FACTORS CONTROLLING PROMOTER-PROXIMAL PAUSING BY RNA POLYMERASE II

A Dissertation
Presented to the Faculty of the graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Nicholas James Fuda
August 2012

FACTORS CONTROLLING PROMOTER-PROXIMAL PAUSING BY RNA POLYMERASE II

Nicholas James Fuda
Cornell University 2012

Most gene expression is regulated at the level of transcription, and the transition from initiation to productive elongation is a key point of regulation. This transition is accompanied by pausing of transcriptionally engaged polymerase in the promoter-proximal region of several heat shock genes. Although this mechanism of regulation was long thought to be limited to a few genes, recent evidence has indicated that pausing is wide-spread in higher eukaryotes. Therefore, it is increasingly important to understand the mechanisms controlling the paused polymerase. I have investigated how the site of pausing on *Hsp70* is specified using high-resolution mapping of polymerase on reporter genes with shifted pausing site sequences. The results indicate that the downstream sequence dictates pause position and the overall level of pausing. I have also used RNAi knock-down in Drosophila cell culture to study the roles of several factors in establishing, maintaining, and releasing the paused polymerase. These experiments have shown GAGA factor is required for pausing on many of its target genes, and the knock-down effects indicate it is involved in establishing the pause. In contrast, Spt5, a protein previously shown to enhance pausing in vitro, reduces pausing genome-wide by increasing levels of elongating polymerase. Two kinases, P-TEFb and CDK12, function in productive elongation. Previously our lab showed that P-TEFb inhibition prevented the transition into elongation, limiting the polymerase to the 5' end of the heat shock-induced *Hsp70* gene. I mapped these polymerases in high resolution to show they occupied sites further downstream than the normal pause sites, suggesting P-TEFb activity may not solely release the

paused polymerase. I also determined the localization of CDK12 on active genes. Its localization downstream of P-TEFb suggests that these kinases may have distinct functions.

Finally, I have examined the role of Fcp1 in *Hsp70* transcription. Our lab previously showed the CTD phosphatase Fcp1 was required for optimum expression of *Hsp70* mRNA. Fcp1 knock-down reduced the heat shock levels of Pol II and increased phosphorylation of non-chromatin bound Pol II, indicating that Fcp1 recycling of RNA polymerase II to an initiation-competent form is required for optimal *Hsp70* heat shock transcription.

BIOGRAPHICAL SKETCH

Nick grew up in Fayetteville, NY, a suburb of Syracuse. Throughout his life, Nick has always wanted to know how things work, and this led him to an interest in science. His first educational memory is of this mother using a globe and table lamp to show him that the day/night pattern and seasons occurred because earth rotated and orbited the sun. After graduating from Fayetteville-Manlius high school in 1989, Nick went to Syracuse University to study biology. He received his B.S. in biology in 1993, and was hired as a lab technician by Dr. David T. Sullivan. This experience reassured him that science was the career for him, and several years later, David retired and urged Nick to pursue a Ph.D. After visits to several schools, Nick decided Cornell fit the best, and in 2003, he finally left the security of Syracuse for the Ithaca unknown. The warm and friendly people quickly made Ithaca feel like home.

To my family and friends

ACKNOWLEDGEMENTS

I would like to thank all of the people who have supported and encouraged my interest in science. This interest was started by my parents, who allowed me to pursue my interests and always told me I could accomplish anything I wanted to do.

I also want all the members of the Lis lab to know how much I appreciate their support and making the long hours in the lab not just bearable, but fun. Of course, this starts with our leader and my adviser, John Lis. His passion for science is obvious and contagious. He is always looking forward toward new methods and experiments, and he has always encouraged my ideas about new ways to explore my research and has freely given his time to help solve problems with experiments. These attributes are reflected in his lab members as well. They are always willing to gives ideas and answer question. This is evident from this dissertation, as I have collaborated with many members of our lab on research.

I would also like to thank my other committee members, Jeff Roberts, Lee Kraus, and Jeff Pleiss. They have always given helpful advice and insightful comments.

Finally, I would like to thank my friends outside of lab for making my time in Ithaca enjoyable and rewarding.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION[1]

The genetic information encoded in the DNA of eukaryotic genes is transcribed into RNA by large molecular machines called RNA polymerases. One of these machines, RNA polymerase II (Pol II), transcribes all the protein-coding genes. The control of Pol II activity is highly modulated at individual genes, and this specific regulation is critical for both the homeostasis of cells and the programmed development of multicellular organisms. The execution of this regulation is dictated by combinatorial molecular interactions of transcription factors with each other and with specific DNA sequences at each gene. They modulate different points in transcription to regulate the timing and level of gene expression.

1.1 The transcription cycle

Transcription can be divided into three general stages: initiation, elongation, and termination. Pol II needs to reproducibly complete several steps in each stage to ensure proper RNA production. Many transcription factors, mainly proteins but also a growing number of RNAs, allow Pol II to complete these steps. They enable Pol II to gain access to the gene's promoter, initiate RNA synthesis at the transcription start site (TSS) of the gene, generate a highly processive elongating transcription complex, and terminate transcription after production of a full-length RNA transcript.

[1] Parts of this chapter, including all of the figures, have been published in Fuda et al. 2009

The largest subunit (Rpb1) of Pol II has long C-terminal domain (CTD) composed of tandem repeats with the consensus sequence YSPTSPS (Prelich, 2002). Various residues within these repeats are targets of post-translational modification (Egloff & Murphy, 2008). These modifications affect its conformation and allow many transcription factors to interact with Pol II (Egloff & Murphy, 2008). Phosphorylation of the CTD dramatically changes as transcription progresses. During initiation, the CTD begins hypophosphorylated (Pol IIa), and is phosphorylated on the Serine 5 and Serine 7 residues of its repeats (Chapman et al., 2007; Egloff & Murphy, 2008; Glover-Cutter et al., 2009). As Pol II enters elongation it is phosphorylated on the Serine 2 residues, which results in a hyperphosphorylated Pol II (Pol IIo) (Hirose & Ohkuma, 2007; Payne, Laybourn, & Dahmus, 1989). The phosphorylation plays a prominent role in the transcription cycle (Hirose & Ohkuma, 2007).

Initiation of transcription requires Pol II to gain access to the TSS and insert the DNA into its active site. These steps are mediated through interactions of specific elements near the TSS (core promoter) with factors, called general transcription factors (GTFs). Eukaryotic DNA is packaged and condensed in nucleosomes, 147bp of DNA around a histone octamer. Nucleosomes can block GTFs and Pol II from accessing the DNA, and need to be removed before Pol II can bind (Saunders, Core, & Lis, 2006). This chromatin opening (Figure 1.1, step 1) can be mediated by several ATP-dependent complexes called nucleosome-remodelers (Bowman, 2010). They use a variety of mechanisms to either shift the nucleosome on or remove the nucleosome from the DNA (Bowman, 2010). Once the core promoter is accessible, the various elements of the core promoter target the assembly of preinitiation complexes (PICs) composed of the

GTFs and Pol II (Figure 1.1, step 2) (Smale & Kadonaga, 2003). Several subunits of the GTF TFIID interact with specific elements. TATA binding protein (TBP) interacts with the TATA box located about 30-35bp upstream of the TSS, the Initiator element (Inr) located at the TSS interacts with the TAF1 and 2 subunits of TFIID, and the motif ten (MTE) and downstream promoter element (DPE) located downstream of the TSS are believed to interact with TAF1, TAF6, and TAF9 (Burke & Kadonaga, 1997; Chalkley & Verrijzer, 1999; D.-H. Lee et al., 2005; C.-H. Wu et al., 2001). The GTF TFIIB interacts with BRE element located just upstream of the TATA box (Smale & Kadonaga, 2003). These interactions between elements and GTFs and GTFs and Pol II recruit Pol II to the core promoter forming the PIC, and define the TSS (Flores et al., 1991).

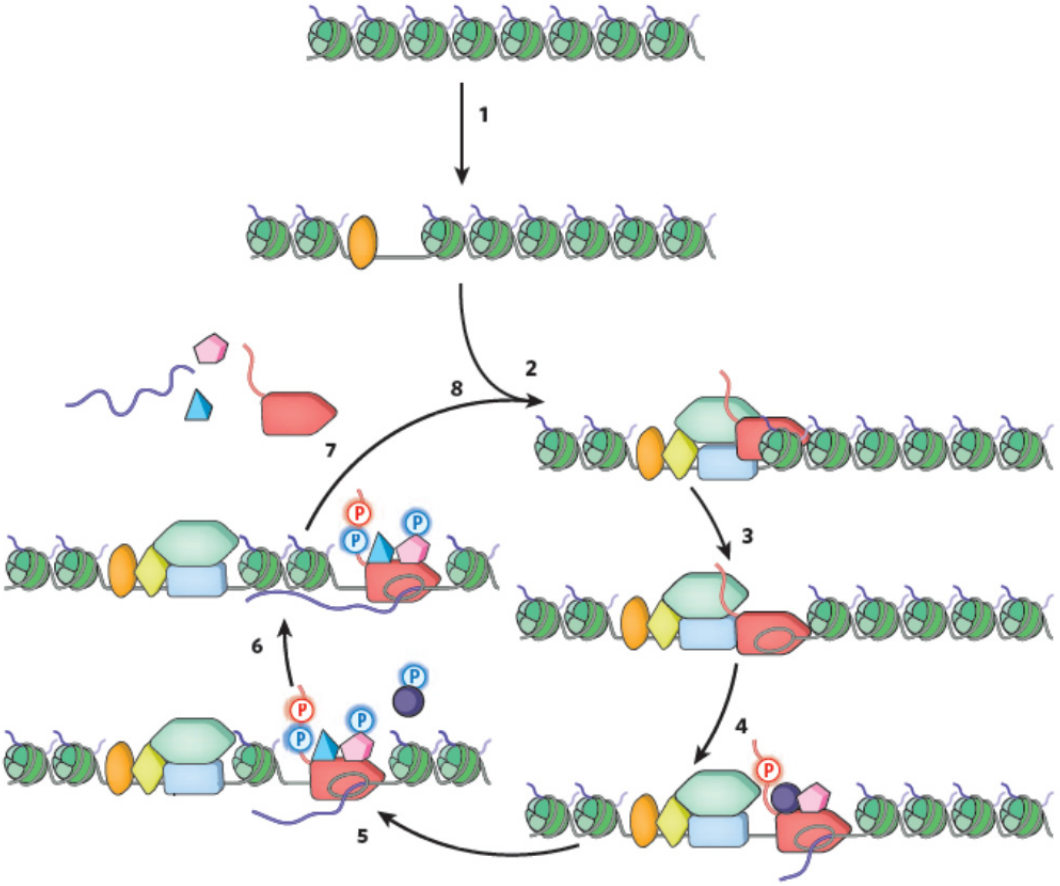


Figure 1.1: The transcription cycle is a multistep process. Step 1: chromatin opening. The repressed gene and regulatory region are entirely packaged as nucleosomes (green). An activator (orange oval) binds and recruits nucleosome remodelers to clear the promoter. **Step 2: PIC formation**. A second activator (yellow diamond) binds, promotes the binding of GTFs (blue rectangle) and recruits coactivators (green hexagon), facilitating Pol II (red 'rocket') entry to the PIC. **Step 3: initiation**. DNA is unwound (oval inside Pol II) at the TSS, and an open complex is formed. **Step 4: promoter escape/clearance**. Pol II breaks contacts with promoter-bound factors, transcribes 20–50 bases downstream of the TSS, produces an RNA (purple line) and pauses, partially mediated by SPT4−SPT5 in Drosophila (pink pentagon) and negative elongation factor (NELF) complex (purple circle). The Serine residues at position 5 (Serine 5) of the Pol II carboxy-terminal domain (CTD) repeats are phosphorylated (red P) during this step. **Step 5: escape from pausing**. P-TEFb (blue triangle) is recruited directly or indirectly by the activator and phosphorylates Serine 2 of the Pol II CTD repeats, SPT5 and the NELF subunits (blue Ps). NELF dissociates from the rest of the complex. Pol II escapes from the pause. **Step 6: productive elongation**. Nucleosomes are disassembled and reassembled as the Pol II elongation complex transcribes through the gene. **Step 7: termination**. After the Pol II complex transcribes the gene, it is removed from the DNA, and the RNA is released. **Step 8: recycling**. The freed Pol II can reinitiate.

After PIC assembly, RNA synthesis is initiated (Figure 1.1, step 3). The ATP hydrolysis by the GTF TFIIH alters the complex conformation to mediate “open” complex formation (Y. C. Lin, Choi, & Gralla, 2005). After the single-stranded DNA has entered the active site, RNA synthesis begins and the transcription bubble grows larger. Stabilization of the short RNA-DNA duplex by TFIIB aids these initial steps, but as the RNA-DNA hybrid grows, this interaction can prevent further synthesis. Collapse of the upstream portion of the extended transcription bubble may mediate rearrangements within the Pol II-GTF complex and allow Pol II to escape from the promoter (Figure 1.1, step 4) (Pal, Ponticelli, & Luse, 2005; Saunders et al., 2006). Within this process, the CTD of Pol II is phosphorylated on Serine 5 by the TFIIH-associated kinase, Cdk7. This modification breaks interactions some promoter bound factors, and may assist promoter escape (Max, Søgaard, & Svejstrup, 2007).

These early elongation complexes (EECs) can then transcribe further into the gene, but efficiency of RNA synthesis is highly dependent on sequence. Pol II complexes pause at certain sequences presumably due to weak RNA-DNA hybrids in the active site (Herbert et al., 2006; Tadigotla et al., 2006). When this occurs, the polymerase can also often translocate backward on the DNA, called backtracking (Landick, 2006). Pol II complexes that have backtracked several nucleotides can arrest, and can only restart synthesis with the help of the transcription factor TFIIS. TFIIS interaction with Pol II positions the backtracked RNA to allow Pol II to cleave the RNA, and creating a new 3’ end in the active site (Fish & Kane, 2002). This pausing is pronounced in higher eukaryotes about 60 bases downstream of the TSS (Core, Waterfall, & Lis, 2008; Min et al., 2011; Muse et al., 2007; Rahl et al., 2010). This

pausing is at least partially mediated through the actions of the Spt5/Spt4 and NELF complexes (Wada et al., 1998; Y Yamaguchi et al., 1999). As Pol II elongates, the growing RNA exits the complexes, and the protruding 5' end is then a substrate of mRNA capping enzyme associated with the Serine 5 phosphorylated CTD (Rasmussen & Lis, 1993).

Pol II escapes this pausing through the actions of P-TEFb (Figure 1.1, step 5). P-TEFb is a cyclin-dependent kinase composed CyclinT1 and the kinase Cdk9. P-TEFb can phosphorylate the CTD on Serine 2 as well as the C-terminal region of Spt5 and the Nelf-E subunit of the NELF complex (Fujinaga et al., 2004; N F Marshall & Price, 1995; Yamada et al., 2006). The NELF complex dissociates from Pol II, but Spt5/Spt4 complex remains associated with the elongating Pol II. Then, additional factors involved in elongation and RNA processing bind, many through interactions with the hyperphosphorylated CTD (Egloff & Murphy, 2008). The elongation factors associated with this complex can facilitate transcription past the nucleosomes obstructing transcription, and other factors can reduce pausing and increase elongation rate. These actions result in the removal of pause sensitivity, and make the complexes more processive during productive elongation (Figure 1.1, step 6) (Saunders et al., 2006).

Once the elongation complex has transcribed to end of a gene, it undergoes termination (Figure 1.1, step 7). Efficient termination is important for maintaining the pool of polymerases that can re-initiate transcription. Although the mechanistic details are not well understood, termination is likely mediated by changes in complex composition and conformation. The current evidence suggests termination is coupled to 3' end processing of the transcript, and Pol II pausing increases after the

polyadenylation site (Kuehner et al., 2011). The recruitment of termination factors and dissociation of elongation factors also appears critical for the process (Lunde et al., 2010). The terminated Pol II cannot begin another cycle of transcription until its CTD is dephosphorylated (Buratowski, 2009). This recycling is mediated by several phosphatases that dephosphorylate specific residues on the CTD repeats (Figure 1.1, step 8). Serine 5 residues are dephosphorylated by Ssu72, SCP1, and Rtr1, and Serine 2 is dephosphorylated by Fcp1 (E. J. Cho et al., 2001; Krishnamurthy et al., 2004; Mosley et al., 2009). Fcp1 and Rtr1 appears to act on the transcribing Pol II in yeast, but there is little in vivo information for higher eukaryotes (E. J. Cho et al., 2001; Mosley et al., 2009). Thus, it remains unclear when and how dephosphorylation is activated.

1.2 Rate-limiting steps in transcription

Any of the steps in the transcription cycle could be rate-limiting, and the distribution of Pol II across a gene can elucidate which step is rate-limiting for that gene. The Pol II density across many genes has been determined in a plethora of individual gene studies (Saunders et al., 2006). Moreover, a wealth of data has been obtained in recent genome-wide chromatin immunoprecipitation (ChIP) studies examining Pol II distribution across the genomes of several organisms: Saccharomyces cerevisiae (Venters & Pugh, 2009), Drosophila melanogaster (Muse et al., 2007), and Homo sapiens (Guenther et al., 2010; Rahl et al., 2010). In each organism, these studies have identified different classes of genes on the basis of their Pol II distribution: no Pol II, Pol II evenly distributed across the gene, and Pol II enriched on the 5´ ends of genes. Genes without Pol II are in an ‘off’ state, and are limited by the clearance of nucleosomes from the promoter (step 1) or the step of PIC assembly (step 2). An even

distribution of Pol II across the gene suggests that Pol II recruitment (step 2) is the rate-limiting step; none of the downstream steps leads to an accumulation of Pol II in other regions of the gene (Ptashne and Gann 1997). An enrichment in Pol II on the 5´ end of the gene suggests that steps downstream of Pol II recruitment (steps 3–5) are rate limiting. Because ChIP localization with a single Pol II-specific antibody cannot distinguish between steps 3-5, more experiments pinpointing the exact rate-limiting step need to be performed. The transition between PIC formation (step 2) and promoter escape (step 4) is marked by the unwinding of DNA, formation of a transcription bubble with a stable RNA–DNA duplex and lengthening of the nascent transcripts associated with Pol II. Transcription bubble formation and RNA length can be distinguished by permanganate mapping of the transcription bubble and nuclear run-on assays, respectively. Permanganate mapping utilizes piperdine cleavage of the permanganate-modified unpaired thymidines within the transcription bubble (Mirkovitch & Darnell, 1992). Nuclear run-on assays utilize the ability of transcriptionally-engaged polymerase in isolated nuclei to continue extending their nascent transcripts when provided exogenous nucleotides (Love & Minton, 1985; Rougvie & Lis, 1988). In addition, the transition between initiation and pausing (step 4) is marked by phosphorylation of the Pol II CTD repeats on Serine 5 by Cdk7, and productive elongation (step 6) is marked by phosphorylation of Pol II CTD repeats on Serine 2 by P-TEFb. Therefore, the rate limiting step on genes with 5´-end Pol II peaks can be distinguished using ChIP with antibodies to specific CTD phosphorylation marks (Boehm et al., 2003).

The Pol II distribution on individual genes and genome-wide differs between species. The amount of Pol II on genes in S. cerevisiae generally reflects the mRNA

level, and Pol II is relatively evenly distributed across genes in S. cerevisiae (Core & Lis, 2008; Peil et al., 2011). PIC formation is rate-limiting for the majority of yeast genes (Venters & Pugh, 2009). In contrast, genome-wide ChIP studies in Drosophila and mammals show a high enrichment of Pol II on the 5' ends of many genes (T. H. Kim et al., 2005; Muse et al., 2007; Rahl et al., 2010; Zeitlinger et al., 2007). Further, these polymerases have Serine 5 phosphorylated CTDs (Rahl et al., 2010). Permanganate mapping of several genes showed unpaired thymdines between 20 and 60 bases downstream of the TSS (C. Lee et al., 2008; Muse et al., 2007; Zeitlinger et al., 2007), and sequencing and mapping of run-on RNAs (GRO-seq) showed a dramatic enrichment for transcriptionally engaged polymerase at the 5' ends of genes (Chopra et al., 2011; Core et al., 2008; Min et al., 2011). This indicates these polymerase have initiated and pause downstream of the TSS. In agreement with this, the transcript levels of these genes do not reflect the level of Pol II (Zeitlinger et al., 2007). Additionally, Pol II distribution changes upon activation (Adelman & Rogatsky, 2010; M J Guertin et al., 2010; Min et al., 2011; Saha et al., 2011). The fraction of paused genes varies between studies depending on species and threshold for enrichment. Drosophila ChIP studies have estimated that 10-30% of genes have this 5' enrichment of Pol II (Muse et al., 2007; Zeitlinger et al., 2007). Mammalian GRO-seq studies have indicated 30-40% of genes have a 5' enrichment of transcriptionally engaged polymerase (Core et al., 2008; Min et al., 2011). These results indicate promoter-proximal pausing may be a wide-spread mechanism of gene regulation in metazoans.

1.3 Promoters and enhancers bind transcription factors that regulate gene expression

Although present evidence suggests that many steps in the transcription process may be rate limiting, the question remains whether these rate-limiting steps are actual points of regulation in gene activation. To meet this criterion, these steps should be regulated by transcription factors in response to particular physiological, environmental or developmental signals. The DNA sequences in and around specific gene promoters provide the code that dictates when, where and at what level specific genes are transcribed. This code comes in three parts: the core promoter, the region proximal to the core promoter, and the more distant enhancer sequences (Figure 1.2). In addition to the various combinations of elements within the core promoter sequence targeting the assembly of PICs through interactions with GTFs (Smale & Kadonaga, 2003), promoter-proximal and more distant enhancer sequences direct the binding of specific transcription factors, called activators or repressors. Although activators or repressors can interact directly with components associated with the core promoter, they execute their regulation predominantly through co-regulators, which are often multiprotein complexes. Some of the co-regulators can interact directly with Pol II and GTFs to influence expression. Others can reorganize nucleosomes or covalently modify chromatin, and change the chromatin architecture of the gene. This can in turn influence transcription factor associations and the transcriptional status of Pol II. Although transcription regulatory factors that act as repressors can also modulate specific steps, I focus here on activators, as they seem to predominate as critical modulators of gene expression in eukaryotes. Thus far, there are two major points of regulation in vivo: Pol II recruitment and escape from the pause.

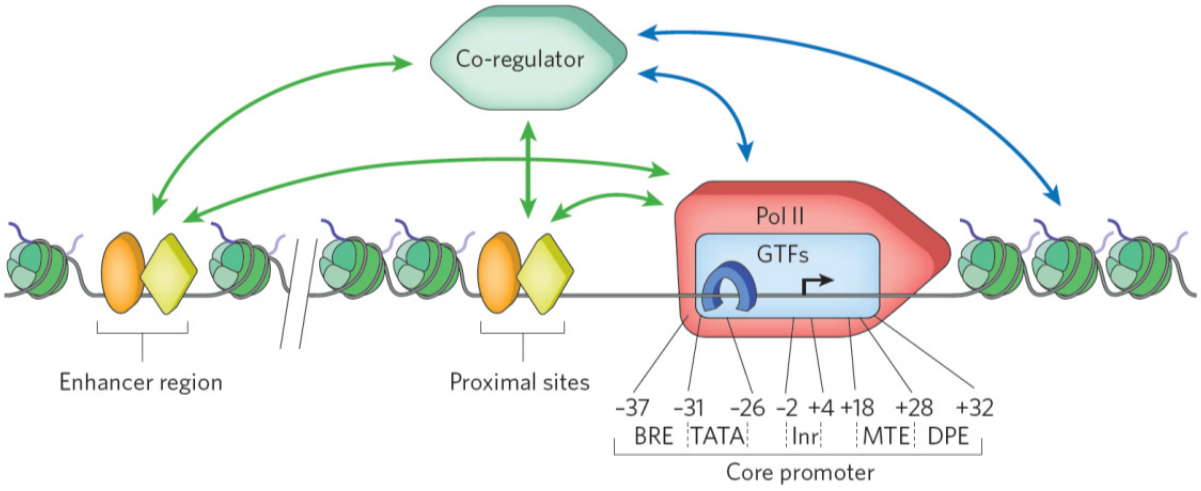


Figure 1.2: Transcription regulatory interactions. General transcription factors (GTFs) bind to specific sequence elements in the promoter. These elements (the B recognition element (BRE), the TATA box (TATA), the initiator (Inr), the motif ten element (MTE) and the downstream promoter element (DPE)) and their approximate locations relative to the transcription start site (TSS, black arrow) are shown2. Transcriptional regulators (orange oval and yellow diamond), which are either activators or repressors, bind to specific DNA sequences located near the core promoter of the gene or various distant regions, called enhancers. The regulators can interact (green arrows) with GTFs, such as TFIID (blue rectangle) and TATA-binding protein (TBP, blue horseshoe), and the Pol II complex (red 'rocket') to enhance or repress transcription. They also interact (green arrows) with co-regulators (green hexagon) that can interact (blue arrows) with the general transcription machinery or chromatin-modifying factors, such as histone modifiers or nucleosome remodelers. The co-regulators can also bind to nucleosomes (green) with various histone modifications, stabilizing the co-regulator binding to the gene. Activators can recruit, stabilize or stimulate these factors, and repressors can disrupt or inhibit these factors.

1.4 Regulating Pol II recruitment

Many genes regulated by the recruitment of Pol II have promoters covered with nucleosomes. Activators at these genes recruit nucleosome remodelers and nucleosome-modifying enzymes to allow GTFs and Pol II access to the promoter (step 1). PHO5 in S. cerevisiae is one of the best studied of the genes regulated in this manner (Figure 1.3). Its activation is dependent on eviction of nucleosome obstructing the core promoter (Svaren & Hörz, 1997). Under uninduced (high PO_4) conditions, acetylation of the nucleosome histone H4 and H2A tails is mediated by Pho2 recruitment of the histone acetyltransferase complex (HAT), NuA4 (Figure 1.3a). Low phosphate conditions ultimately result in nuclear import of the activator Pho4, which binds to its accessible low affinity motifs within the promoter. This binding recruits the coactivator SAGA, and its Gcn5 HAT acetylates the histones tail of the positioned nucleosomes (Reinke & Hörz, 2003; Steger et al., 2003) (Figure 1.3b). The nucleosomes are evicted presumably through the actions of the nucleosome remodelers, Swi/Snf and Ino80, and the histone chaperone Asf1 (M. W. Adkins et al., 2004; Korber et al., 2006) (Figure 1.3c). In other examples, it has been shown that both human and yeast activators interact with the SWI/SNF remodeling complexes and positively stimulate transcription from nucleosome-containing templates (Peterson & Workman, 2000).

In other genes, the promoter is free of nucleosomes, but Pol II recruitment is still rate limiting (step 2). During activated transcription, recruited Pol II quickly progresses into productive elongation and becomes relatively uniformly distributed across the gene (Ahn et al., 2004). At these genes, PIC assembly must be upregulated by activators.

Extensive in vitro studies have shown activators can interact with many GTFs: TBP, TFIID, TFIIA and TFIIB (Stargell & Struhl, 1996). Activators also recruit the coactivator Mediator, which can interact with GTFs and increase expression (Esnault et al., 2008; Y. J. Kim et al., 1994). These interactions might increase the binding of GTFs to the promoter or stabilize the PIC, allowing more efficient recruitment of Pol II. Additionally, activator-dependent recruitment of chromatin-modifying enzymes results in distinctive chromatin marks on promoters. Domains of GTFs can bind to these marks (Jacobson et al., 2000; Vermeulen et al., 2007), and these interactions can further aid in stabilizing PIC formation.

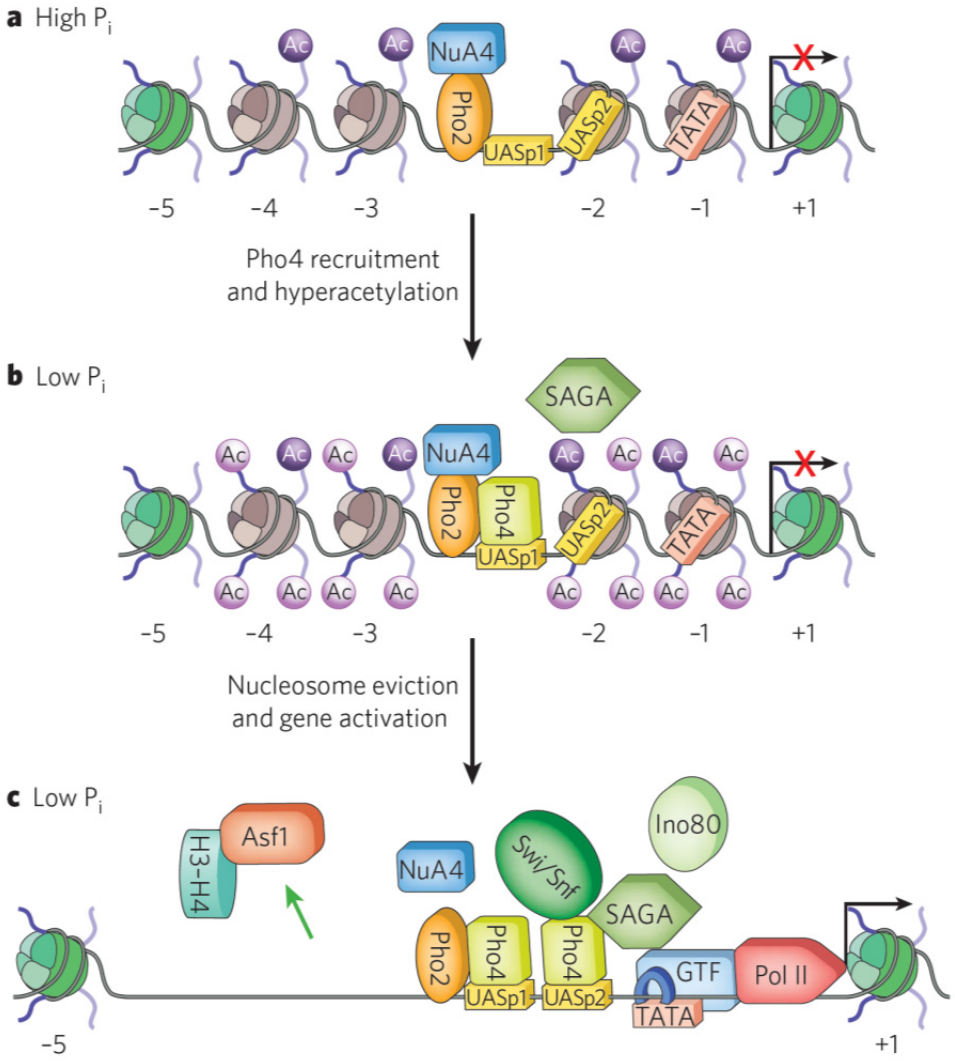


Figure 1.3: The Saccharomyces cerevisiae PHO5 gene is regulated at the chromatin-opening step. Transcription from the Saccharomyces cerevisiae acid-phosphatase gene PHO5 (panel a) is regulated at the level of activator recruitment and eviction of four positioned nucleosomes (gray, −1 to −4) from the upstream regulatory and promoter region. Pho2 recruits NuA4, which acetylates histones H4 and H2A (purple Ac). Phosphate (Pi) starvation (panel b) induces the accumulation of the active unphosphorylated Pho4 activator in the nucleus. Pho4 binds to the low affinity UASp1 within the hypersensitive site flanked by two positioned nucleosomes, cooperatively interacts with Pho2, and triggers disruption of the positioned nucleosomes. Pho4 binds to the high affinity UASp2 and activates transcription. This activation is mediated through the SAGA histone acetyltransferase Gcn5, the Swi/Snf and Ino80 nucleosome remodelers, and H3–H4 histone chaperone Asf1 (panel c).

1.5 Regulating post-recruitment steps

Enrichment in Pol II on the 5´ ends of genes indicates that post-recruitment steps can be rate limiting. Although it is possible that activators may regulate initiation (step 3) or promoter escape (step 4), there is currently no evidence for this in vivo. Permanganate mapping and global run-on sequencing studies (GRO-seq) have shown that the Pol II enriched on the 5´ ends of many genes is already engaged in transcription but is held paused 20-60 bases downstream of the TSS (Core et al., 2008; C. Lee et al., 2008).

Our current understanding of promoter-proximal pausing is derived from many in vitro and in vivo studies. The level of pausing is controlled through the rate of entry into the pause region and the rate of escape from the pause region (Core & Lis, 2008). The entry rate is a function of the steps before pausing: chromatin opening, PIC assembly, initiation, and promoter escape. The escape rate is controlled through the release into productive elongation and any termination of the paused polymerase.

1.5.1 Sequence-dependence of pausing

Pausing of RNA polymerases is dependent on the sequences being transcribed. The sequence of the RNA transcript can induce polymerase pausing in prokaryotes. RNA polymerases transcribing regions with weak RNA-DNA hybrids are more likely to pause and backtrack (Herbert et al., 2006; M Palangat & Landick, 2001). Recent evidence indicates this is also true for eukaryotic Pol II (Nechaev et al., 2010). The calculated melting temperature of 9 nucleotide RNA-DNA hybrids at putative paused genes showed a peak around +25 and progressively decreased until about +60. In

comparison, putative non-paused genes had a constant melting temperature across this region. Sequencing of the short RNAs associated with paused polymerase showed these RNAs increased in length when the transcript cleavage factor TFIIS was RNAi depleted from Drosophila S2 cells. This indicates the promoter-proximal sequences transcribed by Pol II contribute to pausing and backtracking. In addition, prokaryote RNA polymerases can also be induced to pause through interactions between RNA hairpin loops and the polymerase itself that produce a conformational change in the active of site of the polymerase (Toulokhonov et al., 2007), but there is no clear evidence of this mechanism in eukaryotic promoter-proximal pausing (Lee et al. 2008).

1.5.2 Protein factors with roles in pausing

Two protein complexes have central roles in pausing: DSIF and NELF. These factors were identified through experiments examining transcription sensitivity to 5,6-dichloro-1-B-D-ribofuranosylbenzimidazole (DRB). DRB dramatically inhibited transcription both in vivo and in vitro, and this inhibition affected elongation but not initiation (Chodosh et al., 1989). Biochemical fractionation identified two complexes required for this inhibition: DRB-sensitivity-inducing-factor (DSIF) and negative elongation factor (NELF) (Wada et al., 1998; Y Yamaguchi et al., 1999). DSIF is a heterodimer of Spt5 and Spt4 conserved throughout eukaryotes (Martinez-Rucobo et al., 2011). NELF has four subunits: Nelf-A, B, C/D, and E (C.-H. Wu et al., 2005). The NELF subunits are conserved among Drosophila and vertebrates, but are not present in yeast or C. elegans (Saunders et al., 2006; C.-H. Wu et al., 2005). Several in vivo results indicate the Spt5/Spt4 and NELF complexes have a role in pausing. Spt5 and NELF subunits localize to the promoter region of Drosophila NHS *Hsp70* (Andrulis et al.,

2000; C.-H. Wu et al., 2003) and mammalian paused genes (Aida et al., 2006). RNAi depletion of NELF subunits reduced the amount of transcriptionally-engaged polymerase on NHS *Hsp70* (C.-H. Wu et al., 2003) and reduced the 5' enrichment of Pol II at many genes in Drosophila (Gilchrist et al., 2008; Muse et al., 2007).

Although the exact mechanism of Spt5/Spt4 and NELF activity in pausing is not known, there are several studies that suggest possible mechanisms of action. There is evidence that the complexes may act through interactions with the nascent transcript. Pausing is first evident at about +20, when the RNA emerges from Pol II (Missra & Gilmour, 2010; Rasmussen & Lis, 1993). Spt5 and Nelf-E cross-link to the RNA in vitro (Missra & Gilmour, 2010), and Nelf-E has a RNA binding motif that is important for its inhibition of transcription in vitro (Rao et al., 2008; Y Yamaguchi et al., 2002). Spt5/Spt4 and NELF may also aid pausing through interfering with the actions of elongation factors. The Spt5/Spt4 and NELF complexes can compete with TFIIF to accentuate pausing (Renner et al., 2001). Additionally, Spt5/Spt4 and NELF complexes can inhibit TFIIS, and prevent backtracked complexes from elongating (Palangat et al., 2005).

1.5.3 The role of promoter bound factors in pausing

Several studies have investigated the role of GTFs and other promoter associated transcription factors in pausing. In addition to their role in the steps of initiation, two GTFs may also be important for pausing: TFIID and TFIIH. Downstream promoter elements bound by TFIID are enriched in paused genes, and are located within the pause region, around +30 (C. Lee et al., 2008; Smale & Kadonaga, 2003). It is not clear how these elements contribute to pausing. Since they are GC-rich, they may

enhance sequence-specific pausing as discussed previously. Alternatively, initiation on genes with these elements may be efficient, ensuring pause escape is the rate limiting step in transcription. Mutants of the TFIIH kinase Cdk7 reduced pausing in Drosophila and human cells (Glover-Cutter et al., 2009; Schwartz et al., 2003). Although the step targeted by Cdk7 activity is not known, NELF recruitment was reduced by Cdk7 inhibition (Glover-Cutter et al., 2009). Additionally, Cdk7 phosphorylation of the CTD Serine 5 residues alters factor association with the CTD. The interactions between the coactivator Mediator and Pol II during PIC assembly and initiation are disrupted by Cdk7 phosphorylation of the CTD (Max et al., 2007). Capping of the RNA occurs as the transcript emerges from the Pol II complex around +20 (Rasmussen & Lis, 1993). Serine 5 phosphorylated CTD interacts with the mRNA capping enzyme (E. J. Cho et al., 1997; Fabrega et al., 2003). The capping enzyme also interacts with the pausing factor Spt5 (Mandal et al., 2004; Pei & Shuman, 2002; Schneider et al., 2010; Wen & Shatkin, 1999). This interaction stimulates the guanylation by capping enzyme (Wen & Shatkin, 1999), and capping enzyme association with the pause complex can relieve the inhibitory effects of the Spt5/Spt4 and NELF complexes (Mandal et al., 2004). Thus, it has been proposed that pausing may be serving as a checkpoint for capping, but there is currently a little evidence for this in vivo.

1.5.4 The relationship between nucleosomes and pausing

Nucleosomes present a barrier to elongating polymerase. In vitro experiments demonstrate that Pol II alone cannot transcribe through a nucleosome (Izban & Luse, 1991), and other elongation factors are required to transcribe past the nucleosome block (Belotserkovskaya et al., 2003; Orphanides et al., 1998). Genome-wide

distribution of nucleosomes has been determined for several species (Jiang & Pugh, 2009). There is a nucleosome-free region (NFR) on the 5' end of the most genes flanked by a downstream (+1) highly positioned nucleosome. The +1 nucleosome position is dependent on the transcriptional activity of the gene; active and paused genes have +1 nucleosomes shifted downstream as compared to silent genes (Mavrich et al., 2008; Schones et al., 2008). In fact, it has been reported that polymerase is in close proximity to +1 nucleosome in Drosophila (Mavrich et al., 2008). This suggests the +1 nucleosome may provide the barrier to pause release. In contrast, pausing occurs in vitro on templates without nucleosomes (Adelman et al., 2005; Benjamin & Gilmour, 1998), and the +1 nucleosome on Drosophila *Hsp70* is located between +200 and +300, far downstream of the paused polymerase between +20 and +45 (Petesch & Lis, 2008). Thus, the contribution of nucleosomes to pausing is remains unclear.

Nucleosomes can also obstruct access to regulatory elements. As discussed previously, nucleosome remodelers are recruited by activators and coactivators to remove nucleosome from target promoters. Studies of heat shock genes indicate the GAGA factor binding sites are important for pausing (C. Lee et al., 2008; H. Lee et al., 1992; Shopland et al., 1995; Wang, Tang, & Gilmour, 2005). Pause genes are enriched for GAGA elements (Hendrix et al., 2008; C. Lee et al., 2008). In vitro studies show that GAGA factor can recruit the nucleosome remodeler NURF to *Hsp70* and *Hsp26* promoters (Granok, Leibovitch, & Elgin, 2001; T Tsukiyama & Wu, 1996; T Tsukiyama, Becker, & Wu, 1994; Wall et al., 1995). The promoter sequence of many paused genes favors nucleosome occupancy, but the paused polymerase displaces the nucleosome further downstream (Gilchrist et al., 2010). The competition between nucleosomes and

paused polymerase is important for the expression of many genes, indicating the paused polymerase aids in maintaining a nucleosome-free promoter.

1.5.5 Escape from pausing

Several factors are required for the efficient release of polymerase from the paused state. These factors operate on various aspects of the pausing to ensure the efficient transition the polymerase complex into productive elongation.

The kinase P-TEFb appears to play crucial role in pause escape. DRB is an ATP analog that inhibits P-TEFb (N F Marshall & Price, 1995) and dramatically reduces transcription elongation (Chodosh et al., 1989). P-TEFb targets Serine 2 residues on the CTD repeats, the C-terminal region of Spt5, and Nelf-E (Fujinaga et al., 2004; Yamada et al., 2006). Upon escape into productive elongation, NELF dissociates from the elongation complex (C. Lee et al., 2008). Mutation of putative P-TEFb targets on Spt5 impairs elongation in vitro and on paused genes in vivo, suggesting Spt5 phosphorylation acts as a switch from pausing to productive elongation (Yamada et al., 2006). The association of NELF with the paused complex is dependent on Spt5 (Missra & Gilmour, 2010; Y Yamaguchi et al., 2002), suggesting NELF dissociation may be mediated by the phosphorylation of Spt5. Serine 2 phosphorylation of the CTD allows several elongation, RNA processing, and termination factors to associate with the elongating complex (Egloff & Murphy, 2008). These factors are involved in nucleosome displacement and can directly affect elongation rate in vitro (Li et al., 2005; J Liu et al., 2011), suggesting that P-TEFb phosphorylation of the CTD may also mediate pause escape.

The efficiency of pause escape is dependent on the elongation competence of the complex. As discussed previously, the paused complex is prone to backtracking that makes the complex refractory to elongation (Nechaev et al., 2010; Nechaev & Adelman, 2011), and these backtracked complexes become transcriptionally competent through cleavage of the backtracked RNA by TFIIS (Adelman et al., 2005). TFIIS reduced the duration of pausing in vitro (Adelman et al., 2005). TFIIS depletion in Drosophila cells dramatically reduced the initial accumulation of *Hsp70* mRNA, the efficient movement of Pol II into the gene upon heat shock (Adelman et al., 2005), and increased the length of many TSS short RNAs associated with paused polymerases (Nechaev et al., 2010). TFIIS is critical for maintaining the paused polymerases in an elongation competent state for efficient release into productive elongation.

1.6 Drosophila *Hsp70* genes as model of pause regulation

Promoter-proximal pausing has been extensively characterized in focused studies of Drosophila *Hsp70*, and the results observed on *Hsp70* hold true for other paused genes (Core et al. in prep, M J Guertin et al., 2010). As a result, *Hsp70* has served as the model for genes regulated at the step of early elongation. Under uninduced (non-heat shock, NHS) conditions, its promoter resides in a nucleosome free region extending to about 250 bases downstream of the TSS (Petesch & Lis, 2008; C. Wu, 1980). This open promoter is bound by a transcription factor, GAGA factor (GAF) and GTFs (J. Lis, 1998). Studies of heat shock genes indicate the GAGA factor binding sites on *Hsp70* are important for pausing (H. Lee et al., 1992; Wang et al., 2005). In vitro studies show that GAGA factor can recruit the nucleosome remodeler NURF to *Hsp70* and *Hsp26* promoters (T Tsukiyama et al., 1994). GTFs bind to the open

promoter and allow Pol II to initiate and transcribe 20–40 bases downstream of the TSS. Spt5 and NELF association with the complex is important for pausing (C.-H. Wu et al., 2003) (Figure 1.4, panel a). Heat shock results in activation and binding of the activator heat shock factor (HSF) and recruitment of coactivators that modify the paused Pol II complex (Figure 1.4, panel b) (Park et al., 2001). P-TEFb is recruited upon activation, Spt5 CTR and the CTD Serine 2 residues are phosphorylated, and NELF binding is lost (J. T. Lis et al., 2000; C.-H. Wu et al., 2003). Heat shock leads to a rapid general loss of nucleosome protection across the gene (Petesch & Lis, 2008), and release of the paused Pol II into productive elongation (Figure 1.4, panel c). In an optimal heat shock, Pol II is evenly distributed across the gene, indicating pausing is no longer rate-limiting (J. Lis, 1998).

In addition to the conserved mechanisms of pause regulation, *Hsp70* has several attributes that make it an ideal model gene. The Drosophila genome has six copies of the *Hsp70* gene with well-conserved sequences and regulation. *Hsp70* has high levels of pausing, estimated at one polymerase per gene (J. Lis, 1998). This amplification of *Hsp70* and high level of pausing allow its use in assays with limited sensitivity. *Hsp70* is rapidly, robustly, and synchronously induced (Zobeck et al., 2010). *Hsp70* mRNA accumulation increases more than 100-fold just 20-minutes after heat shock, and factors are first recruited within seconds after heat shock (Boehm et al. 2003, Zobeck et al. 2010). The polymerase is stable under NHS conditions. It is estimated that Pol II is released by the paused region once every 10 minutes before heat shock. In contrast, a new Pol II is recruited and released every 4 seconds after heat shock (J. Lis, 1998).

These dramatic changes provide a large range in which to observe effects when the system is perturbed, and make *Hsp70* an ideal system to study pausing.

