FUNCTIONAL EVOLUTION OF PROMOTER-PROXIMAL PAUSING FACTORS IN THE REGULATION OF RNA POLYMERASE II TRANSCRIPTION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Gregory T. Booth

December 2018

FUNCTIONAL EVOLUTION OF PROMOTER-PROXIMAL PAUSING FACTORS IN THE REGULATION OF RNA POLYMERASE II TRANSCRIPTION

Gregory T. Booth, Ph.D.

Cornell University 2018

Promoter-proximal pausing of RNA Polymerase II (Pol II) is now recognized as a ubiquitous mechanism for regulating gene expression in metazoans. By capturing engaged Pol II shortly after transcription initiation, genes are primed for activation of RNA synthesis, enabling cells to rapidly alter global transcription programs. However, despite conservation of many factors involved in establishing this regulatory platform, many eukaryotes do not control gene expression through this process. Here, the examination of the global transcriptional landscape in two distantly related yeast revealed unprecedented divergence in Pol II distributions across genes. Previously undescribed pause-like profiles were identified within promoter-proximal regions of the fission yeast, *Schizosaccharomyces pombe*, that are sensitive to loss of the conserved elongation factor, Spt4. Thus, fission yeast might employ a variant of the system of regulation found in higher eukaryotes

In flies and mammals, Pol II arrested within the promoter proximal region of a gene can only be released through the activity of a positive-transcription elongation factor (P-TEFb), composed of kinase (Cdk9) and cyclin (CycT1/2) subunits. Investigating the functional impact of Cdk9 on transcription in fission yeast revealed that, unlike most metazoan systems, Pol II in *S. pombe* is capable of overcoming the early elongation barrier after kinase inhibition, although not without consequence. However, fission yeast

lack the metazoan-specific negative elongation factor complex (NELF) involved in pausing, perhaps limiting their ability to control the release of Pol II through phosphorylation of the elongation complex.

Ultimately, by depleting pausing factors from cell lines derived from *Drosophila melanogaster,* it was tested whether NELF is required for P-TEFb-regulated pause escape. While global transcription is largely unaffected by the loss of NELF, upon inhibition of Cdk9, a significant amount of Pol II is aberrantly released from the pause, suggesting reduced control of this regulation. These findings suggest that NELF may have evolutionarily refined an ancestral promoter-proximal architecture of the transcription elongation complex, giving rise to a novel mechanism for gene regulation.

BIOGRAPHICAL SKETCH

As a child, born in Gainesville, Florida Gregory was unwittingly allured to the natural world, collecting prehistoric sharks' teeth in a creek near his home in Florida. Being raised by teachers – his father a university professor, and mother, a Montessori school teacher – his parents constantly encouraged this curiosity. In 2008, he enrolled at the University of Toronto in the faculty of Arts and Sciences, later selecting two majors, Ecology and Evolutionary Biology, and Genes, Genetics and Biotechnology.

In Toronto, Gregory's interests and curiosities were scattered amongst the diverse fields he was studying, and he searched for ways to extend his scientific exploration outside of the classroom. During his sophomore year, Gregory applied for a position as a research assistant on a field research expedition in the Peruvian Amazon under the supervision of Dr. Megan Frederickson. Over the course of two months, at a remote research station, Gregory helped to collect data for an experiment Dr. Frederickson's group had been running for a year. They sought to analyze the cost of a mutualistic relationship between the ant *Allomerus octoarticulatus* and its host, the tropical weed, *Cordia nodosa.* In a separate research endeavor the following year, Gregory was selected to attend the National Chao Tung University in Taiwan to conduct a small research project. At NCTU Gregory was situated in a laboratory where he worked closely with a graduate student to perform basic biochemical assays. The goal of his research was to develop a tool for removing sulfate molecules from proteins and polysaccharides, a process that had not been well studied. Upon returning from Taiwan for his final year at the University of Toronto, Gregory began his thesis work in Professor Susannah

Varmuza's laboratory determining the molecular pathways responsible for a spermatogenesis defect in mice lacking an important phosphatase enzyme.

Gregory's interest in research led him to pursue a Ph.D. in the field of Biochemistry, Molecular, and Cell Biology at Cornell University. At Cornell, Gregory paired with Professor John Lis and his laboratory to study the mechanisms that regulate gene expression in different systems. It was in graduate school that he began to appreciate how novel DNA sequencing platforms had transformed the field of molecular biology. Using such tools in Professor Lis' lab he could visualize genome-wide distributions of RNA polymerases, revealing nuances of the genome and its regulation.

I would like to dedicate this work to my parents, James and Rebecca Booth, who have provided unwavering support in all of my endeavors, not least of which were my education and graduate studies.

ACKNOWLEDGEMENTS

My experience at Cornell University would not have been possible without the support of the broad community I have grown to know and love over my years in Ithaca. First and foremost, I owe a wealth of gratitude to John Lis, who mentored me through my graduate school career. John remains such a humble, kind and conscientious person, always looking out for the interests of those working with him, whatever they may be. Thank you, John, for a wonderful experience; it has been an absolute pleasure. I would also like to thank Robert Fisher and Pabitra Parua from the Icahn School of Medicine at Mount Sinai who were instrumental in many of my research projects through the sharing of experimental results, ideas and reagents. I have been truly lucky to be a part of such a fruitful collaboration.

I am also grateful to my committee members Sylvia Lee and Jeff Pleiss. Throughout my studies they challenged me to think about my work from other perspectives, raising novel questions and providing insights often overlooked. I also owe a sincere thank you to Charles Danko and Hojoong Kwak, who have essentially acted as an unofficial committee members and taken great interest in my work. Their unwavering positivity is infectious and has always motivated me to grow and improve.

Through my years at Cornell, John's group has changed a lot and I am grateful to so many past and present of the lab members for helping me along the way. The lab has always felt like a family with each member having their own interests, opinions and personalities. Despite our differences, everyone wants to see each other succeed, and this altruistic property makes it so easy to want to be in lab. I would like to specifically thank Digbjay Mahat, Abdullah Ozer and Fabiana Duarte who encouraged me early on, when I

was intimidated by the general level of talent in the lab. Thank you to my bay-mate Nathaniel Tippens and good friend Benjamin Fair for endless open-ended discussions of science and politics in lab and over beers at the big red barn. Also, thank you to Janis Warner, who manages and organizes our often-disjointed group, providing us with girl scout cookies and homemade sweets throughout the year.

Having grown up in Ithaca, New York I am so fortunate to have my family nearby. The constant love and support from my parents James and Rebecca Booth cannot be overstated in making my time at Cornell so enjoyable. Like my parents, my sisters Carolyn and Jaqueline are remarkably open minded and adventurous, providing a constant source of inspiration for my own endeavors. Long-time friends from Ithaca have made life outside of lab so easy-going and it has been a joy to watch everyone grow and mature together. Outside of Cornell, I owe thanks to the Ithaca United Soccer Club and the various teams I have played on, as well as my musical friends for providing endless fun and creative outlets. I will thoroughly miss the enduring familiarity and nurturing environment of my hometown. Finally, I would like to thank Jessica Schembri who has unconditionally supported me throughout graduate school. With her genuine care and good nature, the pressure and weight of each challenge has been so easily lifted.

Work presented in this thesis was funded through grants from the National Institute of Health, grants GM25232 and HG009393 awarded to John Lis. Gregory Booth was funded in part by the NIH training grant 5T32GM007273-40.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Multicellular organisms often consist of diverse, highly-specialized cell types, which can organize to form sophisticated tissues and organs. Originating from a singular zygotic cell, these various cells share copies of the same genome, frequently equated to an instruction manual for the cell. Exactly how the unique morphologies and functions of trillions of cells in the human body is encoded in a fixed sequence of deoxyribose nucleic acids (DNA) remains an intense subject of research in molecular biology.

Genomes contain discrete units, called genes, which encode the macromolecules required for cellular function. Through the direct synthesis of RNA from genes encoded in the DNA, resulting transcripts are further processed into messenger RNA (mRNA) and transferred to the cytoplasm. Each mRNA can be read by a ribonucleoprotein complex, called the ribosome, which polymerizes the sequence of amino acids necessary to construct the encoded protein. This system of decrypting genes for the production of cellular machinery encompasses gene expression. By tightly controlling the expression of subsets of genes within the genome, cells can mount responses to environmental stimuli(Morimoto, 1993), identify and report infections(Yin Chen et al., 2006), as well as define and alter their identity or functional properties(McGinnis & Krumlauf, 1992).

Levels of gene expression regulation in eukaryotes

Because faithful gene expression is paramount to survival and function of cells, eukaryotes have evolved an impressively complex set of mechanisms for rapidly

regulating the abundance of a specific protein at any given time. Moreover, multiple layers of regulation and feedback between them can amplify effects on gene expression within a cell(Dahan, Gingold, & Pilpel, 2011). At the highest level, proteins themselves can be the target of such regulation. Mediated through covalent addition of ubiquitin molecules to a protein, targeted degradation of polypeptides can serve as a way to rapidly tune or eliminate the abundance of a specific protein from within a cell(Zattas & Hochstrasser, 2015). Other forms of post-translational modification, such as protein phosphorylation, can prevent the need for total degradation or elimination, simply by altering properties or activity of a target protein(Gallego & Virshup, 2007; Seet, Dikic, Zhou, & Pawson, 2006). In another underlying level of regulation, the characteristics and stability of mRNA are also tightly controlled, ultimately to tune protein translation(Baker & Coller, 2006). In fact, many approaches to reduce or eliminate the presence of a specific protein experimentally, target the mRNA rather than the protein. RNA interference (RNAi), for example, exploits an endogenous mechanism to control protein output, where short double stranded RNA guide an endogenous RNA Induced Silencing Complex (RISC) to mRNAs, preventing translation either directly or through degradation of the mRNA molecule(Filipowicz, Bhattacharyya, & Sonenberg, 2008; Fire et al., 1998).

At the most fundamental level, eukaryotic gene expression is controlled within the nucleus, where densely packed DNA must be made accessible by transcription factors for RNA polymerases to read while producing RNA. Eukaryotic DNA is packaged into chromatin (Kornberg, 1977), a higher order structure of nucleosomes consisting of 146 nucleotides of double stranded DNA wrapped around a complex of eight core histone proteins(Luger, Mäder, Richmond, Sargent, & Richmond, 1997). Apart from posing a

physical barrier to the underlying supercoiled DNA(Workman & Kingston, 1998), histones within a nucleosome can be chemically modified in a variety of ways. The complex patterns of histone modifications are now recognized as a dynamic code, which can inform chromatin binding proteins about the status of the contained DNA regarding transcriptional activity(B. Li, Carey, & Workman, 2007). For example, regions of the genome undergoing active transcription frequently harbor nucleosomes with acetylated histones 3 and 4 (H3, H4)(Brown, Lechner, Howe, & Workman, 2000), while transcriptionally silenced stretches of DNA can be ensured through methylation of histone 3 at lysine 27(Beisel & Paro, 2011; Jaenisch & Bird, 2003). Moreover, through a phenomenon described as epigenetics, heritable patterns of gene repression can be achieved through the methylation of cytosines (C) within CpG pairs of the DNA encoding specific genes(Wolffe & Matzke, 1999). To further complicate matters, recent advances for dissecting the three-dimensional organization of the nucleus have expanded our appreciation for how genes and other loci can be regulated by distal sequence elements(Dekker, 2008). Beyond the immediate environment surrounding a given gene, long-range chromatin interactions likely play roles in dictating features of a locus, including transcription factor occupancy, histone modification profiles, and transcriptional activity(Harmston & Lenhard, 2013).

Within the nucleus of a cell, all aspects of this dynamic chemical and physical chromatin landscape conspire to regulate when and where RNA polymerases transcribe DNA into RNA. Unlike prokaryotes and archaea, eukaryotes possess three distinct RNA polymerase complexes, Pol I, Pol II and Pol III(Roeder & Rutter, 1969). Although Pol I, Pol II and Pol III all utilize several common or related subunits and general transcription

factors, such as the TATA binding protein (TBP) during transcription initiation, each polymerase possesses a unique set of core subunits and has evolved to produce transcripts from a specific class of genes(Vannini & Cramer, 2012). Pol I and Pol III account for the majority of RNA synthesis within a cell(Paule & White, 2000), but are highly specialized for the rapid, continual production of ribosomal components, 25s rRNA precursor RNA (Pol I), and 5s rRNA as well as short untranslated RNAs, like transfer RNAs (tRNAs) (Pol III). In contrast, Pol II is solely responsible for the transcription of all protein coding genes in eukaryotes(Vannini & Cramer, 2012).

Organisms mount responses to changes in environment and execute complex developmental programs by dictating the expression of subsets of protein coding genes(Mahat, Salamanca, Duarte, Danko, & Lis, 2016b; Tsankov et al., 2015). The timing, frequency, and location of transcription by Pol II across the genome are all tightly regulated by a series of events that must be occur for a gene to be transcribed(Fuda, Ardehali, & Lis, 2009). Transcription machinery must first gain access to a gene's promoter, a stretch of DNA containing the transcription start site (TSS) upstream of a transcription unit(Juven-Gershon, Hsu, Theisen, & Kadonaga, 2008; Smale & Kadonaga, 2003). The promoter serves as a platform for the binding of general transcription factors (GTFs) and subsequent recruitment of Pol II(Juven-Gershon et al., 2008). In certain cases transcription factors which bind DNA, can act locally to remove, remodel, and modify nucleosomes to maintain a nucleosome free region (NFR) around the promoter(Boeger, Griesenbeck, Strattan, & Kornberg, 2003; Fuda, Guertin, Sharma, Danko, & Martins, 2015; Svaren & Hörz, 1997). Within nucleosome free promoter regions, GTFs assemble to form the pre-initiation complex (PIC)(Rhee & Pugh, 2012).

The formation and stability of the PIC(Yudkovsky, Ranish, & Hahn, 2000), as well as the recruitment of Pol II, can all be further stimulated by transcription activators and cofactors, such as the multi-subunit mediator complex(Kornberg, 2005). Complicating the regulation of promoter use, activators and cofactors can impart their effects locally, within the promoter, or via distally located regulatory elements, called enhancers(Jeronimo et al., 2016; Levine & Tjian, 2003; Petrenko, Jin, Wong, & Struhl, 2016), which share many of the same properties as promoters(Core et al., 2014). Once RNA Pol II has been recruited to the PIC, transcription initiation follows the ATP-dependent unwinding of the DNA by the GTF TFIIH(X. Liu, Bushnell, & Kornberg, 2013) and Pol II begins elongating a nascent RNA chain as it translocates along the template DNA.

Elongating Pol II is an attractive target for regulation, as it is required to coordinate various processes and integrate a variety of signals to produce a full-length transcript. Until fairly recently the regulation of transcription was thought to be primarily controlled via the recruitment of RNA Pol II (Ptashne & Gann, 1997). However, profiles of RNA polymerases in most well-studied metazoans identified an enrichment of transcriptionally engaged Pol II immediately downstream of the TSS many genes(Core, Waterfall, & Lis, 2008; H. Kwak, Fuda, Core, & Lis, 2013; Muse et al., 2007; Nechaev et al., 2010). More recently, this enrichment has been revealed to represent a promoter-proximal regulatory checkpoint during Pol II elongation that has emerged as a major platform for controlling rapid global changes gene expression.

Coordination of transcription elongation-coupled processes

During productive elongation of RNA, Pol II is presented with a series of challenges for which an elaborate system of factor exchange and signal integration and has evolved. While producing RNA at more than 2000 nucleotides per minute(Danko et al., 2013; Jonkers, Kwak, & Lis, 2014; Mason & Struhl, 2005), Pol II dynamically associates with factors to navigate and modify barriers posed by nucleosomes within the gene, chemically modify and protect the 5' end of nascent RNA, selectively identify and remove introns intervening the coding sequences (exons) of RNA, and terminate transcription upon 3' end selection and cleavage(H. Kwak & Lis, 2013; D. W. Zhang, Rodríguez-Molina, Tietjen, Nemec, & Ansari, 2012).

For many coinciding-biochemical processes, it can be difficult to disentangle direct, causative relationships. However, research into the structure and function of RNA Polymerase II has elucidated a fundamental basis for the coupling of transcription and associated operations. Unique to Pol II, the largest core subunit, rpb1 possesses a conserved flexible carboxy-terminal domain (CTD) consisting of repeating units of the seven amino acid sequence Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7 (Corden, 1990). Although the number of repeats varies by organism, with 26 copies in the budding yeast *Saccharomyces cerevisiae*, and 52 in mammals, this unstructured peptide is thought to be capable of orchestrating a multitude of co-transcriptional processes.

The language of orchestration appears in the form of various reversible chemical and structural modifications to residues within each repeat of the CTD. Like a blank canvas, when recruited to a promoter, GTFs of the PIC interact with Pol II CTD lacking post-translational modification(D. W. Zhang et al., 2012). Upon transcription initiation,

the kinase subunit of TFIIH (Kin28 in budding yeast, Cdk7 in fission yeast and metazoans) phosphorylates Ser5 residues of the CTD. RNA capping machinery directly bind the Ser5-phosphorylated CTD and, upon emergence of the nascent RNA from the Pol II exit channel, add a protective 7-methyl guanosine to the 5' end of transcripts(Hsin & Manley, 2012; Rasmussen & Lis, 1993).

Beyond promoter clearance, elongating Pol II frequently encounters nucleosomes which can pose as barriers to its movement(Izban & Luse, 1992; Shaw, Sahasrabuddhe, Hodo, & Saunders, 1978). Factors which engage and modify chromatin are recruited to Pol II CTD through phosphorylated Ser2, enabling a relatively unobstructed passage of the elongation complex(H. Cho et al., 1998; B. Li, Howe, Anderson, Yates, & Workman, 2003; Vojnic, Simon, Strahl, Sattler, & Cramer, 2006). Intron-exon boundaries are encoded by regulatory sequences within pre-mRNA and their recognition and usage by splicing factors frequently occurs co-transcriptionally(Merkhofer, Hu, & Johnson, 2014; Perales & Bentley, 2009). Certain splicing factors have been found to interact with Pol II through the CTD and the elongation rate of Pol II is known to influence alternative splicing(Luco, Allo, Schor, Kornblihtt, & Misteli, 2011; Muñoz, la Mata, & Kornblihtt, 2010). In the final stages of the transcription cycle, pre-mRNA must be cleaved and released from the terminating elongation complex. 3' RNA processing machinery must selectively recognize a cleavage and polyadenylation signal (CPS) present within pre-mRNA and components of this multisubunit complex are known to be recruited to phosphorylated Ser2 within the CTD(Licatalosi et al., 2002; Meinhart & Cramer, 2004). Termination efficiency can also be highly dependent on the rate of Pol II(Fong et al., 2015), and beyond the CPS of most eukaryotic genes Pol II appears to slow down or stall,

apparently to facilitate exonuclease mediated template eviction of Pol II(Proudfoot, 2016). Given the impact of Pol II rate on critical RNA processing steps, understanding how elongation by Pol II is coordinated within an environment and the influence of CTD modification may provide important insights into gene regulation.

Promoter-proximal pausing of RNA Polymerase II

Early investigations of transcription at the heat responsive Hsp70 locus in Drosophila melanogaster revealed that transcriptionally engaged but paused Pol II was present approximately 25 bases downstream of the TSS under non-heat-shock conditions(Rougvie & Lis, 1988). Rather than being unique to one specialized class of genes, similar transcriptionally engaged, but paused Pol II was found downstream of TSSs of other fruit fly housekeeping genes(Rougvie & Lis, 1990). Such Pol II enrichments were also found within the human c-myc gene(Bentley & Groudine, 1986) and even the human immunodeficiency virus (HIV) type 1 long terminal repeat (LTR)(Rosen, 1991). The genes exhibiting such Pol II profiles were called promoter-proximally paused, in what was proposed to be a novel, potentially rate-limiting step in transcription(Krumm, Hickey, & Groudine, 1995).

Early genome-wide profiles of Pol II led to the suggestion that promoter-proximal pausing of Pol II was widespread, likely regulating hundreds or thousands of genes in drosophila(Muse et al., 2007) and humans(Core et al., 2008). Finally, by chemically blocking pause escape in mouse embryonic stem cells, it was observed that ~95% of genes could not elongation beyond ~50 bp downstream of the transcription start site, suggestive of a ubiquitous step in elongation(Jonkers et al., 2014).

In vitro studies lent key insights into the mechanisms of the elongation pause. Initially, work from the Price laboratory found that the ionic detergent sarkosyl, or high salt prevented unknown nuclear factors from blocking elongation of Pol II in a transcription reaction(Kephart, Marshall, & Price, 1992). Shortly thereafter, they uncovered and purified a factor from human nuclear extract, named positive transcription elongation factor b (P-TEFb), that was sensitive to the nucleoside analog, 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) and required for the release of paused elongation complexes(Marshall & Price, 1992; 1995). Subsequently, using the ability to chemically inhibit pause escape with DRB, the first pausing factors were identified. Two human factors, DRB-sensitivity inducing factor (DSIF) and negative elongation factor (NELF) were sufficient to impose the post-initiation block to Pol II elongation *in vitro*(Wada et al., 1998; Yamaguchi et al., 1999).

