Design and Implementation of a Fully Reconstituted Assay to Investigate Mechanisms of Early Human Pol II Transcription

by

Charli B. Fant

B.A. Transylvania University, 2013

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Biochemistry
2019

This thesis entitled:

Design and Implementation of a Fully Reconstituted Assay to Investigate Mechanisms of Early Human Pol II Transcription

Written by Charli B. Fant has been approved for the Department of Biochemistry by:

_____

Dylan J. Taatjes, Ph.D.

_____

Karolin Luger, Ph.D.

Date_____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above-mentioned discipline.

Fant, Charli B. (Ph.D., Biochemistry)

Design and Implementation of a Fully Reconstituted Assay to Investigate Mechanisms of Early Human Pol II Transcription

Thesis directed by Dylan J. Taatjes, PhD

Abstract

Human transcription by RNA polymerase II (pol II) is tightly regulated at multiple steps. During the early stages of transcription, pol II exists within a large 4 MDa assembly called the Pre-Initiation Complex (PIC), which contains TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and Mediator. After initiation, pol II enzymes break contacts with PIC factors in the step of promoter escape but then pause after transcribing less than 100 bases. Mechanistic details about these early stages—PIC assembly, initiation, promoter escape, promoter-proximal pausing, and pause release—are lacking. In this thesis, a fully reconstituted, completely defined pol II transcription assay (i.e. no extracts) is developed to rigorously evaluate mechanisms of transcription initiation and pausing. Successful reconstitution of pol II promoter-proximal pausing enabled a systematic assessment of the contribution of individual PIC factors. Through use of this assay, the striking discovery is made that the PIC is sufficient to establish pol II pausing, and the general transcription factor TFIID is required. Knockdown of TFIID subunits via Trim-Away in human HCT116 cells validate this result by showing genome-wide pause disruption. These results replicate in Drosophila S2 cells and suggest that TFIID may be a genome-wide regulator of pausing.

The reconstituted pausing assay was also used to examine the regulatory roles of three major transcriptional kinases—CDK7, CDK8, and CDK9—on early transcription. Although the data are preliminary, the results suggest that CDK7 may act as an upstream regulator of CDK8 and CDK9 and that CDK8 (as part of the CDK8 module) may activate promoter escape.

Such findings are difficult to obtain in cells due to off-target effects and other confounding issues. However, these results along with several others discussed in this thesis encourage future investigations into early steps of transcription regulation with this newly established reconstituted system.

Acknowledgements

I have only been able to make it through graduate school because of the love and support of others. Family, friends, faculty, staff, and fellow graduate students have made this endeavor achievable and (mostly) enjoyable.

First, I must acknowledge Thomas Buckles. I could not ask for a better partner in life nor a better person with whom to struggle through graduate school. You're a wonderful scientist and a better person. In so many ways, I look up to you. I look forward to taking on life with you by my side.

Ollie and Huck, my sweet pups. I love you as mush as I love anything or anyone in this world. You've been the constants in my life throughout my PhD, and I would not be sane without your devotion and love. People let me down from time to time, but you never do.

Friends. What's life without good friends? Allison Schultz, Meagan Nakamoto, Melissa Hoyer, Verity Johnson, Taylor Chock, Belinda Sly, Kirsten Belinsky, C.P. Santos, Jon Rubin…. I'm lucky to call you all friends. I only hope that I am as good a friend to you as you are to me.

Soccer family. Graduate school is rife with stress and anxiety. My soccer family has given me a productive outlet with competitive people to deal with that stress and anxiety. At the end of a long week, there's nothing I look forward to more than Sunday at Pleasantview Fields.

Lab mates past and present. We are lucky enough to work in an environment where we love and support each other. Sure, science can be competitive, but I think we embrace the collaborative nature of it and that's because each and everyone of you are team players. Thank you for making our lab a great place to work.

Various faculty and staff at CU Boulder Caruthers Biotech. You are all consummate professionals who care about individuals. You are always better than any job description requires you to be and have become supportive friends to me over the years.

Committee members. Karolin Luger, Rob Batey, Jim Goodrich, David Bentley—each of you have contributed to my scientific development and supported me in the professional and personal journey that is grad school. Thank you for your insight and your encouragement along the way.

Dylan Taatjes. You let me choose a hard project and gave me the tools I needed to succeed. I'm thankful for your faith and your optimism, optimism that pulled me through failed experiments that felt personal. Our lab is a fantastic place to work, and that's largely because of your leadership.

Thank you all.

Love,

Charli

Thesis Contents

List of Tables

List of Figures

Chapter I: Background and Significance

A. INTRODUCTION

The term "gene expression" evokes images of flowcharts outlining the central dogma: DNA is converted to RNA, which is then converted into proteins—the molecular machines that build, maintain, and control the cell (**Figure 1**). However oversimplified this image is, it does at least convey the significance of researching gene expression: by studying gene expression, one is studying the processes associated with taking the information rich DNA code from distilled As, Ts, Gs, and Cs into a panoply of molecules that compose life. The importance of fundamental research into these processes cannot be overstated. Unequivocally, the study of gene expression impacts all other biological studies and provides insight into how life is developed, sustained, adapts, and in the end, fails.

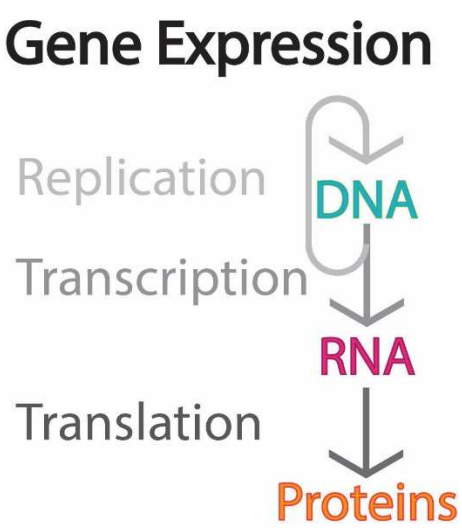


Figure 1: The central dogma of molecular biology. This flowchart shows the path of genetic information in the cell. DNA is converted to RNA by RNA polymerases in the process of transcription. RNA can perform functions itself (ribosomal RNAs, non-coding RNAs, etc.) or it can be translated into proteins by the ribosome. This information flow is generally unidirectional (with the exception of RNA templated DNA enzymes—reverse transcriptases). RNA can rarely template RNA, but DNA templates itself in the process of replication.

This thesis will focus specifically on the process of transcription by the human polymerase RNA polymerase II (pol II). In this process, DNA is converted to RNA, a molecule that both serves as a messenger for the synthesis of proteins and carries out many cellular functions itself. DNA provides the sequence-specific template for the mRNA molecule through base pair interactions. Pol II is the catalytic protein complex that

forms the phosphodiester bonds between nucleotide monophosphates (NMPs) that compose the mRNA transcript. Given that pol II is essential for the transcription of the nearly 20,000 protein coding genes plus an additional unknown number of non-protein coding genes, its regulation is both multi-faceted and intricate. A cohort of proteins, DNA elements, RNAs, and biomolecules exist with a sole purpose: regulation of the pol II enzyme.

B. SETTING THE STAGE FOR POL II

Chromatin architecture regulates transcription of metazoan genes. Textbooks describe metazoan chromatin as being neatly divided into two states: euchromatin and heterochromatin. Heterochromatin contains tightly packed DNA and nucleosomes and is unavailable for transcription; conversely euchromatin is more loosely packed and is available for transcription[1,2]. However, chromatin states likely exist on a continuum with varieties of active and inactive chromatin that cannot be neatly classified as either euchromatin or heterochromatin[3,4]. The nucleosomal histone octamer[5,6] is extensively post-translationally modified, conferring a histone code that is likely cell-type and context specific, thus helping to enact and maintain gene expression programs. Certain post-translational modifications (PTMs) are broadly associated with transcriptional activity (H4K4me3) and others with repression (H3K27me3)[7]. These levels of regulation matter but are not yet fully understood: promoters that are actively transcribed appear to be cleared of nucleosomes, and those nucleosomes that are present are well-organized and enriched for particular chromatin marks[8]. These regions, or nucleosome depleted regions (NDRs) are sites of robust transcriptional activity, likely because the chromatin is most accessible for transcription factor (TF) binding.

C. POL II TAKES THE STAGE

Promoters. Promoters contain the sequences necessary for the recruitment of both general transcription factors (GTFs) and gene-specific TFs. Metazoan promoters seem to be exceedingly diverse with varying degrees of adherence to certain motifs and combinations thereof. Broadly, most metazoan promoters can be described as the region of a gene surrounding the transcription start site (TSS). This site is the position tha encodes the first base of a RNA transcript. The position of the TSS is thus defined as +1. Most promoters contain an Inr element[9,10] that encompasses the TSS (**Figure 2**). As seen in **Figure 2**, additional core elements are represented in mammalian promoters both up and downstream of the TSS[11,12]. Of note is the TATA box, the core promoter element located upstream of the TSS best known for its role in positioning the TATA-binding protein (TBP)[13,14]. Promoters do not require all (or even most) of these elements for transcription, but instead can be classified into two major types: strict TATA box-containing promoters and those with high GC content[15].

Metazoan promoters are also regulated extensively by enhancers in a relationship called enhancer-promoter looping. Enhancers are typically short DNA elements that can be bound by TFs to regulate transcription at their cognate promoter. Enhancers are often located great distances (frequently over 100,000 base pairs) from the promoter they regulate[16,17] but are brought into proximity to their respective promoters by mechanisms not fully understood. However, these enhancer-promoter loops seem to physically reorganize the nucleus in cell-type and context specific manners to facilitate appropriate gene expression programs[18–20].

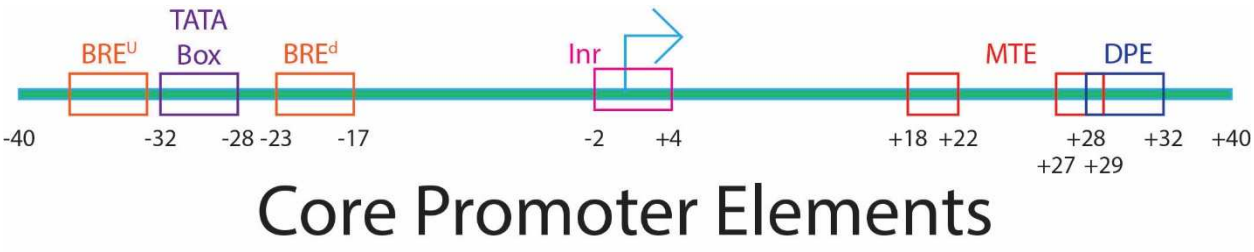


Figure 2: Core promoter elements. The core promoter elements for pol II are indicated by boxes on a linear DNA strand. Their approximate distances are indicated in black with relation to the TSS (black cornered arrow) below the corners of the boxes. Additional general promoter elements such as the TCT, XCPE1, XCPE2, and DCE or not depicted. TBP binds the TATA box; TFIIB binds the BRE (B-recognition elements), and TFIID interacts with the MTE and DPE downstream of the TSS.

Pre-initiation complex (PIC) formation. Assuming that the promoter DNA is available for binding and the appropriate activating TFs have found their cognate promoters and enhancers, a pre-initiation complex (PIC) can form. The PIC is composed of the promoter DNA and GTFs. *In vitro*, a minimally active PIC is composed of the GTFs TBP, TFIIB, TFIIF, and pol II[21–23]. However, the PIC *in vivo* is composed of TFIIB, TBP, TFIIF, TFIIE, TFIIH and pol II, and almost always includes TFIIA, Mediator, and additional promoter-specific TBP-associated factors (TAFs) that, along with TBP, constitute TFIID.

Existing models of PIC formation coalesce around a stepwise assembly model[24–28] (**Figure 3**). In this model, TFIID, IIB, and IIA bind to the DNA upstream of the TSS at loosely conserved promoter elements (BRE and TATA box)[12,13]. TFIIF-bound pol II associates with this upstream promoter complex to form the core PIC. TFIIE then binds and recruits TFIIH, converting the core PIC into the closed PIC, a stable transcriptionally competent complex[29,30]. It is unclear at which stage of this process Mediator associates with the PIC. Mediator does not have known DNA binding functions, although some research supports Mediator association with certain chromatin marks[31]. However, given the role of Mediator in facilitating activated transcription through enhancer-promoter looping[32–35], it is likely that enhancer-bound Mediator binds to the assembled closed PIC, both stabilizing it and

conferring context-specific regulatory information through allosteric structural changes. It is also possible that Mediator could bind after the transition of the closed PIC into the open PIC. This transition occurs in the presence of ATP, when the TFIIH helicase subunit, XPB, melts the duplex DNA, forming an open bubble. At this stage, transcription is primed and ready to begin in the presence of the RNA monomers, NTPs.

Initiation. In the presence of NTPs, pol II begins to transcribe the nascent RNA by matching NTPs with their base pair on the template strand. Pol II catalyzes the formation of phosphodiester bonds to synthesize the growing mRNA chain. At this stage, pol II still maintains strong contacts within the PIC. These contacts compose a tether that must be severed for pol II to advance. Evidence supports that this tether often persists, resulting in abortive transcription cycles in which the short mRNA is dissociated from the polymerase[36]. These cycles continue until pol II breaks contacts with the PIC in a process called promoter escape, the rate limiting step of transcription initiation[37].

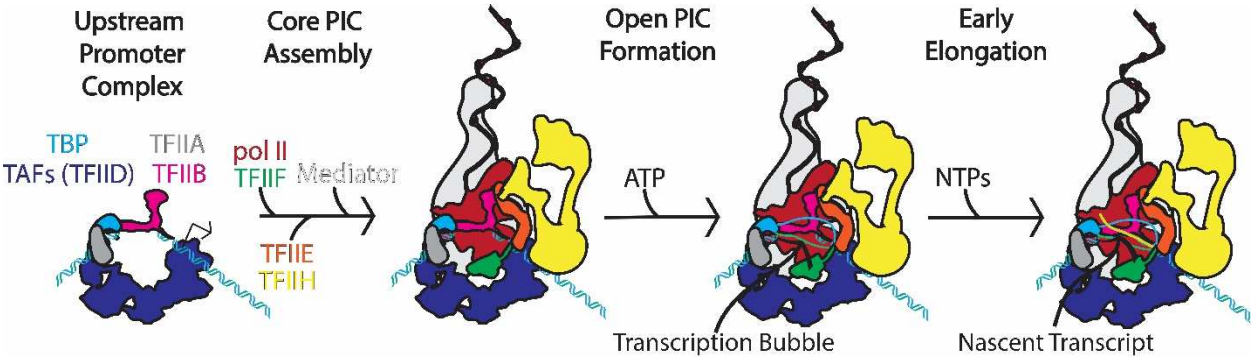


Figure 3: Early steps of pol II transcription. Pol II transcription is tightly regulated by several GTFs (shown here) as well as gene-specific TFs. The upstream promoter is first bound by TBP (at the TATA box) alone or as part of the TFIID complex. Louder *et al.* Nature 2016 and Patel *et al.* Science 2018 suggest that the other TAFs in the TFIID complex may load TBP onto the TATA box. TFIIA and TFIIB the bind, stabilizing the upstream promoter complex. TFIIF and pol II then bind, forming the core PIC. TFIIE and TFIIH subsequently bind converting the complex into the closed PIC. We postulate herein that this is also when Mediator binds. In ATP-dependent manner, TFIIH opens the transcription bubble to form the Open PIC. In the presence of NTPs, early elongation (prior to promoter escape) can begin.

Promoter escape. The likelihood of pol II breaking contacts with the promoter and entering early elongation increases proportionately to the length of the synthesized mRNA chain. Synthesis of three phosphodiester bonds (a four-nucleotide RNA) significantly increases the chances of early elongation commitment[38]. The TFIIB loop that interacts extensively with the non-template DNA strand of the open bubble is displaced by the successful synthesis of a 12-13 base RNA[30,39]. At approximately 14 bases, the mRNA chain is long enough to form a stable ternary complex between the mRNA and the two strands of DNA[40]. At this state of synthesis, the pol II register neatly shifts so that the template DNA strand and 3' end of the mRNA chain are in ideal alignment for catalysis of the next phosphodiester bond. The active site of pol II is ready to enter early stages of elongation. In a TFIIH-dependent manner, pol II must dissociate from the PIC. Previous work has suggested that this step is entirely dependent on the XPB helicase activities of TFIIH[41]. However, recent work suggests that the protein kinase CDK7 plays a unique role in the stage: CDK7 must phosphorylate Ser5 of the pol II CTD heptad repeat in order to electrostatically repel it from the PIC, particularly from Mediator[33,34,42–45]. These changes facilitate promoter escape and entry into early elongation, but as of yet are not fully understood.

Pol II promoter-proximal pausing. Broadly defined, pol II promoter-proximal pausing (hereafter referred to as pausing) occurs between 20 and 100 base pairs downstream of the TSS. This process occurs at almost all human genes and is conserved amongst metazoans[46]. It is unclear which components of the PIC are still associated with pol II during early elongation and pausing, though it is likely that TBP and TFIIA remain at the promoter along with Mediator. It is canonically thought that the rate of elongation of pol II

slows down or halts altogether. An interchange of some combination of pol II associated GTFs (TFIIB, TFIIE, TFIIF)[47] for the highly studied pause-inducing factors DSIF and NELF occurs, stabilizing this paused state. A comprehensive list of factors strongly implicated in pausing are listed in **Table 2**.

Several possible mechanisms for how DSIF (composed of SPT4 and SPT5) and NELF (composed of NELFA, NELFB, NELFC, and NELFE) may stabilize paused pol II have been elucidated through recent cryo-electron microscopy (cryo-EM) structures of the paused complex: NELF may restrict pol II mobility, prohibiting the ratcheting mechanism necessary for mRNA synthesis; additionally, NELF may preclude NTP diffusion into the

Factor	Suggested Role in Pausing/Pause Release
NELF	Binds to pol II/nascent RNA to stabilize paused pol II
DSIF	Binds to pol II/nascent RNA to stabilize paused pol II; part of EC after phosphorylation
TRIM28	Roles in pause mainteneance
Cohesin	Regulates levels of elongating pol II
Integrator	Regulates levels of elongation pol II
PAF1	Regulates levels of elongating pol II
PARP-1	Promotes pause release byADP-ribosylating NELF, evicting it from paused pol II
eRNAs	Promotes pause release by acting as a NELF decoy
BRD4	Promotes pause release/elongation
P-TEFb	Promotes pause release/elongation; part of SEC

Table 1: Factors implicated in pol II pausing and pause release. Along with NELF, DSIF, and P-TEFb, many other factors have been implicated in pause regulation to varying extents. Some of these are listed in this table from top to bottom based on whether they are more involved in pausing (top, red) or pause release/elongation (bottom, green).

active site and/or prevent the factor TFIIS from associated with pol II and cleaving backtracked transcripts, thus stabilizing pausing[47]. Furthermore, DSIF and NELF interact extensively with the nascent mRNA transcript. It has been shown in functional *in vitro* assays that interactions of NELF subunits with the mRNA are critical for pause stabilization. Indeed, hairpin structures formed by the nascent mRNA are enriched at paused sites[48], suggesting that the mRNA may facilitate NELF and/or DSIF recruitment and its interactions with those factors may stabilize pol II pausing. **Figure 4** provides a visual of NELF and DSIF stabilizing pausing.

Pol II pause release and elongation. *In vitro* and cell-based studies point towards the transcriptional cyclin-dependent kinase (CDK) P-TEFb (composed of CDK9 and CCNT1) as being the primary factor required for pause release. P-TEFb phosphorylates DSIF—converting it into an elongation competent form—and phosphorylate NELF, evicting it from the pol II complex (**Figure 4**) and allowing the association of the pre-elongation factor PAF1[49]. P-TEFb also phosphorylates the pol II CTD, promoting the association of elongation-factor SPT6 and various co-transcriptional regulators[50,51]. Further post-translation modification of NELF by PARP-1 (ADP-ribosylation) contribute to NELF eviction[52]; it has also been suggested that NELF may be titrated away through eRNA interactions facilitated by enhancer-promoter looping[53], perhaps facilitated by cohesin[54]. The changes made to this paused pol II complex allow for its transformation into the super elongation complex (SEC)[55]. The SEC is transcriptionally stable and synthesizes mRNA at a rate of approximately 2000 bases per minutes[56], though slowing at intron-exon junctions, presumably to facilitate splicing[57]. Pol II transcribes until the 3' end of genes, where it is

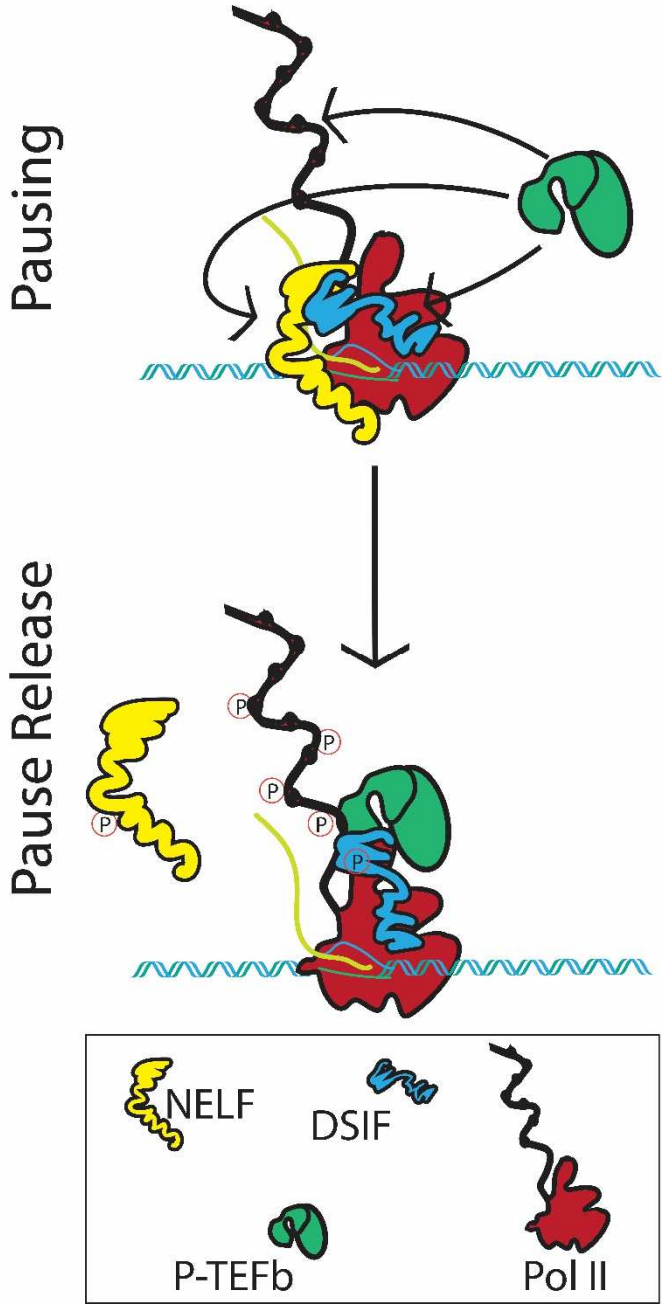


Figure 4: Simplified model of NELF and DSIF dependent pol II pause stabilization and P-TEFb dependent pause release. The canonical model of pol II pausing suggests that NELF and DSIF bind slowed or halted pol II to stabilize pol II pausing. It is likely that NELF and DSIF are recruited by the extrusion of the nascent transcript from the RNA exit channel. When NELF and DSIF are bound to pol II, the simplistic version of the paused complex is formed. P-TEFb is then recruited to the complex where it phosphorylates NELF, evicting it from pol II, and phosphorylates DSIF, converting it into its elongation-competent form. P-TEF also extensively phosphorylates the pol II CTD, specifically at residue Ser2 of the heptad repeat YSPTSPS.

terminated, poly-adenylated, and packaged for translation. Free pol II can the re-enter the transcription cycle, at the stage of PIC formation.

D. OPEN QUESTIONS REGARDING EARLY STEPS OF POL II TRANSCRIPTION

How are contacts between the PIC and pol II broken during promoter escape?

For pol II to begin elongation, extensive contacts with the PIC must be broken. For instance, the pol II CTD has been shown to bind the PIC component Mediator with a sub nanomolar K_d[58]. Phosphorylation of the pol II CTD by TFIIH is reported to decrease the affinity of the CTD for Mediator and, thus, weaken the interactions between pol II and the rest of the PIC[42]. While it is possible that phosphorylation of the pol II CTD by TFIIH alone may be sufficient to activate promoter escape, another factor, the Mediator kinase module (referred to as the CDK8 module), has been implicated in promoter escape through biochemical reconstitution[59] but has not yet been thoroughly characterized. The CDK8 module is a 600kDa complex composed of the transcriptional CDK, CDK8, its cyclin partner, CCNC, and two other large subunits, MED12 and MED13. It is known that the CDK8 module binds to Mediator in a fashion that precludes binding of pol II based on biochemical and cryo-EM data[59,60]. An *in vitro* model has been proposed by which the CDK8 module binds to PIC-bound Mediator, facilitating allosteric shifts in Mediator and the PIC that eject pol II from the PIC[59]. Previous work indicates that the kinase activity of the CDK8 module is not required for its roles in promoter escape[59]. If the effects of the CDK8 module on promoter escape are as large as anticipated, it may be possible to gain a more satisfying view of how pol II breaks such extensive contacts with the PIC.

DSIF and NELF stabilize pol II pausing. How is it established? As articulated in the section on pol II pausing, DSIF and NELF are central to stabilizing pausing. While an abundance of *in vitro* and cell-based data have established clear roles for these protein complexes in maintaining pausing, they have not shown that DSIF and NELF facilitate the establishment of pol II pausing. Furthermore, the methods used to draw conclusions about NELF and DSIF are not sufficient to establish the mechanism by which pausing is established. Cell-based nascent sequencing assays have demonstrated that the depletion of NELF from cells both increases and decreases pausing depending on the promoter[61]. Additionally, *in vitro* transcription experiments using nuclear extracts have shown that NELF or DSIF depletion decreases pausing and that this activity can be titrated back with purified proteins[62,63]. However, these *in vitro* experiments ignore baseline pausing in the absence of NELF and DSIF, leaving open the question of how pausing is established in the first place.

Does TFIIH affect pol II transcription after promoter escape? TFIIH has established roles within the PIC involving promoter opening and promoter escape. This 10-subunit complex contains three enzymatic subunits: two helicases (XPB and XPD) and a CDK. The XPB helicase is necessary to form the transcriptional bubble, making the template DNA strand accessible. The CDK7 kinase phosphorylates the pol II CTD at Ser5 and Ser2, enabling promoter escape. The CDK7 kinase is also known to regulate co-transcriptional processes such as elongation rate, capping, and splicing via additional phosphorylation of the pol II CTD[51,64]. Recent work involving the inhibition or inactivation of CDK7 has shown that it may regulate additional steps of transcription through its kinase activity. CDK7 small molecule inhibition increased pausing genome-wide in human cells[65]. Additionally,

phosphoproteomics have shown that CDK7 directly or indirectly targets the DSIF subunit SPT5 as well as NELF[66–68]. These phosphorylation events are proposed to impact NELF and DSIF recruitment to paused complexes[64,66,69]. Taken together, these data indicate a need for biochemical analyses of post-promoter escape roles for TFIIH in transcription regulation.

How are the transcriptional CDKs regulated? CDK7, CDK8, and CDK9 typically function as part of the larger multi-subunit complexes TFIIH, the CDK8 module, and P-TEFb/SEC, respectively. These complexes play central roles both as kinases and scaffolds throughout early stages of transcription and into elongation. Evidence exists that these kinases have inter-dependent activities that affect pol II processivity. TFIIH has been shown to phosphorylate P-TEFb[66]. The CDK8 module seem to increase elongation of stimulus response genes through interactions with P-TEFb under both hypoxic and starvation conditions[70,71]. Of the three transcriptional CDKs discussed herein, the CDK8 module has the most clearly defined role as a gene-specific TF regulator[72–75]. How these kinases regulate each other has not been well-studied but may provide clues about signaling events that drive transitions from one step of transcription into the next.

E. UPCOMING CHAPTERS

In the upcoming chapters, I will be describing results that may answer some of these "open questions" using a biochemical reconstitution approach accompanied by cell-based experimentation where appropriate and currently possible. While other early steps of transcription have been rigorously described through biochemical approaches, biochemical analysis of pol II pausing in particular has been less exhaustive. Pausing has been primarily studied using cell-based methods. These methods have shown the prevalence of pausing

across the genome and have elucidated some of its roles as a fundamental transcription regulatory mechanism[46,76]. However, these cell-based studies are not capable of assessing the contributions of individual factors on pausing or making causal claims due to side effects associated with perturbations (e.g. off target effects of knockdowns), timing, and confounding issues when dealing with mass averages (millions of cells with slight changes in promoter behavior). While biochemical approaches have been used to examine pol II pausing, extracts have primarily been used to facilitate transcription on artificial promoters. Using an extract rather than purified proteins presents similar challenges to those seen in cells due to complicating effects from the undefined components in the extract. Artificial promoters are alse derived to increase transcription, perhaps at the expense of a later of regulation. This thesis will focus on developing a biologically relevant reconstituted transcription system for the study of pol II pausing in an to help bridge the gaps in our understanding of pol II pausing mechanisms. While a fully reconstituted system certainly has its drawbacks, it is the approach needed to close gaps in the fields current models.

Chapter II, "Innovations in the Reconstitution of Human Transcription and Development of a Pol II Pausing Assay" goes into detail about the strategies and experiments used to develop the pol II pausing assay that will be the workhorse experimental tool used for Chapter III and Chapter IV. It describes how a fully reconstituted system was conceived, developed, and tested in order to detect changes in early steps of transcription, particularly that of pol II pausing.

Chapter III, "TFIID Enables Pol II Promoter Proximal Pausing," describes work that has been submitted for publication. This story focuses on establishment of pol II pausing.

Surprisingly, we found that pausing *in vitro* occurs in absence of DSIF and NELF but is dependent on the GTF TFIID. We further suggest that lobe C of TFIID facilitates pol II pausing and test this hypothesis in both human and *Drosophila* cells using the Trim-Away and siRNA knockdown methods to knockdown TFIID subunits. It represents a "complete" story, but one that should inspire future work.

Chapter IV, "Preliminary Data on the Regulation of Early Steps in Transcription by the Transcription Kinases P-TEFb, the CDK8 Module, and TFIIH" describes preliminary data regarding the roles of transcriptional kinases in regulation of early steps of transcription. Cell-based data from the Taatjes lab and others are discussed, but the focus of the chapter is the application of the reconstituted pausing assay to study promoter escape, pausing, and elongation. Much of the work contained therein represents a limited number of replicates (with a few exceptions) but often corroborates phenomenon observed in cells.

The final chapter, Chapter V, summarizes the findings of Chapter II-IV and suggests future directions for research. The appendix contains multiple sections that will describe the methods used within the research chapters, including protein purifications, pausing assay details, and cell-based methods. Additionally, it will highlight additional interesting but preliminary findings that may prompt further investigation.

In summary, this thesis focuses on human pol II transcription. It builds on previous work across biochemistry and molecular biology to address specific outstanding questions about early steps in transcription at the mechanistic level with a particular impetus on pol II pausing. Cheers, and bonne lecture.

Chapter II: Innovations in Reconstitution of Human Transcription and Development of a Pol II Pausing Assay

A. INTRODUCTION

Over the past several decades, various strategies have been executed to perform cell free RNA pol II transcription. These strategies have varied in almost every way, save the need for a few indispensable component parts: pol II, a limited selection of GTFs, template DNA, and NTPs. However, the source of pol II and the GTFs in the assay varies from the use of crude nuclear extracts to highly purified factors. Additionally, the specific GTFs needed for a given protocol is also tied inextricably to the template DNA. For instance, different combinations of factors are required for a fully chromatinized supercoiled template than for a partially annealed bubble template. Each assay, with its respective use of a template type and source of protein factors, is rich with both advantages and disadvantages. With the goal of studying pol II pausing and elongation, each element of the transcription reaction must be carefully assessed for its potential benefits and limitations.

The DNA template chosen for transcription has varied widely. The adenovirus major late promoter (MLP) has been extensively used in the literature to study early steps of transcription[22,37,77], including pol II pausing and pause release[78]. This promoter is used widely for transcriptional studies due to its high activity *in vitro*[79]. However, the properties that make it an excellent option for producing high levels of transcription suggest that this promoter may be less suited to studies of transcription regulation. Other work undertaken to look at pausing specifically has relied on the HSP70 promoter from *Drosophila*[63] and has been shown to be a model promoter for the study of pol II pausing both *in vitro*[63] and in

cells[80]. The human HSP70 promoter (specifically, the HSPA1B isoform) has been shown to be regulated similarly to its *Drosophila* counterpart[81,82]. Furthermore, this promoter has been shown to be adaptable to studies of pol II pausing in a nucleosome free environment[82,83].

While studies of PIC assembly, transcription initiation, and promoter escape have all been conducted with purified factors over the past several decades[21,22,25,28,37], very little work has been attempted using exclusively purified factors to study pol II pausing. Without a fully defined system to study pol II pausing and pause release, mechanisms of how pol II pauses and then releases remain elusive. The *in vitro* work that has been conducted relies on nuclear extracts as the source of GTFs[63,78,82,84], thus complicating data interpretation and confounding conclusions. Part of the rationale for using nuclear extracts to study pausing is that not only are the purified GTFs that make up the PIC required but pausing and pause release factors must also be purified. The sheer number of factors and difficulties involved in purifying some of them makes the task a formidable challenge. However, given the importance of pausing as a genome-wide regulator of transcription[46,61,76,85], a better understanding of it mechanistically may inform future developments in therapeutics and deepen our understanding of gene expression.

Though options for building a reconstituted pausing assay abound, this chapter will focus on the development of a fully reconstituted and defined pausing assay on the native human HSPA1B promoter. It will detail the various strategies tested and implemented to develop a reliable pausing assay that can detect changes in pausing through direct ^{32}P-NMP incorporation into the transcript. The developed assay should be amenable to further

development (i.e. chromatinized templates or adapted to single molecule studies) but is in and of itself a valuable tool for beginning mechanistic investigations into pol II pausing.

