Stony Brook University

The role of the Kin28 protein kinase in chromatin remodeling and transcription termination

A Thesis Presented

by

Daniel Labuz

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Biochemistry and Cell Biology

Stony Brook University

December 2016

Stony Brook University

The Graduate School

Daniel Labuz

We, the thesis committee for the above candidate for the

Master of Science degree, hereby recommend

acceptance of this thesis.

Ed Luk Ph.D. – Thesis Advisor
Assistant Professor, Biochemistry and Cell Biology

Nancy Hollingsworth – Second Reader
Distinguished Teaching Professor, Biochemistry and Cell Biology

This thesis is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Thesis

The role of the Kin28 protein kinase in chromatin remodeling and transcription termination

by

Daniel Labuz

Master of Science

in

Biochemistry and Cell Biology

Stony Brook University

2016

The Kin28 subunit of the general transcription factor complex, TFIIH, is a kinase that phosphorylates the heptapeptide repeats at serine position 5 (Ser5) on the C-terminal domain of RNA polymerase II (Pol II). Optimal Ser5 phosphorylation is important during transcription for Pol II to transition from the initiation step to the elongation step in a process called promoter escape. During transcription initiation, the nucleosome directly downstream of the promoter, called +1, must be disassembled before Pol II can engage the transcription start site. Previous work showed that the transcription machinery plays a direct role in the disassembly of the +1 nucleosome, but the exact mechanism is currently unknown. We hypothesize that the disassembly of the +1 nucleosome occurs during promoter escape. To test this idea, I attempted to conditionally deplete Kin28, in the yeast, *Saccharomyces cerevisiae*, to block promoter escape. I then applied a modified genomic approach to map nascent Pol II-associated RNA molecules before and after Kin28 depletion. The Kin28 depletion failed to block promoter escape, possibly due to incomplete removal of Kin28 from the nucleus. Instead, I found that the Kin28 depleted strain accumulated Pol II-associated read-through transcripts beyond the transcription termination site suggesting that Kin28 plays a role in the processing of the nascent RNA of protein-coding genes.

Table of Contents

List of Figures

List of Abbreviations

µL – microliter
+1 nucleosome – Nucleosome that is directly downstream of the nucleosome free region
AMP – Adenosine monophosphate
ATP – Adenosine triphosphate
as mutant – Analog sensitive mutant
BAM – Binary file that contains all sequencing data and is aligned to genome
BED – Browser extensible data, file containing 6 or more tab delimited columns with chromosome, start, end, gene name, score, and strand of a given gene
Bp – basepair usually referring to DNA
cDNA – complementary DNA strand, DNA produced by reverse transcriptase
CF IA – Cleavage-polyadenylation factor IA, responsible for cleaving pre-mRNA
ChIP – Chromatin immunoprecipitation
ChIP-seq – Chromatin immunoprecipitation sequencing, Genome-wide mapping of proteins bound to DNA
ChIP-qPCR - Chromatin immunoprecipitation sequencing quantitative PCR
CMK – Chloromethyl-ketone moiety, ATP analog used to irreversibly inhibit engineered kinases
CPF – cleavage-polyadenylation factor, responsible for cleaving pre-mRNA
CTD – C-terminal Domain of RNA Polymerase II
CUT – Cryptic unstable transcript
DMSO – Dimethyl Sulfoxide
DNase I – degrades DNA
Fastq – File that contains all sequencing reads
FKBP – FK506 binding protein
FLAG – Tag used to purify specific proteins
FPKM – Fragment per kilobase million, a normalized measure of gene expression
FRB – FKBP12-rapamycin-binding, tag used on a specific protein to re-localize from nucleus to cytoplasm by anchor away
GAL1 – Yeast gene that is induced by galactose and repressed by glucose
GTF – General Transcription Factors
H3 – Histone 3, Core protein of nucleosomes
H4 – Histone 4, Core protein of nucleosomes
H2A – Histone 2A, Core protein of nucleosomes
H2A.Z – Histone variant of H2A
H2B – Histone 2B, Core protein of nucleosome
HTZ1 – Gene that encodes H2A.Z
INO80 – Protein complex, ATP-dependent chromatin remodeler that slides nucleosomes along DNA
IP – Immunoprecipitation
is mutant – inhibitor sensitive mutant
kDa – kilodaltons, measure of protein molecular weight
Kin28 – Subunit of TFIIH, kinase that phosphorylates Serine-5 of Pol II CTD
L - liters

m^7GpppN – nuclear cap binding complex
mL – milliliter
mM – millimolar
$MnCl_2$ – Manganese Chloride
mRNA – messenger RNA
NET-seq – Nascent elongating transcript sequencing
NFR – Nucleosome free region
NNS – Nrd1, Nab3, Sen1 complex that is required for Nrd1 transcription termination pathway
Nt – nucleotide
OD – optical density, used to measure concentration of yeast cells in liquid
OH – hydroxyl group
ORF – Open reading frame
Pc-genes – Protein-coding genes
PCR – polymerase chain reaction
PIC – Pre-initiation Complex
Pol II – RNA Polymerase II
Rad3 - Subunit of TFIIH, 5'-to-3' helicase
RAP – Rapamycin, chemical that links the FRB tagged protein to the FKBP tagged pre-ribosome
RNA – Ribonucleic Acid
RNA-seq – Ribonucleic acid sequencing, high-throughput method of sequencing millions of short sequences of cDNA created from RNA
Rpb3 – 3rd largest subunit of RNA Polymerase II
RPB3-3xFLAG – RPB3 subunit tagged with 3xFLAG on 3' end
RPM – Reads per million, normalization method for sequencing depth
RSC – Protein complex involved in sliding and disassembling nucleosomes around NFR
SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGR – Sequence graph, a file that contains locations of genome with corresponding expression value from sequenced reads
Ser5 – Serine-5 on RNA Polymerase II C-terminal Domain
snoRNA – small nucleolar RNA
Ssl2 – Subunit of TFIIH, 5'-to-3' translocase responsible for transcription bubble formation
STAR – Spliced Transcripts Alignment to a Reference, Bioinformatics program used to align sequenced reads to genome
SUT – Stable untranslated transcript
SWR1 – Protein complex, ATP-dependent chromatin remodeler that exchanges H2A for H2A.Z
TBP – TATA-Binding protein
TFIIH – Protein complex associated with transcription initiation and DNA repair
tRNA – transfer-RNA
TSS – Transcription Start Site
TTS – Transcription Termination Site
YPD – Yeast extract peptone dextrose

Acknowledgements

First, I would like to thank my mentor Ed Luk for guiding me and being extremely diligent yet helpful at every step of the scientific process, which has ultimately made me a better scientist as well as a better person overall. Also, I appreciate that at all times there were a free flow of ideas between each other and he always appreciated listening to my suggestions however outlandish they may be.

Second, I would like to thank all the current members of the Luk lab (Lu Sun, Chitra Mohan, Tai Li, and Agatha Sleboda) who have been amazing peers and inspiring colleagues throughout my time at Stony Brook University. I would also like to thank those who have graduated (Mike Tramantano and Christy Au) and moved on to new adventures in their scientific careers.

Third, I would like to thank Nancy Hollingsworth for her helpful comments and suggestions throughout my master's degree. Also, I would like to recognize her ability to be critical of data and outspoken, which is an amazing trait to have as a scientist.

Fourth, I would like to thank my family back home for being there to give me a break from the academic rigors of science and their support.

Fifth, I would like to thank Neta Dean for accepting me into the Biochemistry and Cell Biology Master's Program at Stony Brook University. I also like to thank her for all the great advice she has given to me throughout my time here.

Lastly, I would like to thank my best friend Daisy Morocho who has been a great supporter of my aspirations and goals since we met.

I. Introduction

1. The role of Kin28 during the early stages of transcription

The prerequisite for the assembly of the transcription machinery is the binding of TATA-binding protein (TBP) to the TATA box element. This is followed by the recruitment of the general transcription factors (GTFs), including TFIIA, IIB, IID, IIE, IIF, and IIH and RNA Polymerase II (Pol II) to the promoter (Orphanides et al., 1996). The transcription competent complex consists of over 50 proteins and is called the pre-initiation complex (PIC) (Robinson et al., 2016). In yeast, the C-terminal domain (CTD) of Pol II, has a 26-heptapeptide tandem repeat comprised of the consensus sequence, Y_1-S_2-P_3-T_4-S_5-P_6-S_7 (Corden et al., 1985; Jeronimo et al., 2013). The phosphorylation status of the Pol II CTD can change at each of the serine residues, as well as the threonine, that can in turn regulate the recruitment of other proteins to Pol II to modulate the transcription process itself or transcription-coupled activities, such as splicing, RNA processing, and termination (Jeronimo et al., 2013). Overall, the formation of the PIC is critical for transcription to occur and the phosphorylation state of Pol II CTD is critical for efficient transcription.

An essential component of the PIC, TFIIH, is an 11-subunit complex that has three enzymatic activities, two of which are necessary for transcription initiation (Gibbons et al., 2012; Murakami et al., 2012; Svejstrup et al., 1996). Kin28 is a kinase subunit of TFIIH that phosphorylates serine-5 (Ser5) of the heptapeptide repeats of Pol II CTD (Feaver et al., 1994). Ssl2, a subunit of TFIIH is a 5'-to-3' translocase that is responsible for the unwinding of the DNA duplex to form the transcription bubble

(Fishburn et al., 2015). Additionally, the Ssl2 translocase activity may be involved in "scanning" for a transcription start site (TSS) before initiation can occur (Fazal et al., 2015; Murakami et al., 2015). Lastly, the Rad3 subunit of TFIIH is a 5'-to-3' helicase which acts to unwind the DNA duplex during nucleotide excision repair (Feaver et al., 1993). My thesis work focused mainly on the role of Kin28 and the functional consequences of transcription after Kin28 depletion from yeast cells.

During transcription initiation, Kin28 phosphorylates Ser5 on Pol II CTD and this signals Pol II to transition from initiation to elongation in a step called promoter escape (Wong et al., 2014). Two different methods have been used to inactivate Kin28 to understand its role during transcription initiation (Rodríguez-Molina et al., 2016; Wong et al., 2014). One study inactivated Kin28 using the anchor-away technique. The anchor-away technique involves fusing a nuclear protein of interest with the FKBP12-rapamycin-binding (FRB) tag that can be dragged out of the nucleus by interacting with a second tag, called the FK506 binding protein (FKBP) on the pre-ribosome in a rapamycin-dependent manner (Haruki et al., 2008). Pol II chromatin immunoprecipitation sequencing (ChIP-seq) and chromatin immunoprecipitation quantitative PCR (ChIP-qPCR) data showed that Kin28 depletion by anchor-away causes Pol II to accumulate at the promoter of many genes (Wong et al., 2014). In a different study, inactivation of Kin28 kinase acitivity was accomplished using an inhibitor-sensitive allele of *KIN28 (kin28is*), which can irreversibly inhibit Kin28 kinase function by addition of the ATP analog CMK. This approach revealed that Pol II is stalled ~100 bp downstream of the TSS at the +2 nucleosome indicating that Kin28 function is required to move Pol II from this stalled state to productive elongation

(Rodríguez-Molina et al., 2016). Regardless of the exact genomic position relative to promoters of where Pol II would stalls in the *kin28* mutants, one would expect that after Kin28 inactivation that this would cause an accumulation of transcripts at the 5' ends of genes should occur.