Since the initial identification of pausing factors, research into their function and properties have unearthed an intricate network of signal-responsive protein-protein and protein-nucleic acid interactions acting to control the release of Pol II into productive elongation. P-TEFb is a cyclin-kinase complex composed of a kinase, Cdk9 (Bur1 in budding yeast) and cyclin subunit, cyclinT1 or cylclinT2 (CycT1/2) in humans (CycT in flies, Bur2 in budding yeast)(Peng, Zhu, Milton, & Price, 1998). The kinase activity of Cdk9 is demonstrably essential for pause escape, as selective kinase inhibitors such as DRB and flavopiridol globally prevent pause escape in vivo(Chao & Price, 2001; Jonkers et al., 2014; Ni et al., 2008). Global P-TEFb activity can be sequestered and suppressed in mammals through a reversible association with the HEXIM complex consisting of two hexamethylene bisacetamide inducible (HEXIM1/2) proteins and a 7SK small nuclear

RNA(Michels et al., 2004). Locally, transcription factors can influence gene expression through direct recruitment of P-TEFb(Peterlin & Price, 2006). Initially observed as the mechanism used by the viral Tat protein to stimulate transcription of the HIV-1 LTR(Taube, Fujinaga, Wimmer, Barboric, & Peterlin, 1999), human transcription factors like c-Myc, NF-kappaB can promote pause release at nearby genes through direct recruitment of P-TEFb(Barboric, Nissen, Kanazawa, Jabrane-Ferrat, & Peterlin, 2001; Eberhardy & Farnham, 2002).

The mechanism by which P-TEFb triggers release of Pol II into productive elongation remains incompletely understood. However, phosphorylation of substrates within the pausing complex likely reconfigures critical interactions between NELF, DSIF, Pol II and nascent RNA during pause escape(Missra & Gilmour, 2010; Ping & Rana, 2001). Once recruited to the vicinity of the pausing complex Cdk9 phosphorylates Ser2 residues within the Pol II CTD as well as subunits of the NELF and DISF pausing factors(Adelman & Lis, 2012; Fuda et al., 2009; A. Saunders, Core, & Lis, 2006).

NELF is composed of four subunits: NELF-A and B, either NELF-C or D and a smaller, RNA recognition motif (RRM)-containing, NELF-E(Yamaguchi et al., 1999). Cdk9 is proposed to phosphorylate NELF-E near the RRM(Fujinaga et al., 2004), and in doing so , NELF is dissociated from the paused transcription complex(Wu et al., 2003). Rather than being evicted from the transcription complex, DSIF is converted to a positive elongation factor upon phosphorylation, remaining stably associated with Pol II throughout elongation(Yamada et al., 2006).

Apart from roles in promoter-proximal pausing DSIF is considered a general elongation factor that tightly interacts with transcribing Pol II and the nascent

RNA(Ehara et al., 2017). DSIF comprises two subunits, Spt4 and Spt5, which are conserved across archaea and eukaryotes(M. Guo et al., 2008). During the transition from pausing to productive elongation P-TEFb phosphorylates Spt5 within an unstructured, repeated sequence of amino-acids near the C-terminus, called the CTR(Ivanov, Kwak, Guo, & Gaynor, 2000; Yamada et al., 2006). How phosphorylation converts DSIF into a positive elongation factor is an active area of research. DSIF may share properties with that of the Pol II CTD, where modification of specific residues within the Spt5 CTR provide a versatile surface for binding of additional factors. For example, RNA capping components interact directly with the Spt5 CTR (Wen & Shatkin, 1999), but phosphorylation within this domain inhibits this interaction as shown in the fission yeast *Schizosaccharomyces pombe*(Doamekpor, Sanchez, Schwer, Shuman, & Lima, 2014). Also through modification of the CTR, Spt5 may indirectly facilitate manipulation of the chromatin environment during elongation by recruiting Polymerase associated factor 1 complex (Paf1c) (Y. Liu et al., 2009; Wier, Mayekar, Héroux, Arndt, & VanDemark, 2013).

Although NELF and DSIF are sufficient for pausing *in vitro*, additional factors have been implicated in pausing *in vivo*. For example, the paused complex may be further stabilized by the Pol II subunit, GDOWN1, which can block factors that might stimulate premature transcription termination(J. Guo, Turek, & Price, 2014). Additionally, the PAF1 complex assists transcription elongation *in vitro* and has been found to be required for pause escape *in vivo*(Yexi Chen et al., 2009; J. Kim, Guermah, & Roeder, 2010a; Yu et al., 2015). Interestingly, more recently it was reported that depletion of PAF1 leads to global shifts in Pol II into gene bodies and away from the

pause, implying instead, a role in establishing the pause(Fei Xavier Chen et al., 2017). Further functional analysis as well as structural information regarding the paused transcription elongation complex will help to resolve the role of PAF1 and other factors in establishing and regulating this step.

Finally, in addition to factors, which actively bind the transcription elongation complex, contextual features of the promoter environment might also influence properties of paused Pol II. Simply increasing the spacing between promoter elements at the Hsp70 gene in *D. melanogaster* led to a dispersed, downstream-shifted pause(H. Kwak et al., 2013), suggesting GTFs and the core promoter may contribute to post-initiation positioning. Nucleosomes, which are frequently positioned immediately downstream of the pause have encouraged models in which they form a barrier to elongation(H. Kwak & Lis, 2013), though further investigation is required. Nonetheless, each molecular event required for the capture of Pol II at the pause site and control its escape has been shaped through evolution for the regulation of gene expression. However, functions of this process beyond regulating a single locus may have strengthened selection to maintain and regulate such an event.

Functions of promoter-proximal pausing

By minimizing the number of steps remaining to produce RNA from a locus, pausing can provide platform for rapid activation of gene expression programs within a cell. Immediate-early genes, named for their rapid responses to external stimuli such as inflammation, tend to be highly paused. For example, in mammals and flies the most rapidly responsive genes to microbial response through lipopolysaccharide (LPS) stimulation were found to be regulated through pause escape, while slow responding

genes were regulated by Pol II recruitment(Adelman et al., 2009). Beyond synchronous activation of genes within the same pathway during embryonic development, extracellular factors often act to stimulate tissue patterning across many cells with precise spatial and temporal parameters. The simultaneous activation of the *snail* gene in ~1000 cells within a developing fruit fly embryo was found to depend on promoter-proximal pausing. Exchanging the native, pause-inducing promoter with a less, or non-paused gene promoter led to a loss of synchronous *snail* activation, consequently perturbing invagination of the embryonic mesoderm(Lagha et al., 2013). Thus, coordination of cross-cell gene-expression within multi-cellular organisms can utilize promoter-proximal pausing for fine-tuned spatiotemporal regulation.

Besides controlling the timing of production of individual RNA molecules, paused Pol II downstream of the promoter could have important impacts on the local promoter chromatin environment, contributing further to maintained activity of a locus. Particularly in mammals, where pairs of initiation and pausing sites are characteristic of active regulatory sites(Core et al., 2008; Seila et al., 2008), the presence of Pol II flanking transcription factor binding sites or core promoters in divergent orientations could occlude nucleosomes from occupying key sequence elements. Consistent with this idea, the spacing between divergent transcription start site pairs in mammals was shown to dictate the boundaries of nucleosome free regions around promoters(Scruggs, Gilchrist, Nechaev, & Muse, 2015). Moreover, depletion of NELF subunits results in reduced transcription from highly paused genes, coincident with the encroachment of nucleosomes on otherwise nucleosome free promoters in drosophila(Gilchrist et al.,

2010). Overall, these findings suggest addition functions of transcriptionally engaged Pol II on chromatin, beyond simply the production of RNA.

Given that chromatin manipulation and RNA processing must be coordinated through factor exchange during Pol II elongation, promoter-proximal pausing could provide an opportunity to ensure proper maturation of the elongation complex. Conveniently, pausing occurs shortly after RNA has first emerged from the Pol II exit channel, when a variety of elongation factors associate with the complex(Adelman & Lis, 2012). Prior to pausing, through phosphorylation of Ser5 within the CTD of Pol II, capping machinery is recruited to Pol II. Indeed, the vast majority of nascent RNAs are capped by the time 20 to 30 nucleotides of RNA have been synthesized, at or near the position of pausing(Nechaev & Adelman, 2011; Rasmussen & Lis, 1993). Thus, an important series of events required for productive eukaryotic transcription likely takes place within an early stage of Pol II elongation. In addition to the previously mentioned functions, pausing may serve as checkpoint to ensure efficient and productive rounds of transcription.

Ancestral elongation checkpoints related to pausing

Despite nearly all metazoan Pol II transcribed genes requiring the activity of Cdk9 to escape promoter proximal pausing, under normal conditions *in vivo* a large fraction of genes appear to pass through this step with minimal or no impedance, suggesting that they are not rate-limited by pausing(Jonkers et al., 2014). Indeed, many eukaryotes show little, if any, sign of gene regulation via promoter-proximal pausing(Churchman & Weissman, 2011; Kruesi, Core, Waters, Lis, & Meyer, 2013). For example, the large body of research on eukaryotic transcription using the budding yeast *Saccharomyces*

cerevisiae, strongly supports a view of primary gene regulation at the level of Pol II recruitment and initiation(Ptashne & Gann, 1997; Stargell & Struhl, 1996). Nonetheless, early regulatory checkpoints have been proposed in budding yeast after TFIIH kinase inhibition led to accumulation of Pol II at the 5' end of genes(Rodríguez-Molina, Tseng, Simonett, Taunton, & Ansari, 2016).

Interestingly, much of the same machinery utilized in the establishment of promoter-proximal pausing is conserved across all eukaryotes(Peterlin & Price, 2006). DSIF subunits Spt4 and Spt5 were originally identified in budding yeast by screening for factors that could suppress transposable element (Ty)-mediated interference of His4 gene expression(Winston, Chaleff, Valent, & Fink, 1984). Yeast Spt4 and Spt5 were later found to enhance transcription elongation(Mason & Struhl, 2005; Rondón, García-Rubio, González-Barrera, & Aguilera, 2003), possibly by reducing stalling events that lead to Pol II backtracking(Hartzog, Wada, Handa, & Winston, 1998). Moreover, protein sequence conservation revealed a domain of Spt5 resembling the bacterial elongation factor, NusG(Ponting, 2002), further suggesting an ancient origin this complex and its role in gene expression(Burova, Hung, Sagitov, Stitt, & Gottesman, 1995; Werner & Grohmann, 2011). The phosphorylation of the Spt5 CTR implicated in pause escape in metazoans is also conserved across eukaryotes, including yeast (Y. Liu et al., 2009). Moreover, yeast counterparts to DSIF-associating factors in mammals, like capping components, chromatin modifiers, and cleavage and polyadenylation machinery play highly related roles, mediated through Spt5(Hartzog & Fu, 2013). Unlike pausing seen in metazoans, a negative role for Spt4-Spt5 has yet been described in yeast. Nonetheless, Spt5 association with Pol II and its phosphorylation during early Pol II elongation remain

highly coordinated events(Larochelle et al., 2012; Mayer et al., 2010). The conserved importance of Spt5 phosphorylation across studied eukaryotes emphasizes importance of early stages of transcription as a functional platform, even in the absence of pausing.

Transcription coupled kinases, including TFIIH, Mediator and P-TEFb subunits and substrates are also highly conserved amongst eukaryotes. The orthologous Cdk9 and cyclin subunits in yeast (Bur1/Bur2) have likewise been strongly implicated in transcription elongation(Keogh, Podolny, & Buratowski, 2003; Sansó et al., 2012). Initially thought to be a distinct feature of yeast, the phosphorylation of Ser2 within the Pol II CTD appears to be largely accomplished by kinase Ctk1/Cdk12, rather than Cdk9/Bur1. However, recently a metazoan ortholog for Cdk12 was identified with the capacity to phosphorylate Ser2 of the Pol II CTD(Bartkowiak et al., 2010). On the contrary, in all eukaryotes Cdk9/Bur1 appears to have a preserved substrate specificity for the Spt5 CTR. Similar to a function ascribed to promoter-proximal pausing in metazoans, this phosphorylation event has been proposed as an early elongation checkpoint in fission yeast to ensure proper RNA capping and factor assembly(Pei & Shuman, 2002). However, unlike most metazoans, which undergo promoter-proximal pausing, inhibition of Cdk9 orthologs in yeast does not globally inhibit transcription production(Y. Liu et al., 2009; Viladevall et al., 2009). Thus, despite the conservation of various players involved in early elongation, distinct mechanisms must exist, which enable regulation of Pol II through promoter-proximal pausing.

The absence of NELF in organisms lacking pause-like distributions of Pol II has formed the basis for the argument that it is required for the metazoan-specific mechanism of transcription regulation(Peterlin & Price, 2006). After all, NELF is required for

pausing *in vitro*(Yamaguchi et al., 1999), and organisms which lack homologs of NELF subunits tend not to display strong Pol II enrichments within promoter proximal regions(Narita et al., 2003). Moreover, despite clearly impacting transcription elongation, perturbing Cdk9 activity in such organisms does not result in an inability of Pol II to escape the pause(Keogh et al., 2003; Viladevall et al., 2009). Nonetheless, several lines of evidence point to a more complicated picture *in vivo*, where promoter-proximal pausing may not strictly rely on NELF. For example, depletion of the NELF complex *in vivo* fails to completely eliminate pausing of Pol II in both drosophila and mammals(Core et al., 2012; Gilchrist et al., 2008). Additionally, in the nematode *Caenorhabditis elegans*, which also lack NELF, pause-like distributions of Pol II were observed under nutrient limiting conditions(Kruesi et al., 2013).

As described throughout this section, so many of the processes and factors involved in, coordinated with, and/or coinciding with promoter-proximal pausing are shared across eukaryotes. Despite our current understanding of the basis and function of promoter-proximal pausing of transcriptionally engaged Pol II, many questions remain regarding this mechanism for controlling gene expression. Why do some eukaryotes use pausing? In species lacking promoter proximal pausing, what are the functions of these conserved factors and how are the shared processes organized? How did Cdk9 become required for the release Pol II from this early elongation step? By addressing such questions about the origins of regulatory mechanisms, we will unveil how gene regulatory logic can evolve. Although many functions have been ascribed to promoter-proximal pausing for sets of genes, it is still not understood why nearly all Pol II transcription events must undergo this seemingly energetically costly process. Ancestral

systems likely hold answers to the molecular events required for regulated transcription within a nucleus, and Pol II pausing may represent an advancement in control built from a pre-existing architecture.

The body of work presented in this thesis reflects the results of an investigation of the functional origins of promoter-proximal pausing as a pervasive mechanism of gene regulation in metazoans. In the following chapters, I will present the results of three distinct projects. In the first project, using high resolution approaches to map the positions of engaged Pol II, global transcription profiles of two highly divergent yeasts, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* are compared. The results present unprecedented contrasts between yeasts in the way Pol II traverse genes and uncover pause-like Pol II enrichments in fission yeast, which depend on Spt4. The second project examines the role of conserved transcription-coupled kinases in fission yeast using specific chemical-genetic perturbation. Cdk9 is found to be required promoter-proximally in a step that enables maturation and rapid elongating of Pol II, though it is dispensable for pause escape. Finally, I report the results of experiments in *D. melanogaster*, investigating whether NELF enables P-TEFb regulated escape and present a discussion on what lies ahead.

CHAPTER 2

DIVERGENCE OF A CONSERVED ELONGATION FACTOR AND TRANSCRIPTION REGULATION IN BUDDING AND FISSION YEAST[1]

Abstract

Complex regulation of gene expression in mammals has evolved from simpler eukaryotic systems, yet the mechanistic features of this evolution remain elusive. Here, we compared the transcriptional landscapes of the distantly related budding and fission yeast. We adapted the Precision Run-On sequencing (PRO-seq) approach to map the positions of RNA polymerase active sites genome-wide in *S. pombe* and *S. cerevisiae.* Additionally, we mapped preferred sites of transcription initiation in each organism using PRO-cap. Unexpectedly, we identify a pause in early elongation, specific to *S. pombe,* that requires the conserved elongation factor subunit Spt4 and resembles promoter-proximal pausing in metazoans. PRO-seq profiles in strains lacking Spt4 reveal globally elevated levels of transcribing RNA Polymerase II (Pol II) within genes in both species. Messenger RNA abundance, however, does not reflect the increases in Pol II density, indicating a global reduction in elongation rate. Together, our results provide the first base-pair resolution map of transcription elongation in *S. pombe* and identify divergent roles for Spt4 in controlling elongation in budding and fission yeast.

[1] This chapter has been adapted from a published article with associated corrections incorporated (Booth, G. T., Wang, I. X., Cheung, V. G., & Lis, J. T. (2016). Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome Research*, *26*(6), 799–811.)

Introduction

Budding and fission yeast are both widely used model organisms for the study of transcription, yet their genomes are as distantly related as metazoans and budding yeast (Dujon, 2010). Details of transcription and its regulation have been characterized through a variety of high-resolution and genome-wide studies for *S. cerevisiae* (Hahn & Young, 2011; Pelechano & Steinmetz, 2013). However, although location and production of genomic RNA has been investigated in fission yeast (DeGennaro et al., 2013; Wilhelm et al., 2008), various aspects of the transcriptional landscape of *S. pombe* remain more vague. Through our investigation of elongating RNA polymerase distributions in *S. pombe*, we unexpectedly found evidence of promoter-proximal pause-like distributions of RNA polymerase across the *S. pombe* genome. These findings prompted our inquiry into the mechanistic differences between budding and fission yeast at the level of transcription.

Unlike most well-studied metazoans, budding yeast displays a relatively uniform distribution of elongating Pol II across transcription units (E. J. Steinmetz et al., 2006). The lack of significant barriers to elongation in *S. cerevisiae* supports a model in which Pol II recruitment and initiation are the predominant steps regulating gene activity (Ptashne & Gann, 1997; Stargell & Struhl, 1996). By contrast, the majority of genes in metazoans, including mammals and flies, are intricately regulated at a promoter-proximal pausing step during elongation (Adelman & Lis, 2012; Fuda et al., 2009). Characteristics of transcription elongation also differ between metazoans and *S. cerevisiae* beyond the cleavage and polyadenylation signal (CPS) (Porrua & Libri, 2015). In contrast to budding yeast, elongating Pol II in mammals experiences post-CPS slowing or pausing

while continuing to transcribe for several kilobases prior to termination (Core et al., 2008; Gromak, West, & Proudfoot, 2006; Laitem et al., 2015; Proudfoot, 1989). These differences in Pol II dynamics reflect distinct mechanisms for controlling gene transcription that likely reflect the vast evolutionary distance separating these organisms.

How promoter-proximal pausing evolved into a ubiquitous mechanism regulating transcription in mammals is unknown. Metazoans with pervasive pausing possess a four-subunit complex termed negative elongation factor (NELF) with no known orthologs in yeast, *C. elegans*, or plants (Hartzog & Fu, 2013). NELF cooperates with DRB-sensitivity inducing factor (DSIF), composed of the highly conserved Spt4 and Spt5 subunits, to induce stable pausing of RNA polymerase shortly after initiation (Wada et al., 1998; Wu et al., 2003). However, studies in which NELF was highly depleted in vivo have shown that cells still retain substantial promoter-proximal pausing (Brannan et al., 2012; Core et al., 2012), consistent with the idea that DSIF is sufficient to induce promoter-proximal pausing (Wada et al., 1998). In fact, recent studies found that, despite lacking NELF homologs, some *C. elegans* genes are paused under nutrient-limiting conditions (Dujon, 2010; Kruesi et al., 2013; Maxwell et al., 2014). Nonetheless, it remains unknown whether such pausing in vivo is dependent on Spt4-Spt5.

Pause release in metazoans is modulated by the kinase activity of cyclin dependent kinase 9 (CDK9), a complex referred to as positive transcription elongation factor b (pTEFb) (Hahn & Young, 2011; Pelechano & Steinmetz, 2013; Wada, 1998). CDK9 phosphorylates NELF, DSIF, and Pol II to permit pause release and elongation into the gene body (DeGennaro et al., 2013; Peterlin & Price, 2006; Wilhelm et al., 2008). This transition coincides with the release of NELF from the complex, at which

point phosphorylated DSIF is converted into a positive elongation factor (Ping, 2000; E. J. Steinmetz et al., 2006). In all eukaryotes DSIF stably associates with transcribing Pol II throughout the duration of transcription. Mutations in either DSIF subunit (Spt4-Spt5) in *S. cerevisiae* have been found to result in elongation defects (Keogh et al., 2003; Ptashne & Gann, 1997; Stargell & Struhl, 1996; Swanson, Malone, & Winston, 1991), but how Spt4-Spt5 evolved an additional role in promoter-proximal pausing remains unknown.

Our understanding of pausing as well as numerous other transcription-coupled phenomena has been greatly enhanced through the use of nuclear-run on approaches (Adelman & Lis, 2012; Fuda et al., 2009; J. D. Love, Vivino, & Minton, 1985). More recently, Global and Precision-Run-On sequencing (GRO/PRO-seq) have provided genome-wide views of the distribution of engaged RNA Polymerases with strand specificity in metazoan organisms (Core et al., 2008; H. Kwak et al., 2013; Porrua & Libri, 2015). Despite the use of various run-on-based approaches in *S. cerevisiae* (Core et al., 2008; Gromak et al., 2006; Jordán-Pla et al., 2014; Laitem et al., 2015; McKinlay, Araya, & Fields, 2011; Pelechano, Chávez, & Perez-Ortin, 2014; Proudfoot, 1989; Rodríguez-Gil et al., 2010), neither GRO-seq nor PRO-seq has been fully demonstrated in yeast. Here we use a yeast-optimized, Precision Run-On sequencing approach to assay elongating RNA polymerases genome-wide in distantly related budding and fission yeast, providing the first base-pair resolution transcription analysis in *S. pombe*. We further apply this technique to investigate the importance of the Spt4-Spt5 complex in global transcription elongation. Comparing the effect of Spt4 deletion in both *S. cerevisiae* and

S. pombe provides insight into the evolution of its role in regulating Pol II transcription elongation in eukaryotes.

Results

PRO-cap and PRO-seq capture sites of transcription initiation and elongation in *S. cerevisiae* and *S. pombe*. We first set out to map positions of all elongation competent RNA polymerases in both budding and fission yeast using the Precision Run-On Sequencing (PRO-seq) protocol, which we adapted for use in yeast. Furthermore, we used Precision Run-On 5' cap sequencing (PRO-cap), a modification to the PRO-seq protocol which includes selection for short, capped nascent RNAs, to precisely identify the initiation sites for transcription units in both species (Hartzog & Fu, 2013; H. Kwak et al., 2013). PRO-seq signal in *S. pombe* is observed across the transcription unit of active genes, providing the direction, relative amount and distribution of elongating polymerases (Figure 2.1a, top track). By contrast, PRO-cap signal is largely confined to the 5' regions near annotated TSSs, indicating a successful enrichment of nascent RNA 5' cap sites (Figure 2.1a, lower track).