B. INNOVATIONS IN RECONSTITUTION OF HUMAN TRANSCRIPTION

Purification of the GTFs. The Taatjes lab and others have developed protocols for successfully purifying the GTFS. This collection of factors comes from a mix of recombinant human factors generated from *E. coli* (TFIIA, TFIIB, TFIIE, TFIIF, TBP) and factors purified from HeLa nuclear extract (pol II, TFIID, TFIIH, and Mediator). Example purifications of these factors are shown in **Figure 5a**. Protocols for the purification of these factors can be found in the supplement of previous work[59]. Some protocols have been modified for simplicity, enhanced enzymatic activity, increased purity, or clarity. Protocols for purification of those factors can be found in Appendix C. Most labs who have studied pol II pausing *in vitro* have relied on nuclear extracts as their source of GTFs[62,64,78,82,86–89]. By using purified GTFs rather than nuclear extracts, more definitive conclusions can be drawn from the experiments conducted. In a fully designed system, factors can be added, removed, edited, etc. This level of control is not possible with nuclear extracts.

DNA template selection and purification. The human HSPA1B promoter is an ideal promoter for studying promoter proximal pausing. It has been studied in organisms from *Drosophila* to humans[80,90] and has been used in *in vitro* transcription assays previously[63,82]. Importantly, unlike other attempts at *in vitro* reconstitution of pol II pausing, I chose to use a wildtype native template, thus increasing the biological relevance of any findings determined from its use.

The HSPA1B gene contains a single exon and three promoter proximal activator binding elements: the HSE[91] (which binds HSF1) as well as a Myc and Max heterodimer

binding site[82]. I chose to use the -500 to +216 portion of this promoter relative to the TSS as the DNA template in the reconstituted transcription assay. This portion contains all three activator binding sites in the upstream direction as well as the entirety of the 5' UTR in downstream direction. A diagram of the template is shown in **Figure 5b** and **5c**. This segment of the promoter was also what was previously used for *in vitro* transcription by previously[82]. Furthermore, given the definitions of where pol II pauses available in the literature (within the generous range +10 to +100 bases relative to the TSS[76,83,92,93]), this segment of the HSPA1B promoter theoretically allows study of both pausing and early elongation. The template was cloned and purified as described in Appendix B.

It is important to note that a fully annealed template was chosen. While other labs have found success using "bubble" templates in which a DNA mismatches surrounding the TSS promote the formation of a bubble that allows for transcription in the absence of TFIIH and TFIIE[77,94], this choice did not make sense for the development of an assay with the intent of studying pol II pausing. One driving force behind choosing to use the fully annealed template was the knowledge that TFIIH likely plays a role in pol II pausing[45,64,95]. Given that the role of TFIIH was likely to be studied later (and indeed is in Chapter IV), its inclusion during the development of the transcription assay seemed meritorious. Additionally, evidence suggests that ternary structures formed between the nascent RNA and the template DNA impact pol II pausing[47,49]. The stabilization of a permanent "open bubble" at the TSS might enhance such interactions in a non-biologically relevant fashion and thus complicate data interpretation.

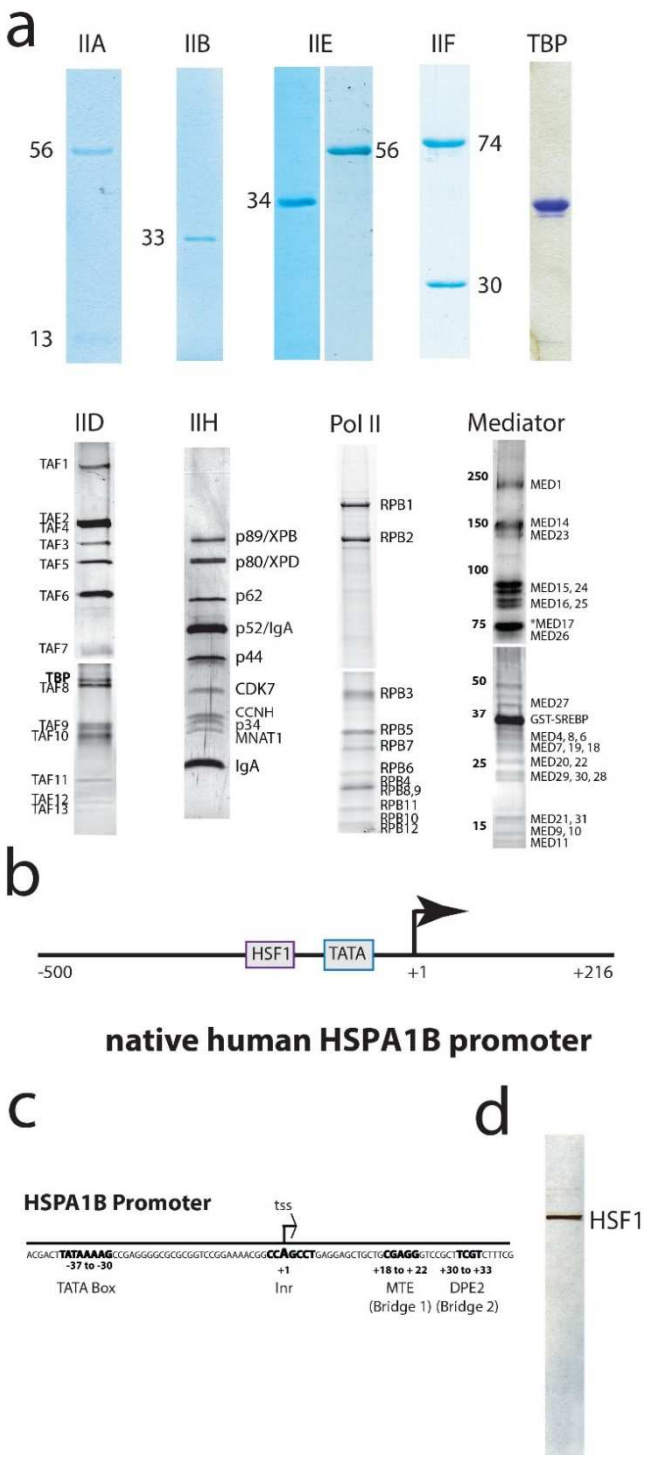


Figure 5: Elements of a fully reconstituted transcription system. a. GTFs are purified as previously described in Kneusel et al. G&D 2009. IIA, IIB, IIE, IIF, and TBP are expressed and purified from E. coli. IID, IIH, pol II, and Mediator are purified from HeLa cells. A modification on the expression of IIE can be found in Appendix C. **b.** The human HSP70 promoter (specifically HSPA1B) was chosen for use in the reconstituted transcription assay. HSPA1B has been highly studied with regards to pol II pausing as described in the text. A portion of the endogenous promoter was selected from -500 to +216 with respect to the TSS. This portion of the promoter contains the HSF1 binding site (HSE) as well as upstream Myc and Max biding sites. The designated +215 3' end of the template corresponds with the end of the intron-less HSPA1B 5'UTR. **c.** The HSPA1B promoter contains idealized promoter elements: a TATA box, Inr, and both the MTE and DPE2 elements that together compose the Bridge according to the ElemeNT Analysis tool generated by Sloutskin et al. Transcription 2015. It can be found at http://lifefaculty.biu.ac.il/gershon-tamar/index.php/resources. **d.** The activator, HSF1 is purified from E. coli via a 6X-His-tag as described in Appendix C.

Activator selection and purification. The HSPA1B promoter contains binding sites for HSF1, Myc, and Max. The Taatjes lab has the plasmids and capability to express and purify all three and have done so according to the protocols described in Appendix C. A representative HSF1 purification is shown in **Figure 5d**. However, to study pausing on the HSPA1B gene, I chose to focus exclusively on the HSF1 activator. The HSPA1B promoter has been most frequently and extensively studied under conditions of heat shock. The role of HSF1 in this process has been thoroughly characterized: upon heat shock, HSF1 dissociates from the HSP90 complex, is extensively post-translationally modified[96–99], translocates to the nucleus, trimerizes, and promotes activation of the HSPA1B gene through binding to the HSE[91] (**Figure 6a**). The full HSP complex is not reconstituted in our system; consequently, HSF1 is available to bind the HSE even in the absence of a heat shock stimulus and evidently does not require PTMs.

It may be counterintuitive that in attempting to study pausing, I have chosen to use a factor that has been shown to promote pause release upon heat shock[76,80,100]. However, as shown through ChIP-seq[101] and nascent sequencing methods[102,103], the promoter-proximal peak indicative of pausing does not decrease with heat shock (**Figure 6b**). Instead, it remains constant or slightly increases while elongation signal increases dramatically. These data suggest pausing is maintained, but that elongation and initiation are also increased upon heat shock. In our system, we expected to see an overall increase in transcription with use of the HSF1 activator, including a proportional increase of transcripts in the paused region. A summary of the key elements in the proposed transcription assay is shown in **Table 2**.

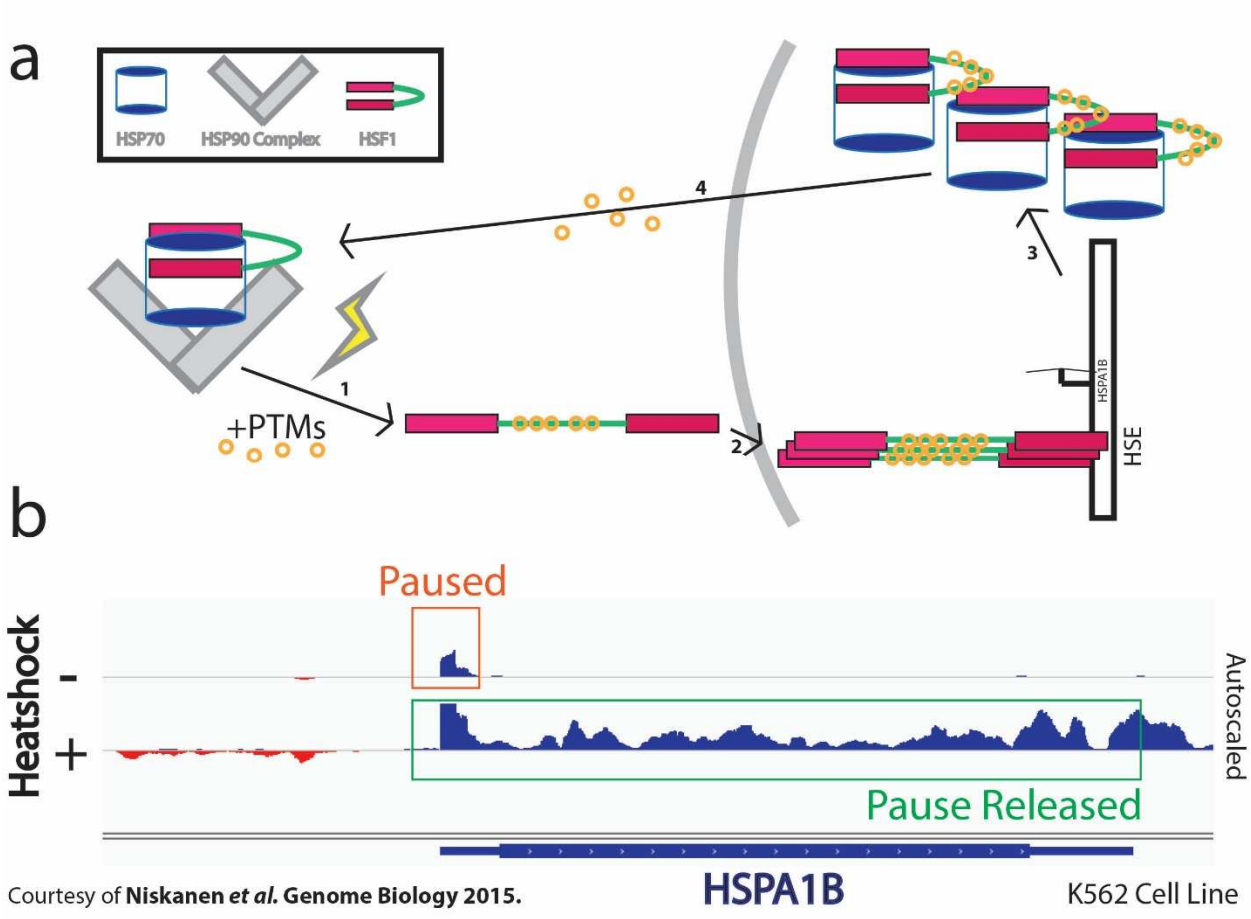


Figure 6: HSPA1B is a quintessential model of pol II pausing. a. Under normal conditions, HSF1 is bound as a monomer to the HSP70 protein as part of the HSP90 chaperone complex. Under conditions of heat shock or stress (1), HSF1 is post-translationally modified (sumoylated, acetylated, and phosphorylated) and translates to the nucleus (2), where it forms a trimer and binds the HSE, driving expression of target genes, such as HSPA1B, which encodes the HSP70 protein. After the stress has been resolved, HSF1 is bound by HSP70 (3) and translocated back into the cytoplasm (4)) where HSF1 and HSP70 re-associates with the HSP90 chaperone complex. **b.** ChIP-seq profile pre- and post-heat shock in K562 cells. Profile shows pol II piled up near the promoter under normal conditions. Upon heat shock, the promoter-proximal pol II peak increase, but pol II is also demonstrated to release into the gene body. Profile is an example of pol II pausing followed by stimulus specific pause release.

Transcription Assay Quick Guide	
DNA Template	Fully Annealed, Linear, Human HSPA1B Promoter 716 bases long (-500 to +216 relative to TSS)
Activator	HSF1 Myc and Max also available for use
GTF Source	Purified Factors (NO EXTRACTS) IIA, IIB, IID, IIE, IIF, IIH, pol II, and Mediator
Transcription Detection Method	^{32}P-CMP Incorporation Into Transcript Derived from α-^{32}P-CTP

Table 2: Transcription assay quick guide. Basic details of the transcription assay that remain constant throughout troubleshooting and development of the pausing assay are indicated as a reference. Additional details, including template sequence and promoter content as well as the activation mechanism of HSF1 and a gel showing its purification are shown in later figures in this chapter. Purification gels of the GTFs are also shown. Descriptions of purifications of the template, activator, and GTFs are given in Appendix C.

Experimental set-up. Previous members of the Taatjes lab and others have engineered a reconstituted minimal transcription assay on a fully annealed linear template[104] based off previous reconstitution methods established in the Tijan lab[21]. This assay differs from the one proposed herein in that it is conducted with TBP rather than TFIID, does not use a transcriptional activator, and is meant to look at runoff transcription (the longest possible transcript originating from the TSS). The general outlines of this assay were used as a starting point for testing the use of TFIID on a naked template and for determining if activated transcription was possible in this context.

The mechanics of the assay are described in detail in Appendix B. This assay will be referred to from hereafter as the "standard run-on assay". Generally, HSF1 is bound to the template in a transcription-friendly buffer. A master mix consisting of the GTFs is made and then added to the activator bound template and allowed to assemble into a PIC. Transcription is then initiated with a full complement of NTPs (ATP, GTP, UTP, and CTP), including ^{32}P-CTP, and allowed to proceed for 30 minutes. The reaction is then stopped. RNA is isolated and prepared for analysis via gel electrophoresis on a sequencing gel. The

sequencing gel is dried, exposed overnight or longer, and imaged on the Typhoon. Transcriptional activity is then analyzed via intensity of the runoff signal—an RNA of ~216 nucleotides in length. **a** shows a diagram of this experimental set-up.

Results. Transcription of the native HSPA1B promoter was observed using fully reconstituted GTFs. A population of RNA molecules migrating slightly above the 200 base DNA marker can be seen in **b**, indicated as "Runoff transcript(s)". Furthermore, as shown in **b**, the addition of HSF1 to the assay increases transcription modestly (19.4% +/- 4.2%), indicating that a small degree of activation is possible on a non-chromatinized template. The runoff transcript can be mapped by primer extension and shown to originate at the HSPA1B annotated TSS (**c**). These results show that the purified GTFs are active, capable of forming a PIC, and that that PIC is transcriptionally competent on a fully annealed linear template. Additionally, these experiments show that a PIC containing purified TFIID rather than TBP is transcriptionally competent and that activation by HSF1 is achievable.

A time course conducted with this standard run-on assay is shown in **d** and shows that transcription increases over time. Interestingly, the appearance of short transcripts (approximately between 20 and 50 nucleotides in length) is observed. However, the relative signal of these products is very low compared to that of the runoff transcripts. This phenomenon is likely due to the increased number of opportunities for ^{32}P-CMP to incorporate into longer transcripts with proportionately more "C's". Attempts at quantitating the ratio of short, potentially paused transcripts at the 30-minute time point were unsuccessful and difficult to replicate.

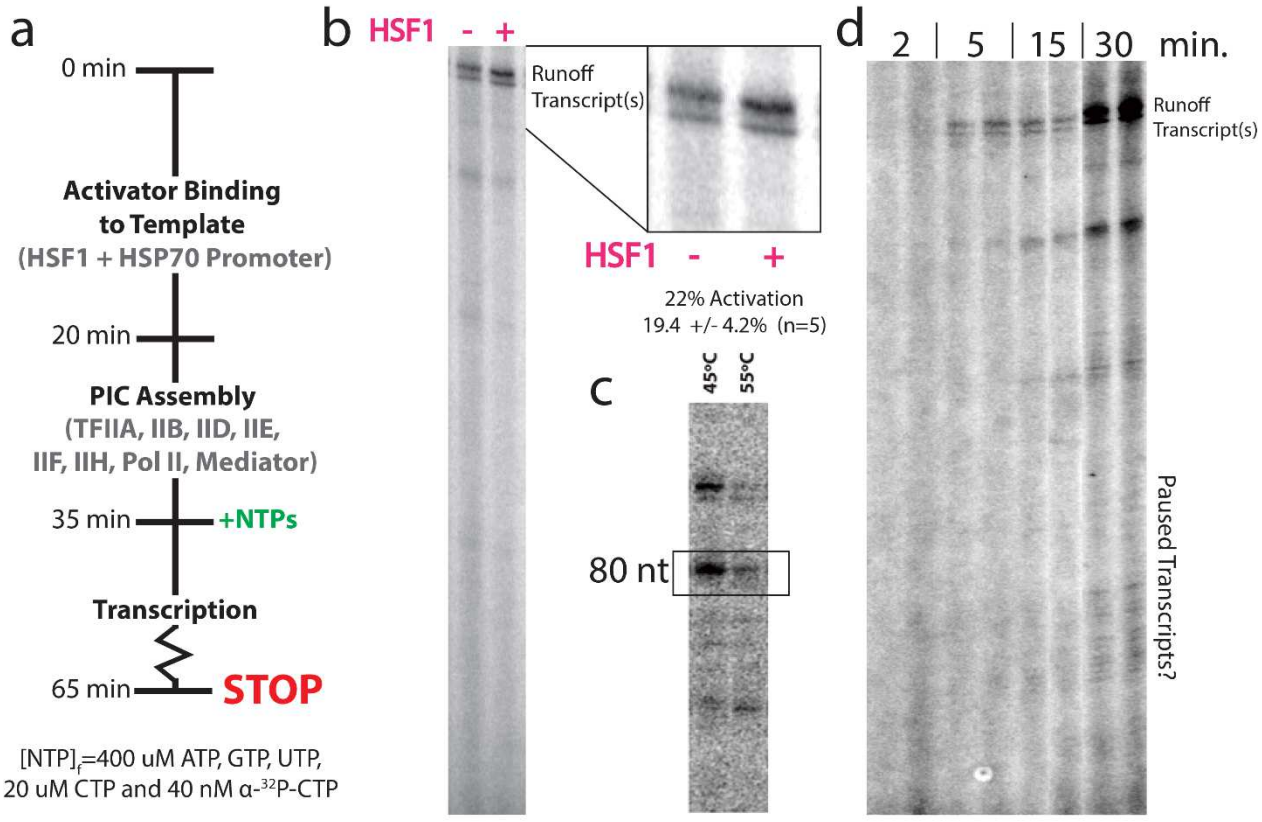


Figure 7: Reconstitution of activated transcription on a linear native human promoter. **a.** Schematic of run-on transcription assay. **b.** HSF1 activates transcription on the HSPA1B promoter in vitro. Activation is modest (~22%), likely to do the non-chromatinized template Radiolabeled transcripts are run on a 6% urea denaturing sequencing gel. **c.** Primer extension maps the TSS of human pol II transcribed HSPA1B RNA to that annotated in vivo. T4-PNK labeled primer is designed to bind the transcribed RNA 80 bases upstream of the annotated TSS. Primer extension then followed and confirmed the TSS. Likely due to the highly structured nature of the HSPA1B RNA, extension temperatures had to be elevated to 45°C rather than the 37°C typically used. **d.** Time course of reconstituted transcription system showing an increase in runoff transcription over time and perhaps the appearance of paused products. Note that total phosphor-signal increases over time, suggesting multi-round transcription or delayed promoter escape by a significant portion of polymerases.

Discussion. The reconstitution of activated human transcription on a fully annealed native template is a major step forward for the biochemical analysis of transcriptional mechanisms. This system is fully dependent on the purity and activity of each of the GTFs—a significant barrier for many labs attempting to study early steps of pol II transcription. The use of an activator mimicking a stimulus response (in this case, heat shock) has not previously been achieved on non-chromatinized templates. Taken together, these results set

us up to study early steps of transcription, including pol II pausing in a physiologically meaningful context.

Variations on this assay have also been useful in collaborations with other labs. The lab of Shimon Weiss at UCLA is currently attempting to translate single molecule investigations into early steps of human transcription from an open-bubble template system to a fully annealed template in which TFIIH and TFIIE are required. Additionally, I have collaborated with the lab of Eva Nogales at UC Berkeley to help them set up an activated *in vitro* transcription system by which they can perform functional assays to validate structural data of activator bound TFIID. These collaborations further demonstrate the demand by investigators for the cell-free reconstituted transcription system enhanced herein.

C. DEVELOPMENT OF A POL II PAUSING ASSAY

Though incredibly useful for answering a wide variety of questions concerning transcription mechanisms, the standard run-on assay is not sufficiently capable of addressing many questions about early steps of transcription, particularly pol II pausing. Most of the components needed to address these questions are there, but short transcripts, perhaps indicative pausing, were not easily observed. On the surface, this was not necessarily surprising—after all, the GTFs were purified and included in the transcription reaction, but neither of the canonical pausing factors (NELF and DSIF) were present. From a biochemical reconstitution standpoint, we were concerned. We hypothesized based on previous data from other labs that these factors would be required to reconstitute pausing,

but other labs were either working in cells or with crude extracts *in vitro*, leaving open the possibility that the addition of NELF and DSIF alone would not be sufficient to cause pol II pausing. One can imagine that the search for additional factors mediating pausing could take several PhDs and still prove fruitless.

Furthermore, while the standard run-on system proved capable of assessing runoff transcription (a proxy for elongating pol II), it was unknown if it could detect transcripts indicative of pol II pausing. Because these transcripts are short (generally defined by literature as being between 10 and 100 bases long), less ^{32}P-CMP would be incorporated into the transcript, thus making their visualization by phosphor imaging more difficult. Based the ratio of radiolabeled ("hot") CTP to unlabeled ("cold") CTP in the standard run-on assay, every time a CMP was incorporated into the transcript, it stood a 1 in 500 chance of incorporating the hot CMP. Longer transcripts that contained more C's had a greater probability of incorporating the hot CMP, but the visualization of shorter transcripts was disadvantaged. Though we had concerns about building a reconstituted pause from a potentially large number of protein factors, without addressing the issue of paused transcript visualization, we would effectively be assaying protein factors for their effects on pausing without any means of detecting the phenomenon.

Detecting short transcripts. The fundamental problem with the standard run-on assay detection scheme was the probability of a hot CMP incorporating into a short transcript compared to that of a cold CMP. In an effort to maintain the overall structure of this assay, the ratio of hot CTP to cold CTP in the initiation step of the transcription reaction was increased to adjust the odds of detecting paused transcripts. Ideally, the same concentration of hot CTP would be used in the assay as there was total CTP in the standard

run-on assay—20 μM. However, given the hazards and costs associated with dealing with radiolabeled nucleotides at that concentration, this strategy was infeasible. Instead, cold CTP was entirely removed from the reaction, apart from that caused by radioactive decay of the source hot CTP. With this alteration, the total concentration of CTP in the assay was shifted from 20 μM to 40 nM, adjusting the probability of hot CMP incorporation at any given site from 0.2% to nearly 100%. This strategy yielded a surprising outcome: visualization of short transcripts in the defined paused region (figure 4b, lane 1 and lane 2). Furthermore, reactions stopped after 10 minutes of elongation and 30 minutes of elongation showed an overall increase in paused transcripts at 30 minutes compared to 10 minutes but little elongation. The protocol for this version of the transcription assay is described in Appendix B and outlined in **Figure 8a.**

As seen in **Figure 8b** lanes 1 and 2, while short transcripts are visible in the paused region, very little runoff product is observed. Concerns about the effects of long-term nucleotide deprivation (effectively, CTP starvation) forced attempts to further improve the assay in an effort to detect both paused and elongated transcripts.

Implementation of the pulse-chase initiation regime. If pol II was indeed stalled due to CTP starvation, it seemed probable that a cold CTP chase at a physiologically relevant concentration might release the stall. The experimental protocol was altered such that the initiation with exclusively hot CTP remained the same to ensure labeling of the short, potentially paused transcripts, but a cold CTP chase was added ten minutes later **(Figure 8b**, lane 3). This "pulse-chase" method ensured radiolabeling of all transcripts initiated within the first ten minutes as well as the possibility of minimizing the confounding effects of nucleotide starvation and subsequent stalling/arrest and a similar

strategy has been implemented previously[89]. Indeed, after the chase, short products seemed to release (**Figure 8b,** lane 3) compared to reactions allowed to proceed for the same duration that were not chased (**Figure 8b**, lane 1), suggesting that the chase was at least in part addressing some of the confounding attributes of the initial hot CTP only trial assay.

Given this promising development, experiments were conducted with a variety of factors implicated in pausing and pause release. Due to the robust signal achieved with this assay, changes were observed in any given experiment but did not replicate well enough to draw meaningful quantitative conclusions. The slightest pipetting error in the addition of the pulse or chase could result in drastically different results. While this issue could be handled through practice, repetition, and slight alterations to the experimental design, additional concerns about the dynamic range of the assay pushed further alterations to the pulse-chase regime in an attempt to develop the best possible version of the pausing assay. Furthermore, and perhaps most alarmingly, even after the cold CTP chase, most of the signal came from the paused region and could not be released even with higher cold CTP concentrations.

Establishment of the pulse-chase pausing assay. Though several factors were unsuccessfully tested to adjust the ratio of short to elongated transcripts in the initial pulse-chase system as well as varying the concentration of cold CTP in the chase, one aspect of the assay had not been altered: the duration of the initiation pulse relative to the chase. We theorized that the long duration of the pulse, in which pol II experienced CTP starvation, might place the polymerase in an irreversibly trapped position[104–106]; by shortening the duration of the pulse, it seemed possible that labeling of short transcripts

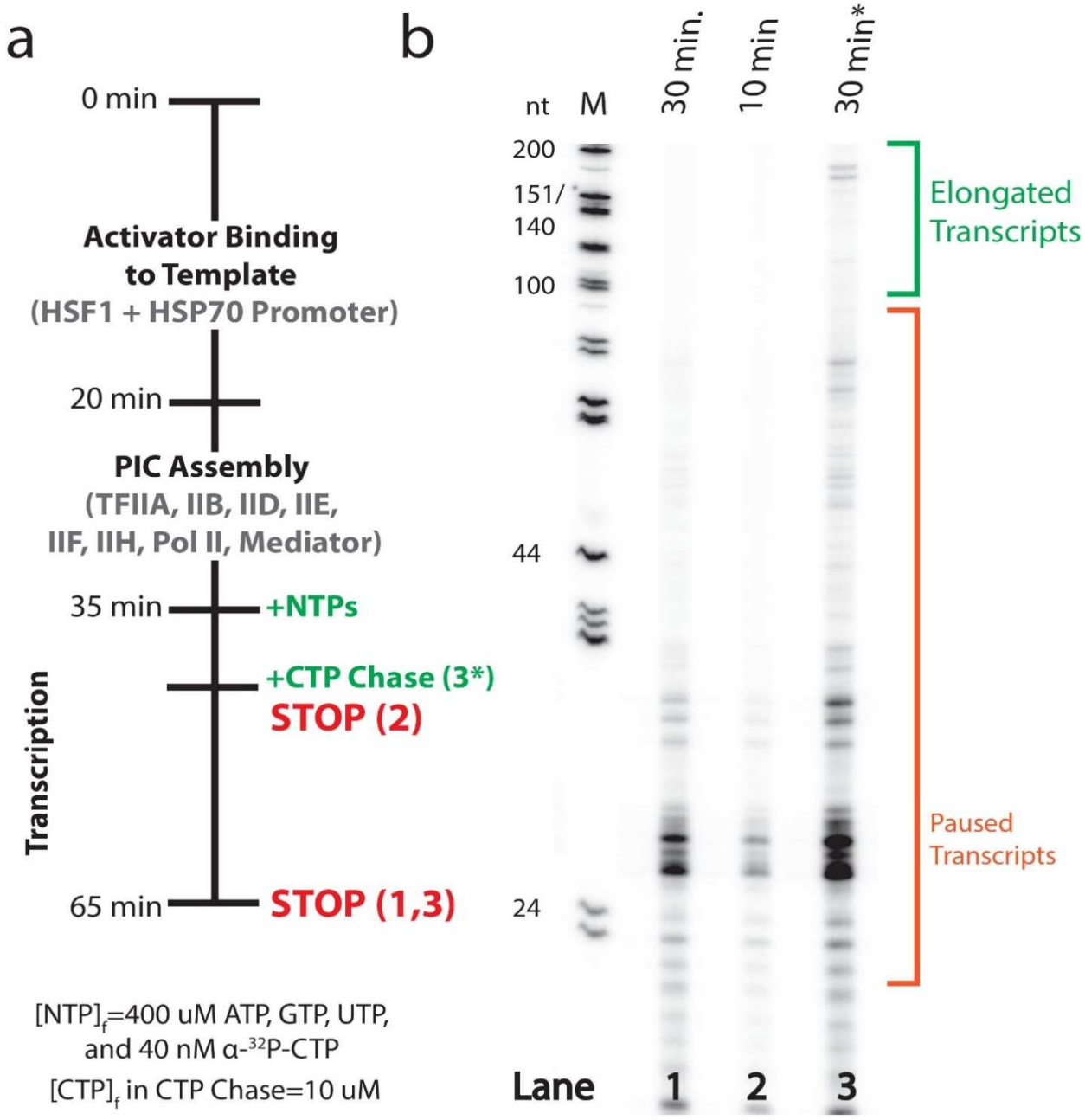


Figure 8: Development of a pulse-chase regime to detect short transcripts. a. Schematic of transcription assay showing time when transcription assay is stopped or chased then stopped. Lane 1 in **b** elongates for 30 minutes under low CTP conditions; lane 2 in **b** is allowed to proceed for 10 minutes under low CTP conditions; lane 3 in **b** is allowed to elongated for 10 minutes under low CTP conditions and then is chased with additional cold CTP for 20 additional minutes before stopping. **b.** 18% urea denaturing sequencing gel showing results of transcription assay outlined in a. Short, perhaps paused products are observed, Runoff transcripts are only observed under pulse-chase conditions in lane 3. Elongated transcripts are described as being between 100 and 250 nucleotides long. Paused transcripts are described as being between 20 and 100 nucleotides long.

could be achieved without "stalling out" pol II. Overall signal would likely be less, but confidence would increase that the short transcripts visualized were due to pol II pausing and not stalling/arrest.

With these considerations in mind, the duration of the pulse was restricted to one minute followed by a chase lasting nine additional minutes (**Figure 9a**). These conditions yielded vastly more reassuring results. A time course demonstrated that one minute after the chase, short paused transcripts peaked in intensity and then were released into longer products indicative of elongation (**Figure 9c**). Furthermore, this short pulse chase regime allows approximation of single-round transcription, with only a 6% increase in signal observed between the five-minute time point and the 10-minute time point. Control experiments confirmed that transcripts detected were driven by the HSP70 promoter (e.g. not any contaminating nucleic acid) and that transcription was dependent on added PIC factors, as expected (**Figure 9b**). In summary, this "final" version of the pausing assay allowed for both visualization of short, potentially paused products as well as elongation. With this strategy, I could confidently begin to assess the activity of pause and pause release factors.

