UNIVERSITY OF CALIFORNIA

SANTA CRUZ

Spt5 and the Pol II Stalk Collaborate to Regulate Co-transcriptional pre-mRNA processing

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY
In

Molecular, Cell and Developmental Biology

By

Zachary Asher Morton

September 2022

The Dissertation of Zachary Asher Morton is approved:

Professor Grant Hartzog, chair

Professor Hinrich Boeger

Professor Rohinton Kamakaka

Professor Doug Kellogg

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures and Tables

Abstract

Spt5 and the Pol II Stalk Collaborate to Regulate Co-transcriptional pre-mRNA processing

Zachary Asher Morton

Transcription of DNA into mRNA by RNA Polymerase II (Pol II) is a highly dynamic and complex process that requires the concurrent collaboration of many factors. One such factor is Spt5 - a multi-domain transcription elongation factor that acts as a component of all Pol II elongation complexes. It is universally conserved and essential for life, playing a central and ancient role in transcription.

Recent structural studies indicate that several of Spt5's central KOW (Kyprides, Ouzounis, Woese) domains lie in close contact with the dissociable stalk of RNA Polymerase II (subunits Rpb4 and Rpb7) (Bernecky *et al.,* 2017). Both Spt4/5 and Rpb4/7 have been previously implicated in polyadenylation (poly(A)) site choice (Cui *et al.,* 2003; Runner *et al.,* 2008) 3' end processing of mRNA (Mayer *et al*., 2012; Runner *et al.,* 2008), mRNA export (Burckin *et al.,* 2005; Farago *et al.,* 2003), and allosteric stabilization of elongating Pol II (Armache *et al.,* 2016; Bernecky *et al.,* 2017). Despite these structural studies that report physical interactions between these proteins, a functional interaction has yet to be revealed.

Here we present evidence that Spt5 and Rpb4/7 collaborate to execute functions relating to 3' end formation, mRNA export, cotranscriptional chromatin maintenance and R-loop formation. We take a genetic and biochemical approach to address the consequences of disrupting the direct interactions of these proteins, revealing a series of allele-specific genetic interactions between *SPT5, RPB4* and *RPB7* that point to a functional cooperation throughout transcription elongation and termination. Affinity chromatography of yeast cell extracts using isolated Spt5 KOW domains as bait revealed a large set of KOW-interacting proteins. Many of these were previously reported to interact with Rpb7 (Mosley *et al.,* 2013). These data point to a direct, functional cooperation between Spt5 and Rpb4/7 to regulate the processing and export of mRNA, revealing new roles for the previously ambiguous central domains of Spt5.

Chapter 1 summarizes the literature and current gaps in our knowledge with regards to the function of Spt5's central domains and the Pol II stalk in transcription.

Chapter 2 presents a genetic analysis of *spt5* and *rpb4/7* mutants. We have identified mutations at multiple points throughout the structure formed by Spt5 KOW2-4 and Rpb4/7, including the juncture of KOW3/Rpb7 and KOW4/Rpb7, as well as multiple solvent exposed regions on the surface of this structure. Many of these mutations genetically interact with each other, resulting in synthetic sickness and enhancement of phenotypes. Interestingly,

mutations in *Rpb4* and *Rpb7* both share the cryptic transcription phenotype that pervades many known *spt5* alleles as well as chromatin remodeling proteins, suggesting that the core structure of Pol II itself evolved to assist in overcoming nucleosomal barriers to transcription.

Chapter 3 reports identification of biochemical interactions between Spt5's central KOW domains and their interacting factors. After establishing that Spt5 KOW2-4 and Rpb4/7 extensively interact genetically, we aim to test the hypothesis that this region functions as a binding platform for tertiary factors. We performed affinity chromatography followed by MudPIT mass spectrometry to identify binding partners of both the KOW2-3 region (K2K3) and the Linker2-KOW4 (L2K4) region of Spt5 (Delahunty, Yates 2007). Supporting the notion that the central KOW domains function in tandem with Rpb4/7, we identified a large number of proteins involved in 3' end formation, RNA processing and chromatin structure maintenance that overlap with previous studies of Rpb7.

Chapter 4 explores the functional relationship between Spt5, Rpb7, Nrd1 and R-loop formation. Together with evidence from previous chapters, we propose a model for the Pol II Stalk region in functioning to maintain R-loop homeostasis by bridging interactions with tertiary factors.

Chapter 5 summarizes the main findings of this work and describes the contribution made to our existing knowledge gaps in transcriptional biology.

Dedication and Acknowledgements

I would first like to express my deepest gratitude to my advisor, Dr. Grant Hartzog, for inspiring me to pursue a higher education in molecular biology research and permitting me to join his lab at the beginning of the program. Your guidance and willingness to help was instrumental in the development of my project, while always inspiring and encouraging me to pursue the questions that I found to be the most interesting. I found a home in the Hartzog Lab and I will always be grateful for your mentorship.

I would also like to thank Professors Hinrich Boeger, Doug Kellogg and Rohinton Kamakaka for their thoughtful advice throughout my thesis committee meetings and guiding the direction of my project. I always enjoyed our stimulating conversations about science and greatly appreciate your contributions to my project. In particular, I express my thanks to Hinrich for suggesting I look into R-loops.

Michael Doody, your support throughout my time in the Hartzog Lab was invaluable. You are a great teacher, friend, and mentor and I would not be the same scientist or teacher without your guidance and friendship throughout the program. I also extend my gratitude to all members of the Hartzog lab for their contributions to the lab and to my dissertation work, as well as being great overall lab mates and friends.

My six years spent in this program would not have been nearly as fun or pleasant without all of my friends that stood by me throughout my time in Santa Cruz. You all know who you are and I am so grateful for the good times we've had together and will always treasure our friendships.

I am so grateful for the support I have received from my family – Mom, Dad, Matt, Aunt Sally, Uncle Chip, Aunt Heather, and my grandma. Your support throughout my life allowed me to flourish as a person and inspired confidence in me to pursue my dreams, and I will always be grateful for that.

Finally, I have to thank Jennifer Long and her family for always being there for me and encouraging me to keep pushing through the challenging times. This work would not be possible without the support by everyone mentioned here and I am very fortunate to have all of you in my life.

Chapter 1

Introduction

The Central Dogma of Biology

The central dogma of biology provides an explanation for the flow of genetic information within our cells, and is comprised of three main assertions (Crick, 1970). First, our genetic information is coded into our cells as DNA and replicated by DNA polymerase during mitosis. Second, RNA polymerase catalyzes the synthesis of RNA from the DNA template, composing a transient “message” that carries the information coded in the DNA. In prokaryotes, there is only one RNA polymerase. However, in eukaryotes, 3 distinct RNA polymerases are required. RNA Polymerase I and III transcribe structural RNAs required for protein translation – ribosomal RNA (rRNA) and transfer RNA (tRNA) respectively. RNA Polymerase II is responsible for transcribing the entirety of our protein-coding genes into messenger RNA (mRNA). Third, the mRNA is read by the ribosome which uses it as a template for protein synthesis. In eukaryotes, this requires the packaging and shuttling of mRNA out of the nucleus and into the cytoplasm where the ribosomes reside. This thesis focuses on transcription of mRNA by Pol II in the model organism *Saccharomyces cerevisiae*.

Fundamentals of Transcription by RNA Polymerase II

Transcription occurs in three steps: initiation, elongation, and termination. Initiation requires the recognition of a specific DNA sequence called the promoter by general transcription factors which recruit Pol II to the beginning of the gene. Pol II must first be assembled around the promoter region. This requires the help of a number of accessory factors, known as transcription initiation factors, which facilitate the targeting of Pol II to the promoter, assembly of the twelve subunits, and the melting of the double-stranded DNA. Following assembly, Pol II escapes the promoter and begins transcription. The next step is transcription elongation, where Pol II catalyzes the addition of ribonucleotides complimentary to the DNA strand. The final step of transcription is termination, wherein Pol II disengages from the DNA and the nascent transcript is cleaved.

Transcription by Pol II is complicated by a number of factors. Nascent mRNA must be appropriately processed and packaged in a way that that serves to stabilize the RNA and prepare it for transit into the cytoplasm. This includes capping the mRNA with a 5’ N7-methylated guanosine, splicing, addition of the 3’ Poly(A) tail, and association with mRNA binding proteins. Much of these processes occur co-transcriptionally. While this is occurring, Pol II must also pass through DNA-bound nucleosomes, which serve to compact the DNA and act as a physical barrier to transcription. All the while, the speed and processivity of Pol II must be maintained as it encounters

these obstacles. As such, RNA Pol II transcription requires the cooperation of many accessory factors to accomplish its function.

Spt5: A Universally Conserved Transcription Elongation Factor

Spt5 is the only known transcription elongation factor that is conserved across all domains of life and is essential for life in every organism (Harris *et al.,* 2003; Werner 2012). Spt5 is a large (~116 kDa in *S. cerevisiae*) protein, with multiple, independently folding domains (Meyer *et al.,* 2015). Spt5's domains include an unstructured acidic N-terminus, the NGN domain which simultaneously binds Spt4 and spans the central cleft of elongating Pol II, 5 consecutive Kyprides, Ouzounis, Woese (KOW) domains and a set of C-terminal repeats (Klein *et al.,* 2011). Spt5's binding partner, Spt4, is a much smaller protein (14 kDa in S. cerevisiae) that is composed of a single zinc-finger domain which serves to stabilize Spt5 (Malone *et al.,* 1993; Ding *et al.,* 2010).

Spt4-Spt5 is implicated in most cotranscriptional processes: it aids in splicing (Maudlin *et al.,* 2019), 5' capping (Lindstrom *et al.,* 2003), regulating Pol II rate and processivity (Quan *et al.,* 2010), mRNA export (Burckin *et al.,* 2005), polyadenylation (Cui *et al.,* 2003), transcription coupled DNA repair (Li *et al.,* 2014) and allosteric stabilization of Pol II (Bernecky *et al.,* 2017). Although bacteria lack a homolog of Spt4, they do encode an Spt5 homolog, NusG, that binds RNA polymerase (Liu *et al.,* 2016, Li *et al.,* 1992) and

regulates transcription elongation (Burova *et al.,* 1995). In contrast to the complexity of Spt5, NusG, is composed of a NGN domain (Ponting 2002) and a single KOW domain (Steiner *et al.,* 2002). The conserved function of the NGN domain is to maintain transcription processivity by clamping down on transcribing polymerase and sealing the DNA within the central cleft (Liu *et al.,* 2016; Bernecky *et al.,* 2016; Bernecky *et al.,* 2017; Martinez-Rucobo *et al.,* 2011). NusG's lone KOW domain directly contacts co-transcriptionally associated ribosomes simultaneously with polymerase, providing a physical link between transcription and translation (Burmann *et al.,* 2010; Washburn *et al.,* 2020).

While the conserved role of the NGN domain has been well-defined, the precise functions of Spt5's central KOW domains remain ambiguous. Previous studies have shown that mutations in and deletions of Spt5's KOW domains result in loss of Spt5 function (Li *et al.,* 2014; Meyer *et al.,* 2015). Recent cryo-EM structures of Spt5 complexed with Pol II show that Spt5's NGN and KOW1 domain form a clamp around the upstream DNA (Ehara *et al.,* 2017; Bernecky *et al.,* 2017, and Spt5's KOW1 has a positively charged patch that appears to contact upstream DNA (Meyer *et al.,* 2015). NusG's NGN domain in *E. coli* also binds the central cleft of bacterial RNA polymerase while the lone KOW domain lies nearby the upstream DNA (Liu *et al.,* 2016). This supports the idea that the function of these domains has been conserved throughout evolutionary time dating back to bacteria – to maintain

transcription processivity by sealing the central cleft of RNA polymerase. However, the remaining KOW domains are not located near the DNA template nor the catalytic core of Pol II, making their role in elongation less clear. It appears that as Spt5's central KOW domains expanded throughout evolutionary time with an ever more complex transcription cycle, so did Spt5's transcription associated responsibilities.

The Structure of the Pol II Elongation Complex

Pol II is a eukaryote-specific 12 subunit holoenzyme complex responsible for transcription of all protein-coding genes and some noncoding RNAs. Pol II is divided into four flexible modules – core, jaw-lobe, shelf, and clamp (Cramer *et al.,* 2001). The catalytic center lies in the core and is comprised of primarily of subunits Rpb1 and Rpb2 and components of Rpb3, Rpb10, Rpb11 and Rpb12. The remaining modules shift conformation in response to DNA being threaded through the central cleft and facilitate DNA melting and RNA chain elongation.

Rpb1 is unique among the Pol II subunits in that it contains a large, unstructured C-terminus named the Carboxy-Terminal repeat Domain (CTD). In *S. cerevisiae,* the CTD is comprised of 26 heptad repeats whose sequence (YSPTSPS) is conserved from yeast to mammals (Allison *et al.,* 1988). The CTD serves as a recruitment platform for transcription associated factors with targets varying depending on the post-translational modifications (PTM's)

associated with it. While many of the residues are subject to PTMs, Ser_2 and Ser_5 are the primary targets and are subject to phosphorylation. Phospho-Ser_5 typically marks Pol II complexes near the promoter, and phospho-Ser_2 marks Pol II complexes near the 3' end (Egloff, Murphy 2008).

Rpb4 and Rpb7 together form a heterodimer that sits near the mRNA exit channel adjacent to the linker region that separates the CTD from the body of Rpb1 (Armache *et al.,* 2003). Rpb4/7 together form the "stalk" of Pol II and are able to reversibly disassociate from Pol II (Edwards *et al.,* 1991). Rpb7 is an essential gene across eukaryotes, while Rpb4 is dispensable in *S. cerevisiae* and essential in *S. pombe* and higher eukaryotes (Sharma, Kumari 2012). Regions of Rpb1, Rpb2 and Rpb6 form a pocket region near the core of Pol II which forms the major binding site for Rpb7. The "tip" of Rpb7 protrudes into the pocket region and aids in locking Pol II into a closed, processive conformation. Rpb4 binds the outside region of Rpb7 but does not form contact points with Pol II itself, and acts to stabilize Rpb7.

Pol II requires a number of accessory factors to facilitate its loading onto template DNA, DNA melting and transcription initiating. These factors include TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and Mediator. Upon successful initiation, Pol II sheds its initiation factors and a new set of factors specific for transcript elongation associate with Pol II (Nechaev, Adelman 2012). The Pol II elongation complex is complex and dynamic, but the main players include TFIIS, Spt4/5, Spt6, the PAF complex, P-TEFb, and Elf1.

Binding of Spt4/5 is mutually exclusive with TFIIB, TFIIE and TFIIF as they bind the same region of Pol II (Li *et al.,* 2014; Vos *et al.,* 2019).

Recent advances in cryo-electron microscopy have allowed the resolution of multiple structures of Pol II in complex with transcription elongation factors (Ehara *et al.,* 2017; Bernecky *et al.,* 2017; Vos *et al.,* 2018a; Filipovski *et al.,* 2022, Xu *et al.,* 2017; Vos *et al.,* 2018b). These structures provide insights into the mechanisms of allosteric regulation of the Pol II active site and maintenance of processive elongation, passage through nucleosomes, cotranscriptional chromatin modifications, mRNA processing and export, and 3' end processing. TFIIS reaches into the mRNA exit channel which explains its capacity for stimulating re-initiation of stalled and backtracked polymerases. Spt4/5 are positioned to impact nucleosomes downstream of the polymerase and seal the DNA within the central cleft. The remaining domains span across the exterior of Pol II from upstream DNA emerging from Pol II to the mRNA exit channel. Spt5 wraps around the stalk domain, appearing to stabilize interactions between the stalk and Pol II and extend the binding interface of the stalk region. Spt6 also contacts the stalk and Spt5 with an unstructured region positioned to contact upstream or downstream nucleosomes, which is consistent with its reported function in chromatin remodeling. The PAF complex spans a large portion of Pol II and mediates processes involving cotranscriptional chromatin modifications, 3' end processing and allosteric regulation of elongation (Crisucci, Arndt 2011).

While existing cryo-EM studies have been informative in acquiring snapshots of the Pol II Elongation Complex (EC) in different stages, a comprehensive model including molecular transitions throughout elongation, and particularly termination, has yet to be constructed.

Rpb4 and Rpb7: Functions of the Pol II Stalk

Rpb4 and Rpb7 are core Pol II subunits that together comprise the stalk region of Pol II in eukaryotes. These Pol II subunits are unique in that they are able to reversibly dissociate from the body of Pol II and exist in sub-stoichiometric amounts in complex with Pol II. Rpb4/7 are involved in every step in transcription, from initiation to termination (Edwards *et al.,* 1991; Kolodziej *et al.,* 1990; Choder, Young 1993; Sharma, Kumari 2012; Choder 2003). *In vitro* experiments have shown that Rpb4/7 are required for promoter-directed transcription initiation. Structurally, the Pol II Stalk protrudes from the foot domain of Pol II. Crystals of Pol II lacking Rpb4/7 are observed to be in an open conformation, and the addition of Rpb4/7 results in a more closed clamp region (Armache *et al.,* 2003). This may provide a mechanism for increased processivity in Rpb4/7 containing polymerases, by restricting the movement of the clamp domain, therefore allowing retention of Pol II on DNA during transcript elongation. While a mechanistic role for the Stalk in termination remains ambiguous, studies implicate Rpb4/7 in recruitment of 3’ processing factors and regulation of polyadenylation site choice.

While Rpb4/7 are absent in prokaryotes, archaeal orthologs exist in RpoE and RpoF, suggesting these Rpb4/7-like heterodimers evolved prior to the separation of *archae* and *eukarya* (Kyrpides, Ouzounis 1999; Todone *et al.,* 2001). Paralogs also exist in Pol I and Pol III (Peyroche *et al.,* 2002; Siaut *et al.,* 2003). Speaking to their functional conservation across species, Rpb4/7 from *Homo sapiens, Candida albicans, Drosophila melanogaster, Schizosaccharomyces pombe* Rpb7 all retain the capability to rescue the lethality of loss of Rpb7 in *Saccharomyces cerevisiae* (Khazak *et al.,* 1995; Singh *et al.,* 2004; Zhou, Lee 2001). Two key characteristics of Rpb4/7 and their paralogs and orthologues is the ability to stably bind each other and to RNA *in vitro* – indeed, the RNA binding fold of Rpb7 is the most highly conserved feature of this heterodimer. In *archae,* RpoE/F assists in DNA melting and stimulates polymerase processivity through allosteric modulations of polymerase structure (Hirtreiter *et al.,* 2010; Grohmann, Werner 2011), enhances formation of full-length gene products, reduces pausing, and facilitates termination *in vitro*. Further, these functions are dependent in part on the ability of RpoE/F to bind RNA. Despite this evidence, however, the *in vivo* relevance of the RNA binding properties remains ambiguous.

Of all the polymerase subunits, the Rpb4 family appears to be the least conserved. While all Rpb4 homologs retain the ability to bind to their cognate Rpb7 homolog, large extensions of the N or C termini can be observed across species. The placement of Rpb4/7 and their homologs away from the body of

Pol II, along with the large extensions of Rpb4-like proteins, may allow the stalk domain to attract diverse functional units to the transcription complex (Jun *et al.,* 2012). Research into the direct identification of the myriad Rpb4-binding proteins remains ongoing.

In addition to Rpb4/7's direct roles in transcription, data in both *S. cerevisiae* and *S. pombe* suggests a role for the heterodimer in transcription-coupled processes such as DNA repair (He, Ramatar 1999; Li, Smerdon 2002; Kumar, Sharma 2019), transcript termination (Runner *et al*., Mitsuzawa *et al.,* 2003; 2008; Sharma, Kumari 2012), ribosome biogenesis (Kumar *et al.,* 2019) and mRNA export (Farago *et al.,* 2003; Harel-Sharvit *et al.,* 2010). Rpb4 was shown to have contrasting roles in both independent transcription coupled repair (TCR) pathways – having a negative role in the Rpb9 mediated pathway and a positive role in the Rad26 mediated pathway. In *S. pombe,* Rpb7 was shown through genome-wide analysis to have roles in nucleotide excision repair, base excision repair as well as TCR. In regards to transcription termination, Rpb7 has been shown to bind Nrd1 which facilitates short-transcript termination as a component of the NNS (Nrd1-Nab3-Sen1) termination complex. Loss of Rpb4 results in failure to recruit the CFI components Rna14 and Rna15, but has no effect on Nrd1 recruitment. Despite these observations, a specific mechanistic role for Rpb7 in termination has yet to be identified.

Rpb4/7 is also thought to play a role in stress response pathways. To this effect, Rpb4 is required for efficient mRNA export under stress. Further, in humans and *S. pombe,* Rpb4/7 has been shown to relocate to the cytoplasm in a stress-specific manner. In *S. cerevisiae,* Rpb7 was shown to localize to P-bodies under stress conditions. These observations hint at a potential cytoplasmic function for the heterodimer. While Rpb4 is not essential under standard growth conditions in *S. cerevisiae,* it becomes essential under stressful conditions (Maillet *et al.,* 1999). Highlighting the essential, stress-related functions of the Pol II stalk, recent research has identified mutations in *rpb7* that increase ethanol tolerance in *S. cerevisiae* (Qiu, Jiant 2017). Interestingly, *RPB4* is an essential gene in *S. pombe* and *Homo sapiens.* Despite the identification of stress-specific roles for Rpb4, a comprehensive model explaining Rpb4's stress specific essentiality remains to be constructed.

While Rpb4/7's roles in 3' processing, TCR and mRNA decay are the subject of much research, more recent studies have focused on the stalk's ability to modulate the Rpb1 CTD phosphorylation state (Allepuz-Fuster *et al.,* 2014,) and stimulate Pol II recycling via the formation of gene loops (Allepuz-Fuster *et al.,* 2019; Calvo 2020). Rpb4 stimulates recruitment of the Ssu72 and Fcp1 phosphatases which de-phosphorylate Pol II's CTD. The stalk's involvement in gene-looping was elucidated via the observation that mutations specific to *rpb4* and *rpb7,* but not other Pol II subunits,

dysregulated gene-looping. In a surprising connection to Pol II phosphorylation regulation, gene-looping was shown in a recent study to be directly dependent on the ability of Rpb4 to mediate interactions between Ssu72 and the general transcription initiation factor TFIIB, highlighting the interconnectivity of Pol II related processes.

Some reports have suggested that the dissociation of Rpb4 and Rpb7 from Pol II is important for their function (Mosley *et al.,* 2014). The authors of this study suggests that Rpb4/7 may act as a shuttle service for cytoplasm-bound mRNA and may also participate in translation initiation and cytoplasmic mRNA decay pathways. The following observations support this model: 1) GFP-tagged Rpb7 localizes to the cytoplasm under stress conditions; 2) Rpb7 co-localizes with P-bodies under stress and stimulates mRNA decay in the cytoplasm as well as cytoplasmic deadenylation; 3) Rpb7 physically and functionally interacts with components of the translation apparatus; and 4) Rpb7 is required for efficient translation initiation. Another report contrasts this with the following lines of evidence: 1) little, if any, free Rpb4/7 exists outside of the context of Pol II in TAP and FLAG tagged purification experiments; 2) a Rpb4-Rpb2 fusion protein can rescue the mRNA degradation and transcription defects resulting from *rpb4Δ (*Schulz *et al.,* 2014). This report has in turn been contradicted by another report that demonstrates the Rpb4-Rpb2 fusion protein: 1) binds mRNA in through the Rpb4 moiety; 2) becomes cleaved into free Rpb4 and Rpb2; and 3) cleaved Rpb4 in this context binds

mRNA in polysomes; and 4) the defect in mRNA degradation observed in strains lacking Rpb4 is *not* in fact significantly rescued by expression of the chimeric protein, arguing in favor of the functional importance of Rpb4 dissociation from Pol II (Duek *et al.,* 2018). Thus, whether dissociation of Rpb4/7 from Pol II has any *in vivo* significance remains highly controversial.

To summarize, Rpb4/7 has known roles in recruitment of 3' end processing factors, stimulating processive elongation, regulation of CTD phosphorylation levels, transcription initiation, gene-looping, Pol II recycling, transcription coupled repair, mRNA decay, and translation. How the function of Rpb4/7 itself is regulated, given such an extraordinary diversity of roles, remains an open question.

Transcription Through Chromatin

If you stretch out all of the DNA in a human nucleus it would measure approximately 1 meter. The compaction of DNA into a nucleus, while maintaining consistent genomic organization presents a major challenge to all eukaryotic cells. This problem is solved partially by wrapping DNA around nucleosomes, an octamer composed of two copies each of four unsique histone proteins (H2A, H2B, H3, H4). The nucleosome functions as the fundamental repeating unit of chromatin, with ~146bp of DNA wrapped around each nucleosome in *S. cerevisiae.* Nucleosomes are subject to a variety of post-translational modifications which modulate the strength of

nucleosome-DNA interactions as well as serving as signals for various chromatin remodelers (Bowman, Poirier 2015). The study of the PTMs that regulate genome structure and expression is known as epigenetics. The modulation of chromatin structure acts as one of the primary modes of regulating the expression of genetic material, serving to silence or activate genes depending on the needs of the cell. As such, the cell employs an extensive network of regulatory factors whose responsibility is the regulation and maintenance of the epigenome.

Nucleosomes act as a barrier to transcription by Pol II. In fact, nucleosomes inhibit Pol II transcription *in vitro* (Lorch *et al.,* 1987; Knezetic, Luse 1986) Passage of Pol II through nucleosomes requires breaking contact between nucleosomes and DNA, followed by reassembly of nucleosome-DNA contacts in the wake of transcribing Pol II. Recent cryo-EM studies including the elongation factors Spt4/5 and Elf1 have found that inclusion of these factors prevents stalling at SHL(-6) and SHL(-2) positions of the incoming nucleosome (Kujirai *et al.,* 2018). These structures showed that a structure is formed by Elf1, Spt4 and the NGN domain of Spt5 which is positioned between the incoming nucleosome and Pol II. In absence of these factors, the incoming nucleosomal DNA becomes trapped between the clamp head and lobe domain of Pol II, resulting in pausing. This demonstrates that Pol II progressively peels DNA from the nucleosome as it transcribes nucleosomal DNA and is dependent on TFIIS, which aids a paused or backtracked

polymerase in reinitiating transcription (Zawel *et al.,* 1995; Luse *et al.,* 2011) Further studies showed that a conserved and essential histone binding motif in the NGN domain of Spt5 is required for the preservation of chromatin structure over transcribed gene bodies, highlighting the mechanism of nucleosome retention in the wake of transcribing polymerase (Filipovski *et al.,* 2022).

Two other factors that have shown to be important for transcription through chromatin are the essential histone chaperones Spt6 and FACT (Spt16/Pob3). Spt6, like Spt5, has a highly acidic N-terminus (Swanson *et al.,* 1990). Spt6 is positioned adjacent to Spt5's KOW2-3 domains and directly contacts Rpb7 of the Pol II Stalk (Vos *et al.,* 2018b). In contrast, FACT does not directly contact the elongation complex, but was shown to bind histones in concert with the N-terminus of Spt5 (Farnung *et al.,* 2021). Both factors have shown to bind Spt5, bind nucleosomes *in vitro* and promote nucleosome retention over transcribed gene bodies (Vos *et al.,* 2018b; Bortvin, Winston 1996; Jamai *et al.,* 2009; Lindstrom *et al.,* 2003; Kato *et al.,* 2013; McCullough *et al.,* 2015). A model based on structural and biochemical studies proposes that FACT binds histones at downstream DNA and facilitates the transfer of nucleosomes to upstream DNA following Pol II passage.

Nucleosomes have an inhibitory effect on transcript initiation. One function of chromatin is to prevent spurious initiation from cryptic internal

promoters that are normally occluded by nucleosomes. Mutations in *spt5, spt6* and *spt16* have all been shown to disrupt chromatin over transcribed gene bodies resulting in initiation from a cryptic internal promoter found in the *FLO8* gene (Cheung *et al.,* 2008; Kaplan *et al.,* 2003). The Winston lab has produced a reporter gene (Cheung *et al.,* 2008) that leverages this phenomenon to identify mutations that disrupt chromatin structure over transcribed gene bodies. Mutations in many genes implicated in chromatin structure maintenance have been identified with this reporter that result in cryptic initiation and have been confirmed to disrupt chromatin structure with MNase assays (Kaplan *et al.,* 2003)

Assembly of the Messenger-Ribonuclear Protein

The process of preparing mRNA for export begins with cotranscriptional messenger ribonucleoprotein particle (mRNP) assembly. The assembly of the mRNP involves chemical modifications made directly to the RNA as well as the binding of tertiary factors that modulate stability and direct transport (Singh *et al.,* 2015). The two most well-known modifications are the 5'-methylguanosine cap and the 3' poly(A) tail. The 5' cap is a 7-methylated guanine nucleotide connected to the 5' end via a 5' to 5' triphosphate linkage. This modification is carried out by mRNA guanylyltransferase and mRNA (guanine-*N7*-)-methyltransferase, both of which bind directly to and display genetic interactions with Spt5. The 5' cap

structure serves to regulate mRNA export, prevent digestion by exonucleases and direct translation initiation.

Addition of the Poly(A) tail is directed by the Poly(A) (pA) signal which is encoded in the 3' end of all protein-coding genes and is coupled to termination. The poly(A) signal sequence is recognized by the CPF (cleavage and polyadenylation factor) complex, which then directs cleavage of the mRNA at the Poly(A) site and facilitates addition of the poly(A) tail via the poly(A) polymerase Pap1. Polyadenylation serves to stabilize the 3' ends of mRNA and protect from degradation. The Poly(A) tail also recruits Poly(A)-binding protein which regulates deadenylation and aids in mRNA export and translation.

Assembly of an export-competent mRNP is complex and dynamic, and is influenced by a number of factors. The THO/TREX complex is thought to be a central coordinator of mRNP assembly, facilitating loading of RNA binding proteins onto the nascent pre-mRNA while itself binding chromatin. THO in particular directly recruits the mRNA export factors Sub2 and Yra1 via the Hpr1 subunit. (Zenklusen *et al.,* 2002). Local chromatin dynamics has also recently been shown to play a direct role in modulating mRNP biogenesis through multiple mechanisms (Meinel, Staber 2015). In *Drosophila,* THO/TREX is recruited by chromatin-associated proteins that recognize H3K9me3 (Hur *et al.,* 2016). Additionally, the chromatin remodeler ISW1 (which is recruited by Spt6) has recently been shown to physically

interact with nascent mRNPs, prolonging their association with chromatin prior to export. Further, if mRNP assembly is defective, their interaction with ISW1 is prolonged, suggesting that ISW1 may act as an mRNP quality control factor (Yoh *et al.,* 2007).

Pre-mRNA secondary structure plays a significant role in determining the fate of the mRNPs. Myriad proteins associate with specific secondary structures and/or facilitate the formation of other secondary structures to modulate mRNA function and destination. Identification of the vast network of proteins that associate with and define mRNPs is an area of active study with significant work to be done. Recently, defects in the mRNP assembly machinery have been associated with various neurodegenerative diseases such as spinal muscular atrophy, amyotrophic lateral sclerosis, frontotemporal dementia, and Alzheimer's disease (Khalil *et al.,* 2018).

Transcript Termination by Pol II

The control of transcript termination by Pol II is governed by different complexes depending on the identity of the RNA being transcribed. Termination of mRNA requires polyadenylation by the CPF complex, which is coupled to termination by the CFI (Cleavage Factor I) complex. Non-coding RNA (ncRNA) termination, however, is governed by the Nrd1-Nab3-Sen1 (NNS) termination complex. Despite our knowledge of the factors involved in

various termination pathways, a comprehensive mechanism remains a topic of debate.

In polyadenylated transcripts, pre-mRNA cleavage is coupled to polyadenylation and is carried out by CPF and CFI (Christofori, Keller 1988). CPF is recruited by the Pol II body and binds to the poly(A) site, which stimulates pausing of the elongation complex prior to termination (Licatalosi *et al.,* 2002; Keller *et al.,* 1991; Enriquiz-Harris *et al.,* 1991; Orozco *et al.,* 2002). CFI, which is recruited by the CTD, then binds to a downstream GU-rich processing site (McCracken *et al.,* 1997; Barilla *et al.,* 2001; Cañadillas, Varani 2003). Following this event, CFI and CPF are released from the transcript in a manner dependent on mRNP assembly and cleavage is stimulated via a concerted effort of CPF and CFI subunits as Pol II transcription continues past the poly(A) site. (Ford, Hsu 1978; Qu *et al.,* 2009; Proudfoot 2011). Disassembly of Pol II is then triggered by the 5' to 3' exonuclease Rat1 and its stimulating partner Rai1 (Kim *et al.,* 2004).

There are two primary models proposed as mechanisms for Poly(A) dependent termination: the torpedo model and the allosteric model (Luo *et al.,* 2004; Epshtein *et al.,* 2007; Kim *et al.,* 2004). The torpedo model proposes a 5' to 3' RNA exonuclease (Rat1 in *S. cerevisiae)* enters at the unprotected 5' end of cleavage site and digests nascent RNA until it chases down the transcribing polymerase and collides with it, resulting in destruction of the transcription complex. Interestingly, Rat1 alone is not sufficient to dissociate a

paused polymerase *in vitro* (Dengl *et al,* 2009). The allosteric model proposes that recognition of the Poly(A) site results in an allosteric change in the transcription machinery, either by loss of an anti-termination factor or an intrinsic conformation change in the polymerase. This then results in disassembly of the transcription apparatus independent of the activity of a 5' to 3' exonuclease. The specific factors that are necessary and sufficient for promoting Pol II disassembly and the allosteric changes required for Pol II release have yet to be defined. It is likely that a combination of both models is required to explain the termination mechanism – allosteric changes likely cause the polymerase to slow down or pause, which both primes it for Rat1-mediated disassembly and allows time for Rat1 to catch up with the polymerase.

An alternative termination pathway exists for the termination of non-coding RNAs, such as snRNA, snoRNA, and Cryptic Unstable Transcripts (CUTs), the large majority of which are carried out by the NNS complex (Steinmetz *et al.,* 2001; Conrad *et al.,* 2000). snRNAs and snoRNAs are not subject to polyadenylation and are usually much shorter in length than polyadenylated mRNAs (50-200bp). CUTS do not have a defined function and are generally understood to be aberrant, non-coding transcripts (often anti-sense) that are targeted for immediate degradation. Interestingly, CUTS possess a poly(A) tail which is added by Trf4, the non-canonical poly(A)

polymerase component of the TRAMP polyadenylation complex (Wyers *et al.,* 2005).

The mechanism by which these transcripts are terminated is not well understood, but many of the major players have been defined. The NNS complex is composed of the two RNA binding proteins Nrd1 and Nab3 along with the DNA and RNA helicase Sen1 (Conrad *et al.,* 2001). Nrd1 is recruited by phospho-Ser5 at the Pol II CTD and recognizes specific RNA motifs in snRNAs, snoRNAs and CUTS (Vasiljeva *et al.,* 2008; Hobor *et al.,* 2011; Carroll *et al.,* 2004; Bacikova *et al.,* 2014). Nrd1 and TRAMP together stimulate the exosome, which simultaneously cleaves and processes the 3' ends of noncoding transcripts (Vasiljeva, Buratowski 2006). Termination is brought about by Sen1 helicase, which is able to induce displacement of stalled Pol II complexes through its helicase activity, forming an intermediate complex with stalled ECs (Hazelbaker *et al.,* 2013; Wang *et al.,* 2019). While the primary factors involved in NNS-dependent termination have been identified, a step-by-step model remains elusive.

Mechanisms of mRNA Export

Packaging of the mRNP for export occurs cotranscriptionally. The THO/TREX complex is one of the major players involved in cotranscriptional mRNP assembly, facilitating loading of mRNA binding proteins that ultimately prepare the mRNP for competent export (Katahira 2012; Strässer *et al.,*

2002). THO is composed of Tho2, Hpr1, Mft1 and Thp2, while TREX is composed of all of the above as well as Yra1 and Sub2. Deletions of TREX components results in nuclear accumulation of mRNA, consistent with an mRNA export defect. Hpr1 is required for recruitment of Sub2 and Yra1, linking THO/TREX to mRNA export (Zenklussen *et al.,* 2002).

After a complete mRNP is assembled, it must be exported out of the nucleus and into the cytoplasm for translation. In yeast, the heterodimeric export receptor complex Mex67-Mtr2 is responsible for the export of virtually all mRNAs (Santos-Rosa *et al.,* 1998). In contrast, All other proteins and RNAs rely on importin/karyopherin-β-type transport receptors. Mex67 contains an RNA recognition motif (RRM) and directly binds mRNA without sequence specificity and also binds directly to nucleoporins, enabling transport out of the nucleus through the nuclear pore complex (NPC) (Strässer *et al.,* 2000). Directionality of export is imposed by two DEAD-box ATPases – Sub2 at the nuclear side and Dbp5 at the cytosolic side (Tran *et al.,* 2007; Xie, Ren 2019). Sub2 facilitates the loading of mRNA onto Mex67 through an adaptor protein, Yra1, while Dbp5 facilitates dissociation of mRNA from Mex67 upon delivery to the cytoplasm (Strässer 2001; Tieg, Krebber 2012).

Overlapping function of Spt5 and Rpb4/7

Recent cryo-EM studies have elucidated the precise location of Rpb4/7 subunits in complex with Spt5 and Pol II (Bernecky *et al.* 2017). Interestingly, Spt5 domains KOW2, KOW3, and KOW4 all apparently reside in direct contact with the base of the Pol II Stalk. These contacts are supported by prior cross-linking studies (Li *et al.,* 2014). Additionally, Rpb4/7 and Spt5 domains KOW4-5 are placed in close proximity to the mRNA exit channel, suggestive of roles in mRNA processing. The structural arrangement of these domains of Spt5 with Rpb4/7 imply a functional relationship. Indeed, Spt5 and Rpb4/7 appear to share several common functions, as Spt5 and Rpb4/7 have both been previously shown to be involved in allosteric polymerase stabilization, 3' end processing and mRNA export (Bernecky *et al.,* 2017, Armache *et al.,* 2003; Runner *et al.,* 2008; Mayer *et al.,* 2012; Farago *et al.,* 2003; Burckin *et al.,* 2005).

One of the first functions identified for NusG/Spt5 was its ability to promote transcription elongation and processivity (Liu, Steitz 2016; Bernecky *et al.,* 2016; Bernecky *et al.,* 2017; Martinez-Rucobo *et al.,* 2011; Burova *et al.,* 1995). This occurs most likely through allosteric effects of the NGN and KOW1 domain, which lie over the central cleft and contact the non-template strand of DNA. In archaea, Spt5 and RpoF/E have also been shown to stimulate processivity in elongating polymerases (Hirtreiter *et al.,* 2010a; Hirtreiter *et al.,* 2010b). Interestingly, despite the lack of KOW domains 2-5 in *archaea,* archaeal Spt5 physically contacts RpoE, the archaeal orthologue of

Rpb7, suggesting that a physical and functional interaction of Spt5 and the Pol II Stalk is conserved from *archaea* to *eukarya* (Zhou *et al.,* 2009). In *S. cerevisiae,* Rpb7 wedges itself into the foot domain of Pol II, promoting a closed conformation of the clamp region, and cells lacking *rpb4,* polymerase density decreases at the 3' end of genes, suggesting the allosteric stabilization and processivity enhancement features are conserved across kingdoms (Armache *et al.,* 2003; Runner *et al.,* 2008).

Both Spt5 and Rpb4/7 have been implicated in 3' end processing and polyadenylation site choice. The protruding tip of Rpb7 is required for binding of the Nrd1 homolog Seb1 in *S. pombe* (Mitsuzawa *et al.,* 2003), and this region is spatially adjacent to Spt5 KOW2. Defects in Spt5 result in enhancement of upstream poly(A) site usage at the *lacZ, ADH2, RNA14* genes (Cui *et al.,* 2003). At the *RNA15* gene, loss of *rpb4* results in altered poly(A) site usage as well (Runner *et al.,* 2008). Further, both Spt5 and Rpb4/7 have been shown to play a role in directly recruiting 3' processing factors, suggesting that these factors influence when to initiate polyadenylation by directly promoting 3' end processing factors (Mayer *et al.,* 2012; Runner *et al.,* 2008). Together, this implicates both Spt5 and the Pol II stalk in recruitment of 3' processing factors in both CPF/CF1 and NNS termination pathways

Loss of Spt5 function has been shown to suppress the mRNA export defect caused by the temperature sensitive *mex67-6* allele (Burckin *et al.,*

2005). Rpb4 has been shown to have a stress-specific role in mRNA export, and more specifically, heat shock genes, at elevated temperatures (Farago *et al.,* 2003). A more controversial study has directly implicated Rpb4/7 as physically accompanying mRNA from transcription complexes outside of the nucleus, linking them to translation (Harel-Sharvit *et al.,* 2010). Thus, both Spt5 and Rpb4/7 have been shown to play roles in mRNA export.

