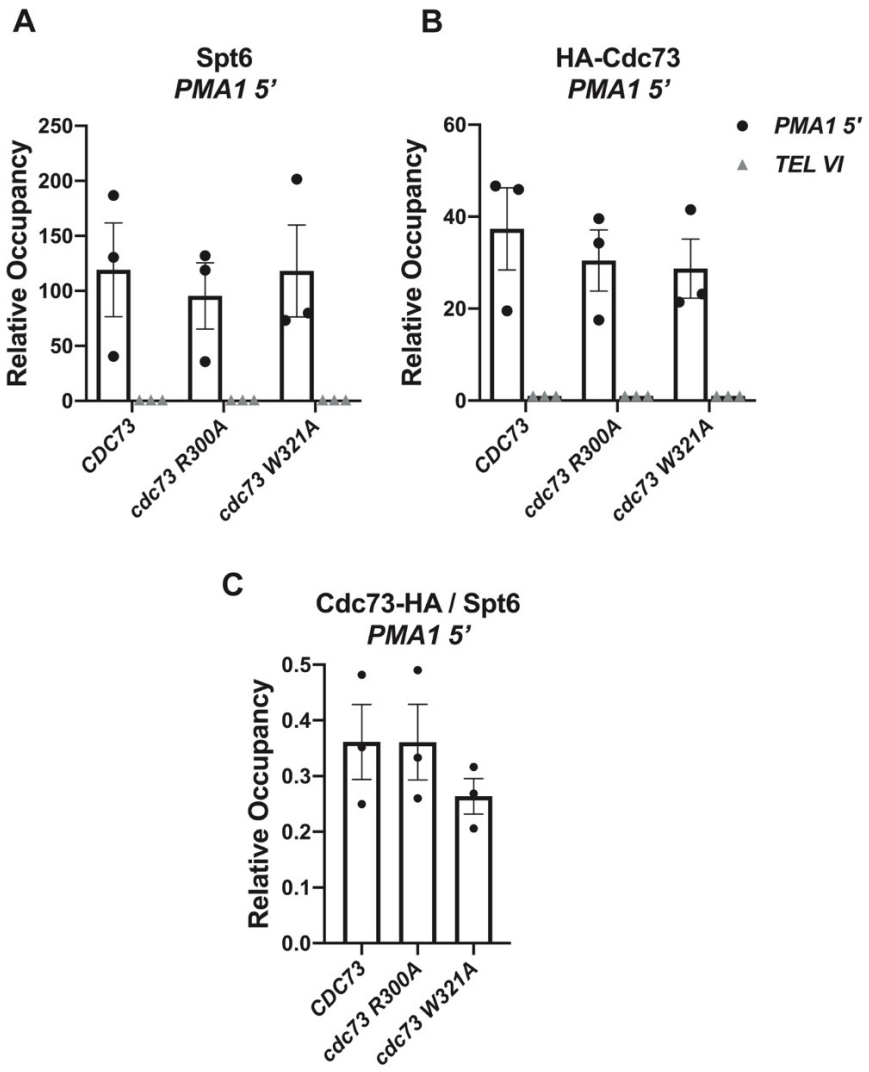


Figure 20. ChIP-qPCR results for Spt6 and Cdc73 in Cdc73 mutants.

A) ChIP-qPCR results for Spt6 showing Spt6 occupancy at the 5' end of the PMA1 locus (highly transcribed) and TELVI (untranscribed) negative control. B) As in A except plotting HA-Cdc73 ChIP-qPCR occupancy. C) ChIP-qPCR data plotting Cdc73 occupancy relative to Spt6 occupancy after normalizing to input and TELVI.All data are normalized to the signal for input and the TELVI region (see methods); therefore, TELVI ChIP signals are set to a value of 1.

3.3.2 Cdc73 and Spt6 interact directly in the absence of all other factors

To assess if the Spt6-Cdc73 interaction is direct, we conducted binding assays with recombinant Spt6 (239-1451) and Cdc73 proteins (Fig 21A and 21B) fused to mRuby2 (red fluorescent protein) or mClover (green fluorescent protein). Gel-shift assay results (Fig 21A) confirm that Spt6 (239-1451) and Cdc73 are sufficient to interact *in vitro* (K_d = 330 nM, Figure 21B). Fluorescent anisotropy data (Figure 21C) collected using mClover Cdc73 and unlabeled Spt6 (239-1451) corroborate the gel-shift results (K_d = 770 nM, Figure 21B). To assess the effects of the Cdc73 W321A mutation, we conducted gel-shift assays (Figure 21A and 21B) and observed a 5-fold reduction in binding affinity in the W321A mutation (Figure 21B). These data suggest that the C-domain participates in the Cdc73-Spt6 interaction and demonstrates the interaction can occur without Rpb1.

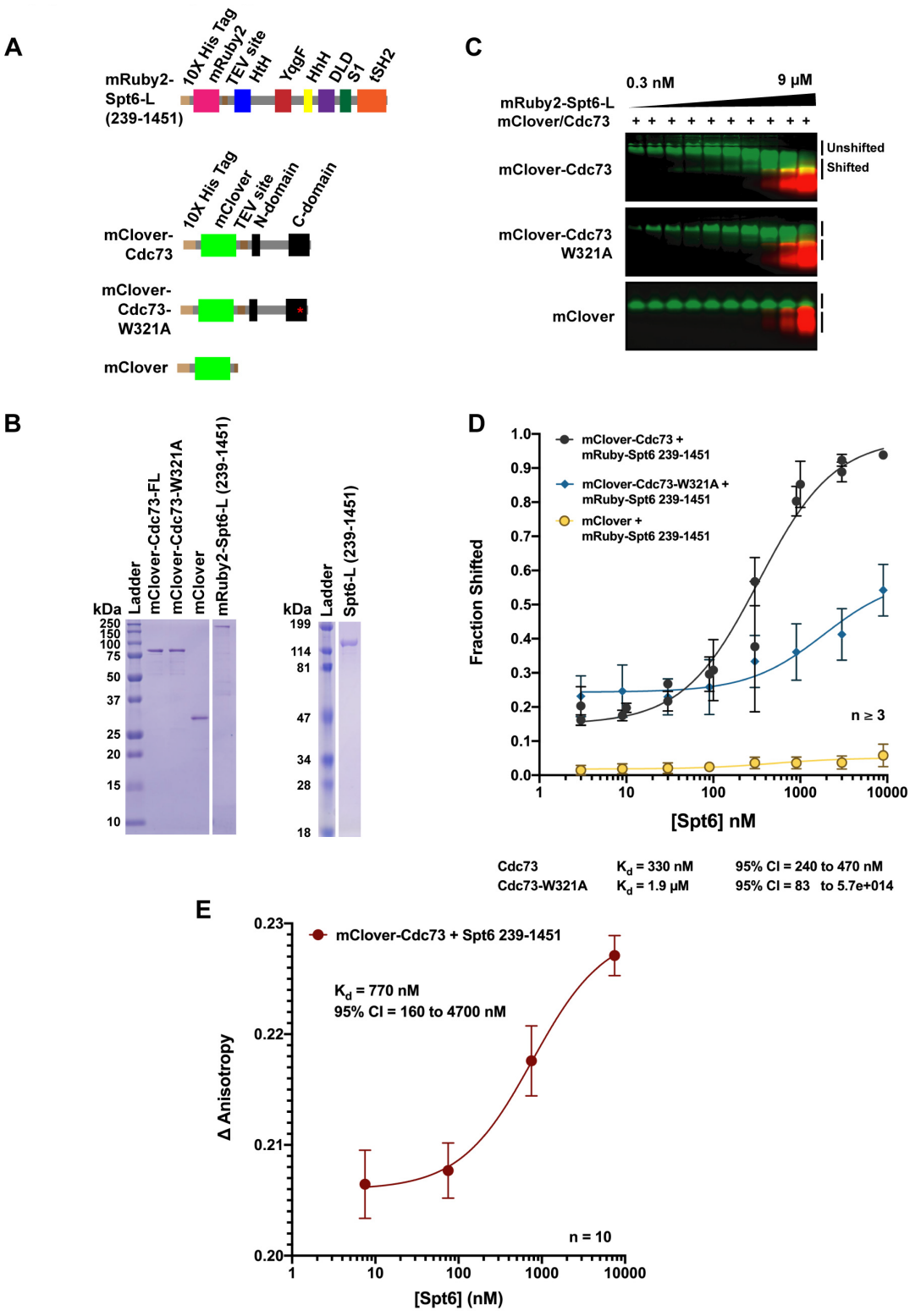

A
mRuby2-Spt6-L (239-1451)
10X His Tag
mRuby2
TEV site
HtH
YqgF
HhH
DLD
S1
tSH2
mClover-Cdc73
mClover
TEV site
N-domain
C-domain
mClover-Cdc73-W321A
mClover
B
kDa
Ladder
mClover-Cdc73-FL
mClover-Cdc73-W321A
mClover
mRuby2-Spt6-L (239-1451)
250
150
100
75
50
37
25
20
15
10
kDa
Ladder
Spt6-L (239-1451)
199
114
81
47
34
28
18
C
0.3 nM
9 µM
mRuby2-Spt6-L
mClover/Cdc73
+ + + + + + + + + +
mClover-Cdc73
Unshifted
Shifted
mClover-Cdc73 W321A
mClover
D
mClover-Cdc73 + mRuby-Spt6 239-1451
mClover-Cdc73-W321A + mRuby-Spt6 239-1451
mClover + mRuby-Spt6 239-1451
Fraction Shifted
n ≥ 3
[Spt6] nM
Cdc73 Kd = 330 nM 95% CI = 240 to 470 nM
Cdc73-W321A Kd = 1.9 µM 95% CI = 83 to 5.7e+014
E
mClover-Cdc73 + Spt6 239-1451
Kd = 770 nM
95% CI = 160 to 4700 nM
Δ Anisotropy
n = 10
[Spt6] (nM)

Figure 21. Cdc73 interacts directly with Spt6 *in vitro*.

A) Diagrams of the primary protein structure of each of the constructs used in the biochemistry experiments. B) SDS PAGE gel stained with Coomassie blue to show protein purity. C) Native gel-shift assays carried out using recombinantly expressed fusion proteins. D) Quantified data from the gel-shift experiment shown in A. Data were quantified by calculating the percentage of mClover-Cdc73 (green band) shifted using the following equation: fraction shifted = (shifted band intensity)/(shifted band intensity + unshifted band intensity). C) Fluorescence anisotropy data collected on a Cytation5 plate reader. Binding curves in D and E were fit to data using the following equation: Y=Bmax*x/(Kd+x)+baseline. Where Bmax, Kd, and baseline were optimized for curve fitting, and X was equal to mean fraction shifted or mean change in anisotropy.

3.3.3 Spt6 tSH2 domain is necessary for Paf1C occupancy on chromatin

We hypothesized that Spt6 helps recruit Paf1C to actively transcribed genes. To test this hypothesis, we performed chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) on *SPT6* and *spt6-50* strains measuring Rpb3, Spt5, Spt6, and Paf1 occupancy. The *spt6-50* mutant contains a nonsense mutation that results in a protein product lacking the tSH2 (Dengl *et al.* 2009; Sun *et al.* 2010; Close *et al.* 2011) domain. This domain is involved in recruiting Spt6 to Pol II and interacts directly with the Pol II CTD linker (Sdano *et al.* 2017; Chun *et al.* 2019). Therefore, if Spt6 recruits Paf1C, we expect a decrease in Paf1 occupancy at genes with reduced Spt6 occupancy in the mutant.

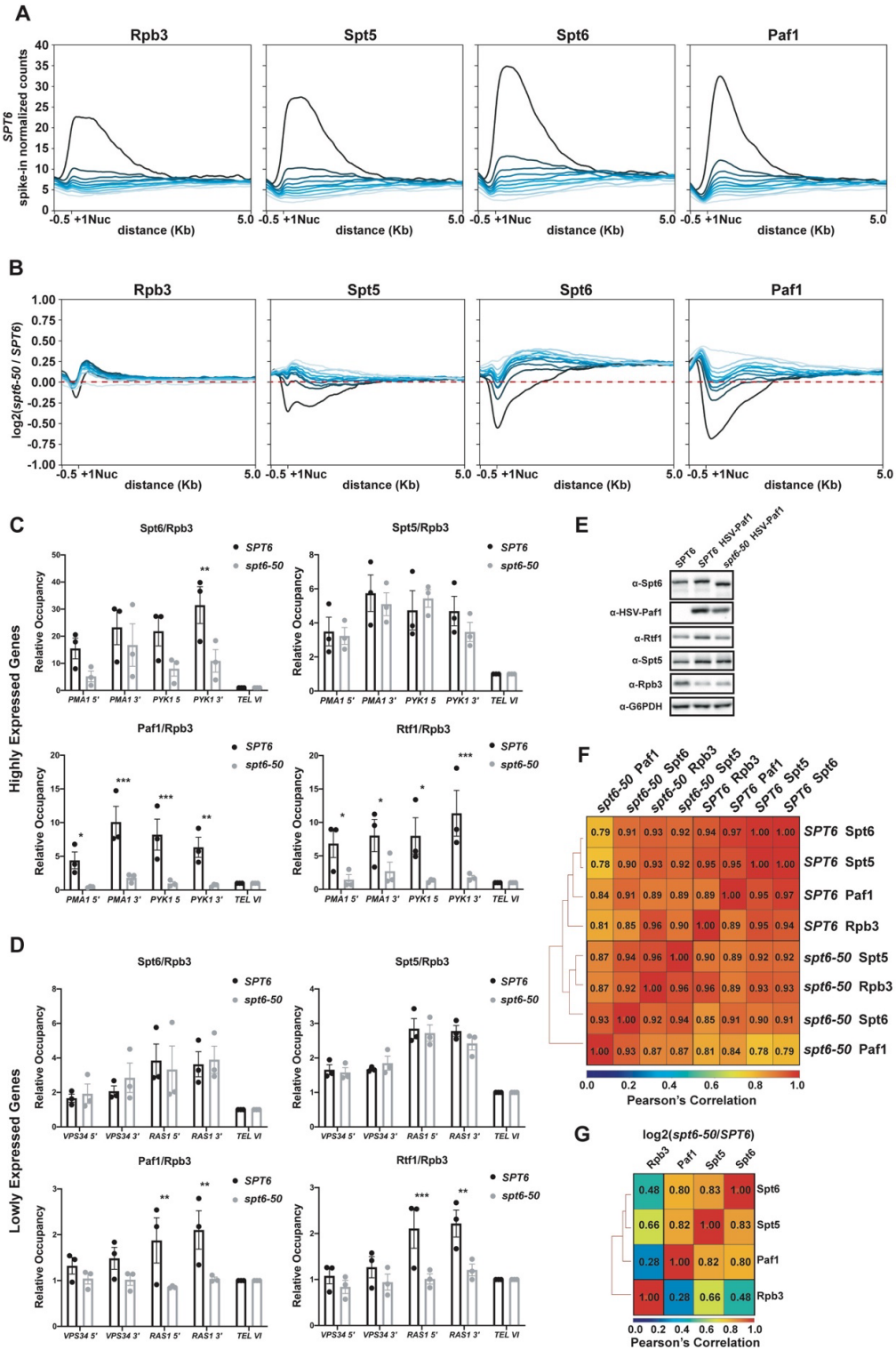


Figure 22. Disruption of Pol II-Spt6 interaction by loss of tSH2 domains reduces Paf1C occupancy.

A) Aggregation plots of Rpb3, Spt5, Spt6, and Paf1 occupancy over mRNA encoding genes with genes broken into deciles after sorting by Rpb3 occupancy. The darkest blue line represents the most highly Pol II (Rpb3) occupied decile (see panel A left plot), and the lightest blue line represents the most lowly occupied decile. B) Aggregate plots of differential Rpb3, Spt5, Spt6, and Paf1 occupancy. Log2 fold difference was calculated between data collected in an *spt6-50* strain and a wild-type *SPT6* strain and plotted as deciles as in A. C & D) ChIP-qPCR for Spt5, Rtf1, HSV-Paf1, and Spt6 plotted relative to Rpb3 for highly expressed genes (C) and lowly expressed genes (D). All ChIP-qPCR data are normalized to both input and an untranscribed region of telomere VI. E) Western analysis showing protein levels of all proteins tested by ChIP-qPCR and G6PDH loading control. F & G) Heatmaps plotting the Pearsons correlation coiefficent calculated using data from all mRNA encoding genes for genome-wide spike-in normalized occupancy (F) and differential occupancy (G) values for all ChIP-seq datasets in this shown in A and B.

Occupancy levels of all proteins tested change with Rpb3 levels (Figure 22A, compare deciles) consistent with their roles as transcription elongation factors (Figure 22A, 22B, 23A and 23B) (Van Oss *et al.* 2016; Sdano *et al.* 2017; Vos *et al.* 2018b). Differential occupancy of the various factors in the *spt6-50* mutant relative to wild type (Figure 22B and 23B) showed a decrease in Paf1 occupancy at genes where decreased Spt6 and Spt5 occupancy was observed. Paf1, Spt5, and Spt6 occupancy decreased dramatically in the top decile and moderately in the second decile in the *spt6-50* strain, while Rpb3 shows little change. Loss of Spt6 in this decile is consistent with this mutation resulting in loss of the Pol II-Spt6 interaction at transcribed regions. Aggregation plots for genes in deciles one and two (Figure 23C, ~1200 genes) support the conclusion that Rpb3 occupancy is unaffected in the *spt6-50* mutant and Spt5, Spt6, and Paf1 decrease across the entire gene in this highly transcribed subset. Surprisingly, in deciles three through eight, Spt6 occupancy increased slightly, particularly at the middle and 3' ends of genes (Fig 22B, 23B and 23D) in the *spt6-50* strain. This result was unexpected but suggests that Spt6 may still be interacting with

transcribed regions via its core and N-terminal unstructured region in this mutant. An additional possibility is that this apparent increase is a result of a normalization artifact because our spike-in, controlling only for library build and sequencing, behaves similar to a counts per million normalization. Indeed this offers an explanation for why we do not see an increase in Spt6 or Paf1 relative to Rpb3 by ChIP-qPCR at lowly expressed genes in this mutant (Figure 22D).

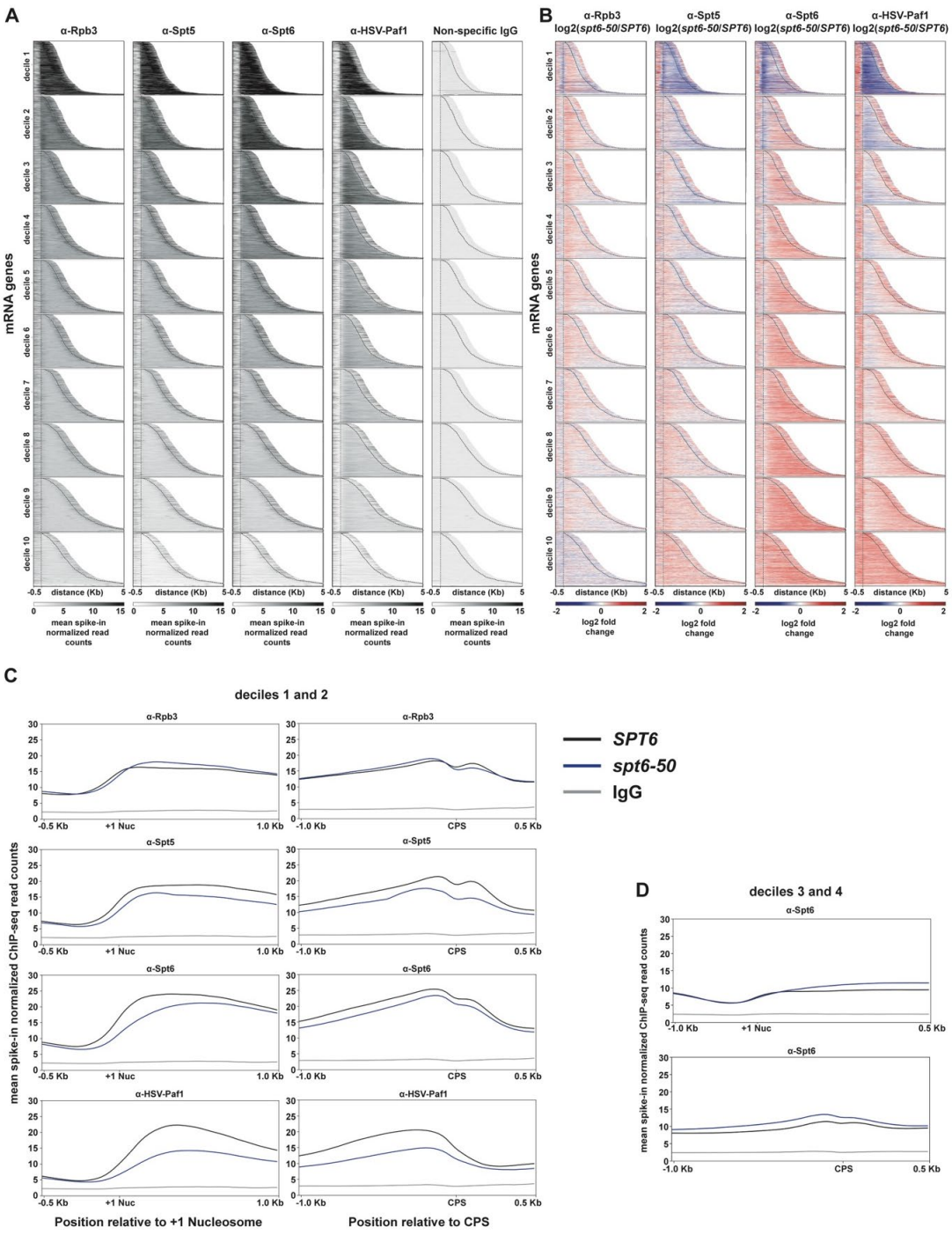


Figure 23. Additional analysis of SPT6 and spt6-50 ChIP-seq data.

A) Heatmaps of Rpb3, Spt5, Spt6, Paf1, and non-specific IgG (Myc) occupancy in an SPT6 strain . Genes are sorted into ten deciles based on Rpb3 occupancy and genes within each decile are sorted by gene length. B) Heatmaps of differential protein occupancy in the *spt6-50* mutant compared to WT (log2(spt6-50/SPT6)).

Data are sorted as in A. C) Aggrigate plots relative to the +1 nucleosome and CPS graphing average occupancy over genes in the top two deciles for Rpb3, Spt5, Spt6 and Paf1 in SPT6 and spt6-50. Aggrigate plots are contain a nonspecific IgG negative control (using a Myc antibody with no epitope Myc in the extract) to show the contribution of nonspecific antibody binding. D) Aggregation plots of average Spt6 occupancy over genes in deciles 3 and 4 plotted as in C.

ChIP-qPCR data for highly expressed (Figure 22C) genes show a more significant effect on Paf1 occupancy than lowly expressed (Figure 22D) genes. Further analysis of Rtf1 (Figure 22C and 22D) suggests that our Paf1 ChIP findings are representative of the entire Paf1C. Western analysis confirms that protein levels of all factors tested are similar between the *SPT6* and *spt6-50* strains that were used for all ChIP-seq and ChIP-qPCR experiments (Figure 22E).

We performed a correlation analysis to test if Paf1 levels agreed more strongly with Spt6, Spt5, or Rpb3 (Figure 22F). Generally, Rpb3, Spt5, Spt6, and Paf1 correlated well at mRNA loci in both *SPT6* and *spt6*-50 strain backgrounds. Spt5 and Spt6 correlated perfectly in the *SPT6* strain background (r = 1.00) and both correlated strongly with Paf1 (Spt5, r = 0.95; Spt6, r = 0.97). Although Rpb3 was highly correlated with Spt5 (r = 0.95) and Spt6 (r = 0.94), it was less well correlated with Paf1 (r = 0.89). In the *spt6-50* strain Spt5 and Rpb3 do not correlate as strongly with Paf1 (Spt5, r = 0.87; Rpb3, r = 0.87) as Spt6 (r = 0.93). To gain further insights, we determined the Pearson's correlation coefficients between differential occupancy datasets (log2(*spt6-50*/*SPT6*)). Differences in the Paf1 datasets correlate well with differences in the Spt5 and Spt6 datasets (Spt5, r = 0.82; Spt6, r = 0.80), but all datasets correlated poorly with Rpb3 (Paf1, r = 0.28; Spt5, r = 0.66; Spt6 r = 0.48). This result makes sense given that Rpb3 occupancy showed little change in the mutant (Figure 22B), and Spt5, Spt6, and Paf1 showed dramatic changes in the

top two deciles. Together these data suggest that Spt5, Spt6, and Paf1 behave similarly upon Spt6 occupancy loss, suggesting that both Spt5 and Spt6 promote genome-wide Paf1C occupancy.

3.3.4 Spt6 is critical to proper genome-wide Paf1 occupancy

Based on the experiments conducted in the *spt6-50* background, it is clear that Spt6 plays a role in Paf1C recruitment. However, Spt6 may still be able to associate with chromatin in the *spt6-50* mutant because it showed an increase in occupancy in the lower Rpb3 occupancy deciles. Additionally, Spt5 decreases where we see Paf1 decrease in that experiment, making it unclear if the effects of Spt6 on Paf1 are direct or indirect (via Spt5). Therefore, to gain additional insights, we appended an auxin-inducible degron tag to the C-terminus of Spt6 to facilitate rapid depletion. This experiment allows us to look at acute effects of its loss without allowing the cells much time to adapt, which we hoped would minimize indirect effects.

To validate the *SPT6-AID* strain, we performed a viability test (Figure 24A) and western analysis (Figure 22B). The viability assay confirmed that auxin treatment over one hour did not result in significant cell death relative to an untreated sample (Figure 24A). Western analysis (Figure 24B) demonstrated loss of Spt6 over the time-course (quantified in Figure 24C) with no adverse effect on Paf1 (quantified in Figure 24D).

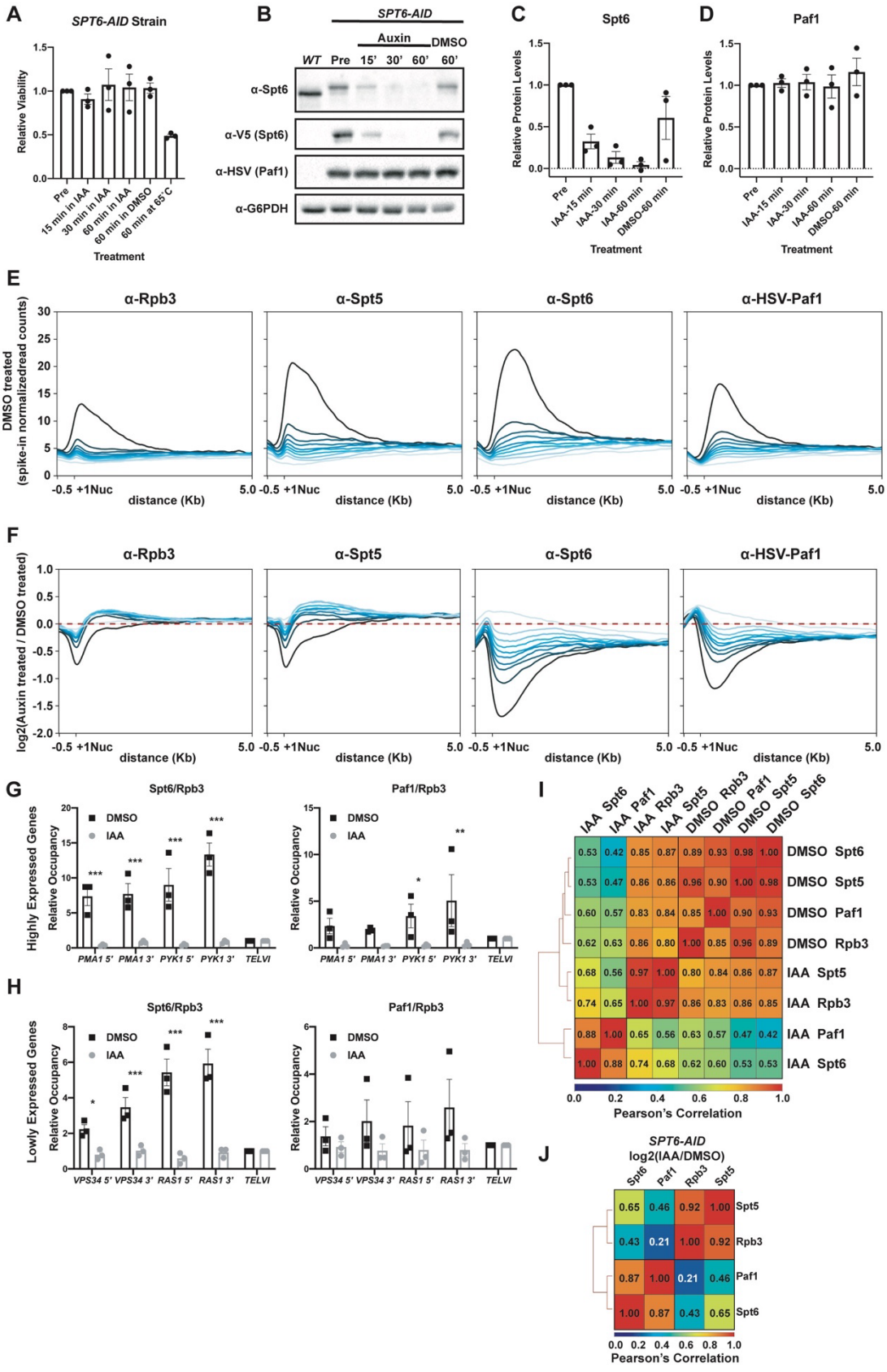

A
SPT6-AID Strain
Relative Viability
Pre
15 min in IAA
30 min in IAA
60 min in IAA
60 min in DMSO
60 min at 65°C
Treatment
B
SPT6-AID
Auxin
DMSO
WT Pre 15' 30' 60' 60'
α-Spt6
α-V5 (Spt6)
α-HSV (Paf1)
α-G6PDH
C
Spt6
Relative Protein Levels
Pre
IAA-15 min
IAA-30 min
IAA-60 min
DMSO-60 min
Treatment
D
Paf1
Relative Protein Levels
Pre
IAA-15 min
IAA-30 min
IAA-60 min
DMSO-60 min
Treatment
E
α-Rpb3
α-Spt5
α-Spt6
α-HSV-Paf1
DMSO treated
(spike-in normalizedread counts)
-0.5 +1Nuc
distance (Kb)
5.0
F
α-Rpb3
α-Spt5
α-Spt6
α-HSV-Paf1
log2(Auxin treated / DMSO treated)
-0.5 +1Nuc
distance (Kb)
5.0
G
Spt6/Rpb3
Paf1/Rpb3
Highly Expressed Genes
Relative Occupancy
DMSO
IAA
PMA1 5'
PMA1 3'
PYK1 5'
PYK1 3'
TELVI
H
Spt6/Rpb3
Paf1/Rpb3
Lowly Expressed Genes
Relative Occupancy
DMSO
IAA
VPS34 5'
VPS34 3'
RAS1 5'
RAS1 3'
TELVI
I
IAA Spt6
IAA Paf1
IAA Rpb3
IAA Spt5
DMSO Rpb3
DMSO Paf1
DMSO Spt5
DMSO Spt6
0.53 0.42 0.85 0.87 0.89 0.93 0.98 1.00 DMSO Spt6
0.53 0.47 0.86 0.86 0.96 0.90 1.00 0.98 DMSO Spt5
0.60 0.57 0.83 0.84 0.85 1.00 0.90 0.93 DMSO Paf1
0.62 0.63 0.86 0.80 1.00 0.85 0.96 0.89 DMSO Rpb3
0.68 0.56 0.97 1.00 0.80 0.84 0.86 0.87 IAA Spt5
0.74 0.65 1.00 0.97 0.86 0.83 0.86 0.85 IAA Rpb3
0.88 1.00 0.65 0.56 0.63 0.57 0.47 0.42 IAA Paf1
1.00 0.88 0.74 0.68 0.62 0.60 0.53 0.53 IAA Spt6
0.0 0.2 0.4 0.6 0.8 1.0
Pearson's Correlation
J
SPT6-AID
log2(IAA/DMSO)
Spt6
Paf1
Rpb3
Spt5
0.65 0.46 0.92 1.00 Spt5
0.43 0.21 1.00 0.92 Rpb3
0.87 1.00 0.21 0.46 Paf1
1.00 0.87 0.43 0.65 Spt6
0.0 0.2 0.4 0.6 0.8 1.0
Pearson's Correlation

Figure 24. Acute depletion of Spt6 results in nearly complete loss of Paf1 occupancy genome-wide. A) Relative viability determined by methylene blue staining across a 60-minute time-course treating samples with IAA or DMSO and sampling at 15, 30, and 60 minutes (IAA), 60 minutes (DMSO) or 60 minutes at 65 ˚C. Viability was calculated relative to the pre-treatment sample. Three biological replicates were performed. B) Representative Western blot of samples taken over a time-course experiment performed as in A. C and D) Quantification of protein levels for three biological replicates assayed by Western blot (representative sample shown in B). Samples are normalized by calculating the following ratio: band intensity/G6PDH-WT band intensity. Fold change was calculated by taking the ratio of each treatment band divided by the pretreatment band. E) Aggregation plots of Rpb3, Spt5, Spt6, and Paf1 occupancy over mRNA encoding genes with genes broken into deciles after sorting by Rpb3 occupancy as in Figure 2. F) Aggregate plots of differential Rpb3, Spt5, Spt6, and Paf1 occupancy. Log2 fold difference was calculated between IAA treated (Spt6 depleted), and DMSO treated (vehicle control) samples and plotted using deciles broken down as in E. G & H) ChIP-qPCR data for Spt6/Rpb3 and HSV-Paf/Rpb3 at highly (G) and lowly (H) expressed genes. I & J) Heatmaps plotting the correlations between genome-wide spike-in normalized occupancy (I) and differential occupancy (J) values for the proteins tested by ChIP-seq.

Aggregation plots for ChIP-seq data collected in DMSO treated control cells showed typical occupancy profiles for all proteins assayed (Figure 24E and 25A). However, occupancy levels were lower than we observed in the *SPT6* strain (compare Fig 22A to Fig 24E). This result may be due to DMSO treatment, or differences between tagged and untagged Spt6. Cells treated with auxin showed little change in Rpb3 and Spt5 occupancy, with each slightly increasing in all deciles, except for decile one, where their occupancy decreased (Figure 24F and 25B). Spt6 occupancy decreased in all but decile ten, which is likely composed of untranscribed genes, confirming that the depletion successfully removed Spt6 from the genome. Importantly, upon Spt6 depletion, Paf1 occupancy decreased in all but decile ten, supporting the hypothesis that Spt6 is required for Paf1C occupancy genome-wide (Figure 24F, 25B and 25C). Further evidence from

ChIP-qPCR experiments measuring Spt6/Rpb3 and Paf1/Rpb3 at highly (Figure 24G) and lowly (Figure 24H) expressed genes supports a decreased occupancy in both gene classes.

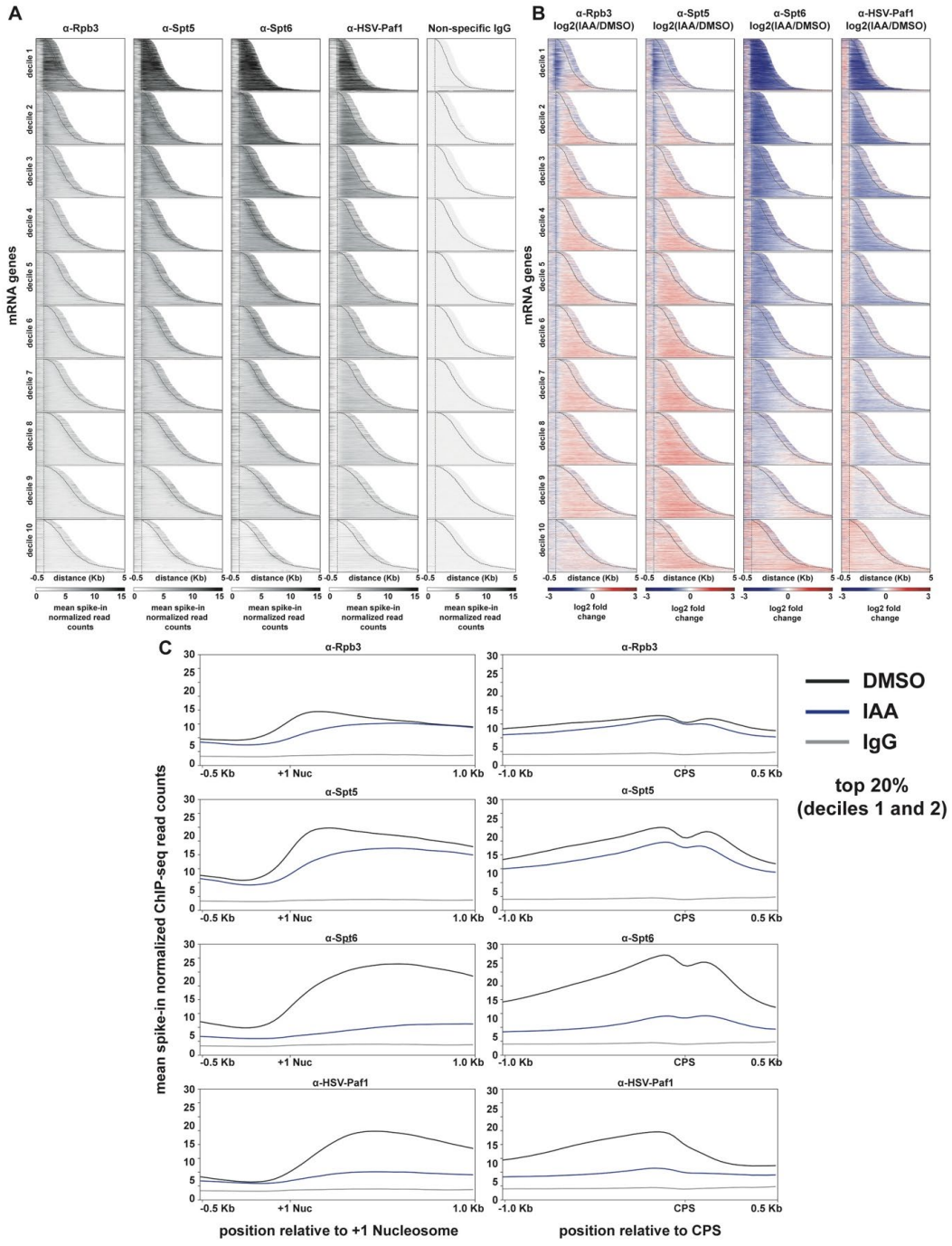


Figure 25. Additional analysis of SPT6-AID ChIP-seq data.

A) Occupancy data as in Figure 23A from *SPT6-AID* strain after a 60-minute exposure to DMSO. B) Differential occupancy heatmaps for (log2(IAA/DMSO)) as in Figure 23B. Aggrigate plots in the top two deciles for IAA treated and DMSO treated plotted as in Figure 23C.

The *SPT6-AID* strain, when treated with DMSO, shows slightly stronger correlations between Spt5, Spt6 and Paf1 (Spt5, r = 0.90; Spt6, r = 0.93) than Paf1 and Rpb3 (r = 0.85) similar to observations in the *SPT6* strain. In auxin treated cells, Spt5 and Rpb3 separate into one group and Paf1 and Spt6 into another, suggesting that Paf1 depends on Spt6 for proper chromatin occupancy through a mechanism that may be at least partially independent of Spt5. Correlation between Paf1 and Spt6 (r = 0.88) is higher than correlation between Paf1 and Spt5 (r = 0.56) or Rpb3 (r = 0.65). Occupancy differences in the Paf1 datasets correlate well with occupancy differences in the Spt6 datasets (r = 0.87), but not the Rpb3 (r = 0.21) or Spt5 (r = 0.46) datasets. This result is surprising given the role the Pol II CTD and Spt5 CTR play in promoting proper Paf1C localization on chromatin (Qiu *et al.* 2012; Mayekar *et al.* 2013; Wier *et al.* 2013). These data implicate Spt6 as a critical factor for promoting proper genome-wide localization of Paf1C.

3.3.5 Identification of the Cdc73-Spt6 binding interface

To understand the interaction between Cdc73 and Spt6, I crosslinked mClover-Cdc73 and Spt6-239-1451 (Figure 21B) with DSS (K-K crosslinking) or EDC (K-E and K-D crosslinking). Proteins exposed to crosslinking agents were resolved on a gel, and bands representing the 1:1 complex were excised and analyzed by mass spectrometry (Figure 26A and 26B, red arrow). Interprotein crosslinks between Cdc73 and Spt6 reveal that Cdc73 interacts with numerous conserved domains within the Spt6 core (Figure 26C, Table 26, and Table 27). The most heavily

crosslinked domains of Spt6 are the Death-like-domain (DLD) and Helix-hairpin-Helix (HhH) domain (Figure 26C).

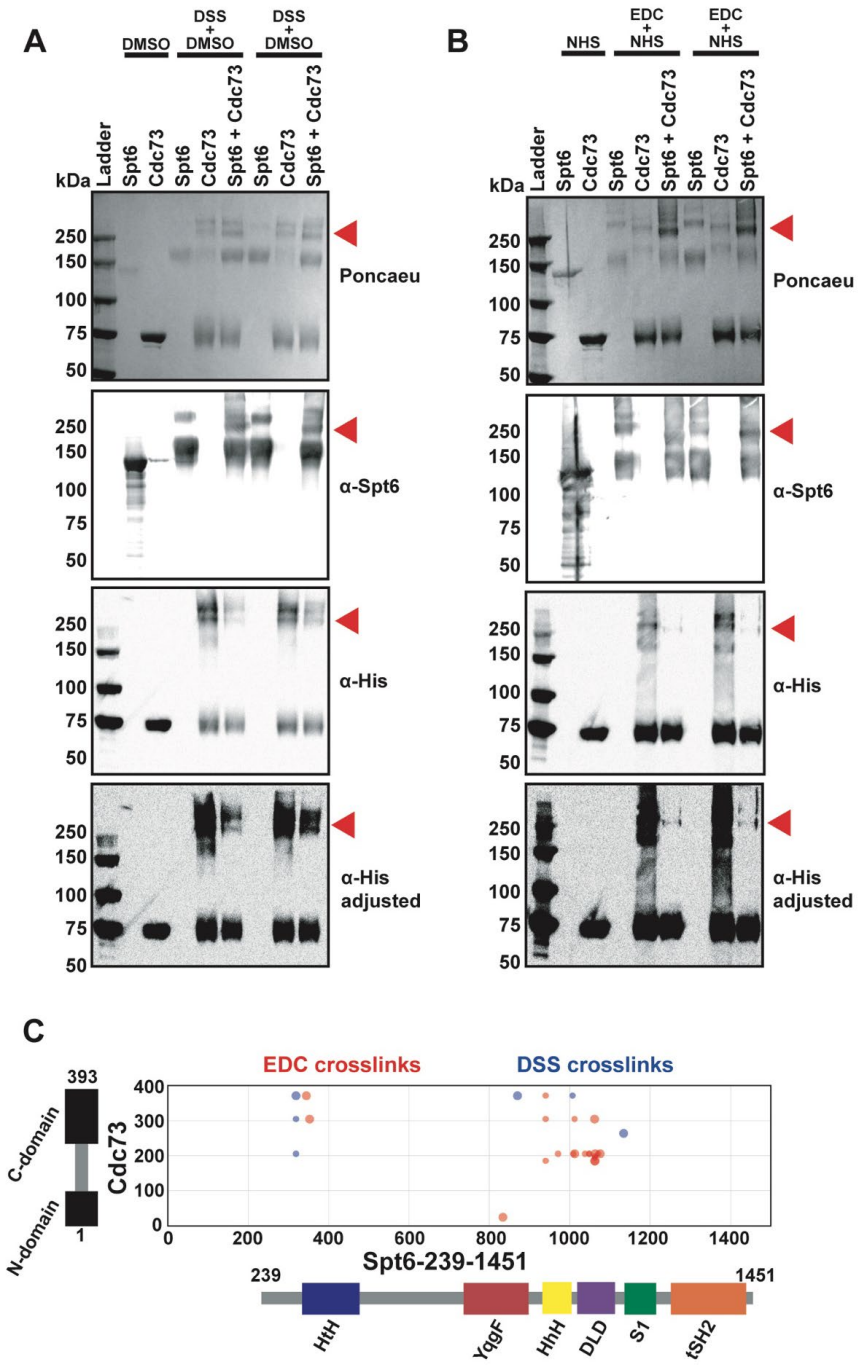


Figure 26. Crosslinking and mass spectrometry analysis of the Spt6-Cdc73 interaction.

A) Ponceau stain and Western analysis of crosslinking reaction products from reactions containing Spt6 (239-1451), Cdc73 full length. DSS dissolved in DMSO was added to each reaction at a final concentration of 625 µM and incubated with agitation for 30 minutes at room temperature. Products were run on an SDS PAGE gel (4%-12% Bis/Tris; Novogene) and transferred to nitrocellulose before Ponceau staining and Western blotting with the indicated antibodies. The red arrows indicate crosslinked products excised from the gel and submitted for mass spec. Adjusted gels (bottom of A and B) have the signal increased by raising the low end of the image intensity. B) Ponceau stain and Western blotting of crosslinking reaction products from reactions containing Spt6 (239-1451), Cdc73 full length, and 12.5 mM EDC with 250 µM NHS-ester. All electrophoresis staining and blotting was carried out as in A. C) Bubble plot of intra-protein crosslinks between Spt6 and Cdc73 identified by mass spectrometry (larger points indicate agreement between datasets; n = 2 for DSS; n = 2 for EDC).

We used interprotein crosslinks from two replicate crosslinking experiments, Spt6 crystal structures (3PSF, 3PSK, and 3PSI (Close *et al.* 2011)), and a model of full-length Cdc73 (Figure 27A-C, Table 28, and Methods) as input for the Integrative Modelling Platform (IMP) (Russel *et al.* 2012).