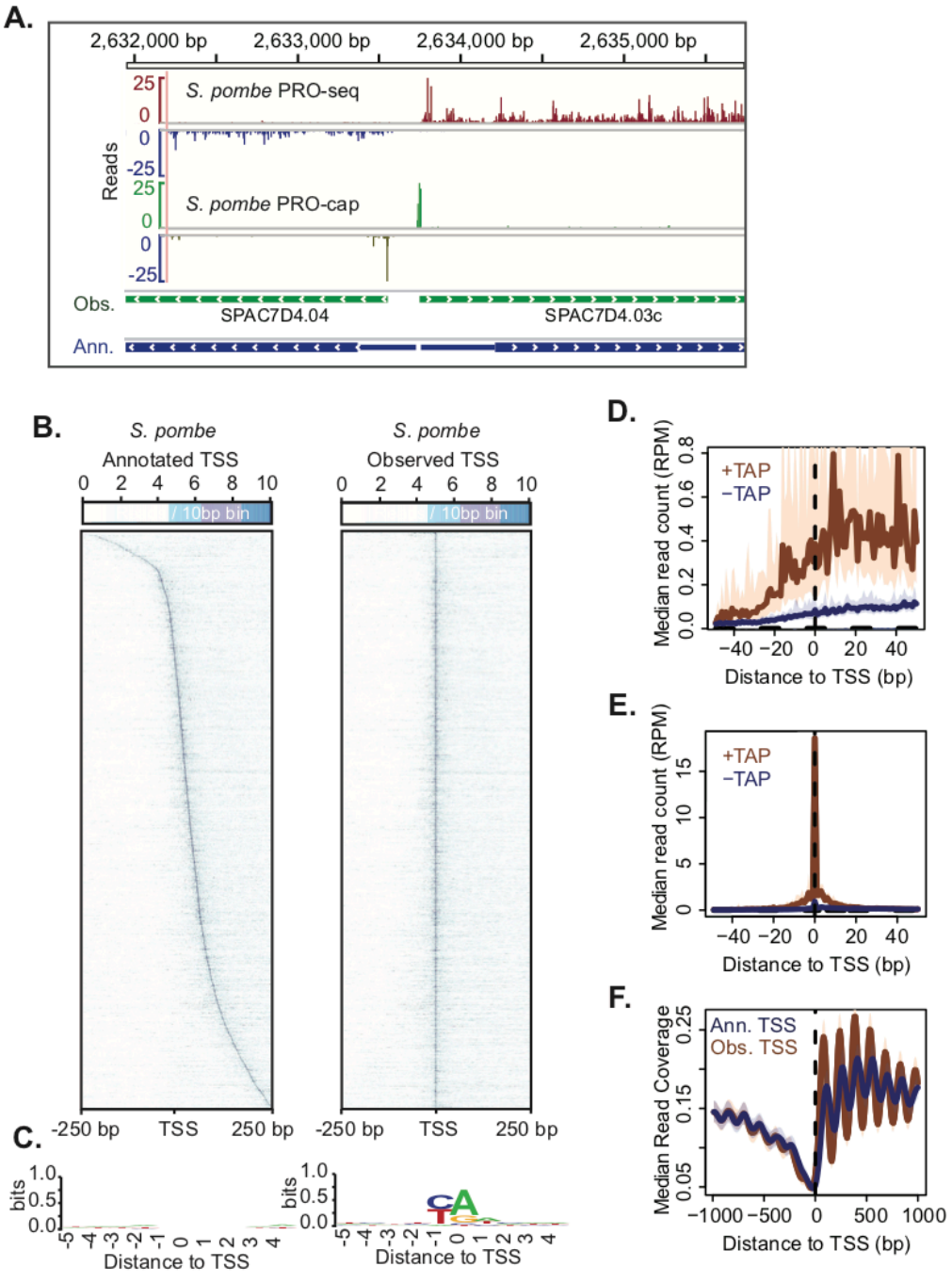

Figure 2.1: PRO-seq and PRO-cap capture transcription elongation and initiation genome-wide in *S. pombe*. **a.** Browser tracks of PRO-seq (plus strand: red; minus strand: blue) and PRO-cap data (plus strand: green; minus strand: yellow) derived from *S. pombe*. Green gene models below the data tracks show the re-annotated, 'observed' transcription start-sites based on PRO-cap data. Blue gene models correspond to the available annotations of genes. **b.** Heatmaps of *S. pombe* PRO-cap signal for each base within +/- 250 bp around the annotated TSS (left) and PRO-cap observed TSS (right) for

all active and filtered genes (N = 3214). Genes within heatmaps are sorted by increasing downstream distance of observed TSS relative to annotated TSS. **c.** Sequence logos of 10 bp sequence centered on either annotated TSS (left) or observed TSS (right) generated using WebLogo (Crooks, Hon, Chandonia, & Brenner, 2004). **d.** PRO-cap signal from samples prepared either with or without Tobacco Acid Pyrophosphatase (TAP) treatment centered on annotated TSS. The TAP-minus samples represent empirical levels of background for each genomic position. **e.** Median PRO-cap signal from samples prepared either with or without TAP treatment centered on observed TSS. **f.** Median MNase-seq coverage centered on annotated TSSs (blue) or observed TSSs (brown). For the meta-gene plots, the y-axis shows the median read counts for each base-pair (d&e), or median read coverage within 10 bp bins (f). In figures d, e and f, the 12.5% and 87.5% quantiles are shown in lightly shaded regions.

To more precisely define the initiation sites of transcription units, we identified the base with the most PRO-cap reads above background within 250 bases upstream or downstream of the annotated TSS (Figure 2.1a, Sup. Figure 2.5a, green models). We refer to the selected base as the 'observed TSS', and to previous annotations as the 'annotated TSS'. Heat maps of PRO-cap read depth surrounding previously 'annotated TSSs' in *S. pombe* show that initiation is distributed around the annotated TSS, but that the annotation rarely matches the base-pair resolution measurement of initiation that we make with PRO-cap (Figure 2.1b, left panel). In contrast, the distribution of PRO-cap reads is centered and highly focused at our 'observed TSSs' in *S. pombe*, indicating a narrow window within which start-site selection occurs (Figure 2.1b, right panel). Moreover, a moderate sequence preference for initiating at an adenosine, two bases downstream of a thiamine, was revealed under the observed TSS, whereas no base preferences underlie the PomBase annotations for the same genes (Figure 2.1c). More recently, groups have used other approaches to re-annotate transcription units across the fission yeast genome (Eser et al., 2016; H. Li et al., 2015; Wada et al., 1998; Wu et al., 2003). Comparing our PRO-cap-defined annotations in *S. pombe* with TSSs determined

by Eser *et. al*. revealed a much closer agreement between these two independent approaches than with PomBase annotations, though some differences still remained (Figure 2.2a,b).

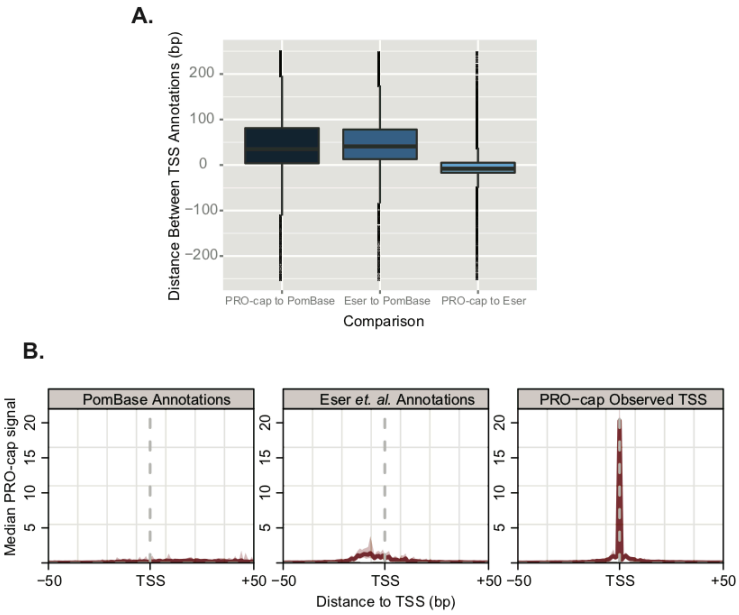

Figure 2.2: Comparison of PRO-cap-observed transcription start sites with previous annotations of *S. pombe* TSSs (Eser et al., 2016). **a.** Box plots showing the distribution of distances to reannotated TSSs from previous positions (positive values are downstream, negative values are upstream of previous annotations). **b.** Composite PRO-cap signal intensity (background subtracted) centered on each set of TSS annotations for the same genes (N = 2646).

For *S. cerevisiae*, we compared our PRO-cap observed TSSs with the 5' ends of the longest major transcript isoforms identified by Pelechano *et al.* (2013) (Figure 2.5). Though a striking enrichment of PRO-cap signal occurred directly over expected 5' ends of these transcript isoforms (Figure 2.5b, d), we found many preferred TSSs to be more downstream, consistent with the existence of many shorter major transcript isoforms in *S. cerevisiae* (Figure 2.5b) (Brannan et al., 2012; Core et al., 2012; Pelechano, Wei, & Steinmetz, 2013).

To assess the amount of signal attributable to background in our PRO-cap experiments, we simultaneously prepared PRO-cap libraries from the same starting material in which nascent RNAs were not treated with Tobacco Acid Pyrophosphatase (TAP) prior to 5' adaptor ligations. These sequenced RNAs represent molecules lacking a 5' cap structure, and therefore are unlikely to represent true transcription start sites. Looking at a 100 bp window around annotated or observed TSSs in *S. pombe*, we observed a highly focused enrichment of reads over background centered on the observed TSS (Figure 2.1e), a trend that does not exist when centering previous annotations (Figure 2.1d). The same analysis in *S. cerevisaie* revealed PRO-cap signal highly consistent with the transcript isoform 5' ends used, suggesting good agreement of our data with previous work (Figure 2.5d) (Pelechano et al., 2013; Wada et al., 1998). Nonetheless, we attained an even greater reduction in signal dispersion by centering PRO-cap signal around the most preferred base (Figure 2.5e). Plotting PRO-seq signal relative to TSSs confirms that average 5' boundaries global RNA polymerase profiles more closely reflect the observed TSS compared with previous annotations in both fission and budding yeast (Figure 2.3). Moreover, PRO-cap signal in *S. cerevisiae* was found to be strictly enriched downstream of previously characterized TATA-like motifs and TFIIB positions (Figure 2.4)(Rhee & Pugh, 2012).

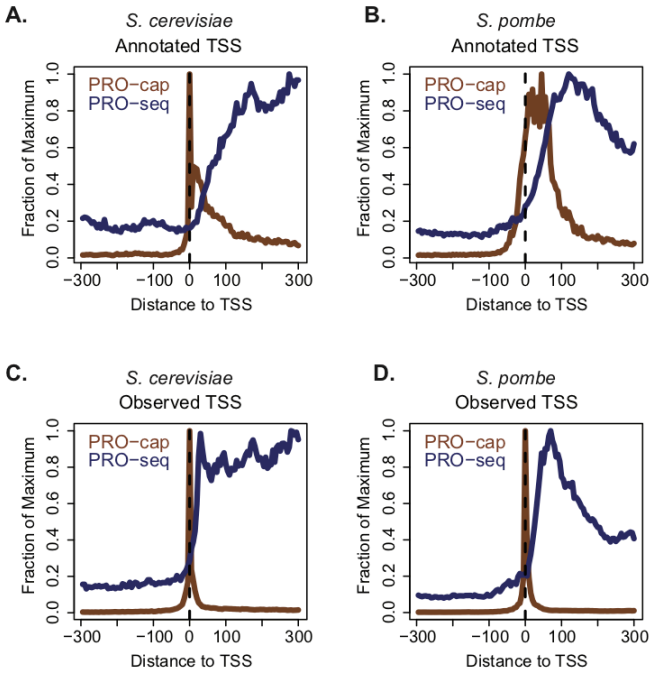

Figure 2.3: Observed transcription start sites define the 5' boundaries of nascent transcription. a, b. Average PRO-cap and PRO-seq signal around the annotated start sites in *S. cerevisiae* (a) and in *S. pombe* (b). **c, d.** Average PRO-cap and PRO-seq signal around the observed TSS in *S. cerevisiae* (c) and in *S. pombe* (d). The y-axis shows the median signal within 5 bp bins around the TSS as a fraction of the maximum bin.

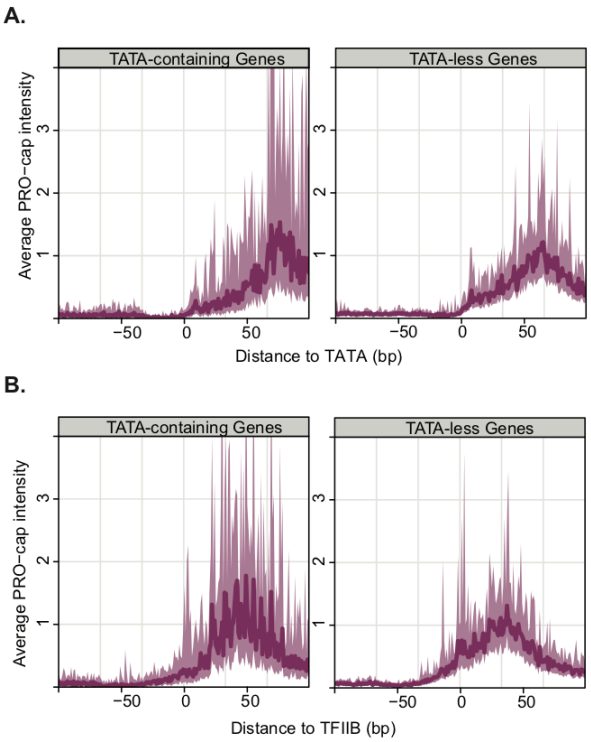

Figure 2.4: PRO-cap signal in *S. cerevisiae* is consistent with characterized TATA box and TFIIB positions. a. Median PRO-cap signal for each base is plotted relative to previously characterized TATA box, or TATA-like motif, coordinates. **b.** Median PRO-cap data at each base relative to TFIIB position for TATA-containing and TATA-like gene groups. Gene groups, as well as motif and TFIIB coordinates were identified previously (Rhee & Pugh, 2012). The y-axis for all plots shows the median read counts for each base. The 12.5% and 87.5% quantiles are shown in lightly shaded regions.

To further evaluate our observed TSSs, we assessed the effect of centering available MNase-sequencing data around the preferred base (DeGennaro et al., 2013; Z. Hu et al., 2014). Composite MNase-seq profiles around the observed TSSs in *S. pombe* show a more highly phased pattern when compared with annotated TSSs (Figure 2.1f). Centering MNase-seq data from *S. cerevisiae* around the observed TSSs produced a modest reduction in phasing pattern of the MNase-seq reads, relative to the longest TIF-seq isoform 5' ends (Figure 2.5f). Together, our results indicate that PRO-cap defined TSSs are truly an accurate representation of transcriptional start sites in *S. pomb*e and *S.*

cerevisiae that can improve our analysis of early transcription elongation in yeast, especially in *S. pombe*, where less mapping has been performed to date.

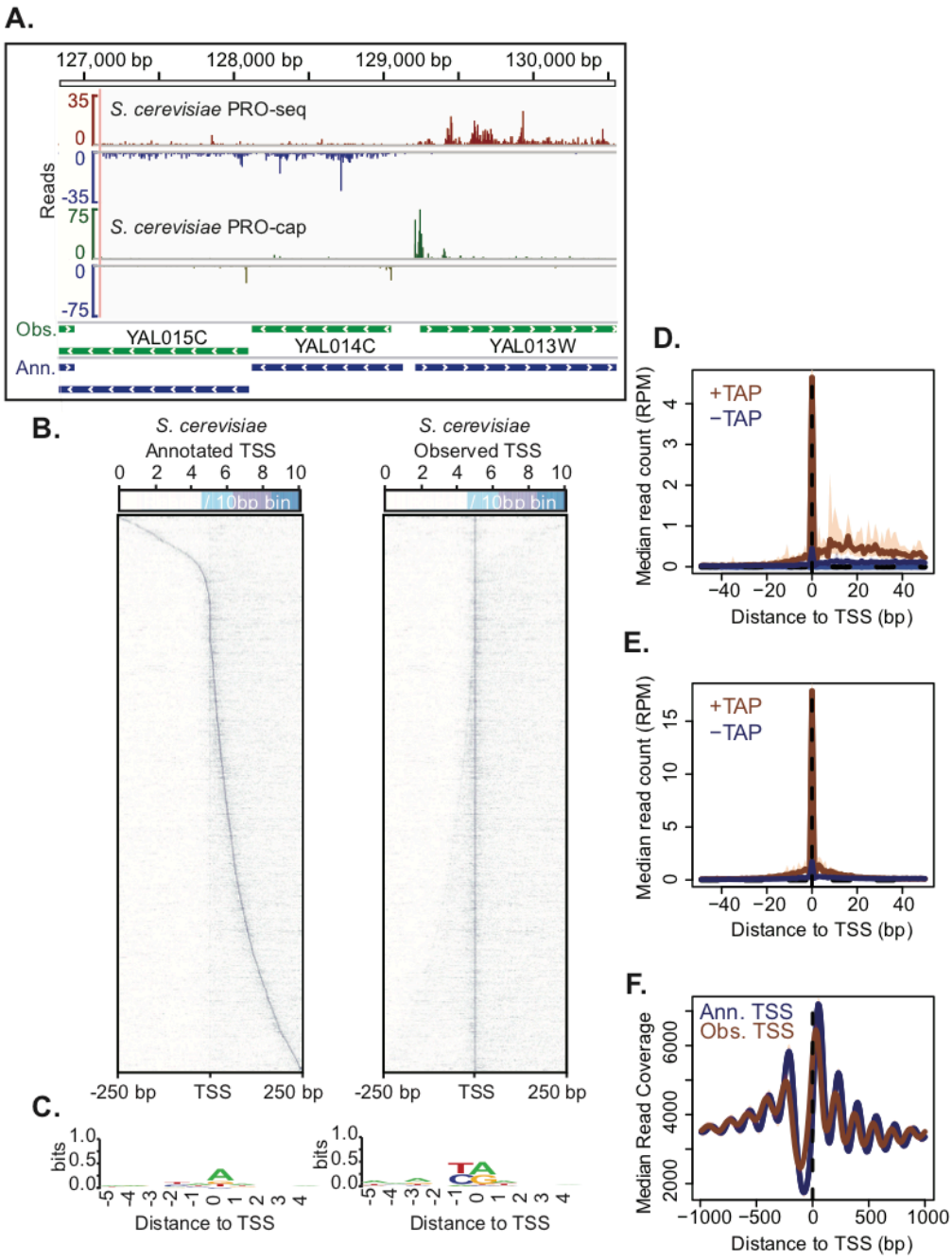

Figure 2.5. PRO-seq and PRO-cap capture transcription elongation and initiation genome-wide in *S. cerevisiae.* **a.** Browser tracks of PRO-seq (plus strand: red; minus strand: blue) and PRO-cap data (plus strand: green; minus strand: yellow) derived from *S. cerevisiae*. Green gene models show the re-annotated, 'observed' transcription start-sites

based on PRO-cap data. Blue gene models correspond to the longest major transcript isoform identified by (Pelechano et al., 2013). **b.** Heatmaps of *S. cerevisiae* PRO-cap signal for each base within +/- 250 bp around either the annotated TSS (left) or PRO-cap-observed TSS (right) for all active and filtered genes (N = 3403). Genes within heatmaps are sorted by increasing downstream distance of observed TSS relative to annotated TSS. **c.** Sequence logos of 10 bp sequence centered on annotated TSS (left) and observed TSS (right) were generated using WebLogo (Crooks et al., 2004). **d.** Median PRO-cap signal from samples prepared either with or without Tobacco Acid Pyrophosphatase (TAP) treatment centered on annotated TSS. The TAP-minus samples represent empirical levels of background for each genomic position. **e.** Median PRO-cap signal from samples prepared either with or without TAP treatment centered on observed TSS. **f.** Median MNase-seq coverage centered on annotated TSSs (blue) or observed TSSs (brown). For the meta-gene plots, the y-axis shows the median read counts for each base-pair (d & e), or median read coverage within 10 bp bins (f). In figures d, e and f, the 12.5% and 87.5% quantiles are shown in lightly shaded regions.

Budding and fission yeast exhibit vastly different polymerase distributions across genes. Using PRO-seq based maps of RNA Pol II in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, we sought to identify general features of transcription that are species-specific. To minimize the effects of transcription emanating from neighboring genes, we identified a set of filtered genes at least 1 kilobase from any other genes on the same strand (*S. pombe*: n = 874; *S. cerevisiae*: n = 1101). We then generated scaled composite profiles in which the region between 300 bp downstream of the observed TSS and 300 bp upstream of the cleavage and polyadenylation signal (CPS) is scaled to the same number of bins, while everything outside this gene-center region is not scaled (10 bp bins). Importantly, PRO-seq signal increases immediately downstream of the observed TSS (Figure 2.6a, b), further indicating PRO-cap–based observed TSSs are genuine start sites for these genes.