Despite a wealth of evidence suggesting that Spt5 and Rpb4/7 participate in overlapping pathways, a direct functional relationship has yet to be defined. These factors appear to form a cooperative structure that may serve to recruit other factors an elongating (or terminating) polymerase. Whether this structure directly supports any *in vivo* function remains to be tested, along with the consequences of disruption of these interactions. The binding of factors to this structure may be dependent on where the polymerase is in the transcription cycle.

R-loops

R-loops, which were first observed *in vitro,* are genomic structures composed of DNA-RNA hybrids formed by nascent mRNA annealing with upstream DNA, displacing the non-template strand and forming a loop-like structure (Thomas *et al.,* 1976). R-loops were previously thought to be artefactual and of no consequence to biological processes. This was challenged by observations in bacteria suggesting that R-loops occurs *in vivo*

and are in fact deleterious (Drolet *et al.,* 1995). R-loops were later revealed to play a positive role in antibody diversification, further complicating our understanding (Chaudhuri *et al.,* 2003). More recent evidence, however, has shown that R-loops are widespread across the genome, and are likely a carefully regulated feature necessary for proper genome maintenance and gene expression. In fact, up to 8% of the yeast genome (and 5% in mammals) is occupied by R-loops at any given time (Chan *et al.,* 2014; Wahba *et al.,* 2016; Sanz *et al.,* 2016). Recent advances R-loop biology have demonstrated R-loops are associated with genome instability, particular chromatin states, poly(A) signals, transposable elements such as Ty1, telomeres and highly expressed genes such as tRNA and rRNA (Sanz *et al.,* 2016; Graf *et al.,* 2017; Zeng *et al.,* 2021; El Hage *et al.,* 2014; Huertas, Aguilera 2003; Pfeiffer *et al.,* 2013). It has been proposed that R-loop induced stalling of Pol II may serve as a trigger for transcript termination (Cristini *et al.,* 2018). On the other hand, R-loops also have the ability to stimulate genome instability (Lin *et al.,* 2010; Huertas, Aguilera 2003). R-loops displace the non-template strand of DNA, exposing the ssDNA to endonucleases and deaminases. This can be a source for DNA nicks, single-stranded and double-stranded DNA breaks. R-loops often form at T-R (Transcription-Replication) collisions, resulting in DNA breaks (Hamperi *et al.,* 2017). It is therefore critically important for the cell to establish proper homeostasis of R-loops to maintain proper gene expression and genome integrity.

R-loops have been shown to display a chromatin signature that is similar to that of chromatin found at promoters and the TSS. This includes the markers H3K4me1, H3K4me3, H3K27ac and H3K36me3 (Sanz *et al.,* 2016; Chen *et al.,* 2015). Depending on the context, R-loops can induce both chromatin decondensation and heterochromatin assembly (Castellano-Pozo *et al.,* 2013, Powell *et al.,* 2013). What the specific triggers are for the interplay between R-loops and chromatin structure remain an open question.

There are many factors that have been shown to have roles in R-loop resolution – typically RNases and RNA helicases. The DNA/RNA helicase Sen1 has been shown to play a role in replication-dependent R-loop resolution at T-R collisions, while components of the THO complex have been shown to prevent transcription-dependent R-loop resolution (Appanah *et al.,* 2020; Gómez-Gonzáles *et al.,* 2011). Loss of Hpr1, a component of the THO complex, results in a genome wide increase in R-loops as well as R-loop associated genome instability (San Martin-Alonso *et al.,* 2021). In general, THO is thought of as a general safeguard against aberrant R-loop formation. FACT mutants have also been shown to upregulate R-loop formation (Herrera-Moyano *et al.,* 2014.

Some of these R-loop associated factors have been recently suggested to be involved in human diseases. Mutations in Senataxin, the human ortholog of Sen1, result in the diseases ataxia with oculomotor apraxia 2 (AOA-2) and amyotrophic lateral sclerosis (ALS4) (Moreira *et al.,* 2004;

Chen *et al.,* 2004). Changes in the levels of human RNase HII have recently been linked to psoriasis (Mehmetbeyoglu *et al.,* 2022). Any defect in R-loop associated factors can promote genome instability, which can lead to an increased risk for cancer (Brambati *et al.,* 2020). Many genetic diseases are a result of gene-specific repeat expansions, including Huntington's disease, myotonic dystrophy type 1, spinocerebellar ataxia type 1, fragile X syndrome, and Friedreich ataxia (Ellerby *et al.,* 2020; Paulson 2018). These repeat expansions have also been associated with increased R-loops (Freudenreich 2018). The well-known *BRCA1* and *BRCA2* genes, mutations in which often lead to breast and ovarian cancers, have been shown to prevent promiscuous R-loop formation (Hatchi *et al.,* 2015; Bhatia *et al.,* 2014). It is therefore of critical importance to understand the formation and regulation of R-loops and their context-specific functions. Increasing our understanding of R-loop biology may lead to the discovery of novel interventions that can have a tremendous impact on our ability to treat genetic diseases.

Summary

In the following chapters we will explore the genetic, physical, and functional relationship between Spt5's central KOW domains and the Pol II subunits Rpb4 and Rpb7. We present an argument that, in the contexts of transcription elongation and termination, Spt5 KOW domains 2-4 and the Pol II Stalk form a functionally relevant substructure within the EC that acts as a platform for protein binding throughout transcription in a context dependent

manner. In Chapter 2, we take a genetic approach to explore the mutant phenotypes in this region and how they relate to cotranscriptional chromatin remodeling and polyadenylation site choice, revealing genetic evidence that Rpb4/7 and Spt5 act synergistically. In Chapter 3, we will explore biochemical interactions between KOW2-3 and Linker2-KOW4 with tertiary factors involved in myriad cotranscriptional processes and how they relate to known functions of the Pol II Stalk, with a focus on proteins that have already been shown to physically interact with Rpb4 or Rpb7. Chapter 4 will explore a role for Spt5 and Rpb7 in maintaining R-loop homeostasis, possibly in concert with the NNS complex. Chapter 5 will summarize our current models of Spt5 and Pol II Stalk function in transcription and address future directions.

Chapter 2

Genetic evidence for a functional cooperation between *SPT5* and *RPB4/7* throughout transcription elongation and termination

Introduction

Structural models generated from cryo-EM experiments have placed the central KOW domains (KOW2/3, Linker2-KOW4) of Spt5 and the Pol II subunits Rpb4/Rpb7 directly adjacent each other in apparent contact (Bernecky *et al.,* 2017). Spt5 and the Pol II stalk have been studied individually fairly extensively, and many overlapping functions have been identified. However, the functional relevance of the structure formed by these factors has yet to be addressed.

In order to test the hypothesis that these factors form a structure with a unified function throughout transcription, we first took a genetic approach. We employed a set of genetic reporters that provide insight into both polyadenylation site choice and the production of cryptic-intragenic transcripts that likely result from the disruption of chromatin. We used a combination of known mutations identified in our lab and others, as well as mutations generated from random mutagenesis experiments and tested them for activity with our genetic reporters. Many of the identified mutations shared overlapping phenotypes and genetically interacted with each other, resulting in synthetic sickness and enhancement of phenotypes.

The results from this genetic analysis, when viewed in light of the published cryo-EM structures, are highly suggestive of a cooperative, functional platform formed by KOW2-4 of Spt5 and Rpb4/7, and the disruption of this platform results in transcriptional defects. Further, these results suggest that Rpb4/7 likely aid in cotranscriptional chromatin remodeling, the first such evidence of core polymerase subunits playing a role in maintaining chromatin structure. This suggests that the structure of the polymerase itself has evolved to overcome the challenge of transcription through chromatin in conjunction with assistance from tertiary factors.

Results

Cryptic intragenic transcripts and Poly(A) site choice defects result from mutations of Spt5's central KOW domains near the Pol II stalk region

Spt4/Spt5 is implicated in regulation of transcription elongation and co-transcriptional processes, such as pre-mRNA capping, splicing, 3' end formation and chromatin dynamics (Li *et al.,* 2014; Maudlin *et al.,* 2019; Cui *et al.,* 2003; Lindstrom *et al.,* 2003; Mayer *et al.,* 2012; Evrin *et al.,* 2022). Biochemical and structural studies show that the NGN and KOW1 domains seal the central cleft of Pol II, associate with both non-template strand DNA and exiting DNA duplex (Meyer *et al.,* 2015; Bernecky *et al.,* 2017; Ehara *et al.,* 2017). The interactions are likely involved in Spt5's roles in promoting Pol

II processivity and elongation rate. In contrast, KOW domains 2-5 lie on the exterior surface of Pol II, far from the DNA template or the catalytic core of Pol II. We hypothesize that these domains allow Spt5 to coordinate transcription elongation with co-transcriptional processes. To test this proposal, we took a genetic approach to investigate the roles of KOW2-5 in gene expression.

Since the function of KOW2-5 is not known, we decided to screen for new *spt5* mutations that fall within these domains. Our approach was to screen for *spt5* mutations that cause transcription from cryptic intragenic promoters. It is thought that these cryptic promoters are normally occluded and transcriptionally repressed by nucleosomes. Perturbation of chromatin relieves this repression, allowing transcription to initiate from the cryptic promoters. To identify mutations that cause this phenotype, we made use of a genetic reporter of cryptic initiation developed in the Winston lab (Fig. 2-1C; Cheung *et al.,* 2008). This reporter is integrated at the *FLO8* gene in a yeast strain with a deletion of the normal HIS3 gene. The reporter consists of *GAL1* promoter driving transcription of a *FLO8::HIS3* fusion gene, in which *HIS3* sits just downstream of a well-characterized *FLO8* cryptic promoter. Transcription initiation from the *GAL1* promoter results in a *FLO8::HIS3* transcript in which *HIS3* is out of frame and not translated. In contrast, transcription initiation from the cryptic internal promoter results in an in-frame *HIS3* transcript, allowing for His3 expression and growth on –His media. We randomly

mutagenized *SPT5* using hydroxylamine and screened for the cryptic transcription phenotype. This screen yielded many mutations in the region immediately 5' to the NGN domain and 3' to the acidic domain – a region we named the NPD (NGN Proximal Domain). This region includes the previously identified *spt5-Cs* allele, *spt5-242*, which also results in cryptic transcription in addition to sensitivity to cold and general growth defects. Because our purpose here is to investigate KOW domains 2-5, these mutations were set aside for future analysis. Our screen also produced three mutations in the KOW2-KOW3 region (Fig. 2-1B). These *spt5* mutations result in the amino acid changes E546K, G587D, and G602S (this mutant also contains S809F, which lies in KOW5), and all three are solvent exposed. Since the function of Spt5's central KOW domains remain unresolved, we decided to focus our efforts on characterizing these mutations for this study.

Prior experiments show that Spt5's central KOW domains fold into stable structures separated by unstructured linker domains (Meyer *et al.,* 2015). Bernecky *et al*. recently solved the structure of human Pol II in complex with DSIF, resolving the location of the KOW domains with respect to individual Pol II subunits (Fig. 2-1A; Bernecky *et al.,* 2017). Spt5 KOW domains 2-4 wrap extensively around the Pol II stalk composed of subunits Rpb4 and Rpb7, and these contacts have been confirmed via cross-linking studies. Spt5 KOW2-3, Rpb4 and Rpb7 appear to form a tripod-like structure that protrudes from the body of Pol II. Interestingly, our three KOW2-3

mutants lie in conspicuous, solvent exposed locations relative to the Pol II Stalk (Fig. 2-1D). Spt5 residues G587 and G602 pack closely at the juncture of Rpb7 and KOW3, while Spt5 E546, whose charge is reversed by the E546K mutation, lies at the tip of KOW2.

Previous studies indicate a role for Rpb4/7 in pre-mRNA processing, and KOW4-5 bridge the mRNA exit channel, which suggest that the central KOW domains may also have a role in pre-mRNA processing (Mitsuzawa *et al.,* 2003; Runner *et al.,* 2008; Bernecky *et al.,* 2017). Additionally, the location and chemical nature of the residues identified in our cryptic initiation screen suggest a functional relationship between this region of Spt5 and the Pol II Stalk. To begin to address KOW2-3's involvement in pre-mRNA processing, we asked if mutations that lie in this region of Spt5 alter polyadenylation site choice, as previous Rpb7 mutations have also been shown to alter polyadenylation site choice. We utilized a reporter generated by Kaplan et al. called *gal10Δ56,* which is based on the *gal10Δ55* construct developed by Greger and Proudfoot (Kaplan *et al.,* 2005; Greger, Proudfoot 1998). This reporter contains a 56bp deletion at the polyadenylation site of the *GAL10* gene, which results in the accumulation of a toxic intermediate that kills otherwise wildtype cells grown on galactose media. Northern blot analysis has shown that this deletion results in a bicistronic transcript containing both *GAL10* and *GAL7,* and reduces expression of *GAL7* from its canonical promoter. Mutations that perturb polyadenylation result in

termination at the truncated polyadenylation site and restoration of growth on galactose media (Fig. 2-2). All three point mutations in KOW2-3 resulted in suppression of *gal10Δ56*, as did a deletion of KOW4 and, to a lesser degree, KOW5. To confirm that this functionality is specific to the KOW domains we also tested mutations in the N and C terminal regions of Spt5 which did not suppress *gal10Δ56*, with the exception of *spt5-194*, which has previously been shown to destabilize Spt5's overall structure (Ding *et al.,* 2010).

Spt5's KOW2-3 region and Rpb4/7 share common mutant phenotypes

Since mutations affecting Spt5's central KOW domains share phenotypes with *rpb4* and *rpb7* mutations, we next asked if the reverse was true as well. As *RPB7* is an essential gene, we mutagenized plasmid-borne Rpb7 and screened for cryptic initiation and for suppression of *gal10Δ55*. Two mutations resulted in strong suppression of *gal10Δ55 – rpb7-G149*D, and rpb7-E100K. Rpb7-G149 lies at the juncture of Rpb7 and KOW3 and sits near Spt5-G587, a residue that resulted in cryptic initiation in our initial Spt5 cryptic initiation screen, which appears to contact KOW4 (Fig. 2-3A, B). We did not identify rpb7 alleles that caused cryptic initiation in this screen. However, we tested a handful of alleles generated in Craig Kaplan's lab shown to exhibit transcriptional defects, one of which, *rpb7-D166G*, did demonstrate strong cryptic initiation in the context of our reporter. Similar to Spt5-E546, Rpb7-D166 is a solvent exposed residue that projects away from the polymerase.

Rpb7-D166 was also shown to result in TSS defects as well as sensitivity to mycophenolic acid, suggestive of defects in transcription initiation. Furthermore, rpb7-L168S, which lies spatially adjacent to rpb7-D166G, exhibited minor cryptic initiation. While rpb7-V101E did not suppress *gal10Δ55* or show cryptic initiation, it did show minor sensitivity to MPA, as well as a slight growth defect. This difference suggests that the cryptic initiation phenotypes seen in rpb7-D166G and L168S are allele specific, rather than resulting from general loss of function.

To assess the evolutionary and functional significance of these residues we generated sequence alignments across model organisms and asked if these residues were highly conserved (Fig. 2-3C). spt5-G602 is universally conserved. spt5-G587 is also universally conserved, as is rpb7-G149, the residue most closely contacting spt5-G587. This suggests that this structure is highly evolutionary relevant. Spt5-E546 appears only as either Glutamate or Aspartate, as does rpb7-D166, indicating preservation of the acidic nature of these residues is universal. rpb7-L168 is also universally conserved highlighting the evolutionary importance of this structure formed by the C terminus of rpb7 and Spt5 KOW2.

Since *RPB4* is not essential, we deleted *RPB4* in strains carrying the cryptic initiation and *gal10Δ56* reporters and observed growth on galactose media with *gal10Δ56* as well as growth in the context of our cryptic initiation

reporter (Fig. 2-3D, E). Loss of *RPB4* in both contexts results suppression of *gal10Δ56* and cryptic initiation. These experiments were performed in the context of overexpressed Rpb7 due to poor growth caused by *rpb4Δ,* but the phenotype persists without Rpb7 overexpression, albeit to a lesser degree (data not shown).

Spt5 KOW2-3 and Rpb4/7 contribute to a common function required for viability

To further assess the functional cooperation of Spt5 KOW2-3 with the Pol II stalk, we generated double mutants of spt5-E546K, spt5-G587D with our previously identified *rpb7* mutants (Fig. 2-4A, B). Since Spt5-G602S is also paired with S809F, we excluded that mutant from this analysis. When paired with *spt5-E546K*, *rpb7-G149D* and *rpb7-E100K* exhibited synthetic lethality, while *rpb7-V101E* exhibited severe growth defects and *rpb7-D166G* exhibited a minor growth defect. When paired with *spt5-G587D*, *rpb7-G149D* and *rpb7-E100K* grew nearly as well as the single *spt5* mutant. *rpb7-D166G* and *rpb7-V101E* exhibited minor growth defects, with *rpb7-V101E* growing slightly less well than rpb7-D166G.

We next tested our *spt5* / *rpb7* double mutants for enhancement of the phenotypes tested for in Figure 2-3. Spt5-G587D was initially identified in our cryptic initiation mutant hunt and displayed a modest cryptic initiation phenotype when plasmid-borne (data not shown). When integrated, however,

this phenotype was reduced significantly. We combined this allele with the cryptic initiation reporter with our *rpb7* mutant alleles and observed strong cryptic initiation in rpb7-G149D, E100K, and D166K (Fig. 2-4A). Interestingly, the cryptic initiation phenotype observed in the rpb7-D166G allele was somewhat suppressed. Additionally, we observed synthetic lethality with spt5-G587D when coupled with rpb7-D166G and V101E on 25ug/mL MPA, and strong synthetic growth defects with rpb7-E100K and G149D. Rpb7-V101E was lethal with spt5-G587D at 39C, and Rpb7-G149D and D166G showed severe growth defects.

Spt5-E546K displays a strong phenotype with the *gal10Δ56* reporter. This phenotype persists when paired with rpb7-D166G and D166K, and is perhaps enhanced to a modest degree. This phenotype is suppressed however when paired with rpb7-L168S. spt5-E546K is nearly as resistant to MPA as wildtype on its own, but when paired with rpb7-D166G and V101E it is MPA-sensitive.

We also tested spt5-G602S+S809F, spt5-E546K, and spt5-G587D against an *rpb4* deletion (Fig. 2-4C). All mutants tested were lethal in combination with *rpb4Δ,* which on its own exhibits growth defects. Together this indicates a significant functional cooperation between the Pol II stalk and Spt5, and provides genetic evidence for the significance of the structure formed by these proteins.

Discussion

Prior evidence has shown loss of function in Spt5 results in the generation of cryptic transcripts from an internal promoter within the *FLO8* gene (Cheung *et al.,* 2008). Previous studies of Rpb4/7 implicate the Pol II stalk in polyadenylation site choice (Runner *et al.,* 2003). We asked if the reciprocal phenotypes could also be seen by utilizing genetic reporters of *spt5* and *rpb4/7* mutants. Indeed, we saw that disruption of the putative structure formed by these domains and the Pol II stalk resulted in overlapping phenotypes. Mutations that appear to disrupt either the juncture of Rpb7 and KOW2-4 or the tip of the structure formed by these factors resulted in altered polyadenylation site choice and cryptic initiation.

There is a wealth of evidence implicating a large number of genes in cryptic transcription – many of these directly involved in chromatin structure maintenance, including *SPT6, SPT16, HHT1, HHT2, HTA1-HTB1, CHD1* (Cheung *et al.,* 2008; Quan, Hartzog 2009). MNase studies show that chromatin structure is indeed disrupted in these mutants, strongly suggesting that other genes that generate cryptic transcripts in the context of this reporter disrupt chromatin as well (Kaplan *et al.,* 2003). The surface of the Stalk/KOW structure is highly acidic in nature, similar to Spt5's acidic NGN domain, which has recently been shown to bind nucleosomes and also displays cryptic initiation when mutated (Cheung *et al.,* 2008; Evrin *et al.,* 2022). Spt6, which has been shown to bind histones and nucleosomes *in vitro*, appears to

directly contact Rpb4/7 on the other side of KOW2-3 according to structural models (Bortvin, Winston 1996; McDonald *et al.,* 2010; Vos *et al.,* 2018b). The placement of Spt6 appears to expand the large acidic platform created by the stalk and KOW domains. It is possible that Rpb4/7 and KOW2-3 play a minor role in nucleosome rearrangement, serving to transiently pass back nucleosomes as they are rearranged in the polymerases wake, as assisted by Spt6.

The *gal10Δ56* reporter has previously been utilized to detect alterations in polyadenylation site choice among resulting from mutating either 3' end processing factors or elongation factors (Kaplan *et al.,* 2005). The primary mode of Galactose toxicity appears to result from transcriptional interference at the *GAL7* promoter from polymerases that fail to terminate at the *gal10* poly(A) site. Mutations that affect 3' end processing factors or perturb elongating Pol II allow for recognition of the truncated poly(A) site, restoring initiation at *GAL7.*

Mutations *in spt6* and other elongation factors have also been shown to suppress *gal10Δ56* by decreasing read through and allowing re-initiation at *GAL7 (*Kaplan *et al.,* 2005, Cui Denis 2003). One model to explain this phenotype is that defects in transcription elongation result in a slower, more pause-prone polymerase, allowing for an increase in processing time at a weakened poly(A) site. This model, however, is disfavored by the observation

that the upstream poly(A) site is equally favored over downstream alternative poly(A) sites. For this phenotype to be strictly explained by a slower elongating Pol II, one would expect a gradient of decreasing poly(A) site usage following the initial Poly(A) site. This suggests an alternative model: suppression of *gal10Δ56* by perturbing elongation factors is caused by an alteration in either conformation or composition of the elongation complex. Supporting the notion that a compositional alteration is the cause of *gal10Δ56* suppression is the observation that the density of Ctr9, a member of the PAF complex, decreases following the poly(A) signal in wildtype *GAL10* as seen by ChIP experiments (Kaplan *et al.,* 2005). Increased occupancy of wild type Spt6 and Ctr9 over the poly(A) site is observed in the *gal10Δ56* mutant, suggesting that the presence of these factors promote read-through, and the removal of these factors, particularly Ctr9, follows recognition of a functional poly(A) site. One appealing model is that disruption of the Stalk/KOW platform indirectly promotes ejection of certain factors such as Ctr9, allowing for polyadenylation at the truncated site. It seems likely that the Pol II EC undergoes a significant structural rearrangement as a prerequisite to polyadenylation and termination. The mutations presented in this study may facilitate this rearrangement by causing a preference for a particular conformation that makes the Pol II EC more amenable to termination.

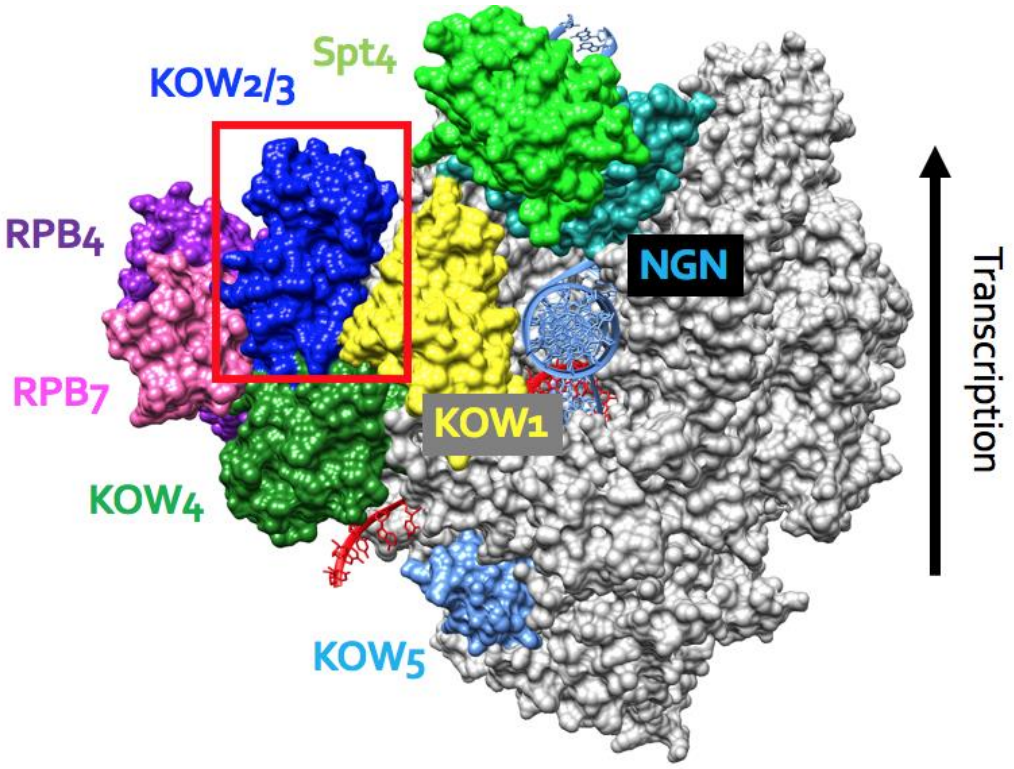
To further confirm a functional cooperation between KOW2-4 and Rpb4/7, we performed crosses to generate double mutants that further disrupt

the binding platform formed by Spt5 and Rpb4/7. Indeed, we saw that further insults to this region resulted in an enhancement in phenotype severity, synthetic growth defects, and even synthetic lethality. The positioning of E546 in appears to be particularly critical for the function of elongating polymerase, as mutations that alter this residue in in *spt5* combined with mutations that appear to alter KOW2/3 or Rpb4/7 position result in synthetic lethality. Taken together, it appears that the Pol II stalk region, including Spt5 KOW domains 2-4, is involved in both cotranscriptional chromatin structure maintenance and polyadenylation site choice.

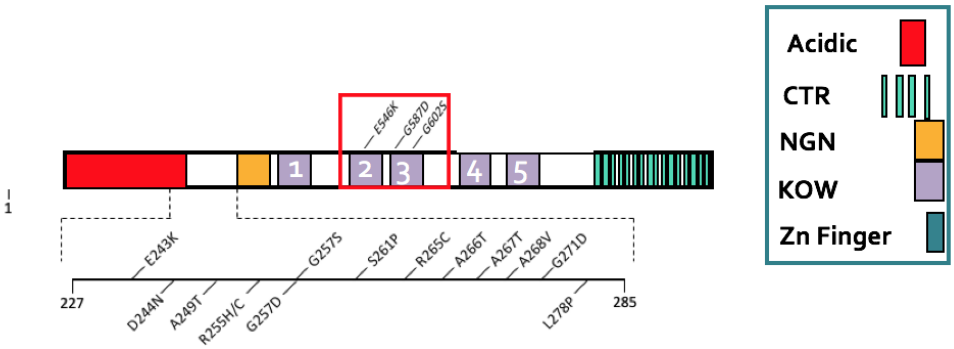
Figure 2-1: *spt5* KOW mutants near Rpb4/7 display cryptic transcription

A) Model of *Bos taurus* Pol II, indicated in grey, with DSIF (human), constructed from cryo-EM data published by Bernecky et al., 2017 (PDB: 5OIK). DSIF domains are indicated by color coded labels. The red box indicates the location of the residues altered in the cryptic initiation screen, shown in 1B.
B) Schematic of Spt5 domains and the residues identified in the cryptic initiation screen. The red box corresponds to the highlighted domains in figure 2-1A.
C) Diagram of cryptic initiation reporter. Red balls are nucleosomes. The blue, grey and green circles are general transcription factors. Red arrow indicates the corresponding mRNA product generated from transcription.
D) A zoomed in view of the structure showed in 1A, focusing on KOW2-3 and Rpb4-Rpb7. KOW domains shown in ribbon. The residues altered in the cryptic initiation screen are indicated as red/orange spheres and labeled.

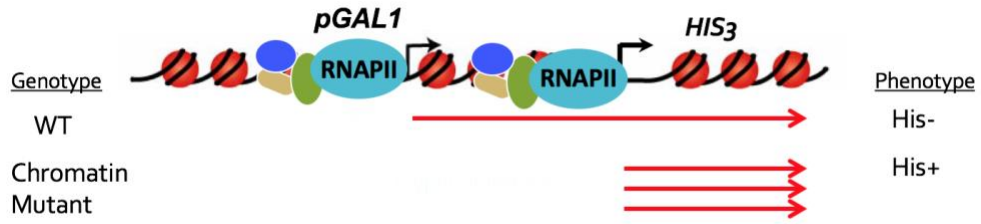
A



B


Acidic
CTR
NGN
KOW
Zn Finger
1
1
2
3
4
5
E546K
G587D
G602S
227
285
E243K
D244N
A249T
R255H/C
G257D
G257S
S261P
R265C
A266T
A267T
A268V
G271D
L278P

C


pGAL1
HIS3
RNAPII
RNAPII
Genotype
WT
Chromatin
Mutant
Phenotype
His-
His+

D

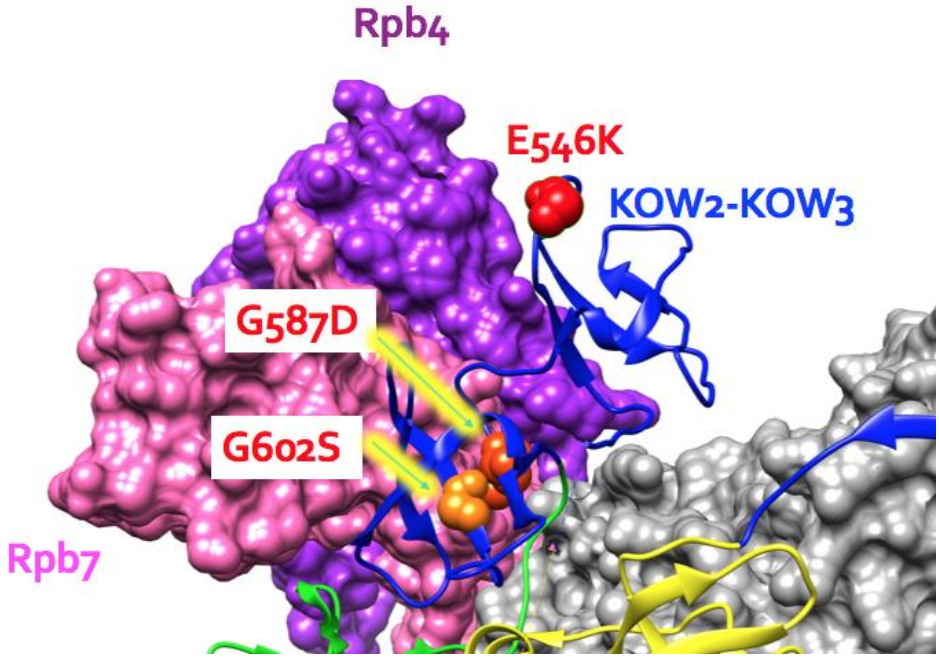


Figure 2-2: Mutations in *spt5* near Rpb4/7 display defects in Poly(A) site choice

On the left is a schematic for the *gal10Δ56* reporter construct. Arrows represent the mRNA product produced while under the presence of absence of galactose as the sole carbon source. *SPT5* shuffle strains (*spt5Δ gal10Δ56* pMS4) were generated and the indicated plasmid born *spt5* mutant was transformed and wild type *SPT5* (pMS4) was counter-selected against 5FOA media prior to dilution spotting onto YPD and YPGal at 30C to measure expression of the *gal10Δ56* reporter.

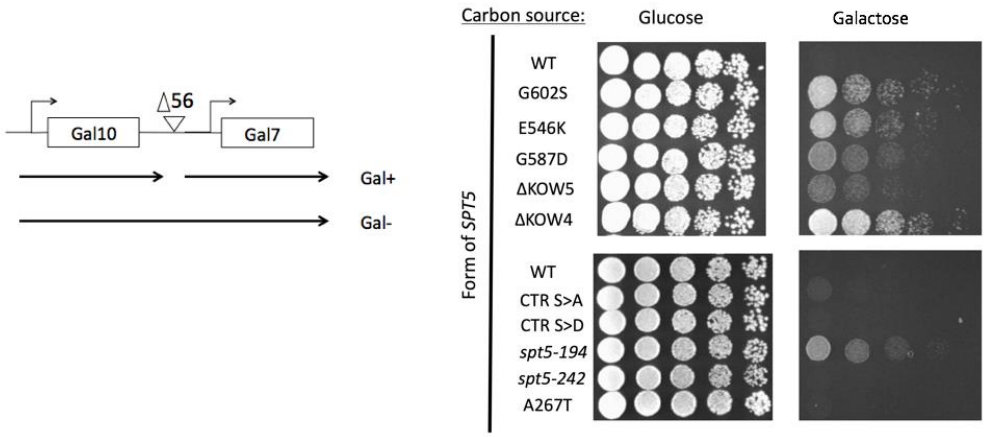
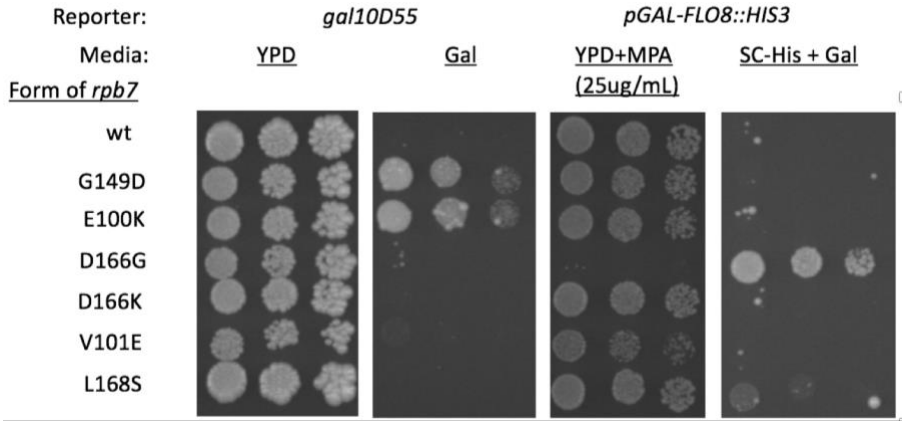

Δ56
Gal10
Gal7
Gal+
Gal-
Carbon source:
Glucose
Galactose
Form of SPT5
WT
G602S
E546K
G587D
ΔKOW5
ΔKOW4
WT
CTR S>A
CTR S>D
spt5-194
spt5-242
A267T

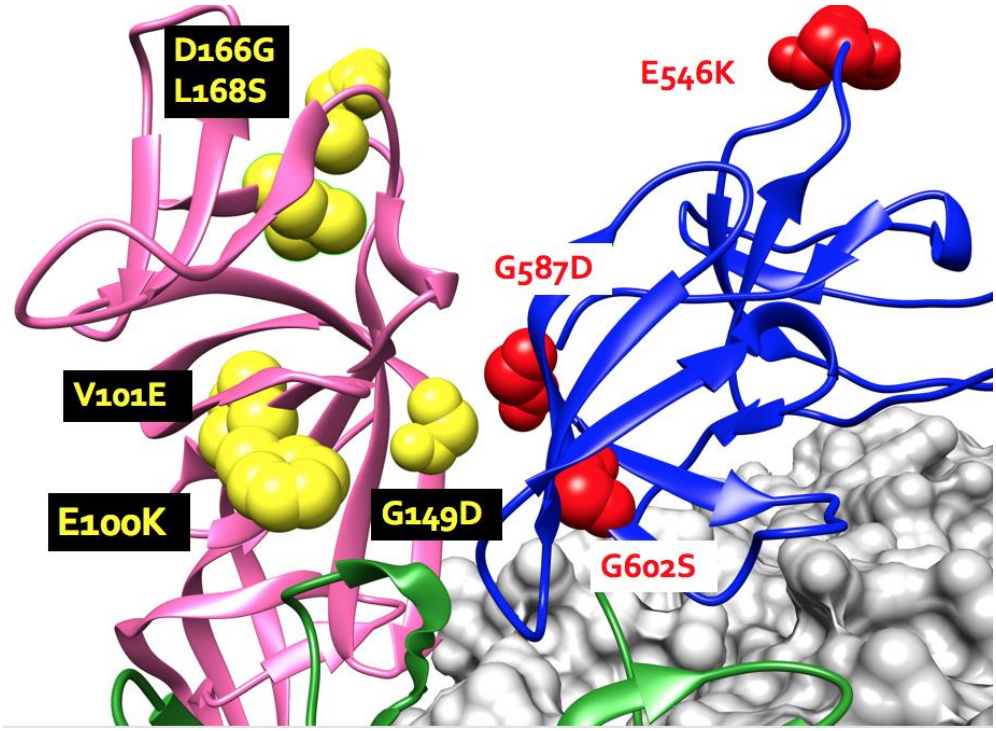
Figure 2-3: *rpb4* and *rpb7* mutations result in transcription defects

A) *RPB7* shuffle strains were generated in either *gal10Δ55* or *pGAL-FLO8::HIS3* background and the indicated *rpb7* allele was transformed and wild type Rpb7 was counter-selected against on 5FOA prior to dilution spotting on the indicated media at 30C.
B) Rpb7 is shown in pink ribbon with indicated amino acid changes as yellow spheres. KOW2-3 is shown in blue ribbon with indicated amino acid changes as red spheres. This image is a close-up generated from the structure shown in figure 2-1A.
C) A multi sequence alignment generated from NCBI BLAST. Red highlighted residues indicate altered conserved amino acids in Spt5 and Rpb7 that lie spatially adjacent at the KOW3 juncture. Blue highlighted residues indicate altered conserved amino acids that are surface exposed at the tip of Spt5 KOW2 and Rpb7.
D) *rpb4Δ* cells in the *pGAL-FLO8::HIS3* background were patched onto YPD, replica plated onto the indicated media and allowed to grow at 30C for 3 days.
E) *rpb4Δ* cells in the *gal10Δ56* background were patched onto YPD, replica plated onto the indicated media and allowed to grow at 30C for 3 days

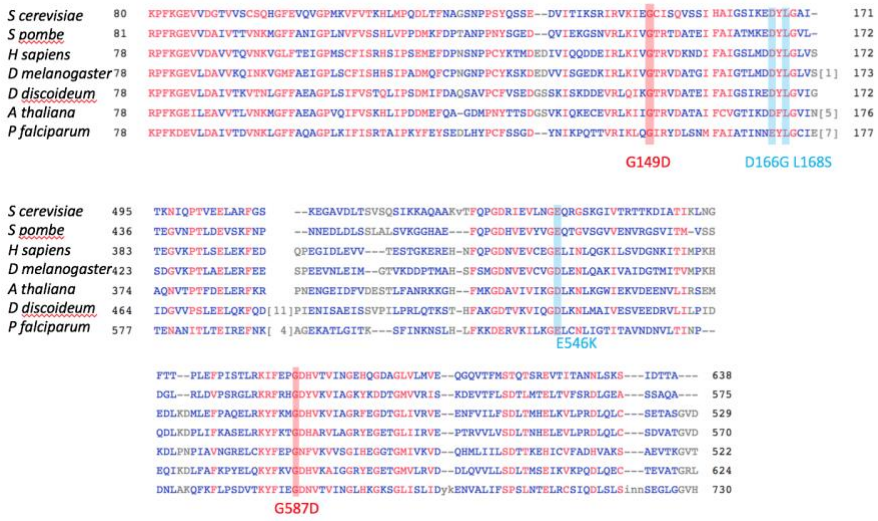
A


Reporter:
gal10D55
pGAL-FLO8::HIS3
Media:
YPD
Gal
YPD+MPA (25ug/mL)
SC-His + Gal
Form of rpb7
wt
G149D
E100K
D166G
D166K
V101E
L168S