2. The role of Pol II CTD phosphorylation after transcription initiation

The Pol II CTD plays an important role in recruiting proteins during transcription termination and processing of the pre-mRNA as these activities are coupled (Porrua and Libri, 2015; Proudfoot, 2011). For example, a truncated Pol II CTD in yeast cells causes both inefficient cleavage of the pre-mRNA and incorrect lengths of poly(A) tail on pre-mRNA (Licatalosi et al., 2002). During transcription elongation, Ser2 of Pol II CTD is phosphorylated by Ctk1 (Bowman and Kelly, 2014). Pol II Ser2 CTD recruits necessary pre-mRNA processing complexes as Pol II travels towards at the end of genes (Jeronimo et al., 2013; Porrua and Libri, 2015). Furthermore, Pol II CTD phosphorylation states appear to help determine which transcription termination pathways will be used as will be discussed below (Jeronimo et al., 2013).

Transcription termination in eukaryotes occurs after Pol II transcribes past the polyadenylation site followed by the cleavage of pre-mRNA by the 4-subunit cleavage-polyadenylation factor IA (CF IA) and the 15-subunit cleavage-polyadenylation factor (CPF) complexes and the dissociation of the elongating Pol II complex from the DNA duplex (Proudfoot, 2011). There are three models for how the elongating Pol II dissociates from the DNA template: the torpedo model the allosteric model, and the NRD1 Pathway (Porrua and Libri, 2015; Rosonina et al., 2006; Figure 1.1). In the

torpedo model after Pol II transcribes past the polyadenylation site, an endonuclease cuts the RNA separating the pre-mRNA and the downstream RNA fragment, which is still tightly associated with the Pol II. A 5'-to-3' exonuclease, called Rat1, engages the 5' end of this Pol II nascent RNA and rapidly degrades it. When Rat1 catches up with the elongating Pol II, Pol II is destabilized, thereby causing Pol II to release from the DNA template (Connelly and Manley, 1988; Porrua and Libri, 2015). Similarly, the allosteric model involves the cleavage of the pre-mRNA after Pol II transcribes past the polyadenylation site. However, this model involves destabilization of the elongating factors, which are important for Pol II to engage in productive elongation, thereby promoting the termination process (Logan et al., 1987; Rosonina et al., 2006). In both models CDF and CF complexes polyadenylate the end of the pre-mRNA and target it for export to the cytoplasm. The CDF and CF complexes interact with Ser2 phosphorylated Pol II CTD to cleave the mRNA after the polyadenylation sequence at the 3' end of protein-coding genes.

A third transcription termination model involves the RNA-binding proteins of the NNS complex (Nrd1, Nab3, Sen1), which interact with the RNA surveillance machinery called the exosome. This termination pathway, called Nrd1-dependent termination pathway primarily targets non-coding transcripts such as snoRNA and cryptic unstable transcripts (CUTs) (Jeronimo et al., 2013; Rondón et al., 2009; Schaughency et al., 2014). In the Nrd1 pathway, the NNS complex is recruited to the RNA through its interaction with Ser5 phosphorylated Pol II CTD (Figure 1.1). Sen1 of the NNS complex is a 5'-to-3' helicase that tracks along the nascent RNA until it contacts Pol II and destabilizes the elongation complex causing termination (Porrua and Libri, 2015). It is

suggested that the Nrd1 pathway may also act as a “fail-safe” mechanism for the termination of protein-coding genes (Rondón et al., 2009). Interestingly, Nrd1 is recruited to Ser5 phosphorylated Pol II CTD, which is the main target for Kin28 kinase activity (Vasiljeva et al., 2008). It has been shown that inactivation of *kin28* function using a temperature sensitive allele leads to inefficient transcription termination of a snoRNAs and CUTs (Gudipati et al., 2008). Furthermore, total RNA-seq analysis showed that Nrd1 depletion by anchor-away causes an accumulation of CUT and snoRNA transcripts beyond the TTS compared to wild type, suggesting that termination was impaired (Schaughency et al., 2014). More recently, it was shown that Kin28 is required for efficient transcription termination of several protein coding genes (e.g. *CHA1, HXT1, ACT1,* and *ASC1*) (Medler and Ansari, 2015). However, whether this is a general phenomenon remains to be determined.

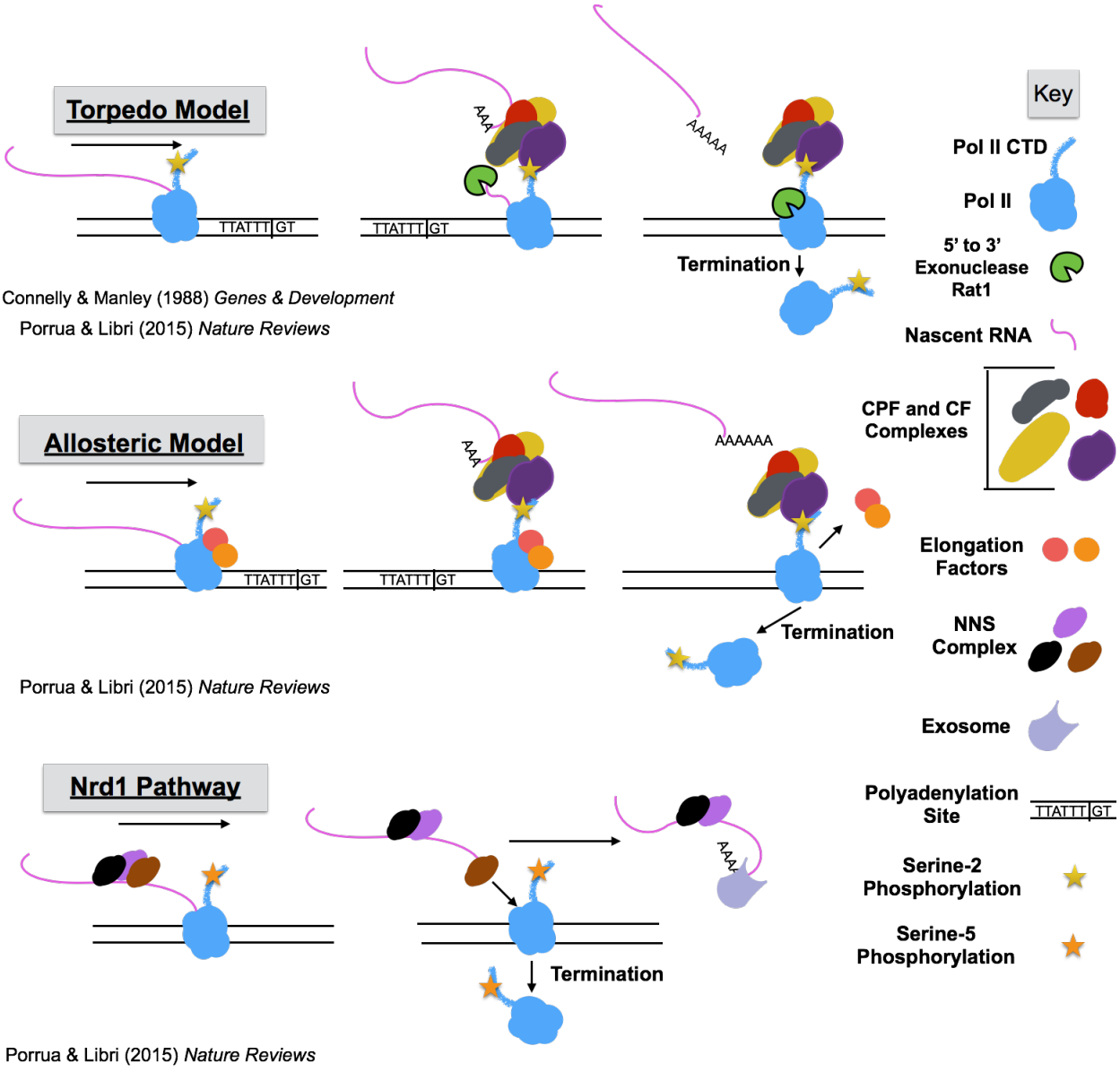


Figure 1.1. Three models for transcription termination in yeast. The torpedo model occurs when the transcribing RNA polymerase II is terminated by a 5'-to-3' exonuclease called Rat1. The allosteric model occurs when the transcribing RNA polymerase II becomes unstable after cleavage of pre-mRNA by CPF and CF complexes causing elongation factors to dissociate causing termination. The Nrd1 pathway occurs when Sen1 (in brown) a 5'-to-3' helicase contacts RNA polymerase II destabilizing the elongating RNA polymerase II causing termination.

3. Chromatin is the *in vivo* substrate of transcription

In eukaryotes, the regulation of all steps of transcription is influenced by the structure of chromatin. The most basic packaging unit of chromatin is a protein-DNA complex called a nucleosome. The protein core of a nucleosome consists of eight histones (two H2A-H2B dimers and an H3-H4 tetramer), around which wraps 147-bp DNA in 1.7 turns to form the full nucleosome. The presence of nucleosomes is generally inhibitory to the transcription process. For example, nucleosomes can block TATA box elements from TBP binding, thereby blocking PIC assembly (Han and Grunstein, 1988; Lorch et al., 1987). But for many genes, PIC assembly is not blocked by nucleosomes because their promoters contain an 80-200 nucleosome free region (NFR) upstream of the transcription start site (TSS). However for these genes, the first nucleosome downstream of the NFR, called +1, covers the TSS, thereby preventing Pol II from engaging the DNA during initiation (Rhee and Pugh, 2012). For Pol II to begin transcription, the +1 nucleosome must be dissembled by some mechanism.

The +1 nucleosome in yeast is enriched with the histone variant H2A.Z (Zhang et al., 2005). Previous studies have shown that H2A.Z is necessary for rapid induction of genes (Dhillon et al., 2006; Halley et al., 2010). For example, when *HTZ1* (the gene encoding H2A.Z) is knocked out, induction of the *GAL1* gene is noticeably slower compared to wild-type *HTZ1* (Halley et al., 2010). Furthermore, yeast cells with *htz1*Δ showed delayed cell cycle progression resulting from delayed induction of the G1-cyclin *CLN2* and the S-cyclin *CLB5,* which are necessary to transition out of G1-phase and S-phase, respectively (Dhillon et al., 2006). This evidence has led to the hypothesis that +1 nucleosomes enriched in H2A.Z aid transcription by allowing quick dissociation of

the histone core from the DNA. Recently, it has been shown that the transcription machinery can cause H2A.Z eviction but the specific mechanism that causes eviction of H2A.Z at promoters is unclear (Tramantano et al., 2016)

More is known about how H2A.Z is incorporated and subsequently positioned on the chromatin. An ATP dependent chromatin remodeler called SWR1 can exchange the H2A-H2B dimer with an H2A.Z-H2B dimer one at a time in a step-wise fashion (Luk et al., 2010). The targeting of SWR1 to the promoters is caused by long stretches of free DNA which are generally found at the NFR (Ranjan et al., 2013). Additionally, SWR1 contains a subunit called Bdf1 that is attracted to acetylated histones (Nguyen et al., 2013; Pamblanco et al., 2001; Ranjan et al., 2013). Furthermore, other ATP-dependent chromatin remodelers, such as RSC and INO80 appears to play a role in sliding the +1 nucleosome into the *in vivo* position (Krietenstein et al., 2016).

The ultimate goal of this study was to understand how the transcription machinery evicts H2A.Z. Previous studies have shown *in vitro* that H2A.Z containing nucleosomes are more unstable than their H2A counterparts (Jin and Felsenfeld, 2007; Placek et al., 2005). However, *in vivo* experiments have shown that after yeast cells are arrested using sodium azide, H2A.Z levels remained constant indicating that an enzymatic activity is required for H2A.Z eviction (E. Luk and M. Tramantano unpublished data). Furthermore, when SWR1 is inactivated *in vivo*, H2A.Z occupancy diminishes drop to baseline levels after 15 minutes further suggesting that there is an active mechanism evicting H2A.Z (Tramantano et al., 2016). Recently, Tramantano et al. (2016) showed that the transcription machinery acts as a chromatin remodeler that is essential in evicting +1 nucleosomes. Specifically, when TBP is conditionally depleted

from the nucleus the PIC is unable to form at the promoter causing H2A.Z levels to increase specifically at the +1 position. Thus, the transcription machinery appears to play a direct role in H2A.Z eviction. At this time, it is unknown which step of transcription is causing H2A.Z eviction. I hypothesized promoter escape is causing H2A.Z eviction. Thus, I tried blocking promoter escape by inactivating Kin28 to understand H2A.Z dynamics.