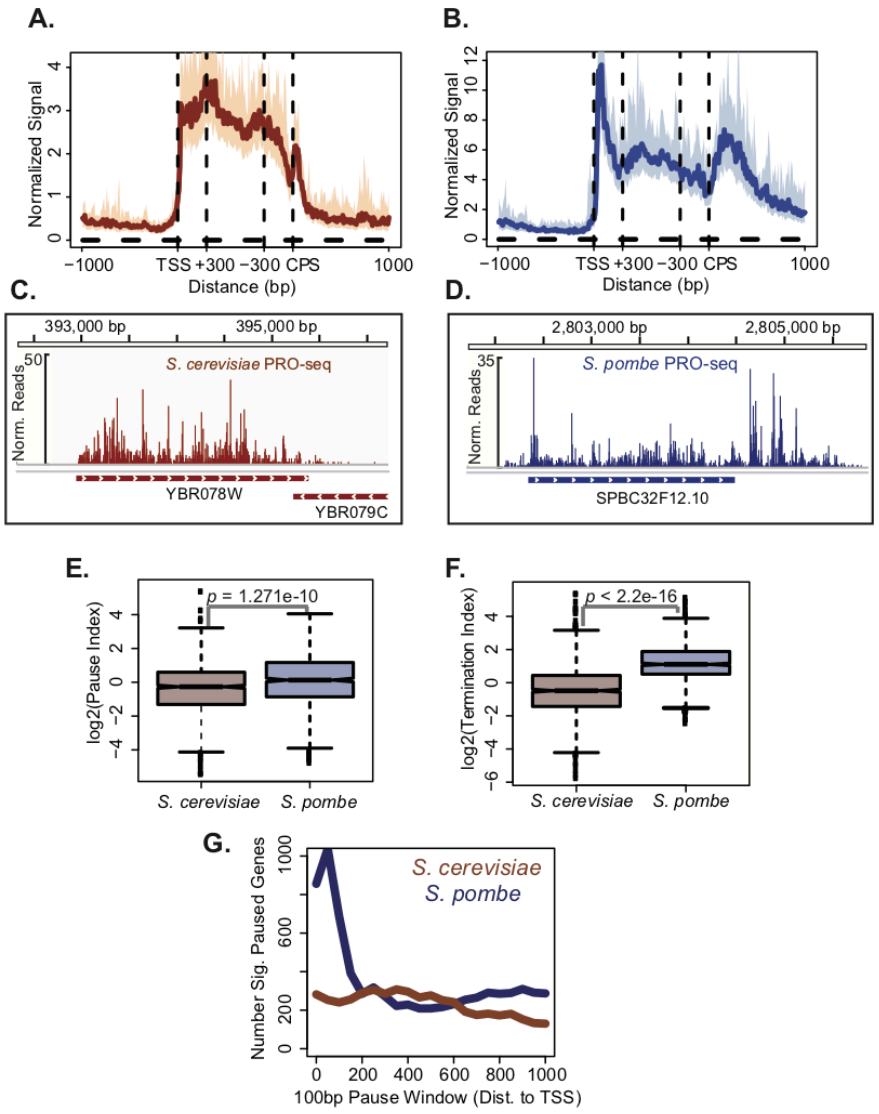

Figure 2.6. PRO-seq reveals distinct transcription elongation profiles in *S. cerevisiae* and *S. pombe*. a, b. Median PRO-seq read count across all active and filtered genes in *S. cerevisiae* (a) or *S. pombe* (b) that are separated from neighboring genes on the strand by at least 1 kb (*S. cerevisaie*: N = 1101; *S. pombe*: N = 874)**.** Bins between the +300 and -300 bp marks are scaled based on gene length, while upstream and downstream of this center region, 10 bp bins were used. The shaded regions around the curves represent the 12.5% and 87.5% quantiles. **c, d.** Representative genes from *S. cerevisiae* (c) and *S. pombe* (d) with PRO-seq read counts plotted above. **e, f.** Box plots of pausing index (e) or termination index (f) values calculated for all genes that were included in a & b. P-values represent results of Student's *t*-test **g.** A test for enrichment of pausing near the promoter versus other gene regions. Reads were counted within a sliding 100 bp window from 0 to 1000 bp from the TSS of all filtered, active genes and divided by the counts within the remaining mappable gene length. Fisher's exact test was used to determine the number of significantly paused genes (adjusted $p < 0.01$).

Similar to previous studies of transcription in *S. cerevisiae* (Churchman & Weissman, 2011; E. J. Steinmetz et al., 2006), we observed a relatively uniform distribution of RNA Pol II across transcription units, with transcription extending approximately 200-300 bp beyond the CPS on average (Figure 2.6a). Notably, NET-seq profiles exhibit a modest decline in read density downstream of the TSS, which was not present in our PRO-seq data (Figure 2.7a). This distinction may reflect differences in the forms of nascent RNA captured by each technique. For instance, NET-seq captures all Pol II-associated RNAs with a 3' OH, such as those in backtracked complexes. Pol II backtracking may occur more frequently during early elongation (Churchman & Weissman, 2011) and not be captured by PRO-seq, which requires complexes capable of running on. Nonetheless, in contrast to the Pol II distributions of *S. cerevisiae* measured by either technique, our composite RNA Pol II profiles across genes in *S. pombe* show two more prominent peaks of Pol II density. The first peak is immediately downstream of the TSS, while a second, broader peak occurs downstream of the 3' CPS. Furthermore, Pol II in *S. pombe* appears to transcribe much farther beyond the CPS when compared with *S. cerevisiae* (Figure 2.6b). Visualization of PRO-seq density for representative genes, YBR078W in *S. cerevisiae* and SPBC32F12.10 in *S. pombe*, illustrates these same characteristics (Figure 2.6c, d). Regarding intron-exon junctions, Pol II density around fission yeast splice sites exhibited features similar to those described in *Drosophila* (H. Kwak et al., 2013), whereas the rarity of introns in budding yeast precluded a more thorough comparison (Figure 2.8a, b).

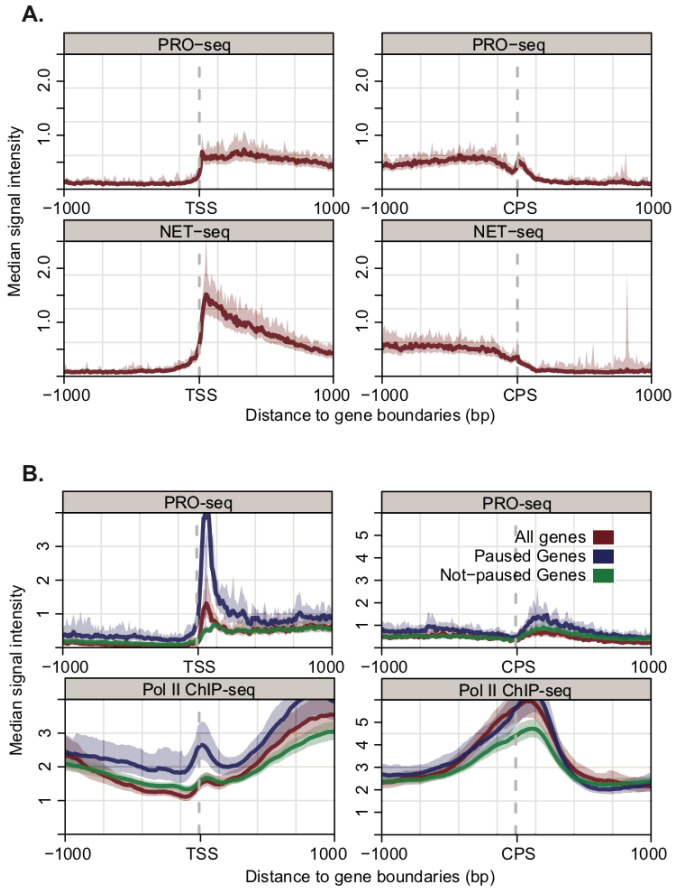

Figure 2.7. **Comparison of PRO-seq in yeast with other methods of mapping Pol II positions genome-wide. a.** Composite profiles of *S. cerevisiae* PRO-seq (top) or NET-seq data (bottom) (GEO accession: GSM617027)(Churchman & Weissman, 2011) relative to the observed TSS or CPS (N = 1614). **b.** Composite profiles of *S. pombe* PRO-seq (top) or Pol II ChIP-seq data (bottom) (GEO accession: GSM1202003)(DeGennaro et al., 2013) for not-paused (N = 2088), paused (N = 656), and all genes (N = 3214) relative to the observed TSS (left) or CPS (right).

To compare the enrichment of RNA Pol II density downstream of the TSS between *S. pombe* and *S. cerevisiae*, we calculated a 5' pausing index for each gene considered in the composite profiles. 5' pausing index was defined as the ratio of read density within the first 100 bp downstream of the TSS over the gene-body (TSS+200 to CPS). On average, *S. pombe* genes exhibited significantly higher pausing indexes when compared with *S. cerevisiae* (Figure 2.6e). Similarly, to assess the amount of pausing or

slowing beyond the 3' CPS, we computed a 3' termination index for these filtered genes. The post-CPS termination index was defined as the ratio of the highest density 100 bp window within 500 bp downstream of the CPS to the gene-body (defined above). We found that *S. pombe* genes on average also have a significantly higher termination index than *S. cerevisiae*.

To investigate the significance of promoter-proximal pause-like distributions observed in *S. pombe*, we systematically classified genes as being paused or not paused. Genes classified as paused were required to show a significant enrichment (Bonferroni corrected $p < 0.01$) of mappable reads within the promoter-proximal region (TSS to +100bp) relative to the gene body, as assessed by Fisher's exact test. Not paused genes were required to have a p-value > 0.99 suggestive of no enrichment of promoter-proximal reads. Based on the criteria above, of the 3214 filtered *S. pombe* genes (see Methods), we identified 901 genes as being paused and 2133 not paused. The relative amount of active genes identified as paused in *S. pombe* (28%) is thus lower than that observed for humans (41%), using the same criteria (Core et al., 2008). Though there were few functional class (Thomas et al., 2003) enrichments for paused or not paused genes, highly paused genes (defined in next results section) were significantly enriched for structural constituents of the ribosome ($p = 3.84 \times 10^{-4}$), which are often highly paused in other species (Min et al., 2011).

Broad peaks of RNA Pol II have been observed beyond the 3' CPS of many genes in *S. pombe* (Castel et al., 2014; Coudreuse et al., 2010; Sansó et al., 2012), yet to our knowledge, no group has identified a prominent peak downstream of the TSS. We thus compared PRO-seq profiles for highly paused, not paused, and all genes with publicly

available Pol II ChIP-seq data in fission yeast (DeGennaro et al., 2013). Both data types showed broad Pol II enrichment downstream of the CPS for all gene types (Figure 2.7b, right). Furthermore, consistent with our gene classification based on PRO-seq, highly paused genes also showed a promoter-proximal Pol II ChIP-seq peak that was greater than that of the other gene groups, further suggesting an enrichment of elongating Pol II downstream of the TSS for these genes (Figure 2.7b, left).

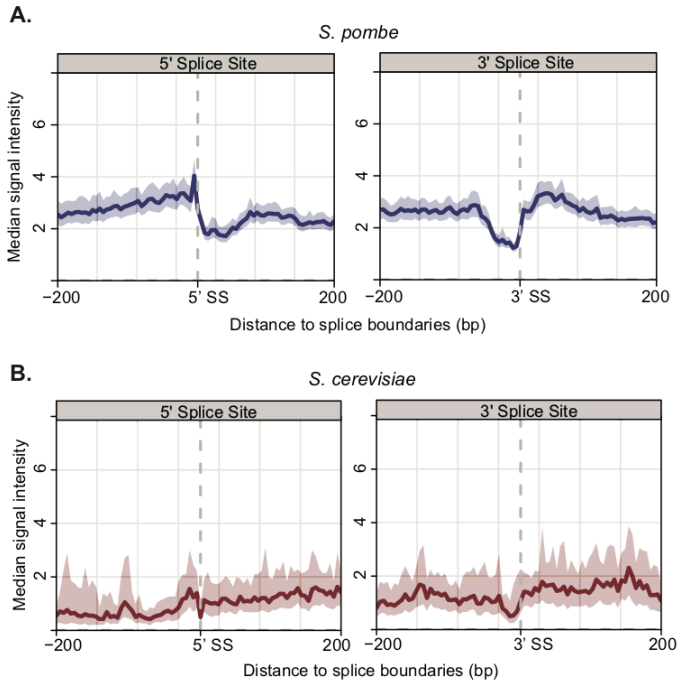

Figure 2.8: PRO-seq signal around splice sites. a. Composite profile of PRO-seq signal around all annotated 5' splice sites (left) and 3' splice sites (right) in *S. pombe* (N = 5044). **b.** Composite profile of PRO-seq signal around all annotated 5' splice sites (left) and 3' splice sites (right) in *S. cerevisiae* (N = 377).

To assess whether the observed pause peaks are specific to the promoter-proximal region, we repeated the calculation of pausing index and significance using a sliding 100 bp window starting every 50 bases between the TSS and 1000 bp downstream (20 measurements per gene). We then tallied the number of genes that were identified as significantly paused within each window relative to the remaining gene length. In *S.*

pombe we found that the most promoter-proximal regions have significantly larger number of genes identified as paused, while in *S. cerevisiae* there is not a significant difference in any given window (Figure 2.6g). These results strongly indicate the existence of a previously undescribed form of promoter-proximal pausing in *S. pombe* that is absent in *S. cerevisiae*.

Average Pol II density across transcription units of fission and budding yeast reveals large, previously uncharacterized, differences in transcription mechanisms between these divergent yeast. While our data in *S. cerevisiae* corroborate the wealth of available information describing Pol II distributions in this system (Churchman & Weissman, 2011; Jordán-Pla et al., 2014), applying the same technique in *S. pombe* has identified Pol II dynamics that resemble mechanisms pervasive in metazoans.

Paused and not paused genes in *S. pombe* exhibit distinct Pol II patterns around their +1 nucleosome. Having observed the distribution of RNA Pol II in fission yeast, we further investigated differences between paused and not paused genes. Paused and not paused genes, identified using the combined replicate data, were further refined as high-confidence paused and not paused genes by requiring that they are also significantly paused ($p < 0.05$) or not paused ($p > 0.95$) in each biological replicate (paused: N = 656; not paused: N = 2088). An example of PRO-seq profiles for a highly paused gene and a not paused gene are displayed in Figure 2.9a and b, respectively. Average PRO-seq profiles centered on the observed TSS in *S. pombe* show a striking signal enrichment, specific to paused genes, largely confined within the first 100 bp downstream of the TSS (Figure 2.9c).

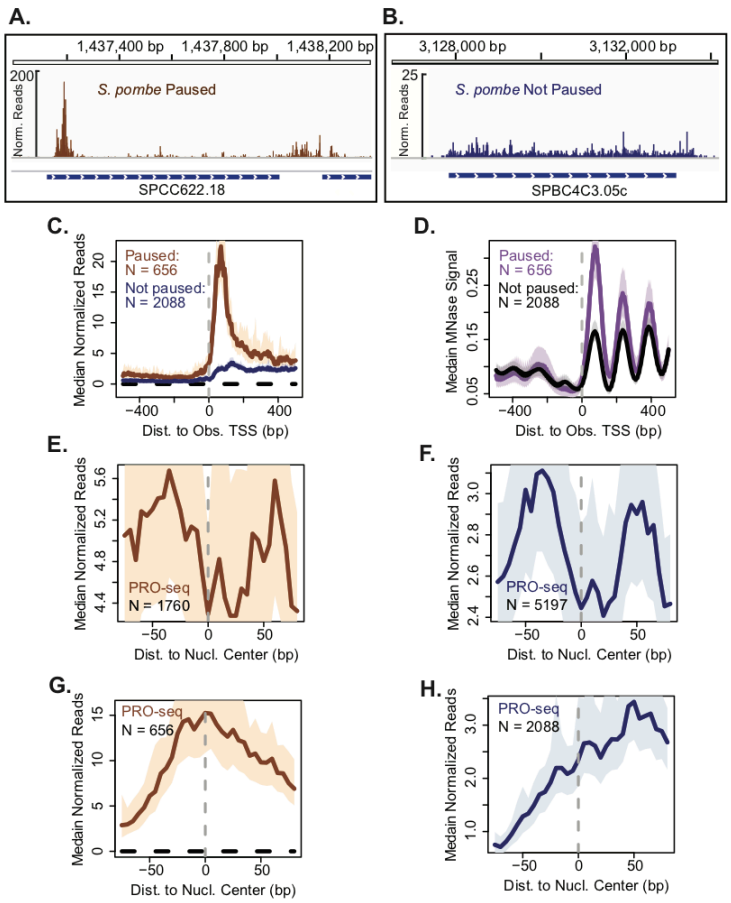

Figure 2.9. Pol II distributions at paused genes in *S. pombe* are coupled with increased nucleosome occupancy or positioning. **a, b.** Browser images of PRO-seq read counts across *S. pombe* genes classified as either high-confidence paused (a), or not paused (b). **c.** Median PRO-seq read counts within 1kb upstream and downstream of the observed TSS of paused and not paused genes. **d.** Median MNase-seq read coverage within 1kb upstream and downstream of the observed TSS of paused and not paused genes. **e, f.** PRO-seq signal around gene-body nucleosome centers for paused (e) and not paused genes (f). **g, h.** PRO-seq signal around +1 nucleosome centers for paused (g) and not paused genes (h). For meta-gene plots in c-h, medians reflect 5 bp bins, and the 12.5% and 87.5% quantiles are shown as lightly colored areas. All panels represent profiles of combined wild type biological replicates.

The transcription process is tightly coupled to the chromatin environment (Venkatesh & Workman, 2015), and nucleosomes are known to form a barrier to transcriptional elongation, slowing or stalling RNA polymerases as they unwind the coiled DNA around the histone octamer (M. A. Hall et al., 2009). We analyzed an

existing MNase-seq dataset to explore the organization of nucleosomes relative to Pol II initiation and pause sites. Interestingly, average MNase profiles relative to the observed TSSs revealed a more highly phased nucleosome pattern for paused genes than for not paused genes (Figure 2.9d). The largest difference in MNase-seq signal was at the +1 nucleosome position, possibly suggesting that this nucleosome is more precisely positioned, or maintains higher occupancy within paused genes.

To assess the distribution of elongating polymerases around nucleosomes in *S. pombe*, we displayed average PRO-seq density relative to previously identified nucleosome centers (Givens et al., 2012). For both paused and not paused genes, we parsed nucleosomes within gene boundaries based on their relative positions to the TSS. Nucleosomes were split into two groups, +1 nucleosomes (first nucleosome downstream of TSS) and gene-body nucleosomes (all nucleosomes downstream of +1 center). Within the gene body, Pol II shows a similar profile around nucleosomes whether the gene is paused or not paused (Figure 2.9e & f). The distributions around gene body nucleosomes reflect known properties of transcription through chromatin, where Pol II slows down when it encounters the strong force of the DNA histone interaction, and speeds up when those contacts are broken as it approaches the DNA near the dyad (M. A. Hall et al., 2009). Plotting the average PRO-seq density within 160 bp around +1 nucleosome centers for paused and not paused genes revealed a remarkable difference in Pol II distribution relative to the dyad axis. Whereas Pol II shows no enrichment upstream of the +1 nucleosome center within not paused genes (Figure 2.9h), Pol II is highly enriched around the dyad axis of the +1 nucleosome of paused genes (Figure 2.9g). This data may suggest a role for histone-DNA interactions in specifying pause sites in *S. pombe*.

Conversely, Pol II pausing may be independent of the nucleosome barrier, but restrict movement of the +1 nucleosome (Struhl & Segal, 2013), inducing a more rigid positioning of downstream nucleosomes as well (Jiang & Pugh, 2009). The Pol II distributions around the +1 and gene body nucleosomes (+2 to n) indicate a general difference between the way Pol II negotiates its transit through the first and subsequent nucleosomes, as well as an interplay between paused Pol II and the +1 nucleosome that distinguishes paused and not paused genes.

Deletion of Spt4 increases Pol II density genome-wide in both budding and fission yeast. The striking resemblance of Pol II distributions in *S. pombe* to promoter-proximal pausing in metazoans led us to postulate that there is a pausing factor conserved in yeast playing a role in this process. As both *S. cerevisiae* and *S. pombe* lack NELF subunits, the next logical candidate was the Spt4-Spt5 complex. To investigate functional divergence of this complex between *S. cerevisiae* and *S. pombe,* we prepared PRO-seq libraries from strains lacking Spt4. Wild type and *spt4Δ* PRO-seq libraries from each species were prepared from two biological replicates. The absence of all transcription and messenger RNA from the Spt4 locus confirms the deletion of the gene in each yeast (Figure 2.10a, b).