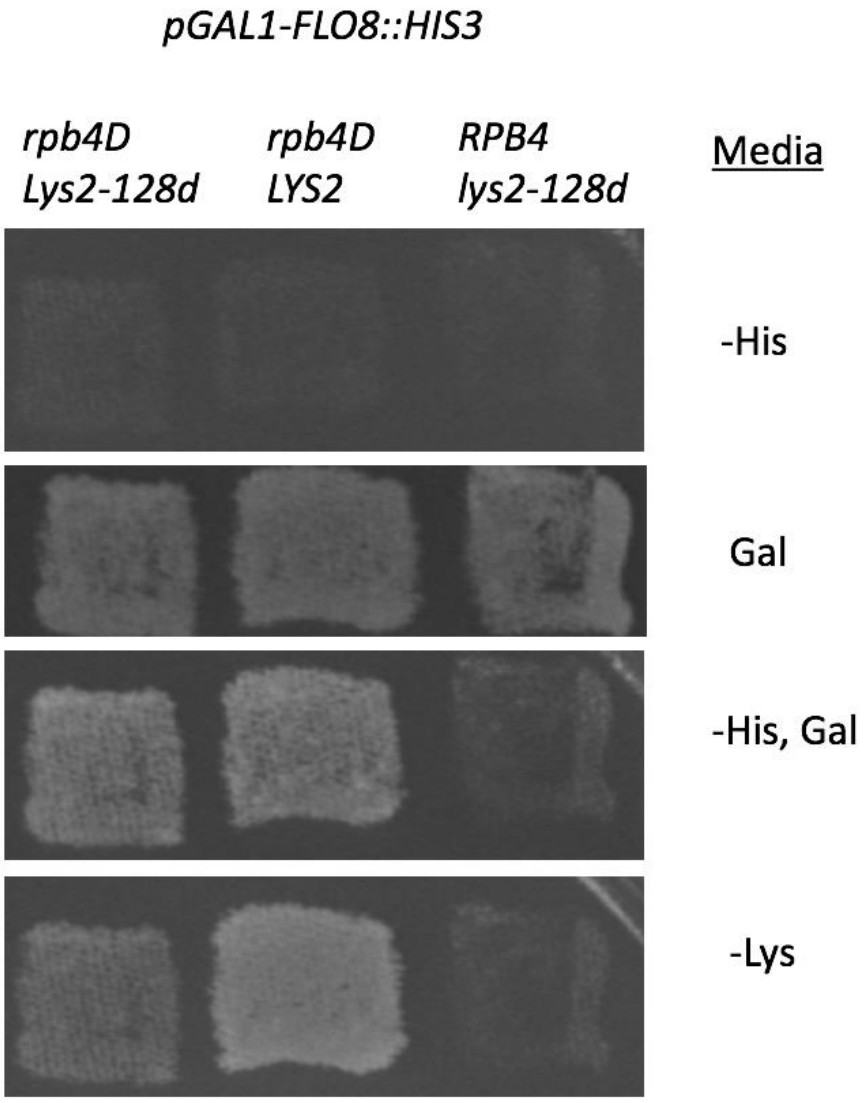
B



C


RPB7
S cerevisiae 80 KPFKGEVVDGTVVSCSQHGFEVQVGPMKVFVTKHLMPQDLTFNAGSNPPSYQSSE--DVITIKSRIRVKIEGCISQVSSI HAIGSIKEDYLGAI- 171
S pombe 81 RPFRGEVVDAIVTTVNKMGFFANIGPLNVFVSSHLVPPDMKFDPTANPPNYSGED--QVIEKGSNVRLKIVGTRTDATEI FAIATMKEDYLGVL- 172
H sapiens 78 RPFKGEVVDAVVTQVNKVGLFTEIGPMSCFISRHSIPSEMEFDPNSNPPCYKTMDEDIVIQQDDEIRLKIVGTRVDKNDI FAIGSLMDDYLGLVS 172
D melanogaster 78 RPFKGEVLDAVVKQINKVGMFAEIGPLSCFISHHSIPADMQFCPNGNPPCYKSKDEDVVISGEDKIRLKIVGTRVDATGI FAIGTLMDDYLGLVS[1] 173
D discoideum 78 KPFKGEVLDAIVTKVTNLGFFAEAGPLSIFVSTQLIPSDMIFDAQSAVPCFVSEDGSSKISKDDEVRLQIKGTRVDATEI FAIGSIREDYLGVIG 172
A thaliana 78 RPFKGEILEAVVTLVNKMGFFAEAGPVQIFVSKHLIPDDMEFQA-GDMPNYTTSDGSVKIQKECEVRLKIIGTRVDATAI FCVGTIKDDFLGVIN[5] 176
P falciparum 78 KPFKDEVLDAIVTDVNKLGFFAQAGPLKIFISRTAIPKYFEYSEDLHYPCFSSGD--YNIKPQTTVRIKLQGIRYDLSNM FAIATINNEYLGCIE[7] 177
G149D
D166G L168S
SPT5
S cerevisiae 495 TKNIQPTVEELARFGS --KEGAVDLTSVSQSIKKAQAAKvTFQPGDRIEVLNGEQRGSKGIVTRTTKDIATIKLNG
S pombe 436 TEGVNPTLDEVSKFNP --NNEDLDLSSLALSVKGGHAE---FQPGDHVEVYVGEQTGVSGVVENVRGSVITM-VSS
H sapiens 383 TEGVKPTLSELEKFED QPEGIDLEVV---TESTGKEREH-NFQPGDNVEVCEGELINLQGKILSVDGNKITIMPKH
D melanogaster 423 SDGVKPTLAELERFEE SPEEVNLEIM--GTVKDDPTMAH-SFSMGDNVEVCVGDLENLQAKIVAIDGTMITVMPKH
A thaliana 374 AQNVTPTFDELERFKR PNENGEIDFVDESTLFANRKKGH--FMKGDAVIVIKGDLKNLKGWIEKVDEENVLIRSEM
D discoideum 464 IDGVVPSLEELQKFQD[11]PIENISAEISSVPILPRLQTKST-HFAKGDTVKVIQGDLKNLMAIVESVEEDRVLILPID
P falciparum 577 TENANITLTEIREFNK[4]AGEKATLGITK---SFINKNSLH-LFKKDERVKILKGELCNLIGTITAVNDNVLTINP--
E546K
FTT--PLEFPISTLRKIFEPGDHVTVINGEHQGDAGLVLMVE--QGQVTFMSTQTSREVTITANNLSKS---IDTTA--- 638
DGL--RLDVPSRGLRKRFRHGDYVKVIAGKYKDDTGMVVRIS--KDEVTFLSDTLMTELTVFSRDLGEA---SSAQA--- 575
EDLKDMLEFPAQELRKYFKMGDHVKVIAGRFEGDTGLIVRVE--ENFVILFSDLTMHELKVLPRDLQLC---SETASGVD 529
QDLKDPLIFKASELRKYFKTGDHARVLAGRYEGETGLIIRVE--PTRVVLVSDLTNHELEVLPRDLQLC---SDVATGVD 570
KDLPNPIAVNGRELCKYFEPGNFVKVVSGIHEGGTGMIVKVD--QHMLIILSDTTKEHICVFADHVAKS---AEVTKGVT 522
EQIKDLFAFKPYELQKYFKVGDHVKAIGGRYEGETGMVLRVD--DLQVVLLSDLTMSEIKVKPQDLQEC---TEVATGRL 624
DNLAKQFKFLPSDVTKYFIEGDNVTVINGLHKGKSGLISLIDykENVALIFSPSLNTELRCSIQDLSLSinnSEGLGGVH 730
G587D

D


pGAL1-FLO8::HIS3
rpb4D Lys2-128d
rpb4D LYS2
RPB4 lys2-128d
Media
-His
Gal
-His, Gal
-Lys

E

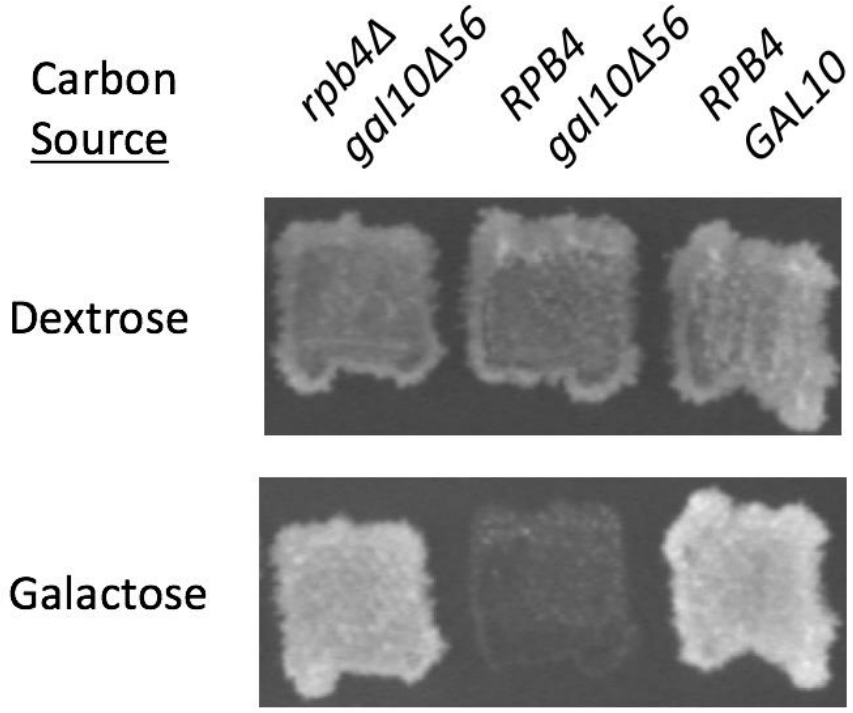

Carbon
Source
rpb4Δ
gal10Δ56
RPB4
gal10Δ56
RPB4
GAL10
Dextrose
Galactose

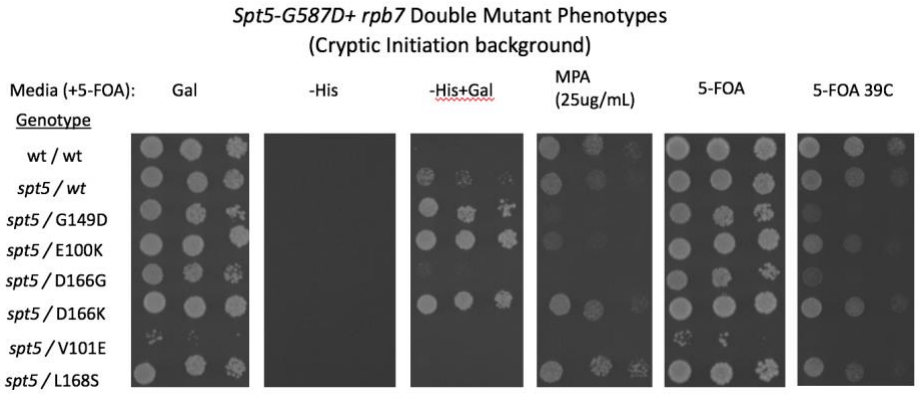
Figure 2-4: *spt5* and *rpb4/7* double mutants display synthetic phenotypes

A) *RPB7* shuffle strain was crossed with a strain containing *spt5-G587D* integrated in the cryptic initiation background. Mutant *rpb7* alleles along with wild type *RPB7* were transformed and passed over 5FOA to remove the *RPB7 URA3* plasmid. Strains were serially diluted onto the indicated media and incubated for 3 days at 30C.

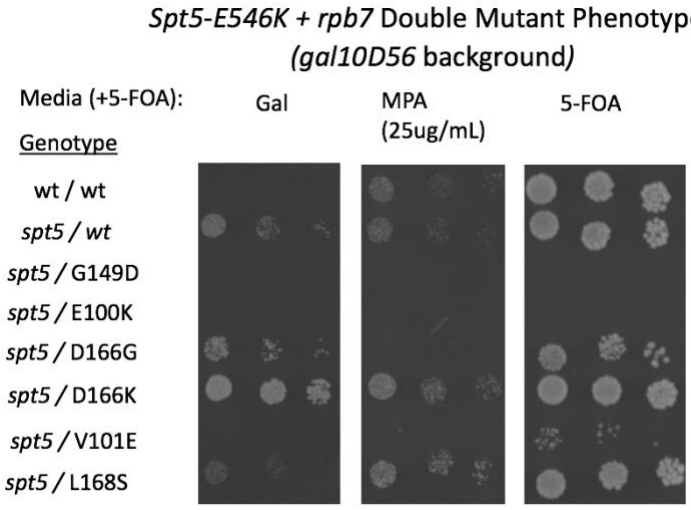
B) *RPB7* shuffle strain was crossed with a strain containing *spt5-E546K* integrated in the *gal10Δ56* background. Mutant *rpb7* alleles along with wild type *RPB7* were transformed and passed over 5FOA to remove the *RPB7 URA3* plasmid. Strains were serially diluted onto the indicated media and incubated for 3 days at 30C.

C) An *rpb4Δ* strain containing a *URA3* marked overexpression vector harboring *RPB7* was crossed with an *SPT5* shuffle strain. *spt5* alleles were transformed and cells were plated onto SC –Leu. Single colonies were selected and struck out on SC –Leu, then replica plated to 5FOA.

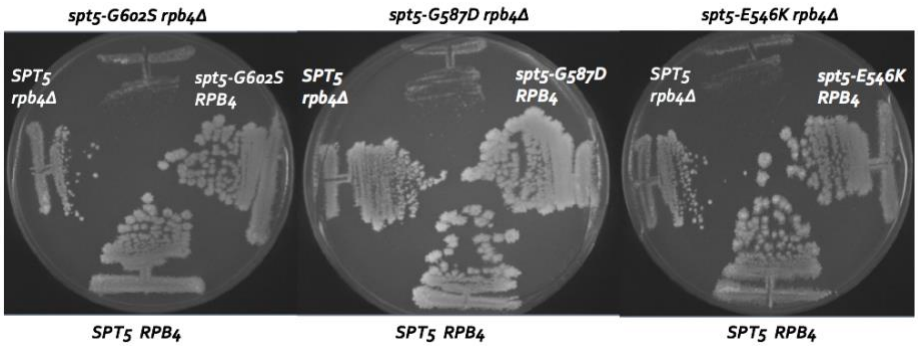
A


Spt5-G587D+ rpb7 Double Mutant Phenotypes
(Cryptic Initiation background)
Media (+5-FOA):
Gal
-His
-His+Gal
MPA (25ug/mL)
5-FOA
5-FOA 39C
Genotype
wt / wt
spt5 / *wt*
spt5 / G149D
spt5 / E100K
spt5 / D166G
spt5 / D166K
spt5 / V101E
spt5 / L168S

B


Spt5-E546K + rpb7 Double Mutant Phenotypes
(gal10D56 background*)*
Media (+5-FOA):
Gal
MPA
(25ug/mL)
5-FOA
Genotype
wt / wt
spt5 / wt
spt5 / G149D
spt5 / E100K
spt5 / D166G
spt5 / D166K
spt5 / V101E
spt5 / L168S

C


spt5-G602S rpb4Δ
spt5-G587D rpb4Δ
spt5-E546K rpb4Δ
SPT5 rpb4Δ
spt5-G602S RPB4
SPT5 rpb4Δ
spt5-G587D RPB4
SPT5 rpb4Δ
spt5-E546K RPB4
SPT5 RPB4
SPT5 RPB4
SPT5 RPB4

Chapter 3

Identification of factors that bind to the central stalk region of Pol II

Introduction

Rpb4/7, KOW2/3 and L2K4 appear to form a large, protruding structure on the Pol II EC. In Chapter 2, we identified residues in *spt5* and *rpb7* that display defects in both cryptic initiation and polyadenylation site choice that exhibit a high degree of evolutionary conservation – *spt5-G587D* and *rpb7-G149D* are universally conserved, while Spt5-E546 and Rpb7-D166 are universally either glutamate or aspartate. A charge distribution map reveals that this region harbors a highly acidic surface throughout Spt5 KOW2-3, Rpb4 and Rpb7 (Fig. 3-1). This is suggestive of the formation of a binding platform for tertiary factors Prior studies on Rpb4/7 have posited that the primary function of the Pol II Stalk is to expand upon the surface area to which transcription-associated factors may bind. Indeed, many protein binding partners of Rpb7 have been previously identified that also physically associate with Spt5. The majority of these factors have known roles in transcription regarding chromatin structure, pre-mRNA processing, 3' end formation and transcription processivity. The significant number of overlapping phenotypes, along with published structural data and sequence conservation analysis support the hypothesis that this structure plays an conserved role in participating in cooperative protein binding events.

Prior biochemical studies have identified a number of interacting factors with Spt5 (Lindstrom *et al.,* 2003; Krogan *et al.,* 2002). Spt5 is a large, multi-domain protein, and the resolution of these binding events to the individual domain of Spt5 remains unsolved. Since the various domains of Spt5 appear to bind to Pol II across disparate locations, it is reasonable to suspect that the various Spt5-interacting factors may bind exclusively to specific domains of Spt5 to support location-specific roles. Spt5's KOW domains are composed of Tudor folds which, in other proteins, are known to support protein-protein and protein-nucleic acid interactions (Meyer *et al.,* 2015; Lasko 2010). To pinpoint the function of the conjoined Spt5 KOW-Stalk region of the elongation complex, it is therefore necessary to identify factors that bind specifically to the KOW domains that participate in this structure.

To accomplish this, we performed affinity chromatography followed by MudPIT mass spectrometry in collaboration with the Yates Lab at UC San Diego to identify binding partners of both the KOW2-3 (K2K3) and the Linker2-KOW4 (L2K4) regions of Spt5. Supporting the notion that the central KOW domains function in tandem with Rpb4/7, we identified many proteins involved in 3' end formation, RNA processing and chromatin structure maintenance that overlap with previous proteomics studies of Rpb7.

Results

Affinity Chromatography of Spt5's KOW domains

To further resolve Spt5's binding partners to the KOW2-4 region, we purified recombinant His-tagged KOW2-3 and Linker2-KOW4 and immobilized them on Afigel. The column was challenged with a yeast protein extract, washed and subjected to an increasing salt gradient. Silver stained gels of the eluates revealed that the KOW2-3 and L2K4 bound a strikingly different pattern of proteins (Fig. 3-1A, B). The three highest salt fractions were subject to MudPIT mass spectrometry, A BSA bound Afigel column was used as control in mass spectrometry experiments (Fig. 3-1C).

We identified many proteins that bound differentially to KOW2-3 and L2K4, as well as overlapping factors that bound to both domains. We excluded a number of proteins from analysis: ribosomal proteins, which are frequently identified nonspecifically in mass spectrometry, cytoplasmic proteins as Spt5 has been shown to reside exclusively in the nucleus, as well as importins and mitochondrial membrane proteins. Spt5 is known to interact with RNAPI, and we did identify a number of RNAPI subunits, however we excluded ribosome biogenesis factors as this study focuses on Pol II transcription. Finally, any proteins that were enriched less than two-fold relative to the BSA control column were excluded. We were left with 80 proteins that bound to either K2K3 or L2K4, presented in Table 3-1. We observed proteins involved in a wide array of nuclear functions, including transcription, RNA polymerase subunits, splicing, replication and RNA processing. Consistent with the phenotypes observed in our KOW2-3

mutants, we identified a number of proteins involved in chromatin structure maintenance as well as 3' end formation.

Identification of Nrd1 and Nab3 as KOW2-3 binding

Prior studies have identified factors that bind to Rpb7. Many of these identified proteins overlap with factors identified in our affinity chromatography experiments, which are indicated in bold. Of particular interest is the binding of the Nrd1-Nab3-Sen1 complex which is involved in non-polyadenylated transcript termination, exosome mediated RNA degradation and mRNA surveillance (Steinmetz *et al.,* 2001; Vasiljeva *et al.,* 2006; Singh *et al.,* 2021). Nrd1 and Nab3 both bound KOW2-3 in our study, while Sen1 and Nrd1 have been identified in prior studies of Rpb7, indicating that the region comprised of Spt5 KOW2-3 and Rpb4/7 is a possible binding target for the Nrd1-Nab3-Sen1 termination complex (Mosley *et al.,* 2013; Mitsuzawa *et al.,* 2003).

A charge distribution map of the Pol II Stalk Region suggests that this structure may be targeting basic proteins (Fig. 3-2; Bernecky *et al.,* 2017). To validate our result of Nrd1 binding to KOW2-3, we utilized the machine learning computational program AlphaFold to predict Nrd1's overall structure (Jumper *et al.,* 2021; Varadi *et al.,* 2021; Fig. 3-3A). We then generated a charge distribution map of Nrd1 to determine if Nrd1 possesses the biochemical characteristics that may facilitate binding to the Pol II Stalk Region including Rpb7 and KOW2-3 (Fig. 3-3B). We found that the N-terminal

region of Nrd1 is largely basic in nature and is surrounded by an unstructured loop region, which is consistent with Rpb7 and KOW2-3 being a potential target for Nrd1.

Recent developments in AlphaFold's algorithm have allowed for the modeling of protein complexes using AlphaFold Multimer. To test our prediction that Nrd1 may bind Rpb7 and KOW2-3 collaboratively, we input sequences for the N-terminal region of Nrd1, which has previously been shown to bind Rpb7, along with Rpb7 and KOW2-3 (Mitsuzawa *et al.,* 2003). Interestingly, the top 5 ranked models output by AlphaFold were all shown to bind Rpb7. In particular, the most highly ranked model showed Nrd1's N-terminal region to bind to the region of Rpb7 that contains Rpb7-D166. However, none of the models showed Nrd1 binding to KOW2-3.

As Nab3 was also shown to bind KOW2-3 in our affinity chromatography experiment, we input the sequence for the entire Nab3 protein with KOW2-3. Surprisingly, all five models output by AlphaFold placed Nab3's RRM directly on top of KOW2 at the location where Spt5-E546 resides. Figure 3-3 shows a composite of the Rpb7, KOW2-3, the Rpb7 binding domain of Nrd1 and Nab3's RRM. These structural models predict that the NNS complex is docked directly onto the Pol II Stalk region, with Nrd1 contacting Rpb7 at D166 through its N-terminal domain and Nab3 targeting KOW2-3 at E546 through its RRM. Further, Nab3-K363 has been previously shown to be a critical residue for RNA binding, and all of the models output by

Alpha fold showed K363 directly contacting Spt5-E546, potentially implicating this region of KOW2 as playing an indirect role in mediating RNA binding through Nab3 (Lee *et al.,* 2020).

Discussion

By performing affinity chromatography and MudPIT mass spectrometry, we sought to identify factors to the hypothesis that Spt5 KOW domains 2-4 act as a platform for recruitment of tertiary factors and provide context for the mutations in *spt5* identified in Chapter 2. We found many factors that were involved in both of these processes, as well as proteins that contribute to general transcription splicing, termination, and surprisingly, replication. Some of these factors were identified in previous mass spectrometry experiments involving full-length Spt5 (Lindstrom *et al.,* 2003). However, the sensitivity of mass spectrometry has evolved significantly since those prior experiments were performed, and we aimed to further resolve Spt5's interactome to specifically the central domains.

Rpb7 and KOW2-3 as a target for the NNS Complex

Both Nrd1 and Nab3, components of the Nrd1-Nab3-Sen1 (NNS) termination complex, bound KOW2-3. Sen1 and Nrd1 have also been identified as an Rpb7 binding protein (Mosley *et al.,* 2013; Mitsuzawa *et al.,* 2003). The NNS complex is responsible for terminating non-polyadenylated RNA such as snRNA, snoRNA, and CUTS (Steinmetz *et al.,* 2001; Conrad *et*

al., 2000), and Sen1 has previously been shown to play a role in resolving R-loops. Seb1, a Nrd1 ortholog found in *S. pombe* also was shown to bind Rpb7 in a two-hybrid assay, as was *S. cerevisiae* Nrd1 to Rpb7, indicating this interaction is conserved evolutionarily and functionally significant (Mitsuzawa *et al.,* 2003). Interestingly, residue D167 in *S. pombe* Rpb7 was identified as a critical residue for binding of Seb1 to Rpb7. This highly conserved residue aligns with D166 in *S. cerevisiae* Rpb7, the same residue featured in this study as Rpb7-D166G, suggesting the Stalk/KOW platform may be a binding site for the NNS termination complex. Prior affinity chromatography studies of Nrd1 also identified Spt5 as a Nrd1 interacting protein (Vasiljeva *et al.,* 2006). Supporting the model that Nrd1 targets both Spt5 KOW2-3 and Rpb7, a charge distribution map of Nrd1 reveals a highly basic interface composed of a region that has previously been identified to be necessary to bind Rpb7, while the structure composed of Rpb7 and KOW2-3 is highly acidic in nature (Mitsuzawa *et al.,* 2003).

AlphaFold Multimer predicts that Nab3's RRM contacts KOW2-3 at Spt5-E546, and Nrd1 targets Rpb7 at D166. It is appealing to speculate that *spt5-E546K* and *rpb7-D166G* break contact with the NNS complex, which may mechanistically explain a subset of the phenotypes observed in these mutants. Future experiments may aim to test the model that these residues are critical for maintaining contact with the NNS complex. Interestingly, Nab3-K363 has been shown to be critical for Nab3's RNA binding capacity, and this

residue is predicted to directly contact Spt5-E546 (Lee *et al.,* 2020. It will be worthwhile in the future to test the role of Spt5-E546 in stimulating or abrogating RNA binding by Nab3 and what the functional consequences of this interaction may be. Further, this interaction may be RNA dependent – it would be worthwhile to test this hypothesis in a pull-down assay or RNase treated affinity chromatography experiments with wildtype *SPT5* and *spt5-E546K.* Nab3-K363 has also been shown to be a target for methylation by Set1 and Set3. Future experiments should aim to test the consequences of methylation of this residue in the context of Nab3's interaction with Spt5 KOW2-3, and also whether Spt5 is a target for methylation by Set1 or Set3.

It is highly likely that the NNS complex docks directly onto Rpb7 and KOW2-3 through acid-base interactions between Nrd1, Nab3 and the Pol II Stalk Region. Prior co-immunoprecipitation experiments have shown that Spt5 and Nrd1 do indeed physically interact with each other, and that this interaction is dependent on Spt5's CTR (Leporé *et al.,* 2011). Mechanistically, it seems likely that Nrd1 is first recruited by Spt5's CTR and then deposited onto Rpb7 where it remains primed for RNA recognition via its RRM. Future studies should aim to define a role for Spt5 and Rpb4/7 in noncoding transcript termination.

Termination

Rat1 and Rai1 were also bound KOW2-3. Rat1 is known to be the "torpedo nuclease" in the "torpedo" model of termination. According to this model, Rat1 (assisted by Rai1) facilitates transcription termination by degrading RNA from the 5'-end that is created by transcript cleavage at the poly-A site, eventually colliding with Pol II to facilitate its removal from the DNA (Luo *et al.,* 2006; Kim *et al.,* 2004). This model of termination is consistent with the function of Rho in *E. coli,* which tracks along nascent RNA and collides with RNAP stimulating termination via a direct interaction with NusG (Mitra *et al.,* 2017). KOW2-3 may serve as a loading dock for Rat1 as it prepares to bind to cleaved mRNA following poly(A) site recognition, or KOW2-3 may serve as the target for Rat1 already engaged with the cleaved transcript, actively tracking the polymerase as it prepares to terminate via collision with the EC.

Dbp2, a dead-box RNA helicase, is involved in coupling of termination to mRNP assembly and is required for loading Yra1, Nab2 and Mex67 onto mRNA (Ma *et al.,* 2013). Interestingly, Dbp2 has also been implicated in suppression of transcription from internal initiation sites (Cloutier *et al.,* 2012). Dbp2 has a strong preference for dsRNA, suggesting a role in preparing mRNA secondary structure for efficient packaging into an export-competent mRNP (Cloutier *et al.,* 2012). KOW2-3 may also serve as a "loading dock" for Dbp2 to target mRNA as it leaves the mRNA exit tunnel.

RNA Processing

A number of components of the TRAMP complex, exosome and SKI complex were identified as interacting with KOW2-4, along with other factors such as nucleases, helicases and RNA binding proteins. Many of these factors are involved in complexes that are known to interact with each other, such as the SKI complex, TRAMP complex and exosome (Keidel *et al.,* 2020; Weir *et al.,* 2010; Jackson *et al.,* 2010; Schneider, Tollervey 2013). It has been proposed that the SKI complex may directly funnel mRNA into the exosome as well as modulating exosome activity (Schneider, Tollervey 2013). The TRAMP complex has also been shown to be involved in exosome-mediated degradation of ncRNAs in a process that is mediated by Nab3 of the Nrd1-Nab3-Sen1 ncRNA termination complex (Mile *et al.,* 2015). This is consistent with the proposed function of KOW2-4 as a nexus for coordinating mRNA processing events due to their proximity to the mRNA exit channel.

Gbp2 has been identified as a novel, selective poly(A) RNA binding protein that facilitates shuttling of mRNA out of the nucleus with a preference for intron-containing genes. Overexpression of Gbp2 causes nuclear retention of polyadenylated RNA (Windgassen, Krebber 2003). Gbp2 also was shown to bind Rpb7-Tap, which may accompany mRNA out of the nucleus with Rpb4 (Mosley *et al.,* 2013; Harel-Sharvit *et al.,* 2010).

Rok1 and Dbp4 are ATP-dependent RNA helicases involved in rRNA processing and participate in early ribosomal small subunit biogenesis (Venema *et al.,* 1997; Koshnevis *et al.,* 2016; Weaver *et al.,* 1997). While

Rok1 is essential, Dbp3 is not (Song *et al.,* 1995; Weaver *et al.,* 1997). Not much is known of the mechanisms behind Rok1 and Dbp3 function, however the fact that Spt5's central KOW domains interact with multiple rRNA processing proteins and have known roles in Pol I function supports the idea that Spt5 may play a significant role in rRNA processing and rRNA biogenesis.

Xrn1, a component of the Xrn1-decaysome, is a 5'-3' exonuclease that has been implicated in both transcriptional activation and termination (Blasco-Morena *et al.,* 2019; Jonas, Izaurralde 2013; Parker 2012). Loss of Xrn1 results in a decrease in Pol II occupancy downstream of promoters and an increase in pol II density near cleavage and polyadenylation sites (Fischer 2020). One potential model of suppression of *gal10Δ56* in our *spt5* and *rpb7* mutants is that these mutations alter residues that result in disruption of interactions with particular transcription factors, favoring a termination-competent EC. One such factor whose interaction may be disrupted could be Xrn1. By severing contact with Xrn1 through mutating the binding platform it relies upon, we may be mirroring the transcriptional impact seen as a result of loss of Xrn1. Future co-IP and ChIP experiments can test this hypothesis.

Sro9 has been shown to chaperone mRNA from the nucleus to the cytoplasm where it then modulates translation (Röther *et al.,* 2010). Rpb4/7 have also been shown to shuttle into and out of the nucleus accompanying mRNA as well (Harel-Sharvit *et al.,* 2010). Interestingly, Sro9 was identified

as a high-copy suppressor of the T^s phenotype of an *rpb4Δ* mutation. Overexpression of Sro9 nearly doubled mRNA levels, which was attributed to a combination of increased transcription and mRNA stability (Tan *et al.,* 2000).

Replication

Previous studies identified Smc1 as an Spt5 interacting protein, and it appears that Smc1 binds specifically to the KOW2-3 region, as well as almost the entire Rfc complex, involved in loading of the PCNA sliding clamp onto chromatin (Lindstrom *et al.,* 2003). Interestingly, many Rfc subunits also have been shown to bind to Rpb7 (Mosley *et al.,* 2013). During S-phase, Pol II and DNA Polymerase compete for the same substrate, which can result in collision events known as T-R collisions (Prado, Aguilera 2005). During a T-R collision, transcribed mRNA can hybridize with unwound DNA, forming R-loops, which are often sources of DNA damage (Hamperi *et al.,* 2017). Sen1, the helicase component of the Nrd1-Nab3-Sen1 termination complex, has been previously shown to specifically unwind DNA-RNA hybrids and resolve R-loops, particularly during S-phase (Appanah *et al.,* 2020). It is possible that the Rfc interactions between Spt5 and Rpb4/7 are a result of T-R collisions, and that their interaction, along with Sen1, helps resolve the collision and prevent DNA damage. Rpb4 and Rpb7 have been shown to mediate the DNA damage response through the independent pathways, as has the Rfc

complex, supporting the idea that this region may mediate resolution of T-R collisions (Li *et al.,* 2002; Kumar *et al.,* 2019; Shiomi, Nishitami 2017).

Splicing

Prp43 is an essential ATP-dependent DEXH/D-box RNA helicase that functions in Pol I and Pol II transcript metabolism (Arenas, Abelson 1997; Combs *et al.,* 2006). Prior studies demonstrate that Prp43 interacts with Rpb7, while our work shows it interacts with L2K4, which lies near the mRNA exit tunnel (Mosley *et al.,* 2013). Prp43's catalytic functions include removing the U2, U5 and U6 snRNPs from the post splicing lariat-intron ribonucleoprotein complex, biogenesis of large and small subunit rRNAs and it also has roles in pre-mRNA processing (Combs *et al.,* 2006; Fourmann *et al.,* 2013). Loss of Prp43 function results in accumulation of unspliced rRNA and, to a lesser degree, unspliced pre-mRNA (Arenas, Abelson 1997). Prior studies have shown that Spt5 loss of function results in accumulation of unspliced mRNA, and that Spt5 functions with Pol I in rRNA metabolism (Lindstrom *et al.,* 2003; Viktorovskaya *et al.,* 2011).

Chromatin and Transcription

Spt16, a component of the FACT (FAcilitates Chromatin Transcription) complex, is a histone chaperone that has known genetic interactions with Spt4/5 and was identified as KOW2-3 interacting in this study (Orphanides *et al.,* 1999; Belotserkovskaya *et al.,* 2003). Recent structural models of

elongating Pol II however do not show Spt16 in direct contact with Spt5's central domains or Rpb7, but rather that it contacts nucleosomes downstream of elongating Pol II (Farnung *et al.,* 2021). Interestingly, Spt16 also was shown to interact with Rpb7 in prior proteomics studies, suggesting FACT may adopt multiple conformations with regards to the Pol II EC (Mosley *et al.,* 2013). Spt6, another known histone chaperone, has been shown to bind near the Pol II stalk, creating together with Rpb4/7 and KOW2-3 a large acidic platform that may attract basic proteins such as histones (Vos *et al.,* 2018b). In fact, a number of histones were identified as interacting with KOW2-4 that have previously been identified to interact with Rpb7 as well. Supporting this notion is the observation that one of the conserved functions of Tudor domain containing proteins, and in particular tandem Tudor domains such as Kow2-3, is to bind post-translationally modified histone tails (Ying, Chen 2012; Botuyan *et al.,* 2006; Huang *et al.,* 2006). It is appealing to speculate a mechanism in which nucleosomes are "rolled" over the body of a transcribing polymerase with the help of Spt6 and Spt16, with the Pol II Stalk region, including KOW2-4, serving as an intermediate area for nucleosome reassembly. Prior studies implicate Spt6 and Spt16 in reassembly of nucleosomes in the wake of transcribing Pol II, and this model has been proposed to explain the cryptic initiation phenotype in both *spt6* and *spt16* mutants (McCullough *et al.,* 2015; Cheung *et al.,* 2008; Kaplan *et al.,* 2003). Tracing a path from where Spt5 contacts the incoming nucleosome to the

binding interface of Spt6, one can speculate that Spt16 may facilitate the transit of nucleosomes downstream of transcribing pol II, over the stalk region comprised of KOW2-3, Rpb4/7 and Spt6, and back over to the upstream DNA. This model is supported by the identification of the novel cryptic initiation mutations in *KOW2-3* and *rpb7-D166G* in Chapter 2. Mechanistically, Spt16 may facilitate a nucleosome hand-off event to Spt6, or it may accompany nucleosomes back to their upstream position, depositing them in Pol II's wake.

Ribosome biogenesis

We identified a number of RNA Polymerase I subunits in our affinity chromatography experiment that were also identified in previous Spt5 full-length affinity chromatography experiments (Lindstrom *et al.,* 2003). Supporting the idea that Spt5 plays roles in ribosome biogenesis, we identified a large number of proteins associated with this function, however we excluded the majority of these factors from analysis due to the focus of this study being Pol II transcription. Hcr1 does remain of interest due to its dual-roles in translation and ribosome biogenesis and prior context in function with Rpb4/7 (Valášek *et al.,* 2001a; Harel-Sharvit *et al.,* 2010). The observation of genetic and physical interactions between Hcr1 and Rpb4/7 makes it an interesting candidate for further study. These experiments have suggested a role for Rpb4/7 in modulating translation through dual functions in facilitating mRNA transport and exerting influence directly on Hcr1. This

suggests two possible models for Spt5's function with Hcr1: Spt5 may function in indirectly modulating Hcr1 activity through Spt5's direct interactions with Rpb4/7, or, Spt5 may modulate Hcr1 function as a complex with RNA Pol I in the context of ribosome biogenesis.

Figure 3-1: Spt5 KOW2-3 and Rpb4/7 form a large acidic structure

Charge distribution map of the Pol II Stalk Region. Red indicates negative charge, blue indicates positive charge (Bernecky *et al.,* 2017; PDB: 5OIK)

A

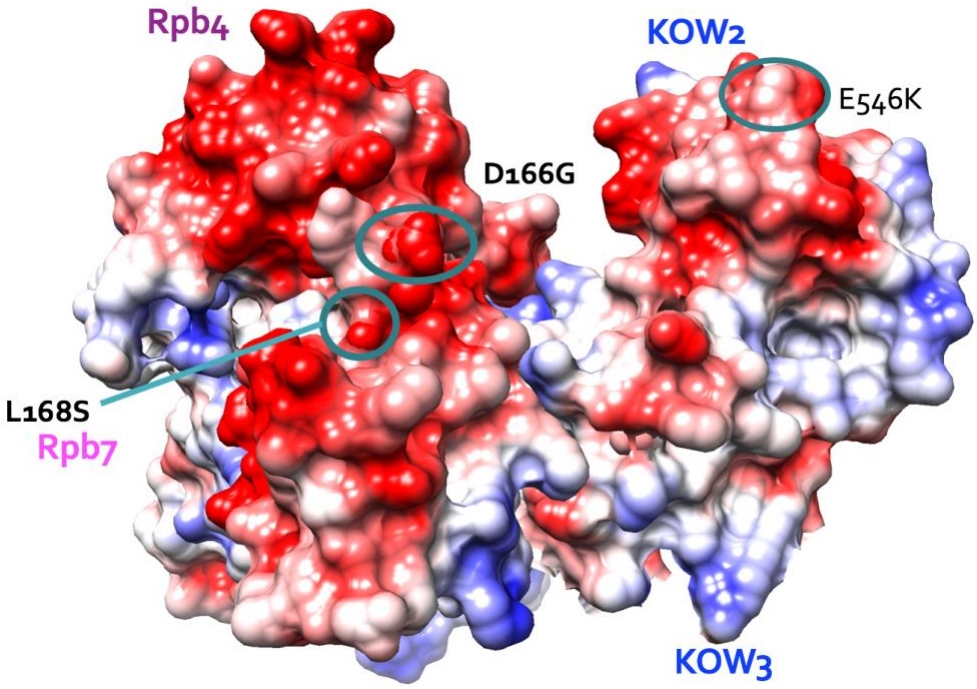


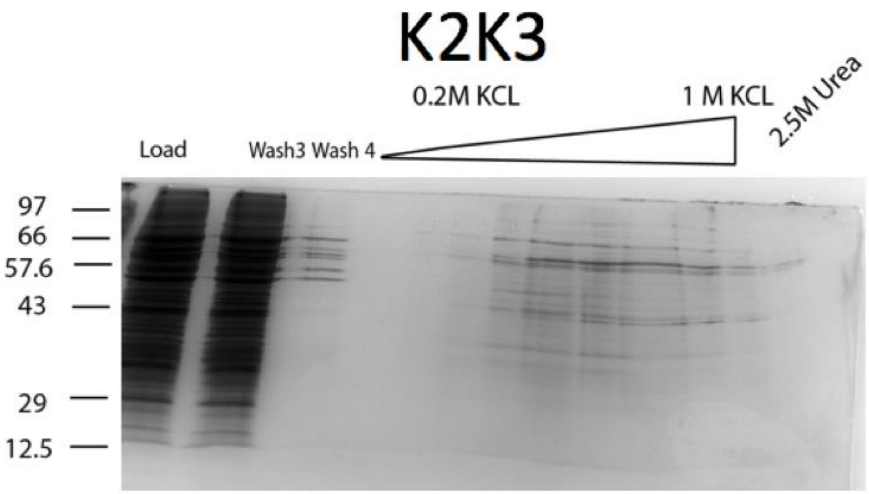
Figure 3-2: Spt5 KOW2-3 and Linker2-KOW4 exhibit unique protein binding fingerprints

A) Silver-stain gradient gel of eluates collected during affinity chromatography experiments of K2K3

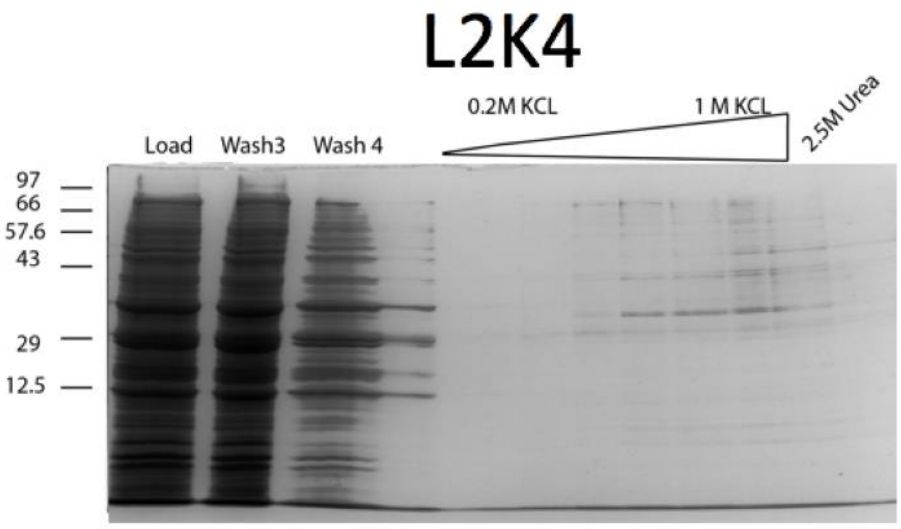
B) Silver stain gradient gel of eluates collected from L2K4 affinity chromatography experiments

C) Eluates obtained from the BSA control column.