D. DISCUSSION

In this chapter, the development of a fully reconstituted pausing assay on a native human promoter was described. The system developed is biologically relevant: it responds to the addition of a transcriptional activator (HSF1) and contains the complete PIC, including TFIID, Mediator, TFIIH and TFIIE—GTFs often "engineered around" in the development of reconstituted systems. While the achievement in developing this system is

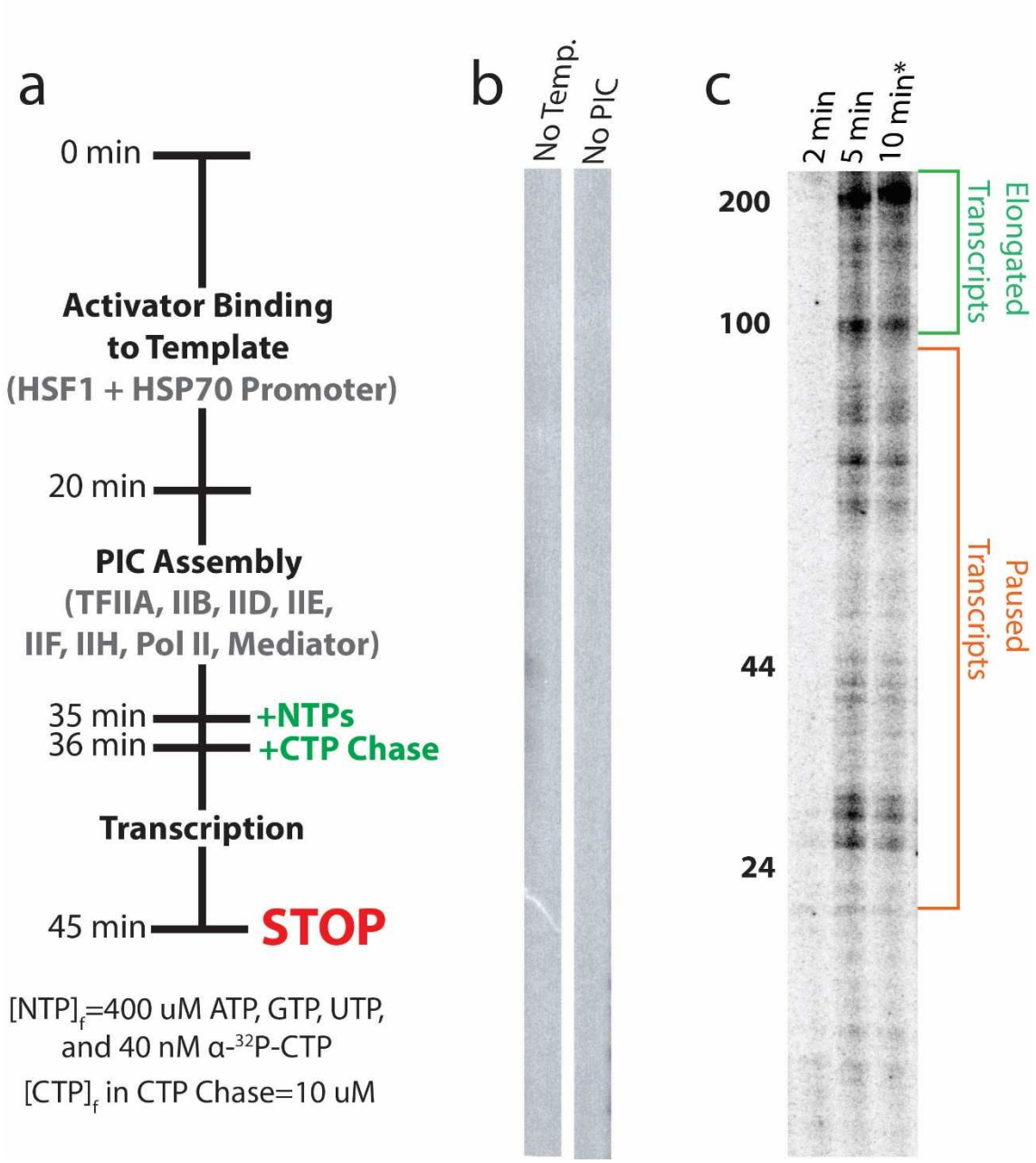


Figure 9: **Development of the pol II pausing assay. a.** Schematic of the pol II pausing assay. Transcription is initiated with a low CTP NTP pulse and allowed to proceed for one minute. Reactions are then chased with additional cold CTP, and transcription is allowed to proceed for an additional nine minutes, bringing the total elongation time 10 minutes. **b.** Controls show no transcription in the absence of the PIC or the template, as expected. **c.** Time course based off of assay described in a. The variation is that transcripts are stopped one minute after the pulse, four minutes after the pulse, or nine minutes after the pulse, bringing the total elongation time to two minutes, five minutes, or 10 minutes, respectively. Paused transcripts are most abundant five minutes after initiation and then release into elongation, which peaks at 10 minutes. Transcription is approximately single round (6% increase in signal between five-minute time point and 10-minute time point), though stimulated by addition of 0.02% sarkosyl. A 10-fold lower concentration than what is typically used to ensure single round transcription.

novel and represents a small step forward in the biochemical study of early steps in transcription, what makes this system worth developing is that it can now be used to ask fundamental questions about initiation, promoter escape, pol II pausing, and early elongation.

The work in this chapter sets-up the experiments mentioned in Chapters III and IV. Chapter III will expand on a fascinating discovery that pol II pausing does not require DSIF and NELF, but that the GTF TFIID establishes pol II pausing. These *in vitro* data were further tested in cells using innovative knockdown methods and advanced RNA-sequencing tools. Chapter IV will delve into transcriptional regulation by the cyclin-dependent kinases (CDKs) TFIIH, P-TEFb, and the CDK8 module. It will show preliminary data that is inspired by existing exploratory cell-based studies and may encourage future rounds of cell-based studies with a narrower focus. Chapter IV shows that the best biochemistry is both inspired by and inspires *in vivo* work.

Fundamentally, Chapter II sets the stage for future work described within this thesis and for future investigators. With this fully-defined, fully-controlled system—and some creativity—the number of questions and that can be asked about early steps of transcription is both motivating and intimidating. It's like a starter-LEGO kit: interesting to manipulate and play with, but with the addition of more and more pieces it can become unwieldly and. All the better to start putting the pieces together now…

Please see Appendices I-II for **Methods.**

Chapter III: TFIID Enables Pol II Promoter-Proximal Pausing

This chapter is derived from a manuscript that has been submitted for publication.

Fant, CB[1]; Levandowski, CB[1]; Gupta, K[2]; Maas, Z[1]; Moir, J[1]; Rubin JD[1]; Sawyer, A[3]; Esbin, MN[1]; Rimel, JK[1]; Marr, MT[3]; Berger, I[2]; Dowell, RD[4,5]; Taatjes, DJ[1]

[1]Dept. of Biochemistry, University of Colorado, Boulder, CO, USA
[2]School of Biochemistry, Bristol Research Centre for Synthetic Biology, University of Bristol, UK
[3]Dept. of Biology, Brandeis University, Waltham, MA, USA
[4]Dept. of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO, USA [5]BioFrontiers Institute, University of Colorado, Boulder, CO, USA

A. INTRODUCTION

After initiation, pol II enzymes typically pause after transcribing less than 100 bases [107], representing a common regulatory intermediate. Accordingly, paused pol II has been implicated in enhancer function [93,108], development and homeostasis [109,110] and diseases ranging from cancer [111,112] to viral pathogenesis [113,114]. Precisely how pol II promoter-proximal pausing is established, enforced, and regulated remains unclear. Previous work shows that protein complexes such as NELF and DSIF increase pausing whereas the activity of P-TEFb (CDK9 and CCNT1) correlates with pause release[107]. To address specific mechanistic questions about pol II pausing and its regulation, we reconstituted human pol II promoter-proximal pausing *in vitro*, entirely with purified factors. As expected, NELF and DSIF increased pol II pausing *in vitro*, whereas P-TEFb promoted pause release. Unexpectedly, the PIC alone was sufficient to reconstitute pol II pausing, suggesting that pausing is an inherent property of the PIC. In agreement, pol II pausing was lost upon replacement of the TFIID complex with TATA-binding protein (TBP); moreover, pausing was dependent upon TFIID subunits TAF1 and TAF2. TAF1/2 bind genomic DNA

downstream of the pol II initiation site[12,13], invoking a "complex interaction" model for pausing[83]. Consistent with this model, PRO-Seq experiments revealed increased transcription upon acute depletion (t=60 min) of TAF1 and TAF2 in human cells, and pol II pausing was disrupted at thousands of genes. Similar results were obtained in TAF1-depleted *Drosophila* S2 cells. Collectively, these data establish the general transcription factor TFIID as a genome-wide regulator of pol II promoter-proximal pausing.

B. RECONSTITUTION OF PAUSING

We sought to reconstitute promoter-proximal pausing entirely from purified human factors (no extracts). Past results in *Drosophila* and mammalian cells and extracts implicated the NELF, DSIF, and P-TEFb complexes as regulators of pol II pausing[88,103,115]. We purified these factors (**Figure 10a**) in addition to the GTFs discussed and shown in Chapter II. Experiments were completed with the native human HSP70 promoter (HSPA1B gene) with the HSF1 activator. Because chromatin *per se* does not appear to be an essential regulator of pol II pausing in *Drosophila* or mammalian cells[78,83,88,89,116], the *in vitro* transcription assays were completed on naked DNA templates.

Following PIC assembly, transcription was initiated by adding ATP, GTP, and UTP at physiologically relevant concentrations, and a low concentration of CTP, primarily ^{32}P-CTP as described in Chapter II. After one minute, reactions were chased with a physiologically relevant concentration of cold CTP and transcription proceeded for an additional nine minutes. By directly labeling all transcripts with ^{32}P-CTP, the method is highly sensitive and allowed detection of transcripts of varied lengths; furthermore, the ^{32}P-CTP "pulse-chase"

protocol ensured that ^{32}P-labeled transcripts resulted almost exclusively from single-round transcription (as described in Chapter II).

A variety of methods have established that pol II pauses after transcribing 20-100 bases in *Drosophila* and mammalian cells[83,85,88,92,107,116–119]. The HSPA1B promoter sequence used in our assays extended 216 base pairs beyond the transcription start site TSS); thus, elongated transcripts would migrate on a sequencing gel between 100 and 216 nucleotides and paused transcripts would be observed between 20 and 100 nucleotides.

Prior to testing DSIF/NELF and P-TEFb, we completed experiments with the PIC alone (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, Mediator, pol II). As expected, elongated transcripts were prevalent; however, we observed short transcripts, between 20 and 100 nucleotides, consistent with promoter-proximal pol II pausing (**Figure 10b**, lane 2, **Figure** quantified in **10c**). Potentially, these short transcripts could represent active pol II complexes transcribing through the promoter-proximal region rather than paused complexes. However, time course experiments showed that these shorter transcripts build up and then release over time (as shown in Chapter II), suggesting a transient pause followed by release into elongation.

Canonical pausing factors perform as expected. Addition of NELF/DSIF to the reconstituted transcription system increased the levels of the short transcripts (20-100 nucleotides) while decreasing the elongated products (**Figure 10b**, lane 3); these data were consistent with established roles for NELF/DSIF in pol II pausing[107] and further suggested that the short transcripts represented promoter-proximal paused products. Addition of P-TEFb to reactions containing NELF/DSIF largely reversed the promoter-proximal pausing induced by NELF/DSIF (**Figure 10b**, lane 4); thus, P-TEFb appeared to

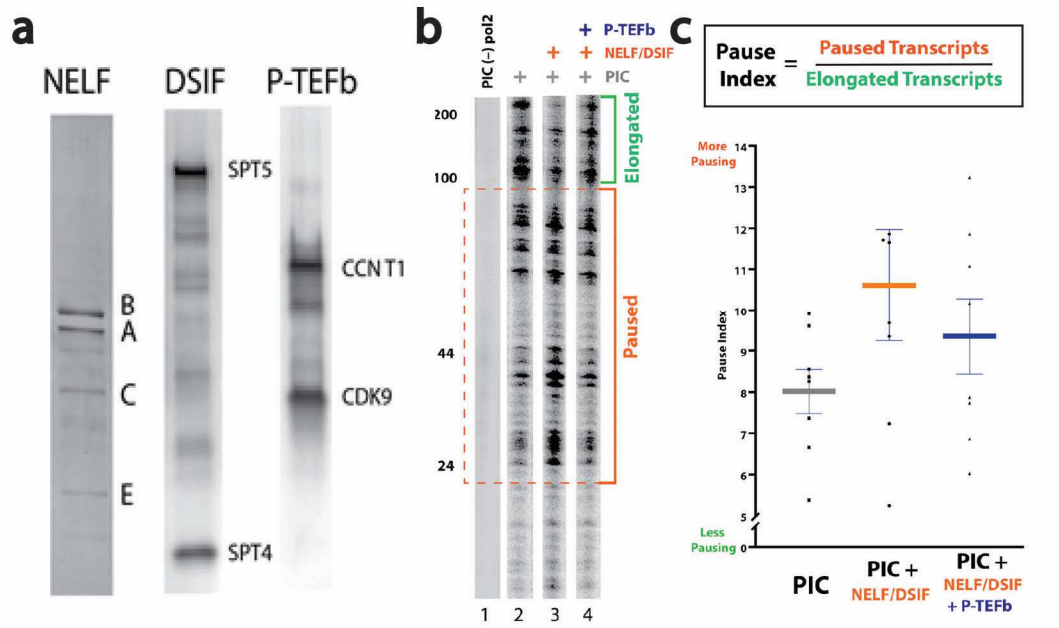


Figure 10: Biochemical reconstitution of promoter-proximal pol II pausing with purified human factors. a. Gels showing purified pausing factors. **b.** Representative data from *in vitro* transcription reactions with the complete PIC (TFIIA, IIB, IID, IIE, IIF, IIH, Mediator, pol II; lane 2) or supplemented with NELF/DSIF (lane 3) and P-TEFb (lane 4). A "no pol II" control experiment is shown in lane 1. At left are approximate lengths (in bases) of the RNA transcripts, with paused and elongated transcript regions highlighted in orange or green, respectively. Note greater numbers of transcripts in paused region coupled with reduced transcripts in elongation region upon addition of NELF and DSIF (lane 2 vs. lane 3) (p=0.03); addition of P-TEFb reverses this trend, though not significantly (lane 3 vs. lane 4) **c.** Calculation of an *in vitro* pause index (PI) at the HSPA1B promoter (n = 8). As expected, PI increased upon addition of NELF and DSIF and P-TEFb alleviated NELF/DSIF-induced pausing. Bars represent mean ± standard error.

increase pol II pause release *in vitro*, also consistent with current models[1]. A pause index (PI) was calculated and averaged across replicate experiments (n=8; **Figure 10c**).

Pausing is enabled by TFIID. Because we were able to recapitulate pause enhancement with NELF/DSIF and pause release with P-TEFb at the native human HSPA1B promoter, this *in vitro* system appeared to reliably reconstitute basic mechanistic aspects of pol II promoter-proximal pausing. Whereas many potential questions could be addressed with this system, we focused on the unexpected result that promoter-proximal pausing was recapitulated with the PIC alone. We next tested whether pol II pausing would be dependent on a specific PIC factor. Although some factors could not be reliably

evaluated given their requirement for transcription in this assay, removal of TFIIA, TFIIH, HSF1, or Mediator still supported transcription *in vitro*, although at reduced levels.

We also addressed a potential dependence on the large, multi-subunit TFIID complex. Whereas pol II transcription was not supported by removal of TFIID, TBP can substitute for TFIID *in vitro*, provided that the DNA templates are not assembled into chromatin[120]. Strikingly, we observed that when PICs were assembled with TBP instead of TFIID, promoter-proximal pol II pausing was lost (**Figure 11a**). In particular, transcription initiation and elongation were still supported with TBP, but levels of transcripts in the promoter-proximal region were markedly reduced. These data implicated TFIID as a key PIC factor that enables pol II promoter-proximal pausing. To test further, we replaced endogenous purified human TFIID with a complete TFIID complex generated by recombinant expression (**Figure 11b**). As shown in **Figure 11c**, the recombinant human TFIID complex performed similarly to endogenous TFIID, confirming that TFIID was required for pol II promoter-proximal pausing *in vitro*.

TFIID lobe C impacts pausing. Having established a TFIID dependence for pol II pausing, we sought to determine whether this activity could be attributed to any specific TFIID subunits. Human TFIID is approximately 1.4 MDa in size and contains TBP plus 13 different TBP-associated factors (TAFs), which are present in one or two copies each. The structures of human TFIID bound to promoter DNA reveal that lobe C—containing TAF1, TAF2, and TAF7—binds downstream DNA[12,13]. In particular, TAF1/2 interact with the Downstream Promoter Element (DPE) and the Motif Ten Element (MTE; **Figure 11d**). At the HSPA1B promoter, these elements reside at template position +18 to +33 relative to the TSS[11]. Because the DPE and MTE encompass part of the pol II pause region, we

hypothesized that lobe C subunits might be important in regulation of pol II promoter-proximal pausing. To test this hypothesis, we expressed and purified TFIID complexes that contained only a subset of TAFs (**Figure 11E**). As shown in **Figure 11F**, the 7-TAF complex was less capable of supporting pol II pausing, whereas partial TFIID complexes that contained TAF1 (S-TAF) or TAF2 (8-TAF) increased pol II pausing. The 7-TAF and 8-TAF complexes lack TBP, which was added separately (**Figure 12**).

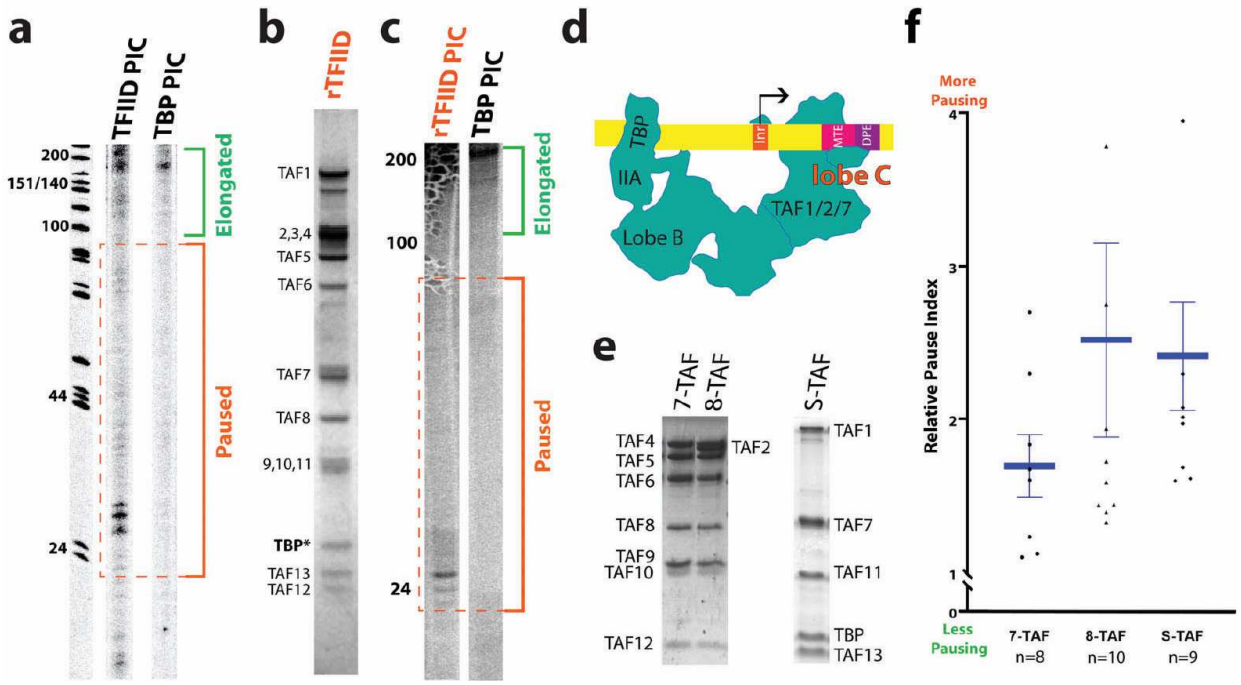


Figure 11: TFIID is required to establish pol II promoter-proximal pausing *in vitro*. a, Reconstituted transcription reactions with PICs containing TFIID or TBP (i.e. reactions contain TFIIA, IIB, IIE, IIF, IIH, Mediator, and pol II, plus either TFIID or TBP). Note that PICs with TBP still support elongated transcription, but paused products are absent. **b**, Coomassie-stained gel of the complete human TFIID complex, generated by recombinant expression[37]; *core TBP (residues 155-335). **c**, As with endogenously purified human TFIID, PICs with recombinant TFIID support transcription and pol II promoter-proximal pausing. Elongated products are obscured in rTFIID lane because of a cracked gel. **d**, Schematic of human TFIID structure bound to promoter DNA[10,11]. **e**, Coomassie-stained gels of partial TFIID complexes lacking TAF1 and TAF2 (7-TAF) or containing TAF2 (8-TAF) or TAF1 (S-TAF). **f**, Plot of PI comparing *in vitro* transcription experiments with PICs containing partial TAF complexes. Note that PI increases in presence of TAF1 or TAF2; because 7-TAF and 8-TAF lack TBP, experiments with these complexes contained added TBP.

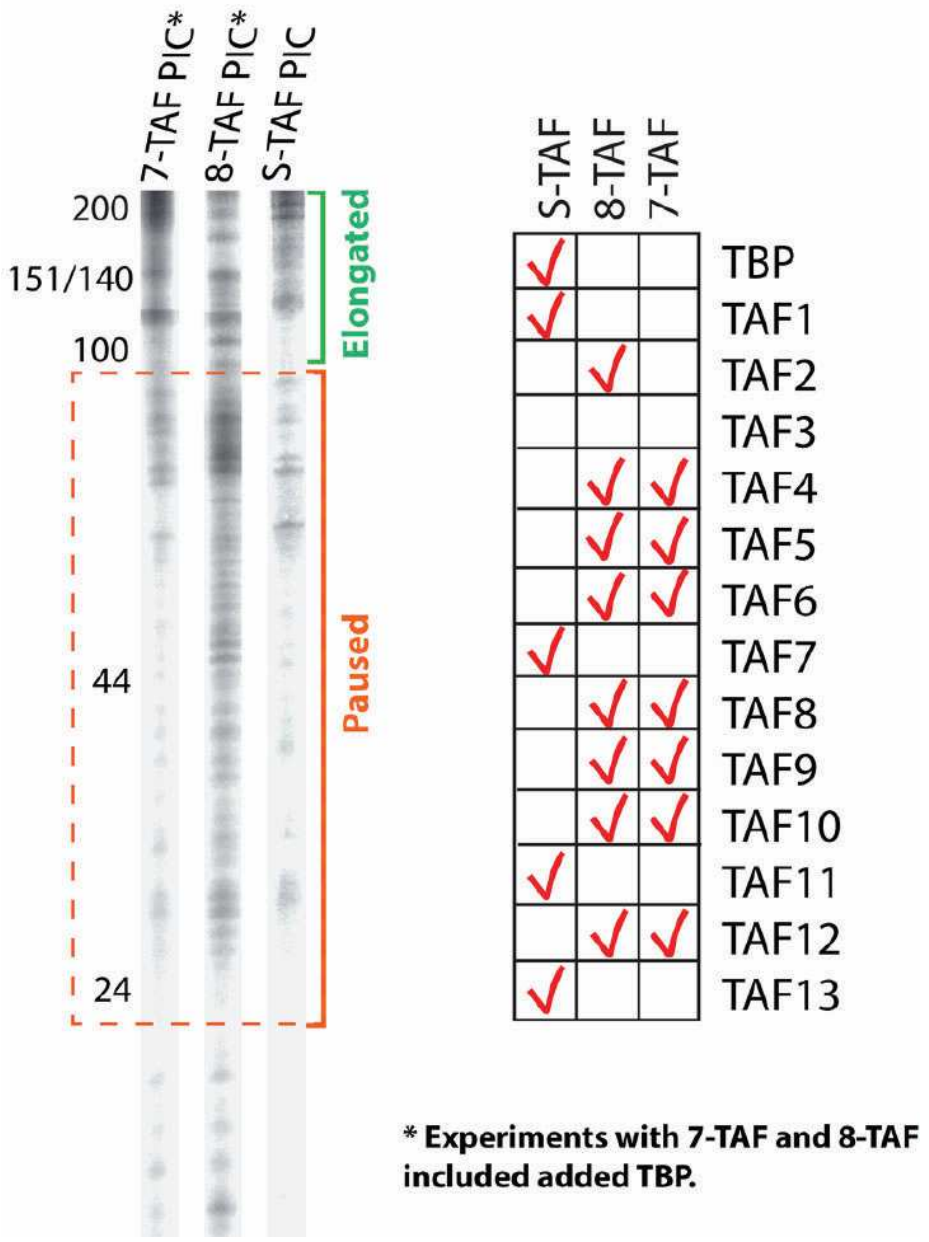


Figure 12: Partial TFIID complexes can support transcription and vary in ability to support pol II promoter-proximal pausing. (left) Representative in vitro transcription data (all data quantified in Figure 2F) for PICs assembled with 7-TAF, 8-TAF, or S-TAF complexes instead of TFIID. Note that free TBP was added to 7-TAF and 8-TAF experiments because these complexes lack TBP. A summary of the composition of each complex is shown at right.

C. CELL-BASED INVESTIGATIONS OF TFIID-DEPENDENT POL II PAUSING

Knockdown of TFIID subunits in human HCT116 cells. To further test the hypothesis that TFIID enables pol II promoter-proximal pausing, we turned to cell-based assays. To circumvent confounding issues with prolonged knockdown of essential TFIID subunits, we utilized the Trim-Away method[121], which enabled rapid (t=60 min) TAF subunit depletion (**Figure 13a** and **13b**, **Figure 14**). With this approach, the effect of TFIID could be evaluated with minimal compensatory and/or cytotoxic consequences. TAF1 and TAF2, together with TAF7, comprise TFIID lobe C[12,13]. Indicative of a direct TAF1-TAF2 interaction in lobe C, Trim-Away experiments targeting TAF1 also depleted TAF2 (TAF7

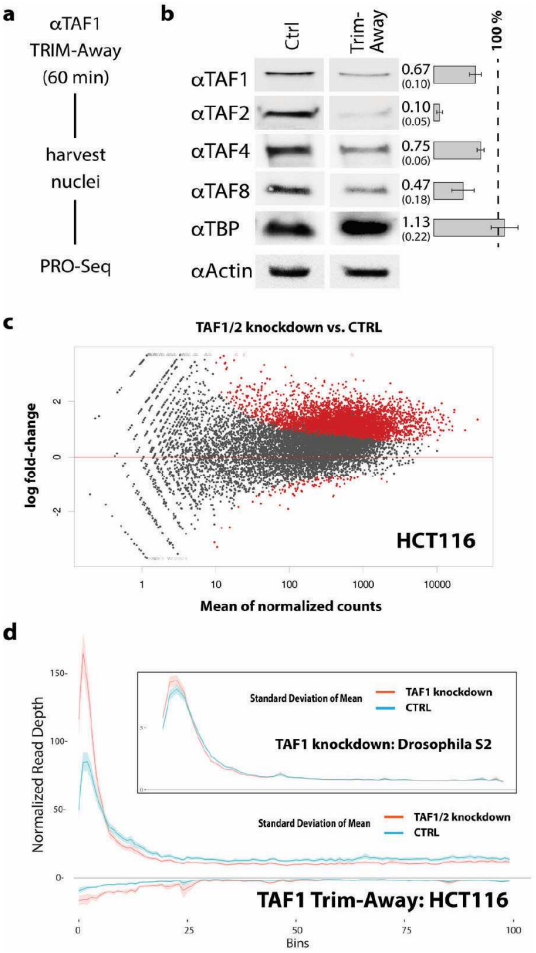


Figure 13: Loss of TAF1 and/or TAF2 disrupts transcription in human and *Drosophila* cells, especially at promoter-proximal pause regions. a, Workflow for TRIM-Away and PRO-Seq. **b,** Representative western blots and quantitation (at right) for TFIID subunits. Bar plots represent mean and standard error, with actin as a loading control. **c**, MA plot showing a genome-wide up-regulation of transcription in TAF1 TRIM-Away cells compared with controls (TSS region -500 to +500). As shown in the metagene plots (**d**), much of this increased transcription was localized to gene 5'-ends, which coincides with the pol II pause region. Metagene generated from top 500 expressed genes. Similar results were observed in TAF1 knockdown *Drosophila* S2 cells vs. controls (inset; n = 10995).

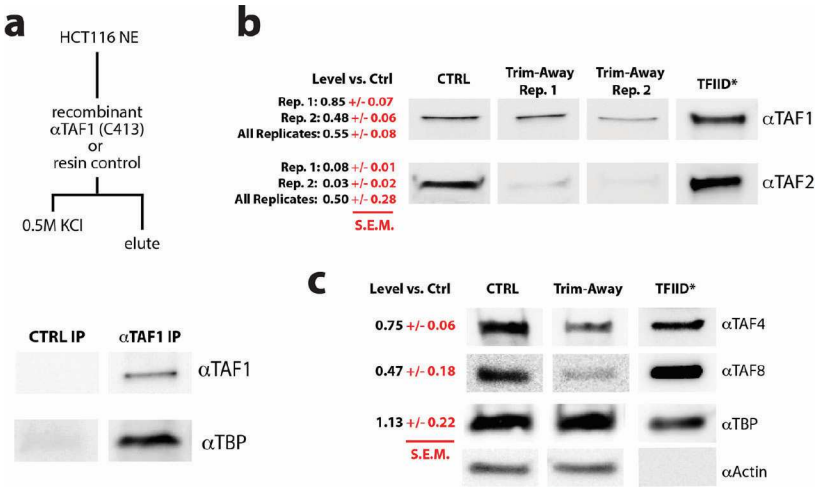
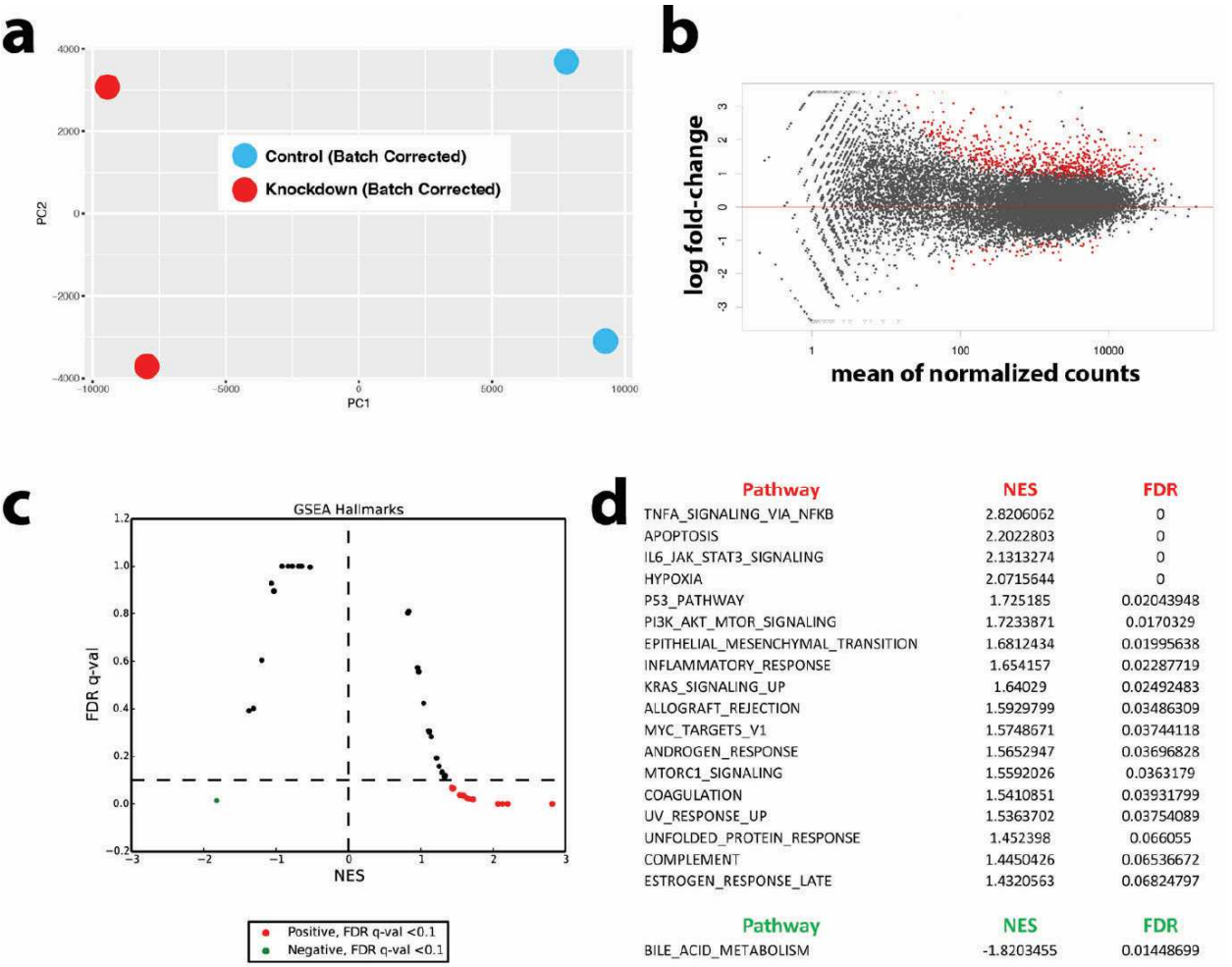


Figure 14: Additional control experiments for TAF1 TRIM-Away in HCT116 cells. a. Human TAF1 antibody validation. TAF1 antibodies were immobilized onto protein A resin, incubated with HCT116 nuclear extract, washed extensively with 0.5M KCl buffer, and eluted. Eluted material was probed for TFIID subunits by western. **b**. Quantitative westerns to compare control vs. TAF1 TRIM-Away experiments. Data for biological replicate 1 & 2 represent samples used for PRO-Seq experiments (2 technical replicates for each biological replicate). Additional TRIM-Away experiments were completed to estimate reproducibility of knockdown and to assess effects on other TFIID subunits (see panel **c**). For TAF1 "all replicates" includes 4 biological and 13 technical replicates; TAF2 = 4 biological and 8 technical replicates. Numbers in red represent the standard error of the mean. **c**. Quantitative westerns to probe TAF1 TRIM-Away effects on other TFIID subunits. Data for TAF4 represent 2 biological and 4 technical replicates; TAF8 = 1 biological and 3 technical; TBP: 2 biological and 5 technical. Numbers is red represent the standard error of the mean. *Purified human TFIID was included as a positive control for antibody specificity.

was not probed due to lack of reliable antibodies), and other TFIID subunits were depleted to varying degrees, except for TBP (**Figure 13b and Figure 14c**).