4. Thesis Statement: What are the specific mechanisms that evict +1 nucleosomes to allow for transition from initiation to elongation?

As mentioned above, the specific mechanistic step of transcription involved in the eviction of the +1 H2A.Z nucleosome is unknown. One driving force of H2A.Z containing nucleosome disassembly could involve the transition of Pol II from the initiation state to the elongation state in a process called promoter escape. Since promoter escape of Pol II requires Kin28 activity, if Kin28 is inactivated, an increase in H2A.Z levels at the +1 nucleosome should occur (Figure 1.2). This was the motivation to investigate whether inactivating Kin28 to stall Pol II promoter escape has any effect on H2A.Z removal.

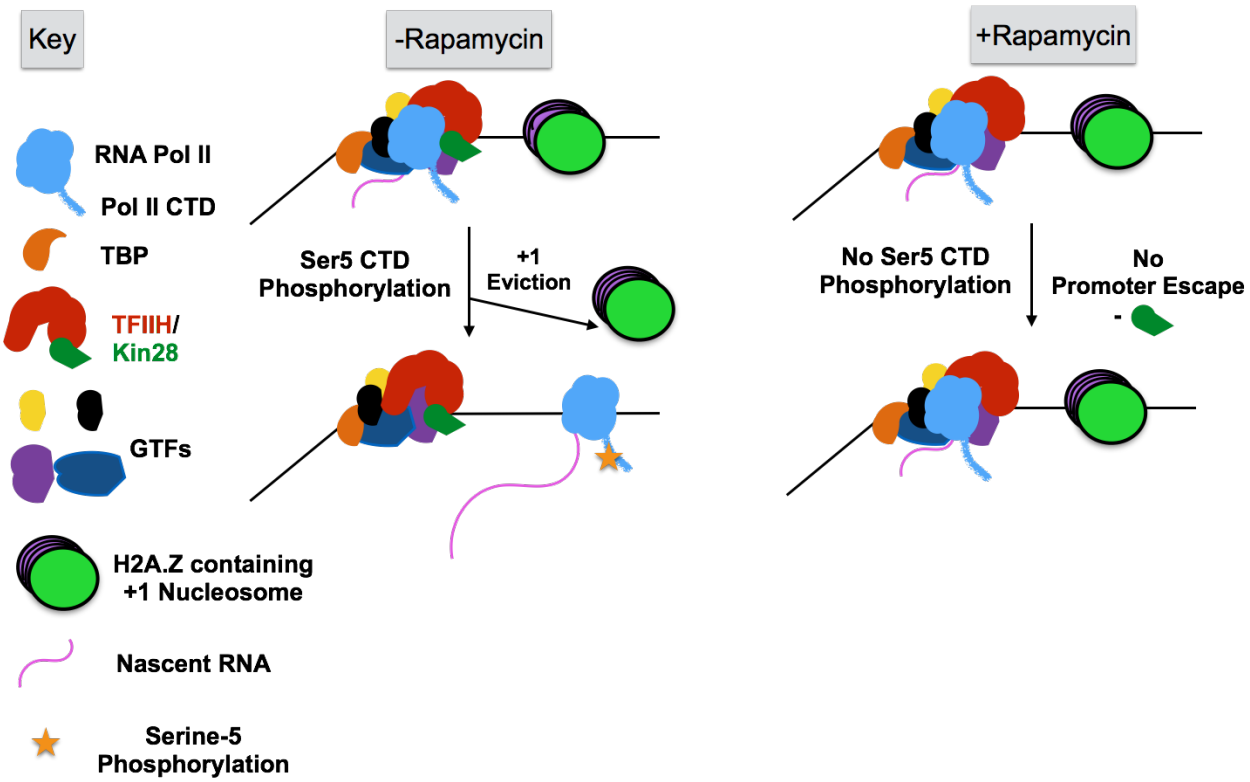


Figure 1.2. Model of inhibition of promoter escape causing an abundance of transcripts at the promoter. In the untreated rapamycin case promoter escape and +1 nucleosome eviction occurs and RNA polymerase II begins productive elongation. In the rapamycin treated case, Kin28 is depleted and promoter escape is blocked causing the RNA polymerase II to stall at the promoter with small RNA attached. TBP: TATA Binding Protein, GTFs: General Transcription Factors.

To test if the anchor-away approach can deplete Kin28 and block promoter escape I examined the distribution of nascent RNA associated with Pol II using a modified version of a protocol called nascent Pol II-associated RNA sequencing or nascent RNA-seq (Churchman and Weissman, 2012). Pol II-associated transcripts are predicted to accumulate if promoter escape is successfully blocked. If this is successful, H2A.Z levels can be analyzed before and after promoter escape. If promoter escape is

driving H2A.Z eviction, then Kin28 inactivation may cause an increase in H2A.Z at the +1 nucleosome. However H2A.Z eviction occurs at a step before promoter escape, blocking Kin28 should not cause any defects in H2A.Z eviction.

In the current study, I attempted to block Kin28 function by the anchor away approach. However, the expected accumulation of transcripts at the promoter was not observed, suggesting that promoter escape was not blocked. Therefore, I did not continue to measure H2A.Z occupancy at +1 nucleosome position. Instead, I found that Kin28 inactivation caused many Pol II-associated transcripts (~100) to extend beyond the transcription termination site, supporting the idea that Kin28 plays a role in the RNA processing step during transcription termination.

II. Results

1. Purification of the chromatin-bound RNA polymerase II complexes before and after Kin28 depletion

To deplete Kin28 conditionally by anchor-away, a yeast strain was used in which the 3' end of *KIN28* gene is fused in-frame with the FRB tag provided by M. Tramantano, Stony Brook University. As previously shown, the *KIN28-FRB* strain exhibited Pol II accumulation at promoters in a rapamycin-dependent manner (Wong et al., 2014). To verify that our *KIN28-FRB* allele responds to rapamycin, yeast strains

expressing either the *KIN28-FRB* or an untagged *KIN28* allele were spotted onto YPD plates with and without 1 µg/mL rapamycin. As seen in Figure 2.1A, the *KIN28-FRB* strain exhibited a slow growth phenotype compared to the wild-type control in the presence of rapamycin, consistent with the published results (Wong et al., 2014).

To allow the purification of the Pol II complex, a 3xFLAG tag was added on the 3' end of *RPB3*, which encodes a subunit of Pol II. The resulting *RPB3-3xFLAG* strain allows immunoprecipitation of the Pol II complex by the use of anti-FLAG agarose. The choice of tagging *RPB3* rather than other Pol II subunits was based on a previous study, which showed that intact Pol II and its associated RNA can be purified with the Rpb3-3xFLAG construct (Churchman and Weissman, 2011). Using the *RPB3-3xFLAG* and *KIN28-FRB* strain, two yeast cultures were grown to a final optical density 600 (OD_{600}) of 1.5. To deplete Kin28, rapamycin solubilized in dimethylsulfoxide (DMSO) was added to one culture at a final concentration of 1 µg/mL and DMSO was added to the other culture as a negative control. After 1 hour of rapamycin or DMSO treatment the cells were harvested, washed and then lysed using a freezer mill (E. Luk unpublished data).

Since my goal was to isolate nascent Pol II transcripts it was important to fragment the chromatin before the FLAG pull down to allow the purification of the chromatin-bound Pol II and its associated RNA. Therefore, the total lysate was incubated with DNase I for 30 minutes in the presence of 10 mM $MnCl_2$ at 4 ºC to liberate the Pol II from the chromatin. To assess the efficiency of the DNase I treatment, the DNA in the lysate was treated with proteinase K and the proteins extracted using phenol-chloroform and concentrated by ethanol-sodium acetate precipitation. As seen in

Figure 2.1B, ~90% of the genomic DNA was fragmented to below 1000 bp after DNase I treatment.

Next, the solubilized Pol II was subjected to anti-FLAG affinity chromatography. To evaluate if a sufficient amount of anti-FLAG agarose beads was used to bind the Rpb3-3xFLAG in the extract, IP efficiency was determined by comparing the amount of Rpb3-3xFLAG in the input and the supernatant after pull down by anti-FLAG western analysis. Densitometry showed that 70% of Rpb3 was immune-depleted from the input fraction, suggesting the IP was efficient (Figure 2.1C). The bead-bound Rpb3 complex was then eluted by incubation with 3xFLAG peptide (Churchman and Weissman, 2012).

To evaluate what other Pol II subunits co-purified with Rpb3, the FLAG peptide eluate was analyzed by SDS-PAGE and SyPRO Ruby staining. All twelve subunits of RNA polymerase II were successfully purified by anti-FLAG IP of Rpb3-3xFLAG (Figure 2.1E). The Pol II subunit bands were identified based on published SDS-PAGE analysis showing the staining pattern of the Pol II complex (Sayre et al., 1992). In addition to the expected Pol II subunits, many other proteins co-eluted with Rpb3. Indeed, previous studies have shown that Pol II co-purifies with other protein complexes, such as the mediator complex and the GTFs (Gustafsson et al., 1997). Some of the protein bands in my gel could belong to these other protein complexes.

Interestingly, two bands around 60 kDa and 39 kDa were diminished in the rapamycin-treated sample (Figure 2.1E, +RAP). One explanation could be that Kin28 depletion interfered with elongation and the bands correspond to elongation factors. Alternatively, since CTD phosphorylation by Kin28 is required to recruit pre-mRNA

processing factors, depletion of Kin28 could cause a decrease in those proteins (Jeronimo et al., 2013).

A more direct approach to confirm if the anchor-away experiment targeted Kin28 depletion was to analyze the Ser5 phosphorylation status of Pol II CTD. The Pol II complex purified from the *KIN28-FRB RPB3-3xFLAG* strain grown in the presence or absence of rapamycin was analyzed by western blot analysis using a Ser5 CTD specific antibody (CTD4H8, Santa Cruz Bio) (Figure 2.1D). As expected, there was a 50% reduction of Ser5 phosphorylation in the rapamycin-treated sample, indicating that the kinase activity of the *KIN28-FRB* cells was only partially inhibited by anchor away (Figure 2.1D). One possibility is that the one-hour rapamycin treatment is not sufficient to completely deplete Kin28 from the nucleus. In fact, Rodríguez-Molina et al., (2016) recently showed that inactivation of Kin28-is with CMK, results in nearly complete dephosphorylation of Ser5 CTD. Alternatively, it is possible that another Ser5 CTD kinase may be present. Bur1, Ctk1, and Srb10 have previously been shown to phosphorylate Ser5 CTD *in vitro* (Jeronimo et al., 2013).