A


K2K3
0.2M KCL
1 M KCL
2.5M Urea
Load
Wash3 Wash 4
97
66
57.6
43
29
12.5

B


L2K4
0.2M KCL
1 M KCL
2.5M Urea
Load
Wash3
Wash 4
97
66
57.6
43
29
12.5

C

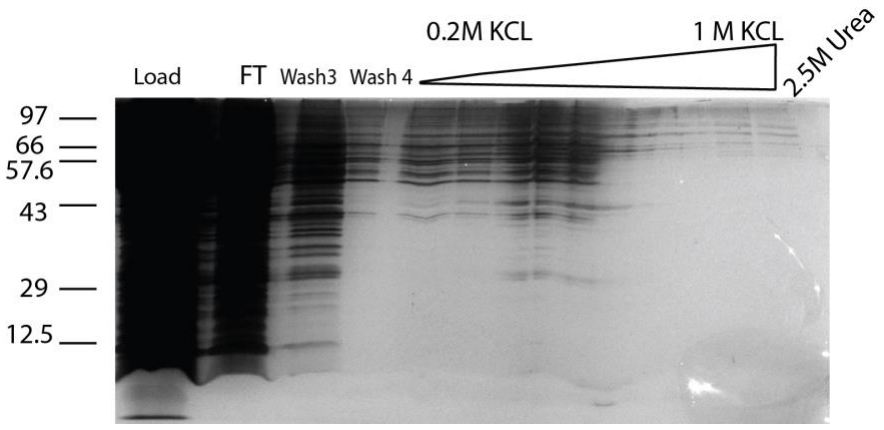
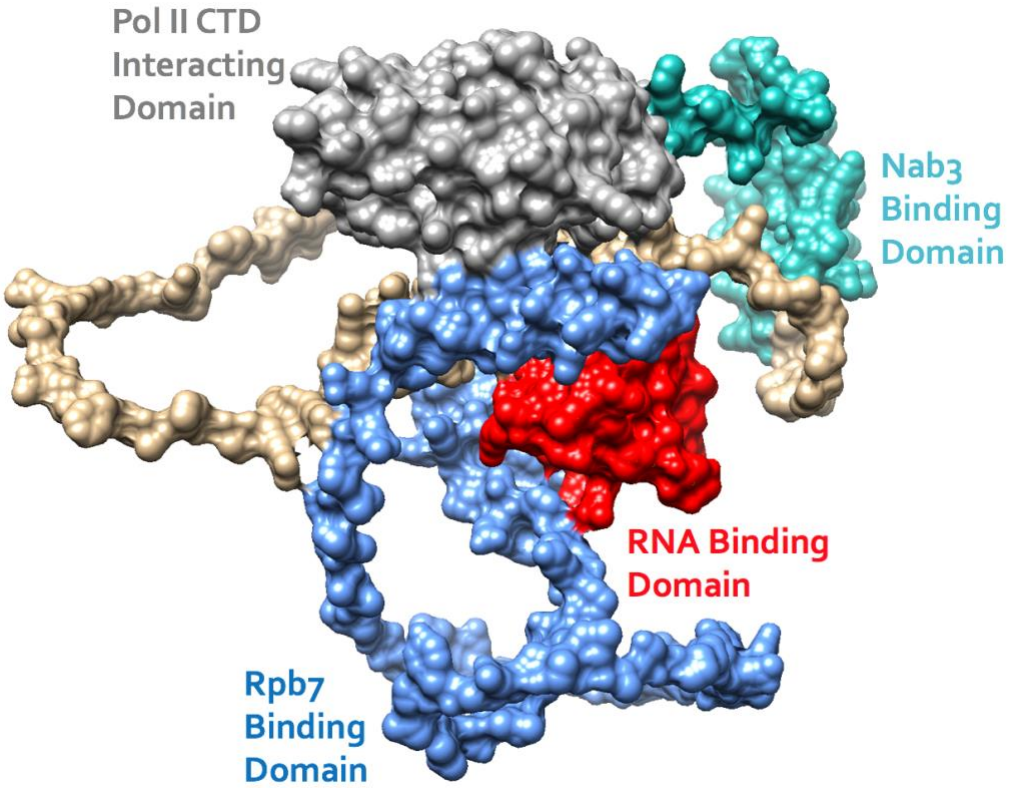

BSA
0.2M KCL
1 M KCL
2.5M Urea
Load
FT
Wash3
Wash 4
97
66
57.6
43
29
12.5

Figure 3-3: Nrd1's Rpb7 binding domain is largely basic in nature

A) Model of Nrd1 structure generated from AlphaFold with functional domains labeled (Jumper *et al.,* 2021; Varadi *et al.,* 2021; Vasiljeva *et al.,* 2008; Mitsuzawa *et al.,* 2003; Conrad *et al.,* 2000).

B) Charge distribution map of Nrd1 generated from AlphaFold

A


Pol II CTD
Interacting
Domain
Nab3
Binding
Domain
RNA Binding
Domain
Rpb7
Binding
Domain

B

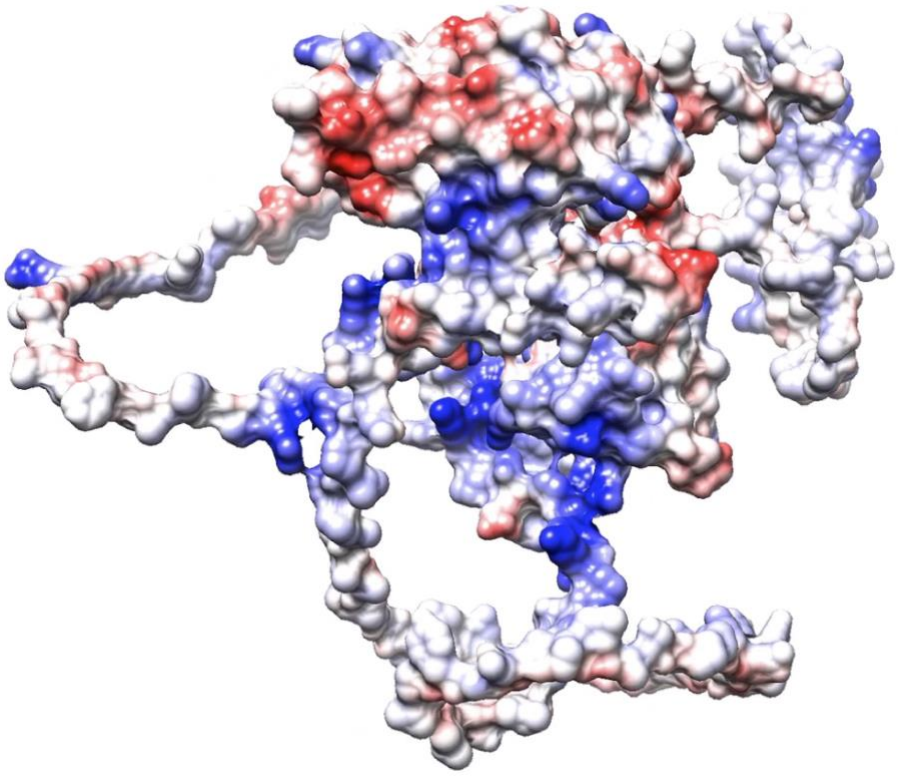


Figure 3-4: AlphaFold Multimer predictions for a complex containing Rpb7, KOW2-3, Nrd1 and Nab3

Yellow residue in Rpb7 indicates Rpb7-D166. Red residue in KOW2-3 indicates Spt5-E546. Orange residue in Nab3 indicates Nab3-K363. Nrd1 Rpb7 Binding Domain (7BD) contains residues 318-490. Nab3 displays residues 311-414. KOW2-3 displays residues 534-632. Entire sequence of Rpb7 shown.

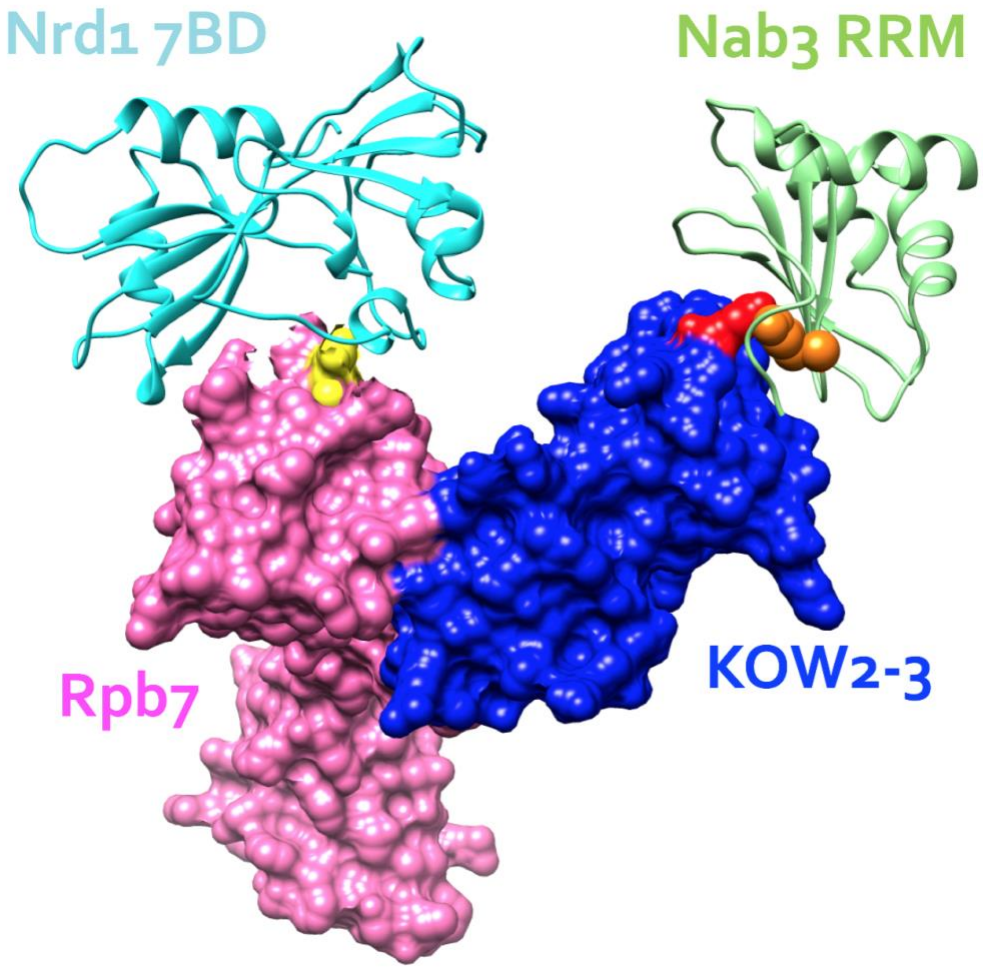

Nrd1 7BD
Nab3 RRM
Rpb7
KOW2-3

Table 3-1: Proteins identified via MudPIT mass spectrometry as interacting with K2K3, L2K4 domains of Spt5

Proteins indicated in bold were previously identified as Rpb7-interacting (Mosley *et al.,* 2013). Underlined proteins were previously identified as Spt5-interacting (Lindstrom *et al.,* 2003).

Protein Type and Gene	Notes	Domain of Spt5
Polymerase subunits		
Rpc10	RNAPI, II, III subunit	K2K3, L2K4
Rpc19	RNAPI, RNAPIII subunit	K2K3
Rpa49	RNAPI subunit	K2K3
Rpa135	RNAPI subunit	K2K3
Rpa190	RNAPI subunit	K2K3
Rpa34	RNAPI subunit	K2K3
Ret1	RNAPIII subunit	K2K3
Rpb8	RNAPI, II, III subunit	K2K3
Rpc344	RNAPIII subunit	K2K3
Termination		
Nrd1	Subunit of Nrd1 complex	K2K3
Nab3	Subunit of Nrd1 complex	K2K3
Rat1	5'-to-3' ssRNA exonuclease	K2K3
Rai1	binds to and stabilizes Rat1	K2K3
Dbp2	RNA Helicase	K2K3
RNA Processing		
Csl4	Exosome component	K2K3
Rok1	RNA helicase	K2K3
Pap2	Component of TRAMP	K2K3
Air1	Component of TRAMP	K2K3
Dbp3	RNA helicase	K2K3, L2K4
Xrn1	5'-3' exoribonuclease	K2K3, L2K4
Gbp2	Poly(A+) RNA-binding protein	K2K3, L2K4
Mtr4	Component of TRAMP	K2K3
Sro9	RNA-binding protein	K2K3
Ski2	SKI complex component	K2K3, L2K4
Ski3	SKI complex component	K2K3, L2K4
Ski8	SKI complex component	K2K3, L2K4
Chromatin		
Nap1	Histone chaperone	K2K3
Nbp2	Nap1 binding protein	K2K3
Hht1	Histone H3	L2K4
Htz1	Histone H2A.Z	K2K3, L2K4
Htb2	Histone H2B.1	L2K4
Fpr3	Role in nucleosome assembly	K2K3, L2K4
Fpr4	Role in nucleosome assembly	K2K3, L2K4
Itc1	Isw2-Itc1 complex subunit	L2K4
Euc1	DNA binding protein	K2K3
Spt16	FACT complex subunit	K2K3
Splicing		
Prp43	RNA helicase	L2K4
Lsm5	Component of U6 snRNP	L2K4
Lsm3	Component of U4/U6-U5 snRNP	K2K3

Protein Type and Gene	Notes	Domain of Spt5
Transcription		
Ncb2	Subunit of NC2 complex	K2K3
Elp3	Elongator subunit	K2K3, L2K4
Elp2	Elongator subunit	K2K3, L2K4
Iki3	Elongator subunit	K2K3, L2K4
Elp4	Elongator subunit	K2K3
Dig1	Inhibitor of Ste12	K2K3
Spt5	Transcription Elongation	K2K3, L2K4
Brf2	TFIIIB B-related factor	K2K3, L2K4
Mss16	Mitochondrial txn factor	K2K3, L2K4
Replication		
Smc1	cohesin subunit	K2K3
Rfc3	Replication factor C subunit	K2K3, L2K4
Rfc1	Replication factor C subunit	K2K3, L2K4
Rfc2	Replication factor C subunit	K2K3, L2K4
Rfc5	Replication factor C subunit	K2K3, L2K4
Rfc4	Replication factor C subunit	K2K3, L2K4
Other		
Ppn1	Endopolyphosphatase	K2K3, L2K4
Ckb1	Casein Kinase II subunit beta	K2K3
Cka2	Casein Kinase II subunit alpha'	K2K3
Cka1	Casein Kinase II subunit alpha	K2K3, L2K4
Hrr25	Casein Kinase I homolog	K2K3
Ckb2	Casein Kinase II subunit beta'	K2K3, L2K4
Sip5	Snf1 interacting protein	L2K4
Yhp1	Transcriptional repressor	L2K4
Cdc13	ssDNA binding protein	L2K4
Pho81	CDK inhibitor	L2K4
TY1B-LR3, LR1, PR3	TY1-LR3 Gag-Pol polyprotein	L2K4
TY1B-ER2, LR1, MR2	TY1-LR2 Gag-Pol polyprotein	L2K4
TY1B-PR3	TY1-PR3 Gag-Pol polyprotein	K2K3
TY1B-NL1	TY1-NL1 Gag-Pol polyprotein	K2K3
Sbp1	Found in P-bodies	K2K3
Pby1	Associated with P-bodies	K2K3
Ptc1	PP2C, dephosphorylates Hog1	K2K3
YML108W	Localizes to nucleus upon DNA	K2K3
Pci8	Cop9 component	K2K3
Tyw1	Wybutosine synthesis	K2K3
Sda1	Binds Nap1	K2K3
Kip3	Motor protein	K2K3
Msh1	mitochondrial DNA repair	K2K3
YCR051W	ankyrin-repeat containing	K2K3
Mak5	DEAD-box RNA helicase	K2K3, L2K4
Hcr1	EIF3 component	K2K3

Chapter 4

Investigating a Role for Spt5 and Rpb7 in R-loop Resolution

Introduction

Thus far, we have established that Spt5 and Rpb4/7 share many genetic interactions including synthetic sickness, synthetic lethality, and enhancement of mutant phenotypes pertaining to suppression of cryptic intragenic transcripts and polyadenylation site choice. We have also identified physical interactions between Spt5 and proteins that function in 3' end processing/termination, mRNA processing/export, ribosome biogenesis and R-loop resolution. Many of these factors were previously identified to interact physically and genetically with Rpb4/7 (Mosley *et al.,* 2013; Harel-Sharvit *et al.,* 2010). This chapter will focus on genetic interactions of Spt5 and the Pol II stalk with tertiary factors related to cotranscriptional processes including 3' end processing and R-loop homeostasis. In these experiments we aim to test the functional relevancy of the physical interactions identified in Chapter 3. Through this analysis we will construct a model that supports Spt5 KOW2-4 and Rpb4/7's cooperative function throughout the transcription cycle.

Nrd1 is an RNA binding protein that supports transcript termination of non-coding RNA and CUTs (Vasiljeva *et al.,* 2008; Hobor *et al.,* 2011; Carroll *et al.,* 2004; Bacikova *et al.,* 2014). Rpb7 was previously shown to physically interact with Seb1, the *S. pombe* ortholog of Nrd1, and this interaction is

conserved in *S. cerevisiae.* Interestingly, in *S. pombe,* Rpb7 residue D167, which is conserved in *S. cereivisiae* Rpb7, was shown to be critical for maintaining interactions with Seb1 (Mitsuzawa *et al.,* 2003). In our studies, a mutation that altered D166 suppressed cryptic intragenic transcription and caused inviability with combined with *spt5-E546K,* which maps to a similarly solvent-exposed region of Spt5 KOW3. The identification of Nrd1 as both Rpb7 and KOW2-3 interacting, as well as Nab3 as KOW2-3 interacting, is highly suggestive of Rpb7 and KOW2-3 cooperatively forming a binding platform for the NNS termination complex. This is further supported by identification of genetic interactions between *spt5* and Nrd1 overexpression. We found that overexpression of Nrd1 results in suppression of the Spt- phenotype and *gal10⊿56* phenotype observed in *spt5* mutants.

Prior studies have shown that Spt5 and Rpb7 interacting factors have known roles in R-loop regulation, such as Spt16 and the NNS complex component Sen1 (Herrera-Moyano *et al.,* 2014; Martin-Alonso *et al.,* 2021). R-loops have been shown to be influenced by a number of cotranscriptional events, such as chromatin remodeling and termination (Al-Hadid, Yang 2016; Skourti-Stathaki *et al.,* 2011). Additionally, R-loops have been observed to be enhanced at Ty1 retrotransposons (Zeng *et al.,* 2021). These observations, coupled with the physical location of the Pol II Stalk region being between the mRNA exit channel and upstream DNA, led us to suspect that this region may play a role in mediating R-loop resolution or regulating R-loop homeostasis

(Ehara *et al.,* 2017; Bernecky *et al.,* 2017). As chromatin structure and 3' end formation have known connections to R-loop formation, we asked if dissolution of R-loops would influence the phenotypes observed in *rpb4, rpb7* and *spt5* mutants in the *gal10Δ56,* cryptic initiation and *lys2-128δ* backgrounds. Human RNaseH1 has previously been demonstrated to resolve R-loops *in vivo* (Parajuli *et al.,* 2017). Therefore, we utilized overexpression of *RNASEH1* to uncover the influence of R-loops on the mutant phenotypes analyzed in Chapter 2. Our results suggest a function for Rpb4/7 and Spt5 KOW2-3 in R-loop homeostasis.

The genetic interactions discussed in this chapter validate the physical interactions reported in Chapter 3 and support a model that Spt5's central KOW domains and Rpb4/7 function cooperatively from transcription elongation through termination. Thus, the Pol II Stalk region appears to function as a single unit, conducting processes involved in cotranscriptional chromatin remodeling, 3' end formation, and R-loop homeostasis by recruitment of tertiary factors involved in these processes such as the NNS complex. These results highlight the idea that all cotranscriptional processes likely depend on each other, and the Pol II Stalk region acts as a conductor of these processes throughout transcription. If you disrupt a single component of the transcription apparatus, you necessarily disrupt all processes that are intertwined with transcription.

Results

Nrd1 overexpression modestly suppresses Spt- and *gal10Δ56* phenotypes in *spt5-E546K*

Nrd1 was identified in our KOW domain affinity chromatography, full-length Spt5 chromatography as well as 2-hybrid and pull-down experiments with Rpb7. To test the functional relevancy of this physical interaction we transformed *spt5* mutants with a Nrd1 overexpression vector and assayed for changes in phenotypes observed in *spt5* single mutants. We observed modest suppression of both the *gal10Δ56* and the Spt- phenotype in *spt5-E546K.* (Figure 4-1).

Human RNaseH1 overexpression sensitizes cells to galactose, which is suppressed by mutations in the Pol II stalk region

We initially took a genetic approach to determine if the Pol II stalk and Spt5's central KOW domains play a role in R-loop regulation. If this region is indeed involved in regulating R-loop formation, it is possible that the phenotypes observed for *spt5* and *rpb4/7* mutations may be influenced R-loops as well. We transformed a high-copy plasmid expressing *RNASEH1* into strains harboring our *spt5* or *rpb4/7* mutations and assayed for suppression or enhancement of mutant phenotypes observed in the *gal10Δ56* and cryptic initiation reporters.

When assessing the impact of RNaseH1 overexpression cryptic initiation, we observed RNaseH1 overexpression induced galactose toxicity in

our wildtype control. To ensure this phenotype was not an artifact of the *pGAL1-FLO8::HIS3* cryptic initiation reporter, we transformed our overexpression plasmid into an old *OY* strain lacking this reporter and observed the same effect (data not shown). Strikingly, *spt5-G587D* and *spt5-G587D* + *rpb7-D166G* both strongly suppressed the galactose induced toxicity of RNaseH1 overexpression (Fig. 4-2A). Both of these mutations also moderately suppressed the slow-growth phenotype caused by RNaseH1 overexpression. The *rpb7* single mutant was not assayed due to auxotrophic marker incompatibility between plasmids.

As another control, we obtained strains from the W303 background and transformed them with either an RNaseH1 overexpression vector or empty vector as well. We assayed for growth on media containing either galactose or a mixture of galactose and raffinose. On media containing galactose only, we observed no growth. However, on media containing galactose and raffinose, we observed that RNaseH1 overexpression no longer conferred lethality (data not shown).

We also tested our *rpb4Δ* strain for suppression of galactose induced toxicity of RNaseH1 overexpression and observed the same effect, in addition to the partial suppression of the slow growth phenotype (Fig. 4-2B). There was minimal effect on the cryptic initiation phenotype in both cases. We also tested for suppression of the *gal10Δ56* phenotype observed in *spt5-E546K.*

Overexpression of RNaseH1 resulted in suppression of the *gal10Δ56* phenotype of *spt5-E546K.* The effect on *gal10Δ56* was less apparent (Fig. 4-2C, D)

Human RNaseH1 overexpression enhances Spt- phenotypes observed in *spt5* mutants and *rpb4Δ*

As R-loops tend to be enriched over Ty1 retrotransposons, we asked if RNaseH1 overexpression would have an influence on the Spt- phenotype observed at *lys2-128δ* in *spt5, rpb7* and *rpb4Δ* mutants (Zeng *et al.,* 2021). The *lys2-128δ* construct confers inability to grow on SC–Lys media in an otherwise wild-type cell due to the insertion of the delta element (i.e., the long terminal repeat) of a TY transposon in the 5' noncoding region of *LYS2* (Farabaugh, Fink 1980; Simchen *et al.,* 1984). Mutations that result in transcriptional defects cause suppression of the Lysine auxotrophy, although the precise mechanism remains undefined. Genome-wide screens for suppression of transcription defects caused by TY insertions at *LYS2* and *HIS4* were utilized to initially isolate the *spt* (Suppressor of TY) mutations (Winston *et al.,* 1983).

While *spt5-G587D* mutants display a very mild Spt- phenotype, the *spt5-G587D rpb7-D166G* double mutant displays a more enhanced Spt- phenotype (Figure 4-2A). When comparing the growth of the double mutant on SC –Leu –Lys to SC –Leu media, it is clear that the RNaseH1

overexpression further enhances the Spt- phenotype in *spt5-G587D, rpb4Δ* and *spt5-E546K*. We repeated this analysis with *rbp4Δ* and *spt5-E546K, rpb7-D166G* double mutants. The enhancement of Spt- phenotype remained consistent across all mutants tested.

To assess the question of whether suppression of RnaseH1 overexpression induced galactose toxicity was directly related to suppression of the *gal10Δ56* phenotype, or is instead related to general loss of Spt5 function, we assayed for suppression of this phenotype with *spt5-194, spt5-242,* and the chromatin remodeler *chd1Δ* (Figure 4-2E). Of these alleles, *spt5-242* is the only one that does not display a *gal10Δ56* phenotype (data now shown). Interestingly, all tested alleles suppressed the poor growth phenotype observed on galactose media when RNaseH1 is overexpressed.

Mutations in *spt5, rpb7* increase the formation of R-loops

While the observation that RNaseH1 overexpression exerts pleiotropic effects on our *spt5* and *rpb4/7* mutants supports the idea that this region is involved in R-loop homeostasis, it does not directly test the hypothesis that loss of function in this region results in increased R-loop formation. In order to address this question directly, we utilized the S9.6 monoclonal antibody specific to RNA/DNA hybrids to directly observe R-loops in lysed yeast nuclei via immunofluorescence. While the S9.6 antibody indiscriminately detects RNA:DNA hybrids, prior studies have shown that most instances of DNA:RNA

hybrids are in fact R-loops (García-Pichardo *et al.,* 2017; Yu *et al.,* 2003; Li, Manley 2017; Malig *et al.,* 2020). Recently, studies using human cells showed that cellular imaging studies utilizing the S9.6 antibody may be compromised by the antibody's tendency to recognize rRNA (Smolka *et al.,* 2021). A structural and biochemical analysis of the S9.6 antibody showed that DNA:RNA hybrids outcompeted double-stranded RNA (dsRNA) by over 100 fold (Bou-Nader *et al.,* 2022). However, the fact that dsRNA exists in a vast excess to R-loops in any given cell can account for a significant cytoplasmic S9.6 signal. To account for this, performed chromosome spreads. Prior studies have utilized a chromosome spread technique to reduce cytoplasmic background and expand the yeast nucleus (Loidl *et al.,* 1991; Lafuente-Barquero *et al.,* 2020; Yang, Zhang 2022). We utilized a modification of these protocols to observe R-loops *in vivo* in two *spt5, rpb7* double mutants.

Asynchronous yeast cultures were grown to mid-log phase in SC-Leu to select for maintenance of either a RNaseH1 overexpression vector or an empty vector. Cells were lysed and fixed on a microscope slide and stained with the S9.6 primary antibody followed by a Cy5 conjugated secondary antibody and imaged using a spinning-disc confocal microscope at 100x. Nuclei were then assayed for the presence of any amount of S9.6+ foci (Fig. 4-3A). The percent-positive S9.6 nuclei were calculated in two independent transformants with n>250 (Fig. 4-3B). We utilized *hpr1Δ* as a positive control which has previously been reported to display ~40% positive nuclei (Lafuente-

Barquero *et al.,* 2020). In wild type cells, we observed 22% positive nuclei, with RNaseH1 overexpression reducing this figure to 15%. This is consistent with previous reports in the literature (Lafuente-Barquero *et al.,* 2017). With *hrp1Δ,* we observed 45% positive nuclei reduced to 33% with RNaseH1 overexpression. When the *spt5-E546K, rpb7-D166G* double mutant was assayed, we observed a striking 61% S9.6+ nuclei reduced to 37% with RNaseH1 overexpression. In the *spt5-G587D, rpb7-D166G* double mutant we observed a similar effect with 57% of nuclei showing S9.6 signal reduced to 32% with *RNASEH1* overexpression.

Discussion

In this chapter we performed a genetic analysis of the mutations in *spt5* and *rpb4/7* identified in previous chapters with factors involved in noncoding transcript termination and R-loop resolution. We uncovered suppression of the *gal10Δ56* and spt- phenotypes observed in *spt5-E546K* conferred by of overexpression of Nrd1. We further demonstrate that the R-loop specific RnaseH1 sensitizes cells to galactose and that *rpb4Δ* and *spt5* mutations that suppress *gal10Δ56* also suppress the galactose sensitivity induced by RnaseH1 overexpression. We observed only a modest impact on the cryptic initiation phenotypes, but *spt5* / *rpb7* double mutants were still able to suppress the Galactose sensitivity caused by *gal10Δ56* with concurrent RnaseH1 overexpression. Further, RnaseH1 overexpression enhanced the

Spt- phenotype observed in *spt5* mutants. These results were validated by the direct observation of R-loop formation in *spt5* and *rpb7* mutants and subsequent abrogation of R-loops upon RnaseH1 overexpression.

Nrd1 overexpression influences transcription through Ty1 retrotransposons and polyadenylation site choice

Spt5 has previously been shown to play a role in the recruitment of 3' processing factors for polyadenylated transcripts, and in this study is shown to play a role in polyadenylation site choice (Mayer *et al.,* 2012). A role for Spt5 in noncoding transcript termination, however, is less well-defined. Nrd1 is a main component of the NNS noncoding RNA termination complex and has previously been shown to interact physically with Rpb7 at the solvent-exposed tip region at D167 in *S. pombe* (Mitsuzawa *et al.,* 2003). Prior proteomics studies of Nrd1 have also demonstrated a physical interaction with Spt5, and we identified both Nrd1 and Nab3 as binding to Spt5 KOW2-3 (Vasiljeva, Buratowki 2006). A recent study in *Paramecium tetraurelia* has demonstrated that depletion of Spt4/5 results in a reduction in scan RNAs (scnRNAs) in the germline, suggesting that a role for Spt5 in noncoding transcript termination may be conserved (Gruchota *et al.,* 2016). Our genetic results suggest, taken together with previous studies with Rpb7 and Nrd1 as well as structures predicted by AlphaFold, that the Pol II Stalk region composed of KOW2/3 and Rpb4/7 likely serves as a docking site for the NNS

complex, and the NNS complex functions in transcription through Ty1 retrotransposons and polyadenylation site choice.

Overexpression of Nrd1 resulted in suppression of the Spt- phenotype and the *gal10Δ56* phenotype observed in *spt5-E546K.* As the Spt- phenotype is purportedly caused by transcriptional defects, it stands to reason that overexpression of Nrd1 partially restores the transcriptional function that is disrupted by the *spt5* mutants. The mechanism behind the Spt- phenotype remains obscure. The galactose sensitivity caused by *gal10Δ56* results from transcriptional interference at the *GAL7* promoter (Kaplan *et al.,* 2005). It is possible that overexpression of Nrd1 results in enhanced termination via the NNS pathway resulting in fewer polymerases reaching *GAL7,* allowing polymerases to form at the *GAL7* promoter and formation of the independent, *GAL10, GAL7* gene products. This supports the idea that Spt5 KOW2-3 and Rpb4/7 form a platform recognized by the NNS complex and that this region supports proper termination of transcripts.

The genetic results we report from overexpression of Nrd1 are mild. This may be because Nrd1 does not directly dock onto Spt5 KOW2-3 as predicted by AlphaFold Multimer, but rather directly interacts with Rpb7 alone. AlphaFold Multimer predicted Nab3, however, to bind directly to KOW2-3 through its RRM. This may explain the mild genetic suppression we observe here, and it would be worthwhile to test for suppression of *spt5* phenotypes

with overexpression of Nab3, rather than Nrd1. If the predicted model is correct, we would expect to see stronger suppression of *spt5-E546K* phenotypes with Nab3 overexpression over Nrd1.

The Pol II Stalk Region Regulates R-Loop Formation

Two observations led us to consider the role of the Pol II stalk in R-loop regulation. Based on structural models, it appears that the Pol II stalk may serve as a buffer region between nascent mRNA and upstream DNA, analogous to bumpers in a bowling lane, serving to physically separate the mRNA from upstream DNA (Bernecky *et al.,* 2017; Ehara *et al.,* 2017). This is supported by the mRNA binding properties of these factors as well (Meka *et al.,* 2005; Orlicky *et al.,* 2001; Zuber *et al.,* 2018; Gilmour *et al.,* 2017; Meyer *et al.,* 2015). The second observation was the identification of KOW-associated factors that were revealed by protein affinity chromatography – many of these proteins are helicases, exonucleases, and RNA binding proteins. Some of these factors have been identified in prior studies to be involved in R-loop regulation – such as Dbp2, Mtr4, Xrn1, Pap2, FACT, Pbp1, and Sen1 (Cloutier *et al.,* 2016; Herrera-Moyano *et al.,* 2014; San Martin-Alonso *et al.,* 2021; Chan *et al.,* 2014; Lim *et al.,* 2017; Salvi *et al.,* 2014; Gavalda *et al.,* 2013). Thus, it is possible that the Pol II Stalk region both acts to prevent aberrant R-loops from forming as well as primes the elongation complex for timely resolution of R-loops as they occur.

Our surprising finding that RnaseH1 overexpression results in galactose sensitivity may be a reflection of R-loops' roles in mediating transcript termination. It has been shown previously that R-loops are enriched over polyadenylation sites, and loss of function in 3' processing factors increase R-loops (Stirling *et al.,* 2012; Skourti-Stathaki *et al.,* 2011). This would be consistent with R-loop dependent termination at *GAL10.* R-loops over the *GAL10* poly(A) site may slow down the polymerase and prime it for termination. By attenuating the levels of R-loops of the *GAL10* poly(A) site, we may be promoting read-through of the poly(A) site, which results in transcriptional interference at the *GAL7* promoter, similar to the effect of *gal10Δ56* (Kaplan *et al.,* 2005). If this were the case, however, one would expect that cells overexpressing RNaseH1 would also be unable to grow on media containing a mixture of galactose and rafinose due to the presence of the toxic galactose 1-phosphate, which is metabolized by the *GAL7* gene product, as is the case with *gal10Δ56* (Greger, Proudfoot 1998; Douglas, Hawthorne 1964; Kaplan *et al.,* 2005). This was not the case however, as we observed cells overexpressing RNaseH1 were able to grow, albeit slowly, on media containing galactose and raffinose, suggesting an alternative mechanism explaining the cells inability to grow on galactose when RNaseH1 is overexpressed. Interestingly, all tested alleles in *spt5* that suppress *gal10Δ56* also suppressed RNaseH1 overexpression induced galactose toxicity. In addition to these alleles, we also observed the cold sensitive allele

spt5-242 also suppressed RnaseH1 overexpression induced galactose toxicity, which does not suppress *gal10Δ56.* This suggests that RNaseH1 overexpression induced galactose toxicity is being suppressed by an alternative mechanism independent of reducing transcriptional interference at *GAL7*, as is the case with *gal10Δ56*, and further suggests that the *spt5* and *rpb7* alleles that suppress *gal10Δ56* may do so by a separate mechanism more directly related to the polyadenylation machinery. The observation that *chd1Δ* also suppresses RnaseH1 overexpression induced galactose toxicity, in addition to *spt5-242,* argues in favor of local chromatin state playing a role in the regulation of R-loops. Indeed, it has been shown previously that formation of R-loops can influence local chromatin state, and vice versa (Skourti-Stathaki *et al.,* 2014; Al-Hadid *et al.,* 2016; García-Pichardo *et al.,* 2017). R-loops have also been shown to promote transcript initiation, both at sense and antisense genes, by promoting an open local chromatin state (Rondón, Aguilera 2019; Manzo *et al.,* 2018; Chen *et al.,* 2015). RNaseH1 overexpression may be attenuating transcription initiation across the entire galactose regulon by promoting closed chromatin at the promoters, and mutations in *spt5* and *chd1Δ* may be disrupting promoter-proximal chromatin to restore transcript initiation. This model is supported by the fact that all tested mutations that suppress RNaseH1 induced galactose toxicity also result in in cryptic initiation, but not all tested cryptic initiation mutants suppress *gal10Δ56*. An alternative, more trivial model may be that mutations

causing defects in transcription may interfere with the overexpression of RNaseH1. This is supported by the observation that in our S9.6 immunofluorescence experiments, as RNaseH1 overexpression does not suppress the appearance of S9.6 foci completely to wild type levels. However, they do significantly reduce the appearance of S9.6 foci, suggesting that transcriptional defects interfering with RNaseH1 overexpression cannot completely explain the phenotypes we observe. Further experiments on R-loops' influence on the galactose regulon are warranted and will be necessary to parse out the mechanism behind this phenomenon.

It has been reported previously that R-loops are enriched over Ty1 transposons (Zeng *et al.,* 2021). Insertions of Ty1 transposons or their long terminal repeats in the 5' regions of auxotrophic markers have been used to assay for disruptions in transcription such as in the Spt class of genes (Winston *et al.,* 1983; Simchen *et al.,* 1984). Generally speaking, mutations that result in transcriptional defects are reported to suppress the auxotrophy caused by Ty1 insertions at *HIS3* and *LYS2*. In this work, we report that *rpb4Δ, spt5-E546K, spt5-G587D,* and *spt5/rpb7* double mutants also exhibit this phenotype. Interestingly, our data shows that RnaseH1 overexpression enhances the Spt- phenotype observed in these mutants. Although the mechanism of the Spt- phenotype has yet to be fully described, these data do argue in favor of a role for R-loops in transcription through Ty1 elements.

Taken together, our results support a model in which Spt5 and Rb7 collaboratively regulate the formation of R-loops, which influence termination and polyadenylation at the *GAL10* gene as well as transcription through Ty1 retrotransposons. We have identified Nrd1 as a factor that binds to the Pol II Stalk region and have attributed this interaction to suppression of phenotypes that are exhibited by *spt5* and *rpb7* mutants that are also effected by RNaseH1 overexpression, such as Spt- and *gal10Δ56.* The location and biochemical nature of the *spt5* and *rpb7* mutations used in this experiment suggest disruption of interactions with tertiary factors. Taken with the predictions from Chapter 3, it is likely that *rpb7-D166G* and *spt5-E546K* are together breaking a contact with the NNS complex, resulting in failure to recruit the Sen1 R-loop specific helicase. This does not preclude the notion that other factors may be associating with this region in a temporally separated manner. Differential recruitment of factors is likely regulated by the phosphorylation cycle of the Pol II CTD and Spt5's CTR or other PTMs on Spt5 or Rpb7 that have yet to be characterized. Future studies should aim to test a direct role for Nrd1 and Nab3 in regulating R-loop homeostasis, either alone or in conjunction with Sen1. As the NNS complex is preferentially recruited by the phospho-Ser_5 form of the Pol II CTD, it is likely that other R-loop resolvases associate with the Pol II EC in a phosphor-Ser_2 specific manner (Vasiljeva *et al.,* 2008).