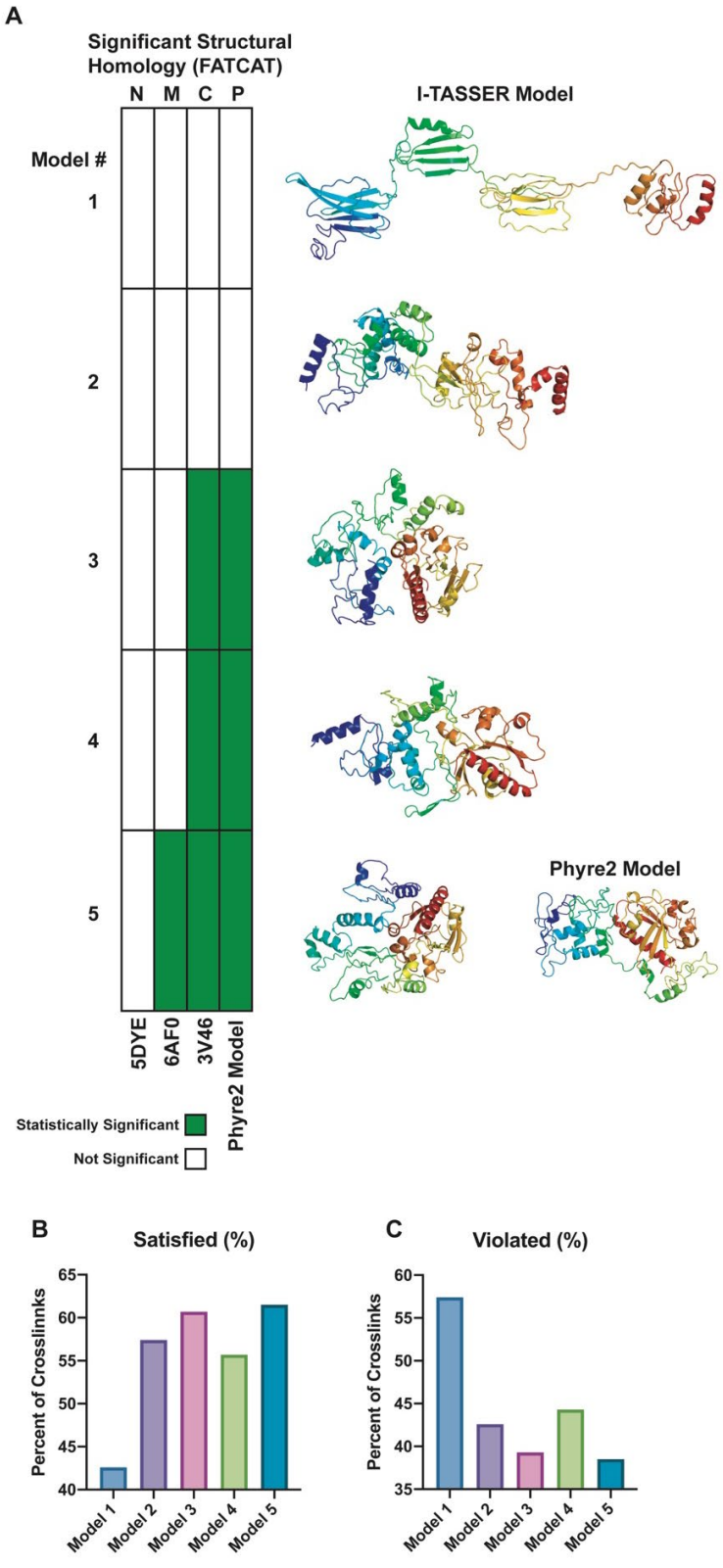

A
Significant Structural Homology (FATCAT)
N M C P
Model #
1
2
3
4
5
5DYE
6AF0
3V46
Phyre2 Model
Statistically Significant
Not Significant
I-TASSER Model
Phyre2 Model
B
Satisfied (%)
Percent of Crosslinnks
65
60
55
50
45
40
Model 1
Model 2
Model 3
Model 4
Model 5
C
Violated (%)
Percent of Crosslinks
60
55
50
45
40
35
Model 1
Model 2
Model 3
Model 4
Model 5

Figure 27. Cdc73 full-length predictive modeling.

A) Structural models generated the I-TASSER (Zhang 2008; Roy et al. 2010; Yang et al. 2014) threading algorithm with a heatmap showing statically significant agreement (green) determined using the FATCAT server (Ye and Godzik 2003, 2004a; b; Li 2006) between the five I-TASSER models and published crystal structures (3V46 (Amrich et al. 2012), 5YDE (Sun et al. 2017) and 6AF0 (Deng et al. 2018)) or a Phyre2 model. B and C) Percentage of crosslinks satisfied (B) and violated (C) when Cdc73 is exposed to crosslinker in the absence of Spt6.

The top 4 structural models produced aligned with an RMSD < 16 (Figure 28B) and were used to generate a model from their combined clustering densities (Figure 28C). The top-scoring model showed Cdc73 interacting with the Spt6 core in a depressed region between the DLD, HhH, and YqgF domains (Figure 28D). Consistent with results from our previous experiments, residues W321 and R300 face in toward the Spt6 core in this model.

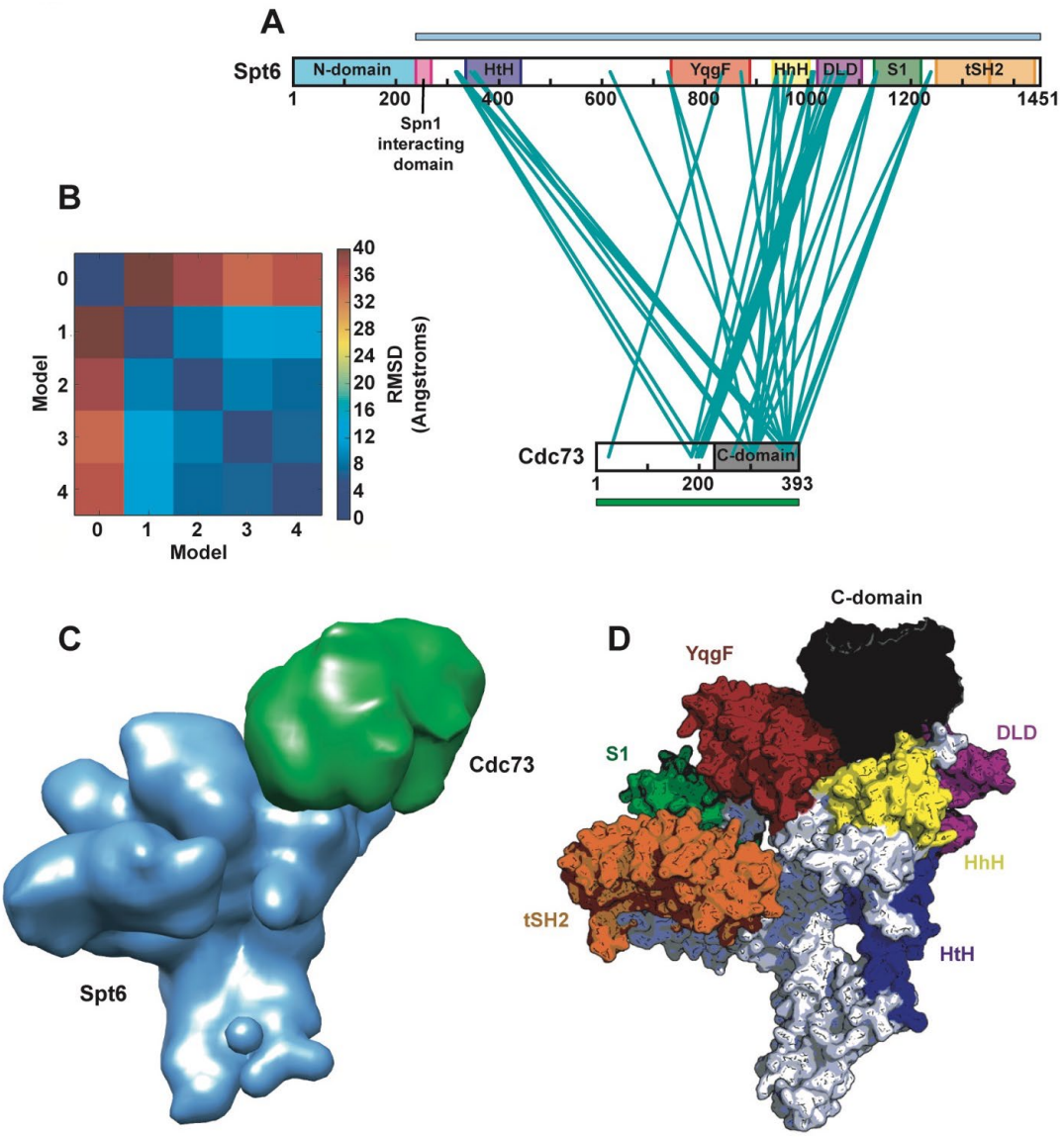


Figure 28. The structural model of the Cdc73 and Spt6 interaction interface suggests that Cdc73 C-domain interacts directly with several domains in the Spt6 core.

A) Protein crosslinking network with protein diagrams, with conserved domains, indicated, as nodes and both EDC and DSS crosslinks as edges using xiView. B) Heatmap of route-mean-square deviation (RMSD) describing the distance between carbon alpha atoms in the top 5 models generated by the Integrative Modelling Platform's (IMP) Monte-Carlo simulation after 100,000 iterations with 10 steps per iteration. C) Molecular model built using data from the top 4 models from B. Note that the model of full-length *Saccharomyces cerevisiae* Cdc73 shown here was predicted using the I-TASSER threading algorithm and validated by two metrics: 1) alignment to various Cdc73 structures (3V46, 5YDE and 6AF0) and 2) the

number of valid crosslink lengths in our XL-MS data for Cdc73 alone. This model was then used along with a composite PDB file made from aligning published Spt6 structures 3PSF and 3PSI. The published structure for the tSH2 domain of Spt6 (3PSK) was also included. Combined with our DSS and EDC crosslinking data these structures both real (from PDB) and predicted (from I-TASSER) were used to generate the model shown in C. D) Model constructed by aligning input structures (described above) back to the carbon alpha trace generated by IMP with conserved domains colored as in A.

As a quality control metric, we mapped all crosslinks (inter and intra-protein, Table 26) back to the model. The majority (105/162 or 64.8%) of cross-linked residues are within a carbon alpha to carbon alpha distance of 30 Å (Figure 29A). If we look at only the crosslinks used as input for IMP (Table 27 and Methods), we see that half of the crosslinks (17/34 or 50%) meet the standard of a 30 Å crosslinking distance (Fig 29B). Crosslinks at greater distances than the 30 Å are found near the HhH or S1 domain, at the N-terminus (construct starts at residue 239), or the region between the S1 domain and the tHS2 domains (a highly flexible linker (Close *et al.* 2011; Vos *et al.* 2018b)), respectively. Both regions are highly mobile, probably allowing for crosslinking to occur between residues that are much farther apart in the model, which shows only a single conformation. The I-TASSER (Zhang 2008; Roy *et al.* 2010; Yang *et al.* 2014) threading algorithm (see Figure 27, Table 28, and Methods) generated all residues of Cdc73 N-terminal to residue 230. Therefore, some crosslinks cannot be accurately measured if the structural model is not entirely correct. In summary, crosslinks exceeding the 30 Å cutoff are likely due to artifacts of our *in vitro* system, or errors in our model. Both the conformation observed in the model, and the fact that crosslinks to the YqgF, HhH, and DLD domains are generally within the accepted 30 Å range, suggest that the Cdc73-Spt6 interaction interface is composed of the C-domain of Cdc73 and the YqgF, HhH and DLD domains of Spt6.

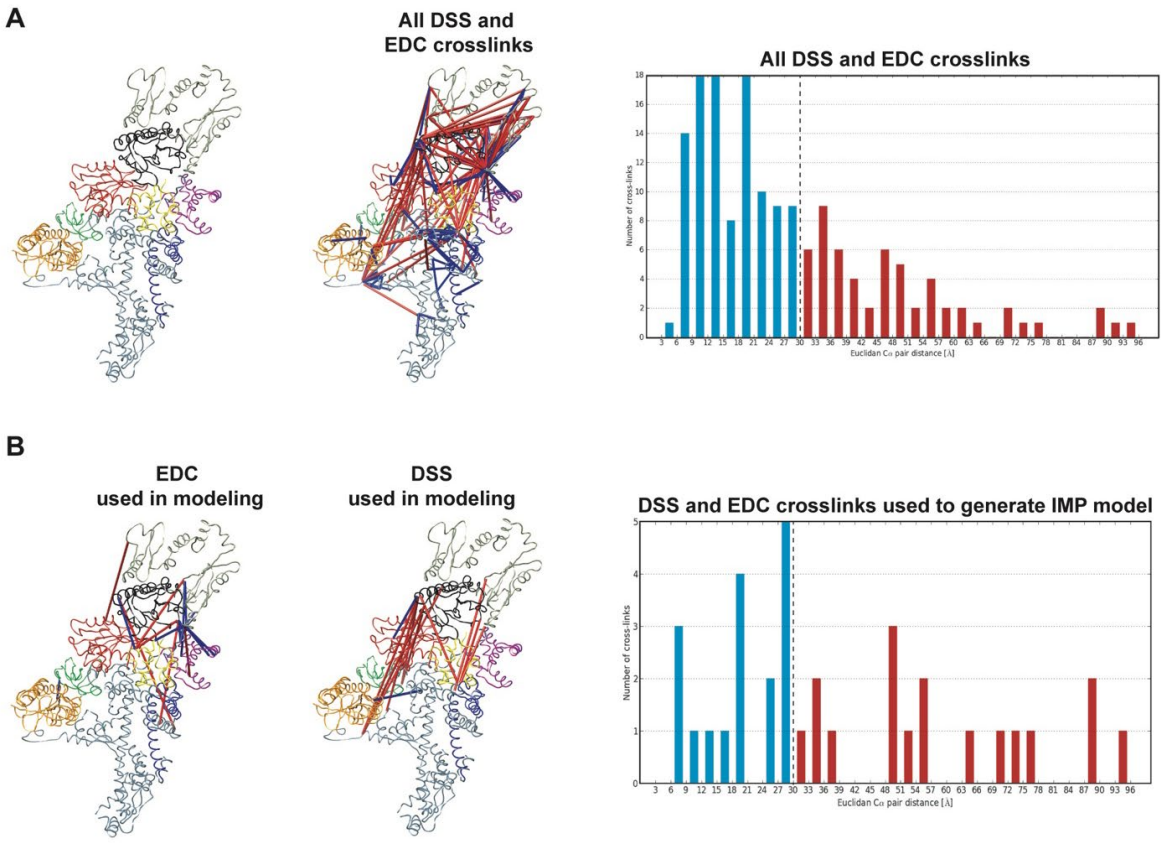


Figure 29. Crosslink mapping and length distribution for the Cdc73-Spt6 structural model.

A) All crosslinking data mapped to the Spt6-Cdc73-FL binding model. From left to right: Carbon alpha trace of the Spt6-Cdc73-FL binding model colored as in Figure 28D with the middle and N-terminal regions of Cdc73 (modeled by I-TASSER) shown in grey (top) and non-domain regions of Spt6 shown in silver (near the bottom); Carbon alpha trace with crosslinks from DSS and EDC crosslinking experiments mapped between carbon alphas of cross-linked residues with blue lines representing crosslinks equal to or less than the 30 Å cutoff and red representing crosslink lengths exceeding the 30 Å cutoff; Histogram of crosslink distances with blue bars for crosslinks at or under the 30 Å cutoff and red for these exceeding the cutoff. B) Crosslinks between Cdc73 and Spt6 and Spt6 core and tSH2 domain only mapped to the Spt6-Cdc73-FL binding model. From left to right as in A: DSS crosslinks; EDC crosslinks, Histogram of crosslink distances.

3.3.6 Mutational analysis of Spt6 at its Cdc73 interacting region

To assess the accuracy of our model, we performed a mutational analysis on Spt6 (Figure 30). We generated plasmid-borne copies of Spt6 tagged with a V5 epitope tag that either lack a domain or contain alanine scanning mutations at the proposed binding site on each domain. We mutated multiple binding site residues at once due to the low resolution of our XL-MS-based model (Figure 30A), which does not allow us to identify residue-residue contacts precisely.

We found that a ΔYqgF mutant was inviable, however all alanine scanning mutants and the ΔHhH and ΔDLD deletions were viable. All mutants except ΔHhH grew similar to wild-type (Figure 30B) on rich media; however, they showed a broad range of Spt- phenotypes. Numerous mutants have a weak (NEP 970-972 AAA and ELLRE 1058-1062 AAAAA) or moderate (DFI 775-777 AAA, NKATD 936-940 AAAAA and KRQK 1004-1007 AAAA) Spt- phenotype, but many are no worse than strains shuffled with the parent plasmid containing *3XV5-SPT6*. We observed a mutant phenotype in the ΔHhH mutant consistent with results from an *spt6-1004* mutant, which is known to have an Spt- phenotype (Cheung *et al.* 2008). Many mutants with moderate phenotypes are central to the proposed binding site near the HhH. By far, the strongest mutants were the ΔDLD mutant and the PEDY 1023-1026 AAAA mutant, which show phenotypes similar to the *spt6-14* positive control. Interestingly, the PEDY 1023-1026 AAAA mutant, which resides at a conserved patch of unknown function on the DLD (Close *et al.* 2011), outside of the proposed binding interface, phenocopies the complete deletion of the DLD.

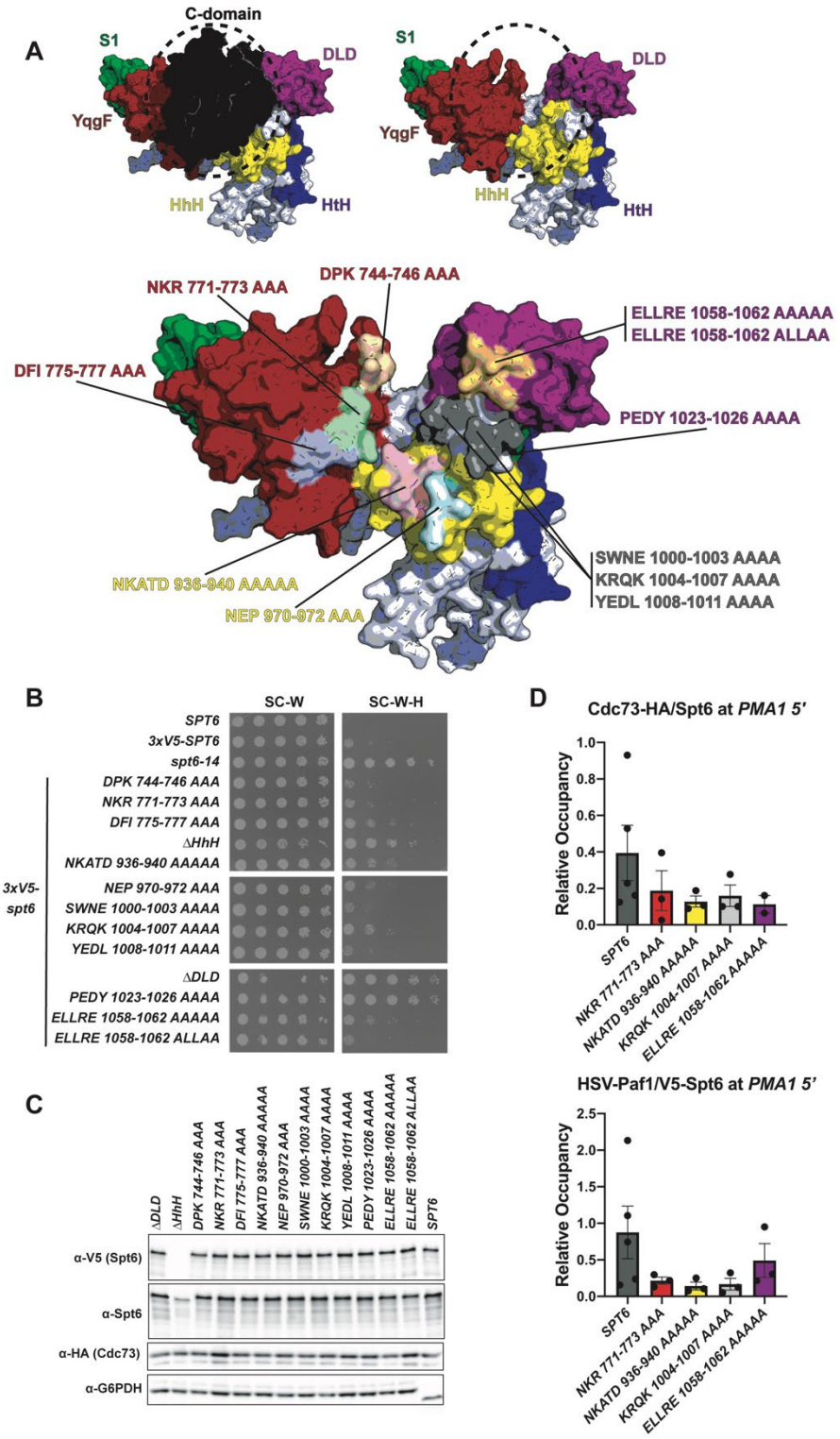

A
C-domain
S1
DLD
YqgF
HhH
HtH
S1
DLD
YqgF
HhH
HtH
NKR 771-773 AAA
DPK 744-746 AAA
ELLRE 1058-1062 AAAAA
ELLRE 1058-1062 ALLAA
DFI 775-777 AAA
PEDY 1023-1026 AAAA
SWNE 1000-1003 AAAA
KRQK 1004-1007 AAAA
YEDL 1008-1011 AAAA
NKATD 936-940 AAAAA
NEP 970-972 AAA
B
SC-W
SC-W-H
SPT6
3xV5-SPT6
spt6-14
DPK 744-746 AAA
NKR 771-773 AAA
DFI 775-777 AAA
ΔHhH
NKATD 936-940 AAAAA
3xV5-spt6
NEP 970-972 AAA
SWNE 1000-1003 AAAA
KRQK 1004-1007 AAAA
YEDL 1008-1011 AAAA
ΔDLD
PEDY 1023-1026 AAAA
ELLRE 1058-1062 AAAAA
ELLRE 1058-1062 ALLAA
C
ΔDLD
ΔHhH
DPK 744-746 AAA
NKR 771-773 AAA
DFI 775-777 AAA
NKATD 936-940 AAAAA
NEP 970-972 AAA
SWNE 1000-1003 AAAA
KRQK 1004-1007 AAAA
YEDL 1008-1011 AAAA
PEDY 1023-1026 AAAA
ELLRE 1058-1062 AAAAA
ELLRE 1058-1062 ALLAA
SPT6
α-V5 (Spt6)
α-Spt6
α-HA (Cdc73)
α-G6PDH
D
Cdc73-HA/Spt6 at PMA1 5'
Relative Occupancy
1.0
0.8
0.6
0.4
0.2
0.0
SPT6
NKR 771-773 AAA
NKATD 936-940 AAAAA
KRQK 1004-1007 AAAA
ELLRE 1058-1062 AAAAA
HSV-Paf1/V5-Spt6 at PMA1 5'
Relative Occupancy
2.5
2.0
1.5
1.0
0.5
0.0
SPT6
NKR 771-773 AAA
NKATD 936-940 AAAAA
KRQK 1004-1007 AAAA
ELLRE 1058-1062 AAAAA

Figure 30. Spt6 binds to Cdc73 C-domain via multiple conserved domains.

A) The structural model of the Cdc73-Spt6 interaction focusing in on the interaction interface on the surface of Spt6. Conserved protein domains are colored and labeled as in Figure 5A and 5D. Alanine scanning mutations are shown in various colors and labeled with colors corresponding to the domain to which the mutated residues belong. B) Serial dilution assay results for Spt6 mutants assessing Spt- phenotype (SC-W-H) and growth on control media (SC-W). C) Western analysis of Spt6 mutants assessing protein levels of Spt6, Cdc73, and G6PDH (loading control). D) ChIP-qPCR data (as in Figure 1D) plotting Cdc73 occupancy relative to Spt6 occupancy after normalizing to input and TELVI. All data in this figure are preliminary and pending further replication. Statistical tests will be performed upon completion of other replicates.

To ensure that protein expression defects were not driving the observed phenotypes, we analyzed all mutants by western. All mutants were stably expressed except for the ΔHhH mutant (Fig 30C). The reduction of protein levels in the ΔHhH mutant comes as no surprise since reduced levels of Spt6 were reported previously in an *spt6-1004* strain (Kaplan *et al.* 2005; Dronamraju and Strahl 2014). These data suggest that the phenotypes observed in the mutants are a functional consequence of the mutated residues and not due to reduced protein levels, except for ΔHhH, where a reduction in protein levels likely contributes to the phenotype.

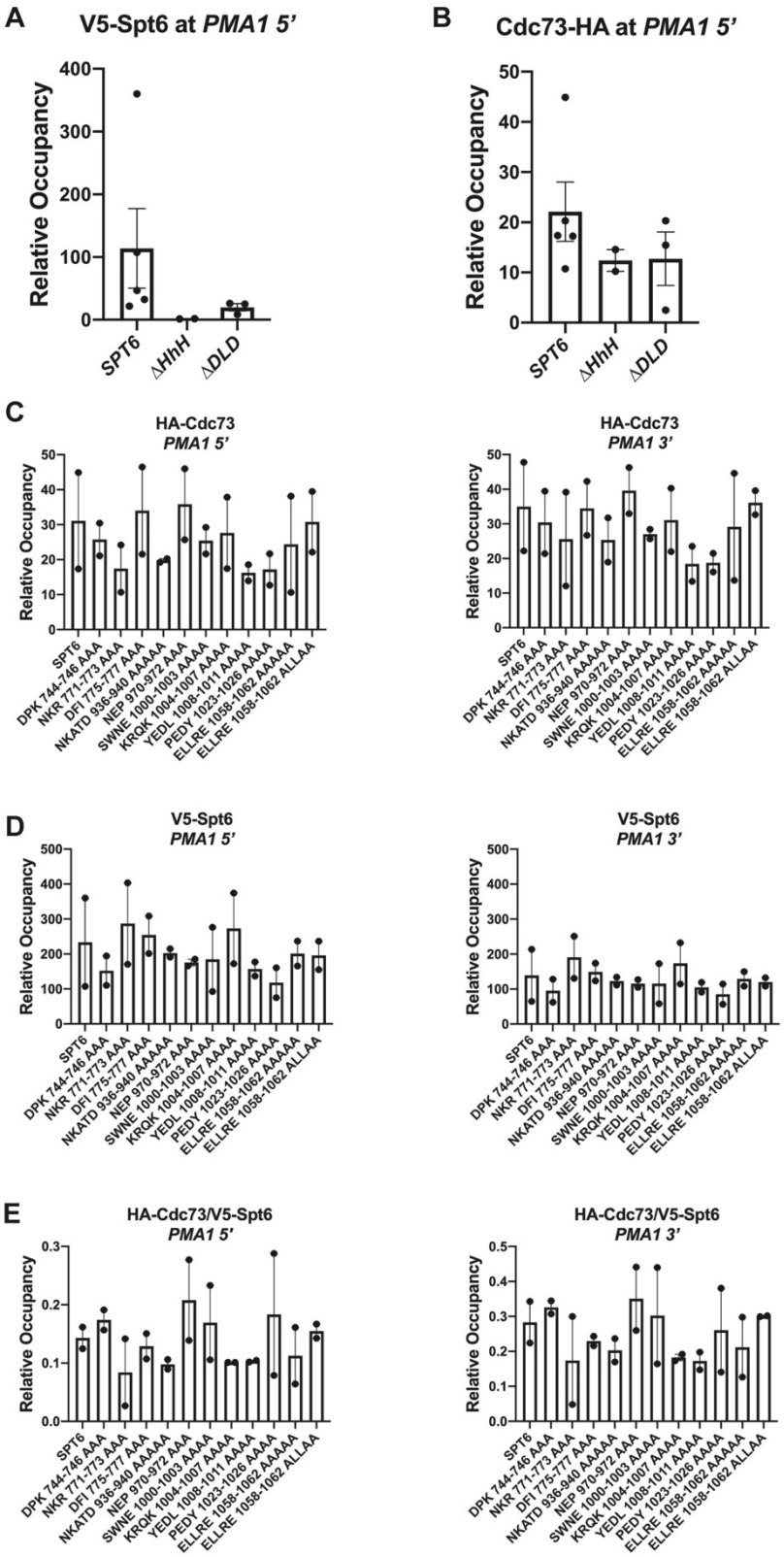


Figure 31. Spt6 and Cdc73 occupancy in Spt6 mutants.

A) ChIP-qPCR results for Spt6 showing Spt6 occupancy at the 5' end of the *PMA1* locus (highly transcribed). B) As in A except plotting HA-Cdc73 ChIP-qPCR occupancy. C) HA-Cdc73 occupancy as in A for *PMA1 5'* (left) and *5'* end (right). D) Spt6 occupancy as in C. Cdc73/Spt6 occupancy as in C. All data is normalized to input and then to the TELVI region (see methods). All data in this figure are preliminary and pending further replication and no statistical tests have been performed on these data.

To further analyze these mutants, we performed ChIP-qPCR to assess chromatin occupancy of both Cdc73 and Spt6 using HA and V5 epitope tags, respectively. Results for full-domain deletions (ΔHhH and ΔDLD) showed reduced Spt6 (Figure 31A) and Cdc73 (Figure 31B) occupancy making it challenging to determine if loss of Cdc73 was due to loss of the interaction or decreased Spt6 levels on chromatin. An integrated *spt6-1004* mutant showed similar results at highly expressed genes (Figure 32A), with less dramatic changes observed at lowly expressed genes (Fig 32B), similar to what we observed in the *spt6-50* mutant (Figure 22). Oddly, when we plot Cdc73/Spt6, we observe an increase in the ratio (Figure 32C and 32D) which is counter to what we would expect for a domain involved in the Cdc73-Spt6 interaction, however, we believe, that the substantial decrease in Spt6 occupancy makes this metric less reliable for the full domain deletion mutants. Therefore, we turned to our alanine scanning mutants to assess the binding interface more directly.

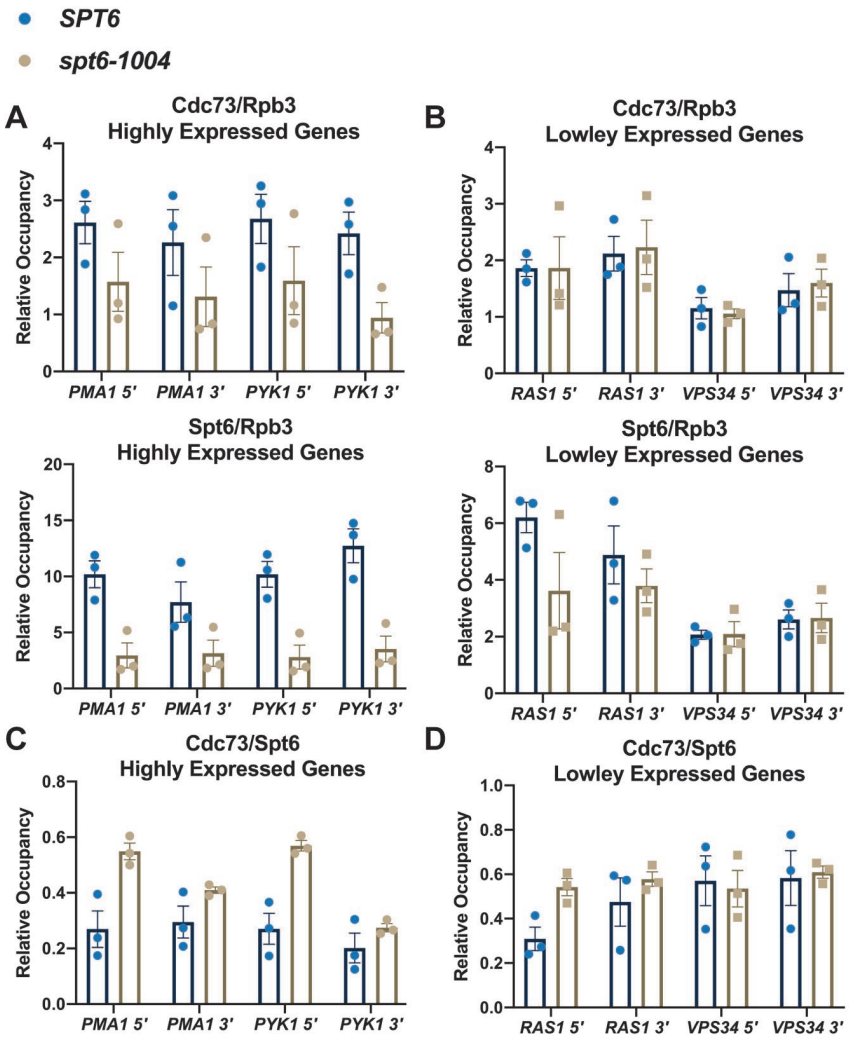


Figure 32. Cdc73 and Spt6 occupancy in an integrated *spt6-1004* mutant.

A) Cdc73/Rpb3 occupancy levels at *PMA1* and *PYK1* in *SPT6* and *spt6-1004* strain backgrounds for highly expressed (top) and lowly expressed (bottom) genes. Spt6/Rpb3 occupancy as in A. C and D) Cdc73/Spt6 occupancy levels at *PMA1* and *PYK1* in *SPT6* and *spt6-1004* strain backgrounds for highly expressed (left) and lowly expressed (right) genes. No statistical tests have been performed on these data.

We predicted that Spt6 mutants that lie at the hypothesized binding interface would show a reduction in the Cdc73/Spt6 ratio at the *PMA1* locus, similar to Cdc73 W321A (Figure 20C). We performed our initial screen of these mutants in biological duplicate (Figure 31). We observed

slightly reduced Cdc73 levels (Figure 31C) in many of these mutants, some with a subsequent decrease in Spt6 (Figure 31D). We observed small decreases in Spt6 levels in a few mutants, which enabled us to use the Spt6/Cdc73 occupancy as a proxy for disruption of the Spt6-Cdc73 interaction on chromatin.

We observe a reduction in mean Cdc73/Spt6 occupancy in a handful of our Spt6 mutants (Figure 31E; NKR 771-773 AAA, DFI 775-777 AAA, NKATD 936-940 AAAAA, KRQK 1004-1007 AAAA, YEDL 1008-1011 AAAA, ELLRE 1058-1062 AAAAA). Again, we note greater effects in mutants in and near the HhH domain. For four of these mutants, one from each interacting domain, replicates were analyzed (Figure 30D). All of the selected mutants had a lower mean Cdc73/Spt6 ratio than wild-type, suggesting that they disrupt the Cdc73-Spt6 binding interface. We assayed Paf1 levels and calculated the mean Paf1/Spt6 ratio. All mutants result in a reduction in mean Paf1/Spt6 ratio suggesting this interface is important for Paf1C recruitment. These results support our XL-MS structural model and reinforce the conclusion that multiple domains of Spt6 work together to coordinate the binding of the Cdc73 C-domain.

3.3.7 Mutational analysis of Cdc73 at its Spt6 interacting region

To test the Cdc73-Spt6 interaction from the Cdc73 side, we again chose to use mutational analysis (Figure 33). Here we expect key residues like W321 and R300 in the C-domain of Cdc73 to be part of the Cdc73-Spt6 interface (Figure 4 G), and indeed they are (Figure 33A). The model agrees with both *in vivo* (Figure 19E and 20C), and *in vitro* (22C and 22D) results suggest that it is accurate. Furthermore, the fact that the C-domain appears to be the main point of interaction agrees with previously published results (Amrich *et al.* 2012). Nevertheless, we chose to design additional mutants to assess the model further.

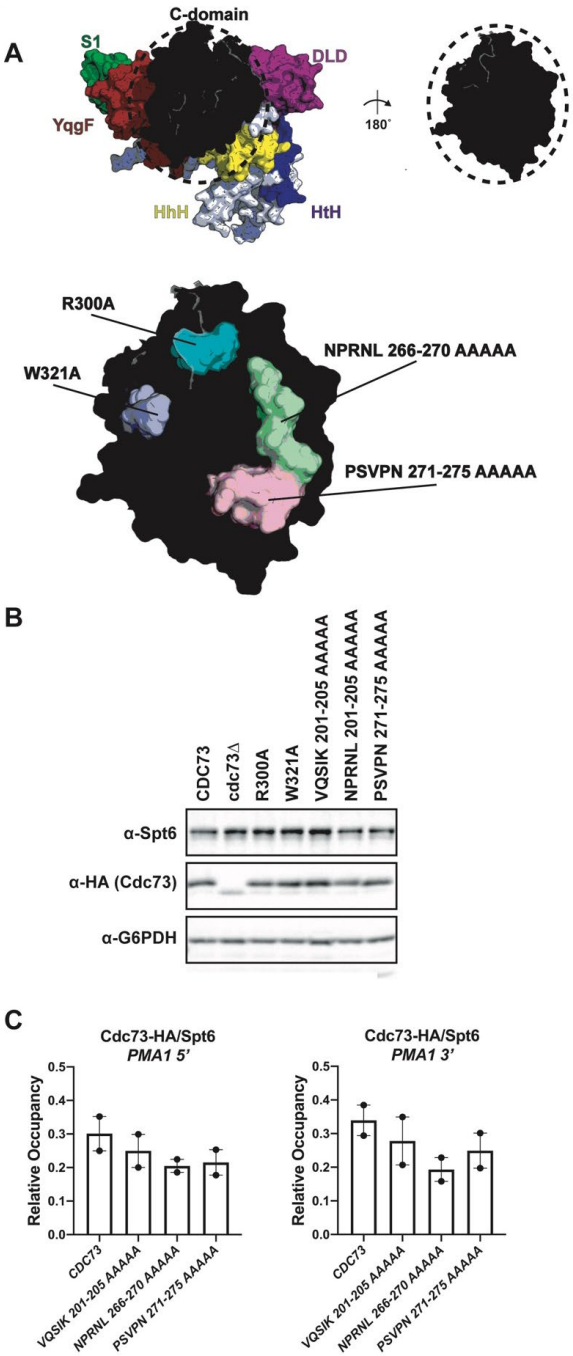


Figure 33. Cdc73 interacts with Spt6 via its conserved C-domain.

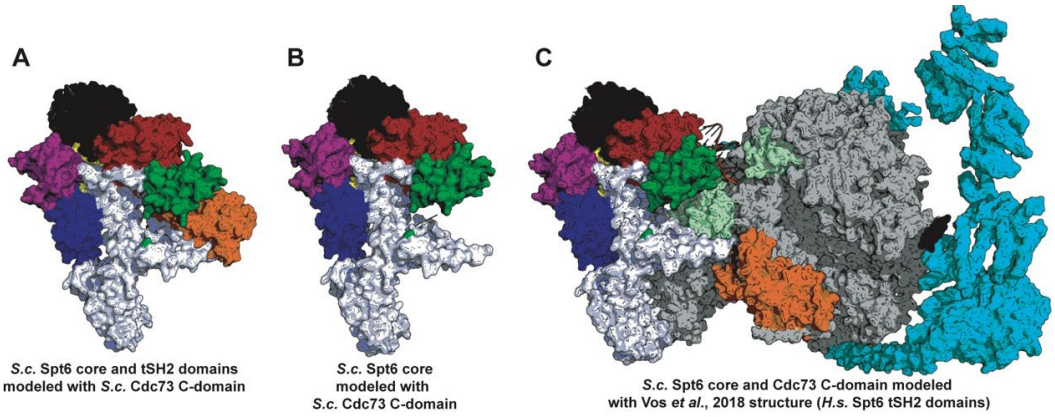
A) The structural model of the Cdc73-Spt6 interaction (as in Figure 30A) focusing on the interaction interface of the Cdc73 C-domain. Conserved protein domains are colored and labeled as in Figure 28A, 28D, and 30A. Alanine scanning mutations are shown in various colors and labeled with colors corresponding to the domain to which the mutated residues belong. B) Western analysis of Spt6 mutants assessing protein levels of Spt6, Cdc73, and G6PDH (loading control). C) ChIP-qPCR data (as in Figure 19D and 30D) plotting Cdc73 occupancy relative to Spt6 occupancy after normalizing to input and TELVI. All data in this figure are preliminary and pending further replication and no statistical tests have been performed on these data.

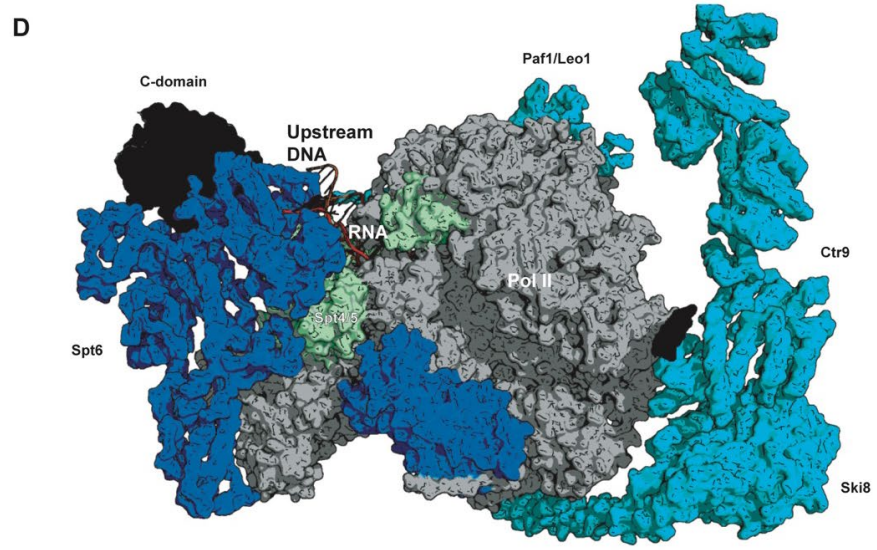
We made two mutations in the C-domain, at the interaction interface, but opposite W321A (Figure 33A, NPRNL 266-270 AAAAA and PSVPN 271-275 AAAAA), and one mutation in a region outside of the C-domain (VQSIK 201-205 AAAAA) where numerous cross-links were observed in our EDC crosslinking dataset (at K205). We performed western analysis on all mutants and found that all proteins were expressed to a similar extent as the *CDC73* control (Figure 33B). ChIP-qPCR results show a decrease in mean Cdc73/Spt6 occupancy at the 3' and 5' ends of *PMA1*, suggesting that these mutated regions of Cdc73 are participating in the Cdc73-Spt6 interaction. The C-domain mutants (NPRNL 266-270 AAAAA and PSVPN 271-275 AAAAA) result in a similar decrease in occupancy to what we observed for W321A. However, Cdc73/Spt6 ratio defects in VQSIK 201-205 AAAAA is less severe. These results further validate the model and support the conclusion that the C-domain of Cdc73 directly interacts with Spt6 and promotes proper Paf1C occupancy.

3.3.8 Positioning Cdc73 C-domain on the active elongation complex

To place our structural model into the context of what is known about the structure of the active Pol II elongation complex (EC*), we aligned our model to the recently published model

(Vos *et al.* 2018b). We considered only the core of Spt6 and the C-domain of Cdc73 from our model (Figure 34A and B) and aligned yeast Spt6 to human Spt6. In the resulting model, the C-domain of Cdc73 (in black) is nearby the exiting RNA, Spt5 KOWx-KOW4 domains, and the Stalk of Pol II (Fig 34C and 34D).





Vos *et al.*, 2018 structure *H.s.* Spt6, Spt5 and Paf1C and *S.s.* Pol II modeled with *S.c.* Cdc73 C-domain

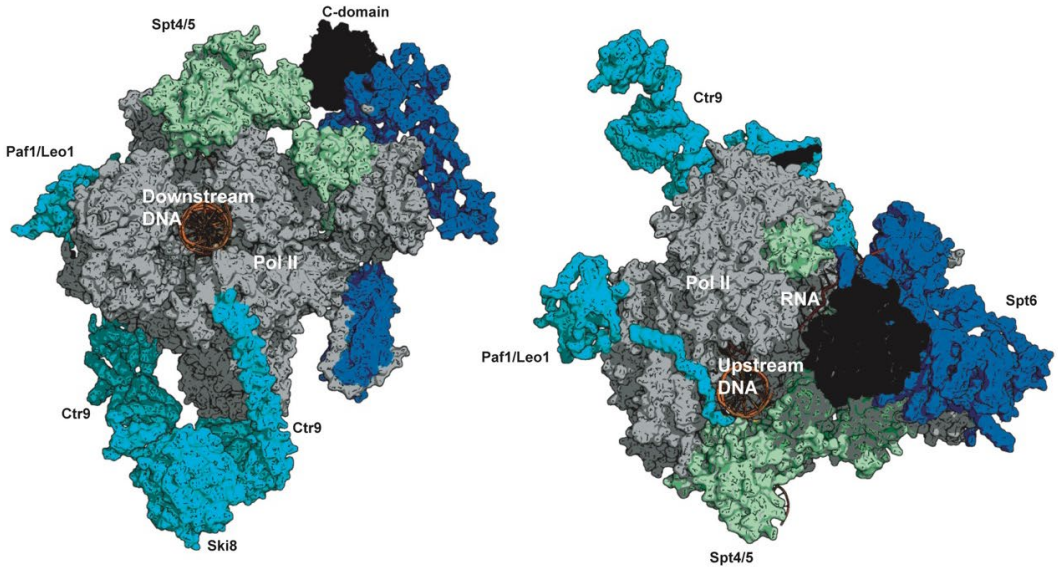


Figure 34. Additional structural modeling data.

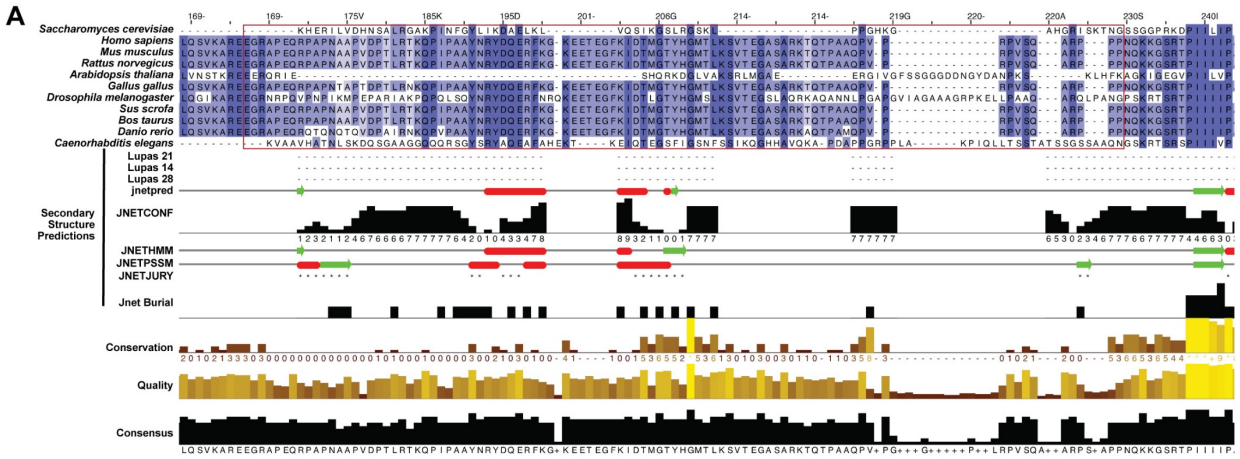
A) *S. cerevisiae* Spt6-Cdc73 model with tSH2 domain included (same as Figure 28D except for orientation). B) Spt6-Cdc73 model as in A except lacking tSH2 domain. C) Spt6-Cdc73-EC* model generated (see Methods) by aligning the yeast Spt6 core to the human Spt6 core in the EC* structure with the human tSH2 domain shown in orange. D) EC* structure with C-domain incorporated as in C except showing human Spt6 core instead of yeast (top), oriented to show downstream DNA (bottom left) and oriented to show upstream DNA (bottom right).

Only a single helix of Cdc73 is visible in the published EC* structure, which was assigned to residues 249-262 in the human protein, based on crosslinking and secondary structure predictions (Vos *et al.* 2018b). This helix interacts with Ctr9 and is termed the “anchor helix” (Vos *et al.* 2018b). Residues in the hCdc73 "anchor helix" are part of a region of hCdc73 that is important for its interaction with Paf1, Leo1, and Ctr9 (Rozenblatt-Rosen *et al.* 2005). Residues 249-262 in the human protein correspond to a gap in the yeast protein in our Clustal O results (Figure 35A) between residues 168-169 in the yeast protein. This region of Cdc73 was observed interacting with Ctr9 and Paf1 in data collected by another research group (Deng *et al.* 2018). When measurements are taken from the C-terminus of the hCdc73 “anchor helix” to the N-terminus of the yCdc73 C-domain in our model, we find that they are ~181 Å apart. There are 62 amino acids in the Pol II spanning region of Cdc73 which is between residue 168 and 230 in the yeast protein. The polypeptide backbone from Cα to Cα can span up to 3.5 Å. This distance allows for a maximum distance of 217 Å (Figure 35B), which means that the Cdc73 protein can span the distance between the "anchor helix" (Vos *et al.* 2018b) and C-domain (our model).

I reasoned that there may have been secondary structure within this region, so we performed secondary structural predictions on our Clustal O alignment results (Figure 35A). I found that there was minimal secondary structure predicted in this region of Cdc73. There are two

short alpha helices consisting of a combined ~15 amino acids. In an alpha helix the distance between residues drops to 1.5 Å meaning that we can only span 187 Å if the helices are taken into account (47 aa x 3.5 Å/aa + 15 aa x 1.5 Å/aa = 187 Å). This is still more than sufficient to cover the distance required for this model to be plausible. I calculated distances using 3.0 Å and 3.5 Å in Figure 35B in order to estimate the distance spanned by this region with and without secondary structure being considered. In all organisms except for *Arabidopsis thaliana* this region is long enough to span the gap and even *A. thaliana* is very close to the proper distance, so it is possible that in plants differences in the proteins involved allow for a shorter distance to be accommodated. Furthermore the *A. thaliana* protein lacks some of the regions with predicted secondary structure so it may be reaching a distance closer to that predicted by the 3.5 Å residue length estimate which allows for its shorter Pol II spanning region to reach an acceptable distance of 189 Å. Finally, it is also possible that members of EC* undergo conformational changes *in vivo* that facilitate the Cdc73-Spt6 interaction.