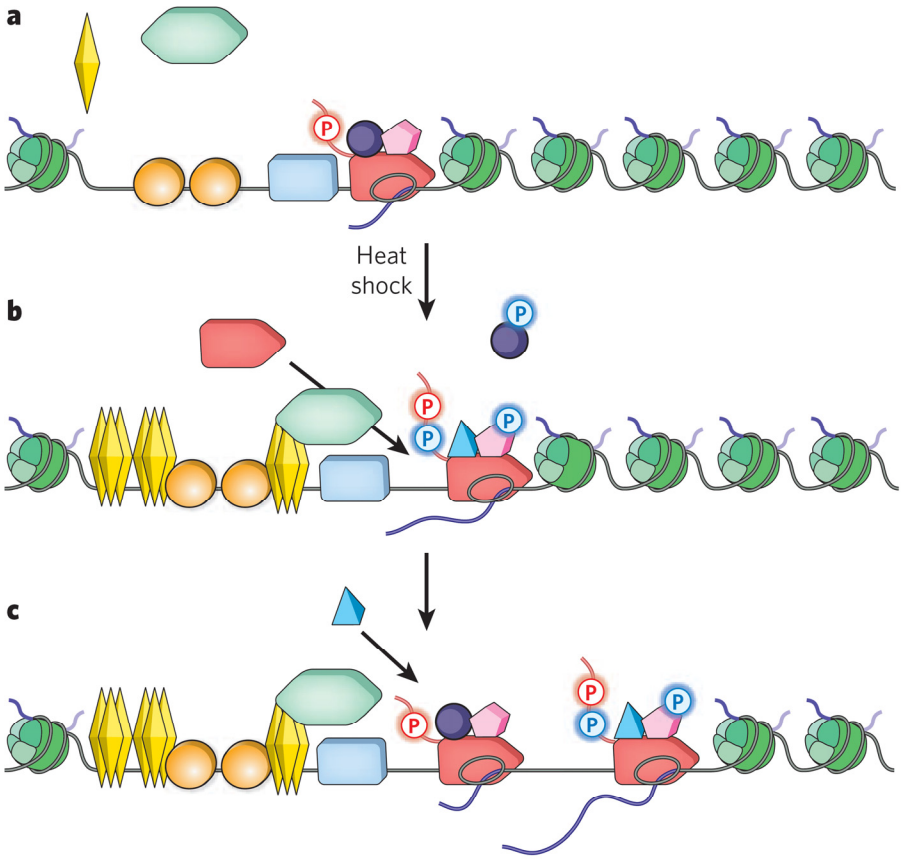


Figure 1.4: The Drosophila Hsp70 gene is regulated at the pause-escape step. a) The promoter resides in a nucleosome free region extending to about 250 bases downstream of the TSS, and is bound by GAGA factor (GAF, orange circles) and GTFs (blue rectangle). Pol II (red 'rocket') is recruited and initiates transcription, is partially phosphorylated on Serine 5 (red P), and transcribes 20-40 nucleotides, where it is held paused, at least partially mediated by the SPT4−SPT5 (pink pentagon) and the NELF (purple circle) complexes. b) Upon heat shock, the transcriptional activator HSF (yellow diamonds) trimerizes, binds to its upstream elements, and coactivators (green hexagon) and P-TEFb (blue triangle) are recruited. P-TEFb phosphorylates (blue P) the CTD, SPT5 and NELF subunits, NELF dissociates, Pol II releases from the pause sites, and new Pol II is rapidly recruited to the gene (black arrow). **c)** The new Pol II initiates transcription to begin the cycle again. Pol II still resides in the canonical pause sites for a much shorter duration, and NELF is also present after heat shock, but at lower levels.

1.7 Benefits of regulating at pause escape

Genes with promoter-proximal pausing have several characteristics that are beneficial to regulation of gene expression. Promoter-proximal paused Pol II provide a means of achieving a rapid and synchronous activation of gene expression (Boettiger & Levine, 2009; Boettiger, Ralph, & Evans, 2011). The paused Pol II has already progressed through multiple processes that can be slow and stochastic. A transcriptional activator, acting on a preloaded paused Pol II, allows a rapid transition into productive elongation. Genes with paused Pol II are not in a completely transcriptionally 'off ' state (Core et al., 2008). Therefore, regulation of pausing may sacrifice tight control of RNA production in favor of the uniform and rapid response of a gene. The heat-shock genes are a classic example of this regulation: their rapid induction seems critical in responding to a stress that is normally lethal. Other stress-response genes, such as those responsible for DNA-damage, unfolded-protein and immune-response pathways, are also enriched in paused Pol II (Adelman & Rogatsky, 2010; Saha et al., 2011). In the early embryo, narrow bands of cells must respond rapidly and uniformly to developmental signals, and genes that respond to these signals are also highly enriched in paused Pol II at the developmental stage at which they must be turned on (Zeitlinger et al., 2007). This may also be related to paused polymerase ability to maintain promoters open chromatin conformation (Gilchrist et al., 2010).

Pausing could also serve as a checkpoint. The proper assembly and modification of the elongation complex is critical to ensure efficient transcription and processing of the transcript (Sims et al., 2004). Additionally, the paused polymerase may also serve

as rapid method for attenuating transcription. A recent study has indicated that pausing is important in shutting off induced transcription (Ghosh, Missra, & Gilmour, 2011).

1.8 Dissertation outline

Although promoter-proximal pausing was identified many years ago and has been extensively researched, there are still many questions that remain. Recent studies have demonstrated that transcriptionally-engaged polymerase is enriched on the 5' ends of many genes, suggesting pausing may be a wide-spread point of regulation. Pausing appears to play a central role in the regulation of many genes involved in development and response to environmental changes. This dissertation presents my research into the factors and mechanisms controlling promoter-proximal pausing of RNA polymerase II in Drosophila melanogaster.

Chapter 2 contains the materials and methods used in this dissertation.

Chapter 3 investigates the role of GAGA factor in the establishing pausing. I depleted GAGA factor in S2 cell culture and examined the effects of pausing on specific genes using chromatin immunoprecipitation and genome-wide using global run-on sequencing. This chapter shows that GAGA factor depletion changes polymerase levels on many of its target genes. The large majority of genes have reductions in polymerase levels across the gene, but larger in the promoter-proximal region. In addition, GAGA factor bound genes that are unaffected by the depletion are enriched for transcription factors associated with insulator elements, suggesting these factors are protecting the genes from the effects of the depletion. Further analysis of specific genes demonstrates

the effects of GAGA factor depletion on polymerase levels are not due to changes in the histone or general transcription factor levels at most of these genes.

Chapter 4 investigates the role of downstream sequence in pause site specification. I have constructed transgene variants of the Drosophila *Hsp70* gene which shift the pause site sequence either five or ten bases downstream. The distribution of Pol II on these transgenes is determined in base pair resolution with precision run-on sequencing (PRO-seq). The five base insertion shifts pausing about 5 base downstream, suggesting pausing occurs at a specific sequence. In contrast, the gene with 10 bases inserted shows approximately a 15-fold reduction in pause Pol II, and the remaining pausing is more dispersed than the wild type genes. This suggests the spacing of downstream sequences determines the level of pausing that occurs. These results demonstrate the spacing of the underlying sequence is critical for specifying the pause pattern on *Hsp70*.

Chapter 5 investigates the factors contributing to pausing, in vitro. I used salt extraction of chromatin to examine which factors prevent the paused polymerase from transcribing. I find that moderate levels of salt allowed the paused polymerase on *Hsp70* to run-on, and dramatically dissociate the pausing factors Spt5 and NELF from chromatin, but not GTFs or histones. This suggests that the pausing factors are responsible for preventing the *Hsp70* paused Pol II from transcribing and not the +1 nucleosome. I also discuss possible uses of this technique to investigate mechanisms of pausing.

Chapter 6 investigates the role of Spt5 in maintaining the pause. I depleted Spt5 in S2 cell culture and examined the effects of pausing on *Hsp70* and genome-wide using global run-on sequencing. The Spt5 depletion reduces levels of promoter-proximal polymerase at most paused genes, and increased levels of polymerase in the gene body of many genes. The effects of depletion are similar to those seen previously when the pausing factor NELF is depleted.

Chapter 7 investigates the role of kinases in transcription elongation. I have investigated the effects of P-TEFb inhibition on RNA polymerase II distribution, and determined the localization of dCDK12 on actively transcribed genes. The results demonstrate that although P-TEFb is required for efficient release of the paused polymerase into productive elongation, there may be an additionaly mechanism functioning. CDK12 localizes to the gene body of transcriptionally active genes, but downstream of P-TEFb, suggesting a distinction functions for these kinases. I also discuss the possible roles of these kinases in controlling progression through the transcription cycle.

Appendix A presents work on a CTD phosphatase Fcp1 that is important for *Hsp70* transcription. I have depleted Fcp1 in S2 cell culture and examined the effects on RNA polymerase II distribution on induced *Hsp70*, and the phosphorylation status of the CTD. Fcp1 depletion reduces Pol II levels 2-fold on induced *Hsp70*, and phosphorylation accumulates on the CTD of free, non-chromatin Pol II. This result indicates Fcp1 is required for efficient initiation on *Hsp70* during an optimal heat shock induction.

CHAPTER 2

MATERIALS AND METHODS

2.1 dsRNA production:

The dsRNAs used in RNAi treatments were selected to target 400-1000bp of the desired target RNA, avoiding intron sequence as much as possible, using the online resource: http://www.dkfz.de/signaling2/e-rnai/. T7 promoter sequence (TAATACGACTCACTATAGGGA) was added to both the forward and reverse primers. The dsRNA template was amplified from Drosophila genomic DNA, and purified using a PCR purification kit. One microgram of template DNA was used per 25µl T7 polymerase reaction (400mM Tris-Cl pH 8.0, 100mM DTT, 20mM spermidine-HCl, 200mM $MgCl_2$, 0.1µl T7 polymerase (lab stock)). The reaction was incubated at 37°C for 4-16 hours, the DNA was digested with DNaseI. After extraction with phenol:chloroform, the RNAs were precipitated with NH_4Ac and 2 volumes of ethanol. The RNAs were resuspended in DEPC water, denatured at 80°C for 3 minutes, and reannealed on ice. The RNAs were quantified, and diluted to 1mg/ml. See Table2.1 for primers.

2.2 RNAi:

Drosophila S2 cells were grown in M3+BPYE+10% serum to a density between 3-5x10^6cells/ml. After splitting to 1x10^6cells/ml in serum-free M3 media (at least a 1:3 split), the desired volume of cells were mixed to 10µg/ml dsRNA, incubated at 25°C for 45 minutes, and an equal volume of M3+BPYE+20% serum was added. After 5 days, the cells were harvested for the experiments.

2.3 RT-qPCR:

RNA was isolated using Omega E.Z.N.A. Total RNA kit I (R6834), and quantified using NanoDrop 1000 spectrophotometer. Duplicate reverse transcription reactions were performed with 200ng of total RNA using SuperscriptIII reverse transcriptase (Invitrogen 18080) with oligo(dT) primer. After the reactions were diluted 10-fold with 10mM Tris-Cl (pH 8.0), 2µl was used in 10µl qPCR reactions to quantify the cDNAs using the following primer sets.

endogenous Hsp70:

Hsp70Ab+2155F primer GGTCGACTAAGGCCAAAGAGTCTA

Hsp70Ab+2266R primer TCGATCGAAACATTCTTATCAGTCTCA

Hsp70 transgenes:

Hsp70+1649F GGGTGTGCCCCAGATAGAAG

Hsp70+1754R TGTCGTTCTTGATCGTGATGTTC

Hsp26

Hsp26+580F CAAGGTTCCCGATGGCTACA

Hsp26+667R CTGCGGCTTGGGAATACTGA

RpL32 (Rp49):

Rp49+549F CCCAAGGGTATCGACAACAGA

Rp49+613R CGATGTTGGGCATCAGATACTG

Actin5C:

Actin5C+1781F 5'-GGAAATCCGCATTCTTTCCA

Actin5C+1848R 5'-CGACAACCAGAGCAGCAACTT

The qPCR was run on the Roche LightCycler480, and the level of each mRNA was calculated relative Rp49 using $2^{-\Delta C(t)}$.

2.4 Chromatin immunoprecipitation (ChIP):

Drosophila S2 cells were grown in M3+BPYE+10% serum to approximately $6x10^6$cells/ml. To prepare the heat shock chromatin, an equal volume of 48°C M3+BPYE (no serum) was added to the cells, and incubated at 36.5°C for the desired time. Then the same volume of 4°C M3+BPYE (no serum) was added to bring the cells to room temperature, and formaldehyde was immediately added to 1% final concentration. After 2 minutes of mixing at room temperature, the cross-linking was quenched with 2.5M glycine added to 125mM final concentration. After 2 minutes of mixing at room temperature, the cells were cooled on ice for 2 minutes, and centrifuged for 5 minutes at 4°C. After thoroughly removing the media, the cells were resuspended to $1x10^8$cells/ml in sonication buffer (20 mM Tris-Cl pH 8.0, 2 mM EDTA, 0.5 mM EGTA, 0.5% SDS, 0.5 mM PMSF, protease inhibitor cocktail [Roche catalog no. 05 056 489 001]). The cells were sonicated 12 times for 20 seconds each time with a 1 minute rest in between at 4°C using a Bioruptor sonicator (Diagenode) on the highest setting.

The sonicated material was centrifuged at 20K for 10 min at 4°C, and the supernatant was saved for the immunoprecipitation (IP). For each IP, 25µl of material

was mixed with 1ml IP buffer (20mM Tris-Cl pH 8.0, 150mM NaCl, 2mM EDTA, 10% glycerol, 0.5% TritonX-100) and 30ul 50% Protein A agarose (Millipore) at 4°C for 1-2 hours. The cleared material was mixed with the antisera at 4°C overnight (proteins were immunoprecipitated with 4µl of rabbit anti-Rpb3 antisera (Lis laboratory stock), 2µl of rabbit anti-HSF antisera (Lis laboratory stock), 2.5µl guinea pig anti-Spt5 (Lis laboratory stock), 7µl affinity purified rabbit Anti-dCDK12 (Greenleaf lab stock), 10µl rabbit anti-Fcp1 (Reinberg lab stock), 10µl mouse monoclonal IgM H5 antibody (Covance) for phosphorylated Serine 2 CTD, 10µl mouse monoclonal IgM H14 antibody (Covance) for phosphorylated Serine 2 CTD, 10µl affinity purified rabbit anti-GAF (Lis lab stock), 2µl rabbit anti-histone H3 (Abcam ab1791), 4µl rabbit anti-TFIIA (Kadonaga lab stock), 4µl rabbit anti-TFIIB(Kadonaga lab stock). The non-heat shock chromatin was prepared in the same manner, except 2 volumes of room temperature M3+BPYE (no serum) was added to the cells before cross-linking with formaldehyde. The immunoprecipitated DNA and a standard curve of 10%, 1%, 0.1%, and 0.01% of input DNA was quantified using a Roche LightCycler 480, and the standard curve was used to determine the amount of DNA immunoprecipitated.

2.5 Terminated nuclear run-ons:

This method was adapted from previous work (Rasmussen & Lis, 1993) (Figure 2.1).

After treatment, cells were rapidly cooled by addition of treated cells to 150ml ice-cold media. Cells were washed and resuspended in Buffer A (10mM Tris-Cl pH 8.0, 300mM sucrose, 3mM $CaCl_2$, 2mM $MgAc_2$, 0.1% TritonX-100, 0.5mM DTT) at $1x10^8$cells/ml. The cells were homogenized with 20 strokes in a 2ml teflon dounce

homogenizer. The nuclei were extracted and washed with Buffer A, and resuspended to $1x10^8$cells/ml in Buffer D (50mM Tris-Cl pH 8.0, 25% glycerol, 5mM $MgAc_2$, 0.1mM EDTA, 5mM DTT).

Nuclei were mixed with 115µl Run-on Buffer (7.2mM Tris-Cl pH 8.0, 4.4mM $MgAc_2$, 1.4mM $MnCl_2$, 150mM KCl, 3mM ATP, 3mM CTP, 0.4mM 3'-deoxy-GTP (TriLlink BioTechnologies), 100µCi ^{32}P-UTP (3000Ci/mmol), 40U Superase-In (Ambion), 0.6% N-lauroyl-sarcosine). Run-ons proceeded for 10 minutes at room temperature before a 30 minutes RQ1 DNase digestion at 37°C, followed by addition of 235µl Stop Solution (20mM Tris-Cl pH 7.4, 2% SDS, 10mM EDTA) and 40µl RNA grade Proteinase K with incubation at 55°C for 30 minutes. Purified RNAs were mixed with the biotinylated probe (the biotin was covalently attached to the 3' end of the oligo via a 12carbon spacer),

Hsp70 probe: GCTTTCGCTTAGCGACGTGTTCACTTTGCTTGTTTGATTT-3'BioTEG

Hsp26 probe: CTTTGAGTTGTTCACTGCTCGATTTTTGAATTCGATCTGT-3'BioTEG

After denaturing at 80°C for 5 minutes, the RNA/probe mix was chilled on ice for 2 minutes, and allowed to hybridize at room temperature for at least 12 hours. Hybridized RNA was captured with 300µg strepavidin-coated magnetic beads (Dynal Biotech). Beads were washed four times with Wash Buffer (10mM Tris-Cl pH 7.5, 50mM NaCl, 5mM EDTA, 0.5mg/ml yeast tRNA), and resuspended in 6µl Gel Loading Buffer II (Ambion). RNAs were separated by electrophoresis through 8% acrylamide/7M urea sequencing gels. To calculate the 1.5 – 2.5 fold increase in the number of Pol II molecules on the promoter proximal region from the NHS to the HS + FP condition, we

estimated that this pair of labeled and chain-terminating nucleotides will detect 35% of paused Pol II in NHS cells and 55% of the Pol II (if randomly distributed) through the first 120 bases of *Hsp70* (Ni et al., 2008).

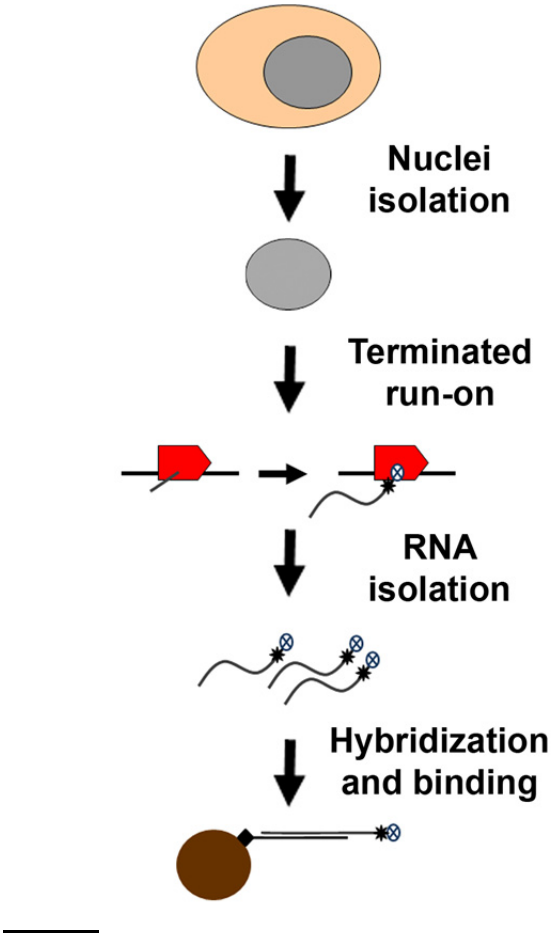


Figure 2.1 Schematic for the terminated run-on methods.

2.6 Cellular fractionation

The fractionation of proteins (free versus chromatin bound) adapted from previous work (Aygün, Svejstrup, & Liu, 2008).

Cells were centrifuged the cell at 1000rpm for 5 minutes, and resuspended in nuclei lysis buffer (20mM Tris-Cl pH 7.5, 3mM EDTA, 10% glycerol, 150mM KAc, 1.5mM $MgCl_2$, 1mM DTT, 0.1% NP-40, and protease inhibitors [roche Roche catalog no. 05 056 489 001]) to $1x10^8$cells/ml. The cells were immediately homogenized with 60 strokes in a 2ml Teflon dounce homogenizer, and centrifuged at 15K for 5 minutes at 4°C, the supernatant (free fraction) was transferred to a new tube, and the pellet was resuspended in nuclei lysis buffer to the equivalent of $1x10^8$cells/ml (chromatin fraction). SDS loading buffer was added to each fraction to 1x final concentration.

2.7 Construction of GRO-seq libraries

2.7.1 Nuclei isolation (GRO-seq)

Cells were centrifuged at 3000rpm for 5 minutes, washed once with cold PBS, and resuspended in 1 ml Buffer S (10mM Tris-Cl pH 7.5, 10% glycerol, 3mM $CaCl_2$, 2mM $MgAc_2$, 0.5mM DTT, protease inhibitors [Roche catalog no. 05 056 489 001], 4U/ml Superase-In [Ambion]). After 5 minutes at 4°C, 9ml Buffer L (10mM Tris-Cl pH 7.5, 300mM sucrose, 10mM NaCl, 3mM $CaCl_2$, 2mM $MgAc_2$, 0.1% TritonX-100, 0.5mM DTT, protease inhibitors [Roche catalog no. 05 056 489 001], 4U/ml Superase-In [Ambion]) was added, and the cells were homogenized with 20 strokes in a 15ml teflon dounce homogenizer. The cells were centrifuged at 3000rpm for 5 minutes at 4°C, washed once with 10ml of buffer L, centrifuged at 3000rpm for 5 minutes at 4°C, and

the nuclei were resuspended in 1ml cold storage buffer (50mM Tris-Cl pH 8.0, 25% glycerol, 5mM $MgAc_2$). The nuclei were immediately counted, centrifugation at 1000g for 5min, resuspended in storage buffer to $2x10^8$nuclei/ml, aliquots of 100µl were transferred into 1.5 ml tubes, flash frozen in liquid nitrogen, and stored at -80°C until used.

2.7.2 Nuclear run-on

The nuclear run-on should result in a run-on of about 100 nucleotides (Core et al., 2008). $1x10^7$ nuclei (100µl) were mixed with 100µl Run-on Buffer (10mM Tris-Cl pH 8.0, 5mM $MgCl_2$, 300mM KCl, 500uM ATP, 500uM GTP, 2uM CTP (cold), 1mCi/ml ^{32}P-CTP (100uCi/ run-on), 500uM Br-UTP, 0.4 units Superase-In, 1mM DTT, 40U Superase-In (Ambion), 0.6% N-lauroyl-sarcosine), incubated for 10 minutes at 30°C, the reaction was stopped with 1.5ml Trizol and 200µl chloroform, the aqueous phase was extracted once with 1ml acid phenol:chloroform, and once with 1ml chloroform. The aqueous phase was mixed with 1.5µl glycoblue (for carrier) and 2.5 volumes cold ethanol, precipitated at room temperature for 5 mnutes, centrifuged at 12000rpm for 30 minutes at 4°C, washed with 500µl 70% ethanol, and resuspended in 20µl DEPC-treated ddH_2O.

2.7.3 Base hydrolysis

To get high resolution mapping of the position of the polymerase, the run-on RNAs should be hydrolyzed to distance equivalent to the run-on distance (about 100 nucleotides (Core et al., 2008). This hydrolyzed is performed with NaOH, and was optimized empirically. I used 200mM final concentration on ice for 18 minutes. The

hydrolysis was stopped with Tris-Cl (pH 6.8) at a final concentration of 500mM and buffer exchange using a Bio-spin P30 column.

2.7.4 BrU purification

The RNAs were brought up to a final volume for 100µl with DEPC-treated ddH_2O, denatured at 70°C for 5 minutes, and chilled on ice for 2 minutes. The RNAs were brought up to 450µl with binding buffer (0.25x SSPE, 1uM EDTA, 0.05% Tween-20, 37.5mM NaCl, 0.1% polyvinylpyrrolidone, 1ug/ml BSA) and bound to 50ul Anti-Br-dUTP beads preblocked with to each Blocking buffer (0.25x SSPE, 1uM EDTA, 0.05% Tween-20, 37.5mM NaCl, 0.1% polyvinylpyrrolidone, 1ug/ml BSA) at room temperature for at least 20 minutes. The supernatant was saved to check the binding efficiency (Figure 2.2B). The beads were washed once with 500µl binding buffer, once with 500µl Low salt buffer (0.2x SSPE, 1mM EDTA, 0.05% Tween-20), once with 500µl High salt wash (0.25x SSPE, 1mM EDTA, 137.5mM NaCl, 0.05% Tween-20), and twice with 500µl TET wash (10mM Tris-Cl pH 7.5, 1mM EDTA, 0.05% Tween-20). The RNAs were eluted the beads three times for 10 minutes with 125µl elution buffer (50mM Tris-Cl pH 7.5, 150mM NaCl, 1mM EDTA, 0.1% SDS, 20mM DTT), the RNAs were precipitated as before. The RNAs were resuspended in 20µl DEPC-treated ddH_20, and 1µl was sampled to check binding efficiency (Figure 2.2B). Typically greater than 80% of counts are eluted from the beads.

2.7.5 PNK treatment

The run-on RNAs are treated with T4 polymucleotide kinase (PNK) without ATP to ensure the 3'end has a hydroxyl group. After the reaction, the RNAs were put through two more bead bindings.

2.7.6 Polyadenylation, reverse transcription, and PCR amplification

Previous GRO-seq libraries used linker ligations to construct the libraries (Core et al., 2008; Min et al., 2011). The RNA ligation can add sequence bias to the libraries. To avoid any biases, the Weismann lab created a method to eliminate this bias (Ingolia et al., 2009). This method uses polyadenylation of the RNAs and reverse transcription with a poly(dT)-3'linker covalently attached to the 5' linker with a 18 carbon spacer (sequences below, Figure 2.2C).

The RNAs were polyadenylated with E. coli polyA polymerase at room temperature for 20 minutes in 10µl reactions. Samples were taken before and after polyadenylation and run on an 8% TBE-urea polyacrylamide gel (Figure 2.2C).

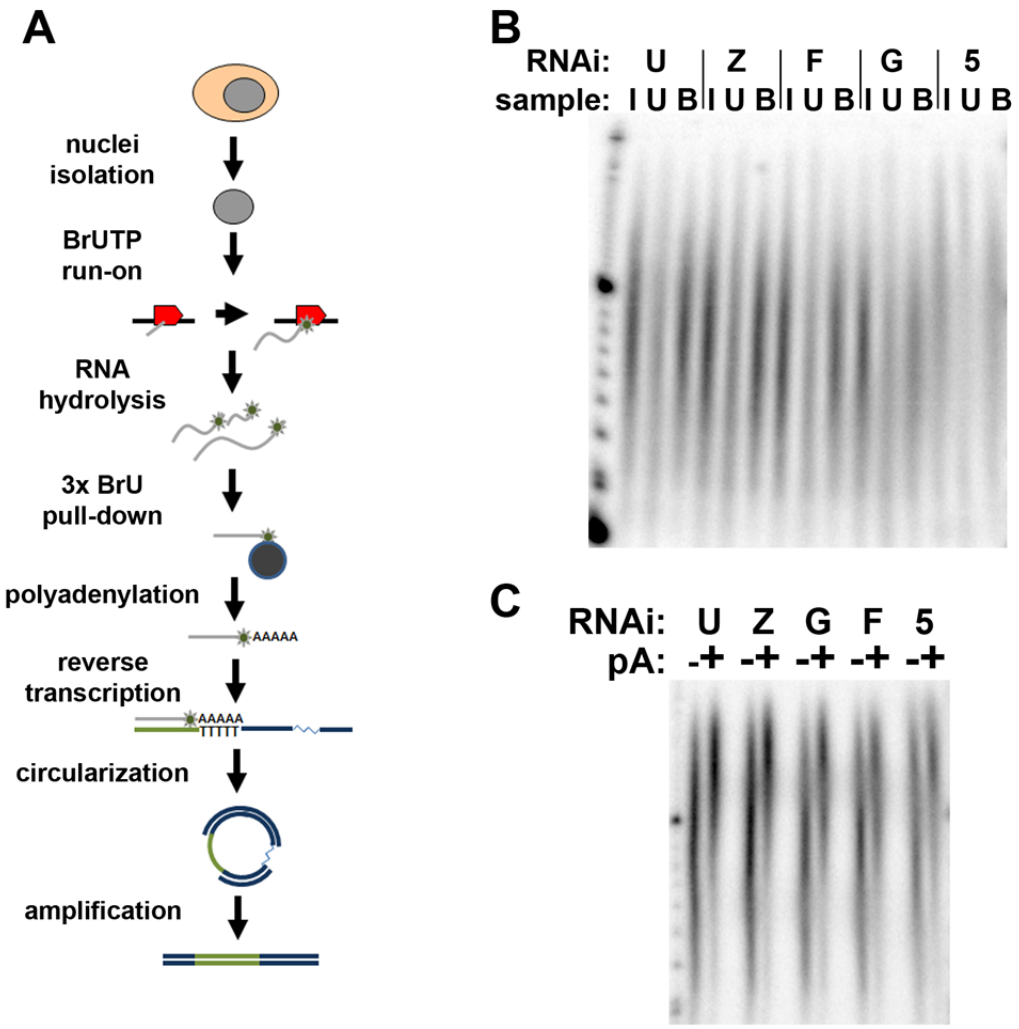


Figure 2.2 Construction of GRO-seq libraries. **(A)** Schematic for methods. **(B)** First BrU bead binding for run-ons from Untreated (U), LacZ-RNAi (Z), GAF-RNAi (G), FCP1-RNAi (F), and Spt5-RNAi (5) nuclei. One percent of each run-on (I), 5% of each unbound fraction (U), and 1% of each eluted material (B) was run on a 8% PAGE TBE-urea gel. **(C)** Check from the run-on RNA polyadenylation for run-ons from Untreated (U), LacZ-RNAi (Z), GAF-RNAi (G), FCP1-RNAi (F), and Spt5-RNAi (5) nuclei. Five percent of each eluted material from the third BrU bead binding and 5% of each polyadenylated RNA was run on a 8% PAGE TBE-urea gel.

Each library was made in biological replicates, and bar-coded using specific reverse transcription primers.

INOO3:

5’-pTAGAGATCGTCGGACTGTAGAACTCT-iSp18-
CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTTTVN

INOO4:

5’-pTGATGATCGTCGGACTGTAGAACTCT-iSp18-
CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTTTVN

The reverse transcription was performed in 20µl reactions, and the cDNA was purified from a polyacrylamide gel to remove the reverse oligomer. The cDNA was circularized using Circligase (Epicentre catalog # CL4111K) in 20µl reactions to connect the 5’ linker to the 5’ end of the cDNA. PCR amplification was performed on this material. The correct number of cycles to use in the amplification was determined empirically with 1/10 of each library because over-amplification can also lead to sequence bias. After full-scale amplification of the remaining libraries, the amplified libraries were gel purified away from primers. Each replicate library was combined in equal amounts and sequenced on one lane of an Illumina sequencer.

2.8 Nuclei isolation (Chapter 5):

Nuclei were isolated as previously (Love & Minton, 1985). Cells were resuspended in 7.5ml buffer A (10mM Tris-Cl pH 8.0, 300mM sucrose, 3mM $CaCl_2$, 2mM $MgAc_2$, 0.1% TritonX-100, 0.5mM DTT), transferred to a 15ml glass dounce homogenizer, and homogenized with 20 strokes of a tight-fitting pestle. The cells were mixed with an equal volume of buffer B (10mM Tris-Cl pH 8.0, 2M sucrose, 5mM $MgAc_2$, 0.5mM DTT), and layered over 10ml buffer B in centrifuge tubes. The nuclei were centrifuged through the buffer B cushion at 12K for 25 minutes at 4°C in a SW28 swinging bucket rotor. The supernatants were thoroughly removed, and the nuclei were resuspended in 1ml buffer C (50mM Tris-Cl pH 8.0, 25% glycerol, 5mM $MgAc_2$, 0.1mM EDTA, 5mM DTT). The nuclei were counted with a hemocytometer, centrifuged at 100rpm for 5 minutes at 4°C, resuspended to $1x10^8$nuclei/ml, and 100µl aliquots were flash frozen in liquid nitrogen, and stored at -80°C until used.

2.9 Adult fly nuclei isolation (Chapter 4):

The nuclear isolation from adult flies was adpted from (H. Lee et al., 1992).

One gram of flies was homogenized in 15ml cold Buffer A (10mM Tris-Cl pH 8.0, 300mM sucrose, 3mM $CaCl_2$, 2mM $MgAc_2$, 0.1% TritonX-100, 0.5mM DTT) for 1 minute straight using the Omni-mixer, the homogenate was filtered through 100µm nylon mesh into a 15ml Dounce homogenizer, and dounce 40 times (40ml dounce homogenizer). The dounced homogenate was filtered through 35µm nylon mesh into a 50ml tube, mixed with an equal volume of Buffer B (10mM Tris-Cl pH 8.0, 2M sucrose, 5mM $MgAc_2$, 0.5mM DTT), and layered over 10ml Buffer B in 35ml Ultracentrifuge tube.

The nuclei were centrifuged through the buffer B cushion at 12K for 25 minutes at 4°C in a SW28 swinging bucket rotor. The supernatants were thoroughly removed, and the nuclei were resuspended in 1ml buffer C (50mM Tris-Cl pH 8.0, 25% glycerol, 5mM $MgAc_2$, 0.1mM EDTA, 5mM DTT). The nuclei were counted with a hemocytometer, centrifuged at 100rpm for 5 minutes at 4°C, resuspended to $1x10^8$nuclei/ml, and 100µl aliquots were flash frozen in liquid nitrogen, and stored at -80°C until used.

CHAPTER 3

GAGA FACTOR ESTABLISHES PROMOTER-PROXIMAL PAUSING ON A SUBSET OF BOUND GENES

3.1 Introduction

GAGA factor (GAF) is encoded by the gene *Trithorax-like* (*Trl*), binds GA repeats, and was first identified as a regulator of developmental genes (Farkas et al., 1994; Hagstrom, Muller, & Schedl, 1997; Horard et al., 2000; Omichinski et al., 1997; Wilkins & Lis, 1998). Genome-wide studies have identified as many as 1566 genes bound by GAF (Granok et al., 2001; C. Lee et al., 2008; Nègre et al., 2006; van Steensel, Delrow, & Bussemaker, 2003). The bound regions generally have clusters of GAGA elements, suggesting cooperative binding (van Steensel et al., 2003). GAF is composed of three functional domains: an N-terminal BTB/POZ domain, a DNA-binding domain, and a C-terminal polyglutamine region (Adkins, Hagerman, & Georgel, 2006). The DNA binding domain is composed of a region rich in basic residues followed by a C2-H2 zinc finger. GAF can bind as little as GAG through its zinc finger, but interactions with the adjacent basic residues allows it to recognize the larger sequence of GAGAG (Omichinski et al., 1997; Pedone et al., 1996; Wilkins & Lis, 1998). The BTB/POZ domain mediates interactions with other proteins, and allows GAF to dimerize or interact with other POZ-containing factors (Espinás et al., 1999; Katsani, Hajibagheri, & Verrijzer, 1999; Pagans et al., 2002; Read et al., 2000; Schwendemann & Lehmann, 2002). The function of the polyQ domain is not well-understood, but has been reported

to act as a transcription activator (Vaquero et al., 2000; Vaquero et al., 2008) and form multimers with itself (Agianian et al., 1999; Wilkins & Lis, 1999).

A wide variety of activities has been associated with GAF. There have been several reports of GAF functioning as an anti-repressor for many genes (Adkins et al., 2006). GAF anti-repressor function is proposed to maintain bound regions in an accessible state, which is possibly mediated through interactions with nucleosome remodelers (Okada & Hirose, 1998; T Tsukiyama & Wu, 1996; T Tsukiyama et al., 1994). Additionally, GAF can relieve the repressive effects of histone H1 (Croston et al., 1991). In contrast to its positive role in expression, GAF can also associate with repressors. GAF binding overlaps with the transcriptional silencer, polycomb complex (PcG) (Nègre et al., 2006), and PcG may be recruited by GAF through mutual interactions with Corto (Salvaing, 2003). GAF can also mediate repression by recruiting the Sin3-histone deacetylase (HDAC) corepressor complex through mutual interactions with SAP18 (Espinás et al., 2000). Additionally, GAF is located at many insulators, and several studies indicate GAGA elements are necessary for insulator function (Belozerov et al., 2003; Nègre et al., 2006, 2010; Ohtsuki & Levine, 1998; O'Donnell, Chen, & Wensink, 1994).

GAGA elements and GAF binding are enriched on paused genes (Hendrix et al., 2008; C. Lee et al., 2008), and studies using reporter genes have shown that the presence of transcriptionally-engaged polymerase under uninduced conditions increases in GAGA element-containing reporter genes (H. Lee et al., 1992; Wang et al., 2005). These results suggest that GAF has a role in pausing, but this has not been well-studied in vivo. Here, I examined the role of GAF in promoter-proximal pausing, and

show that RNAi depletion of GAF reduced paused polymerase levels on NHS *Hsp70*, as well as on many GAF-bound genes. These changes in promoter-proximal polymerase levels were accompanied by similar changes polymerase on the gene body. I observed changes in histones and GTFs on only a subset of these promoters of genes of genes examined, and from the modENCODE ChIP datasets, I found insulator-associated factors were enriched on unaffected GAF-bound genes.

3.2 Results

3.2.1 RNAi depletion of GAF in S2 cells

To deplete GAF, I used a dsRNA targeting all mRNA isoforms for knock-down. Drosophila S2 cells were treated with the dsRNA for 5 days, and whole cell extracts were immunoblotted with an affinity purified antibody to GAF. GAF levels were typically reduced greater than 10-fold in GAF-RNAi cells compared to Untreated and control cells treated with LacZ dsRNA (Figure 3.1A). The effect of knock-down on GAF binding to NHS *Hsp70* was examined using chromatin immunoprecipitation (ChIP). Untreated and LacZ-RNAi cells showed the expected distribution of GAF with high levels on the promoter (Hsp70-154) and low levels in the gene body (Gilmour et al., 1989), and GAF knock-down decreased GAF levels about 4-fold (Figure 3.1B).

Next, I assayed the effect of GAF reduction on the levels of paused polymerase using ChIP for the Pol II subunit, Rpb3. In Untreated and LacZ-RNAi cells, Pol II levels were high at the 5' end of the uninduced NHS gene (Hsp70+96) and low in the gene body and a background region 32 Kb away from *Hsp70* (Figure 3.1C). GAF knock-down resulted in 2-fold reduction in Pol II at the 5' end with no discernible change in the gene

body. These results show GAF has a role in maintaining the level of Pol II on the 5’ end of NHS *Hsp70*.

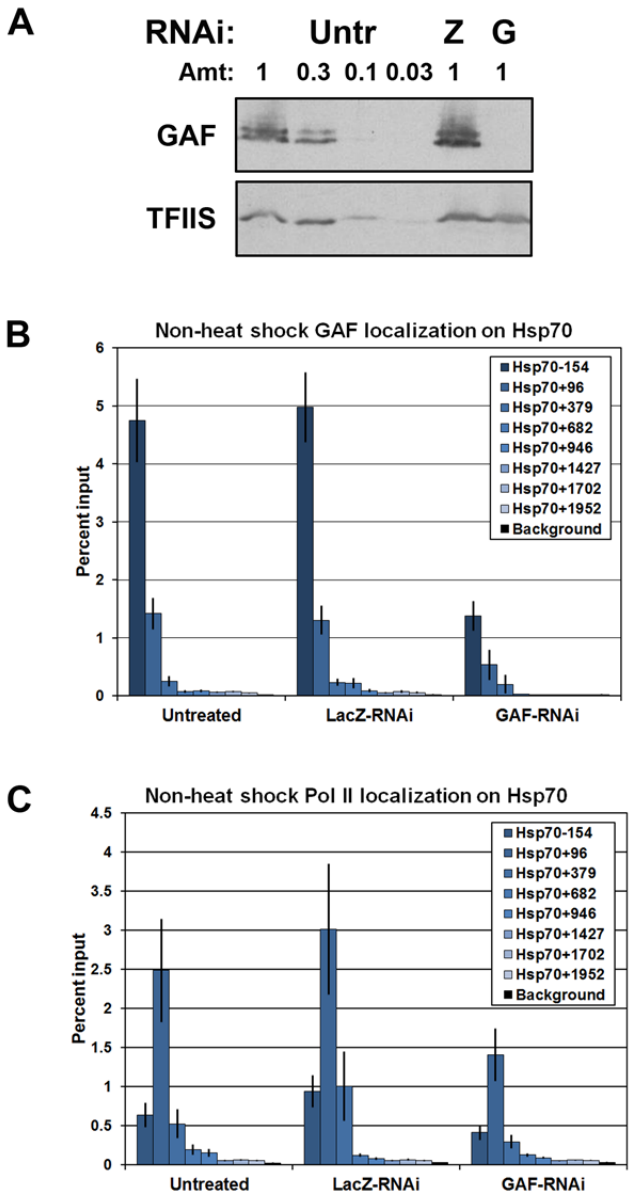


Figure 3.1: Depletion of GAF reduces paused polymerase on NHS Hsp70. (A) Western blots of whole cell extracts from Untreated (Untr), LacZ-RNAi (Z), and GAF-RNAi (G) cells for GAF and a loading control, TFIIS (1 is equivalent to $1x10^6$ cells). **(B)** ChIP for GAF on Hsp70 in non-heat shock (NHS) Untreated, LacZ-RNAi, and GAF-RNAi cells. **(C)** ChIP for Pol II subunit, Rpb3, on Hsp70 in NHS Untreated, LacZ-RNAi, and GAF-RNAi cells. The legend indicates the center of each primer set relative to the TSS. Error bars represents the SEM from at least 3 experiments.

3.2.2 Genome-wide distribution of polymerase in GAF knock-down cells

The wealth of data on GAF binding motivated me to examine the polymerase distribution genome-wide in control and GAF knock-down cells. Although ChIP can identify Pol II bound regions of the genome, its sensitivity is limited and it does not determine if the polymerase is transcriptionally engaged. To obtain a comprehensive view of the distribution of transcriptionally-engaged polymerase, Leighton Core and I performed global run-on sequencing (GRO-seq) in control and GAF knock-down cells (Core et al., 2008). Nuclear run-ons make use of the ability of transcriptionally-engaged polymerases to extend their nascent transcripts when provided with exogenous nucleotide-triphosphates. By performing the run-on with bromo-UTP (BrUTP), the incorporation of BrUTP into the extended nascent transcripts allows them to be purified away from the abundant non-extended RNAs by successive immunoaffinity pull-downs. After the library is sequenced, the sequence reads are mapped to the genome, and the density of reads within a region indicates the level of engaged.

The density of GRO-seq reads was used to determine the distribution of engaged polymerase in Untreated, LacZ-RNAi, and GAF-RNAi NHS cells. Colin Waters performed the mapping and some of the preliminary analysis. The distribution of normalized reads for each library was graphed relative to the TSS of all genes. In agreement with previous GRO-seq results in Drosophila (Chopra et al., 2011; Larschan et al., 2011, Core et al. in prep) and genome-wide ChIP data (Muse et al., 2007), each library displayed a peak of engaged polymerase on 5' ends, and the GAF knock-down profile looked very similar to the Untreated and LacZ-RNAi control (Figure 3.2). Next, we examined the levels of polymerase on the promoter region (100bp region within

250bp of the TSS containing the most reads) and gene body (500bp downstream of the TSS to the polyA site) for each gene. The promoter and gene body read counts for each gene correlated well between the Untreated and LacZ-RNAi libraries (Figure 3.3A-B, promoter ρ=0.979, gene body ρ=0.978) and between the LacZ-RNAi and GAF-RNAi libraries (Figure 3.3C-D, promoter ρ=0.972 and gene body ρ=0.973). Although the overall correlation is high, the read counts in many regions differ between the LacZ-RNAi and GAF-RNAi libraries, as evident from the larger variation in GAF-RNAi to LacZ-RNAi read ratios (Figure 3.3C-D). The genes with significantly different reads between the LacZ-RNAi and GAF-RNAi libraries were determined using edgeR, which determines these differences based on the variation between the replicate libraries of each treatment (Robinson, McCarthy, & Smyth, 2010). EdgeR called 139 genes with significantly different read counts (p-value<0.01) in the promoter-proximal region (Figure 3.3C, red points); virtually all were reduced (136 decreased and 3 increased). Although these genes showed large GAF knock-down effects on promoter-proximal polymerase, there were many more genes with moderate non-significant changes to promoter-proximal reads. We also examined the change in gene body reads between LacZ-RNAi and GAF-RNAi libraries. The GAF-RNAi library had 97 genes with gene body read levels significantly different (21 increased and 76 decreased, p-value<0.01) from LacZ-RNAi (Figure 3.3D, red points). The direction of gene body read change for genes with significantly different prmoter-proximal reads correlated with the direction of promoter-proximal read change (Figure 3.3E, red points ρ=0.403). These results extend the role of GAF in maintaining levels of Pol II on the 5' end of genes beyond *Hsp70*, and

suggest GAF plays a role in establishing or maintaining the levels of paused polymerase.

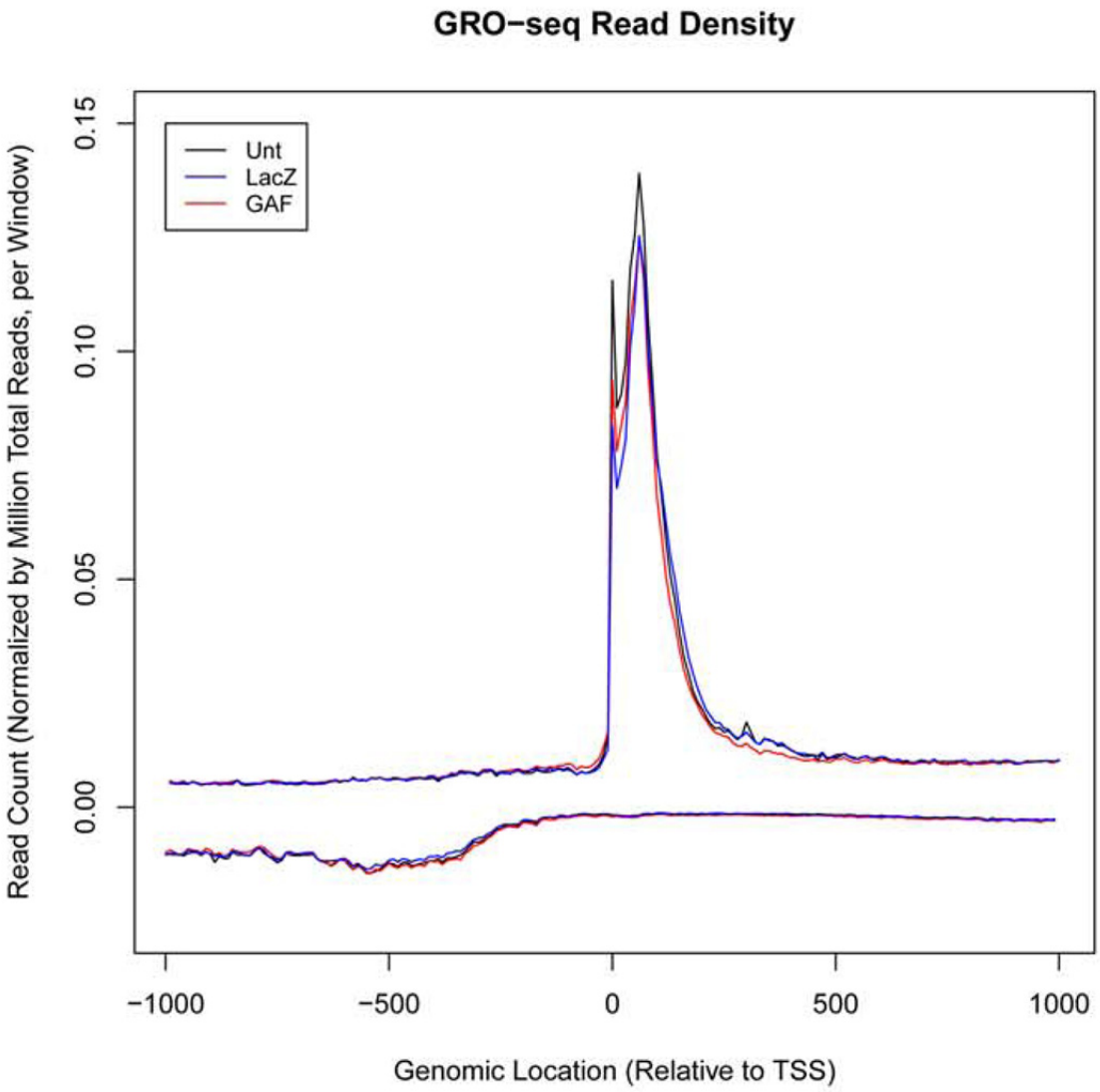


Figure 3.2: GRO-seq read density shows a peak of engaged polymerase on the 5' end of genes. GRO-seq reads from libraries for all RefSeq genes +/-1Kb relative to the TSS binned by 10bp and averaged per gene. The reads from the sense strand are plotted above zero and the reads from the anti-sense strand are plotted below zero. The x-axis indicates the position relative to the TSS, and the y-axis indicates reads per million mappable reads in the library per 10bp window.