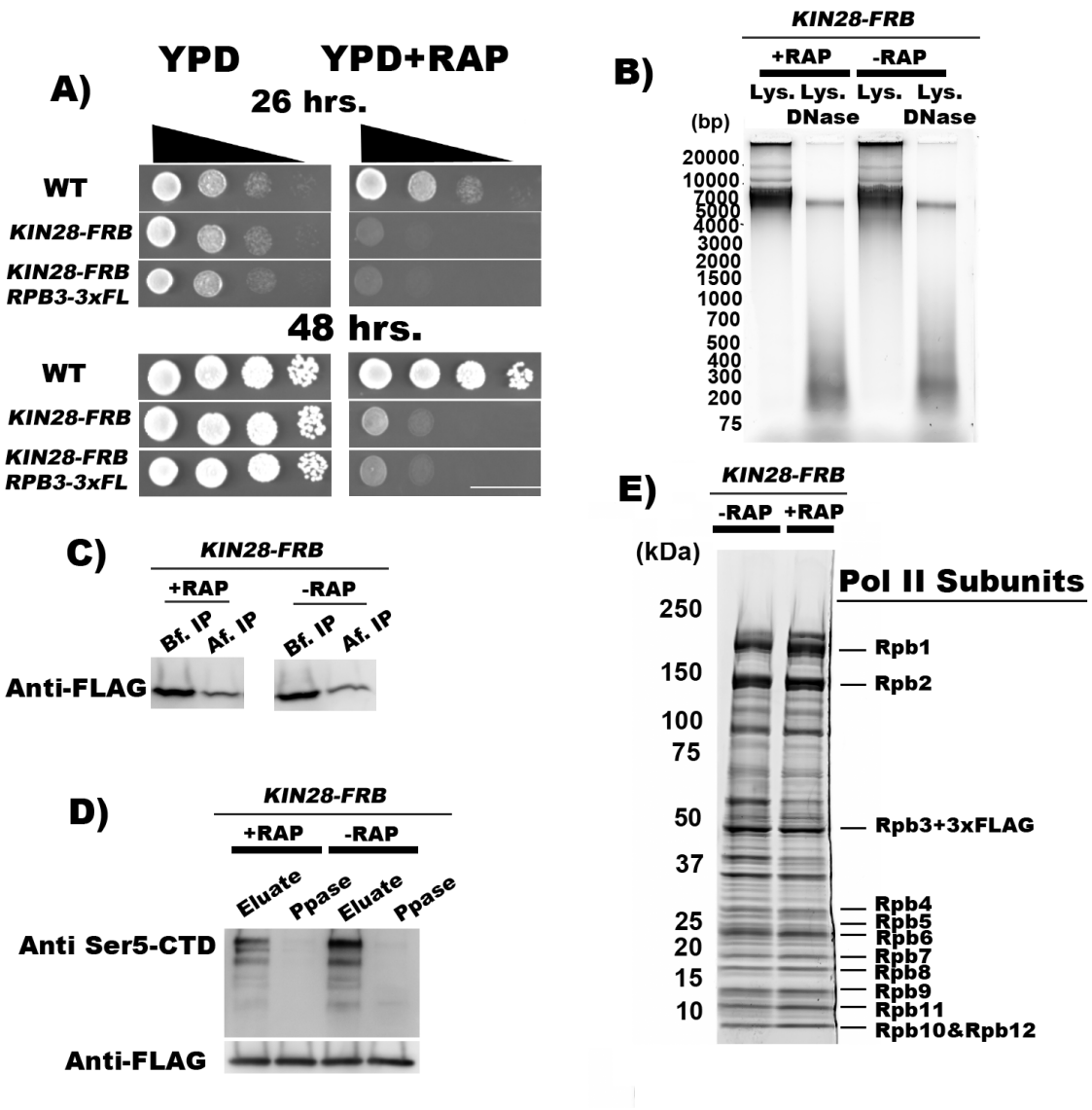


Figure 2.1. Pol II successfully purified by *RPB3-3xFLAG* immunoprecipitation. A) Plate test of strains used in absence or presence of 1 µg/mL of rapamycin. B) Comparison of chromatin before and after DNase digestion of –RAP and +RAP lysates. Lys.: Lysate. C) IP efficiency of –RAP and +RAP samples, Bf. IP: Before IP, Af. IP: After IP. D) Comparison of Ser5 phosphorylation of Rpb1 CTD with and without rapamycin treatment, Ppase = λ-protein phosphatase (NEB) treatment. E) SDS-PAGE gel comparing the eluates of *RPB3-3xFLAG* purification with and without rapamycin treatment.

2. Nascent Pol II small RNAs are mainly tRNA, snoRNA, and degraded mRNA

The function of Kin28 was at least partially inhibited upon rapamycin treatment. As such, the promoter escape step of Pol II could be inhibited as previously shown (Wong et al., 2014). If true, we would expect that nascent Pol II transcripts should be more enriched in the rapamycin-treated sample (Figure 1.1). Previous work by Fazal et al., (2015) indicated that Pol II can transcribe up to ~85 nt of RNA before promoter escape. To see if Kin28 depletion leads to the accumulation of promoter-proximal nascent transcripts, I attempted to isolate Pol II-associated RNA between 70-80 nt for the sequencing analysis. As shown in the denaturing 6% polyacrylamide gel in Figure 2.2B, which was stained with SYBR Gold, RNA species corresponding to 70-80 nt RNA were detected (Figure 2.2B, pink highlight). Additionally, other RNA species were purified that were associated with Pol II. Mainly, Pol II-associated nascent RNA above 300 nt in the 6% polyacrylamide gel can be seen a smear in the 1.2% agarose gel (Figure 2.2A). Furthermore, bands at 121 nt and 156 nt shown on the polyacrylamide gel (Figure 2.2A, left) and bands at 2000 nt and 3800 nt on the 1.2% agarose gel (Figure 2.2A, right) were identified as rRNA and depletion of these rRNA will be elaborated on in the next section. Finally, to isolate the Pol II-associated small RNAs, the band around 70-80 nt was excised, minced, incubated with water, filtered and then ethanol precipitated (Figure 2.2B).

To sequence the isolated Pol II-associated small RNA, adapters were ligated to both ends of the small RNAs to allow priming of cDNA synthesis. First, the pre-adenylated 5’ end of the first adapter was ligated to the 3’-OH of the small RNA using a truncated T4 RNA Ligase (Illumina). The purpose of using a truncated T4 RNA ligase is

to prevent self-ligation or tandem-ligation of the small RNA as this enzyme is unable to activate the 5' phosphate of the small RNA with AMP whereas the 5' end of the adapter is already attached to an AMP. Secondly, the 3'-OH of the second adapter as ligated to the 5' phosphate of the small RNA using a full-length T4 RNA Ligase in the presence of ATP. This ligation reaction adds an AMP group to the 5' phosphate on the small RNA and allow the 3'-OH of the adapter to ligate. However, the 5' end of the nascent Pol II RNAs could potentially be capped with the m^7GpppN 5' structure and for those that are not capped will have a triphosphate. But in both cases, these 5' structures interfere with the ligation of the 5' adapter. To overcome this problem, acid pyrophosphatase (CellScript) was used to remove the 5' cap, gamma, and beta phosphates leaving only an alpha phosphate remaining on the 5' end of nascent Pol II RNA (Figure 2.2B). After PCR amplification with primers complementary to the adapter sequences, cDNAs of the expected size of (~180 bp) were detected but not in the no RNA input control (Figure 2.2 C & D). Since barcodes were included in the adapter sequences, multiplexing (i.e. combining multiple samples into one sequencing run) was performed. Unfortunately, the no rapamycin control was lost because the barcode being used was the reverse complementary to that of another sample in the same sequencing reaction and the data from the two samples coud not be deconvoluted by the bioinformatics pipeline. Therefore, the results shown represent only that of the rapamycin-treated sample.

The sequencing results showed that the aligned cDNA reads have a length distribution of 50 nt to 90 nt which is consistent with the sizes on the excised RNA from the gel (Figure 2.2 B & E). However, many reads were mapped to the entire open

reading frame (ORF) but not to the promoter regions. In addition, highly transcribed genes were overly represented. These results suggest that the small RNAs could simply be degradation products of the mRNA (Figure 2.2F, left panel dark green highlighted). Additionally, many reads represent tRNAs and snoRNAs and therefore could be contaminants of the Pol II immunoprecipitates (Figure 2.2F, right panel light green highlighted & Figure 2.2G). Overall, the nascent small RNA originally thought to be specific to promoters of genes contains a mixture of tRNA, snoRNA, and degraded highly transcribed mRNA.

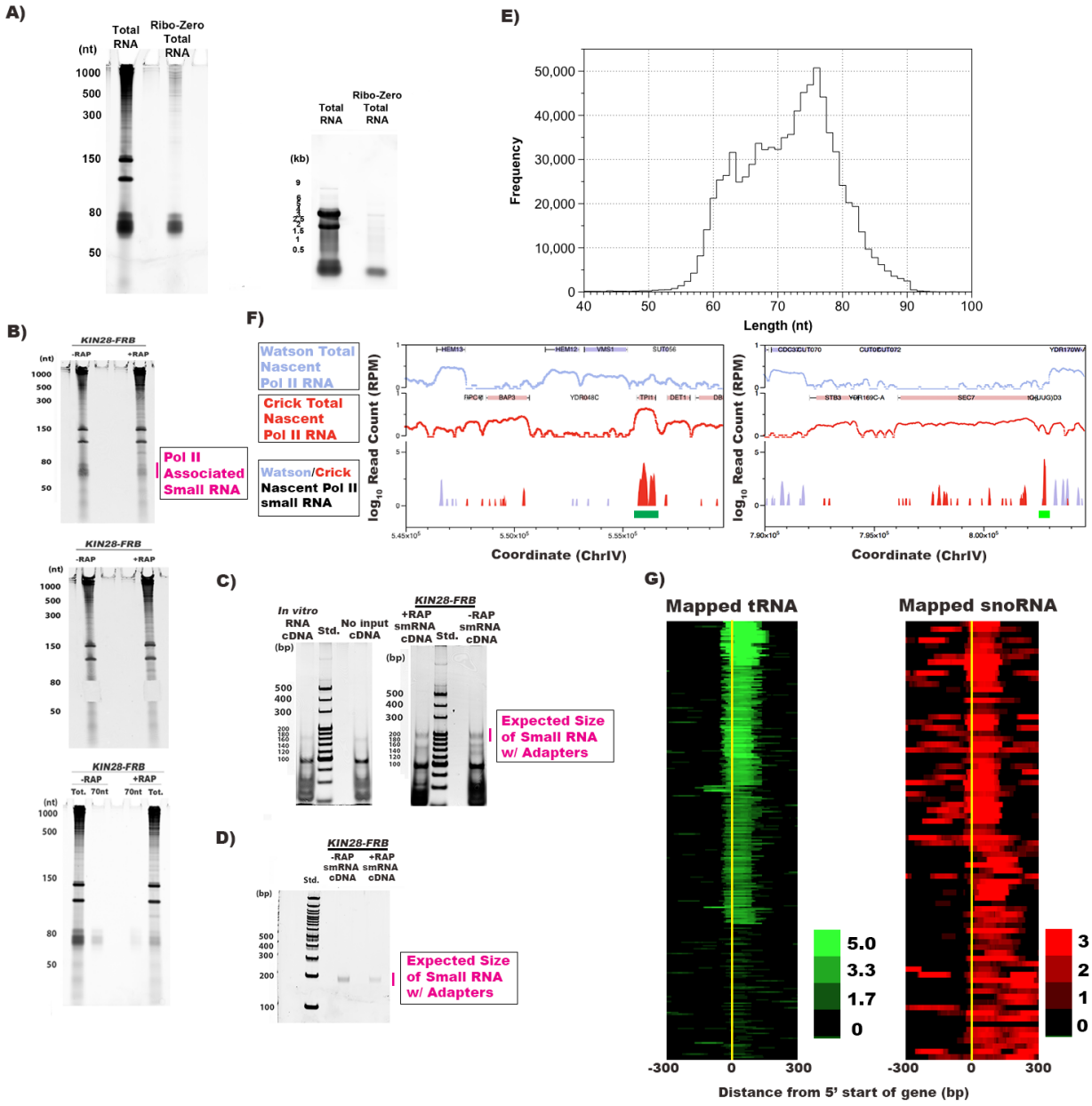


Figure 2.2. Purified nascent Pol II small RNA is a mixture of different RNA species. A) Pol II-associated RNA before and after Ribo-Zero treatment B) Purification of Pol II nascent small RNA by gel extraction. Tot.: Total Pol II-associated RNA, 70: 70 nt purified small RNA C) cDNA libraries generated using the Illumina Small RNA library prep kit, input RNA is noted as: *In vitro*, No RNA input, +RAP small RNA, and –RAP small RNA. D) Gel purified ligated –RAP and +RAP cDNA libraries from small RNA. E) Histogram of read lengths of Pol II-associated small RNA, bins = 1 nt. F) Traces of Pol II-associated small RNA (bottom panel) with separated strand of total nascent Pol II-associated RNA (from Figure 2B), blue represent watson strand traces, red represents crick strand traces. Dark green highlight indicates highly transcribed gene TPI1 and bright green highlights tRNA. G) Heatmaps of tRNA and snoRNA present in Pol II nascent small RNA sample centered at 5' end of annotated gene

3. Kin28 depletion did not cause specific accumulation of nascent RNA at promoters

Since the 70 nt was not enriched for promoter specific RNA species, a more unbiased approach was used to analyze the Pol II-associated RNA species without size selection. However, since rRNA contamination was significant, the rRNA was removed first by incubating the RNA samples with magnetic beads coated with nucleic acids with complementary sequences to rRNA (Illumina, Ribo-Zero kit). After depletion the majority (~90%) of the original Pol II-associated RNA species were discovered to be rRNA (Figure 2.2A). After rRNA depletion, the enriched Pol II-associated RNA samples were processed to obtain cDNA libraries for sequencing (Illumina stranded mRNA library prep kit).