Figure 4-1: Nrd1 overexpression suppresses Spt- and *gal10Δ56* phenotypes in *spt5-E546K*

NRD1 high copy plasmid was transformed into a strain containing integrated *spt5-E546K.* Strains were spotted onto the indicated media and allowed to grow for 3 days.

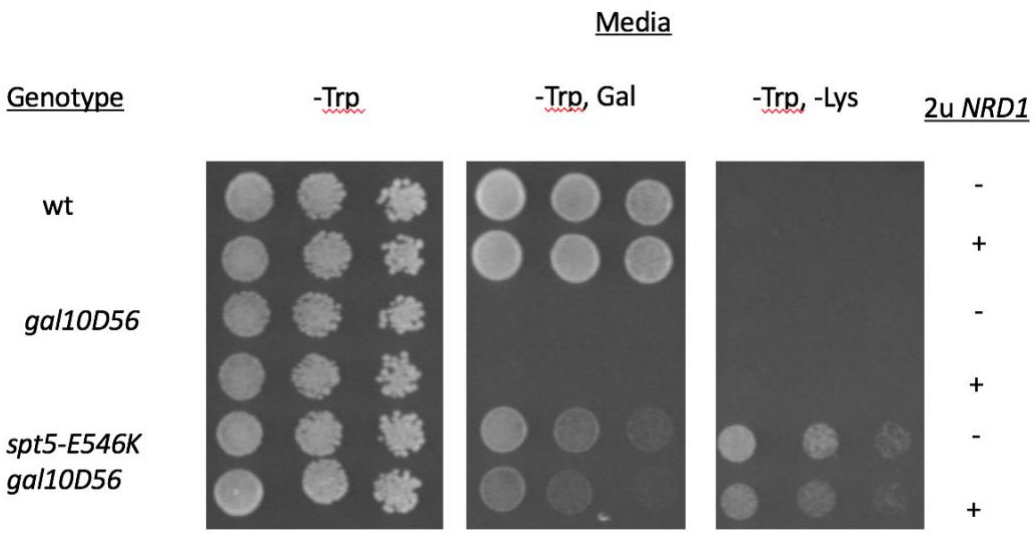

Media
Genotype
-Trp
-Trp, Gal
-Trp, -Lys
2u NRD1
wt
gal10D56
spt5-E546K
gal10D56
-
+
-
+
-
+

Figure 4-2: RNaseH1 overexpression sensitizes cells to galactose media and enhances Spt- phenotypes, while *spt5, rpb4/7* mutations suppress RNaseH1 induced galactose sensitivity.

A) *RPB7* shuffle strains with integrated *spt5-G587D, pGAL-FLO8::HIS3* and *lys2-128δ* were transformed with *rpb7-D166G LEU2* and subject to counter-selection against *RPB7 URA3* with 5FOA. They were then transformed with a linear *TRP1* marker-switch plasmid generated via PCR and selected for Trp+ colonies, and then confirmed to be stable Trp+, Leu- via restreak onto YPD and replica plating onto SC-Leu and SC-Trp. Strains were then transformed with a *LEU2* marked overexpression vector containing human *RNASEH1* or empty vector and serially diluted on the indicated media. Colonies were allowed to grow for 3 days at 30C.

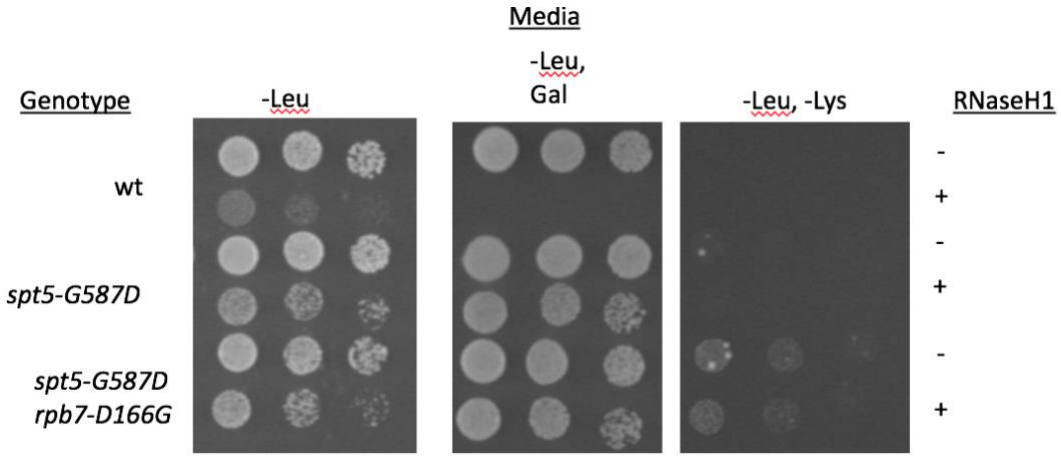
B) A strain containing *rpb4Δ* in the *pGAL-FLO8::HIS3, lys2-128δ* background with *RPB7 URA3* on an overexpression vector was transformed with a *LEU2* marked overexpression vector containing *RNASEH1* or empty vector, serially diluted onto the indicated media and allowed to grow for 3 days at 30C.

C) Strains were generated using the same strategy as in A), but in the *gal10Δ56, spt5-E546K, lys2-128δ* background. Strains were serially diluted onto the indicated media and allowed to grow for 4 days.

D) A strain containing *rpb4Δ* in the *gal10Δ56* background with *RPB7 URA3* on an overexpression vector was transformed with a *LEU2* marked overexpression vector containing *RNASEH1* or empty vector, serially diluted onto the indicated media and allowed to grow for 3 days at 30C.

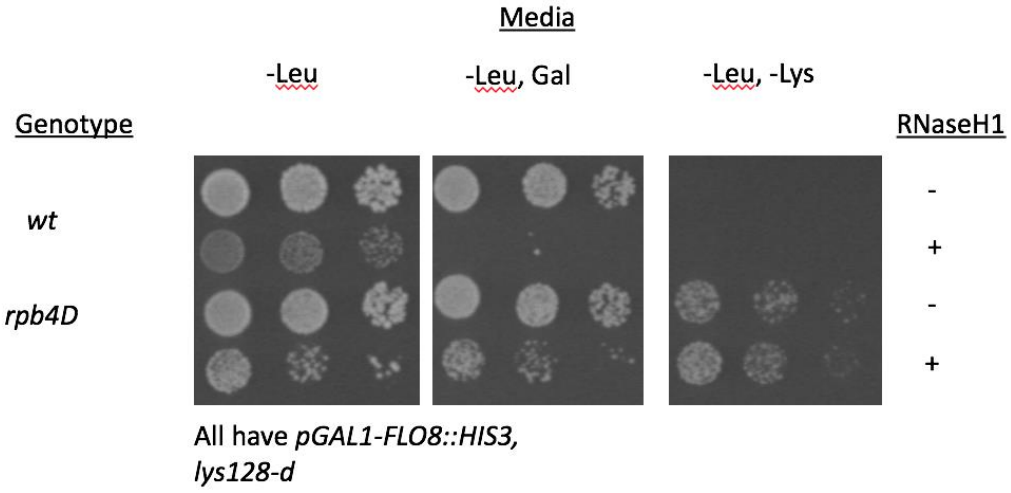
E) Strains with the indicated genotype were transformed with a *LEU2* marked overexpression vector containing *RNASEH1* or empty vector, patched onto SC –Leu and replica plated onto SC –Leu and SC –Leu + Galactose and allowed to grow for 2 days.

A

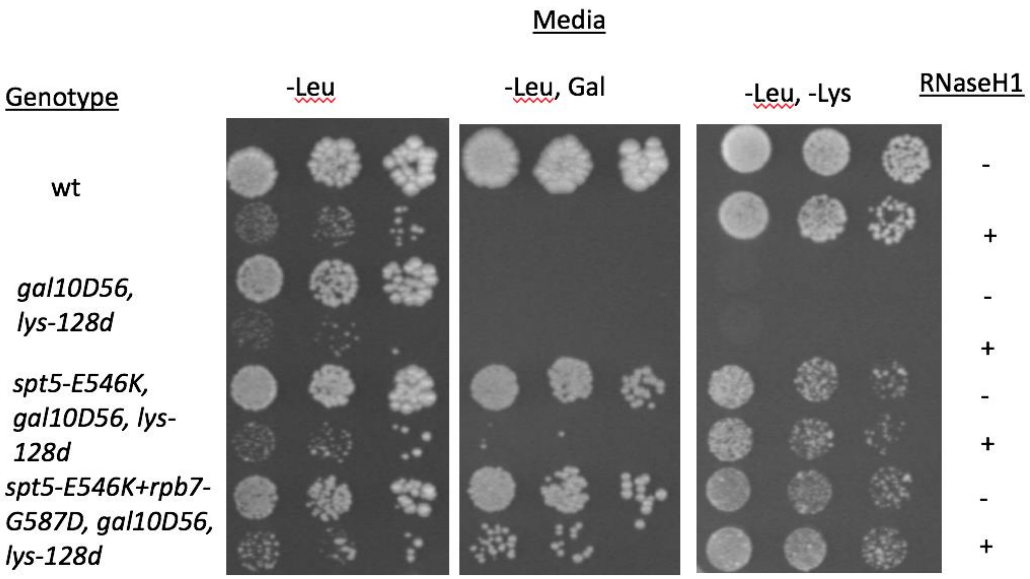


All have *pGAL1-FLO8::HIS3, lys128-d*

B

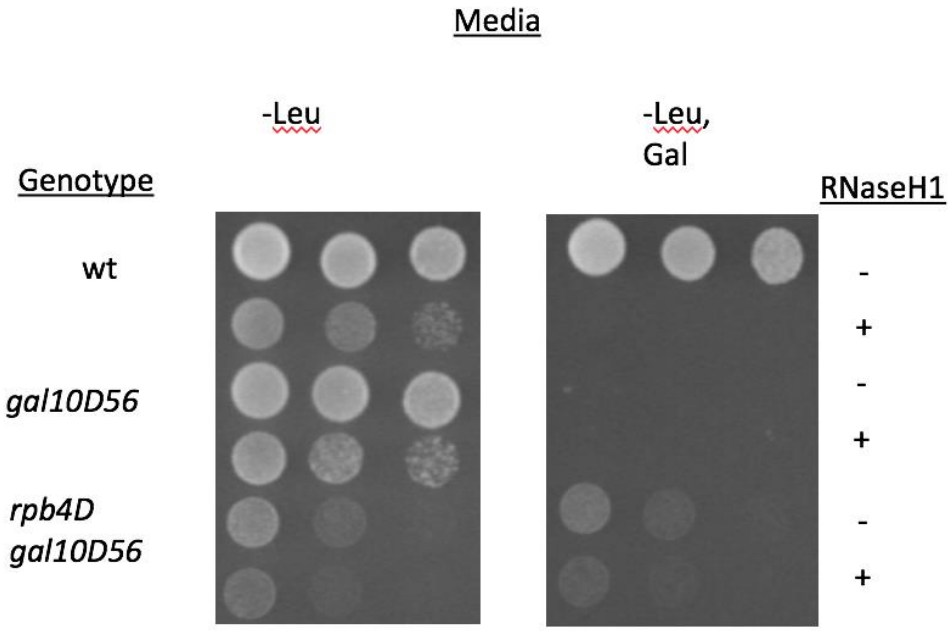

Media
-Leu
-Leu, Gal
-Leu, -Lys
Genotype
RNaseH1
wt
rpb4D
-
+
-
+
All have pGAL1-FLO8::HIS3, lys128-d

C



All have *lys2-128d*

D


Media
-Leu
-Leu, Gal
Genotype
RNaseH1
wt
gal10D56
rpb4D gal10D56
-
+
-
+
-
+

E

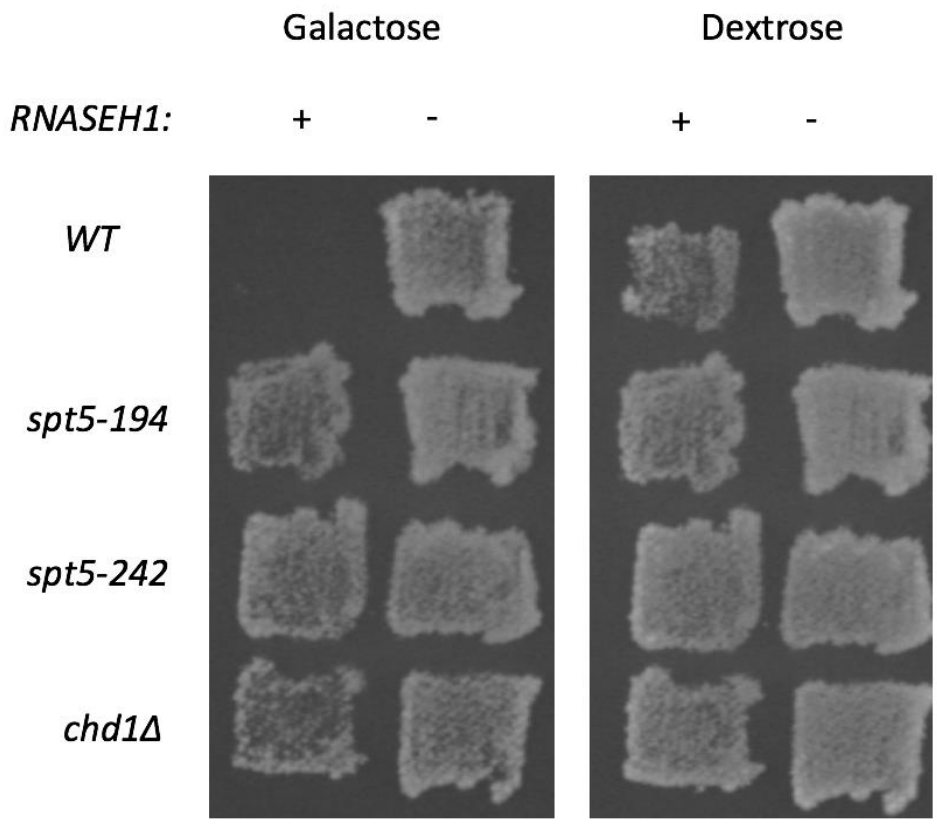
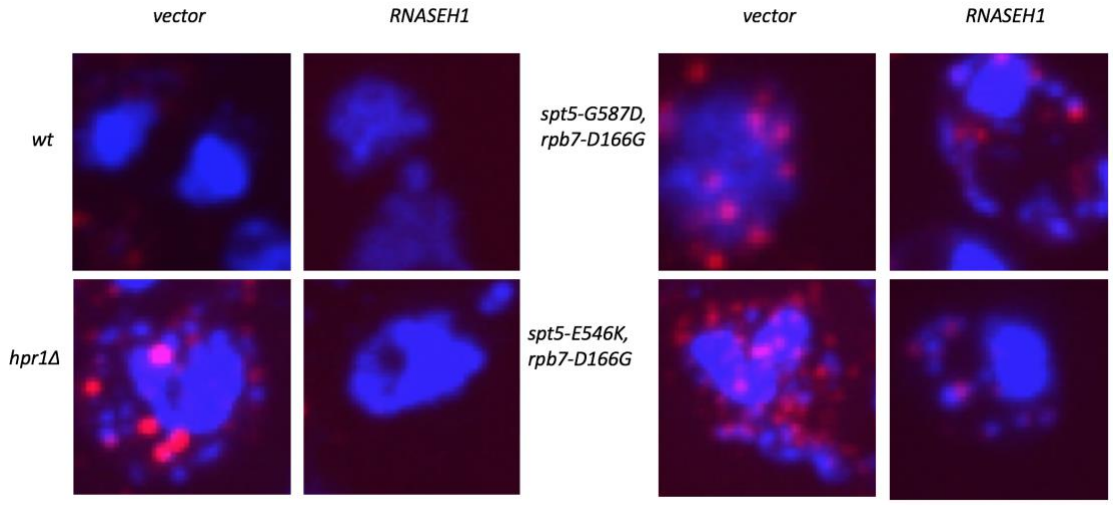


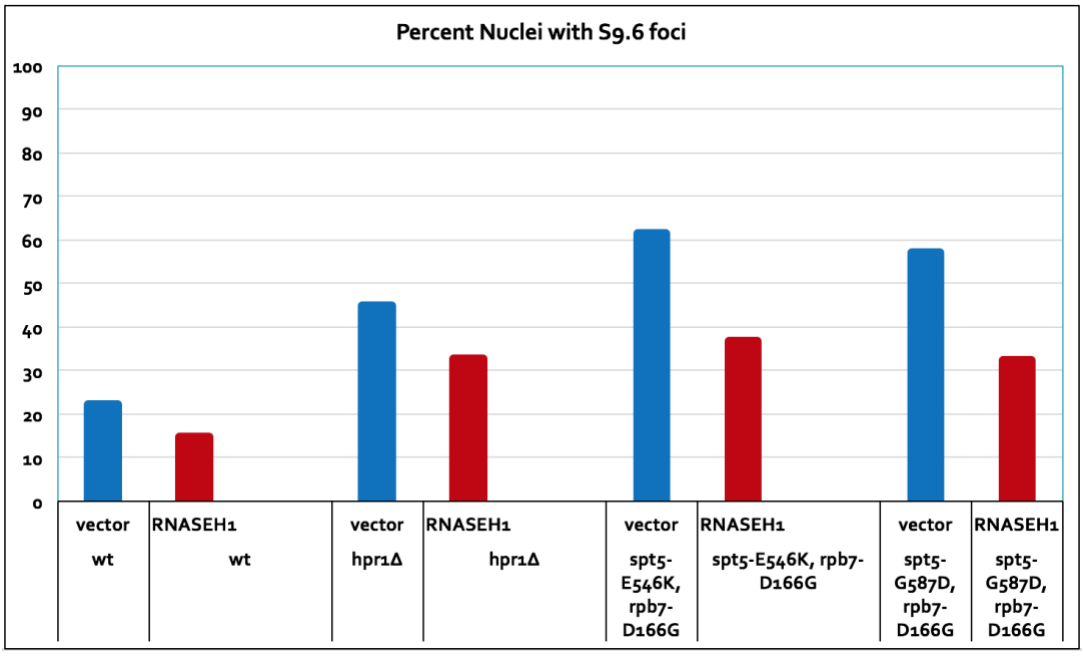
Figure 4-3: *spt5, rpb7* double mutants result in increased R-loop accumulation

A) S9.6 foci detected by immunofluorescence on yeast chromosome spreads in WT, *hcr1Δ, spt5-E546k* + *rpb7-D166G* and *spt5-G587D* + *rpb7-D166G* strains harboring an *RNASEH1* overexpression vector or empty plasmid. Red foci represent DNA:RNA hybrids detected by the S9.6 antibody, DNA is stained in blue (DAPI). Images were all taken during the same imaging session on a confocal microscope under the same settings.
B) Quantification of nuclei containing S9.6 foci is shown from two biological replicates with n>250 for each replicate.

A


vector
RNASEH1
vector
RNASEH1
wt
spt5-G587D,
rpb7-D166G
hpr1Δ
spt5-E546K,
rpb7-D166G

B


Percent Nuclei with S9.6 foci
100
90
80
70
60
50
40
30
20
10
0
vector
wt
RNASEH1
wt
vector
hpr1Δ
RNASEH1
hpr1Δ
vector
spt5-E546K, rpb7-D166G
RNASEH1
spt5-E546K, rpb7-D166G
vector
spt5-G587D, rpb7-D166G
RNASEH1
spt5-G587D, rpb7-D166G

Chapter 5

Concluding Remarks

All Spt5 KOW domains adopt a stable fold containing the beta-barrel characteristic of Tudor domains. This structural motif is thought to facilitate protein-protein and protein-nucleic acid interactions (Meyer *et al.,* 2015; Lasko 2010; Ying, Chen 2012. Structural studies have shown that the KOW domains form stable and independently folding sub-structures in KOW1-Linker1, KOW2-KOW3, Linker2-KOW4, and KOW5 that are distributed across the surface of Pol II (Meyer *et al.,* 2015; Bernecky *et al.,* 2017). The KOW1-Linker1 region functionally overlaps with Spt4 in its nucleic acid binding properties through a positively charged patch on its surface and is structurally primed to aid in sealing the central cleft of Pol II. Deletion of KOW2 results in a severe growth defect at 25C and lethality at 30C (Viktorovskaya *et al.,* 2011). Deletion of KOW4-5 also results in defects that include temperature sensitivity, sensitivity to mycophenolic acid, slow-growth, and a loss of binding to Rpb4/7 (Li *et al.,* 2014). This study also shows that a KOW4-5 deletion results in lower Pol II occupancy at the 3' end of *RPB2*, a phenotype similar to that shown in *rpb4Δ.* Further, Spt5 KOW4-5 were shown to extensively cross link with Rpb4 and Rpb7. Despite these lines of evidence, evidence for a direct functional cooperation between Spt5 and Rpb4/7 has not previously been reported.

This thesis presents a genetic and biochemical analysis of the functional interaction between Spt5's central KOW domains and the Pol II stalk, aiming to elucidate the phenotypic consequences of disruption of the structure composed of KOW2-4 and Rpb4/7 that was observed in previous cryo-EM studies and cross-linking experiments. We present genetic and biochemical evidence that Spt5's central KOW domains functionally overlap with that of the Pol II stalk, Rpb4 and Rpb7. We have identified mutations that alter amino acids at multiple points throughout the Pol II Stalk/KOW structure, including the juncture of KOW3/Rpb7 and KOW4/Rpb7, as well as multiple solvent exposed regions on the surface of this structure. We used a set of genetic reporters that provide insight into both polyadenylation site choice and the production of cryptic-intragenic transcripts that likely result from the disruption of chromatin. Many of these mutations genetically interact with each other, resulting in synthetic sickness and enhancement of phenotypes. Interestingly, mutations in *rpb4* and *rpb7* both share the cryptic transcription phenotype, suggesting that the core structure of Pol II itself evolved to assist in overcoming nucleosomal barriers to transcription.

We identified a large set of Spt5 interacting proteins that specifically interact with the KOW2-3 and Linker2-KOW4 regions of Spt5, many of which have also been previously identified in Rpb7 proteomics studies (Mosley *et al.,* 2013). Many of these factors have been implicated in R-loop resolution. We found mutations that alter solvent-exposed amino acids in Spt5 and Rpb7

result in increased R-loops, supporting the idea that this region regulates R-loop homeostasis possibly through the recruitment of tertiary factors.

Prior structural models suggested a physical interaction of KOW2-4 with the dissociable core polymerase subunits Rpb4 and Rpb7, which together comprise the Pol II stalk, and possibly form a cooperative binding platform (Bernecky *et al.,* 2017). The location of KOW2-4 near the mRNA exit channel argue against a direct modulation of Pol II's catalytic cycle by the central KOW domains and suggest a potential role in mRNA processing. Combining a genetic and biochemical approach, we sought to elucidate the phenotypic consequences of disrupting the structure formed by Rpb4/7 and KOW2-4 and further resolve Spt5's function in this particular region. Consistent with the structural proximity to Rpb4 and Rpb7, we observed overlapping phenotypes for transcription elongation, polyadenylation site choice and cryptic internal transcription. MudPIT mass spectrometry, performed in collaboration with John Yates, identified a number of factors that bind to both Rpb7 and KOW2-3/L2K4, further suggesting that this region forms a cooperative binding platform whose maintenance is required for elongation integrity.

We propose a model in which Rpb4, Rpb7 and Spt5 KOW2-4 form a structure that acts as a direct extension of the Pol II Elongation Complex, recruiting proteins that function during transcription elongation and termination. It is likely that this structure targets varying factors in response to

Spt5/Rpb1 phosphorylation state and to the stage of the transcription cycle in which the elongation complex currently resides. Nrd1 in particular binds the Spt5-CTR – it is likely that both the Spt5 and Rpb1 C-terminal domains recruit elongation and termination factors to the Pol II EC, which in turn are deposited onto the Spt5/Stalk structure where they execute their functions (Leporé *et al.,* 2011). Nrd1 may serve to anchor the NNS complex in the early stage of elongation, priming the elongation complex for termination of short-noncoding transcripts, monitoring for aberrant transcription, or actively regulating the formation of R-loops through the Sen1 DNA/RNA helicase.

As Pol II traverses the length of the gene, the Stalk Region and its interacting factors may be remodeled to better serve the elongation complex in response to other transcriptionally relevant challenges, such as chromatin remodeling, mRNP packaging, and polyadenylation. Part of this remodeling procedure may involve a shift in conformation of Spt5's central KOW domains or transient dissociation of Rpb4/7. Supporting this idea is our observation that *rpb7-G149D* does not share the same phenotype as *spt5-G587D.* As Rpb7-G149 and Spt5-G587 are in apparent direct contact, if this were the only relevant conformation, one would expect disruption of this binding event by mutating either *rpb7* or *spt5* would result in the same mutant phenotype. However, we observe that *rpb7-G149D* only results in a *gal10Δ56* phenotype, while *spt5-G587D* confers defects both in *gal10Δ56* and cryptic initiation.

While it remains controversial whether Rpb4/7 have a role outside of their context with Pol II, there are multiple lines of evidence to suggest that Rpb4/7 reversibly dissociate from transcription complexes. One study demonstrates that Pol II complexes that lack Rpb4/7 preferentially contain a subset of elongation factors including Rtr1, Npa3, Set2 and Asr1 (Mosley *et al.,* 2013). Another study showed that Rpb4/7 function in the cytoplasm in mRNA export and translation (Harel-Sharvit *et al.,* 2010). It was further shown that dissociation of Rpb4/7 support mRNA decay, cell proliferation and stress response (Duek *et al.,* 2018). Under optimal growing conditions, Rpb4 exists in complex with Pol II in a sub-stoichiometric amount, and Rpb4 preferentially associates with Pol II during stationary phase (Rosenheck, Choder 1998; Choder, Young 1993; Kolodziej *et al.,* 1990). Finally, ChIP studies examining Rpb4/7 show Rpb7 having a strong association at transcription initiation sites with reduced of association along gene bodies, and further increased association at the 3' extremity (Cojocaru *et al.,* 2008). During heat shock, Rpb4/7 are detected at equal levels relative to Rpb11, indicating a role for the heterodimer during stress. They interpret this as Rpb4/7, under normal conditions, engages in alternate conformations across the body of the gene. Another study also showed that Rpb4 shows variable occupancy patterns relative to Rpb3, further supporting the model that Rpb4/7 may engage in alternate conformations across the body of genes (Verma-Gaur *et al.,* 2008). Taken in the context of the mutations we identified, *rpb7-G149D* and *spt5-*

G587D may promote a transition to a Stalk-less polymerase by breaking the interaction between Rpb7 and Spt5 KOW2-3.

An alternative model to these mutations promoting dissociation of Rpb4/7 could be that they are facilitate an alternate conformation of Spt5 within the Pol II EC. This is supported by two observations: 1) Spt5's KOW2-3 domain has remained unresolved in at least one cryo-EM study, suggesting this domain may be mobile (Ehara *et al.,* 2017). 2) Unpublished crystallographic results from Jinhua Fu suggest that Spt5's central domains may exist in multiple different conformations relative to Pol II (Jianhua Fu, unpublished results). Further studies are necessary to differentiate between these two models.

Future studies should aim to elucidate the precise molecular rearrangements that are likely to occur throughout elongation, with the Pol II Stalk Region serving as a nexus for many of these rearrangements. Certain factors are likely to associate with the Stalk Region only under specific conditions, and future experiments ought to take into consideration the local environment and how the elongation complex responds to cis and trans stimuli. Such factors may include sequence-specific motifs in the mRNA, the phosphorylation state of Spt5, Rpb1 and other components of the EC, gene length and local chromatin state. In metazoans, it is likely that the Pol II Stalk Region is targeted by various context-dependent transcriptional modulators that act as developmental regulators. It will be important to consider the

stable interaction between Spt5 KOW2-4 and Rpb4/7 as a functionally relevant target for these factors, and future studies should consider that this structure collaboratively acts to provoke interactions with tertiary factors. Since Rpb4/7 are known to act in transcription initiation, while Spt5 is not, Rpb4/7 likely bridges further interactions with tertiary factors prior to elongation that are sterically occluded once elongation begins.

This study has identified many proteins that interact physically with Spt5's central domains and additionally display genetic interactions that are relevant to the Pol II Stalk region. We argue that the Pol II Stalk region likely plays roles throughout transcription elongation, including early termination via NNS, cotranscriptional chromatin structure maintenance, mRNP packaging, polyadenylation and R-loop homeostasis by bridging interactions with factors identified in this study. It will be important in the future to integrate the many pleiotropic functions in this region and further resolve the specific contexts in which functions take place. Future studies can utilize this data to temporally separate the association of factors as a function of the transcription cycle.

Appendix – 1

Allele specific interactions between *rpb7* and *hcr1Δ.*

Hcr1 is a component of Eukaryotic Initiation Factor 3 (eIF3). eIF3 functions in translation initiation, termination, and ribosomal recycling (Valášek *et al.,* 2001a, 2001b, 2017). Hcr1 has dual, spatially separated roles; it facilitates eIF3's function in translation and ribosomal biogenesis in the nucleus. In the translational context, Hcr1 functions to bridge interactions between eIF1, 2, 3, 5 and $tRNA^{Met}$ (Phan *et al.,* 1998, 2001). In ribosome biogenesis, Hcr1 aids in 20S pre-rRNA processing (Valášek *et al.,* 2001a). Prior studies have shown that Rpb4 and Rpb7 physically interact with Hcr1 and *rpb4Δ* interacts with *hcr1Δ*, arguing a case for Rpb4/7's role in stimulating translation (Harel-Sharvit *et al.,* 2010). This is consistent with the model that Rpb4/7 evolved as a tool to recouple transcription and translation as eukaryotes diverged from archaea. As we have demonstrated an intimate linkage between Spt5 KOW2-4 and Rpb4/7, we wanted to ask

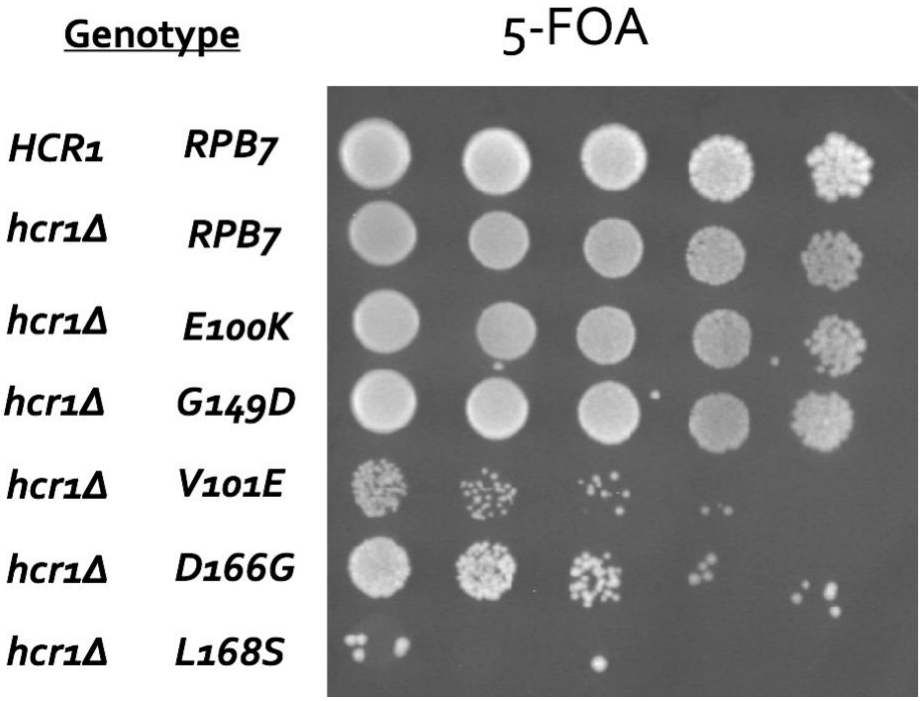
To further characterize Rpb7's function with Hcr1 we created and *hcr1Δ rpb7Δ* containing a *RPB7 URA3* plasmid. Mutant alleles of *rpb7* were transformed into these strains and loss of the wildtype plasmid was selected for on 5FOA. Deletion of *hcr1* results in a slight growth defect on its own. We observed synthetic sickness in 2 *rpb7* alleles (*rpb7-D166G, -L168S)* and synthetic lethality in *rpb7-V101E* (Fig. A1-1).

NusG, the prokaryotic ancestor of Spt5, physically couples transcription and translation (Burmann *et al.,* 2010). As eukaryotes diverged from prokaryotes, the evolution of the nucleus necessitated the invention of new processes to recouple the previously physically linked steps in gene expression. Rpb4/7 play roles in mRNA export, mRNA decay, and mRNA processing (Harel-Sharvit *et al.,* 2010; Mitsuzawa *et al.,* 2003). It is appealing to speculate that Rpb4/7 and Spt5's expanded central KOW domains evolved as components of the transcription complex to aid in the recoupling of transcription and translation. One study suggests that Rpb4/7 plays a role in stimulating translation by virtue of *rpb4/7* mutants with the following evidence: 1) *rpb4/7* mutants display defects in facilitating the transfer of mRNA from P-bodies to polysomes during transition out of stationary phase. 2) the same mutants display sensitivity to the translational inhibitors paramomycin and cycloheximide 3) *rpb4/7* mutants cannot tolerate overexpression of the translation initiation repressor Caf20p 4) Rpb4/7 and Hcr1 interact with each other physically via pull-down experiments. 5) Rpb4 was also shown to physically interact with Nip1, another component of eIF3. 6) Rpb4/7 can be found localized in the cytoplasm with P-bodies (Harel-Sharvit *et al.,* 2010). This evidence supports their hypothesis that Rpb4/7's interaction with Hcr1 promotes translation. Hcr1 also functions in ribosome biogenesis, as *hcr1Δ* results in a delay in the processing of 20S pre-rRNA to 18S rRNA (Valášek *et al.,* 2001a). It is unlikely, however, that Rpb7 functions with Hcr1 in the

context of ribosome biogenesis as Rpb4/7 are not known to interact with RNA Polymerase I, which has its own paralogs of Rpb4/7. It is therefore likely that Rpb4/7's physical interaction with Hcr1 occurs in the cytoplasm to support translation.

Figure A1-1: *rpb7* displays allele specific genetic interactions with *hcr1Δ*

Diploid strain harboring *hcr1Δ* from Stanford Yeast Deletion Collection was sporulated and crossed with *RPB7* shuffle strain. Plasmids containing *rpb7* mutants were transformed in and subject to 5FOA counter selection against *RPB7 URA3* plasmid. Strains were serially diluted onto 5FOA and allowed to grow at 30C for three days.


Genotype
5-FOA
HCR1 RPB7
hcr1Δ RPB7
hcr1Δ E100K
hcr1Δ G149D
hcr1Δ V101E
hcr1Δ D166G
hcr1Δ L168S

Appendix – 2

Spt5 and Rpb7 impact mRNA export through Mex67

Previous evidence has shown that general loss of function *spt5* mutants suppress the mRNA export defect of the temperature sensitive *mex67-6* allele (Burckin *et al.,* 2005). In this allele, elevated temperatures result in nuclear accumulation of Poly(A)+ mRNA as seen by FISH. This export defect is suppressed by an allele in *spt5* that abolishes Spt5 binding, which is thought to destabilize Spt5's structure (Ding *et al.,* 2010). Mex67 has previously been shown to physically interact with Rpb7, and Rpb4/7 have a well-documented role in participating in mRNA export pathways (Mosley *et al.,* 2013; Harel-Sharvit *et al.,* 2010). Because of this, and due to Rpb4/7's proximity to the mRNA exit channel along with Spt5 KOW2-5, we took a genetic approach to determining if *SPT5's* interaction with *mex67-6* could be pinpointed to the central KOW domains (Bernecky *et al.,* 2017; Ehara *et al.,* 2017). We also performed a similar genetic analysis with *rpb7* loss of function alleles.

First, we created double mutants by crossing strains containing integrated *spt5-E546K, spt5-194, spt5-242* and *spt5D5* (which deletes Spt5's CTR) with a *mex67* deletion strain. Since *MEX67* is an essential gene, the deletion strain is covered by a wild type *MEX67* plasmid (Segref *et al.,* 1997).

We then used a plasmid shuffle strategy to assay interactions between our integrated *spt5* mutants and *mex67-6.* We saw mild suppression of the temperature sensitive phenotype at 35C in all tested double mutants compared to *mex67-6* alone (Fig. A2-1A). The suppression of temperature sensitivity does not extend to suppression of the mRNA export defect seen in *mex67-6,* suggesting that suppression of the Ts phenotype either occurs through an indirect mechanism or is too mild to be observed via FISH (data not shown).

To further assess the overlapping function of Spt5 and Rpb7, we used a plasmid shuffle strategy to assess for genetic interactions with our *rpb7* mutants and *mex67-6* at various temperatures (Fig. A2-1B). Rpb7-V101E showed severe synthetic sickness at 30C with *mex67-6,* and lethality at all other temperatures tested. We found that Rpb7-G149D, -E100K, and –D166G all partially suppressed the temperature sensitive phenotype at 37C. None of the tested mutants were able to suppress the temperature sensitivity of *mex67-6* at 39C. Very mild suppression was seen in rpb7-D166K at 37C. Similar to our *spt5* alleles, we did not observe that the Ts suppression extended to the mRNA export phenotype (data now shown).

Previously proposed models suggest that the Rpb4/7 module plays a direct role in mediating export of mRNA through its RNA binding and cytoplasmic shuttling properties (Harel-Sharvit *et al.,* 2010; Farago *et al.,* 2003). Physical interactions have also been reported between Rpb7 and

Mex67 (Mosley *et al.,* 2013). We uncovered allele specific genetic interactions between *rpb7* and *mex67-6* that support a functional interaction between Rpb7 and Mex67. Interestingly, the alleles of *rpb7* that demonstrate overlapping phenotypes with the *spt5* alleles identified in this study are the same alleles that suppress the temperature sensitivity of *mex67-6.* One of the alleles, *rpb7-V101E,* demonstrated an increased degree of temperature sensitivity with *mex67-6.* While this allele resulted in synthetic lethality with *spt5-E546K,* this allele on its own does not result in a *gal10Δ56* or cryptic initiation phenotype. This may reflect a separation of function between Spt5 and Rpb7. This region in Rpb7 does not appear to directly contact Spt5, and may be an indication of its role directly in mRNA export.

Interestingly, we found that mutations in the NGN Proximal domain of Spt5, which harbors the cold sensitive *spt5-242* allele, also strongly suppresses the temperature sensitivity of *mex67-6,* similar to the loss-of-function allele *spt5-194* and the *spt5-E546K* allele characterized in this study. A deletion of Spt5's C-terminal Repeat Region did not confer suppression of temperature sensitivity, suggesting that factors recruited by Spt5's CTR likely do not play a direct role in mRNA export. A common factor between all alleles of *spt5* that we tested against *mex67-6* is the cryptic initiation phenotype, suggesting that the local chromatin state may directly influence the packaging and export of mRNA.

While only one of the *rpb7* alleles that suppresses *mex67-6* results in cryptic initiation on its own, all of the alleles that suppress *mex67-6* result in enhanced cryptic initiation when paired with *spt5-G587D.* Furthermore, while some of the *spt5* alleles that suppress *mex67-6* result in a *gal10⊿56* phenotype, all of them result in cryptic initiation.

This suggests that the primary mode of suppression of *mex67-6* via Spt5 may be a result of modulations of local chromatin state. However, since the *rpb7* alleles that interact with *mex67-6* do not themselves display cryptic initiation, and Mex67 has been shown to physically interact with Rpb7, the mechanism of suppression here may be distinct. This may be a reflection of Rpb7's reported roles in shuttling mRNA out of the nucleus, or there could also be indirect impacts on export via Rpb7's role in transcription elongation with Spt5.