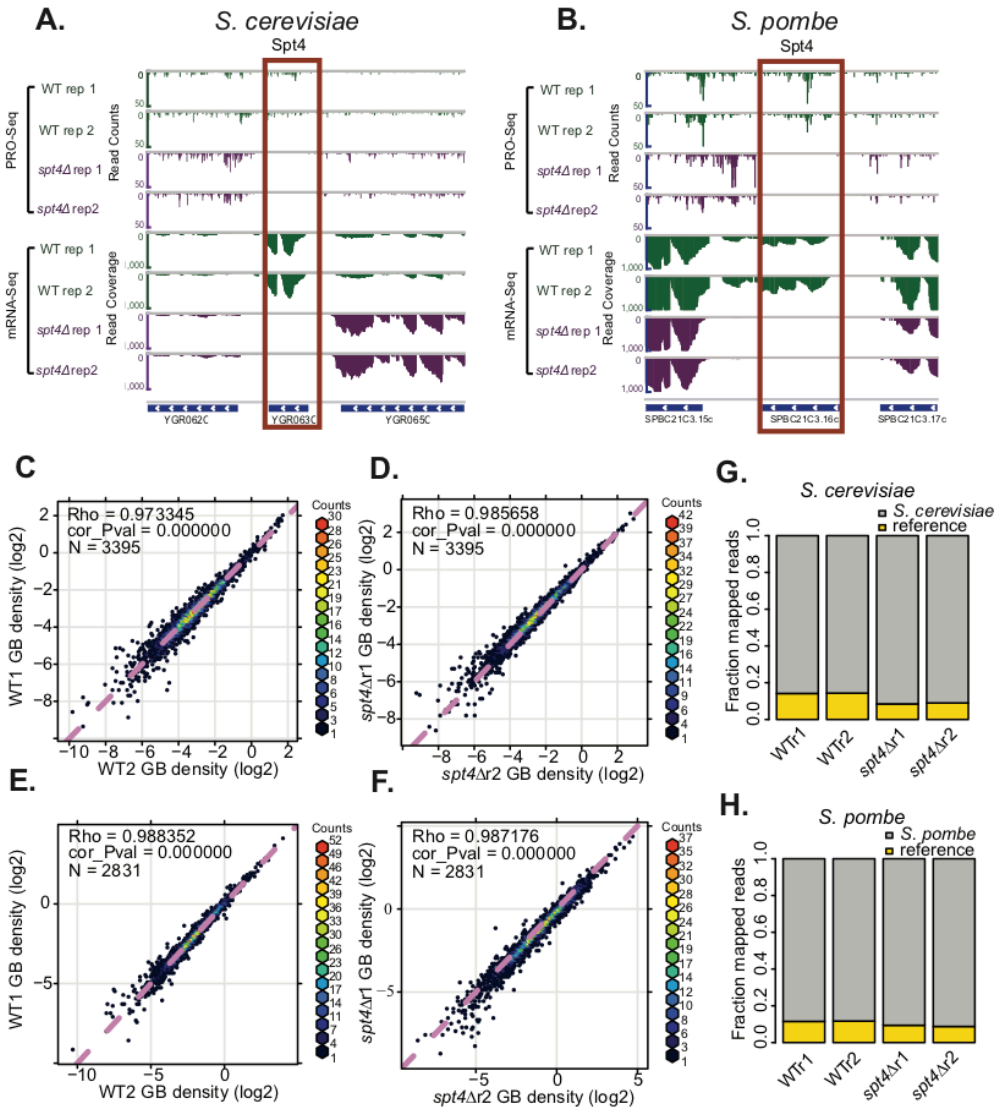

Figure 2.10. Spike-in libraries provide a method for accurate normalization between samples. a, b. Browser shot of the Spt4 locus in *S. cervisiae* (a) and *S. pombe* (b). Displayed are the PRO-seq (top four tracks) and mRNA-seq data (lower four tracks) for two biological replicates of WT cells (green) and *spt4Δ* (purple). The gene upstream of the Spt4 locus in *S. cerevisiae* (YGR0465C) was found to be significantly changed at the mRNA level in *spt4Δ* cells. **c-f.** Scatter plots of PRO-seq gene body read density (reads per mappable base) between biological replicates. Counts are normalized based on the relative amount of reads that align uniquely to the spiked-in genome for each library. **g, h.** Stacked bar plots displaying the fraction of uniquely mapped reads aligning to the genome of interest or the spike-in (reference) genome for each PRO-seq sample in *S. cerevisiae* (g) and *S. pombe* (h).

Anticipating the possibility of global transcriptional changes upon deletion of Spt4, we used an approach to normalize sequencing depth between samples based on a spike-in reference control. In this approach, we took advantage of the fact that over 99% of mappable 36 nt sequences derived from *S. pombe* and *S. cerevisiae* can be distinguished. Similar to an approach developed previously for micro-array analysis of labeled mRNA (M. Sun et al., 2012), we included an exogenous, whole yeast spike-in control in our experiments, which was carried out as follows: all samples and biological replicates to be compared were required to have precisely the same starting OD_{600} before harvesting. Immediately prior to harvesting, a fixed volume of the divergent "spike-in" yeast culture was added to each sample. After preparing and sequencing PRO-seq libraries from the mixed samples, we aligned sequencing reads to a combined genome of *S. cerevisiae* and *S. pombe* with the requirement that sequences align to only one location to remove any ambiguous reads. The proportion of spike-in to experimental reads was highly similar between biological replicates (Figure 2.10g, h), and normalization factors were calculated from the relative number of uniquely mapped reads from the spiked-in genome between samples (Table 2.1). To verify that this approach accurately scales the sequencing depth of each sample, we compared the normalized gene-body read densities between biological replicates, resulting in Spearman's $\rho > 0.97$ (Figure 2.10c-f). Importantly, replicate comparisons are nearly symmetrical about the diagonal line $x = y$, indicating this normalization accurately accounts for any technical differences in library preparation and sequencing.

Experiment	Sample	spike In Genome	Ribosomal	Genome	Unmapped	Total Reads	Norm. Factor
PRO-Seq	SC_WTr1	348939	7314357	2113476	1351857	11128629	1
PRO-Seq	SC_WTr2	321265	7097561	1917056	1144824	10480706	0.920691009
PRO-Seq	SC_Spt4KOr1	309441	6780503	3337629	1073766	11501339	0.88680543
PRO-Seq	SC_Spt4KOr2	323311	7680125	3268891	1886762	13159089	0.926554498
mRNA-seq	SC_WTr1	1631698	NA	35165891	6766362	43563951	0.856646354
mRNA-seq	SC_WTr2	1829123	NA	37399417	6702414	45930954	0.960295073
mRNA-seq	SC_Spt4KOr1	1619275	NA	35026042	6040124	42685441	0.850124242
mRNA-seq	SC_Spt4KOr2	1904751	NA	41592838	7227292	50724881	1
PRO-Seq	pombeWTr1	722562	7231534	5535779	1073132	14563007	0.837361789
PRO-Seq	pombeWTr2	540885	4659402	4079360	744834	10024481	0.626820164
PRO-Seq	pombeSpt4KOr1	405405	5511285	3898856	789841	10605387	0.469815263
PRO-Seq	pombeSpt4KOr2	862903	12837838	8960587	2101368	24762696	1
mRNA-seq	pombeWTr1	3546756	NA	34199493	3786757	41533006	0.750287962
mRNA-seq	pombeWTr2	5144621	NA	51375901	5183180	61703702	1.088303566
mRNA-seq	pombeSpt4KOr1	4107511	NA	37414945	4194282	45716738	0.868911212
mRNA-seq	pombeSpt4KOr2	4727193	NA	35433485	3977657	44138335	1
PRO-Cap	SC_CAP+Tap	NA	7930727	1865241	818844	10614812	NA
PRO-Cap	SC_CAP-Tap	NA	15093960	2268916	1148714	18511590	NA
PRO-Cap	SP_CAP+Tap	NA	10428693	11373221	2246184	24048098	NA
PRO-Cap	SP_CAP-Tap	NA	15663591	8337945	1967933	25969469	NA

Table 2.1: Sequencing alignment statistics for each experiment performed. Normalization factors refer to the relative amount of reads mapping to the spike in organism genome within each experiment (set of 4 libraries). In the PRO-cap experiments reads per million (RPM) was used for normalization between plus-TAP and minus-TAP samples.

To assess the effects of deletion of Spt4 on transcription in *S. cerevisiae* and *S. pombe*, we performed a differential expression analysis using read counts within the gene-body region. After spike-in normalization, we observed a global increase in the gene-body read density in *spt4Δ* strains in both yeasts. Using the DESeq2 R package (M. I. Love, Huber, & Anders, 2014) we identified 1,572 genes in *S. cerevisiae* (46% of filtered, active genes); and 2,291 genes in *S. pombe* (67%) as having significantly higher gene-body Pol II levels ($p < 0.01$) (Figure 2.11a & b). By contrast, only 47 (1.4%) and 156 (4.8%) genes were found to have significantly lower read counts in *S. cerevisiae* and *S. pombe*, respectively. On average, Pol II density within gene bodies of *S. pombe* and *S. cerevisiae* transcription units increased by a factor of 1.81 and 1.47, respectively.

Looking more specifically at the YGR204W locus in *S. cerevisiae* (Figure 2.11c) and SPCC622.18-1 locus in *S. pombe* (Figure 2.11d), we clearly observed a reproducible increase in the signal intensity within the gene body in *spt4Δ* biological replicates compared with wild type. Thus, the use of spike-in controls in the preparation of these data is invaluable. This approach circumvents the assumptions of an equal number of RNA sequence reads per cell, or that genes on average are unchanged, both of which are violated in mutants globally affecting transcription (M. Sun et al., 2012).

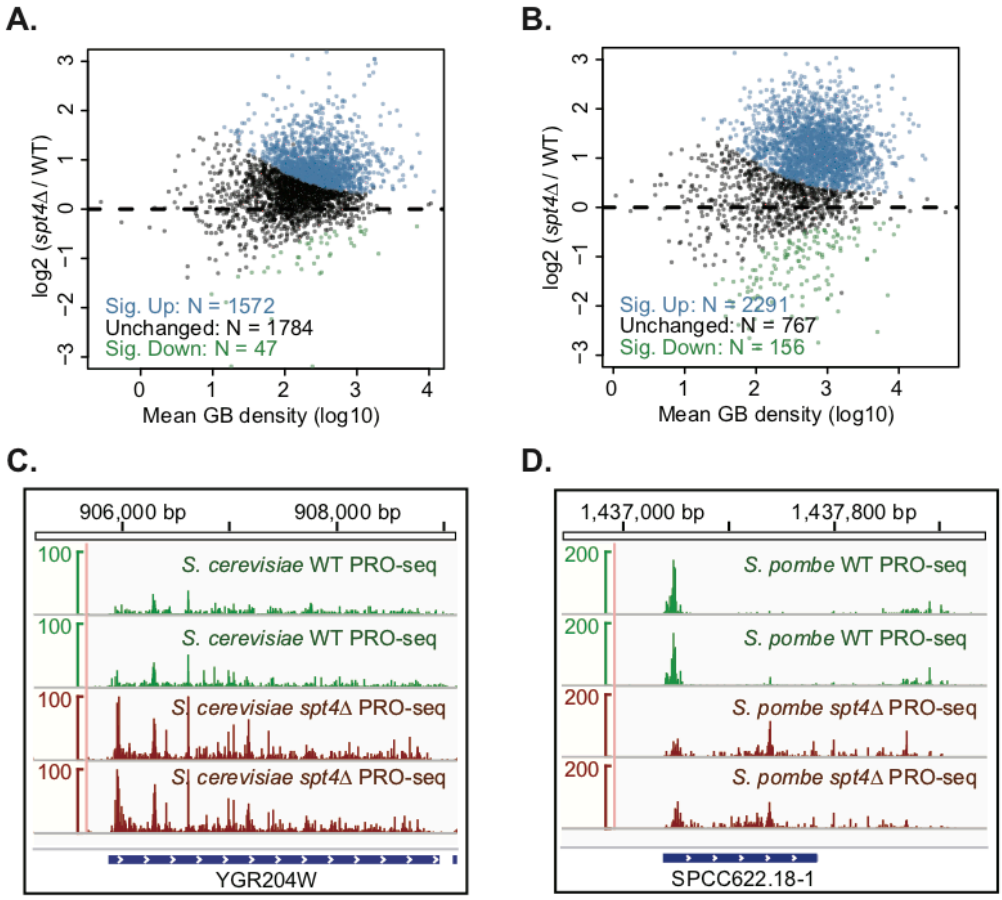

Figure 2.11. Deletion of Spt4 results in genome-wide increase in Pol II density within gene bodies of *S. cerevisiae* and *S. pombe*. a & b. MA plots showing the DESeq2-based differential expression analysis of spike-in normalized PRO-seq read-counts within the gene bodies of *S. cerevisiae* (a) and *S. pombe* genes (b). **c & d.** Example of *spt4Δ* affected gene in *S. cerevisiae* (c) and *S. pombe* (d). Browser tracks correspond to PRO-

seq data derived separately from two biological replicates of WT (green) and *spt4Δ*(red) in budding and fission yeast.

Our spike-in-normalized PRO-seq data also lends itself to a precise analysis of non-coding RNA. Antisense transcription within the gene body, as well as upstream, divergent transcription at promoters has become increasingly appreciated in yeast with the advent of high-throughput assays of RNA (Neil et al., 2009; Wilhelm et al., 2008; Xu et al., 2009). Evidence for alterations in such non-coding transcription have been reported for mutations in the elongation factor, Spt6, in *S. pombe* (DeGennaro et al., 2013). Subsets of genes used in the analysis of upstream, antisense or intragenic, antisense transcription were filtered to remove possible influence of nearby functional transcription. Unlike what has been observed in mammals (Core et al., 2008), composite profiles of PRO-cap and PRO-seq density around transcription start sites showed little indication of widespread upstream, divergent transcription at promoters in either yeast (Figure 2.12a, b, e, f), though *S. cerevisiae* tended to have higher ratios of upstream, antisense to downstream, sense transcription than *S. pombe* (compare Figure 2.12c and G). Interestingly, intragenic, antisense to sense transcription ratios were more affected in *S. pombe* than in *S. cerevisiae* by the deletion of Spt4 (compare Figure 2.12d and h). Indeed, we found that most levels of antisense transcription in fission yeast were unchanged (71%) or decreased (19%) in *spt4Δ*, suggesting that in *S. pombe,* antisense Pol II transcription may not depend on Spt4 in the same manner.

The general increase in sense-transcribing Pol II levels within a large proportion of genes likely reflects the ubiquitous role of Spt4 in positively affecting transcription processivity and elongation rate in these divergent yeasts (Hartzog & Fu, 2013; Rondón

et al., 2003; Wada et al., 1998). Elevated levels of RNA abundance are often interpreted as a higher level of gene activity or increased initiation rates. However, increases in Pol II density could also reflect a reduction of Pol II elongation rate in the case of a fixed initiation rate (Danko et al., 2013; Jonkers et al., 2014).

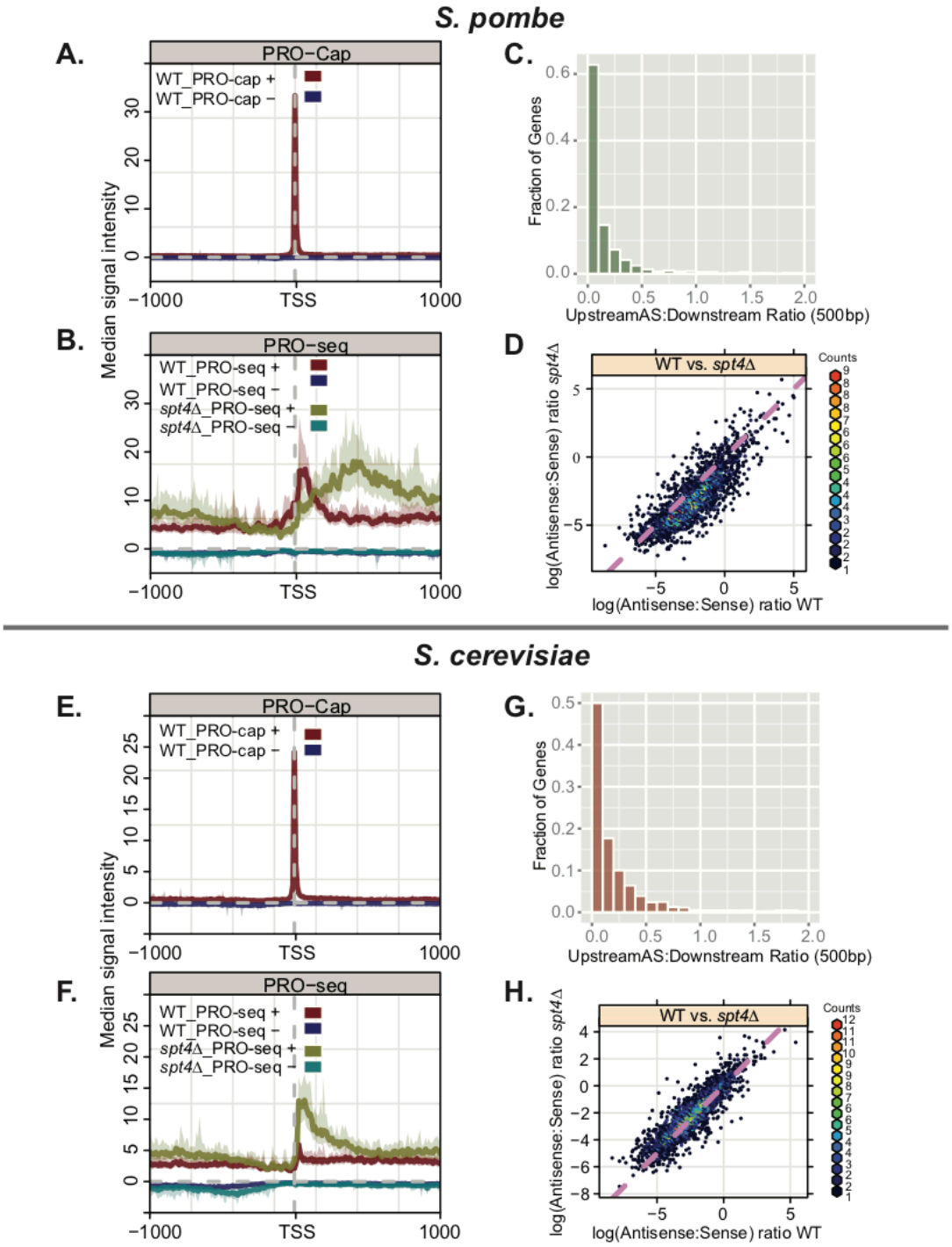

Figure 2.12: Analysis of upstream, divergent and intragenic, antisense transcription. **a-d.** Correspond to analysis of *S. pombe* PRO-seq/PRO-cap data. **a.** Composite profile of PRO-cap signal (background subtracted) for plus (+) and minus (-) strands around *S. pombe* TSSs of genes filtered for analysis of upstream, antisense transcription (described below). **b.** Composite profile of PRO-seq data on the plus and minus strand for wild type and *spt4Δ* strains around the TSS of genes used in a. **c.** Histogram of upstream, antisense

to downstream, sense ratios calculated as reads per mappable base-pair on the antisense strand from -500 to TSS over reads per mappable base-pair on sense strand from TSS to +500. **d.** Scatter plot comparing the ratio of reads per mappable base-pair on the antisense strand over that of the sense strand across the entire transcription unit between wild type and *spt4Δ* strains. Genes used in scatter plot were filtered for analysis of intragenic, antisense transcription (described below). **e-h.** Correspond to the same set of analyses as performed in a-d, but for *S. cerevisiae* PRO-seq and PRO-cap data. Genes used for intragenic, antisense analysis were selected because they did not contain functional transcription units on the antisense strand within 500 bp of either boundary (*S. cerevisiae*, n = 2432; *S. pombe*, n = 2407). Genes used for analysis of divergent (upstream, antisense) promoter transcription were additionally required to be tandemly located downstream of a gene on the same strand (*S. cerevisiae*, n = 1329; *S. pombe*, n = 1250).

Genome-wide increases in nascent transcription in *spt4Δ* are not reflected in mRNA abundance. To test the possibility of a global elongation rate defect in *S. cerevisiae* and *S. pombe* lacking Spt4, we investigated the effects of the mutation on mRNA abundance. We postulated that under conditions in which Pol II experiences a compromised elongation rate, with little change in recruitment and initiation, messenger RNA would continue to be made at roughly the same rate. To assay the relative abundance of all expressed transcripts, we prepared strand-specific poly-A selected RNA-seq libraries from the same wild type and *spt4Δ* strains of budding and fission yeast. Again, to control for the possibility of genome-wide differences in transcript abundance per cell, we used the same culture-level divergent yeast spike-in approach as used for the PRO-seq experiments (Sup. Table S1). The deletion of Spt4 was confirmed by absence of any reads from the locus (Figure 2.10a, b) and mRNA levels were found to be highly reproducible between biological replicates (Figure 2.13a-d).