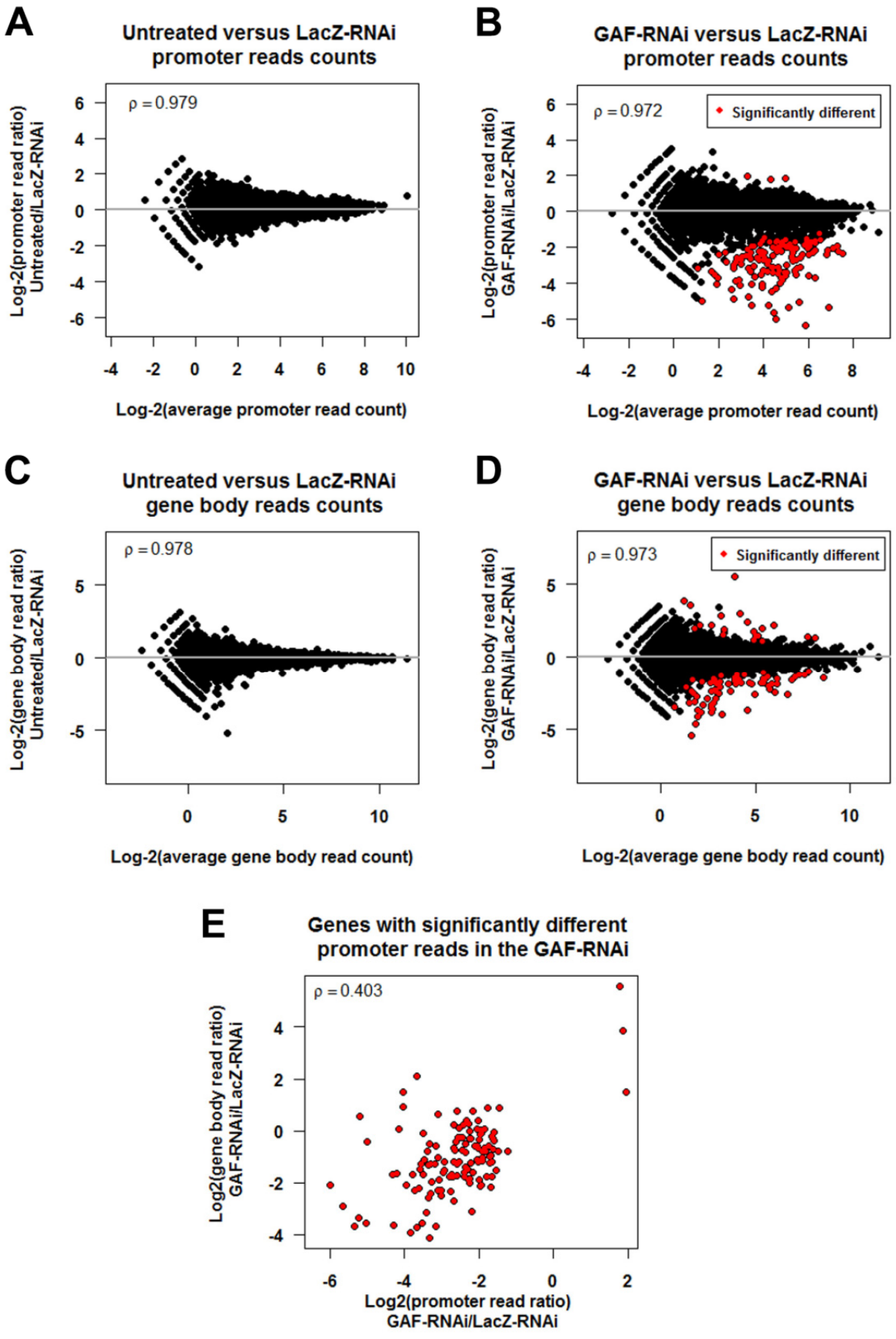

A
Untreated versus LacZ-RNAi
promoter reads counts
ρ = 0.979
Log-2(promoter read ratio)
Untreated/LacZ-RNAi
Log-2(average promoter read count)
B
GAF-RNAi versus LacZ-RNAi
promoter reads counts
ρ = 0.972
Significantly different
Log-2(promoter read ratio)
GAF-RNAi/LacZ-RNAi
Log-2(average promoter read count)
C
Untreated versus LacZ-RNAi
gene body reads counts
ρ = 0.978
Log-2(gene body read ratio)
Untreated/LacZ-RNAi
Log-2(average gene body read count)
D
GAF-RNAi versus LacZ-RNAi
gene body reads counts
ρ = 0.973
Significantly different
Log-2(gene body read ratio)
GAF-RNAi/LacZ-RNAi
Log-2(average gene body read count)
E
Genes with significantly different
promoter reads in the GAF-RNAi
ρ = 0.403
Log2(gene body read ratio)
GAF-RNAi/LacZ-RNAi
Log2(promoter read ratio)
GAF-RNAi/LacZ-RNAi

Figure 3.3: GAF depletion changes polymerase levels on a subset of genes. (A-B) Promoter-proximal reads (100bp window with the most reads +/-250bp of the TSS) and gene body (500bp downstream of the TSS to the polyadenylation site) for each gene are plotted comparing the ratio of Untreated reads to LacZ-RNAi reads versus the average Untreated and LacZ-RNAi reads for each region, respectively. **(C-D)** Promoter-proximal and gene body reads for each gene are plotted comparing the ratio of GAF-RNAi reads to LacZ-RNAi reads versus the average GAF-RNAi and LacZ-RNAi reads for each region, respectively. Significantly different read counts between libraries are indicated by the red points (p-value<0.01). **(E)** The change in promoter-proximal and gene body reads represented as log-2 of the GAF-RNAi to LacZ-RNAi ratio.

3.2.3 Genes with GAF bound promoters are enriched for promoter-proximal pausing

To evaluate whether GAF was important for promoter-proximal pausing, we determined the GAF binding sites genome-wide. Although two recent studies have examined GAF binding genome-wide (Kharchenko et al., 2011; C. Lee et al., 2008) using ChIP-chip (hybridization of immunoprecipitated DNA to a oligo microarray chip), ChIP-seq (massive parallel sequencing of immunoprecipitated DNA) generally has higher signal-to-noise ratios and more defined regions of enrichment than ChIP-chip (Ho et al., 2011). Sumeet Sharma and Michael Guertin from our lab performed ChIP-seq for GAF in S2 cells to determine GAF binding in high resolution (Sharma in prep). Both control (mock RNAi) and GAF-RNAi NHS ChIP material was immunoprecipitated with and without (mock IP) the affinity purified Anti-GAF antibody. To ensure the resulting peaks represented true GAF-binding sites, peaks were selected using the following criteria: 1) the peak reads were reduced in the GAF-RNAi ChIP-seq, 2) a GAGA element was contained within the peak, and 3) the peak was located within 500bp upstream of a TSS. From this, 758 genes had RNAi-sensitive peaks with GAGA elements within 500bp upstream of the TSS of genes that we designated as high confidence GAF bound genes (hc-GAF).

To examine the role of GAF in pausing, paused genes were called based on a significant enrichment of read density in the promoter compared to the gene body read density(Fisher exact test p-value<0.01). The 758 hc-GAF genes are significantly enriched for paused genes as compared to other actively-transcribed genes (92.7% versus 77.8%, fisher exact p-value<0.005). In agreement with this, hc-GAF genes have

higher levels of promoter reads than non-hc-GAF genes (Figure 3.4A). Since there are differences in the level of transcription between the hc-GAF and the non-hc-GAF genes, genes were divided into quartiles of expression based on their gene body read counts in the Untreated library, and then split into hc-GAF and non-hc-GAF genes. The hc-GAF genes had higher promoter reads than non-hc-GAF genes in each quartile (Figure 3.4B).

We next used the hc-GAF genes to evaluate the effect of GAF knock-down on polymerase levels. Of the 139 genes with significantly different promoter reads in GAF-RNAi, 117 were hc-GAF genes (blue points, Figure 3.5A), and upon closer inspection, 19 of these 22 non-hc-GAF genes had some indication of GAF binding on their promoter (data not shown). This indicates that nearly all of the effects were on primary targets of GAF, and the number of genes affected due to RNAi secondary effects was probably low.

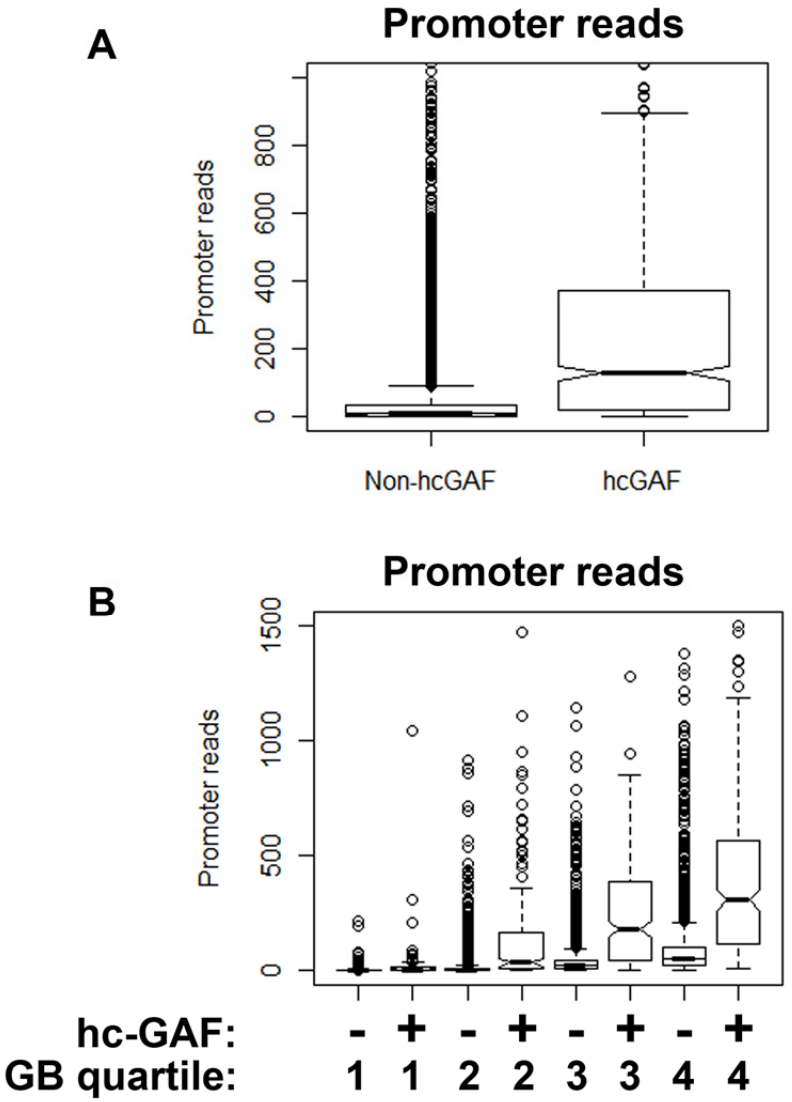


Figure 3.4: GAF-bound genes are enriched for paused genes. (A) Boxplot of the promoter-proximal read counts for high-confidence GAF-bound (hcGAF), and all other genes (Non-hcGAF). **(B)** Boxplot of the promoter-proximal read counts divided in transcription level quartiles based on gene body read density (1=lowest and 4=highest) and then split by high-confidence GAF-bound (hcGAF), and all other genes (Non-hcGAF). The lower and upper edges of the box marks the first and third quartiles of the distribution, respectively, the thick line within each box marks the median, the whiskers mark 1.5 times the interquartile range (first to third quartile) outside the box, and the open circles marks any points lying outside the whiskers.

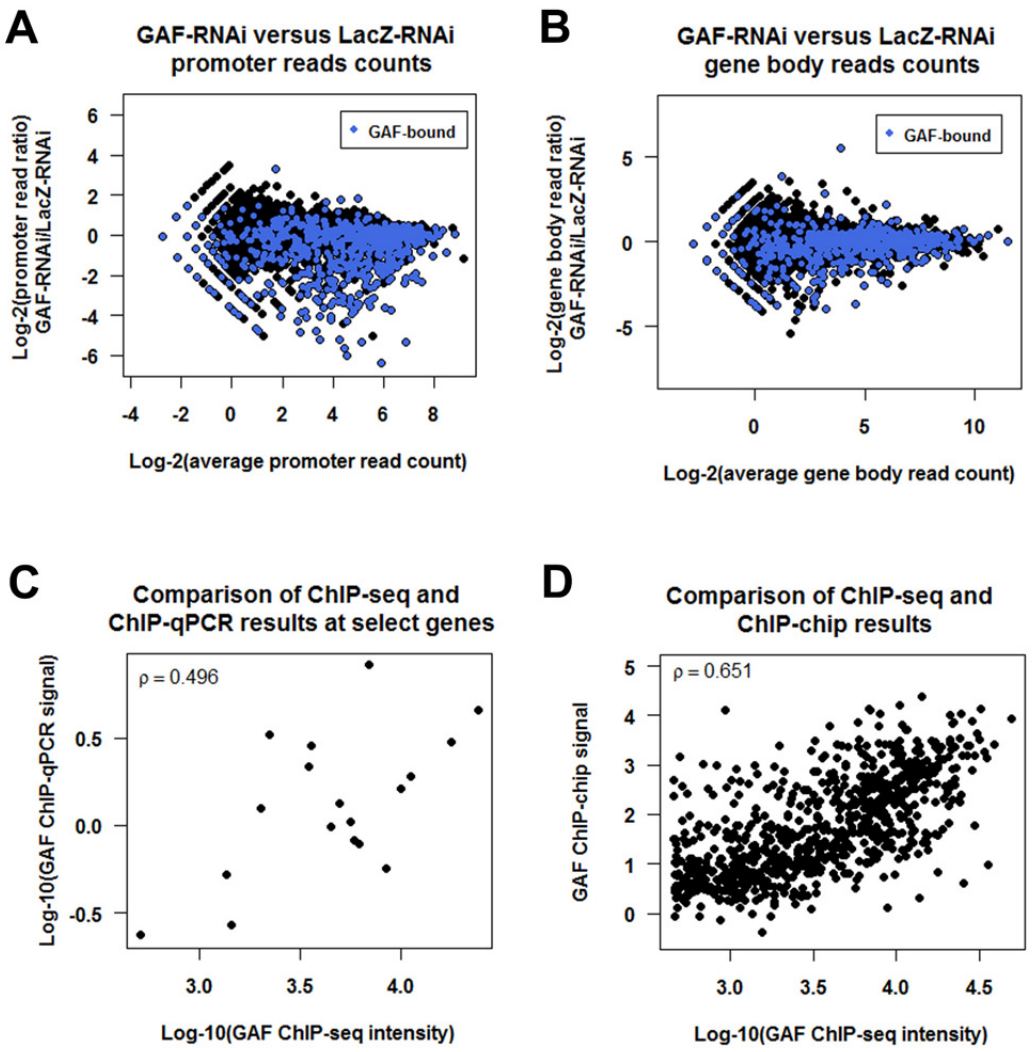


Figure 3.5: Significantly different genes are enriched for GAF-bound genes. (A) Promoter proximal reads (100bp window with the most reads +/-250bp of the TSS) for each gene are plotted comparing the ratio of GAF-RNAi reads to LacZ-RNAi reads versus the average GAF-RNAi and LacZ-RNAi read counts for each region. **(B)** Gene body reads (500bp downstream of the TSS to the polyadenylation site) for each gene are plotted comparing the ratio of GAF-RNAi reads to LacZ-RNAi reads versus the average GAF-RNAi and LacZ-RNAi reads for each region. High-confidence GAF-bound genes are indicated by the blue points. **(C)** The level of GAF on selected genes determined by ChIP-qPCR is plotted compared to GAF levels determined by ChIP-seq read counts in GAGA element-containing peaks. **(D)** The level of GAF determined by ChIP-chip (Kharechenko Nature 2011) is plotted compared to GAF levels determined by ChIP-seq read counts in GAGA element-containing peaks.

3.2.4 Differential effects of GAF knock-down are correlated with GAF binding

Although the vast majority of genes with significantly different promoter reads in GAF-RNAi library were hc-GAF genes, most hc-GAF genes had comparable levels of promoter reads between control and GAF-RNAi libraries (Figure 3.5A). To investigate if these differential effects are due to differences in GAF binding on the genes, the levels of GAF were determined based on the ChIP-seq read counts in the promoter peaks containing GAGA elements. The ChIP-seq levels correlated well with the levels determined by ChIP qPCR from a select set of genes (ρ=0.496, Figure 3.5C) and the genome-wide ChIP-chip levels determined by the modENCODE consortium (ρ=0.651, Figure 3.5D) (Kharchenko et al., 2011). First, I investigated whether the differential effects were due to differences in the reduction of GAF on the genes in GAF-RNAi cells. The genes with non-significant and significant promoter read changes had similar GAF reductions (black and red line, respectively), indicating the differences are not due to some genes being more resistant to the effects of knock-down than others (Figure 3.6A). Although all of the genes are bound by GAF, there is a large range in the level of GAF binding that relate to functionality of GAF at these genes. The genes with significant promoter read changes (red line) had higher levels of GAF binding (Figure 3.6B). In spite of this, many of the unaffected genes (black line) had comparably high levels of GAF, indicating GAF level does not completely explain the differential effects (Figure 3.6B). Additionally, GAF function may be dependent on its distance from the TSS. Therefore, I examined the distribution of GAGA elements on the hc-GAF genes. Although the distance between the GAGA elements and the TSS was comparable for the genes with non-significant and significant promoter read changes (black and red

lines, respectively), the GAGA elements on the genes with significant promoter read changes had a narrower distribution (Figure 3.6C). These results suggest that some the differential effects on GAF-bound genes may be due to differences in the level and position of GAF binding.

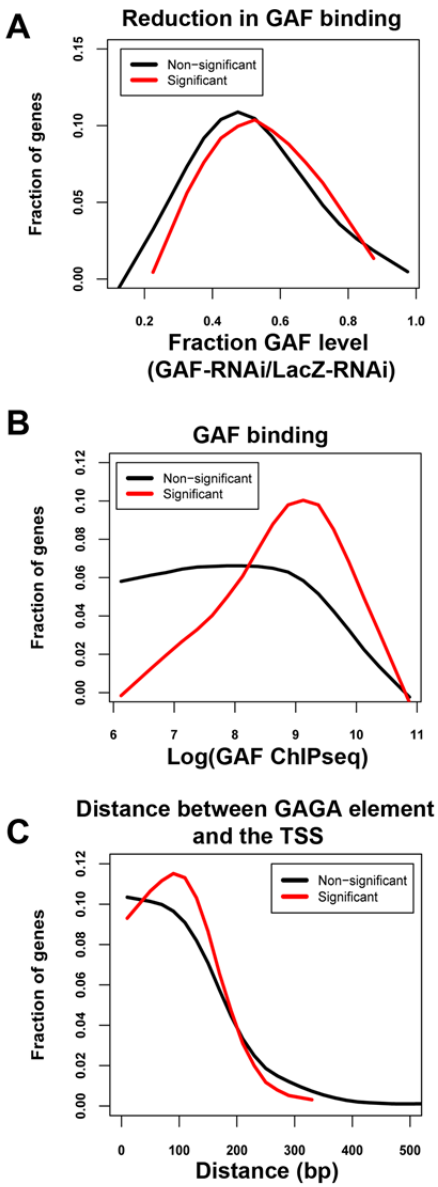


Figure 3.6: GAF binding is similar for affected and unaffected genes. (A) The change in promoter-proximal reads represented by log of the ratio of GAF-RNAi to LacZ-RNAi reads plotted compared to GAF reduction represented by ratio of GAF-RNAi to mock RNAi GAF ChIP-seq read counts in GAGA element-containing peaks. Genes with significantly different promoter-proximal reads are indicated by red points. **(B)** The change in promoter-proximal reads represented by log of the ratio of GAF-RNAi to LacZ-RNAi reads plotted compared to GAF level represented by GAF ChIP-seq read counts in GAGA element-containing peaks. **(C)** The cumulative distribution of GAGA element to TSS distance for high-confidence GAF-bound genes with significantly different promoter-proximal reads (red points) and all other high-confidence GAF-bound genes (black points).

3.2.5 Histone and GTF levels on promoters correlate with GAF knock-down effects

GAF has been shown to affect promoter accessibility through interactions with nucleosome remodelers (Kerrigan, Croston, Lira, & Kadonaga, 1991; Okada & Hirose, 1998; Toshio Tsukiyama et al., 1994). To investigate whether the differential effects of GAF knock-down are due to changes in promoter accessibility, I performed ChIP for histone H3 and GTFs (TFIIA and TFIIB) on promoters of several genes with varying degrees of promoter read reduction in GAF-RNAi. As controls, two regions without genes (102D and 87C backgrounds) had high histone H3 levels and no GTF binding, *Hsp83* (a gene not bound by GAF) did not have any changes in histone H3 or GTF levels in GAF-RNAi cells, and none of the unaffected genes had any differences between control and GAF-RNAi cells in either histone H3 or GTF levels (Figure 3.7A-C). The severity of the promoter read reduction generally correlated with the levels of GTFs and histone H3. The genes with significantly reduced promoter reads had the highest levels of histone H3 (Figure 3.7A) and lowest levels of GTFs on their promoters (Figure 3.7B-C). This may reflect the fact that more of the genes with significantly reduced promoter reads have lower levels of polymerase (Figure 3.3B), and suggests that GAF is necessary to maintain the polymerase levels on these genes. Several genes with reduced promoter reads showed a modest increase histone H3 levels upon GAF knock-down (Figure 3.7A). To further identify changes in promoter accessibility, I examined the levels of the GTFs, TFIIA and TFIIB. Although a few genes with reduced promoter reads had reduced GTF levels, most genes had little or no change in GTF levels (Figure 3.7B-

C). Thus, changes in promoter-proximal polymerase levels are not due to reductions in GTF levels, and this indicates changes in promoter accessibility only explain a subset of the GAF-RNAi effects.

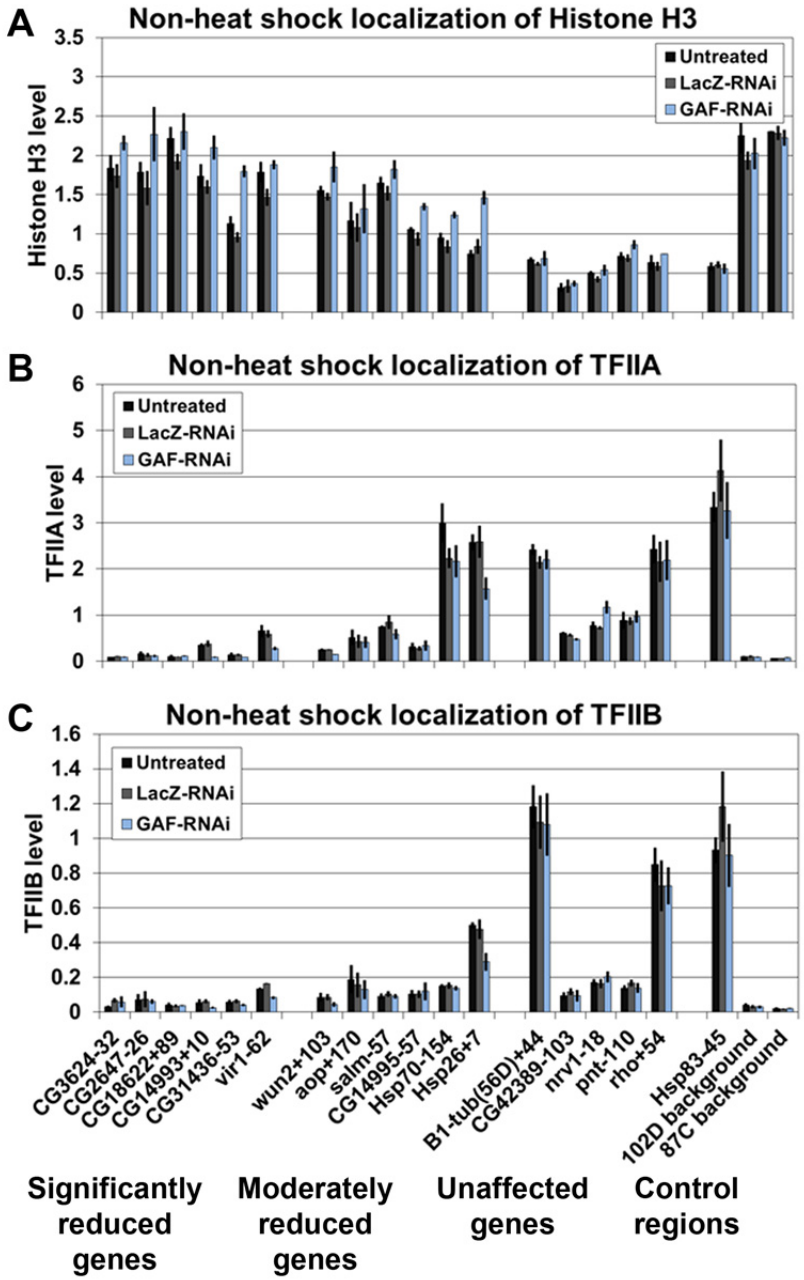
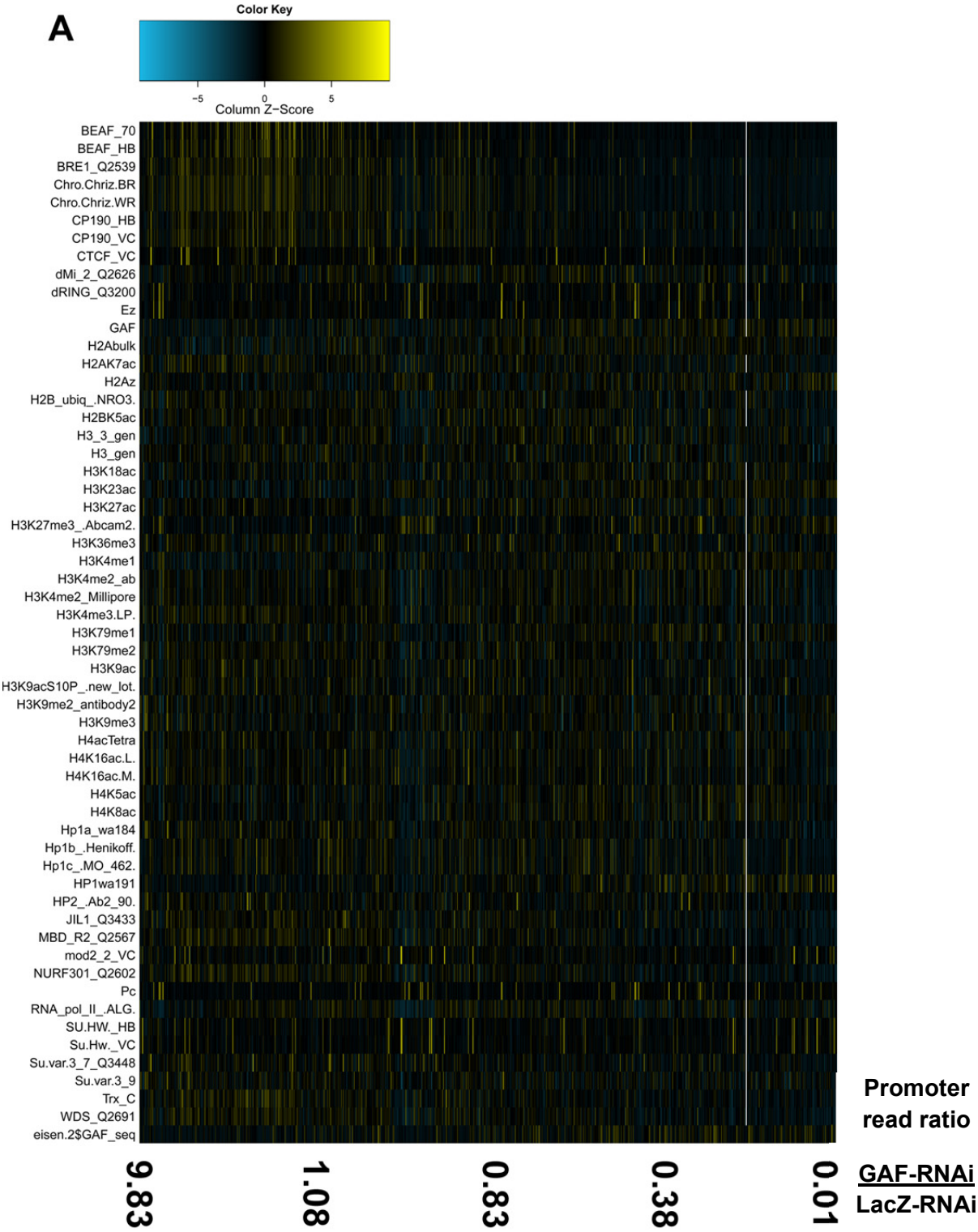


Figure 3.7: Genes with significantly different promoter reads have higher histone H3 and lower GTF on their promoters. (A) ChIP-qPCR for histone H3 on the promoter of select significantly reduced, moderately reduced, and unaffected GAF-bound genes. **(B)** ChIP-qPCR for TFIIA on the promoter of select significantly reduced, moderately reduced, and unaffected GAF-bound genes. **(C)** ChIP-qPCR for TFIIB on the promoter of select significantly reduced, moderately reduced, and unaffected GAF-bound genes. The genes and the center of each primer set relative to the TSS are indicated below. Error bars represents the SEM from at least 3 experiments.

3.2.6 Insulator-associated factors are enriched on promoters of unaffected genes

The modENCODE consortium has generated genome-wide ChIP-chip datasets for many chromatin bound factors and histone marks in various model organisms, including Drosophila (Kharchenko et al., 2011). This information can be used to correlate the various marks and factors with particular effects to achieve a better understanding of the underlying mechanisms (Guertin & Lis, 2010). I used the modENCODE datasets to examine the enrichment or depletion of various factors and chromatin marks on promoters of the hc-GAF genes. The promoter levels of each factor on hc-GAF genes were determined and results were displayed in the form of a heat map (Figure 3.8). The levels of several factors were higher in unaffected genes (BEAF32, Chriz, CP190, MBD-R2, WDS, dMi-2, and NURF 301), and a larger fraction of unaffected genes were enriched for these factors (Figure 3.9). MBD-R2 and WDS are subunits of the transcription regulator nonspecific lethal complex (NSL) (Mendjan et al., 2006; Raja et al., 2010). dMi-2 and Nurf 301 are both nucleosome remodelers (Murawska et al., 2008; T Tsukiyama & Wu, 1995), and Nurf 301 interacts with GAF (T Tsukiyama et al., 1994). BEAF, Chriz, and CP190 are known to bind with GAF to insulators (Nègre et al., 2010). Since insulators have been shown to block transcription factor actions when placed between the transcription factor binding site and the TSS, I determined the location of the BEAF binding sites relative to the GAGA element and the TSS. The DNA sequence motif bound by BEAF-32 has been identified in previous studies (Cuvier, Hart, & Laemmli, 1998; Nègre et al., 2010). I identified the location of BEAF-32 motifs around the hc-GAF genes. Only about 4% of unaffected genes had a

BEAF motif in between the GAGA elements and the TSS, indicating these insulator factors were not preventing GAF knock-down from affecting the gene through enhancer-blocking activity.


A
Color Key
−5
0
5
Column Z−Score
BEAF_70
BEAF_HB
BRE1_Q2539
Chro.Chriz.BR
Chro.Chriz.WR
CP190_HB
CP190_VC
CTCF_VC
dMi_2_Q2626
dRING_Q3200
Ez
GAF
H2Abulk
H2AK7ac
H2Az
H2B_ubiq_.NRO3.
H2BK5ac
H3_3_gen
H3_gen
H3K18ac
H3K23ac
H3K27ac
H3K27me3_.Abcam2.
H3K36me3
H3K4me1
H3K4me2_ab
H3K4me2_Millipore
H3K4me3.LP.
H3K79me1
H3K79me2
H3K9ac
H3K9acS10P_.new_lot.
H3K9me2_antibody2
H3K9me3
H4acTetra
H4K16ac.L.
H4K16ac.M.
H4K5ac
H4K8ac
Hp1a_wa184
Hp1b_.Henikoff.
Hp1c_.MO_462.
HP1wa191
HP2_.Ab2_90.
JIL1_Q3433
MBD_R2_Q2567
mod2_2_VC
NURF301_Q2602
Pc
RNA_pol_II_.ALG.
SU.HW._HB
Su.Hw._VC
Su.var.3_7_Q3448
Su.var.3_9
Trx_C
WDS_Q2691
eisen.2$GAF_seq
9.83
1.08
0.83
0.38
0.01
Promoter read ratio
GAF-RNAi
LacZ-RNAi

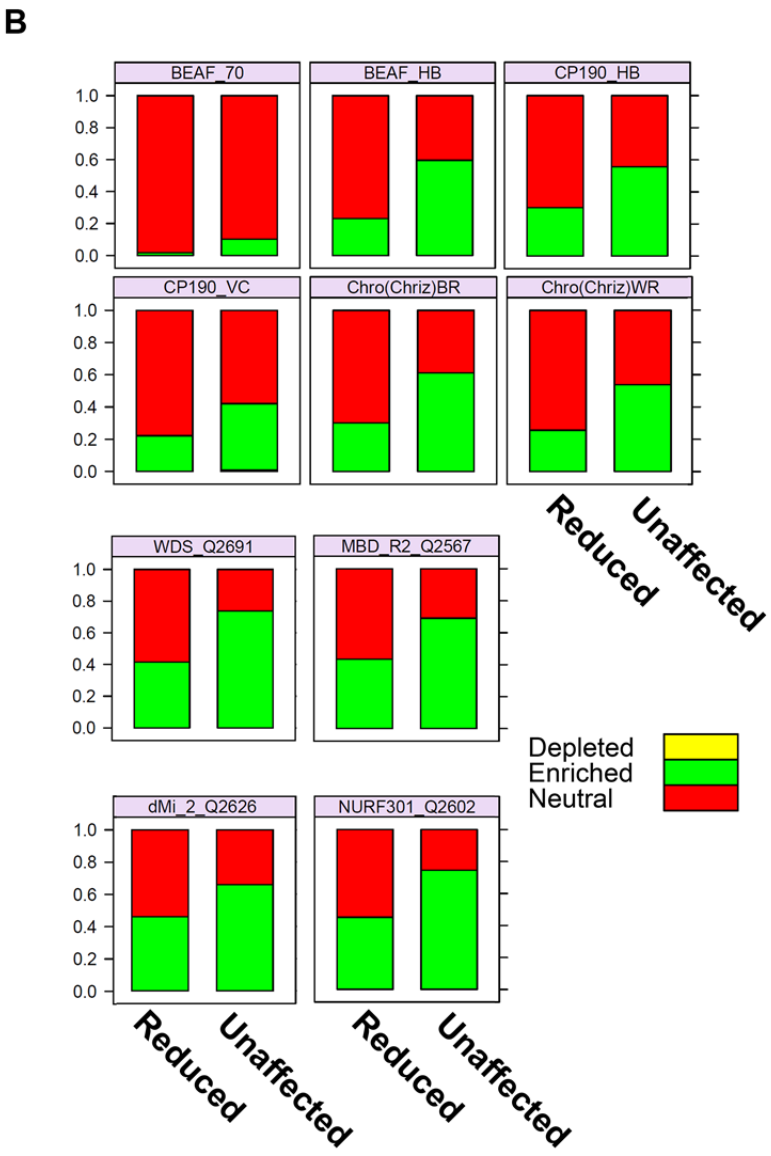


Figure 3.8: Levels of insulator-associated factors are highest on unaffected genes. (A) ChIP-chip levels of chromatin factors and histone modifications on GAF-bound genes. Genes are arranged in rows by increasing GAF-RNAi to LacZ-RNAi promoter reads ratio indicated to the right of the heat map. Columns represent the average microarray intensity of all the probes within 500bp upstream of the TSS for each factor or histone modification (yellow indicates higher levels and blue indicates lower levels). **(B)** Fraction of genes with significantly reduced promoter reads (left bar charts) and gene with unaffected promoter reads (right bar charts). Green bar indicates the fraction of genes within each set with significantly enriched probes, yellow bar indicates the fraction of genes within each set with significantly depleted probes, and red bar indicates the fraction of genes within each set without enriched or depleted probes.

3.3 Discussion

Although a correlation between GAF binding and pausing has been well demonstrated (Hendrix et al., 2008; C. Lee et al., 2008), the direct evidence for its role in pausing is limited. Previous studies have shown the GAGA elements are important for pausing on reporter genes containing the *Hsp70* core promoter (H. Lee et al., 1992; Wang et al., 2005). To investigate the role of GAF in pausing, I depleted GAF in S2 cell culture and examined its effects on Pol II distribution. The paused polymerase on NHS *Hsp70* is reduced 2-fold in GAF-RNAi cells (Figure 3.1C), indicating GAF is affecting pausing on the endogenous *Hsp70* genes. The effect of GAF depletion on the distribution of transcriptionally-engaged Pol II genome-wide was examined using GRO-seq. GAF depletion significantly changed promoter-proximal polymerase levels on subset of GAF bound genes, and virtually all of these genes had reduced polymerase levels (Figure 3.5A). Furthermore, the change in promoter-proximal polymerase was accompanied by a similar change on the gene body (Figure 3.3E). This correlation of polymerase changes is in stark contrast to the effects of Spt5 and NELF depletion, where many of the genes with reduced promoter-proximal Pol II increase in expression (Gilchrist et al., 2008; Rahl et al., 2010, Chapter 6 of this dissertation), and indicates GAF plays a role in establishing pausing rather than maintaining the pause, like NELF and Spt5/Spt4.

GAF depletion had a range of effects on GAF bound genes. Although these differential effects do correlate with differences in GAF binding (Figure 3.6), levels of GAF binding do not fully explain the differential effects because there are many unaffected genes with comparably high GAF levels. ModENCODE ChIP data showed the insulator-associated factors BEAF-32, CP-190, and Chriz were all enriched on

unaffected genes (Figure 3.9). This indicates insulators are protecting these genes from the effects of reduced GAF binding. In addition, other unknown transcription factors may function redundantly with GAF to maintain paused polymerase levels.

Previous studies have indicated GAF has a role in maintaining promoters in a chromatin conformation favorable to expression. The nucleosome-free promoter allows sequence-specific transcription factors and GTFs to bind and control gene expression. I assessed promoter accessibility using ChIP for histone H3 and the GTFs TFIIA and TFIIB at select genes with significant, moderate, or no change in promoter-proximal polymerase levels. Genes with significantly reduced polymerase had the highest levels of histone H3, and conversely, GTF levels were lowest on these genes (Figure 3.7). Many of the largest reductions in promoter reads occur on genes with moderate levels of polymerase, and this suggests GAF is particularly important for maintain levels of promoter-proximal polymerase on these genes. GAF knock-down changed both histone H3 and GTF levels on only a few genes, indicating nucleosomes repositioning is not "closing" most promoters of affected genes in GAF depleted cells. Interestingly, one of genes with changes in both histone and GTF levels is *Hsp26*. *Hsp26* has a positioned nucleosome centered about 250bp upstream of its TSS, and its positioning is dependent on GAF binding (Thomas & Elgin, 1988; Wall et al., 1995). In contrast, the paused polymerase on *Hsp70* is reduced 2-fold (Figure 3.1), but a previous study has shown there is no rearrangement of nucleosomes on the NHS *Hsp70* promoter upon GAF knock-down (Petesch & Lis, 2008). These results demonstrate changes in promoter nucleosome distribution can only explain the promoter-proximal polymerase change on a subset of genes. Consistent with this, there is evidence that GAF can also

more directly affect transcription. GAF and TFIID physically and genetic interact (Chopra et al., 2008), and the polyQ domain can function as an activator both in vitro and in vivo (Vaquero et al., 2000; Vaquero et al., 2008). Regardless of the mechanism, these results have demonstrated GAF does have a role in establishing promoter-proximal pausing at many of its target genes.

CHAPTER 4

POSITIONING OF DOWNSTREAM PROMOTER SEQUENCE DETERMINES PAUSE SITE SELECTION AND PAUSING LEVEL

4.1 Introduction

Until recently, the exact site of polymerase pausing was known for very few genes (Giardina, Perez-Riba, & Lis, 1992; Krumm et al., 1992; Mirkovitch & Darnell, 1992; Rasmussen & Lis, 1993; Rougvie & Lis, 1988). Our lab mapped the positions for the paused polymerases on a few Drosophila genes in single nucleotide resolution based on the lengths of their nascent transcripts (Rasmussen & Lis, 1993). These experiments indicated pausing was not uniform across the 5' region, but instead, polymerases occupied defined positions within the region. On the genes analyzed, polymerases paused in two sites spaced by about one turn of the helix. In particular, the paused polymerases on *Hsp70* occupied either of two pause sites, a narrow focused peak of polymerase around +21 or a broader peak around +30.

These distinct sites of paused polymerase may have implications on the mechanics of promoter-proximal pausing. The spacing of these peaks locates them on the same side of the DNA helix, and suggests the polymerases may be interacting with a factor bound to a specific site on the DNA (Rasmussen & Lis, 1993). One set of attractive candidate factors is the general transcription factors (GTFs) that are bound to the promoter. Some GTFs interact with specifically spaced DNA elements in the promoter to form the pre-initiation complex (PIC) with Pol II and aid initiation (Juven-Gershon & Kadonaga, 2010). Subunits of TFIID can interact with several different

promoter elements. TATA binding protein (TBP) binds to the TATA element located about 30bp upstream of the transcription start site (TSS). TAF1 interacts with initiator (Inr) elements located at the TSS (Metcalf & Wassarman, 2006), and TAF1, TAF6, and TAF9 can interact with the DPE (downstream core promoter element) around 30bp downstream of the TSS (Burke & Kadonaga, 1997; D.-H. Lee et al., 2005; Zhang, Wu, & Gilmour, 2004). Detailed information of the mechanisms surrounding initiation and promoter escape has suggested that interactions between polymerases and some GTFs are broken upon promoter escape (Fuda, Ardehali, & Lis, 2009), but it remains possible that interactions between promoter elements and GTFs and between GTFs and Pol II still anchor the complex to the TSS. The large TFIID complex could provide enough flexibility to allow substantial transcription (Saunders et al., 2006).

A recent genome-wide study indicated pausing is dependent on the sequence being transcribed (Nechaev et al., 2010). Short 5' capped nuclear RNAs derived from the paused polymerases were isolated and sequenced. The positions of paused polymerases were inferred from the 3' ends of these RNAs. These positions correlated with regions where the RNA-DNA hybrid would be relatively weak compared to neighboring sequences. These characteristics lead to pausing and backtracking of RNA polymerases in vitro (Herbert et al., 2006; Komissarova & Kashlev, 1997; Tadigotla et al., 2006). Thus, this study suggests that it is the transcribed sequence specifying the location of pausing.

These two possible determinants of pausing can be tested by altering the orientation of the pause sites relative to the TSS, while maintaining the sequence. To do this, I created transgenes with altered the spacing between the promoter and pause

sites by inserting 5 or 10 bases downstream of the TSS. Using a new method developed in our laboratory, the polymerases were then mapped with nucleotide resolution to determine the pausing sites on the transgenes.

4.2 Results

4.2.1 Strategy to identify determinants specifying sites of pausing.

To distinguish between sequence and orientation, the spacing between the TSS and pause sites was varied by creating three *Hsp70* genes: an unchanged wild type version (Hsp70wt), a version with 5 bases inserted at +10 (Hsp70+5), and a version with 10 bases inserted at +10 (Hsp70+10) (red lettering in Figure 4.1). The +10 position was chosen to avoid mutating the Inr or downstream sequences that are important for full expression of *Hsp70* (C.-H Wu et al., 2001), and the insertion duplicated the *Hsp70* sequence in order to keep any possible elements intact. The Hsp70+5 and Hsp70+10 transgenes shift the pause site sequence relative to the upstream promoter elements by a half or full turn of the DNA, respectively (asterisks in Figure 4.1). If the locations are being defined solely by the sequence being transcribed, the paused polymerase locations will shift downstream by the same amount as the insert size. In contrast, a result where the paused polymerase location does not change relative to the TSS may mean the pause locations are being defined by interactions with some upstream promoter-bound factors. Of course, both may be important, and more complicate alteration to the pattern of pausing may occur.

A previous study from our lab investigated the role of promoter elements in specifying pausing using multiple P-element insertion lines for each transgene to control

for the different insertion sites in each line (H. Lee et al., 1992). Additionally, they placed the YP1 gene downstream to the modified promoters to distinguish the trangene from the endogenous *Hsp70* genes. Since YP1 is transcriptionally silent in male flies, they hybridized the run-on RNAs, performed with adult male nuclei, to a probe for YP1 to determine the level of paused polymerase on the transgenes.

I attempted to simplify the experiment by taking advantage of two recent advances. First, a technique utilizing bacteriophage ΦC31 integrase was developed that allows constructs to be integrated a specific locations within the Drosophila genome (Venken & Bellen, 2005). Using this system, all of my constructs were integrated into the 22A locus on the X chromosome. Because each construct should be in the same location, position-dependent differences are avoided. Second, a procedure allowing homologous recombination in Drosophila was developed, and used to delete all of the *Hsp70* genes to create an *Hsp70* null fly (Gong & Golic, 2004). Although these flies have reduced thermotolerance, they are perfectly viable and develop normally under standard growth conditions (Gong & Golic, 2006). Therefore, I crossed my transgenic lines into this *Hsp70* null background. This should allow the use of complete *Hsp70* leader region in the transgenes, creating a gene nearly identical to the endogenous genes. After integration of the transgenes, the transgene presence was followed by expression of the wild type *white* gene present in the constructs. All lines had orange colored eyes indicating expression from the constructs was uniform and comparable across the various fly lines.