While we observed both *spt5* and *rpb7* mutants that suppressed the temperature sensitive phenotype of the *mex67-6* allele, the mechanism of suppression as yet remains unclear. However, we can begin to construct a model of suppression based on the factors identified through mass spectrometry. Mex67 was shown to directly bind Rpb7 in prior proteomics studies, and we found that Dbp2 binds directly to KOW2-3 (Mosley *et al.,* 2013, this work). Dbp2 also plays a role in loading nascent mRNA onto Mex67 (Ma *et al.,* 2013). Two other RNA-binding factors identified as KOW2-3 interacting have been shown to shuttle mRNA out of the nucleus in a

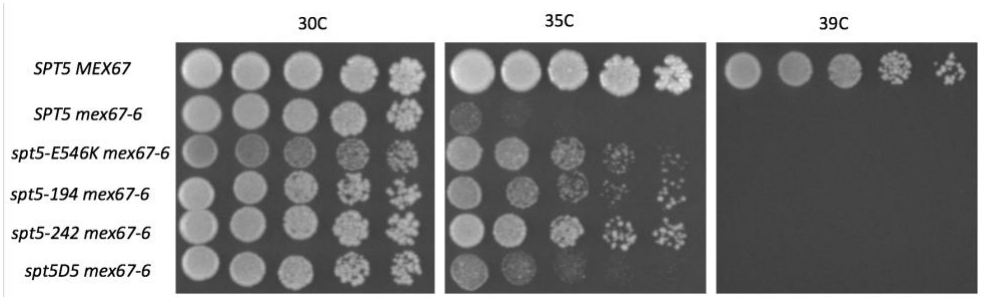
transcription dependent manner – Gbp2 and Sro9 (Windgassen *et al.,* 2003; Röther *et al.,* 2010). Interestingly, Gbp2 overexpression results in nuclear retention of polyadenylated mRNA, similar to the *mex67-6* allele. Further, overexpression of Sro9 was shown to suppress the temperature sensitivity of *rpb4Δ* and dramatically increase mRNA stability. It is possible that KOW2-4 facilitate loading of export factors onto the nascent mRNP and prepare Rpb4/7 and other RNA binding factors for shuttling out of the nucleus. By breaking contact with the central KOW domains, these RNA binding proteins may be prematurely or promiscuously loaded onto nascent mRNPs and increasing mRNP stability, compensating for the loss of *mex67* function in the temperature sensitive allele.

Figure A2-1: *spt5* and *rpb7* suppress temperature sensitivity of *mex67-6* in an allele specific manner

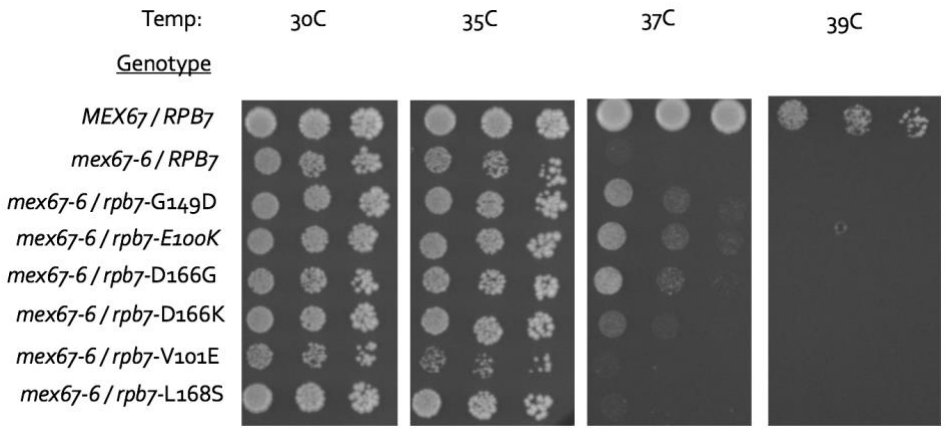
A) Integrated *spt5* mutant strains were crossed into a *MEX67* shuffle strain and a plasmid shuffle strategy was used to transform the *mex67-6* allele. *MEX67 URA3* was counter-selected against using 5FOA and strains were serially diluted onto YPD at the indicated temperatures and allowed to grow for 4 days.

B) An *RPB7* shuffle strain was transformed with *RPB7 LEU2* and *RPB7 URA3* was counter-selected against using 5FOA. This strain was then crossed into a *MEX67* shuffle strain, transformed with *mex67-6 TRP1* and *MEX67 URA3* was counter-selected against using 5FOA. Strains were then transformed with *RPB7 URA3* and transformants were re-streaked onto SC-Ura and replica plated onto SC-Leu to select for loss of *RPB7 LEU2.* Leu- colonies were then transformed with plasmids containing *rpb7* mutants marked with *LEU2* and *RPB7 URA3* was counter-selected against using 5FOA. Strains were serially diluted onto YPD at the indicated temperature and allowed to grow for 4 days.

A


30C
35C
39C
SPT5 MEX67
SPT5 mex67-6
spt5-E546K mex67-6
spt5-194 mex67-6
spt5-242 mex67-6
spt5D5 mex67-6

B


Temp:
30C
35C
37C
39C
Genotype
MEX67 / RPB7
mex67-6 / RPB7
mex67-6 / rpb7-G149D
mex67-6 / rpb7-E100K
mex67-6 / rpb7-D166G
mex67-6 / rpb7-D166K
mex67-6 / rpb7-V101E
mex67-6 / rpb7-L168S

Appendix - 3

Spt5 KOW5 Influences Polyadenylation Site Choice

While Spt5's N- and C-termini are fairly well characterized, the function of Spt5 KOW5 remains largely unknown. Structural studies demonstrate that KOW5 is wedged in a pocket between RNA polymerase II subunits Rpb1, Rpb2, Rpb3, Rpb8 and Rpb12 (Ehara *et al.,* 2017). The positioning of KOW5 is unambiguous in structural models and appears to anchor the position of the C-terminal region of Spt5. Prior studies in humans have shown that the linker region on the N-terminal side of KOW5 has RNA binding capability, while KOW5 itself does not enhance binding to RNA, suggesting a role for KOW5 in mRNA processing (Zuber *et al.,* 2012). In order to gain further understanding of KOW5's function, we took a structure-guided targeted PCR-based mutagenesis approach to alter highly conserved residues in KOW5 that are in apparent direct contact with other highly conserved residues in Pol II subunits.

Upon performing a structural analysis of KOW5 based on cryo-EM models, we determined that a three amino acid stretch in KOW5 appeared to hold functional significance based on position, conservation, and contact with conserved residues in Pol II subunits (Ehara *et al.,* 2017; Bernecky *et al.,* 2017, Fig. A3-1). These amino acids are Y812, K813 and G814. Y812 appears to contact conserved residues in Rpb11, K813 appears to coordinate

a cysteine-rich zinc loop in Rpb3, while G814 appears to contact the highly conserved and essential C-terminus of Rpb12 (Rubbi *et al.,* 1999).

We chose to alter these residues one at a time by changing them to alanine and then assaying for common phenotypes associated with Spt5 such as growth defects, cold sensitivity, temperature sensitivity, caffeine sensitivity, MPA sensitivity, spt- phenotype, suppression of *gal10Δ56,* and cryptic initiation. The only phenotype that Spt5-Y812A displayed was suppression of *gal10Δ56,* which is consistent with a role for Spt5 KOW5 in 3' end processing and polyadenylation site choice. This is in line with the model that KOW5's proximity to exiting mRNA and to the RNA binding linker region between KOW4 and KOW5 positions it for functions related to mRNA processing as opposed to chromatin structure maintenance. Interestingly, K813A, which apparently coordinates a zinc loop in Rpb3, did not display any phenotypes. This may be a reflection of redundant function between this residue in KOW5 and the many cysteines that encircle the predicted zinc residue.

Concurrent with the targeted mutagenesis of G814, we began a screen for *spt* mutants that display suppression of *gal10Δ56,* as we had previously determined that *spt5* cryptic initiation mutants in KOW2-3 resulted in *gal10Δ56* suppression. Interestingly, this screen resulted in the identification of Spt5-G814D as *gal10Δ56* suppressing. Therefore, we did not continue with the PCR-based mutagenesis experiment of Spt5-G814. The identification of

two residues in KOW5 that result in suppression of *gal10Δ56* further supports a model for this domain in 3' end processing and polyadenylation site choice.

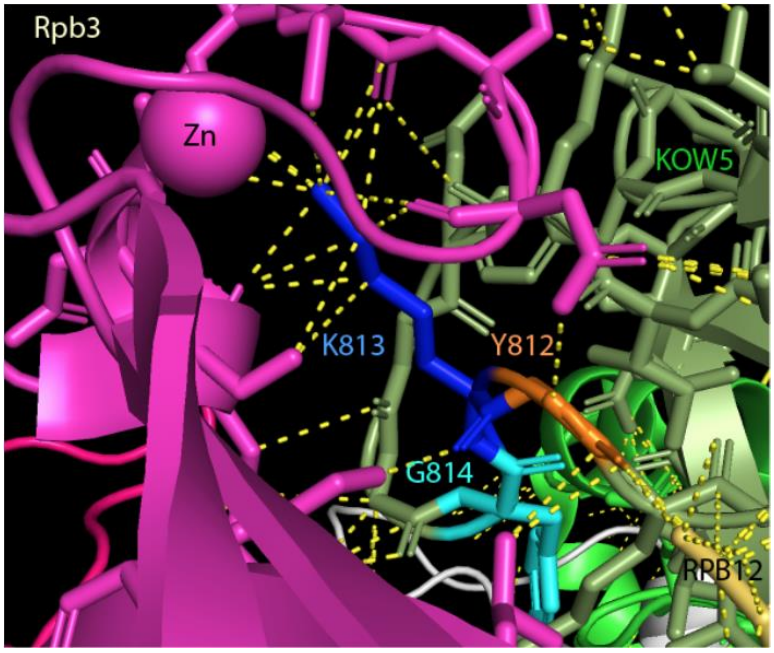
Taken together with prior mutagenesis studies of *SPT5,* we can begin to construct a model for the function of Spt5's central domains. Importantly, there are many overlapping functions and phenotypes in common between the disparate domains of Spt5. Many mutations have been identified in the N-terminal region of Spt5 that result in cryptic initiation, while mutations near the C-terminus in KOW5 result in altered polyadenylation site choice. Interestingly, mutations that lie in the intermediary region of KOW2-3 result in both cryptic initiation and polyadenylation site choice. This suggests a separation of function in Spt5 from N to C terminus, with the N terminus having roles dedicated to chromatin structure maintenance and the C-terminus having roles in mRNA processing. Consistent with this notion, the CTR has known functions in recruiting mRNA processing factors such as the capping apparatus and components of CFI, while the N-terminal region of Spt5, where the majority of Spt5's cryptic initiation mutations lie, has recently been shown to directly binds histones (Evrin *et al.,* 2022; Mayer *et al.,* 2012; Lidschreiber *et al.,* 2013). Spt5's central KOW domains, therefore, likely play roles in supporting functions of both termini of Spt5. Supporting this idea are the following observations: 1) *spt5-E546K* enhances the growth defect observed in *spt5-242,* which displays cryptic initiation and cold sensitivity on its own (Hartzog Lab, unpublished results) 2) Spt5 KOW2-3 displays genetic

and physical interactions with Nrd1, which has previously been shown to interact physically with Rpb7 and Spt5's CTR (Mosley *et al,* 2013; Leporé, Lafontaine 2011).

Two mechanistic models may be gleaned from these observations. First, Spt5 KOW2-3 may play a role in the pass-back of histones as Pol II traverses a chromatin template. This is supported by Spt5 KOW2-3 interacting with Spt16, a known histone chaperone, and Spt6 binding to the Pol II Stalk as well as mutations in *spt6* and *spt5* at KOW2-3 displaying cryptic initiation (Vos *et al.,* 2018b). Second, Spt5 KOW2-3 receives proteins that are initially recruited by the CTR such as Nrd1 (Lidschreiber *et al.,* 2013). In this model, KOW2-3 is mechanistically supporting the functions of both termini of Spt5 in both chromatin structure maintenance as well as mRNA processing. These functions may be separated temporally as Pol II traverses a gene body, or they may be separated spatially on different interfaces of the largely solvent-exposed KOW2-3 and Rpb4/7 or in concert with their interacting factors. Taken together, KOW2-3 likely serves as a nexus for coordinating Spt5's overall function across transcription.

Figure A3-1: Identifying conserved and functionally relevant residues in Spt5 KOW5 based on published structural models

Spt5-Y812, K813, and G814 contact Rpb11, Rpb3 and Rpb12 respectively. Based on structure published by Ehara *et al.,* 2017 (PDB: 5XON)


Rpb3
Zn
KOW5
K813
Y812
G814
RPB12

Appendix – 4

Rpb7-G149D disrupts binding of Spt5 KOW2-3

According to cryo-EM data, Rpb7-G149 appears to directly contact Spt5-G587 (Bernecky *et al.,* 2017). The identification of the two mutations *rpb7-G149D* and *spt5-G587D* suggests that the observed mutant phenotypes are a result of disruption of a binding event between these two factors. The importance of these residues is highlighted by their universal conservation (Fig. 2-3C). To test this hypothesis, we performed pull-down experiments with GST-tagged Rpb7 and Spt5 KOW2-3 with either wildtype Rpb7 or Rpb7-G149D. As Rpb4 stabilizes Rpb7, we co-expressed Rpb4 with Rpb7-GST (Runner *et al.,* 2007).

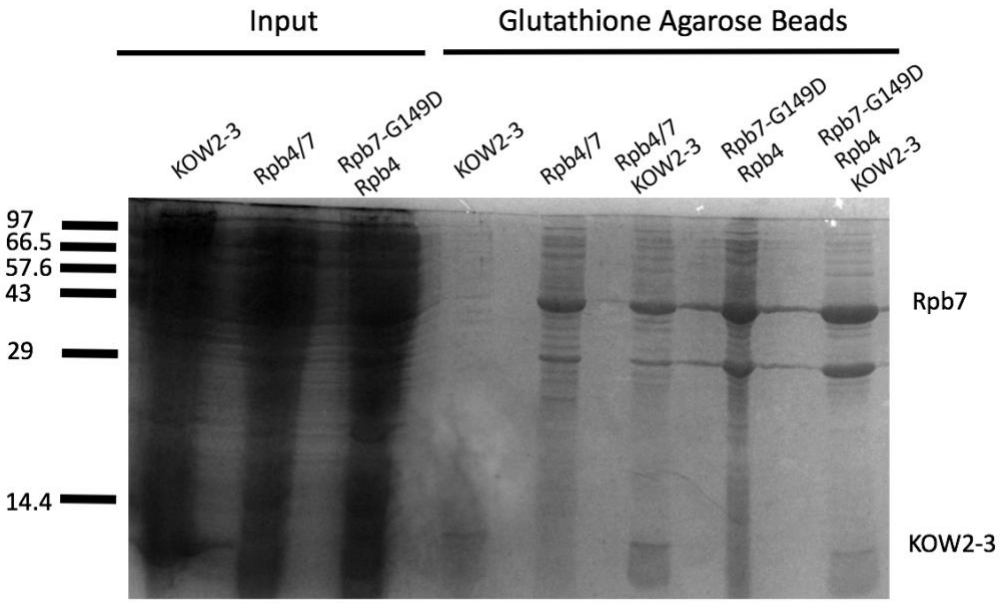
Bacterial lysates containing either Rpb4 and GST-Rpb7 or Rpb4 and Gst-Rpb7-G149D as bait were incubated on a glutathione-agarose column and washed. The glutathione-agarose column was then incubated with bacterial lysate containing Spt5 KOW2-3, washed, and boiled in SDS sample buffer. Despite an increased amount of Rpb7-G149D relative to wildtype in the pull-down, we observed a decrease in the amount of KOW2-3 that was retained in the Rpb7-G149D column (Fig. A4-1). In fact, the amount of Spt5 KOW2-3 retained in Rpb7-G149D was similar to the amount detected as background in the control column containing Spt5 KOW2-3 alone, indicating that Rpb7-G149D disrupts binding with Spt5 KOW2-3. The presence of Rpb7 and Rpb4 were both confirmed via Western Blot (data not shown).

Despite these promising results, there remains the issue of rather significant background binding of Spt5 KOW2-3 to the glutathione-agarose column. In addition, while repeat experiments appear promising, the results are not definitive and require further optimization of the experiment. It is possible that Spt5 KOW2-3 and Rpb7 are anchored to the Pol II EC primarily through other Rpb subunits, or that Spt5 KOW2-3 remains flexible to participate in alternate conformations, which could be a reflection of the weak interaction we see between Spt5 KOW2-3 and Rpb7.

Disruption of the interaction between Rpb7 and KOW2-3 suggests that the phenotypes we observe at these residues are directly related to loss of binding between Spt5 and Rpb7. By disrupting this binding event, we may be inducing a preference for an alternate conformation for Spt5 as it relates to the EC, or we may be disrupting the placement of Rpb7. There is evidence to suggest both of these models may be plausible. It has been reported that Rpb4/7 reversibly dissociates with elongation complexes, and this is relevant to Rpb4/7's function (Mosley *et al.,* 2013; Duek *et al.,* 2018). Additionally, one cryo-EM study was unable to resolve KOW2-3 in complex with the Pol II EC, suggesting alternate conformations may be at play (Ehara *et al.,* 2017).

Figure A4-1: Rpb7-G149D disrupts binding of Spt5 KOW2-3

Bacterial lysates containing either Rpb4 + GST-Rpb7 or Rpb4 + GST-Rpb7(G149D) were used as bait on a glutathione-agarose column, washed and then incubated with bacterial lysate containing 6x-His-Spt5-KOW2-3. Bait and prey alone were used as controls. Beads were boiled in 2x SDS Sample buffer for 5 minutes and 25ul was loaded on a 15% polyacrylamide gel and subsequently stained with Coomassie Brilliant Blue.


Input
Glutathione Agarose Beads
KOW2-3
Rpb4/7
Rpb7-G149D
Rpb4
KOW2-3
Rpb4/7
Rpb4/7
KOW2-3
Rpb7-G149D
Rpb4
Rpb7-G149D
Rpb4
KOW2-3
97
66.5
57.6
43
29
14.4
Rpb7
KOW2-3

Appendix - 5

List of Plasmids

plasmid	genotype	source	ch
pGH233	*CEN AmpR LEU2 SPT5*	Hartzog Lab	2
pMD11	CEN AmpR LEU2 spt5-E546K	Hartzog Lab	2
pMD18	*CEN AmpR LEU2 spt5-G587D*	Hartzog Lab	2
pMD15	*CEN AmpR LEU2 spt5-G602S, S809F*	Hartzog Lab	2
pSPT5ΔKOW4	*CEN AmpR LEU2 spt5-KOW4Δ*	Hartzog Lab	2
pSPT5ΔKOW5	*CEN AmpR LEU2 spt5-KOW5Δ*	Hartzog Lab	2
pGH235	*CEN AmpR LEU2 spt5-194*	Hartzog Lab	2
pMD23	*CEN AmpR LEU2 spt5-A267T*	Hartzog Lab	2
pJO1	*CEN AmpR LEU2 spt5-242*	J. Ortiz	2
pHQ1876	*CEN AmpR LEU2 spt5-(CTR-S1-15A)-3xHA*	A. Hinnebusch	2
pHQ1894	*CEN AmpR LEU2 spt5-(CTR-S1-15D)-3xHA*	A. Hinnebusch	2
pCK223	*CEN AmpR LEU2 RPB7*	C. Kaplan	2, A1, A2
pCK228	*CEN AmpR URA3 RPB7*	C. Kaplan	2, A1, A2
CKB310	*CEN AmpR LEU2 rpb7-D166G*	C. Kaplan	2, A1, A2
pZM6	*CEN AmpR LEU2 rpb7-D166K*	Hartzog Lab	2, A1, A2
pZM2	*CEN AmpR LEU2 rpb7-G149D*	Hartzog Lab	2, A1, A2
pZM1	*CEN AmpR LEU2 rpb7-E100K*	Hartzog Lab	2, A1, A2
CKB311	*CEN AmpR LEU2 rpb7-L168S*	C. Kaplan	2, A1, A2
CKB309	*CEN AmpR LEU2 rpb7-V101E*	C. Kaplan	2, A1, A2
pGH382	*KanR T7p-Spt5-kow2-kow3(aa534-632)His6*	Hartzog Lab	3
pGH378	*KanR T7p-Spt5-Linker2-KOW4(aa633-751)-His6*	Hartzog Lab	3
KB886	*AmpR TRP1 NRD1 2μ*	D. Brow	4
p425-GPD-RNAseH	*AmpR LEU2 GPDp-RNASEH1 (human) 2μ*	P. Hieter	4
p425-GPD	*AmpR LEU2 GPDp 2μ*	P. Hieter	4
pRS426-RPB7	*AmpR URA3 RPB7 2μ*	Hartzog Lab	2, 4
pZM11	*rpb7-D166G TRP*	Hartzog Lab	4
pURA3 MEX67	*CEN AmpR URA3 MEX67*	Ed Hurt	A2
pRS314-mex67-6	*CEN AmpR TRP1 mex67-6*	Ed Hurt	A2
pMM1	*CEN AmpR LEU2 spt5-Y812A*	Hartzog Lab	A3
pZM4	*CEN AmpR LEU2 spt5-G814D*	Hartzog Lab	A3
pZM7	*KanR GST-RPB7+RPB4*	GenScript	A4
PZM9	*KanR GST-RPB7(G149D)+RPB4*	GenScript	A4

Appendix - 6

List of Strains

name	#	a/b	genotype	source	ch
GHY	3246	a	*his3Δ200 lys2-128δ leu2Δ(0 or 1) ura3-52 trp1Δ63 gal10Δ56 spt5Δ*4 [pMS4]	Hartzog Lab	2, A3
GHY	1555	a	*rpb7D::CLONAT leu2D(0 or 1) lys2-128δ ura3-52 trp1D63 his3D200 met15D0 RPB3-TAP::KlactisTRP1 gal10Δ55* [pCK228 = *RPB7* pRS316]	C. Kaplan	2
GHY	3243	b	*rpb7D::CLONAT leu2D(0 or 1) lys2-128δ ura3-52 trp1D63 his3D200 pGAL1-FLO8-HIS3::KANMX RPB3-TAP::KlactisTRP1* [pCK228 = *RPB7* pRS316]	Hartzog Lab	2
GHY	2741	a	*his3Δ200 lys2-128δ leu2Δ1 ura3-52 trp1Δ63 spt5Δ4::TRP1 pGAL-FLO8-HIS3::KANMX* [pMS4]	Hartzog Lab	2
GHY	3248	b	*his3Δ(1 or 200) lys2-128δ leu2Δ(1 or 0) ura3(0 or -52) trp1Δ63 rpb4Δ::KANMX spt5Δ4::TRP1* [pMS4]	Hartzog Lab	2
GHY	3348	a	*rpb7D::CLONAT his3D200 lys2-128d leu2D(0 or 1) ura3-52 trp1D63 spt5-G587D pGAL1-FLO8-HIS3::KANMX* [pCK228 = *RPB7* pRS316 *URA]*	Hartzog Lab	2, 4
GHY	3349	b	*rpb7D::CLONAT his3D200 lys2-128d leu2D(0 or 1) ura3-52 trp1D63 spt5-E546K gal10Δ56* [pCK228 = *RPB7* pRS316 *URA*]	Hartzog Lab	2, 4
GHY	4005	a	*lys2-128d met15D0 his3Δ(1 or 200) leu2Δ(0 or 1) ura3Δ(0 or -52) rpb4Δ::KANMX gal10Δ56* [pRS426 2**μ** *RPB7*]	Hartzog Lab	2
GHY	4008	b	*his3D(0 or 200) leu2D(0 or 1) ura3D(0 or -52) rpb4Δ::KANMX pGAL-FLO8-HIS3::KANMX* [pRS426 2**μ** *RPB7*]	Hartzog Lab	2
GHY	4009	a	*his3D(0 or 200) leu2D(0 or 1) ura3D(0 or -52) LYS2-129δ rpb4Δ::KANMX pGAL-FLO8-HIS3::KANMX* [pRS426 2**μ** *RPB7*]	Hartzog Lab	2
GHY	610	b	*his4-912δ lys2-128δ leu2D1 pep4D::LEU2 ura3-52*	Hartzog Lab	3
OY	98	b	*hisΔ200 lys2-128δ leu2Δ1 trp1Δ63 ura3-52*	F. Winston	4
GHY	1068	b	*his3D0 lys2D0 leu2D0 ura3D0 hpr1Δ::KANMX4*	Hartzog Lab	4
GHY	3335	a	*his3Δ200 lys2-128δ leu2Δ(0 or 1) ura3-52 trp1Δ63 spt5-E546K gal10Δ56*	Hartzog Lab	4
GHY	3336	a	*his3Δ200 lys2-128δ leu2Δ1 ura3-52 trp1Δ63spt5-G587D pGAL-FLO8-HIS3::KANMX*	Hartzog Lab	4
GHY	4006	b	*rpb7D::CLONAT his3D200 lys2-128d leu2D(0 or 1) ura3-52 trp1D63 spt5-E546K gal10Δ56* [*rpb7-D166G, TRP*]	Hartzog Lab	4
GHY	4007	a	*rpb7D::CLONAT his3D200 lys2-128d leu2D(0 or 1) ura3-52 trp1D63 spt5-G587D pGAL1-FLO8-HIS3::KANMX [rpb7-D166G, TRP]*	Hartzog Lab	4
GHY	2014	b	*his3Δ200 leu2Δ1 ura3-52 trp1Δ63 chd1Δ::URA3 pGAL1-FLO8-HIS3::KANMX*	Hartzog Lab	4
GHY	3296	b	*his3Δ200 lys2-128δ leu2Δ1 ura3-52 trp1Δ63 spt5-242 pGAL1-FLO8-HIS3::KANMX*	Hartzog Lab	4
GHY	1007	b	*his4-912δ lys2-128δ leu2D1 ura3-52 trp1Δ63 spt5-194*	Hartzog Lab	4
ROY*	1309	a	*ADE2 his3 leu2 lys2 trp1 ura3 hmrD::URA3 gal4D::HIS3*	R. Kamakaka	4

ROY*	4203	a	*hmr1D::URA3 ADE2 lys2D hcr1Δ::KANMX rpb7Δ::CLONAT leu2Δ(0 or 1) his3Δ(0 or 200) lys2-128**δ** ura3-52 trp1Δ63 gal10Δ56 RPB3-TAP::KlactisTRP1*	R. Kamakaka	4
GHY	3345	a	*his3Δ(1 or 200) lys2-128**δ** leu2Δ(0 or 1) ura3Δ(0 or -52) trp1Δ63*	Hartzog Lab	A1
GHY	3347	b	*mex67Δ::KANMX spt5-E546K [MEX67 URA3] rpb7Δ::CLONAT his3Δ(1 or 200) lys2-128**δ** leu2Δ0 ura3Δ0 trp1Δ63 gal10Δ55 mex67Δ::KANMX* [pRS314 *mex67-6 TRP1*	Hartzog Lab	A2
GHY	3348	a	CEN] [pRS316 *RPB7 URA3* CEN] *his3Δ(1 or 200) lys2-128**δ** leu2Δ(0 or 1) ura3(Δ0 or -52) trp1Δ63*	Hartzog Lab	A2
GHY	1537	b	*mex67Δ::KANMX spt5-242* [*mex67-6* pRS314] *his3Δ(1 or 200) lys2-128**δ** leu2Δ(0 or 1) ura3(Δ0 or -52) trp1Δ63*	Hartzog Lab	A2
GHY	1536	b	*mex67Δ::KANMX spt5-194* [*mex67-6* pRS314] *his3Δ(1 or 200) lys2-128**δ** leu2Δ(0 or 1) trp1Δ63 ura3 (Δ0 or -52)*	Hartzog Lab	A2
GHY	4014	b	*mex67Δ::KANMX spt5D5::NAT [MEX67 URA3] his3Δ(1 or 200) lys2-128**δ** leu2Δ0 ura3Δ0 trp1Δ63*	Hartzog Lab	A2
GHY	1493	a	*mex67Δ::KANMX* [*MEX67 URA3*]	Hartzog Lab	A2
			* indicates W303 background		

Appendix – 7

Materials and Methods

Media and yeast genetic methods

All yeast media was made as described previously and all strain construction methods were standard methods (Rose *et al.,* 1990). *S. cerevisiae* strains used in this study are isogenic to S288C and are *GAL2*+ unless indicated otherwise and are listed in Appendix 6.

GHY4006 and GHY4007 were generated via a marker switch of GHY3349 and GHY3348 respectively harboring CKB310. PCR amplification of pRS404 was performed using OGH473 and OGH474 (1x Mango Mix (Bioline), 30 cycles of 95C 30sec, 51C 1min, 72C 90sec). High efficiency transformation was performed and plated onto –Trp, then replica plated to –Trp –Leu to select for loss of Leu marked CKB310.

Isolation of *spt5* and *rpb7* cryptic initiation and *gal10Δ56* mutants were performed with hydroxylamine mutagenized plasmids (Rose, Fink 1987)

To integrate *spt5-E546K* and *spt5-G587D*, a high efficiency transformation was performed with plasmids pMD11 and pMD18 that were digested with *Eag*I and *Hind*III into GHY3246 and GHY2741 respectively. Cells were allowed to grow for 1 day on YPD at 30C then replica plated to F-FOA, grown for 2 days and replica plated again to 5-FOA and grown for 2 days. Plates

were then replicated to 5-FOA –Trp and Trp- colonies were selected and restreaked to assess for stability.

hcr1Δ was generated via sporulation from the Stanford deletion collection (Giaever, Nislow 2014).

Plasmids

A detailed list of the plasmids used in this study is provided in Appendix 5.

Plasmid pZM6 was generated by PCR mutagenesis. Plasmid CKB223 was amplified with either OGH1630+OGH1631 or OGH1629+OGH1632 (Phusion polymerase, 1x Phusion HF Buffer (Thermo Scientific), .1 mM dNPTs; 30 cycles 95C 2min, 62C for 1min, 70C for 30sec). Gel excision was performed on the products using Nucleospin Gel and PCR cleanup kit (Macherey-Nagel). 5 μl of each product was added to a new reaction under the same conditions lacking primers and cycled 5 times (95C 30sec, 65C 1min, 70C 45 seconds). OGH 1631 and OGH1632 were then added and reaction was cycled 30x. CKB233 and the PCR product were individually digested with *Xho*I and *Eag*I and the digested PCR product was ligated into the CKB233 backbone.

Plasmid pZM11 was generated via a plasmid rescue from GHY 4006

Plasmid pMM1 was generated by PCR mutagenesis. Using the same strategy outlined above with pGH233, OGH1545+OGH166, OGH1546+OGH168 (30

cycles 95C 30sec, 52C 1min, 68C 45sec). Gel-excised products were amplified with OGH166+168 (40 cycles 95C 30sec, 50C 1min, 68C 1min). Products were ligated into a pGH233 backbone digested with *Mlu*I and *Apa*I.

Pull-down Assays

GST Pulldown assays were performed via a modification from a protocol from the Kellogg Lab.

For GHB1557 and GHB159 harboring Rpb4/7 expression constructs: 10mL LB+Kan was inoculated from a single colony and grown overnight at 37C. 1.3L of LB+Kan was inoculated with 5mL of overnight culture and grown to OD 0.7, supplemented with IPTG to 1mM and grown overnight at RT. Cells were harvested via centrifugation and ground into a powder with a mortar and pestle under liquid nitrogen. Cells were frozen in liquid nitrogen and stored at -80C.

Cells were thawed on ice and lysis buffer (1X PBS, 1M NaCl, 1 mM PMSF added fresh) was added in a ratio of 1 μl per mg of cell powder. Cells were sonicated twice for 1 minute with 1 minute on ice in between rounds and then centrifuged for 10 minutes at 10,000 rpm at 4C. 33mg protein was added to 50 μl glutathione agarose resin (Thermo Scientific Pierce) and brought to 1 mL with lysis buffer, and incubated for 2 hours at 4C on an inverter. Column was washed once with 10CV wash buffer (1X PBS, 0.25 M KCl) and then once with 1 CV Spt5 lysis buffer (1x PBS 150 mM K(Ac)). Control columns

that were not subject to Spt5 KOW extract were resuspended in 50 μl SDS sample buffer and boiled at 100C for 5 minutes and reserved for SDS PAGE analysis.

For GHB1095 harboring Spt5 KOW2-3: cell extract was generated as above except induction was performed at 0.5 mM IPTG and Spt5 lysis buffer was used.

100 μl of Spt5 KOW extract was added to the column and brought to 500 μl volume in Spt5 lysis buffer and incubated at 4C on an inverter for 30 minutes. Columns were washed 2x with 10 CV Spt5 lysis buffer and then resuspended in 50 μl SDS sample buffer and boiled at 100C for 5 minutes and analyzed via SDS PAGE on a 15% gel.

Spt5 KOW Domain Affinity Chromatography

BL21-DE3 cells harboring plasmids XX and XX were inoculated from a single colony in 10mL LB+Kan at 37C and grown overnight on a rotator. 1mL of culture was used to inoculate 1.3L of LB+Kan at 37C and grown to OD 0.7 on an orbital shaker. Cultures were then induced with IPTG to 0.5mM and shaken at 20C overnight, harvested and frozen in liquid nitrogen. Frozen cell pellet was grown to a fine powder with a mortar and pestle frozen with liquid nitrogen and scooped into screw-cap tubes and frozen again in liquid nitrogen.

8 vials of cells were allowed to thaw on ice and mixed each with 0.5mL of phosphate buffer (P-Buffer) (50 mM phosphate pH7.8, 10% glycerol, 50 mM NaCl, 50 mM Arginine, 50 mM Glutamic acid, PMSF 1x added fresh). Cells were then subjected to three rounds of sonication at 4C, each round being 30 seconds followed by 30 seconds on ice. Cells were spun for 10 minutes at 10,000rpm in a microfuge at 4C. 1 mL HisPur Ni-NTA beads (Thermo Scientific) were equilibrated in P-Buffer and then placed in a chromatography column along with 4mL of cell extract and placed on a rocker at 4C for 30 minutes. Column was washed with 20CV P-Buffer at 250 mM NaCl, followed by 20CV P-Bufer with 0.1% NP-40, followed by 20CV P-Buffer with 100mM NaCl and 4mM imidazole. Column was eluted with 10 500 μl fractions of P-Buffer with 200 mM imidazole and frozen in liquid nitrogen. Fractions were then analyzed via SDS-PAGE.

5 Fractions containing the highest quantity of protein as determined by SDS-PAGE were pooled and dialyzed three times with 4mL each of 25 mM MES utilizing Amicon 3k dialysis centrifugal filters at 4,000 rpm for 40 minutes and concentrated to 500 μl in 25 mM MES.

1mL μl of Affi-Gel 10 (BioRad) was pipetted into a Falcon tube and washed three times with 10 mL mQ H_2O. ~20mg of purified protein (either K2K3, L2K4 or BSA) was added directly to the resin and placed on a rocker at 4C for 4 hours. Coupling proceeded until supernatant read OD 0.01 at 595 nm on a

Bradford assay (2.5 hr). 100 μl of 1M ethanolamine was added for 1 hour at 4C to stop the reaction. Resin was washed 3 times with 10 CV 25 mM MES and resuspended in 5 mL 25 mM MES with 0.2% sodium azide and stored at 4C.

Afi-gel was transferred into a chromatography column, drained and washed with 10CV of yeast lysis buffer. 25 mg crude yeast extract generated from GHY610 was diluted to 50mL and loaded onto the column over a period of 8 hours. Column was washed with 10CV yeast lysis buffer with 0.5% NP-40 at 0.2 M K(Ac). Column was washed with 5CV yeast lysis buffer. 8 400 μl elutions were collected with yeast lysis buffer with an increasing gradient of K(Ac), starting with 0.3M and ending in 1M, with a ninth fraction collected with yeast lysis buffer at 150 mM K(Ac) and 2.5M urea. Fractions were split into two aliquots, frozen in liquid nitrogen and stored at -80C. Column was washed with 10CV yeast lysis buffer with 0.05% sodium azide and stored at -20C.

Half of each fraction was TCA precipitated and analyzed via SDS-PAGE on a silver stained gradient gel. Fractions containing 0.8M, 0.9M and 1M K(Ac) were sent to the Yates lab for mass spectrometry analysis.

Yeast Extract Preparation

5mL YPD was inoculated with a single colony and grown on a rotator at 30C overnight. 1.3L of YPD was inoculated with 5mL of overnight culture and

grown to mid-log phase, harvested via centrifugation and frozen in liquid nitrogen. Cell pellets were ground into a find powder using a mortar and pestle frozen with liquid nitrogen. Cell powder was allowed to thaw on ice and then mixed with an equal volume of yeast lysis buffer (30 mM HEPES [pH 7.4], 200 mM potassium acetate, 1 mM magnesium acetate, 1 mM EGTA, 0.05% Tween-20, 10% glycerol, 1 mM PMSF, 0.02 μM Pepstain A, 0.02 μM Chymostatin 6 nM, 0.02 mM Benzamidine HCl, added fresh). Cells were then centrifuged at 14,000rpm for 10 minutes and supernatant collected. Supernatant was clarified using an ultracentrifuge at 30,000rpm for 30 minutes at 4C. Extract was frozen in liquid nitrogen and stored at -80C, or used fresh.

Mass Spectrometry

MudPIT mass spectrometry performed in collaboration with John R Yates III (Delahunty, Yates 2007)

Chromosome Spreads and S9.6 Immunofluorescence

Chromosome spreads and S9.6 immunofluorescence were performed using a modification of published protocols (Loidl *et al.,* 1991; Lafuente-Barquero *et al.,* 2020; Yang, Zhang 2022).

2 mL SC-Leu was inoculated with a single colony and grown overnight at 30C on a rotator. Overnight culture was used to inoculate 5 mL SC-Leu and grown overnight to mid-log phase. 3 mL of yeast culture was harvested and washed

once in Buffer B (1.2 M sorbitol, 0.1 M potassium phosphate, pH 7.5). Pellet was resuspended in 500 μl Buffer B. 10 μl 1 M DTT was added along with 10 μl of Zymolyase 100T 10mg / mL and incubated in a 37C water bath. Cells were monitored for lysis by mixing 10 μl of cells with 100 μl of mq H_2O and viewed under a microscope for lysis. Cells were removed from the water bath after 1 hour and centrifuged for 5 minutes at 800 x g, then resuspended in 200 μl MES stop-wash buffer (0.1 M MES, 1 M sorbitol, 1 mM EDTA, 0.5 mM $MgCl_2$, pH 6.4). Glass microscope slides were scrubbed with soap and water, dipped in 70% ethanol, allowed to dry completely and polished with lens paper. 20 μl cells were added to a microscope slide, followed by 40 μl fixative (4% formaldehyde, 3.4% sucrose) and swirled until Schlieren lines disappear. 80 μl lysis solution was added (1% NP-40 in mq H_2O), swirled and monitored under a microscope until >90% of cells lysed (~3 minutes). 80 μl fixative was added, swirled, and liquid was spread across the surface of the slide using a Pasteur pipette and allowed to dry in the fume hood overnight.

Slides were dipped in 0.2% Photo-Flo solution (Kodak) for 30 seconds and allowed to dry completely. Slides were washed in 1x PBS for 15 minutes and then blocked in blocking buffer (5% BSA, 0.2% milk, 1x PBS) for 15 minutes. R-loop specific primary antibody (clone S9.6, EMD Millipore) was added 1:500 in blocking buffer for 1 hour. Slides were washed in 1x PBS for 1 hour. Cy5 donkey anti-mouse secondary antibody was applied 1:1000 in blocking

buffer for 1 hour and then slides were washed for 15 minutes in 1x PBS and allowed to dry completely. Cover slips were scrubbed with soap and water, dipped in ethanol, dried and polished with lens paper. Vecta Shield mounting media with DAPI was applied and cover slips were sealed with nail polish.

Slides were imaged using a spinning-disc confocal microscope at 100X. All images were taken under the same microscope settings in the same imaging session. >250 nuclei were counted for two technical replicates each of two biological replicates. Nuclei were considered S9.6+ when Cy5 foci directly overlapped DAPI signal.

Prediction of Complex containing Rpb7, Nrd1, Nab3, KOW2-3

Complex prediction was carried out utilizing AlphaFold Multimer through the COSMIC^2 cloud computing platform, which utilizes Comet, a cluster at the San Diego Supercomputer Center. Primary amino acid sequences were input into ColabFold.