In summary, the structural data collected for the Cdc73-Spt6 interaction suggests that the Cdc73 C-domain interacts with Spt6 via multiple conserved domains, and this interaction positions the C-domain near the Pol II stalk, Spt5 and the exiting RNA in the actively transcribing Pol II elongation complex. In order to accomplish this structural arrangement, given what we know about the structure of EC* the length of the Cdc73 Pol II spanning region appears to be constrained and suggesting that the length of this sequence is evolutionarily conserved even if its sequence is not.



B

Organism	Start	End	Residues	Å (if 3 Å/Residue)	Å (if 3.5 Å/Residue)
Saccharomyces cerevisiae	168	230	62	186	217
Homo sapiens	260	352	92	276	322
Mus musculus	260	352	92	276	322
Rattus norvegicus	260	352	92	276	322
Arabidopsis thaliana	191	245	54	162	189
Gallus gallus	260	352	92	276	322
Drosophila melanogaster	260	371	111	333	388.5
Sus scrofa	260	352	92	276	322
Bos taurus	260	352	92	276	322
Danio rerio	259	351	92	276	322
Caenorhabditis elegans	246	346	100	300	350

Figure 35. Conservation of the length of the Pol II spanning region of Cdc73

A) From top to bottom: Clustal O alignment of 12 organisms colored by percent identity with darker colors indicating higher percent identity and a red box drawn around the region spanning from the anchor helix to the C-domain, secondary structural predictions for coil-coil regions (Lupas_21, Lupas_14, Lupas_28), secondary structure prediction (jnetpred) with alpha helices in red and beta sheets in green, confidence estimate for secondary structure prediction (JNETCONF), hiden markov model (HMM) based secondary structure predictions (JNETHMM), position-specific scoring matrix (PSSM) based secondary structure predictions (JNETPSSM), JNETJURY where * indicate that the JNETJURY algorithm was invoked to rationalize differences in primary predictions (HMM and PSSM) to arrive at the final prediction (jnetpred), prediction of likelihood a residue is buried (Jnet Burial), Conservation, Quality, and Consensus. B) Shows data tabulated for each organism in the alignment listing residues at the start and end of the region in Cdc73 that would have to span between the anchor helix to the C-domain and listing distance in residues and angstroms that can be reached in each organism

3.4 Discussion

I investigated an interaction between Cdc73 and Spt6 that I suspected to be involved in Paf1C recruitment to chromatin. Spt6 is an essential elongation factor and histone chaperone (Duina 2011) known to promote transcription (Keegan *et al.* 2002; Endoh *et al.* 2004) and prevent cryptic initiation (Kaplan 2003). It is involved in chaperoning H3-H4 histone dimers (Bortvin and Winston 1996) during transcription, where it facilitates their placement back onto the DNA after Pol II passage (Cheung *et al.* 2008; Jensen *et al.* 2008). A mutant form of Spt6 (*spt6-1004*) is synthetically lethal with *cdc73*Δ*, paf1*Δ *,* and *ctr9*Δ (Kaplan *et al.* 2005), results in changes to chromatin architecture, loss of H3K36me3 (Dronamraju and Strahl 2014; Doris *et al.* 2018) and produces aberrant histone modification patterns genome-wide (discussed below) (Jeronimo *et al.* 2019). In this mutant, cells are unable to properly recruit Ctr9 to the *GAL10* and *GAL7* locus (Kaplan *et al.* 2005), suggesting a role for Spt6 in promoting proper Paf1C levels at transcribed regions.

Using *in vivo* BPA crosslinking, we identified a direct interaction between Cdc73 and Spt6, and further analysis *in vitro* confirmed that Spt6 and Cdc73 can interact without any other EC* members. In a mutant lacking the domain of Spt6 responsible for its interaction with the Pol II CTD linker, Paf1 occupancy correlates with Spt6 occupancy. Acute depletion of Spt6 leads to genome-wide loss of Paf1 occupancy, confirming that Spt6 is a critical factor in promoting Paf1C occupancy on chromatin. The lack of an effect on Spt5 after acute depletion of Spt6 suggests that Spt5 does not globally require Spt6 for its recruitment meaning the effects on Paf1 do not merely result from disruption of the known Paf1C recruitment pathway involving Spt5 (Mayekar *et al.* 2013; Wier *et al.* 2013). These results demonstrate that Spt6 is a critical factor in promoting proper

genome-wide localization of Paf1C. This new interaction adds a third tethering point between the Paf1C and the active Pol II elongation complex (EC*).

The two known Paf1C-EC* tethering points are the Pol II C-terminal domain (CTD) and the Spt5 C-terminal repeat sequence (CTR). Interactions between the Spt5 CTR and the ORF associated region (OAR/Plus 3 domain) of Rtf1 are well-characterized and known to play a critical role in Paf1C recruitment (Mayekar *et al.* 2013; Wier *et al.* 2013). Interactions between Paf1C and the phosphorylated Pol II CTD have been characterized *in vitro*, and decreases in Paf1C member occupancy are observed at coding regions in strains that are unable to place specific phosphorylation marks on the CTD repeats (Qiu *et al.* 2012).

Previous data suggested that Cdc73 interacted with the Pol II CTD via its C-domain (Amrich *et al.* 2012; Qiu *et al.* 2012), and there was no evidence of a direct interaction between Cdc73 and Spt6 except for a few crosslinks reported in the mammalian proteins (Vos *et al.* 2018b). Although direct interaction between the CTD and Cdc73 C-domain had not been observed *in vivo*, it was demonstrated *in vitro* through peptide binding experiments (Qiu *et al.* 2012). *In vivo* analysis of yeast strains lacking the C-domain of Cdc73 (Amrich *et al.* 2012) demonstrate that this domain is critical for the Paf1C recruitment, but not required for Paf1C formation. Structural analysis of the Cdc73 C-domain (residues 230-393, PDB-ID 3V46) revealed that this highly conserved portion of Cdc73 has a Ras-like fold with a flattened out GTP binding pocket, which may have been evolutionarily repurposed to participate in a protein-protein interaction (Amrich *et al.* 2012).

We present further evidence for the interaction between the Cdc73 C-domain and the Rpb1 CTD (Qiu *et al.* 2012) and report the first evidence for a direct interaction *in vivo.* These experiments implicate the Cdc73 C-domain residues R300 and W321 in a direct interaction with

the Rpb1 CTD and Spt6. Our data support a role for the Rpb1 CTD and Spt6 in recruiting Cdc73 to EC*.

Interestingly, the CTD linker, Spt5 CTR and Pol II CTD (S2P) all require Bur1 kinase and its cyclin partner Bur2 (Bur1/2 or pTEFB in metazoans) (Qiu *et al.* 2009, 2012; Liu *et al.* 2009; Suh *et al.* 2016). Inhibition of Bur1, or deletion/mutation of Bur2, results in a reduction of phosphorylation, leading to loss of Paf1C occupancy *in vivo* (Zhou *et al.* 2009; Liu *et al.* 2009; Qiu *et al.* 2012; Mayekar *et al.* 2013; Wier *et al.* 2013; Chun *et al.* 2019) due to loss of Cdc73-CTD and Rtf1-CTR interactions. However, our data suggest that loss of the Spt6-CTD linker interaction in this context also contributes. Thus our results provide an additional molecular explanation for the importance of Bur1/2 in controlling Paf1C recruitment (Qiu *et al.* 2012; Mayekar *et al.* 2013).

While older data show Paf1C at promoters and the 5' end genes, more recent high resolution data show that it is present at much higher levels in the middle and 3' end of genes (Krogan *et al.* 2002; Pokholok *et al.* 2002) with occupancy peaking near the +3 nucleosome (Van Oss *et al.* 2016). This occupancy pattern likely occurs because the known recruitment mechanisms for Paf1C involve phosphorylation events that are observed in these same regions (Qiu *et al.* 2009, 2012; Mayer *et al.* 2010; Bataille *et al.* 2012; Cortazar *et al.* 2019). Co-localization of many factors is observed in this region, including, but not limited to, Spt4-Spt5, Spt6, and Paf1C (Mayer *et al.* 2010). Genome-wide occupancy patterns of these factors suggest that they join Pol II very early in transcription elongation (Mayer *et al.* 2010; Van Oss *et al.* 2016). So it is not necessarily surprising that we see an interplay between these factors as they all occupy EC* at the same time (Vos *et al.* 2018b).

We were surprised to observe that Spt6 occupancy increased in the middle and 3’ ends of the lowly expressed genes in the *spt6-50* mutant. However, our results for highly expressed genes generally agree with other *spt6* strains with different tSH2 deletion boundaries (*spt6-50* = 1-1273, *spt6*Δ*C* = 1-1249, *spt6*Δ*202* = 1-1250) where decreases in Spt6 are noted, and only the data at highly expressed genes are presented (Mayer *et al.* 2010; Burugula *et al.* 2014). The persistent decrease near the 5’ end in all expression classes is consistent with a reduction in initial recruitment to EC* in the *spt6-50* mutant. Decreases in Spt6 occupancy at the 5' end of genes were reported in strains where kinases were deleted or inhibited (*ctk1*Δ *bur1as*) (Burugula *et al.* 2014). Now we understand that this is due to loss of Bur1/2 dependent phosphorylation of the CTD linker (Chun *et al.* 2019), which directly interacts with the tSH2 domains of Spt6 (Sdano *et al.* 2017), so it makes sense we would observe a similar phenotype in this *spt6-50* mutant even if a secondary retention or recruitment mechanism were to exist. Perhaps interactions between Spt5 and Spt6 may allow Spt6 to continue to occupy the 3’ of the gene but are not sufficient to support high levels of Spt6 at the 5’ end. Indeed structural studies show interactions between the Spt6 core and both the Spt5 KOWx-KOW4 domain and the Rpb4-Rpb7 stalk of Pol II (Vos *et al.* 2018b). However, this does not explain the increase in Spt6 levels at the 3’ end of the gene.

Interactions between the histone deacetylase complex Rpd3S (specifically the Rco1 and Rpb3 subunits) and Spt6 can support occupancy in the middle of the gene (Burugula *et al.* 2014). Spt6 occupancy near the middle and 3' end is promoted by interactions between the tSH2 domain and S2P CTD, and occupancy at the extreme 3' end of the gene is dependent upon interactions with tyrosine one phosphorylated Y1P CTD (Burugula *et al.* 2014). However, in a strain lacking the tSH2 domain, it is unclear how these patterns persist.

It may be possible that in the absence of the tSH2 domain Spt6 relies on interactions with Spn1 and the histone proteins to recruit it to actively transcribed regions. The N-terminus of Spt6 (Spn1/histone binding region; ~300 amino acids) promotes Spt6 occupancy likely through a nucleosome or Spn1 dependent mechanism. Indeed, Spn1 has a similar ChIP-seq profile to Spt6 in the *spt6-50* mutant, with levels starting low at the 5' end and increasing toward the 3' end (Mayer *et al.* 2010). ChIP-seq results for a phosphomimetic Spt6 mutant (spt6 S8-E8 mutated at casein kinase II (CKII) targeted residues), which imitates constitutive phosphorylation of residues that promote the interaction between Spt6 and Spn1, results in decreased Spt6 occupancy at the 5' end and only a slight increase at the 3' end of highly expressed genes (Dronamraju *et al.* 2018c). There is very little difference observed when all genes are considered in aggregate, which may indicate that, in this context, the tSH2-CTD linker interaction is sufficient to maintain the interaction with Pol II at most genes even in the presence of increased Spn1 binding (Sdano *et al.* 2017; Dronamraju *et al.* 2018c). Spt6-Spn1 and Spt6-histone interactions may play a more prominent role in the context of the *spt6-50* mutant. However, further studies will be required to elucidate how the tSH2 domain and N-terminus of Spt6 collaborate to promote its interaction with chromatin.

In addition to the genomic data, I developed a structural model of the Spt6-Cdc73 interaction. The structural model indicates that the YqgF, HhH, and DLD domains of Spt6 work together to coordinate the binding of the Cdc73 C-domain. Spt6 residues that, when mutant, have the most potent effects on Paf1C occupancy are located near or in the HhH domain. Perhaps this is not surprising given the previously reported reduction of Ctr9 occupancy in the *spt6-1004* (ΔHhH) mutant (Kaplan *et al.* 2005). However, because of the considerable reduction in Spt6 levels both in the cell and on chromatin in this strain, it was not possible to conclude that this domain is explicitly contributing to the Cdc73-Spt6 interaction without the detailed mutational

analysis performed here. Importantly, *spt6-1004* was synthetically lethal with deletion of Paf1C member Cdc73, as well as, Ctr9 and Paf1 suggesting that supporting Paf1C recruitment is not the only role the HhH is playing, which makes sense given the multifunctional nature Paf1C and Spt6 (Duina 2011; Tomson and Arndt 2013).

The tSH2 domain is positioned on the side of the Spt6 core that interacts with Pol II in our model. It is tempting to consider the possibility that the tSH2 domain could enable Spt6 to switch between Pol II bound and unbound states by an autoinhibitory binding event such as this, however far more work would be necessary to confirm such a hypothesis. It is also possible that this is an artifact of our *in vitro* system, although if the tSH2 domain were promiscuously interacting with the core, we would expect to see crosslinks between the tSH2 domain and various locations on the Spt6-core structure, and we do not.

When we place our model into the context of the EC* structure (Vos *et al.* 2018b), we find that the Cdc73 C-domain is located near the Spt5 KOWx-KOW4 domains and the Stalk of Pol II near the nascent RNA. Indeed, Cdc73 has been shown to interact with mRNA and possesses a similar PAR-CLIP profile to Spt5 and Spt6 (Battaglia *et al.* 2017) supporting its proposed positioning. While Cdc73 showed strong interactions with RNA in previously published electrophoretic mobility shift (EMSA) assays, it was unclear if this was a legitimate interaction because the complex did not enter the gel, which often indicates aggregation (Dermody and Buratowski 2010). The role that the Cdc73 C-domain plays during recruitment of Paf1C is evident from our work, but the role it is playing in the EC* beyond that, especially concerning its positioning near the nascent RNA, is unclear at this time.

Recent work by Jeronimo *et al*. (2019) demonstrated that loss of Spt6 broadly affects Paf1C- promoted histone modifications genome-wide (Jeronimo *et al.* 2019), possibly in part due

to its role in promoting Paf1C occupancy. Mutational analysis of Paf1C members in yeast show that loss of Cdc73 or its C-domain only slightly alters Paf1C- promoted histone modification levels, which may be because Rtf1 is still able to remain in contact with EC*. Consistent with that idea, when Cdc73 C-domain is deleted in Rtf1 OAR mutants, a more significant decrease in chromatin occupancy is observed. Additionally, in these double mutations, H3K4me2, H3K4me3, H3K36me3, and H3K79me2/3 marks are lost. This failure to achieve a complete loss of Paf1 occupancy or histone modifications in previously published Cdc73 mutants agrees with our mutational analysis data, where only small decreases in Cdc73/Spt6 were observed.

The Rtf1-OAR helps to maintain Paf1C associated co-transcriptional histone modifications such as H3K4me2, H3K4me3, and H3K79me2/3 (Mayekar *et al.* 2013). However, a domain of Rtf1 termed the HMD (histone modification domain) is far more critical to these modifications as it is required for the upstream H2BK123ub mark. Interestingly, HMD alone is sufficient to target itself to chromatin, likely through its interaction with the acidic patch of the nucleosome (Van Oss *et al.* 2016; Cucinotta *et al.* 2019). However, in this context, the HMD is no longer restricted to coding regions, resulting in aberrant genome-wide H2BK123ub patterns (Van Oss *et al.* 2016). Thus, both the Rtf1 OAR-Spt5 CTR* interactions and HMD-acidic patch interactions are required for proper H2BK123ub, demonstrating how the loss of tethering of the major histone modification promoting component of Paf1C to Pol II can lead to aberrant histone modification patterns.

In conclusion, our findings support roles for both Spt6 and Rpb1 in Paf1C recruitment during transcription elongation via interactions with its Cdc73 subunit. The addition of the role of Spt6 in the Paf1C recruitment process marks a significant advance in our understanding of Paf1C recruitment. The more we learn about the mechanism by which Paf1C is targeted to genes, the

better we will understand how it can influence co-transcriptional histone modification placement and patterning.

4.0 Conclusions and future directions

My thesis work has focused primarily on two scientific questions: 1) how does Paf1C regulate the transcriptome, and 2) how is Paf1C recruited to the transcription elongation complex (more detailed descriptions of specific aims can be found in section 1.4.5). To begin to answer the first question I analyzed data from a transcriptomics experiment that had been collected in the Arndt lab before I arrived. I used my bioinformatic expertise to analyze these data in great detail. I performed follow-up experiments on the most affected and well-studied genes that I identified as differentially expressed in cells deleted for the *PAF1* gene. These model gene sets helped to uncover mechanisms by which Paf1 regulates the transcriptome and illustrated the complexity that exists in its modes of regulation. Numerous Paf1C-regulated non-coding RNAs were identified. I showed that antisense transcription was altered in cells lacking *PAF1* and many of the affected antisense transcripts were known to be regulated by H3K36me3. My data suggest that the loss of H3K36me3 upon *PAF1* deletion contributes strongly to the transcriptome-wide effects we observed. This work connected Paf1's effects on a specific histone modification with changes in transcript levels and demonstrated the requirement for Paf1C in proper transcription regulation genome-wide.

There was already a lot known about the second question posed above, but a new interaction potentially involved in Paf1C recruitment was discovered prior to the onset of my thesis work. I decided to pursue the interaction between the Paf1C component Cdc73 and the essential elongation factor Spt6 as a portion of my thesis work. This project began with a single *in vivo* crosslinking result and ended with a genome-wide and structural understanding of the interaction between Cdc73 and Spt6. Spt6 turned out to be a key factor in Paf1C recruitment. This finding

added a completely new layer of mechanistic understanding to the Paf1C recruitment process and assigned functions to conserved domains in both Cdc73 and Spt6 that were previously unknown.

Both of these questions have been addressed and the answers obtained are summarized in the conclusions section below (4.1). New questions can now be asked based off these results. We currently have students in the lab embarking on their own thesis projects that stem from this work. In the future directions section below (4.2) I will describe the directions that they are heading and numerous additional ideas I have for future projects in the lab.

4.1 Conclusions

This section contains summaries of the major conclusions drawn from the work presented in Chapters 2 and 3 and points out how these findings advance the field of transcription regulation and our understanding of Paf1C function and recruitment. A more detailed account of results and conclusions and their place in the context of the literature can be found in the results and discussion sub-sections of Chapter 2 (2.3 and 2.4) and Chapter 3 (3.3 and 3.4).

4.1.1 Paf1C regulates coding and non-coding RNA levels

Prior to our transcriptomic analysis of a *paf1*Δ strain, little was known about Paf1Cs role in ncRNA transcription beyond its role in snoRNA transcription termination (Sheldon *et al.* 2005; Terzi *et al.* 2011; Tomson *et al.* 2011, 2013). There were few studies in yeast that measured global RNA levels in a *paf1*Δ strain and most of those focused on only mRNA transcripts (Shi *et al.* 1996; Penheiter *et al.* 2005; Porter *et al.* 2005; Batta *et al.* 2011; van Bakel *et al.* 2013; Cao *et al.* 2015;

Chen *et al.* 2015; Yang *et al.* 2016; Harlen and Churchman 2017b; Fischl *et al.* 2017). However, in recent years many papers had been published that defined classes of ncRNA transcripts, many of which were the result of pervasive transcription, and it was unknown what role, if any, Paf1 played in regulating their transcription (Xu *et al.* 2009; van Dijk *et al.* 2011; Schulz *et al.* 2013; Venkatesh *et al.* 2016).

The data presented here demonstrates how important Paf1 and by extension the Paf1C is for both coding and non-coding RNA production in yeast. The data recapitulated the results of other labs with respect to coding transcripts or mRNAs where both up and down regulated transcripts were observed (Shi *et al.* 1996; Penheiter *et al.* 2005; Porter *et al.* 2005; Batta *et al.* 2011; van Bakel *et al.* 2013; Cao *et al.* 2015; Chen *et al.* 2015; Yang *et al.* 2016; Harlen and Churchman 2017b; Fischl *et al.* 2017). However, my results vastly extended our knowledge of the Paf1C's role in ncRNA regulation from previously reported snoRNA extension, to its involvement in both up- and down-regulation of SUTs, CUTs, NUTs, SRATs, and XUTs. This study is the first to report the effects of Paf1 on SUTs, NUTs, SRATs, and XUTs. It also was the first to characterize the relationship between Paf1 and the upstream CUTs that regulate the *FET4* locus.

Antisense transcripts were found to make up a large portion of both annotated and *de-novo* identified ncRNAs upon *PAF1* loss. Among SUTs, CUTs, NUTs, and XUTs there are numerous transcripts that are oriented antisense to a coding region (Xu *et al.* 2009; van Dijk *et al.* 2011; Schulz *et al.* 2013) and by definition SRATs are antisense transcripts (Venkatesh *et al.* 2016). Some of these antisense transcripts were stable and could be detected in the *paf1*Δ strain, although many were unstable and required additional deletion of *TRF4* to enable them to be measured by the tiling array assay. I observed both increased and decreased levels of annotated antisense ncRNAs. Clustering analysis revealed that very few of these antisense transcripts affected mRNA

transcript levels in this bulk cell steady state experiment. Although it is likely that in a single cell antisense transcription interferes with sense transcription, results on a population of cells show this pattern at only a few genes. Among these genes is the well-studied *PHO84* gene, which is known to be regulated by an antisense transcript (Castelnuovo *et al.* 2013). Indeed, this gene grouped with those that show an anticorrelation in our tiling array data. Before this work the effects of Paf1C on antisense transcripts had been demonstrated only at the *ARG1* locus (Crisucci and Arndt 2012). Our study revealed that misregulation of antisense transcription is a wide-spread phenomenon in cells deleted for *PAF1*.

4.1.2 Paf1 loss promotes expression of H3K36me3 repressed transcripts

Loss of Paf1 was known to result in a global decrease in H3K36me3 in yeast (Krogan *et al.* 2003; Chu *et al.* 2007), but how that affected the transcriptome specifically in the context of a *paf1*Δ mutant was not understood. Our work demonstrated de-repression of H3K36me3-repressed transcripts in a *paf1*Δ strain. This informed us on the requirement of Paf1 for this regulatory histone modification and demonstrates one way that loss of Paf1 can lead to transcriptional defects. This was apparent in SRATs, or Set2 repressed antisense transcripts, which generally increased expression upon Paf1 loss (Venkatesh *et al.* 2016). Additionally, our analysis of *FIT3* and *SIT1* transcript levels in *paf1*Δ, *set2*Δ, and *paf1*Δ *set2*Δ strains suggest that the effect of Paf1 loss of these transcripts is likely to be due to loss of H3K36me3. Comparison of our data to NET-seq data further confirmed this result, suggesting that H3K36me3 and the Rpd3S histone deacetylase complex governed the transcription of these transcripts (Churchman and Weissman 2011; Harlen and Churchman 2017b). It was surprising that H3K36me3 was the only modification we could link to transcriptional changes in *paf1*Δ given that Paf1 promotes the levels of other histone

modifications such as H2BK123ub, H3K4me2/3, and H3K79me2/3. Perhaps further analysis of nascent transcripts coupled with ChIP-seq analysis of histone modification profiles in cells lacking Paf1C members will provide more insights in the future.

4.1.3 The Rpb1 CTD and Spt6 interact directly with the Cdc73 C-domain *in vitro* and *in vivo*

We are the first to demonstrate an interaction between Cdc73 and both the Rpb1 CTD (*in vivo*) and Spt6 core (*in vitro* and *in vivo*). This is a very important finding because it begins to explain the mechanism of Paf1C recruitment by Cdc73. As discussed earlier (section 1.4.4.2) much was known about the mechanism of Paf1C recruitment by Rtf1 prior to the work presented here (Mayekar *et al.* 2013; Wier *et al.* 2013). It was also known that deletion of either *CDC73* or *RTF1* decreased Paf1C occupancy (Amrich *et al.* 2012). Importantly deletion or mutation of these two domains together results in an even stronger defect (Amrich *et al.* 2012; Wier *et al.* 2013). This additive phenotype suggested that the two proteins work in separate recruitment pathways. Thus, I set out to characterize the Cdc73-based recruitment pathway.

Results from previous studies demonstrated that the Cdc73 C-domain could bind to Pol II CTD peptides *in vitro* (Qiu *et al.* 2012). Further studies showed that when the C-domain of Cdc73 is deleted, other Paf1C members are improperly recruited to Pol II; however no direct interaction involving the Cdc73 C-domain had been demonstrated *in vivo* before now (Amrich *et al.* 2012). Undergraduates in the lab were able to identify crosslinks between Cdc73 and both Spt6 and the Pol II CTD *in vivo* using the BPA crosslinking technique. Due to the extremely short crosslinking range for BPA we were able to conclude that these interactions were both direct. The CTD

interaction was already shown to occur in the absence of additional factors *in vitro* in previous studies using peptides, so we did not perform further biochemical analysis of that interaction (Qiu *et al.* 2012).

Additional biochemical characterization of the Spt6-Cdc73 interaction was performed using both *in vitro* binding assays and crosslinking and mass spectrometry analysis (XL-MS). XL-MS revealed an interaction between the C-domain of Cdc73 and the Spt6 core, which I describe more in the section below (4.1.5). Additional binding assays were performed to obtain binding affinity information and to rule out the possibility that Cdc73 crosslinked to Spt6 only in the context of Pol II CTD binding. I performed binding assays with recombinant Spt6 and Cdc73. These binding assays revealed a micromolar binding affinity for the Cdc73-Spt6 interaction and confirmed that the two proteins could interact without any other factors present, thus confirming the direct interaction and sufficiency.

A substitution in Cdc73, W321A that causes phenotypes associated with defective transcription greatly reduced the Cdc73-Spt6 and Cdc73-Rpb1 interactions *in vivo*, as measured by BPA crosslinking. I also tested this mutant further *in vitro* and found that it resulted in a 5-fold decrease in Cdc73-Spt6 binding affinity. Chromatin immunoprecipitation (ChIP) results revealed a decrease in chromatin occupancy at *PMA1* and *PYK1* for the Cdc73 W321A mutant protein, suggesting impaired Cdc73 recruitment. The results of the ChIP experiment and the results of the binding assays with the Cdc73 W321A mutant protein are consistent with the results obtained by BPA crosslinking and suggested that the C-domain was the domain of Cdc73 interacting with Spt6.

The Spt6-Cdc73 interaction results by themselves represented a large step forward in the study of Paf1C recruitment because a new interaction was uncovered and the domain of Cdc73 that is responsible for the interaction was identified. The ChIP results demonstrate that this

interaction is relevant to Cdc73 recruitment at highly transcribed genes. Additionally, the work our undergraduates performed identified the CTD of Rpb1 as a C-domain interaction partner and represents a significant finding; providing strong evidence for the Cdc73-CTD interaction that was assumed to exist, but had never been rigorously tested *in vivo*. These findings advanced the field by identifying two key interactions involved in recruitment of Paf1C by way of its Cdc73 subunit. Furthermore, the Cdc73-Spt6 interaction is novel and had never been considered to have a role in Paf1C recruitment before this work.

4.1.4 Spt6 plays a prominent role in Paf1C recruitment during transcription elongation

Having discovered that the Cdc73 C-domain bound to Spt6 and confirmed that this interaction occurs *in vitro* in the absence of all other factors I set out to assess the effects of loss of this interaction on genome-wide Paf1C occupancy. As an initial assessment, I performed a ChIP-qPCR experiment in an Spt6 mutant that lacks the tSH2 domain (*spt6-50)*, which is known to interact with the Rpb1 CTD linker region. My preliminary analysis at the *PMA1* gene suggested that Spt6 played a major role in Paf1C recruitment with both Paf1 and Rtf1 occupancy levels dramatically decreased in the mutant. When I tested this mutant by ChIP-seq I noticed that Spt6 levels were not reduced at all genes and in fact increased at moderately to lowly transcribed genes. At the most highly expressed genes (top 10-20% most Pol II occupied) there was a large decrease in both Spt6 and Paf1 occupancy. Although the Spt6 result was surprising, it was clear that Spt6 and Paf1 occupancies were correlated at all genes regardless of expression level.

These results were encouraging so I decided to test Paf1 occupancy in a strain that was engineered to rapidly deplete the essential Spt6 protein. For this I used the auxin inducible degron system and was able to achieve complete loss of Spt6 in one hour as measured by western blot.

ChIP-qPCR and ChIP-seq results after one hour of auxin treatment confirmed that Spt6 plays a major role in Paf1 recruitment. I observed a genome-wide loss of Spt6 at all coding regions and a concomitant decrease in Paf1 occupancy in Spt6 depleted cells. Interestingly, I did not observe a decrease in Pol II or Spt5 occupancy at most genes, although slight occupancy decreases were observed for both at them most highly expressed genes. I tested Pol II and Spt5 because they both participate in Paf1C recruitment and I wanted to rule out the possibility that reduced occupancy of Pol II or Spt5 upon depletion of Spt6 was causing the observed decrease in genome-wide Paf1 occupancy.

The results of the ChIP-seq experiments in both the *spt6-50* mutant and AID contexts demonstrated the important role that Spt6 plays in Paf1C recruitment. It was surprising to see that Spt5 and Pol II interactions were not sufficient to maintain high Paf1 levels upon depletion of Spt6. This suggests that Spt6 might play a larger role in Paf1C recruitment than the known recruitment mechanism involving Spt5 and Rtf1 (Mayekar *et al.* 2013; Wier *et al.* 2013). These findings are significant and inform the field on the role of Spt6 in Paf1C recruitment on a genome-wide scale, presenting a novel function for Spt6 and uncovering further mechanistic details in the Paf1C recruitment process.

4.1.5 Spt6 HhH, DLD, and YqgF domains interact with the Cdc73 C-domain near the RNA exit site of Pol II

Spt6 contains multiple conserved domains many of which have not been assigned a function, although there are hints that many may be involved in protein-protein or protein-nucleic acid interactions based on structural and sequence conservation (Close *et al.* 2011). I wanted to determine which of these domains interacted with the Cdc73 C-domain and to do that I performed

XL-MS. XL-MS is a low-resolution structural technique that allows for a structural model to be generated based off chemical crosslinking data (Shi *et al.* 2015). I was able to perform this analysis and obtain useful information through collaboration with the lab of Dr. Yi Shi. Structural modeling was greatly aided by the crystal structures for the Spt6 core and tSH2 domains and the Cdc73 C-domain published by the labs of Dr. Christopher Hill (Close *et al.* 2011) and Dr. Andrew VanDemark (Amrich *et al.* 2012).

Modelling carried out using the Integrative Modelling Platform placed the C-domain in a large valley in the Spt6 core in between the YqgF, HhH, and DLD domains. These domains appear to coordinate binding of the Cdc73 C-domain. Alanine scanning substitutions, changing three to five residues at a time were made to each domain at the interaction interface. These mutations result in a small reduction in Cdc73 occupancy as measured by ChIP-qPCR, which is more clearly observed when the ratio of Cdc73 to Spt6 is considered. This small but reproducible reduction in Paf1C member occupancy is similar to what is observed in individual Cdc73 or Rtf1 point mutations. Additionally, mutations to the Cdc73 C-domain also show reductions in Cdc73 occupancy. Although this is not a high-resolution model and orientation of the two proteins could be slightly off, I believe that the general positioning of the C-domain on the Spt6 core is accurate based off this mutational analysis.

When we align the Spt6 core in our model with that of the current model of the Pol II elongation complex (Vos *et al.* 2018b) we are able to place the Cdc73 C-domain near the RNA exit tunnel of Pol II. It is important to note that the binding interface we propose is completely free in the current model of the elongation complex. The only region of Cdc73 shown in the current model is quite far from where we place the C-domain but there are more than enough residues to reach between the two locations. Thus, I believe that our model is likely to reflect the true

positioning of the Cdc73 C-domain on the Pol II elongation complex. This work extends our knowledge of the Pol II elongation complex by placing the Cdc73 C-domain into the model and raises more questions about this domain's functional role near the RNA exit site.

4.2 Future directions

This section contains suggested future experiments to continue to advance the work presented here. The ideas presented all come from observations raised in the previous chapters. I describe the rationale for asking the question and experimental methods that could be used to answer the question. Some of these future directions are underway in the lab with some of the newer graduate students heading up the projects, but most are completely new investigations that I hope someone might want to undertake someday.

4.2.1 Nascent transcriptome sequencing of Paf1C upon rapid depletion

Although we uncovered large effects on the transcriptome from assessing strains deleted for *PAF1*, deletion strains such as these are able to adapt to the loss of this key component of the Paf1C. It might be interesting to investigate the transcriptome of a strain where Paf1 can be rapidly depleted from the cell or nucleus using techniques such as auxin inducible degradation (AID) or anchor-away, respectively. It is possible that the deletion strain has adapted to loss of Paf1 very well and we are only able to observe the minimal effect of Paf1 loss in our tiling array experiments. If that is the case, then what is the full and immediate response to Paf1 loss? In preparation to test this hypothesis, I generated AID strains for each Paf1C member tagged at their C-terminus with

IAA7 (an auxin responsive protein) to allow for rapid depletion of individual Paf1C members. I also pioneered the use of AID in the lab, establishing protocols for strain validation and usage. Therefore, this kind of experiment could easily be pursued in the future.

We measure steady-state RNA levels by tiling array and RT-qPCR, which in many cases is not the best measure of effects on transcription as RNA stability can skew the results in this kind of experiment. It is likely that combining rapid depletion with a nascent transcript sequencing technique such as 4-tU-seq or NET-seq would be ideal to capture the true effect of loss of Paf1 on unadapted cells. By comparing steady state RNA results to nascent RNA results it is possible to determine which RNAs are affected directly by Paf1 and which are affected indirectly by post transcriptional mechanisms such as RNA stability.

I observed independent effects of individual Paf1C member deletions by RT-qPCR in my work. This is extremely interesting and begs the question: what are the functions of the individual Paf1C members? Toward answering this question, it may be of interest to test all Paf1C members using rapid depletion and nascent transcriptome sequencing in order to determine genome-wide subunit-specific effects on the transcriptome. Indeed, a new member of the lab, Alex Francette, is investigating this question in this manner and I look forward to seeing his results published in the future.

4.2.2 Investigate the mechanism of snoRNA termination at Paf1-dependent and independent genes

We identified two classes of snoRNAs in a *paf1*Δ strain, those that are upregulated and 3' extended and those that are not. We have known for some time now that snoRNAs experience read-through transcription resulting in extended 3' ends in individual Paf1C complex member

deletions and mutants affecting H2B K123ub (Tomson *et al.* 2011, 2013a). The mechanism by which this phenotype occurs is still unclear, although it is presumed to result from a loss of physical or genetic interactions between Paf1 and termination factors in the deletion mutant context.

RNA-seq and tiling array datasets for many factors involved in transcription elongation and termination have been published and exist in the Gene Expression Omnibus (GEO) database. It is possible that an in-depth analysis of the snoRNAs across numerous datasets could provide a hint at the mechanism. We have seen different snoRNAs affected differently in different contexts in the lab. For example, Dr. Elizabeth Hildreth showed numerous snoRNAs were upregulated and had extended 3' ends in her thesis work studying mutants of histone H3 involved in transcription termination. However, the affected snoRNAs in her work and mine are not all the same, although there is some overlap. It may be possible to identify other factors that phenocopy Paf1 by analyzing gene expression data at the snoRNA loci and performing hierarchical or k-means clustering. This could reveal clusters of snoRNAs that behave similarly in termination mutants and Paf1C mutants, for example. Termination factors behaving similarly could then be studied further and mutated in combination with Paf1 in order to determine genetic relationships and begin to uncover the mechanisms by which loss of Paf1 results in extended snoRNA 3' ends.

4.2.3 Investigate the role of Spt5 in recruitment of Spt6 to Pol II using rapid Spt5 depletion and CTR mutants

Spt5 is known to play a large role in Paf1C recruitment, but in the experiments conducted here it appears as though Spt6 is playing a larger role. My results clearly show that upon auxin depletion of Spt6, Spt5 occupancy is unaffected. We do not know if Spt5 affects Spt6 occupancy. Additionally, my work and previously published work suggests that Spt5 and Spt6 are both

necessary for proper Paf1C recruitment and that neither alone is sufficient to sustain normal levels of Paf1C on chromatin. These results suggest two testable models: 1) Spt5 and Spt6 are recruited at the same time, but independently and then together they recruit Paf1C to Pol II, or 2) Spt5 recruits Spt6 and then together they recruit Paf1C to Pol II.

The known recruitment mechanism for Spt6 is via an interaction between its tSH2 domain and the Pol II CTD linker (Sdano *et al.* 2017). However, based on our Spt6 results and results for Spt6 N-terminal domain mutants (Dronamraju *et al.* 2018c) it appears that the tSH2 domain is not the only domain of Spt6 that allows it to associate with chromatin. Indeed, in the current structure of the active Pol II elongation complex the Spt6 core is interacting with the Spt5 KOW domains and the Pol II stalk (Vos *et al.* 2018b). ChIP-seq (presented here) and ChIP-chip profiles (Mayer *et al.* 2010) for Spt5 and Spt6 appear to start in the same location relative to the transcription start site.

In a recent rotation project Kayla Komondor created and validated the Spt5-AID strain and performed ChIP-qPCR on Rpb3, Spt5, Spt6, Paf1 and Rtf1. Her preliminary data appeared to support model two, but the work needs to be repeated. By performing ChIP-qPCR and ChIP-seq on the Spt5-AID strain with and without auxin treatment and assaying Rpb1, Spt5, Spt6, and Paf1 we will be able to identify if Spt5 is necessary for recruitment of Spt6 (model two) or not (model one). Our newest student Sanchirmaa Namjilsuren is currently spearheading this project.

Strains with integrated copies of Spt5 alleles containing mutant forms of the Spt5 CTR that are either unable to be phosphorylated or phosphomimetic (imitating constitutive phosphorylation) exist in the lab. If the Spt5-AID experiment proposed above reveals that Spt6 depends on Spt5 for its recruitment, in addition to its known interactions with chromatin via the histones and the CTD linker, then it will be important to assay these mutants for Spt6 occupancy levels to determine if

the Spt5 CTR is involved in Spt6 recruitment. It is important to note here that a recent publication demonstrated that the tSH2 domain binds promiscuously to peptides phosphorylated in many different configurations, meaning that although the CTD and its linker are the canonical targets of the tSH2 domain this does not necessarily rule out the Spt5 CTR playing a role in its initial recruitment (Brázda *et al.* 2020).

4.2.4 Investigate chromatin occupancy patterns of Spt6 when key interacting domains are deleted

The results obtained in the *spt6-50* strain lacking the tSH2 domain clearly demonstrate that the tSH2 domain is not the only way that Spt6 is able to interact with chromatin. When the N-terminus is deleted Spt6 loses its ability to interact with nucleosomes and Spn1 (McDonald *et al.* 2010) and when mutated its genome-wide occupancy profiles are altered (Dronamraju *et al.* 2018b). Additionally, the Spt6 core is clearly shown to interact with the Pol II stalk and the KOW domains of Spt5 in the most recent Pol II elongation complex model (Vos *et al.* 2018b). Therefore, at least three regions of Spt6 are important for its interactions with Pol II and transcribed chromatin. Most studies consider the tSH2 domain to be the primary region involved in recruitment of Spt6, which often appears to be the case. Various tSH2 truncation mutants have been tested with slightly different boundaries, however, in all these reports data for only the most highly expressed genes are presented. Consistent with these published studies, in my *spt6-50* dataset, I observed a large reduction in Spt6 occupancy at highly expressed genes. However, when I looked at the moderate to lowly expressed genes, I saw an increase in Spt6 occupancy. Some hypotheses as to how this occurs at low to moderately expressed genes are: 1) the tSH2 domain is necessary to prevent Spt6 from remaining attached to histone proteins or Pol II, or 2) Spt6 cycles between

primarily histone bound and primarily Pol II bound states and loss of the tSH2 domain throws that balance out of equilibrium. Another hypothesis I have is that highly expressed genes behave differently because a different or modified mechanism is required to deal with the high levels of Pol II.

These hypotheses can be at least partially addressed by performing comprehensive genomics analyses of *SPT6, spt6-ΔN300, spt6-ΔtSH2, and spt6-ΔN300-ΔtSH2* strains. ChIP-seq for Spt6 and Rpb1 along with MNase-seq or histone H3 ChIP-seq could be performed on these strains. This experimental design will allow for the contribution of the N-terminal 300 amino acids (deleted in *spt6-ΔN300*), tSH2 domain and the core to be assessed with respect to Spt6 occupancy and allow for Spt6 occupancy patterns to be compared to Pol II and histone occupancy levels genome-wide. By performing correlation and clustering analyses on all mRNA encoding genes across these three datasets insights will be gained into how Spt6 interacts with the chromatin landscape by way of Pol II and histone proteins. Additionally, candidate loci can be identified for follow-up studies and, if appropriate, selected for use in building genetic reporters.

4.2.5 Determine if the tSH2 domain serves as an autoinhibitory domain for the Spt6-Pol II interaction

In the structural model of the Spt6-Cdc73 interaction I developed in collaboration with Dr. Yi Shi's lab the tSH2 domain is placed on the side of the Spt6 core in a way that would sterically hinder the Pol II-Spt6 core interaction (discussed briefly in 3.4). I hypothesize that the Spt6 tSH2 domain acts as an autoinhibitory domain functioning to prevent the Spt6 core from interacting with Spt5 and the Pol II stalk under certain conditions. The tSH2-core interaction may even promote rapid binding of the core to the Pol II stalk upon recruitment if, for example, the tSH2-CTD

interaction releases the core in close proximity to the stalk of Pol II allowing for a rapid transition. I think that it is likely that Spt6 undergoes some conformational changes to allow for interactions with both Pol II and the histone proteins during transcription elongation and I think that the tSH2 domain might be helping to control these exchanges. At this stage this is simply a hypothesis and I do not yet have a good way to test it other than the experiments described in section 4.2.4, but I think it is a hypothesis worth considering given the ChIP-seq and XL-MS results presented here.

We know that losing the tSH2 domain at highly expressed genes does not result in retention of Spt6 and in-fact results in a reduction of Spt6 occupancy. In contrast, at the lowly expressed genes we observe an increase in Spt6 occupancy, perhaps supporting the possibility that loss of this domain results in an uninhibited interaction between Pol II and the Spt6 core or a shifted equilibrium that favors the Pol II-Spt6 core interaction or N-terminus-histone or -Spn1 interaction. In either case the tSH2 domain may be playing a role that has not been considered previously which could prove to be an interesting avenue of investigation.

4.2.6 Determine the function of the DLD in transcription regulation

During my work studying the binding interface shared between Spt6 and Cdc73 I generated a mutant of Spt6 lacking the death-like domain, referred to as *spt6-ΔDLD,* using the plasmid shuffle technique. This strain showed a very strong Spt- phenotype and little is known about the DLD of Spt6. Therefore, I propose that experiments be conducted to attempt to characterize this domain.

One way to do this is to perform further phenotypic analysis of this mutant. One easy assay would be spot tests in a strain containing a cryptic initiation reporter. More difficult, but well within the Arndt labs skill set would be to perform Spt6 ChIP-seq, MNase-seq, and 4tU-seq in this

mutant in order to determine the genome- and transcriptome-wide effects of loss of the DLD. This would likely produce publication quality results relatively quickly and help to identify the mechanism by which the DLD is affecting gene expression at the *his4-912∂* locus.

Another way to do this would be to try to identify a binding partner other than Cdc73 interacting with Spt6 at the DLD. This may take a while but could potentially be very rewarding if a novel interaction is uncovered. Identification of an interacting partner could be accomplished by placing BPA in and near the DLD in an otherwise wild-type Spt6 context and checking to see if crosslinks can be detected by western blotting. This is a difficult experiment because of the size of Spt6 and the limits of polyacrylamide gel electrophoresis, but because we were able to capture the Cdc73-Spt6 interaction by BPA crosslinking I think it is feasible. If a crosslink is discovered the band can be excised from the gel and sent out for mass spectrometry analysis to identify the interacting protein. Alternatively, a yeast 2-hybrid (Y2H) screen could be conducted using a Y2H plasmid library for prey and the DLD as bait. This would identify target proteins for follow-up analyses. Y2H is much easier than the BPA method, but also has a much higher potential for false positives. If an interaction partner is discovered follow-up analyses can be performed by conducting co-IPs. Characterization of the phenotypes associated with the DLD and identification of binding partners in addition to Cdc73 will add to our understanding of this highly conserved domain of Spt6.

4.2.7 Revisit the role of Bur1/2 in Paf1C recruitment

Bur1/2 phosphorylates all relevant proteins for Cdc73 and Rtf1 based recruitment of Paf1C including the Spt5 CTR, the Rpb1 CTD (at serine 2), and its linker region. We know that *BUR2* deletion or Bur1 inhibition leads to reduced recruitment of Paf1C and prior to this work it was

believed that this occurred through loss of phosphorylation of the CTD and the CTR (Amrich *et al.* 2012; Mayekar *et al.* 2013; Wier *et al.* 2013). Recently studies have been published showing that Spt6 binds to the phosphorylated CTD linker and that Bur1/2 kinase is responsible for phosphorylation of the linker (Sdano *et al.* 2017; Chun *et al.* 2019). In addition, recent advances kinase in inhibition have allowed for the generation of irreversibly sensitized or IS versions of all the major CTD kinases including Bur1. Dramatic effects are observed after only 15 minutes of treatment with CMK, which is an inhibitory ATP analog. This allows for the immediate effects of loss of kinase activity to be observed (Chun *et al.* 2019).

I propose that Bur1-IS strains be used to determine the effects of loss of phosphorylation on Paf1C recruitment factors using ChIP-seq as the readout. ChIP-seq of + and – CMK treated cells on Pol II (Rpb1), Spt5, Spt6, and Paf1 occupancy genome-wide can be used to determine the effects of loss of Bur1 activity on protein factors relevant to Paf1C recruitment. The effects on Spt6 in this experiment should at least be as strong as the effect on Spt6 occupancy in the *spt6-50* mutant. To determine CTR dependent recruitment and Spt6 dependent recruitment the Bur1-IS strain can be crossed with a *spt6-50* strain or a strain containing an unphosphorylatable CTR, where all serine's have been substituted for alanine. In these contexts, ChIP-qPCR and ChIP-seq experiments can be performed to measure the effect of loss of Bur1 activity on Paf1C recruitment.

Bur1-IS +CMK will likely abolish occupancy of Paf1C if the by reducing Spt6 levels due to loss of Rpb1 linker phosphorylation and by removing Spt5 CTR phosphorylation marks. We can confirm that the observed occupancy decrease is indeed the result of loss of these two binding events by demonstrating that occupancy profiles do not change if this same experiment is done in a background also containing mutant forms of Spt6 and Spt5 lacking the tSH2 and phosphorylatable CTR respectively. Furthermore, if Spt6 is not only recruited by its tSH2 domain,

it appears not to be based of the *spt6-50* data, we will be able to determine other Bur1 kinase targets play a role by inhibiting the kinase in an *spt6-50* background. We will also be able to take this a step further to determine if the second kinase target is the Spt5 CTR by performing the same Bur1 inhibition experiment in a strain lacking the tSH2 and a phosphorylatable CTR. This type of experiment will provide a more detailed mechanistic understanding of the Paf1C recruitment mechanism with respect to the roles of Spt5, Spt6, Rpb1, and Bur1/2 kinase.