To determine the role of Kin28 depletion on promoter specific transcript accumulation the cDNA libraries of two biological replicates of the Pol II-associated RNA isolated from *KIN28-FRB* cells with and without rapamycin treatment were analyzed by deep sequencing using as Illumina MiSeq instrument. The sequenced reads were aligned to the R64 yeast genome (Ensembl, R64-1-1), the percentage of reads aligned ranged from 68% to 83% (Figure 2.3A). To evaluate the reproducibility of the RNA-seq data, the fragments per kilobase million (FPKM) value of 5,300 protein-coding (pc) genes were compared between the replicates (denoted R1 and R2) (Rodríguez-Molina et al., 2016; Trapnell et al., 2012). Significant differences between the samples with and without rapamycin treatment were defined as values larger than two standard deviations (1.456) of the datapoints between the R1 and R2 of the –RAP

controls. We found that the 5,300 pc-genes measured have similar expression values for the untreated replicates (median = 0.126; σ = 0.728; Figure 2.3C, panel 1). Similarly, the rapamycin treated replicates showed comparable FPKM values (median = -0.165, σ = 0.752, Figure 2.3D). Next, to see if Kin28 depletion caused an effect on global nascent transcription levels the untreated and treated rapamycin samples were compared (Figure 2.3C, panels 2 and 3). The first replicate of –RAP and +RAP showed the –RAP sample had 28.3% of pc-genes that were significantly higher in FPKM value than the +RAP pc-genes (Figure 2.3C, 2nd panel). The second replicate of –RAP and +RAP showed the –RAP sample had 12.9% of pc-genes significantly higher in FPKM value than the +RAP sample. These data suggest that Kin28 depletion has a negative effect on nascent transcript levels consistent with an earlier study, although the defect observed here is less severe (Rodríguez-Molina et al., 2016). Furthermore, based on the single traces of the nascent Pol II-associated RNA-seq data, Kin28 depletion does not appear to cause accumulation of transcripts at the promoters (Figure 2.3D).

A)

Strain	Replicate	Treatment	All Reads	Aligned	% Aligned
KIN28-FRB	1	DMSO	7,508,809	6,282,497	83.67
KIN28-FRB	1	Rapamycin	7,376,415	5,043,667	68.38
KIN28-FRB	2	DMSO	9,974,848	8,090,523	81.11
KIN28-FRB	2	Rapamycin	8,970,404	6,854,914	76.42

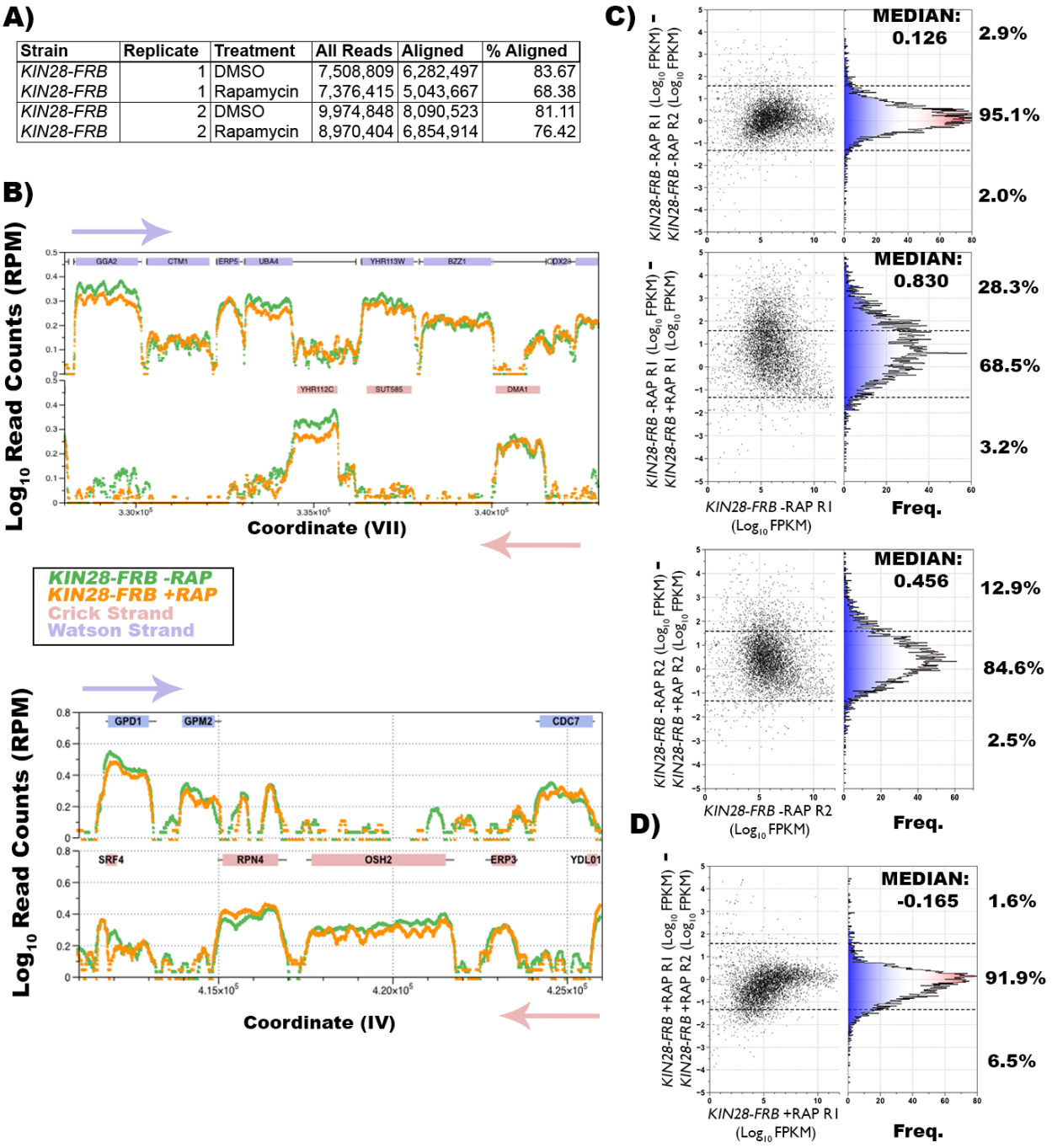


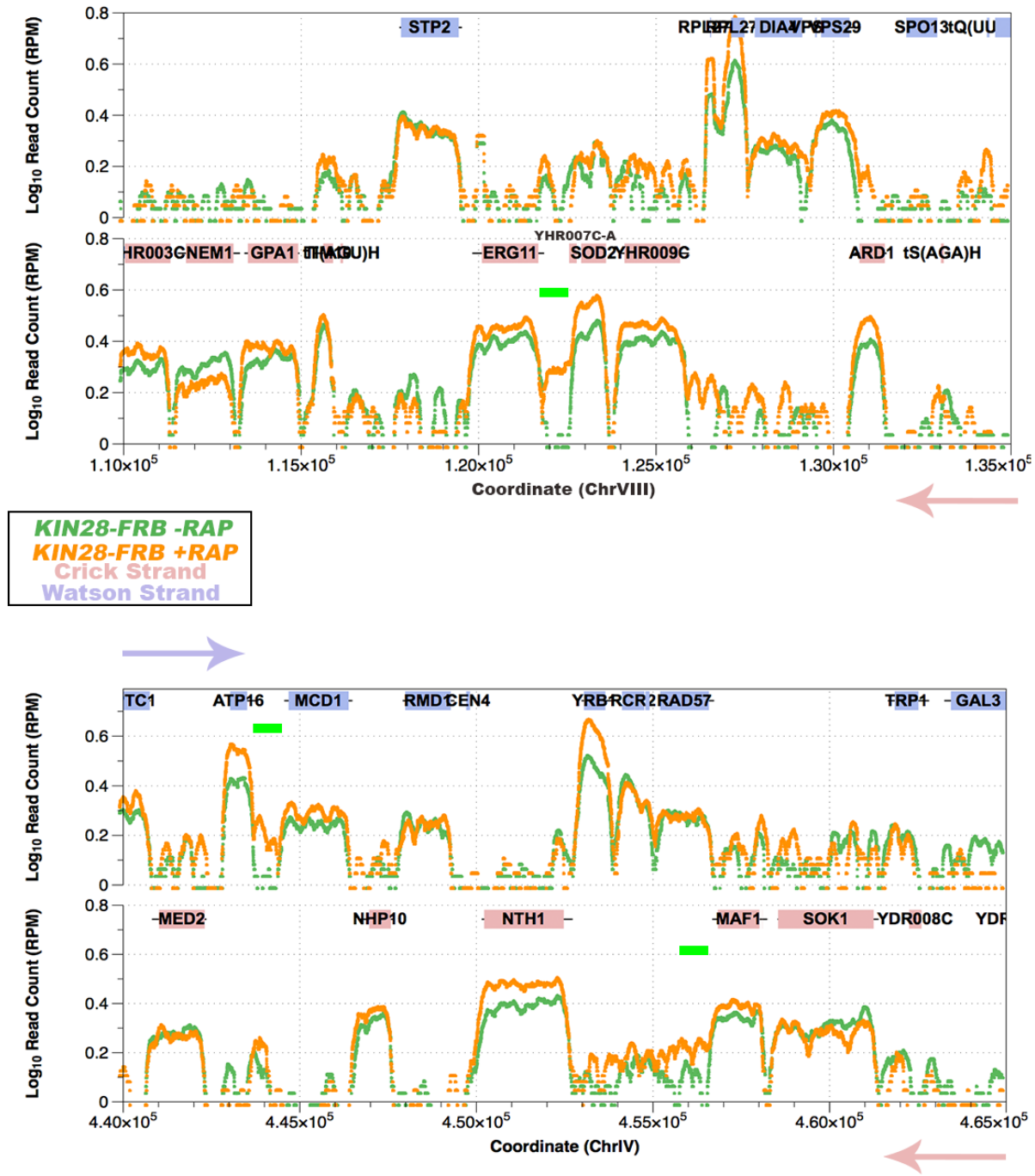
Figure 2.3. Kin28 depletion causes minor decrease in gene expression. A) Read counts for both total nascent Pol II-associated RNA replicates used in this thesis. B) Traces of total nascent Pol II-associated RNA data after RPM normalization, green traces indicate –RAP Rep. 1, orange traces represent +RAP Rep. 1, purple arrows indicate the Watson strands, pink arrows indicate the Crick strands. C) Comparison of

FPKM values of 5,300 protein-coding genes between –RAP R1 and R2 (top panel) and comparison of –RAP and +RAP of both replicates (bottom 2 panels), Black dotted line indicates two standard deviations from the median of the –RAP R1 and –RAP R2 histogram. C) Comparison of FPKM values of 5,300 protein-coding genes between +RAP and +RAP of both replicates.