The structure displayed in Figure 3-4 was generated by merging two AlphaFold Multimer complex predictions. The first complex was Nrd1, Rpb7 and Spt5 KOW2-3. The following amino acids were input to generate this complex: Nrd1(299-490), Spt5(534-632), Rpb7(1-171). The second complex was Spt5 KOW2-3 and Nab3. Spt5(534-632) and the full sequence of Nab3 were used, however, in this structure only Nab3(311-415) are displayed.

References

Al-Hadid, Qais, and Yanzhong Yang. 2016. “R-Loop: An Emerging Regulator of Chromatin Dynamics.” *Acta Biochimica et Biophysica Sinica* 48 (7): 623–31. https://doi.org/10.1093/abbs/gmw052.

Allepuz-Fuster, Paula, Verónica Martínez-Fernández, Ana I. Garrido-Godino, Sergio Alonso-Aguado, Steven D. Hanes, Francisco Navarro, and Olga Calvo. 2014. “Rpb4/7 Facilitates RNA Polymerase II CTD Dephosphorylation.” *Nucleic Acids Research* 42 (22): 13674–88. https://doi.org/10.1093/nar/gku1227.

Allepuz-Fuster, Paula, Michael J. O’Brien, Noelia González-Polo, Bianca Pereira, Zuzer Dhoondia, Athar Ansari, and Olga Calvo. 2019. “RNA Polymerase II Plays an Active Role in the Formation of Gene Loops through the Rpb4 Subunit.” *Nucleic Acids Research* 47 (17): 8975–87. https://doi.org/10.1093/nar/gkz597.

Allison, L. A., J. K. Wong, V. D. Fitzpatrick, M. Moyle, and C. J. Ingles. 1988. “The C-Terminal Domain of the Largest Subunit of RNA Polymerase II of Saccharomyces Cerevisiae, Drosophila Melanogaster, and Mammals: A Conserved Structure with an Essential Function.” *Molecular and Cellular Biology* 8 (1): 321–29. https://doi.org/10.1128/mcb.8.1.321-329.1988.

Appanah, Rowin, Emma Claire Lones, Umberto Aiello, Domenico Libri, and Giacomo De Piccoli. 2020. “Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability.” *Cell Reports* 30 (7): 2094-2105.e9. https://doi.org/10.1016/j.celrep.2020.01.087.

“Architecture of the RNA Polymerase–Spt4/5 Complex and Basis of Universal Transcription Processivity.” 2011. *The EMBO Journal* 30 (7): 1302–10. https://doi.org/10.1038/emboj.2011.64.

Arenas, Jaime E., and John N. Abelson. 1997. “Prp43: An RNA Helicase-like Factor Involved in Spliceosome Disassembly.” *Proceedings of the National Academy of Sciences of the United States of America* 94 (22): 11798–802.

Armache, Karim-Jean, Hubert Kettenberger, and Patrick Cramer. 2003. “Architecture of Initiation-Competent 12-Subunit RNA Polymerase II.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (12): 6964–68.

https://doi.org/10.1073/pnas.1030608100.

Bacikova, Veronika, Josef Pasulka, Karel Kubicek, and Richard Stefl. 2014. "Structure and Semi-Sequence-Specific RNA Binding of Nrd1." *Nucleic Acids Research* 42 (12): 8024–38. https://doi.org/10.1093/nar/gku446.

Barillà, D., B. A. Lee, and N. J. Proudfoot. 2001. "Cleavage/Polyadenylation Factor IA Associates with the Carboxyl-Terminal Domain of RNA Polymerase II in Saccharomyces Cerevisiae." *Proceedings of the National Academy of Sciences of the United States of America* 98 (2): 445–50. https://doi.org/10.1073/pnas.98.2.445.

Belotserkovskaya, Rimma, Sangtaek Oh, Vladimir A. Bondarenko, George Orphanides, Vasily M. Studitsky, and Danny Reinberg. 2003. "FACT Facilitates Transcription-Dependent Nucleosome Alteration." *Science (New York, N.Y.)* 301 (5636): 1090–93. https://doi.org/10.1126/science.1085703.

Bennett, Craig L., Bryce L. Sopher, and Albert R. La Spada. 2020. "Tight Expression Regulation of Senataxin, Linked to Motor Neuron Disease and Ataxia, Is Required to Avert Cell-Cycle Block and Nucleolus Disassembly." *Heliyon* 6 (6): e04165. https://doi.org/10.1016/j.heliyon.2020.e04165.

Bernecky, Carrie, Franz Herzog, Wolfgang Baumeister, Jürgen M. Plitzko, and Patrick Cramer. 2016. "Structure of Transcribing Mammalian RNA Polymerase II." *Nature* 529 (7587): 551–54. https://doi.org/10.1038/nature16482.

Bernecky, Carrie, Jürgen M. Plitzko, and Patrick Cramer. 2017. "Structure of a Transcribing RNA Polymerase II-DSIF Complex Reveals a Multidentate DNA-RNA Clamp." *Nature Structural & Molecular Biology* 24 (10): 809–15. https://doi.org/10.1038/nsmb.3465.

Bhatia, Vaibhav, Sonia I. Barroso, María L. García-Rubio, Emanuela Tumini, Emilia Herrera-Moyano, and Andrés Aguilera. 2014. "BRCA2 Prevents R-Loop Accumulation and Associates with TREX-2 MRNA Export Factor PCID2." *Nature* 511 (7509): 362–65. https://doi.org/10.1038/nature13374.

Blasco-Moreno, Bernat, Leire de Campos-Mata, René Böttcher, José García-Martínez, Jennifer Jungfleisch, Danny D. Nedialkova, Shiladitya Chattopadhyay, et al. 2019. "The Exonuclease Xrn1 Activates Transcription and Translation of MRNAs Encoding Membrane

Proteins.” *Nature Communications* 10 (1): 1298. https://doi.org/10.1038/s41467-019-09199-6.

Bortvin, A., and F. Winston. 1996. “Evidence That Spt6p Controls Chromatin Structure by a Direct Interaction with Histones.” *Science (New York, N.Y.)* 272 (5267): 1473–76. https://doi.org/10.1126/science.272.5267.1473.

Botuyan, Maria Victoria, Joseph Lee, Irene M. Ward, Ja-Eun Kim, James R. Thompson, Junjie Chen, and Georges Mer. 2006. “Structural Basis for the Methylation State-Specific Recognition of Histone H4-K20 by 53BP1 and Crb2 in DNA Repair.” *Cell* 127 (7): 1361–73. https://doi.org/10.1016/j.cell.2006.10.043.

Bou-Nader, Charles, Ankur Bothra, David N. Garboczi, Stephen H. Leppla, and Jinwei Zhang. 2022. “Structural Basis of R-Loop Recognition by the S9.6 Monoclonal Antibody.” *Nature Communications* 13 (1): 1641. https://doi.org/10.1038/s41467-022-29187-7.

Bowman, Gregory D., and Michael G. Poirier. 2015. “Post-Translational Modifications of Histones That Influence Nucleosome Dynamics.” *Chemical Reviews* 115 (6): 2274–95. https://doi.org/10.1021/cr500350x.

Brambati, Alessandra, Luca Zardoni, Eleonora Nardini, Achille Pellicioli, and Giordano Liberi. 2020. “The Dark Side of RNA:DNA Hybrids.” *Mutation Research. Reviews in Mutation Research* 784 (June): 108300. https://doi.org/10.1016/j.mrrev.2020.108300.

Burckin, Todd, Roland Nagel, Yael Mandel-Gutfreund, Lily Shiue, Tyson A. Clark, Jean-Leon Chong, Tien-Hsien Chang, Sharon Squazzo, Grant Hartzog, and Manuel Ares. 2005. “Exploring Functional Relationships between Components of the Gene Expression Machinery.” *Nature Structural & Molecular Biology* 12 (2): 175–82. https://doi.org/10.1038/nsmb891.

Burmann, Björn M., Kristian Schweimer, Xiao Luo, Markus C. Wahl, Barbara L. Stitt, Max E. Gottesman, and Paul Rösch. 2010. “A NusE:NusG Complex Links Transcription and Translation.” *Science (New York, N.Y.)* 328 (5977): 501–4. https://doi.org/10.1126/science.1184953.

Burova, E., S. C. Hung, V. Sagitov, B. L. Stitt, and M. E. Gottesman. 1995a. “Escherichia Coli NusG Protein Stimulates Transcription Elongation Rates in Vivo and in Vitro.” *Journal of Bacteriology* 177 (5): 1388–92.

https://doi.org/10.1128/jb.177.5.1388-1392.1995.

Burova, E, S C Hung, V Sagitov, B L Stitt, and M E Gottesman. 1995b. "Escherichia Coli NusG Protein Stimulates Transcription Elongation Rates in Vivo and in Vitro." *Journal of Bacteriology* 177 (5): 1388–92.

Calvo, Olga. 2020. "RNA Polymerase II Phosphorylation and Gene Looping: New Roles for the Rpb4/7 Heterodimer in Regulating Gene Expression." *Current Genetics* 66 (5): 927–37. https://doi.org/10.1007/s00294-020-01084-w.

Carroll, Kristina L., Dennis A. Pradhan, Josh A. Granek, Neil D. Clarke, and Jeffry L. Corden. 2004. "Identification of Cis Elements Directing Termination of Yeast Nonpolyadenylated SnoRNA Transcripts." *Molecular and Cellular Biology* 24 (14): 6241–52. https://doi.org/10.1128/MCB.24.14.6241-6252.2004.

Castellano-Pozo, Maikel, José M. Santos-Pereira, Ana G. Rondón, Sonia Barroso, Eloisa Andújar, Mónica Pérez-Alegre, Tatiana García-Muse, and Andrés Aguilera. 2013. "R Loops Are Linked to Histone H3 S10 Phosphorylation and Chromatin Condensation." *Molecular Cell* 52 (4): 583–90. https://doi.org/10.1016/j.molcel.2013.10.006.

Chan, Yujia A., Maria J. Aristizabal, Phoebe Y. T. Lu, Zongli Luo, Akil Hamza, Michael S. Kobor, Peter C. Stirling, and Philip Hieter. 2014. "Genome-Wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with DRIP-Chip." *PLOS Genetics* 10 (4): e1004288. https://doi.org/10.1371/journal.pgen.1004288.

Chaudhuri, Jayanta, Ming Tian, Chan Khuong, Katrin Chua, Eric Pinaud, and Frederick W. Alt. 2003. "Transcription-Targeted DNA Deamination by the AID Antibody Diversification Enzyme." *Nature* 422 (6933): 726–30. https://doi.org/10.1038/nature01574.

Chen, Poshen B., Hsiuyi V. Chen, Diwash Acharya, Oliver J. Rando, and Thomas G. Fazzio. 2015. "R Loops Regulate Promoter-Proximal Chromatin Architecture and Cellular Differentiation." *Nature Structural & Molecular Biology* 22 (12): 999–1007. https://doi.org/10.1038/nsmb.3122.

Chen, Ying-Zhang, Craig L. Bennett, Huy M. Huynh, Ian P. Blair, Imke Puls, Joy Irobi, Ines Dierick, et al. 2004. "DNA/RNA Helicase Gene Mutations in a Form of Juvenile Amyotrophic Lateral Sclerosis (ALS4)." *American Journal of Human Genetics* 74 (6): 1128–35.

Cheung, Vanessa, Gordon Chua, Nizar N. Batada, Christian R. Landry, Stephen W. Michnick, Timothy R. Hughes, and Fred Winston. 2008. “Chromatin- and Transcription-Related Factors Repress Transcription from within Coding Regions throughout the Saccharomyces Cerevisiae Genome.” *PLoS Biology* 6 (11): e277. https://doi.org/10.1371/journal.pbio.0060277.

Choder, M., and R. A. Young. 1993. “A Portion of RNA Polymerase II Molecules Has a Component Essential for Stress Responses and Stress Survival.” *Molecular and Cellular Biology* 13 (11): 6984–91. https://doi.org/10.1128/mcb.13.11.6984-6991.1993.

Choder, Mordechai. 2004. “Rpb4 and Rpb7: Subunits of RNA Polymerase II and Beyond.” *Trends in Biochemical Sciences* 29 (12): 674–81. https://doi.org/10.1016/j.tibs.2004.10.007.

Christofori, G., and W. Keller. 1988. “3’ Cleavage and Polyadenylation of MRNA Precursors in Vitro Requires a Poly(A) Polymerase, a Cleavage Factor, and a SnRNP.” *Cell* 54 (6): 875–89. https://doi.org/10.1016/s0092-8674(88)91263-9.

Cloutier, Sara C., Wai Kit Ma, Luyen T. Nguyen, and Elizabeth J. Tran. 2012. “The DEAD-Box RNA Helicase Dbp2 Connects RNA Quality Control with Repression of Aberrant Transcription.” *The Journal of Biological Chemistry* 287 (31): 26155–66. https://doi.org/10.1074/jbc.M112.383075.

Cloutier, Sara C., Siwen Wang, Wai Kit Ma, Nadra Al Husini, Zuzer Dhoondia, Athar Ansari, Pete E. Pascuzzi, and Elizabeth J. Tran. 2016. “Regulated Formation of LncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation.” *Molecular Cell* 61 (3): 393–404. https://doi.org/10.1016/j.molcel.2015.12.024.

Cojocaru, Marilena, Célia Jeronimo, Diane Forget, Annie Bouchard, Dominique Bergeron, Pierre Côte, Guy G. Poirier, Jack Greenblatt, and Benoit Coulombe. 2008. “Genomic Location of the Human RNA Polymerase II General Machinery: Evidence for a Role of TFIIF and Rpb7 at Both Early and Late Stages of Transcription.” *The Biochemical Journal* 409 (1): 139–47. https://doi.org/10.1042/BJ20070751.

Combs, D. Joshua, Roland J. Nagel, Manuel Ares, and Scott W. Stevens. 2006. “Prp43p Is a DEAH-Box Spliceosome Disassembly Factor Essential for Ribosome Biogenesis.” *Molecular and Cellular Biology* 26

(2): 523–34. https://doi.org/10.1128/MCB.26.2.523-534.2006.

Conrad, N K, S M Wilson, E J Steinmetz, M Patturajan, D A Brow, M S Swanson, and J L Corden. 2000. “A Yeast Heterogeneous Nuclear Ribonucleoprotein Complex Associated with RNA Polymerase II.” *Genetics* 154 (2): 557–71.

Cramer, P., D. A. Bushnell, and R. D. Kornberg. 2001. “Structural Basis of Transcription: RNA Polymerase II at 2.8 Angstrom Resolution.” *Science (New York, N.Y.)* 292 (5523): 1863–76. https://doi.org/10.1126/science.1059493.

Crick, Francis. 1970. “Central Dogma of Molecular Biology.” *Nature* 227 (5258): 561–63. https://doi.org/10.1038/227561a0.

Cristini, Agnese, Matthias Groh, Maiken S. Kristiansen, and Natalia Gromak. 2018. “RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage.” *Cell Reports* 23 (6): 1891–1905. https://doi.org/10.1016/j.celrep.2018.04.025.

Crisucci, Elia M., and Karen M. Arndt. 2011. “The Roles of the Paf1 Complex and Associated Histone Modifications in Regulating Gene Expression.” *Genetics Research International* 2011: 707641. https://doi.org/10.4061/2011/707641.

Cui, Yajun, and Clyde L. Denis. 2003. “In Vivo Evidence That Defects in the Transcriptional Elongation Factors RPB2, TFIIS, and SPT5 Enhance Upstream Poly(A) Site Utilization.” *Molecular and Cellular Biology* 23 (21): 7887–7901. https://doi.org/10.1128/MCB.23.21.7887-7901.2003.

Delahunty, Claire M., and John R. Yates. 2007. “MudPIT: Multidimensional Protein Identification Technology.” *BioTechniques* 43 (5): 563, 565, 567 passim.

Dengl, Stefan, and Patrick Cramer. 2009. “Torpedo Nuclease Rat1 Is Insufficient to Terminate RNA Polymerase II in Vitro.” *The Journal of Biological Chemistry* 284 (32): 21270–79. https://doi.org/10.1074/jbc.M109.013847.

Ding, Baojin, Danielle LeJeune, and Shisheng Li. 2010. “The C-Terminal Repeat Domain of Spt5 Plays an Important Role in Suppression of Rad26-Independent Transcription Coupled Repair.” *The Journal of Biological Chemistry* 285 (8): 5317–26.

https://doi.org/10.1074/jbc.M109.082818.

Douglas, H. C., and D. C. Hawthorne. 1964. “Enzymatic Expression and Genetic Linkage of Genes Controlling Galactose Utilization in Saccharomyces.” *Genetics* 49 (5): 837–44.

Drolet, M., P. Phoenix, R. Menzel, E. Massé, L. F. Liu, and R. J. Crouch. 1995. “Overexpression of RNase H Partially Complements the Growth Defect of an Escherichia Coli Delta TopA Mutant: R-Loop Formation Is a Major Problem in the Absence of DNA Topoisomerase I.” *Proceedings of the National Academy of Sciences of the United States of America* 92 (8): 3526–30. https://doi.org/10.1073/pnas.92.8.3526.

Duek, Lea, Oren Barkai, Ron Elran, Isra Adawi, and Mordechai Choder. 2018. “Dissociation of Rpb4 from RNA Polymerase II Is Important for Yeast Functionality.” *PLoS ONE* 13 (10): e0206161. https://doi.org/10.1371/journal.pone.0206161.

Edwards, A. M., C. M. Kane, R. A. Young, and R. D. Kornberg. 1991. “Two Dissociable Subunits of Yeast RNA Polymerase II Stimulate the Initiation of Transcription at a Promoter in Vitro.” *The Journal of Biological Chemistry* 266 (1): 71–75.

Egloff, Sylvain, and Shona Murphy. 2008. “Cracking the RNA Polymerase II CTD Code.” *Trends in Genetics: TIG* 24 (6): 280–88. https://doi.org/10.1016/j.tig.2008.03.008.

Ehara, Haruhiko, Takeshi Yokoyama, Hideki Shigematsu, Shigeyuki Yokoyama, Mikako Shirouzu, and Shun-Ichi Sekine. 2017. “Structure of the Complete Elongation Complex of RNA Polymerase II with Basal Factors.” *Science (New York, N.Y.)* 357 (6354): 921–24. https://doi.org/10.1126/science.aan8552.

El Hage, Aziz, Sarah L. French, Ann L. Beyer, and David Tollervey. 2010. “Loss of Topoisomerase I Leads to R-Loop-Mediated Transcriptional Blocks during Ribosomal RNA Synthesis.” *Genes & Development* 24 (14): 1546–58. https://doi.org/10.1101/gad.573310.

El Hage, Aziz, Shaun Webb, Alastair Kerr, and David Tollervey. 2014. “Genome-Wide Distribution of RNA-DNA Hybrids Identifies RNase H Targets in TRNA Genes, Retrotransposons and Mitochondria.” *PLoS Genetics* 10 (10): e1004716. https://doi.org/10.1371/journal.pgen.1004716.

Ellerby, Lisa M. 2019. “Repeat Expansion Disorders: Mechanisms and Therapeutics.” *Neurotherapeutics* 16 (4): 924–27. https://doi.org/10.1007/s13311-019-00823-3.

Enriquez-Harris, P, N Levitt, D Briggs, and N J Proudfoot. 1991. “A Pause Site for RNA Polymerase II Is Associated with Termination of Transcription.” *The EMBO Journal* 10 (7): 1833–42.

Epshtein, Vitaly, Christopher J. Cardinale, Andrei E. Ruckenstein, Sergei Borukhov, and Evgeny Nudler. 2007. “An Allosteric Path to Transcription Termination.” *Molecular Cell* 28 (6): 991–1001. https://doi.org/10.1016/j.molcel.2007.10.011.

Evrin, Cecile, Albert Serra-Cardona, Shoufu Duan, Progya P. Mukherjee, Zhiguo Zhang, and Karim P. M. Labib. 2022a. “Spt5 Histone Binding Activity Preserves Chromatin during Transcription by RNA Polymerase II.” *The EMBO Journal* 41 (5): e109783. https://doi.org/10.15252/embj.2021109783.

Farabaugh, Philip J., and Gerald R. Fink. 1980. “Insertion of the Eukaryotic Transposable Element Ty1 Creates a 5-Base Pair Duplication.” *Nature* 286 (5771): 352–56. https://doi.org/10.1038/286352a0.

Farago, Marganit, Tal Nahari, Christopher Hammel, Charles N. Cole, and Mordechai Choder. 2003. “Rpb4p, a Subunit of RNA Polymerase II, Mediates MRNA Export during Stress.” *Molecular Biology of the Cell* 14 (7): 2744–55. https://doi.org/10.1091/mbc.E02-11-0740.

Farnung, Lucas, Moritz Ochmann, Maik Engeholm, and Patrick Cramer. 2021. “Structural Basis of Nucleosome Transcription Mediated by Chd1 and FACT.” *Nature Structural & Molecular Biology* 28 (4): 382–87. https://doi.org/10.1038/s41594-021-00578-6.

Fasken, Milo B., R. Nicholas Laribee, and Anita H. Corbett. 2015. “Nab3 Facilitates the Function of the TRAMP Complex in RNA Processing via Recruitment of Rrp6 Independent of Nrd1.” *PLoS Genetics* 11 (3): e1005044. https://doi.org/10.1371/journal.pgen.1005044.

Filipovski, Martin, Jelly H M Soffers, Seychelle M Vos, and Lucas Farnung. 2022. “Structural Basis of Nucleosome Retention during Transcription Elongation.” *Science (New York, N.Y.)* 376 (6599): 1313–16. https://doi.org/10.1126/science.abo3851.

Fischer, Jonathan, Yun S. Song, Nir Yosef, Julia di Iulio, L. Stirling

Churchman, and Mordechai Choder. 2020. “The Yeast Exoribonuclease Xrn1 and Associated Factors Modulate RNA Polymerase II Processivity in 5’ and 3’ Gene Regions.” *The Journal of Biological Chemistry* 295 (33): 11435–54. https://doi.org/10.1074/jbc.RA120.013426.

Ford, J. P., and M. T. Hsu. 1978. “Transcription Pattern of in Vivo-Labeled Late Simian Virus 40 RNA: Equimolar Transcription beyond the MRNA 3’ Terminus.” *Journal of Virology* 28 (3): 795–801. https://doi.org/10.1128/JVI.28.3.795-801.1978.

Fourmann, Jean-Baptiste, Jana Schmitzová, Henning Christian, Henning Urlaub, Ralf Ficner, Kum-Loong Boon, Patrizia Fabrizio, and Reinhard Lührmann. 2013. “Dissection of the Factor Requirements for Spliceosome Disassembly and the Elucidation of Its Dissociation Products Using a Purified Splicing System.” *Genes & Development* 27 (4): 413–28. https://doi.org/10.1101/gad.207779.112.

Freudenreich, Catherine H. 2018. “R-Loops: Targets for Nuclease Cleavage and Repeat Instability.” *Current Genetics* 64 (4): 789–94. https://doi.org/10.1007/s00294-018-0806-z.

García-Pichardo, Desiré, Juan C Cañas, María L García-Rubio, Belén Gómez-González, Ana G Rondón, and Andrés Aguilera. 2017. “Histone Mutants Separate R Loop Formation from Genome Instability Induction.” *Molecular Cell* 66 (5): 597-609.e5. https://doi.org/10.1016/j.molcel.2017.05.014.

Gavaldá, Sandra, Mercedes Gallardo, Rosa Luna, and Andrés Aguilera. 2013. “R-Loop Mediated Transcription-Associated Recombination in Trf4Δ Mutants Reveals New Links between RNA Surveillance and Genome Integrity.” *PLOS ONE* 8 (6): e65541. https://doi.org/10.1371/journal.pone.0065541.

Giaever, Guri, and Corey Nislow. 2014. “The Yeast Deletion Collection: A Decade of Functional Genomics.” *Genetics* 197 (2): 451–65. https://doi.org/10.1534/genetics.114.161620.

Gómez-González, Belén, María García-Rubio, Rodrigo Bermejo, Hélène Gaillard, Katsuhiko Shirahige, Antonio Marín, Marco Foiani, and Andrés Aguilera. 2011. “Genome-Wide Function of THO/TREX in Active Genes Prevents R-Loop-Dependent Replication Obstacles.” *The EMBO Journal* 30 (15): 3106–19. https://doi.org/10.1038/emboj.2011.206.

Graf, Marco, Diego Bonetti, Arianna Lockhart, Kamar Serhal, Vanessa Kellner, André Maicher, Pascale Jolivet, Maria Teresa Teixeira, and Brian Luke. 2017. “Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle.” *Cell* 170 (1): 72-85.e14. https://doi.org/10.1016/j.cell.2017.06.006.

Greger, I H, and N J Proudfoot. 1998a. “Poly(A) Signals Control Both Transcriptional Termination and Initiation between the Tandem GAL10 and GAL7 Genes of Saccharomyces Cerevisiae.” *The EMBO Journal* 17 (16): 4771–79. https://doi.org/10.1093/emboj/17.16.4771.

Grohmann, Dina, and Finn Werner. 2011. “Cycling through Transcription with the RNA Polymerase F/E (RPB4/7) Complex: Structure, Function and Evolution of Archaeal RNA Polymerase.” *Research in Microbiology*, Archaea and the Tree of Life, 162 (1): 10–18. https://doi.org/10.1016/j.resmic.2010.09.002.

Gruchota, Julita, Cyril Denby Wilkes, Olivier Arnaiz, Linda Sperling, and Jacek K. Nowak. 2017. “A Meiosis-Specific Spt5 Homolog Involved in Non-Coding Transcription.” *Nucleic Acids Research*, January, gkw1318. https://doi.org/10.1093/nar/gkw1318.

Halbach, Felix, Michaela Rode, and Elena Conti. 2012. “The Crystal Structure of S. Cerevisiae Ski2, a DExH Helicase Associated with the Cytoplasmic Functions of the Exosome.” *RNA* 18 (1): 124–34. https://doi.org/10.1261/rna.029553.111.

Hamperl, Stephan, Michael J. Bocek, Joshua C. Saldivar, Tomek Swigut, and Karlene A. Cimprich. 2017a. “Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses.” *Cell* 170 (4): 774-786.e19. https://doi.org/10.1016/j.cell.2017.07.043.

Harel-Sharvit, Liat, Naama Eldad, Gal Haimovich, Oren Barkai, Lea Duek, and Mordechai Choder. 2010. “RNA Polymerase II Subunits Link Transcription and MRNA Decay to Translation.” *Cell* 143 (4): 552–63. https://doi.org/10.1016/j.cell.2010.10.033.

Harris, J. Kirk, Scott T. Kelley, George B. Spiegelman, and Norman R. Pace. 2003. “The Genetic Core of the Universal Ancestor.” *Genome Research* 13 (3): 407–12. https://doi.org/10.1101/gr.652803.

Hatchi, Elodie, Konstantina Skourti-Stathaki, Steffen Ventz, Luca Pinello,

Angela Yen, Kinga Kamieniarz-Gdula, Stoil Dimitrov, et al. 2015. “BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair.” *Molecular Cell* 57 (4): 636–47. https://doi.org/10.1016/j.molcel.2015.01.011.

Hazelbaker, Dane Z., Sebastian Marquardt, Wiebke Wlotzka, and Stephen Buratowski. 2013. “Kinetic Competition between RNA Polymerase II and Sen1-Dependent Transcription Termination.” *Molecular Cell* 49 (1): 55–66. https://doi.org/10.1016/j.molcel.2012.10.014.

He, Chuan Hua, and Dindial Ramotar. 1999. “An Allele of the Yeast RPB7 Gene, Encoding an Essential Subunit of RNA Polymerase II, Reduces Cellular Resistance to the Antitumor Drug Bleomycin.” *Biochemistry and Cell Biology* 77 (4): 375–82. https://doi.org/10.1139/o99-039.

Herrera-Moyano, Emilia, Xénia Mergui, María L. García-Rubio, Sonia Barroso, and Andrés Aguilera. 2014. “The Yeast and Human FACT Chromatin-Reorganizing Complexes Solve R-Loop-Mediated Transcription–Replication Conflicts.” *Genes & Development* 28 (7): 735–48. https://doi.org/10.1101/gad.234070.113.

Hirtreiter, Angela, Gerke E. Damsma, Alan C. M. Cheung, Daniel Klose, Dina Grohmann, Erika Vojnic, Andrew C. R. Martin, Patrick Cramer, and Finn Werner. 2010. “Spt4/5 Stimulates Transcription Elongation through the RNA Polymerase Clamp Coiled-Coil Motif.” *Nucleic Acids Research* 38 (12): 4040–51. https://doi.org/10.1093/nar/gkq135.

Hirtreiter, Angela, Dina Grohmann, and Finn Werner. 2010. “Molecular Mechanisms of RNA Polymerase—the F/E (RPB4/7) Complex Is Required for High Processivity in Vitro.” *Nucleic Acids Research* 38 (2): 585–96. https://doi.org/10.1093/nar/gkp928.

Hobor, Fruzsina, Roberto Pergoli, Karel Kubicek, Dominika Hrossova, Veronika Bacikova, Michal Zimmermann, Josef Pasulka, Ctirad Hofr, Stepanka Vanacova, and Richard Stefl. 2011. “Recognition of Transcription Termination Signal by the Nuclear Polyadenylated RNA-Binding (NAB) 3 Protein.” *The Journal of Biological Chemistry* 286 (5): 3645–57. https://doi.org/10.1074/jbc.M110.158774.

Huang, Ying, Jia Fang, Mark T. Bedford, Yi Zhang, and Rui-Ming Xu. 2006. “Recognition of Histone H3 Lysine-4 Methylation by the Double Tudor Domain of JMJD2A.” *Science (New York, N.Y.)* 312 (5774): 748–51. https://doi.org/10.1126/science.1125162.

Huertas, Pablo, and Andrés Aguilera. 2003. “Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination.” *Molecular Cell* 12 (3): 711–21. https://doi.org/10.1016/j.molcel.2003.08.010.

Hur, Junho K., Yicheng Luo, Sungjin Moon, Maria Ninova, Georgi K. Marinov, Yun D. Chung, and Alexei A. Aravin. 2016. “Splicing-Independent Loading of TREX on Nascent RNA Is Required for Efficient Expression of Dual-Strand PiRNA Clusters in Drosophila.” *Genes & Development* 30 (7): 840–55. https://doi.org/10.1101/gad.276030.115.

Jackson, Ryan N., A. Alejandra Klauer, Bradley J. Hintze, Howard Robinson, Ambro van Hoof, and Sean J. Johnson. 2010. “The Crystal Structure of Mtr4 Reveals a Novel Arch Domain Required for RRNA Processing.” *The EMBO Journal* 29 (13): 2205–16. https://doi.org/10.1038/emboj.2010.107.

Jamai, Adil, Andrea Puglisi, and Michel Strubin. 2009. “Histone Chaperone Spt16 Promotes Redeposition of the Original H3-H4 Histones Evicted by Elongating RNA Polymerase.” *Molecular Cell* 35 (3): 377–83. https://doi.org/10.1016/j.molcel.2009.07.001.

Jonas, Stefanie, and Elisa Izaurralde. 2013. “The Role of Disordered Protein Regions in the Assembly of Decapping Complexes and RNP Granules.” *Genes & Development* 27 (24): 2628–41. https://doi.org/10.1101/gad.227843.113.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. “Highly Accurate Protein Structure Prediction with AlphaFold.” *Nature* 596 (7873): 583–89. https://doi.org/10.1038/s41586-021-03819-2.

Jun, Sung-Hoon, Matthew J. Reichlen, Momoko Tajiri, and Katsuhiko S. Murakami. 2011. “Archaeal RNA Polymerase and Transcription Regulation.” *Critical Reviews in Biochemistry and Molecular Biology* 46 (1): 27–40. https://doi.org/10.3109/10409238.2010.538662.

Kaplan, Craig D., Michael J. Holland, and Fred Winston. 2005. “Interaction between Transcription Elongation Factors and MRNA 3’-End Formation at the Saccharomyces Cerevisiae GAL10-GAL7 Locus.” *The Journal of Biological Chemistry* 280 (2): 913–22. https://doi.org/10.1074/jbc.M411108200.

Katahira, Jun. 2012. “MRNA Export and the TREX Complex.” *Biochimica et*

Biophysica Acta (BBA) - Gene Regulatory Mechanisms, Nuclear Transport and RNA Processing, 1819 (6): 507–13. https://doi.org/10.1016/j.bbagrm.2011.12.001.

Kato, Hiroaki, Kosuke Okazaki, Tetsushi Iida, Jun-ichi Nakayama, Yota Murakami, and Takeshi Urano. 2013. "Spt6 Prevents Transcription-Coupled Loss of Posttranslationally Modified Histone H3." *Scientific Reports* 3 (1): 2186. https://doi.org/10.1038/srep02186.

Keidel, Achim, Elena Conti, and Sebastian Falk. 2020. "Purification and Reconstitution of the S. Cerevisiae TRAMP and Ski Complexes for Biochemical and Structural Studies." *Methods in Molecular Biology (Clifton, N.J.)* 2062: 491–513. https://doi.org/10.1007/978-1-4939-9822-7_24.

Keller, W, S Bienroth, K M Lang, and G Christofori. 1991. "Cleavage and Polyadenylation Factor CPF Specifically Interacts with the Pre-MRNA 3' Processing Signal AAUAAA." *The EMBO Journal* 10 (13): 4241–49.

Khalil, Bilal, Dmytro Morderer, Phillip L. Price, Feilin Liu, and Wilfried Rossoll. 2018. "MRNP Assembly, Axonal Transport, and Local Translation in Neurodegenerative Diseases." *Brain Research* 1693 (Pt A): 75–91. https://doi.org/10.1016/j.brainres.2018.02.018.

Khazak, V, P P Sadhale, N A Woychik, R Brent, and E A Golemis. 1995. "Human RNA Polymerase II Subunit HsRPB7 Functions in Yeast and Influences Stress Survival and Cell Morphology." *Molecular Biology of the Cell* 6 (7): 759–75.

Khoshnevis, Sohail, Isabel Askenasy, Matthew C. Johnson, Maria D. Dattolo, Crystal L. Young-Erdos, M. Elizabeth Stroupe, and Katrin Karbstein. 2016. "The DEAD-Box Protein Rok1 Orchestrates 40S and 60S Ribosome Assembly by Promoting the Release of Rrp5 from Pre-40S Ribosomes to Allow for 60S Maturation." *PLoS Biology* 14 (6): e1002480. https://doi.org/10.1371/journal.pbio.1002480.

Kim, Minkyu, Nevan J. Krogan, Lidia Vasiljeva, Oliver J. Rando, Eduard Nedea, Jack F. Greenblatt, and Stephen Buratowski. 2004. "The Yeast Rat1 Exonuclease Promotes Transcription Termination by RNA Polymerase II." *Nature* 432 (7016): 517–22. https://doi.org/10.1038/nature03041.

Klein, Brianna J., Daniel Bose, Kevin J. Baker, Zahirah M. Yusoff, Xiaodong Zhang, and Katsuhiko S. Murakami. 2011. "RNA Polymerase and Transcription Elongation Factor Spt4/5 Complex Structure."

Proceedings of the National Academy of Sciences of the United States of America 108 (2): 546–50. https://doi.org/10.1073/pnas.1013828108.

Knezetic, J. A., and D. S. Luse. 1986. “The Presence of Nucleosomes on a DNA Template Prevents Initiation by RNA Polymerase II in Vitro.” *Cell* 45 (1): 95–104. https://doi.org/10.1016/0092-8674(86)90541-6.

Kolodziej, P A, N Woychik, S M Liao, and R A Young. 1990a. “RNA Polymerase II Subunit Composition, Stoichiometry, and Phosphorylation.” *Molecular and Cellular Biology* 10 (5): 1915–20.
———. 1990b. “RNA Polymerase II Subunit Composition, Stoichiometry, and Phosphorylation.” *Molecular and Cellular Biology* 10 (5): 1915–20.

Krogan, Nevan J., Minkyu Kim, Seong Hoon Ahn, Guoqing Zhong, Michael S. Kobor, Gerard Cagney, Andrew Emili, Ali Shilatifard, Stephen Buratowski, and Jack F. Greenblatt. 2002. “RNA Polymerase II Elongation Factors of Saccharomyces Cerevisiae: A Targeted Proteomics Approach.” *Molecular and Cellular Biology* 22 (20): 6979–92. https://doi.org/10.1128/MCB.22.20.6979-6992.2002.

Kujirai, Tomoya, Haruhiko Ehara, Yuka Fujino, Mikako Shirouzu, Shun-Ichi Sekine, and Hitoshi Kurumizaka. 2018. “Structural Basis of the Nucleosome Transition during RNA Polymerase II Passage.” *Science (New York, N.Y.)* 362 (6414): 595–98. https://doi.org/10.1126/science.aau9904.

Kumar, Deepak, and Nimisha Sharma. 2019. “Genome-Wide Transcriptional Response to Altered Levels of the Rpb7 Subunit of RNA Polymerase II Identifies Its Role in DNA Damage Response in Schizosaccharomyces Pombe.” *FEMS Yeast Research* 19 (1): foy118. https://doi.org/10.1093/femsyr/foy118.

Kumar, Deepak, Swati Varshney, Shantanu Sengupta, and Nimisha Sharma. 2019. “A Comparative Study of the Proteome Regulated by the Rpb4 and Rpb7 Subunits of RNA Polymerase II in Fission Yeast.” *Journal of Proteomics* 199 (March). https://doi.org/10.1016/j.jprot.2019.03.007.

Kyrpides, Nikos C., and Christos A. Ouzounis. 1999. “Transcription in Archaea.” *Proceedings of the National Academy of Sciences of the United States of America* 96 (15): 8545–50.

Lafuente-Barquero, Juan, Maria Luisa García-Rubio, Marta San Martin-Alonso, Belén Gómez-González, and Andrés Aguilera. 2020. “Harmful DNA:RNA Hybrids Are Formed in Cis and in a Rad51-Independent

Manner." *ELife* 9 (August): e56674. https://doi.org/10.7554/eLife.56674.

Lafuente-Barquero, Juan, Sarah Luke-Glaser, Marco Graf, Sonia Silva, Belén Gómez-González, Arianna Lockhart, Michael Lisby, Andrés Aguilera, and Brian Luke. 2017. "The Smc5/6 Complex Regulates the Yeast Mph1 Helicase at RNA-DNA Hybrid-Mediated DNA Damage." *PLoS Genetics* 13 (12): e1007136. https://doi.org/10.1371/journal.pgen.1007136.

Leporé, Nathalie, and Denis L. J. Lafontaine. 2011. "A Functional Interface at the RDNA Connects RRNA Synthesis, Pre-RRNA Processing and Nucleolar Surveillance in Budding Yeast." *PLoS ONE* 6 (9): e24962. https://doi.org/10.1371/journal.pone.0024962.

Li, J., R. Horwitz, S. McCracken, and J. Greenblatt. 1992. "NusG, a New Escherichia Coli Elongation Factor Involved in Transcriptional Antitermination by the N Protein of Phage Lambda." *The Journal of Biological Chemistry* 267 (9): 6012–19.

Li, Shisheng, and Michael J. Smerdon. 2002. "Rpb4 and Rpb9 Mediate Subpathways of Transcription-Coupled DNA Repair in Saccharomyces Cerevisiae." *The EMBO Journal* 21 (21): 5921–29. https://doi.org/10.1093/emboj/cdf589.

Li, Wentao, Cristina Giles, and Shisheng Li. 2014. "Insights into How Spt5 Functions in Transcription Elongation and Repressing Transcription Coupled DNA Repair." *Nucleic Acids Research* 42 (11): 7069–83. https://doi.org/10.1093/nar/gku333.

Li, Xialu, and James L. Manley. 2005. "Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability." *Cell* 122 (3): 365–78. https://doi.org/10.1016/j.cell.2005.06.008.

Licatalosi, Donny D, Gabrielle Geiger, Michelle Minet, Stephanie Schroeder, Kate Cilli, J Bryan McNeil, and David L Bentley. 2002. "Functional Interaction of Yeast Pre-MRNA 3' End Processing Factors with RNA Polymerase II." *Molecular Cell* 9 (5): 1101–11. https://doi.org/10.1016/s1097-2765(02)00518-x.

Lidschreiber, Michael, Kristin Leike, and Patrick Cramer. 2013. "Cap Completion and C-Terminal Repeat Domain Kinase Recruitment Underlie the Initiation-Elongation Transition of RNA Polymerase II." *Molecular and Cellular Biology* 33 (19): 3805–16.

https://doi.org/10.1128/MCB.00361-13.