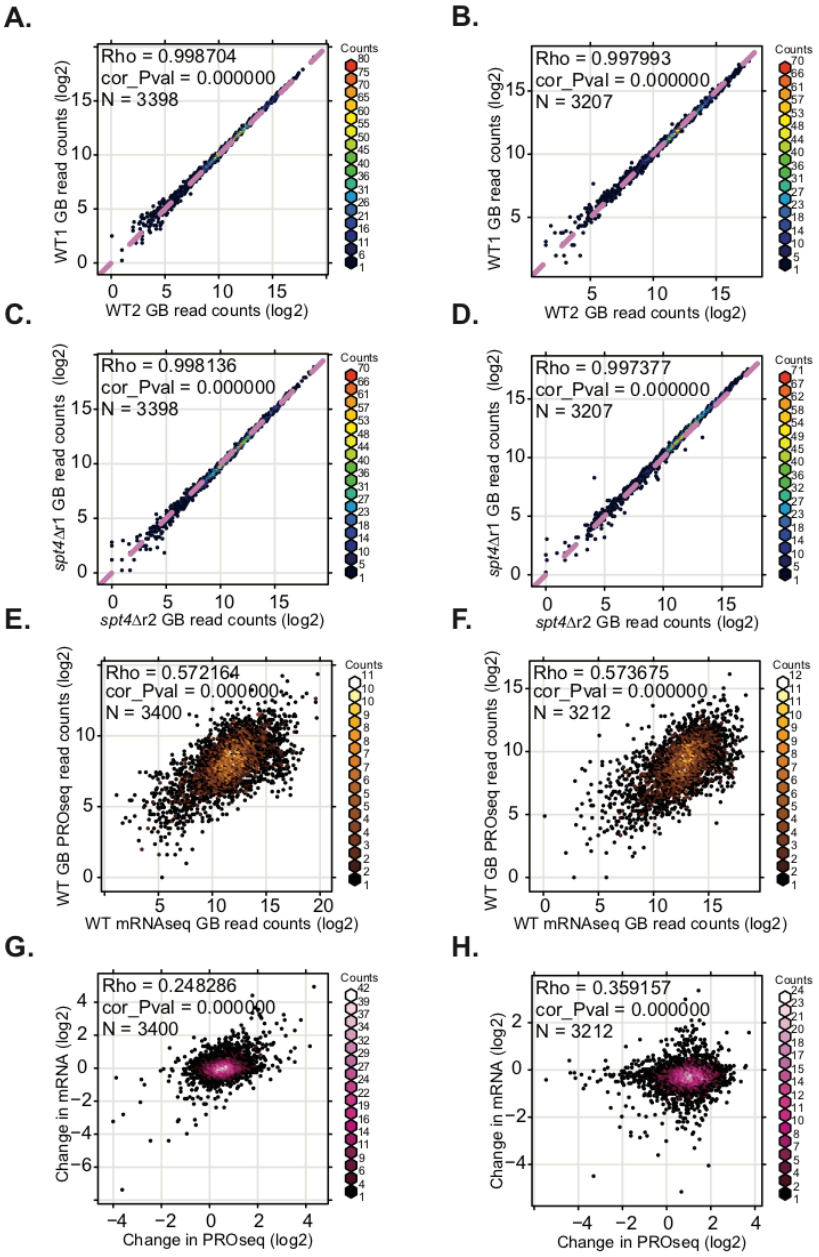

Figure 2.13: mRNA abundance is not as affected by deletion of Spt4 as nascent transcription in *S. cerevisiae* and *S. pombe*. **a-d.** Scatter plots of mRNA-seq gene-body read counts between biological replicates. Read counts were normalized based on the relative amount of reads that align uniquely to the spiked-in genome for each library. **e, f.** Correlations between nascent RNA production (PRO-seq) and mRNA abundance (mRNA-seq), assayed as normalized read counts within the gene bodies of *S. cerevisiae* (e) and *S. pombe* genes (f). **g, h.** Scatter plots comparing $\log_2$ fold change *spt4Δ* vs. wild type in gene body read density as assayed by mRNA-seq and PRO-seq for *S. cerevisaie* (g) and *S. pombe* (h). All correlation values reflect Spearman's ρ.

Using the gene-body region to quantify gene activity from PRO-seq or mRNA-seq data, we observed a moderate correlation between nascent transcription and mRNA abundance in wild-type strains (*S. cerevisiae*: Spearman's ρ = 0.572; *S. pombe*: Spearman's ρ = 0.574) (Figure 2.13e, f). We next asked if changes in transcription, resulting from the deletion of Spt4, consequently affect mRNA abundance. Here, we compared the distribution of fold changes in gene-body read-density from *spt4Δ* to wild-type libraries between PRO-seq and mRNA-seq experiments. While PRO-seq signal increases within gene bodies upon deletion of Spt4 in both species, changes in mRNA levels were either largely unchanged (*S. cerevisiae*) or slightly reduced (*S. pombe*) (Figure 2.14a & b). Although the deletion of Spt4 altered mRNA abundance of some genes in both species (Figure 2.15a, b), these changes in gene-body read density were poorly correlated with changes in PRO-seq density (*S. cerevisiae*: Spearman's ρ = 0.248; *S. pombe*: Spearman's ρ = 0.359) (Figure 2.13g, h). Overall, the effects of deleting Spt4 on Pol II density within a gene is in stark contrast with the messenger RNA abundance, supporting a model where Spt4 increases the elongation rate of RNA Pol II transcription in vivo without necessarily altering Pol II recruitment or initiation.

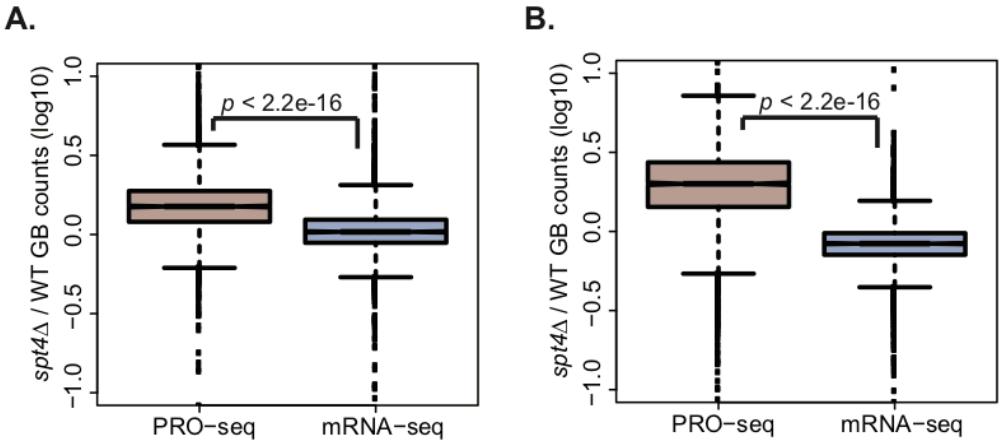

Figure 2.14. Global increase in Pol II density in *spt4ΔΔ*does not result in increased transcript abundance in *S. cerevisiae* or *S. pombe*. a & b. Boxplots comparing the fold change in gene body PRO-seq density resulting from the deletion of Spt4 with the corresponding change in poly-A selected RNA-seq in *S. cerevisiae* (a) and *S. pombe* (b). P-values represent the results of Student's *t*-test.

To determine if the effects of Spt4 deletion in *S. pombe* depended on the extent of observed promoter proximal pausing, we binned genes in deciles based on the magnitude of their pausing index. Gene characteristics, such as expression level and termination index, appeared to be un-related to pausing index (Figure 2.16a, b, e). Furthermore, there were no considerable relationships between how paused a gene is and the effect of Spt4 deletion on gene body Pol II density (Figure 2.16c) or mRNA abundance (Figure 2.16d). We surmise that Spt4, as a complex with Spt5, influences various transcriptional features independent of the degree of promoter-proximal pausing.

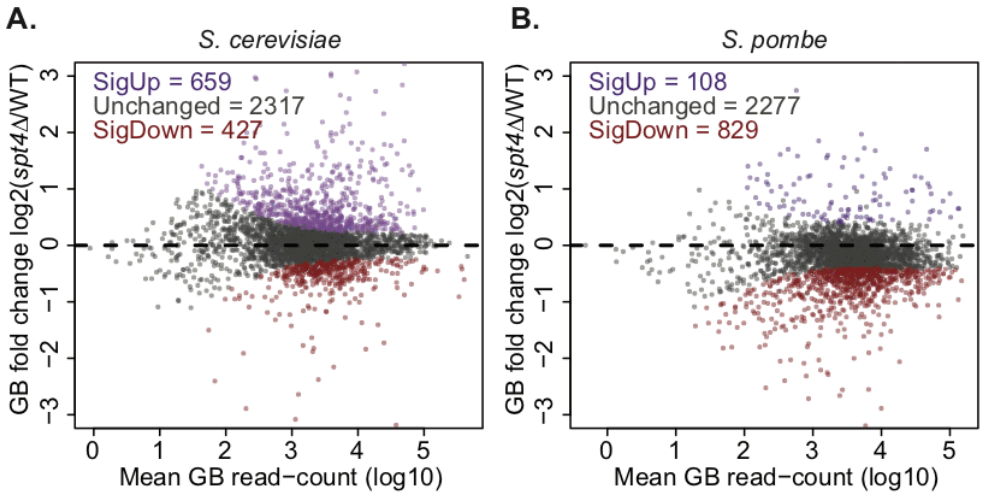

Figure 2.15: mRNA abundance is affected by the deletion of Spt4. a-b. MA plots describing the DESeq2-based differential expression analysis of spike-in normalized mRNA-seq read-counts within the gene bodies of *S. cerevisiae* (a) and *S. pombe* genes (b).

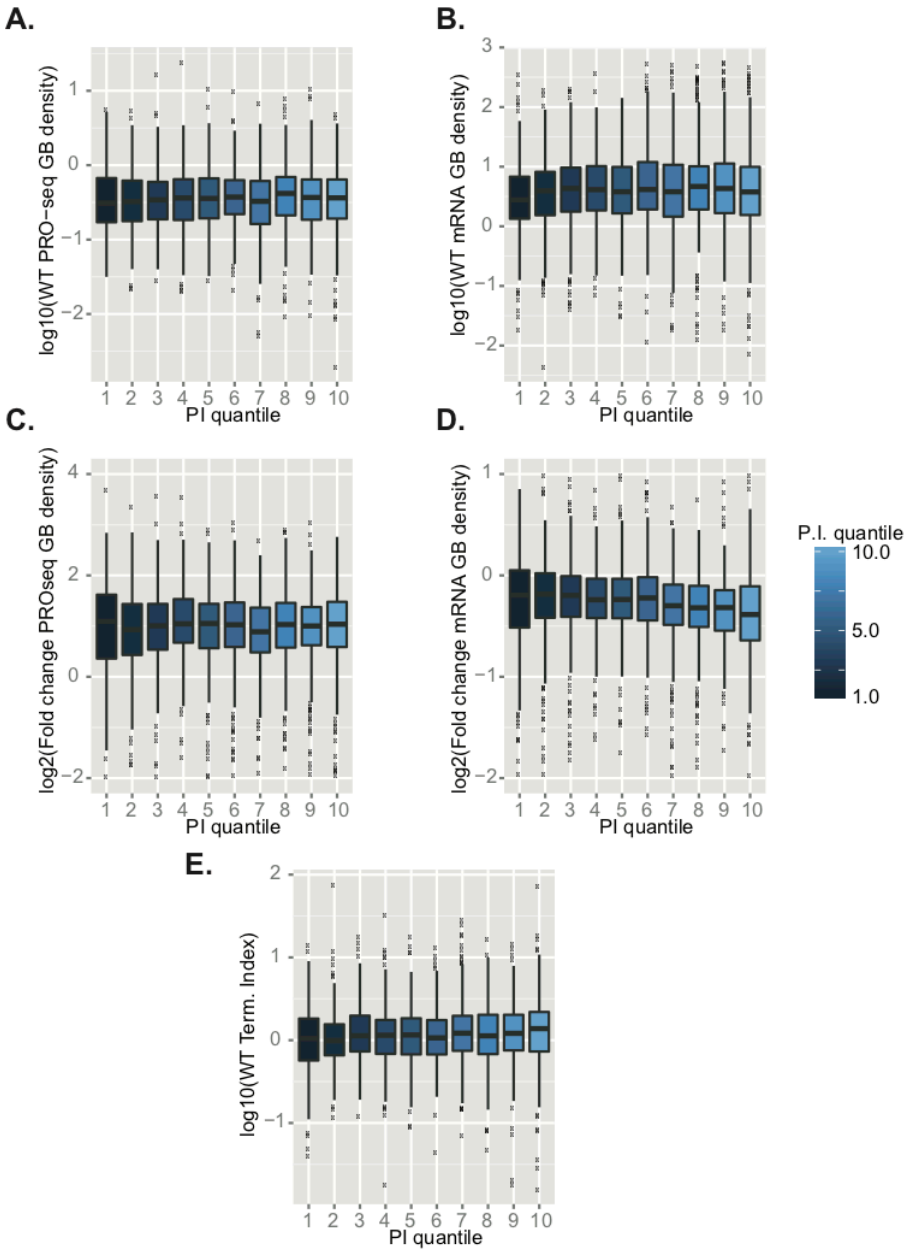

Figure 2.16: Analysis of transcriptional features as a function of increasing pausing index in *S. pombe*. a. Expression level measured as PRO-seq gene-body density. **b.** Expression level measured as mRNA-seq gene-body density. **c.** Fold change in gene-body PRO-seq density $\log_2$ (*spt4Δ*/WT). **d.** Fold change in gene-body mRNA-seq density $\log_2$ (*spt4Δ*/WT). **e.** $\log_{10}$ termination index in wild type *S. pombe*. Number of genes per decile (N = 321). Median WT pause index by decile (dec. 1 = 0.25, dec. 2 = 0.45, dec. 3 = 0.68, dec. 4 = 0.89, dec. 5 = 1.13, dec. 6 = 1.45, dec. 7 = 1.85, dec. 8 = 2.40, dec. 9 = 3.38, dec. 10 = 5.66).

Deletion of Spt4 reduces promoter proximal pausing and elongation beyond the cleavage and polyadenylation signal in *S. pombe*. While in metazoans, the Spt4-Spt5 complex acts as a pausing factor as well as a positive elongation factor (Wada et al., 1998; Wu et al., 2003), most research in yeast has focused on its role in positively influencing transcription (Hartzog & Fu, 2013). To assess the existence of additional roles for the Spt4-Spt5 complex in yeast, we investigated the effect of Spt4 deletion on the distribution of Pol II at gene 5' and 3' ends in both species. To limit the effects of nearby genes, we restricted all analyses in this section to genes greater than 1 kb in length and at least 1 kb away from neighboring genes on the same strand (*S. pombe*: n = 714; *S. cerevisiae*: n = 816). Plotting the median PRO-seq signal around the observed TSS of *S. pombe* genes for wild-type and *spt4Δ* strains revealed a dramatic shift of the Pol II density downstream in mutants (Figure 2.17a). Whereas wild-type PRO-seq profiles display the pause-like distributions immediately adjacent to the TSS, *spt4ΔΔ*strains lack this focused, promoter-proximal peak. Instead, they display a more gradual accumulation of density that is much greater than that of the wild-type within the gene body. To evaluate the generality of this phenotype for all 714 genes, we generated heatmaps displaying the fold change in PRO-seq read density ($\log_2$(*spt4Δ*/WT)) within 10 bp bins from -1 kb to +1 kb around the observed TSS (Figure 2.17b). We found that the majority of genes in the mutant show a decreased level of Pol II relative to wild-type solely within the first ~150 bases. Beyond this promoter-proximal region, *spt4Δ* strains showed heightened PRO-seq density compared with wild type. Furthermore, this mutant enrichment is most pronounced early within the gene-body, as if early elongation complexes were leaking into the downstream regions rather than undergoing a pause. To

quantify the significance of this shift in Pol II density, we calculated a pausing index for each gene in wild-type and *spt4Δ* strains. Consistent with the changes in average PRO-seq profiles in Figure 2.17a, we observed a highly significant decrease in average pausing index for all genes used in this analysis (Figure 2.17c).

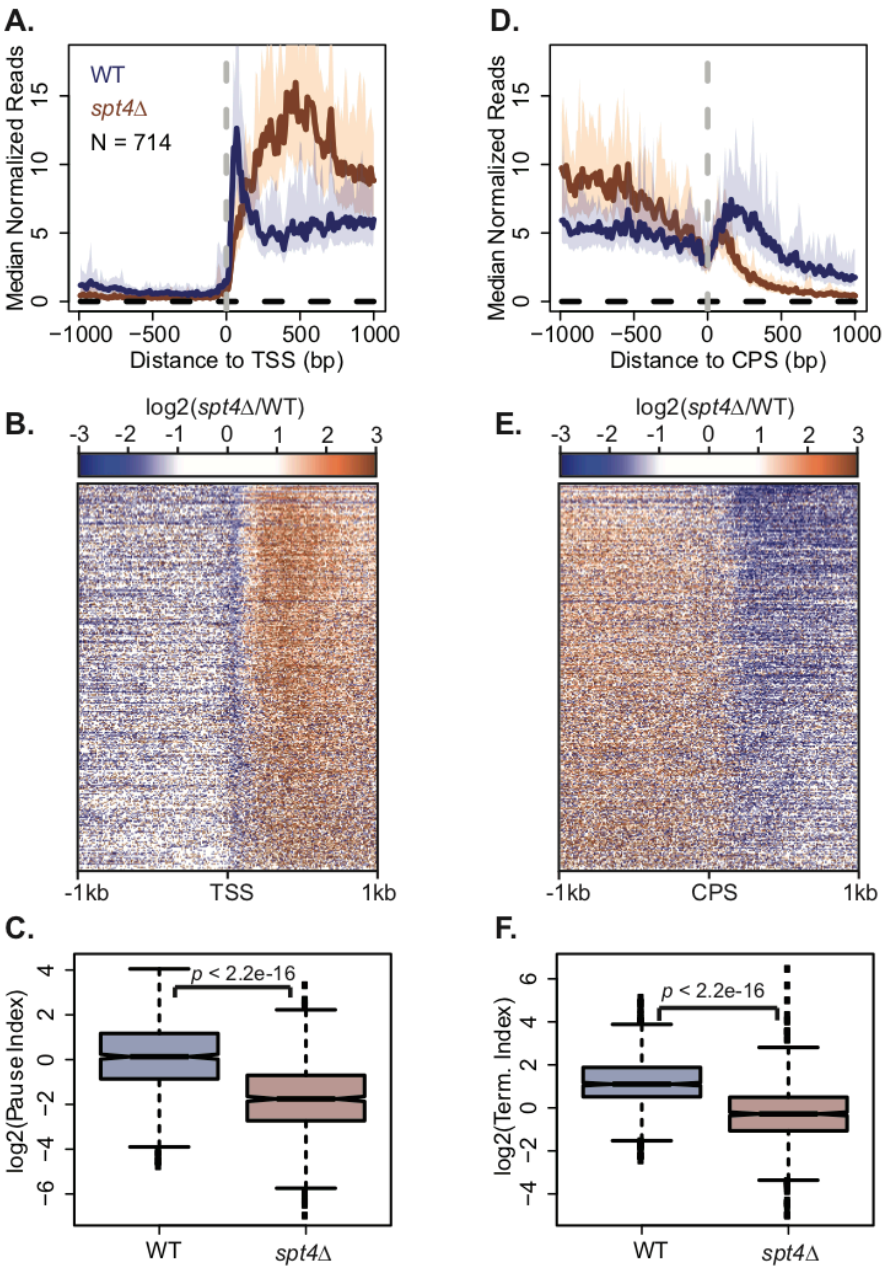

Figure 2.17. 5' and 3' ends of genes exhibit a loss of Pol II density as a result of Spt4 deletion in *S. pombe*. a & d. Median PRO-seq signal around the observed TSS (a) or

CPS (b) of active, filtered genes longer than 1 kb and separated from the boundaries of neighboring genes on the same strand by at least 1 kb. Medians reflect 10 bp bins, and the 12.5% and 87.5% quantiles are shown in light shaded regions. **b & e.** Heatmaps of $\log_2$ fold change of mutant vs. wild type per 10 bp bin around the TSS (b) or CPS (e) for all genes used in a and d. Genes within heatmaps are sorted by decreasing amount of wild-type PRO-seq reads within the first 500 bp downstream of the TSS. **c.** Box plots showing the distribution of pausing index values for WT and *spt4Δ* in *S. pombe*. **f.** Box plots showing the distribution of termination index values for WT and *spt4Δ* in *S. pombe*. P-values represent the results of Student's *t*-test.

In light of the fact that Spt4-Spt5 tracks with elongating Pol II across the entire length of genes in eukaryotes (Mayer et al., 2010; Rahl, Lin, Seila, Flynn, & McCuine, 2010), we were curious about the effect of Spt4 deletion on transcription around the cleavage and polyadenylation signal (CPS). Similar to our analysis of PRO-seq signal around the observed TSS, we calculated the median read density around the annotated CPS for all 714 genes. Strikingly, though the *spt4Δ S. pombe* strain shows elevated levels of elongating Pol II leading up to the CPS, immediately after this element, we begin to observe a much more rapid decline in polymerase density in the mutants relative to wild-type (Figure 2.17d). Looking at fold change in polymerase density across this surrounding region for each individual gene shows that this decreased transcription beyond the CPS in *spt4Δ* is broadly observable (Figure 2.17e). Finally, we assessed the significance of this change in post-CPS transcription by calculating a termination index for each gene in wild type and *spt4Δ S. pombe* strains. The combined effect of higher read density within the gene-body and lower signal beyond the CPS in *spt4Δ,* produced a significantly lower average termination index relative to wild-type (Figure 2.17f). Thus, in fission yeast, transcription beyond the CPS is dramatically reduced in the absence of Spt4.

Profiles of transcription elongation differ greatly between budding and fission yeast, yet upon deletion of Spt4 both show a similar increase in Pol II levels within gene bodies. To address whether *S. cerevisiae* is similarly affected by deletion of Spt4 at the 5' and 3' gene boundaries, we performed the same set of analyses on our PRO-seq samples from wild-type and *spt4Δ* strains (Figure 2.18). Surprisingly, genes in *S. cerevisiae* exhibit an enormous increase in PRO-seq density at the 5' ends of genes upon deletion of Spt4 (Figure 2.18a, b). This is precisely the opposite effect of that observed in *S. pombe* and results in a significant increase in average pausing index (Figure 2.18c). Beyond the CPS), somewhat similar to *S. pombe,* there appears to be a slight reduction in PRO-seq signal as a result of Spt4 deletion (Figure 2.18d, e). However, this modest reduction does not register as a significant reduction in average termination index (Figure 2.18f).

The highly contrasting effect of Spt4 deletion in budding and fission yeast at the 5' ends of genes indicates divergent roles for this elongation factor in early transcription elongation. Whereas Spt4 appears to be solely required to enhance productive elongation in *S. cerevisiae*, our results in *S. pombe* reveal an additional function of Spt4 in promoting early Pol II pausing. Perhaps as evidence of shared defects in elongation, Pol II requires Spt4 for extended transcription beyond the CPS in both species.