4.3 Concluding remarks

The thesis work presented here has uncovered the widespread impact of Paf1 on the yeast transcriptome and identified a previously unrecognized mechanism for Paf1C recruitment to the Pol II elongation complex. The transcriptomics work taught us that Paf1C regulates ncRNA transcription in addition to mRNA transcription and demonstrated that when Paf1 is not available to promote H3K36me3 we see rampant pervasive transcription genome-wide. The recruitment study has added a significant amount of information to our understanding of the mechanisms that constrain Paf1C to active genes, where it functions to regulate chromatin structure. Prior to this work Spt6 was not thought to play a role in Paf1C recruitment and now it appears to be as important, if not more important, than the previously known recruitment factors. Both of these projects have answered key questions and unearthed novel functions for transcription elongation factors and raised new and interesting questions. I look forward to seeing how this work progresses as future Arndt lab members tackle the new questions generated here.

Appendix A

This appendix has been accepted for publication at the journal *G3* or *Genes Genomes and Genetics* as a Software Resource article entitled "MutantHuntWGS: A Pipeline for Identifying *Saccharomyces cerevisiae* Mutations" and is presented in altered form here. This is in accordance with the journals permissions statement which states: "G3 articles are published as open-access articles distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited."

Appendix A.1 Introduction

Saccharomyces cerevisiae is a powerful model system for understanding the complex processes that direct cellular function and underpin many human diseases (Birkeland *et al.* 2010; Botstein and Fink 2011; Kachroo *et al.* 2015; Hamza *et al.* 2015, 2020; Wangler *et al.* 2017; Strynatka *et al.* 2018). Mutant hunts (*i.e.*, genetic screens and selections) in yeast have played a vital role in the discovery of many gene functions and interactions (Winston and Koshland 2016). A classical mutant hunt produces a phenotypically distinct colony derived from an individual yeast cell with at most a small number of causative mutations. However, identifying these mutations using traditional genetic methods (Lundblad 1989) can be difficult and time-consuming (Gopalakrishnan and Winston 2019).

Whole-genome sequencing (WGS) is a powerful tool for rapidly identifying mutations that underlie mutant phenotypes (Smith and Quinlan 2008; Irvine *et al.* 2009; Birkeland *et al.* 2010). As sequencing technologies improve, the method is becoming more popular and cost-effective (Shendure and Ji 2008; Mardis 2013). WGS is particularly powerful when used in conjunction with lab-evolution (Goldgof *et al.* 2016; Ottilie *et al.* 2017) or mutant-hunt experiments, both with (Birkeland *et al.* 2010; Reavey *et al.* 2015) and without (Gopalakrishnan *et al.* 2019) bulk segregant analysis.

Analysis methods that identify sequence variants from WGS data can be complicated and often require bioinformatics expertise, limiting the number of investigators who can pursue these experiments. There is a need for an easy-to-use, data-transparent tool that allows users with limited bioinformatics training to identify sequence variants relative to a reference genome. To address this need, we created MutantHuntWGS, a bioinformatics pipeline that processes data from WGS experiments conducted in *S. cerevisiae*. MutantHuntWGS first identifies sequence variants in both control and experimental (*i.e.* mutant) samples, relative to a reference genome. Next, it filters out variants that are found in both the control and experimental samples while applying a variant quality score-cutoff. Finally, the remaining variants are annotated with information such as the affected gene and the predicted impact on gene expression and function. The program also allows the user to inspect all relevant intermediate and output files.

To enable quick and easy installation and to ensure reproducibility, we incorporated MutantHuntWGS into a Docker container (https://hub.docker.com/repository/docker/mellison/mutant_hunt_wgs). With a single command, users can download and install the software. A second command runs the analysis, performing all

steps described above. MutantHuntWGS allows researchers to leverage WGS for the efficient identification of causal mutations, regardless of bioinformatics experience.

Appendix A.2 Methods

Appendix A.2.1 Pipeline overview

The MutantHuntWGS pipeline integrates a series of open-source bioinformatics tools and Unix commands that accept raw sequencing reads (compressed FASTQ format or .fastq.gz) and a text file containing ploidy information as input, and produces a list of sequence variants as output. The user must provide input data from at least two strains: a control strain and one or more experimental strains. The pipeline uses (1) Bowtie2 to align the reads in each input sample to the reference genome (Langmead and Salzberg 2012), (2) SAMtools to process the data and calculate genotype likelihoods (Li *et al.* 2009), (3) BCFtools to call variants (Li *et al.* 2009), (4) VCFtools (Danecek *et al.* 2011) and custom shell commands to compare variants found in experimental and control strains, and (5) SnpEff (Cingolani *et al.* 2012) and SIFT (Vaser *et al.* 2016) to assess where variants are found in relation to annotated genes and the potential impact on the expression and function of the affected gene products (Figure 36).

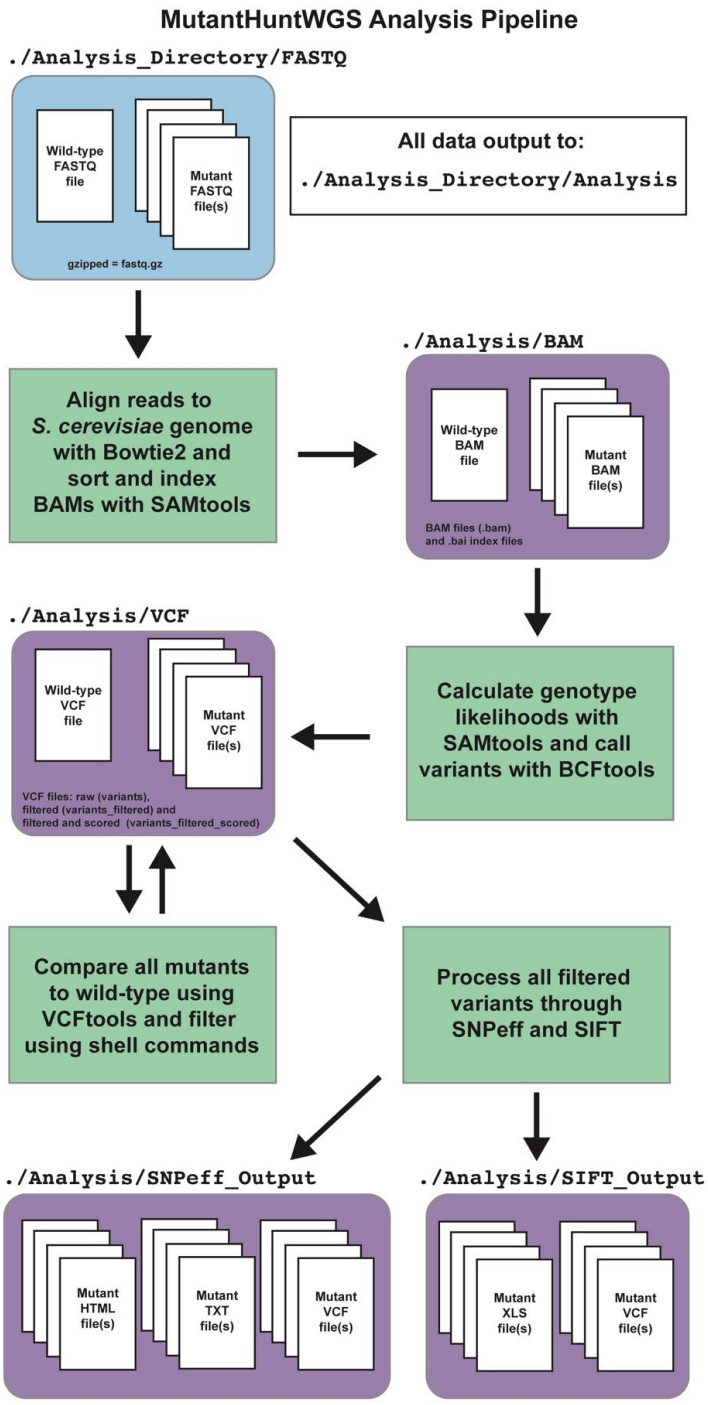


Figure 36. Flow chart of the MutantHuntWGS pipeline.

Input data are colored in blue, the various bioinformatics tools in the pipeline are colored in green, and output data are colored in purple. Arrows identify the path of the workflow at each step of the pipeline.

Appendix A.2.2 Sequence alignment

MutantHuntWGS uses Bowtie2 version 2.2.9 (Langmead and Salzberg 2012) to first align the raw reads present in the input FASTQ files to the *S. cerevisiae* genome (S288C version = R64-2-1) (Cherry *et al.* 2012; Engel *et al.* 2014). As the default, we set Bowtie2 to search for a maximum of two distinct alignments per read (-k 2 option), which reduces alignment time and multiple mapping. The pipeline only retains sequencing reads that align (--no-unal option) in the SAM (Sequence Alignment/Map Format) output to help reduce file size. MutantHuntWGS uses SAMtools version 1.3.1 to convert the aligned-read output from Bowtie2 (SAM format) into the BAM (Binary Alignment/Map) format (view -bS options) (Li *et al.* 2009). SAMtools then sorts (sort option) and indexes (index option) the BAM file to prepare the data for variant calling. Users can view the sorted and indexed BAM files in a genome browser such as IGV (Integrative Genome Viewer) to examine the aligned reads (Thorvaldsdottir *et al.* 2013).

Appendix A.2.3 Variant calling

Based on the aligned reads, SAMtools outputs genotype likelihoods as BCF (Binary Call Format) files (mpileup -g -f options) using the BAM file as input (Li *et al.* 2009). BCFtools version 1.3.1 then uses the genotype likelihoods recorded in the BCF file to call single nucleotide polymorphisms (SNPs), as well as insertions and deletions (INDELs) (-c -v --samples-file –ploidy-file options) (Li 2011). This variant information is saved in the Variant Call Format (VCF), the format used by the 1000 Genomes Project (Danecek *et al.* 2011). IGV can again be used to view the VCF files that are output from MutantHuntWGS (Thorvaldsdottir *et al.* 2013). At the

variant calling step, BCFtools also considers a user-specified input ploidy file to account for genome copy number.

Appendix A.2.4 Identifying candidate variants

VCFtools version 0.1.14 compares VCF files (--diff-site option) from the control and experimental samples (Danecek *et al.* 2011). To retain the variants that are found only in the experimental dataset, MutantHuntWGS uses the Unix awk command (Aho *et al.* 1979) to remove variants from the VCFtools output that have VCF scores lower than a user-defined variant-quality-score cutoff. For each experimental dataset, it then uses the Unix grep, head, and cat commands to construct new VCF files that contain only the variants specific to the experimental strain. These VCF files can also be viewed in IGV.

Appendix A.2.5 Variant effect prediction

SnpEff version 4.3p (Cingolani *et al.* 2012) and SIFT4G (i.e., SIFT) (Vaser *et al.* 2016) are useful programs for (1) determining whether sequence variants are located in or near an annotated coding region and (2) predicting the effect the variant might have on gene expression or function of the protein product. SnpEff determines the locations of sequence variants relative to protein-coding genes and the severity of each variant based on how likely it is to disrupt gene expression or function (Cingolani *et al.* 2012). SnpEff also annotates variants in 5' and 3' UTRs as well as promoter regions. This information is vital if the causal mutation disrupts a ncRNA or DNA element rather than altering a protein-coding sequence. SIFT uses the EF4.74 library for *S. cerevisiae* to score variants found in protein-coding genes in order to predict the impact of the

resulting amino acid changes (Vaser *et al.* 2016). MutantHuntWGS saves all SnpEff and SIFT output files so the user can further filter the results to reduce the number of candidate sequence variants identified.

Appendix A.2.6 Analysis of previously published data

To demonstrate utility, we used MutantHuntWGS to analyze published datasets from paired-end sequencing experiments with DNA prepared from bulk segregants or lab-evolved strains (Birkeland *et al.* 2010; Goldgof *et al.* 2016; Ottilie *et al.* 2017). These data were downloaded from the sequence read archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra; project accessions: SRP003355, SRP074482, SRP074623) and decompressed using the SRA toolkit (https://github.com/ncbi/sra-tools/wiki). MutantHuntWGS was run from within the Docker container, and each published mutant (experimental) file was compared to its respective published control. When processing data from bulk segregant analysis, we reduced the number of candidate variants by additionally using more stringent cutoffs: variant quality score > 130, SnpEff impact score > Moderate, and SIFT score < 0.05 (deleterious).

Appendix A.3 Results and discussion

Appendix A.3.1 Easy installation and reproducible analysis through containerization

To facilitate distribution and maximize reproducibility, we implemented MutantHuntWGS in a Docker container (Boettiger 2015; Di Tommaso *et al.* 2015). The container houses the pipeline

and all of its dependencies in a Unix/Linux environment. To download and install the MutantHuntWGS Docker container, users need only install Docker Desktop (https://docs.docker.com/get-docker/), open a command-line terminal, and execute the following command:

```
$docker run -it -v
/PATH_TO_DESKTOP/Analysis_Directory:/Main/Analysis_Directory
mellison/mutant_hunt_wgs:version1
```

After download and installation, the command opens a Unix terminal running in the Docker container so users can begin their analysis.

Reproducibility in genomic analyses is a growing concern (Kim *et al.* 2018), and containerization is one solution (Boettiger 2015; Di Tommaso *et al.* 2015). By running the MutantHuntWGS pipeline within the standardized software environment that our Docker container provides, we ensure that all users reproducibly apply the same bioinformatics approach to their data (Kim *et al.* 2018).

Appendix A.3.2 Running the MutantHuntWGS pipeline

From the Unix terminal running in the Docker container, users need only execute the following command to run the MutantHuntWGS pipeline:

```
$ MutantHuntWGS.sh \
  -n FILENAME \
  -g /Main/MutantHuntWGS/S_cerevisiae_Bowtie2_Index_and_FASTA/genome \
```

```
-f /Main/MutantHuntWGS/S_cerevisiae_Bowtie2_Index_and_FASTA/genome.fa \
-r single \
-s 100 \
-p /Main/MutantHuntWGS/S_cerevisiae_Bowtie2_Index_and_FASTA/ploidy_n1.txt \
-d /Main/Analysis_Directory \
-o /Main/Analysis_Directory/NAME_YOUR_OUTPUT_FOLDER
-a YES
```

A detailed description of installation and usage is available on the MutantHuntWGS Git repository (https://github.com/mae92/MutantHuntWGS/blob/master/README.md). Briefly, the user must provide separate sequencing data files in FASTQ format (gzipped; filename.fastq.gz) for the control and each experimental strain. Directory paths to genome sequences (-f), indexes (-g), and ploidy files (-p) are required, along with the FASTQ file directory (-d) and output directory path (-o). The control file prefix is given as input (-n) to allow the user to specify a reference strain. A variant quality-score cutoff is set to allow the user to adjust the analysis stringency (-s). The score is calculated as $-10 \times log_{10}(P)$, where P is the probability that the alternate allele (the sequence variant or mutation) is called incorrectly. The user must select whether to process paired-end or single-end sequencing reads (-r). Finally, we added an option (-a) to bypass the alignment and variant calling steps when running MutantHuntWGS a second time. For example, users should specify (-a NO) if they wish to simply reset the score option and rerun the analysis, without repeating the more time-consuming steps of the analysis protocol.

Appendix A.3.3 Utility of the MutantHuntWGS pipeline

MutantHuntWGS processes WGS data through a standard alignment/variant-calling pipeline and compares each experimental strain to a control strain (Figure 19, described in detail in the Methods). The pipeline's constituent tools are often used for WGS analysis (Reavey *et al.* 2015; Gopalakrishnan and Winston 2019; Gopalakrishnan *et al.* 2019). However, MutantHuntWGS ensures ease of use by assembling these tools in a Docker container and requiring only one command to run them all in sequence. This approach combines the best aspects of previously published pipelines (discussed below) while allowing even inexperienced users to install the software and reproducibly apply popular methods.

MutantHuntWGS also ensures that the output data files are well organized and easy to locate. Output files include aligned reads (BAM format), alignment statistics (TXT format), pre- and post-filtering variants (VCF format), SnpEff output (HTML, VCF, and TXT formats), and SIFT output (VCF, XLS formats). The user thus has all the information needed to identify and visually inspect sequence variants, and to generate figures and tables for publication.

Appendix A.3.4 Manual inspection of data using Integrative Genomics Viewer (IGV)

To manually inspect variants and associated raw data, and to identify larger whole-gene deletions or copy-number variations, users can load the VCF and BAM output files into a genome browser (IGV) (https://software.broadinstitute.org/software/igv/download). Deleted genes and common auxotrophic markers can serve as internal controls as these will lack read coverage in the BAM file. IGV visualization allows users to generate publication-quality graphics for individual genomic loci. Users can also access the statistics stored in the VCF and BAM files by simply

selecting a read or variant to inspect. Information such as allele frequency (within the sequencing dataset), variant-quality score, and read-quality scores can also be viewed.

Appendix A.3.5 MutantHuntWGS combines versatility and simplicity

Our goal in creating MutantHuntWGS was to simplify the installation and usage of robust bioinformatics tools while maintaining flexibility by allowing users to specify certain critical options. Examples of this, discussed below, include (1) enabling use with additional organisms, (2) allowing users to specify ploidy, (3) filtering by a user-specified variant-quality score, and (4) exposing all intermediate and final output files to facilitate additional filtering and quality control.

MutantHuntWGS is designed for use with *S. cerevisiae* by default but can be adapted to analyze WGS data from any organism. At present, only the necessary reference files for *S. cerevisiae* are included in the MutantHuntWGS download. Investigators who wish to analyze data from an organism other than *S. cerevisiae* need to provide, at minimum, new Bowtie2 indices, a genome FASTA file, and a ploidy file. Bowtie2 indices and genome FASTA files for many model organisms are available at https://support.illumina.com/sequencing/sequencing_software/igenome.html. A FASTA index file (genome.fasta.fai) that can be easily converted into a ploidy file is also available at this link. Unfortunately, performing the SnpEff and SIFT analysis would require slight alterations to the SnpEff and SIFT commands in the pipeline script and a copy of the SIFT library for the organism of interest. We chose not to include reference files and SIFT libraries for other organisms within the Docker container due to the large size of these files. If users encounter difficulties when analyzing non-*S. cerevisiae* WGS data, we encourage them to seek assistance by opening an issue on the MutantHuntWGS Git repository.

Experiments in yeast are often performed in a haploid background, but can also be performed in diploid or occasionally aneuploid backgrounds. The MutantHuntWGS download includes two ploidy files, one for diploids and one for haploids. The user can specify either ploidy file when running the pipeline. MutantHuntWGS will automatically provide this file to BCFtools during the variant-calling step.

Users may also set variant-quality-score cutoffs (described in detail in the Methods) to tune the stringency of the analysis. They can also toggle the alignment step to save time when resetting the stringency. This option re-subsets variant calls with a higher or lower stringency cutoff, skipping the more time-consuming upstream steps of the pipeline. Although MutantHuntWGS does not allow users to specify additional cutoffs that filter the output per SnpEff/SIFT effect predictions and scores, users can separately apply such filters to the MutantHuntWGS output files after the fact—thus allowing for increased stringency. Increasing stringency is particularly important when analyzing experimental and control strains from particularly heterogeneous strain backgrounds.

Appendix A.3.6 Assessing MutantHuntWGS performance on a bulk segregant dataset

To assess MutantHuntWGS performance, we applied it to bulk segregant analysis data (Birkeland *et al.* 2010) with ploidy set to haploid. MutantHuntWGS identified 188 variants not present in the control strain that passed the variant-quality-score cutoff of 100. Thus only 1.95% of all variants detected in the experimental strain passed the filtering steps (Table 24). Among these was the same *PHO81* (*VAC6*) mutation found in the Birkeland *et al.* (2010) study, which results in an R701S amino acid substitution in the Pho81 protein (Birkeland *et al.* 2010). Our pipeline thus identified the same published causal variant described in the original study.

We were surprised by how many sequence variants (relative to the reference genome) remained, even after we removed variants present in the control dataset from the experimental dataset and filtered by a stringent variant-quality score. Given our variant-quality-score cutoff of 100, it is unlikely that these variants were called in error; instead, they likely reflect high sequence heterogeneity in the genetic backgrounds of the experimental and control strains. Regardless, only 1.95% of the variants found in the experimental strain passed all filtering steps, and, of those, only 152 (1.58% of the total) were coding variants present in the SIFT output. MutantHuntWGS thus removed 98.48% of all variants, yielding a more manageable dataset.

To further reduce the length of this list, we experimented with additional cutoffs, including (1) more stringent variant-quality-score, (2) SIFT score, and (3) SnpEff impact score cutoffs. A SIFT-score cutoff of <0.05 (deleterious) reduced the number of variants in the final output from 152 to 6 while retaining the causal variant (Table 24). An increased variant-quality-score stringency (> 130) reduced the number of variants to 21 while retaining the causal variant. A SnpEff impact-score cutoff of > Moderate reduced the number of variants to 55, again retaining the causal variant. Finally, a variant quality-score cutoff of > 130 and a SnpEff score of > Moderate, used together, reduced the number of variants to only 6 and again retained the causal variant. These post-hoc tests demonstrate how users might similarly narrow their lists of potential candidates.

Appendix A.3.7 Assessing MutantHuntWGS performance using lab evolution datasets

To test MutantHuntWGS performance on strains that did not undergo bulk-segregant analysis, we analyzed nine datasets from lab evolution experiments (Goldgof *et al.* 2016; Ottilie *et al.* 2017), again setting ploidy to haploid and using a variant-quality-score cutoff of 100. In each

of these studies, yeast cells were allowed to evolve resistance to a drug. WGS then identified mutations that altered the amino acid sequence of putative drug-binding proteins in or near well-defined binding pockets (Goldgof *et al.* 2016; Ottilie *et al.* 2017).

In this analysis, a small number of mutations were detected after comparing to control and filtering by variant-quality score (>100). Shortlists of only 4 to 11 (1.33% to 3.00%) of the variants originally detected in the experimental strain(s) were obtained for each dataset. Out of these variants, only 2 to 5 (0.66% - 1.37% of called variants in the experimental strain) were present in the SIFT output for each dataset. In each case, the list of variants generated by MutantHuntWGS included the mutation identified in the published study (SRR3480237: Pma1 N291K, Yrr1 T623K, SRR3480212: Pma1 P339T, Yrr1 L611F, SRR3480251: Pma1 L290S, Yrr1 T623K, SRR3480267: Pma1 G294S, SRR3490304: Erg11 V154G, SRR3490397: Erg11 T318N, SRR3490399: Erg25 D234E, SRR3490425: Erg25 H156N). These test cases confirm that MutantHuntWGS can identify yeast-sequence variants from WGS sequencing samples and can accurately filter out background mutations.

Table 24. Analysis of previously published bulk-segregant and lab-evolution WGS datasets using MutantHuntWGS demonstrates the effectiveness of the pipeline.

Birkeland *et al.* (2010)

SRA ID	SRR064545	SRR064546			
Total Reads (% Mapped)	19,782,779 (92.80%)	20,015,390 (89.57%)	**Additional Output Filtering**		
Variants Called in Control	10,022		**Filtering by**	**Cutoff**	**Variant count (%)**
Variants Called in Experimental		9,646	Variant quality score	>130	21 (0.21%)
Unique Variants in Experimental (% of total)		188 (1.95%)	SnpEff Impact	>Moderate	55 (0.57%)
CDS Variants in SIFT output (% of total)		152 (1.58%)	Variant Quality Score + Impact	>130 + >Moderate	6 (0.06%)
Published mutation (s) Identified		Yes	SIFT	Deleterious	6 (0.06%)

Ottilie *et al.* (2016)

SRA ID	SRR3480136	SRR3490425	SRR3490399	SRR3490397	SRR3490304
Total Reads (Percent Mapped)	14,684,843 (93.61%)	7,935,729 (98.11%)	3,629,049 (97.53%)	5,611,439 (97.80%)	6,904,333 (97.91%)
Variants Called in Control	526				
Variants Called in Experimental		367	298	336	377
Unique Variants in Experimental (% of total)		11 (3.00%)	4 (1.34%)	7 (2.08%)	8 (2.12%)
CDS Variants in SIFT output (% of total)		4 (1.09%)	2 (0.67%)	4 (1.19%)	5 (1.33%)
Published Mutation(s) Identified		Yes	Yes	Yes	Yes

Goldgof *et al.* (2016)

SRA ID	SRR3480136	SRR3480251	SRR3480237	SRR3480212	SRR3480267
Total Reads (Percent Mapped)	14,684,843 (93.61%)	6,347,816 (94.49%)	7,480,005 (64.04%)	3,058,951 (98.11%)	2,545,805 (97.51%)
Variants Called in Control	526				
Variants Called in Experimental		487	316	292	301
Unique Variants in Experimental (% of total)		10 (2.05%)	8 (2.53%)	6 (2.05%)	4 (1.33%)
CDS Variants in SIFT output (% of total)		4 (0.82%)	3 (0.95%)	4 (1.37%)	2 (0.66%)
Published Mutation(s) Identified		Yes	Yes	Yes	Yes

Appendix A.3.8 Existing WGS analysis pipelines

Other platforms exist that perform similar analyses. Each possesses a subset of the features enabled by MutantHuntWGS and has notable strengths. However, MutantHuntWGS is unique in its ability to combine the best attributes of these published tools while including additional functionality and providing output data in standard formats, such as BAM and VCF.

One user-friendly program, Mudi (Iida *et al.* 2014), uses BWA (Jo and Koh 2015), SAMtools (Li *et al.* 2009), and ANNOVAR (Wang *et al.* 2010) for sequence alignment, identification, and annotation of sequence variants, respectively. Like MutantHuntWGS, it performs numerous filtering steps before returning a list of putative causal variants. However, MutantHuntWGS predicts variant effects and maps variants to annotated *S. cerevisiae* genes using SnpEff and SIFT instead of ANNOVAR and allows access to all intermediate data files.

Another program, VAMP, consists of a series of Perl scripts that build and query an SQL database made from user-provided short-read sequencing data. VAMP identifies sequence variants, including large insertions and deletions. It also has built-in functionality that allows for manual inspection of the data (Birkeland *et al.* 2010). However, we feel MutantHuntWGS is more user-friendly than VAMP, which does not adhere to common data formats.

A recent article describing WGS in yeast samples includes a bioinformatics pipeline, referred to as wgs-pipeline (Gopalakrishnan and Winston 2019). It is built in a Snakemake framework (Köster and Rahmann 2012) that runs in a Conda environment (https://docs.conda.io/en/latest/), similar to the container-based analysis environment we used for MutantHuntWGS. This pipeline uses Bowtie2 (Langmead and Salzberg 2012), SAMtools (Li *et al.* 2009), Picard (Toolkit 2016), and GATK (McKenna *et al.* 2010) to align, process, and compare

datasets. This pipeline performs variant calling and comparison steps similar to MutantHuntWGS and maps variants to affected genes. However, MutantHuntWGS, which runs both SnpEff and SIFT on the candidate variants, provides a more comprehensive analysis of the predicted effects of the variants.

The Galaxy platform (Giardine *et al.* 2005; Blankenberg *et al.* 2010) provides a user-friendly, online interface for building bioinformatics pipelines. Galaxy also offers access to intermediate files. However, analysis with this platform requires the user to select the tools and parameters to incorporate, so some knowledge of the tools themselves is essential. Implementation is straightforward after those decisions are made, and the user need not have any understanding of Unix/Linux. The advantage of MutantHuntWGS over the Galaxy platform and pipelines such as CloudMap (Minevich *et al.* 2012) is that the user does not need to make decisions about the data analysis workflow.

In summary, the MutantHuntWGS pipeline is among the most user-friendly of these programs. It combines the most useful features of the existing WGS analysis programs while also enabling the user to account for ploidy. The use of containerization streamlines the installation of MutantHuntWGS and enhances its reproducibility. Thus, MutantHuntWGS offers ease of use, functionality, and data-transparency setting it apart from existing WGS pipelines.

Appendix A.3.9 Conclusions

Processing data generated from next-generation sequencing platforms requires significant expertise, and so is inaccessible to many investigators. We have developed a highly effective differential variant-calling pipeline capable of identifying causal variants from WGS data. Our

pipeline allows the user to access all relevant intermediate and output data files. We demonstrate the utility of MutantHuntWGS by analyzing previously published datasets. In all cases, our pipeline successfully identified the causal variant. We offer this highly reproducible and easy-to-implement bioinformatics pipeline to the *Saccharomyces cerevisiae* research community. A copy can be downloaded free of charge from https://github.com/mae92/MutantHuntWGS.

Appendix B

Appendix B.1 Investigating the role of iron responsive transcription factors in *paf1*Δ

This appendix contains analysis and experimental data that was not published in the *Genetics* paper or shown in chapter 2. These experiments and analyses were aimed at determining the role of sequence specific transcription factors (TFs), important for genes involved in iron homeostasis, in the transcriptional defects observed in a *paf1*Δ. The results in Figures 37, 38, and 38 are negative data and the experiments in Figures 38 and 39 have technical issues. I chose to present the data in Figures 38 and 39 to discuss the technical issues we had during the process and the experimental premise, which I thought was an interesting hypothesis. Figure 40 contains an experiment using strains I constructed containing deletions of iron TFs *AFT1* and *AFT2*, but some results are a bit confusing.

Indirect immunofluorescence experiments presented here were performed by Chelsea Guan, an undergraduate in the lab. I performed the other experiments presented. Additional experiments looking at the iron regulon were performed by another undergraduate I mentored, Alex Lederer, and he presented the results in his Honors thesis. Additionally, before Alex, Chelsea, and I began working on the iron regulon, Dr. Brett Tomson (as a post-doc in the lab) and Dr. Celeste Shelton (as an undergraduate student in the lab) had worked together to produce the HA tagged Aft1 strains used in the indirect immunofluorescence experiments. Brett and Celeste also performed spot tests and many northern blots to assess RNA levels and iron-starvation related phenotypes in *paf1*Δ strains.

Appendix B.1.1 Metal-responsive transcription factors and *paf1*Δ transcriptome defects

Transcription factors involved in metal homeostasis in yeast are reviewed along with their target genes in (Rutherford and Bird 2004). I hypothesized that metal-responsive TFs might be playing a role in the upregulation of the iron regulon in cells deleted for *PAF1*. I performed a descriptive analysis of tiling array data (using log2(*paf1*Δ/WT) values) to determine if all known targets of Aft1 and Aft2 (transcription factors for genes involved in iron homeostasis) were upregulated in a *paf1*Δ strain. Log2 fold change values for genes activated by each TF as listed in (Rutherford and Bird 2004) were plotted as histograms to see how their values were distributed (Figure 37). All appear to have a normal distribution which is clear in all cases except for Mac1 even with a low number of samples. There is a slight right-hand skew in the Zap1, Aft2 and Aft1 histograms suggesting that the targets of these TFs tend to be upregulated in the *paf1*Δ cells. However, many genes in each group show decreased or unchanged expression, which is not consistent with the hypothesis that all targets of these TFs are upregulated in a *paf1*Δ strain.

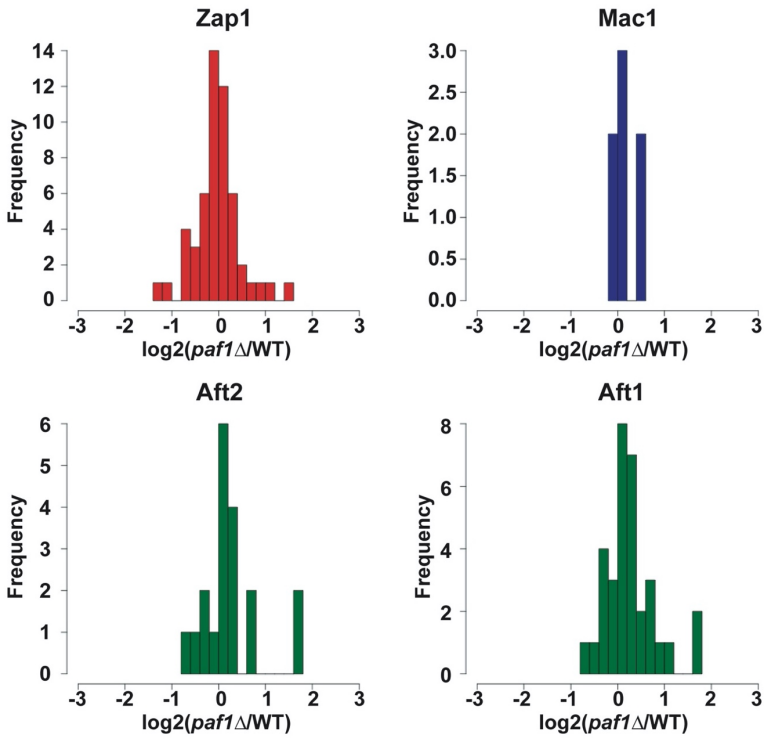


Figure 37. Histograms of log2(*paf1*Δ/WT) for metal homeostasis transcription factor target genes.

Appendix B.1.2 Assessing nuclear localization of Aft1 in a *paf1*Δ strain

I mentored an undergraduate student, Chelsea Guan, as she conducted indirect immunofluorescence (IDF) experiments and her work is presented in Figures 38 and 39. Because Aft1 is only able to activate genes from within the nucleus, I hypothesized that the reason we see increased expression of iron homeostasis genes in *paf1*Δ is because Aft1 is constitutively nuclear in these cells. I had conducted IDF experiments to this end in the past without much success, for technical reasons, but Chelsea was eager to attempt the new technique, so we analyzed *paf1*Δ and WT strains with and without HA tagged Aft1 in an attempt to troubleshoot the assay. We obtained protocols and advice from Chris Gurrero in the Brodsky lab who had performed this technique

successfully in the past, but we still only achieved limited success and as a result were not able to draw any conclusions from our data.

I thought my past technical issues might have been due to the antibody not being able to enter the cells properly so we conducted a zymolyase incubation time titration in WT and *paf1*Δ cells grown in media containing the iron chelator BPS (starves the cell for iron) in hopes of finding a concentration that would work well for our experiments (Figure 38). The reason I chose to do the experiment in this way is so that we could try to identify nuclear localization of Aft1 (which occurs in iron starved conditions) in the WT strain as a way to identify our Aft1 signal more easily. Aft1 is present in cells at a low concentration and it is hard to tell if we are observing autofluorescence or true signal. We do not see proper signal for Kar2, which is a positive control target that associates with the endoplasmic reticulum, in any of the conditions. We do not observe HA signal in the nucleus in WT cells and the incubation time with zymolyase does not appear to make a difference. This suggested that the zymolyase step was not the problem.

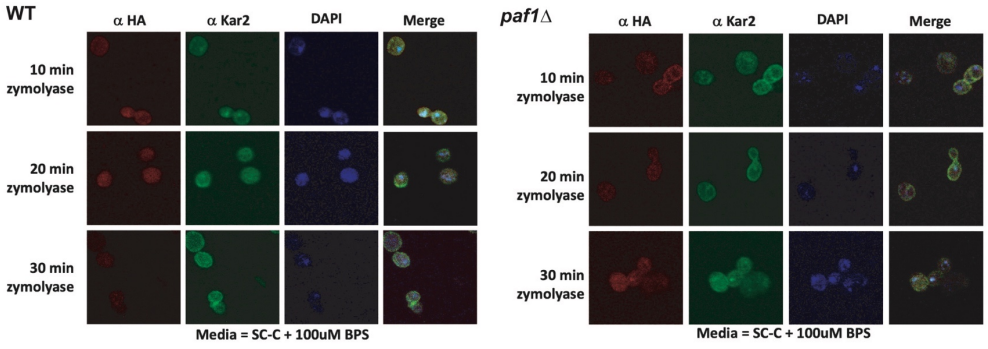


Figure 38. Indirect immunofluorescence data for WT and *paf1*Δ strains.

Indirect immunoflurescence was performed targeting the HA epitope (fused to Aft1) and the endoplasmic reticulam bound protein Kar2 was used as a positive control. DAPI shows staining of the nuclear DNA.

Chelsea and I decided to use the 30-minute zymolyase treatment and work through the protocol one more time with HA tagged Rtf1, which was successfully used in this assay in the

past, as a positive control and an untagged strain as a negative control. This time we did not treat the cells with BPS to induce the iron starvation response. These data are presented in Figure 39 where it is clear that we are not properly detecting Rtf1 but may be detecting Aft1 in the cytoplasm. We also have clear signal here for Kar2 in all, except the HA-tagged Rtf1 sample and its negative control. More optimization experiments such as these can be done. However, I chose to end these experiments here, because the bioinformatic analysis presented in Figure 37 indicated that Aft1 was not as important as I had originally thought in the context of a *paf1*Δ strain.

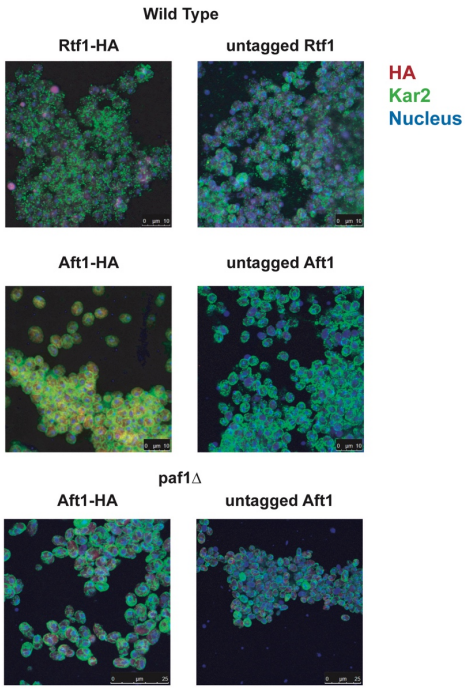


Figure 39. Indirect immunofluorescence results for WT and *paf1*Δ cells with HA tagged Rtf1 or Aft1. Combined data from HA detection, Kar2 detection and DAPI nuclear stain are presented.

Appendix B.1.3 Analysis of *AFT1* and *AFT2* mutant yeast strains

Figure 40 contains RT-qPCR data from an experiment aimed at determining the roles of Aft1 and Aft2 (iron gene activating TFs) in the changes in gene expression observed in the iron

regulon in a *paf1*Δ strain. I was surprised to find that in *aft1*Δ and *aft2*Δ strains an increase in expression was observed at *FIT2*, *FIT3*, and *ARN1*. The *aft1*Δ result at *SIT1* is what was expected for all of these genes. These data suggest that Aft1 and Aft2 may be acting as repressors of these iron genes in iron replete conditions. Because I do not understand how this could be occurring and because these data do not align with published data in other strain backgrounds, I chose to leave them out of the *Genetics* paper. I think that the idea of Aft2 acting as a repressor could potentially be interesting considering it is similar to Aft1, but less well understood, or the two TFs could possibly cross-regulate each other.

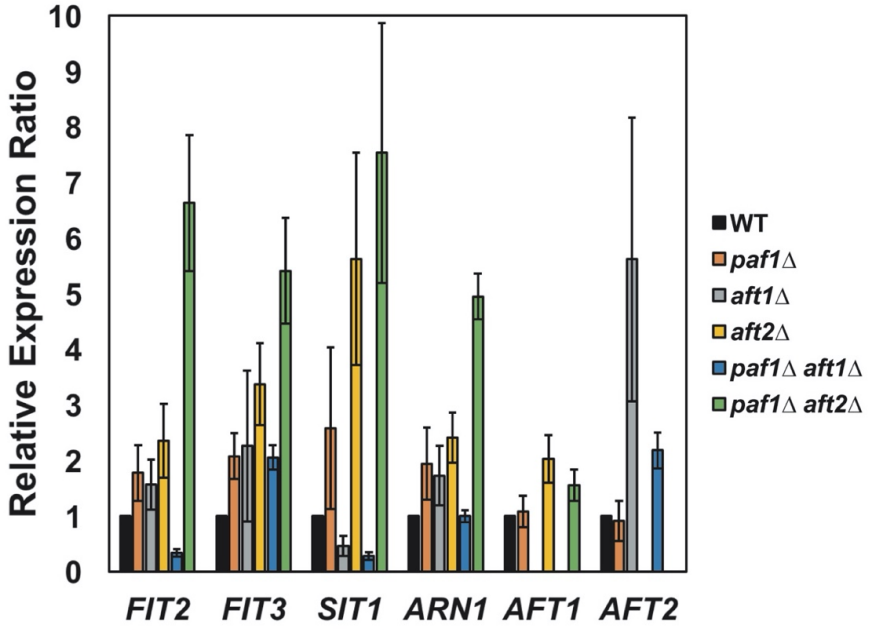


Figure 40. RT-qPCR data for *AFT1*, *AFT2*, and *PAF1* deletion mutants.

When the deletions of the iron-responsive TFs are combined with a *PAF1* deletion (*aft1*Δ *paf1*Δ and *aft2*Δ *paf1*Δ) we can see that loss of *AFT1* together with *PAF1* (*aft1*Δ *paf1*Δ) suppresses the effect of a *paf1*Δ mutation at *FIT1*, *SIT1*, and *ARN1*, but not *FIT3* whereas deletion of *AFT2* with *PAF1* (*aft2*Δ *paf1*Δ) has an additive phenotype. The double mutant data suggest that Aft1 acts

as a positive regulator and Paf1 as a negative regulator. However, in the *aft1*Δ strain many of these transcripts are de-repressed which goes against what we know about Aft1 function. The double mutant data suggest that Aft2 and Paf1 act in parallel pathways to negatively regulate the iron genes assayed here. This makes sense given that TFs are important at the initiation phase of transcription and Paf1 is important at the elongation phase.

Appendix B.2 Materials and methods

Appendix B.2.1 Bioinformatic analysis

Bioinformatic analysis of tiling array data was performed as described in 2.2.5-2.2.9 and log2(*paf1*Δ/WT) was calculated from those data. Lists of genes regulated by Mac1, Zap1, Aft1, and Aft2 were assembled from data presented in Table 1 of (Rutherford and Bird 2004) and were used to subset the gene expression data prior to plotting histograms. Histograms were generated using R Studio.

Appendix B.2.2 Indirect immunofluorescence

Cells were grown to mid log phase (OD_{600} = 0.4-0.8) in YPD or YPD + 100 µM BPS before harvesting 5 mL of cells and washing once in KM solution (40 mM KPO_4 pH 6.5 and 500 µM $MgCl_2$) and again in KMS (1.2 M Soribtol, 16 mM KPO_4 pH 6.5, and 200 µM $MgCl_2$). Cells were treated with 30 µL of 10 mg/mL zymolyase per 0.5 mL of cell suspension and incubated at 37 ˚C for the indicated amount of time before washing and resuspending cells in 200 µL of KMS. Cells

were loaded onto prepared microscope slides by adding 20 µL of cells per slide-well and incubating at room temperature for 10 minutes. Excess KMS was aspirated off and the slide it was submerged into ice cold methanol and incubated at -20 ˚C for 6 minutes. Upon removal from the methanol solution the slide was submerged in ice cold acetone for 30 seconds before removal and evaporation of acetone using a slanted heat block. Wells were blocked in 30 µL of blocking solution (1X PBS pH 7.4, 0.5% BSA, 0.5% Ovalbunum, a dab of fish gelatin, 0.1% Triton X 100) for one hour before aspirating off blocking solution and adding primary antibodies diluted 1:250 in blocking solution and incubated overnight at 37 ˚C overnight in a humidified chamber constructed using a pipet tip box and Watman's paper. Primary antibody was aspirated off and wells were washed with 30 µL of blocking solution three times for five minutes each before applying 30 µL of secondary antibodies diluted 1:500 with blocking solution and incubated at 37 ˚C for 2 hours before aspirating off the secondary antibody solution. Slides were prepared for visualization by addition of a few drops of aqueous mounting medium containing DAPI, addition of a coverslip and application of pressure in the dark overnight (accomplished by placing a heavy object such as a textbook on slides wrapped in paper towels). Slides were visualized on the departmental confocal microscope with the help of Tom Harper. Strains used in these experiments: KY2322, KY2320, KY2318, KY2316, KY2808.

Appendix B.2.3 RT-qPCR

Cell culture, RNA isolation, reverse transcription and qPCR were conducted as described in 2.2.1, 2.2.2, and 2.2.4. All primers used in qPCR are shown in Table 13. Strains used in this experiment: KY3260, KY3258, KY3259, KY3262, KY3257, KY3263.

Appendix C

Appendix C.1 Additional Cdc73 biochemistry experiments

I performed numerous biochemistry experiments trying to identify binding partners for the Cdc73 C-domain and investigate the stability of full-length Cdc73 in various contexts. Most of these experiments failed or were set aside in favor of more fruitful avenues of investigation. Some of the work presented here was conducted as a rotation project in the VanDemark lab and other work was conducted during my years in the Arndt lab. Much of the data presented is negative, but I believe having these results curated here will offer an advantage to any new student that might want to continue this work.

Appendix C.1.1 Validating the thermal shift assay using Cdc73 C-domain, MBP, and GST

When I began my rotation in the VanDemark lab, working with Dr. Joel Rosenbaum, I was tasked with re-developing the labs thermal shift assay and using it to study interactions between the C-domain of Cdc73 and Pol II CTD peptides. I was successfully able to re-develop the assay and now it is regularly used by many people in the department.

Initially I tested the Cdc73 C-domain along with positive controls MBP and GST and each with their preferred ligand to see if the assay could detect the increased stabilization of the control proteins upon ligand binding and to get a baseline melting temperature (Tm) for the Cdc73 C-domain (Figure 41). The C-domain had a melting temperature (Tm) of ~61 ˚C in this initial test (Figure 41A and 41B), but may appear more stable (Tm ~63 ˚C) in assays presented later due to

better buffer conditions. All control proteins behaved as expected with increased stability and thus a higher Tm in their ligand bound states (Figure 41C).

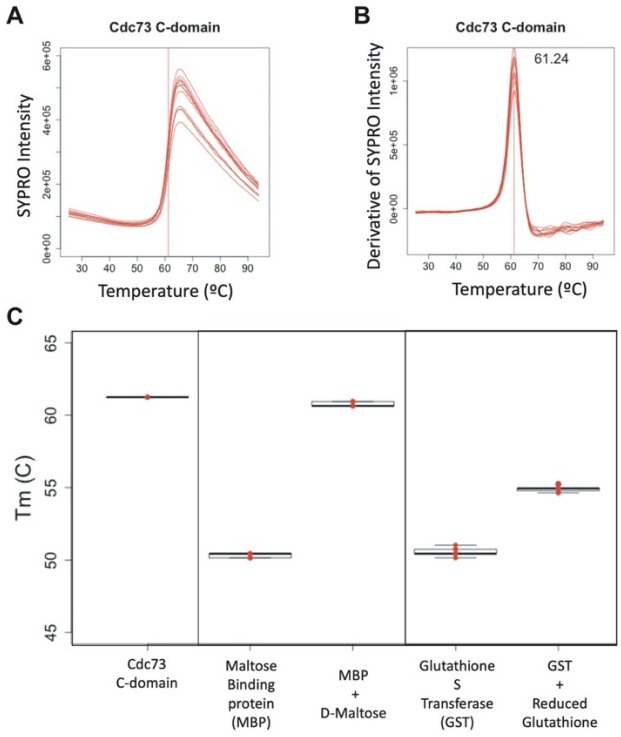


Figure 41. Fluorescence thermal shift assay results for Cdc73 C-domain and MBP and GST control proteins. A) Fluorescent intensity and B) derivative of fluorescence intensity of SYPRO orange dye upon denaturation of Cdc73 C-domain. C) Box and whisker plots of fluorescent thermal shift data for Tm of three replicate reactions for Cdc73 C-domain and control proteins.