4. Depletion of Kin28 reveals a transcription termination defect at specific genes

Interestingly, my data showed that Kin28 depletion caused some genes to exhibit a defect in termination (e.g. *SOD2* and *VPS29* Figure 2.4A top panel; and *ATP16 and MAF1* Figure 2.4A, bottom panel), while the majority of genes were unaffected (e.g. *STP2, ARD1* Figure 2.4A, top panel; *MED2, NHT1* Figure 2.4A, bottom panel). This termination defect could be associated with impaired cleavage of the pre-mRNA consistent with the idea that Ser5 phosphorylation of the CTD is required for the recruitment of factors involved in pre-mRNA processing. To evaluate how general this phenotype is, genes with termination defects were manually chosen based on the criterion that the transcript signal extends beyond the transcription termination site (TTS) in the rapamycin-treated sample but not the untreated control (Figure 2.4A). As such 99 genes were identified to exhibit a termination defect upon Kin28 depletion. These genes were aligned at the TTS to allow better visualization of a trend (Jiang and Pugh, 2009; Figure 2.4B). The sum of the read coverage within the regions -1000 to +1000 bp from the TTS in the rapamycin-treated and untreated samples was compared (Figure 2.4C, purple traces). As shown in Figure 2.4C, the rapamycin-treated samples had higher amounts of transcripts between the 0-500 bp region suggesting that Kin28 depletion caused a termination defect in the selected genes.

A)


STP2
SPO13
YHR007C-A
ERG11
SOD2
ARD1
Log10 Read Count (RPM)
Coordinate (ChrVIII)
KIN28-FRB -RAP
KIN28-FRB +RAP
Crick Strand
Watson Strand
MCD1
CEN4
RAD57
TRP1
GAL3
MED2
NHP10
NTH1
MAF1
SOK1
YDR008C
Coordinate (ChrIV)

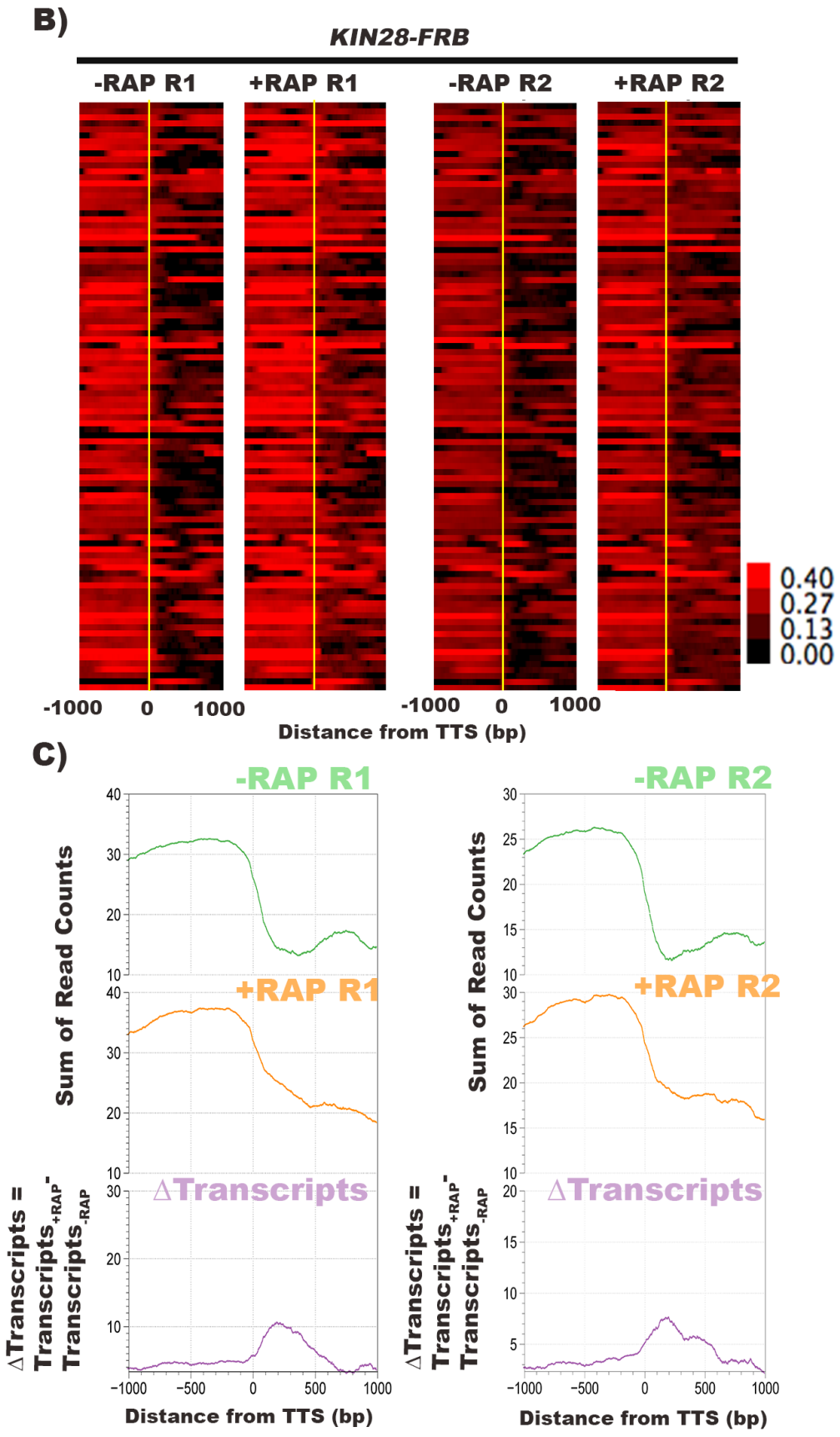

B)
KIN28-FRB
-RAP R1
+RAP R1
-RAP R2
+RAP R2
0.40
0.27
0.13
0.00
-1000
0
1000
-1000
0
1000
Distance from TTS (bp)
C)
-RAP R1
+RAP R1
ΔTranscripts
Sum of Read Counts
ΔTranscripts = Transcripts+RAP - Transcripts-RAP
Distance from TTS (bp)
-RAP R2
+RAP R2
ΔTranscripts
Sum of Read Counts
ΔTranscripts = Transcripts+RAP - Transcripts-RAP
Distance from TTS (bp)

Figure 2.4. Kin28 depletion shows read-through transcripts after the transcription termination site at specific genes. A) Traces of replicate 1 (R1) of total nascent Pol II-associated RNA data that has highlighted in bright green the termination defect observed. B) Heatmaps of 99 genes of –RAP and +RAP replicate 1 (R1) and 2 (R2) observed to have termination defect centered at TTS (Red indicates high read counts, black indicates no read counts). C) Composite of 99 genes showing sharp increase in transcripts past TTS between –RAP and +RAP samples in both replicates, green represents –RAP, orange represents +RAP, and purple is the difference of +RAP and -RAP.

III. Discussion

My thesis project investigated the function of Kin28 during transcription. The original goal was to deplete Kin28 by anchor away to block promoter escape and then evaluate the role of the transcription machinery on +1 H2A.Z nucleosome disassembly. However, a promoter escape defect was not observed. Instead, Kin28 depletion caused a transcription termination defect at many protein-coding genes. In the following section, I discuss the genomic approach that was used to study nascent Pol II RNA and how the methodology can be improved. Second, I discuss why Kin28 depletion did not cause the expected promoter escape defect and provide alternative approaches to obtain more efficient Kin28 inactivation. Third, I speculate on how Kin28 is causing the observed termination defect at protein-coding genes Lastly, I discuss another subunit of TFIIH, Ssl2, and its potential role in the eviction of +1 H2A.Z containing nucleosome.

1. Nascent RNA-seq is a powerful tool to enrich for Pol II-associated transcripts

To determine the transcription distribution before and after Kin28 depletion on transcription, the previously described protocol was adapted and modified for mapping nascent Pol II-associated RNA (Churchman and Weissman, 2011). In contrast to the original protocol, the standard Illumina stranded mRNA kit was used to prepare the cDNA sequencing libraries, circumventing a more arduous library construction procedure in the original protocol (Churchman and Weissman, 2012). Overall, our procedure was equally effective as the Churchman protocol in identifying Pol II transcribed regions of protein coding genes as well as non-coding genes. In addition, unlike the standard RNA-seq procedure, which does not enrich for chromatin-bound Pol II, our nascent Pol II-associated RNA-seq data revealed cryptic RNAs that are normally subjected to degradation by the exosome. Furthermore, our nascent Pol II-associated RNA-seq provided much higher resolution transcript map than microarray data. These data have provided important information that was applied in another study for identifying an upstream SUT (Stable Untranslated Transcript) that was obscuring the bona fide promoter of the *HSE1* gene (Tramantano et al., 2016).

In this study, the original motivation for developing a procedure to capture and map the Pol II-associated RNA was to assess the effect of Kin28 depletion on promoter escape. However, how to normalize the sequencing data to allow a fair comparison of the +RAP and –RAP sample remains unclear. For example, there was only a minor decrease in nascent transcript levels upon Kin28 depletion (Figure 2.3C), in contradiction to a previous study indicating that nascent transcription globally decreased 4-fold after Kin28 inactivation (Rodríguez-Molina et al., 2016). If the small difference

observed in our experiment is due to a normalization artifact, the analysis could be improved by addition of an exogenous DNA spike-in control. For example, *Schizosaccharomyces pombe* spike-in has previously been used in *Saccharomyces cerevisiae* RNA-seq experiments to normalize for cell numbers assuming that *S. pombe* cells lyse with similar efficiency as *S. cerevisiae* (Bonnet et al., 2014).

2. The inability to block promoter escape could be due to incomplete inactivation of Ser5 CTD kinase activities

Normalization artifacts notwithstanding, if Kin28 depletion can indeed block promoter escape, a redistribution of transcripts from the gene bodies to promoters should have been observed. However, the redistribution of transcripts was not seen and is contrary to previous results showing that Pol II occupancy at promoters increases after Kin28 depletion (Wong et al., 2014). One explanation for this discrepancy is that Kin28 depletion was incomplete as indicated by anti-Ser5 CTD gel that showed a substantial level (~50%) of Ser5-phosphorylation remaining after Kin28 depletion (Figure 2.1D). Alternatively, it is possible that Kin28 is not solely responsible for promoter escape In fact, other kinases such as: Bur1, Ctk1, and Srb1 have previously been shown to phosphorylate Pol II Ser5 CTD *in vitro* (Jeronimo et al., 2013). Future experiments that involve simultaneously inactivating multiple Pol II Ser5 CTD kinases could potentially address this issue.

As an alternative to Kin28 anchor away, inactivation of the *kin28-is* allele which has been shown in a previous study to cause almost complete dephosphorylation of Pol II Ser5 CTD (Rodríguez-Molina et al., 2016). Using the *kin28-is* allele also avoids

possible problem involving nuclear depletion of other components of the transcription machinery in response to Kin28 anchor away, as Kin28-FRB could drag other proteins out of the nucleus in response to rapamycin treatment. In fact, Kin28 anchor away showed a decrease of Pol II at the promoters, suggesting that some of the phenotype observed in the *KIN28-FRB* mutant could be caused by the depletion of other proteins of the transcription machinery (Wong et al., 2014). The use of the *kin28-is* mutant could potentially circumvent this problem.