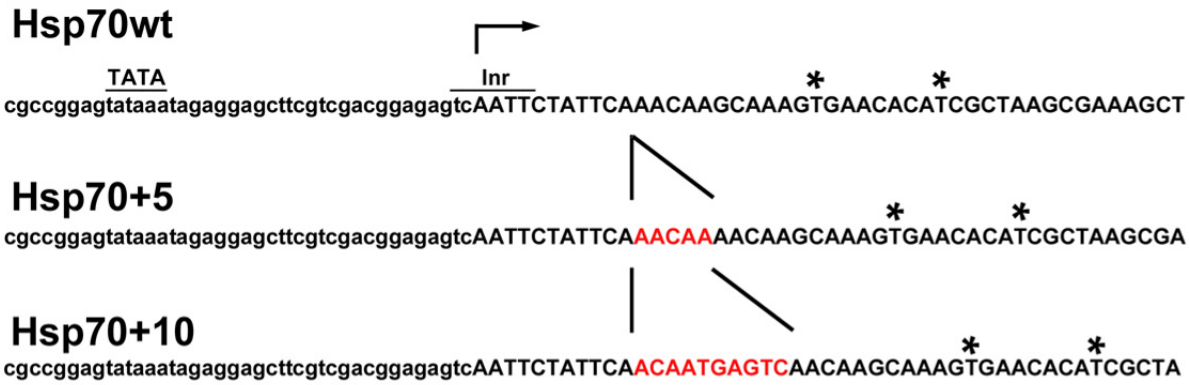
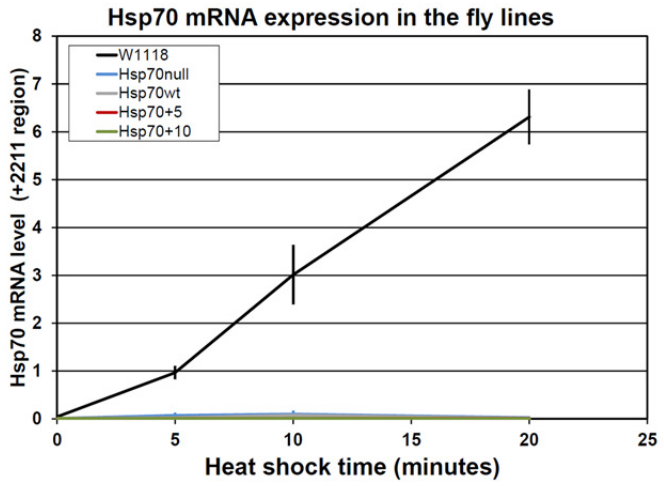


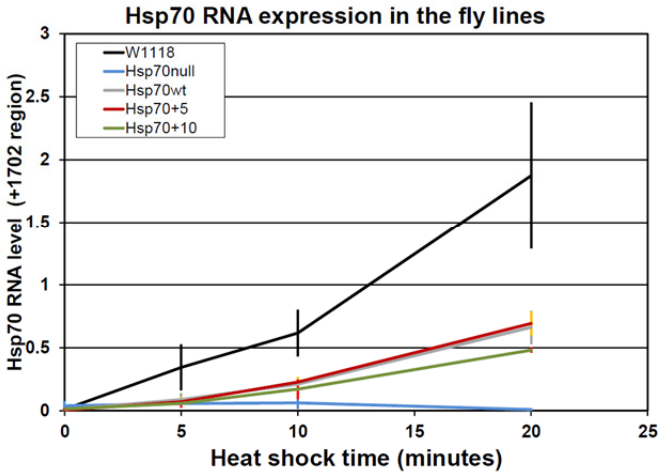
Figure 4.1: Construction of Hsp70 genes with 5 or 10 base pair shifts. Wild type (Hsp70wt), *Hsp70* with 5 bases (red letters) inserted at +10 position (Hsp70+5), or *Hsp70* with 10 bases (red letters) inserted at +10 position (Hsp70+10) genes with the *Hsp70* sequence from -245 to +1863 (relative to the TSS) were cloned to a pCasper plasmid with AttB site. The well-annotated promoter elements are shown (TATA and Inr), the TSS is designated by the arrow, and the predominant endogenous paused sites (Rasmussen & Lis, 1993) are indicated with the asterisks.

To assess the function of each transgene, quantitative RT-PCR was performed on RNA isolated from larvae shifted to 37°C for 60 minutes, and RNA was isolated. My constructs contained the *Hsp70* sequence from the upstream and core promoter (245 bases upstream of the TSS) to the middle of the *Hsp70* gene (1863 base downstream of the TSS), short of the endogenous polyadenylation site. Truncation of the gene provides a method to differentiate between the transgenes and the endogenous *Hsp70* genes. I assayed for mRNA from the endogenous *Hsp70* genes by reverse transcribing the RNA with an oligo(dT) primer and amplifying with a primer set contained only in the endogenous genes. Reverse transcription and quantitative PCR (RT-qPCR) showed that wild type (W1118) flies induced *Hsp70* mRNA to high levels, but *Hsp70* null line and all of the transgene lines did not express any mRNA containing this region (Figure 4.2A). In addition, all lines expressed the small heat shock gene *Hsp26* to similar levels (Figure 4.2C). In contrast, when the RNA was reverse transcribed with a primer to a region contained in the transgenes, all lines expressed RNA with this region, albeit at a lower level than the wild type line due to their lower copy number (Figure 4.2B). This demonstrates the transgenes are transcriptionally similar to the endogenous *Hsp70* genes.

A


Hsp70 mRNA expression in the fly lines
W1118
Hsp70null
Hsp70wt
Hsp70+5
Hsp70+10
Hsp70 mRNA level (+2211 region)
Heat shock time (minutes)

B


Hsp70 RNA expression in the fly lines
W1118
Hsp70null
Hsp70wt
Hsp70+5
Hsp70+10
Hsp70 RNA level (+1702 region)
Heat shock time (minutes)

C

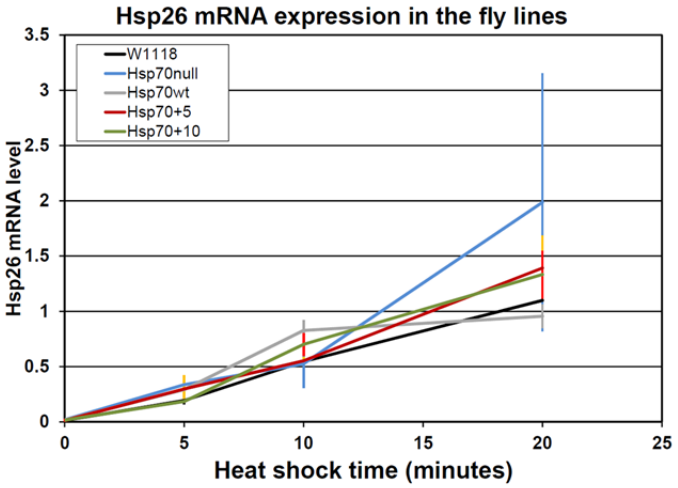

Hsp26 mRNA expression in the fly lines
W1118
Hsp70null
Hsp70wt
Hsp70+5
Hsp70+10
Hsp26 mRNA level
Heat shock time (minutes)

Figure 4.2 Hsp70 fly lines express the transgenes similar to the endogenous Hsp70 genes. (A) RT-qPCR results for total RNA from larvae heat shocked for the indicated times, reversed transcribed with oligo(dT), amplified with Hsp70Ab +2155/+2266 and RpL32+204/+304 primer sets (+2211). **(B)** RT-qPCR results for total RNA from larvae heat shocked for the indicated times, reversed transcribed with the Hsp70+1754 and RpL32+304 reverse primers, amplified with Hsp70Ab+1649/+1754 and RpL32+204/+304 primer sets (+1702), and then Hsp70 values were normalized to the RpL32 values. **(C)** RT-qPCR results for total RNA from larvae heat shocked for the indicated times, reversed transcribed with oligo(dT), amplified with Hsp26 +580+667 and RpL32+204/+304 primer sets.

Terminated run-ons (tROs) can be used to map the polymerase location with high resolution, based on the length of the nascent RNA (Figure 4.3A) (Rasmussen & Lis, 1993). Nuclear run-on assays allow transcriptionally-engaged polymerases from isolated nuclei to further extend their nascent transcripts by the exogenous addition of nucleoside triphosphates (NTPs) to the isolated nuclei. The addition of the detergent sarkosyl prevents initiation by unengaged polymerases and removes any impediments to polymerase elongation. A tRO reaction includes a radiolabeled NTP to detect the RO RNA and a chain terminating NTP, 3'-deoxyribonucleoside (3'-dNTP). The polymerases will extend until they incorporate the 3'-dNTP. Hybridization of the run-on RNAs to a biotinylated complementary oligonucleotide probe is used to pull out the RNA of interest, and the length of radiolabeled RO RNAs are determined by running them on a gel. This technique allows the polymerase location to be determined within several nucleotides.

Terminated run-ons were used to determine the location of pausing on each transgene. The nuclei were isolated from non-heat shocked adult flies, and run-on in the presence of radiolabeled ATP, CTP, and UTP as well as chain terminating 3'-dGTP. The wild type nuclei gave the expected pattern for *Hsp70* (Figure 4.3C, lane 1), transcripts of 21, 23, 32, 39, and 43 nucleotides. Unfortunately, the *Hsp70* null nuclei gave a pattern similar to wild type (Figure 4.3C, lane 1 and 2, respectively). This indicates the presence of regions with homology to the *Hsp70* pause region in the null flies. Tandem repeats with homology to the 5' end of *Hsp70* are present between the Hsp70 genes at the 87C locus. Although these repeats should have been excised by the recombination that excised the *Hsp70* genes, it is possible these repeats are also

present in other regions of the genome. In spite of the presence of these signals in the *Hsp70* null flies, it was interestingly to see that additional longer labeled RNAs were pulled out of Hsp70+5 nuclei (lane 4), suggesting the 5 base shift in sequence results in a shift in pausing. Interestingly, there was little labeled RNA pulled out of the Hsp70+10 nuclei (lane 5), suggesting the 5 base shift in sequence results in dramatically less pausing. As a control, labeled *Hsp26* run-on transcripts were pulled out. Because all of these nuclei have the same size *Hsp26* labeled transcripts, the changes in pausing seen at *Hsp70* are not due to some systematic difference in the run-on (Figure 4.3A).

I investigated the source of the RNAs to determine if I could remove them. After re-crossing the flies, I checked for the presence of the endogenous *Hsp70* sequences. RT-qPCR showed RNA with the *Hsp70* leader sequence was still present, but full-length RNAs were not present (data not shown). The 87C locus contains many repeats of the *Hsp70* 5' end arranged tandemly in between the *Hsp70* genes located at this locus (Hackett & Lis, 1981). Although these should have been excised in the homologous recombination, qPCR from genomic DNA from the *Hsp70* null flies indicated low levels of 5' end sequence, estimated at 2-3 copies in comparison to the 10-15 copies in the wild type lines (data not shown). This suggests that the *Hsp70* null flies contain some of these repeats elsewhere in the genome.

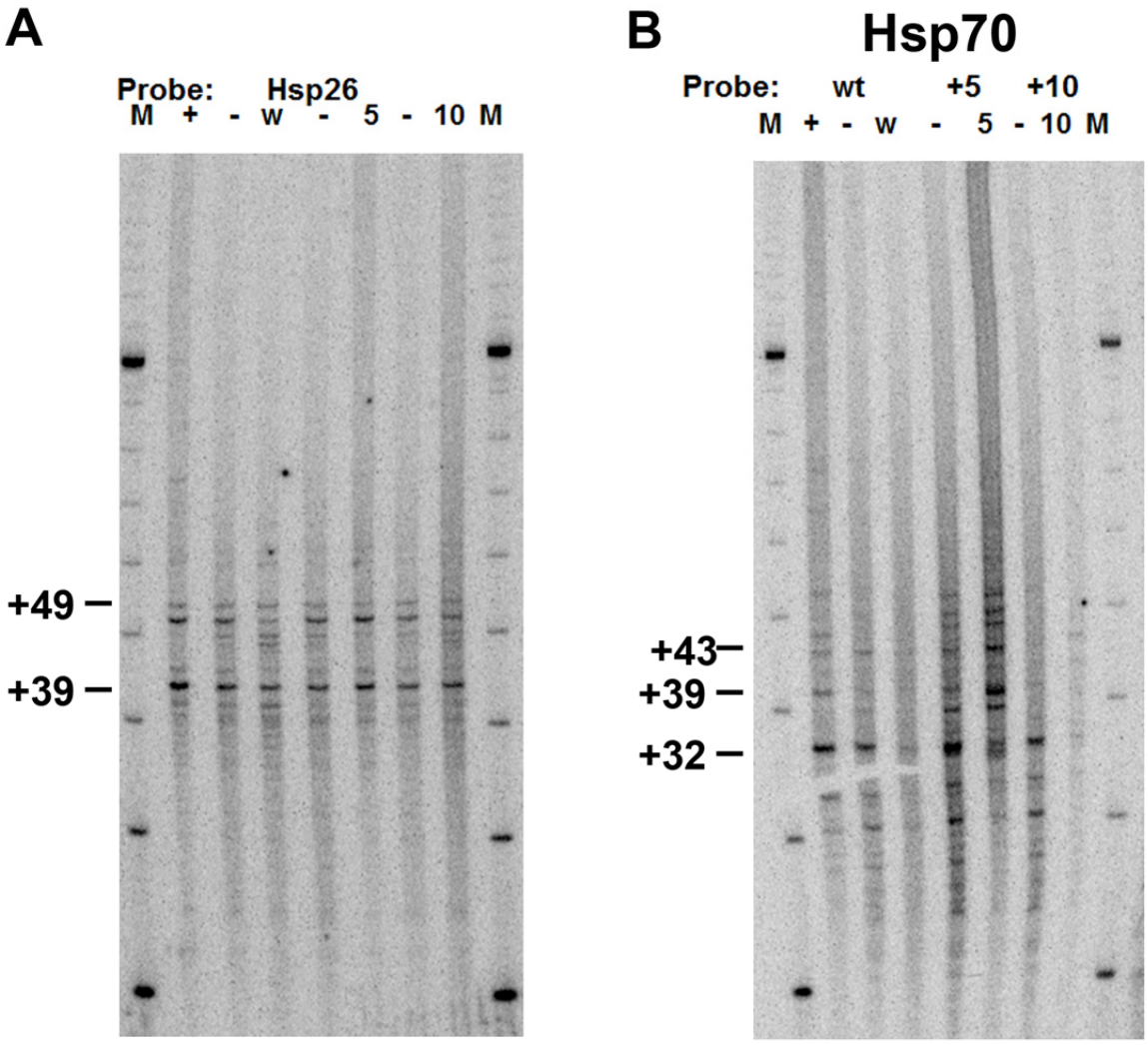


Figure 4.3 Terminated run-ons from Hsp70 null flies have Hsp70 sequences. (A) Terminated run-on performed with nuclei from W1118 (+), Hsp70null (-), Hsp70wt; Hsp70 null (w), Hsp70+5; Hsp70 null (5), Hsp70+10; Hsp70 null (10) adult flies. The Hsp26 run-on RNAs were pulled out with a biotinylated probe complementary to the 5' end of the gene, and run on a 10% PAGE TBE-urea gel. **(B)** The *Hsp70* run-on RNAs were pulled out with a biotinylated probe complementary to the 5' end of the gene, and run on a 10% PAGE TBE-urea gel.

The presence of these repeats in the *Hsp70* null flies prevents tROs from being used to map the polymerase on the transgenes. One solution to this dilemma is to use the sequence of the RNA to distinguish the transgenes with inserts from any RNAs with endogenous *Hsp70* sequences. To do this, we used PRO-seq (Precision Run-On), a new method developed in our lab by Hojoong Kwak. This method performs run-ons with biotinylated nucleotides, which are not easily extended by the polymerase. As a result, the polymerases generally incorporate only one biotinylated nucleotide onto the 3' ends of the nascent transcripts. This allows these RNAs to be specifically and efficiently isolated from all contaminating RNAs, and analogous to tROs, provides nucleotide resolution of polymerases by based on the sequence of 3' end of the biotinylated RNAs.

Hojoong has performed PRO-seq with nuclei from *Hsp70* null and the transgenic lines. He mapped the reads with *Hsp70* sequences from each line. The RNAs from the endogenous *Hsp70* genes in all of the PRO-seq libraries had 5' ends originating from the TSS and the expected 3' end locations around +32 (Figure 4.4B-C). The reads from the Hsp70wt library showed a similar pattern to the pattern to the reads from endogenous genes, indicating the wild type transgene does not have a dramatically different pattern from the endogenous genes (Figure 4.4D). Reads from the Hsp70+5 and Hsp70+10 were mapped to the transgenes. Although there are fewer reads than the endogenous genes, the 5' ends of the Hsp70+5 reads originated from the TSS, but 3' ends reads mapped about 5bp further downstream, around +38 (Figure 4.4E). In contrast, there were 15-fold less reads from the Hsp70+10 transgene, but the 5' ends of the Hsp70+10 reads originated from the TSS and the 3' ends were spread between +20 to +55 (Figure 4.3D-E). .

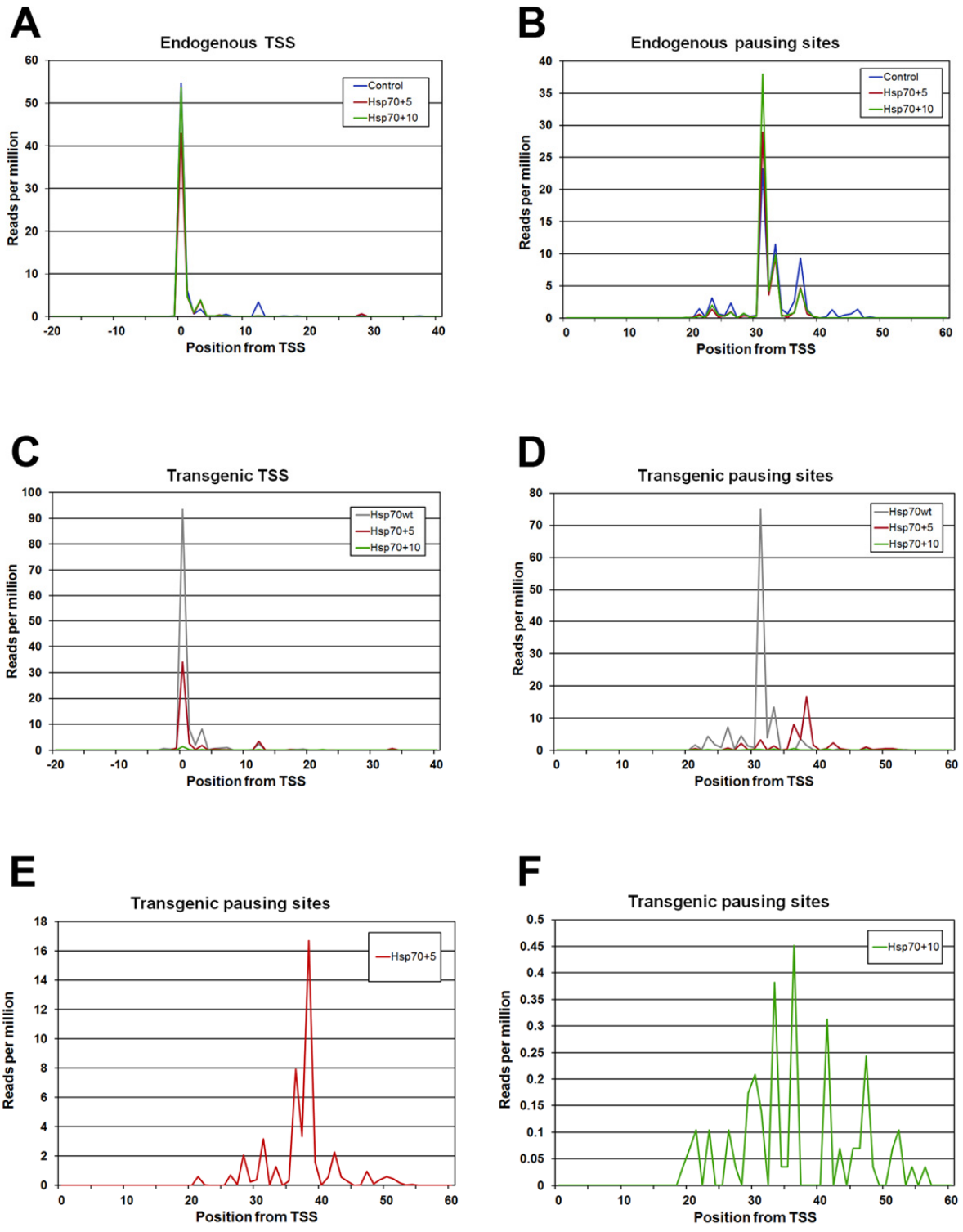

A
Endogenous TSS
Reads per million
Position from TSS
Control
Hsp70+5
Hsp70+10
B
Endogenous pausing sites
Reads per million
Position from TSS
Control
Hsp70+5
Hsp70+10
C
Transgenic TSS
Reads per million
Position from TSS
Hsp70wt
Hsp70+5
Hsp70+10
D
Transgenic pausing sites
Reads per million
Position from TSS
Hsp70wt
Hsp70+5
Hsp70+10
E
Transgenic pausing sites
Reads per million
Position from TSS
Hsp70+5
F
Transgenic pausing sites
Reads per million
Position from TSS
Hsp70+10

Figure 4.4 The Hsp70+5 polymerases pause at the same sequence as on the endogenous *Hsp70* genes. (A) The 5' end of reads mapping to the endogenous *Hsp70* genes. **(B)** The 3' ends of reads mapping to the endogenous *Hsp70* genes. **(C)** The 5' end of reads mapping to the transgenes *Hsp70* genes. (Note: The Hsp70wt reads also contain reads from endogenous *Hsp70* sequences) **(D)** The 3' ends of reads mapping to the transgenes *Hsp70* genes. (Note: The Hsp70wt reads also contain reads from endogenous *Hsp70* sequences) **(E)** The 3' ends of reads from Hsp70+10 run-ons mapping to the Hsp70+5 transgenes. **(F)** The 3' ends of reads from Hsp70+10 run-ons mapping to the Hsp70+10 transgenes.

4.3 Conclusions

The endogenous *Hsp70* genes pause at two distinct sites at +21 and +32, which are located on the same side of the DNA helix. I examined whether the site of pausing was dependent on this orientation using *Hsp70* transgenes that maintain the pause site sequence in the same position (Hsp70wt), shift the sequence to the opposite side of the helix with a 5 base insertion, or shift the sequence one full helical turn with a 10 base shift. Hojoong Kwak performed PRO-seq on nuclei from each line to map polymerase with nucleotide resolution based on the sequence of their nascent transcripts. Although the wild type transgene could not be distinguished from the endogenous *Hsp70*-like sequences, reads from the Hsp70+5 transgene showed the polymerases were located about 5 bases further downstream than polymerase on the wild type genes. Thus, pausing still occurs on the same sequence, in spite of its location on the opposite side of the helix, suggesting either the transcribed sequence specifies the pausing site or there is enough flexibility to allow pausing to occur 5 bases further downstream.

These results demonstrate a specific orientation of sequence on one side of the double helix is not necessary to specify pausing sites. There are lower numbers of reads mapping to the transgenes than the endogenous sequences; the reads mapped to the Hsp70+10 transgene were 15-fold lower than reads mapped to the Hsp70+5 transgene. This dramatic decrease in read numbers between the Hsp70+5 and Hsp70+10 transgenes indicates both orientation and distance of these downstream sequences is critical for determining pausing level. However, the increased distance of the pause sites in the Hsp70+10 transgene does not preclude pausing because the *Hsp26* and *Hsp27* pause sites are about the same distance from the TSS as the

sequence of the *Hsp70* pause sites in the Hsp70+10 transgene (Rasmussen & Lis, 1993). Additionally, the differences in pausing level are probably not due to changes in PIC assembly and initiation because heat shock RNA expression is comparable between the three transgenes (Figure 4.2). There is evidence that perturbations to the region around +30 will affect polymerase levels. Previous experiments using *Hsp70* transgenes showed that mutating the sequence between +23 and +62 reduced in the level of pausing and heat-shock induced expression (H. Lee et al., 1992). In addition, mutations in the downstream regions of *Hsp70* between +24 and +29 moderately reduced transcription in vitro and in vivo (C.-H Wu et al., 2001). The results of this study demonstrate disrupting the spacing of these pausing sites affects the level of pausing.

Further experiments could discriminate between these possibilities and provide a greater understanding of the role of downstream sequence in pausing. Systematic mutation of specific bases within the pause sites could circumvent the problems from the insertions, but the reads would be difficult to distinguish from the endogenous *Hsp70* sequences. This complication could be circumvented by including mutations that distinguish the transgene from the endogenous *Hsp70* sequences without disrupting transcription. The region between +14 and +19 may tolerate mutations because mutation of the region did not affect transcription in vitro (C.-H Wu et al., 2001). Certainly, our understanding of the mechanisms controlling pausing would benefit from additional focused studies into the role of promoter and downstream sequence that can directly test the hypotheses drawn from the recent genome-wide information on pausing.

CHAPTER 5

PAUSING FACTORS DISSOCIATE DURING SALT EXTRACTION OF CHROMATIN

5.1 Introduction

RNA polymerase II (Pol II) transitions through many steps in process of transcribing a gene. In particular, pause escape is a rate-limiting step (Core et al., 2008; Min et al., 2011; Muse et al., 2007; Zeitlinger et al., 2007) and likely a point of regulation at many genes in Drosophila and mammals (Adelman & Rogatsky, 2010; J. Lis, 1998; Rahl et al., 2010; Saha et al., 2011). This is most evident on Drosophila *Hsp70*. Upon activation, the paused polymerase is released from pausing and transcribes into the gene. Activation results in a dramatic change in the efficiency of pause escape. Before heat shock, Pol II is estimated to escape into productive elongation once every 10 minutes, but during an optimal heat shock, Pol II escapes every 4 seconds and Pol II is evenly distributed across the gene, indicating pausing is no longer the rate-limiting step (Fuda et al., 2009; J. Lis, 1998).

Pol II undergoing promoter-proximal pausing is distinguished from polymerase in other steps of the transcription cycle based on several attributes: it is located on the 5' end of genes, it is phosphorylated only on Serine 5 of the CTD repeats, it is transcriptionally engaged, and it is prevented from elongating further into the gene (Fuda et al., 2009; Nechaev & Adelman, 2011). These latter two criteria can be tested using nuclear run-on assays, which allow transcriptionally-competent polymerases to extend their transcripts by providing exogenous nucleotides to isolated nuclei (Love & Minton, 1985; Rougvie & Lis, 1988). Paused Pol II cannot efficiently run-on without

treatment of the nuclei with detergent or high salt, but polymerases transcribing in the body of the gene can readily incorporate these nucleotides independent of high salt or detergent treatments (Rougvie & Lis, 1988). The inability of pause polymerases to run-on is presumably due to an inhibitory factor or impediment to elongation unique to promoter-proximal pausing.

Based on current evidence, a few candidates for this inhibitory factor have been proposed. The most obvious candidates are Spt5/Spt4 and NELF complexes and first downstream (+1) nucleosome (Gilmour, 2009). Spt5/Spt4 and NELF complexes were first identified as factors required for the DRB-dependent inhibition of transcription in vitro (Wada et al., 1998; Y Yamaguchi et al., 1999). DRB-treatment reduced the production of full-length transcripts and short RNAs accumulated (N F Marshall & Price, 1995; Wada et al., 1998). Subsequent studies have shown these factors are localized to the 5' ends of genes (Andrulis et al., 2000; Gilchrist et al., 2010; C. Lee et al., 2008; Rahl et al., 2010), and RNAi depletion of NELF and Spt5 from Drosophila S2 cells globally reduced paused polymerase peaks (Gilchrist et al., 2008, Chapter 6 of this dissertation). As the primary component of chromatin, nucleosomes are located throughout the transcribed gene, and experiments using highly purified polymerase demonstrate nucleosomes can strongly block polymerase elongation in vitro (Izban & Luse, 1991; Orphanides et al., 1998). Recent studies have indicated the +1 nucleosome is positioned near the TSS, in close proximity to the paused Pol II (Mavrich et al., 2008; Schones et al., 2008). Together, this evidence suggests the +1 nucleosome could be a major obstacle blocking further elongation by the paused polymerase. It is not currently clear the extent to which each of these proposed mechanisms contribute to pausing.

To gain a greater understanding of the mechanisms involved in promoter-proximal pausing, we took advantage of the dependence of paused polymerases on sarkosyl or high salt in nuclear run-on assays. Presumably the factor or factors responsible for preventing the paused polymerase from transcribing will dissociate from the chromatin at the sarkosyl or high salt concentrations that allow the paused polymerase to run-on. Centrifugation allows the proteins associated with chromatin to be purified away from soluble nucleoplasmic proteins and proteins dissociated in the sarkosyl or high salt treatment. Therefore, the proteins released from chromatin can be identified to help elucidate the factors preventing pause polymerases from running on.

5.2 Results

I tested the ability to isolate DNA-associated proteins at various sarkosyl concentrations by incubating nuclei in run-on buffer with increasing amounts of sarkosyl and centrifuging the samples at high speed. After isolating the nucleic acids from both the supernatant and pellet, the DNA content of each fraction was quantified. As expected, in low salt buffer without sarkosyl, all of the genomic DNA was contained in the pellet fraction, indicating that it was still well packaged into chromatin (Figure 5.1A, Table 5.1). As the sarkosyl concentration increased, the amount of DNA in the supernatant increased, indicating that the chromatin is so dramatically disrupted that the DNA could not be pelleted by centrifugation. Even as little as 0.13% sarkosyl solubilized the majority of the DNA (Figure 5.1A, Table 5.1). Since sarkosyl disrupted chromatin at rather low concentrations, I tested the ability to isolate chromatin at various salt concentrations. The KCl concentrations used were less disruptive to chromatin; only the 800mM KCl treatment had a large portion of the DNA in the supernatant, as previously

reported (Love & Minton, 1985). Thus, KCl can be titrated across a broad range and allow the DNA and associated proteins to be easily isolated (Figure 5.1B, Table 5.2).

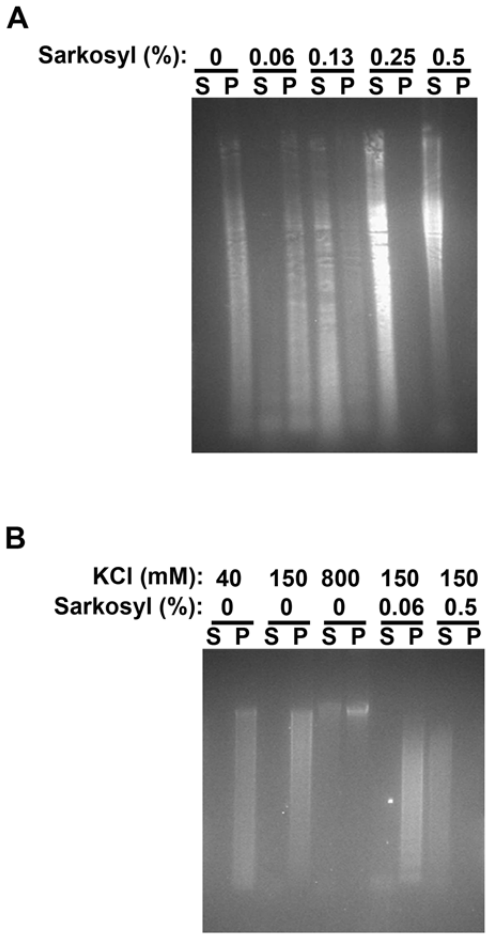


Figure 5.1 Fractionation of chromatin at various sarkosyl and KCl concentrations. **(A)** DNA isolated from the soluble (S) and insoluble (P) fractions of nuclei treated at various sarkosyl concentrations. The nucleic acids from each fraction were extracted, precipitated, resuspended in water, and the RNA was digested with RNase cocktail (Ambion). The DNA was run on a 1% agarose gel and also quantified using the Qubit (Table 5.1). **(B)** DNA isolated from the soluble (S) and insoluble (P) fractions of nuclei treated at various KCl or sarkosyl concentrations. DNA was isolated and examined as in A. The DNA is quantified in Table 5.2.

Table 5.1 The DNA content of the fractions from sarkosyl-treated nuclei.

DNA content (percent of total)		
Buffer	**Supernatant**	**Pellet**
0% sarkosyl	**16%**	**84%**
0.06% sarkosyl	**28%**	**72%**
0.13% sarkosyl	**67%**	**33%**
0.25% sarkosyl	**93%**	**7%**
0.5% sarkosyl	**93%**	**7%**

Table 5.2 The DNA content of the fractions from KCl-treated nuclei.

DNA content (percent of total)		
Buffer	**Supernatant**	**Pellet**
40mM KCl	**0%**	**100%**
150mM KCl	**2%**	**98%**
800mM KCl	**86%**	**14%**
0.06% sarkosyl	**13%**	**87%**
0.5% sarkosyl	**100%**	**0%**

I next determined the salt concentration that allowed the paused polymerase to run-on. Nuclear run-ons were performed with radiolabeled CTP, as a means to detect the run-on RNAs, at various KCl concentrations. The run-on RNAs were hybridized to single-stranded DNA probes to either strand of the promoter, pause region, and gene body of *Hsp70*. *Hsp70* is the prototypical example of a gene regulated at the pausing stage, and recent studies have indicated that the characteristics of *Hsp70* paused polymerase are typical of paused polymerases genome-wide (C. Lee et al., 2008; Muse et al., 2007), including run-on dependence on sarkosyl (Core et al. in prep). Thus, the results obtained from *Hsp70* should be applicable to paused polymerases genome-wide. As controls, I used probes to the histone H1 gene (non-paused Pol II-transcribed

gene), the Pol I-transcribed 18S ribosomal RNA, and the Pol III-transcribed U6 snRNA. As has been seen in previous studies (Rougvie & Lis, 1988), the *Hsp70* paused polymerase incorporated low amounts of radiolabel in the 40mM KCl run-on and high amounts in the 800mM KCl run-on (Figure 4.2A-B). There was a dramatic increase in incorporation from the *Hsp70* pause region with increases in salt concentration that began to plateau around 400mM KCl. In contrast, the non-paused histone H1 and U6 snRNA regions had consistent incorporation over the whole KCl range. In agreement with previous results, the 18S control had good incorporation and was moderately sensitive to KCl concentration (Love & Minton, 1985). This indicates that the factors responsible blocking paused polymerase run-on dissociate at fairly low KCl concentrations.

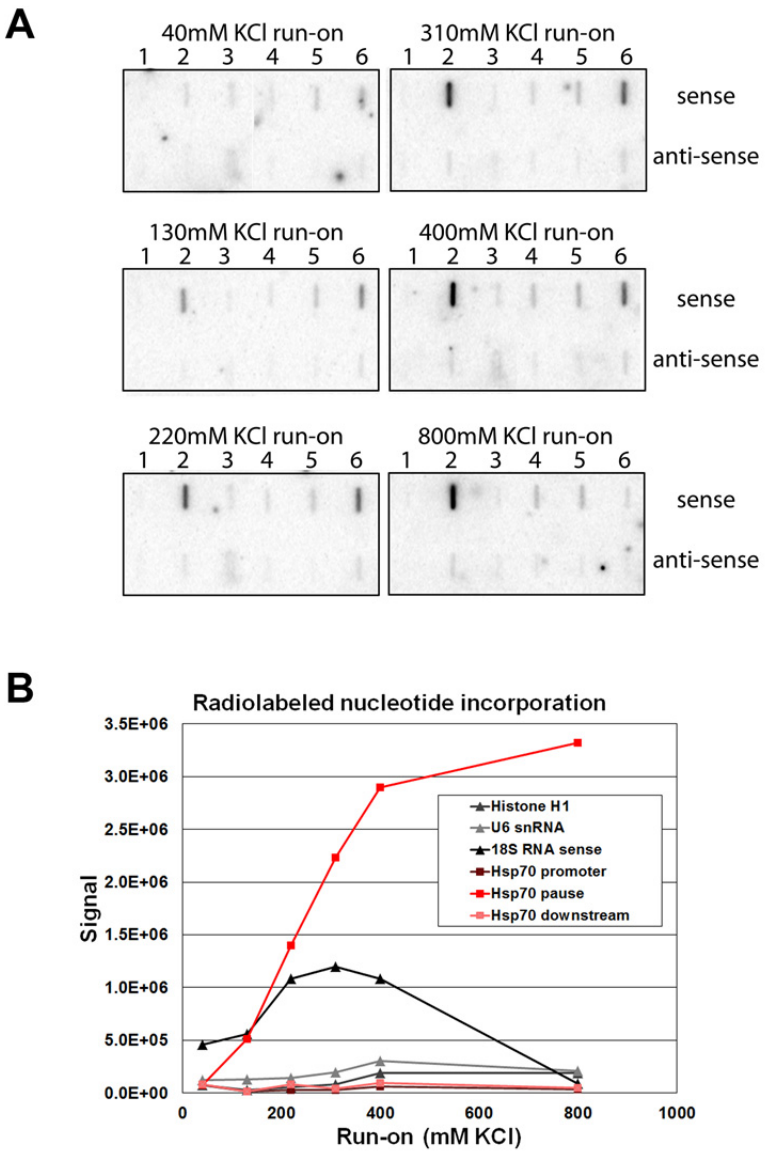


Figure 5.2 Hsp70 paused polymerase runs on at moderate salt concentrations. (A) Nuclear run-ons at the indicated KCl concentrations were performed on nuclei isolated from NHS S2 cells, the run-ons RNAs were isolated, and hybridized to single-stranded DNA probes to the indicated regions. **(B)** Graph showing the quantification of the signals in A.

To examine the effects on protein dissociation, I isolated chromatin after run-ons at the 40mM and 220mM KCl, and assayed the protein in the free (supernatant) and chromatin (pellet) fractions by Western blotting. The signal from the 220mM KCl fractions was compared to the signal from serial dilutions of the 40mM KCl fractions to determine the change in protein levels (Figure 5.3). The amount of Spt5 in the chromatin fraction after a 220mM KCl run-on was reduced to approximately 10% of the 40mM KCl levels. Nelf-E was also dramatically reduced at this KCl concentration. In contrast, histone H3 and the histone variant H2AvD present in the +1 nucleosome were stably associated with the chromatin at 220mM KCl. In addition, GTFs TFIIA and TBP are moderately affected by the higher salt concentration. These results are consistent with the hypothesis that the inability of *Hsp70* paused polymerase to run-on at low salt concentrations is due to the continued inhibition by the Spt5/Spt4 and NELF complexes, and not presence of the downstream +1 nucleosome.

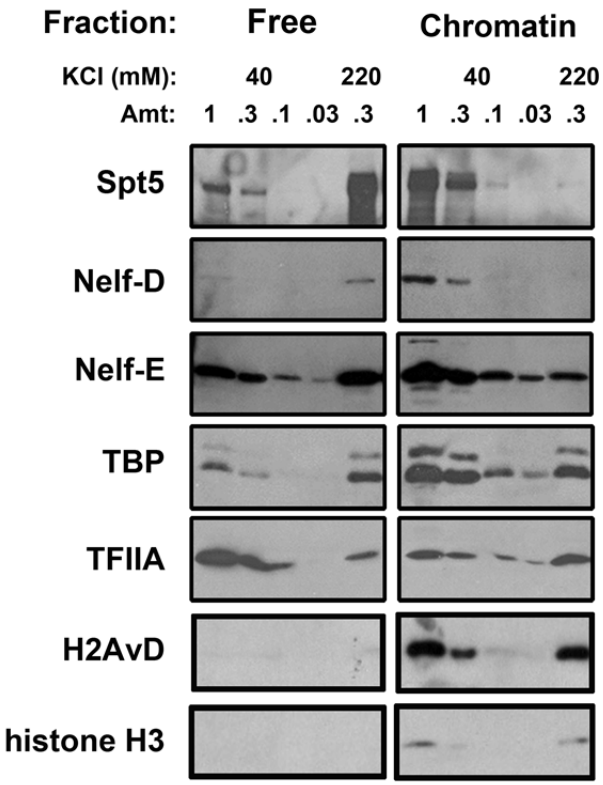


Figure 5.3 Spt5 and Nelf-E dissociate at 220mM KCl. Proteins from chromatin treated at the indicated salt concentrations was separated by centrifugation into soluble (free) and chromatin-bound (Chromatin). Western blots were performed for the factors indicated to the left of the blots. The relative amount loaded of each sample is indicated above the lane.

5.3 Conclusions

These results are consistent with the idea that Spt5/Spt4 and NELF have a larger contribution to promoter-proximal pausing than the barrier to elongation provided by the downstream +1 nucleosome. There is ample evidence for this hypothesis in the literature. First, in vitro assays that show the Spt5/Spt4 and NELF complexes inhibit transcription in the presence of the kinase inhibitor, DRB (Wada et al., 1998; Y Yamaguchi et al., 1999). In addition, ChIP signal for NELF at the 5' end of *Hsp70* is high under NHS condition when Pol II is stably paused, but low during heat shock when pausing is no longer rate-limiting (C.-H. Wu et al., 2003). RNAi depletion of either Spt5 or NELF reduces paused polymerase levels genome-wide, including on *Hsp70* (Gilchrist et al., 2008; Rahl et al., 2010; C.-H. Wu et al., 2003). In contrast, the +1 nucleosome on *Hsp70* is centered near 330bp downstream of the TSS (Petesch & Lis, 2008), well downstream of the paused polymerase, and in vitro transcription assays can recapitulate pausing on templates that do not contain nucleosomes (Adelman et al., 2005; Benjamin & Gilmour, 1998). The paused polymerase occupies the preferred +1 nucleosome position at many Drosophila genes, suggesting the +1 nucleosome can be displaced by the paused polymerase (Gilchrist et al., 2010).

5.3.1 Future directions

Using chromatin isolated from nuclei could be an ideal in vitro system for testing the mechanisms controlling transcription. Many experiments have used nucleosomes assembled in vitro on specific templates, but it is difficult to recreate in vivo chromatin environment with these reconstituted chromatin templates. Our results suggest the

chromatin from isolated nuclei is generally intact, and thus, is likely the best in vitro representation of the in vivo chromatin environment. Therefore, chromatin isolated from nuclei can avoid the problems plaguing chromatinized templates and allow investigation of the mechanisms regulating transcription at specific endogenous genes using either conventional run-on hybridizations or genome-wide using global run-on sequencing (GRO-seq).

In particular, future experiments could more directly test the hypothesis that the dissociation of Spt5/Spt4 and NELF are releasing the paused polymerase in the run-on. It will be interesting to see if paused polymerases from salt-extracted nuclei can run-on when returned to lower salt concentration or provided with exogenous recombinant or biochemically purified Spt5/Spt4 and NELF complexes. Additionally, chromatin from isolated nuclei could be treated with exogenous P-TEFb to see if treated paused polymerases can run-on in low salt conditions. Thus, this system has the potential to further clarify the mechanisms controlling pausing as well as other steps in transcription.

CHAPTER 6

SPT5 MAINTAINS PAUSED POLYMERASE BY PREVENTING ELONGATION

6.1 Introduction

The level of promoter-proximal pausing is dependent on the entry rate of Pol II into the pause sites and the rate of escape (Core & Lis, 2008). While the entry rate of Pol II is dependent on the many factors that control the recruitment, initiation, and promoter escape steps of the transcription cycle (Fuda et al., 2009), three factors control the escape from pausing: the Spt5/Spt4 and NELF complexes and the kinase P-TEFb.

The role of the Spt5/Spt4 and NELF complexes in control of elongation was first identified in vitro. These complexes inhibited elongation when P-TEFb was inhibited (Wada et al., et al., 1998; Y Yamaguchi et al., 1999). In addition, the Spt5/Spt4 complex enhanced elongation when nucleosides were limiting, and Spt5 mutants reduce expression (Guo et al., 2000; Swanson, Malone, & Winston, 1991; Wada et al., 1998). The Spt5/Spt4 and NELF complexes inhibit elongation through enhancing the intrinsic pausing of Pol II (Renner et al., 2001; Y Yamaguchi et al., 1999). They can inhibit TFIIS from facilitating the cleavage of backtracked RNAs, an activity required for maintaining paused polymerases in transcriptionally-competent state (Adelman et al., 2005; Murali Palangat et al., 2005). In addition, Spt5/Spt4 and NELF complexes inhibit the activity of TFIIF (Cheng & Price, 2007; Renner et al., 2001).

P-TEFb releases the negative effects of these complexes and allows the polymerases enter productive elongation (Cheng & Price, 2007; Ni et al., 2008). P-TEFb

can phosphorylate multiple components of the paused polymerase complex: Serine 2 residues on the CTD repeats, Spt5 CTR, and the Nelf-E subunit of NELF (Fujinaga et al., 2004; Ivanov et al., 2000; N F Marshall & Price, 1995). In particular, the threonine residues of the Spt5 CTR repeats are phosphorylated by P-TEFb, and this phosphorylation is important for the transition to productive elongation (Yamada et al., 2006). Spt5 phosphorylation may act as a switch, releasing the negative effects on the pause complex by dissociation of NELF and contributing to new interactions with positive elongation factors (Yamada et al., 2006).

Consistent with this dual role, Spt5 localizes to both the paused and elongation complexes in vivo (Andrulis et al., 2000; Rahl et al., 2010). Its localization overlaps with that of NELF, and is required for NELF to associate with Pol II (Missra & Gilmour, 2010; Y Yamaguchi et al., 2002). Spt5 can also interact with mRNA capping enzyme, and stimulate its guanyltransferase activity (Mandal et al., 2004; Pei & Shuman, 2002). In vivo, capping occurs coincident with pausing (Nechaev et al., 2010; Rasmussen & Lis, 1993). In addition, capping enzyme can counteract the negative effects of Spt5/Spt4 and NELF complexes on elongation during in vitro transcription reactions (Mandal et al., 2004). This suggests pausing and capping are interconnected, and may form a checkpoint at the transition into productive elongation.

Several recent studies have investigated the role of NELF in vivo (Gilchrist et al., 2008, 2010; Rahl et al., 2010; C.-H. Wu et al., 2003). NELF depletion in Drosophila reduces the level of pausing at many genes, but contrary to its in vitro effects, expression of the majority of genes affected by NELF depletion actually decreased (Gilchrist et al., 2008). This reduced transcription was a consequence of changes in

chromatin environment preventing initiation (Gilchrist et al., 2010). NELF depletion in mammalian stem cells had similar effects on paused polymerase level, especially at less active “non-productive” genes (Rahl et al., 2010), but little effect on genes with higher levels of transcription.

The role of Spt5 in transcription has recently been investigated in mouse embryonic stem cells (Rahl et al., 2010). When Spt5 was RNAi depleted, active genes had little reduction in promoter-proximal level of Pol II, but Pol II levels clearly increased on the body of the gene. This somewhat surprising finding that Spt5 depletion had little effect on promoter-proximal Pol II levels was attributed to high initiation rates. Genes with initiaition rates higher than pause escape rates can maintain paused polymerase levels when pause rates increase upon Spt5 depletion. To my knowledge, this is the only genome-wide study on the role of Spt5 in transcription. Here, I investigate the role of Spt5 in pausing in Drosophila cells.

6.2 Results

6.2.1 Spt5 depletion reduces paused Pol II levels on NHS *Hsp70*

To examine the role of Spt5 in pausing, I depleted Spt5 using RNAi in S2 cell culture, and examined Pol II levels on NHS *Hsp70*. After 5 days of depletion, cellular levels of Spt5 protein were reduced by 90% (Figure 6.1A). ChIP for the Pol II subunit Rpb3 showed that NHS Pol II on the 5’ end of *Hsp70* was decreased by 40%, compared to control cells treated with the non-targeting LacZ dsRNA, but levels of Pol II in downstream regions were comparable (Figure 6.1B). When I examined localization of Spt5 by ChIP, levels were reduced by 80% compared to control cells (Figure 6.1C).