Following acute TAF depletion using Trim-Away, we isolated nuclei and performed replicate PRO-seq experiments (TAF1/2 knockdown vs. controls; **Figure 15**). An expectation based upon our *in vitro* results (**Figure 11a** and **Figure 11c**) and cryo-EM structural data[12,13] was that TFIID might serve as a "brake" for promoter-associated pol II complexes and that removal of this brake would enhance pause release and increase transcription. This expectation was largely confirmed by the PRO-Seq data. Transcription increased



d

Pathway	NES	FDR
TNFA_SIGNALING_VIA_NFKB	2.8206062	0
APOPTOSIS	2.2022803	0
IL6_JAK_STAT3_SIGNALING	2.1313274	0
HYPOXIA	2.0715644	0
P53_PATHWAY	1.725185	0.02043948
PI3K_AKT_MTOR_SIGNALING	1.7233871	0.0170329
EPITHELIAL_MESENCHYMAL_TRANSITION	1.6812434	0.01995638
INFLAMMATORY_RESPONSE	1.654157	0.02287719
KRAS_SIGNALING_UP	1.64029	0.02492483
ALLOGRAFT_REJECTION	1.5929799	0.03486309
MYC_TARGETS_V1	1.5748671	0.03744118
ANDROGEN_RESPONSE	1.5652947	0.03696828
MTORC1_SIGNALING	1.5592026	0.0363179
COAGULATION	1.5410851	0.03931799
UV_RESPONSE_UP	1.5363702	0.03754089
UNFOLDED_PROTEIN_RESPONSE	1.452398	0.066055
COMPLEMENT	1.4450426	0.06536672
ESTROGEN_RESPONSE_LATE	1.4320563	0.06824797

Pathway	NES	FDR
BILE_ACID_METABOLISM	-1.8203455	0.01448699

Figure 15: Details from PRO-Seq experiments in HCT116 cells. a, PCA from each replicate (control TRIM-Away vs. TAF1 TRIM-Away knockdown) shows controls cluster separately from TAF1 knockdown samples. **b**, Gene body MA plot (TSS to polyA site) showing up-regulation of transcription in TAF1 TRIM-Away cells. **c**, Moustache plot of false discovery rate (FDR-q) vs. normalized enrichment score (NES) of hallmark gene sets analyzed by GSEA. Differences (increase = red; decrease = green) are highlighted in TAF1-knockdown cells vs. controls. **d**, Listing of Hallmark gene sets highlighted in **b**; up-regulated hallmarks suggest onset of stress responses in TAF1/2-knockdown cells.

genome-wide upon TAF1/2 knockdown, especially at gene 5'-ends (**Figure 16c** and **Figure 16d**; **Figure 17a**). Analysis of individual genes reflected these genome-wide trends (Figure 9a; **Figure 17b**), as did an assessment of pause index in TAF1/2-depleted vs. control cells (**Figure 18b**). These data are consistent with increased pol II release from promoter-proximal pause regions. Others have shown that pol II pause release allows additional pol II enzymes to re-initiate transcription[48,122]; in agreement, increased re-initiation was observed in TAF1/2-depleted cells, which resulted in more gene 5'-end reads and an increased pause index (**Figure 18b**). The increased pause index was not promoter element-dependent (TATA and/or Inr, **Figure 17c** and **Figure 17d**).

Knockdown of TAF1 in Drosophila S2 cells. TAF1 knockdown in *Drosophila* S2 cells has minimal impact on other TFIID subunits[123,124]. To further test the impact of TAF1 on pol II pausing, TAF1 was knocked down in S2 cells and PRO-Seq experiments were completed in triplicate (**Figure 17b**). Consistent with TAF1/2-depleted human cells, TAF1 knockdown in *Drosophila* S2 cells showed a characteristic promoter-proximal increase in transcription, genome-wide (**Figure 16d**, inset; **Figure 17c**). Furthermore, pausing was similarly disrupted in TAF1-knockdown S2 cells, compared with TAF1 Trim-Away HCT116 cells (**Figure 16c**) and PI was not promoter-element dependent (**Figure 17c-h**). These data suggest a conserved role for TFIID in the regulation of pol II promoter-proximal pausing. It is noted that the Drosophila S2 data shows less of a stark contrast between the control and knockdown conditions. In Drosophila S2 cells, other TFIID subunits (specifically TAFs 2, 4, and 8) have previously been shown to be unaffected by TAF1 knockdown[124]. These findings differ from that seen in human TAF1 knockdown in human HCT116 cells (**Figure 13b, Figure 14c**).

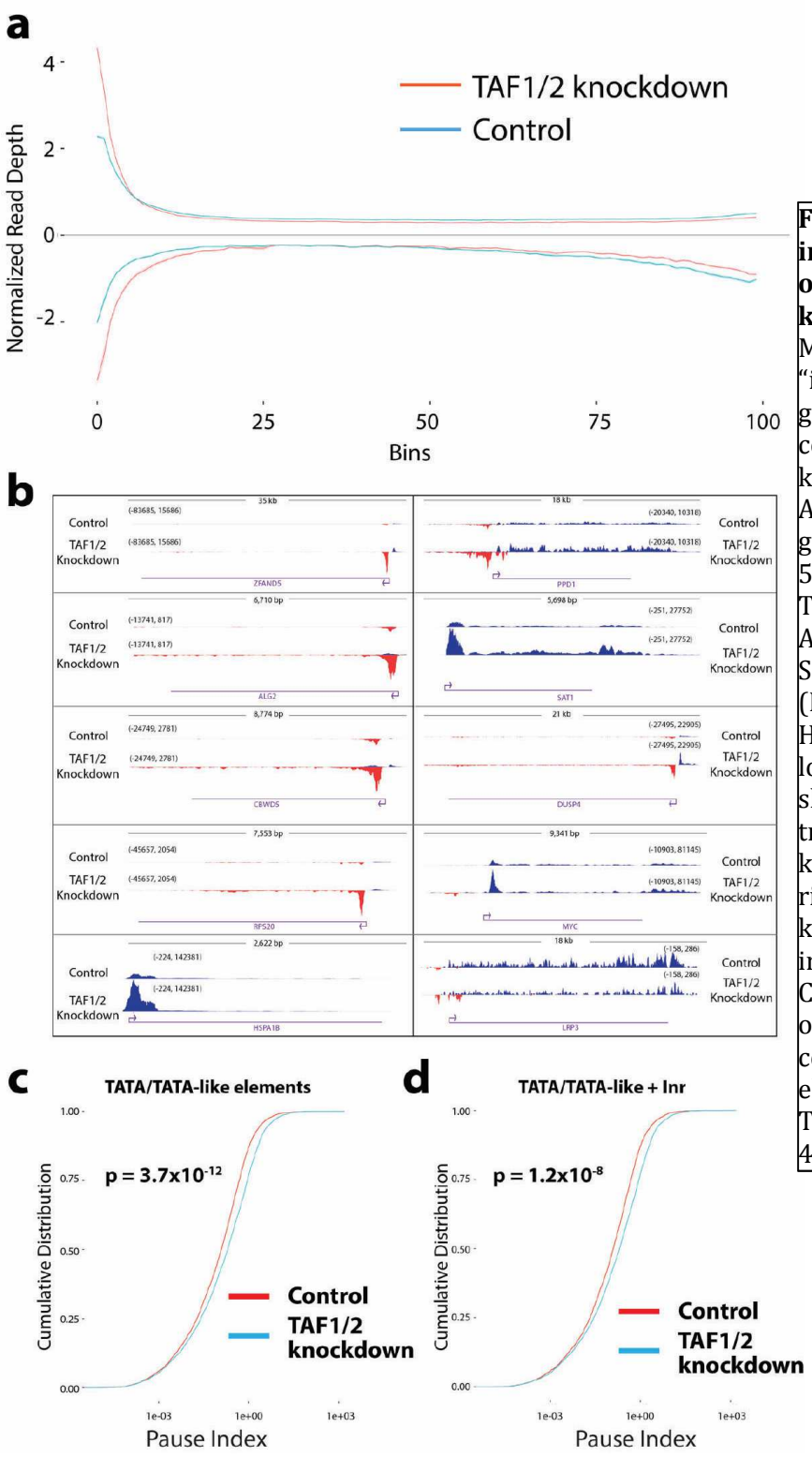


Figure 16: PRO-Seq data indicate global disruption of pol II pausing in TAF1/2-knockdown cells. a. Metagene plot for "intermediate" expressed genes (n = 16908) in HCT116 cells, comparing TAF1/2 knockdown with control cells. As with top 500 expressed genes (**Fig. 3d**), reads at gene 5'-ends are increased with TAF1/2 knockdown. **b.** Additional examples of PRO-Seq data from HCT116 cells (IGV traces). Note that the HSPA1B locus is shown at lower left and not all genes show evidence of increased transcription upon TAF1/2 knockdown (LRP3, lower right). **c/d.** TAF1/2 knockdown increases pause index (PI) in HCT116 cells. Cumulative distribution plots of PI are shown for all genes containing TATA or TATA-like elements (n = 5892; **c**) or TATA + Inr elements (n = 4159; **d**).

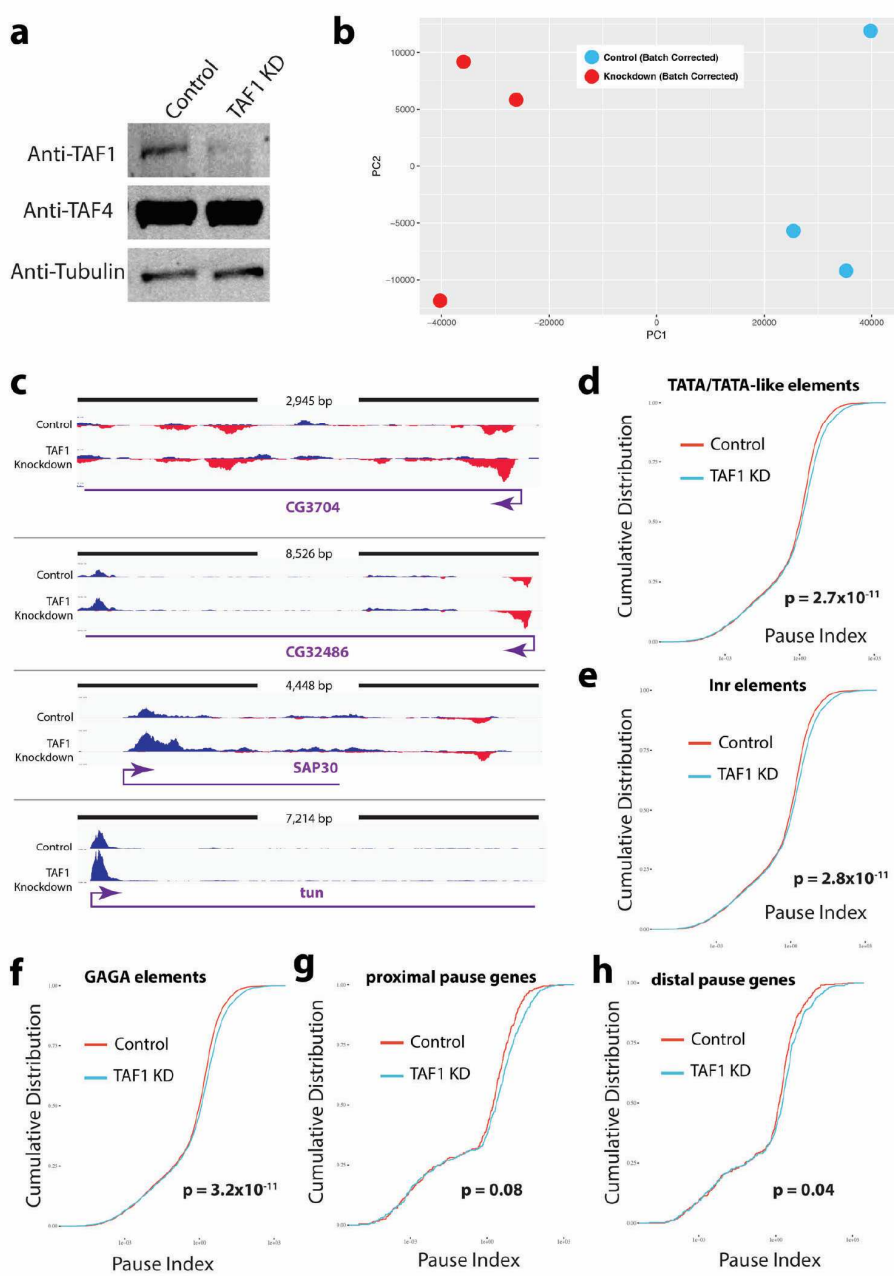


Figure 17: PRO-Seq data implicate global disruption of pol II pausing in TAF1 knockdown *Drosophila* S2 cells. a. Representative western blots showing TAF1 knockdown in S2 cells. TAF1 knockdown was determined to be 88% (± 6.7) from 3 biological replicates. **b**. PCA from each replicate (control RNAi vs. TAF1 knockdown) shows controls cluster separately from TAF1 knockdown samples. **c**. Example PRO-Seq data from S2 cells (IGV traces) showing results similar to TAF1/2-knockdown HCT116 cells. Increased transcription is observed specifically at the promoter-proximal pause region (gene 5'-ends). **d-h**. TAF1 knockdown increases pause index (PI) at all classes of genes in S2 cells. Cumulative distribution plots of PI are shown for all genes containing TATA or TATA-like elements (**d**, n = 4566), Inr elements (**e**, n = 4545), GAGA elements (**f**, n = 4513), and at genes with proximal (**g**, n = 395) or distal (**h**, n = 411) pause sites, as defined previously[44].

These differences may be attributed to the knockdown method or to differences in TFIID architecture between the two species.

D. DISCUSSION

Pol II pausing has been studied extensively in cells and *in vitro* through the use of nuclear extracts. However, because of the number of factors that must be purified (both GTFs and pausing factors), reconstitution of pol II pausing has been unsuccessful or not attempted. The *in vitro* work shown in this chapter represents the first successful attempt at fully reconstituting pol II pausing, providing a steppingstone to better understand mechanisms of pausing.

The canonical pausing factors (NELF and DSIF) increase pausing as has been previously reported in the literature[63,78,88]. P-TEFb releases NELF and DSIF dependent pausing, also in accordance with published work[93,107,125]. These findings show that the reconstituted pausing assay we have developed can reconstitute well-reported effects and increase our confidence that the assay can be reliably used to measure the effects of a wide variety of factors on early steps of transcription, particularly pol II pausing and release.

We were surprised to observe pol II pausing in the absence of NELF and DSIF. Because of the defined nature of the reconstituted pausing assay, we were able to deduce that pausing must be an inherent property of the PIC or early transcribing pol II complexes. Past studies have suggested a role for TFIID in pol II pausing in *Drosophila* based upon promoter mutagenesis[83] and correlations among paused genes and DNA sequence elements bound by TFIID[88,90,118,122,124]. Specifically, it has previously been shown that

increased spacing between the Inr element and the downstream promoter elements results in decreased pausing[83] and that TFIID lobe C binds these downstream elements[12,13]. Since we have a level of mechanistic control that is not possible with cell-based assays or in vitro systems using nuclear extracts, we were able to determine that TFIID is indeed a key PIC factor that establishes pausing by replacing the multi-subunit TFIID complex with the single subunit required for transcription on a linear template, TBP (**Figure 11** and **Figure 18d**). Our results suggest a modest revision to the canonical model of pol II pausing: TFIID lobe C first facilitates pol II pausing through its lobe C subunits. NELF and DSIF can then act to stabilize TFIID-enabled pol II pausing, and P-TEFb may then release the paused complex.

Upon TAF depletion in cells, either by Trim-Away in HCT116 cells or by shRNA in *Drosophila S2* cells, we observe pause disruption genome-wide. This result validates our *in vitro* data that show that TFIID is essential for pause regulation. However, rather than observing less transcription at the 5' ends of genes as we might have expected based on the *in vitro* data, promoter-proximal transcripts increase upon TAF depletion. While on its surface this result seems directly counter to the result obtained *in vitro*, it makes biological sense when two considerations are made: first, the reconstituted pausing assay is essentially a single-round transcription assay (Chapter II), whereas PRO-seq detects transcription at multiple stages and picks up re-initiation events. Essentially, re-initiation cannot be observed in the reconstituted pausing assay, but it can be detected using PRO-seq. Secondly, previous work has shown that paused pol II prevents re-initiation[125]. In TAF-depleted cells, we hypothesize that re-initiation, not pol II pausing, accounts for the increase in short transcripts that map to the 5' ends of genes and increases the "pausing index". Indeed, this data further highlights the effects of pause regulation on transcription

regulation as a whole: by dysregulating pol II pausing, transcription initiation is also dysregulated.

While TFIID is not known to interact directly with pol II during early steps of transcription, TFIID possesses multiple domains that bind specific chromatin marks, including H3K4me3[126,127]. Promoter recruitment of TFIID, and therefore pol II pausing, may thus be regulated in part through epigenetic mechanisms. The role of the interactions between TFIID and downstream promoter elements is also ripe for interrogation as a potential mechanism through which TFIID may enforce pausing. Additionally, in the paused region, the nascent RNA is long enough to protrude from the pol II RNA exit channel and potentially interact with TFIID, perhaps slowing or halting the polymerase in a manner that would allow for pause stabilization by DSIF and NELF. While mechanisms for how TFIID enables pol II pausing remain hypothetical, further work with the reconstituted pausing assay may provide additional mechanistic insight.

Defects in TFIID function are linked to numerous diseases, including cancer[128] and neurodegenerative disorders[129]. Its requirement for pol II promoter-proximal pause regulation may underlie these and other biological functions.

.

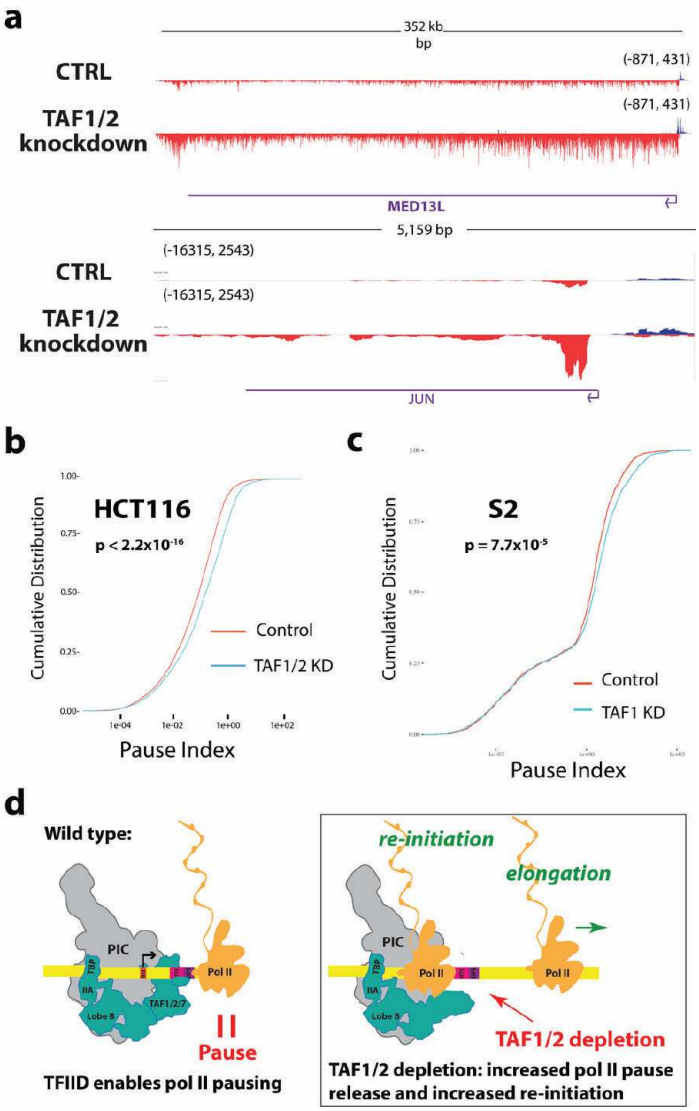


Figure 18: Pol II promoter-proximal pausing is disrupted genome-wide upon TAF knockdown in human or Drosophila cells. a, Example PRO-Seq data from HCT116 cells (IGV traces). **b**, TAF knockdown increases PI in human (HCT116) and **c**, *Drosophila* (S2) cells. Cumulative distribution plots of PI are shown for all "expressed" genes (HCT116; n=5303) as defined and all "paused" genes (S2 cells; n=3225), as defined[12]. **d**, Model. TFIID is required to establish pol II promoter-proximal pausing; disruption of TFIID correlates with increased pol II release from promoter-proximal pause regions. Release of paused pol II, in turn, enables additional pol II complexes to re-initiate transcription[27,28]. Collectively, this causes increased transcription around the pause region (i.e. at gene 5'-ends). Pol II pausing may be especially dependent on TFIID lobe C subunits TAF1 and TAF2 because they bind Inr and downstream promoter elements; however, the precise molecular mechanism remains unclear. Although pause release and increased 5'-end reads were evident at thousands of genes in HCT116 cells, most transcripts abruptly decreased after a few hundred bases (e.g. +300), suggesting additional regulatory mechanisms downstream of the pause site.

Please see Appendices I-III for **Methods**

Chapter IV: Regulation of Early Steps in Transcription by the Kinases P-TEFb, the CDK8 Module, and TFIIH

A. INTRODUCTION

The transcriptional kinases function as part of large, multi-subunit machines. P-TEFb (CDK9 and CCNT1) is thought to regulate pause release through phosphorylation of pausing factors and the pol II CTD as part of the Super Elongation Complex (SEC). The Mediator kinase module (herein called the CDK8 module) is composed of CDK8, CCNC, MED12, and MED13 and may function independently or as part of the larger Mediator complex. Final, TFIIH is a 10-subunit complex with CDK7 and CCNH composing the CDK-cyclin pair. Two other CDKs—CDK12 and CDK13—also play roles in transcription regulation, though at later steps in transcription than are the focus of this thesis.

P-TEFb. P-TEFb is best understood for its role in pause release[115,130]. P-TEFb is thought to phosphorylate NELF, thus evicting it from the paused pol II complex, and phosphorylate the SPT5 subunit of DSIF[87], thereby converting it into an elongation competent form. It then phosphorylates Ser2 of the pol II CTD heptad repeat, promoting the association of elongation factors such as SPT6[49]. In many ways, P-TEFb can be thought of as a paused pol II complex remodeler: it exchanges NELF for the pre-elongation factor PAF1[49,131], reconfigures DSIF so that it promotes elongation, and alters the landscape of the pol II CTD so that it can accommodate SPT6. While it remains unclear if P-TEFb directly phosphorylates each of these factors, it is clear that the kinase activity of P-TEFb is an important regulator of pol II elongation.

It comes as no surprise that previous work in *Drosophila* and human cells have shown that kinase inhibition of CDK9 results in increased pol II pausing genome wide. However, because these experiments were initially done using the CDK9 inhibitor flavopiridol[86,117], the results are difficult to interpret due to inhibitor promiscuity: flavopiridol also inhibits the kinase activities of CDK7, CDK2, CDK4, CDK7 at similar concentrations to that needed to inhibit CDK9 and reportedly CDK12 and CDK13 to a lesser extent[132–134]. Thus, many of the existing conclusions drawn about the role of P-TEFb in pause release and elongation must be carefully evaluated. It should no longer be assumed that increased pausing associated with cell treatment with flavopiridol is exclusively due to inactivation of CDK9. Recent work has begun to more stringently assess the kinase activities of CDK9 by creation of analogue sensitive CDK9 cell lines ($CDK9^{as}$). These cell lines contain a mutation that increases access to the active site of the kinase by a bulky, covalent ATP-analogue but does not affect kinase activity until the ATP-analogue is added[135]. When the analogue is added to these cells, CDK9 is covalently inhibited, with minimal off target effects. In Raji B human lymphocyte cells, CDK9 inhibition by this method demonstrated a clear role for P-TEFb in releasing the pause, thus releasing a block on re-initiation[48]. In budding yeast, pol II could still escape the pause site under CDK9 inhibition, but elongated at a slower rate[136]. Together, these data suggest that P-TEFb indeed regulates pol II promoter-proximal pause release, and by extension, re-initiation. However, prior to the data shown in Chapter III, mechanistic *in vitro* data supporting this cell-based observation has not been collected. I will show in this chapter that while P-TEFb does indeed release pol II from a NELF and DSIF stabilized pause, it also releases pol II

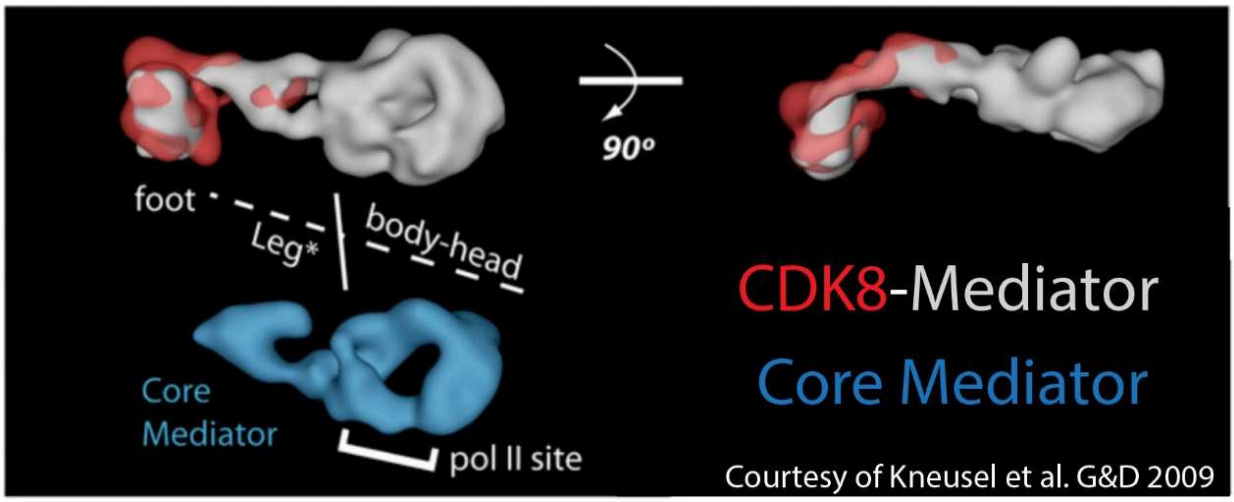


Figure 19: **The CDK8 module binding to Mediator occludes pol II binding to Mediator[59].** Cryo-EM structures of the CDK8-Mediator complex show that binding of the CDK8 module to Mediator results in an allosteric shift that disrupts the pol II docking site. The CDK8 module, shown in red, interfaces extensively with the foot and leg domains of Mediator through the MED13 module subunit and yet undetermined Mediator subunit(s).

from a weaker pause established by TFIID, indicating an expanding role for P-TEFb in pause release and elongation.

The CDK8 module. By comparison to P-TEFb, little is known about the roles of the CDK8 module in early steps of transcription, and specifically in pol II pausing. The four-subunit complex associates reversibly with the 26-subunit core Mediator complex, an essential complex for the expression of all protein-coding and most non-coding genes. It has been established that the CDK8 module and pol II cannot both associate with core Mediator at the same time (**Figure 19**)[59,60]. Previous *in vitro* work shows that promoter bound CDK8-Mediator precludes pol II association with the PIC, inhibiting transcription initiation[59,137]. However, if the CDK8 module is added after transcription initiation, levels of transcription are unaffected. These data suggest that timing of CDK8 module association with the PIC is important.

Cell-based studies show that the kinase module plays central roles in facilitating stimulus-specific changes in gene expression[71,74,138]. Phosphorylation of gene-specific transcription factors by the CDK8 module likely promotes activation and repression of genes as part of the stimulus response in both development and stress [71,73,138–140]. Furthermore, it has been speculated that enhancer bound CDK8 module may facilitate the formation of liquid-liquid phase separated (LLPS) domains that support transcriptional bursting[75].

How the CDK8 module regulates transcription (activation or repression) may be dependent on the time at which it associates with the PIC. When the CDK8 module is associated with the PIC prior to transcription initiation, it represses transcription. However, if it associates with the PIC after initiation, it may increase transcription. This chapter will show that addition of the CDK8 module after transcription initiation in the reconstituted pausing assay activates promoter escape, underscoring the importance of timing in CDK8 recruitment to the PIC.

TFIIH. The 10-subunit TFIIH complex can be parsed out into two distinct subcomplexes: the CAK (CDK-activating complex, composed of CDK7, CCNH, and MAT1) and the core[45]. The core contains two helicases—XPB and XPD. XPB is essential for opening promoter DNA[141]. Recent structural data and genome wide kinase data have pushed the CAK complex, and in particular CDK7, into the spotlight. It has been shown in human cells that CDK7 phosphorylates the CTD at Ser5 and Ser7 of the pol II CTD heptad repeat[95], thus driving pol II promoter escape through a proposed mechanism involving electrostatic exile of the CTD from Mediator associated with the PIC[137]; cryo-EM data of the TFIIH complex in the context of the PIC (comprised of both human[29,142] and yeast[143] proteins separately) has

supported this hypothesis. CDK7 phosphorylation of the CTD also seems to regulate the transcription elongation rate, subsequently altering co-transcriptional processing behavior[64,95], and potentially termination[66]. However, the most controversial and interesting role for CDK7 in transcription concerns its potential functions in regulating pol II pausing and release. In unpublished work from the Taatjes lab in collaboration with other labs, SILAC-mass spectrometry results using a derivative of THZ1, a highly specific covalent CDK7 inhibitor[144], highlight DSIF and NELF amongst potential kinase targets in human cells—targets that have long been assumed to be within the province of P-TEFb (CDK9) phospho-regulation. It is possible that redundancy in kinases capable of phosphorylating these targets exist, emphasizing both the essential nature of pol II pausing and the robustness by which it is regulated. Alternatively, it is plausible that the transcriptional kinases regulate each other, and that these regulatory mechanisms would be perturbed under kinase inhibition conditions of any of the CDKs involved in the regulatory hierarchy.

CDK inter-regulation. Each of the three transcriptional kinases—CDK7 (TFIIH), CDK8 (CDK8 module), and CDK9 (P-TEFb)—regulate steps of early transcription. Broadly speaking, TFIIH and the CDK8 module regulate initiation and promoter escape, and TFIIH and P-TEFb regulate pausing and elongation. These statements vastly oversimplify the roles of these kinases, particularly because these kinases likely regulate each other and thus impact transcription generally. For instance, previous work suggests that the CDK8 module inactivates TFIIH through phosphorylation of its cyclin subunit CCNH at Ser5 and Ser304[145]. Additionally, the CDK8 module has been suggested to activate P-TEFb[71]. As highlighted above, CDK7 phosphorylates Thr186 on the T-loop of P-TEFb[66], thereby

activating it and promoting directly or indirectly phosphorylation of the canonical pausing factors DSIF and NELF and promoting elongation. However, phosphorylation of this site can also promote P-TEFb binding to the inhibitory 7SK snRNP complex[146,147], rendering it inactive. Taken together, these data highlight the need for careful *in vitro* analyses coupled with next-generation cell-based studies to parse out the roles of the transcriptional CDKs in early steps of transcription.

The reconstituted pausing assay provides a unique means to understand the contributions of each kinase in varying stages of transcription. It is nearly impossible to study the effects of CDK7, CDK8, and CDK9 in cells due to off target effects and the inability to discern direct kinase targets from indirect targets. The fully defined nature of the *in vitro* reconstituted pausing assay provides the ability to reliably and systematically assess the roles of each kinase. CDK7 can be inhibited with THZ1, CDK8 can be inhibited with Cortistatin A (CA), and P-TEFb can be inhibited with CDK9-IN-2 without concerns about off-target inhibition of one of the other kinases[44,72,144,148,149]. Additionally, CDK7 and CDK8 can be inhibited through mutations that render the kinase inactive[139,150,151]. The various strategies for kinase inhibition make robust study of the kinase activities of P-TEFb, the CDK8 module, and TFIIH relatively simple *in vitro*.

Work described in this chapter represents preliminary *in vitro* data that attempts to describe the functions of both the kinase-dependent and independent activities of TFIIH, the CDK8 module, and P-TEFb. This work is performed using the reconstituted pausing assay described in detail in Chapter II. Some cell-based studies that complement this work are ongoing by other members of the Taatjes lab and its collaborators. These *in vitro* data begin to take advantage of novel and specific kinase inhibition methods that have been

shown to have *in vivo* relevance. While these experiments are ongoing, the data provided herein provides encouragement for further studying the roles of these kinases *in vitro*, including the study of interdependent kinase regulation.

B. RESULTS

P-TEFb promotes pause release. Chapter III presented data that supported P-TEFb dependent pause release of NELF and DSIF stabilized pausing. This data validated the existing canonical pausing and pause release model developed from previous cell-based studies and *in vitro* work completed with nuclear extracts[76]. Surprisingly, in control experiments in which P-TEFb was added in the absence of NELF and DSIF, P-TEFb also facilitated increased pause release/elongation (**Figure 20**). Quantification of several identical experiments is also shown in **Figure 20** and shows a significant decrease in the pause index upon P-TEFb addition. Given the defined nature of the *in vitro* pausing assay, these data suggest that P-TEFb can promote pause release in the absence of NELF and DSIF, possibly supporting a broader, non-canonical role for P-TEFb in pause release.