Lim, Junghyun, Pankaj Kumar Giri, David Kazadi, Brice Laffleur, Wanwei Zhang, Veronika Grinstein, Evangelos Pefanis, et al. 2017. “Nuclear Proximity of Mtr4 to RNA Exosome Restricts DNA Mutational Asymmetry.” *Cell* 169 (3): 523-537.e15. https://doi.org/10.1016/j.cell.2017.03.043.

Lin, Yunfu, Sharon Y.R. Dent, John H. Wilson, Robert D. Wells, and Marek Napierala. 2010. “R Loops Stimulate Genetic Instability of CTG·CAG Repeats.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (2): 692–97. https://doi.org/10.1073/pnas.0909740107.

Lindstrom, D. L., S. L. Squazzo, N. Muster, T. A. Burckin, K. C. Wachter, C. A. Emigh, J. A. McCleery, J. R. Yates, and G. A. Hartzog. 2003. “Dual Roles for Spt5 in Pre-MRNA Processing and Transcription Elongation Revealed by Identification of Spt5-Associated Proteins.” *Molecular and Cellular Biology* 23 (4): 1368–78. https://doi.org/10.1128/MCB.23.4.1368-1378.2003.

Liu, Bin, and Thomas A. Steitz. 2017. “Structural Insights into NusG Regulating Transcription Elongation.” *Nucleic Acids Research* 45 (2): 968–74. https://doi.org/10.1093/nar/gkw1159.

Loidl, Josef, Knud Nairz, and Franz Klein. 1991. “Meiotic Chromosome Synapsis in a Haploid Yeast.” *Chromosoma* 100 (4): 221–28. https://doi.org/10.1007/BF00344155.

Lorch, Yahli, Janice W. LaPointe, and Roger D. Kornberg. 1987. “Nucleosomes Inhibit the Initiation of Transcription but Allow Chain Elongation with the Displacement of Histones.” *Cell* 49 (2): 203–10. https://doi.org/10.1016/0092-8674(87)90561-7.

Luo, Weifei, and David Bentley. 2004. “A Ribonucleolytic Rat Torpedoes RNA Polymerase II.” *Cell* 119 (7): 911–14. https://doi.org/10.1016/j.cell.2004.11.041.

Luo, Weifei, Arlen W. Johnson, and David L. Bentley. 2006. “The Role of Rat1 in Coupling MRNA 3’-End Processing to Transcription Termination: Implications for a Unified Allosteric-Torpedo Model.” *Genes & Development* 20 (8): 954–65. https://doi.org/10.1101/gad.1409106.

Luse, Donal S, Lisa C Spangler, and Andrea Újvári. 2011. “Efficient and

Rapid Nucleosome Traversal by RNA Polymerase II Depends on a Combination of Transcript Elongation Factors." *The Journal of Biological Chemistry* 286 (8): 6040–48. https://doi.org/10.1074/jbc.m110.174722.

Ma, Wai Kit, Sara C. Cloutier, and Elizabeth J. Tran. 2013. "The DEAD-Box Protein Dbp2 Functions with the RNA-Binding Protein Yra1 to Promote MRNP Assembly." *Journal of Molecular Biology* 425 (20): 3824–38. https://doi.org/10.1016/j.jmb.2013.05.016.

Maillet, Isabelle, Jean Marie Buhler, André Sentenac, and Jean Labarre. 1999. "Rpb4p Is Necessary for RNA Polymerase II Activity at High Temperature *." *Journal of Biological Chemistry* 274 (32): 22586–90. https://doi.org/10.1074/jbc.274.32.22586.

Malig, Maika, Stella R. Hartono, Jenna M. Giafaglione, Lionel A. Sanz, and Frederic Chedin. 2020. "Ultra-Deep Coverage Single-Molecule R-Loop Footprinting Reveals Principles of R-Loop Formation." *Journal of Molecular Biology* 432 (7): 2271–88. https://doi.org/10.1016/j.jmb.2020.02.014.

Malone, E. A., J. S. Fassler, and F. Winston. 1993. "Molecular and Genetic Characterization of SPT4, a Gene Important for Transcription Initiation in Saccharomyces Cerevisiae." *Molecular & General Genetics: MGG* 237 (3): 449–59. https://doi.org/10.1007/BF00279450.

Manzo, Stefano G., Stella R. Hartono, Lionel A. Sanz, Jessica Marinello, Sara De Biasi, Andrea Cossarizza, Giovanni Capranico, and Frederic Chedin. 2018. "DNA Topoisomerase I Differentially Modulates R-Loops across the Human Genome." *Genome Biology* 19 (1): 100. https://doi.org/10.1186/s13059-018-1478-1.

Maudlin, Isabella E., and Jean D. Beggs. 2019. "Spt5 Modulates Cotranscriptional Spliceosome Assembly in Saccharomyces Cerevisiae." *RNA (New York, N.Y.)* 25 (10): 1298–1310. https://doi.org/10.1261/rna.070425.119.

Mayer, Andreas, Amelie Schreieck, Michael Lidschreiber, Kristin Leike, Dietmar E. Martin, and Patrick Cramer. 2012. "The Spt5 C-Terminal Region Recruits Yeast 3' RNA Cleavage Factor I." *Molecular and Cellular Biology* 32 (7): 1321–31. https://doi.org/10.1128/MCB.06310-11.

McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S.

D. Patterson, M. Wickens, and D. L. Bentley. 1997. “The C-Terminal Domain of RNA Polymerase II Couples MRNA Processing to Transcription.” *Nature* 385 (6614): 357–61. https://doi.org/10.1038/385357a0.

McCullough, Laura, Zaily Connell, Charisse Petersen, and Tim Formosa. 2015. “The Abundant Histone Chaperones Spt6 and FACT Collaborate to Assemble, Inspect, and Maintain Chromatin Structure in Saccharomyces Cerevisiae.” *Genetics* 201 (3): 1031–45. https://doi.org/10.1534/genetics.115.180794.

McDonald, Seth M., Devin Close, Hua Xin, Tim Formosa, and Christopher P. Hill. 2010. “Structure and Biological Importance of the Spn1-Spt6 Interaction, and Its Regulatory Role in Nucleosome Binding.” *Molecular Cell* 40 (5): 725–35. https://doi.org/10.1016/j.molcel.2010.11.014.

Mehmetbeyoglu, Ecmel, Leila Kianmehr, Murat Borlu, Zeynep Yilmaz, Seyma Basar Kılıc, Hassan Rajabi-Maham, Serpil Taheri, and Minoo Rassoulzadegan. 2022. “Decrease in RNase HII and Accumulation of LncRNAs/DNA Hybrids: A Causal Implication in Psoriasis?” *Biomolecules* 12 (3): 368. https://doi.org/10.3390/biom12030368.

Meinel, Dominik M., and Katja Sträßer. 2015. “Co-Transcriptional MRNP Formation Is Coordinated within a Molecular MRNP Packaging Station in S. Cerevisiae.” *BioEssays* 37 (6): 666–77. https://doi.org/10.1002/bies.201400220.

Meka, Hedije, Finn Werner, Suzanne C. Cordell, Silvia Onesti, and Peter Brick. 2005. “Crystal Structure and RNA Binding of the Rpb4/Rpb7 Subunits of Human RNA Polymerase II.” *Nucleic Acids Research* 33 (19): 6435–44. https://doi.org/10.1093/nar/gki945.

Meyer, Peter A., Sheng Li, Mincheng Zhang, Kentaro Yamada, Yuichiro Takagi, Grant A. Hartzog, and Jianhua Fu. 2015. “Structures and Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5.” *Molecular and Cellular Biology* 35 (19): 3354–69. https://doi.org/10.1128/MCB.00520-15.

Mitra, Pallabi, Gairika Ghosh, Md Hafeezunnisa, and Ranjan Sen. 2017. “Rho Protein: Roles and Mechanisms.” *Annual Review of Microbiology* 71 (September): 687–709. https://doi.org/10.1146/annurev-micro-030117-020432.

Mitsuzawa, Hiroshi, Emi Kanda, and Akira Ishihama. 2003. “Rpb7 Subunit of

RNA Polymerase II Interacts with an RNA-Binding Protein Involved in Processing of Transcripts." *Nucleic Acids Research* 31 (16): 4696–4701.

Mosley, Amber L., Gerald O. Hunter, Mihaela E. Sardiu, Michaela Smolle, Jerry L. Workman, Laurence Florens, and Michael P. Washburn. 2013. "Quantitative Proteomics Demonstrates That the RNA Polymerase II Subunits Rpb4 and Rpb7 Dissociate during Transcriptional Elongation." *Molecular & Cellular Proteomics: MCP* 12 (6): 1530–38. https://doi.org/10.1074/mcp.M112.024034.

Nechaev, Sergei, and Karen Adelman. 2011. "Pol II Waiting in the Starting Gates: Regulating the Transition from Transcription Initiation into Productive Elongation." *Biochimica et Biophysica Acta* 1809 (1): 34–45. https://doi.org/10.1016/j.bbagrm.2010.11.001.

Nielsen, Klaus H., Leos Valášek, Caroah Sykes, Antonina Jivotovskaya, and Alan G. Hinnebusch. 2006. "Interaction of the RNP1 Motif in PRT1 with HCR1 Promotes 40S Binding of Eukaryotic Initiation Factor 3 in Yeast." *Molecular and Cellular Biology* 26 (8): 2984–98. https://doi.org/10.1128/MCB.26.8.2984-2998.2006.

Orlicky, S. M., P. T. Tran, M. H. Sayre, and A. M. Edwards. 2001. "Dissociable Rpb4-Rpb7 Subassembly of Rna Polymerase II Binds to Single-Strand Nucleic Acid and Mediates a Post-Recruitment Step in Transcription Initiation." *The Journal of Biological Chemistry* 276 (13): 10097–102. https://doi.org/10.1074/jbc.M003165200.

Orozco, Ian J., Steven J. Kim, and Harold G. Martinson. 2002. "The Poly(A) Signal, without the Assistance of Any Downstream Element, Directs RNA Polymerase II to Pause in Vivo and Then to Release Stochastically from the Template." *The Journal of Biological Chemistry* 277 (45): 42899–911. https://doi.org/10.1074/jbc.M207415200.

Orphanides, G., W. H. Wu, W. S. Lane, M. Hampsey, and D. Reinberg. 1999. "The Chromatin-Specific Transcription Elongation Factor FACT Comprises Human SPT16 and SSRP1 Proteins." *Nature* 400 (6741): 284–88. https://doi.org/10.1038/22350.

Parajuli, Shankar, Daniel C. Teasley, Bhavna Murali, Jessica Jackson, Alessandro Vindigni, and Sheila A. Stewart. 2017. "Human Ribonuclease H1 Resolves R-Loops and Thereby Enables Progression of the DNA Replication Fork." *The Journal of Biological Chemistry* 292 (37): 15216–24. https://doi.org/10.1074/jbc.M117.787473.

Parker, Roy. 2012. “RNA Degradation in Saccharomyces Cerevisae.” *Genetics* 191 (3): 671–702. https://doi.org/10.1534/genetics.111.137265.

Paulson, Henry. 2018. “Repeat Expansion Diseases.” *Handbook of Clinical Neurology* 147: 105–23. https://doi.org/10.1016/B978-0-444-63233-3.00009-9.

Pérez Cañadillas, José Manuel, and Gabriele Varani. 2003. “Recognition of GU-Rich Polyadenylation Regulatory Elements by Human CstF-64 Protein.” *The EMBO Journal* 22 (11): 2821–30. https://doi.org/10.1093/emboj/cdg259.

Peyroche, Gérald, Erwann Levillain, Magali Siaut, Isabelle Callebaut, Patrick Schultz, André Sentenac, Michel Riva, and Christophe Carles. 2002. “The A14–A43 Heterodimer Subunit in Yeast RNA Pol I and Their Relationship to Rpb4–Rpb7 Pol II Subunits.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (23): 14670–75. https://doi.org/10.1073/pnas.232580799.

Pfeiffer, Verena, Jérôme Crittin, Larissa Grolimund, and Joachim Lingner. 2013. “The THO Complex Component Thp2 Counteracts Telomeric R-Loops and Telomere Shortening.” *The EMBO Journal* 32 (21): 2861–71. https://doi.org/10.1038/emboj.2013.217.

Phan, Lon, Lori W. Schoenfeld, Leoš Valášek, Klaus H. Nielsen, and Alan G. Hinnebusch. 2001. “A Subcomplex of Three EIF3 Subunits Binds EIF1 and EIF5 and Stimulates Ribosome Binding of MRNA and TRNAiMet.” *The EMBO Journal* 20 (11): 2954–65. https://doi.org/10.1093/emboj/20.11.2954.

Phan, Lon, Xiaolong Zhang, Katsura Asano, James Anderson, Hans-Peter Vornlocher, Jay R. Greenberg, Jun Qin, and Alan G. Hinnebusch. 1998. “Identification of a Translation Initiation Factor 3 (EIF3) Core Complex, Conserved in Yeast and Mammals, That Interacts with EIF5.” *Molecular and Cellular Biology* 18 (8): 4935–46.

Ponting, Chris P. 2002. “Novel Domains and Orthologues of Eukaryotic Transcription Elongation Factors.” *Nucleic Acids Research* 30 (17): 3643–52. https://doi.org/10.1093/nar/gkf498.

Powell, Weston T., Rochelle L. Coulson, Michael L. Gonzales, Florence K. Crary, Spencer S. Wong, Sarrita Adams, Robert A. Ach, et al. 2013.

"R-Loop Formation at Snord116 Mediates Topotecan Inhibition of Ube3a-Antisense and Allele-Specific Chromatin Decondensation." *Proceedings of the National Academy of Sciences* 110 (34): 13938–43. https://doi.org/10.1073/pnas.1305426110.

Prado, Félix, and Andrés Aguilera. 2005. "Impairment of Replication Fork Progression Mediates RNA PolII Transcription-Associated Recombination." *The EMBO Journal* 24 (6): 1267–76. https://doi.org/10.1038/sj.emboj.7600602.

Proudfoot, Nick J. 2011. "Ending the Message: Poly(A) Signals Then and Now." *Genes & Development* 25 (17): 1770–82. https://doi.org/10.1101/gad.17268411.

Qiu, Yijun, and David S. Gilmour. 2017. "Identification of Regions in the Spt5 Subunit of DRB Sensitivity-Inducing Factor (DSIF) That Are Involved in Promoter-Proximal Pausing." *The Journal of Biological Chemistry* 292 (13): 5555–70. https://doi.org/10.1074/jbc.M116.760751.

Qiu, Zilong, and Rongrong Jiang. 2017. "Improving Saccharomyces Cerevisiae Ethanol Production and Tolerance via RNA Polymerase II Subunit Rpb7." *Biotechnology for Biofuels* 10 (1): 125. https://doi.org/10.1186/s13068-017-0806-0.

Qu, Xiangping, Søren Lykke-Andersen, Tommy Nasser, Cyril Saguez, Edouard Bertrand, Torben Jensen, and Claire Moore. 2009. "Assembly of an Export-Competent MRNP Is Needed for Efficient Release of the 3'-End Processing Complex after Polyadenylation." *Molecular and Cellular Biology* 29 (August): 5327–38. https://doi.org/10.1128/MCB.00468-09.

Quan, Tiffani Kiyoko, and Grant Ashley Hartzog. 2010. "Histone H3K4 and K36 Methylation, Chd1 and Rpd3S Oppose the Functions of Saccharomyces Cerevisiae Spt4-Spt5 in Transcription." *Genetics* 184 (2): 321–34. https://doi.org/10.1534/genetics.109.111526.

Radford, Alan. 1991. "Methods in Yeast Genetics — A Laboratory Course Manual by M Rose, F Winston and P Hieter. Pp 198. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. 1990. $34 ISBN 0-87969-354-1." *Biochemical Education* 19 (2): 101–2. https://doi.org/10.1016/0307-4412(91)90039-B.

Rondón, Ana G., and Andrés Aguilera. 2019. "R-Loops as Promoters of Antisense Transcription." *Molecular Cell* 76 (4): 529–30.

https://doi.org/10.1016/j.molcel.2019.11.001.

Rose, M. D., and G. R. Fink. 1987. “KAR1, a Gene Required for Function of Both Intranuclear and Extranuclear Microtubules in Yeast.” *Cell* 48 (6): 1047–60. https://doi.org/10.1016/0092-8674(87)90712-4.

Rosenheck, Sonia, and Mordechai Choder. 1998. “Rpb4, a Subunit of RNA Polymerase II, Enables the Enzyme To Transcribe at Temperature Extremes In Vitro.” *Journal of Bacteriology* 180 (23): 6187–92.

Röther, Susanne, Cornelia Burkert, Katharina M. Brünger, Andreas Mayer, Anja Kieser, and Katja Strässer. 2010. “Nucleocytoplasmic Shuttling of the La Motif-Containing Protein Sro9 Might Link Its Nuclear and Cytoplasmic Functions.” *RNA (New York, N.Y.)* 16 (7): 1393–1401. https://doi.org/10.1261/rna.2089110.

Rubbi, L., S. Labarre-Mariotte, S. Chédin, and P. Thuriaux. 1999. “Functional Characterization of ABC10alpha, an Essential Polypeptide Shared by All Three Forms of Eukaryotic DNA-Dependent RNA Polymerases.” *The Journal of Biological Chemistry* 274 (44): 31485–92. https://doi.org/10.1074/jbc.274.44.31485.

Runner, Vanessa M., Vladimir Podolny, and Stephen Buratowski. 2008. “The Rpb4 Subunit of RNA Polymerase II Contributes to Cotranscriptional Recruitment of 3’ Processing Factors.” *Molecular and Cellular Biology* 28 (6): 1883–91. https://doi.org/10.1128/MCB.01714-07.

Salvi, Jayesh S., Janet N. Y. Chan, Kirk Szafranski, Tony T. Liu, Jane D. Wu, Jonathan B. Olsen, Nurussaba Khanam, Betty P. K. Poon, Andrew Emili, and Karim Mekhail. 2014. “Roles for Pbp1 and Caloric Restriction in Genome and Lifespan Maintenance via Suppression of RNA-DNA Hybrids.” *Developmental Cell* 30 (2): 177–91. https://doi.org/10.1016/j.devcel.2014.05.013.

San Martin-Alonso, Marta, María E. Soler-Oliva, María García-Rubio, Tatiana García-Muse, and Andrés Aguilera. 2021. “Harmful R-Loops Are Prevented via Different Cell Cycle-Specific Mechanisms.” *Nature Communications* 12 (1): 4451. https://doi.org/10.1038/s41467-021-24737-x.

Santos-Rosa, Helena, Horacio Moreno, George Simos, Alexandra Segref, Birthe Fahrenkrog, Nelly Panté, and Ed Hurt. 1998. “Nuclear MRNA Export Requires Complex Formation between Mex67p and Mtr2p at the Nuclear Pores.” *Molecular and Cellular Biology* 18 (11): 6826–38.

Sanz, Lionel A., Stella R. Hartono, Yoong Wearn Lim, Sandra Steyaert,

Aparna Rajpurkar, Paul A. Ginno, Xiaoqin Xu, and Frédéric Chédin. 2016. “Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals.” *Molecular Cell* 63 (1): 167–78. https://doi.org/10.1016/j.molcel.2016.05.032.

Schneider, Claudia, and David Tollervey. 2013. “Threading the Barrel of the RNA Exosome.” *Trends in Biochemical Sciences* 38 (10): 485–93. https://doi.org/10.1016/j.tibs.2013.06.013.

Schulz, Daniel, Nicole Pirkl, Elisabeth Lehmann, and Patrick Cramer. 2014. “Rpb4 Subunit Functions Mainly in MRNA Synthesis by RNA Polymerase II♦.” *The Journal of Biological Chemistry* 289 (25): 17446–52. https://doi.org/10.1074/jbc.M114.568014.

Segref, A., K. Sharma, V. Doye, A. Hellwig, J. Huber, R. Lührmann, and E. Hurt. 1997. “Mex67p, a Novel Factor for Nuclear MRNA Export, Binds to Both Poly(A)+ RNA and Nuclear Pores.” *The EMBO Journal* 16 (11): 3256–71. https://doi.org/10.1093/emboj/16.11.3256.

Sharma, Nimisha, and Rashmi Kumari. 2013. “Rpb4 and Rpb7: Multifunctional Subunits of RNA Polymerase II.” *Critical Reviews in Microbiology* 39 (4): 362–72. https://doi.org/10.3109/1040841X.2012.711742.

Shiomi, Yasushi, and Hideo Nishitani. 2017. “Control of Genome Integrity by RFC Complexes; Conductors of PCNA Loading onto and Unloading from Chromatin during DNA Replication.” *Genes* 8 (2): 52. https://doi.org/10.3390/genes8020052.

Siaut, Magali, Cécile Zaros, Emilie Levivier, Maria-Laura Ferri, Magali Court, Michel Werner, Isabelle Callebaut, Pierre Thuriaux, André Sentenac, and Christine Conesa. 2003. “An Rpb4/Rpb7-Like Complex in Yeast RNA Polymerase III Contains the Orthologue of Mammalian CGRP-RCP.” *Molecular and Cellular Biology* 23 (1): 195–205. https://doi.org/10.1128/MCB.23.1.195-205.2003.

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.” *Proceedings of the National Academy of Sciences of the United States of America* 81 (8): 2431–34.

Singh, Guramrit, Gabriel Pratt, Gene W. Yeo, and Melissa J. Moore. 2015.

"The Clothes Make the MRNA: Past and Present Trends in MRNP Fashion." *Annual Review of Biochemistry* 84: 325–54. https://doi.org/10.1146/annurev-biochem-080111-092106.

Singh, Pragyan, Anusha Chaudhuri, Mayukh Banerjea, Neeraja Marathe, and Biswadip Das. 2021. "Nrd1p Identifies Aberrant and Natural Exosomal Target Messages during the Nuclear MRNA Surveillance in Saccharomyces Cerevisiae." *Nucleic Acids Research* 49 (20): 11512–36. https://doi.org/10.1093/nar/gkab930.

Singh, Sunanda R., Nambudiry Rekha, Beena Pillai, Vijender Singh, Aruna Naorem, Vinaya Sampath, N. Srinivasan, and Parag P. Sadhale. 2004. "Domainal Organization of the Lower Eukaryotic Homologs of the Yeast RNA Polymerase II Core Subunit Rpb7 Reflects Functional Conservation." *Nucleic Acids Research* 32 (1): 201–10. https://doi.org/10.1093/nar/gkh163.

Skourti-Stathaki, Konstantina, Kinga Kamieniarz-Gdula, and Nicholas J. Proudfoot. 2014. "R-Loops Induce Repressive Chromatin Marks over Mammalian Gene Terminators." *Nature* 516 (7531): 436–39. https://doi.org/10.1038/nature13787.

Skourti-Stathaki, Konstantina, Nicholas J. Proudfoot, and Natalia Gromak. 2011. "Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination." *Molecular Cell* 42 (6): 794–805. https://doi.org/10.1016/j.molcel.2011.04.026.

Smolka, John A., Lionel A. Sanz, Stella R. Hartono, and Frédéric Chédin. 2021. "Recognition of RNA by the S9.6 Antibody Creates Pervasive Artifacts When Imaging RNA:DNA Hybrids." *The Journal of Cell Biology* 220 (6): e202004079. https://doi.org/10.1083/jcb.202004079.

Song, Y., S. Kim, and J. Kim. 1995. "ROK1, a High-Copy-Number Plasmid Suppressor of Kem1, Encodes a Putative ATP-Dependent RNA Helicase in Saccharomyces Cerevisiae." *Gene* 166 (1): 151–54. https://doi.org/10.1016/0378-1119(96)80010-2.

Steiner, Thomas, Jens T. Kaiser, Snezan Marinkoviç, Robert Huber, and Markus C. Wahl. 2002. "Crystal Structures of Transcription Factor NusG in Light of Its Nucleic Acid- and Protein-Binding Activities." *The EMBO Journal* 21 (17): 4641–53. https://doi.org/10.1093/emboj/cdf455.

Steinmetz, E. J., N. K. Conrad, D. A. Brow, and J. L. Corden. 2001. "RNA-

Binding Protein Nrd1 Directs Poly(A)-Independent 3'-End Formation of RNA Polymerase II Transcripts." *Nature* 413 (6853): 327–31. https://doi.org/10.1038/35095090.

Stirling, Peter C., Yujia A. Chan, Sean W. Minaker, Maria J. Aristizabal, Irene Barrett, Payal Sipahimalani, Michael S. Kobor, and Philip Hieter. 2012. "R-Loop-Mediated Genome Instability in MRNA Cleavage and Polyadenylation Mutants." *Genes & Development* 26 (2): 163–75. https://doi.org/10.1101/gad.179721.111.

Strässer, K., J. Bassler, and E. Hurt. 2000. "Binding of the Mex67p/Mtr2p Heterodimer to FXFG, GLFG, and FG Repeat Nucleoporins Is Essential for Nuclear MRNA Export." *The Journal of Cell Biology* 150 (4): 695–706. https://doi.org/10.1083/jcb.150.4.695.

Strässer, K., and E. Hurt. 2001. "Splicing Factor Sub2p Is Required for Nuclear MRNA Export through Its Interaction with Yra1p." *Nature* 413 (6856): 648–52. https://doi.org/10.1038/35098113.

Strässer, Katja, Seiji Masuda, Paul Mason, Jens Pfannstiel, Marisa Oppizzi, Susana Rodriguez-Navarro, Ana G. Rondón, et al. 2002. "TREX Is a Conserved Complex Coupling Transcription with Messenger RNA Export." *Nature* 417 (6886): 304–8. https://doi.org/10.1038/nature746.

Swanson, M. S., M. Carlson, and F. Winston. 1990. "SPT6, an Essential Gene That Affects Transcription in Saccharomyces Cerevisiae, Encodes a Nuclear Protein with an Extremely Acidic Amino Terminus." *Molecular and Cellular Biology* 10 (9): 4935–41. https://doi.org/10.1128/mcb.10.9.4935-4941.1990.

Swanson, M. S., E. A. Malone, and F. Winston. 1991. "SPT5, an Essential Gene Important for Normal Transcription in Saccharomyces Cerevisiae, Encodes an Acidic Nuclear Protein with a Carboxy-Terminal Repeat." *Molecular and Cellular Biology* 11 (6): 3009–19. https://doi.org/10.1128/mcb.11.6.3009-3019.1991.

Tan, Qian, Xin Li, Parag P. Sadhale, Takenori Miyao, and Nancy A. Woychik. 2000. "Multiple Mechanisms of Suppression Circumvent Transcription Defects in an RNA Polymerase Mutant." *Molecular and Cellular Biology* 20 (21): 8124–33.

The UniProt Consortium. 2021. "UniProt: The Universal Protein Knowledgebase in 2021." *Nucleic Acids Research* 49 (D1): D480–89. https://doi.org/10.1093/nar/gkaa1100.

Thomas, M, R L White, and R W Davis. 1976. “Hybridization of RNA to Double-Stranded DNA: Formation of R-Loops.” *Proceedings of the National Academy of Sciences of the United States of America* 73 (7): 2294–98.

Tieg, Bettina, and Heike Krebber. 2013. “Dbp5 — From Nuclear Export to Translation.” *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, SI: The Biology of RNA helicases - Modulation for life, 1829 (8): 791–98. https://doi.org/10.1016/j.bbagrm.2012.10.010.

Todone, Flavia, Peter Brick, Finn Werner, Robert O. J. Weinzierl, and Silvia Onesti. 2001. “Structure of an Archaeal Homolog of the Eukaryotic RNA Polymerase II RPB4/RPB7 Complex.” *Molecular Cell* 8 (5): 1137–43. https://doi.org/10.1016/S1097-2765(01)00379-3.

Tran, Elizabeth J., Yingna Zhou, Anita H. Corbett, and Susan R. Wente. 2007. “The DEAD-Box Protein Dbp5 Controls MRNA Export by Triggering Specific RNA:Protein Remodeling Events.” *Molecular Cell* 28 (5): 850–59. https://doi.org/10.1016/j.molcel.2007.09.019.

Valásek, L., J. Hasek, K. H. Nielsen, and A. G. Hinnebusch. 2001. “Dual Function of EIF3j/Hcr1p in Processing 20 S Pre-RRNA and Translation Initiation.” *The Journal of Biological Chemistry* 276 (46): 43351–60. https://doi.org/10.1074/jbc.M106887200.

Valášek, Leoš, Lon Phan, Lori W. Schoenfeld, Věra Valášková, and Alan G. Hinnebusch. 2001. “Related EIF3 Subunits TIF32 and HCR1 Interact with an RNA Recognition Motif in PRT1 Required for EIF3 Integrity and Ribosome Binding.” *The EMBO Journal* 20 (4): 891–904. https://doi.org/10.1093/emboj/20.4.891.

Valášek, Leoš Shivaya. 2012. “‘Ribozoomin’ – Translation Initiation from the Perspective of the Ribosome-Bound Eukaryotic Initiation Factors (EIFs).” *Current Protein & Peptide Science* 13 (4): 305–30. https://doi.org/10.2174/138920312801619385.

Valášek, Leoš Shivaya, Jakub Zeman, Susan Wagner, Petra Beznosková, Zuzana Pavlíková, Mahabub Pasha Mohammad, Vladislava Hronová, Anna Herrmannová, Yaser Hashem, and Stanislava Gunišová. 2017. “Embraced by EIF3: Structural and Functional Insights into the Roles of EIF3 across the Translation Cycle.” *Nucleic Acids Research* 45 (19): 10948–68. https://doi.org/10.1093/nar/gkx805.

Varadi, Mihaly, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy

Natassia, Galabina Yordanova, David Yuan, et al. 2022. “AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models.” *Nucleic Acids Research* 50 (D1): D439–44. https://doi.org/10.1093/nar/gkab1061.

Vasiljeva, Lidia, and Stephen Buratowski. 2006. “Nrd1 Interacts with the Nuclear Exosome for 3’ Processing of RNA Polymerase II Transcripts.” *Molecular Cell* 21 (2): 239–48. https://doi.org/10.1016/j.molcel.2005.11.028.

Vasiljeva, Lidia, Minkyu Kim, Hannes Mutschler, Stephen Buratowski, and Anton Meinhart. 2008. “The Nrd1-Nab3-Sen1 Termination Complex Interacts with the Ser5-Phosphorylated RNA Polymerase II C-Terminal Domain.” *Nature Structural & Molecular Biology* 15 (8): 795–804. https://doi.org/10.1038/nsmb.1468.

Venema, J, C Bousquet-Antonelli, J P Gelugne, M Caizergues-Ferrer, and D Tollervey. 1997. “Rok1p Is a Putative RNA Helicase Required for RRNA Processing.” *Molecular and Cellular Biology* 17 (6): 3398–3407.

Verma-Gaur, Jiyoti, Sudha Narayana Rao, Toshiki Taya, and Parag Sadhale. 2008. “Genomewide Recruitment Analysis of Rpb4, a Subunit of Polymerase II in Saccharomyces Cerevisiae, Reveals Its Involvement in Transcription Elongation.” *Eukaryotic Cell* 7 (6): 1009–18. https://doi.org/10.1128/EC.00057-08.

Viktorovskaya, Olga V., Francis D. Appling, and David A. Schneider. 2011. “Yeast Transcription Elongation Factor Spt5 Associates with RNA Polymerase I and RNA Polymerase II Directly.” *The Journal of Biological Chemistry* 286 (21): 18825–33. https://doi.org/10.1074/jbc.M110.202119.

Vos, Seychelle M., Lucas Farnung, Marc Boehning, Christoph Wigge, Andreas Linden, Henning Urlaub, and Patrick Cramer. 2018. “Structure of Activated Transcription Complex Pol II-DSIF-PAF-SPT6.” *Nature* 560 (7720): 607–12. https://doi.org/10.1038/s41586-018-0440-4.

Vos, Seychelle M., Lucas Farnung, Henning Urlaub, and Patrick Cramer. 2018. “Structure of Paused Transcription Complex Pol II-DSIF-NELF.” *Nature* 560 (7720): 601–6. https://doi.org/10.1038/s41586-018-0442-2.

Wahba, Lamia, Lorenzo Costantino, Frederick J. Tan, Anjali Zimmer, and Douglas Koshland. 2016. “S1-DRIP-Seq Identifies High Expression

and PolyA Tracts as Major Contributors to R-Loop Formation." *Genes & Development* 30 (11): 1327–38. https://doi.org/10.1101/gad.280834.116.

Wang, S., Z. Han, D. Libri, O. Porrua, and T. R. Strick. 2019. "Single-Molecule Characterization of Extrinsic Transcription Termination by Sen1 Helicase." *Nature Communications* 10 (1): 1545. https://doi.org/10.1038/s41467-019-09560-9.

Washburn, Robert S., Philipp K. Zuber, Ming Sun, Yaser Hashem, Bingxin Shen, Wen Li, Sho Harvey, et al. 2020. "Escherichia Coli NusG Links the Lead Ribosome with the Transcription Elongation Complex." *IScience* 23 (8): 101352. https://doi.org/10.1016/j.isci.2020.101352.

Weaver, P. L., C. Sun, and T. H. Chang. 1997. "Dbp3p, a Putative RNA Helicase in Saccharomyces Cerevisiae, Is Required for Efficient Pre-RRNA Processing Predominantly at Site A3." *Molecular and Cellular Biology* 17 (3): 1354–65. https://doi.org/10.1128/MCB.17.3.1354.

Weir, John R., Fabien Bonneau, Jendrik Hentschel, and Elena Conti. 2010. "Structural Analysis Reveals the Characteristic Features of Mtr4, a DExH Helicase Involved in Nuclear RNA Processing and Surveillance." *Proceedings of the National Academy of Sciences of the United States of America* 107 (27): 12139–44. https://doi.org/10.1073/pnas.1004953107.

Werner, Finn. 2012. "A Nexus for Gene Expression-Molecular Mechanisms of Spt5 and NusG in the Three Domains of Life." *Journal of Molecular Biology* 417 (1–2): 13–27. https://doi.org/10.1016/j.jmb.2012.01.031.

Windgassen, Merle, and Heike Krebber. 2003. "Identification of Gbp2 as a Novel Poly(A)+ RNA-Binding Protein Involved in the Cytoplasmic Delivery of Messenger RNAs in Yeast." *EMBO Reports* 4 (3): 278–83. https://doi.org/10.1038/sj.embor.embor763.

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 (2): 179–97. https://doi.org/10.1093/genetics/107.2.179.

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 (Chichester, England)* 11 (1): 53–55. https://doi.org/10.1002/yea.320110107.

Wyers, Françoise, Mathieu Rougemaille, Gwenaël Badis, Jean-Claude Rousselle, Marie-Elisabeth Dufour, Jocelyne Boulay, Béatrice Régnault, et al. 2005. “Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase.” *Cell* 121 (5): 725–37. https://doi.org/10.1016/j.cell.2005.04.030.

Xie, Yihu, and Yi Ren. 2019. “Mechanisms of Nuclear MRNA Export: A Structural Perspective.” *Traffic (Copenhagen, Denmark)* 20 (11): 829–40. https://doi.org/10.1111/tra.12691.

Xu, Youwei, Carrie Bernecky, Chung-Tien Lee, Kerstin C. Maier, Björn Schwalb, Dimitry Tegunov, Jürgen M. Plitzko, Henning Urlaub, and Patrick Cramer. 2017. “Architecture of the RNA Polymerase II-Paf1C-TFIIS Transcription Elongation Complex.” *Nature Communications* 8 (June): 15741. https://doi.org/10.1038/ncomms15741.

Yang, Xuan, and Liangran Zhang. 2022. “Detection of RNA-DNA Hybrids by Immunostaining in Meiotic Nuclei of Saccharomyces Cerevisiae.” *STAR Protocols* 3 (2): 101325. https://doi.org/10.1016/j.xpro.2022.101325.

Ying, Muying, and Dahua Chen. 2012. “Tudor Domain-Containing Proteins of Drosophila Melanogaster.” *Development, Growth & Differentiation* 54 (1): 32–43. https://doi.org/10.1111/j.1440-169x.2011.01308.x.

Yoh, Sunnie M., Helen Cho, Loni Pickle, Ronald M. Evans, and Katherine A. Jones. 2007. “The Spt6 SH2 Domain Binds Ser2-P RNAPII to Direct Iws1-Dependent MRNA Splicing and Export.” *Genes & Development* 21 (2): 160–74. https://doi.org/10.1101/gad.1503107.

Yu, Kefei, Frederic Chedin, Chih-Lin Hsieh, Thomas E. Wilson, and Michael R. Lieber. 2003. “R-Loops at Immunoglobulin Class Switch Regions in the Chromosomes of Stimulated B Cells.” *Nature Immunology* 4 (5): 442–51. https://doi.org/10.1038/ni919.

Zawel, L., K. P. Kumar, and D. Reinberg. 1995. “Recycling of the General Transcription Factors during RNA Polymerase II Transcription.” *Genes & Development* 9 (12): 1479–90. https://doi.org/10.1101/gad.9.12.1479.

Zeng, Chao, Masahiro Onoguchi, and Michiaki Hamada. 2021. “Association Analysis of Repetitive Elements and R-Loop Formation across

Species.” *Mobile DNA* 12 (1): 3. https://doi.org/10.1186/s13100-021-00231-5.

Zenklusen, Daniel, Patrizia Vinciguerra, Jean-Christophe Wyss, and Françoise Stutz. 2002. “Stable MRNP Formation and Export Require Cotranscriptional Recruitment of the MRNA Export Factors Yra1p and Sub2p by Hpr1p.” *Molecular and Cellular Biology* 22 (23): 8241–53. https://doi.org/10.1128/MCB.22.23.8241-8253.2002.

Zhou, H., and K. A. Lee. 2001. “An HsRPB4/7-Dependent Yeast Assay for Trans-Activation by the EWS Oncogene.” *Oncogene* 20 (12): 1519–24. https://doi.org/10.1038/sj.onc.1204135.

Zhou, Huihao, Qi Liu, Yongxiang Gao, Maikun Teng, and Liwen Niu. 2009. “Crystal Structure of NusG N-Terminal (NGN) Domain from Methanocaldococcus Jannaschii and Its Interaction with RpoE″.” *Proteins: Structure, Function, and Bioinformatics* 76 (4): 787–93. https://doi.org/10.1002/prot.22465.

Zuber, Philipp K., Lukas Hahn, Anne Reinl, Kristian Schweimer, Stefan H. Knauer, Max E. Gottesman, Paul Rösch, and Birgitta M. Wöhrl. 2018. “Structure and Nucleic Acid Binding Properties of KOW Domains 4 and 6–7 of Human Transcription Elongation Factor DSIF.” *Scientific Reports* 8 (1): 11660. https://doi.org/10.1038/s41598-018-30042-3.

Evans, Richard, Michael O’Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Zidek, et al. 2021. “Protein Complex Prediction with AlphaFold-Multimer.” *BioRxiv*. https://doi.org/10.1101/2021.10.04.463034v1.

Mirdita, Milot, Lars von den Driesch, Clovis Galiez, Maria J. Martin, Johannes S"oding, and Martin Steinegger. 2017. “Uniclust Databases of Clustered and Deeply Annotated Protein Sequences and Alignments.” *Nucleic Acids Res.* 45 (D1): D170–76. https://doi.org/10.1093/nar/gkw1081.

Mirdita, Milot, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. 2022. “ColabFold: Making Protein Folding Accessible to All.” *Nature Methods*. https://doi.org/10.1038/s41592-022-01488-1.

Mirdita, Milot, Martin Steinegger, and Johannes S"oding. 2019. “MMseqs2 Desktop and Local Web Server App for Fast, Interactive Sequence Searches.” *Bioinformatics* 35 (16): 2856–58.

https://doi.org/10.1093/bioinformatics/bty1057.

Mitchell, Alex L, Alexandre Almeida, Martin Beracochea, Miguel Boland, Josephine Burgin, Guy Cochrane, Michael R Crusoe, et al. 2019. “MGnify: The Microbiome Analysis Resource in 2020.” *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkz1035.

Lee, Kwan Yin, Anand Chopra, Giovanni L Burke, Ziyan Chen, Jack F Greenblatt, Kyle K Biggar, and Marc D Meneghini. 2020. “A Crucial RNA-Binding Lysine Residue in the Nab3 RRM Domain Undergoes SET1 and SET3-Responsive Methylation.” *Nucleic Acids Research* 48 (6): 2897–2911. https://doi.org/10.1093/nar/gkaa029.