These results indicate Spt5 depletion is impairing Pol II pausing on NHS *Hsp70*. Since Spt5 negatively affects transcription by inhibiting elongation in vitro (Renner et al., 2001; Wada et al., 1998) and Spt5 depletion in embryonic stem cells increased Pol II on the gene body (Rahl et al., 2010), it was surprising that Spt5 depletion did not increase downstream levels of Pol II. To investigate this further, I looked at the levels of NHS *Hsp70* RNA by reverse transcription and quantitative PCR (RT-qPCR). When total RNA from control and Spt5-RNAi cells was reverse transcribed using a gene specific primer just 3' to the *Hsp70* pause region (Hsp70+112), RNA levels were about 10-fold higher than control cells (Figure 6.2). Reverse transcription using primers downstream (Hsp70+1754) shows RNA from this region only increases 4-fold (Figure 6.2). These results suggest Spt5 depletion decreases pausing on *Hsp70* and allows more Pol II to elongate into the gene, but the polymerase does not efficiently reach the 3' end of the gene. In fact, reverse transcription with an oligo(dT) primer shows NHS *Hsp70* mRNA levels only increase 2-fold in Spt5-RNAi cells (Figure 6.2).

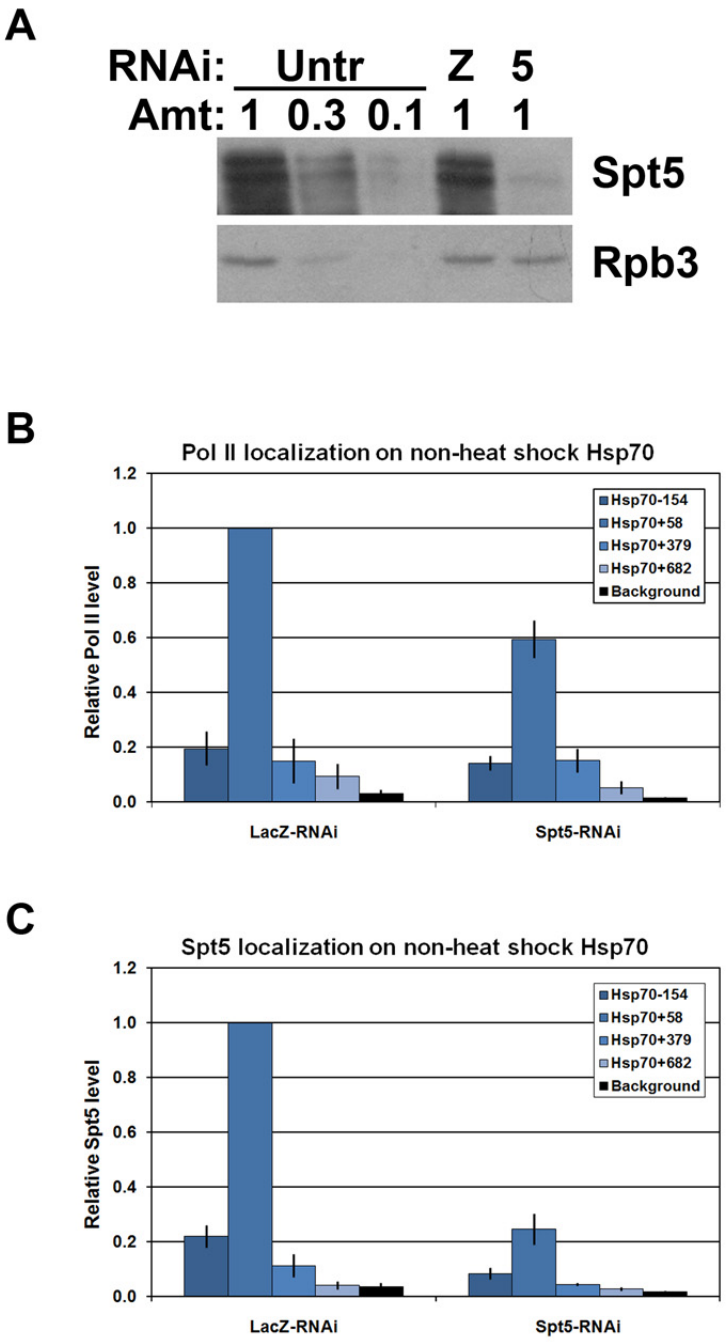


Figure 6.1 Depletion of Spt5 reduces paused polymerase on NHS *Hsp70*. (A) Western blots for Spt5 and TFIIS for whole cell extracts from Untreated, LacZ-RNAi, and GAF-RNAi cells (1 is equivalent to $1x10^6$ cells). **(B)** ChIP for Pol II subunit, Rpb3, on *Hsp70* in non-heat shock (NHS) LacZ-RNAi and Spt5-RNAi cells. **(C)** ChIP for Spt5 on *Hsp70* in NHS LacZ-RNAi and GAF-RNAi cells. The legend indicates the center of each primer set relative to the TSS. Error bars represents the SEM from at least 3 experiments.

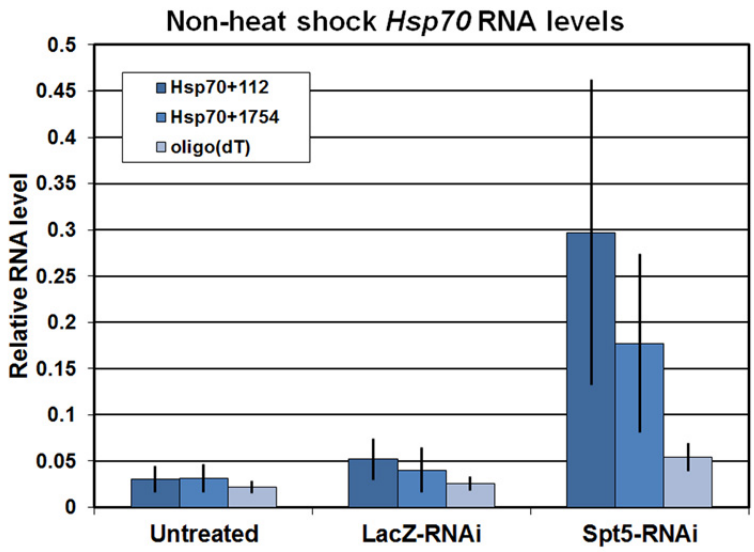


Figure 6.2 Spt5 depletion increase NHS *Hsp70* RNA levels. Total RNA of NHS Untreated, LacZ-RNAi, and Spt5-RNAi cells was reversed transcribed with Hsp70+112 and RpL32+204 (Hsp70+112), Hsp70+1754 and RpL32+204 (Hsp70+1754), or oligo(dT) (oligo(dT) reverse primers. The resulting cDNA was quantified by qPCR and the Hsp70 values were normalized to RpL32 values.

6.2.2 Spt5 depletion reduces pausing at many genes

Pause escape is rate-limiting at many Drosophila genes (Chopra et al., 2011; Core et al., 2008; Larschan et al., 2011; Min et al., 2011; Muse et al., 2007). To investigate the dependence of these paused polymerases on Spt5, I examined the distribution of polymerase genome-wide using global run-on sequencing (GRO-seq). Leighton Core helped with GRO-seq library construction, Colin Waters mapped the GRO-seq reads to the Drosophila genome, and both Colin and Leighton helped with the preliminary analysis. The density of GRO-seq reads was used to determine the distribution of engaged polymerase in Untreated, LacZ-RNAi, and Spt5-RNAi NHS cells, and the distribution of normalized reads for each library was graphed relative to the TSS. Spt5-RNAi library had decreased promoter-proximal reads and increased gene body reads when compared to either the Untreated or LacZ-RNAi libraries (Figure 6.3A). The number of promoter and gene body reads was determined for each unique non-overlapping gene (9452 genes), and compared between the LacZ-RNAi and Spt5-RNAi libraries (Figure 6.3B-C, promoter ρ=0.957 and gene body ρ=0.958). In agreement with the composite profile, a plot of promoter reads shows Spt5-RNAi reads are reduced at many genes (Figure 6.3B). The plot of gene body reads shows that there are more reads on many genes in the Spt5-RNAi library (Figure 6.3C). The genes with significantly different reads between the LacZ-RNAi and Spt5-RNAi libraries were determined using edgeR. There were 205 genes with significantly different promoter read levels (182 lower, 23 higher), and 855 genes with significantly different gene body reads (146 lower, 737 higher) (Figure 6.3B-C, red points).

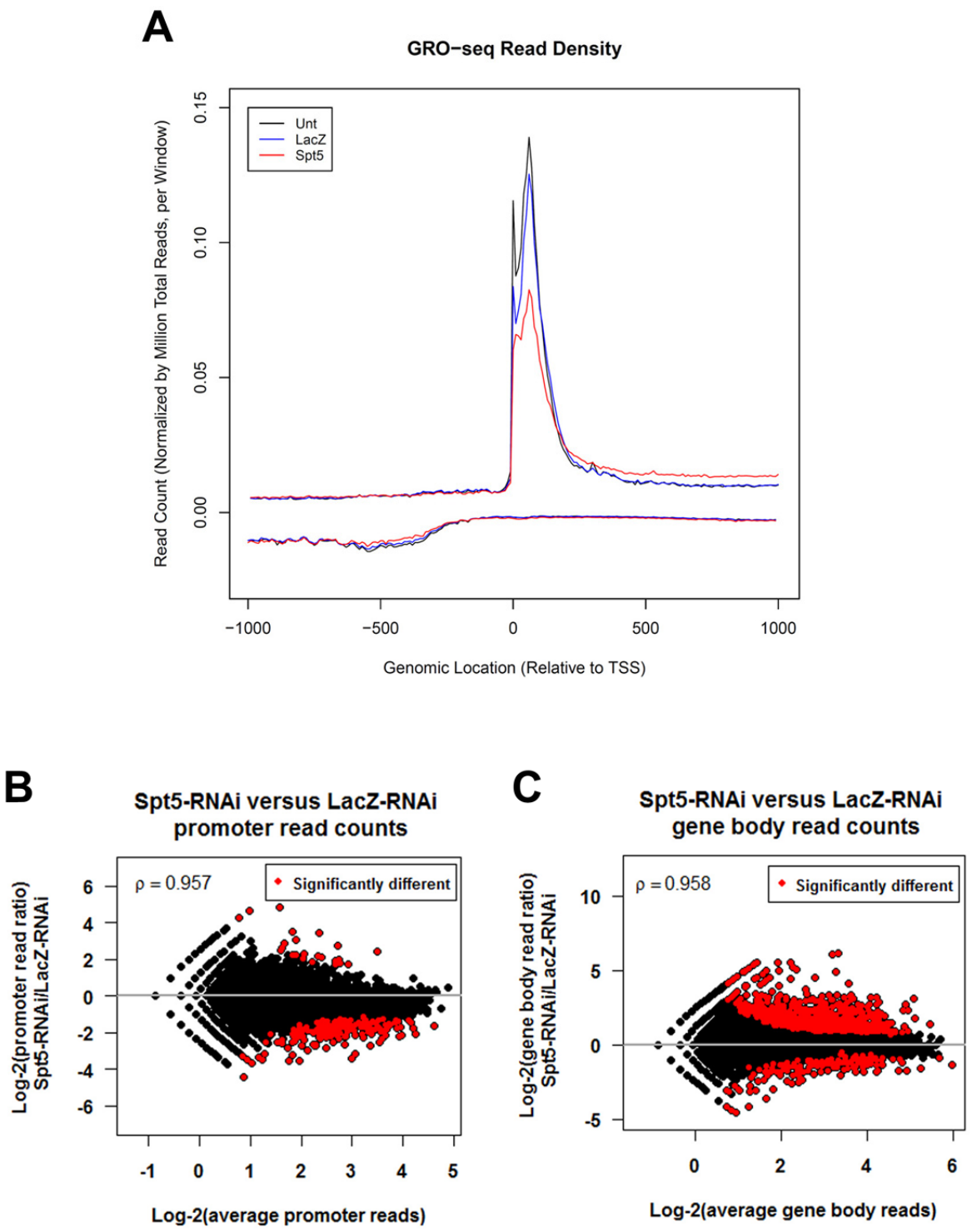

A
GRO−seq Read Density
Unt
LacZ
Spt5
Read Count (Normalized by Million Total Reads, per Window)
0.00
0.05
0.10
0.15
−1000
−500
0
500
1000
Genomic Location (Relative to TSS)
B
Spt5-RNAi versus LacZ-RNAi
promoter read counts
ρ = 0.957
Significantly different
Log-2(promoter read ratio)
Spt5-RNAi/LacZ-RNAi
6
4
2
0
-2
-4
-6
-1
0
1
2
3
4
5
Log-2(average promoter reads)
C
Spt5-RNAi versus LacZ-RNAi
gene body read counts
ρ = 0.958
Significantly different
Log-2(gene body read ratio)
Spt5-RNAi/LacZ-RNAi
10
5
0
-5
0
2
4
6
Log-2(average gene body reads)

Figure 6.3 Spt5 depletion changes polymerase levels on many genes. (A) GRO-seq reads from libraries for all RefSeq genes +/-1Kb relative to the TSS binned by 10bp and averaged per gene. The reads from the sense strand are plotted above zero and the reads from the anti-sense strand are plotted below zero. The x-axis indicates the position relative to the TSS, and the y-axis indicates reads per million mappable reads in the library per 10bp window. **(B)** Promoter proximal reads (100bp window with the most reads +/-250bp of the TSS) for each gene are plotted comparing the ratio of Spt5-RNAi reads to LacZ-RNAi reads versus the average Spt5-RNAi and LacZ-RNAi read counts for each region. Significantly different read counts between libraries are indicated by the red points. **(C)** Gene body reads (500bp downstream of the TSS to the polyadenylation site) for each gene are plotted comparing the ratio of GAF-RNAi reads to LacZ-RNAi reads versus the average GAF-RNAi and LacZ-RNAi read counts for each region. Significantly different read counts between libraries are indicated by the red points.

The pausing index (PI) is the ratio of promoter-proximal polymerase density to the gene body polymerase density, and measures the magnitude of the 5' enrichment of polymerase on a gene. The PI can be used to identify paused genes, based on a significant enrichment of read density in the promoter compared to the gene body read density, based on a Fisher exact test. Of the 205 genes with significantly different promoter read levels, 196 of these genes were paused and Spt5 depletion reduced pausing indices of these genes (Figure 6.4). Therefore, Spt5 depletion was not just affecting Pol II levels uniformly across these genes, but had different effects on the promoter and gene body levels of polymerase. These results agree with the ChIP and RT-qPCR experiments on NHS *Hsp70,* and suggest Spt5 depletion is allowing more polymerase to escape from pausing and transcribe into the gene.

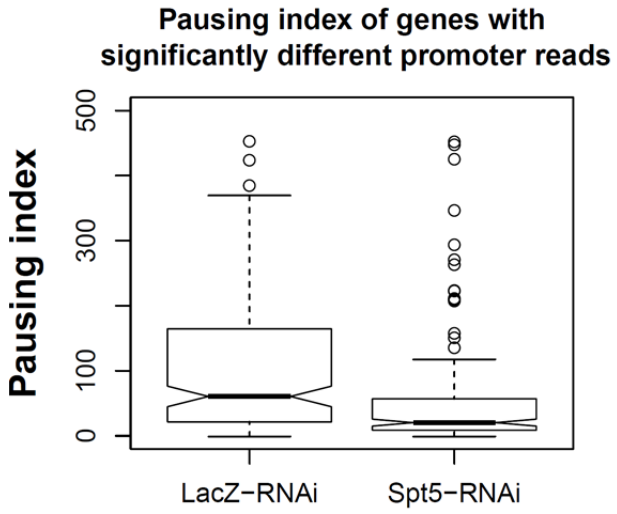


Figure 6.4 Spt5-RNAi changes the polymerase distribution across genes with significantly different promoter reads. The distribution of the pausing indices in the LacZ-RNAi and Spt5-RNAi libraries for the genes with significantly different promoter reads between LacZ-RNAi and GAF-RNAi displayed in box-and-whisker plots. The lower and upper edges of the box marks the first and third quartiles of the distribution, respectively, the thick line within each box marks the median, the whiskers mark 1.5 times the interquartile range (first to third quartile) outside the box, and the open circles marks any points lying outside the whiskers.

Although the relative levels of polymerase measured by the PI demonstrate pausing is affected, direct measures of polymerase levels are also important for a clear understanding of the role of pausing in controlling transcription. For example, NELF depletion reduced the PI of many genes in Drosophila cells, but it also reduced expression of many genes (Gilchrist et al., 2008). This led to the observation that the paused polymerase at these genes was maintaining the promoter in an accessible state by preventing nucleosomes from covering the TSS (Gilchrist et al., 2010).

GRO-seq provides a direct measure of transcribing polymerase across the genome based on GRO-seq read count. This easily demonstrates difference between genes within a GRO-seq library, but due to differences in library size, each library must be normalized in some manner to compare it to other libraries. Thus far, the reads for each library have been normalized to all mapped Pol II reads. Since most Pol II is located within the promoter-proximal region (Figure 6.3A), this normalization may lead to biases if Spt5 depletion reduces a large portion of Pol II pausing. This would mean we have underestimated the effect of Spt5-RNAi on promoter reads and overestimated the effect on gene body reads by using this normalization. To address this, I determined the effect of Spt5 depletion on the overall level of Pol II transcription. Since low levels of alpha-amanitin specifically inhibit Pol II (Chao & Price, 2001; Lindelle tal., 1970), I performed nuclear run-on assays with Untreated, LacZ-RNAi, GAF-RNAi, and Spt5-RNAi nuclei in the presence or absence of 1µg/ml alpha-amanitin. The total radiolabel incorporation was measured by scintillation counting. The level of Pol II transcription was calculated by subtracting the counts in run-ons with amanitin from the counts in run-ons without amanitin. The fraction of total signal contributed by Pol II from each

treatment was normalized to the Untreated level. Assuming Spt5-RNAi does not affect either Pol I or Pol III transcription, Spt5 knock-down reduced Pol II transcription by 60% (Figure 6.5A). In contrast, GAF knock-down only reduced Pol II transcription by 15%. To assess the effect that this reduction in Pol II transcription has on the libraries, we assumed Pol I or Pol III transcription was not affected by knock-down, and corrected the mapped reads by either the total number of Pol I or Pol III reads in each library. Both corrections shifted the reads in Spt5-RNAi library lower relative to the LacZ-RNAi library (Figure 6.5B-E). The size of the Pol III correction fit the results of the alpha-amanitin experiments the best. This normalization lowered the Spt5-RNAi reads the most, but even after this normalization, many of the genes with significantly different gene body reads still showed an increase in reads over the LacZ-RNAi library. This strongly suggests that Spt5 depletion reduces the level of promoter-proximal Pol II and increases the level of polymerase in the gene body for many genes.

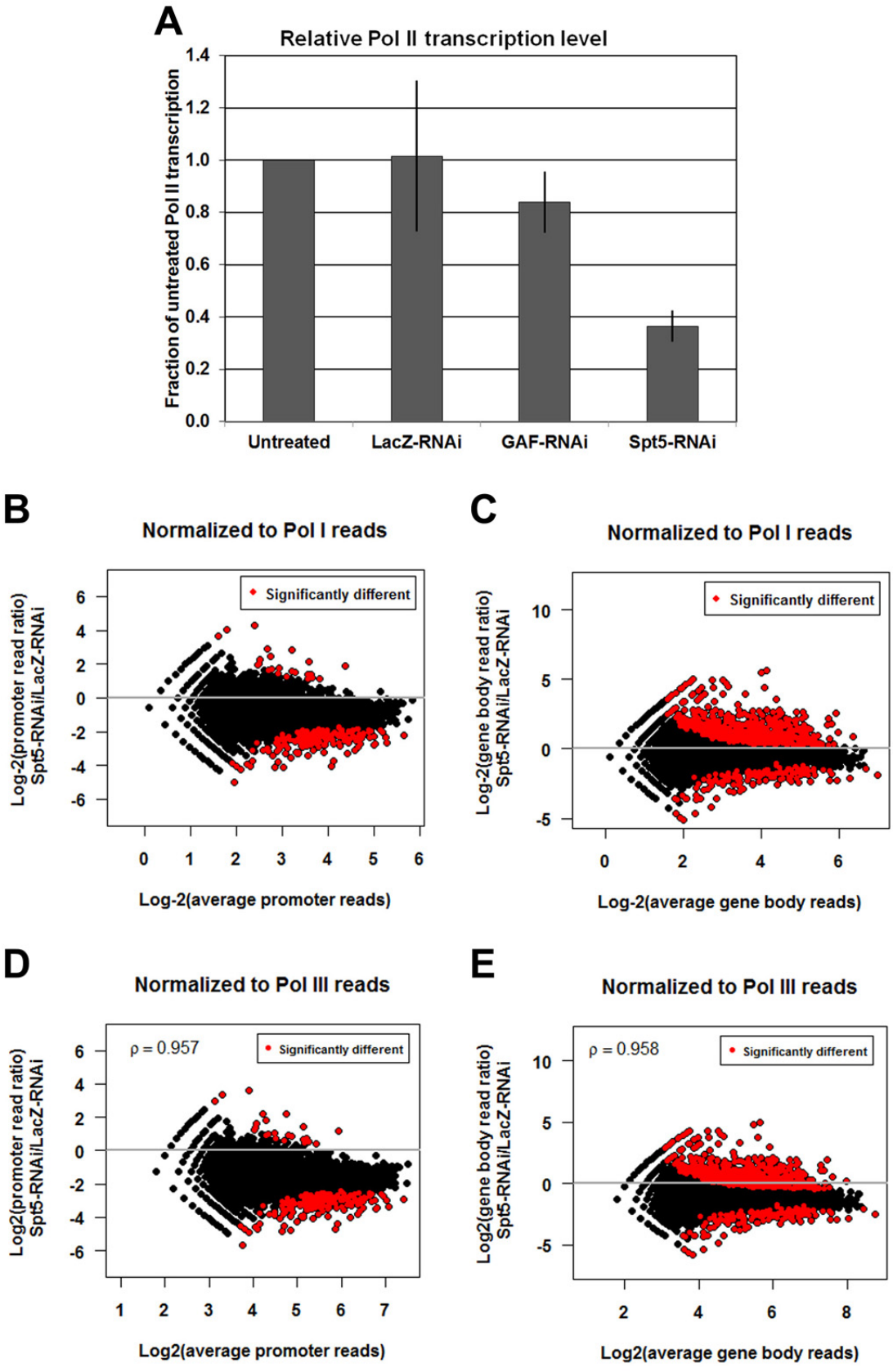

A
Relative Pol II transcription level
Fraction of untreated Pol II transcription
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Untreated
LacZ-RNAi
GAF-RNAi
Spt5-RNAi
B
Normalized to Pol I reads
Log-2(promoter read ratio) Spt5-RNAi/LacZ-RNAi
Significantly different
Log-2(average promoter reads)
C
Normalized to Pol I reads
Log-2(gene body read ratio) Spt5-RNAi/LacZ-RNAi
Significantly different
Log-2(average gene body reads)
D
Normalized to Pol III reads
Log2(promoter read ratio) Spt5-RNAi/LacZ-RNAi
ρ = 0.957
Significantly different
Log2(average promoter reads)
E
Normalized to Pol III reads
Log2(gene body read ratio) Spt5-RNAi/LacZ-RNAi
ρ = 0.958
Significantly different
Log2(average gene body reads)

Figure 6.5: Spt5 depletion reduces all Pol II transcription by 60%. (A) The fraction of the run-on signal contributed by Pol II was computed based on the radiolabel incorporation for run-on +/- alpha-amanitin, and graphed relative to Untreated. The error bars represent the range from 2 biological replicates. **(B)** Promoter proximal reads (100bp window with the most reads +/-250bp of the TSS) for each gene are normalized to total Pol I reads in the library, and plotted comparing the ratio of Spt5-RNAi reads to LacZ-RNAi reads versus the average Spt5-RNAi and LacZ-RNAi read counts for each region. Significantly different read counts between libraries are indicated by the red points. **(C)** Gene body reads (500bp downstream of the TSS to the polyadenylation site) for each gene are normalized to total Pol I reads in the library, and plotted comparing the ratio of GAF-RNAi reads to LacZ-RNAi reads versus the average GAF-RNAi and LacZ-RNAi read counts for each region. Significantly different read counts between libraries are indicated by the red points. **(D)** Promoter proximal reads (100bp window with the most reads +/-250bp of the TSS) for each gene are normalized to total Pol III reads in the library, and plotted comparing the ratio of Spt5-RNAi reads to LacZ-RNAi reads versus the average Spt5-RNAi and LacZ-RNAi read counts for each region. Significantly different read counts between libraries are indicated by the red points. **(E)** Gene body reads (500bp downstream of the TSS to the polyadenylation site) for each gene are normalized to total Pol III reads in the library, and plotted comparing the ratio of GAF-RNAi reads to LacZ-RNAi reads versus the average GAF-RNAi and LacZ-RNAi read counts for each region. Significantly different read counts between libraries are indicated by the red points.

6.2.3 The effects of Spt5 depletion are similar to the effects of NELF depletion.

Both Spt5/Spt4 and NELF complexes inhibit elongation, but the Spt5/Spt4 complex can also have a positive effect on elongation. It has been proposed that Spt5 serves as a scaffold for binding other factors to the elongation complex, NELF when it is unphosphorylated and positive elongation factors when it is phosphorylated (Nechaev & Adelman, 2011; Yamada et al., 2006). In vitro evidence indicates Spt5 is required for NELF association with the Pol II complex (Missra & Gilmour, 2010; Y Yamaguchi et al., 2002). P-TEFb phosphorylates the Spt5 CTR (Ivanov et al., 2000; Yamada et al., 2006) and phosphorylation relieves the negative effect of Spt5/Spt4 on elongation (Cheng & Price, 2007). Therefore, I was interested to determine if Spt5-RNAi had effects distinct from NELF-RNAi. Recent genome-wide experiments have shown NELF depletion has a range of effects on pausing at individual genes in vivo using ChIP (Gilchrist et al., 2008, 2010) or GRO-seq (Core et al. in prep). I compared the results from the Spt5-RNAi GRO-seq with those of a NELF-RNAi GRO-seq library. The levels of promoter and gene body reads for each gene in the Spt5-RNAi and NELF-RNAi libraries were similar (Figure 6.6E-F, promoter ρ=0.806 and gene body ρ=0.867). The correlation between the Spt5-RNAi and NELF-RNAi libraries was less than the correlation of Spt5-RNAi with its corresponding LacZ-RNAi control library (see above). This analysis is complicated by the fact that the correlation between the LacZ-RNAi libraries from each experiment was lower as well (Figure 6.6A-D, promoter ρ=0.833 and gene body ρ=0.905), and suggests some of the variation is likely due to experimental variation, such as cell culture differences and differences in library construction. Although a more thorough analysis is needed, these preliminary results suggest there are no major differences between Spt5

and NELF knock-down effects on polymerase distribution and effects seen in Spt5 knock-down are due primarily to effects on pausing.

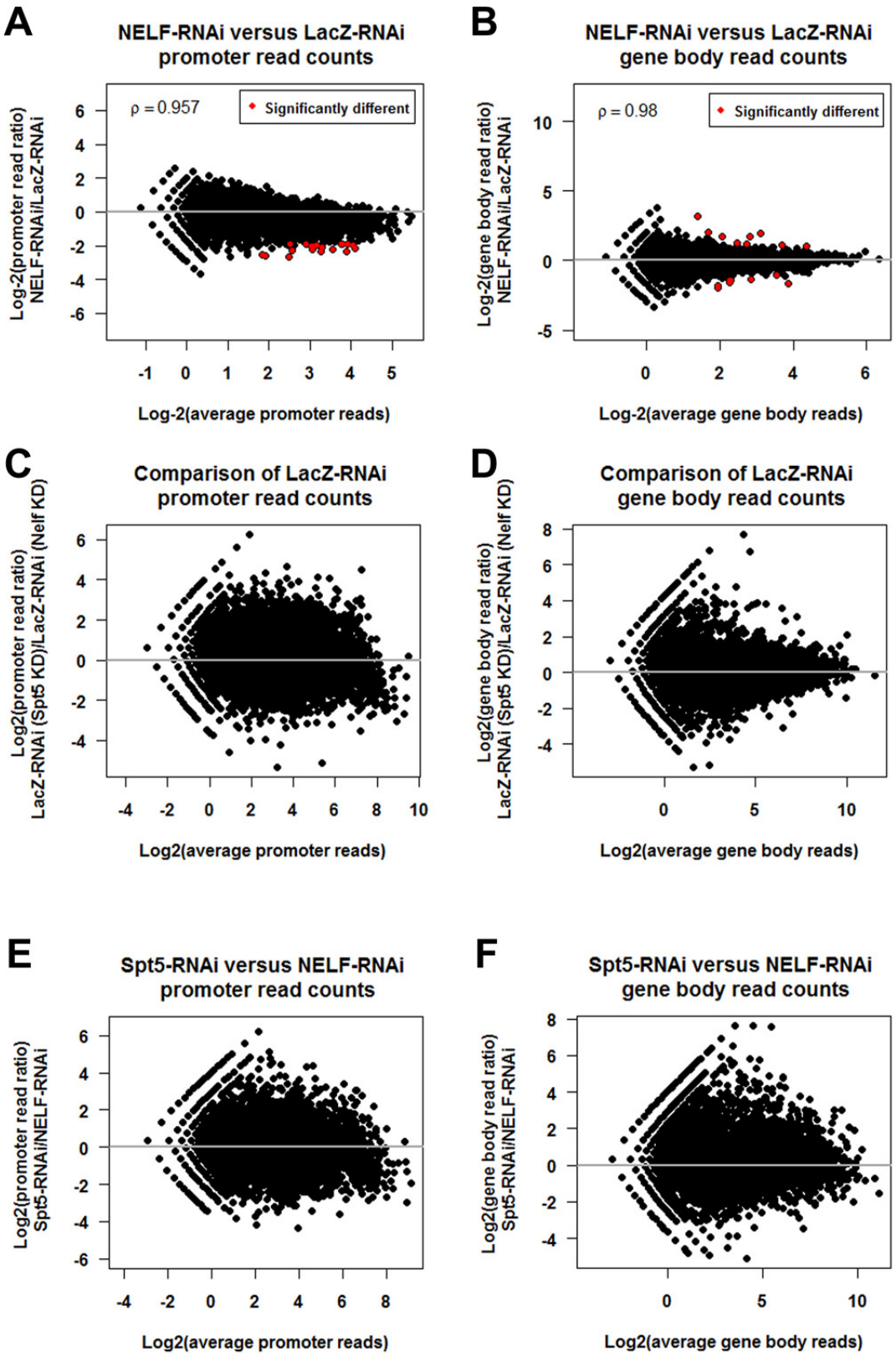

A
NELF-RNAi versus LacZ-RNAi
promoter read counts
ρ = 0.957
Significantly different
Log-2(promoter read ratio)
NELF-RNAi/LacZ-RNAi
Log-2(average promoter reads)
B
NELF-RNAi versus LacZ-RNAi
gene body read counts
ρ = 0.98
Significantly different
Log-2(gene body read ratio)
NELF-RNAi/LacZ-RNAi
Log-2(average gene body reads)
C
Comparison of LacZ-RNAi
promoter read counts
Log2(promoter read ratio)
LacZ-RNAi (Spt5 KD)/LacZ-RNAi (Nelf KD)
Log2(average promoter reads)
D
Comparison of LacZ-RNAi
gene body read counts
Log2(gene body read ratio)
LacZ-RNAi (Spt5 KD)/LacZ-RNAi (Nelf KD)
Log2(average gene body reads)
E
Spt5-RNAi versus NELF-RNAi
promoter read counts
Log2(promoter read ratio)
Spt5-RNAi/NELF-RNAi
Log2(average promoter reads)
F
Spt5-RNAi versus NELF-RNAi
gene body read counts
Log2(gene body read ratio)
Spt5-RNAi/NELF-RNAi
Log2(average gene body reads)

Figure 6.6 Spt5 depletion effects on polymerase distribution are similar to NELF-RNAi. (A) Promoter proximal reads (100bp window with the most reads +/-250bp of the TSS) for each gene are plotted comparing the ratio of NELF-RNAi reads to LacZ-RNAi reads versus the average NELF-RNAi and LacZ-RNAi read counts for each region (Core et al. in prep). Significantly different read counts between libraries are indicated by the red points. **(B)** Gene body reads (500bp downstream of the TSS to the polyadenylation site) for each gene are plotted comparing the ratio of NELF-RNAi reads to LacZ-RNAi reads versus the average NELF-RNAi and LacZ-RNAi read counts for each region. **(C)** Promoter proximal reads (100bp window with the most reads +/-250bp of the TSS) for each gene are plotted comparing the ratio of LacZ-RNAi libraries from the Spt5-RNAi and NELF-RNAi experiment. **(D)** Gene body reads (500bp downstream of the TSS to the polyadenylation site) for each gene are plotted comparing the ratio of LacZ-RNAi libraries from the Spt5-RNAi and NELF-RNAi experiment. **(E)** Distribution of promoter-proximal read ratios for genes with significantly reduced promoter proximal reads in the Spt5-RNAi libraries. **(F)** Distribution of gene body read ratios for genes with significantly reduced gene body reads in the Spt5-RNAi libraries. **(F)** Distribution of gene body read ratios for genes with significantly increased gene body reads in the Spt5-RNAi libraries.

6.3 Discussion

There is extensive evidence in vitro demonstrating the Spt5/Spt4 and NELF complexes inhibit transcription elongation in the absence of P-TEFb activity and indicating that these complexes at least partially mediate promoter-proximal pausing in vivo (N F Marshall & Price, 1995; Wada et al., 1998; Y Yamaguchi et al., 1999). I have examined the role of Spt5 in transcription in vivo by depleting Spt5 from Drosophila S2 cells and examining its effect of Pol II distribution. Spt5 knock-down reduced Pol II levels on the 5' end of NHS *Hsp70* by 40%. Although this reduction in 5' Pol II was not accompanied by an increase in Pol II on the body of the gene, RT-qPCR demonstrated NHS *Hsp70* RNA increased. The distribution of transcriptionally-engaged polymerase was examined genome-wide using GRO-seq. As on NHS *Hsp70*, many of genes had reduced promoter-proximal polymerase and significant increases of polymerase in the gene body. Although a more in depth analysis is necessary, the distribution of Spt5-RNAi GRO-seq reads on genes as similar to the distribution of reads from NELF-RNAi GRO-seq, and these preliminary results do not show obvious differences between the Spt5-RNAi and NELF-libraries that would indicate an additional positive role for Spt5 in vivo.

6.3.1 Future directions

My comparison of Spt5-RNAi and NELF-RNAi is very preliminary and rather simplistic. It indicates the effects are fairly similar, and there were no obvious differential effects due to the positive role of Spt5 in elongation. These effects may be masked by effects on pausing or hidden by the variation between the NELF-RNAi and Spt5-RNAi

libraries, as exemplified by the comparison between the LacZ-RNAi libraries generated by Leighton Core and me, respectively. Additionally, long genes, or genes with other specific characteristics may be more dependent on the positive activity of Spt5. For example, ChIP-seq for Pol II in mouse embryonic stem cells knocked-down for either NELF or Spt5 found that Spt5 knock-down had different effects on genes with higher levels of transcription (Rahl et al., 2010). Thus, I plan to examine whether the effects of Spt5 knock-down are different for genes with higher levels of gene body reads.

There is extensive evidence that the early elongation is linked to 5' mRNA capping. Capping occurs coincident with pausing (Nechaev et al., 2010; Rasmussen & Lis, 1993), and Spt5 and Serine 5 phosphorylated CTD interact with and stimulate mRNA capping enzymes (Mandal et al., 2004; Pei & Shuman, 2002; Wen & Shatkin, 1999). Intriguingly, capping enzyme association can counteract the negative effects of Spt5/Spt4 and NELF complexes on elongation during in vitro transcription reactions (Mandal et al., 2004). This suggests a possible role for 5' capping in pause release, but the importance of these interactions in vivo is unclear. Therefore, it is of interest to examine the effects of Spt5 depletion on capping.

The presence of the 5' cap on RNA can be identified from cDNA synthesis (Schmidt & Mueller, 1999; Shibata et al., 2001). Reverse transcriptase frequently adds an extra cytosine to the cDNA from capped transcripts due a terminal transferase-like activity, but this activity varies with reaction conditions (Schmidt & Mueller, 1999). The procedure I used to construct the GRO-seq libraries should allow identification of capped RNAs in LacZ-RNAi and Spt5-RNAi libraries. These reads should be readily identified by an enrichment of an untemplated guanine at the beginning of reads around

the TSS. If these reads are present in my libraries, I will determine depletion of Spt5 changes their levels.

CHAPTER 7

THE ROLE OF TWO KINASES, P-TEFB AND CDK12, IN PAUSE ESCAPE AND PRODUCTIVE ELONGATION[2]

7.1 Introduction

RNA polymerase II (Pol II) transitions through various stages in the course of synthesizing mRNA. The various stages not only ensure that the RNA synthesis begins and ends at the correct place, but also the proper processing of the transcript into mature mRNA (Saunders et al., 2006). During transcription, the transcription complex changes dramatically in both factor composition and levels of post-translational modification, and Pol II itself is a major target of these modifications. The C-terminal domain of the largest subunit of Pol II (CTD) is composed of heptapeptide repeats with the consensus sequence YSPTSPS. Phosphorylation of the serines within the repeats occurs as Pol II progresses through the transcription cycle (Laybourn & Dahmus, 1990; Payne et al., 1989). Pol II is recruited and initiates transcription in a hypophosphorylated form (Laybourn & Dahmus, 1989). As Pol II transitions from PIC to early elongation, the CTD is phosphorylated on Serine 5 and 7 by the Cdk7 kinase subunit of TFIIH (Chapman et al., 2007; Egloff et al., 2007; Glover-Cutter et al., 2009). In metazoans, these early elongation complexes often encounter a rate-limiting step (Core & Lis, 2008; Guenther et al., 2010; Muse et al., 2007; Nechaev & Adelman, 2011). After transcribing 20-65 nucleotides the polymerase pauses or stalls, and this pausing is at least partially

[2] The data and figures in this chapter are published in Ni et al. 2008 and Bartkowiak et al. 2010.

dependent on the actions of the Spt5/Spt4 and NELF complexes (Wada et al., 1998; Y Yamaguchi et al., 1999). The transition from the pausing into productive elongation is marked by phosphorylation of the Serine 2 in the CTD, residues in the C-terminal region of Spt5, and the Nelf-E subunit of the NELF complex by the Cdk9 kinase of P-TEFb (Fujinaga et al., 2004; Ivanov et al., 2000; Ni et al., 2008; Yamada et al., 2006). This hyperphosphorylated CTD interacts with elongation, mRNA processing, and termination factors to ensure efficient transcription and processing of the transcript into mRNA (Ahn et al., 2004; Bird et al., 2004; McCracken et al., 1997; Ni et al., 2004).

There is a wealth of evidence that this phosphorylation is crucial for progression through the transcription cycle. The kinase inhibitors impair transcription elongation in vitro and in vivo (Chao & Price, 2001; Chodosh et al., 1989; Kephart et al., 1992; Sehgal et al., 1976). The transcription is dependent on TFIIH-mediated phosphorylation of the Pol II complex (Glover-Cutter et al., 2009; Schwartz et al., 2003). Serine 5 phosphorylation may allow Pol II to break its contacts with promoter-bound factors and aid promoter escape. Interactions between the co-activator Mediator and the CTD are disrupted by phosphorylation (Max et al., 2007). Serine 5 phosphorylation is readily detected on the transcriptionally-engaged Pol II on the 5' end of genes (Boehm et al., 2003; Muse et al., 2007), and it is important for the transition to the promoter-proximal pausing (Glover-Cutter et al., 2009; Schwartz et al., 2003). Pausing is a target of regulation for several genes (Adelman & Rogatsky, 2010; J. Lis, 1998; Saha et al., 2011). The transition from the pausing into productive elongation is dependent on phosphorylation by the Cdk9 kinase, P-TEFb (Ni et al., 2008). P-TEFb phosphorylation relieves the inhibitory actions of Spt5/Spt4 and NELF in vitro (Cheng & Price, 2007; N F

Marshall & Price, 1995), and artificial recruitment of P-TEFb to the paused *Hsp70* gene stimulates transcription (Lis et al., 2000). Several activators have been shown to recruit P-TEFb to these paused genes directly or indirectly (Peterlin & Price, 2006).

The role of CTD phosphorylation in transcription is generally conserved among eukaryotes. Although the transition from initiation to productive elongation is not rate-limiting in yeast, the progression through these stages is still marked by Serine5, and subsequently Serine 2, phosphorylation (Buratowski, 2009). Interestingly, Serine2 phosphorylation is mediated by two kinases in yeast, the Cdk9-ortholog Bur1 and another kinase Ctk1. Interestingly, these two kinases have non-redundant roles in transcription (Cho et al., 2001; Keogh et al., 2003; Yao & Prelich, 2002). Ctk1 mutants dramatically reduce Serine 2 phosphorylation of the CTD, but have little effect on Pol II levels (Cho et al., 2001). In contrast, Bur1 mutants reduce elongating Pol II on the 3' end of genes, but the affect of the mutant on CTD phosphorylation was less dramatic (Keogh et al., 2003). These Bur 1 mutants display similar phenotypes to Spt5 mutants, suggesting Bur1 may target Spt5 and Ctk1 may target the CTD (Zhou et al., 2009). P-TEFb appears to be the predominant Serine 2 kinase in metazoans, but other kinases that are similar to Ctk1 do exist in metazoans (Liu & Kipreos, 2000).

Given the importance of Serine 2 phosphorylation in the progression through the transcription cycle and its potentially pivotal role in metazoan transcriptional regulation, it is important to not only identify the kinases involved, but also obtain a detailed understanding the role of these kinases in the release of paused polymerase. Herein, I will highlight my work from two studies addressing these topics, one investigating the role of P-TEFb in releasing the paused polymerase (a collaboration with Drs. Zhuoyu Ni

and Abbie Saunders from our laboratory) (Ni et al., 2008) and the other examining the localization of the Drosophila ortholog of CTDK1, CDK12 (a collaboration with Bartlomiej Bartkowiak from Dr. Arno Greenleaf's laboratory) (Bartkowiak et al., 2010).

7.2 Results

7.2.1 Flavopiridol inhibition of P-TEFb prevents efficient entry into productive elongation.

The stimulation of transcription when P-TEFb is artificially recruited to the uninduced non-heat shock (NHS) *Hsp70* gene indicates release of the paused polymerase is dependent on P-TEFb activity (Lis et al., 2000), but the question remained whether P-TEFb activity was required for the pause release or just coincident. To address this question, the effects of P-TEFb inhibition by flavopiridol (FP) were examined on transcription of *Hsp70* in Drosophila S2 cells (Ni et al., 2008). ChIP for Pol II showed a fairly even distribution of Pol II across the gene in heat-shocked control cells. As early as 1 minute after FP treatment, Pol II levels dropped in the body of the gene, but not at the 5' end. This initial loss was most dramatic in the middle of the gene. After 3 minutes of treatment, Pol II signal from the 3' end of gene was lost as well, and the Pol II signal was completely restricted to the 5' end of the gene. The pattern and kinetics of Pol II loss suggested P-TEFb activity is necessary for Pol II release into productive elongation, but not Pol II already in productive elongation. Consistent with this, FP treatment immediately before a heat shock prevented Pol II from entering into the body of the gene (Ni et al., 2008).

Although these results suggested P-TEFb activity is necessary for Pol II release into productive elongation, the exact nature of this stable 5' Pol II was not known. ChIP cannot determine whether this 5' polymerase is transcriptionally-engaged, and if so, whether it occupies the same sites as the paused polymerase under NHS conditions. To answer these questions, I used terminated run-ons (tRO) to map the polymerase locations on *Hsp70* in high resolution during NHS and HS +/- FP treatment, based on the length of their nascent RNAs (Rasmussen & Lis, 1993). Polymerases on NHS *Hsp70* are localized predominantly between 20 and 45 nucleotides downstream of the TSS, similar to previous results (Figure 7.1B, lane1) (Rasmussen & Lis, 1993). After 2.5 minutes of heat shock, polymerase is distributed across *Hsp70* (lane 2). Results from cells treated with 500nM FP immediately before a 2.5 minute heat shock show these 5' polymerases were restricted to within 150 nucleotides downstream of the TSS, but occupied sites downstream of the NHS pause sites. This difference from NHS sites is not likely due to residual P-TEFb activity because treatment with 4-fold higher concentrations of FP did not change the distribution (lane 4). Quantification of the signal showed a 1.5-2.5 times more signal on FP-treated HS *Hsp70* than NHS *Hsp70*. Since previous estimates indicate there is one polymerase per gene under NHS conditions, this indicates there is likely one additional polymerase occupying this region after FP treatment (Ni et al., 2008).

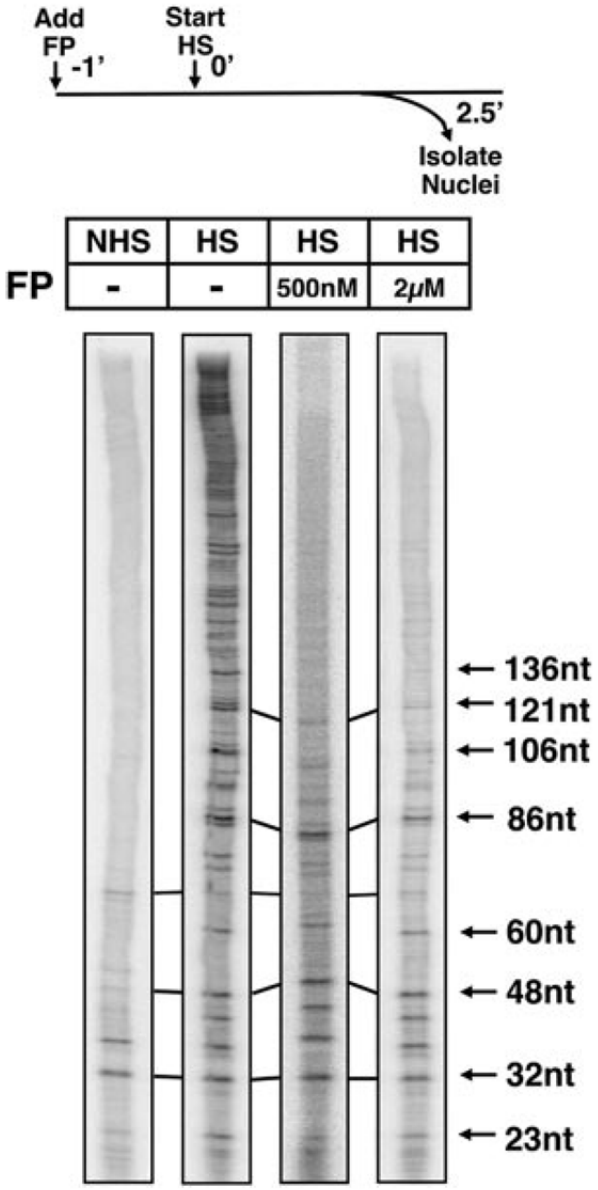


Figure 7.1 Polymerases remaining on *Hsp70* 3.5 min post-FP treatment are promoter-proximally restricted. Terminating nuclear run-ons were performed under non-heat shock conditions (lane 1), or following 2.5 min of heat shock in the absence (lane 2) or presence of 500nM (lane 3) or 2μM (lane 4) FP added 1 min before heat shock. The experimental scheme is illustrated (top). The sizes of the *Hsp70* RNA transcripts are indicated to the left.