Appendix C.1.2 Testing Cdc73 C-domains ability to interact with S2P-S5P CTD peptides

Cdc73 C-domain and residues upstream have been shown to interact with S2P-S5P CTD peptides (Qiu *et al.* 2012). However, CTD peptides phosphorylated at S2 and S5 positions can be found in two different configurations (Figure 42A) each presenting a different ligand to the C-domain. The S5P-S2P version probably is what is naturally found in the cell because most of the time there is only one phosphorylation mark per repeat despite that there are multiple phosphor-

competent residues in each repeat (Suh *et al.* 2016). Therefore, I hypothesized that the Cdc73 C-domain bound to CTD peptides in the S5P-S2P configuration. First, I tested full length Cdc73 and did not observe a stabilization indicating that the interaction was not occurring (Figure 42B). I went on to test all of the Cdc73 C-domain containing constructs in the VanDemark lab and only in the Cdc73 190-393 construct did we observe increases in Tm and in that case the unmodified CTD peptide produced the strongest Tm shift. These results do not agree with the data published in (Qiu *et al.* 2012).

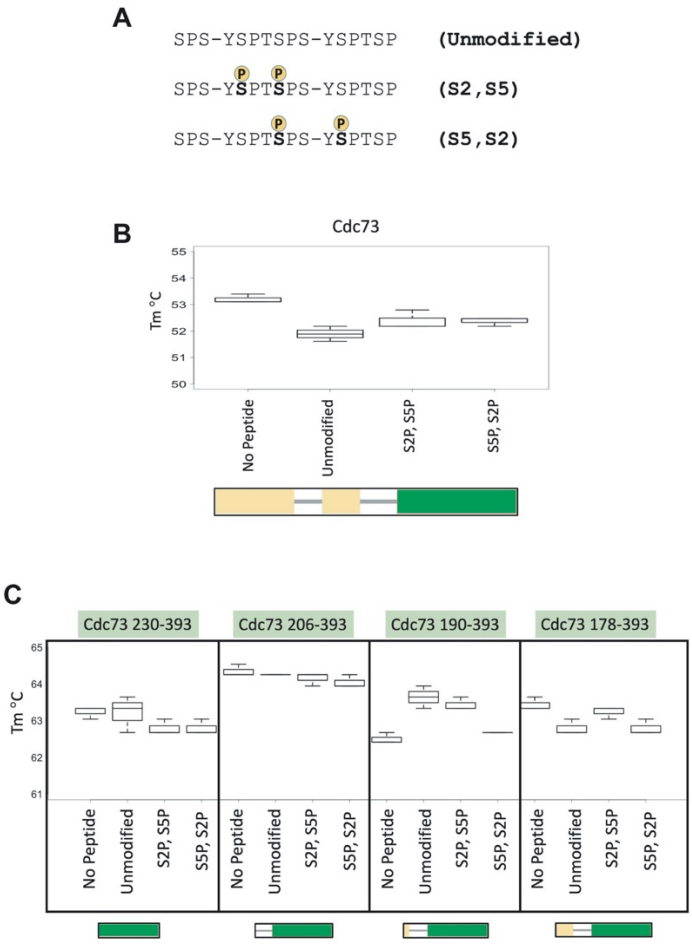


Figure 42. Fluorescent thermal shift assay for Cdc73 constructs and CTD peptides.

A) Descriptions of peptides tested. B) Fluorescent thermal shift assay results for full length Cdc73 and C) constructs truncated from the N-terminus, but containing the C-domain.

In an attempt to replicate the data published in (Qiu *et al.* 2012) Dr. Joel Rosenbaum and I performed two more binding assays. The first was a biotin pulldown assay (Figure 43A) where we biotinylated the peptides and bound them to streptavidin beads before incubating the beads with full length Cdc73. These results again indicate no binding over background (compare no-peptide bound lane to all other bound lanes).

The next assay was a native PAGE gel-shift assay using fluorescently labeled peptides and full length Cdc73. In this experiment we observed only weak interactions between Cdc73 and each peptide at the highest concentration of peptide (385 μM), but no difference was observed between CTD peptides.

Now we know that the Cdc73-Spt6 interaction has a ~1 μM Kd so perhaps the titration of peptide in this experiment was not taken far enough to capture binding information, although the fact that there does not appear to be any difference between unmodified and phosphorylated peptides is still concerning given the published data. We know from the data presented in chapter 3 that Cdc73 C-domain interacts directly with the CTD. It is hard to say why we do not obtain a positive shift in Tm in the thermal shift assay or better binding in each of these binding assays. Furthermore, it is unclear why we are not able to recapitulate the results published in (Qiu *et al.* 2012). Perhaps with further optimization and/or correspondence with the authors the published results could be reproduced. However, the peptides that Joel and I were using were very expensive and in short supply, so we were not able to do very much optimization.

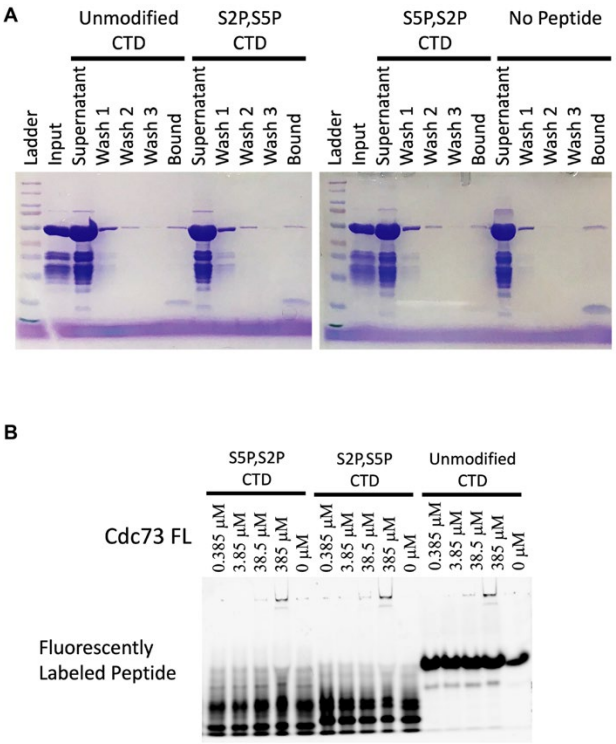


Figure 43. Additional binding assays to test for Cdc73 binding to CTD peptides.

A) Biotin pulldown results using peptides as bait and full length Cdc73 as prey. B) Native PAGE gel-shift assay assessing binding between CTD peptides and full length Cdc73.

Appendix C.1.3 Screen to identify compounds that bind to full length Cdc73

The full structure of Cdc73 is of interest to the field, thus we sought to identify stabilizing agents that might allow for full length Cdc73 to be crystalized. I used freshly purified full length Cdc73 to conduct a chemical library screen using the silver-bullets-biochemical panel (Hampton Research, Aliso Viejo, CA). Results of the screen are shown in Figure 44A and 44B. Outliers with extremely high or low Tm values present in Figure 44A are due to technical issues often resulting from the protein being denatured upon addition of the compound or a compound that is able to fluoresce at the same wavelength. When we zoom in on the y-axis to look at only a few degrees on either side of full length Cdc73s Tm (~53 ˚C) we can see that well G2 contains a reagent that

stabilizes the protein increasing the Tm by ~1 ˚C. This compound is termed Tacsimate and contains a number of small molecules. All of which stabilize Cdc73 (Figure 44C). Any of these compounds might be good candidates to help stabilize Cdc73 during crystallization. However, the concentrations present in the silver bullet assay are quite high and in the crystallization trial we performed with malic acid at the end of my rotation we only obtained salt crystals. It may be worth trying this again with lower concentrations of the compounds listed or using tacsimate itself at low concentration.

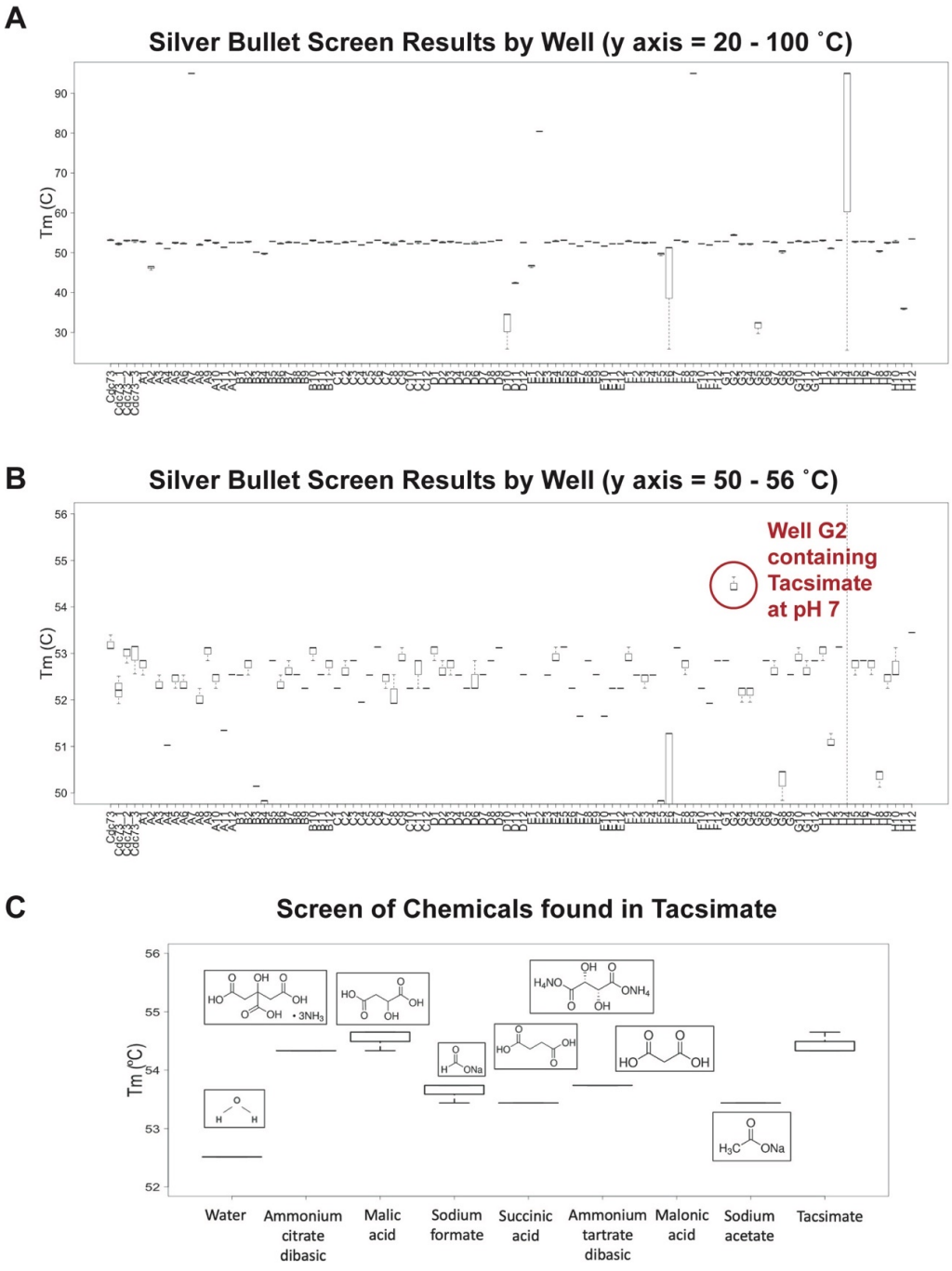


Figure 44. Screening for reagents that stabalize full length Cdc73.

A and B) Box and whisker plots of results from screening Cdc73 with the Silver Bullets Bio 96 well plate screen by thermal shift assay with expanded y-axis range in A and narrow y-axis range in B. C) Box and whisker plots of the followup screen of compounds found in the tacsimate reagent which offered the most stability to Cdc73 in the initial screen.

Appendix C.1.4 Small molecule binding assay

In collaboration with the VanDemark and Durrant labs a compound screen was conducted, first *in silico* by Dr. Jacob Durrant, and then *in vitro* by me. I ran into significant technical obstacles while trying to conduct this compound screen and want to discuss these here and describe the state of this project as of the last time I worked on it.

The *in silico* screen resulted in a list of 19 compounds, which after the order was placed decreased to 17 because ChemBridge was out of two of the compounds and it would have cost much more if we had asked them to re-synthesize these two compounds. Therefore compounds 4 and 15 are never analyzed but are shown here (red in Table 25) among the top hits in the *in silico* screen.

Table 25. Compounds identified in the *in silico* screen for small molecules that interact with the C-domain of Cdc73

Item	ChemBridge Cat#	Description	Smiles nomenclature
1	69584745	7-(4-cyclopentylpyrimidin-2-yl)-3-(pyridin-3-ylmethyl)-6,7,8,9-tetrahydro-5H-[1,2,4]triazolo[4,3-d][1,4]diazepine	N12CCN(CCC1=NN=C2CC1C=CC=NC=1)C1=NC=CC(C2CCCC2)=N1
2	16295036	1-(1-cyclopentyl-4-piperidinyl)-N-[(1-isopropyl-1H-pyrazol-4-yl)methyl]-N-(3-pyridinylmethyl)methanamine	N(C(C)C)1C=C(C=N1)CN(CC1CCN(CC1)C1CCCC1)CC1=CC=CN=C1
3	43035740	2-(3-{[4-(4-methoxyphenyl)-1H-1,2,3-triazol-1-yl]methyl}piperidin-1-yl)pyrimidine	N1N(CC2CCCN(C3N=CC=CN=3)C2)C=C(C2C=CC(=CC=2)OC)N=1
4	84585898	5-(3-fluorophenyl)-3-{4-[(2E)-3-phenyl-2-propen-1-yl]-1-piperazinyl}-1,2,4-triazine	C1(N=NC=C(C2C=CC=C(F)C=2)N=1)N1CCN(CC1)C/C=C/C1C=CC=CC=1

5	76476244	4-[2-(4-{[2-(2-pyridinyl)-1-piperidinyl]carbonyl}-1H-1,2,3-triazol-1-yl)ethyl]morpholine	C1(=CN(N=N1)CCN1CCOCC1)C(=O)N1CCCCC1C1=CC=CC=N1
6	48772328	N-(3-{4-[(2,2-dimethyltetrahydro-2H-pyran-4-yl)amino]-1-piperidinyl}phenyl)-3-(trifluoromethyl)benzamide	C(F)(F)(F)C1=CC=CC(=C1)C(=O)NC1=CC=CC(=C1)N1CCC(CC1)NC1CCOC(C)(C)C1
7	64057760	2-{4-[(4-cyclohexyl-1-piperazinyl)carbonyl]-5-methyl-1H-pyrazol-1-yl}-4-(2-fluorophenyl)pyrimidine	C1(C=NN(C2N=CC=C(C3C=CC=CC=3F)N=2)C=1C)C(=O)N1CCN(CC1)C1CCCCC1
8	7749708	2-chloro-N-[3-(6-methyl[1,2,4]triazolo[3,4-a]phthalazin-3-yl)phenyl]benzamide	N12N=C(C)C3=C(C=CC=C3)C1=NN=C2C1=CC=CC(=C1)NC(=O)C1C=CC=CC=1Cl
9	9016282	4-(1H-indol-3-yl)-6-methyl-4,6-dihydro-2H-pyrano[3,2-c]quinoline-2,5(3H)-dione	C12C(CC(OC=1C1C=CC=CC=1N(C)C2=O)=O)C1C2C=CC=CC=2NC=1
10	9279070	2-(4-benzyl-1-piperazinyl)-2'-(4-ethyl-1-piperazinyl)-4'-methyl-4,5'-bipyrimidin-6(1H)-one	C(N1CCN(CC1)CC1C=CC=CC=1)1NC(=O)C=C(C2=CN=C(N=C2C)N2CCN(CC2)CC)N=1
11	78565823	N-{2-[(5-fluoro-4-morpholin-4-ylpyrimidin-2-yl)amino]ethyl}acetamide	N1=C(NCCNC(C)=O)N=CC(F)=C1N1CCOCC1
12	98653539	N-{1-[(4-ethylphenyl)sulfonyl]-3-piperidinyl}-N,N',N'-trimethyl-1,3-propanediamine	S(=O)(=O)(C1C=CC(=CC=1)CC)N1CCCC(N(C)CCCN(C)C)C1
13	9248461	2-({2-[4-(4-methoxyphenyl)-1-piperazinyl]-4-quinazolinyl}amino)ethanol	C(N1CCN(CC1)C1C=CC(=CC=1)OC)1=NC2=CC=CC=C2C(NCCO)=N1
14	9075976	6-[(4-benzyl-1-piperazinyl)methyl]-N-(4-methoxyphenyl)-1,3,5-triazine-2,4-diamine	C(NC1=CC=C(C=C1)OC)1=NC(CN2CCN(CC2)CC2C=CC=CC=2)=NC(N)=N1
15	97728749	N-(3'-methoxy-3-biphenylyl)-1-[3-(1H-1,2,4-triazol-1-yl)propanoyl]-2-piperidinecarboxamide	N1(CCCCC1C(=O)NC1=CC=CC(C2C=CC=C(OC)C=2)=C1)C(=O)CCN1C=NC=N1
16	9120356	3-phenyl-5-(5-phenyl-1,3,4-oxadiazol-2-yl)-2,1-benzisoxazole	C12C(C=CC(=C1)C1=NN=C(C3C=CC=CC=3)O1)=NOC=2C1C=CC=CC=1

17	5923520	3-fluoro-N'-(6-hydroxy-7-methyl-3,4-dihydro-1(2H)-naphthalenylidene)benzohydrazide	C12/C(/CCCC=1C=C(C(C)=C2)O)=N/NC(=O)C1C=CC=C(F)C=1
18	7611694	2-(1,3-benzoxazol-2-ylamino)-4,6-dimethyl-N-(2-methylphenyl)-5-pyrimidinecarboxamide	C1(OC2=C(C=CC=C2)N=1)NC1=NC(C)=C(C(C)=N1)C(=O)NC1=CC=CC=C1C
19	69552614	N-{[7-fluoro-5-(3-quinolinyl)-2,3-dihydro-1-benzofuran-2-yl]methyl}-1-cyclopentene-1-carboxamide	C12=C(F)C=C(C=C1CC(CNC(=O)C1CCCC=1)O2)C1C=NC2=C(C=CC=C2)C=1

Undergraduate student Matthew Blacksmith and I conducted this screen. I spent a long time trying to figure out which solvent would be the best for each compound and finally decided to dissolve some in ethanol and some in DMSO after testing ethanol, isopropanol and DMSO for their effect on Cdc73 C-domain stability by fluorescent thermal shift assay (Figure 45A). In hindsight if I were to do this again, I would dissolve all compounds in DMSO at the exact same molar concentration from the start because most compounds are soluble or miscible in DMSO and it makes the work considerably easier and more directly comparable if the same solvent is used for all compounds. After testing multiple buffer components and protein concentrations Matt performed a test on all compounds in the two best conditions. In most cases results agreed, but there were some reproducibility issues. Here Matt added 1 mM of each compound and we observed negative Tm shifts for nearly all samples (Figure 45B-45D).

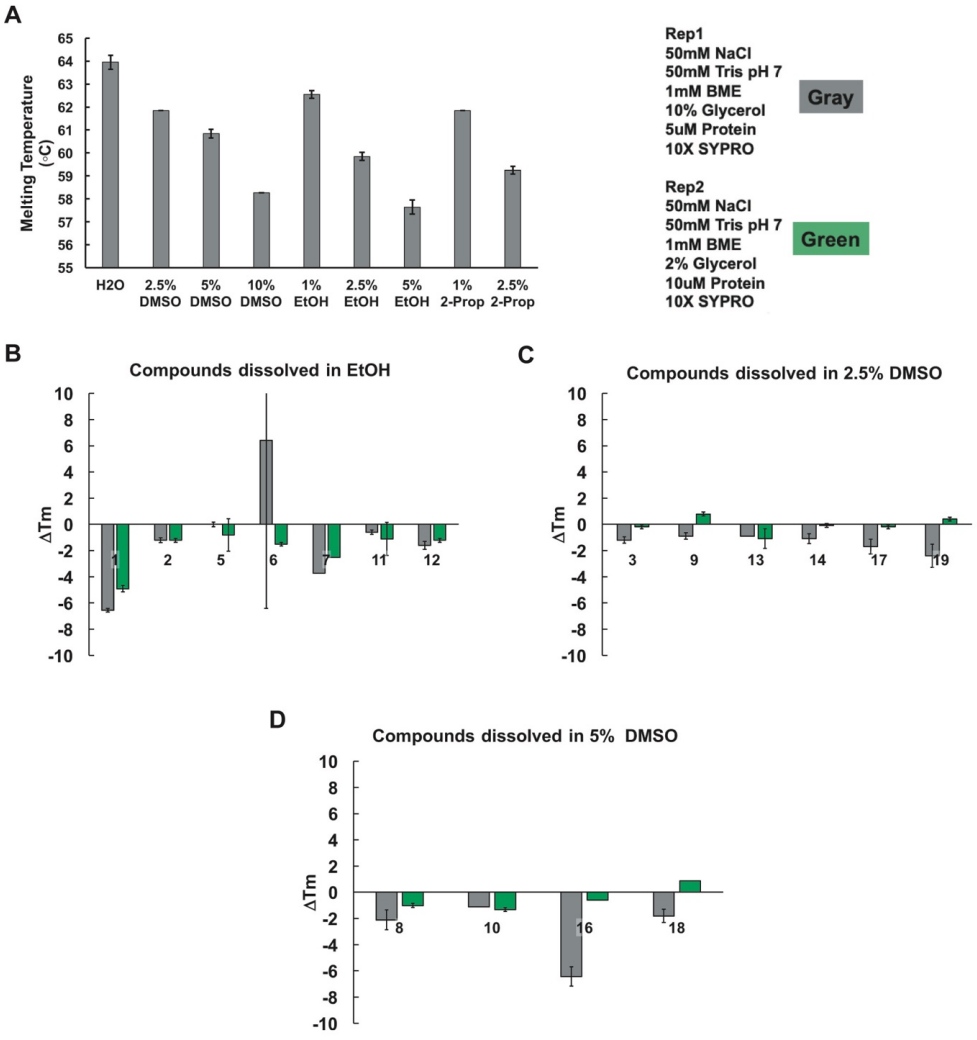


Figure 45. Fluorescent thermal shift assay results for 17 putative C-domain interacting compounds. A) Bar graphs of Tm assessing Cdc73 C-domain stability in the presence of various solvents. B-D) Bar graphs of change in Tm testing 1 mM of compound against 10 µM C-domain for stabalization of the C-domain.

Next a titration was performed, but only for the samples that were dissolved in ethanol because we wanted to test a subset to determine if we could identify positive temperature shifts with addition of less compound. We performed a 5-fold dilution series starting at 1 mM and ending at 8 µM. These results looked very promising (Figure 46A) but proved difficult to reproduce (Figure 46B).

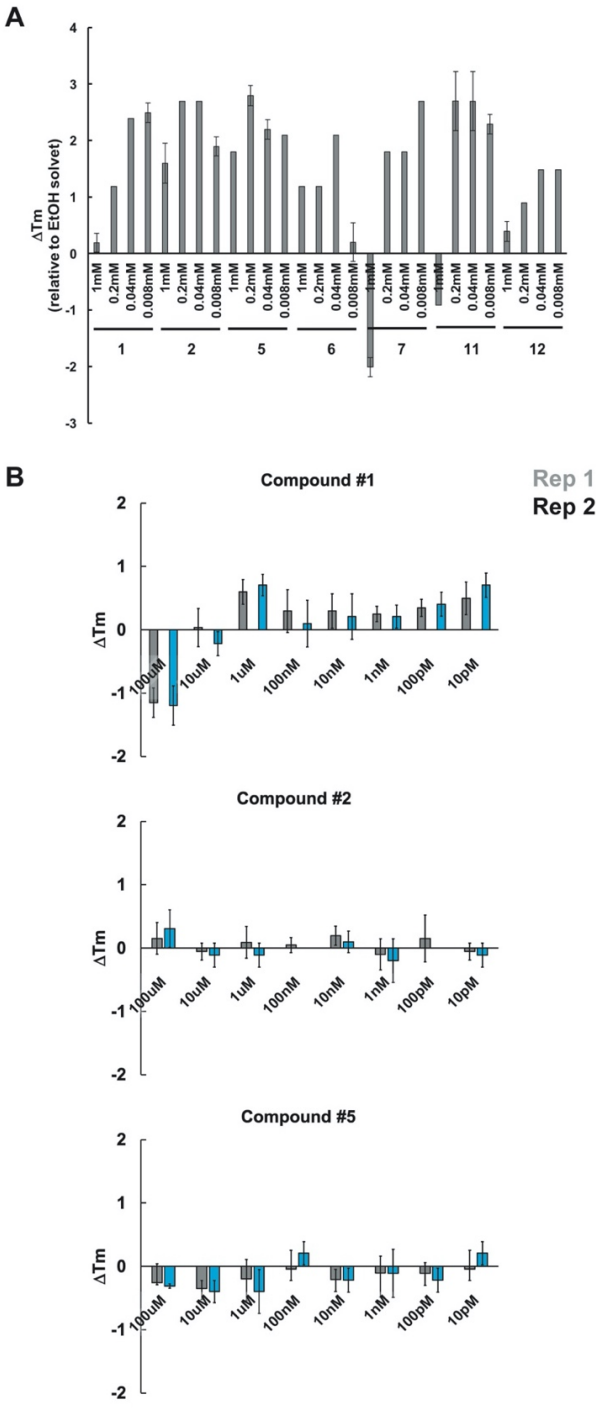


Figure 46. Fluorescent thermal shift assay results for titrations of a subset of putative Cdc73 interacting compounds.

A and B) Bar graphs of ΔTm of Cdc73 C-domain when treated with increasing concentrations of a subset of the compounds screened in Figure 45.

I identified the problem to be, at least partly, the use of various disposable tubes during the reaction preparation. I think that a combination of the solvent leaching chemicals out of the plastic and the molecule of interest sticking to the tubes is largely causing the reproducibility issues Matt and I observed. Figure 47A shows the results of testing the C-domain and compound #2 while using epp tubes during the reaction setup. Figure 47B shows the results of using PCR strip tubes instead. When these tubes were used, I was able to recapitulate the results from earlier experiments (Figure 46A) in which Matt had prepared his reactions using PCR strip tubes. I conducted two more replicates with compound #2 and found that in all conditions even down to picomolar amounts it caused a ~2˚C shift in Tm of the C-domain. This seems far to low of an amount required to stabilize all of the C-domain in the reaction and I believe it indicates that the solvent is contaminated with impurities from the plastic tubes.

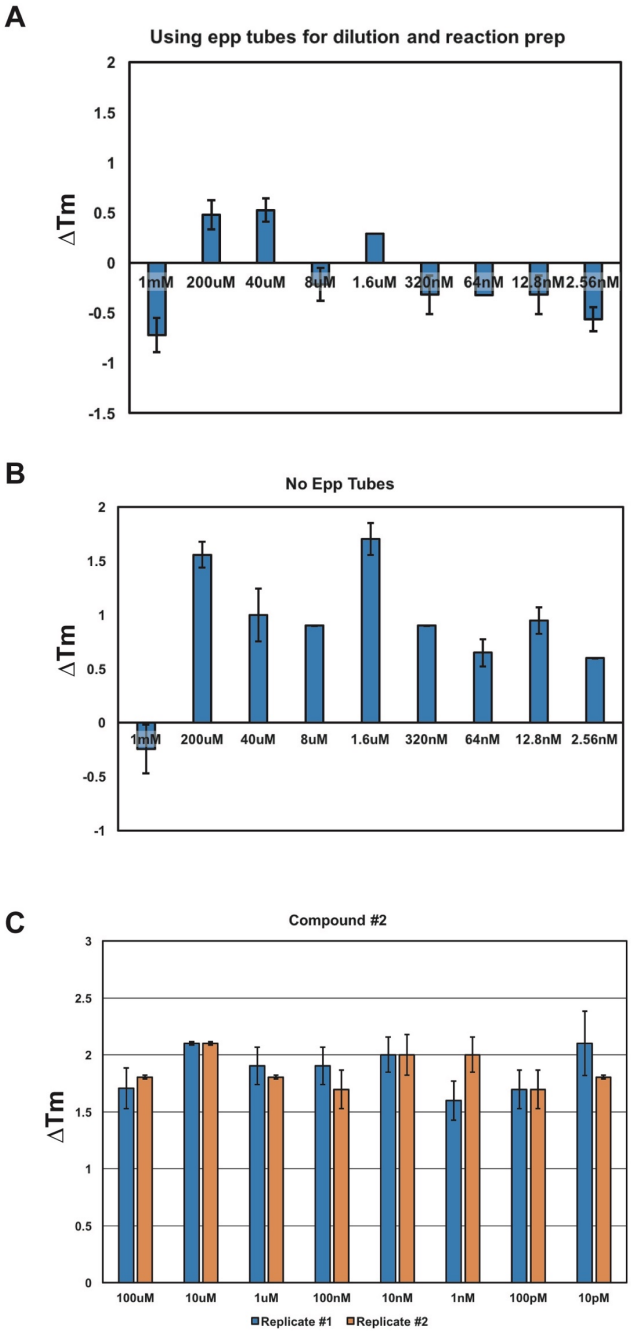


Figure 47. Troubleshooting the use of various consumables for compound screening using compound 2 as a test case.

At this point Dr. Jacob Durrant suggested that we use a small amount of detergent in an aqueous buffer to help keep these compounds from sticking to the inside of our tubes. I looked

into this and performed the experiment shown in Figure 48 to test various detergents. Based on these results TritonX 100 appears to be the best detergent to use in the future if these experiments are to be continued.

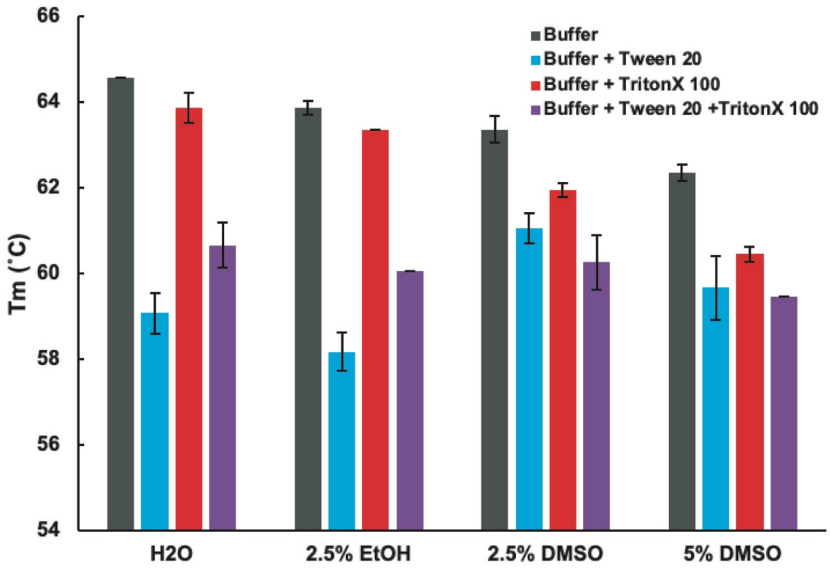


Figure 48. Cdc73 C-domain stability in the presence of various solvents and detergents.

Bar graph of fluorescent thermal shift assay results.

As another method to try to identify an interaction between the C-domain and the compounds from the *in silico* screen I spotted a small amount of each compound onto a piece of nitrocellulose incubated them with mRuby2-C-domain fusion protein (Figure 49A and 49B). Unfortunately, these data were uninterpretable due to the high autofluorescence of some compounds and the high level of background signal in spite of my attempts to block the membrane. No further attempts to optimize this assay were made because of the level of fluorescence generated by the compounds themselves, but if this assay were to be applied in another context I would suggest using a mClover fusion instead of mRuby2 because mRuby2 is more likely to

participate in nonspecific interactions. Also, mClover is a brighter fluorophore. I would also suggest washing for longer than 2 minutes.

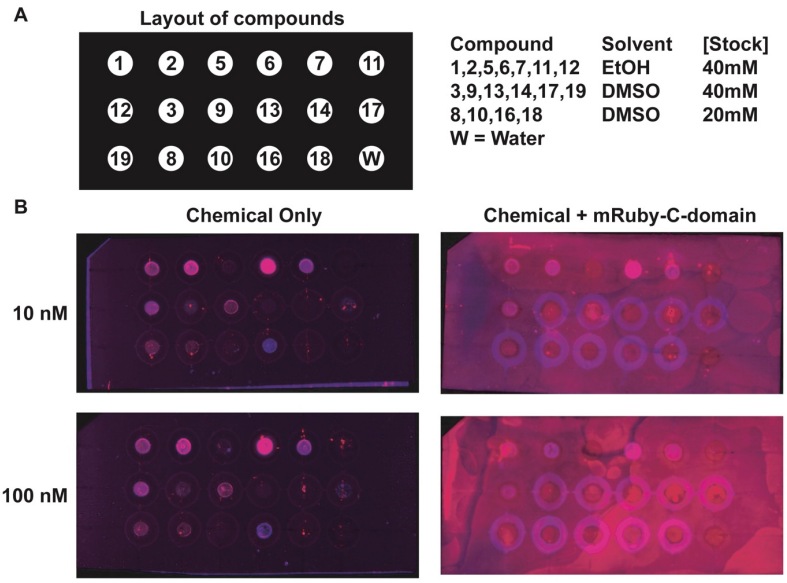


Figure 49. Failed dot blot assay testing 17 putiative C-domain interacing compounds.

Appendix C.1.5 W321A Limited proteolysis results

During my work in the Arndt lab investigating the Cdc73 W321A mutant and the Cdc73-Spt6 interaction I generated a model of the structure of full length Cdc73, which is discussed in chapter 3. When I looked more deeply at this model of the full-length structure it led me to hypothesize that Cdc73 W321A may be breaking up a hydrophobic face of Cdc73 that is coordinated by W321 (Figure 50). I thought it was possible given the inherent mobility of loops that the hydrophobic residues (yellow) relied on W321 (orange tryptophan) to coordinate the loop structures forming a binding interface at W321 and nearby R300.

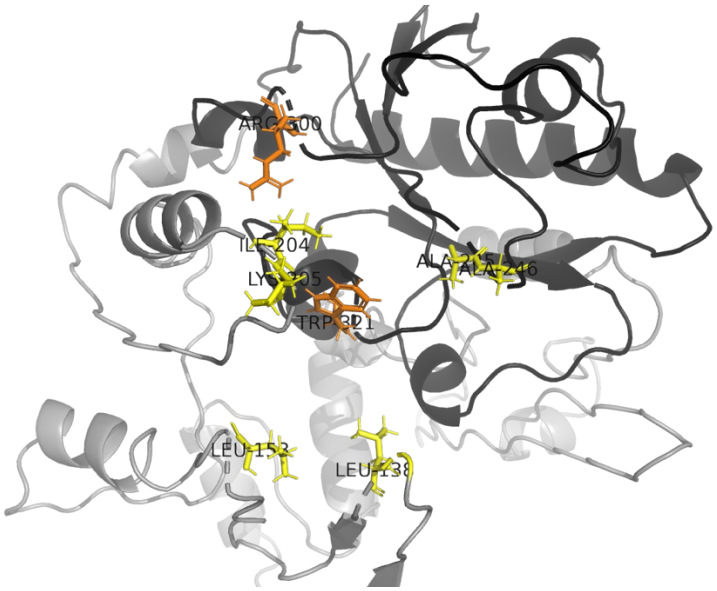


Figure 50. Structural model of full length Cdc73 highlighting a possible role for W321A in cordinating multiple hydrophobic residues.

Residues R300 and W321 are colored orange and hydrophobic residues with the potential to interact with W321, assuming loops are somewhat mobile, are colored in yellow.

To test this hypothesis, I used limited proteolysis which is a classic biochemistry technique for assessing protein stability and identifying unprotected and protected regions of a protein by treating it with a protease. In this case I used trypsin as the protease which cuts at lysine residues. Unfortunately, I could not detect a difference between Cdc73 and Cdc73 W321A by this assay, which suggests that my hypothesis is incorrect (Figure 51).

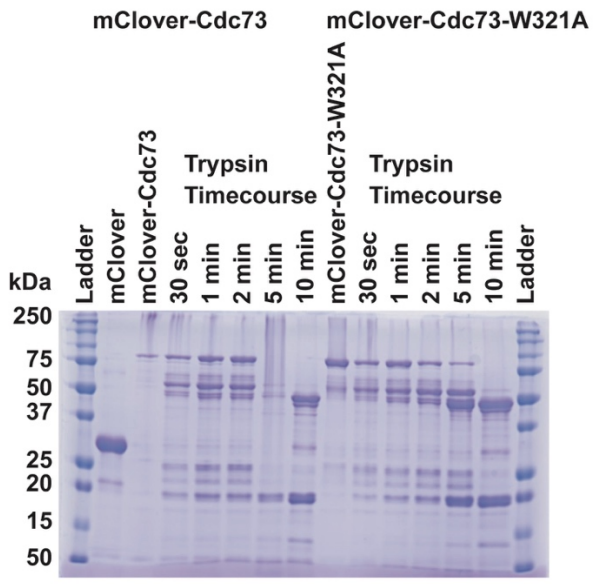


Figure 51. Representative limited proteolysis results for Cdc73 and Cdc73-W321A. Reaction products from a timecourse of trypsin digesttion resolved on an SDS-PAGE gel and stained with Commassie blue. The gel shown is represnetative of three independent replicate experiments.

Appendix C.2 Materials and methods

Appendix C.2.1 Fluorescent thermal shift assay

Protein (5-10 μM) along with the compound or peptide to be tested were added to a buffer containing 10X SYPRO orange dye, 50 mM Tris pH 7, 50 mM NaCl, 10% Glycerol, 1 mM BME in a 96 well plate. Plates were centrifuged briefly and placed into a qPCR machine where they were gradually heated from 20 ˚C to 100 ˚C. During heat denaturation the SYPRO dye is measured over short temperature intervals offering multiple datapoints per degree. Data are plotted as either SYPRO intensity, derivative of SYPRO intensity, or melting temperature (Tm). The Tm is

determined by finding the temperature at the maximum value of the derivative of SYPRO intensity and indicates the point at which 50% of the protein in the reaction has been denatured.

Appendix C.2.2 Biotin pulldown assay

All peptides were labeled with biotin, purified by desalting, and mixed with full length Cdc73 in binding buffer containing 20 mM Tris pH 7, 100 mM NaCl, 1 mM BME, 10% Glycerol. Biotinylated peptides were bound to magnetic streptavidin beads and washed once in binding buffer. Full length Cdc73 was added to the reaction and allowed to incubate at room temperature before washing three times and eluting with SDS PAGE loading dye and heat. Samples were run on a 10% SDS PAGE gel and gels were imaged using an Amersham imager. Refer to Dr. Joel Rosenbaum and my notebooks in the VanDemark lab for further experimental details.

Appendix C.2.3 Peptide gel shift assay

All peptides were labeled with fluorescein, purified by desalting, and mixed with full length Cdc73 in binding buffer containing 20 mM Tris pH 7, 100 mM NaCl, 1 mM BME, 10% Glycerol. Reactions were allowed to incubate at room temperature before loading onto a 15% native PAGE gel and running at 120 volts in cold running buffer. Gel was imaged using an Amersham imager. Refer to Joel Rosenbaum and my notebooks in the VanDemark lab for further experimental details.

Appendix C.2.4 Small molecule compound handing

All compounds were dissolved into solvents at concentrations indicated in Figure 49A (right hand side) and stored at -20 ˚C. All dilution series were made using the solvents listed. As stated above, if this type of screen is done again, I would recommend dissolving all compounds into DMSO at the same molar concentration.

Appendix C.2.5 Dot blot

All compounds were added to a solution of 100mM Tris pH 7.5 and 1% TritonX-100 before loading 2 microliters onto a nitrocellulose membrane. Membrane with compounds added was imaged in all color channels revealing high background for some compounds. Membranes were then blocked for 1hr in 25mM Tris pH 7.5, 100mM NaCl, 10% Glycerol, 1mM BME, and 5% milk. A 10nM or 100nM final concentration of mRuby2-Cdc73-C-domain was added to the blocker and allowed to incubate for an additional 30 minutes. Blots were washed three times (2 minutes each) in 25mM Tris pH 7.5, 100mM NaCl, 10% Glycerol, and 1mM BME before fluorescence imaging.

Appendix C.2.6 Structural modelling

A model of full-length Cdc73 was predicted as described in section 3.2.16. Briefly, the model was created using the I-TASSER threading algorithm and validated by two metrics 1) alignment to various Cdc73 structures (3V46, 5YDE and 6AF0) and 2) the number of valid crosslink lengths in our XL-MS data for Cdc73-10xHis.

Appendix C.2.7 Limited proteolysis

Prior to running the assay 4 µL of 0.5 M PMSF was added to 0.2 mL strip tubes to enable proteolysis of samples to be immediately stopped upon sampling. Cdc73 and trypsin were each added to one large reaction at 1 mg/mL in a reaction buffer containing 20 mM Tris pH 8, 100 mM NaCl, 10% Glycerol, 1 mM BME and incubated for 10 minutes. Over the time course, samples were taken at 30 sec, and 1, 2, 5, and 10 min. Sampling was conducted by pipetting 12 µL of the reaction into a prepared tube containing PMSF and pipetting up and down to mix. Before boiling samples 2 µL of 1 M Tris pH 8 was and 4 µL of 5x SDS page loading dye was added to the samples. Additional Tris was added because in an early replicate of this assay the bromophenol blue in the SDS PAGE loading dye turned yellow upon boiling indicating that the reaction was no longer able to be properly buffered at high temperature with the amount of Tris used. This is likely cause by a combination of 1) trypsin addition, because it is stored in an acidic storage buffer to prevent autoproteolysis, and 2) the fact that Tris buffer pH decreases as temperature increases (see https://www.neb.com/tools-and-resources/usage-guidelines/ph-vs-temperature-for-tris-buffer).

Appendix D

Appendix D.1 Additional Spt6-Cdc73 interaction data

Appendix D.1.1 AID tagging

I made multiple strains containing a form of the *SPT6* gene that produces a protein product tagged with an auxin inducible degron (AID) tag. I initially did this by C-terminally tagging Spt6 with the full *IAA7* gene from *Arabidopsis thaliana* and a 3XV5 epitope tag (the strain used in Chapter 2). Although, this allowed for rapid depletion of the Spt6 protein it had a slow growth phenotype and an *spt-* phenotype (see Figure 52A-52D). I next generated a mini-AID tag based off boundaries ascertained by Dr. Sarah Hainer for a minimal degron tag and tagged Spt6 C-terminally with this smaller version of the *IAA7* protein also containing the 3XV5 epitope tag. This strain grew slightly less slowly, but still had the *spt-* phenotype (Figure 52A-52C). The mini-AID strain also depleted Spt6 more slowly compared to the full-length tag (Figure 52E). Finally, I decided to generate the appropriate plasmid to allow me to integrate the mini-AID tag at the N-terminus or internally at the *SPT6* locus. I attempted to integrate this tag at both the N-terminus and internally between the S1 domain and the tSH2 domain of Spt6 in a large linker region. I was not able to get the N-terminal integration completed, but was able to complete the internally tagged strain, which to my surprise grew much like wild type yeast and had a much less sever *spt-* phenotype (Figure 52A-52C). Unfortunately, depletion of Spt6 in the internally tagged strain was unsatisfactory with only about a 2-fold reduction in Spt6 observed after 60 minutes in auxin treated media (Figure 52F). During the process of validating these strains I grew ChIP cultures for the C-

terminal mini39AID tagged strain and conducted a ChIP-qPCR experiment after a 60-minute auxin or vehicle (EtOH) treatment which demonstrated that using this method I can recapitulate the Paf1 occupancy results obtained in the *spt6-50* strain at the *PMA1* gene (Figure 52G). After much discussion it was finally decided that I should use the original strain containing the full-length AID tag because rapid and complete depletion was favored over prevention of the slow growth and *spt-* phenotypes. Because the strain is always compared to itself (DMSO treated versus IAA treated), any effects of the tag itself should be negated.

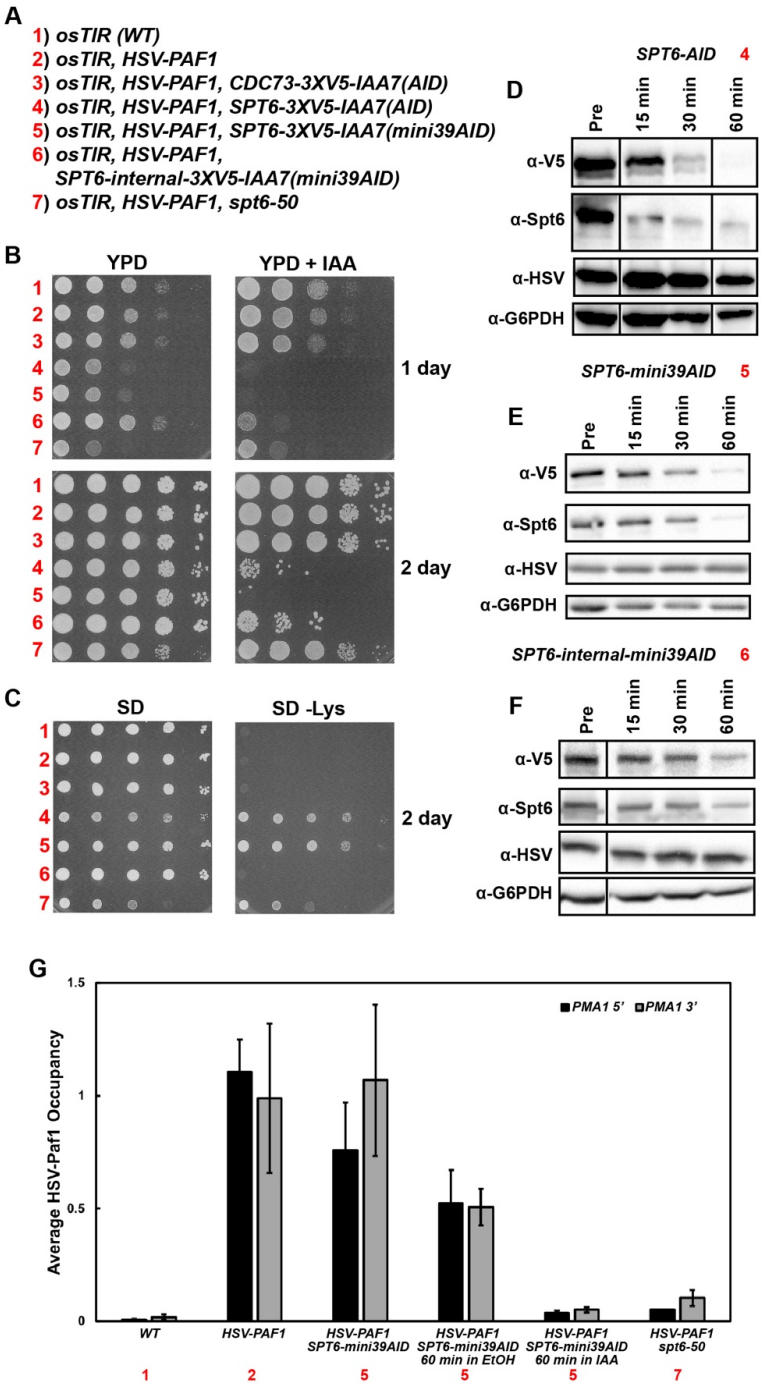


Figure 52. Spt6 auxin inducible degron strain validation and comparison.

A) List of relevant genotypic information for strains assayed in this figure. B) Serial dilution assays comparing growth on YPD and YPD+IAA media. Serial dilution assays were conducted by spotting 2 µL each of 10-fold dilutions (108-104 cells/mL) of stationary phase yeast cells from the strains indicated in A. C)

Serial dilution assays comparing growth on SD and SD-Lys media to assess Spt- phenotype. D) Western blot carried out on protein extracts from *SPT6-AID* strain #4 in panel A). Protein levels for Spt6-AID (α-V5 and α-Spt6), HSV-Paf1 (α-HSV), and a loading control (α-G6PDH) without auxin treatment and after 15, 30, and 60 minutes post- treatment with auxin (500 μM IAA dissolved in 100% EtOH). E) As in D, but assaying the Spt6-mini39AID strain (#5 in panel A). F) As in D and E, but measuring levels of SPT6-internal-AID strain (#6 in panel A) and using DMSO as a vehicle instead of EtOH. G) ChIP-qPCR results from a pilot experiment showing HSV-Paf1 levels after the SPT6-mini39AID strain with IAA or EtOH (vehicle) for 60 minutes and comparing those results to control strains and an spt6-50 strain.