The CDK8 module increases promoter escape. Previous work in the Taatjes lab and others has generated hypotheses about the role of the CDK8 module in transcription[152,153]. The CDK8 module inhibits pol II initiation, according to published *in vitro* work, but seems to have no effect if added post-initiation[59]. However, in cells, inhibition or depletion of he CDK8 kinase results in transcriptional defects under stimulus response (stress or development)[71,138,139]. In examining existing *in vitro* work, it is clear that the model proposed by the authors in which CDK8 module association with Mediator inhibited pol II incorporation into the PIC (**Figure 19**) was highly dependent on the timing

of CDK8 module addition[59]. In these experiments, when the CDK8 module was added along with the other GTFs, transcription was repressed. This timing may have allowed CDK8 module to GTFs, transcription was repressed. This timing may have allowed CDK8 module to compete with pol II for association with Mediator and incorporation into the PIC perhaps

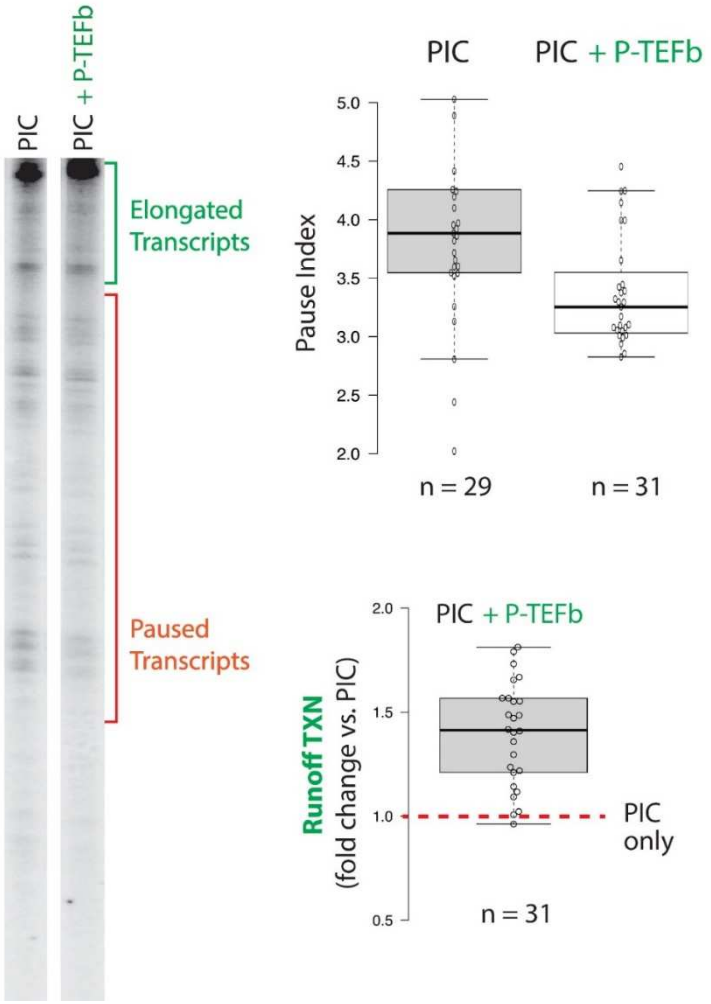


Figure 20: P-TEFb promotes pol II pause release/elongation in the absence of NELF and DSIF. Using the reconstituted pol II pausing assay discussed in Chapters II and III, P-TEFb appears to alleviate the TFIID-enabled pause. Representative data in shown in the left-most part of the panel followed by aggregate data represented through box plots on the right. Pause indices are calculated as described in Chapter III and are a ratio of paused to elongated transcripts; Runoff TXN Is the intensity of radioactivity of the elongated transcripts (range labeled on gel). The middle 50% of data is defined by the box with the whiskers extending to the upper and lower 25% of the data. The middle bar represents the data median. P=0.0012Box plots are made with the free online tool BoxPlotR: http://shiny.chemgrid.org/boxplotr/.

suggesting a greater affinity of the module for Mediator or that the concentration of the CDK8 module was high enough to compete away pol II. The authors also added the CDK8 module 15 minutes after transcription initiation. When added at this time point the CDK8 module had no effect on transcription, perhaps due to its late addition. Indeed, time courses with the reconstituted pausing assay show pausing (a later stage of transcription) two minutes after initiation (Chapter II). In order to test the potential role of the CDK8 module in promoter escape, the module would need to be added soon enough after initiation to affect this early stage of transcription.

A time point 15 seconds after initiation was chosen for the addition of the CDK8 module. When added 15 seconds after transcription initiation, the CDK8 module supported more transcription than did the PIC alone, suggesting a role for the CDK8 module in promoter escape (**Figure 21a**, lane 1 and lane 2). These data show a proportional increase in both paused and elongated transcripts. Since this increase is proportionate between paused and elongated transcripts, it is unlikely that addition of the CDK8 module is affecting pause release.

Single molecule *in vitro* transcription experiments show low levels of template usage (that is templates onto which a PIC assembled and transcription occurs)—around 10-15%[77]. Addition of the CDK8 module to PICs after the addition of NTPs and corresponding transcription initiation may facilitate increased template usage, an indicator of increased promoter escape.

Furthermore, limited experiments with the highly specific CDK8 kinase inhibitor CA

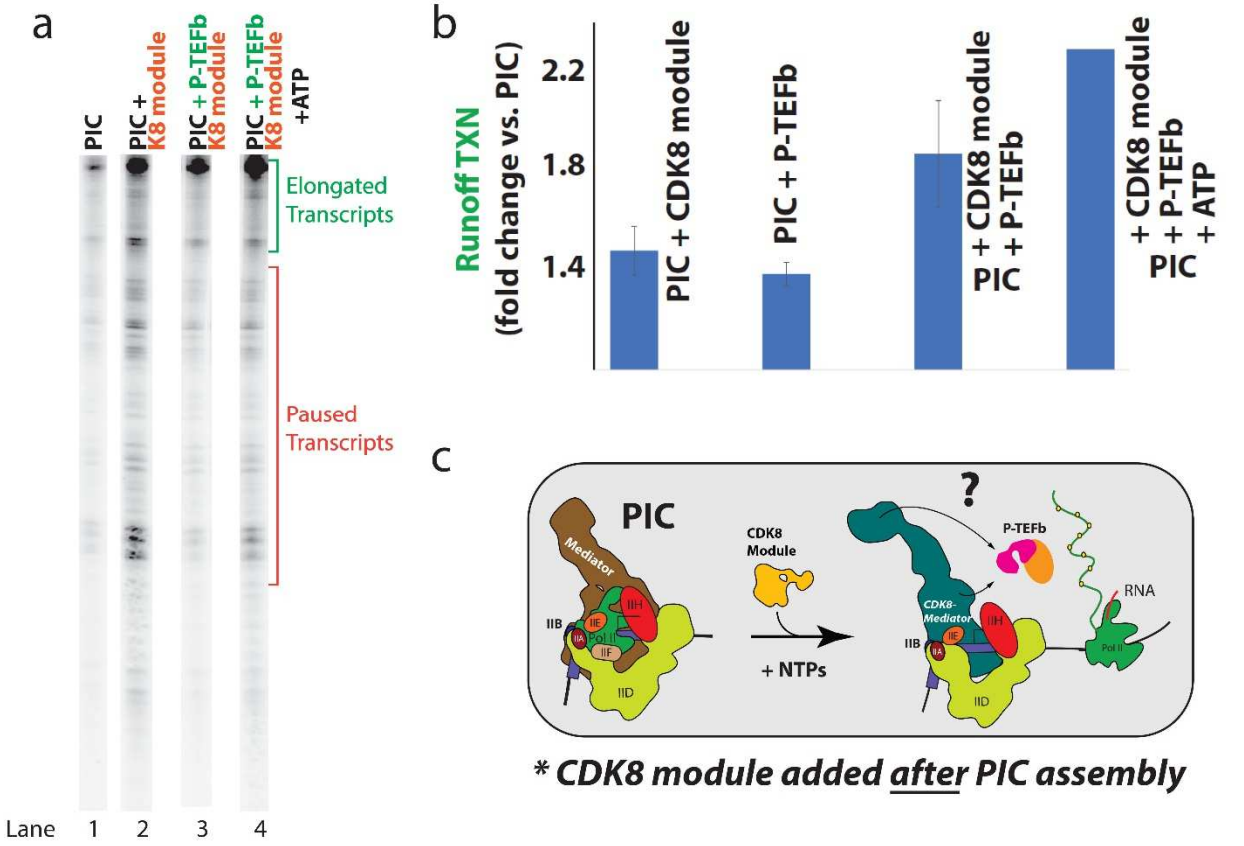


Figure 21: The CDK8 module increases promoter escape and likely works in a kinase dependent manner to facilitate elongation with P-TEFb. a. Representative data from reconstituted pausing assay. Lane 2 shows elevated levels of transcription (both paused and elongated transcripts) when the CDK8 module is added 15 seconds after transcription initiation with NTPs. Lane 3 shows that while elongated transcripts remain constant, paused transcripts visibly decrease when both P-TEFb and the CDK8 module are added post-initiation, suggesting a general mechanism by which the CDK8 module and P-TEFb activate each other. In lane 4, the CDK8 module and P-TEFb are pre-incubated with each other in the presence of ATP before addition to the transcription assay. Increased overall transcription is observed compared to the addition of the two kinases without preincubation in lane 3, suggesting kinase dependent inter-regulation of these transcriptional CDKs. **b.** Quantification of runoff TXN of aggregate data with SEM indicated by error bars for conditions shown in **a. c.** Model demonstrating how the CDK8 module may increase promoter escape and activate P-TEFb to promote pause release. Modified from a previously published Taatjes lab review[33].

(K_d=0.2 nM[148]) were performed at a final concentration of 100 nM (**Figure 22a**, lane 3). These extremely preliminary data show a slight decrease in pause release and elongation that is reflected in ChIP-seq experiments in HCT116 cells after treatment with CA[72]. The *in vitro* transcription data supporting this phenomenon are abnormal (pause signal is uncharacteristic compared with other data) and should be scrutinized and repeated; **Figure 22c** shows quantification of runoff transcripts for this limited data set.

P-TEFb and the CDK8 module increase elongation. Previous work has highlighted the likely indirect interaction between the CDK8 module and P-TEFb, based on co-IP western blots[71]. Additionally, the same studies found that the CDK8 module promoted elongation under oxidative stress[70] and after serum addition[71] under starved conditions. Furthermore, CDK8 depletion severely inhibited the recruitment of both Mediator and P-TEFb to activated genes, thus reducing the overall stimulus response[71].

Data obtained from the *in vitro* pausing assay suggest that, at least under simulated activated conditions (heat shock mimic by addition of HSF1), the CDK8 module and P-TEFb indeed increase transcription elongation as shown by a marked increase in elongated transcripts (**Figure 21a** and **21b**). The CDK8 module was added post-initiation. While this conclusion must be considered appropriately given the limited number of replicates completed, they do reflect the observations made in cells. Furthermore, when the CDK8 module and P-TEFb were pre-incubated with ATP and added post-initiation, elongation was further increased as assayed with the *in vitro* pausing assay (Figure 3a and 3b). Taken together, these data suggest a role for P-TEFb and CDK8 module inter-regulation that is likely kinase dependent (**Figure 21c**).

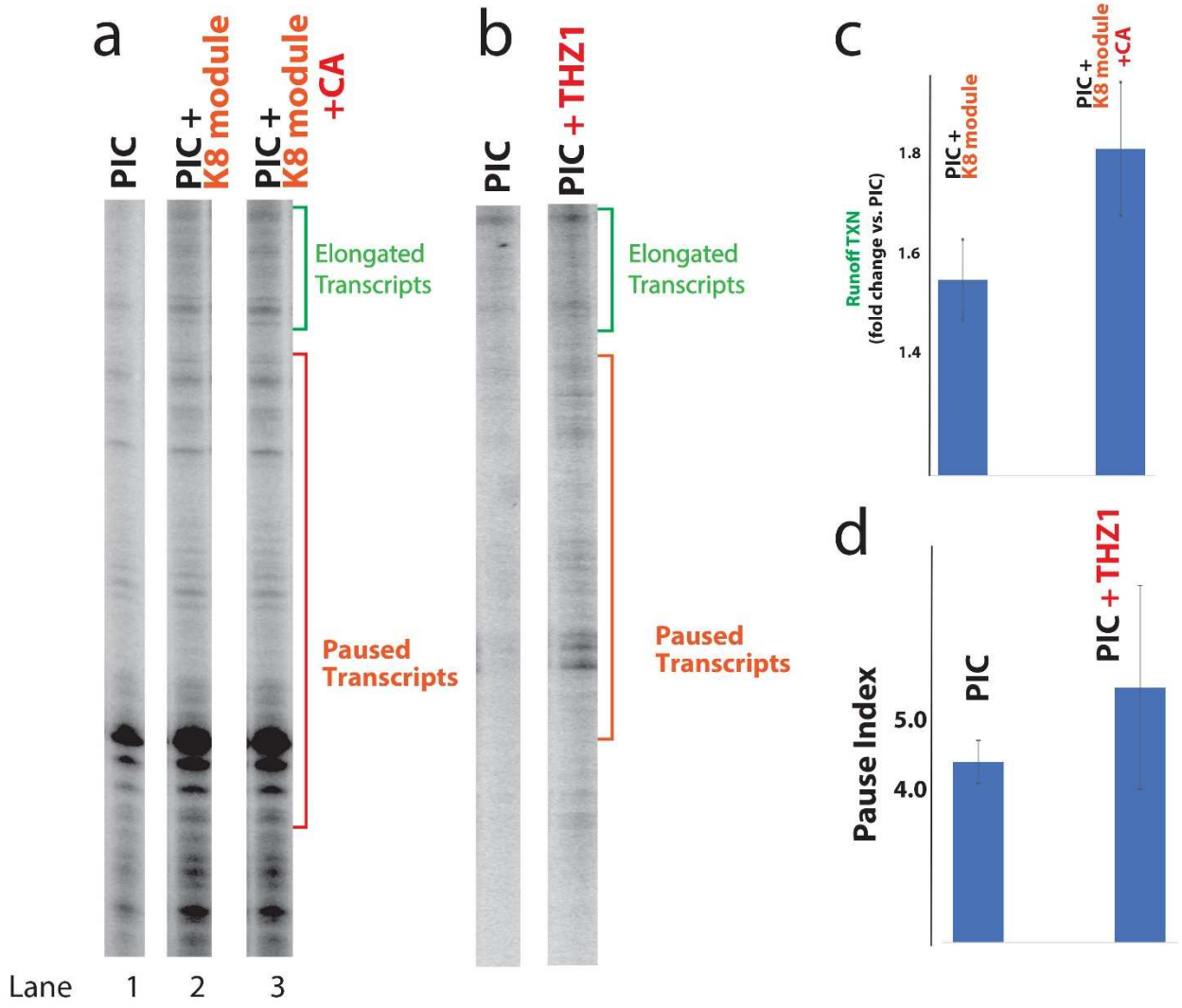


Figure 22: **Preliminary results with small molecule kinase inhibitors. a.** Reconstituted pausing assay results showing perhaps a slight decrease in pause release when the CDK8 module is preincubated with CA prior to addition to the transcription reaction. Transcript patterns are unusual and call into question the validity of these results. **b.** Pre-incubation of TFIIH with THZ1 at 1 uM results in an increase in pausing. **c.** Quantification of runoff transcription of aggregate data represented in **a** and **b**. **d.** Quantification of pause indices calculated from aggregate data represented in **a** and **b**.

TFIIH kinase activity may enhance pause release. Because of the fully annealed DNA template used in the transcription assay, the removal of TFIIH is not feasible. The necessity of TFIIH underscores its essential roles in initiation, but makes distinguishing its activities (kinase activity, helicase activity, translocase activity, structural/scaffolding roles) difficult. However, *in vitro* pausing assay experiments were performed in which the TFIIH kinase CDK7 was inhibited by 1 μM THZ1, a potent, selective inhibitor with an IC_{50}=3.2 nM that has been used in previous in vitro and cell-based studies (**Figure 22b**)[64,144]; quantification over a limited number of replicates is shown in **Figure 22c** and **Figure 22d**. These experiments support a modest increase in pausing upon CDK7 inhibition, agreeing with ChIP-seq results obtained by THZ1 inhibition in HCT116 cells[95]. As with CDK8 module results after treatment with CA, these results were obtained from a very small sample size and require replication and further investigation.

TFIIH kinase inhibition affects P-TEFb and the CDK8 module. TFIIH was pre-incubated with 1 μM THZ1, as was done in the previous section, and then tested in the pausing assay under two additional conditions: with the addition of P-TEFb and P-TEFb plus the CDK8 module pre-incubated with ATP. As usual, any addition of the CDK8 module was 15 seconds post-initiation. Under these conditions, inhibition of the TFIIH kinase decreases P-TEFb pause release (**Figure 23**, lane 3) and disrupts CDK8 module/P-TEFb cooperativity in elongation (**Figure 23**, lane 5). These preliminary data suggest that the TFIIH kinase may serve as a "master regulator" of P-TEFb and the CDK8 module. If the TFIIH kinase is inhibited, the "downstream kinases" may be blocked from functioning. Quantification of these data is shown in **Figure 23b.**

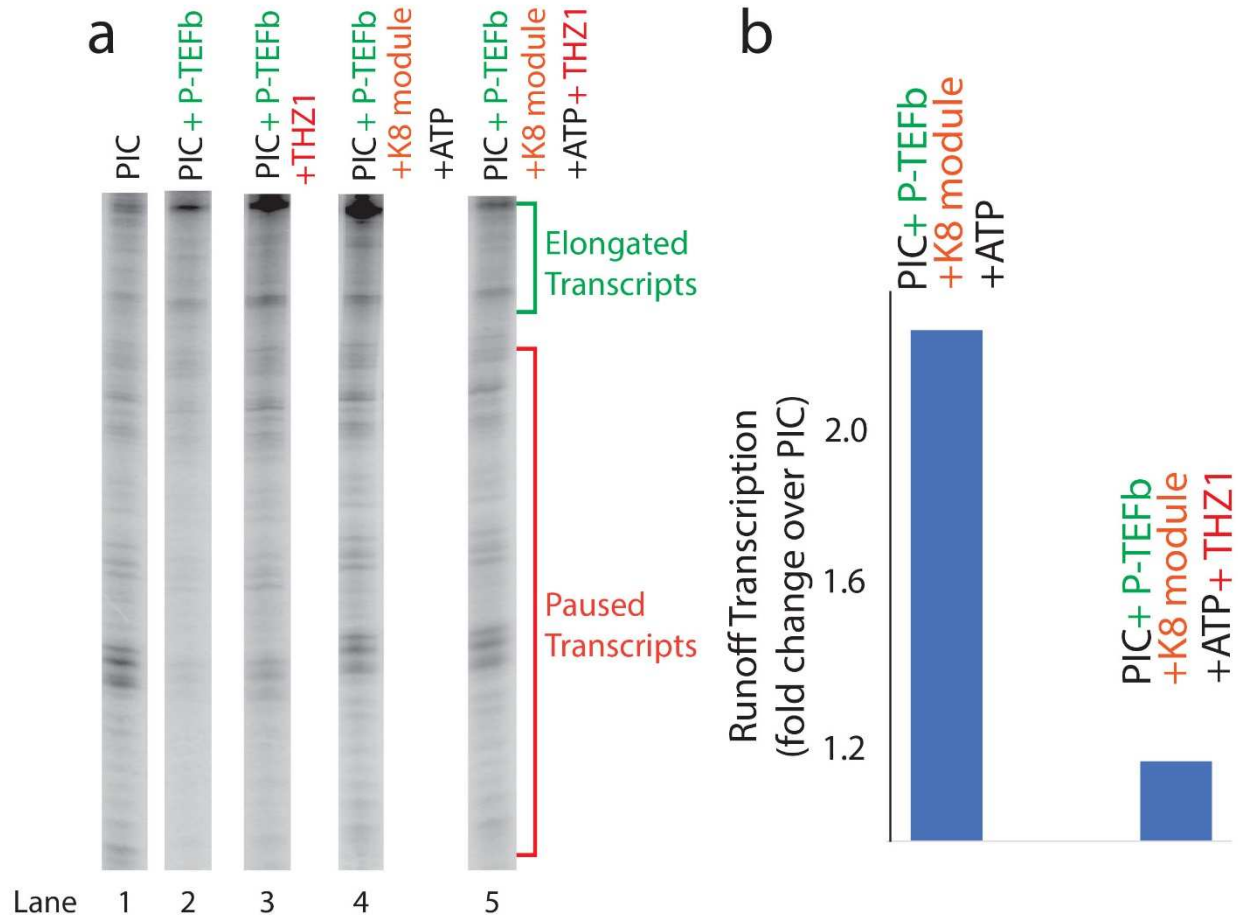


Figure 23: Preliminary results of TFIIH kinase inhibition show potential downstream regulatory effects on the other kinases P-TEFb and the CDK8 module a. When the TFIIH kinase is inhibited, P-TEFb-dependent pause release is decreased (lane 3) compared to uninhibited conditions (lane 2). Cooperativity between the CDK8 module and P-TEFb in promoting elongation is also reduced (lane5) compared to when the TFIIH kinase is uninhibited (lane 4) **b.** Quantification of average of small number of replicates (n=3) showing the cooperative effects of P-TEFb and the CDK8 module in promoting elongation (via a measurement of runoff transcription). Additional quantification (n=2) shows that THZ1in inhibition of the TFIIH kinase decreases that effect.

C. DISCUSSION

The data presented in this chapter establish a basis for better understanding the functions of the transcriptional kinases TFIIH, P-TEFb, and the CDK8 module during early steps of transcription. These data take advantage of the reconstituted pausing assay

developed in Chapter II and first utilized in Chapter III. By building on the strengths of this assay (a fully defined system capable of assessing multiple steps of transcription), this chapter goes further into elucidating both kinase dependent and kinase independent functions of TFIIH, the CDK8 module, and P-TEFb and encourages future *in vitro* work.

P-TEFb was shown to increase pause release in agreement with previously published data[48,78], although it was able to do so in the absence of NELF and DSIF. This finding represents additional insight into the role of P-TEFb as a pause release factor. Because of the defined nature of the reconstituted pausing assay, P-TEFb must be working mechanistically with other components of the fully defined system to increase pause release. Given these restrictions, four potential novel hypotheses for how P-TEFb could increase pause release in the absence of NELF and DSIF need to be tested: 1.) CDK9 phosphorylation of Ser2 (and possibly Ser5[154]) of the pol II CTD is sufficient to facilitate pause release, 2.) CDK9 targets another, yet unidentified, substrate in this defined system (one of the GTFs or the activator HSF1), 3.) P-TEFb plays a binding/scaffolding role that facilitates pause release, or 4.) some combination of these proposed mechanisms. To test these hypotheses, kinase inhibition experiments are a valuable start.

While it is standard to use flavopiridol to inhibit CDK9, flavopiridol promiscuously inhibits several other CDKs, including CDK7 (TFIIH)—a necessary factor in the *in vitro* transcription assay[133]. Next-generation inhibitors exist, including CDK9-IN-2. Such an inhibitor could be pre-incubated with P-TEFb, then inhibited P-TEFb could be added to the transcription assay to measure the effects of the kinase activity on this phenomenon. As an interesting experiment, it would also be insightful to inhibit the P-TEFb kinase in an assay with NELF and DSIF. While our data and that of others support P-TEFb dependent pause

release of NELF and DSIF stabilized pausing, it does not clearly demonstrate the essentiality of the kinase function to elicit these effects.

Kinase inhibition could also be achieved through purification of a CDK9^{as} mutant, similar to that tested in cells[125,135,136]. Additionally, a kinase dead version of P-TEFb (P-TEFb-KD) could be engineered, purified, and assayed. Previous work has shown the efficacy of these chemical and chemical genetic tools in ablating CDK9 kinase activity specifically.

Our data suggest that the CDK8 module seems to increase promoter escape, thus establishing *in vitro* evidence for a potential role for the CDK8 module as a transcriptional activator, as has previously been demonstrated using cell-based approaches[70–72,138]. In this model, CDK8 module association with Mediator is favorable over pol II association given either assay conditions (disparate concentrations) or biophysical properties (CDK8 module-Mediator K_d<pol II-Mediator K_d), and that binding of Mediator to both pol II and the CDK8 module is not possible. These factors indicate a possible tuning mechanism by which the CDK8 module can regulate pol II occupancy at promoters and the subsequent step of promoter escape[155]. Furthermore, these data agree with cell-based studies which show that the CDK8 module enhances transcription of genes associated with super enhancers: bringing the CDK8 module to PICs through enhancer looping would increase pol II promoter escape and consequently drive expression of associated genes (**Figure 24**). The role of the CDK8 module kinase activity in these processes remains elusive, but seems to be a context-specific activity based on cell-based work rather than an activity that regulates transcription generally[72,75,139].

This chapter also shows that the CDK8 module and P-TEFb may work in concert to promote elongation in a manner that is at least in part a function of their kinase activities. This finding has also been supported in cells[71]. These data are preliminary but should encourage more rigorous biochemical analyses in which each kinase is inhibited individually and then tested alongside the other for effects on transcription to provide better mechanistic insight. Furthermore, the cooperativity between the CDK8 module and P-TEFb appears to be disrupted when the TFIIH kinase is inhibited by THZ1. As has been postulated in a recent review[45], these data portend a role for the TFIIH kinase as a master regulator of P-TEFb and the CDK8 module in early steps of transcription. These findings are summarized in **Table 3**.

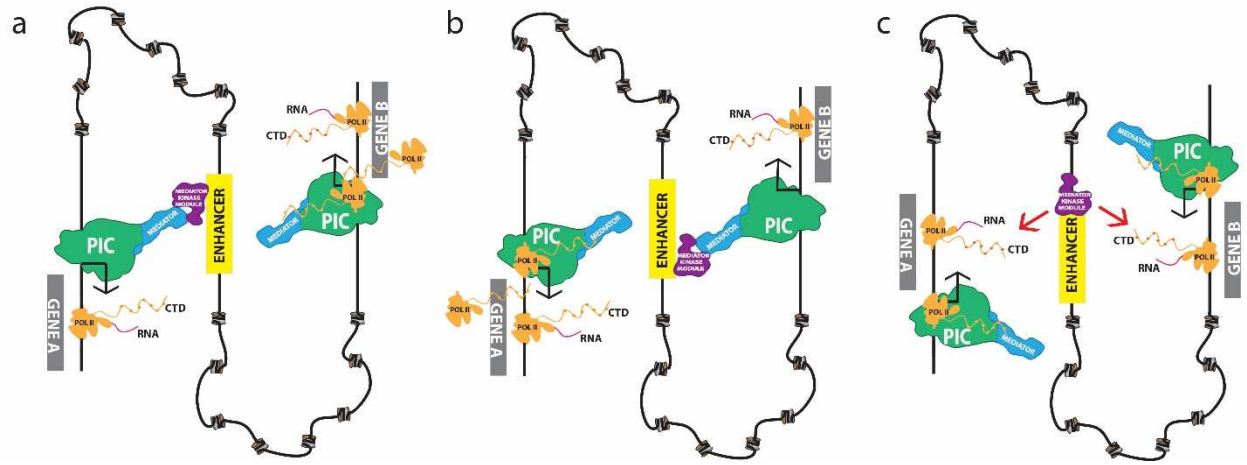


Figure 24: **Hypothesized models for how the CDK8 module may regulate transcription at human enhancers. a.** The CDK8 module (Mediator Kinase Module) in purple may regulate transcription initiation at the stage of promoter escape through establishing a link between enhancers and promoters. Enhancer bound module can associate with Mediator, forcing pol II escape from the promoter, and subsequent gene transcription. Promoters at which CDK8 is not bound to Mediator are occupied by pol II. **b.** Activated promoters are switched from that contained in a. **c.** Alternatively or in addition to, the CDK8 module can associate with actively transcribing pol II at both promoters and perhaps with P-TEFb (as part of the Super Elongation Complex, SEC)(not shown) to facilitate elongation. The roles of the kinase activity of the module are unclear, but in vitro data suggests that the CDK8 module may target the pol II CTD (Figure 1). Cell-based data has not confirmed this target in human cells but suggests that the CDK8 module may phosphorylate (directly or indirectly) AFF4 of the SEC (Poss *et al.* Cell Reports 2016).

	Inititation	Promoter Escape	Pause Release	Elongation
P-TEFb			Increases	Increases (with CDK8 module)
CDK8 Module	Decreases (previously shown)	Increases		Increases (with P-TEFb)
TFIIH	Essential (previously shown)	Increases	Increases	Increases (perhaps through other kinases)

Table 3: Summarized results of the roles of P-TEFb, the CDK8 module, and TFIIH in early steps of transcription. Generalized findings are shown from both previous work and work within this chapter. The CDK8 module has previously shown to decrease initiation if added prior to initiation[59]. Additionally, TFIIH has previously been shown to be essential for inititation[21].

Looking forward, these data will pair nicely with ongoing and recently published cell-based experiments that utilize next generation kinase inhibition strategies[72,95], highlighting the biological relevance of the reconstituted pausing assay. This circular process of biochemical assessment and cell-based validation (and vice versa) is particularly important when researching the transcriptional kinases: P-TEFb, the CDK8 module, and TFIIH all are hot areas of interest in cancer therapeutic development as well as therapeutics for other human diseases[156–159]. Highly targeted kinase inhibition strategies seem to be a way forward in the treatment of these diseases, but given the central roles of P-TEFb, the CDK8 module, and TFIIH in gene expression, any inhibitor should be carefully examined both biochemically and *in vivo* for its effects on transcription.

Please see Appendices I-III for **Methods.**

Chapter V: Future Directions

The research chapters of this thesis are composed of a tool development chapter (Chapter II) and two chapters that use the developed tool to test hypotheses (Chapters III-IV). Now that the tool is developed, preliminary data may be further assessed and additional hypotheses may be tested. Some possibilities are outlined below.

A. INVESTIGATING THE ROLES OF TFIIH, THE CDK8 MODULE, AND P-TEFb IN EARLY STEPS OF TRANSCRIPTION

Inter-regulation of the transcription-associated kinases. The reconstituted pausing assay could be used to complete the work on inter-regulation of the kinases TFIIH (CDK7), the CDK8 module (CDK8), and P-TEFb (CDK9). The effects of kinase inhibition presented in Chapter IV are preliminary but promising. Additional experiments may show the effects of the kinase activities of each of these complexes on early steps of transcription and, through combinatorial inhibition, may elucidate a regulatory hierarchy amongst the kinases (**Figure 25a**). For instance, we could test if CDK7 or CDK8 are required to activate P-TEFb[45,66] by inhibiting CDK7 with THZ1 and CDK8 with CA and testing their effects on P-TEFb-dependent pause release. If either CDK7 or CDK8 helps regulate pause release by P-TEFb, pausing should increase or decrease. Similar approaches could be used to test the activation/repression of both CDK7 and CDK8 kinase activities by the other two kinases.

Distinguishing direct and indirect targets of CDK7 and CDK9 in pause release. The above strategy could also be used to look at primary versus secondary targets of CDK7 and CDK9 on NELF and DSIF stabilized pause complexes. In cells, it is difficult to distinguish primary kinase targets from secondary targets. NELF and DSIF were shown in Chapter III to enhance pol II pausing. These complexes were released into elongation with the addition

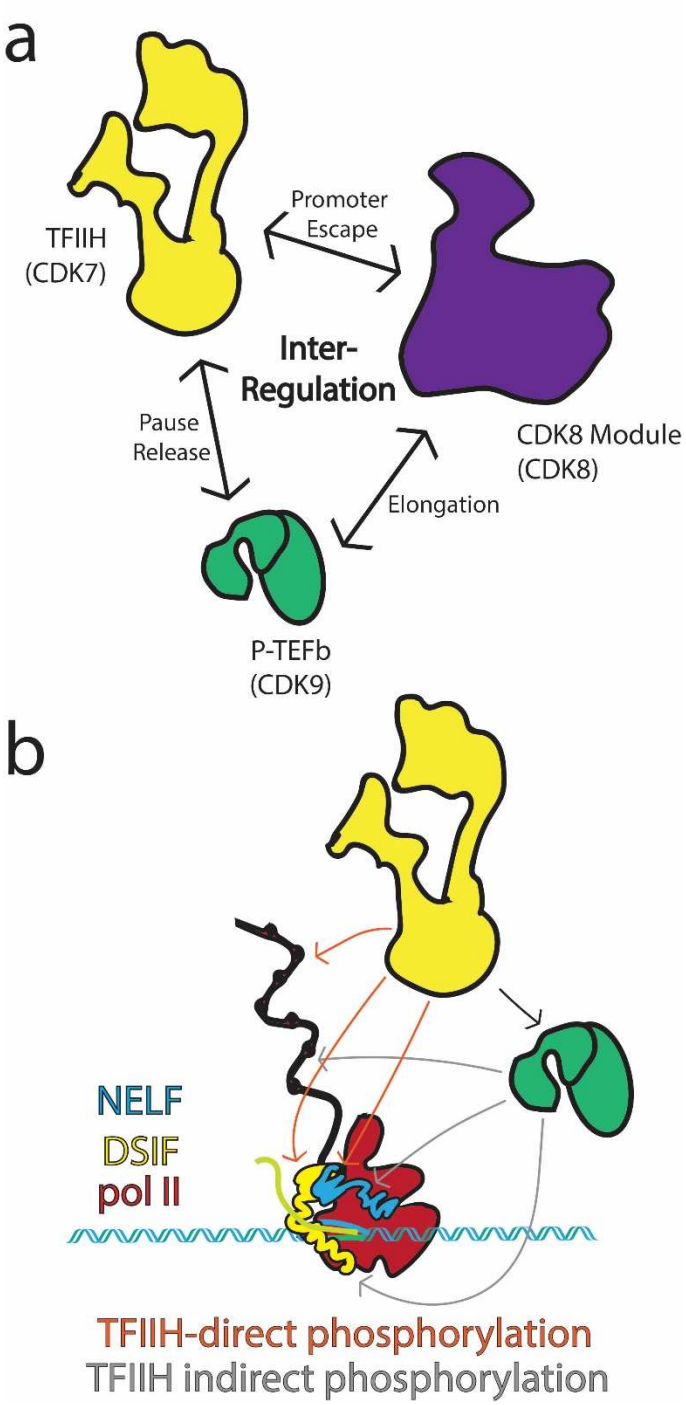


Figure 25: Figure 1: Inter-regulation of TFIIH, the CDK8 module, and P-TEFb. a. Kinase inter-regulation may play important roles in early steps of transcription. The reconstituted pausing assay can be used to test how these kinases regulate each other, particularly at the steps of promoter escape, pause release, and elongation. **b.** TFIIH (CDK7) and P-TEFb (CDK9) have some overlapping kinase targets. Due to complications in cells, it is difficult to distinguish indirect from direct kinase targets. While it is established that CDK7 activates CDK9 through phosphorylation (black arrow), it is unclear if the remaining kinase targets on DSIF, NELF, and the pol II CTD are redundantly phosphorylated by both CDK7 and CDK9.

of P-TEFb. CDK7 and CDK9 have both been shown to directly or indirectly phosphorylate NELF and DSIF in cell-based experiments[45,65,136], but it is unclear if these kinases function redundantly or cooperatively to modify NELF and DSIF to promote pause release (**Figure 25b**). Reconstituted transcription assays could be performed in which CDK7 or CDK9 were alternatively inhibited with THZ1[144] or CDK9-IN-2[149], respectively, in the presence of DSIF and NELF. Analysis of the nascent transcripts would be informative about the overall effects on transcription. A subsequent kinase experiment within the context of a pol II transcription assay could be performed in order to distinguish direct from indirect kinase targets in a biologically meaningful context. In this experiment, α-^{32}P-NTPs would not be used, and therefore the transcripts would not be detectable by phosphorimaging. However, ɣ-^{32}P-ATP would be added to the reactions, facilitating detection of kinase targets. In this experiment, proteins rather than RNA would be isolated. NELF and DSIF could be probed by western blot and the chemiluminescent signal overlaid with phosphor signal to look at NELF and DSIF phosphorylation events. Alternatively, a mass-spectrometry based approach could be used to identify sites of phosphorylation, similar to methods previously published[160]. If activation of CDK9 by CDK7 or CDK8 is required for pause release as predicted earlier, this could be controlled for by pre-phosphorylation of CDK9 with the needed kinase followed by its subsequent inhibition before addition to the transcription assay. Completing such experiments would answer important questions about primary versus secondary targets of CDK7 and CDK9 in pause release (Figure 1b).