7.2.2 The homolog to S. cerevisiae CTDK1, CDK12, localizes to actively transcribed genes.

In yeast, Serine 2 phosphorylation is performed by two kinases, Bur1 (Cdk9) and CTDK1 (Ctk1), but P-TEFb was the only identified Serine 2 kinase in metazoans (Peterlin & Price, 2006). The metazoan orthologs of CTDK1 were identified as CDK12/CDK13 (Bartkowiak et al., 2010). A chimeric Ctk1 with the kinase homology domain of human CDK12 substituted in the yeast Ctk1 is functional, and cannot rescue temperature sensitive mutants of Bur1. In addition, CDK12 co-localizes with hyperphosphorylated Pol II on Drosophila polytene chromosomes, and RNAi depletion of the Drosophila or human CDK12 altered levels of CTD phosphorylation (Bartkowiak et al., 2010). Although these results indicated CDK12 is recruited to actively transcribing genes, the exact localization on genes was not known. I determined the localization on genes in higher resolution using ChIP for Drosophila dCDK12. Since the robust induction of *Hsp70* makes this gene ideal for examining the recruitment of factors (Boehm et al., 2003; Zobeck et al., 2010), I examined whether dCDK12 localized to *Hsp70* under NHS and HS conditions. Under NHS condition, the paused Pol II is present at the 5' end of the gene, but dCDK12 did not show enrichment in any region examined (Figure 7.2A). At 10 minutes of heat shock, Pol II is fairly evenly distributed across the gene (Figure 7.2B). dCDK12 shows enrichment in all regions occupied by Pol II (Figure 7.2A). This indicates dCDK12 interacts with the elongation complex. To determine if this pattern of localization is shared with other non-inducible genes, I examined dCDK12 localization under NHS conditions on constitutively active genes. At each gene examined, CDK12 was enriched over mock IPs (No Ab) (Figure 7.2D).

Interestingly, CDK12 distribution did not have the exact same pattern of localization as Pol II (Figure 7.2F). Pol II distribution showed a higher level on the 5' ends of the genes (Figure 7.2E), but CDK12 was relatively evenly distributed across the genes (Figure 7.2D). This is particularly evident on β-1-tubulin (56D) gene, which has a large peak of Pol II on the 5' end of the gene that has previously shown to be transcriptionally-engaged (Figure 7.2D-F). These results demonstrate that CDK12 is localized to actively transcribing genes in regions occupied by productively elongating Pol II.

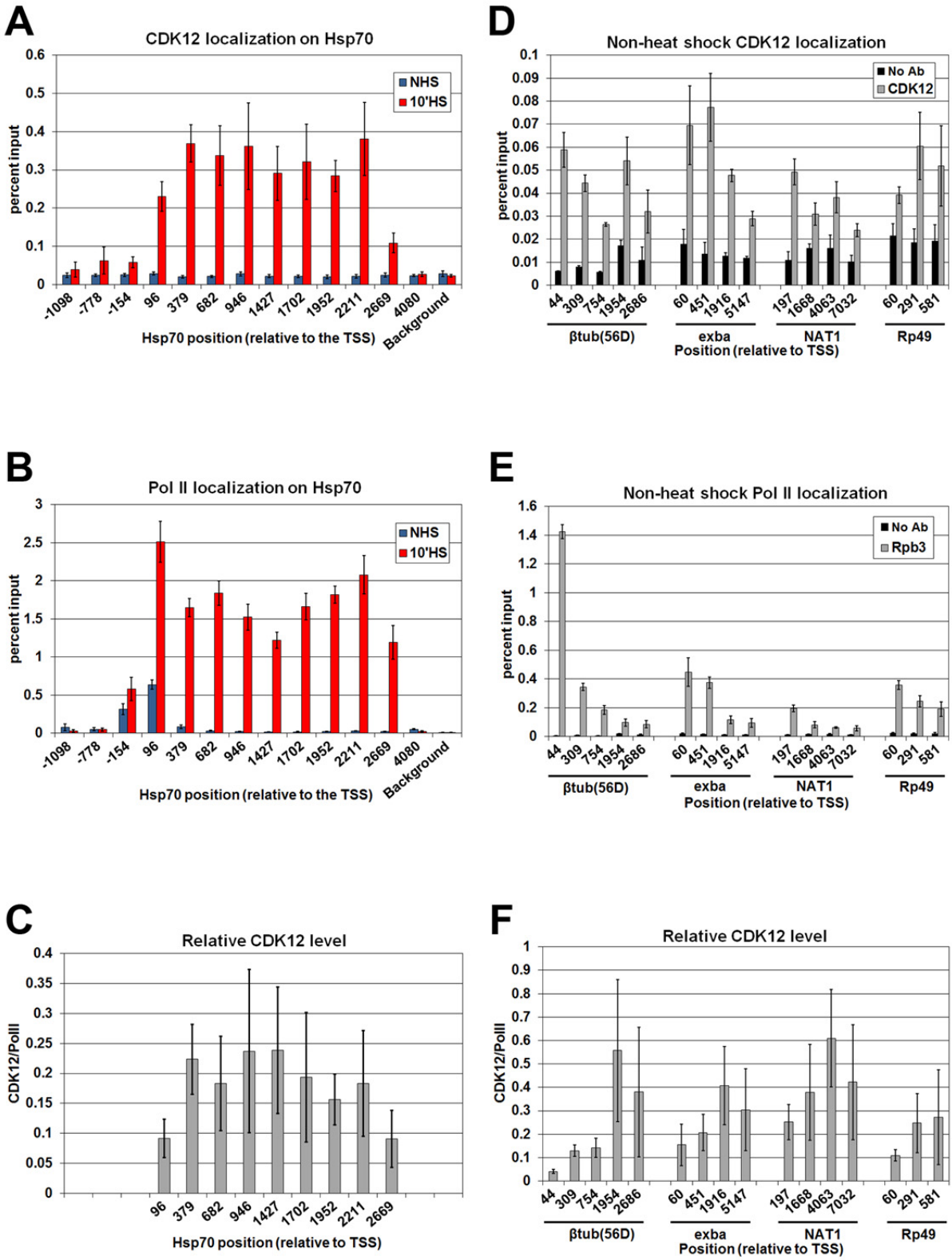

A
CDK12 localization on Hsp70
NHS
10'HS
percent input
Hsp70 position (relative to the TSS)
-1098
-778
-154
96
379
682
946
1427
1702
1952
2211
2669
4080
Background
B
Pol II localization on Hsp70
NHS
10'HS
percent input
Hsp70 position (relative to the TSS)
C
Relative CDK12 level
CDK12/PolII
Hsp70 position (relative to TSS)
D
Non-heat shock CDK12 localization
No Ab
CDK12
percent input
44
309
754
1954
2686
60
451
1916
5147
197
1668
4063
7032
60
291
581
βtub(56D)
exba
NAT1
Rp49
Position (relative to TSS)
E
Non-heat shock Pol II localization
No Ab
Rpb3
percent input
βtub(56D)
exba
NAT1
Rp49
Position (relative to TSS)
F
Relative CDK12 level
CDK12/PolII
βtub(56D)
exba
NAT1
Rp49
Position (relative to TSS)

Figure 7.2 ChIP analysis of Pol II and dCDK12. (A) ChIP analysis of dCDK12 on Hsp70 under non-heat shock (NHS) and 10 minute heat-shock (10'HS) conditions. **(B)** ChIP analysis of Pol II (anti-Rpb3) on *Hsp70*, as in A. **(C)** Ratios of dCDK12/RNAPII values on 10'HS *Hsp70*. **(D)** ChIP analysis of dCDK12 on four constitutively active genes under NHS conditions. **(E)** ChIP analysis of Pol II on four constitutively active genes under NHS conditions. **(F)** Ratios of dCDK12/RNAPII values on four constitutively active genes under NHS conditions. Measurements in A–F are averages of three biological replicates with standard errors.

7.3 Discussion

P-TEFb is critical for transition from pausing to productive elongation. Flavopiridol inhibition before heat shock restricts Pol II to the 5' end of the gene. Treatment after heat shock reduces Pol II on the body, but not the 5' end, of the gene. The pattern of Pol II and kinetics of Pol II loss after FP treatment suggest elongating polymerases continue to transcribe, but newly initiating polymerases cannot enter into productive elongation. The pattern of Pol II distribution after FP suggested these Pol II complexes may be identical to paused polymerases. Terminated run-ons demonstrated that these 5' polymerases were transcriptionally-engaged. Surprisingly, they did not just occupy the NHS pause positions, but were further downstream as well. Although these additional downstream sites occupied by Pol II after FP treatment may be alternate sites of pausing used under HS conditions, these polymerases may have been released from the NHS pause sites and be transcribing very inefficiently. This would indicate the transition into productive elongation is a two-step process, another factor releases the paused polymerase and then P-TEFb-modification of the elongation complex allows the complex to efficiently transcribe.

The metazoan ortholog of yeast Ctk1 was identified as CDK12/CDK13. CDK12 can phosphorylate the CTD in vitro, and RNAi depletion in vivo altered CTD phosphorylation in human and Drosophila (Bartkowiak et al., 2010). ChIP for dCDK12 in Drosophila S2 cells showed it localizes to transcriptionally active genes (Figure 7.2). Its recruitment to and distribution on *Hsp70* after heat shock induction suggests it localizes to elongating Pol II complexes. Interestingly, its pattern of localization is different from P-TEFb. dCDK12 shows an even distribution across HS *Hsp70*, but levels of P-TEFb

are highest at the 5' end of *Hsp70* (Boehm et al., 2003). The important role of P-TEFb in transition into productive elongation may be reflected in its enrichment towards the 5' end. Thus, the enrichment of dCDK12 on the body of gene may reflect a function distinct from P-TEFb that has been conserved from yeast. Yeast bur1 mutants have greater effects on Pol II levels than Serine 2 phosphorylation (Keogh et al., 2003), and the large effects on Serine 2 phosphorylation of ctk1 mutants do not reduce Pol II levels (Cho et al., 2001; Wood & Shilatifard, 2006).

Given the orthologous relationship, it will be interesting to see if P-TEFb and CDK12 have distinct targets in metazoans. P-TEFb can phosphorylate the Serine 2 residues on the CTD as well as Spt5 and Nelf-E subunit of the NELF complex (Fujinaga et al., 2004; Ivanov et al., 2000). The Spt5/Spt4 and NELF complexes are important for pausing (Gilchrist et al., 2008; C.-H. Wu et al., 2003), and the phosphorylation of Spt5 and NELF is proposed to trigger Pol II release from pausing (Wada et al., 1998; Yamada et al., 2006). Thus, P-TEFb may be responsible for triggering the transition from pausing to productive elongation, and CDK12 may be responsible for allowing efficient elongation by maintaining or adding to CTD hyperphosphorylation.

APPENDIX A

FCP1 DEPHOSPHORYLATION OF THE RNA POLYMERASE II C-TERMINAL DOMAIN IS REQUIRED FOR EFFICIENT TRANSCRIPTION OF HEAT SHOCK GENES[3]

A.1 Introduction

Proper temporal and spatial expression of RNA transcripts is vital to the development and health of all organisms. At the heart of eukaryotic transcription is RNA Polymerase II (Pol II), the enzyme that catalyzes the synthesis of RNA from a DNA template for protein-coding genes. Transcription is a cyclic process that can be divided into three distinct phases: initiation, elongation and termination (Saunders et al., 2006). During initiation, the Pol II complex assembles around the DNA at promoters and catalyzes the synthesis of the first phosphodiester bond in the gene's RNA transcript. Elongation involves the processive synthesis of the RNA transcript. Termination of the transcription cycle results in the release of both the nascent transcript and Pol II from the DNA template, and terminated Pol II can then be recycled for subsequent rounds of transcription.

The C-terminal domain (CTD) of the largest subunit of Pol II contains a series of heptad repeats (YSPTSPS) that are differentially modified during distinct phases of the

[3] The work in this chapter was a collaboration with Martin Buckley and has been accepted to Mol Cell Biol. (in press).

transcription cycle. CTD residues are targets of various modifications including methylation, phosphorylation, glycosylation and proline isomerization (Egloff & Murphy, 2008). The best studied of these CTD modifications is phosphorylation. In particular, phosphorylation on Serine 5, Serine 7, and Serine 2 of the CTD repeats is readily apparent during Pol II's progression though the transcription cycle (Chapman et al., 2007; Egloff & Murphy, 2008; Payne et al., 1989). Phosphorylation of Serine 5 occurs early in the cycle, between initiation and elongation, and is predominantly catalyzed by the Cdk7 kinase associated with the general transcription factor (GTF) TFIIH (Akhtar et al., 2009; Watanabe et al., 2000). Serine 7 is also phosphorylated early in the transcription cycle by Cdk7, but phosphorylation of this residue further increases toward the 3' end of genes (Chapman et al., 2007), mediated by the kinase Cdk9 (Akhtar et al., 2009; Tietjen et al., 2010). Serine 2 phosphorylation occurs at the transition into productive elongation, and can be catalyzed by two kinases: Cdk9 of P-TEFb (BUR1 in yeast) and Cdk12 (CTDK-I in yeast) (Bartkowiak et al., 2010; Nick F Marshall, Peng, Xie, Price, & Chem, 1996).

As a result of these modifications, the unphosphorylated Pol II (Pol IIa) that initiates transcription is radically transformed to the hyperphosphorylated Pol II (Pol IIo) that transcribes through the gene body during productive elongation (Laybourn & Dahmus, 1990; Payne et al., 1989). Importantly, these marks serve as a platform for the

recruitment of factors with functions relevant to particular stages in the transcription cycle (Egloff & Murphy, 2008). For example, early in the transcription cycle, the Serine 5 phosphorylated CTD is bound by the mRNA capping enzymes (Fabrega et al., 2003; Ghosh et al., 2011), and during elongation, the Serine 2 phosphorylated CTD is bound by several factors, including elongation factors (Liu et al., 2011), RNA processing factors (Morris & Greenleaf, 2000), and termination factors (Licatalosi et al., 2002; Meinhart & Cramer, 2004). Thus, the phosphorylated CTD serves as a scaffold for the timely recruitment of factors during the transcription cycle to ensure proper mRNA biogenesis.

Since unphosphorylated Pol II forms the pre-initiation complex, dephosphorylation of the CTD is critical for the recycling of terminated Pol II into a form that can initiate transcription (H. Cho et al., 1999). The mechanistic details of how termination interfaces with Pol II dephosphorylation are unknown (Buratowski, 2005), but the conversion of Pol IIo back to Pol IIa is catalyzed by CTD phosphatases. These phosphatases target different phosphorylated residues of the CTD repeat (Meinhart et al., 2005). The CTD phosphatases Rtr1, SCP1 and Ssu72 all target Serine 5 phosphorylation, and abrogation of Ssu72 leads to defects in transcription in yeast (Krishnamurthy et al., 2004; Mosley et al., 2009; Reyes-Reyes & Hampsey, 2007; Yeo et al., 2003). Fcp1 is an essential CTD phosphatase in yeast and *Drosophila*, and

although there is detailed information about how it binds Pol II (Chambers et al., 1995; Kamenski et al., 2004; Kimura et al., 2002; Suh et al., 2005), the target of Fcp1 is less clear. In vitro assays have implicated both Serine 2 and Serine 5 as possible targets (Hausmann et al., 2005; Lin et al., 2002), and Serine 2 has been shown to be the in vivo target in yeast (Cho et al., 2001).

Several studies have indicated that Fcp1 has a direct role in transcription. Both in vitro biochemical studies and in vivo studies in yeast have shown that Fcp1 dephosphorylation increases transcription (Cho et al., 2001; H. Cho et al., 1999; Kobor et al., 1999), and expression of *Drosophila* Fcp1 affected luciferase expression from reporter genes (Tombácz et al., 2009). Moreover, a role for Fcp1 in metazoan gene transcription in vivo is supported by a study showing FLAG-tagged Fcp1 co-localizes with bulk Rpb1 on *Drosophila* polytene chromsomes (Tombácz et al., 2009), and an RNAi screen identified *Drosophila* Fcp1 as an important factor in optimal *Hsp70* mRNA accumulation after heat shock (Ardehali et al., 2009). However, another study could not observe localization of the FLAG-tagged Fcp1 on the induced *Hsp70* gene by ChIP (Tombácz et al., 2009). To reconcile and extend these studies, we examine the role of Fcp1 in *Hsp70* gene regulation in vivo. Using immunostaining and chromatin immunoprecipitation (ChIP), we show that Fcp1 colocalizes with phosphorylated Pol II at active sites of transcription, including the induced *Hsp70* gene, in *Drosophila* polytene

chromosomes and S2 cell culture. Moreover, RNAi depletion of Fcp1 in S2 cells results in the loss of Pol II in the coding region of heat shock induced *Hsp70*. Intriguingly, this loss of Pol II signal correlates with a dramatic increase in phosphorylation of the non-chromatin bound Pol II, both of which are dependent on the catalytic activity of Fcp1. These findings indicate that the decrease in Pol II levels at *Hsp70* in Fcp1 depleted cells are a consequence of free phosphorylated Pol II that cannot be recycled for additional rounds of transcription.

A.2 Results

A.2.1 Fcp1 localizes to sites of active transcription

To investigate the role of Fcp1 in transcription, we generated an antibody to the previously identified *Drosophila* Fcp1 homolog, CG12252 (Tombácz et al., 2009). Immunoblots using the antibody detected one major band at the predicted size of 97 kDa (Figure A.4A). Additionally, this protein is depleted in Fcp1-RNAi cells, demonstrating the antibody recognizes Fcp1 (Figure A.4A). In order to assess the global distribution of Fcp1 at gene loci in vivo, *Drosophila* polytene chromosomes were immunostained for Fcp1 and phosphorylated Pol II (H14 monoclonal antibody). Fcp1 co-localized with phosphorylated Pol II at many interband loci, including developmental puffs at 2B, 23E, 74E, and 75B (Figure A.1A), although not always with the same

intensity. This agrees with previous results indicating Fcp1 localizes to most sites of active transcription (Tombácz et al., 2009).

Our previous work found that Fcp1 depletion decreases *Hsp70* transcript levels by ~50% as compared to RNAi control cells treated with β-galactosidase dsRNA (LacZ RNAi) (Ardehali et al., 2009). To further study the role of Fcp1 in HS gene regulation, Fcp1 and phosphorylated Pol II localization was examined on polytene chromosomes derived from salivary glands heat shocked at 37°C for various times. The fixed polytene chromosomes show that Fcp1 was recruited to the endogenous *Hsp70* genes at the 87A and C loci after heat shock, as well as to a transgenic *Hsp70* gene inserted at the 87E locus (arrows in Figure A.1B). Fcp1 immunostaining can be observed after 2 and 10 minutes of heat shock, albeit with reduced signal at 10 minutes (Figure A.1B). The weaker Fcp1 signal at 10 minutes of HS is likely due to decondensation of the loci, as 60 minutes of recovery after heat shock results in the return of a strong immunofluorescence signal (Figure A.2A). A similar pattern of recruitment was observed for GFP-Fcp1 in living cells (Figure A.2B,C). Interestingly, despite the fact that our previous work showed RNAi depletion of SCP1 and Ssu72 results in a modest effect on *Hsp70* gene expression (Ardehali et al., 2009), both factors are also recruited to active *Hsp70* loci (Figure A.3). Taken together, these immunostaining and live-cell imaging experiments indicate that Fcp1 localizes to transcribing *Hsp70* loci.

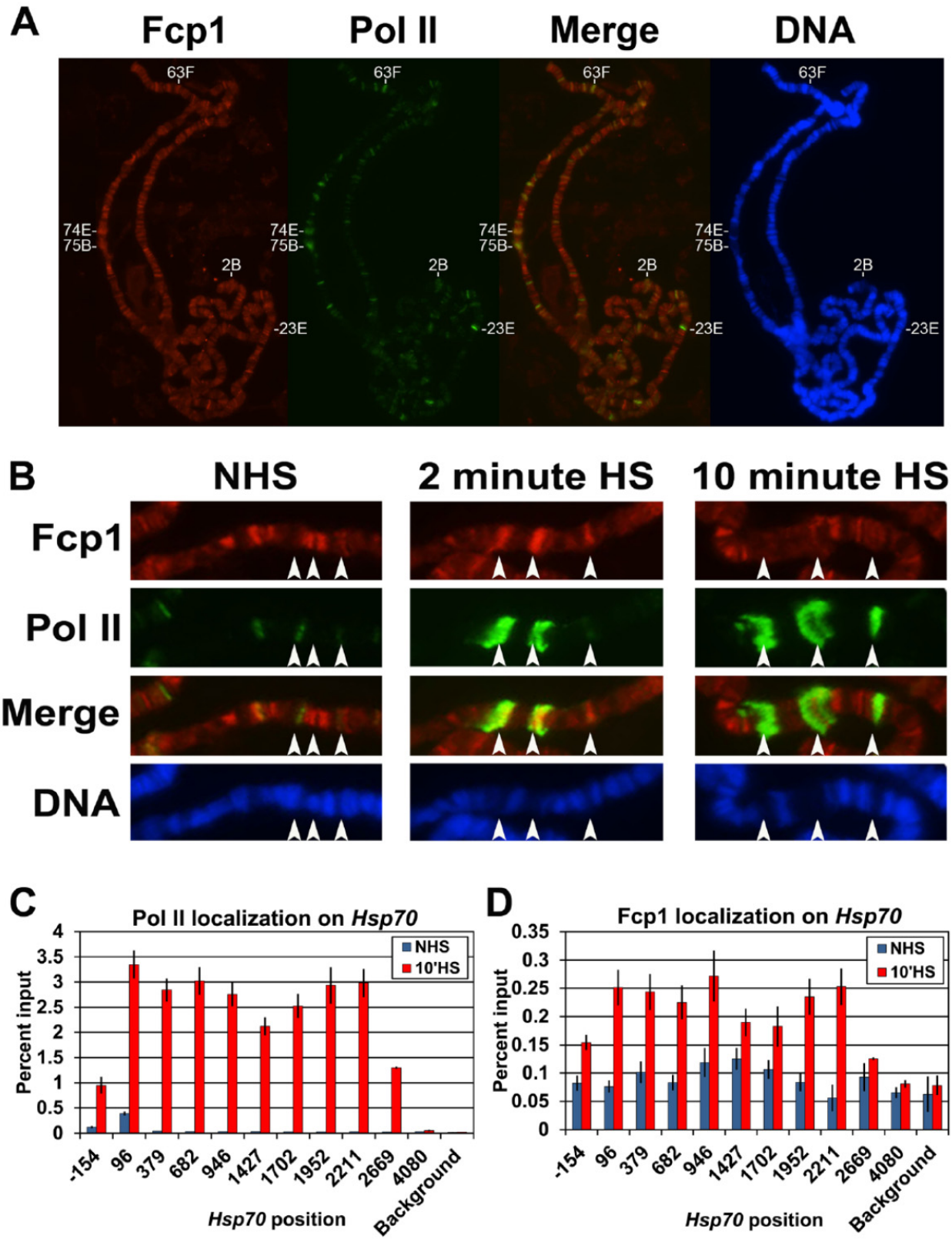

A
Fcp1
Pol II
Merge
DNA
63F
74E-
75B-
2B
-23E
B
NHS
2 minute HS
10 minute HS
Fcp1
Pol II
Merge
DNA
C
Pol II localization on Hsp70
Percent input
NHS
10'HS
-154
96
379
682
946
1427
1702
1952
2211
2669
4080
Background
Hsp70 position
D
Fcp1 localization on Hsp70
Percent input
NHS
10'HS
Hsp70 position

Figure A.1 Fcp1 localizes to transcriptionally active loci. (A-B) *Drosophila* spread polytene chromsomes immunostained with antibodies to Fcp1 (red) and Serine 5 phosphorylated Pol II CTD (H14 antibody, green). The DNA is stained with DAPI (blue). Merge is an overly of Fcp1 and Serine 5 phosphorylated Pol II CTD. **(A)** Chromosomes from NHS salivary glands. **(B)** *Hsp70* loci (87A and 87C (endogenous) and a single *Hsp70* transgene at 87E) from left to right in NHS, 2'HS, and 10'HS salivary glands are marked by arrows. **(C)** ChIP results showing the enrichment of Pol II (Rpb3) at the *Hsp70* gene in *Drosophila* S2 cells under NHS (blue bars) and 10'HS (red bars). **(D)** ChIP results of the Fcp1 enrichment on the *Hsp70* gene in *Drosophila* S2 cells under NHS (blue bars) and 10'HS (red bars). The x axes show the midpoint of each PCR fragment along *Hsp70* gene and the y axes shows the percentage of input DNA immunoprecipitated (error bars indicate the SEM of at least four biological replicates).

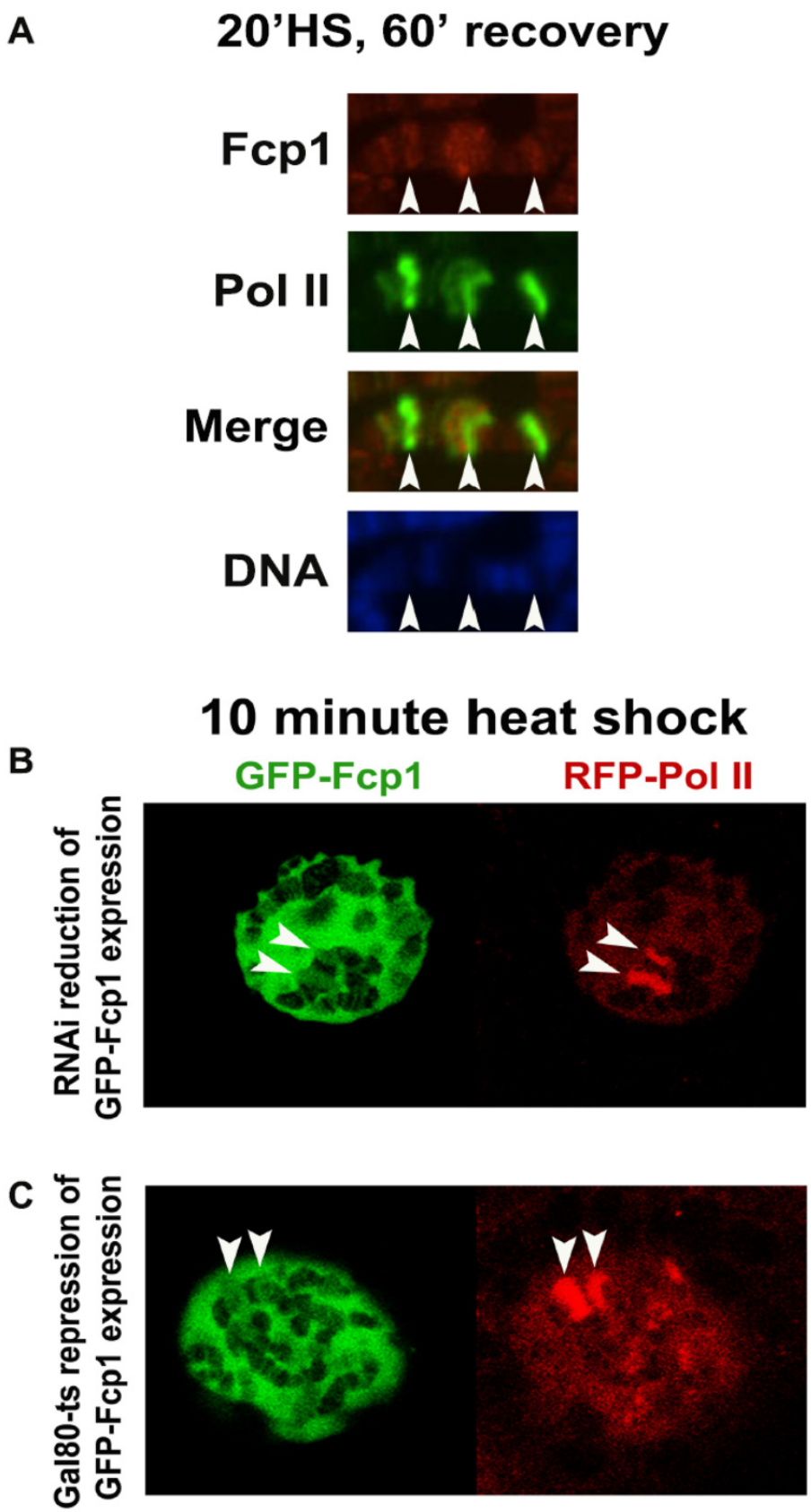

A
20'HS, 60' recovery
Fcp1
Pol II
Merge
DNA
10 minute heat shock
B
GFP-Fcp1
RFP-Pol II
RNAi reduction of GFP-Fcp1 expression
C
Gal80-ts repression of GFP-Fcp1 expression

Figure A2 Localization of Fcp1 at *Hsp70* loci after recovery from heat shock. (A) Immunofluorescence staining of polytene chromosomes with antibodies specific to Fcp1 (red) and Serine 5 phosphorylated Pol II CTD (H14 antibody, green) after a 10 minute HS and 60 minutes recovery at room temperature. The arrows indicate the *Hsp70* loci (87A and 87C (endogenous) and a single *Hsp70* transgene at 87E) from left to right. The DNA is stained with DAPI (blue). Merge is an overly of Fcp1 and Pol II. **(B,C)** Laser scanning confocal microscopy images of polytene nuclei co-expressing GFP-Fcp1 (green) and mRFP-Pol II (red) at 10 minute HS. Overexpression of GFP-Fcp1 in salivary glands using the Gal4-UAS system results in a small salivary gland phenotype that does not allow imaging (data not shown). In order to image GFP-Fcp1 in living cells, Fcp1 protein levels were reduced by RNAi depletion (**B**, GFP-Fcp1/+;Rpb3RN, 6983/WizFcp1RNAi or using the Gal80-ts repressor (**C**, GFP-Fcp1/+;Rpb3RN, 6983/7017) grown at room temperature. Arrows highlight the *Hsp70* loci.

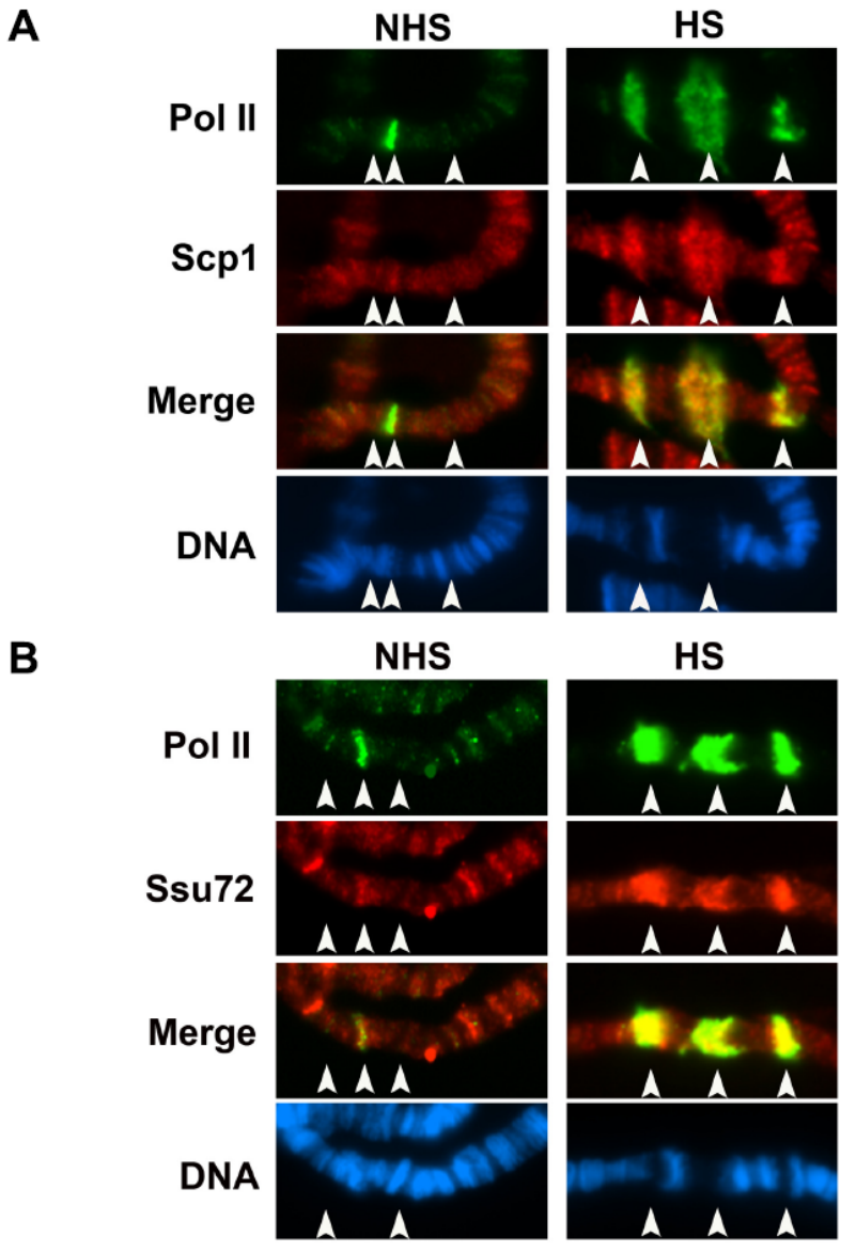


Figure A.3 CTD phosphatases Scp1 and Ssu72 localize to active *Hsp70* loci. Immunostaining of polytene chromosomes with **(A)** Scp1 or **(B)** Ssu72 antibodies (red) and Serine 5 phosphorylated Pol II CTD (H14 antibody, green) at the *Hsp70* loci (87A and 87C (endogenous) and a single *Hsp70* transgene at 87E) from left to right (indicated with the arrows) under NHS and HS conditions. The DNA is stained with DAPI (blue). Merge is an overlay of Ssu72 or Scp1 with Pol II.

In order to assess the distribution of Fcp1 at higher resolution, we performed ChIP for Fcp1 at the uninduced and induced *Hsp70* gene in *Drosophila* S2 cells. Under NHS conditions, Fcp1 is not enriched on the *Hsp70* gene in comparison to the signal in a background region 30Kb away from the *Hsp70* gene. In contrast, at 10 minutes of HS, Fcp1 is enriched on the transcribed region of *Hsp70* as compared to either region downstream of the transcribing Pol II (4080) or the background region (Figure A.1D). Fcp1 localizes evenly across the *Hsp70* gene at 10 minutes of heat shock, and the pattern of Fcp1 enrichment is similar to that of Pol II (Figure A.1C). These findings suggest that Fcp1 associates with the elongating Pol II complex, similar to results from *Saccharomyces cerevisiae* (Cho et al., 2001).

A.2.2 Fcp1 depletion affects transcription of *Hsp70* during heat shock

Previously, we showed that RNAi knock-down of *Drosophila* Fcp1 resulted in a 2-fold reduction in *Hsp70* mRNA accumulation after 20 minutes of heat shock (Ardehali et al., 2009). To further characterize this effect, we performed a heat shock time course in control and Fcp1 knock-down cells, and examined the level of mRNA from three heat shock genes: *Hsp70*, *Hsp26*, and *Hsp83*. RNAi-depletion of Fcp1 was performed using a dsRNA targeting the fifth exon of Fcp1 (region A in Figure A.5A). RNAi knock-down reduced Fcp1 protein levels by at least 90% as assayed by Western blot (Figure A.4A). In agreement with the previous work, Fcp1 knock-down reduces *Hsp70* mRNA levels 2-

3 fold at heat shock time points of 5 minutes or longer (Figure A.4B). *Hsp26* mRNA accumulation is similarly affected by Fcp1 knock-down (Figure A.4C), and *Hsp83* mRNA accumulation is reduced, but less so (Figure A.4D).

The localization of Fcp1 on *Hsp70* during heat shock suggests that these Fcp1 knock-down effects on *Hsp70* mRNA levels may be due to direct effects on transcription. To investigate this, we used ChIP to assay Pol II localization at the active *Hsp70* gene in control and Fcp1-depleted cells. Compared to Untreated or LacZ-RNAi control cells, Fcp1 knock-down results in a reduction of Pol II throughout the *Hsp70* transcription unit at 10 minutes of heat shock (Figure A.4E). The Pol II ChIP signal is slightly more reduced toward the 3' end of the gene (from about 40% at 5' end to about 60% at the 3' end). Additionally, ChIP for Fcp1 showed the knock-down reduced Fcp1 levels on the *Hsp70* gene close to levels at the background region (Figure A.5D). Fcp1 knock-down also results in a reduction of Pol II levels on the transcription unit of induced *Hsp26* and *Hsp83* (Figure A.4F,G). Notably, the loss of Pol II signal is similar to the decrease in *Hsp70* mRNA levels observed in Fcp1-RNAi cells (Ardehali et al., 2009). The comparable decrease in *Hsp70* mRNA accumulation and Pol II levels indicates that Fcp1 knock-down affects transcription directly.

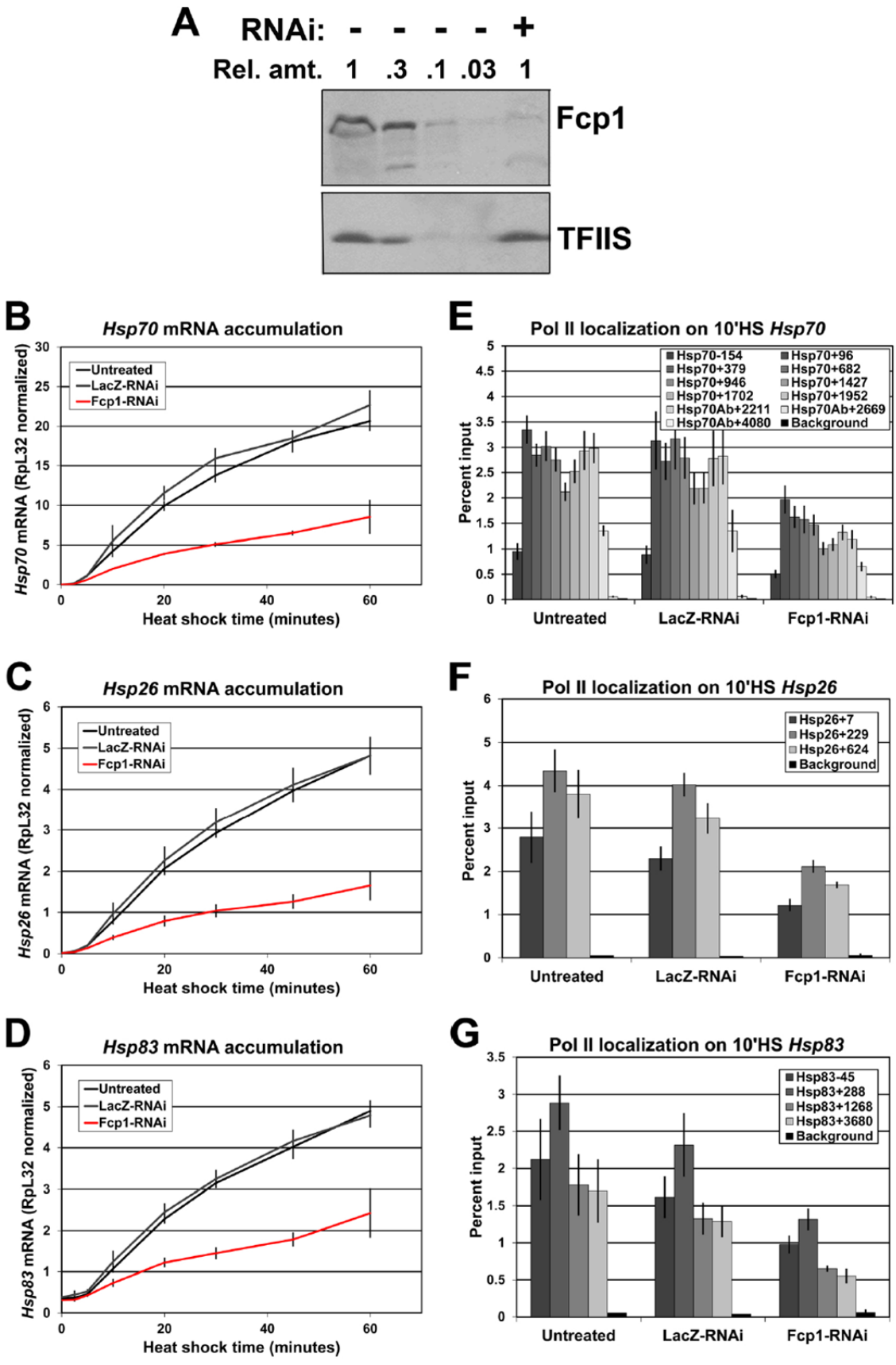
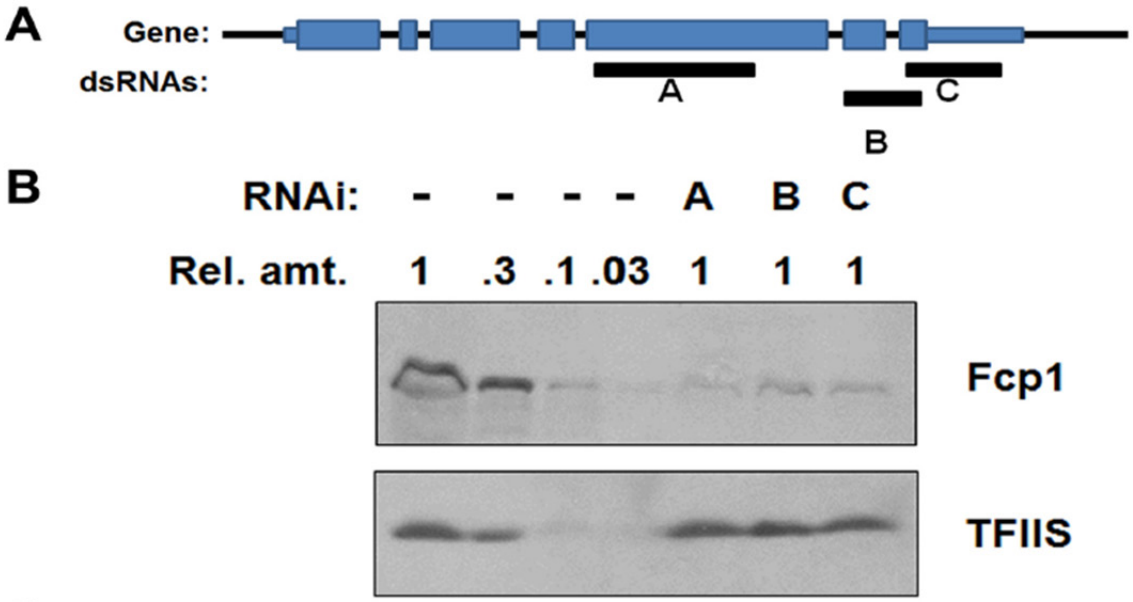
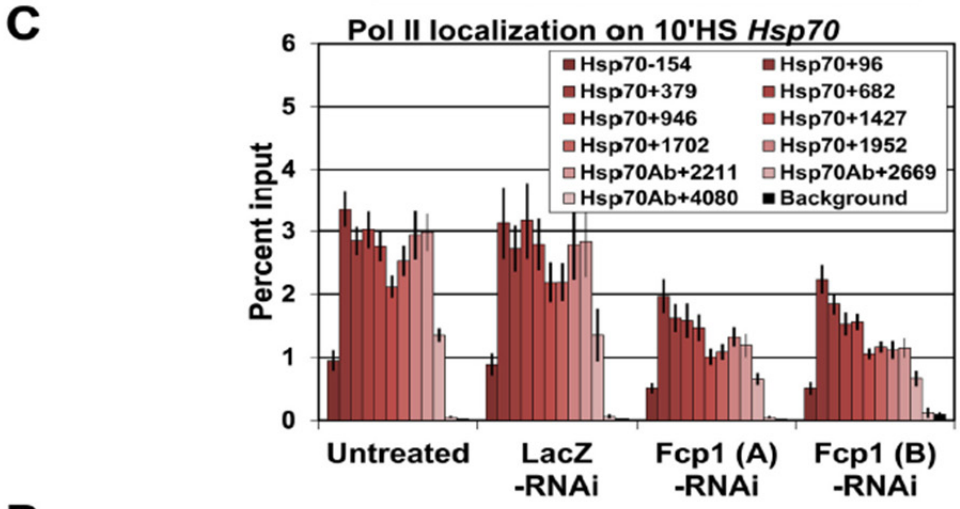

A
RNAi: - - - - +
Rel. amt. 1 .3 .1 .03 1
Fcp1
TFIIS
B
Hsp70 mRNA accumulation
Hsp70 mRNA (RpL32 normalized)
Heat shock time (minutes)
Untreated
LacZ-RNAi
Fcp1-RNAi
C
Hsp26 mRNA accumulation
Hsp26 mRNA (RpL32 normalized)
Heat shock time (minutes)
Untreated
LacZ-RNAi
Fcp1-RNAi
D
Hsp83 mRNA accumulation
Hsp83 mRNA (RpL32 normalized)
Heat shock time (minutes)
Untreated
LacZ-RNAi
Fcp1-RNAi
E
Pol II localization on 10'HS Hsp70
Percent input
Hsp70-154
Hsp70+96
Hsp70+379
Hsp70+682
Hsp70+946
Hsp70+1427
Hsp70+1702
Hsp70+1952
Hsp70Ab+2211
Hsp70Ab+2669
Hsp70Ab+4080
Background
Untreated
LacZ-RNAi
Fcp1-RNAi
F
Pol II localization on 10'HS Hsp26
Percent input
Hsp26+7
Hsp26+229
Hsp26+624
Background
Untreated
LacZ-RNAi
Fcp1-RNAi
G
Pol II localization on 10'HS Hsp83
Percent input
Hsp83-45
Hsp83+288
Hsp83+1268
Hsp83+3680
Background
Untreated
LacZ-RNAi
Fcp1-RNAi

Figure A.4 Fcp1 depletion by RNAi diminishes the levels of Pol II on *Hsp* genes. **(A)** Western blots of whole cell extracts from control (-) and Fcp1-RNAi (+) cells probed with antibodies for Fcp1 (lab stock, 1:1000) and TFIIS (lab stock loading control, 1:3000). The relative amount loaded is indicated (1=1x10^6cells). **(B-D)** RT-qPCR results for heat shock time course in Untreated, LacZ-RNAi, and Fcp1-RNAi cells. Total RNA was reverse transcribed with olgo(dT) and amplified with primer sets to the **(B)** *Hsp70*, **(C)** *Hsp26*, and **(D)** *Hsp83* genes. (E-G) ChIP results for the Pol II subunit Rpb3 in Untreated, LacZ-RNAi and Fcp1-RNAi S2 cells at 10'HS on the **(E)** *Hsp70*, **(F)** *Hsp26*, and **(G)** *Hsp83* genes. The legend indicates the midpoint of each PCR fragment. The y axes shows the percentage of input DNA immunoprecipitated (error bars indicate SEM of at least three biological replicates).