Appendix D.1.2 Spt5 and Spt6 levels in strains deleted for genes encoding Paf1C subunits

To test the possibility that loss of all Paf1C subunits or individual Paf1C subunits would result in decreased Spt5 and Spt6 levels I performed a western blot on protein extracts from strains lacking all or individual Paf1C members (Figure 53). The results indicate that Spt5 and Spt6 remain at wild type levels in all genetic backgrounds tested. This experiment was performed in biological duplicate and similar results were obtained from each replicate (Figure 53, compare Rep 1 to Rep 2).

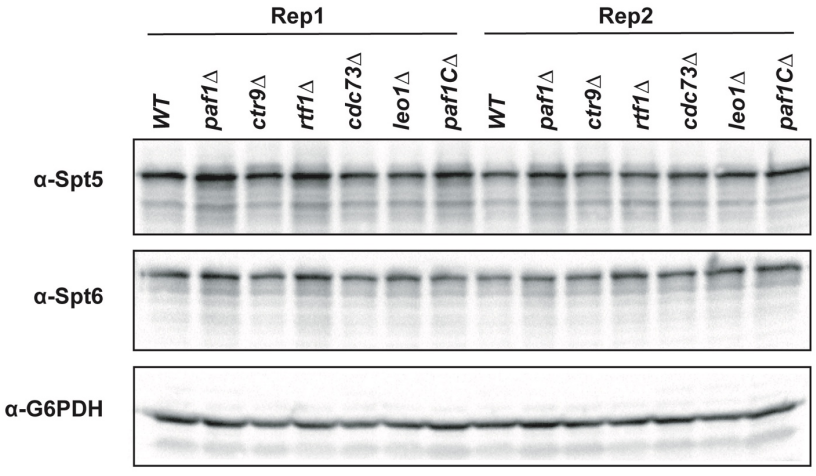


Figure 53. Western analysis of Spt5 and Spt6 stability in Paf1C member deletion strains.

Appendix D.1.3 *spt6-50* long gene ChIP results

When I obtained the ChIP-seq results for the *spt6-50* mutant it appeared that there were increases in Spt6 occupancy at lowly transcribed genes which raised the question: is Spt6 interacting with nucleosomes when it is missing the tSH2 domain? In other words, does Spt6 require Pol II passage to interact with chromatin in the absence of its tSH2 domain? To test this hypothesis, I performed the long-gene assay which is often used to measure Pol II elongation rate. The assay measures Pol II or other elongation factors at various locations on a "long gene" by ChIP-qPCR before and at multiple timepoints after turning off the gene (see methods section below for more details). I measured Rpb3 (to show occupancy loss upon gene shutoff), Spt16 (as a positive control for Pol II dependent nucleosome interaction) and Spt6 occupancy in galactose (ON) and glucose (OFF) media conditions by ChIP-qPCR (Figure 54). The results indicate that Spt6 requires Pol II passage to interact with chromatin even in the absence of its tSH2 domain (*spt6-50* strain). This may indicate that it can work like FACT by responding to disrupted nucleosomes or may be the result of the Spt6 cores interaction with Pol II.

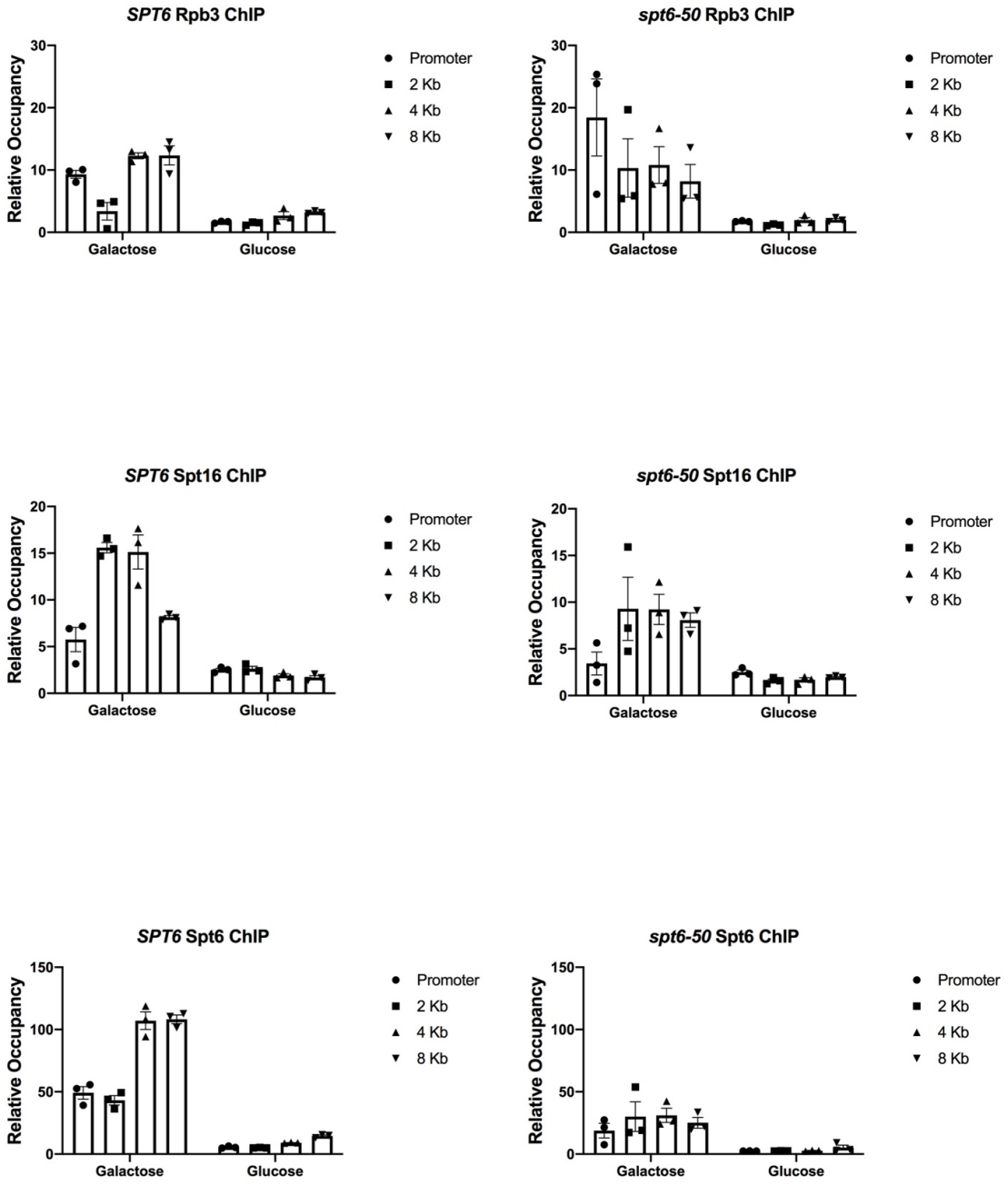


Figure 54. Assessing the requirement for Pol II in the association between Spt6 and chromatin by ChIP-qPCR. Glucose samples were exposed to glucose for 8 minutes to provide enough time for all Pol II to clear the gene body.

Appendix D.1.4 Spt16 ChIP-qPCR results in Spt6-AID

Given the ability to rapidly deplete Spt6 I wanted to test the hypothesis that when the chaperoning activity of Spt6 is lost FACT compensates for it. I predicted that, if this hypothesis was true, I would observe an increase in FACT occupancy as measured by its Spt16 subunit. I was surprised to find that FACT occupancy decreased upon loss of Spt6 (Figure 55). These data indicate that Spt6 is required for proper FACT occupancy, at least at the 5' end of the *PMA1* gene.

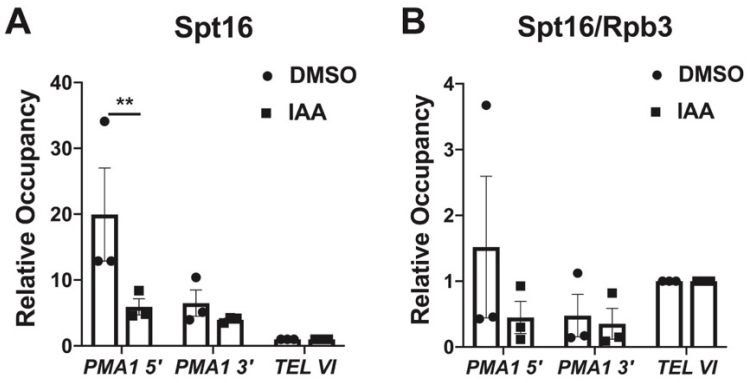


Figure 55. ChIP-qPCR results for Spt16 upon Spt6 depletion.

Appendix D.1.5 Effects of Spt6-AID depletion on Paf1C-promoted histone modifications

To test the hypothesis that Spt6 depletion affects Paf1C-dependent histone modifications, I performed ChIP-qPCR on histone modifications stimulated by Paf1C. As a control, H2B levels do not appear to change, although the H2B replicates do not all agree well in all regions tested (Figure 56A). However, H3 levels decrease (Figure 57A) at the highly transcribed gene *PMA1* upon Spt6 depletion which is not surprising given that Spt6 chaperones H3. Both H2B K123ub (Figure 56B) and H3 K4me2/3 (Figure 57B and 57C) decrease upon loss of Spt6. When the signals

for the histone marks are compared to the appropriate total histone levels, the trend holds (Figures 56C, 57D, and 57E). This suggests that Spt6 depletion alters levels of Paf1C-promoted histone modifications. This is not surprising given recently published results (Jeronimo *et al.* 2019) and the results presented here.

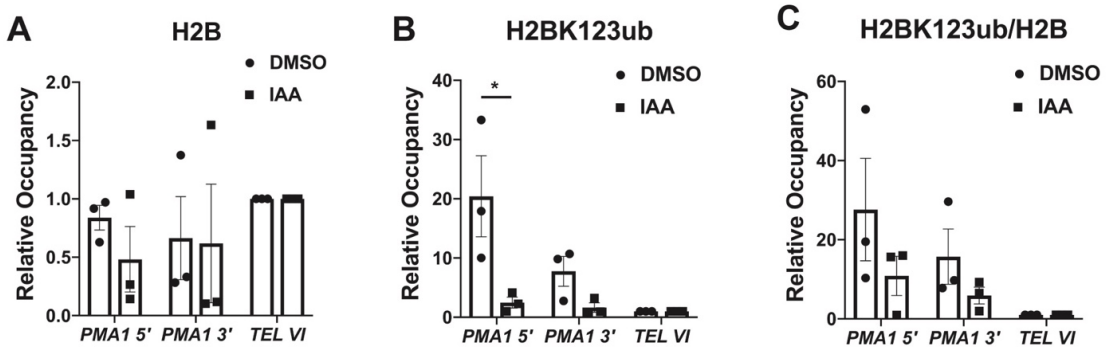


Figure 56. ChIP-qPCR results for H2B and H2B K123ub upon Spt6 depletion

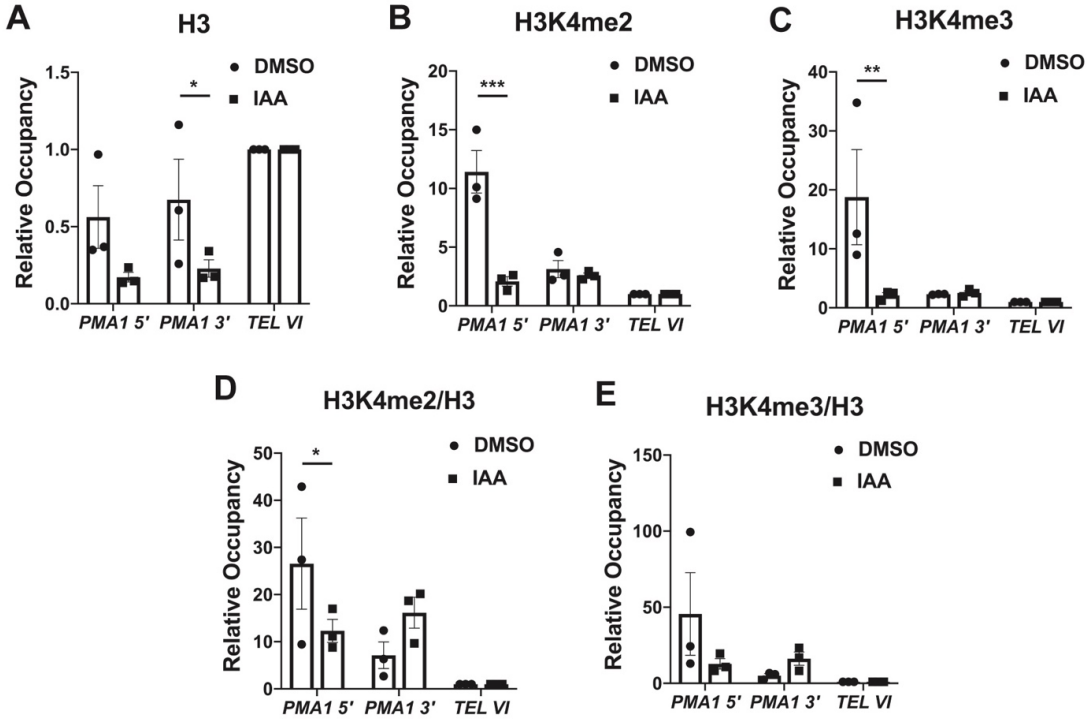


Figure 57. ChIP-qPCR results for H3 and H3 K4me2/3 upon Spt6 depeltion.

Appendix D.1.6 tRNA occupancy data

I observed an increase in Spt6 at tRNA genes after depletion of Spt6 in the *SPT6-AID* strain. This was surprising and lead me to perform additional analyses on the tRNA genes in all of my ChIP-seq datasets. I found that Spt6 levels do not change at tRNAs in the *spt6-50* mutant, but Paf1 levels do increase. In the *SPT6-AID* strain after one-hour IAA treatment I observe an increase in both Spt6 and Paf1 levels at tRNA genes. The tRNA genes are transcribed by Pol III, begging the question what are the Pol II elongation factors Spt6 and Paf1 doing at these genes. Additionally, why is Spt6 retained here upon AID depletion?

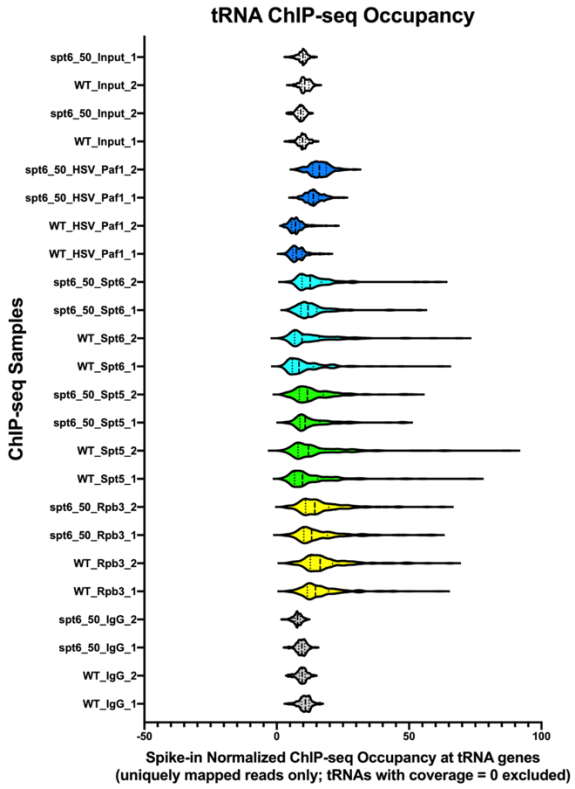


Figure 58. Violin plots of occupancy of Rpb3, Spt5, Spt6, and Paf1 at tRNA genes in WT and *spt6-50* strains

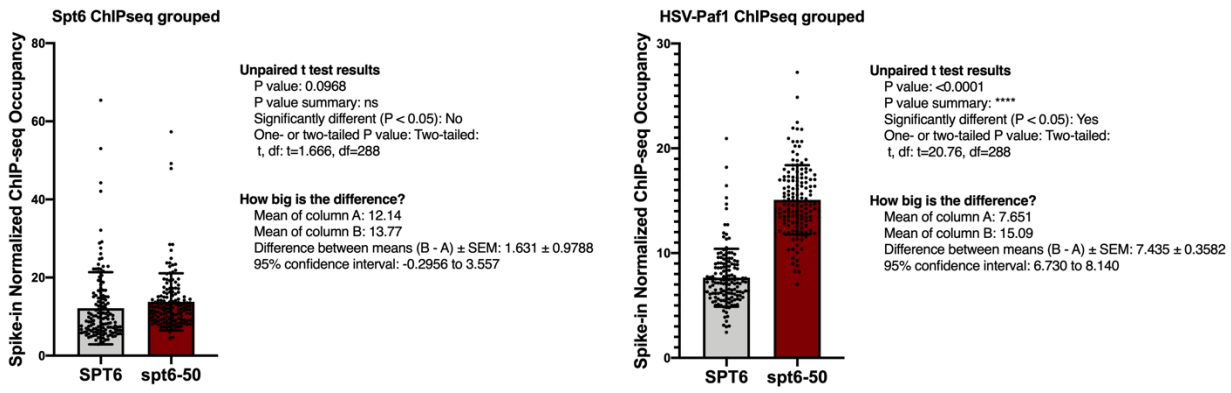


Figure 59. Bar graphs and statistics for Spt6 and Paf1 in WT and *spt6-50* strains.

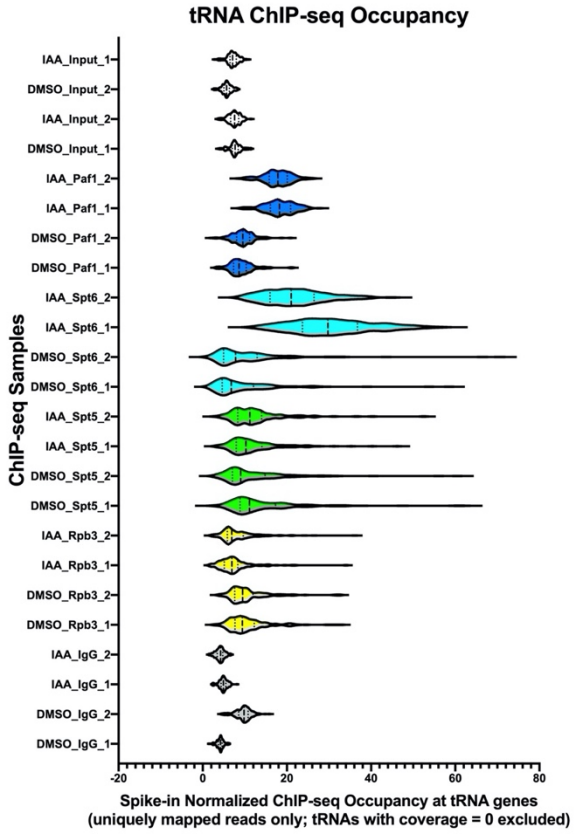


Figure 60. Violin plots of occupancy of Rpb3, Spt5, Spt6, and Paf1 at tRNA genes the *SPT6-AID* strain treated with DMSO or IAA.

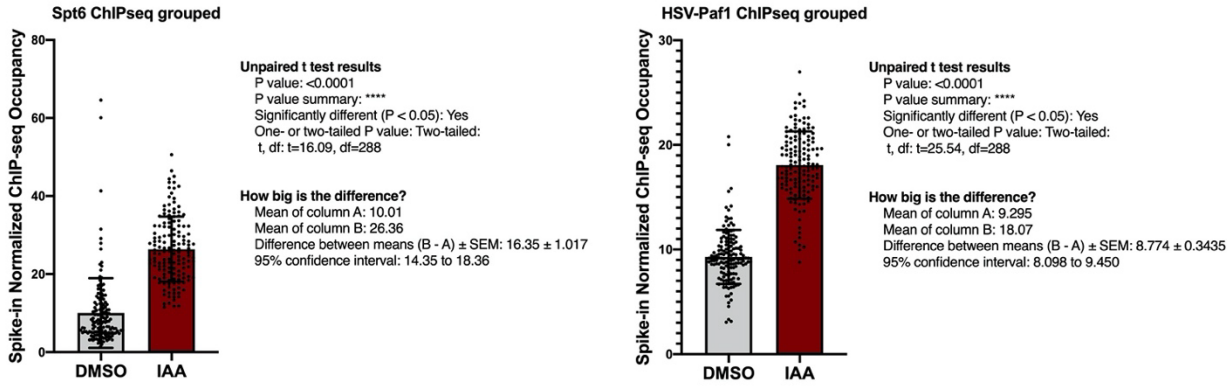


Figure 61. Bar graphs and statistics for Spt6 and Paf1 in the *SPT6-AID* strain treated with DMSO or IAA.

Appendix D.1.7 FRET attempts

I attempted to use FRET to assess the Cdc73-Spt6 interaction numerous times and was unable to get reproducible data at the higher titration points in any of my experiments (Figure 61A). I also attempted to choose a single titration point to try to measure the difference between two proteins, in this case full length Cdc73 and the C-domain, and the results initially looked promising (Figure 62B). However, in subsequent experiments where each fluorescent protein tag, mRuby2 or mClover, was assayed with mClover-Cdc73 or mRuby-Spt6 239-1451, respectively, as negative controls. The results showed that binding curves generated from reactions containing mClover-Cdc73 + mRuby-Spt6 239-1451, mClover + mRuby-Spt6 239-1451, and mClover-Cdc73 or mRuby were equivalent. This suggested that the assay could not work for these two proteins (Figure 63). With the linker and large fluorophores used in these fusion proteins it is difficult to obtain true binding signal because of the mobility of the two fluorophores and this leaves us only measuring background signal or noise which is why we see our binding curves generated from Cdc73 and Spt6 matching our negative controls. We get what look like the beginning of binding curves for the controls simply from the crowding effects of adding the high concentrations of protein, which leads to some background FRET.

I think this assay is still promising for use with different proteins. First, the Cytation5 plate reader could easily be used to perform a FRET experiment with proteins labeled with small molecule fluorophores. Second, it may be possible to use this method with mRuby2 and mClover if the linker used is very short or the Kd of the interaction is much lower than that of the Spt6-Cdc73 interaction, requiring the addition of less donor molecule, and thus eliminating the background FRET signal observed at high donor concentrations. In short, this assay is promising,

but could not work in the specific context I was using it. That does not mean that future studies should automatically rule it out though.

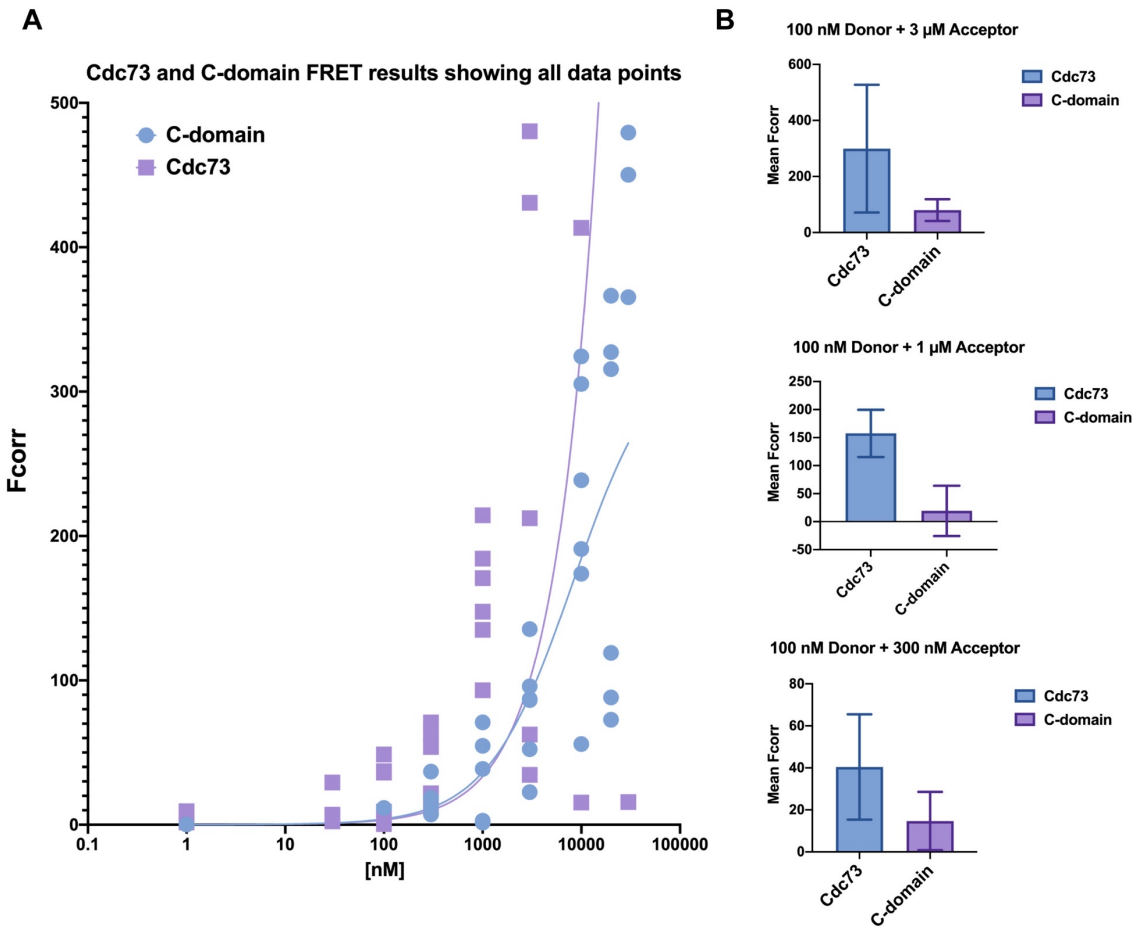


Figure 62. FRET appears promising for full length Cdc73 and Cdc73 C-domain.

A) Fcorr values for all datapoints in all replicates and B) mean and standard deviation of Fcorr values at select titration points. Cdc73 constructs are tagged with mClover (donor) and Spt6 constructs are tagged with mRuby2 (acceptor)

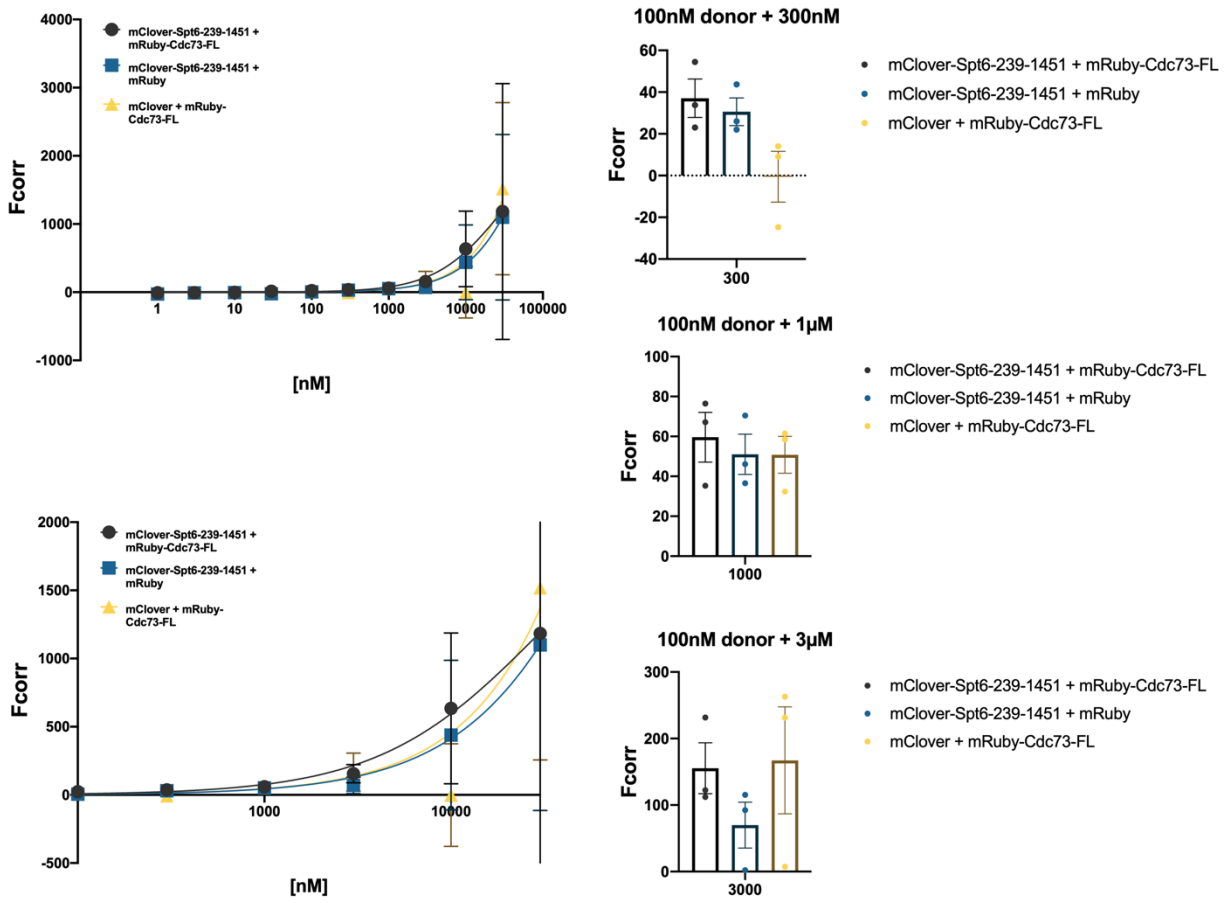


Figure 63. FRET is inconclusive for Spt6 239-1451 and full length Cdc73 with mRuby2 and mClover controls. Binding curves and bar graphs are presented as in Figure 62. Top binding curve contains all data points assayed and the bottom one is truncated at the y-axis to zoom in on the datapoints relevant for the bar graphs. Fluorescent tags are the same as in Figure 62.

Appendix D.1.8 Crosslinking and mass spectrometry data for full Paf1C and Pol II

Using protein sent to us by the Cramer lab I performed crosslinking (Figure 64A and 64B) and mass spectrometry on Spt6 and the full Paf1C. Due to limited time and resources in our collaborator's lab, we were unable to see this experiment through to completion. However, we were able to obtain a sparse DSS crosslinking dataset for Paf1C + Spt6 samples (Figure 65; top

network). From this it appears as though Spt6 crosslinks to the Plus3 domain/OAR of Rtf1 (colored brown in the Rtf1 diagram). In this experiment I was unable to detect the Cdc73 Spt6 interaction. I am confident that this interaction would show up if I were able to process more sample, but this was not possible. If I combine our Spt6-Cdc73 crosslinking data with the data for the full complex (Figure 65; middle network) it offers a more complete picture of the interaction. Finally, a network generated from the Cramer labs BS3 data (Xu *et al.* 2017) for the full Paf1C looks very similar to the data in my sparse DSS dataset suggesting that the data I collected for the full Paf1C is accurate (Figure 65; bottom network).

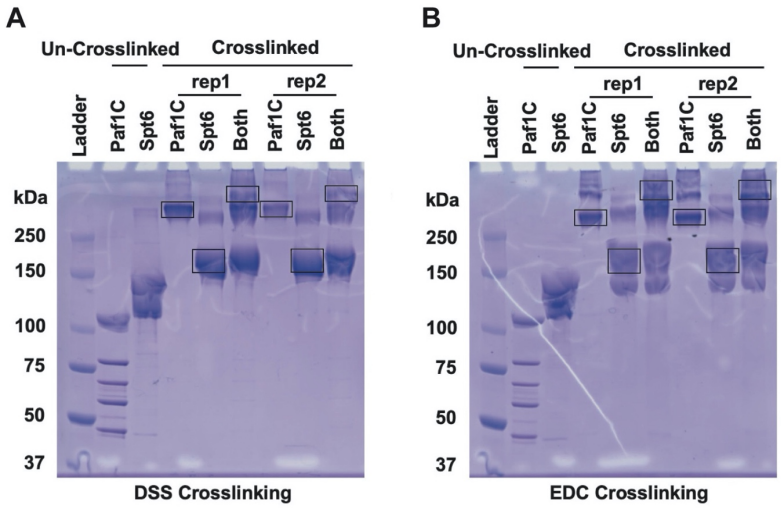


Figure 64. EDC and DSS crosslinking products for full Paf1C and Spt6.

A) EDC and B) DSS crosslinking products run on 4-12% Bis-Tris SDS PAGE and stained with coomassie blue.

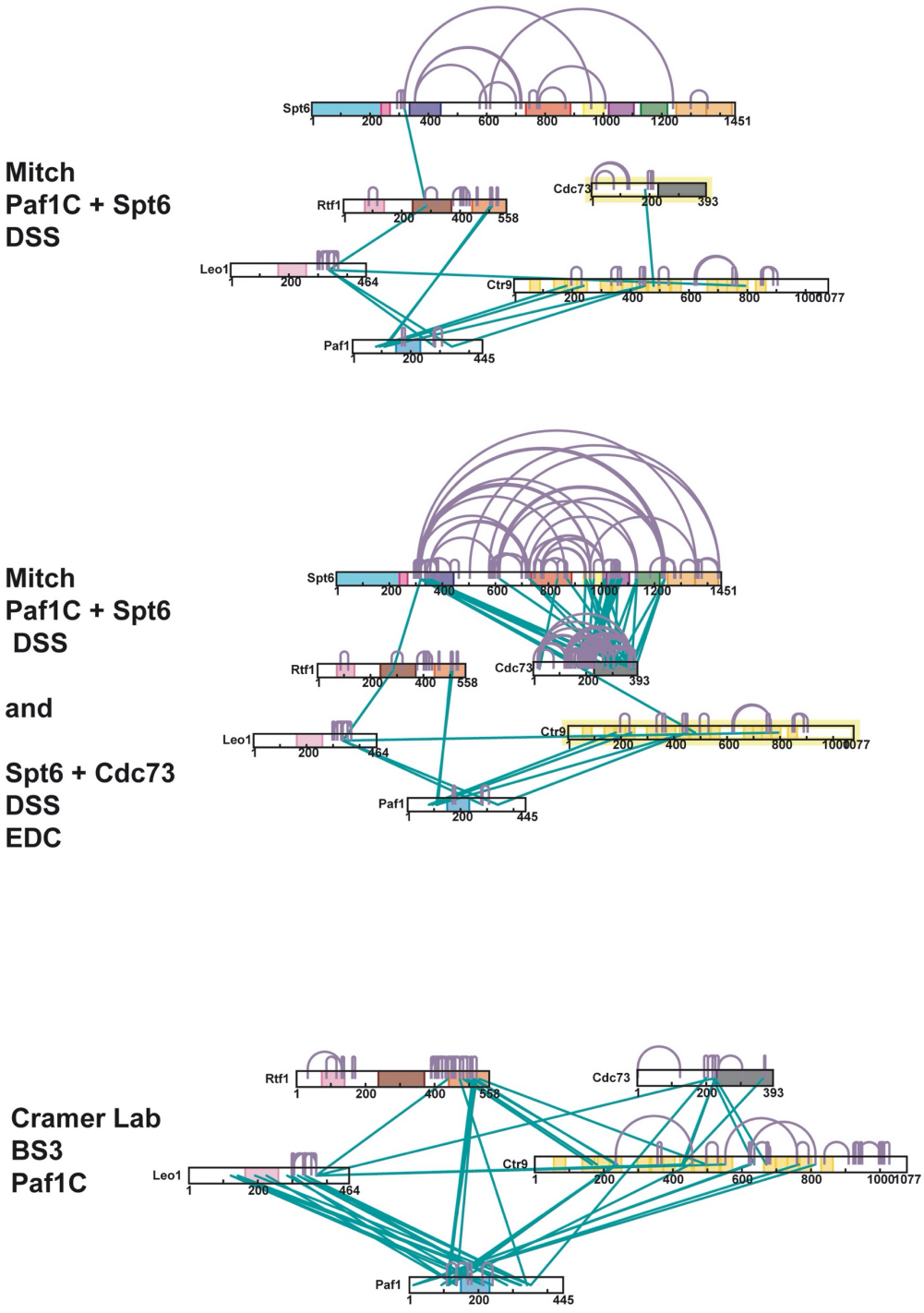


Figure 65. Crosslinking networks for Paf1C members and Spt6.

Purple lines represent intra-protein crosslinks and green lines represent inter-protein crosslinks (Xu et al. 2017).

Appendix D.1.9 Recombinant Spt6-Cdc73 pulldowns

Initially I had looked at the Spt6-Cdc73 interaction using nickel pulldowns with Cdc73 as bait and Spt6 as prey. I was able to detect the interaction above background Spt6 binding in some but not all replicates. The replicates shown here are among the best data I collected. The one conclusion that can be made from this assay is that the tSH2 domain does not interact with Cdc73 or the C-domain. At the time this assay was conducted this was very surprising because I had first hypothesized Cdc73 interacts with the tSH2 domain. This seemed logical at the time because the Spt6-tSH2 domain interacts with the CTD/CTD linker (Sdano *et al.* 2017; Brázda *et al.* 2020) and the *in vivo* BPA crosslinking data (Figure 19) showed Cdc73 binding to both the CTD and Spt6.

The Spt6 core has a fairly strong negative charge. This may cause nonspecific binding to the positively charged nickel causing this assay to have high background. It may be possible to get better results using Spt6 as the bait and Cdc73 as the prey. This assay is relatively straight forward and only takes a short time to run. Therefore, it may be worth the effort to troubleshooting for this protein-protein interaction or other interactions studied in the lab.

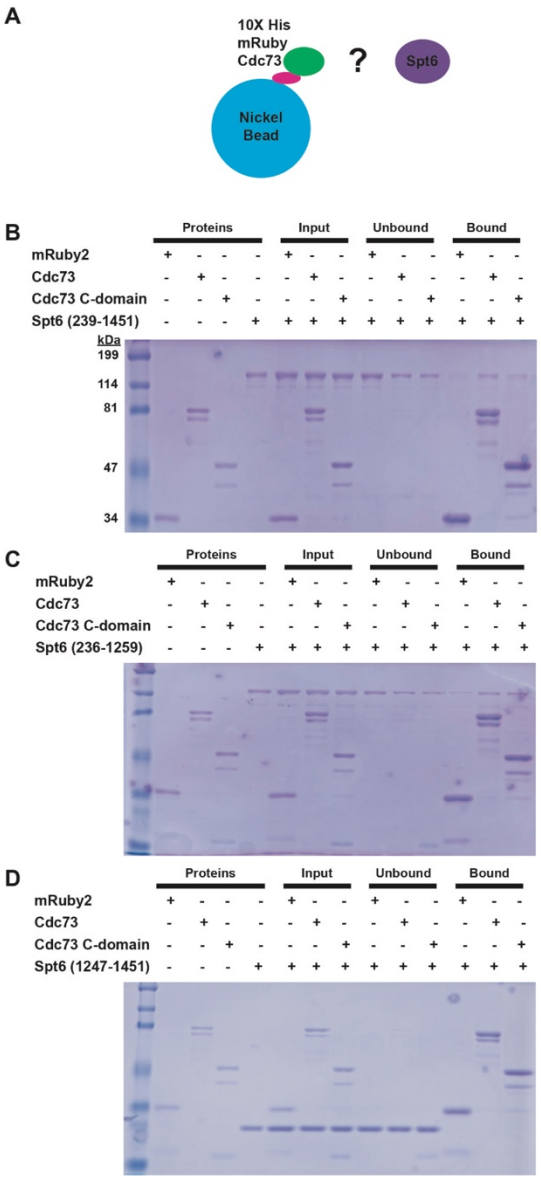


Figure 66. Nickel affinity pulldown assay results.

A) Diagram of nickel affinity pulldown assay showing mRuby2 Cdc73 (bait) bound to nickel beads via its 10xHis tag and Spt6 in solution (prey). B-D) Binding assay results for three Spt6 constructs: Spt6 239- 1451 (B), Spt6 236-1259 (C), and Spt6 1247-1451 (D). Each version of Spt6 is tested against mRuby2 Cdc73, mRuby2 Cdc73 C-domain, and mRuby2 alone (negative control).

Appendix D.1.10 Spt6 truncation gel-shift data

To further dissect the Spt6-Cdc73 interaction, Matthew Blacksmith cloned, expressed, and purified truncated versions of the mRuby2-Spt6 fusion protein (Figure 67A). Matt conducted native gel-shift assays (Figure 67B) and his results indicate that the S1 domain and perhaps even the tSH2 domain are necessary for the Spt6-Cdc73 interaction. This is most likely because their presence confers stability to the Spt6 core.

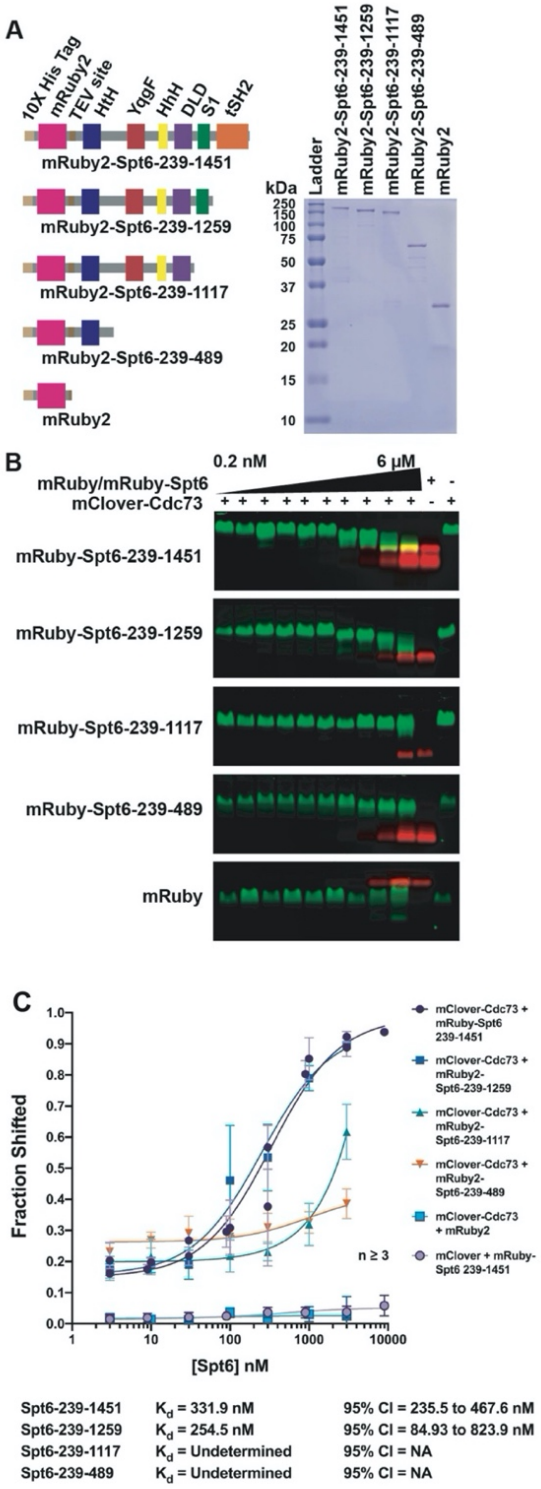
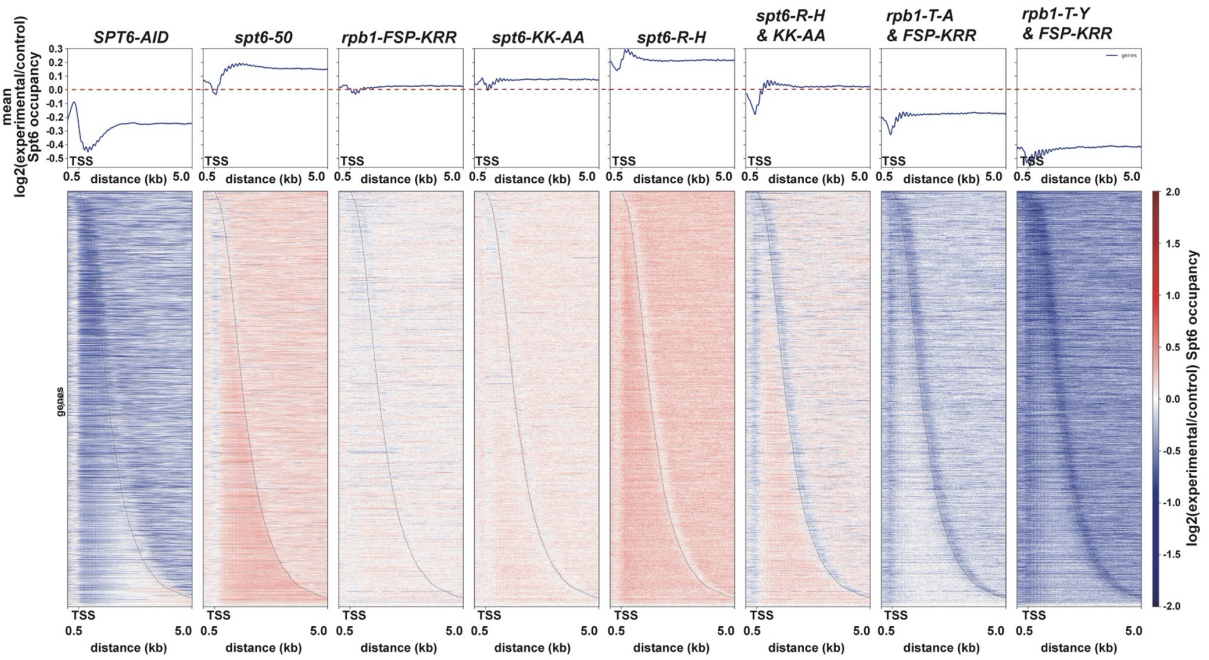


Figure 67. Spt6 core domains are critical for Cdc73 binding.

(A) Diagrams of truncated Spt6 fusion proteins purified from *E.coli* used in native gel-shift assays in B (left) and a 15% SDS PAGE gels stained with Coomassie blue showing final purity of the proteins used in the assays in B (right). (B) Native gel-shift assays conducted as described in Figure 1C. C) Quantified data from the gel-shift experiment shown in B. Data were quantified as in Figure 21. At least 4 replicates experiments were performed for each Spt6 construct.

Appendix D.1.11 Sdano *et al* 2017 data re-analysis

I re-analyzed data from (Sdano *et al.* 2017) because my *spt6-50* (mutant lacking the tSH2 domain) data showed increased Spt6 occupancy at many genes, while their analysis of tSH2 domain mutants (mutating phosphate binding pockets; *spt6 R-H & KK-AA*) showed a global decrease in Spt6 occupancy. I also noticed that the authors performed a reads per million normalization on the data before applying the reference adjusted reads per million method (Orlando *et al.* 2014). The two normalizations are not meant to be used together and when they are used in this way it is possible to obtain incorrect results. Upon re-analysis I found that all results from the published paper held except for the result for the *spt6 R-H & KK-AA* mutant, which went from the reported global decrease to a global increase similar to what I observed in the *spt6-50* strain, but with clear decreases in occupancy outside of the 5' and 3' ends of genes.