Mechanism of CDK8-Module activation of promoter escape. Chapter IV provided evidence that the CDK8 module activates promoter escape but only postulated a

mechanism that is dependent on its interaction with Mediator. When the CDK8 module binds to Mediator, it extensively remodels its structure[59]. We have suggested that this structural rearrangement, in addition to pol II CTD phosphorylation by TFIIH, may be sufficient to remodel the CTD and dissociate it from Mediator. This is no small feat given the high affinity of Mediator for the pol II CTD ($K_d \cong 530$ pM)[161] and emphasizes the potential importance of the CDK8 module in promoter escape. The structural and/or scaffolding roles of the CDK8 module in promoter escape and elongation could be better defined through use of CDK8 module mutants that lack MED13, the subunit that facilitates association with Mediator[59]. The Taatjes lab can purify a three-subunit CDK8 module that lacks MED13 or versions of the four-subunit module with a MED13 truncation that also is unable to associate with Mediator[59]. Both are kinase active and permit the decoupling of CDK8 module kinase versus structural/conformational functions. Previous work showed that titration of the three-subunit module prior to or during PIC assembly inhibited transcription whereas the CDK8-CCNC dimer had no effect[59]. This result suggests an additional role for MED12 in PIC assembly. Unlike with the four-subunit wild-type CDK8 module, these variant modules were not added after transcription initiation. By purifying and testing these CDK8 module variants in the reconstituted pausing assay, the stage at which these variant modules inhibit transcription could be determined. The findings from these experiments may better develop our understanding of promoter escape.

Mechanism of HSF1 activation of P-TEFb. Chapter II demonstrated that we can achieve activated transcription on a non-chromatinized linear HSPA1B promoter DNA template. In Chapter IV, I suggested that P-TEFb may enhance pause release even in the absence of NELF and DSIF. Preliminary data show that HSF1 and P-TEFb may cooperatively

increase pause release (**Figure 26**). HSF1 has previously been suggested to recruit P-TEFb to promoters through unknown mechanisms[162]. Using the reconstituted pausing assay, we can test the hypothesis that HSF1 is required for elongation by P-TEFb and determine the

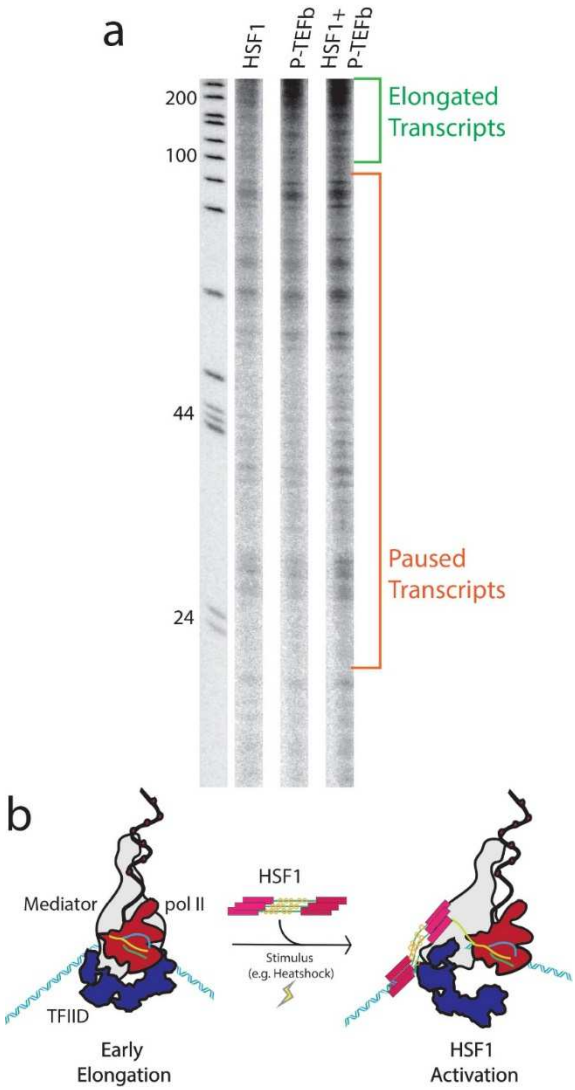


Figure 26: HSF1 and P-TEFb cooperate to promotes pause release. a. Preliminary reconstituted pausing data that shows that HSF1 and P-TEFb increase elongated transcripts and decrease the pausing index. **b.** A simplified model for how HSF1 may activate P-TEFb-dependent pause release. PIC factors other than those directly relevant are not shown. Upon HSF1 activation during stress (e.g. heat shock), the HSF1 DBD binds to the HSE, and the TAD binds to Mediator. These binding events likely facilitate a conformational change in Mediator[60,163], and perhaps also in associating TFIID. This conformational change may facilitate pol II pause release and elongation.

mechanism by which HSF1 activates P-TEFb. Early during elongation or during pausing, it has been proposed that HSF1 binds Mediator, triggering a conformational shift[163,164] that dissociates Mediator from TFIID and permits the association of P-TEFb and the SEC with pol II (**Figure 26**)[165]. This mechanism could be tested by removing Mediator from the assay, as it is not required for transcription on naked DNA templates *in vitro*. If this hypothesized mechanism is correct, the effect of HSF1 on P-TEFb activation will be lost.

B. INVESTIGATING THE FATE OF PAUSED POL II

Half-lives of paused complexes. The pausing assay could also be used to determine key parameters for pausing *in vitro* such as the half-life of the pause and how it may change upon addition of relevant factors. These experiments could be accomplished through time courses (as shown in Chapter II). After experimentally determining the time point after initiation at which pausing is strongest, time points could continually to be taken to determine the time after initiation at which 50% of the pause is released. Such data may not be entirely physiologically relevant due to NTP deprivation in the pausing assay[104–106], but will provide relative effects that are difficult to achieve in cells. Additionally, these experiments may provide some insight into the contentious issue of the duration of pausing in cells[56,166], though insights are limited by the defined nature of the reconstituted pausing assay—additional factors that may alter the stability of paused pol II in cells are not necessarily present *in vitro*.

Pol II turnover at promoter-proximal sites. The pausing assay could be used to test the degree of pol II turnover at the pause site. Recent work suggests that the fate of most paused polymerases is termination, with high pol II turnover at the promoter[167,168].

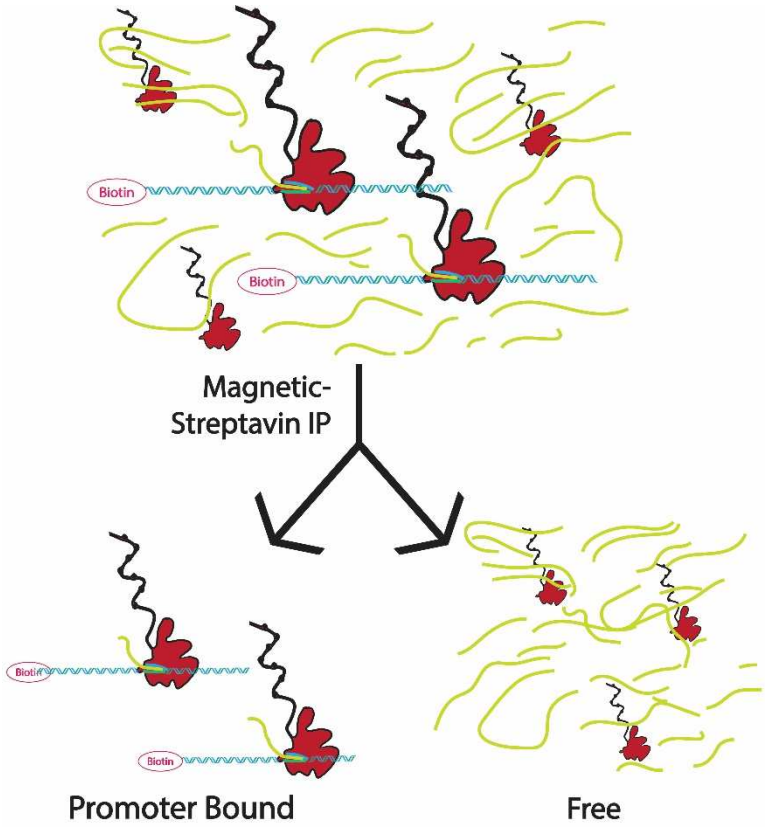


Figure 27: Schematic of immobilized template assay to test pol II turnover at promoter proximal sites. The reconstituted pausing assay can be performed on 5'-biotinylated templates. At various time points during the reaction, the template can be isolated from the buffer using magnetic streptavidin beads. The transcription reaction can be separated into a promoter bound fraction and into a free fraction. Ratios of free versus promoter-bound pol II can be correlated with paused versus elongated transcripts to assess pol II turnover.

By biotinylating the 5' end of the HSPA1B template, it is possible to isolate promoter bound factors from free/released factors using magnetic streptavidin beads[169]. Such templates have been shown to be transcriptionally active by others[78,89]. In principle, any factor associated with the promoter should be able to be isolated from unbound factors in the buffer. Pol II exists in several transcriptional stages during transcription: promoter bound, paused, elongating, backtracked, or terminated. To look at the fate of paused pol II (remaining, released from the pause, or terminated), one could separate promoter bound complexes from the buffer and run isolated RNA from both samples to gain a more complete picture of the fate of paused pol II (**Figure 27**). If the predominant fate of paused pol II *in vitro* is termination, we would hypothesize that the ratio of paused transcripts in the buffer would be greater than those in promoter-bound complexes. One could also probe for promoter bound versus free pol II over time by Western blot. Future versions of

this experiment could be conducted using a cold recombinant transcription assay followed up by analysis via mass spectrometry to look at the composition of promoter bound complexes over time paired with those in the buffer. These experiments may be challenging but could answer pervasive current questions into pol II promoter-proximal dynamics.

C. INVESTIGATING TFIID-DEPENDENT PAUSING MECHANISMS

Role of TFIID-promoter interactions. The data presented in Chapter III are interpreted to constitute a model in which TFIID enables pol II pausing, perhaps through its lobe C subunits TAF1, TAF2, and TAF7. We do not, however, test how lobe C achieves this feat. TFIID is not known to interact with pol II, although if this interaction were transient, it would be difficult to detect. Cryo-EM structures of the TFIID complex show that TAF1 and TAF2 of TFIID lobe C interact with downstream promoter elements[12,13]. These structures were performed with TFIID bound to the Super Core Promoter (SCP), an artificial promoter designed with idealized promoter elements[170]. Surprisingly, the promoter elements

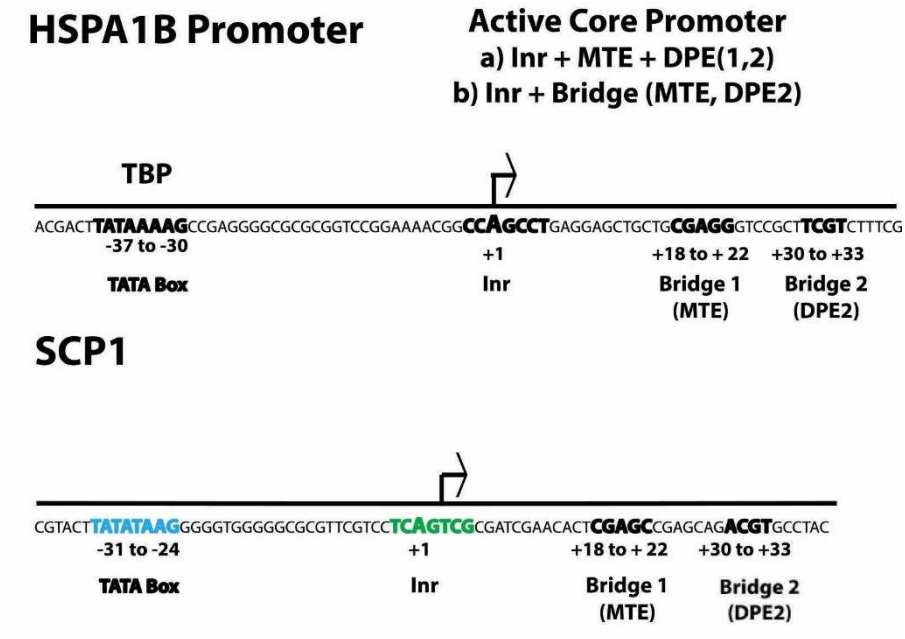


Figure 28: The HSPA1B promoter and SCP are similar. The HSPA1B promoter contains the same promoter elements as those contained in the SCP. The first version of the SCP is shown, SCP1. In particular, the downstream promoter elements are nearly identical and are positioned in the same place with respect to the TSS.

between the SCP and the HSPA1B promoter used for the reconstituted pausing assays are strikingly similar in terms of promoter element content, sequence, and spacing (**Figure 28**)[170]. Additionally, it has been shown that the positioning of the downstream promoter elements can affect both the strength and position of paused pol II[107]. Taken together, these data suggest that the interactions between TFIID and the downstream DNA are important.

Replicating the work done previously done in cells by the Lis lab *in vitro* would be an appropriate first step in addressing this hypothesis[83]. In these experiments the distance between the TSS and downstream promoter elements was altered through the insertion of five or 10 additional base pairs between the TSS and MTE at the native HSP70 locus in *Drosophila S2* cells. A similar template could be designed for analysis in the reconstituted pausing assay (Figure XX). In these experiments, we would expect that paused transcripts would be correspondingly longer depending on the length of the template insert (i.e. a five base pair DNA insert would correspond to an RNA five nucleotides longer) and that pausing would be lost with the 10 base pair insertion. These results could then be validated in cells using the same approach used by the Lis lab[83] under both normal and TAF1/2 knockdown conditions. We would expect that under TAF1/2 knockdown conditions, sensitivity to the positioning of the downstream promoter elements would be lost.

TFIIH, the CDK8 module, and P-TEFb regulation of TFIID function. It is important to note that due to the requirement for TFIID in pol II pausing, any mechanism by which pol II pausing is released likely involves TFIID. Factors that promote pause release may interact directly or indirectly with TFIID or may function through TFIID phosphorylation. For instance, TAF subunits have been identified as kinase targets of CDK7 (Rimel *et al.* unpublished), a TFIIH subunit that is present in the reconstituted pausing assay.

Additionally, TAFs are also confirmed to be high confidence kinase targets of CDK8[148]. P-TEFb has also been shown to interact with TFIID directly or indirectly through the TFIID lobe C subunit TAF7[171,172]. The potential functional importance of this interaction could be tested by using the 7-TAF or 8-TAF TFIID subcomplexes described in Chapter III to test the effects of TAF7 depletion on TFIID on P-TEFb enhanced pause release. If the interaction between TAF7 and P-TEFb is important for pause release, elongation would be expected to decrease upon its depletion. This hypothesis could also be tested in cells using a similar strategy to that for TAF1/2 depletion by Trim-Away in Chapter III or by the use of auxin-inducible degrons (AIDs)[173]. These findings further necessitate *in vitro* analysis of the roles of transcriptional-associated kinases in pausing and pause release.

D. INVESTIGATING PROPOSED ROLES FOR ADDITIONAL FACTORS IN PAUSING

Given the capabilities of the reconstituted pausing assay to assess early steps of transcription (e.g. initiation, promoter escape, pausing, pause release, and elongation), it can be used to test the effects of a variety of proteins on pol II functions. Some of the factors described in Chapter I (Table 1) are attractive candidates (e.g. BRD4[174], PAF1[175], and TRIM28[82]). Each of these factors can be purified and tested in the pausing assay to determine if they are indeed affecting pause regulation as suggested.

Enhancer RNAs (eRNAs) have also been implicated in pol II pause release by acting as a NELF decoy[53]. Their study would make for a challenging but exciting use of the *in vitro* transcription assay and may elucidate mechanisms by which enhancer-promoter looping regulates transcription. A current graduate student is currently working on designing and testing a promoter for this purpose. This study is a prime example of the merits of a reconstituted pausing assay—eRNAs have notoriously short half-lives[176] and are

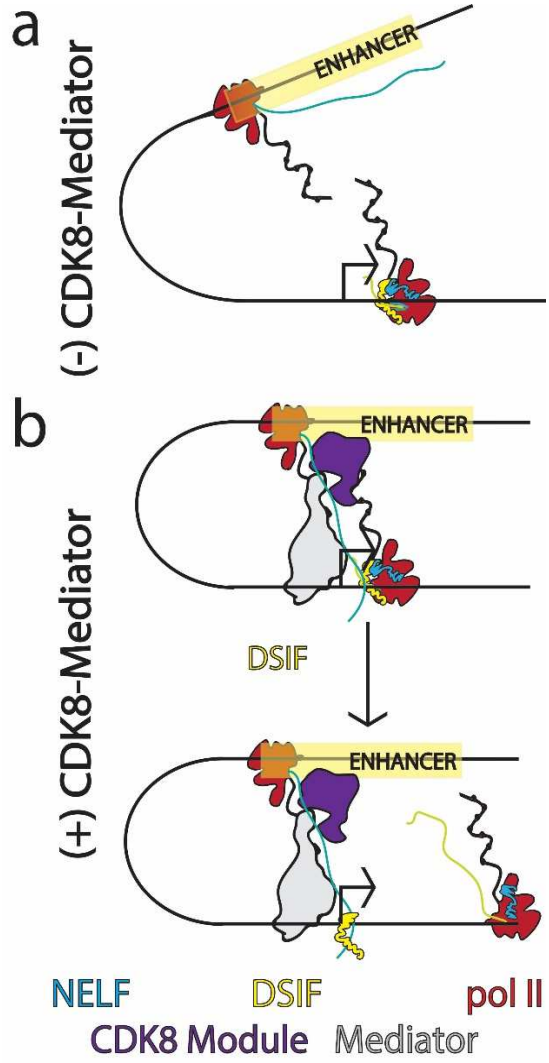


Figure 29: Speculative model for pause regulation by CDK8-Mediator, NELF, and eRNAs. a. In the absence of CDK8-Mediator, eRNA is transcribed, but is not proximal to paused pol II complexes. **b.** CDK8-Mediator facilitates enhancer-promoter looping, which juxtaposes the nascent eRNA with paused pol II. The nascent eRNA can bind NELF, promoting its disassociation from the pol II and facilitating pause release.

challenging to study in cells. Furthermore, cell-based assays cannot reliably assess direct versus indirect effects. However, in the reconstituted pausing assay, there are no RNAses to degrade the transcripts, thus artificially stabilizing the processes they regulate and making them accessible for experimentation.

The CDK8 module and Mediator may be important co-factors for eRNAs in pause regulation. As discussed in Chapter IV, the CDK8 module has been proposed to regulate eRNA transcription through yet unresolved mechanisms[72,75], though likely through interactions with Mediator, as Mediator has well-established roles for facilitation of enhancer-promoting looping[35,43]. The addition of the CDK8 module and Mediator may be necessary for bringing the enhancer into the proximity of promoter-proximal paused pol II, forming a bridge between enhancer-associated factors such as eRNAs. In forming this bridge, actively transcribed eRNAs may be brought into proximity to paused pol II, thus increasing the local concentration of eRNAs and increasing the likelihood of eRNA tethering-away of NELF in a

process that could facilitate pause release (**Figure 29**). This project represents an exciting new front onto which the reconstituted pausing assay can be deployed.

Bibliography

1. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–53 (2003).
2. Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science* **293**, 1074–1080 (2001).
3. Filion, G. J. *et al.* Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* **143**, 212–224 (2010).
4. Roy, S. *et al.* Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. *Science* **330**, 1787–1797 (2010).
5. Li, G. & Reinberg, D. Chromatin higher-order structures and gene regulation. *Current Opinion in Genetics & Development* **21**, 175–186 (2011).
6. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
7. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell research* **21**, 381–95 (2011).
8. Bai, L., Ondracka, A. & Cross, F. R. Multiple Sequence-Specific Factors Generate the Nucleosome-Depleted Region on CLN2 Promoter. *Molecular Cell* **42**, 465–476 (2011).
9. Gershenzon, N. I. & Ioshikhes, I. P. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics* **21**, 1295–1300 (2005).
10. Smale, S. T. & Baltimore, D. The "initiator" as a transcription control element. *Cell* **57**, 103–13 (1989).
11. Vo ngoc, L., Wang, Y.-L., Kassavetis, G. A. & Kadonaga, J. T. The punctilious RNA polymerase II core promoter. *Genes & Development* **31**, 1289–1301 (2017).
12. Louder, R. K. *et al.* Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604–609 (2016).
13. Patel, A. B. *et al.* Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362**, eaau8872 (2018).
14. Baptista, T. *et al.* SAGA Is a General Cofactor for RNA Polymerase II Transcription. *Molecular Cell* **68**, 130-143.e5 (2017).
15. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics* **13**, 233–245 (2012).
16. Li, G. *et al.* Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell* **148**, 84–98 (2012).

17. Souaid, C., Bloyer, S. & Noordermeer, D. Promoter–Enhancer Looping and Regulatory Neighborhoods: Gene Regulation in the Framework of Topologically Associating Domains. *Nuclear Architecture and Dynamics* 435–456 (2018). doi:10.1016/B978-0-12-803480-4.00018-1

18. Kieffer-Kwon, K.-R. *et al.* Interactome Maps of Mouse Gene Regulatory Domains Reveal Basic Principles of Transcriptional Regulation. *Cell* **155**, 1507–1520 (2013).

19. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).

20. Zhang, Y. *et al.* Chromatin connectivity maps reveal dynamic promoter–enhancer long-range associations. *Nature* **504**, 306–310 (2013).

21. Goodrich, J. A. & Tjian, R. Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* **77**, 145–56 (1994).

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

23. Tyree, C. M. *et al.* Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. *Genes & development* **7**, 1254–65 (1993).

24. Conaway, R. C. & Conaway, J. W. General Initiation Factors for RNA Polymerase II. *Annual Review of Biochemistry* **62**, 161–190 (1993).

25. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends in biochemical sciences* **21**, 327–35 (1996).

26. Grünberg, S. & Hahn, S. Structural insights into transcription initiation by RNA polymerase II. *Trends in Biochemical Sciences* **38**, 603–611 (2013).

27. Reinberg, D. *et al.* The RNA polymerase II general transcription factors: past, present, and future. *Cold Spring Harbor symposia on quantitative biology* **63**, 83–103 (1998).

28. Kornberg, R. D. Mechanism and regulation of yeast RNA polymerase II transcription. *Cold Spring Harbor symposia on quantitative biology* **63**, 229–32 (1998).

29. Greber, B. J., Toso, D. B., Fang, J. & Nogales, E. The complete structure of the human TFIIH core complex. *eLife* **8**, (2019).

30. He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013).

31. Zhu, X. *et al.* Histone modifications influence mediator interactions with chromatin. *Nucleic acids research* **39**, 8342–54 (2011).

32. Malik, S. & Roeder, R. G. Previews Mediator: A Drawbridge across the Enhancer-Promoter Divide. *Molecular Cell* **64**, 433–434 (2016).

33. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription.

Nature reviews. Molecular cell biology **16**, 155–66 (2015).

34. Petrenko, N., Jin, Y., Wong, K. H. & Struhl, K. Mediator Undergoes a Compositional Change during Transcriptional Activation. *Molecular Cell* **64**, 443–454 (2016).

35. Jeronimo, C. *et al.* Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. *Molecular Cell* **64**, 455–466 (2016).

36. Jiang, Y., Yan, M. & Gralla, J. D. Abortive initiation and first bond formation at an activated adenovirus E4 promoter. *The Journal of biological chemistry* **270**, 27332–8 (1995).

37. Kugel, J. F. & Goodrich, J. A. Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA. *Proceedings of the National Academy of Sciences* **95**, 9232–9237 (1998).

38. Kugel, J. F. & Goodrich, J. A. Translocation after synthesis of a four-nucleotide RNA commits RNA polymerase II to promoter escape. *Molecular and cellular biology* **22**, 762–73 (2002).

39. He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359–365 (2016).

40. Dvir, A. Promoter escape by RNA polymerase II. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1577**, 208–223 (2002).

41. Moreland, R. J. *et al.* A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. *The Journal of biological chemistry* **274**, 22127–30 (1999).

42. Wong, K. H., Jin, Y. & Struhl, K. TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Molecular Cell* **54**, 601–612 (2014).

43. Petrenko, N., Jin, Y., Wong, K. H. & Struhl, K. Evidence that Mediator is essential for Pol II transcription, but is not a required component of the preinitiation complex in vivo. *eLife* **6**, (2017).

44. Ebmeier, C. C. *et al.* Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. *Cell Reports* **20**, 1173–1186 (2017).

45. Rimel, J. K. & Taatjes, D. J. The essential and multifunctional TFIIH complex. *Protein Science* **27**, 1018–1037 (2018).

46. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes & development* (2019). doi:10.1101/gad.325142.119

47. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* **560**, 601–606 (2018).

48. Gressel, S. *et al.* CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* **6**, (2017).

49. Vos, S. M. *et al.* Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* **560**, 607–612 (2018).

50. Brès, V., Yoh, S. M. & Jones, K. A. The multi-tasking P-TEFb complex. *Current opinion in cell biology* **20**, 334–40 (2008).

51. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nature reviews. Genetics* **15**, 163 (2014).

52. Gibson, B. A. *et al.* Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. *Science* **353**, 45–50 (2016).

53. Schaukowitch, K. *et al.* Enhancer RNA Facilitates NELF Release from Immediate Early Genes. *Molecular Cell* **56**, 29–42 (2014).

54. Fay, A. *et al.* Cohesin selectively binds and regulates genes with paused RNA polymerase. *Current biology : CB* **21**, 1624–34 (2011).

55. Luo, Z., Lin, C. & Shilatifard, A. The super elongation complex (SEC) family in transcriptional control. *Nature Reviews Molecular Cell Biology* **13**, 543–547 (2012).

56. Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3**, (2014).

57. Kawamura, Y., Koyama, S. & Yoshida, R. Statistical inference of the rate of RNA polymerase II elongation by total RNA sequencing. *Bioinformatics* **35**, 1877–1884 (2019).

58. Näär, A. M., Taatjes, D. J., Zhai, W., Nogales, E. & Tjian, R. Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. *Genes & development* **16**, 1339–44 (2002).

59. Knuesel, M. T., Meyer, K. D., Bernecky, C. & Taatjes, D. J. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes & Development* **23**, 439–451 (2009).

60. Bernecky, C., Grob, P., Ebmeier, C. C., Nogales, E. & Taatjes, D. J. Molecular Architecture of the Human Mediator–RNA Polymerase II–TFIIF Assembly. *PLoS Biology* **9**, e1000603 (2011).

61. Williams, L. H., Fromm, G., Hu, G. & Adelman, K. Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks. *Molecular Cell* **58**, 311–322 (2015).

62. Wu, C.-H. *et al.* NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes & Development* **17**, 1402–1414 (2003).

63. Qiu, Y. & Gilmour, D. S. Identification of Regions in the Spt5 Subunit of DRB Sensitivity-inducing Factor (DSIF) That Are Involved in Promoter-proximal Pausing. *The Journal of biological chemistry* **292**, 5555–5570 (2017).

64. Nilson, K. A. *et al.* THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and

Pausing. *Molecular Cell* **59**, 576–587 (2015).

65. Ebmeier, C. C. *et al.* Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. *Cell Reports* **20**, 1173–1186 (2017).

66. Larochelle, S. *et al.* Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nature Structural & Molecular Biology* **19**, 1108–1115 (2012).

67. Larochelle, S. *et al.* Dichotomous but stringent substrate selection by the dual-function Cdk7 complex revealed by chemical genetics. *Nature Structural & Molecular Biology* **13**, 55–62 (2006).

68. Patel, S. A. & Simon, M. C. Functional analysis of the Cdk7.cyclin H.Mat1 complex in mouse embryonic stem cells and embryos. *The Journal of biological chemistry* **285**, 15587–98 (2010).

69. Glover-Cutter, K. *et al.* TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. *Molecular and Cellular Biology* **29**, 5455–5464 (2009).

70. Galbraith, M. D. *et al.* HIF1A Employs CDK8-Mediator to Stimulate RNAPII Elongation in Response to Hypoxia. *Cell* **153**, 1327–1339 (2013).

71. Donner, A. J., Ebmeier, C. C., Taatjes, D. J. & Espinosa, J. M. CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nature Structural & Molecular Biology* **17**, 194–201 (2010).

72. Pelish, H. E. *et al.* Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature* **526**, 273–276 (2015).

73. Chen, M. *et al.* CDK8/19 Mediator kinases potentiate induction of transcription by NFκB. *Proceedings of the National Academy of Sciences* **114**, 10208–10213 (2017).

74. Bancerek, J. *et al.* CDK8 Kinase Phosphorylates Transcription Factor STAT1 to Selectively Regulate the Interferon Response. *Immunity* **38**, 250–262 (2013).

75. Fant, C. B. & Taatjes, D. J. Regulatory functions of the Mediator kinases CDK8 and CDK19. *Transcription* **10**, 76–90 (2019).

76. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews. Genetics* **13**, 720–31 (2012).

77. Horn, A. E., Kugel, J. F. & Goodrich, J. A. Single molecule microscopy reveals mechanistic insight into RNA polymerase II preinitiation complex assembly and transcriptional activity. *Nucleic Acids Research* **44**, gkw321 (2016).

78. Renner, D. B., Yamaguchi, Y., Wada, T., Handa, H. & Price, D. H. A Highly Purified RNA Polymerase II Elongation Control System. *Journal of Biological Chemistry* **276**, 42601–42609 (2001).

79. Reach, M., Xu, L. X. & Young, C. S. Transcription from the adenovirus major late

promoter uses redundant activating elements. *The EMBO journal* **10**, 3439–46 (1991).

80. Buckley, M. S., Kwak, H., Zipfel, W. R. & Lis, J. T. Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation. *Genes & development* **28**, 14–9 (2014).

81. Samarakkody, A. *et al.* RNA polymerase II pausing can be retained or acquired during activation of genes involved in the epithelial to mesenchymal transition. *Nucleic acids research* **43**, 3938–49 (2015).

82. Bunch, H. *et al.* TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. *Nature structural & molecular biology* **21**, 876–83 (2014).

83. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science* **339**, 950–953 (2013).

84. Mandal, S. S. *et al. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II*. **101**, (PNAS, 2004).

85. Muse, G. W. *et al.* RNA polymerase is poised for activation across the genome. *Nature Genetics* **39**, 1507–1511 (2007).

86. Chao, S.-H. & Price, D. H. Flavopiridol Inactivates P-TEFb and Blocks Most RNA Polymerase II Transcription *in Vivo*. *Journal of Biological Chemistry* **276**, 31793–31799 (2001).

87. Yamada, T. *et al.* P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats Is Critical for Processive Transcription Elongation. *Molecular Cell* **21**, 227–237 (2006).

88. Li, J. *et al.* Kinetic Competition between Elongation Rate and Binding of NELF Controls Promoter-Proximal Pausing. *Molecular Cell* **50**, 711–722 (2013).

89. Qiu, Y. & Gilmour, D. S. Identification of Regions in the Spt5 Subunit of DRB Sensitivity-inducing Factor (DSIF) That Are Involved in Promoter-proximal Pausing. *Journal of Biological Chemistry* **292**, 5555–5570 (2017).

90. Hendrix, D. A., Hong, J.-W., Zeitlinger, J., Rokhsar, D. S. & Levine, M. S. Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7762–7 (2008).

91. Trinklein, N. D., Murray, J. I., Hartman, S. J., Botstein, D. & Myers, R. M. The Role of Heat Shock Transcription Factor 1 in the Genome-wide Regulation of the Mammalian Heat Shock Response. *Molecular Biology of the Cell* **15**, 1254–1261 (2004).

92. Nechaev, S. *et al.* Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila. *Science* **327**, 335–338 (2010).

93. Henriques, T. *et al.* Widespread transcriptional pausing and elongation control at enhancers. *Genes & Development* **32**, 26–41 (2018).

94. CPHolstege, F., Cvan der Viiet, P. & Timmers, Ht. *Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH.; Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. The EMBO Journal* **15**, (1996).

95. Ebmeier, C. C. *et al.* Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. *Cell reports* **20**, 1173–1186 (2017).

96. Sarge, K. D., Murphy, S. P. & Morimoto, R. I. Activation of heat shock gene transcription by heat shock factor 1 involves oligomerization, acquisition of DNA-binding activity, and nuclear localization and can occur in the absence of stress. *Molecular and cellular biology* **13**, 1392–407 (1993).

97. Baler, R., Dahl, G. & Voellmy, R. Activation of human heat shock genes is accompanied by oligomerization, modification, and rapid translocation of heat shock transcription factor HSF1. *Molecular and cellular biology* **13**, 2486–96 (1993).

98. Holmberg, C. I. *et al.* Phosphorylation of serine 230 promotes inducible transcriptional activity of heat shock factor 1. *The EMBO Journal* **20**, 3800–3810 (2001).