To ensure the effects seen are a result of Fcp1 knock-down and not due to depletion of an unintended target, we depleted Fcp1 using a different dsRNA targeting a non-overlapping region of the gene (region B in Figure A.5A). The new dsRNA showed comparable knock-down of Fcp1 (Figure A.5B, A compared to B), and a similar reduction in Pol II levels (Figure A.5C). Given the highly unlikely overlap of any possible unintended targets for these two dsRNAs, this indicates the effects seen are due to Fcp1 depletion.

We next investigated whether Fcp1 depletion also perturbs levels of the promoter-proximally paused Pol II. To test this, we used ChIP to examine the distribution of Pol II on *Hsp70* in Fcp1 depleted cells under non-heat shock (NHS) conditions. We did not observe any effect of Fcp1 knockdown on the level of paused Pol II at *Hsp70* in uninduced cells (Figure A.6A).


A
Gene:
dsRNAs:
A
B
C
B
RNAi: - - - - A B C
Rel. amt. 1 .3 .1 .03 1 1 1
Fcp1
TFIIS


C
Pol II localization on 10'HS Hsp70
Percent input
Hsp70-154
Hsp70+96
Hsp70+379
Hsp70+682
Hsp70+946
Hsp70+1427
Hsp70+1702
Hsp70+1952
Hsp70Ab+2211
Hsp70Ab+2669
Hsp70Ab+4080
Background
Untreated
LacZ -RNAi
Fcp1 (A) -RNAi
Fcp1 (B) -RNAi

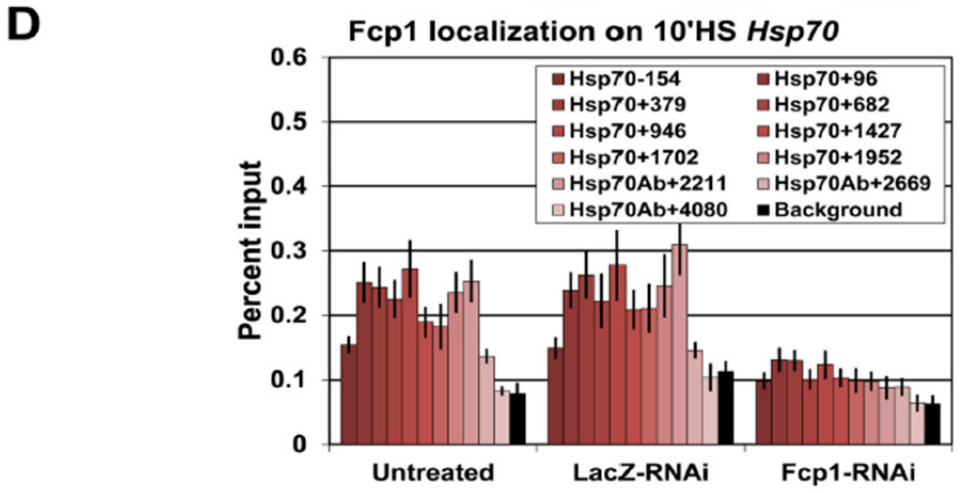

D
Fcp1 localization on 10'HS Hsp70
Percent input
Hsp70-154
Hsp70+96
Hsp70+379
Hsp70+682
Hsp70+946
Hsp70+1427
Hsp70+1702
Hsp70+1952
Hsp70Ab+2211
Hsp70Ab+2669
Hsp70Ab+4080
Background
Untreated
LacZ-RNAi
Fcp1-RNAi

Figure A.5 Multiple dsRNAs result in similar depletion of Fcp1. (A) The location of the three dsRNAs used to knock-down Fcp1 (A: +1502 to +2290, B: +2720 to +3103, C: +3017 to +3486, relative to the TSS). The blue areas represent the region included in the mRNA, with the thicker regions indicating the coding region and the narrow regions indicating the UTRs. **(B)** Western blots of whole cell extracts from control (-) and each Fcp1 RNAi (A-C) probed with antibodies for Fcp1 (lab stock, 1:1000) and TFIIS (lab stock loading control, 1:3000). The relative amount loaded is indicated (1=1x10^6cells). **(C)** ChIP results for the Pol II subunit Rpb3 enrichment on the *Hsp70* gene in Untreated, LacZ-RNAi and Fcp1-RNAi (A and B dsRNAs) S2 cells at 10 minute HS. **(D)** ChIP results for the Fcp1 enrichment on the *Hsp70* gene in Untreated, LacZ-RNAi and Fcp1-RNAi (A dsRNA) S2 cells at 10 minute HS. The legends indicate the midpoint of each PCR fragment (*Hsp70Ab* indicates primer sets specific for the Ab copy of *Hsp70*). The error bars indicate the SEM of at least three biological replicates.

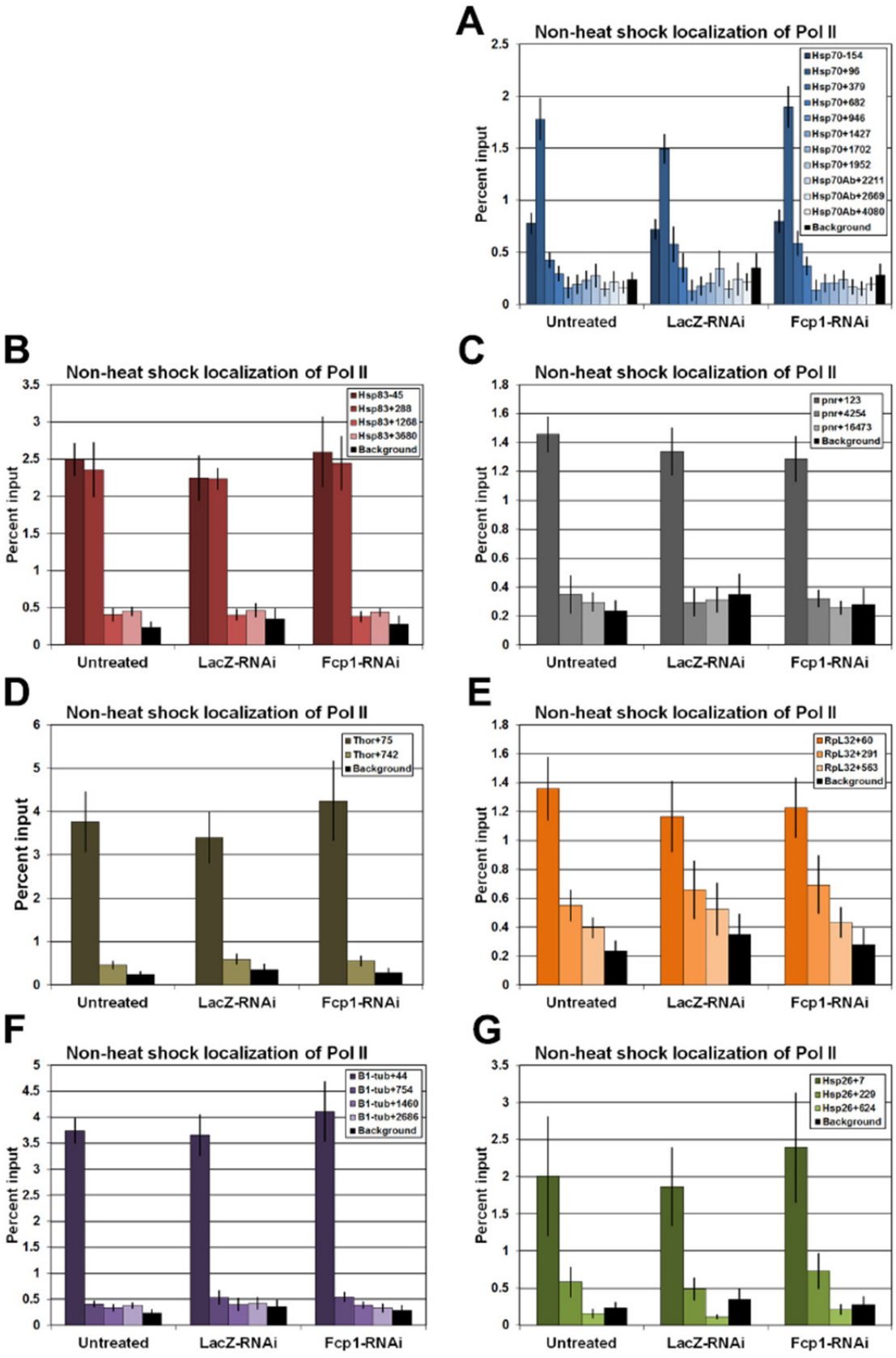

A
Non-heat shock localization of Pol II
Percent input
Hsp70-154
Hsp70+96
Hsp70+379
Hsp70+682
Hsp70+946
Hsp70+1427
Hsp70+1702
Hsp70+1952
Hsp70Ab+2211
Hsp70Ab+2669
Hsp70Ab+4080
Background
Untreated
LacZ-RNAi
Fcp1-RNAi
B
Non-heat shock localization of Pol II
Percent input
Hsp83-45
Hsp83+288
Hsp83+1268
Hsp83+3680
Background
Untreated
LacZ-RNAi
Fcp1-RNAi
C
Non-heat shock localization of Pol II
Percent input
pnr+123
pnr+4254
pnr+16473
Background
Untreated
LacZ-RNAi
Fcp1-RNAi
D
Non-heat shock localization of Pol II
Percent input
Thor+75
Thor+742
Background
Untreated
LacZ-RNAi
Fcp1-RNAi
E
Non-heat shock localization of Pol II
Percent input
RpL32+60
RpL32+291
RpL32+563
Background
Untreated
LacZ-RNAi
Fcp1-RNAi
F
Non-heat shock localization of Pol II
Percent input
B1-tub+44
B1-tub+754
B1-tub+1460
B1-tub+2686
Background
Untreated
LacZ-RNAi
Fcp1-RNAi
G
Non-heat shock localization of Pol II
Percent input
Hsp26+7
Hsp26+229
Hsp26+624
Background
Untreated
LacZ-RNAi
Fcp1-RNAi

Figure A.6 Constitutively expressed genes are not detectably affected by Fcp1 knock-down. (A-E) ChIP results for the Pol II subunit Rpb3 enrichment in Untreated, LacZ-RNAi and Fcp1-RNAi S2 cells at the **(A)** *Hsp70*, **(B)** *Hsp83*, **(C)** *pnr*, **(D)** *Thor*, **(E)** *RpL32*, **(F)** *β-1-tubulin*, and **(G)** *Hsp26* genes under NHS conditions. The legends indicate the midpoint of each PCR fragment (*Hsp70Ab* indicates primer sets specific for the Ab copy of *Hsp70*). The error bars indicate the SEM of at least three biological replicates.

A.2.3 Constitutively expressed genes are not detectably affected by Fcp1 depletion

Given the effect of Fcp1 depletion on transcription of *Hsp70* during heat shock, we investigated whether Fcp1 depletion also affects transcription of constitutively expressed genes under NHS conditions. We performed ChIP for Pol II in NHS control and Fcp1-RNAi cells. Surprisingly, we failed to see significant changes in Pol II levels on any genes in Fcp1 depleted cells, even at highly expressed genes (*Hsp83*, *Thor*) or moderately expressed genes (*RpL32*, *β-1-tubulin*, *pnr*, *Hsp26*, *Hsp70*) (Figure S5B-G). To exhaustively investigate constitutively expressed genes, we also performed GRO-seq in control LacZ-RNAi and Fcp1-RNAi cells to comprehensively quantify the transcriptionally-engaged polymerases genome-wide. Comparison of biological replicates for LacZ-RNAi control and Fcp1-RNAi cells failed to identify any genes with significantly reduced polymerase levels in Fcp1 depleted cells, and only 7 genes (*T48*, *Appl*, *mfas*, *GlcAT-P*, *amon*, *corn*, *Rgk1*) had increased polymerase levels (Figures A.7). Moreover, it was also surprising to find that under NHS conditions Fcp1 depletion did not influence the expression of highly transcribed genes (according to GRO-seq gene body read density). The observed effects on heat shock-induced genes could be due to a requirement for Fcp1 under heat shock conditions. To investigate this, we

examined Pol II on the constitutively expressed genes at 10 minutes of heat shock. Although the levels of Pol II on these genes is lower due to a general shut down of transcription during heat shock, the Pol II levels are comparable for control and Fcp1 depleted cells (Figure A.8). Thus, at this level of Fcp1 depletion, transcription is impaired on only the extremely highly expressed heat shock-induced genes.

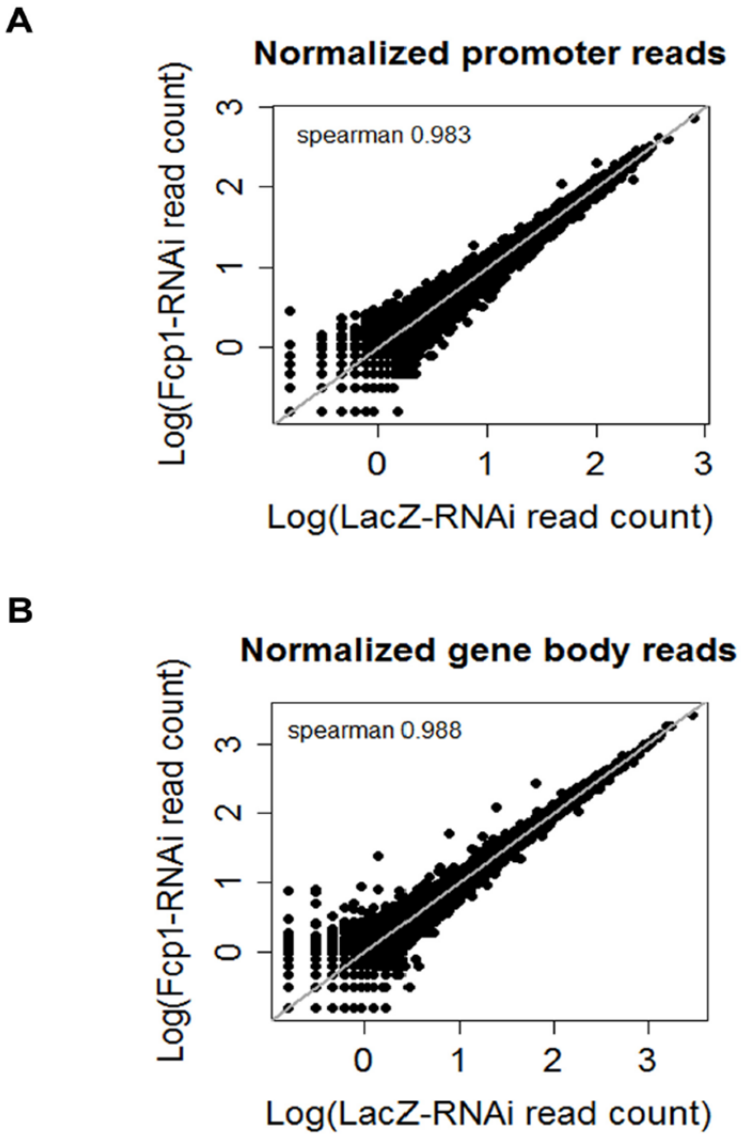


Figure A.7 A genome-wide assay shows constitutively expressed genes are not detectably affected by Fcp1 knock-down. GRO-seq was performed in LacZ-RNAi and Fcp1-RNAi cells under NHS conditions to determine the levels of transcriptionally engaged polymerase. **(A)** Normalized read density (log10) from global run-on sequencing (GRO-seq) in promoter regions from LacZ-RNAi and Fcp1-RNAi cells under NHS conditions. The gray line indicates a one-to-one ratio of reads. The spearman correlation between the two datasets is 0.983. **(B)** Normalized GRO-seq read density (log10) in the gene body from LacZ-RNAi and Fcp1-RNAi cells under NHS conditions. The gray line indicates a one-to-one ratio of reads. The spearman correlation between the two datasets is 0.988. Edge-R was used to identify genes with significantly different read counts between LacZ-RNAi and Fcp1-RNAi, and no genes had significantly lower read counts in either the promoter or gene body regions of the Fcp1-RNAi library.

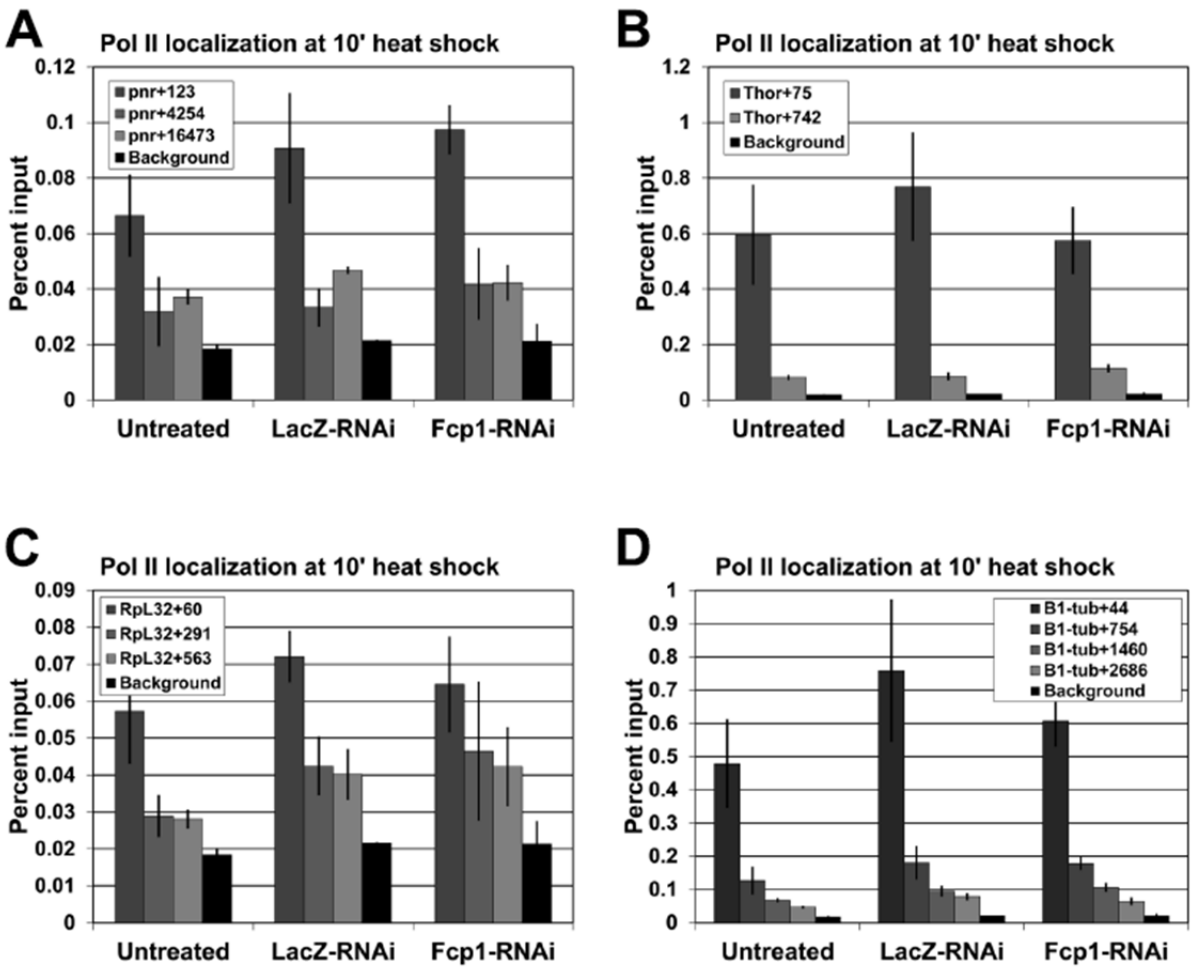


Figure A.8 Constitutively expressed genes are not detectably affected by Fcp1 knock-down during heat shock. ChIP results for the Pol II subunit Rpb3 enrichment on the **(A)** *pnr*, **(B)** *Thor*, **(C)** *RpL32*, **(D)** *β-1-tubulin* genes at 10 minutes of heat shock in Untreated, LacZ-RNAi, and Fcp1-RNAi S2 cells. The legends indicate the midpoint of each PCR fragment. The error bars indicate the SEM of at least three biological replicates.

A.2.4 Fcp1 depletion results in an increase in CTD phosphorylation of unengaged Pol II

Next, we used Western blots to examine the CTD phosphorylation level in Fcp1-RNAi cells for Serine 5 and Serine 2 phosphorylation (using the 3E10 and 3E8 monoclonal antibodies, respectively). Surprisingly, there were no dramatic changes in the overall levels of either epitope in whole cell extracts (Figure A.9A). Previous studies have shown that hypophosphorylated Pol IIa initiates transcription (Chesnut et al., 1992; Kang & Dahmus, 1993; Laybourn & Dahmus, 1990; Lu et al., 1991). Therefore, we next investigated whether the level of terminated non-chromatin bound (free) unphosphorylated Pol II is reduced when Fcp1 is depleted. To do this, we examined free and chromatin-bound fractions of Pol II for changes in CTD phosphorylation in Fcp1-RNAi treated cells. The free (cytoplasmic and nucleoplasmic) proteins were separated from chromatin-bound proteins with a modified version of a previously developed procedure (Aygün et al., 2008). As expected, Histone H3 is enriched on chromatin fraction, and triose phosphate isomerase (TPI) is enriched in the free fraction (Figure A.9B). In addition, chromatin-bound Pol II in control cells had high levels of both Serine 5 and Serine 2 phosphorylation, and free Pol II had extremely low levels of phosphorylation (Figure A.9B lanes 5 and 1 respectively). Although Fcp1 knock-down did not dramatically change levels of chromatin-bound phosphorylated Pol II (Figure

A.9B lane 9 compared to lanes 6 and 8), knock-down did increase the levels of free phosphorylated Pol II (Figure A.9B lanes 3 and 4 compared to lanes 1 and 2). Similar levels of the Pol II subunit, Rpb3, show that the increase in phosphorylated CTD was not due to an increase in overall free Pol II in the Fcp1-RNAi cells (Figure A.9B lanes 1 through 3). Interestingly, both Serine 5 and Serine 2 phosphorylation increased, indicating Fcp1 is important for dephosphorylation of Serine 2 and Serine 5 in vivo (Figure A.9B). These results indicate that changes in the phosphorylation of free Pol II constitute a small fraction of the total phosphorylated Pol II in the cell.

Although fractionation showed that the level of chromatin-bound Pol II was unaffected under NHS conditions, we next investigated whether the Pol II reduction on the *Hsp70* gene body in Fcp1 depleted cells might be associated with abnormal Pol II phosphorylation levels on the gene during heat shock. ChIP using antibodies to Serine 5 and Serine 2 phosphorylated CTD (using the H14 and H5 monoclonal antibodies, respectively) showed reduced levels of phosphorylated Pol II across *Hsp70* at 10 minutes of heat shock, comparable to the Pol II reduction. We also see a similar reduction in Serine 5 phosphorylated CTD and Serine 2 phosphorylated CTD using the 3E8 and 3E10 antibodies, respectively (data not shown). Therefore, Pol II-normalized phosphorylation levels of both Serine 5 and Serine 2 showed no significant change in any region of *Hsp70* (Figure A.9C,D). The relatively uniform reduction of all forms of Pol

II across *Hsp70* in Fcp1 knock-down cells indicates that Pol II modifications during elongation occurred normally and suggests that it is the Pol II initiation rate that is affected in induced cells by Fcp1 knockdown (Figure A.4D).

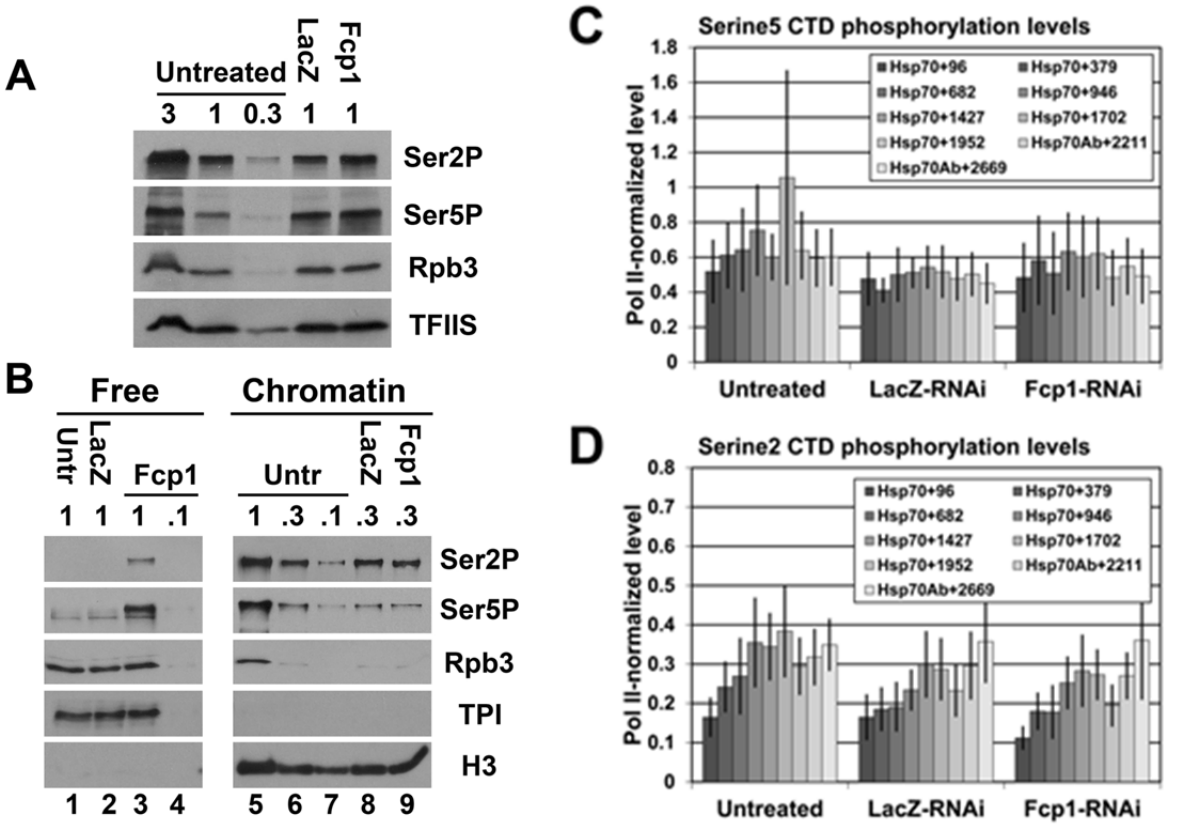


Figure A.9 Fcp1 knock-down does not significantly change phosphorylation level of Pol II on the *Hsp70* gene. (A) Western blots of whole cell extracts from Untreated, LacZ-RNAi, and Fcp1-RNAi cells probed with antibodies for phosphorylated CTD Serine 2 (EMD Millipore 3E10, 1:250), phosphorylated CTD Serine 5 (EMD Millipore 3E8, 1:250), Rpb3 (lab stock loading control, 1:1000) and TFIIS (lab stock loading control, 1:3000). The relative amount loaded is indicated (1=6x10^5 cells). **(B)** Western blots of free and chromatin-bound protein fractions from Untreated, LacZ-RNAi, and Fcp1-RNAi cells probed with antibodies for phosphorylated CTD Serine 5 (EMD Millipore 3E8, 1:250), phosphorylated CTD Serine 2 (EMD Millipore 3E10, 1:250), Rpb3 (lab stock loading control , 1:1000), Triosphosphatase isomerase (lab stock loading control, 1:1000) and Histone H3 (abcam ab1791, 1:500). The relative amount loaded is indicated (1=1x10^6 cells). **(C-D)** ChIP results of the **(C)** Serine 5 and **(D)** Serine 2 phosphorylated Pol II CTD enrichment relative to Pol II enrichment on the *Hsp70* gene in Untreated, LacZ-RNAi and Fcp1-RNAi S2 cells at 10'HS. The legend indicates the midpoint of each PCR fragment. The y axes shows the percentage of input DNA immunoprecipitated (error bars indicate SEM of at least three biological replicates).

A.2.5 Transcription defects of Fcp1 depletion are dependent on Fcp1 phosphatase activity

Previous work has shown that some functions of yeast Fcp1 can occur independently of its catalytic activity (H. Cho et al., 1999). Therefore, the effects seen in our various assays could be due to the loss of the Fcp1 phosphatase activity or loss of the protein itself, independent of its catalytic function. Fcp1 is the founding member of the FCPH family of phosphatases, which contain a highly conserved DXDX(T/V) active site. Mutation of either aspartate residue abolishes Fcp1 phosphatase activity (Hausmann & Shuman, 2002). To test if the effects we see were dependent on the phosphatase activity of Fcp1, we stably transfected a copper-inducible FLAG-tagged transgene with either a wild type Fcp1 or a catalytically dead version (second catalytic aspartate was mutated to asparagine) into S2 cells. A dsRNA targeting the Fcp1 3'UTR (region C in Figure A.5A) was used to knock-down endogenous Fcp1 to similar levels as the other dsRNAs (Figure A.5B). RNAi-resistant wild type or mutant versions of Fcp1 were then re-expressed by addition of $CuSO_4$ to the cell culture media (Figure A.10A-C). We examined Pol II distribution in Untreated and RNAi +/-$CuSO_4$ cultures. In all cell lines, 3'UTR-RNAi depletion reduced Pol II levels on heat shock-induced *Hsp70* similar to the other dsRNAs (Figure A.10D-F). Re-expression of the wild type Fcp1 partially restored the Fcp1 knock-down 10'HS *Hsp70* Pol II to untreated levels (Figure A.10E). In

contrast, neither an empty vector control nor the catalytically dead version restored Pol II levels (Figure A.10D,F). Both *Hsp26* and *Hsp83* show some rescue upon re-expression of the wild type Fcp1, but not the catalytically dead mutant (Figure A.11). Similar to the other dsRNAs, 3'UTR-RNAi depletion also increased the phosphorylated free Pol II (Figure A.10G-I, middle lane). Cells re-expressing the wild type had levels of phosphorylated free Pol II similar to Untreated cells (Figure A.10H, right lane), but interestingly, cells re-expressing the catalytically dead Fcp1 further increased the amount of phosphorylated free Pol II above Fcp1 knock-down alone (Figure A.10I, right lane). To determine if this additional increase in phosphorylation of free Pol II has an effect on the transcription of constitutively expressed genes under NHS conditions, we performed ChIP for Pol II under NHS conditions in cells re-expressing the mutant Fcp1, but we did not see any effect (Figure A.12). The rescue of Pol II levels on heat shock induced genes by wild type Fcp1, but not the catalytic mutant, demonstrates that the effects of Fcp1 knock-down are due to loss of the Fcp1 phosphatase activity.

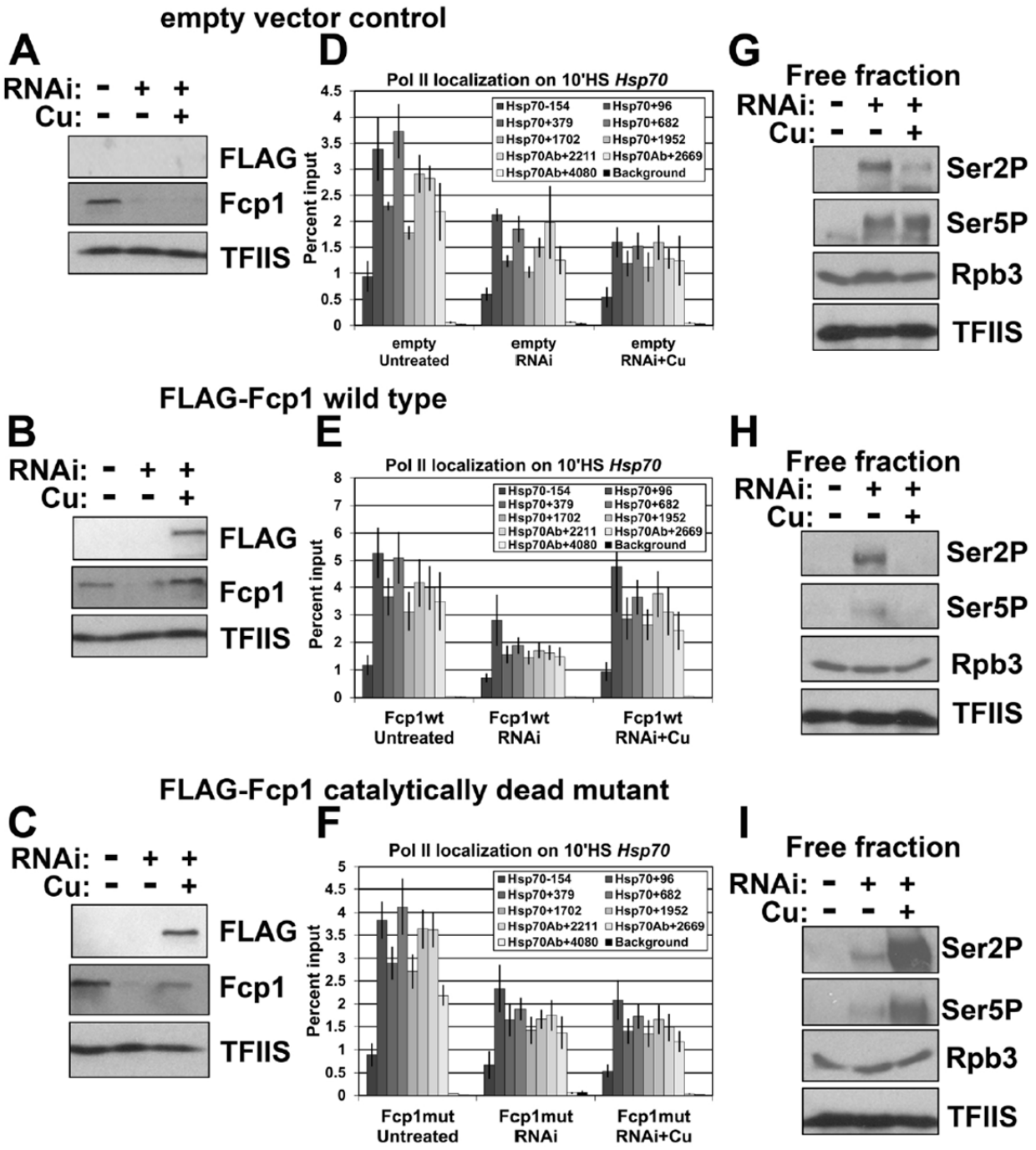

empty vector control
A
RNAi: - + +
Cu: - - +
FLAG
Fcp1
TFIIS
D
Pol II localization on 10'HS Hsp70
Percent input
Hsp70-154
Hsp70+96
Hsp70+379
Hsp70+682
Hsp70+1702
Hsp70+1952
Hsp70Ab+2211
Hsp70Ab+2669
Hsp70Ab+4080
Background
empty Untreated
empty RNAi
empty RNAi+Cu
G
Free fraction
RNAi: - + +
Cu: - - +
Ser2P
Ser5P
Rpb3
TFIIS
FLAG-Fcp1 wild type
B
RNAi: - + +
Cu: - - +
FLAG
Fcp1
TFIIS
E
Pol II localization on 10'HS Hsp70
Percent input
Fcp1wt Untreated
Fcp1wt RNAi
Fcp1wt RNAi+Cu
H
Free fraction
RNAi: - + +
Cu: - - +
Ser2P
Ser5P
Rpb3
TFIIS
FLAG-Fcp1 catalytically dead mutant
C
RNAi: - + +
Cu: - - +
FLAG
Fcp1
TFIIS
F
Pol II localization on 10'HS Hsp70
Percent input
Fcp1mut Untreated
Fcp1mut RNAi
Fcp1mut RNAi+Cu
I
Free fraction
RNAi: - + +
Cu: - - +
Ser2P
Ser5P
Rpb3
TFIIS

Figure A.10 Re-expression of wild type Fcp1 rescues Pol II levels on heat shock *Hsp70*. (A-C) Western blots of whole cell extracts +/- Fcp1-RNAi and +/- Cu induction of the transgenic Fcp1 from **(A)** control (empty vector), **(B)** FLAG-tagged wild type Fcp1 transgenes cells, and **(C)** FLAG-tagged catalytically dead mutant Fcp1 transgenes cells probed with antibodies for FLAG (Stratagene, 1:5000), Fcp1 (lab stock, 1:1000), and TFIIS (lab stock loading control, 1:3000) on. **(D-F)** ChIP results for the Pol II subunit Rpb3 enrichment on the *Hsp70* gene at 10'HS for **(D)** control (empty vector), **(E)** FLAG-tagged wild type Fcp1 transgenes (Fcp1wt) cells, and **(F)** FLAG-tagged catalytically dead mutant Fcp1 transgenes (Fcp1mut) cells. The legend indicates the midpoint of each PCR fragment. The y axes shows the percentage of input DNA immunoprecipitated (error bars indicate SEM of three biological replicates). **(G-I)** Western blots of Serine 2 phosphorylated CTD (EMD Millipore 3E10, 1:250), Serine 5 phosphorylated CTD (EMD Millipore 3E8, 1:250), Rpb3 (lab stock loading control, 1:1000), and TFIIS (lab stock loading control, 1:3000) on the free fraction +/- Fcp1-RNAi and +/- Cu induction of the transgene from **(G)** control (empty vector), **(H)** FLAG-tagged wild type Fcp1transgenes cells, and **(I)** FLAG-tagged catalytically dead mutant Fcp1transgenes cells. The relative amount loaded is indicated (1=1x10^6 cells).

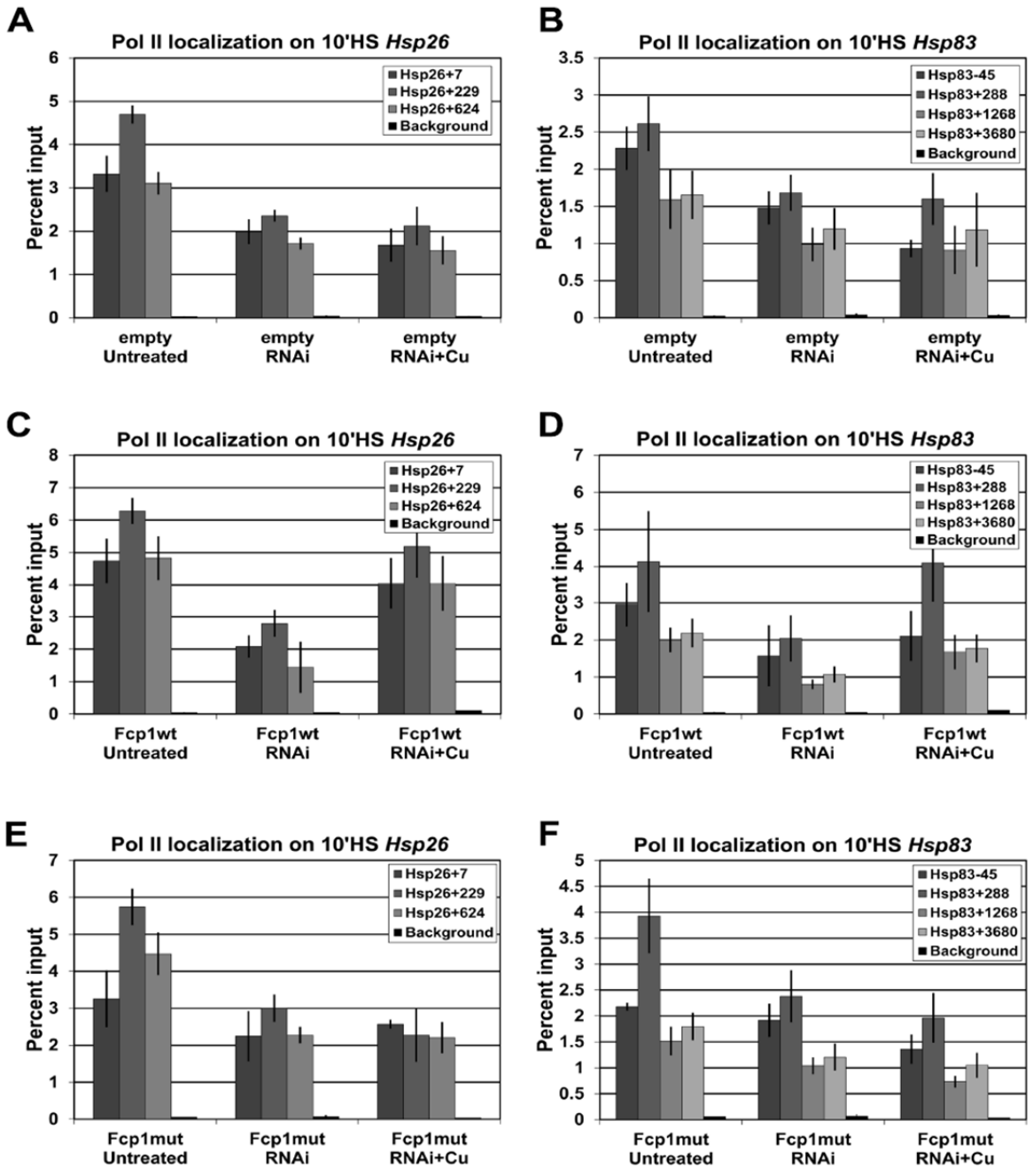


Figure A.11 Re-expression of wild type, but not catalytically dead, Fcp1 partially restores Pol II levels on heat shock induced *Hsp26* and *Hsp83*. ChIP results for the Pol II subunit Rpb3 enrichment on **(A,C,E)** *Hsp26* and **(B,D,F)** *Hsp83* at 10 minutes of heat shock +/- Fcp1-RNAi and +/- Cu induction of the transgene from **(A,B)** control (empty vector), **(C,D)** FLAG-tagged wild type Fcp1 transgenic (Fcp1wt) cells, and **(E,F)** FLAG-tagged catalytically dead mutant Fcp1 transgenic (Fcp1mut) cells. The legends indicate the midpoint of each PCR fragment. The error bars indicate the SEM of at least three biological replicates.

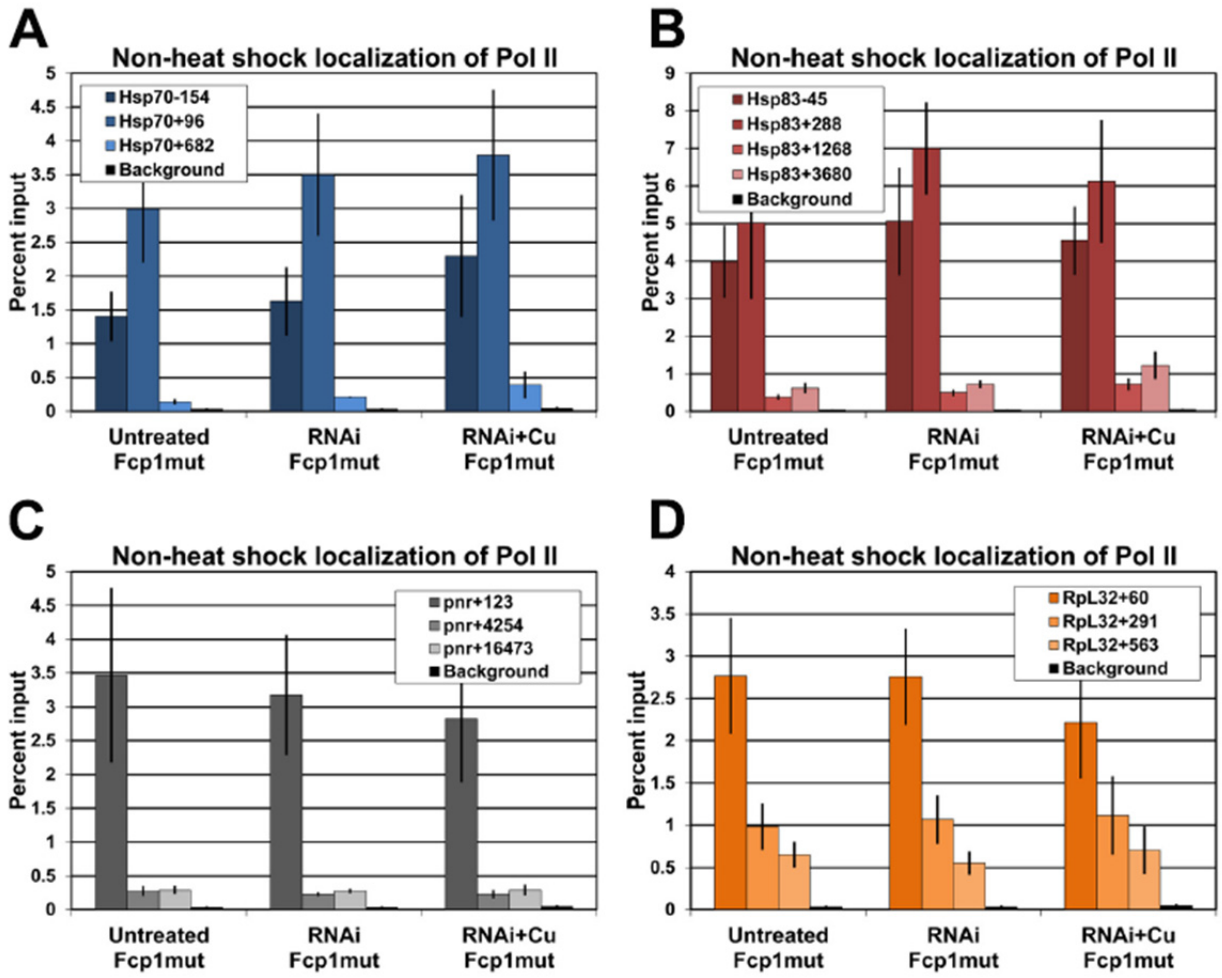


Figure A.12 Re-expression of the catalytically dead Fcp1 mutant does change levels of Pol II on constitutively expressed genes. ChIP results for the Pol II subunit Rpb3 enrichment on **(A)** *Hsp70*, **(B)** *Hsp83*, **(C)** *pnr,* and **(D)** *RpL32* genes under non-heat shock conditions +/- Fcp1-RNAi and +/- Cu induction of the FLAG-tagged catalytically dead mutant Fcp1 transgene The legends indicate the midpoint of each PCR fragment. The error bars indicate the SEM of at least three biological replicates.