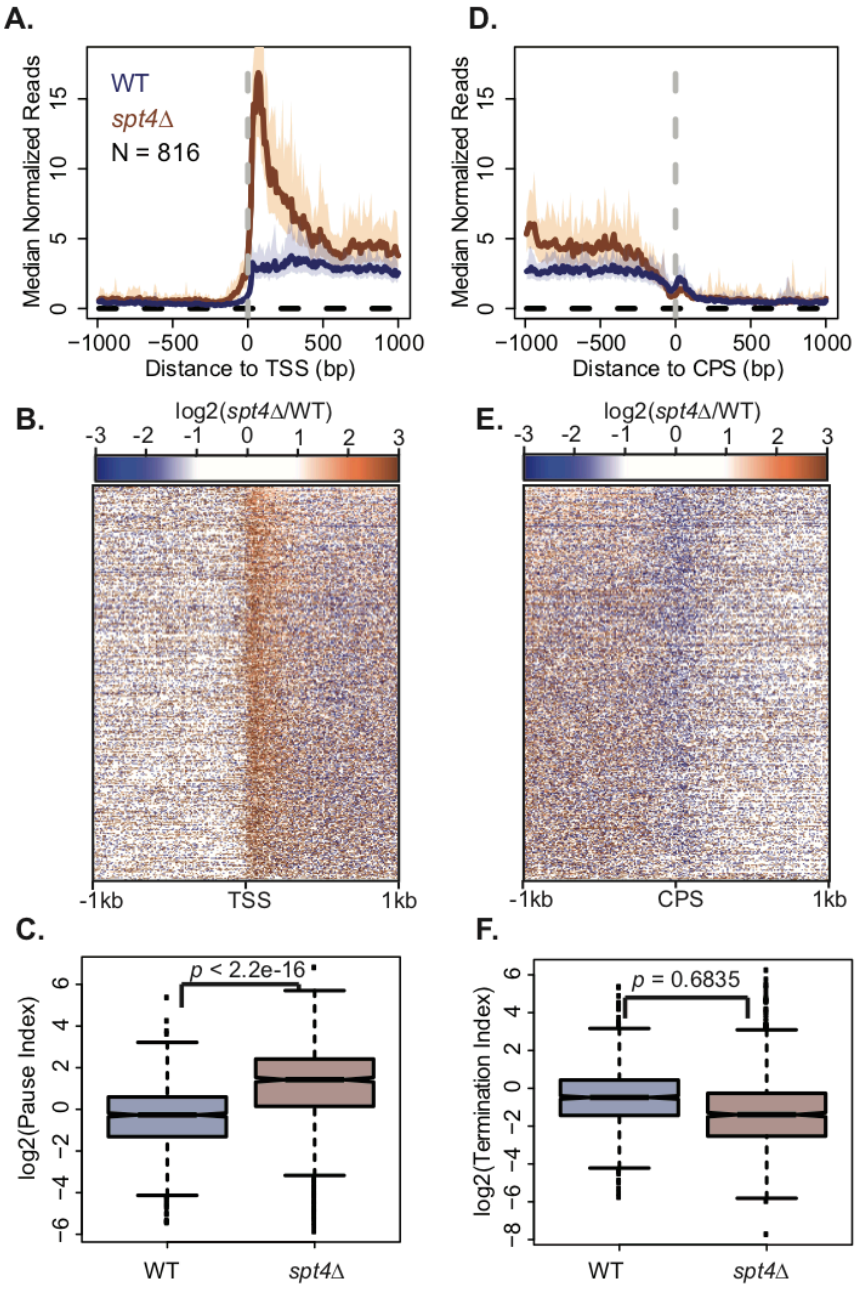

Figure 2.18: Promoter proximal regions of genes exhibit large increase in Pol II density as a result of Spt4 deletion in *S. cerevisiae*. a, d. Median PRO-seq signal around the observed TSS (a) or CPS (d) of active and filtered genes that are greater than 1 kb in length and separated from neighboring genes on the same strand by at least 1 kb. Medians reflect 10 bp bins, and the 12.5% and 87.5% quantiles are shown in lightly shaded regions. **b, e.** Heatmaps of of $\log_2$ fold change of mutant vs. wild-type per 10 bp bin around the TSS (b) or CPS (e) for all genes used in a and d. Heatmaps are sorted by decreasing amount of wild type PRO-seq reads within the first 500 bp downstream of the TSS. **c.** Box plots showing the distribution of pausing index values for WT and *spt4Δ* in *S. cerevisiae*. **f.** Box plots showing the distribution of termination index values for WT and *spt4Δ* in *S. cerevisiae*.

Discussion

We have adapted the precision run-on sequencing (PRO-seq) and 5' cap sequencing (PRO-cap) protocols for use in both budding and fission yeast. Precisely mapping 5' ends of capped nascent RNAs reveals highly preferred transcription start-sites for a large fraction of coding genes and non-coding transcription units, enhancing current annotations for many applications. Our PRO-seq profiles provide comprehensive base-pair resolution maps of elongating RNA polymerases genome-wide in two evolutionarily distant yeast strains. These data provide the highest resolution, strand-specific look at transcription across the fission yeast genome to date.

The detailed transcriptional landscapes of *S. cerevisiae* and *S. pombe* prove to be very dissimilar, with metazoan-like features of transcription unique to fission yeast. Consistent with previous investigations of transcription in *S. cerevisiae*, Pol II across annotated genes is largely confined within the well annotated 5' and 3' gene boundaries, with a relatively uniform distribution across transcription units (E. J. Steinmetz et al., 2006). In contrast, we found *S. pombe* to exhibit elongation continuing well beyond the 3' cleavage and polyadenylation signal of most genes. Immediately downstream of the *S. pombe* 3' CPS, Pol II appears to experience an elongation rate-reducing step, resulting in heightened polymerase density over a broad region. Though delayed termination is less well characterized in yeast, studies in a wide variety of metazoans have revealed similar patterns (Core et al., 2008; 2012; Jonkers et al., 2014; Kruesi et al., 2013). Our findings support the conclusion that the kinetics of transcription termination and its regulation in *S. pombe* is more similar to that of metazoans than *S. cerevisiae.*

Another intriguing characteristic of transcription profiles in fission yeast is the prominent enrichment in Pol II immediately downstream of the TSS, resembling promoter-proximal pausing in metazoans (Adelman & Lis, 2012; Fuda et al., 2009). We found that 28% of our filtered, actively transcribed *S. pombe* genes could be classified as significantly paused when considering the first 100 transcribed bases. These paused genes tend to show a much more highly positioned or high-occupancy +1 nucleosome, with Pol II density peaking near the dyad axis. Interestingly, this near-overlapping association of Pol II with the +1 nucleosome dyad is unlike that observed in *Drosophila*, where most pausing occurs upstream of the +1 nucleosome (H. Kwak et al., 2013). This difference could be the result of NELF, which might have evolved in metazoans to capture Pol II rapidly in a more promoter-proximal paused state (J. Li et al., 2013) and thereby strengthen the potential for upstream factor regulation. Considering the proximity of the +1 nucleosome to the TSS and the tight relationship between the Pol II and nucleosome occupancy within this region in *S. pombe*, it is tempting to speculate that they are connected (Figure 2.9g). On one hand, the first nucleosome may act as a barrier to early transcription elongation, causing a slower rate of transcription shortly after initiation (Weber, Ramachandran, & Henikoff, 2014). Alternatively, a frequently paused polymerase at this position could restrict the localization of the +1 nucleosome (Struhl & Segal, 2013). A rigidly positioned +1 nucleosome could then restrict the mobility of downstream, gene-body nucleosomes, resulting in the observed increase in phasing (Figure 2.9d) (Jiang & Pugh, 2009).

In metazoans, two complexes, NELF and DSIF (Spt4-Spt5), are both critical for pausing (Wu et al., 2003). The observation of pause-like distributions in fission yeast is

thus curious, as *S. pombe* lack any recognizable NELF subunits that might help to establish a pause (Narita et al., 2003). Both *S. pombe* and *S. cerevisiae*, however, possess Spt4-Spt5 and CDK9 (Bur1) homologs (Hartzog & Fu, 2013). Intriguingly, deleting Spt4 in *S. pombe* causes a dramatic reduction in Pol II density within the promoter-proximal region. Moreover, heightened Pol II density shifts into the gene-body, suggesting Spt4, most likely in complex with Spt5, is necessary for preventing the premature release of the Pol II from these sites during early transcription elongation. This possibility is supported by findings in metazoans indicating pausing is only modestly reduced upon RNAi knock down of NELF (Core et al., 2012; Gilchrist et al., 2012).

The apparent similarities between *S. pombe* and metazoan pausing raise many additional mechanistic and functional questions. As in metazoans, Spt4-Spt5 mediated promoter-proximal pausing in *S. pombe* might provide a regulatory step or checkpoint to ensure proper capping of the nascent RNA or modification of the elongation complex (Pei & Shuman, 2002; Schneider, Pei, Shuman, & Schwer, 2010; Schwer, Schneider, Pei, Aronova, & Shuman, 2009; Viladevall et al., 2009). Indeed, CDK9 and its Cyclin partner Pch1 in fission yeast interacts with nascent RNA 5' capping machinery (Pei et al., 2006; Pei, Schwer, & Shuman, 2003). Alternatively, paused Pol II could be capable of holding certain *S. pombe* genes poised for expression, synchronizing the expression of sets of genes, or allowing the integration of regulatory signals at distinct steps of recruitment and release of paused Pol II (Adelman & Lis, 2012). In the future, determining whether pause release in *S. pombe* relies on the kinase activity of CDK9 will be imperative for comparing this phenomenon with that of metazoans.

As Spt4 is a subunit of a ubiquitous elongation factor and can affect transcription broadly, we found that proper normalization of our PRO-seq and mRNA-seq data was critical for accurately assessing genome-wide changes in transcription between wild type and mutant yeast strains. Using a cell culture spike-in approach, we observed global increases in sense-transcribing RNA polymerase density across genes in *spt4Δ*strains of *S. cerevisiae* and *S. pombe* relative to wild type. However, effects on antisense transcription appeared to be species specific. In *S. cerevisiae* antisense to sense ratios were minimally affected by the deletion of Spt4. In contrast, the relative amount of antisense transcription in *S. pombe*, appeared to globally decrease, perhaps indicating little influence of Spt4-Spt5 over cryptic, intragenic, antisense transcription events in fission yeast. We found that mRNA levels in budding and fission were largely either unchanged, or slightly decreased in the absence of Spt4, despite the increase in gene body Pol II (Sup. Fig S9). One explanation could be that RNAs may in fact be increasingly produced in *spt4Δ*strains, yet buffered through surveillance mechanisms such as preRNA degradation. However, in light of several lines of evidence, we favor a hypothesis where, in the absence of Spt4, Pol II exhibits a reduced elongation rate. First, early experiments demonstrated that recombinant human Spt4 and Spt5, but not Spt5 alone, could increase rates of in-vitro transcription reactions (Wada et al., 1998). Second, reduced elongation rates are known to decrease the distance transcribed by Pol II beyond the 3' cleavage and polyadenylation signal (CPS) (Fong et al., 2015). We found that in *S. pombe,* transcription termination occurs within a much shorter distance from the CPS in the absence of Spt4 (This was also observed to a lesser extent in *S. cerevisiae,* but may be more difficult to detect because termination occurs normally only a short distance from

the CPS). Compatible with the torpedo model of transcription termination (Connelly & Manley, 1988), 5'-3' exonucleases, Dhp1 (*S. cerevisae*: Rat1) may be better able to catch up to and displace the slower Pol II from the DNA. Importantly, since Spt4 acts in complex with Spt5 to affect elongation, we surmise that many of the phenotypes we have observed in *spt4Δ*strains are a consequence of perturbation of this elongation factor complex and it's association with Pol II, rather than independent action of Spt4. Further experiments are required to distinguish independent roles of each subunit of the Spt4-spt5 complex in transcription elongation.

Apart from having a role in maintaining transcription elongation rates, the Spt4-Spt5 complex is also thought to affect the processivity of Pol II (Hartzog & Fu, 2013; Mason & Struhl, 2005). The increase in gene body Pol II in *spt4Δ*strains does not preclude the possibility of defects in Pol II processivity, but rather indicate that the more dominant affect of Spt4 deletion is on Pol II elongation rate. The modest decreases in levels of many mRNAs in *S. pombe* could reflect a combination of decreased rates of transcription and reduced processivity by Pol II in the absence of Spt4.

Using PRO-seq, adapted for determining the transcription profiles in budding and fission yeast, provides novel insight into the evolution of mechanisms governing their transcriptional landscapes. In this study, we identified divergence in the role of Spt4-Spt5 between budding and fission yeast. *S. cerevisiae* appears to rely on Spt4 primarily to maintain a steady and rapid elongation rate. In contrast, *S. pombe* additionally possesses a previously undescribed role for this complex in slowing or stalling Pol II between the TSS and the +1 nucleosome. The ability with which PRO-seq can be applied in yeast ultimately enhances our ability to combine genome-wide

snapshots of transcription elongation with the powerful genetic tools inherent to these model systems.

Methods

Yeast Strains:

S. cerevisiae strains, WT and *spt4Δ,* are in the background W303-1a and W303-1α, respectively. WT and *spt4Δ* strains in *S. pombe* are in the background 972 h-.

Precision Run-On sequencing in yeast:

The protocol described in (H. Kwak et al., 2013) was modified slightly to accommodate the physiology of yeast (García-Martínez, Aranda, & Pérez-Ortín, 2004). Instead of nuclei isolation, both *S. cerevisiae* and *S. pombe* were permeabilized to allow for the efflux of endogenous NTPs and uptake of biotin-11-NTPs (epicentre). Yeast cultures, grown overnight, were diluted to an $OD_{600} = 0.1$ and grown to mid-log phase ($OD_{600} = 0.5 – 0.6$) in either YPD (*S. cerevisiae*) or YES media (*S. pombe*). 10 mL cultures of equal cell concentration were spun down and media was removed. Cells were then resuspended in 10 mL cold H_20. Cultures were pelleted again, resuspended in 10 mL 0.5% sarkosyl at 4 °C, and incubated on ice for 20 minutes. Cells were then spun at a reduced RCF (400 x g) for 5 minutes at 4 °C. After removal of permeablization buffer, yeast pellets were resuspended in 120 µL of 2.5X transcription buffer [50 mM Tris –HCl, pH 7.7, 500mM KCL, 12.5 mM $MgCl_2$] with 6 µL 0.1 M DTT and 3.75 µL of each 1 mM biotin-NTP added immediately before use. After suspending yeast in the above transcription mix, the volume was brought to 285 µL with DEPC-treated H_20. Finally, 15

μL 10% sarkosyl was added and the reaction was placed at 30 °C and allowed to run on for 5 minutes. RNA was extracted using a hot phenol approach (Collart & Oliviero, 1993); after the run-on reaction cells were pelleted at 400 x g for 5 minutes at 4 °C and quickly resuspended in 500 μL acid phenol. An equal volume of AES buffer [50 mM NaAc pH 5.3, 10mM EDTA, 1% SDS] was added and placed at 65 °C for 5 minutes with periodic vortexing, followed by 5 minutes on ice. 200 μL chloroform was added and mixed followed by centrifugation at 14000 x g for 5 minutes (4 °C). 3 M NaOAc was added to the aqueous layer (200 mM) followed by ethanol precipitation with 3x volume of 100% ethanol. The RNA pellet was air dried before being resuspended in 20 μL DEPC-treated water. The PRO-seq or PRO–Cap protocol (H. Kwak et al., 2013) was then followed, beginning with base-hydrolysis, through to sequencing.

Spike-In Approach for Run-On (PRO-seq) or RNA isolation (RNA-seq)

Preparation of libraries with a spiked in reference organism was carried out by first growing a culture of the reference organism separate from the experimental cultures to OD_{600} = 0.5. After experimental samples had reached the desired OD_{600} and all samples were adjusted to have precisely equal numbers of cells, a fixed amount of the reference culture was added to each sample for a given experiment. All samples and biological replicates for a given experiment were prepared side by side such that all replicates received exactly the same amount of reference culture. In the case of either RNA-seq or PRO-seq, the spike-in approach described above was applied in order to control for every step of the sequencing library preparation during which technical variation may be introduced, beginning with the pelleting of cell cultures.

PRO-seq/Cap Sequencing Alignment:

Raw sequencing files were processed by first removing reads that do not pass Illumina's quality filter. Full-length or partial adapter sequences (5'-TGGAATTCTCGGGTGCCAAGG -3) were then removed using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Reads were then trimmed to a maximum length of 36 nt (minimum: 15 nt) and, in the case of PRO-seq, reverse-complemented using FASTX-Toolkit. Finally, all nascent sequencing alignments were performed using Bowtie (version 1.0.0) (Langmead, Trapnell, Pop, & Salzberg, 2009). Processed reads were aligned allowing for 2 mismatches, while requiring a unique alignment to the genome. For PRO-seq experiments, due to the existence of reads originating from separate species resulting from our spike in procedure, reads were initially aligned to ribosomal DNA sequences from each species. Due to the high level of conservation of these sequences, ribosomal mapping reads were not considered in a species-specific manner. The remaining reads were then mapped to a combined genome consisting of all chromosomes from both *S. cerevisiae* (sacCer3 = S288C_reference_genome_R64-1-1_20110203) and *S. pombe* (version: ASM294v2) facilitating the removal of reads with an ambiguous origin. Although we are aligning reads of various lengths (15-36 nt), the majority of reads in each library were 36 nt long. The method of aligning to a combined *S. pombe* and *S. cerevisiae* genome prevents genome-indistinguishable, shorter reads from being included in downstream analysis. Unique reads were parsed based on their species of origin. Bedgraph files were created by recording only the most 3' base of each read, which represents the position of the Pol II active site, for PRO-seq data, or the most

5’ base of reads from PRO-cap data. The counts at each position in the bedgraph file were normalized based on the relative amount of reads aligning to the spike-in genome (PRO-seq), or based on the number of reads per million mapped reads (PRO-cap) (for displaying in genome browsers). Bedgraphs were ultimately converted to bigwig-formatted files for downstream analysis.

Preparation and Alignment of RNA-seq libraries:

Total RNA was isolated from each sample culture (with spike-in reference) using the hot phenol approach followed by ethanol precipitation. The integrity of each sample was assessed on a denaturing poly-acrylamide gel (7 M urea PAGE) and the quantity of total RNA was measured. 3 μg of total RNA from each sample was used for library preparation. Sequencing libraries were prepared using the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina). Raw sequencing of 75 nt reads was conducted and raw sequencing files were aligned using the STAR alignment suite (version 2.3.0) (Dobin et al., 2013). Uniquely mapping reads with no more than 2 mismatches were reported and used in this analysis. Sequences were aligned to a combined genome of *S. pombe* (version: ASM294v2) and *S. cerevisiae* (sacCer3 = S288C_reference_genome_R64-1-1_20110203.tgz), and reads were separated by species-specific chromosome names. Bedgraph files were created by recording only the most 3' base of each read. For calculating the number of reads within desired regions, custom scripts were used to count the amount of read 3’ ends within provided coordinates. The counts at each position in a bedgraph file were normalized based on the relative amount of reads aligning to the spike-in genome.

Mappability Tracks:

When determining the read density for genes or other transcription units, the length of the region was defined as the number of "mappable" bases within the unit. Uniquely mappable positions for 36-mer sequence alignments were determined by, first computing all possible 36-mers across the genome of interest, and then aligning them back to the genome. By restricting the reported hits to perfect and unique alignments, we were able to generate mappability files that describe the mappable regions of the genome. In the case of samples receiving spiked in libraries (i.e. *S. pombe* cultures spiked into *S. cerevisiae*), mappability tracks were created by taking all possible 36-mers derived from the *S. cerevisiae* genome and aligning them to a species-combined genome. By aligning *S. cerevisiae* 36-mers to a composite genome and reporting only unique hits, we are able to identify all 36-mers that (1) uniquely align to the *S. cerevisiae* genome and are thus not repeat elements, and (2) are distinguishable from the spiked-in genome.

Gene Sets:

For *S. cerevisiae*, the original (annotated) gene list consisted of all transcription units identified by (Pelechano et al., 2013), considering only the longest major transcript isoform for each model. For *S. pombe,* a general transfer format (version: ASM294v1.16) file for gene annotations was downloaded from ftp://ftp.ensemblgenomes.org/pub/fungi/release-16/gtf/schizosaccharomyces_pombe/. Gene models used in this study were filtered and refined based on the following criteria. Genes had to be non-overlapping with any other annotated transcription units on the same

strand. Gene transcription was defined as the normalized read depth across mappable bases within the gene body. A background read depth was estimated based on the read density within a set of intergenic regions. Genes were only considered active if their gene body read density was greater than expected ($P<0.01$) based on a poisson distribution with a background sequencing rate. Inactive genes were omitted from downstream analysis. Moreover, only genes for which we were able to identify an observed TSS were used in our analysis (*S. cerevisiae*: n = 4398 (84% of not-overlapping transcription units); *S. pombe*: n = 4674 (80% of not-overlapping transcription units)). In order to avoid instances where upstream transcription interferes with a gene's perceived transcription, we calculated a run-through index for each gene. The run-through index was defined as the ratio of reads within the upstream window from -300 bases to TSS, relative to a downstream region +250 to +550. Genes were only used if their upstream run-through index was < 1. While the figures in this work are based on PRO-cap-based re-annotation of early annotations (ASM294v1.16), we found no effect of using an updated set of annotations for *S. pombe* (ASM294v2.30) on our results and figures. Only 23 of the shared genes appear to have 5' or 3' coordinate changes. All genes for which we called an observed TSS are listed in Supplemental Table S2 (*S. cerevisiae*) and Supplemental Table S3 (*S. pombe*). In supplemental tables S2 and S3, normalized, background-subtracted PRO-cap signal for each gene is provided and filtered genes used in this work are distinguished from unused genes.