99. Hilgarth, R. S., Hong, Y., Park-Sarge, O.-K. & Sarge, K. D. Insights into the regulation of heat shock transcription factor 1 SUMO-1 modification. *Biochemical and biophysical research communications* **303**, 196–200 (2003).

100. Duarte, F. M. *et al.* Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes & development* **30**, 1731–46 (2016).

101. Niskanen, E. A. *et al.* Global SUMOylation on active chromatin is an acute heat stress response restricting transcription. *Genome Biology* **16**, 153 (2015).

102. Tome, J. M., Tippens, N. D. & Lis, J. T. Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nature Genetics* **50**, 1533–1541 (2018).

103. Core, L. J. *et al.* Defining the Status of RNA Polymerase at Promoters. *Cell Reports* **2**, 1025–1035 (2012).

104. Lerner, E. *et al.* Backtracked and paused transcription initiation intermediate of *Escherichia coli* RNA polymerase. *Proceedings of the National Academy of Sciences* **113**, E6562–E6571 (2016).

105. Duchi, D. *et al.* RNA Polymerase Pausing during Initial Transcription. *Molecular cell* **63**, 939–50 (2016).

106. Rudd, M. D. & Luse, D. S. Amanitin greatly reduces the rate of transcription by RNA polymerase II ternary complexes but fails to inhibit some transcript cleavage modes. *The Journal of biological chemistry* **271**, 21549–58 (1996).

107. Kwak, H. & Lis, J. T. Control of transcriptional elongation. *Annual review of genetics*

47, 483–508 (2013).

108. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).

109. Adelman, K. *et al.* Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. *Proceedings of the National Academy of Sciences* **106**, 18207–18212 (2009).

110. Lagha, M. *et al.* Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. *Cell* **153**, 976–87 (2013).

111. Lin, C. *et al.* AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Molecular cell* **37**, 429–37 (2010).

112. Miller, T. E. *et al.* Transcription elongation factors represent in vivo cancer dependencies in glioblastoma. *Nature* **547**, 355–359 (2017).

113. Wei, P., Garber, M. E., Fang, S.-M., Fischer, W. H. & Jones, K. A. A Novel CDK9-Associated C-Type Cyclin Interacts Directly with HIV-1 Tat and Mediates Its High-Affinity, Loop-Specific Binding to TAR RNA. *Cell* **92**, 451–462 (1998).

114. Yamaguchi, Y. Stimulation of RNA Polymerase II Elongation by Hepatitis Delta Antigen. *Science* **293**, 124–127 (2001).

115. Marshall, N. F. & Price, D. H. Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Molecular and cellular biology* **12**, 2078–90 (1992).

116. Lai, W. K. M. & Pugh, B. F. Genome-wide uniformity of human 'open' pre-initiation complexes. *Genome Research* **27**, 15–26 (2017).

117. Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3**, e02407 (2014).

118. Lee, C. *et al.* NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. *Molecular and cellular biology* **28**, 3290–300 (2008).

119. Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nature Genetics* **39**, 1512–1516 (2007).

120. Näär, A. M. *et al.* Chromatin, TAFs, and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP-1a in vitro. *Genes & development* **12**, 3020–31 (1998).

121. Clift, D. *et al.* A Method for the Acute and Rapid Degradation of Endogenous Proteins. *Cell* **171**, 1692-1706.e18 (2017).

122. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nature genetics* **49**, 1045–1051 (2017).

123. Pennington, K. L., Marr, S. K., Chirn, G.-W. & Marr, M. T. Holo-TFIID controls the magnitude of a transcription burst and fine-tuning of transcription. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 7678–83 (2013).

124. Wright, K. J., Marr, M. T. & Tjian, R. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proceedings of the National Academy of Sciences* **103**, 12347–12352 (2006).

125. Gressel, S. *et al.* CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* **6**, (2017).

126. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and Function of a Human TAFII250 Double Bromodomain Module. *Science* **288**, 1422–1425 (2000).

127. Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).

128. Xu, Y. *et al.* A TFIID-SAGA Perturbation that Targets MYB and Suppresses Acute Myeloid Leukemia. *Cancer cell* **33**, 13-28.e8 (2018).

129. Aneichyk, T. *et al.* Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly. *Cell* **172**, 897-909.e21 (2018).

130. Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D. & Handa, H. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *The EMBO Journal* **17**, 7395 (1998).

131. Yu, M. *et al.* RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350**, 1383–1386 (2015).

132. Liang, K. *et al.* Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Molecular and cellular biology* **35**, 928–38 (2015).

133. Sedlacek, H. *et al.* Flavopiridol (L86 8275; NSC 649890), a new kinase inhibitor for tumor therapy. *International Journal of Oncology* **9**, 1143–68 (1996).

134. Asghar, U., Witkiewicz, A. K., Turner, N. C. & Knudsen, E. S. The history and future of targeting cyclin-dependent kinases in cancer therapy. *Nature Reviews Drug Discovery* **14**, 130–146 (2015).

135. Bishop, A. C. *et al.* A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* **407**, 395–401 (2000).

136. Booth, G. T., Parua, P. K., Sansó, M., Fisher, R. P. & Lis, J. T. Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. *Nature Communications* **9**, 543 (2018).

137. Tsai, K.-L. *et al.* A conserved Mediator–CDK8 kinase module association regulates Mediator–RNA polymerase II interaction. *Nature Structural & Molecular Biology* **20**,

611–619 (2013).

138. Donner, A. J., Szostek, S., Hoover, J. M. & Espinosa, J. M. CDK8 Is a Stimulus-Specific Positive Coregulator of p53 Target Genes. *Molecular Cell* **27**, 121–133 (2007).

139. Audetat, K. A. *et al.* A Kinase-Independent Role for Cyclin-Dependent Kinase 19 in p53 Response. *Molecular and Cellular Biology* **37**, (2017).

140. Westerling, T., Kuuluvainen, E. & Makela, T. P. Cdk8 Is Essential for Preimplantation Mouse Development. *Molecular and Cellular Biology* **27**, 6177–6182 (2007).

141. Tirode, F., Busso, D., Coin, F. & Egly, J. M. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Molecular cell* **3**, 87–95 (1999).

142. Greber, B. J. *et al.* The cryo-electron microscopy structure of human transcription factor IIH. *Nature* **549**, 414–417 (2017).

143. Schilbach, S. *et al.* Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204–209 (2017).

144. Kwiatkowski, N. *et al.* Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616–620 (2014).

145. Akoulitchev, S., Chuikov, S. & Reinberg, D. TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature* **407**, 102–106 (2000).

146. Peterlin, B. M. & Price, D. H. Controlling the Elongation Phase of Transcription with P-TEFb. *Molecular Cell* **23**, 297–305 (2006).

147. Chen, R., Yang, Z. & Zhou, Q. Phosphorylated Positive Transcription Elongation Factor b (P-TEFb) Is Tagged for Inhibition through Association with 7SK snRNA. *Journal of Biological Chemistry* **279**, 4153–4160 (2004).

148. Poss, Z. C. *et al.* Identification of Mediator Kinase Substrates in Human Cells using Cortistatin A and Quantitative Phosphoproteomics. *Cell Reports* **15**, 436–450 (2016).

149. Zawistowski, J. S. *et al.* Enhancer Remodeling during Adaptive Bypass to MEK Inhibition Is Attenuated by Pharmacologic Targeting of the P-TEFb Complex. *Cancer Discovery* **7**, 302–321 (2017).

150. Firestein, R. *et al.* CDK8 is a colorectal cancer oncogene that regulates β-catenin activity. *Nature* **455**, 547–551 (2008).

151. Ganuza, M. *et al.* Genetic inactivation of Cdk7 leads to cell cycle arrest and induces premature aging due to adult stem cell exhaustion. *The EMBO journal* **31**, 2498–510 (2012).

152. Nemet, J., Jelicic, B., Rubelj, I. & Sopta, M. The two faces of Cdk8, a positive/negative regulator of transcription. *Biochimie* **97**, 22–27 (2014).

153. Dannappel, M. V., Sooraj, D., Loh, J. J. & Firestein, R. Molecular and in vivo Functions

of the CDK8 and CDK19 Kinase Modules. *Frontiers in cell and developmental biology* **6**, 171 (2018).

154. Bösken, C. A. *et al.* The structure and substrate specificity of human Cdk12/Cyclin K. *Nature Communications* **5**, 3505 (2014).

155. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology* **16**, 155–166 (2015).

156. Philip, S., Kumarasiri, M., Teo, T., Yu, M. & Wang, S. Cyclin-Dependent Kinase 8: A New Hope in Targeted Cancer Therapy? *Journal of Medicinal Chemistry* **61**, 5073–5092 (2018).

157. Franco, L. C., Morales, F., Boffo, S. & Giordano, A. CDK9: A key player in cancer and other diseases. *Journal of Cellular Biochemistry* **119**, 1273–1284 (2018).

158. Chipumuro, E. *et al.* CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer. *Cell* **159**, 1126–1139 (2014).

159. Hu, S. *et al.* Discovery and Characterization of SY-1365, a Selective, Covalent Inhibitor of CDK7. *Cancer research* **79**, 3479–3491 (2019).

160. Joo, Y. J., Ficarro, S. B., Marto, J. A. & Buratowski, S. In vitro assembly and proteomic analysis of RNA polymerase II complexes. *Methods* **159–160**, 96–104 (2019).

161. Robinson, P. J. *et al.* Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166**, 1411-1414.e16 (2016).

162. Yoon, Y. J. *et al.* KRIBB11 inhibits HSP70 synthesis through inhibition of heat shock factor 1 function by impairing the recruitment of positive transcription elongation factor b to the hsp70 promoter. *The Journal of biological chemistry* **286**, 1737–47 (2011).

163. Taatjes, D. J., Näär, A. M., Andel, F., Nogales, E. & Tjian, R. Structure, Function, and Activator-Induced Conformations of the CRSP Coactivator. *Science* **295**, 1058–1062 (2002).

164. Meyer, K. D., Lin, S., Bernecky, C., Gao, Y. & Taatjes, D. J. p53 activates transcription by directing structural shifts in Mediator. *Nature Structural & Molecular Biology* **17**, 753–760 (2010).

165. Takahashi, H. *et al.* Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* **146**, 92–104 (2011).

166. Steurer, B. *et al.* Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proceedings of the National Academy of Sciences* **115**, E4368–E4376 (2018).

167. Erickson, B., Sheridan, R. M., Cortazar, M. & Bentley, D. L. Dynamic turnover of paused Pol II complexes at human promoters. *Genes & development* **32**, 1215–1225 (2018).

168. Krebs, A. R., Imanci, D., Hoerner, L., Gaidatzis, D. & Burger, L. Genome-wide Single-

Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular Cell* **67**, 411-422.e4 (2017).

169. Carey, M. F., Peterson, C. L. & Smale, S. T. Protein complex binding to promoter DNA: immobilized template assay. *Cold Spring Harbor protocols* **2010**, pdb.prot5465 (2010).

170. Juven-Gershon, T., Cheng, S. & Kadonaga, J. T. Rational design of a super core promoter that enhances gene expression. *Nature Methods* **3**, 917–922 (2006).

171. Yadav, D., Ghosh, K., Basu, S., Roeder, R. G. & Biswas, D. Multivalent Role of Human TFIID in Recruiting Elongation Components at the Promoter-Proximal Region for Transcriptional Control Poly-Ser domain aids in TFIID-dependent SEC recruitment for transcriptional activation Article Multivalent Role of Human TFIID in Recruiting Elongation Components at the Promoter-Proximal Region for Transcriptional Control. *CellReports* **26**, 1303-1317.e7 (2019).

172. Gegonne, A. *et al. TFIID component TAF7 functionally interacts with both TFIIH and P-TEFb*.

173. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods* **6**, 917–922 (2009).

174. Winter, G. E. *et al.* BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. *Molecular Cell* **67**, 5-18.e19 (2017).

175. Chen, F. X. *et al.* PAF1 regulation of promoter-proximal pause release via enhancer activation. *Science (New York, N.Y.)* **357**, 1294–1298 (2017).

176. Rusk, N. The human transient transcriptome. *Nature Methods* **13**, 612–612 (2016).

Appendix I: Transcription Protocols

A. TEMPLATES AND TEMPLATE PREPARATION

HSPA1B template

TCTCAGGCAGACTAGGCCATTAGGTGCCTCGGAGAAAGGACCCAAGGCTGCTCCGTCCTTCACAG

ACACAGTCCAATCAGAGTTTCCCAGGCACATCGATGCACCGCCTCCTTCGAGAAACAAGGTAACT

TTCGGGTTCTGGTTGTCTCCAAAGTCATCCGACCAATCTCGCACCGCCCAGAGCGGGCCCTTCCTG

TCAATTACCTACTGAAGGGCAGGCGGCCAGCATCGCCATGGAGACCAACACCCTTCCCACCACCA

CTCCCCCTTTCTCTCAGGGCCCCTGTCCCCTCCAGTGAATCCCAGAAGACTCTGGAGAGTTCTGAG

CAGAGGGCGGCACCCTGCCCTCTGATTGGTCCAAGGAAGGCTGGGGGGCAGGACGGGAGGCGAAA

CCCCTGGAATATTCCCGACCTGGCAGCCTCATCGAGCTTGGTGATTGGCTCAGAAGGGGAAAGGC

GGGTCTCCACGACGACTTATAAAAGCCGAGGGGCGCGCGGTCCGGAAAACGGCCAGCCTGAGGAG

CTGCTGCGAGGGTCCGCTTCGTCTTTCGAGAGTGACTCCCGCGGTCCCAAGGCTTTCCAGAGCGA

ACCTGTGCGGCTGCAGGCACCGGCGTGTTGAGTTTCCGGCGTTCCGAAGGACTGAGCTCTTGTCG

CGGATCCCGTCCGCCGTTTCCAGCCCCCAGTCTCAGAGCGGAGCCCACAGAGCAGGGCACCGGC

+1 to +216

SCP1 Template #1

TCTCAGGCAGACTAGGCCATTAGGTGCCTCGGAGAAAGGACCCAAGGCTGCTCCGTCCTTCACAG

ACACAGTCCAATCAGAGTTTCCCAGGCACATCGATGCACCGCCTCCTTCGAGAAACAAGGTAACT

TTCGGGTTCTGGTTGTCTCCAAAGTCATCCGACCAATCTCGCACCGCCCAGAGCGGGCCCTTCCTG

TCAATTACCTACTGAAGGGCAGGCGGCCAGCATCGCCATGGAGACCAACACCCTTCCCACCACCA

CTCCCCCTTTCTCTCAGGGCCCCTGTCCCCTCCAGTGAATCCCAGAAGACTCTGGAGAGTTCTGAG

CAGAGGGCGGCACCCTGCCCTCTGATTGGTCCAAGGAAGGCTGGGGGGCAGGACGGGAGGCGAAA

CCCCTGGAATATTCCCGACCTGGCAGCCTCATCGAGCTTGGTGATTGGCTCAGAAGGGGAAAGGC

GGGTCTCGTACTTATATAAGGGGGTGGGGGCGCGTTCGTCCTCAGTCGCGATCGAACACTCGAGC

CGAGCAGACGTGCCTACGGACCGTTTCGAGAGTGACTCCCGCGGTCCCAAGGCTTTCCAGAGCGA

ACCTGTGCGGCTGCAGGCACCGGCGTGTTGAGTTTCCGGCGTTCCGAAGGACTGAGCTCTTGTCG

CGGATCCCGTCCGCCGTTTCCAGCCCCCAGTCTCAGAGCGGAGCCCACAGAGCAGGGCACCGGC

+1 to +216

SCP1 Template #2

TCTCAGGCAGACTAGGCCATTAGGTGCCTCGGAGAAAGGACCCAAGGCTGCTCCGTCCTTCACAG

ACACAGTCCAATCAGAGTTTCCCAGGCACATCGATGCACCGCCTCCTTCGAGAAACAAGGTAACT

TTCGGGTTCTGGTTGTCTCCAAAGTCATCCGACCAATCTCGCACCGCCCAGAGCGGGCCCTTCCTG

TCAATTACCTACTGAAGGGCAGGCGGCCAGCATCGCCATGGAGACCAACACCCTTCCCACCACCA

CTCCCCCTTTCTCTCAGGGCCCCTGTCCCCTCCAGTGAATCCCAGAAGACTCTGGAGAGTTCTGAG

CAGAGGGCGGCACCCTGCCCTCTGATTGGTCCAAGGAAGGCTGGGGGGCAGGACGGGAGGCGAAA

CCCCTGGAATATTCCCGACCTGGCAGCCTCATCGAGCTTGGTGATTGGCTCAGAAGGGGAAAGGC

GGGTCTCGTACTTATATAAGGGGGTGGGGGCGCGTTCGTCCTCAGTCGCGATCGAACACTCGAGC

CGAGCAGACGTGTCCGCTTCGTCTTTCGAGAGTGACTCCCGCGGTCCCAAGGCTTTCCAGAGCGA

ACCTGTGCGGCTGCAGGCACCGGCGTGTTGAGTTTCCGGCGTTCCGAAGGACTGAGCTCTTGTCG

CGGATCCCGTCCGCCGTTTCCAGCCCCCAGTCTCAGAGCGGAGCCCACAGAGCAGGGCACCGGC

+1 to +216

SCP1 Template #3

GAGTGGAGGAGCCGAGTTTATTGGAGGACATTCTTCCAGAATTTCCGCCGGCGTCCTGATCGTCA

TTCGTCACTAACGTGCTCGGTGCGTGCAGCACATGCCGGCGAATCCGCCCCCGTTAGCCGAAGCC

ATGCTAGACCACTCGCAATTCCGACTTAGCGATCAGAGCCACCACGGAAGGACTGGTACCAGGTC

GGGCCTTCGCCCGTCAGTCCCCCCCAGGGTCCCACTCATCAGCGCGCAGCACCAGGATCCATCAAC

TACCCTAACTCGACCCAGAGCCTCATACGAAGCACTACCGCGGCGAGGGAGGTATCAAGGTCTTC

CTGTGTTATTAGCCACAGCCAGTCCCGTACGGAAGAGGATCTGTTCCGGAGTGTTAGCGCTCGGC

CATCAAAGCAGTGCCCGCCCCTCAGTGAGGCAGGCCCACGCCAACATAGAGTTCAGTGTAGCTAC

GGCTCAAGGTACTTATATAAGGGGGTGGGGGCGCGTTCGTCCTCAGTCGCGATCGAACACTCGAG

CCGAGCAGACGTGCCTACGGACCGCCGCCGCCAGGTGTTCGGGTGGCGTCTGCCTCCCATCGTGG

AGCTGCAGCACCAGGGGGAATTTACCCGTCTGCGAGCTAGAACCGCGGAGTACTGCCCCGTTGCC

CGTCTAGCCGCACTCGGAGCACGTATATCTGGTGCCCCGAGTAACCGACTCGCCGGGATATGGCC

+1 to +216

Template preparation

The native human HSPA1B promoter was amplified from genomic DNA (HeLa) by PCR (forward primer: CTCCTT CCCATT AAGACG GAAAAA ACATCC GGGAGA GCCGGT CCG; reverse primer: ACCTTG CCGTGT TGGAAC ACCCCC ACGCAG GAGTAG GTGGTG CCCAGGTC) and cloned into a pCR-Blunt-TOPO plasmid. The HSPA1B promoter corresponding to -500 to +216 base pairs relative to the transcription start site was amplified off this plasmid using Phusion polymerase (Thermo-Fisher Catalogue #F530S). The resulting PCR product was then purified using the E.Z.N.A. Gel Extraction kit (Omega BioTek Catalogue #D2500). The DNA was then ethanol precipitated, washed, resuspended to 100 nM in milliQ water, and stored frozen in single-use aliquots.

B. STANDARD RUNOFF ASSAY

When to Use

The standard runoff assay is typically used to titrate GTFs. It is also useful for assessing the effects of inhibitors on overall transcription and for cleanly demonstrating the dependence of the assay on particular factors. The total signal in this assay is less than that in the pausing assay because the ratio of cold CTP to hot CTP is higher throughout the run-on, allowing for cold CTP to compete with hot CTP through the duration of transcription rather than merely after the cold CTP chase.

Buffers and Reagents

DB(100): 10% glycerol, 10 mM Tris pH 7.9, 180 mM KCl, 1 mM DTT

Template Mix: 5 nM template DNA, 20 mM HEPES pH 7.6, 1 mM DTT, 8 mM $MgCl_2$, plus 0.2 uL RNase OUT/reaction (activator may be added to template mix after being appropriately titrated)

Stock NTP Mix: 100 mM ATP, GTP, and UTP; 5 mM CTP

NTP Mix with Hot CTP: one Stock NTP Mix, two parts DB(100), and one part alpha-labeled ^{32}P-CTP

Stop Buffer: 20 mM EDTA, 200 mM NaCl, 1% SDS

Formamide Loading Buffer (FLB): 950 uL formamide, 5 uL 0.5 M EDTA, and 130 uL 10 mg/mL bromophenol blue, 130 uL 10 mg/ml xylene cyanol, and 130 uL 10 mg/mL amaranth; prior to use, add 1 part 1M NaOH to 29 parts FLB.

Protocol

Heat water bath to 30°C. Mix template mix and the appropriate volume of DB(100) so that the final volume of the transcription reaction is 20 uL into 1.6 mL low-retention

tubes. Incubate at 30°C for 15 minutes (approximately the amount of time to prepare the PIC Mix). Prepare PIC Mix (some combination of IIA, IIB, IID, IIE, IIF, IIH, pol II, and Mediator depending on experiment) on ice. Remove template mix from incubator. At room temperature, add PIC mix to the template mix. Incubate at 30°C for 15 minutes. Prepare Stop Buffer by adding 1 uL 20 mg/mL glycoblue and 5 uL 7.5 M ammonium acetate to 150 uL Stop Buffer per reaction. Initiate transcription with 2 uL NTP Mix with Hot CTP per reaction. Allow reactions to proceed for 30 minutes before stopping with 150 uL prepared Stop Buffer. Add 450 uL cold 100% ethanol to each reaction. Mix vigorously and allow to precipitate for as little as 30 minutes and as much as overnight at -20°C. Pellet by centrifugation at 4°C for 20 minutes. Remove ethanol. Wash with 200 uL cold 75% ethanol. Remove ethanol. Dry pellet using speedvac for 5 minutes. Resuspend pellet via vigorous pipetting in 6 uL FLB. Boil pellet for 5 minutes at 90°C. Load onto sequencing gel (percentage of gel is dependent on purpose of assay).

C. PAUSING ASSAY

When to Use

The pausing assay has uses that are largely redundant with the pausing assay with pulse-chase. It is an assay that can assess changes in initiation, promoter escape, and pausing upon addition of a certain factor, but is less capable of assessing elongation and is complicated by exacerbated effects of nucleotide starvation. For these reasons, I recommend that this assay is only used when assessing short time points post-initiation (i.e. <5 minutes, preferably <2 minutes).

Buffers and Reagents

Make the same buffers as described for standard run-on assay, except for these changes to the Stock NTP Mix and NTP Mix with Hot CTP:

Stock NTP Mix (-CTP): 100 mM ATP, GTP, and UTP

NTP Mix with Hot CTP: one Stock NTP Mix (-CTP), two parts DB(100), and one part alpha-labeled ^{32}P-CTP

Protocol

Perform assay as described for the standard run-on assay, except that the NTP Mix is replaced with NTP Mix (-CTP). Additionally, elongation time should be reduced to 5 minutes or less. The RNA isolated from these assays is typically run on an 18% gel.

D. PULSE-CHASE PAUSING ASSAY

When to Use

This assay is the workhorse assay for measuring pausing and elongation. By almost exclusively using hot CTP in the pulse (or initiation) and then chasing with cold CTP, the pulse-chase pausing assay allows for approximation of single round transcription, visualization of almost all transcripts (including paused transcripts), and observation of elongation products. This assay is more dynamic than the similar assay without the cold CTP chase: time courses completed using this regime demonstrate pause release into elongation whereas the pulse only assay shows very little elongation, likely due to nucleotide starvation.

Buffers and Reagents

Make the same buffers as described for the pausing assay, except for the addition of the Cold CTP Chase:

Cold CTP Chase: 100 uM CTP in DB(100)

Protocol

Perform assay as described for standard run-on assay, except that the NTP Mix is replaced with NTP Mix (-CTP). One minute after initiation with NTP Mix with Hot CTP, add 2 uL of Cold CTP Chase per reaction. Allow reaction to proceed for an additional 9 minutes. Stop reactions and work-up as described in the standard run-on protocol.

E. PRIMER EXTENSION ON LINEAR TEMPLATE

When to Use

The primer extension assay is used to determine the transcription start site of a particular template. Furthermore, it also can be used to infer the presence or absence of RNA secondary structures.

Buffers and Reagents

Make the same buffers as described for the standard run-on assay, except for this change to the Stock NTP Mix:

Stock NTP Mix (Full NTPs): 100 mM ATP, GTP, UTP, and CTP

Protocol

Transcription assay is carried out as described in the standard run-on assay except that to initiate transcription, 2 uL of an NTP Mix containing 100 mM each ATP, GTP, UTP and CTP is added. Reactions are allowed to proceed for 30 minutes. After stopping the

reactions with Stop Buffer, as described in the standard run-on assay protocol RNA was then phenol:chloroform extracted prior to ethanol precipitation and RNA work-up. After RNA isolation, primer extension is carried out using AMV reverse transcriptase as described (https://www.promega.com/products/pcr/rt-pcr/amv-reverse-transcriptase/?catNum=M5101) at 45°C-55°C. This particular RT is thermostable and is capable of transcribing through RNA secondary structure, a particularly important feature when dealing with the highly structured HSPA1B RNA.

F. BIOTIN-IMMUNOPRECIPITATION OF ^{32}P-RADIOLABELED TRANSCRIPTS

When to Use

The biotin-IP protocol of radiolabeled transcripts is used to determine a portion of the sequence of transcripts. Essentially, one can use the protocol to ensure that all transcripts being assessed originate from a particular tss or if a transcript contains a particular sequence. The strength of conclusions drawn from this assay are entirely dependent on the specificity of the biotinylated oligo binding to its complementary RNA transcript.

In troubleshooting this assay, we struggled with annealing the highly folded HSPA1B transcripts to the biotinylated oligo. By incorporating a few tricks used by RNA structural biologists (addition of guanidinium hydrochloride to the assay and designing a probe with a high melting temperature) we were able to gain information on start sites and identity of products from this assay, although it could certainly be further adapted.

This assay could be used in the future to look at transcription of eRNAs *in vitro* or perhaps for assays that used nuclear extracts as the GTF source rather than purified proteins, as was done in Qiu and Gilmour 2017[89].

Buffers and Reagents

Biotin Oligo Buffer: 30 mM Tris pH 7.5, 500 mM NaCl, 1.6 pmol biotinylated oligo/reaction (added immediately prior to use from frozen stocks)

4M GnHCl

Wash Buffer: 10 mM NaCl, 10 mM Tris HCl pH 7.5, 5 mM EDTA, 0.1 mg/mL yeast tRNA (or total RNA, added immediately prior to use from frozen stocks)

Dynabeads M-280 Streptavidin Beads

Protocol

Stop transcription with Stop Buffer as usual. Phenol:chloroform extract reactions, transferring aqueous phase to a new 500 uL low-retention tube. Prepare Biotin Oligo Buffer for use by adding 68.8 uL Biotin Oligo Buffer and 31.2 uL 4M GnHCl per reaction together. Add 100 uL of this solution to each transcription reaction. Anneal using thermocycler with some variation of the following program, dependent on the particular melting T_m of the biotinylated oligo for the radiolabeled transcripts: 95° for 5 min, 80° for 10 min, 80° to 40° over 40 min, 40° for 20 min, 25° for 45 min. Add annealed RNA-Biotin Oligo solution to 150 ug washed Dynabeads M-280 Streptavidin, prepared in a 1.6 mL low-retention tube; the Dynabeads have been washed with the Wash Buffer described above. Nutate at room temperature for 10 minutes. Wash beads twice on magnetic stand with 300 uL Wash Buffer, allowing beads to sit in Wash Buffer for one minute prior to gently pipetting off. Dry beads in speedvac for 5 minutes. Elute radiolabeled RNA off biotinylated

oligo-beads by resuspending beads in 10 uL FLB and boiling for 5 minutes. Load onto sequencing gel for analysis.

G. Labeling ϕX174 Ladder w/ Hot ATP

Mix 6 uL ladder, 1 uL PNK buffer, 1 uL T4 PNK, and 2 uL hot ATP for a total volume of 10 uL. Incubate at 37° C for 30 minutes. Run through a G-25 column at ~735*g (low speed). In FLB, make 1/5, 1/10, 1/50, 1/100, and 1/1000 dilutions. Heat for three minutes at 95°C.

Appendix II: Purification Protocols

A. pol II

Pol II is purified from HeLa nuclear extract. One can either combine nuclear extract from previous preparations or make some for the express purpose of purifying pol II. In this section, I will not describe how to generate nuclear extract, as it is explicitly described in general lab protocols.

Prepare the following buffers and reagents:

- 4 M ammonium sulfate
- 0.1M $ZnCl_2$
- 1L 0 M AS BUFFER R: 200 mL 50% glycerol, 25 mL 2M Tris pH 7.9, 2 mL 0.5 M EDTA pH 8, 100 uL 0.1 M $ZnCl_2$, 2.5 mL 10% NP-40 alternative
- 1L 0.7 M AS Buffer R: 200 mL 50% glycerol, 25 mL 2 M Tris pH 7.9, 2 mL 0.5 M EDTA pH 8, 100 uL 0.1 M $ZnCl_2$, 2.5 mL 10% NP-40 alternative, 175 mL 4 M ammonium sulfate

FILTER ALL BUFFERS PRIOR TO USE; Combine 0 M AS Buffer R and 1 M AS Buffer R appropriately to reflect various molarities of ammonium sulfate.

Washed two aliquots of 800 uL Protein G sepharose beads (1600 uL 50% slurry) in 1.5 mL tubes with PBS twice. Centrifuged two 1600 uL anti-8WG16 antibody at 14K rpm at 4°C for 30 minutes. Split Protein G sepharose into 4X 1.5 mL low retention tubes. Added 800 uL anti-8WG16 supernatant to beads. Mixed for 4 hours on nutator at 4°C. Meanwhile, thawed 40 mL HeLa nuclear extract (NE). Added 80 uL 4M ammonium sulfate and 40 uL

benzonase to NE. Nutated at 4°C for 2 hours then spun down at 14K rpm for 30 minutes. Collected clarified NE and kept on ice. Washed immobilized anti-8WG16 beads twice with 1 mL 0.1 M AS Buffer R. Divided NE between 4X 15 mL falcon tubes (10 mL each). Divided immobilized anti-8WG16 beads between the 4 falcon tubes. Nutate at 4°C O/N.

Centrifuged at 4°C and 800 rpm on floor centrifuge. Collected flowthrough. Washed 5X with 1 mL 0.5 M AS Buffer R. Washed 2X with 1 mL 0.15 M AS Buffer R. Combined immobilized anti-8WG16 beads into 2X 1.6 mL low retention tubes. Prepared 5 mL of 2 mg/mL elution peptide (3X YSPTSPS) in 0.15 M AS Buffer R+100 mM Tris pH 7.9. Eluted each sample 3X with 800 at 4°C for 40 minutes on nutator. Sampled each elution and combined for a total of 4.8 mL eluate. Ran through centrifugal filter to remove any residual beads.

Washed and equilibrated UNO-Q column with 0.15 M AS Buffer R. Injected and loaded sample onto column. Washed column thoroughly with 0.15 M AS Buffer R. Ran an elution gradient from 0.15 M AS Buffer R to 0.7 M AS Buffer R over 5 CV. Collected 0.2 mL fractions. *

*The UV detector on the AKTA effectively detect pol II due to amino acid composition or experiences interference due to the composition of the AS Buffer regime. Therefore, it is essential to test each fraction by Bradford to determine where pol II is eluting. Fractions that appear to contain protein should be analyzed by SDS-PAGE.

Additional Note: If desired, a heparin column may be ran prior to the UNO-Q column. The column is run similarly to the UNO-Q column. Pol II that is run over both columns is slightly purer and contains less GDOWN1. It is, however, less transcriptionally active.

B. TFIIE34

TFIIE34 is expressed in E. coli via transformation into DE3 BL21 pLysS cells. Importantly, TFIIE34 is purportedly 6X-His tagged but does NOT bind Ni-NTA resin well despite a variety of attempts at doing so. However, it is simple to purify via DEAE column followed by a 1 mL SPFF column.

TFIIE34 was transformed into DE3 BL21 pLysS cells for protein expression and selected with ampicillin. 10 mL O/N cultures were grown from selected colonies. The following day, 1 L of LB was inoculated with the O/N culture and allowed to reach an OD_{600} of 0.4. The temperature of the culture was then dropped to 16°C. Protein expression was induced by addition of IPTG (final concentration=1 mM). Expression continued O/N at 16°C. Cells were spun down at 5K rpm for 30 minutes at 4°C on the GS3 rotor (note: protein expression may also be induced with 1 mM IPTG at OD_{600} of 0.4 followed by incubation at 30°C for 2-3 hours). Cells were then resuspended in 20 mL of H/E buffer and sonicated 3X for 30 seconds (1 second on followed by one second off 15X times). If at this point the lysate is too viscous, additional H/E Buffer may be added as well as 1 uL Benzonase per mL of lysate followed by a 30-minute nutation at 4°C. 0M HEMG is then added to bring the salt concentration to 300 mM. The sample is then loaded onto 45 mL of DE52 resin that has been prepared, pre-equilibrated with 0.3 M HEMG, and packed into an appropriate column. The flowthrough from the column is then collected, diluted to 0.1M HEMG, and then loaded onto a pre-equilibrated 1 mL SPFF column. After washing thoroughly with 0.1M HEMG, the protein is eluted via a step-gradient: 0.2 M HEMG, 0.3M HEMG, 0.4 M HEMG, and 0.5 M HEMG. Fractions should be collected from each sample and analyzed by SDS-PAGE.

We observed that well-purified E34 eluted in the 0.4 M fraction as well as slightly less pure material in the 0.3 M fraction. Approximately 15 mL of useable material can be found in the IIE34 box in the -80 to be used in transcription reactions at 0.5 uL per reaction after being diluted 1:125.