A.2.6 Co-depletion of Fcp1 and P-TEFb restores Pol II levels at the 5' end of *Hsp70*

Levels of Pol II on a gene are controlled at multiple steps during the transcription cycle. For example, the level of Pol II on the 5' ends of genes depend upon both the rate of initiation and the rate of pause escape (Core & Lis, 2008). This is exemplified at the *Hsp70* gene, where under uninduced conditions, Pol II initiation rate is higher than the pause escape rate, and thus the 5' end is highly occupied by a transcriptionally-engaged Pol II. In contrast, during an optimal heat shock; Pol II is efficiently released into productive elongation and the *Hsp70* genes are fully occupied with a transcribing Pol II complex every 80bp (Lis, 1998). Thus, if our hypothesis is that Fcp1 knock-down diminishes levels of Pol II on *Hsp70* by reducing initiation, we predict that the level of Pol II on the 5' end of *Hsp70* during Fcp1 knock-down will increase back to its fully-occupied, induced levels by reducing the pause escape rate.

Since P-TEFb activity is required for pause escape (Ni et al., 2008), we reasoned that the pause escape rate could be reduced by depleting the P-TEFb subunit, Cyclin T1. We therefore used RNAi to deplete the Cyclin T1 alone or in combination with Fcp1, and performed Pol II ChIP at 10 minutes of heat shock. Cyclin T1 is reduced by about 90% when knocked-down alone or in combination with Fcp1, and Fcp1 depletion is similar when knocked-down alone or in combination with Cyclin T1 (Figure A.13A). As

we expected, ChIP for Pol II showed the rate of pause escape was reduced in cells depleted for Cyclin T1. Levels of Pol II in the pause region at the 5' end of the gene were unaffected, remaining fully occupied with paused Pol II, but levels of Pol II in the downstream gene body region were reduced indicative of the lower rate of pause escape. Consistent with previous experiments, Fcp1 knock-down reduced Pol II levels in the pause region (Hsp70+96) to approximately half of control levels. As we hypothesized, depletion of Cyclin T1 in conjunction with Fcp1 increased the Pol II level in the paused region as compared to Fcp1 depletion alone, restoring full Pol II occupancy on the 5' end of the gene in cells depleted for both Fcp1 and Cyclin T1, similar to control or Cyclin T1 knock-down alone (Figure A.13B). Similar results were seen on both *Hsp26* and *Hsp83* (Figure A.13D,E).

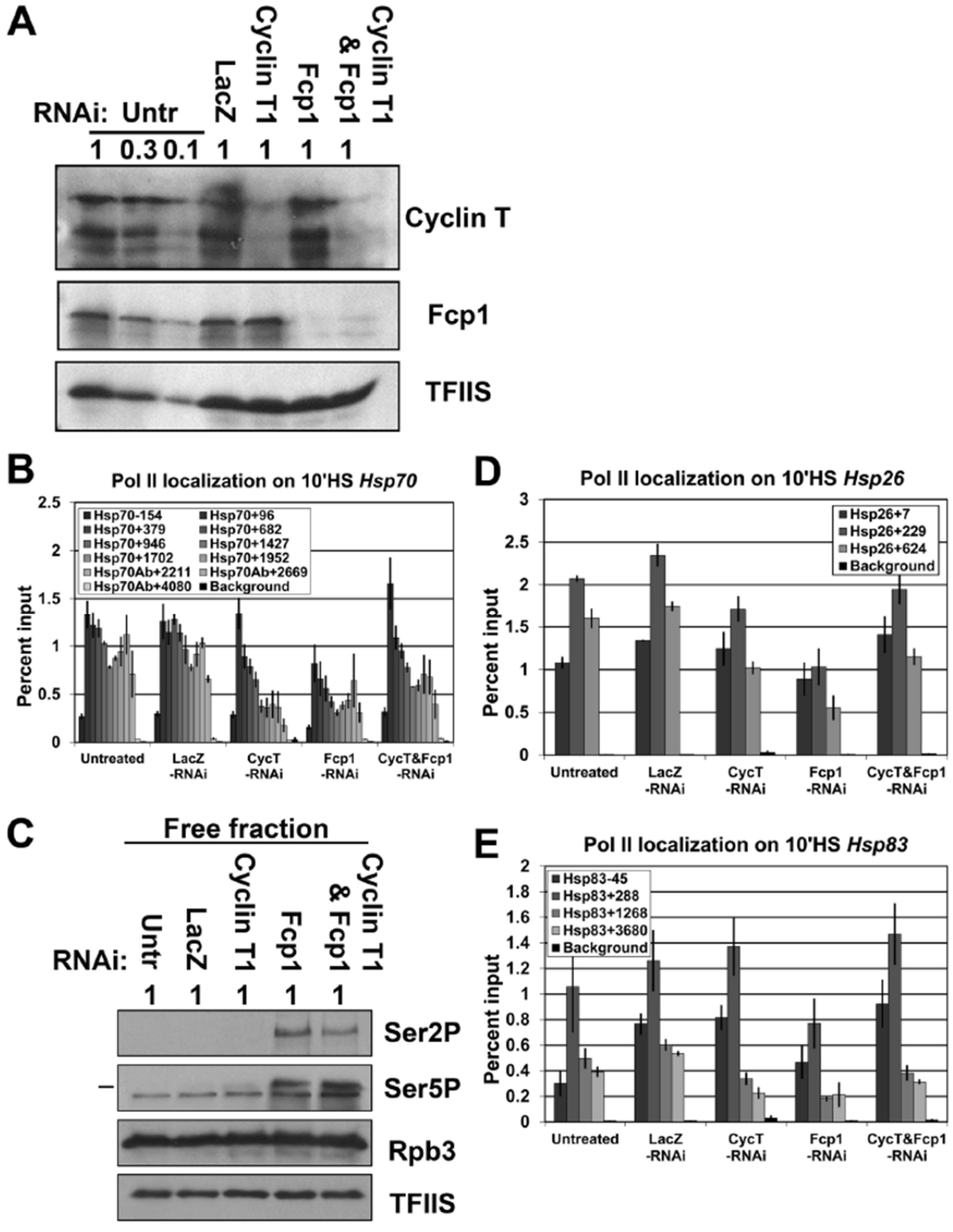

A
RNAi: Untr
1 0.3 0.1
LacZ
Cyclin T1
Fcp1
Cyclin T1 & Fcp1
1 1 1 1
Cyclin T
Fcp1
TFIIS
B
Pol II localization on 10'HS Hsp70
Percent input
Hsp70-154
Hsp70+96
Hsp70+379
Hsp70+682
Hsp70+946
Hsp70+1427
Hsp70+1702
Hsp70+1952
Hsp70Ab+2211
Hsp70Ab+2669
Hsp70Ab+4080
Background
Untreated
LacZ -RNAi
CycT -RNAi
Fcp1 -RNAi
CycT&Fcp1 -RNAi
D
Pol II localization on 10'HS Hsp26
Percent input
Hsp26+7
Hsp26+229
Hsp26+624
Background
Untreated
LacZ -RNAi
CycT -RNAi
Fcp1 -RNAi
CycT&Fcp1 -RNAi
C
Free fraction
RNAi:
Untr
LacZ
Cyclin T1
Fcp1
Cyclin T1 & Fcp1
1 1 1 1 1
Ser2P
Ser5P
Rpb3
TFIIS
E
Pol II localization on 10'HS Hsp83
Percent input
Hsp83-45
Hsp83+288
Hsp83+1268
Hsp83+3680
Background
Untreated
LacZ -RNAi
CycT -RNAi
Fcp1 -RNAi
CycT&Fcp1 -RNAi

Figure A.13 Cyclin-T1, Fcp1 double knock-down increases Pol II levels at the 5' end of *Hsp70* during heat shock. (A) Western blots of whole cell extracts from Untreated, LacZ-RNAi, Cyclin-T1-RNAi, Fcp1-RNAi, and Cyclin-T1+Fcp1-RNAi cells probed with antibodies for Cyclin T (lab stock, 1:1000), Fcp1 (lab stock, 1:1000), and TFIIS (lab stock loading control, 1:3000). The relative amount loaded is indicated (1=1.5x10^6cells). **(B)** ChIP results for the Pol II subunit Rpb3 enrichment on the *Hsp70* gene in Untreated, LacZ-RNAi, Cyclin-T1-RNAi, Fcp1-RNAi, and both Cyclin-T1+Fcp1-RNAi cells at 10 minutes of heat shock. The x axes show the midpoint of each PCR fragment along *Hsp70* gene and the y axis shows the percentage of input DNA immunoprecipitated (error bars indicate the SEM of at least four biological replicates). **(C)** Phosphorylated CTD Serine5 (EMD Millipore 3E8, 1:250), Serine2 (EMD Millipore 3E10, 1:250), and TFIIS (lab stock loading control, 1:3000) Western blots of non-chromatin (free) fractions from Untreated, LacZ-RNAi, Cyclin-T1-RNAi, Fcp1-RNAi, and both CyclinT1+Fcp1-RNAi cells. The relative amount loaded is indicated (1=1x10^6 cells). **(D,E)** ChIP results for the Pol II subunit Rpb3 enrichment on **(D)** *Hsp26* and **(E)** *Hsp83* genes at 10 minutes of heat shock in Untreated, LacZ-RNAi, Cyclin T1-RNAi, Fcp1-RNAi, and CyclinT1+Fcp1-RNAi S2 cells. The legends indicate the midpoint of each PCR fragment. The error bars indicate the SEM of at least three biological replicates.

P-TEFb phosphorylates the CTD on Serine 2, the presumed target of Fcp1 in *Drosophila*; therefore, we investigated if the co-depletion's rescue of the 5' levels of Pol II on induced HS genes could be caused by the co-depletion reducing the high level of phosphorylated free Pol II seen in the Fcp1 knock-down. Westerns show that the levels of phosphorylated free Pol II (both Serine 5 and Serine 2) remained high in the co-depleted cells, similar to Fcp1 knock-down alone (Figure A.13C). Taken together, our results support the model that the most highly expressed genes depend on Fcp1 phosphatase activity to provide sufficient levels of unphosphorylated Pol II to support correspondingly high initiation rates.

A.3 Discussion

Our previous studies showed that Fcp1 depletion in *Drosophila* S2 cells results in reduced *Hsp70* mRNA accumulation after heat shock (Ardehali et al., 2009). In this study, we set out to further investigate the role of Fcp1 in transcription in vivo. Consistent with a direct role in transcription, we have demonstrated that *Drosophila* Fcp1 localizes to actively elongating Pol II complexes. In particular, Fcp1 co-localizes with Pol II at many loci on polytene chromosomes under NHS conditions. Although the ratio between Pol II and Fcp1 signals varied at different loci, the variation in relative signal may represent differences in the transcription level at each locus.

Immunostaining at 2 minutes of heat shock, when Pol II is being recruited to the 87A and C loci, showed strong Fcp1 signal, but the Fcp1 signal was more diffuse at 10 minutes of heat shock, when the loci are saturated with Pol II and maximally decondensed (Zobeck et al., 2010). At higher resolution, our ChIP experiments showed that Fcp1 localization on heat shock *Hsp70* was evenly distributed across the gene in the same pattern as Pol II. These findings are consistent with previous in vitro and ChIP experiments in yeast showing Fcp1 co-localizes with elongating Pol II (Calvo & Manley, 2005; Cho et al., 2001; Kong et al., 2005).

Fcp1 temperature-sensitive mutants in yeast have increased Serine 2 phosphorylation on genes at the restrictive temperature (Cho et al., 2001). Therefore, it was surprising to find that Pol II-normalized CTD phosphorylation levels on heat shock *Hsp70* did not change in Fcp1 depleted cells. There are several possible explanations. First, RNAi treated cells may still contain enough Fcp1 to transiently associate with the elongation complex and prevent abnormal phosphorylation levels. Second, CTD phosphorylation may be maximal on heat shock *Hsp70*, and therefore cannot increase further in Fcp1-RNAi cells. Finally, Fcp1 may not catalyze CTD dephosphorylation of the elongating complex. Although in vivo experiments in *S. cerevisiae* found evidence for Fcp1 catalytic activity during transcription and post-termination, an in vitro study indicated that free Pol II is the preferred substrate of Fcp1 (Kong et al., 2005). Our ChIP

results are consistent with dephosphorylation occurring after elongation. In addition, Fcp1 depletion does not change the amount of phosphorylation or total Pol II in the chromatin fraction, but dramatically increased the amount of phosphorylated Pol II in the free fraction.

Strikingly, Fcp1 depletion resulted in a reduction of Pol II levels across all regions of the induced *Hsp70, Hsp26 and Hsp83 genes,* similar in magnitude to the decrease in the corresponding mRNAs (Ardehali et al., 2009). However, we failed to see significant changes in Pol II levels on any genes in Fcp1 depleted cells under non-heat shock conditions by ChIP or GRO-seq. Although we cannot eliminate the possibility that Fcp1 depletion affects heat shock signaling, we believe it is unlikely because recruitment of the activator HSF to *Hsp70* is unaffected (data not shown). The detection of a two-fold reduction in *Hsp70* transcription with no detectable changes in transcription of constitutively expressed genes may be explained by the extremely high levels of transcription on induced heat shock genes compared to NHS genes. It has been estimated that an optimally induced *Hsp70* gene has Pol II complexes every 80bp (Lis, 1998). This translates into a high turnover of Pol II with a new Pol II initiating about every 4 seconds, corresponding to greater than a 100 fold increase in transcription (Gilmour & Lis, 1985). Thus, a decrease in unphosphorylated free Pol II, the form which is required for initiation (Laybourn & Dahmus, 1990), may slow initiation on induced

Hsp70 sufficiently to cause an increase in the spacing between elongating Pol II complexes on heat shock induced *Hsp70*. This increased spacing would cause a corresponding decrease in Pol II ChIP along the *Hsp70* transcription unit (Figure A.14, B versus A). Based on GRO-seq gene body reads, both *Hsp83* and *Thor* were among the highest expressed in uninduced cells, but notably, Hsp83 is known to be transcribed at a 11-fold higher level in induced cells based on pulse-labeling measurements in vivo (O'Connor & Lis, 1981). Thus, no constitutively expressed gene in S2 cells has a density of Pol II approaching that of induced *Hsp70, Hsp26, or Hsp83*. The fact that only super-highly expressed HS genes are affected indicate that the concentration of unphosphorylated Pol II, which is required for initiation, is not limiting for the vast majority of gene expression (Figure A.14, B versus A). In agreement with this model, slowing the rate of pause escape by co-depletion of the P-TEFb subunit Cyclin T1 with Fcp1 depletion restored Pol II levels on the 5' end of induced *Hsp70* to control levels by making pause escape sufficiently slow that the reduced initiation rate could still fill the pause site to its normal level.

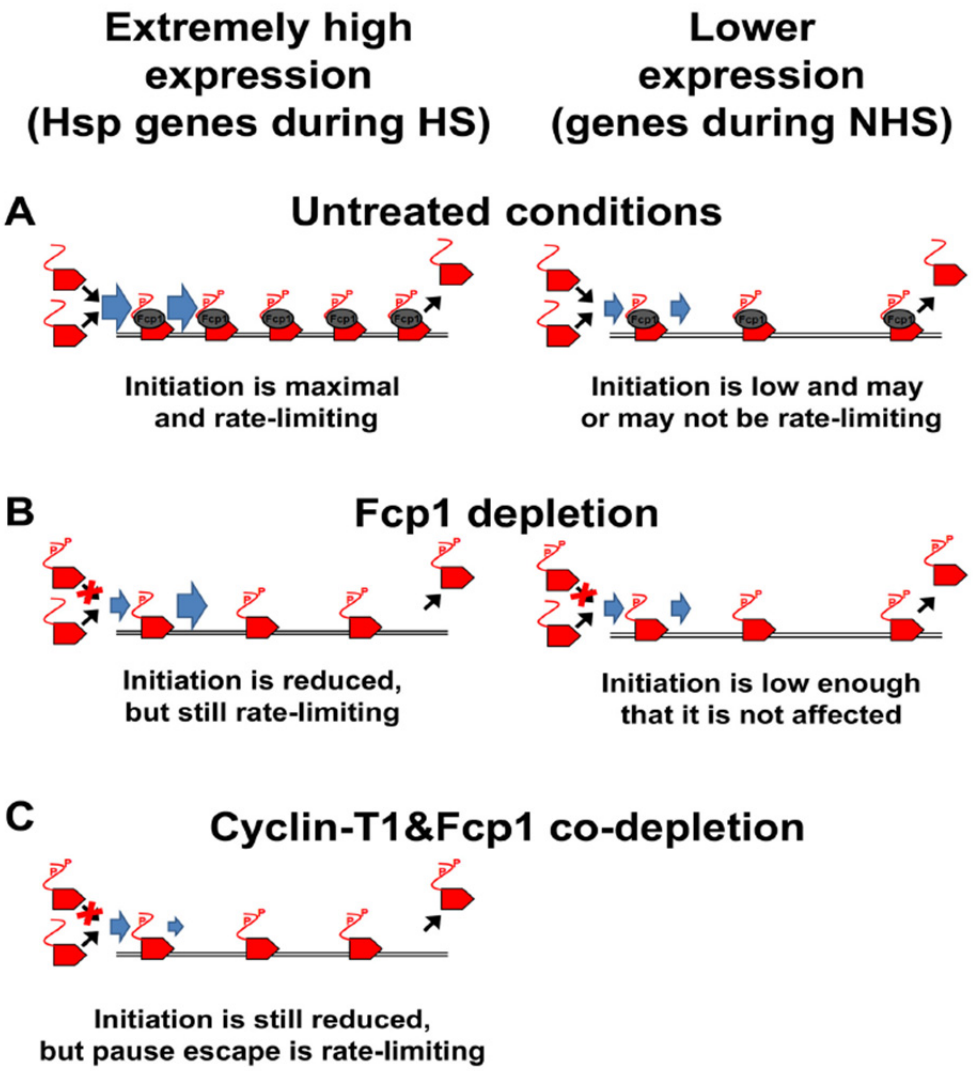


Figure A.14 Model for results. (A) Highly expressed heat shock genes rapidly recruit and release Pol II, and other genes have much lower rates of Pol II recruitment and release (Pol II is represented by the red rocket with CTD tail, and the rates of initiation and pause release are indicated by the size of the blue arrows). **(B)** Fcp1 knock-down increases the amount of phosphorylated free Pol II (represented by Ps on the CTD tail). Because the phosphorylated Pol II cannot initiate transcription, this reduces Pol II initiation on heat shock genes, but the lower initiation rates on other genes allows enough time to bind an initiation-competent Pol II. **(C)** Cyclin-T1 knock-down lowers the release rate, and the pause region is fully occupied by Pol II in co-depleted cells because the reduced release rate allows the heat shock genes enough time to bind an initiation-competent Pol II.

Overall, our study demonstrates that Fcp1 depletion causes reduced HS gene expression and a corresponding reduction of Pol II on induced *Hsp70,* and it also causes a dramatic increase in phosphorylation of both Serine 2 and Serine 5 on free Pol II. Although these results suggest Fcp1 dephosphorylates both Serine 2 and Serine 5, we can't rule out that its activity is coupled to a second phosphatase. Further studies are required to determine if both residues are direct targets of *Drosophila* Fcp1 or if dephosphorylation of these different residues is indeed coupled. Taken together, our results are consistent with Fcp1 depletion impairing the ability to recycle Pol II, reducing the pool of initiation-competent polymerase, and leading to reduced levels of transcribing Pol II on the highly transcribed heat shock genes.

REFERENCES

Adelman K, Marr MT, Werner J, Saunders A, Ni Z, Andrulis ED, Lis JT. 2005. Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol Cell* **17**: 103-112.

Adelman K, Rogatsky I. 2010. RNA polymerase II stalling mediates cytokine gene expression. *Cell Cycle* **9**: 630-631.

Adkins MW, Howar SR, Tyler JK. 2004. Chromatin disassembly mediated by the histone chaperone Asf1 is essential for transcriptional activation of the yeast PHO5 and PHO8 genes. *Mol Cell* **14**: 657-666.

Adkins NL, Hagerman TA, Georgel P. 2006. GAGA protein: a multi-faceted transcription factor. *Biochem Cell Biol* **84**: 559-567.

Agianian B, Leonard K, Bonte E, Van der Zandt H, Becker PB, Tucker PA. 1999. The glutamine-rich domain of the Drosophila GAGA factor is necessary for amyloid fibre formation in vitro, but not for chromatin remodelling. *J Mol Biol* **285**: 527-544.

Ahn SH, Kim M, Buratowski S. 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* **13**: 67-76.

Aida M, Chen Y, Nakajima K, Yamaguchi Y, Wada T, Handa H. 2006. Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the junB gene. *Mol Cell Biol* **26**: 6094-6104.

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

Andrulis ED, Guzman E, Doring P, Werner J, Lis JT. 2000. High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes Dev* **14**: 2635-2649.

Ardehali MB, Yao J, Adelman K, Fuda NJ, Petesch SJ, Webb WW, Lis JT. 2009. Spt6 enhances the elongation rate of RNA polymerase II in vivo. *EMBO J* **28**: 1067-1077.

Artsimovitch I, Landick R. 1998. Interaction of a nascent RNA structure with RNA polymerase is required for hairpin-dependent transcriptional pausing but not for transcript release. *Genes Dev* **12**: 3110-3122.

Aygun O, Svejstrup J, Liu Y. 2008. A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. *Proc Natl Acad Sci U S A* **105**: 8580-8584.

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

Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Studitsky VM, Reinberg D. 2003. FACT facilitates transcription-dependent nucleosome alteration. *Science* **301**: 1090-1093.

Belozerov VE, Majumder P, Shen P, Cai HN. 2003. A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of Drosophila. *EMBO J* **22**: 3113-3121.

Benjamin LR, Gilmour DS. 1998. Nucleosomes are not necessary for promoter-proximal pausing in vitro on the Drosophila hsp70 promoter. *Nucleic Acids Res* **26**: 1051-1055.

Bird G, Zorio DA, Bentley DL. 2004. RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol* **24**: 8963-8969.

Boehm AK, Saunders A, Werner J, Lis JT. 2003. Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. *Mol Cell Biol* **23**: 7628-7637.

Boettiger AN, Ralph PL, Evans SN. 2011. Transcriptional regulation: effects of promoter proximal pausing on speed, synchrony and reliability. *PLoS Comput Biol* **7**: e1001136.

Bowman GD. 2010. Mechanisms of ATP-dependent nucleosome sliding. *Curr Opin Struct Biol* **20**: 73-81.

Buratowski S. 2005. Connections between mRNA 3' end processing and transcription termination. *Curr Opin Cell Biol.* **17**(3):257-61.

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

Burke TW, Kadonaga JT. 1997. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes Dev* **11**: 3020-3031.

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

Chalkley GE, Verrijzer CP. 1999. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J* **18**: 4835-4845.

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

Chao SH, Price DH. 2001. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. *J Biol Chem* **276**: 31793-31799.

Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E, Eick D. 2007. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**: 1780-1782.

Cheng B, Price DH. 2007. Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *J Biol Chem* **282**: 21901-21912.

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

Cho EJ, Takagi T, Moore CR, Buratowski S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* **11**: 3319-3326.

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

Chodosh LA, Fire A, Samuels M, Sharp PA. 1989. 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. *J Biol Chem* **264**: 2250-2257.

Chopra VS, Hendrix DA, Core LJ, Tsui C, Lis JT, Levine M. 2011. The polycomb group mutant esc leads to augmented levels of paused Pol II in the Drosophila embryo. *Mol Cell* **42**: 837-844.

Chopra VS, Srinivasan A, Kumar RP, Mishra K, Basquin D, Docquier M, Seum C, Pauli D, Mishra RK. 2008. Transcriptional activation by GAGA factor is through its direct interaction with dmTAF3. *Dev Biol* **317**: 660-670.

Core LJ, Lis JT. 2008. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* **319**: 1791-1792.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845-1848.

Croston GE, Kerrigan LA, Lira LM, Marshak DR, Kadonaga JT. 1991. Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. *Science* **251**: 643-649.

Cuvier O, Hart CM, Laemmli UK. 1998. Identification of a class of chromatin boundary elements. *Mol Cell Biol* **18**: 7478-7486.

Gilmour DS. 2009. Promoter proximal pausing on genes in metazoans. *Chromosoma* **118**(1):1-10.

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

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

Espinas ML, Canudas S, Fanti L, Pimpinelli S, Casanova J, Azorin F. 2000. The GAGA factor of Drosophila interacts with SAP18, a Sin3-associated polypeptide. *EMBO Rep* **1**: 253-259.

Espinas ML, Jimenez-Garcia E, Vaquero A, Canudas S, Bernues J, Azorin F. 1999. The N-terminal POZ domain of GAGA mediates the formation of oligomers that bind DNA with high affinity and specificity. *J Biol Chem* **274**: 16461-16469.

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

Farkas G, Gausz J, Galloni M, Reuter G, Gyurkovics H, Karch F. 1994. The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature* **371**: 806-808.

Fish RN, Kane CM. 2002. Promoting elongation with transcript cleavage stimulatory factors. *Biochim Biophys Acta* **1577**: 287-307.

Flores O, Lu H, Killeen M, Greenblatt J, Burton ZF, Reinberg D. 1991. The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proc Natl Acad Sci U S A* **88**: 9999-10003.

Fuda NJ, Ardehali MB, Lis JT. 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**: 186-192.

Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. 2004. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol* **24**: 787-795.

Ghosh A, Shuman S, Lima CD. 2011. Structural insights to how mammalian capping enzyme reads the CTD code. *Mol Cell* **43**: 299-310.

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

Ghosh SK, Missra A, Gilmour DS. 2011. Negative elongation factor accelerates the rate at which heat shock genes are shut off by facilitating dissociation of heat shock factor. *Mol Cell Biol* **31**: 4232-4243.

Gilchrist DA, Dos Santos G, Fargo DC, Xie B, Gao Y, Li L, Adelman K. 2010. Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**: 540-551.

Gilchrist DA, Nechaev S, Lee C, Ghosh SK, Collins JB, Li L, Gilmour DS, Adelman K. 2008. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev* **22**: 1921-1933.

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

Gong WJ, Golic KG. 2006. Loss of Hsp70 in Drosophila is pleiotropic, with effects on thermotolerance, recovery from heat shock and neurodegeneration. *Genetics* **172**: 275-286.

Gong WJ, Golic KG. 2004. Genomic deletions of the Drosophila melanogaster Hsp70 genes. *Genetics* **168**: 1467-1476.

Granok H, Leibovitch BA, Elgin SC. 2001. A heat-shock-activated cDNA encoding GAGA factor rescues some lethal mutations in the Drosophila melanogaster Trithorax-like gene. *Genet Res* **78**: 13-21.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**: 77-88.

Guertin MJ, Lis JT. 2010. Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet* **6**: e1001114.

Guertin MJ, Petesch SJ, Zobeck KL, Min IM, Lis JT. 2010. Drosophila heat shock system as a general model to investigate transcriptional regulation. *Cold Spring Harb Symp Quant Biol* **75**: 1-9.

Guo S, Yamaguchi Y, Schilbach S, Wada T, Lee J, Goddard A, French D, Handa H, Rosenthal A. 2000. A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**: 366-369.

Hackett RW, Lis JT. 1981. DNA sequence analysis reveals extensive homologies of regions preceding hsp70 and alphabeta heat shock genes in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **78**: 6196-6200.

Hagstrom K, Muller M, Schedl P. 1997. A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the Drosophila bithorax complex. *Genetics* **146**: 1365-1380.

Hausmann S, Koiwa H, Krishnamurthy S, Hampsey M, Shuman S. 2005. Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases. *J Biol Chem* **280**: 37681-37688.

Hausmann S, Shuman S. 2002. Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. *J Biol Chem* **277**: 21213-21220.

Hendrix DA, Hong JW, Zeitlinger J, Rokhsar DS, Levine MS. 2008. Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. *Proc Natl Acad Sci U S A* **105**: 7762-7767.

Herbert KM, La Porta A, Wong BJ, Mooney RA, Neuman KC, Landick R, Block SM. 2006. Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell* **125**: 1083-1094.

Hirose Y, Ohkuma Y. 2007. Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *J Biochem* **141**: 601-608.

Ho JW, Bishop E, Karchenko PV, Negre N, White KP, Park PJ. 2011. ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. *BMC Genomics* **12**: 134.

Horard B, Tatout C, Poux S, Pirrotta V. 2000. Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. *Mol Cell Biol* **20**: 3187-3197.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218-223.

Ivanov D, Kwak YT, Guo J, Gaynor RB. 2000. Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol Cell Biol* **20**: 2970-2983.

Izban MG, Luse DS. 1991. Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes Dev* **5**: 683-696.

Jacobson RH, Ladurner AG, King DS, Tjian R. 2000. Structure and function of a human TAFII250 double bromodomain module. *Science* **288**: 1422-1425.

Jiang C, Pugh BF. 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* **10**: 161-172.

Juven-Gershon T, Kadonaga JT. 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol* **339**: 225-229.

Kamenski T, Heilmeier S, Meinhart A, Cramer P. 2004. Structure and mechanism of RNA polymerase II CTD phosphatases. *Mol Cell* **15**: 399-407.

Katsani KR, Hajibagheri MA, Verrijzer CP. 1999. Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. *EMBO J* **18**: 698-708.

Keogh MC, Podolny V, Buratowski S. 2003. Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Mol Cell Biol* **23**: 7005-7018.

Kephart DD, Marshall NF, Price DH. 1992. Stability of Drosophila RNA polymerase II elongation complexes in vitro. *Mol Cell Biol* **12**: 2067-2077.

Kerrigan LA, Croston GE, Lira LM, Kadonaga JT. 1991. Sequence-specific transcriptional antirepression of the Drosophila Kruppel gene by the GAGA factor. *J Biol Chem* **266**: 574-582.

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**: 480-485.

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

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

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436**: 876-880.

Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**: 599-608.

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

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

Komissarova N, Kashlev M. 1997. RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. *J Biol Chem* **272**: 15329-15338.

Kong SE, Kobor MS, Krogan NJ, Somesh BP, Sogaard TM, Greenblatt JF, Svejstrup JQ. 2005. Interaction of Fcp1 phosphatase with elongating RNA polymerase II holoenzyme, enzymatic mechanism of action, and genetic interaction with elongator. *J Biol Chem* **280**: 4299-4306.

Korber P, Barbaric S, Luckenbach T, Schmid A, Schermer UJ, Blaschke D, Horz W. 2006. The histone chaperone Asf1 increases the rate of histone eviction at the yeast PHO5 and PHO8 promoters. *J Biol Chem* **281**: 5539-5545.

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

Krumm A, Meulia T, Brunvand M, Groudine M. 1992. The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. *Genes Dev* **6**: 2201-2213.

Kuehner JN, Pearson EL, Moore C. 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* **12**: 283-294.

Landick R. 2006. The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* **34**: 1062-1066.

Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, Kuroda MI. 2011. X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. *Nature* **471**: 115-118.

Laybourn PJ, Dahmus ME. 1990. Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. *J Biol Chem* **265**: 13165-13173.

Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, Jiang C, Li J, Pugh BF, Gilmour DS. 2008. NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. *Mol Cell Biol* **28**: 3290-3300.

Lee DH, Gershenzon N, Gupta M, Ioshikhes IP, Reinberg D, Lewis BA. 2005. Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Mol Cell Biol* **25**: 9674-9686.

Lee H, Kraus KW, Wolfner MF, Lis JT. 1992. DNA sequence requirements for generating paused polymerase at the start of hsp70. *Genes Dev* **6**: 284-295.

Li M, Phatnani HP, Guan Z, Sage H, Greenleaf AL, Zhou P. 2005. Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proc Natl Acad Sci U S A* **102**: 17636-17641.

Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL. 2002. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell* **9**: 1101-1111.

Lin PS, Dubois MF, Dahmus ME. 2002. TFIIF-associating carboxyl-terminal domain phosphatase dephosphorylates phosphoserines 2 and 5 of RNA polymerase II. *J Biol Chem* **277**: 45949-45956.

Lin YC, Choi WS, Gralla JD. 2005. TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. *Nat Struct Mol Biol* **12**: 603-607.

Lindell TJ, Weinberg F, Morris PW, Roeder RG, Rutter WJ. 1970. Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science* **170**: 447-449.

Lis J. 1998. Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. *Cold Spring Harb Symp Quant Biol* **63**: 347-356.

Lis JT, Mason P, Peng J, Price DH, Werner J. 2000. P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev* **14**: 792-803.

Liu J, Kipreos ET. 2000. Evolution of cyclin-dependent kinases (CDKs) and CDK-activating kinases (CAKs): differential conservation of CAKs in yeast and metazoa. *Mol Biol Evol* **17**: 1061-1074.

Liu J, Zhang J, Gong Q, Xiong P, Huang H, Wu B, Lu G, Wu J, Shi Y. 2011. Solution structure of tandem SH2 domains from Spt6 protein and their binding to the phosphorylated RNA polymerase II C-terminal domain. *J Biol Chem* **286**: 29218-29226.

Love JD, Minton KW. 1985. Screening of lambda library for differentially expressed genes using in vitro transcripts. *Anal Biochem* **150**: 429-441.

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

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

Mandal SS, Cho H, Kim S, Cabane K, Reinberg D. 2002. FCP1, a phosphatase specific for the heptapeptide repeat of the largest subunit of RNA polymerase II, stimulates transcription elongation. *Mol Cell Biol* **22**: 7543-7552.

Mandal SS, Chu C, Wada T, Handa H, Shatkin AJ, Reinberg D. 2004. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc Natl Acad Sci U S A* **101**: 7572-7577.

Marshall NF, Price DH. 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem* **270**: 12335-12338.

Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P. 2011. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J* **30**: 1302-1310.

Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC et al. 2008. Nucleosome organization in the Drosophila genome. *Nature* **453**: 358-362.

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

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

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

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

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

Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, Vermeulen M, Buscaino A, Duncan K, Mueller J et al. 2006. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. *Mol Cell* **21**: 811-823.

Metcalf CE, Wassarman DA. 2006. DNA binding properties of TAF1 isoforms with two AT-hooks. *J Biol Chem* **281**: 30015-30023.

Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT. 2011. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev* **25**: 742-754.

Mirkovitch J, Darnell JE,Jr. 1992. Mapping of RNA polymerase on mammalian genes in cells and nuclei. *Mol Biol Cell* **3**: 1085-1094.

Missra A, Gilmour DS. 2010. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proc Natl Acad Sci U S A* **107**: 11301-11306.

Morris DP, Greenleaf AL. 2000. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* **275**: 39935-39943.

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

Murawska M, Kunert N, van Vugt J, Langst G, Kremmer E, Logie C, Brehm A. 2008. dCHD3, a novel ATP-dependent chromatin remodeler associated with sites of active transcription. *Mol Cell Biol* **28**: 2745-2757.

Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K. 2007. RNA polymerase is poised for activation across the genome. *Nat Genet* **39**: 1507-1511.

Nechaev S, Adelman K. 2011. Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim Biophys Acta* **1809**: 34-45.

Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K. 2010. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science* **327**: 335-338.

Negre N, Brown CD, Shah PK, Kheradpour P, Morrison CA, Henikoff JG, Feng X, Ahmad K, Russell S, White RA et al. 2010. A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet* **6**: e1000814.

Negre N, Hennetin J, Sun LV, Lavrov S, Bellis M, White KP, Cavalli G. 2006. Chromosomal distribution of PcG proteins during Drosophila development. *PLoS Biol* **4**: e170.

Ni Z, Saunders A, Fuda NJ, Yao J, Suarez JR, Webb WW, Lis JT. 2008. P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo. *Mol Cell Biol* **28**: 1161-1170.

Ni Z, Schwartz BE, Werner J, Suarez JR, Lis JT. 2004. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol Cell* **13**: 55-65.

O'Donnell KH, Chen CT, Wensink PC. 1994. Insulating DNA directs ubiquitous transcription of the Drosophila melanogaster alpha 1-tubulin gene. *Mol Cell Biol* **14**: 6398-6408.

Ohtsuki S, Levine M. 1998. GAGA mediates the enhancer blocking activity of the eve promoter in the Drosophila embryo. *Genes Dev* **12**: 3325-3330.

Okada M, Hirose S. 1998. Chromatin remodeling mediated by Drosophila GAGA factor and ISWI activates fushi tarazu gene transcription in vitro. *Mol Cell Biol* **18**: 2455-2461.

Omichinski JG, Pedone PV, Felsenfeld G, Gronenborn AM, Clore GM. 1997. The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nat Struct Biol* **4**: 122-132.

Orphanides G, LeRoy G, Chang CH, Luse DS, Reinberg D. 1998. FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92**: 105-116.

Pagans S, Ortiz-Lombardia M, Espinas ML, Bernues J, Azorin F. 2002. The Drosophila transcription factor tramtrack (TTK) interacts with Trithorax-like (GAGA) and represses GAGA-mediated activation. *Nucleic Acids Res* **30**: 4406-4413.

Pal M, Ponticelli AS, Luse DS. 2005. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol Cell* **19**: 101-110.

Palangat M, Landick R. 2001. Roles of RNA:DNA hybrid stability, RNA structure, and active site conformation in pausing by human RNA polymerase II. *J Mol Biol* **311**: 265-282.

Palangat M, Renner DB, Price DH, Landick R. 2005. A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. *Proc Natl Acad Sci U S A* **102**: 15036-15041.

Park JM, Werner J, Kim JM, Lis JT, Kim YJ. 2001. Mediator, not holoenzyme, is directly recruited to the heat shock promoter by HSF upon heat shock. *Mol Cell* **8**: 9-19.

Payne JM, Laybourn PJ, Dahmus ME. 1989. The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit IIa. *J Biol Chem* **264**: 19621-19629.

Pedone PV, Ghirlando R, Clore GM, Gronenborn AM, Felsenfeld G, Omichinski JG. 1996. The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding. *Proc Natl Acad Sci U S A* **93**: 2822-2826.

Pei Y, Shuman S. 2002. Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J Biol Chem* **277**: 19639-19648.

Peil K, Varv S, Looke M, Kristjuhan K, Kristjuhan A. 2011. Uniform distribution of elongating RNA polymerase II complexes in transcribed gene locus. *J Biol Chem* **286**: 23817-23822.

Peterlin BM, Price DH. 2006. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23**: 297-305.

Peterson CL, Workman JL. 2000. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Curr Opin Genet Dev* **10**: 187-192.

Petesch SJ, Lis JT. 2008. Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. *Cell* **134**: 74-84.

Prelich G. 2002. RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. *Eukaryot Cell* **1**: 153-162.

Ptashne M, Gann A. 1997. Transcriptional activation by recruitment. *Nature* **386**: 569-577.

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

Raja SJ, Charapitsa I, Conrad T, Vaquerizas JM, Gebhardt P, Holz H, Kadlec J, Fraterman S, Luscombe NM, Akhtar A. 2010. The nonspecific lethal complex is a transcriptional regulator in Drosophila. *Mol Cell* **38**: 827-841.

Rao JN, Schweimer K, Wenzel S, Wohrl BM, Rosch P. 2008. NELF-E RRM undergoes major structural changes in flexible protein regions on target RNA binding. *Biochemistry* **47**: 3756-3761.

Rasmussen EB, Lis JT. 1995. Short transcripts of the ternary complex provide insight into RNA polymerase II elongational pausing. *J Mol Biol* **252**: 522-535.

Rasmussen EB, Lis JT. 1993. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc Natl Acad Sci U S A* **90**: 7923-7927.

Read D, Butte MJ, Dernburg AF, Frasch M, Kornberg TB. 2000. Functional studies of the BTB domain in the Drosophila GAGA and Mod(mdg4) proteins. *Nucleic Acids Res* **28**: 3864-3870.

Reinke H, Horz W. 2003. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol Cell* **11**: 1599-1607.

Renner DB, Yamaguchi Y, Wada T, Handa H, Price DH. 2001. A highly purified RNA polymerase II elongation control system. *J Biol Chem* **276**: 42601-42609.

Reyes-Reyes M, Hampsey M. 2007. Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation. *Mol Cell Biol* **27**: 926-936.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Rougvie AE, Lis JT. 1988. The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell* **54**: 795-804.

Saha RN, Wissink EM, Bailey ER, Zhao M, Fargo DC, Hwang JY, Daigle KR, Fenn JD, Adelman K, Dudek SM. 2011. Rapid activity-induced transcription of Arc and other IEGs relies on poised RNA polymerase II. *Nat Neurosci* **14**: 848-856.

Salvaing J, Lopez A, Boivin A, Deutsch JS, Peronnet F. 2003. The Drosophila Corto protein interacts with Polycomb-group proteins and the GAGA factor. *Nucleic Acids Res* **31**: 2873-2882.

Saunders A, Core LJ, Lis JT. 2006. Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* **7**: 557-567.

Schmidt WM, Mueller MW. 1999. CapSelect: a highly sensitive method for 5' CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucleic Acids Res* **27**: e31.

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

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**: 887-898.

Schwartz BE, Larochelle S, Suter B, Lis JT. 2003. Cdk7 is required for full activation of Drosophila heat shock genes and RNA polymerase II phosphorylation in vivo. *Mol Cell Biol* **23**: 6876-6886.

Schwendemann A, Lehmann M. 2002. Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. *Proc Natl Acad Sci U S A* **99**: 12883-12888.

Sehgal PB, Darnell JE,Jr, Tamm I. 1976. The inhibition by DRB (5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole) of hnRNA and mRNA production in HeLa cells. *Cell* **9**: 473-480.

Shibata Y, Carninci P, Watahiki A, Shiraki T, Konno H, Muramatsu M, Hayashizaki Y. 2001. Cloning full-length, cap-trapper-selected cDNAs by using the single-strand linker ligation method. *BioTechniques* **30**: 1250-1254.

Shopland LS, Hirayoshi K, Fernandes M, Lis JT. 1995. HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. *Genes Dev* **9**: 2756-2769.

Sims RJ,3rd, Mandal SS, Reinberg D. 2004. Recent highlights of RNA-polymerase-II-mediated transcription. *Curr Opin Cell Biol* **16**: 263-271.

Smale ST, Kadonaga JT. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**: 449-479.

Steger DJ, Haswell ES, Miller AL, Wente SR, O'Shea EK. 2003. Regulation of chromatin remodeling by inositol polyphosphates. *Science* **299**: 114-116.

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

Svaren J, Horz W. 1997. Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. *Trends Biochem Sci* **22**: 93-97.

Swanson MS, Malone EA, Winston F. 1991. SPT5, an essential gene important for normal transcription in Saccharomyces cerevisiae, encodes an acidic nuclear protein with a carboxy-terminal repeat. *Mol Cell Biol* **11**: 4286.

Tadigotla VR, O Maoileidigh D, Sengupta AM, Epshtein V, Ebright RH, Nudler E, Ruckenstein AE. 2006. Thermodynamic and kinetic modeling of transcriptional pausing. *Proc Natl Acad Sci U S A* **103**: 4439-4444.

Thomas GH, Elgin SC. 1988. Protein/DNA architecture of the DNase I hypersensitive region of the Drosophila hsp26 promoter. *EMBO J* **7**: 2191-2201.

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

Tsukiyama T, Becker PB, Wu C. 1994. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**: 525-532.

Tsukiyama T, Wu C. 1996. Purification of GAGA factor of Drosophila and its role in nucleosome disruption. *Methods Enzymol* **274**: 291-299.

Tsukiyama T, Wu C. 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**: 1011-1020.

van Steensel B, Delrow J, Bussemaker HJ. 2003. Genomewide analysis of Drosophila GAGA factor target genes reveals context-dependent DNA binding. *Proc Natl Acad Sci U S A* **100**: 2580-2585.

Vaquero A, Blanch M, Espinas ML, Bernues J. 2008. Activation properties of GAGA transcription factor. *Biochim Biophys Acta* **1779**: 312-317.

Vaquero A, Espinas ML, Azorin F, Bernues J. 2000. Functional mapping of the GAGA factor assigns its transcriptional activity to the C-terminal glutamine-rich domain. *J Biol Chem* **275**: 19461-19468.

Venken KJ, Bellen HJ. 2005. Emerging technologies for gene manipulation in Drosophila melanogaster. *Nat Rev Genet* **6**: 167-178.

Venters BJ, Pugh BF. 2009. A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome Res* **19**: 360-371.

Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**: 58-69.

Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, Sugimoto S, Yano K, Hartzog GA, Winston F et al. 1998a. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev* **12**: 343-356.

Wada T, Takagi T, Yamaguchi Y, Watanabe D, Handa H. 1998b. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *EMBO J* **17**: 7395-7403.

Wall G, Varga-Weisz PD, Sandaltzopoulos R, Becker PB. 1995. Chromatin remodeling by GAGA factor and heat shock factor at the hypersensitive Drosophila hsp26 promoter in vitro. *EMBO J* **14**: 1727-1736.

Wang YV, Tang H, Gilmour DS. 2005. Identification in vivo of different rate-limiting steps associated with transcriptional activators in the presence and absence of a GAGA element. *Mol Cell Biol* **25**: 3543-3552.

Watanabe Y, Fujimoto H, Watanabe T, Maekawa T, Masutani C, Hanaoka F, Ohkuma Y. 2000. Modulation of TFIIH-associated kinase activity by complex formation and its relationship with CTD phosphorylation of RNA polymerase II. *Genes Cells* **5**: 407-423.

Wen Y, Shatkin AJ. 1999. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev* **13**: 1774-1779.

Wilkins RC, Lis JT. 1999. DNA distortion and multimerization: novel functions of the glutamine-rich domain of GAGA factor. *J Mol Biol* **285**: 515-525.

Wilkins RC, Lis JT. 1998. GAGA factor binding to DNA via a single trinucleotide sequence element. *Nucleic Acids Res* **26**: 2672-2678.

Wood A, Shilatifard A. 2006. Bur1/Bur2 and the Ctk complex in yeast: the split personality of mammalian P-TEFb. *Cell Cycle* **5**: 1066-1068.

Wu C. 1980. The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286**: 854-860.

Wu CH, Lee C, Fan R, Smith MJ, Yamaguchi Y, Handa H, Gilmour DS. 2005. Molecular characterization of Drosophila NELF. *Nucleic Acids Res* **33**: 1269-1279.

Wu CH, Madabusi L, Nishioka H, Emanuel P, Sypes M, Arkhipova I, Gilmour DS. 2001. Analysis of core promoter sequences located downstream from the TATA element in the hsp70 promoter from Drosophila melanogaster. *Mol Cell Biol* **21**: 1593-1602.

Wu CH, Yamaguchi Y, Benjamin LR, Horvat-Gordon M, Washinsky J, Enerly E, Larsson J, Lambertsson A, Handa H, Gilmour D. 2003. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev* **17**: 1402-1414.

Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T, Handa H. 2006. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell* **21**: 227-237.

Yamaguchi Y, Inukai N, Narita T, Wada T, Handa H. 2002. Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. *Mol Cell Biol* **22**: 2918-2927.

Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H. 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**: 41-51.

Yao S, Prelich G. 2002. Activation of the Bur1-Bur2 cyclin-dependent kinase complex by Cak1. *Mol Cell Biol* **22**: 6750-6758.

Yeo M, Lin PS, Dahmus ME, Gill GN. 2003. A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. *J Biol Chem* **278**: 26078-26085.

Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA. 2007. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet* **39**: 1512-1516.

Zhou K, Kuo WH, Fillingham J, Greenblatt JF. 2009. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc Natl Acad Sci U S A* **106**: 6956-6961.

Zobeck KL, Buckley MS, Zipfel WR, Lis JT. 2010. Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. *Mol Cell* **40**: 965-975.