Mutant	Mutant Shorthand	RRPM (my analysis) Direction of Spt6 Occupancy Change	RPM+RRPM (Hill analysis) Direction of Spt6 Occupancy Change
Spt6-AID (after 60 min in IAA)		-	
spt6-50		+	
rpb1-LMFSPLV1490KMKRRRK	*rpb1-FSP-KRR*	+	+
spt6-K1355A,K1435A	*spt6-KK-AA*	+	+
spt6-R1282H	*spt6-R-H*	+	+
spt6-R1282H,K1355A,K1435A	*spt6-R-H & KK-AA*	+	-
rpb1-T1471A, LMFSPLV1490KMKRRRK	*rpb1-T-A & FSP-KRR*	-	-
rpb1-Y1473A, LMFSPLV1490KMKRRRK	*rpb1-T-Y & FSP-KRR*	-	-

Figure 68. Results from re-analysis of Sdano *et al.*, 2017 data

Heatmaps showing log2 fold change in experimental strain or treatment (IAA) versus wild type control strain or treatment (DMSO) for all yeast coding regions (top) and a table summarizing the results (bottom).

Appendix D.1.12 Crosslinking results tables

Table 26. Table containing all DSS and EDC crosslink sites for mClover-Cdc73 and Spt6 (239-1451)

Protein 1	Residue 1	Protein 2	Residue 2	Crosslinker
Cdc73	77	Cdc73	17	DSS
Cdc73	119	Cdc73	130	DSS
Cdc73	119	Cdc73	136	DSS
Cdc73	119	Cdc73	205	DSS
Cdc73	119	Cdc73	371	DSS
Cdc73	124	Cdc73	304	DSS
Cdc73	130	Cdc73	185	DSS
Cdc73	130	Cdc73	371	DSS
Cdc73	169	Cdc73	185	DSS
Cdc73	169	Cdc73	205	DSS
Cdc73	169	Cdc73	304	DSS
Cdc73	169	Cdc73	371	DSS
Cdc73	185	Cdc73	199	DSS
Cdc73	185	Cdc73	205	DSS
Cdc73	185	Cdc73	371	DSS
Cdc73	194	Cdc73	205	DSS
Cdc73	205	Cdc73	17	DSS
Cdc73	205	Cdc73	212	DSS
Cdc73	205	Cdc73	371	DSS
Cdc73	226	Cdc73	205	DSS
Cdc73	236	Cdc73	205	DSS
Cdc73	236	Cdc73	226	DSS
Cdc73	236	Cdc73	304	DSS
Cdc73	263	Cdc73	366	DSS
Cdc73	263	Cdc73	371	DSS
Cdc73	282	Cdc73	205	DSS
Cdc73	282	Cdc73	371	DSS
Cdc73	304	Cdc73	17	DSS
Cdc73	304	Cdc73	130	DSS
Cdc73	304	Cdc73	185	DSS
Cdc73	304	Cdc73	205	DSS
Cdc73	304	Cdc73	218	DSS
Cdc73	304	Cdc73	226	DSS
Cdc73	304	Cdc73	371	DSS

Cdc73	366	Cdc73	371	DSS
Cdc73	368	Cdc73	371	DSS
Cdc73	371	Cdc73	366	DSS
Cdc73	371	Cdc73	368	DSS
Cdc73	385	Cdc73	226	DSS
Cdc73	385	Cdc73	371	DSS
Cdc73	205	Spt6	319	DSS
Cdc73	263	Spt6	1134	DSS
Cdc73	304	Spt6	319	DSS
Cdc73	371	Spt6	319	DSS
Cdc73	371	Spt6	870	DSS
Cdc73	371	Spt6	1007	DSS
Spt6	306	Spt6	316	DSS
Spt6	306	Spt6	319	DSS
Spt6	306	Spt6	715	DSS
Spt6	306	Spt6	956	DSS
Spt6	306	Spt6	1357	DSS
Spt6	307	Spt6	319	DSS
Spt6	307	Spt6	956	DSS
Spt6	316	Spt6	319	DSS
Spt6	316	Spt6	956	DSS
Spt6	354	Spt6	715	DSS
Spt6	354	Spt6	719	DSS
Spt6	354	Spt6	723	DSS
Spt6	379	Spt6	319	DSS
Spt6	504	Spt6	1444	DSS
Spt6	576	Spt6	1444	DSS
Spt6	598	Spt6	354	DSS
Spt6	598	Spt6	719	DSS
Spt6	616	Spt6	1134	DSS
Spt6	616	Spt6	1239	DSS
Spt6	728	Spt6	723	DSS
Spt6	746	Spt6	773	DSS
Spt6	749	Spt6	773	DSS
Spt6	773	Spt6	746	DSS
Spt6	834	Spt6	773	DSS
Spt6	870	Spt6	781	DSS
Spt6	937	Spt6	773	DSS
Spt6	956	Spt6	319	DSS
Spt6	958	Spt6	319	DSS

Spt6	958	Spt6	723	DSS
Spt6	1007	Spt6	746	DSS
Spt6	1007	Spt6	749	DSS
Spt6	1007	Spt6	773	DSS
Spt6	1047	Spt6	1007	DSS
Spt6	1228	Spt6	1239	DSS
Spt6	1239	Spt6	611	DSS
Spt6	1239	Spt6	612	DSS
Spt6	1299	Spt6	1357	DSS
Spt6	1355	Spt6	1439	DSS
Spt6	1355	Spt6	1444	DSS
Spt6	1366	Spt6	1357	DSS
Cdc73	24	Cdc73	205	EDC
Cdc73	24	Cdc73	371	EDC
Cdc73	119	Cdc73	126	EDC
Cdc73	119	Cdc73	141	EDC
Cdc73	130	Cdc73	141	EDC
Cdc73	156	Cdc73	205	EDC
Cdc73	158	Cdc73	205	EDC
Cdc73	176	Cdc73	205	EDC
Cdc73	176	Cdc73	371	EDC
Cdc73	185	Cdc73	176	EDC
Cdc73	205	Cdc73	141	EDC
Cdc73	263	Cdc73	24	EDC
Cdc73	24	Spt6	834	EDC
Cdc73	185	Spt6	940	EDC
Cdc73	185	Spt6	1062	EDC
Cdc73	185	Spt6	1063	EDC
Cdc73	194	Spt6	1062	EDC
Cdc73	199	Spt6	1062	EDC
Cdc73	205	Spt6	971	EDC
Cdc73	205	Spt6	1009	EDC
Cdc73	205	Spt6	1012	EDC
Cdc73	205	Spt6	1038	EDC
Cdc73	205	Spt6	1048	EDC
Cdc73	205	Spt6	1049	EDC
Cdc73	205	Spt6	1063	EDC
Cdc73	205	Spt6	1071	EDC
Cdc73	205	Spt6	1076	EDC
Cdc73	304	Spt6	353	EDC

Cdc73	304	Spt6	940	EDC
Cdc73	304	Spt6	1012	EDC
Cdc73	304	Spt6	1062	EDC
Cdc73	371	Spt6	345	EDC
Cdc73	371	Spt6	940	EDC
Spt6	301	Spt6	307	EDC
Spt6	306	Spt6	312	EDC
Spt6	306	Spt6	861	EDC
Spt6	314	Spt6	319	EDC
Spt6	335	Spt6	306	EDC
Spt6	335	Spt6	383	EDC
Spt6	342	Spt6	431	EDC
Spt6	384	Spt6	379	EDC
Spt6	462	Spt6	431	EDC
Spt6	699	Spt6	1239	EDC
Spt6	940	Spt6	773	EDC
Spt6	1007	Spt6	971	EDC
Spt6	1007	Spt6	1055	EDC
Spt6	1007	Spt6	1058	EDC
Spt6	1007	Spt6	1062	EDC
Spt6	1007	Spt6	1063	EDC
Spt6	1009	Spt6	773	EDC
Spt6	1069	Spt6	1038	EDC
Spt6	1184	Spt6	1278	EDC
Spt6	1230	Spt6	319	EDC
Spt6	1232	Spt6	611	EDC
Spt6	1232	Spt6	1239	EDC
Spt6	1235	Spt6	1239	EDC
Spt6	1236	Spt6	611	EDC
Spt6	1237	Spt6	611	EDC

Table 27. Table containing only DSS and EDC crosslink sites for mClover-Cdc73 and Spt6 (239-1451) used in modeling

Protein 1	Residue 1	Protein 2	Residue 2	Crosslinker
Cdc73	185	Spt6	316	DSS
Cdc73	205	Spt6	319	DSS
Cdc73	263	Spt6	1134	DSS
Cdc73	304	Spt6	319	DSS
Cdc73	366	Spt6	1239	DSS
Cdc73	371	Spt6	319	DSS
Cdc73	371	Spt6	616	DSS
Cdc73	371	Spt6	870	DSS
Cdc73	371	Spt6	1007	DSS
Cdc73	371	Spt6	1134	DSS
Cdc73	371	Spt6	1239	DSS
Cdc73	385	Spt6	1239	DSS
Spt6	306	Spt6	1357	DSS
Spt6	504	Spt6	1444	DSS
Spt6	576	Spt6	1444	DSS
Cdc73	24	Spt6	834	EDC
Cdc73	185	Spt6	940	EDC
Cdc73	185	Spt6	1062	EDC
Cdc73	185	Spt6	1063	EDC
Cdc73	194	Spt6	1062	EDC
Cdc73	199	Spt6	1062	EDC
Cdc73	205	Spt6	971	EDC
Cdc73	205	Spt6	1009	EDC
Cdc73	205	Spt6	1012	EDC
Cdc73	205	Spt6	1038	EDC
Cdc73	205	Spt6	1048	EDC
Cdc73	205	Spt6	1049	EDC
Cdc73	205	Spt6	1063	EDC
Cdc73	205	Spt6	1071	EDC
Cdc73	205	Spt6	1076	EDC
Cdc73	304	Spt6	353	EDC
Cdc73	304	Spt6	940	EDC
Cdc73	304	Spt6	1012	EDC
Cdc73	304	Spt6	1062	EDC
Cdc73	304	Spt6	1062	EDC
Cdc73	371	Spt6	345	EDC

Cdc73	371	Spt6	940	EDC
Spt6	1184	Spt6	1278	EDC

Table 28. Table of DSS and EDC crosslinks from Cdc73-10xHis used for determining which Cdc73 structural model to use

Protein 1	Residue 1	Protein 2	Residue 2	Crosslinker
Cdc73	21	Cdc73	13	DSS
Cdc73	77	Cdc73	17	DSS
Cdc73	119	Cdc73	13	DSS
Cdc73	119	Cdc73	130	DSS
Cdc73	119	Cdc73	136	DSS
Cdc73	119	Cdc73	185	DSS
Cdc73	119	Cdc73	205	DSS
Cdc73	119	Cdc73	304	DSS
Cdc73	124	Cdc73	13	DSS
Cdc73	124	Cdc73	185	DSS
Cdc73	124	Cdc73	304	DSS
Cdc73	124	Cdc73	371	DSS
Cdc73	124	Cdc73	385	DSS
Cdc73	130	Cdc73	185	DSS
Cdc73	130	Cdc73	205	DSS
Cdc73	130	Cdc73	263	DSS
Cdc73	136	Cdc73	124	DSS
Cdc73	136	Cdc73	185	DSS
Cdc73	136	Cdc73	263	DSS
Cdc73	136	Cdc73	304	DSS
Cdc73	136	Cdc73	371	DSS
Cdc73	169	Cdc73	205	DSS
Cdc73	169	Cdc73	371	DSS
Cdc73	169	Cdc73	385	DSS
Cdc73	185	Cdc73	199	DSS
Cdc73	185	Cdc73	205	DSS
Cdc73	185	Cdc73	371	DSS
Cdc73	194	Cdc73	205	DSS
Cdc73	205	Cdc73	212	DSS
Cdc73	205	Cdc73	371	DSS
Cdc73	212	Cdc73	371	DSS
Cdc73	218	Cdc73	205	DSS

Cdc73	236	Cdc73	185	DSS
Cdc73	236	Cdc73	205	DSS
Cdc73	236	Cdc73	212	DSS
Cdc73	236	Cdc73	218	DSS
Cdc73	236	Cdc73	226	DSS
Cdc73	236	Cdc73	304	DSS
Cdc73	263	Cdc73	130	DSS
Cdc73	263	Cdc73	205	DSS
Cdc73	263	Cdc73	366	DSS
Cdc73	263	Cdc73	368	DSS
Cdc73	263	Cdc73	371	DSS
Cdc73	282	Cdc73	185	DSS
Cdc73	282	Cdc73	205	DSS
Cdc73	282	Cdc73	371	DSS
Cdc73	282	Cdc73	385	DSS
Cdc73	304	Cdc73	130	DSS
Cdc73	304	Cdc73	185	DSS
Cdc73	304	Cdc73	205	DSS
Cdc73	304	Cdc73	212	DSS
Cdc73	304	Cdc73	218	DSS
Cdc73	304	Cdc73	226	DSS
Cdc73	304	Cdc73	371	DSS
Cdc73	361	Cdc73	371	DSS
Cdc73	366	Cdc73	17	DSS
Cdc73	366	Cdc73	371	DSS
Cdc73	368	Cdc73	366	DSS
Cdc73	371	Cdc73	13	DSS
Cdc73	371	Cdc73	205	DSS
Cdc73	371	Cdc73	212	DSS
Cdc73	371	Cdc73	366	DSS
Cdc73	371	Cdc73	368	DSS
Cdc73	385	Cdc73	17	DSS
Cdc73	385	Cdc73	130	DSS
Cdc73	385	Cdc73	185	DSS
Cdc73	385	Cdc73	205	DSS
Cdc73	385	Cdc73	212	DSS
Cdc73	385	Cdc73	218	DSS
Cdc73	385	Cdc73	263	DSS
Cdc73	385	Cdc73	366	DSS
Cdc73	385	Cdc73	368	DSS

Cdc73	385	Cdc73	371	DSS
Cdc73	13	Cdc73	141	EDC
Cdc73	24	Cdc73	185	EDC
Cdc73	103	Cdc73	205	EDC
Cdc73	113	Cdc73	13	EDC
Cdc73	113	Cdc73	185	EDC
Cdc73	113	Cdc73	205	EDC
Cdc73	113	Cdc73	385	EDC
Cdc73	119	Cdc73	126	EDC
Cdc73	119	Cdc73	132	EDC
Cdc73	119	Cdc73	141	EDC
Cdc73	124	Cdc73	141	EDC
Cdc73	126	Cdc73	185	EDC
Cdc73	126	Cdc73	199	EDC
Cdc73	126	Cdc73	212	EDC
Cdc73	130	Cdc73	141	EDC
Cdc73	130	Cdc73	176	EDC
Cdc73	134	Cdc73	368	EDC
Cdc73	136	Cdc73	126	EDC
Cdc73	136	Cdc73	132	EDC
Cdc73	136	Cdc73	176	EDC
Cdc73	141	Cdc73	212	EDC
Cdc73	141	Cdc73	371	EDC
Cdc73	154	Cdc73	185	EDC
Cdc73	154	Cdc73	385	EDC
Cdc73	156	Cdc73	185	EDC
Cdc73	156	Cdc73	205	EDC
Cdc73	156	Cdc73	218	EDC
Cdc73	156	Cdc73	385	EDC
Cdc73	158	Cdc73	185	EDC
Cdc73	158	Cdc73	205	EDC
Cdc73	158	Cdc73	218	EDC
Cdc73	166	Cdc73	205	EDC
Cdc73	169	Cdc73	141	EDC
Cdc73	176	Cdc73	205	EDC
Cdc73	176	Cdc73	212	EDC
Cdc73	185	Cdc73	126	EDC
Cdc73	185	Cdc73	134	EDC
Cdc73	185	Cdc73	141	EDC
Cdc73	185	Cdc73	176	EDC

Cdc73	194	Cdc73	141	EDC
Cdc73	197	Cdc73	205	EDC
Cdc73	199	Cdc73	176	EDC
Cdc73	205	Cdc73	134	EDC
Cdc73	205	Cdc73	141	EDC
Cdc73	218	Cdc73	132	EDC
Cdc73	218	Cdc73	141	EDC
Cdc73	226	Cdc73	141	EDC
Cdc73	236	Cdc73	261	EDC
Cdc73	261	Cdc73	236	EDC
Cdc73	263	Cdc73	16	EDC
Cdc73	263	Cdc73	24	EDC
Cdc73	348	Cdc73	366	EDC
Cdc73	371	Cdc73	132	EDC
Cdc73	385	Cdc73	126	EDC
Cdc73	385	Cdc73	129	EDC
Cdc73	385	Cdc73	132	EDC
Cdc73	385	Cdc73	134	EDC
Cdc73	385	Cdc73	141	EDC

Appendix D.2 Materials and methods

Appendix D.2.1 Auxin inducible degron strain construction and testing

Construction and testing of the various Spt6-AID strains discussed here were carried out as described in 3.2.1, 3.3.3-3.3.5, 3.3.7, and 3.3.8. Strains used in these experiments: KY3208, KY3343, KY3383, KY3291, KY3344, KY3352, and KY3447.

Appendix D.2.2 Western blotting

All protein extractions and western blots were performed as described in 3.2.4 and 3.2.5. Strains used for this experiment: KY1021, KY2239, KY2241, KY2243, KY2244, KY2271, KY2466.

Appendix D.2.3 Long gene assay

Rpb3, Spt6, and Spt16 occupancy was measured as described in 3.3.7 and 3.3.8 in a strain containing a "long gene" (*YLR454W*) driven by the *GAL1* promoter allowing the gene to be rapidly turned on or off by addition of galactose or glucose, respectively. Cells were grown to log phase (OD_{600} = 0.6-0.8) in YPGal (gene ON condition) before addition of glucose to turn off the *GAL1* promoter. I performed a full time-course collecting pre glucose, two, four, and eight-minute samples. Using chromatin from only the pre (galactose) and 8 min (glucose) time points I performed ChIP-qPCR on Rpb3, Spt16, and Spt6 at the *YLR454W* locus. Strains used in this experiment: KY3533, KY3535.

Appendix D.2.4 Spt6-AID ChIP-qPCR experiments

Experiments were carried out as described in 3.2.1, 3.2.7 and 3.2.8 using strain KY3291.

Appendix D.2.5 tRNA analysis

ChIP-seq data was aligned using HISAT2 (only allowing for one alignment per read) and converted to BAM using Samtools before using the featureCounts function from the Rsubread package to count the reads over tRNAs. Counts lists for each file were merged using the merge function in R and written out of R as a CSV file. The CSV file was opened in Excel and the reference adjusted reads per million method was used to scale the tRNA read counts based off an endogenous *K. lactus* spike in control (similar to what is described in 3.2.17 but using a value of one million instead of one hundred thousand in the scaling equation). Plots and statistics were performed in Prism8.

Appendix D.2.6 FRET

Förster resonance energy transfer (FRET) binding assay methods were developed based off previously published methodology (Hieb *et al.* 2012). FRET experiments were carried out in binding buffer (50 mM HEPES, 100 mM KOAc, 2 mM MgOAc, 10% glycerol, and 1 mM BME) using 200 nM of mClover tagged protein as the FRET donor and a titration of the mRuby2 tagged acceptor protein ranging from 1 nM to 30 μM per reaction. Stocks of donor and acceptor were prepared in 1.5 mL tubes at twice the concentration required for a FRET reaction and then 20 μL of each was mixed to generate a 40 μL reaction in a black 384-well plate. Reactions were pipetted up and down to mix acceptor and donor being careful to not create bubbles and allowed to incubate at room temperature for 10 minutes prior to measuring fluorescent intensities on a Cytation5 plate reader. Fcorr values were calculated as follows:

$$Ff = FRET\ value\ (ex470 - em600)\ for\ Donor + Acceptor\ Sample$$

$$Df = Donor\ value\ (470 - 515)\ for\ Donor + Acceptor\ Sample$$

$$Af = Acceptor\ value\ (559 - 600)\ for\ Donor + Acceptor\ sample$$

$$Xd = Fd / Dd$$

$$Xa = Fa / Aa$$

$$Fcorr = Ff - (Xd * Df) - (Xa * Af)$$

Binding curves were fit to data as described in 3.2.14.

Appendix D.2.7 Crosslinking

Crosslinking was performed as described in 3.2.12

Appendix D.2.8 Crosslinking network modelling

Crosslinking network models were generated using xView (https://xiview.org/xiNET_website/index.php)

Appendix D.2.9 Pulldowns

Binding was carried out by mixing 20 µL of washed magnetic nickel beads (NiNTA magnetic agarose beads, Qiagen) with binding buffer (25 mM Tris pH 7.5, 100 mM NaCl, 10% Glycerol, 5 mM imidazole) containing 6 µg of prey (10xHis-mRuby2, 10xHis-mRuby2-Cdc73, or

10xHis-mRuby2-Cdc73-C-domain) and 6 µg of bait (Spt6-239-1451, Spt6-236-1259, or Spt6-1247-1451) in a 1.5 mL tube and incubated for 1 hr at room temperature on an end over end roller. Two washes were carried out for 5-minutes each in 200 µL of 25 mM imidazole wash buffer (25 mM Tris pH 7.5, 100 mM NaCl, 10% Glycerol, 25 mM imidazole) and the first was sampled. Protein was eluted from beads using 50 µL of 500 mM imidazole elution buffer (25 mM Tris pH 7.5, 100 mM NaCl, 10% Glycerol, 500 mM imidazole). Protein samples were boiled in SDS PAGE loading dye before running a 12% SDS PAGE gel and staining with Coomassie blue.

Appendix D.2.10 Fluorescent gel shift assay

Experiments conducted as described in 3.2.14.

Appendix D.2.11 Published ChIP-seq data analysis

ChIP-seq datasets were procured from the SRA database and analyzed as described in 3.2.17.

Works Cited

Adelman K., and J. T. Lis, 2012 Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. 13: 720–731. https://doi.org/10.1038/nrg3293

Ahn S. H., M. Kim, and S. Buratowski, 2004 Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3′ End Processing. Mol. Cell 13: 67–76. https://doi.org/10.1016/S1097-2765(03)00492-1

Aho A. V., B. W. Kernighan, and P. J. Weinberger, 1979 Awk a pattern scanning and processing language. Softw. Pract. Exp. 9: 267–279. https://doi.org/10.1002/spe.4380090403

Akanuma T., S. Koshida, A. Kawamura, Y. Kishimoto, and S. Takada, 2007 Paf1 complex homologues are required for Notch-regulated transcription during somite segmentation. EMBO Rep. 8: 858–63. https://doi.org/10.1038/sj.embor.7401045

Amrich C. G., C. P. Davis, W. P. Rogal, M. K. Shirra, A. Heroux, *et al.*, 2012 Cdc73 subunit of Paf1 complex contains C-terminal Ras-like domain that promotes association of Paf1 complex with chromatin. J. Biol. Chem. 287: 10863–10875. https://doi.org/10.1074/jbc.M111.325647

Arndt K. M., and C. M. Kane, 2003 Running with RNA polymerase: Eukaryotic transcript elongation. Trends Genet. 19: 543–550. https://doi.org/10.1016/j.tig.2003.08.008

Arndt K. M., and D. Reines, 2014 Termination of transcription of short noncoding RNAs by RNA polymerase II. Annu. Rev. Biochem. 84: 150306093657004. https://doi.org/10.1146/annurev-biochem-060614-034457

Ashburner M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, *et al.*, 2000 Gene Ontology: tool for the unification of biology. Nat. Genet. 25: 25–29. https://doi.org/10.1038/75556

Audia J. E., and R. M. Campbell, 2017 Histone Modifications and Cancer. Cold spring Harb. Perspect. Biol. 8: 1–32.

Baejen C., J. Andreani, P. Torkler, S. Battaglia, B. Schwalb, *et al.*, 2017 Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. Mol. Cell 66. https://doi.org/10.1016/j.molcel.2017.02.009

Bakel H. van, K. Tsui, M. Gebbia, S. Mnaimneh, T. R. Hughes, *et al.*, 2013 A Compendium of Nucleosome and Transcript Profiles Reveals Determinants of Chromatin Architecture and Transcription. PLoS Genet. 9. https://doi.org/10.1371/journal.pgen.1003479

Baker S. P., and P. A. Grant, 2007 The SAGA continues: Expanding the cellular role of a transcriptional co-activator complex. Oncogene 26: 5329–5340.

https://doi.org/10.1038/sj.onc.1210603

Balakrishnan R., J. Park, K. Karra, B. C. Hitz, G. Binkley, *et al.*, 2012 YeastMine: An integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. Database 2012: 1–8. https://doi.org/10.1093/database/bar062

Baluapuri A., J. Hofstetter, N. Dudvarski Stankovic, T. Endres, P. Bhandare, *et al.*, 2019 MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. Mol. Cell 74: 674-687.e11. https://doi.org/10.1016/j.molcel.2019.02.031

Bao Y., and X. Shen, 2007 SnapShot: Chromatin Remodeling Complexes. Cell 129: 632.e1-632.e2. https://doi.org/10.1016/j.cell.2007.04.018

Baptista T., S. Grünberg, N. Minoungou, M. J. E. Koster, H. T. M. Timmers, *et al.*, 2017 SAGA Is a General Cofactor for RNA Polymerase II Transcription. Mol. Cell 68: 130-143.e5. https://doi.org/10.1016/j.molcel.2017.08.016

Barbaric S., T. Luckenbach, A. Schmid, D. Blaschke, W. Hörz, *et al.*, 2007 Redundancy of chromatin remodeling pathways for the induction of the yeast PHO5 promoter in vivo. J. Biol. Chem. 282: 27610–27621. https://doi.org/10.1074/jbc.M700623200

Barnett D. W., E. K. Garrison, A. R. Quinlan, M. P. Strömberg, and G. T. Marth, 2011 BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27: 1691–1692.

Bataille A. R., C. Jeronimo, P. É. Jacques, L. Laramée, M. È. Fortin, *et al.*, 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. https://doi.org/10.1016/j.molcel.2011.11.024

Batta K., Z. Zhang, K. Yen, D. B. Goffman, and B. F. Pugh, 2011 Genome-wide function of H2B ubiquitylation in promoter and genic regions. Genes Dev. 25: 2254–2265. https://doi.org/10.1101/gad.177238.111

Battaglia S., M. Lidschreiber, C. Baejen, P. Torkler, S. M. Vos, *et al.*, 2017 RNA-dependent chromatin association of transcription elongation factors and pol II CTD kinases. Elife 6: 1–26. https://doi.org/10.7554/eLife.25637

Becker D. M., and V. Lundblad, 1994 Introduction of DNA into Yeast Cells. Curr. Protoc. Mol. Biol. 27: 13–7.

Benjamini Y., and Y. Hochberg, 1995 Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. 57: 289–300.

Berger S. L., 2007 The complex language of chromatin regulation during transcription. Nature 447: 407–412. https://doi.org/10.1038/nature05915

Bernecky C., J. M. Plitzko, and P. Cramer, 2017 Structure of a transcribing RNA polymerase II-

DSIF complex reveals a multidentate DNA-RNA clamp. Nat. Struct. Mol. Biol. 24: 809–815. https://doi.org/10.1038/nsmb.3465

Betz J. L., M. Chang, T. M. Washburn, S. E. Porter, C. L. Mueller, *et al.*, 2002 Phenotypic analysis of Paf1/RNA polymerase II complex mutations reveals connections to cell cycle regulation, protein synthesis, and lipid and nucleic acid metabolism. Mol. Genet. Genomics 268: 272–285. https://doi.org/10.1007/s00438-002-0752-8

Birkeland S. R., N. Jin, A. C. Ozdemir, R. H. Lyons, L. S. Weisman, *et al.*, 2010 Discovery of Mutations in Saccharomyces cerevisiae by Pooled Linkage Analysis and Whole-Genome Sequencing. Genetics 186: 1127–1137. https://doi.org/10.1534/genetics.110.123232

Blankenberg D., G. Von Kuster, N. Coraor, G. Ananda, R. Lazarus, *et al.*, 2010 Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. 89: 10–19.

Boettiger C., 2015 An introduction to Docker for reproducible research. Oper. Syst. Rev. 49: 71–79. https://doi.org/10.1145/2723872.2723882

Bortvin A., and F. Winston, 1996 Evidence That Spt6p Controls Chromatin Structure by a Direct Interaction with Histones. Science (80-.). 272: 1473–1476. https://doi.org/10.1126/science.272.5267.1473

Botstein D., and G. R. Fink, 2011 Yeast: An Experimental Organism for 21st Century Biology. Genetics 189: 695–704. https://doi.org/10.1534/genetics.111.130765

Bourgeois C. F., Y. K. Kim, M. J. Churcher, M. J. West, and J. Karn, 2002 Spt5 Cooperates with Human Immunodeficiency Virus Type 1 Tat by Preventing Premature RNA Release at Terminator Sequences. Mol. Cell. Biol. 22: 1079–1093. https://doi.org/10.1128/mcb.22.4.1079-1093.2002

Brázda P., M. Krejčíková, A. Kasiliauskaite, E. Šmiřáková, T. Klumpler, *et al.*, 2020 Yeast Spt6 Reads Multiple Phosphorylation Patterns of RNA Polymerase II C-Terminal Domain In Vitro. J. Mol. Biol. https://doi.org/10.1016/j.jmb.2020.05.007

Buratowski S., 2003 The CTD code. Nat. Struct. Mol. Biol. 10: 679–680. https://doi.org/10.1038/nsb0903-679

Burugula B. B., C. Jeronimo, R. Pathak, J. W. Jones, F. Robert, *et al.*, 2014 Histone Deacetylases and Phosphorylated Polymerase II C-Terminal Domain Recruit Spt6 for Cotranscriptional Histone Reassembly. Mol. Cell. Biol. 34: 4115–4129. https://doi.org/10.1128/MCB.00695-14

Cao Q.-F., J. Yamamoto, T. Isobe, S. Tateno, Y. Murase, *et al.*, 2015 Characterization of the Human Transcription Elongation Factor Rtf1: Evidence for Nonoverlapping Functions of Rtf1 and the Paf1 Complex. Mol. Cell. Biol. 35: 3459–3470. https://doi.org/10.1128/MCB.00601-15

Carrozza M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson, *et al.*, 2005 Histone H3

methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123: 581–592. https://doi.org/10.1016/j.cell.2005.10.023

Castelnuovo M., S. Rahman, E. Guffanti, V. Infantino, F. Stutz, *et al.*, 2013 Bimodal expression of PHO84 is modulated by early termination of antisense transcription. Nat. Struct. Mol. Biol. 20: 851–858. https://doi.org/10.1038/nsmb.2598

Castelnuovo M., J. B. Zaugg, E. Guffanti, A. Maffioletti, J. Camblong, *et al.*, 2014 Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. Nucleic Acids Res. 42: 4348–4362. https://doi.org/10.1093/nar/gku100

Castelnuovo M., and F. Stutz, 2015 Role of chromatin, environmental changes and single cell heterogeneity in non-coding transcription and gene regulation. Curr. Opin. Cell Biol. 34: 16–22. https://doi.org/10.1016/j.ceb.2015.04.011

Chanfreau G., M. Buckle, and A. Jacquier, 2000 Recognition of a conserved class of RNA tetraloops by Saccharomyces cerevisiae RNase III. Proc. Natl. Acad. Sci. 97: 3142–3147. https://doi.org/10.1073/pnas.070043997

Chang M., D. French-Cornay, H. Y. Fan, H. Klein, C. L. Denis, *et al.*, 1999 A complex containing RNA polymerase II, Paf1p, Cdc73p, Hpr1p, and Ccr4p plays a role in protein kinase C signaling. Mol. Cell. Biol. 19: 1056–67.

Chaudhary K., S. Deb, N. Moniaux, M. P. Ponnusamy, and S. K. Batra, 2007 Human RNA polymerase II-associated factor complex: dysregulation in cancer. Oncogene 26: 7499–507. https://doi.org/10.1038/sj.onc.1210582

Chen F. X., A. R. Woodfin, A. Gardini, R. A. Rickels, S. A. Marshall, *et al.*, 2015 PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. Cell 162: 1003–1015. https://doi.org/10.1016/j.cell.2015.07.042

Chen F. X., E. R. Smith, and A. Shilatifard, 2018 Born to run: control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 19: 464–478. https://doi.org/10.1038/s41580-018-0010-5

Chen Z. L., J. M. Meng, Y. Cao, J. L. Yin, R. Q. Fang, *et al.*, 2019 A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. Nat. Commun. 10. https://doi.org/10.1038/s41467-019-11337-z

Cherry J. M., C. Adler, C. Ball, S. A. Chervitz, S. S. Dwight, *et al.*, 1998 SGD: Saccharomyces genome database. Nucleic Acids Res. 26: 73–79. https://doi.org/10.1093/nar/26.1.73

Cherry J. M., E. L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, *et al.*, 2012 Saccharomyces Genome Database: The genomics resource of budding yeast. Nucleic Acids Res. 40: 1–6. https://doi.org/10.1093/nar/gkr1029

Chervona Y., and M. Costa, 2012 Histone modifications and cancer : biomarkers of prognosis ? 2: 589–597.

Cheung V., G. Chua, N. N. Batada, C. R. Landry, S. W. Michnick, *et al.*, 2008 Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces cerevisiae genome. PLoS Biol. 6: 2550–2562. https://doi.org/10.1371/journal.pbio.0060277

Chin J. W., A. B. Martin, D. S. King, L. Wang, and P. G. Schultz, 2002 Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. Proc. Natl. Acad. Sci. U. S. A. 99: 11020–4. https://doi.org/10.1073/pnas.172226299

Chu Y., R. Simic, M. H. Warner, K. M. Arndt, and G. Prelich, 2007 Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. EMBO J. 26: 4646–4656. https://doi.org/10.1038/sj.emboj.7601887

Chun Y., Y. J. Joo, H. Suh, G. Batot, C. P. Hill, *et al.*, 2019 Selective Kinase Inhibition Shows That Bur1 (Cdk9) Phosphorylates the Rpb1 Linker In Vivo. Mol. Cell. Biol. 39: 1–13. https://doi.org/10.1128/MCB.00602-18

Churchman L. S., and J. S. Weissman, 2011 Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469: 368–73. https://doi.org/10.1038/nature09652

Cingolani P., A. Platts, M. Coon, T. Nguyen, L. Wang, *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 6: 80–92.

Clapier C. R., J. Iwasa, B. R. Cairns, and C. L. Peterson, 2017 Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat. Rev. Mol. Cell Biol. 18: 407–422. https://doi.org/10.1038/nrm.2017.26

Clark-Adams C., and F. Winston, 1987 The SPT6 gene is essential for growth and is required for delta-mediated transcription in Saccharomyces cerevisiae. Mol. Cell. Biol. 7: 679–686.

Close D., S. J. Johnson, M. A. Sdano, S. M. McDonald, H. Robinson, *et al.*, 2011 Crystal structures of the S. cerevisiae Spt6 core and C-terminal tandem SH2 domain. J. Mol. Biol. 408: 697–713. https://doi.org/10.1016/j.jmb.2011.03.002

Collart M., and S. Oliviero, 1993 Preparation of Yeast RNA. Curr. Protoc. Mol. Biol. 13: 13.12. https://doi.org/10.1002/0471142727.mb1312s23

Corden J. L., 2013 RNA Polymerase II C-Terminal Domain: Tethering Transcription to Transcript and Template. Chem. Rev. 113: 8423–8455. https://doi.org/10.1021/cr400158h

Cortazar M. A., R. M. Sheridan, B. Erickson, N. Fong, K. Glover-cutter, *et al.*, 2019 Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a Article Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a ‘“ Sitting Duck Torpedo ”’ Mechanism. Mol. Cell 1–13. https://doi.org/10.1016/j.molcel.2019.09.031

Courel M., S. Lallet, J.-M. Camadro, and P.-L. Blaiseau, 2005 Direct activation of genes involved

in intracellular iron use by the yeast iron-responsive transcription factor Aft2 without its paralog Aft1. Mol. Cell. Biol. 25: 6760–6771. https://doi.org/10.1128/MCB.25.15.6760-6771.2005

Cox J. S., R. E. Chapman, and P. Waltert, 1997 The Unfolded Protein Response Coordinates the Production of Endoplasmic Reticulum Protein and Endoplasmic Reticulum Membrane. 8: 1805–1814.

Cramer P., K.-J. J. Armache, S. Baumli, S. Benkert, F. Brueckner, *et al.*, 2008 Structure of eukaryotic RNA polymerases. Annu Rev Biophys 37: 337–352. https://doi.org/10.1146/annurev.biophys.37.032807.130008

Crisucci E. M., and K. M. Arndt, 2011a The Roles of the Paf1 Complex and Associated Histone Modifications in Regulating Gene Expression. Genet. Res. Int. 2011: 1–15. https://doi.org/10.4061/2011/707641

Crisucci E. M., and K. M. Arndt, 2011b The Paf1 Complex Represses ARG1 Transcription in Saccharomyces cerevisiae by Promoting Histone Modifications. Eukaryot. Cell 10: 712–723. https://doi.org/10.1128/EC.05013-11

Crisucci E. M., and K. M. Arndt, 2012 Paf1 Restricts Gcn4 Occupancy and Antisense Transcription at the ARG1 Promoter. Mol. Cell. Biol. 32: 1150–1163. https://doi.org/10.1128/MCB.06262-11

Cucinotta C. E., A. E. Hildreth, B. M. McShane, M. K. Shirra, and K. M. Arndt, 2019 The nucleosome acidic patch directly interacts with subunits of the Paf1 and FACT complexes and controls chromatin architecture in vivo. Nucleic Acids Res. 47: 8410–8423. https://doi.org/10.1093/nar/gkz549

Cyert M. S., and C. C. Philpott, 2013 Regulation of cation balance in Saccharomyces cerevisiae. Genetics 193: 677–713. https://doi.org/10.1534/genetics.112.147207

Danecek P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

David L., W. Huber, M. Granovskaia, J. Toedling, C. J. Palm, *et al.*, 2006 A high-resolution map of transcription in the yeast genome. Proc. Natl. Acad. Sci. U. S. A. 103: 5320–5. https://doi.org/10.1073/pnas.0601091103

Deng P., Y. Zhou, J. Jiang, H. Li, W. Tian, *et al.*, 2018 Transcriptional elongation factor Paf1 core complex adopts a spirally wrapped solenoidal topology. Proc. Natl. Acad. Sci. 115: 9998–10003. https://doi.org/10.1073/pnas.1812256115

Dengl S., A. Mayer, M. Sun, and P. Cramer, 2009 Structure and in Vivo Requirement of the Yeast Spt6 SH2 Domain. J. Mol. Biol. 389: 211–225. https://doi.org/10.1016/j.jmb.2009.04.016

Dermody J. L., and S. Buratowski, 2010 Leo1 subunit of the yeast Paf1 complex binds RNA and contributes to complex recruitment. J. Biol. Chem. 285: 33671–33679.

https://doi.org/10.1074/jbc.M110.140764

Dijk E. L. van, C. L. Chen, Y. D'Aubenton-Carafa, S. Gourvennec, M. Kwapisz, *et al.*, 2011 XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature 475: 114–7. https://doi.org/10.1038/nature10118

Ding L., M. Paszkowski-Rogacz, A. Nitzsche, M. M. Slabicki, A. K. Heninger, *et al.*, 2009 A Genome-Scale RNAi Screen for Oct4 Modulators Defines a Role of the Paf1 Complex for Embryonic Stem Cell Identity. Cell Stem Cell 4: 403–415. https://doi.org/10.1016/j.stem.2009.03.009

Dix D. R., J. T. Bridgham, M. A. Broderius, C. A. Byersdorfer, and D. J. Eide, 1994 The FET4 gene encodes the low affinity Fe(II) transport protein of Saccharomyces cerevisiae. J. Biol. Chem. 269: 26092–26099.

Donczew R., L. Warfield, D. Pacheco, A. Erijman, and S. Hahn, 2020 Two roles for the yeast transcription coactivator SAGA and a set of genes redundantly regulated by TFIID and SAGA. Elife 9: 1–27. https://doi.org/10.7554/eLife.50109

Doris S. M., J. Chuang, O. Viktorovskaya, M. Murawska, D. Spatt, *et al.*, 2018 Spt6 Is Required for the Fidelity of Promoter Selection. Mol. Cell 72: 687-699.e6. https://doi.org/10.1016/j.molcel.2018.09.005

Dover J., J. Schneider, M. A. Tawiah-Boateng, A. Wood, K. Dean, *et al.*, 2002 Methylation of Histone H3 by COMPASS Requires Ubiquitination of Histone H2B by Rad6. J. Biol. Chem. 277: 28368–28371. https://doi.org/10.1074/jbc.C200348200

Dronamraju R., and B. D. Strahl, 2014 A feed forward circuit comprising Spt6, Ctk1 and PAF regulates Pol II CTD phosphorylation and transcription elongation. Nucleic Acids Res. 42: 870–881. https://doi.org/10.1093/nar/gkt1003

Dronamraju R., A. J. Hepperla, Y. Shibata, A. T. Adams, T. Magnuson, *et al.*, 2018a Spt6 Association with RNA Polymerase II Directs mRNA Turnover During Transcription. Mol. Cell 70: 1054-1066.e4. https://doi.org/10.1016/j.molcel.2018.05.020

Dronamraju R., J. L. Kerschner, S. A. Peck, A. J. Hepperla, A. T. Adams, *et al.*, 2018b Casein Kinase II Phosphorylation of Spt6 Enforces Transcriptional Fidelity by Maintaining Spn1-Spt6 Interaction. Cell Rep. 25: 3476-3489.e5. https://doi.org/10.1016/j.celrep.2018.11.089

Dronamraju R., J. L. Kerschner, S. A. Peck, A. J. Hepperla, A. T. Adams, *et al.*, 2018c Casein Kinase II Phosphorylation of Spt6 Enforces Transcriptional Fidelity by Maintaining Spn1-Spt6 Interaction. Cell Rep. 25: 3476-3489.e5. https://doi.org/10.1016/j.celrep.2018.11.089

Duina A. A., 2011 Histone Chaperones Spt6 and FACT: Similarities and Differences in Modes of Action at Transcribed Genes. Genet. Res. Int. 2011: doi:10.4061/2011/625210. https://doi.org/10.4061/2011/625210

Ehara H., T. Kujirai, Y. Fujino, M. Shirouzu, H. Kurumizaka, *et al.*, 2019 Structural insight into

nucleosome transcription by RNA polymerase II with elongation factors. Science (80-.). 363: 744–747. https://doi.org/10.1126/science.aav8912

Eick D., and M. Geyer, 2013 The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. Chem. Rev. 113: 8456–8490. https://doi.org/10.1021/cr400071f

Ellison M. A., A. R. Lederer, M. H. Warner, T. N. Mavrich, E. A. Raupach, *et al.*, 2019 The Paf1 Complex Broadly Impacts the Transcriptome of Saccharomyces cerevisiae. Genetics 212: 711–728. https://doi.org/10.1534/genetics.119.302262

Endoh M., W. Zhu, J. Hasegawa, H. Watanabe, D.-K. Kim, *et al.*, 2004 Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. Mol. Cell. Biol. 24: 3324–36. https://doi.org/10.1128/MCB.24.8.3324

Engel S. R., F. S. Dietrich, D. G. Fisk, G. Binkley, R. Balakrishnan, *et al.*, 2014 The reference genome sequence of Saccharomyces cerevisiae: then and now. G3 (Bethesda). 4: 389–98. https://doi.org/10.1534/g3.113.008995

Faghihi M. A., and C. Wahlestedt, 2009 Regulatory roles of natural antisense transcripts. Nat. Rev. Mol. Cell Biol. 10: 637–43. https://doi.org/10.1038/nrm2738

Fernandez-Martinez J., S. J. Kim, Y. Shi, P. Upla, R. Pellarin, *et al.*, 2016 Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. Cell 167: 1215-1228.e25. https://doi.org/10.1016/j.cell.2016.10.028

Fischl H., F. S. Howe, A. Furger, and J. Mellor, 2017 Paf1 Has Distinct Roles in Transcription Elongation and Differential Transcript Fate. Mol. Cell 65: 685-698.e8. https://doi.org/10.1016/j.molcel.2017.01.006

Fox M. J., H. Gao, W. R. Smith-Kinnaman, Y. Liu, and A. L. Mosley, 2015 The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway. PLoS Genet. 11: e1004999. https://doi.org/10.1371/journal.pgen.1004999

Fuchs S. M., J. Mellor, H. Histone, S. M. Fuchs, E. Duro, *et al.*, 2008 Roles for Ctk1 and Spt6 in Regulating the Different Methylation States of Histone H3 Roles for Ctk1 and Spt6 in Regulating the Different Methylation. https://doi.org/10.1128/MCB.00001-08

Gentleman R., V. Carey, D. Bates, B. Bolstad, M. Dettling, *et al.*, 2004 Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5: R80. https://doi.org/10.1186/gb-2004-5-10-r80

Giardine B., C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski, *et al.*, 2005 Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 15: 1451–1455.