C. CDK8 Module

The CDK8 module is expressed in Sf9 cells. We order the infection and expression of the module from the University of Colorado Protein Production, Monoclonal Antibody, and Tissue Culture Shared Resource Center (PPSR). CDK8 is tagged with a Glu-tag. Any variation on the 4-WT module may be generated and purified as follows so long as CDK8 interactions are unaffected.

Resuspended cells in 3.75 mLs cold WCE Buffer (50 mM HEPES pH7, 150 mM NaCl, 0.1% NP-40) with protease inhibitors. Thawed and resuspended pellet in WCE Buffer via incubation in a cold-water bath. Dounced the homogenized mixture in a pre-chilled douncer with pestle B 25 times. Combined lysates. Spun down at 14K in 1.6 mL tubes at 4°C. Pulled off supernatant and combined.*

Meanwhile, washed 200 uL of Protein G beads (400 uL 50% slurry) 2X with 0.15 M HEGN. Added an additional 100 uL of 0.15 M HEGN after last wash as well as 120 uL Covance anti-Glu antibody. Nutated for 2 hours at 4°C. Washed beads 3X with 0.15 M HEGN. Divided beads and lysate equally among 1.6 mL low retention tubes (note: could likely be done in a single 15 mL conical). Nutated for 3 hours at 4°C. Washed 4X with 0.5 M HEGN and 2X with 0.15 M HEGN, combining beads into two tubes after the first wash. Eluted each sample twice with 1X CV (100 uL 2X per sample) 1 mg/mL Glu peptide

resuspended in 0.15 M HEGN. Combined elutions from each separate sample (200 uL total elution per sample, kept separately). Loaded onto a 15-40% glycerol gradient and ran 4°C for 6 hours at 55K rpm. Collected 100 uL samples and analyzed via SDS-PAGE.

*At this step, one can choose to further clean-up the end purification by treating with 1 uL benzonaze per mL of lysate for 1 hour at 4°C, dialyzing the lysate against 0.1 M HEMG and then running over a Heparin FF column in order to further clean-up the end purification. The column is equilibrated in 0.1 M HEMG, loaded with lysate, and then eluted with a step gradient of 0.3M HEMG followed by 0.5 M HEMG. The 0.3 M HEMG elution is most enriched for the CDK8 module and can be used as the input material for the anti-Glu IP.

D. P-TEFb

P-TEFb is purified from Sf9 cells that are generated and ordered from the University of Colorado Protein Production, Monoclonal Antibody, and Tissue Culture Shared Resource Center (PPSR). It is composed of subunits CDK9 and CCNT1.

Prepare the following buffers and reagents:

- 200 mL 1 M imidazole, pH to 7.6
- 10 mL Lysis Buffer: 2.5 mL B-PER II, 25 uL 1 M imidazole, 150 uL 5 M NaCl, 250 uL 10% Triton-X 100, 10 uL 1 M $MgCl_2$
- 20 mL 0.5 M Lysis Buffer: 100 uL 1 M imidazole, 2 mL 5 M NaCl, 200 uL 2 M Tris pH 7.9, 2 mL 10% Triton-X 100, 40 uL 1 $MgCl_2$
- 20 mL 0.5 M Wash Buffer: 100 uL 1 M imidazole, 2 mL 5 M NaCl, 200 uL 2 M Tris pH 7.9, 2 mL 10% Triton-X 100

- 20 mL 1.0 M Wash Buffer: 500 uL 1 M imidazole, 4 mL 5 M NaCl, 200 uL 2 M Tris pH 7.9, 2 mL 10% Triton-X 100
- 20 mL 70 HGKE Wash Buffer: 500 uL 1 M imidazole, 468 uL 3 M KCl, 500 uL 1 M HEPES pH 7.6, 6 mL 50% glycerol, 4 uL 0.5 M EDTA, 2 mL 10% Triton-X 100
- 10 mL 300 mM Imidazole Elution Buffer: 3 mL 1 M imidazole, 234 uL 3 M KCl, 250 uL 1 M HEPES pH 7.6, 3 mL 50% glycerol, 2 uL 0.5 M EDTA, 1 mL 10% Triton-X 100

Filter all buffers prior to use and add protease inhibitors.

Added 8 mL of Lysis Buffer to an insect cell pellet generated from 500 mL of Sf9 cells. Thawed cells in cold water bath, resuspended by vortexing, and dounced cells with pre-chilled pestle B 25 times. Added 11 uL 5 M NaCl and 8 uL benzonase to lysate. Nutated at 4°C for 2 hours. Centrifuged lysate at 15K rpm for 30 minutes at 4°C. Washed 2 mL of Ni-NTA resin (4 mL 50% slurry) with milliQ H_2O followed by 0.5 Lysis Buffer. Packed column with Ni-NTA resin and further equilibrated with 0.5 Lysis Buffer. Added cleared lysate to column. Washed column with 10 CV 0.5 Wash Buffer, followed by 10 CV 1.0 Wash Buffer. Performed one additional 10 CV wash with 70 HGKE Wash Buffer. Eluted with 10 mL 300 mM Imidazole Elution Buffer, collecting 0.5 mL fractions. Analyzed via SDS-PAGE.

E. DSIF

DSIF is expressed in E. coli via transformation of a single plasmid encoding both Spt4 and Spt5. Spt4 is His-tagged at the N-terminus. This plasmid is a gift from Miriam Sanso, formerly of the Fisher Lab at the Icahn School of Medicine at Mount Sinai.

The DSIF plasmid was transformed into DE3 BL21 cells for protein expression and selected with ampicillin. A 10 mL O/N culture was grown from a selected colony. The

following day, 1 L of LB was inoculated with the O/N culture and allowed to reach an OD_{600} of 0.6. The temperature of the culture was then dropped to 16°C. Protein expression was induced by addition of IPTG (final concentration=0.5 mM). Expression continued O/N at 16°C. Cells were spun down at 5K rpm for 30 minutes at 4°C on the GS3 rotor. Cells were resuspended in 20 mL of Lysis Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole adjusted to pH 8) and 1 mL B-PER. Cells were sonicated 3X for 30 seconds (1 second on followed by one second off 15X times). Lysate was then clarified by a 30 minute spin at 15K rpm at 4°C. 500 uL Ni-NTA resin (1 mL of 50% slurry) was then washed with milliQ H_2O and then Lysis Buffer. Clarified Lysate and washed and equilibrated Ni-NTA resin were then added to a 50 mL conical and nutated at 4°C for 1 hour. Resin was then pelleted via centrifugation at 800 rpm in floor centrifuge at 4°C. Resin was washed with 3X with 10 CVs Wash Buffer (50 mM NaH_2PO_4, 1 M NaCl, 100 mM imidazole adjusted to pH 8). Resin was transferred to a 1.6 mL low retention tube and 3X 500 uL elutions were performed with Elution Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole adjusted to pH 8). Elutions were collected, kept separate, and individually ran through centrifugal filter to remove any residual beads.

F. NELF

NELF is expressed in E. coli via transformation of two plasmids: one contains NELFA and NELFB with the other containing NELC and NELF E. NELFA is FLAG-tagged at the C-terminus; NELFB is Sumo-tagged at the N-terminus; NELFC is 10X His-tagged at the N-terminus. These plasmids are a gift from Brian Gibson, formerly of the Kraus Lab at UT-Southwestern.

The two NELF plasmids were transformed into DE3 BL21 pLysS cells for protein expression and selected via chloramphenicol, ampicillin, and kanamycin. 10 mL O/N cultures were grown from selected colonies. The following day, 1 L of LB was inoculated with the O/N culture and allowed to reach an OD_{600} of 0.2. The temperature of the culture was then dropped to 16°C. Protein expression was induced by addition of IPTG (final concentration=0.5 mM). Expression continued O/N at 16°C. Cells were spun down at 5K rpm for 30 minutes at 4°C on the GS3 rotor. Cells were resuspended in 20 mL of Lysis Buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% NP-40, 5 mM BME, 1 mM PMSF, 10 mM imidazole), incubated for 30 minutes at 4°C, and sonicated 3X for 30 seconds (1 second on followed by one second off 15X times). Lysate was then clarified by a 30 minute spin at 15K rpm at 4°C. Clarified lysate was then flowed over a pre-equilibrated and packed 2 mL Ni-NTA column. Column was washed with 30 CV Wash Buffer (50 mM Tris pH 7.5, 500 mM NaCl, 5 mM BME, 1 mM PMSF, and 10 mM imidazole). Protein was eluted with 5 CV Elution Buffer (20 mM Tris pH 7.5, 150 mM NaCl, 1 mM BME, 1 mM PMSF, 500 mM imidazole). 0.5 mL fractions should be collected, sampled, and pooled accordingly to the results of SDS-PAGE analysis.

Depending on the purity of the NELF obtained after the Ni-NTA column, a further purification step may be desired. If so, dialyze the sample against 0.1 M HEMG, and run over a 1 mL Heparin FF column. Elute with a 0.1 to 1.0 M HEMG gradient, collecting fractions throughout. Sample elution fractions and determine the peak fraction via SDS-PAGE analysis.

G. Activators: HSF1, Myc, and Max

HSF1, Myc, and Max are all tagged with a 6X His-tags. They were independently transformed into DE3 BL21 cells and selected via ampicillin. 10 mL O/N cultures were grown from selected colonies. The following day, 1 L of LB was inoculated with the O/N culture and allowed to reach an OD_{600} of 0.4. Protein expression was induced by addition of IPTG (final concentration=1 mM). Expression continued for 3 hours at 30°C. Cells were spun down at 5K rpm for 30 minutes at 4°C on the GS3 rotor. Cells were resuspended in 20 mL of Lysis Buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% NP-40, 5 mM BME, 1 mM PMSF, 10 mM imidazole), incubated for 30 minutes at 4°C, and sonicated 3X for 30 seconds (1 second on followed by one second off 15X times). Lysate was then clarified by a 30-minute spin at 15K rpm at 4°C. 2 mL of Ni-NTA (4 mL of 50 % slurry) was washed with water and then Lysis Buffer. 10 mL of clarified lysate was then added to the Ni-NTA resin in a 15 mL falcon tube and allowed to nutate at 4°C for 1 hour. Resin was then spun down at 800 rpm in floor centrifuge at 4°C. Resin was then washed 3X with 10 CV Wash Buffer (50 mM Tris pH 7.5, 500 mM NaCl, 5 mM BME, 1 mM PMSF, and 50 mM imidazole). Protein was then eluted with 5X with 1 CV Elution Buffer (20 mM Tris pH 7.5, 150 mM NaCl, 1 mM BME, 1 mM PMSF, 500 mM imidazole).

H. TFIIS

TFIIS is purified by the same protocol as the Activators.

I. TAF1 C413 Antibody

The antibody plasmid was purchased from the Recombinant Antibody Network (https://recombinant-antibodies.org. The Anti-TAF1 expression plasmid was transformed

into OverExpress C43(DE3) chemically competent cells and expressed according to standard protocols. Protein was then isolated via batch purification over Protein A beads according to the Recombinant Antibody Network (RAN) protocol (**https://recombinant-antibodies.org/protocols/psfv4-avi**).

Appendix III: Cell-Based Experimental Conditions and Analysis Protocols

A. TRIM-Away

Cell culture. Cells were grown in McCoy's media (Gibco, 16600082) with Gibco 100x Antibiotic-Antimycotic (Fisher Sci, 15240062) penicillin-streptomycin and 10% fetal bovine serum (FBS) supplementation.

Cell preparation and transfection. The method used was adapted from [Clift/Schuh Cell 2017]. HCT116 cells cultured in McCoy's 5A medium were grown to approximately 70% confluency. Media was aspirated off, and the cells were washed with PBS. 2ml of trypsin per plate were used to harvest adherent cells, after which an equal volume of Opti-MEM was added to each plate to neutralize the trypsin. Cells were combined in a 50ml centrifuge tube and spun down at 2,000xg for 5 minutes, then washed in PBS and spun down again at 2,000xg for 5 minutes. Cells were counted using a hemocytometer and diluted to 25 million cells/mL. 100 µl reactions were prepared, and cells were re-suspended in Buffer R and anti-TAF1 C413 antibody. A pulse only control was prepared, which consisted of cells suspended only in Buffer R. Transfections were performed using the Neon Transfection Kit (1530V, 1ms width, 1 pulse). Transfected cells were then pipetted into 1 mL of Opti-MEM in a 35mm dish and incubated at 37˚C for 1 hour. The Opti-Mem media (containing some suspended cells) was then pipetted off and saved. 500 µl of PBS was added to the cells on the plates, which were then harvested and centrifuged at 6,000xg for 5 minutes. Supernatant was aspirated off, and cell nuclei were

subsequently isolated. A small sample of cells (50 µl) were saved for analysis via western blot.

Western blot antibodies. TAF1 (1:1000, sc-735 X, Santa Cruz Biotechnology), TAF2 (1:500, ab103468, abcam), TAF4 (1:250, 07-1803, Millipore Sigma), TAF8 (1:250, ab204894, abcam), and TBP (1:2000, sc-273, Santa Cruz Biotechnology), actin (1:1000, sc-47778, Santa Cruz Biotechnology). Antibodies against *Drosophila* proteins were monoclonals 30H9 (Taf1)[Weinzierl Nature 1993 511] and 3E12 (Taf4)[Marr G&D 2006 1458].

B. TAF1 KNOCKDOWN IN DROSOPHILA S2 CELLS (Courtesy of Marr lab, Brandeiss University)

Drosophila cell culture and RNAi. *D. melanogaster* Schneider line 2 (S2) cells were maintained at 25°C in Schneider's medium containing 10% (vol/vol) Fetalplex (Gemini), 100 units/mL penicillin, and 0.1 mg/mL streptomycin. RNAi was performed as described [Clemens, 2000] using 20-40 µg dsRNA. Cells were incubated with dsRNA for 2.5 d.

Drosophila S2 nuclei isolation. Following RNAi with either TAF1 dsRNA or a LacI dsRNA control, cells were processed using the nuclei isolation steps as described [Mahat BD, 2016] before flash-freezing and storing at -80ºC.

Measuring TAF1 knockdown from S2 cells. Following RNAi treatment samples were run on a SurePAGE Bis-Tris 4-12% gel (GenScript) at 200V for 70 min. Protein was then transferred onto a nitrocellulose membrane (80V for 2hrs). For imaging and quantitation, membranes were exposed for sub-saturated times (BIO-RAD Chemidoc MP).

C. PRO-SEQ

Nuclei preparation. After treatment, HCT116 cells (control or TAF1 TRIM-Away) were washed 3x with ice cold PBS, and then treated with 10 ml (per 15 cm plate) ice-cold lysis buffer (10 mM Tris–HCl pH 7.4, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 0.5% NP-40, 10% glycerol, 1 mM DTT, 1x Protease Inhibitors (1mM Benzamidine (Sigma B6506-100G), 1mM Sodium Metabisulfite (Sigma 255556-100G), 0.25mM Phenylmethylsulfonyl Fluoride (American Bioanalytical AB01620), and 4U/mL SUPERase-In). Cells were centrifuged with a fixed-angle rotor at 1000×*g* for 15 min at 4°C. Supernatant was removed and pellet was resuspended in 1.5 mL lysis buffer to a homogenous mixture by pipetting 20-30X before adding another 8.5 mL lysis buffer. Suspension was centrifuged with a fixed-angle rotor at 1000×*g* for 15 min at 4°C. Supernatant was removed and pellet was resuspended in 1 mL of lysis buffer and transferred to a 1.7 mL pre-lubricated tube (Costar cat. No. 3207). Suspensions were then pelleted in a microcentrifuge at 1000×*g* for 5 min at 4°C. Next, supernatant was removed and pellets were resuspended in 500 µL of freezing buffer (50 mM Tris pH 8.3, 40% glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA, 4U/ml SUPERase-In). Nuclei were centrifuged 2000×*g* for 2 min at 4°C. Pellets were resuspended in 100 µL freezing buffer. To determine concentration, nuclei were counted from 1 µL of suspension and freezing buffer was added to generate 100 µL aliquots of 10×10^6 nuclei. Aliquots were flash frozen in liquid nitrogen and stored at −80°C.

Nuclear run-on and RNA preparation. Nuclear run-on experiments (HCT116 and S2 cells) were performed as described with the following modifications: the final concentration of non-biotinylated CTP was raised from 0.25 µM to 25 µM, and the final library clean-up and size selection was accomplished using 1X AMPure XP beads (Beckman).

D. SEQUENCING DATA PROCESSING

Sequencing. Sequencing was performed at the BioFrontiers Sequencing Facility (UC-Boulder). Single-end fragment libraries (75 bp) were sequenced on the Illumina NextSeq 500 platform (RTA version: 2.4.11, Instrument ID: NB501447), demultiplexed and converted BCL to fastq format using bcl2fastq (bcl2fastq v2.20.0.422); sequencing data quality was assessed using FASTQC (v0.11.5) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and FastQ Screen (v0.11.0, https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/). Trimming and filtering of low-quality reads was performed using BBDUK from BBTools (v37.99) and FASTQ-MCF from EAUtils (v1.05). Alignment to the human reference genome (GRCh37/hg19) was carried out using Hisat2 (v2.1.0) in unpaired, no-spliced-alignment mode with a GRCh37/hg19 index, and alignments were sorted and filtered for mapping quality (MAPQ>10) using Samtools (v1.5). Gene-level count data for transcription start site (TSS, -30 to +300) and gene body (+301 to end) regions were obtained using featureCounts from the Subread package (v1.6.2) with custom annotation files for single unique TSS and gene body regions per gene. Custom annotation files with single unique TSS and gene body regions per gene were generated as follows: 1) hg19 RefSeqCurated transcript-level annotation was downloaded from the UCSC genome table browser (09-07-2018), transcripts shorter than 1500bp and non-standard chromosome were removed, and only transcripts with unique start/stop coordinates per gene were retained; 2) Sense and anti-sense counts were tabulated and each candidate TSS region was ranked by sense and antisense reads to obtain a single 'most-active' TSS per gene; 3) Finally, per gene, the TSS was combined with the shortest gene body to avoid the influence of alternative

transcription termination/polyadenylation sites. Analysis of pol II pausing was carried out using a custom R script (R v3.5.1 / RStudio v1.1.453) with the ggplot2 package (v3.1.0) used for visualizations. Gene level TSS and gene body counts were normalized by counts-per-million and by region length (cpm/bp) and Pausing Index (PI) calculated as the ratio of normalized reads in the TSS (cpm/bp) to normalized reads in the gene body (cpm/bp). Genes with <0.5 cpm in all samples were excluded from analysis. Means of replicate values were used for plots and Wilcoxon/Mann-Whitney U tests. For genome browser snapshots, aligned reads were downsampled to the lower aligned read count per replicate using Samtools, to ensure equal contributions from each replicate, followed by merging of replicates and generation of coverage tracks in the bedgraph format using HOMER (v4.9.1). Genome browser snapshots were then generated from the bedgraph files using a custom R script (R v3.5.1 / RStudio v1.1.453 / Bioconductor v3.7) and the Gviz package (v1.26.4).

Sequencing data processing. The initial processing of all sequencing data was performed using the NascentFlow Pipeline (doi: 10.17605/OSF.IO/NDHJ2), a data processing pipeline written in the Groovy programming language. The code for this pipeline can be found at https://github.com/Dowell-Lab/Nascent-Flow, with analysis for this experiment performed at commit 3fe1b7. Data were mapped to the hg38 reference genome for human cells, and to the dm6 reference genome for *Drosophila* S2 cells. For the remainder of the analysis, only the maximally expressed isoform of each gene was considered, which was determined by calculating the RPKM normalized expression over each isoform and selecting the one with the maximum RPKM expression. When different isoforms were determined across samples, the isoform from the first control sample was

selected. In HCT116 cells, this was sample PO_1_S1_R1_001 whereas in S2 cells this was sample Control_1_S1_R1_001.

Pause index calculation. Pause indices were calculated using a fixed-window approach. From -30 to +300 base pairs around the annotated transcription start site (TSS) was defined as the paused region, and from +301kb to the annotated polyA site was defined as the elongation region. Pause index was calculated as the ratio of length-normalized reads in the paused region to length-normalized reads in the elongation region. P-value determination for pause indices was performed by comparing the distribution of pause indices between control and knockdown samples using a Kolomogorov-Smirnov (KS) test, with the built-in implementation of the test in the R programming language. Subsets of genes containing promoter elements were found by searching across the reference sequence of each gene, for promoter elements in their expected positions relative to the TSS. The following motifs were used for each promoter element: TATA-like: WWWW; Initiator: BBCABW; Motif Ten Element: CGANC….CGG; Downstream Promoter Element: RGWYVT; GAGA Element: NVNVMGNRMR.

Metagene analysis. Each gene in the isoform-resolved reference sequence was divided into a fixed number of bins, and utility featurecounts[12] was used to determine the total counts in those regions. The mean count and standard deviation of the mean were calculated, and all bins were then plotted along with the standard deviation.

Principal component analysis. PCA was performed using the standard prcomp function provided by the sva package for the R programming language[13]. Batch effects from replicates completed on different days replicates were corrected using the

removeBatchEffect function provided by the limma package from the R programming language.

Differential expression analysis. Differential expression analysis was performed using the DESeq2 package for the R programming language. Counts were generated using the utility featurecounts. Initial analysis using counts across the full annotated gene showed significant skew, indicating that the baseline assumptions of the differential expression model did not hold. To correct, counts in the region from +500 of the TSS to -500 from the TES (Transcription End Site) were used to obtain suitable model weights. Those model weights were then used when performing differential expression across the full gene, which corrected the skew effect.

Gene set enrichment analysis. GSEA was performed with the Broad Institute's GSEA software on the GenePattern Server using the pre-ranked module. Log(2) fold-change values were used as the rank metric for all genes and compared against the Hallmark gene sets database for enrichment.

Appendix IV: Additional Data

A. TSS MAPPING

Primer Extension. In addition to the primer extension experiment shown in Chapter II, additional attempts at TSS mapping of in vitro transcription products from the HSPA1B promoter were made. The method is described in Appendix I. These had various degrees of success, likely due to the strong secondary structures predicted to be adopted by the nascent HSPA1B transcript. I only had success with primer extension once I began to use the AMV RT rather than MLV (**Figure 30**). The AMV is marketed as an RT enzyme that transcribes well through secondary structures. Based on my experience, I highly recommend it when analyzing the TSSs of highly paused promoters, and likely G-C rich,

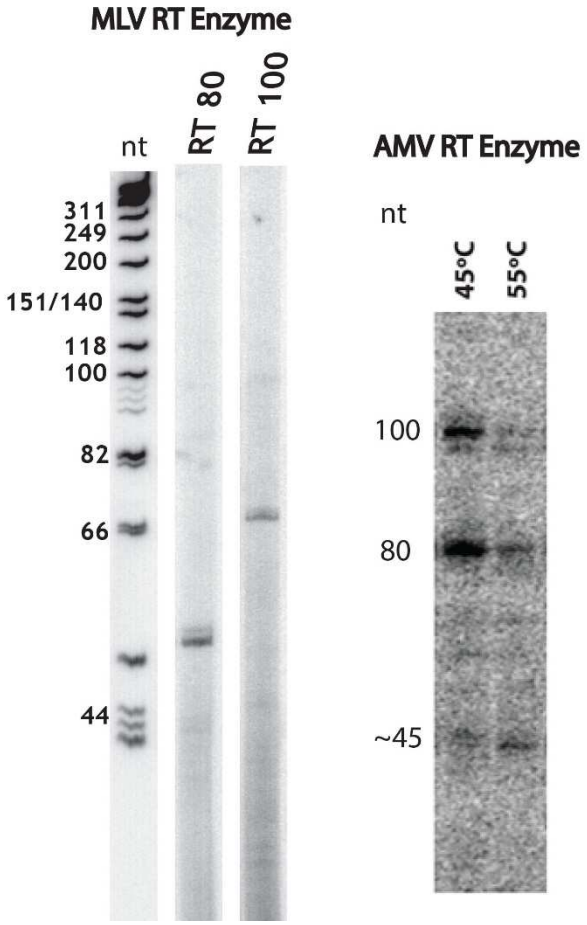


Figure 30: Comparison of primer extension strategies. The AMV RT appears to be superior to MLV enzyme for reverse-transcribing the HSPA1B RNA. The MLV enzyme consistently produces a major product that is ~40 bases shorter than the expected product but would be consistent with in being unable to transcribe through the highly structured HSPA1B RNA 5' predicted motifs. Additionally, the AMV RT enzyme is capable of working at higher temperatures (up to 55°C compared to 37°C for the MLV RT.

promoters. Additionally, increasing the temperature of the primer extension reaction increased the product corresponding due to the annotated TSS. This result is also likely due to the secondary structure of the HSPA1B RNA.

Biotinylated oligo-IP of in vitro transcribed transcripts. I developed an assay to immunoprecipitate hot transcripts via binding to complementary magnetic streptavidin beads. This method is described in Appendix I. The IP yields low total RNA and is still not highly specific. However, it also helps to demonstrate that the runoff transcript produced in the reconstituted transcription system is from the annotated HSPA1B TSS (**Figure 31**).

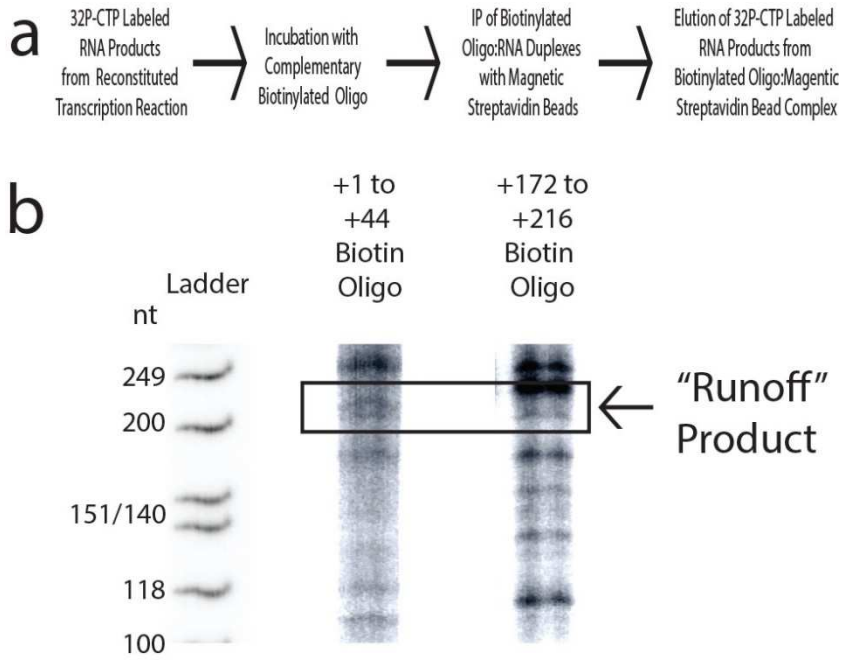


Figure 31: Biotin oligo pulldown of complementary transcripts. a. Schematic of the biotinylated oligo-IP. b. The biotinylated oligo pulldown using either an oligo complementary to the 5' or 3' ends of the nascent HSPA1B transcript immunoprecipitates products of ~216 nucleotides. These products are consistent with transcription from the annotated HSPA1B TSS.

B. THE SUPER CORE PROMOTER IS PAUSE-REGULATED

Given the similarities between the Super Core Promoter (SCP) and the HSPA1B promoter, it came as no surprise that the SCP1 promoter seems to support pausing in the reconstituted pausing assay (**Figure 32**). Most surprising is the apparent similarities in positioning between the SCP1 promoter and HSPA1B. However, if TFIID does indeed enable pausing (as demonstrated in Chapter III) its interactions with downstream promoter elements may play a key role. The HSPA1B promoter and SCP1 promoter share identical positioning of their downstream promoter elements (see Chapter V).

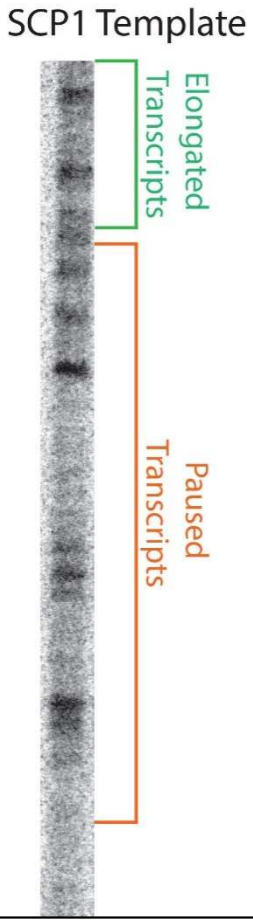


Figure 32: The SCP is pause regulated. Reconstituted transcription was performed on the SCP1 promoter. Pausing patterns are nearly identical to those of the HSPA1B promoter shown throughout this thesis.

C. PRELIMINARY FACTOR TESTS USING THE STANDARD RUNOFF ASSAY

Prior to successful purification of DSIF and NELF (and before the advent of the pausing assay), I tried to reconstitute pausing using the standard runoff assay described in Chapter II. The methods for this assay is thoroughly described in Appendix I. While unsuccessful at detecting paused transcripts, it did provide some interesting data about total transcription under various conditions.

Addition of phosphocellular fractionated nuclear extract impacts transcription.

We typically fractionate HeLa nuclear extract with phosphocellulose (P-cell) resin as a first step in many purification protocols. For instance, Mediator is enriched in the 1.0M salt eluate (P1.0M) of the P-cell resin; TFIIH is enriched in the 0.5M salt elution (P0.5M). We were interested to see if these P-cell fractions might be useful in unbiasedly reconstituting pol II pausing. While these attempts were unsuccessful, we did discover that they affect transcription in interesting, if not entirely explained ways (**Figure 33**).

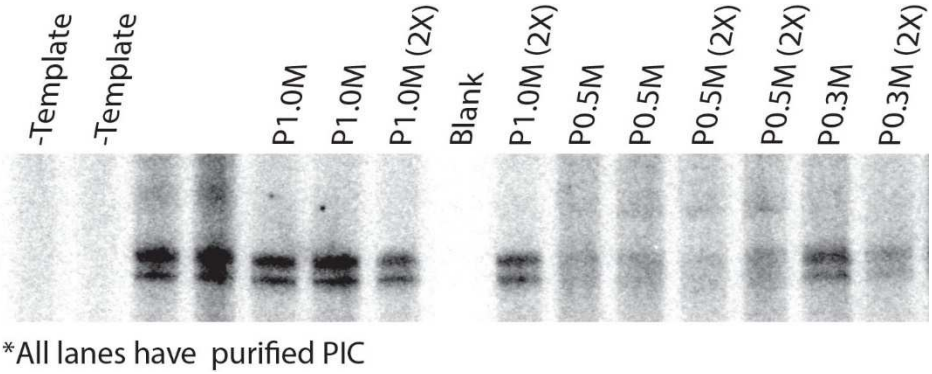


Figure 33: Effects of P-cell fractions on transcription. The addition of P-cell fractions generally seems to inhibit transcription in a titratable manner. This suggests that the P-cell fractions may contain inhibitory regulators of transcription.

CDK8 modules lacking MED13 may activate transcription. Data from the reconstituted pausing assay suggests that the CDK8 module may play a role in promoter escape (Chapter IV). When the CDK8 module was added post-initiation in the standard runoff assay, similar results were observed: an increase in overall transcription (**Figure 5**). We have hypothesized in Chapter IV that CDK8 module may activate promoter escape through association with Mediator. However, in a somewhat surprising result, when a 3-WT version of the CDK8 module was added (lacking MED13, the subunit necessary for Mediator association but still kinase active) transcription was also increased (**Figure 5**). This result suggests that the CDK8 module may activate promoter escape through alternative mechanisms or in addition to the mechanism postulated in Chapter IV.

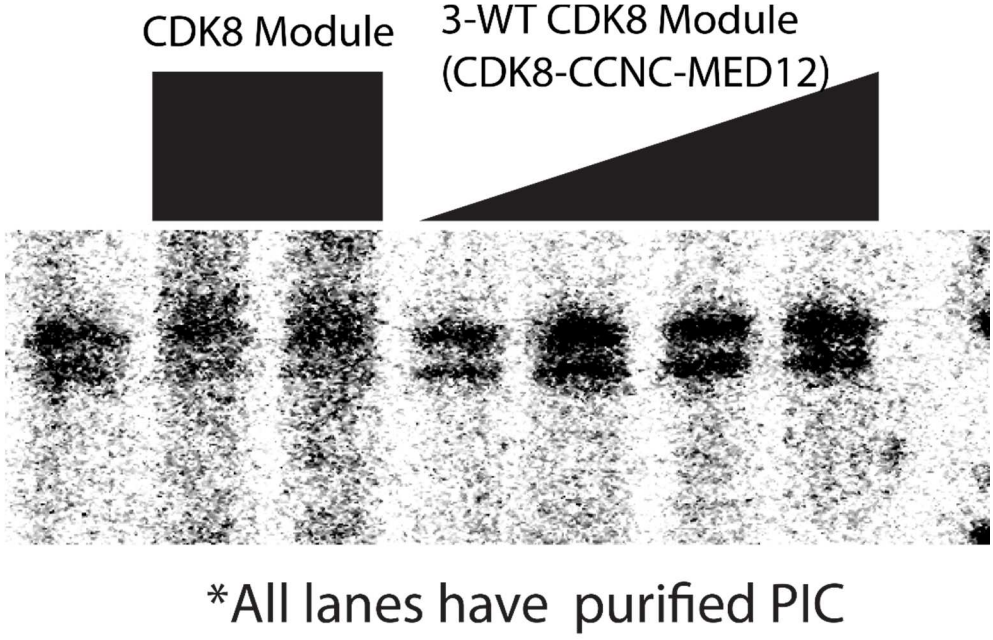


Figure 34: The CDK8 module increases transcription. Both the 4-WT (CDK8, CCNC, MED12, and MED13) and 3-WT (CDK8, CCNC, and MED12) versions of the CDK8 module activate transcription, through yet undescribed mechanisms.