3. The role of Kin28 in transcription termination of protein-coding genes

My data suggested that Kin28 plays a role in transcription termination of protein-coding genes genome-wide consistent with several earlier findings (Medler and Ansari, 2015) First, Kin28 ChIP-chip revealed that Kin28 is enriched not only at the 5' ends of genes but also at the 3' ends of genes, indicating that Kin28 might functionally interacting with factors involved in transcription termination (Mayer et al., 2010). Second, Kin28 phosphorylation of Ser5 on Pol II CTD is required for the recruitment of the 3' mRNA processing protein Pta1, which is part of the CDF complex required for both the allosteric and torpedo termination pathways (Rodriguez et al., 2000; Zhao et al., 1999; Figure 1.1). Third, Kin28 physically interacts with Ssu72, a Ser5 phosphatase involved in gene-looping and transcription termination (Medler and Ansari, 2015; Tan-Wong et al., 2012; Zhang et al., 2012). After Kin28 inhibition ChIP-qPCR showed decreased occupancy of Ssu72 at 5' and 3' end of genes compared to wild-type (Medler and Ansari, 2015). More recently, ChIP-qPCR experiments showed that several protein-coding genes (*CHA1*, *HXT1*, *ACT1*, and *ASC1*) exhibited aberrant transcript signals beyond their termination sites in an ATP analog sensitive *kin28-as* allele after 1-NA-PP1

treatment (Medler and Ansari, 2015). Our data provide further genome-wide evidence by showing that at least 100 genes exhibited a similar termination defect, supporting the idea that Kin28 may play a more general role in pre-mRNA processing during transcriptional termination.

The termination defect observed in my Kin28 depletion experiment raises the question of why a different study that also inactivated Kin28 and looked at nascent transcripts did not see similar results (Rodríguez-Molina et al., 2016). Rodriguez-Molina et al., (2016) purified nascent transcripts by metabolically labeling nascent RNA with thiouracil followed by biotinylation and streptavidin pulldown. However, this approach to purify nascent RNA enriches for the soluable RNA including mature mRNA that are not associated with Pol II. Therefore, the Pol II associated "read-through" transcripts that accumulate beyond the TTS we observed in my data should be underrepresented in their data.

4. Future experiments to understand TFIIH role in +1 nucleosome eviction

The original hypothesis was that during promoter escape Pol II drives +1 nucleosome disassembly. However, an alternative possibility is that +1 nucleosome disassembly occurs before promoter escape. Specifically, the torsional stress that occurs during promoter melting could be the mechanism of +1 nucleosome disassembly. Promoter melting is mediated by another component of the TFIIH complex, namely Ssl2, which is a DNA translocase (Fishburn et al., 2015). During promoter melting, the 5'-to-3' translocase activity of Ssl2 acts on the non-template strand of DNA, thereby feeding downstream DNA of the PIC into the active cleft of Pol II

and simultaneously untwisting the DNA duplex towards the DNA bend restricted by TBP (Fishburn et al., 2015). However, Ssl2 translocase activity will causes overwinding of the DNA downstream of the +1 nucleosome. Since the DNA wraps around a nucleosome in a left-handed turn (i.e. negative supercoil) the overwinding will destabilize the nucleosome structure. Previous studies have shown that Ssl2 is the main protein required for transcription bubble formation *in vitro* (Fishburn et al., 2015). Interestingly, since Ssl2 has limited movement due to being bound to the PIC it suggests that as Ssl2 tracks along the coiled sugar-phosphate backbone it will simultaneously cause the DNA to untwist leading to the formation of the transcription bubble formation. Thus, the +1 nucleosome downstream of the PIC and transcription bubble formation may experience torsional stress resulting from Ssl2 translocase activity. It has been shown that nucleosomes dissociate when DNA is being "unzipped" and H2A.Z-containing nucleosomes require less force and are more mobile relative to their H2A counterparts (Rudnizky et al., 2016). An outstanding question is whether the stress induced by Ssl2 translocase activity is enough to dissociate +1 nucleosomes containing H2A.Z? There are a few ways to approach this question through *in vivo* and *in vitro* experiments.

To address the role of Ssl2 in +1 H2A.Z nucleosome disassembly, we aim to use a protein-swap strategy could be used to replace the wild-type Ssl2 with an ATPase mutant form of Ssl2. Briefly, this approach involves anchoring away the endogenous wild-type Ssl2 while simultaneously having an *ssl2* ATPase mutant constitutively expressed in a plasmid (Jeronimo and Robert, 2014). As such, after rapamycin treatment, the mutant form of Ssl2 will at the promoters and it can therefore be

determine whether inactivation of Ssl2 causes an increase in H2A.Z levels by ChIP-seq. Second, another approach to monitor the putative H2A.Z eviction function of Ssl2 consists of using an *in vitro* assay with purified components of the transcription machinery and histones assembled on DNA. This approach involves reconstituting minimal transcription components for *in vitro* transcription which requires: TBP, TFIIB, TFIID, TFIIE, TFIIF, Pol II, and TFIIH (Parvin et al., 1992). In this case, mutant and wild-type Ssl2 proteins can be used to compare the nucleosome eviction efficiency by electro-mobility shift assay. As a whole, understanding the transcription machinery's specific chromatin remodeling activity will further our knowledge of a key regulatory step of H2A.Z eviction during transcription initiation in eukaryotes.

IV. Materials and Methods

1. Strain construction and growth conditions

The yeast strains used in this study containing the anchor away tags were constructed as previously described (Haruki et al., 2008; Tramantano et al., 2016). A previously made *KIN28-FRB* strain was used and tagged with 3xFLAG at the 3' end of *RPB3* by transforming with a 3xFLAG PCR product by homologous recombination and selected using *kanMX*. Successful integration of the 3xFLAG tag was confirmed by PCR and Sanger sequencing.

For Pol II purification and subsequent nascent Pol II RNA purification, 6 L of *KIN28-FRB RPB3-3xFLAG* cells were grown at 30ºC to 1.5 OD_{600} followed by addition

of rapamycin to a final concentration of 1 µg/mL or equivalent volume of DMSO for one hour after which cells were harvested.

2. Purification of RNA Polymerase II by immunoprecipitation of RPB3-3xFLAG

After the 1-hour rapamycin treatment, the cells were harvested by centrifuging at 3500 rpm at 4ºC for 5 minutes (Beckman J20-I). Cells were washed once with H_2O and once with a buffer containing 40 mM HEPES-KOH (pH 7.6), 40% Glycerol, 2 mM EDTA (Sigma-Aldrich), and 100 mM KOAc. Cells were lysed under LN_2 by freezer mill and re-suspended in lysis buffer containing 20 mM HEPES-KOH (pH 7.6), 20% Glycerol, 1 mM EDTA (Sigma-Aldrich), 110 M KOAc, 10mM $MnCl_2$, 2 mM DTT (Bio-Rad), 0.0085 g PMSF (Sigma-Aldrich), 0.0165 g benzamidine (Sigma-Aldrich), 0.0685 mg pepstatin A (Roche), 0.0145 mg leupeptin (Roche), and 0.01 mg chymostatin (Roche). The following were added to the re-suspended lysate and incubated at 4ºC for 30 minutes: 50 U/mL DNase I (Worthington, DPRF), 50 U/mL SUPERase.In. (Ambion). After DNase digestion, lysates were centrifuged at 20,000 rpm for 30 minutes at 4ºC. The cleared supernatant was stored at -80ºC until read to perform anti-FLAG affinity chromatography.

Eight-hundred µL slurry volume of anti-FLAG agarose beads (Sigma-Aldrich) were incubated with 12 mL of cleared supernatant (~1 L culture equivalent) and incubated for 2.5 hours at 4ºC with gentle mixing. After 2.5 hr of binding the anti-FLAG beads to *Rpb3-3xFLAG*, beads were washed 5 times with a wash buffer containing 20 mM HEPES-KOH (pH 7.6), 10% glycerol, 110 mM KOAc, 0.1% NP-40, 0.0085 g PMSF (Sigma-Aldrich), 0.0165 g benzamidine (Sigma-Aldrich), 0.0685 mg pepstatin A

(Roche), 0.0145 mg leupeptin (Roche), and 0.01 mg chymostatin (Roche). Afterwards, the beads (~400 µL) were incubated in 400 µL of wash buffer containing 200 µg of anti-FLAG peptide for 2 hours at 4 ºC (Sigma Aldrich). The *Rpb3-3xFLAG* enriched eluate was separated from the beads by centrifuging for 2 minutes at 12,000 g in an 800 µL 0.5 µM centrifuge column (Millipore).

3. DNase I Digestion, Western Blot, and SDS-PAGE Analyses

To digest chromatin, re-suspended lysates were incubated with 20 U/mL DNase I in the presence of 10 mM $MnCl_2$ (Worthington, DPRF) for 30 minutes at 4 ºC. To confirm the DNase I digestion of chromatin was successful, the DNase I digested chromatin was deproteinized 400 µg proteinase K (Thermo) at 55 ºC for 2 hours followed by standard phenol-chloroform extraction and ethanol-sodium acetate precipitation. After DNA digestion, the DNA was analyzed by electrophoresis on a 1.3% agarose, 0.5xTBE gel stained with Sybr Green I (Invitrogen).

To evaluate IP efficiency, the input and the supernatant after IP were analyzed by anti-FLAG (Sigma-Aldrich) western to detect the Rpb3-3xFLAG. To compare the Ser5-CTD phosphorylation status of Pol II, the purified eluate and lambda-phosphatase treated purified eluates were probed with anti-Ser5-CTD (Santa Cruz Biotech) and anti-FLAG (Sigma-Aldrich) followed by a standard western blot procedure. Additionally, purified eluates were visualized on Novex 8-16% Tris-Glycine gradient gel (Life Technologies) by SDS-PAGE and stained with SyPRO ruby (Invitrogen).

4. Purification of nascent RNA from purified RNA Polymerase II and cDNA library construction

Qaigen mRNeasy mini kit was used to separate the proteins and FLAG peptide from the Pol II-associated RNA. The purified Pol II-associated RNA was analyzed by 1X TBE 6% polyacrylamide 8M-urea gel and stained with Sybr gold (Invitrogen). RNA concentration was determined by A260 spectrometric reading and the RiboGreen RNA assay (Invitrogen). Gel extraction was used to purify small RNA by cutting out the associated band around 70 nt. The gel pieces ware minced using gel breaker tubes (Fisher) and then incubated with RNase-free H_2O overnight at 4 ºC. The mixture of gel, H_2O, and RNA was passed through a 5 µM filter tube (Fisher) to obtain Pol II-associated small RNA enriched H_2O this was followed by ethanol precipitation to concentrate the Pol II-associated small RNA further.

For library construction of Pol II-associated RNA, Illumina Ribo-Zero Gold Yeast Kit was used to deplete rRNAs and Illumina stranded mRNA library Prep kit was used to prepare cDNA libraries of the Pol II-associated RNA. For the preparation of the cDNA of the small RNA, the Illumina Small RNA library prep kit was then used with the few following modifications. First, an acid pyrophosphatase (CellScript, C-CC15011H) was used after the 3' ligation of the adapters for 1 hr to modify the small RNA to accept 5' adapter ligation. Secondly, to purify the cDNA library after PCR amplification, the cDNA were separated on 6% native polyacrylamide gel and stained using Sybr Green I (Invitrogen) followed by overlaying the gel with cDNA on a gel image to cut out the appropriate sized bands. After library preparation, an Illumina MiSeq sequencing instrument was used to conduct paired-end sequencing (with 33 bp read length).