Goldgof G. M., J. D. Durrant, S. Ottilie, E. Vigil, K. E. Allen, *et al.*, 2016 Comparative chemical genomics reveal that the spiroindolone antimalarial KAE609 (Cipargamin) is a P-type ATPase inhibitor. Sci. Rep. 6: 27806. https://doi.org/10.1038/srep27806

Gopalakrishnan R., S. K. Marr, R. E. Kingston, and F. Winston, 2019 A conserved genetic interaction between Spt6 and Set2 regulates H3K36 methylation. Nucleic Acids Res. 47: 3888–3903. https://doi.org/10.1093/nar/gkz119

Gopalakrishnan R., and F. Winston, 2019 Whole-Genome Sequencing of Yeast Cells. Curr. Protoc. Mol. Biol. 128: e103. https://doi.org/10.1002/cpmb.103

Govind C. K., H. Qiu, D. S. Ginsburg, C. Ruan, K. Hofmeyer, *et al.*, 2010 Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes. Mol. Cell 39: 234–246. https://doi.org/10.1016/j.molcel.2010.07.003

Greer E. L., and Y. Shi, 2012 Histone methylation : a dynamic mark in health , disease and inheritance. 13. https://doi.org/10.1038/nrg3173

Guenther M. G., R. G. Jenner, B. Chevalier, T. Nakamura, C. M. Croce, *et al.*, 2005 Global and Hox-specific roles for the MLL1 methyltransferase. Proc. Natl. Acad. Sci. 102: 8603–8608. https://doi.org/10.1073/pnas.0503072102

Guo M., F. Xu, J. Yamada, T. Egelhofer, Y. Gao, *et al.*, 2008 Core Structure of the Yeast Spt4-Spt5 Complex: A Conserved Module for Regulation of Transcription Elongation. Structure 16: 1649–1658. https://doi.org/10.1016/j.str.2008.08.013

Hammond C. M., C. B. Strømme, H. Huang, D. J. Patel, and A. Groth, 2017 Histone chaperone networks shaping chromatin function. Nat. Rev. Mol. Cell Biol. 18: 141–158. https://doi.org/10.1038/nrm.2016.159

Hamza A., E. Tammpere, M. Kofoed, C. Keong, J. Chiang, *et al.*, 2015 Complementation of Yeast Genes with Human Genes as an Experimental Platform for Functional Testing of Human Genetic Variants. Genetics 201: 1263–1274. https://doi.org/10.1534/genetics.115.181099

Hamza A., M. R. M. Driessen, E. Tammpere, N. J. O'Neil, and P. Hieter, 2020 Cross-Species Complementation of Nonessential Yeast Genes Establishes Platforms for Testing Inhibitors of Human Proteins. Genetics 214: 735–747. https://doi.org/10.1534/genetics.119.302971

Harlen K. M., and L. S. Churchman, 2017a The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat. Rev. Mol. Cell Biol. https://doi.org/10.1038/nrm.2017.10

Harlen K. M., and L. S. Churchman, 2017b Subgenic Pol II interactomes identify region-specific transcription elongation regulators. Mol. Syst. Biol. 13: 900. https://doi.org/10.15252/msb.20167279

Hartzog G. A., T. Wada, H. Handa, and F. Winston, 1998 Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev. 12: 357–369. https://doi.org/10.1101/gad.12.3.357

Hartzog G. A., and J. Fu, 2013 The Spt4-Spt5 complex: A multi-faceted regulator of transcription

elongation. Biochim. Biophys. Acta - Gene Regul. Mech. 1829: 105–115. https://doi.org/10.1016/j.bbagrm.2012.08.007

Hieb A. R., S. D'Arcy, M. A. Kramer, A. E. White, and K. Luger, 2012 Fluorescence strategies for high-throughput quantification of protein interactions. Nucleic Acids Res. 40: 1–13. https://doi.org/10.1093/nar/gkr1045

Hou L., Y. Wang, Y. Liu, N. Zhang, I. Shamovsky, *et al.*, 2019 Paf1C regulates RNA polymerase II progression by modulating elongation rate. Proc. Natl. Acad. Sci. U. S. A. 116: 14583–14592. https://doi.org/10.1073/pnas.1904324116

Hsu P. L., F. Yang, W. Smith-Kinnaman, W. Yang, J. E. Song, *et al.*, 2014 Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr1 and Ser5 on the RNA polymerase II CTD. J. Mol. Biol. 426: 2970–2981. https://doi.org/10.1016/j.jmb.2014.06.010

Huang H., B. R. Sabari, B. A. Garcia, C. David Allis, and Y. Zhao, 2014 SnapShot: Histone modifications. Cell 159: 458-458.e1. https://doi.org/10.1016/j.cell.2014.09.037

Huber W., A. von Heydebreck, H. Sültmann, A. Poustka, and M. Vingron, 2002 Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18 Suppl 1: S96-104. https://doi.org/10.1093/bioinformatics/18.suppl_1.S96

Huber W., J. Toedling, and L. M. Steinmetz, 2006 Transcript mapping with high-density oligonucleotide tiling arrays. Bioinformatics 22: 1963–1970. https://doi.org/10.1093/bioinformatics/btl289

Iida N., F. Yamao, Y. Nakamura, and T. Iida, 2014 Mudi, a web tool for identifying mutations by bioinformatics analysis of whole-genome sequence. Genes to Cells 19: 517–527. https://doi.org/10.1111/gtc.12151

Irvine D. V., D. B. Goto, M. W. Vaughn, Y. Nakaseko, W. R. McCombie, *et al.*, 2009 Mapping epigenetic mutations in fission yeast using whole-genome next-generation sequencing. Genome Res. 19: 1077–1083. https://doi.org/10.1101/gr.089318.108

Jaehning J. A., 2010 The Paf1 complex: Platform or player in RNA polymerase II transcription? Biochim. Biophys. Acta - Gene Regul. Mech. 1799: 379–388. https://doi.org/10.1016/j.bbagrm.2010.01.001

Jensen M. M., M. S. Christensen, B. Bonven, and T. H. Jensen, 2008 Requirements for chromatin reassembly during transcriptional downregulation of a heat shock gene in Saccharomyces cerevisiae. FEBS J. 275: 2956–2964. https://doi.org/10.1111/j.1742-4658.2008.06451.x

Jeronimo C., A. R. Bataille, and F. Robert, 2013 The writers, readers, and functions of the RNA polymerase II C-terminal domain code. Chem. Rev. 113: 8491–8522. https://doi.org/10.1021/cr4001397

Jeronimo C., C. Poitras, and F. Robert, 2019 Histone Recycling by FACT and Spt6 during

Transcription Prevents the Scrambling of Histone Modifications. Cell Rep. 28: 1206-1218.e8. https://doi.org/10.1016/j.celrep.2019.06.097

Jo H., and G. Koh, 2015 Faster single-end alignment generation utilizing multi-thread for BWA. Biomed. Mater. Eng. 26 Suppl 1: S1791-6. https://doi.org/10.3233/BME-151480

Joshi A. A., and K. Struhl, 2005 Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to pol II elongation. Mol. Cell 20: 971–978. https://doi.org/10.1016/j.molcel.2005.11.021

Juneau K., C. Palm, M. Miranda, and R. W. Davis, 2007 High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. Proc. Natl. Acad. Sci. 104: 1522–1527. https://doi.org/10.1073/pnas.0610354104

Kachroo A. H., J. M. Laurent, C. M. Yellman, A. G. Meyer, C. O. Wilke, *et al.*, 2015 Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science 348: 921–5. https://doi.org/10.1126/science.aaa0769

Kaplan C. D., J. R. Morris, C. T. Wu, and F. Winston, 2000 Spt5 and Spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. Genes Dev. 14: 2623–2634. https://doi.org/10.1101/gad.831900

Kaplan C. D., 2003 Transcription Elongation Factors Repress Transcription Initiation from Cryptic Sites. Science (80-.). 301: 1096–1099. https://doi.org/10.1126/science.1087374

Kaplan C. D., M. J. Holland, and F. Winston, 2005 Interaction between Transcription Elongation Factors and mRNA 3′-End Formation at the Saccharomyces cerevisiae GAL10-GAL7 Locus. J. Biol. Chem. 280: 913–922. https://doi.org/10.1074/jbc.M411108200

Kaplan C. D., and J. Kaplan, 2009 Iron acquisition and transcriptional regulation. Chem. Rev. 109: 4536–4552. https://doi.org/10.1021/cr9001676

Karmakar S., P. Dey, A. P. Vaz, S. R. Bhaumik, M. P. Ponnusamy, *et al.*, 2018 PD2/PAF1 at the Crossroads of the Cancer Network. Cancer Res. 78: 313–319. https://doi.org/10.1158/0008-5472.CAN-17-2175

Keegan B. R., J. L. Feldman, D. H. Lee, D. S. Koos, R. K. Ho, *et al.*, 2002 The elongation factors Pandora/Spt6 and Foggy/Spt5 promote transcription in the zebrafish embryo. Development 129: 1623–32.

Kent W., C. Sugnet, T. Furey, and K. Roskin, 1976 The Human Genome Browser at UCSC W. [available from: http://genome.ucsc.edu/]. J. Med. Chem. 19: 1228–31. https://doi.org/10.1101/gr.229102.

Keogh M. C., S. K. Kurdistani, S. A. Morris, S. H. Ahn, V. Podolny, *et al.*, 2005 Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell 123: 593–605. https://doi.org/10.1016/j.cell.2005.10.025

Kim J., M. Guermah, and R. G. Roeder, 2010 The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. Cell 140: 491–503. https://doi.org/10.1016/j.cell.2009.12.050

Kim D., B. Langmead, and S. L. Salzberg, 2015 HISAT: A fast spliced aligner with low memory requirements. Nat. Methods 12: 357–360. https://doi.org/10.1038/nmeth.3317

Kim J. H., B. B. Lee, Y. M. Oh, C. Zhu, L. M. Steinmetz, *et al.*, 2016 Modulation of mRNA and lncRNA expression dynamics by the Set2–Rpd3S pathway. Nat. Commun. 7: 13534. https://doi.org/10.1038/ncomms13534

Kim Y. M., J. B. Poline, and G. Dumas, 2018 Experimenting with reproducibility: A case study of robustness in bioinformatics. Gigascience 7: 1–8. https://doi.org/10.1093/gigascience/giy077

Klemm S. L., Z. Shipony, and W. J. Greenleaf, 2019 Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. 20: 207–220. https://doi.org/10.1038/s41576-018-0089-8

Komarnitsky P., E. J. Cho, and S. Buratowski, 2000 Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. 14: 2452–2460. https://doi.org/10.1101/gad.824700

Korber P., and S. Barbaric, 2014 The yeast PHO5 promoter: From single locus to systems biology of a paradigm for gene regulation through chromatin. Nucleic Acids Res. 42: 10888–10902. https://doi.org/10.1093/nar/gku784

Kosinski J., A. von Appen, A. Ori, K. Karius, C. W. Müller, *et al.*, 2015 Xlink analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. J. Struct. Biol. 189: 177–183. https://doi.org/10.1016/j.jsb.2015.01.014

Köster J., and S. Rahmann, 2012 Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28: 2520–2522.

Krogan N. J., M. Kim, S. H. Ahn, G. Zhong, M. S. Kobor, *et al.*, 2002 RNA Polymerase II Elongation Factors of Saccharomyces cerevisiae : a Targeted Proteomics Approach. Society 22: 6979–6992. https://doi.org/10.1128/MCB.22.20.6979

Krogan N. J., M. Kim, A. Tong, A. Golshani, G. Cagney, *et al.*, 2003 Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Mol. Cell. Biol. 23: 4207–4218. https://doi.org/10.1128/MCB.23.12.4207-4218.2003

Kuehner J. N., E. L. Pearson, and C. Moore, 2011 Unravelling the means to an end: RNA polymerase II transcription termination. Nat. Rev. Mol. Cell Biol. 12: 283–294. https://doi.org/10.1038/nrm3098

Kuras L., and K. Struhl, 1999 Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. Nature 399: 609–613. https://doi.org/10.1038/21239

Kurtz S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, *et al.*, 2004 Versatile and open software for comparing large genomes. Genome Biol. 5: R12. https://doi.org/10.1186/gb-2004-5-2-r12

LaCava J., J. Houseley, C. Saveanu, E. Petfalski, E. Thompson, *et al.*, 2005 RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell 121: 713–724. https://doi.org/10.1016/j.cell.2005.04.029

Lai W. K. M., and B. F. Pugh, 2017 Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat. Rev. Mol. Cell Biol. 18: 548–562. https://doi.org/10.1038/nrm.2017.47

Langenbacher A. D., C. T. Nguyen, A. M. Cavanaugh, J. Huang, F. Lu, *et al.*, 2011 The PAF1 complex differentially regulates cardiomyocyte specification. Dev. Biol. 353: 19–28. https://doi.org/10.1016/j.ydbio.2011.02.011

Langmead B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359. https://doi.org/10.1038/nmeth.1923

Lee A., A. K. Henras, and G. Chanfreau, 2005 Multiple RNA surveillance pathways limit aberrant expression of iron uptake mRNAs and prevent iron toxicity in S. cerevisiae. Mol. Cell 19: 39–51. https://doi.org/10.1016/j.molcel.2005.05.021

Li Z., 2006 Flexible Structural Neighborhood--a database of protein structural similarities and alignments. Nucleic Acids Res. 34: D277–D280. https://doi.org/10.1093/nar/gkj124

Li B., M. Carey, and J. L. Workman, 2007 The Role of Chromatin during Transcription. Cell 128: 707–719. https://doi.org/10.1016/j.cell.2007.01.015

Li H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Liu Y., L. Warfield, C. Zhang, J. Luo, J. Allen, *et al.*, 2009 Phosphorylation of the Transcription Elongation Factor Spt5 by Yeast Bur1 Kinase Stimulates Recruitment of the PAF Complex. Mol. Cell. Biol. 29: 4852–4863. https://doi.org/10.1128/MCB.00609-09

Liu L., N. M. Oliveira, K. M. Cheney, C. Pade, H. Dreja, *et al.*, 2011 A whole genome screen for HIV restriction factors. Retrovirology 8: 94. https://doi.org/10.1186/1742-4690-8-94

Liu X., W. L. Kraus, and X. Bai, 2015 Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. Trends Biochem. Sci. 40: 516–525. https://doi.org/10.1016/j.tibs.2015.07.003

Liu Y., K. Zhou, N. Zhang, H. Wei, Y. Z. Tan, *et al.*, 2020 FACT caught in the act of manipulating the nucleosome. Nature 577: 426–431. https://doi.org/10.1038/s41586-019-1820-0

Lu X., X. Zhu, Y. Li, M. Liu, B. Yu, *et al.*, 2016 Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. Nucleic Acids Res. 44: 6853–6867. https://doi.org/10.1093/nar/gkw571

Lundblad V., 1989 Cloning Yeast Genes by Complementation. Curr. Protoc. Mol. Biol. 5: 5–8. https://doi.org/10.1002/0471142727.mb1308s05

Lundblad V., G. Hartzog, and Z. Moqtaderi, 2001 Manipulation of cloned yeast DNA. Curr. Protoc. Mol. Biol. 13: 13.10. https://doi.org/10.1002/0471142727.mb1310s39

Marazzi I., J. S. Y. Ho, J. Kim, B. Manicassamy, S. Dewell, *et al.*, 2012 Suppression of the antiviral response by an influenza histone mimic. Nature 483: 428–433. https://doi.org/10.1038/nature10892

Mardis E. R., 2013 Next-Generation Sequencing Platforms. Annu. Rev. Anal. Chem 6: 287–303. https://doi.org/10.1146/annurev-anchem-062012-092628

Martens J. A., L. Laprade, and F. Winston, 2004 Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature 429: 571–574. https://doi.org/10.1038/nature02538

Matera A. G., R. M. Terns, and M. P. Terns, 2007 Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol 8: 209–220. https://doi.org/10.1038/nrm2124

Mayekar M. K., R. G. Gardner, and K. M. Arndt, 2013 The recruitment of the Saccharomyces cerevisiae Paf1 complex to active genes requires a domain of Rtf1 that directly interacts with the Spt4-Spt5 complex. Mol. Cell. Biol. 33: 3259–73. https://doi.org/10.1128/MCB.00270-13

Mayer A., M. Lidschreiber, M. Siebert, K. Leike, J. Söding, *et al.*, 2010 Uniform transitions of the general RNA polymerase II transcription complex. Nat. Struct. Mol. Biol. 17: 1272–1278. https://doi.org/10.1038/nsmb.1903

McDonald S. M., D. Close, H. Xin, T. Formosa, and C. P. Hill, 2010 Structure and Biological Importance of the Spn1-Spt6 Interaction, and Its Regulatory Role in Nucleosome Binding. Mol. Cell 40: 725–735. https://doi.org/10.1016/j.molcel.2010.11.014

McGinty R. K., and S. Tan, 2015 Nucleosome structure and function. Chem. Rev. 115: 2255–2273. https://doi.org/10.1021/cr500373h

McKenna A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, *et al.*, 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

Meyer P. A., S. Li, M. Zhang, K. Yamada, Y. Takagi, *et al.*, 2015 Structures and Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5. Mol. Cell. Biol. 35: 3354–3369. https://doi.org/10.1128/mcb.00520-15

Minevich G., D. S. Park, D. Blankenberg, R. J. Poole, and O. Hobert, 2012 CloudMap: A Cloud-Based Pipeline for Analysis of Mutant Genome Sequences. Genetics 192: 1249–1269. https://doi.org/10.1534/genetics.112.144204

Mueller C. L., and J. A. Jaehning, 2002 Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. Mol. Cell. Biol. 22: 1971–80. https://doi.org/10.1128/MCB.22.7.1971

Mueller C. L., S. E. Porter, M. G. Hoffman, and J. A. Jaehning, 2004 The Paf1 complex has functions independent of actively transcribing RNA polymerase II. Mol. Cell 14: 447–456. https://doi.org/10.1016/S1097-2765(04)00257-6

Muntean A. G., J. Tan, K. Sitwala, Y. Huang, J. Bronstein, *et al.*, 2010 The PAF Complex Synergizes with MLL Fusion Proteins at HOX Loci to Promote Leukemogenesis. Cancer Cell 17: 609–621. https://doi.org/10.1016/j.ccr.2010.04.012

Murray S. C., S. Haenni, F. S. Howe, H. Fischl, K. Chocian, *et al.*, 2015 Sense and antisense transcription are associated with distinct chromatin architectures across genes. Nucleic Acids Res. 43: 7823–7837. https://doi.org/10.1093/nar/gkv666

Nakanishi S., S. L. Jung, K. E. Gardner, J. M. Gardner, Y. H. Takahashi, *et al.*, 2009 Histone H2BK123 monoubiquitination is the critical determinant for H3K4 and H3K79 trimethylation by COMPASS and Dot1. J. Cell Biol. 186: 371–377. https://doi.org/10.1083/jcb.200906005

Neil H., C. Malabat, Y. D'Aubenton-Carafa, Z. Xu, L. M. Steinmetz, *et al.*, 2009 Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457: 1038–1042. https://doi.org/10.1038/nature07747

Ng H. H., S. Dole, and K. Struhl, 2003a The Rtf1 Component of the Paf1 Transcriptional Elongation Complex Is Required for Ubiquitination of Histone H2B. J. Biol. Chem. 278: 33625–33628. https://doi.org/10.1074/jbc.C300270200

Ng H. H., F. Robert, R. A. Young, and K. Struhl, 2003b Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell 11: 709–719. https://doi.org/10.1016/S1097-2765(03)00092-3

Nguyen C. T., A. Langenbacher, M. Hsieh, and J. N. O. Chen, 2010 The PAF1 complex component Leo1 is essential for cardiac and neural crest development in zebrafish. Dev. Biol. 341: 167–175. https://doi.org/10.1016/j.ydbio.2010.02.020

Nishimura K., T. Fukagawa, H. Takisawa, T. Kakimoto, and M. Kanemaki, 2009 An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat. Methods 53: 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Nordick K., M. G. Hoffman, J. L. Betz, and J. A. Jaehning, 2008 Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II. Eukaryot. Cell 7: 1158–1167. https://doi.org/10.1128/EC.00434-07

Ohdate H., C. R. Lim, T. Kokubo, K. Matsubara, Y. Kimata, *et al.*, 2003 Impairment of the DNA Binding Activity of the TATA-binding Protein Renders the Transcriptional Function of Rvb2p / Tih2p , the Yeast RuvB-like Protein , Essential for Cell Growth. 278: 14647–14656. https://doi.org/10.1074/jbc.M213220200

Orlando D. A., M. W. Chen, V. E. Brown, S. Solanki, Y. J. Choi, *et al.*, 2014 Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Rep. 9: 1163–1170. https://doi.org/10.1016/j.celrep.2014.10.018

Oss S. B. Van, M. K. Shirra, A. R. Bataille, A. D. Wier, K. Yen, *et al.*, 2016 The Histone Modification Domain of Paf1 Complex Subunit Rtf1 Directly Stimulates H2B Ubiquitylation through an Interaction with Rad6. Mol. Cell 64: 815–825. https://doi.org/10.1016/j.molcel.2016.10.008

Oss S. B. Van, C. E. Cucinotta, and K. M. Arndt, 2017 Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation. Trends Biochem. Sci. 42: 788–798. https://doi.org/10.1016/j.tibs.2017.08.003

Ottilie S., G. M. Goldgof, C. M. Calvet, G. K. Jennings, G. Lamonte, *et al.*, 2017 Rapid Chagas Disease Drug Target Discovery Using Directed Evolution in Drug-Sensitive Yeast. ACS Chem. Biol. 12: 422–434. https://doi.org/10.1021/acschembio.6b01037

Pelechano V., and L. M. Steinmetz, 2013 Gene regulation by antisense transcription. Nat. Rev. Genet. 14: 880–93. https://doi.org/10.1038/nrg3594

Penheiter K. L., T. M. Washburn, S. E. Porter, M. G. Hoffman, and J. A. Jaehning, 2005 A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. Mol. Cell 20: 213–223. https://doi.org/10.1016/j.molcel.2005.08.023

Perocchi F., Z. Xu, S. Clauder-Münster, and L. M. Steinmetz, 2007 Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. Nucleic Acids Res. 35. https://doi.org/10.1093/nar/gkm683

Pettersen E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, *et al.*, 2004 UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. 25: 1605–1612. https://doi.org/10.1002/jcc.20084

Pfaffl M. W., 2001 A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29: e45. https://doi.org/10.1093/nar/29.9.e45

Piro A. S., M. K. Mayekar, M. H. Warner, C. P. Davis, and K. M. Arndt, 2012 Small region of Rtf1 protein can substitute for complete Paf1 complex in facilitating global histone H2B

ubiquitylation in yeast. Proc. Natl. Acad. Sci. U. S. A. 109: 10837–42. https://doi.org/10.1073/pnas.1116994109

Plaschka C., K. Nozawa, and P. Cramer, 2016 Mediator Architecture and RNA Polymerase II Interaction. J. Mol. Biol. 428: 2569–2574. https://doi.org/10.1016/j.jmb.2016.01.028

Pokholok D. K., N. M. Hannett, and R. A. Young, 2002 Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Mol. Cell 9: 799–809. https://doi.org/10.1016/S1097-2765(02)00502-6

Ponnusamy M. P., S. Deb, P. Dey, S. Chakraborty, S. Rachagani, *et al.*, 2009 RNA polymerase II associated factor 1/PD2 maintains self-renewal by its interaction with Oct3/4 in mouse embryonic stem cells. Stem Cells 27: 3001–3011. https://doi.org/10.1002/stem.237

Porrua O., and D. Libri, 2015 Transcription termination and the control of the transcriptome: why, where and how to stop. Nat. Rev. Mol. Cell Biol. 16: 190–202. https://doi.org/10.1038/nrm3943

Porter S. E., K. L. Penheiter, and J. A. Jaehning, 2005 Separation of the Saccharomyces cerevisiae Paf1 complex from RNA polymerase II results in changes in its subnuclear localization. Eukaryot. Cell 4: 209–220. https://doi.org/10.1128/EC.4.1.209-220.2005

Puig S., E. Askeland, and D. J. Thiele, 2005 Coordinated remodeling of cellular metabolism during iron deficiency through targeted mRNA degradation. Cell 120: 99–110. https://doi.org/10.1016/j.cell.2004.11.032

Qiu H., C. Hu, C.-M. Wong, and A. G. Hinnebusch, 2006 The Spt4p Subunit of Yeast DSIF Stimulates Association of the Paf1 Complex with Elongating RNA Polymerase II. Mol. Cell. Biol. 26: 3135–3148. https://doi.org/10.1128/MCB.26.8.3135-3148.2006

Qiu H., C. Hu, and A. G. Hinnebusch, 2009 Phosphorylation of the Pol II CTD by KIN28 Enhances BUR1/BUR2 Recruitment and Ser2 CTD Phosphorylation Near Promoters. Mol. Cell 33: 752–762. https://doi.org/10.1016/j.molcel.2009.02.018

Qiu H., C. Hu, N. A. Gaur, and A. G. Hinnebusch, 2012 Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. EMBO J. 31: 3494–505. https://doi.org/10.1038/emboj.2012.188

Quinlan A. R., and I. M. Hall, 2010 BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033

Quinlan A. R., 2014 BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinforma. 47: 11.12.1-11.12.34. https://doi.org/10.1002/0471250953.bi1112s47

Ramírez F., F. Dündar, S. Diehl, B. A. Grüning, and T. Manke, 2014 DeepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42: 187–191. https://doi.org/10.1093/nar/gku365

Ramírez F., D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, *et al.*, 2016 deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44: W160–W165. https://doi.org/10.1093/nar/gkw257

Rando O. J., and H. Y. Chang, 2009 Genome-wide views of chromatin structure. Annu. Rev. Biochem. 78: 245–271. https://doi.org/10.1146/annurev.biochem.78.071107.134639

Rando O. J., and F. Winston, 2012 Chromatin and transcription in yeast. Genetics 190: 351–387. https://doi.org/10.1534/genetics.111.132266

Ransom M., B. K. Dennehey, and J. K. Tyler, 2010 Chaperoning Histones during DNA Replication and Repair. Cell 140: 183–195. https://doi.org/10.1016/j.cell.2010.01.004

Raupach E. A., J. A. Martens, and K. M. Arndt, 2016 Evidence for Regulation of ECM3 Expression by Methylation of Histone H3 Lysine 4 and Intergenic Transcription in Saccharomyces cerevisiae. G3 Genes, Genomes Genet. 6: 2971–2981. https://doi.org/10.1534/g3.116.033118

Reavey C. T., M. J. Hickman, K. C. Dobi, D. Botstein, and F. Winston, 2015 Analysis of Polygenic Mutants Suggests a Role for Mediator in Regulating Transcriptional Activation Distance in Saccharomyces cerevisiae. Genetics 201: 599–612. https://doi.org/10.1534/genetics.115.181164

Rhee H. S., and B. F. Pugh, 2012 Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature 483: 295–301. https://doi.org/10.1038/nature10799

Ritchie M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, *et al.*, 2015 Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43: e47. https://doi.org/10.1093/nar/gkv007

Rose M. D., F. Winston, and P. Heiter, 1991 Methods in yeast genetics: a laboratory course manual. Biochem. Educ. 19: 101–102.

Rowley M. J., and V. G. Corces, 2018 Organizational principles of 3D genome architecture. Nat. Rev. Genet. 19: 789–800. https://doi.org/10.1038/s41576-018-0060-8

Roy A., A. Kucukural, and Y. Zhang, 2010 I-TASSER: A unified platform for automated protein structure and function prediction. Nat. Protoc. 5: 725–738. https://doi.org/10.1038/nprot.2010.5

Rozenblatt-Rosen O., C. M. Hughes, S. J. Nannepaga, K. S. Shanmugam, T. D. Copeland, *et al.*, 2005 The Parafibromin Tumor Suppressor Protein Is Part of a Human Paf1 Complex. Mol. Cell. Biol. 25: 612–620. https://doi.org/10.1128/MCB.25.2.612-620.2005

Rozenblatt-Rosen O., T. Nagaike, J. M. Francis, S. Kaneko, K. A. Glatt, *et al.*, 2009 The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3′ mRNA processing factors. Proc. Natl. Acad. Sci. 106: 755–760. https://doi.org/10.1073/pnas.0812023106

Russel D., K. Lasker, B. Webb, J. Velázquez-Muriel, E. Tjioe, *et al.*, 2012 Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol. 10: 1–5. https://doi.org/10.1371/journal.pbio.1001244

Rutherford J. C., and A. J. Bird, 2004 Metal-Responsive Transcription Factors That Regulate Iron , Zinc , and Copper Homeostasis in Eukaryotic Cells. Eurkaryotic Cell 3: 1–13. https://doi.org/10.1128/EC.3.1.1–13.2004

Sainsbury S., C. Bernecky, and P. Cramer, 2015 Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 16: 129–143. https://doi.org/10.1038/nrm3952

Saldi T., N. Fong, and D. L. Bentley, 2018 Transcription elongation rate affects nascent histone pre-mRNA folding and 3 ′ end processing. 1–12. https://doi.org/10.1101/gad.310896.117.min

Samara N. L., and C. Wolberger, 2011 A new chapter in the transcription SAGA. Curr. Opin. Struct. Biol. 21: 767–774. https://doi.org/10.1016/j.sbi.2011.09.004

Saunders A., L. J. Core, and J. T. Lis, 2006 Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol 7: 557–567. https://doi.org/10.1038/nrm1981

Schlessinger J., 1994 SH2/SH3 signaling proteins. Curr. Opin. Genet. Dev. 4: 25–30. https://doi.org/10.1016/0959-437X(94)90087-6

Schmid M., and T. H. Jensen, 2008 The exosome: a multipurpose RNA-decay machine. Trends Biochem. Sci. 33: 501–510. https://doi.org/10.1016/j.tibs.2008.07.003

Schreieck A., A. D. Easter, S. Etzold, K. Wiederhold, M. Lidschreiber, *et al.*, 2014 RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. Nat. Struct. Mol. Biol. 21: 175–179. https://doi.org/10.1038/nsmb.2753

Schrödinger, LLC, 2015 *The {PyMOL} Molecular Graphics System, Version~1.8*.

Schüller R., I. Forné, T. Straub, A. Schreieck, Y. Texier, *et al.*, 2016 Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol. Cell 61: 305–314. https://doi.org/10.1016/j.molcel.2015.12.003

Schulz D., B. Schwalb, A. Kiesel, C. Baejen, P. Torkler, *et al.*, 2013 Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell 155: 1075–1087. https://doi.org/10.1016/j.cell.2013.10.024

Sdano M. A., J. M. Fulcher, S. Palani, M. B. Chandrasekharan, T. J. Parnell, *et al.*, 2017 A novel SH2 recognition mechanism recruits Spt6 to the doubly phosphorylated RNA polymerase II linker at sites of transcription. 1–24.

Selth L. a, S. Sigurdsson, and J. Q. Svejstrup, 2010 Transcript Elongation by RNA Polymerase II. Annu. Rev. Biochem. 79: 271–293. https://doi.org/10.1146/annurev.biochem.78.062807.091425

Shattuck T., S. Valimaki, T. Obara, R. D. Gaz, O. H. Clark, *et al.*, 2003 Somatic and Germ-Line Mutations of the HRPT2 Gene in Sporadic Parathyroid Carcinoma. N. Engl. J. Med. 349: 1722–1729. https://doi.org/10.1056/NEJMoa031237

Sheldon K. E., D. M. Mauger, and K. M. Arndt, 2005 A Requirement for the Saccharomyces cerevisiae Paf1 Complex in snoRNA 3′ End Formation. Mol. Cell 20: 225–236. https://doi.org/10.1016/j.molcel.2005.08.026

Shendure J., and H. Ji, 2008 Next-generation DNA sequencing. Nat. Biotechnol. 26: 1135–1145. https://doi.org/10.1038/nbt1486

Shetty A., S. P. Kallgren, C. Demel, P. Cramer, P. J. Park, *et al.*, 2017 Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation Article Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. Mol. Cell 66: 77-88.e5. https://doi.org/10.1016/j.molcel.2017.02.023

Shi X., A. Finkelstein, a J. Wolf, P. a Wade, Z. F. Burton, *et al.*, 1996 Paf1p, an RNA polymerase II-associated factor in Saccharomyces cerevisiae, may have both positive and negative roles in transcription. Mol. Cell. Biol. 16: 669–676. https://doi.org/10.1128/MCB.16.2.669

Shi X., M. Chang, a J. Wolf, C. H. Chang, a a Frazer-Abel, *et al.*, 1997 Cdc73p and Paf1p are found in a novel RNA polymerase II-containing complex distinct from the Srbp-containing holoenzyme. Mol. Cell. Biol. 17: 1160–1169.

Shi Y., J. Fernandez-Martinez, E. Tjioe, R. Pellarin, S. J. Kim, *et al.*, 2014 Structural Characterization by Cross-linking Reveals the Detailed Architecture of a Coatomer-related Heptameric Module from the Nuclear Pore Complex. Mol. Cell. Proteomics 13: 2927–2943. https://doi.org/10.1074/mcp.M114.041673

Shi Y., R. Pellarin, P. C. Fridy, J. Fernandez-Martinez, M. K. Thompson, *et al.*, 2015 A strategy for dissecting the architectures of native macromolecular assemblies. Nat. Methods 12: 1135–1138. https://doi.org/10.1038/nmeth.3617

Shirra M. K., S. E. Rogers, D. E. Alexander, and K. M. Arndt, 2005 The Snf1 protein kinase and Sit4 protein phosphatase have opposing functions in regulating TATA-binding protein association with the Saccharomyces cerevisiae INO1 promoter. Genetics 169: 1957–1972. https://doi.org/10.1534/genetics.104.038075

Siegel R. L., K. D. Miller, and A. Jemal, 2016 Cancer Statistics , 2016. 66: 7–30. https://doi.org/10.3322/caac.21332.

Sikorski R. S., and J. D. Boeke, 1991 [20] In vitro mutagenesis and plasmid shuffling: From cloned gene to mutant yeast, pp. 302–318 in *Methods in enzymology*, Elsevier.

Simchen G., F. Winston, C. A. Styles, and G. R. Fink, 1984 Ty-mediated gene expression of the LYS2 and HIS4 genes of Saccharomyces cerevisiae is controlled by the same SPT genes. Proc. Natl. Acad. Sci. U. S. A. 81: 2431–4.

Simic R., D. L. Lindstrom, H. G. Tran, K. L. Roinick, P. J. Costa, *et al.*, 2003 Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. 22.

Smith D., and A. Quinlan, 2008 Rapid whole-genome mutational profiling using next-generation sequencing technologies. Genome Res. 1638–1642. https://doi.org/10.1101/gr.077776.108.

Smith E., C. Lin, and A. Shilatifard, 2011 The super elongation complex (SEC) and MLL in development and disease. Genes Dev. 25: 661–672. https://doi.org/10.1101/gad.2015411

Smolle M., and J. L. Workman, 2013 Transcription-associated histone modifications and cryptic transcription. Biochim. Biophys. Acta - Gene Regul. Mech. 1829: 84–97. https://doi.org/10.1016/j.bbagrm.2012.08.008

Squazzo S. L., P. J. Costa, D. L. Lindstrom, K. E. Kumer, R. Simic, *et al.*, 2002 The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. EMBO J. 21: 1764–1774. https://doi.org/10.1093/emboj/21.7.1764

Srivastava R., and S. H. Ahn, 2015 Modifications of RNA polymerase II CTD: Connections to the histone code and cellular function. Biotechnol. Adv. 33: 856–872. https://doi.org/10.1016/j.biotechadv.2015.07.008

Strikoudis A., C. Lazaris, T. Trimarchi, A. L. Galvao Neto, Y. Yang, *et al.*, 2016 Regulation of transcriptional elongation in pluripotency and cell differentiation by the PHD-finger protein Phf5a. Nat. Cell Biol. 18: 1127–1138. https://doi.org/10.1038/ncb3424

Strynatka K. A., M. C. Gurrola-Gal, J. N. Berman, and C. R. McMaster, 2018 How Surrogate and Chemical Genetics in Model Organisms Can Suggest Therapies for Human Genetic Diseases. Genetics 208: 833–851. https://doi.org/10.1534/genetics.117.300124

Suh H., S. B. Ficarro, U.-B. Kang, Y. Chun, J. A. Marto, *et al.*, 2016 Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Mol. Cell 61: 297–304. https://doi.org/10.1016/j.molcel.2015.12.021

Sun Z.-W., and C. D. Allis, 2002 Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature 418: 104–108. https://doi.org/10.1038/nature00883

Sun M., L. Larivière, S. Dengl, A. Mayer, and P. Cramer, 2010 A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). J. Biol. Chem. 285: 41597–41603. https://doi.org/10.1074/jbc.M110.144568

Sun W., X. Kuang, Y. Liu, L. Tian, X. Yan, *et al.*, 2017 Crystal structure of the N-terminal domain of human CDC73 and its implications for the hyperparathyroidism-jaw tumor (HPT-JT) syndrome. Sci. Rep. 1–9. https://doi.org/10.1038/s41598-017-15715-9

Svejstrup J. Q., 2004 The RNA polymerase II transcription cycle: Cycling through chromatin. Biochim. Biophys. Acta - Gene Struct. Expr. 1677: 64–73.

https://doi.org/10.1016/j.bbaexp.2003.10.012

Takahashi A., R. Tsutsumi, I. Kikuchi, C. Obuse, Y. Saito, *et al.*, 2011 SHP2 Tyrosine Phosphatase Converts Parafibromin/Cdc73 from a Tumor Suppressor to an Oncogenic Driver. Mol. Cell 43: 45–56. https://doi.org/10.1016/j.molcel.2011.05.014

Team R. C., 2016 R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria.

Terzi N., L. S. Churchman, L. Vasiljeva, J. Weissman, and S. Buratowski, 2011 H3K4 Trimethylation by Set1 Promotes Efficient Termination by the Nrd1-Nab3-Sen1 Pathway. Mol. Cell. Biol. 31: 3569–3583. https://doi.org/10.1128/MCB.05590-11

Thiebaut M., E. Kisseleva-Romanova, M. Rougemaille, J. Boulay, and D. Libri, 2006 Transcription Termination and Nuclear Degradation of Cryptic Unstable Transcripts: A Role for the Nrd1-Nab3 Pathway in Genome Surveillance. Mol. Cell 23: 853–864. https://doi.org/10.1016/j.molcel.2006.07.029

Thorvaldsdottir H., J. T. Robinson, and J. P. Mesirov, 2013 Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14: 178–192. https://doi.org/10.1093/bib/bbs017

Toesca I., C. R. Nery, C. F. Fernandez, S. Sayani, and G. F. Chanfreau, 2011 Cryptic transcription mediates repression of subtelomeric metal homeostasis genes. PLoS Genet. 7. https://doi.org/10.1371/journal.pgen.1002163

Tommaso P. Di, E. Palumbo, M. Chatzou, P. Prieto, M. L. Heuer, *et al.*, 2015 The impact of Docker containers on the performance of genomic pipelines. PeerJ 3: e1273. https://doi.org/10.7717/peerj.1273

Tomson B. N., C. P. Davis, M. H. Warner, and K. M. Arndt, 2011 Identification of a role for histone H2B ubiquitylation in noncoding RNA 3'-end formation through mutational analysis of Rtf1 in Saccharomyces cerevisiae. Genetics 188: 273–289. https://doi.org/10.1534/genetics.111.128645

Tomson B. N., and K. M. Arndt, 2013 The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. Biochim. Biophys. Acta - Gene Regul. Mech. 1829: 166–126. https://doi.org/10.1016/j.bbagrm.2012.08.011

Tomson B. N., E. M. Crisucci, L. E. Heisler, M. Gebbia, C. Nislow, *et al.*, 2013a Effects of the Paf1 Complex and Histone Modifications on snoRNA 3'-End Formation Reveal Broad and Locus-Specific Regulation. Mol. Cell. Biol. 33: 170–182. https://doi.org/10.1128/MCB.01233-12

Tomson B. N., E. M. Crisucci, L. Heisler, M. Gebbia, C. Nislow, *et al.*, 2013b Characterization of snoRNA 3'-end formation in Saccharomyces cerevisiae reveals a broad role for the Paf1 complex and locus-specific roles for histone post-translational modifications. FASEB J. 27: lb97–lb97.

Toolkit P., 2016 Broad Institute

Tous C., A. G. Rondón, M. García-Rubio, C. González-Aguilera, R. Luna, *et al.*, 2011 A novel assay identifies transcript elongation roles for the Nup84 complex and RNA processing factors. EMBO J. 30: 1953–1964. https://doi.org/10.1038/emboj.2011.109

Tremethick D. J., 2007 Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. Cell 128: 651–654. https://doi.org/10.1016/j.cell.2007.02.008

Vaňáčová Š., J. Wolf, G. Martin, D. Blank, S. Dettwiler, *et al.*, 2005 A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol. 3: 0986–0997. https://doi.org/10.1371/journal.pbio.0030189

Vandemark A. P., H. Xin, L. McCullough, R. Rawlins, S. Bentley, *et al.*, 2008 Structural and functional analysis of the Spt16p N-terminal domain reveals overlapping roles of yFACT subunits. J. Biol. Chem. 283: 5058–5068. https://doi.org/10.1074/jbc.M708682200

VanDemark A. P., M. Blanksma, E. Ferris, A. Heroux, C. P. Hill, *et al.*, 2006 The Structure of the yFACT Pob3-M Domain, Its Interaction with the DNA Replication Factor RPA, and a Potential Role in Nucleosome Deposition. Mol. Cell 22: 363–374. https://doi.org/10.1016/j.molcel.2006.03.025

Vannini A., and P. Cramer, 2012 Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. Mol. Cell 45: 439–446. https://doi.org/10.1016/j.molcel.2012.01.023

Vaser R., S. Adusumalli, S. N. Leng, M. Sikic, and P. C. Ng, 2016 SIFT missense predictions for genomes. Nat. Protoc. 11: 1–9. https://doi.org/10.1038/nprot.2015.123

Venkatesh S., M. Smolle, H. Li, M. M. Gogol, M. Saint, *et al.*, 2012 Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature 489: 452–455. https://doi.org/10.1038/nature11326

Venkatesh S., and J. L. Workman, 2015 Histone exchange, chromatin structure and the regulation of transcription. Nat. Rev. Mol. Cell Biol. 16: 178–189. https://doi.org/10.1038/nrm3941

Venkatesh S., H. Li, M. M. Gogol, and J. L. Workman, 2016 Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. Nat. Commun. 7: 13610. https://doi.org/10.1038/ncomms13610

Vera J. M., and R. D. Dowell, 2016 Survey of cryptic unstable transcripts in yeast. BMC Genomics 17: 305. https://doi.org/10.1186/s12864-016-2622-5

Vinayachandran V., R. Reja, M. J. Rossi, B. Park, L. Rieber, *et al.*, 2018 Widespread and precise reprogramming of yeast protein–genome interactions in response to heat shock. Genome Res. 28: 357–366. https://doi.org/10.1101/gr.226761.117

Vos S. M., L. Farnung, H. Urlaub, and P. Cramer, 2018a Structure of paused transcription complex

Pol II–DSIF–NELF. Nature 560: 601–606. https://doi.org/10.1038/s41586-018-0442-2

Vos S. M., L. Farnung, M. Boehning, C. Wigge, A. Linden, *et al.*, 2018b Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature 560: 607–612. https://doi.org/10.1038/s41586-018-0440-4

Vos S. M., L. Farnung, A. Linden, H. Urlaub, and P. Cramer, 2020 Structure of complete Pol II–DSIF–PAF–SPT6 transcription complex reveals RTF1 allosteric activation. Nat. Struct. Mol. Biol. https://doi.org/10.1038/s41594-020-0437-1

Wada T., T. Takagi, Y. Yamaguchi, A. Ferdous, T. Imai, *et al.*, 1998 DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev. 12: 343–356. https://doi.org/10.1101/gad.12.3.343

Wade P., W. Werel, R. Fentzke, N. Thompson, J. Leykam, *et al.*, 1996 A novel collection of accessory factors associated with yeast RNA polymerase II. Protein Expr Purif 8: 85–90. https://doi.org/10.1006/prep.1996.0077

Wang P. F., M. H. Tan, C. Zhang, H. Morreau, and B. T. Teh, 2005 HRPT2, a Tumor Suppressor Gene for Hyperparathyroidism−Jaw Tumor Syndrome P. Horm Metab Res 37: 380–383. https://doi.org/10.1055/s

Wang P., M. R. Bowl, S. Bender, J. Peng, L. Farber, *et al.*, 2008 Parafibromin, a component of the human PAF complex, regulates growth factors and is required for embryonic development and survival in adult mice. Mol. Cell. Biol. 28: 2930–40. https://doi.org/10.1128/MCB.00654-07

Wang K., M. Li, and H. Hakonarson, 2010 ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38: 1–7. https://doi.org/10.1093/nar/gkq603

Wangler M. F., S. Yamamoto, H.-T. Chao, J. E. Posey, M. Westerfield, *et al.*, 2017 Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research. Genetics 207: 9–27. https://doi.org/10.1534/genetics.117.203067

Warner M. H., K. L. Roinick, and K. M. Arndt, 2007 Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification. Mol. Cell. Biol. 27: 6103–15. https://doi.org/10.1128/MCB.00772-07

Warren C., and D. Shechter, 2017 Fly Fishing for Histones: Catch and Release by Histone Chaperone Intrinsically Disordered Regions and Acidic Stretches. J. Mol. Biol. 429: 2401–2426. https://doi.org/10.1016/j.jmb.2017.06.005

Weiner A., T. H. S. Hsieh, A. Appleboim, H. V. Chen, A. Rahat, *et al.*, 2015 High-resolution chromatin dynamics during a yeast stress response. Mol. Cell 58. https://doi.org/10.1016/j.molcel.2015.02.002

Werner-Allen J. W., C.-J. Lee, P. Liu, N. I. Nicely, S. Wang, *et al.*, 2011 cis -Proline-mediated

Ser(P) 5 Dephosphorylation by the RNA Polymerase II C-terminal Domain Phosphatase Ssu72. J. Biol. Chem. 286: 5717–5726. https://doi.org/10.1074/jbc.M110.197129

Werner F., and D. Grohmann, 2011 Evolution of multisubunit RNA polymerases in the three domains of life. Nat Rev Microbiol 9: 85–98. https://doi.org/nrmicro2507 [pii]\r10.1038/nrmicro2507

Wier A. D., M. K. Mayekar, A. Heroux, K. M. Arndt, and A. P. VanDemark, 2013 Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin. Proc. Natl. Acad. Sci. 110: 17290–17295. https://doi.org/10.1073/pnas.1314754110

Winston F., D. T. Chaleff, B. Valent, and G. R. Fink, 1984 Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. Genetics 107: 179–197.

Winston F., C. Dollard, and S. L. Ricupero-Hovasse, 1995 Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast 11: 53–55. https://doi.org/10.1002/yea.320110107

Winston F., and D. Koshland, 2016 Back to the future: Mutant hunts are still the way to go. Genetics 203: 1007–1010. https://doi.org/10.1534/genetics.115.180596

Wood K., M. Tellier, and S. Murphy, 2018 DOT1L and H3K79 Methylation in Transcription and Genomic Stability. Biomolecules 8: 11. https://doi.org/10.3390/biom8010011

Wyers F., M. Rougemaille, G. Badis, J. C. Rousselle, M. E. Dufour, *et al.*, 2005 Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 121: 725–737. https://doi.org/10.1016/j.cell.2005.04.030

Xiao T., C. Kao, N. J. Krogan, J. F. Greenblatt, M. A. Osley, *et al.*, 2005 Histone H2B Ubiquitylation Is Associated with Elongating RNA Polymerase II Histone H2B Ubiquitylation Is Associated with Elongating RNA Polymerase II. Society 25: 637–651. https://doi.org/10.1128/MCB.25.2.637

Xu Z., W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Münster, *et al.*, 2009 Bidirectional promoters generate pervasive transcription in yeast. Nature 457: 1033–7. https://doi.org/10.1038/nature07728

Xu Y., C. Bernecky, C. Lee, K. C. Maier, B. Schwalb, *et al.*, 2017 Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. Nat. Commun. 8: 15741. https://doi.org/10.1038/ncomms15741

Yamaguchi-Iwai Y., R. Ueta, A. Fukunaka, and R. Sasaki, 2002 Subcellular Localization of Aft1 Transcription Factor Responds to Iron Status in Saccharomyces cerevisiae. J. Biol. Chem. 277: 18914–18918. https://doi.org/10.1074/jbc.M200949200

Yang J., R. Yan, A. Roy, D. Xu, J. Poisson, *et al.*, 2014 The I-TASSER suite: Protein structure and function prediction. Nat. Methods 12: 7–8. https://doi.org/10.1038/nmeth.3213

Yang Y., W. Li, M. Hoque, L. Hou, S. Shen, *et al.*, 2016 PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation. PLoS Genet. 12: 1–28. https://doi.org/10.1371/journal.pgen.1005794

Yassour M., J. Pfiffner, J. Z. Levin, X. Adiconis, A. Gnirke, *et al.*, 2010 Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. Genome Biol. 11: R87. https://doi.org/10.1186/gb-2010-11-8-r87

Ye Y., and A. Godzik, 2003 Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics 19. https://doi.org/10.1093/bioinformatics/btg1086

Ye Y., and A. Godzik, 2004a Database searching by flexible protein structure alignment. Protein Sci. 13: 1841–1850. https://doi.org/10.1110/ps.03602304

Ye Y., and A. Godzik, 2004b FATCAT: A web server for flexible structure comparison and structure similarity searching. Nucleic Acids Res. 32: 582–585. https://doi.org/10.1093/nar/gkh430

Yu M., W. Yang, T. Ni, Z. Tang, T. Nakadai, *et al.*, 2015a RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science 350: 1383–1386. https://doi.org/10.1126/science.aad2338

Yu M., W. Yang, T. Ni, Z. Tang, T. Nakadai, *et al.*, 2015b RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science 350: 1383–1386. https://doi.org/10.1126/science.aad2338

Zhang Y., 2008 I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 1–8. https://doi.org/10.1186/1471-2105-9-40

Zhao J., A. Yart, S. Frigerio, A. Perren, P. Schraml, *et al.*, 2007 Sporadic human renal tumors display frequent allelic imbalances and novel mutations of the HRPT2 gene. Oncogene 26: 3440–3449. https://doi.org/1210131 [pii]\r10.1038/sj.onc.1210131

Zhou K., W. H. W. Kuo, J. Fillingham, and J. F. Greenblatt, 2009 Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. Proc. Natl. Acad. Sci. U. S. A. 106: 6956–61. https://doi.org/10.1073/pnas.0806302106

Zhou X., and E. K. O'Shea, 2011 Integrated Approaches Reveal Determinants of Genome-wide Binding and Function of the Transcription Factor Pho4. Mol. Cell 42: 826–836. https://doi.org/10.1016/j.molcel.2011.05.025

Zhu B., S. S. Mandal, A. D. Pham, Y. Zheng, H. Erdjument-Bromage, *et al.*, 2005 The human PAF complex coordinates transcription with events downstream of RNA synthesis. Genes Dev. 19: 1668–1673. https://doi.org/10.1101/gad.1292105

Zhu W., T. Wada, S. Okabe, T. Taneda, Y. Yamaguchi, *et al.*, 2007 DSIF contributes to transcriptional activation by DNA-binding activators by preventing pausing during transcription elongation. Nucleic Acids Res. 35: 4064–4075.

https://doi.org/10.1093/nar/gkm430