5. Bioinformatics analyses for nascent Pol II small RNA and nascent Pol II RNA

A workflow detailing the primary files obtained during bioinformatics analyses is provided in Figure 4.1. Briefly, fastq files obtained from Illumina Miseq instrument were trimmed of Illumina adapters and reads under 15 bp were discarded (Trimmomatic v 0.36). Fastq files were aligned to the most recent yeast genome build (Ensembl, 64-1-1) using STAR (v2.5.2a) in the default parameters. To obtain SGR files, genomecoverageBED was used with the following parameters for nascent Pol II-associated RNA: -d, -fs 200 and for Pol II-associated small RNA: -d –fs 100 (Bedtools v2.26.0). Afterwards, reads were scaled by $\log_{10}$ and RPM (reads per million) were plotted using python scripts. The final scaled SGR files were visualized using Datagraph (v4.1). Cufflinks (v2.1.1) was used on the STAR aligned BAM file to obtain FPKM (Fragment per kilobase million) values. The list of 5,300 protein-coding genes to compare FPKM values between replicates in Figure 2.3C was obtained by Rodríguez-Molina et al., (2016). The following parameters were used with cufflinks: --library-type fr-firststrand, --compatible-hits-norm. FPKM value scatterplots were plotted using Datagraph (v4.1). Lastly, to compare post-TTS transcripts, 99 identified genes from respective RPM Log_{10} SGR files were compiled then converted into a BED file then centered at TTS using windowBED (Jiang and Pugh, 2009; Bedtools v2.26.0). The same exact workflow was done to generate heatmap for 3200 genes for comparison. Custom python scripts were used to organize the BED file into a histogram and subsequently a matrix that could be plotted as a heatmap using Java TreeView (v1.1.6r4).

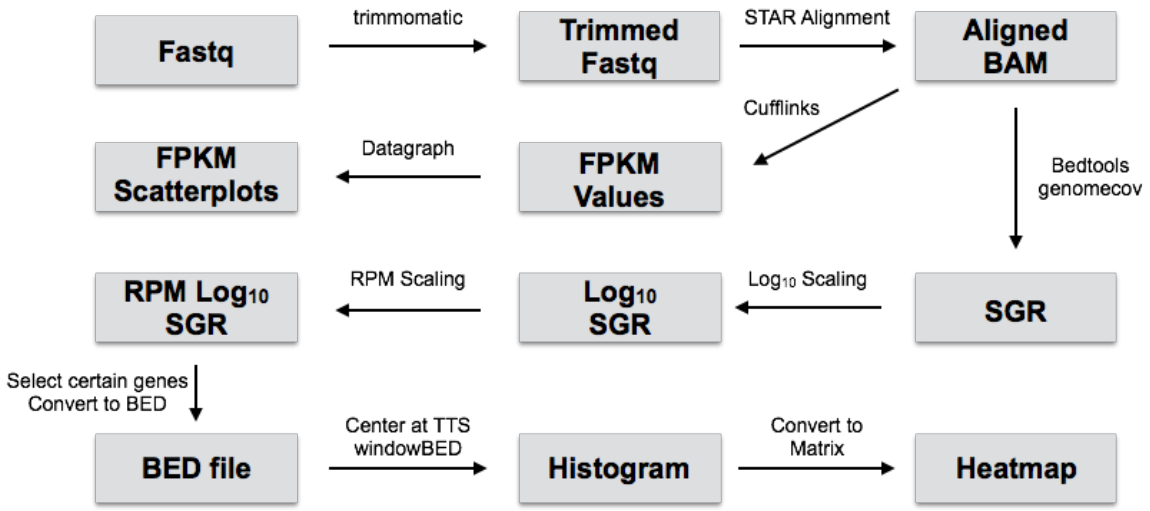


Figure 4.1. Bioinformatics workflow. How FPKM, SGR, and Matrices used for scatterplots, RNA traces and, heatmaps were obtained.

V. References

Bonnet, J., Wang, C.Y., Baptista, T., Vincent, S.D., Hsiao, W.C., Stierle, M., Kao, C.F., Tora, L., and Devys, D. (2014). The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. Genes Dev. 28, 1999–2012.

Bowman, E. a, and Kelly, W.G. (2014). RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation. Nucleus 5, 224–236.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469, 368–373.

Churchman, L.S., and Weissman, J.S. (2012). Native elongating transcript sequencing (NET-seq). Curr. Protoc. Mol. Biol. 1, 1–17.

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

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

Dhillon, N., Oki, M., Szyjka, S.J., Aparicio, O.M., and Kamakaka, R.T. (2006). H2A.Z functions to regulate progression through the cell cycle Mol. Cell. Biol. 26, 489–501.

Fazal, F.M., Meng, C.A., Murakami, K., Kornberg, R.D., and Block, S.M. (2015). Real-time observation of the initiation of RNA polymerase II transcription. Nature 525, 274–277.

Feaver, W.J., Svejstrup, J.Q., Bardwell, L., Bardwell, A.J., Buratowski, S., Gulyas, K.D., Donahue, T.F., Friedberg, E.C., and Kornberg, R.D. (1993). Dual roles of a multiprotein complex from *S. cerevisiae* in transcription and DNA repair. Cell 75, 1379–1387.

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

Fishburn, J., Tomko, E., Galburt, E., and Hahn, S. (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. Proc. Natl. Acad. Sci. 112, 3961–3966.

Gibbons, B.J., Brignole, E.J., Azubel, M., Murakami, K., Voss, N.R., Bushnell, D.A., Asturias, F.J., and Kornberg, R.D. (2012). Subunit architecture of general transcription factor TFIIH. Proc. Natl. Acad. Sci. 109, 1949–1954.

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

Gustafsson, C.M., Myers, L.C., Li, Y., Redd, M.J., Lui, M., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (1997). Identification of Rox3 as a component of mediator and RNA polymerase II holoenzyme. J. Biol. Chem. 272, 48–50.

Halley, J.E., Kaplan, T., Wang, A.Y., Kobor, M.S., and Rine, J. (2010). Roles for H2A.Z and its acetylation in *GAL1* transcription and gene induction, but not *GAL1*-transcriptional memory. PLoS Biol. 8.

Han, M., and Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters in vivo. Cell 55, 1137–1145.

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol. Cell 31, 925–932.

Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. Nat. Struct. Mol. Biol. 21, 449–455.

Jeronimo, C., Bataille, A.R., and Robert, F. (2013). The writers, readers, and functions of the RNA polymerase II C-terminal domain code. Chem. Rev. 113, 8491–8522.

Jiang, C., and Pugh, B.F. (2009). A compiled and systematic reference map of nucleosome positions across the *Saccharomyces cerevisiae* genome. Genome Biol. 10, R109.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev. 21, 1519–1529.

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., and Korber, P. (2016). Genomic nucleosome organization reconstituted with pure proteins. Cell 167, 709–721.e12.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional interaction of yeast pre-mRNA 3’ end processing factors with RNA polymerase II. Mol. Cell 9, 1101–1111.

Logan, J., Falck-Pedersen, E., Darnell, J.E., and Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse β^{maj}-globin gene. Proc. Natl. Acad. Sci. 84, 8306–8310.

Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell 49, 203–210.

Luk, E., Ranjan, A., FitzGerald, P.C., Mizuguchi, G., Huang, Y., Wei, D., and Wu, C. (2010). Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. Cell 143, 725–736.

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

Medler, S., and Ansari, A. (2015). Gene looping facilitates TFIIH kinase-mediated termination of transcription. Sci. Rep. 5, 12586.

Murakami, K., Gibbons, B.J., Davis, R.E., Nagai, S., Liu, X., Robinson, P.J.J., Wu, T., Kaplan, C.D., and Kornberg, R.D. (2012). Tfb6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Ssl2 helicase after transcription initiation. Proc. Natl. Acad. Sci. 109, 4816–4821.

Murakami, K., Mattei, P.J., Davis, R.E., Jin, H., Kaplan, C.D., and Kornberg, R.D. (2015). Uncoupling promoter opening from start-site scanning. Mol. Cell 59, 133–138.

Nguyen, V.Q., Ranjan, A., Stengel, F., Wei, D., Aebersold, R., Wu, C., and Leschziner, A.E. (2013). Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1. Cell 154, 1220–1231.

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

Pamblanco, M., Poveda, A., Sendra, R., Rodríguez-Navarro, S., Pérez-Ortín, J.E., and Tordera, V. (2001). Bromodomain factor 1 (Bdf1) protein interacts with histones. FEBS Lett. 496, 31–35.

Parvin, J.D., Timmers, H.T.M., and Sharp, P.A. (1992). Promoter specificity of basal transcription factors. Cell 68, 1135–1144.

Placek, B.J., Harrison, L.N., Villers, B.M., and Gloss, L.M. (2005). The H2A.Z/H2B dimer is unstable compared to the dimer containing the major H2A isoform. Protein Sci. 14, 514–522.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nat. Rev. Mol. Cell Biol. 16, 190–202.

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

Ranjan, A., Mizuguchi, G., Fitzgerald, P.C., Wei, D., Wang, F., Huang, Y., Luk, E., Woodcock, C.L., and Wu, C. (2013). Nucleosome-free region dominates histone

acetylation in targeting SWR1 to promoters for H2A.Z replacement. Cell 154, 1232–1245.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature 483, 295–301.

Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.-J., Burlingame, A.L., Kornberg, R.D., Allison, L.A., Ingles, C.J., Asturias, F.J., et al. (2016). Structure of a complete mediator-RNA polymerase II pre-initiation complex. Cell 166, 1411–1422.e16.

Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L., and Buratowski, S. (2000). Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. Mol. Cell. Biol. 20, 104–112.

Rodríguez-Molina, J.B., Tseng, S.C., Simonett, S.P., Taunton, J., and Ansari, A.Z. (2016). Engineered covalent inactivation of TFIIH-kinase reveals an elongation checkpoint and results in widespread mRNA stabilization. Mol. Cell 433–444.

Rondón, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. Mol. Cell 36, 88–98.

Rosonina, E., Kaneko, S., and Manley, J.L. (2006). Terminating the transcript: breaking up is hard to do. Genes Dev. 20, 1050–1056.

Rudnizky, S., Bavly, A., Malik, O., Pnueli, L., Melamed, P., and Kaplan, A. (2016). H2A.Z controls the stability and mobility of nucleosomes to regulate expression of the LH genes. Nat. Commun. 7, 12958.

Sayre, M.H., Tschochner, H., and Kornberg, R.D. (1992). Reconstitution of transcription with five purified initiation factors and RNA polymerase II from *Saccharomyces cerevisiae*. J. Biol. Chem. 267, 23376–23382.

Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-wide mapping of yeast RNA polymerase II termination. PLoS Genet. 10.

Svejstrup, J.Q., Vichi, P., and Egly, J.M. (1996). The multiple roles of transcription factor/repair factor TFIIH. Trends Biochem. Sci. 21, 346–350.

Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M., and Proudfoot, N.J. (2012). Gene loops enhance transcriptional directionality. Science 338, 671–675.

Tramantano, M., Sun, L., Au, C., Labuz, D., Liu, Z., Chou, M., Shen, C., and Luk, E. (2016). Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. Elife 5.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578.

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

Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Mol. Cell 54, 601–612.

Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodríguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. (2012). Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. J. Biol. Chem. 287, 8541–8551.

Zhang, H., Roberts, D.N., and Cairns, B.R. (2005). Genome-wide dynamics of *HTZ1*, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. Cell 123, 219–231.

Zhao, J., Kessler, M., Helmling, S., O'Connor, J.P., and Moore, C. (1999). Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor. Mol. Cell. Biol. 19, 7733–7740.