Observed TSS Identification:

Our approach for calling transcription start sites with PRO-Cap data is modeled after a similar analysis used for Start-seq data in mouse (Scruggs et al., 2015). Using the starting lists of TSS annotations, we developed an algorithm to search for the position within +/- 250 bases of each annotation possessing the highest PRO-cap read count on the corresponding strand. Importantly, we assess the background signal of our PRO-cap libraries with a TAP-minus replicate experiment. Thus, to identify the true most preferred base near the annotated TSS, we specifically looked for the position with the highest background subtracted signal. In cases in which two or more bases shared the same signal intensity, the base closest to the annotation was chosen. Though situations arose in which multiple bases appeared to be used as TSSs, we restricted our analysis to the most preferred site to simplify downstream analysis. An observed TSS was required to have at least 5 reads more than the relative background signal.

Pausing and Termination Index Analysis:

Pausing index was quantified as the read density per mappable base within the first 100 bp downstream of the observed TSS divided by the read density per mappable base within the gene body (TSS+200 to CPS). Termination index was calculated by identifying the 100 bp window 0 to +500 bp downstream of the CPS with the maximum read density per mappable base and dividing by the read density per mappable base within the gene body. For determining the number of paused genes within each WT yeast strain, we used the combined data from biological replicates to calculate pausing index. We further tested the significance of pausing within each gene by using Fisher's

exact test to assess the likelihood of obtaining each ratio by chance. Using Bonferroni's correction, Fisher's exact test p-values were adjusted based on the number of genes analysed. We called a gene paused if it had an adjusted p-value < 0.01 and not paused if p > 0.99. For calling high-confidence paused and not paused genes, we further restricted the respective gene sets based on analysis of individual biological replicates. For both biological replicates, a gene was required to have an adjusted p-value < 0.05 (high-confidence paused), or p > 0.095 (high-confidence not paused).

MNase Data:

Aligned and processed MNase-seq files were downloaded from GEO (*S. pombe* accession: GSE49575; *S. cerevisiae* accession: GSM1143089) for use in our analysis (DeGennaro et al., 2013; Z. Hu et al., 2014). Nucleosome centers were previously identified by independent groups (*S. pombe*: (Givens et al., 2012); *S. cerevisiae*: (Weiner, Hughes, Yassour, Rando, & Friedman, 2010)). To ensure that the nucleosome centers used accurately reflected the distinct MNase-seq data, we assessed average MNase profiles relative to the centers. We found that the MNase-seq profiles of data used in this work corresponded nearly perfectly (within ~5 bp from MNase peaks on average) to the lists of nucleosome centers.

Data Access

The raw and processed sequencing files have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession GSE76142.

CHAPTER 3

CDK9 REGULATES A PROMOTER-PROXIMAL CHECKPOINT TO MODULATE RNA POLYMERASE II ELONGATION RATE IN FISSION YEAST[2]

Abstract

Post-translational modifications of the transcription elongation complex provide mechanisms to fine-tune gene expression, yet their specific impacts on RNA polymerase II regulation remain difficult to ascertain. Here, in *Schizosaccharomyces pombe,* we examine the role of *Cdk9*, and related *Mcs6*/*Cdk7* and *Lsk1*/*Cdk12* kinases, on transcription at base-pair resolution with Precision Run-On sequencing (PRO-seq). Within a minute of *Cdk9* inhibition, phosphorylation of Pol II-associated factor, *Spt5* is undetectable. The effects of *Cdk9* inhibition are more severe than inhibition of *Cdk7* and *Cdk12*, resulting in a shift of Pol II towards the transcription start site (TSS). A time course of *Cdk9* inhibition reveals that early transcribing Pol II can escape promoter-proximal regions, but with a severely reduced rate of only ~400 bp/min. Our results in fission yeast suggest the existence of a conserved global regulatory checkpoint that requires *Cdk9* kinase activity.

[2] This chapter has been adapted from a published article (**Booth, G. T.,** Parua, P. K., Sansó, M., Fisher, R. P., & Lis, J. T. (2018). Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. *Nature Communications*, *9*(1), 543.)

Introduction

Multiple kinases modify RNA Polymerase II (Pol II) and associated pausing and elongation factors to regulate Pol II transcription and pre-mRNA processing(Bentley, 2014; Buratowski, 2003; Y. Liu et al., 2009; Pei & Shuman, 2002; Schneider et al., 2010; K. Zhou, Kuo, Fillingham, & Greenblatt, 2009). For example, the concerted action of *Cdk7*, *Cdk9*, and *Cdk12* are required for the early transition of Pol II from an initiating to a productively elongating and RNA processing complex(Bartkowiak et al., 2010; Bösken et al., 2014; Larochelle et al., 2012; Marshall & Price, 1995; Ni et al., 2008; Wada, 1998). *Cdk7*, the kinase subunit of the TFIIH general transcription factor complex (*Mcs6* in *S. pombe*, *Cdk7* in humans, *Kin28* in *S. cerevisiae*), phosphorylates the largest subunit of Pol II within the C-terminal domain (CTD) at the ser5 and ser7 positions of heptad amino acid sequence, YSPTSPS, which is repeated 29 times in *S. pombe*(Akhtar et al., 2009; Azuma, Yamagishi, Ueshima, & Ishihama, 1991; Glover-Cutter et al., 2009; Jonkers et al., 2014). The modification of Pol II by *Mcs6* is critical for recruitment of the 5' RNA capping machinery and release of the mediator complex at the earliest stages of gene transcription(Adelman & Lis, 2012; E. J. Cho, Takagi, Moore, & Buratowski, 1997; Robinson, Bushnell, Trnka, Burlingame, & Kornberg, 2012; Rodriguez et al., 2000; Wong, Jin, & Struhl, 2014). Ser2 of the CTD heptad repeat is also phosphorylated as Pol II elongates further into the gene body(Hartzog et al., 1998; Keogh et al., 2003; Mayer et al., 2010; Swanson et al., 1991), and both *Cdk9* and *Lsk1* (*Cdk12* in humans, *Ctk1* in *S. cerevisiae*) have been suggested to modify this residue, possibly to facilitate 3' RNA processing(Booth, Wang, Cheung, & Lis, 2016; Hsin & Manley, 2012).

Similar to the Pol II CTD, targets within elongation factors can also mediate essential regulation of transcription elongation. *Spt4* and *Spt5* comprise a highly conserved elongation factor (DRB-sensitivity inducing factor, or DSIF in metazoans)(Pei et al., 2003; Wada et al., 1998). *Spt5* possesses an unstructured C-terminal repeat domain (CTR), with specific residues actively targeted by *Cdk9*(Bishop et al., 2000; Pei et al., 2003). Moreover, the structure of eukaryotic *Spt4*/5 bound to transcribing RNA Pol II reveals contacts with upstream DNA and nascent RNA exiting Pol II(Ehara et al., 2017; Viladevall et al., 2009), and depletions of *Spt5* in fission yeast globally reduce levels of elongating Pol II(Rodríguez-Molina et al., 2016; Shetty et al., 2017). *Cdk9* (Bur1 in *S. cerevisiae)* positively impacts transcription elongation(Ebmeier et al., 2017; Keogh et al., 2003; Larochelle et al., 2012), and has been suggested to convert *Spt5* from a negative to positive elongation factor in humans(Mason & Struhl, 2005; Yamada et al., 2006). Such observations may reflect regulated recruitment of elongation coupled factors, as association of capping enzymes and Polymerase associated factor 1 (Paf1) with the transcription complex are influenced by the phosphorylation status of *Spt5* (Bentley, 2014; Buratowski, 2003; Y. Liu et al., 2009; Pei & Shuman, 2002; Schneider et al., 2010; K. Zhou et al., 2009).

In most metazoan systems, the phosphorylation of *Spt5* and negative elongation factor (NELF) by positive transcription elongation factor b (P-TEFb, the *Cdk9/cyclin T1* complex), is thought to be required for the release of elongating Pol II from a promoter-proximal pause site(Bartkowiak et al., 2010; Bösken et al., 2014; Larochelle et al., 2012; Marshall & Price, 1995; Ni et al., 2008; Wada, 1998). This *Cdk9*-regulated process (referred to hereafter as pausing) is now known to occur on the vast majority of genes in

mammals(Akhtar et al., 2009; Azuma et al., 1991; Glover-Cutter et al., 2009; Jonkers et al., 2014), serving to regulate gene expression, or to ensure appropriate processing of RNA and/or maturation of the elongation complex(Adelman & Lis, 2012; E. J. Cho et al., 1997; Robinson et al., 2012; Rodriguez et al., 2000; Wong et al., 2014). In budding yeast, which lack homologs of all NELF subunits, such pausing has not been observed, yet, perturbation of and *Spt5* impair normal transcription by Pol II(Hartzog et al., 1998; Keogh et al., 2003; Mayer et al., 2010; Swanson et al., 1991). In contrast, pause-like distributions of Pol II have been found in the fission yeast *S. pombe*, where they depend on *Spt4*(Booth et al., 2016; Hsin & Manley, 2012). In addition, *S. pombe Cdk9* interacts with the triphosphatase component of the 5' RNA capping apparatus (Pct1), possibly to alleviate a *Spt4-5*-induced checkpoint that ensures proper pre-mRNA processing(Pei et al., 2003; Wada et al., 1998). Still, such a checkpoint in yeast remains largely speculative, and the relation to *Cdk9*-regulated pausing in metazoans is unknown.

Investigations of the direct influence of kinase activity on the dynamics and regulation of transcription in vivo requires highly selective inhibitors, which can be difficult to obtain. Mutant kinases have been designed with exquisite specificity for bulky ATP analogs that selectively inhibit their activity (Bishop et al., 2000; Pei et al., 2003). The use of such analog sensitive (AS) strains in fission yeast revealed correlated changes in global mRNA levels upon inhibition of *Mcs6* or *Cdk9* within minutes of inhibition(Ehara et al., 2017; Viladevall et al., 2009). Inhibitions of an analog sensitive TFIIH kinase resulted in a shift of Pol II levels towards the 5' ends of genes in budding yeast after one hour (Rodríguez-Molina et al., 2016; Shetty et al., 2017) and impaired pausing and termination in humans after a day of inhibition(Ebmeier et al., 2017; Keogh

et al., 2003; Larochelle et al., 2012). However, even within just one hour, the primary effects of kinase inhibition on Pol II dynamics may be masked, especially in yeast where genes are short and transcribed in a matter of minutes(Mason & Struhl, 2005; Yamada et al., 2006).

In this study, to gain an insight into the impact of transcription complex-targeting kinases on Pol II elongation, we use analog sensitive kinase mutants in the fission yeast *Schizosaccharomyces pombe* to rapidly and selectively inhibit the activities of *Cdk9*, *Mcs6*, and *Lsk1* individually or in combination. Using precision run-on sequencing analysis, we observe global changes in transcriptional dynamics within minutes of treatment, indicative of the general influence of these kinases on transcription by Pol II. While almost no instantaneous changes in Pol II distribution result from the loss of $lsk1^{as}$ activity, inhibition of $cdk9^{as}$ and/or $mcs6^{as}$ dramatically alter the transcription landscape within five minutes, despite non-overlapping substrates. *Spt5* phosphorylation becomes undetectable after just 1 minute of $cdk9^{as}$ inhibition, however, unlike metazoans, elongating Pol II is not trapped at the promoter-proximal pause site(Jonkers et al., 2014). Instead, a fine-scale time-course of $cdk9^{as}$ inhibition reveals that promoter-proximal Pol II has a severely compromised elongation rate, while downstream Pol II appears less affected. Our results support the existence of a *Cdk9*-dependent early elongation checkpoint in fission yeast, which may reflect a primitive form of *Cdk9*-regulated promoter-proximal pausing in metazoans.

Results

***Cdk9* is not required for promoter-proximal pause escape.** Recently, we reported promoter-proximal pause-like distributions of Pol II on many genes in fission yeast(Booth et al., 2016), yet whether escape from such pausing is regulated through kinase activity is not known. In metazoans, kinase inhibitors, including 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) and flavopiridol (FP), which primarily inhibit *Cdk9*, have been instrumental in the discovery of promoter-proximal pausing as a major regulatory hurdle for most genes(Jonkers et al., 2014; Marshall & Price, 1995; Ni et al., 2008; Wada, 1998). However, off-target effects of chemical inhibitors make selective ablation of *Cdk9* activity *in vivo* nearly impossible(Bensaude, 2014).

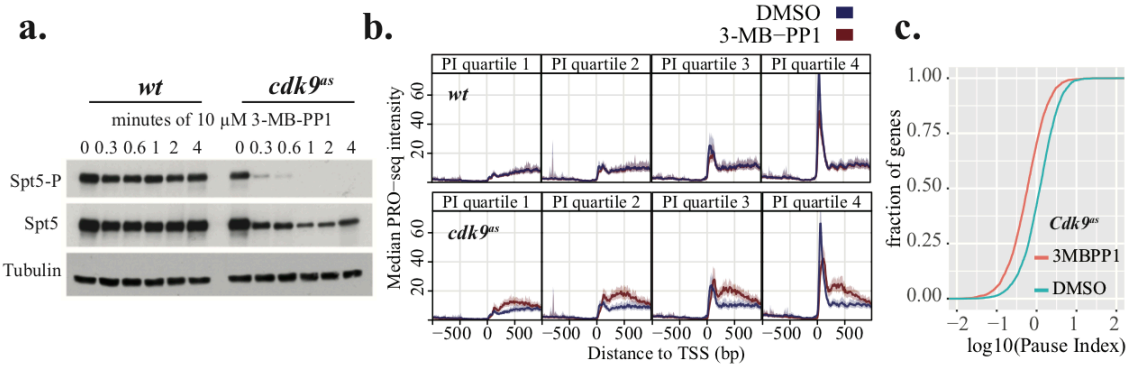

Figure 3.1 | *Cdk9* is not required for promoter-proximal pause escape. **a**. Western blot analysis using antibodies that specifically recognize p*Spt5* compared with total *Spt5* in *wt* and *cdk9*^{as} over a time course of cultures treated with 10 μM 3-MB-PP1. Total *Spt5* represents both phosphorylated and unphosphorylated forms. Tubulin serves as a loading control. **b**. TSS-centered composite profiles of PRO-seq data before (blue) and after (red) 5 minutes of treatment with 10 μM 3-MB-PP1 for genes grouped by quartiles of increasing pausing index from left to right (calculated using untreated wild type data from combined replicates). The top four panels represent wild-type data, while the bottom four panels represent *cdk9*^{as} data. Each quartile contains 841 filtered genes **c.** Cumulative density functions for pausing index ($\log_{10}$) of all filtered genes in treated (red) and untreated (blue) samples for the *cdk9*^{as} strain. Data used in b. and c. reflect the results of combined data from two biological replicates for each treatment.

To isolate the immediate impact of loss of *Cdk9* activity on Pol II pausing and elongation at single nucleotide resolution in fission yeast, we performed PRO-seq in an analog sensitive (AS) mutant strain, *cdk9*^{as}, which is vulnerable to inhibition by the addition of 3-MB-PP1(Cipak et al., 2011; Viladevall et al., 2009). Phosphorylation of *Spt5* became undetectable after one minute of drug addition in *cdk9*^{as} (Figure 3.1a), while bulk measurements of phosphorylation on Pol II CTD residues were largely unaffected, within the timeframe tested (Figure 3.2a). The near-immediate loss of *Spt5* phosphorylation upon *cdk9*^{as} inhibition implies an extremely rapid action of the small molecule 3-MB-PP1 and supports a model of active, competing *Spt5* dephosphorylation(Parua et al., n.d.). Taking advantage of the ability to rapidly inhibit AS kinase activity, PRO-seq libraries were prepared in two biological replicates (Figure 3.2b) from *cdk9*^{as} cells treated for 5 minutes with 10 μM 3-MB-PP1 (treated) or an equivalent volume of DMSO (untreated). As a control, we prepared PRO-seq libraries from equivalently treated wild-type (*wt*) cells. Comparison of untreated *cdk9*^{as} profiles with *wt* profiles revealed no obvious differences (Figure 3.3a) and addition of 3-MB-PP1 to *wt* cells produced almost no transcriptional changes (Figure 3.3b), indicating that this system has little to no basal phenotype of the AS mutation or off-target drug effects. To evaluate the effect of *Cdk9* on promoter-proximal pause-like distributions, genes were divided into quartiles based on their pausing index (PI), which measures the enrichment of Pol II in the promoter-proximal region (TSS to +100 nt) relative to the gene body ('+' indicates downstream and '-' indicates upstream). Although we observed striking increases in early gene body PRO-seq signal (TSS to +1kb) resulting from 5 minutes of *cdk9*^{as} inhibition, pausing did not explain this effect (Figure 3.1b).

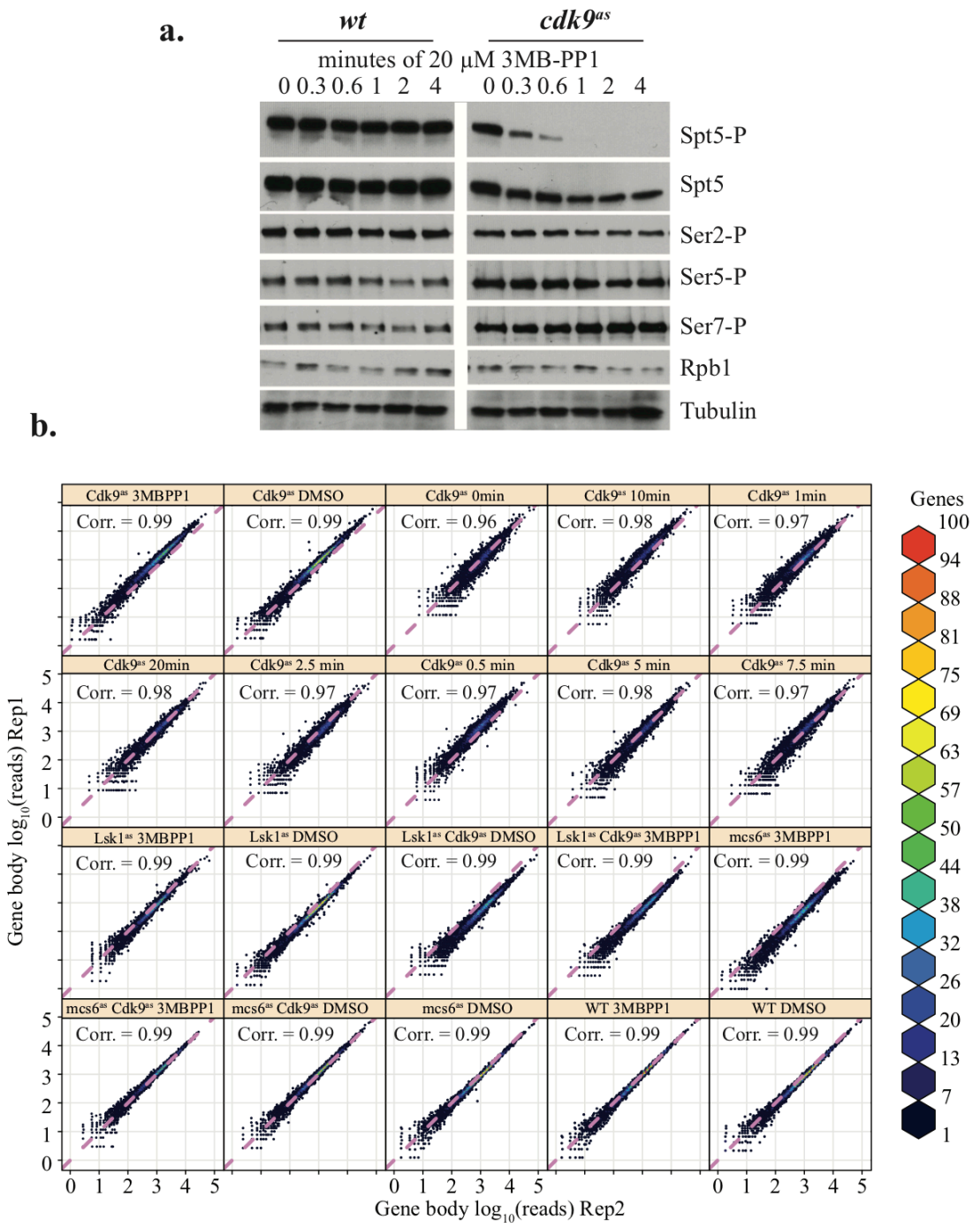

Figure 3.2 | *Cdk9* inhibition primarily affects *Spt5*-P within a short timeframe and PRO-seq experiments are reproducible. a. Western blot analysis of phosphorylated residues within the elongation complex in *wt* and *cdk9*^{as} strains after increasing durations of treatment with 20 µM 3-MB-PP1. From top to bottom, antibodies were raised against, *Spt5*-P, total *Spt5*, Ser2-P, Ser5-P, Ser7-P, Rpb1, or Tubulin (loading control). **b.** Scatter plots displaying a correlation (Spearman's rho) between biological replicate PRO-seq

data for each sample. Data points show spike-in normalized read counts (log_{10}) within the gene body (TSS + 200 bp to annotated CPS, where "+" indicates downstream of TSS) of all filtered genes. Since minimal variation is expected between replicates, accurate spike in-based normalization will produce scatter that is approximately centered on the purple diagonal line (x = y).

Inhibition of pause escape in mammals with drugs that target *Cdk9* is known to result in an increase in engaged, promoter-proximal Pol II coinciding with loss of signal from the gene body(Cheng et al., 2012; Jonkers et al., 2014), yielding a greater PI for many genes. In contrast, we find few genes with significant increases in promoter-proximal Pol II upon *Cdk9* inhibition (Figure 3.3c & d). Moreover, compared to *wt, cdk9*^{as} cells exhibit a global decrease in pausing index as a result of 5 minutes of treatment with 3-MB-PP1 (Figure 3.1c, Figure 3.3e). Together, our results reveal that *Cdk9* activity in *S. pombe* is not required for release of promoter-proximal Pol II into elongation, but it is nonetheless critical for efficient elongation across gene bodies.

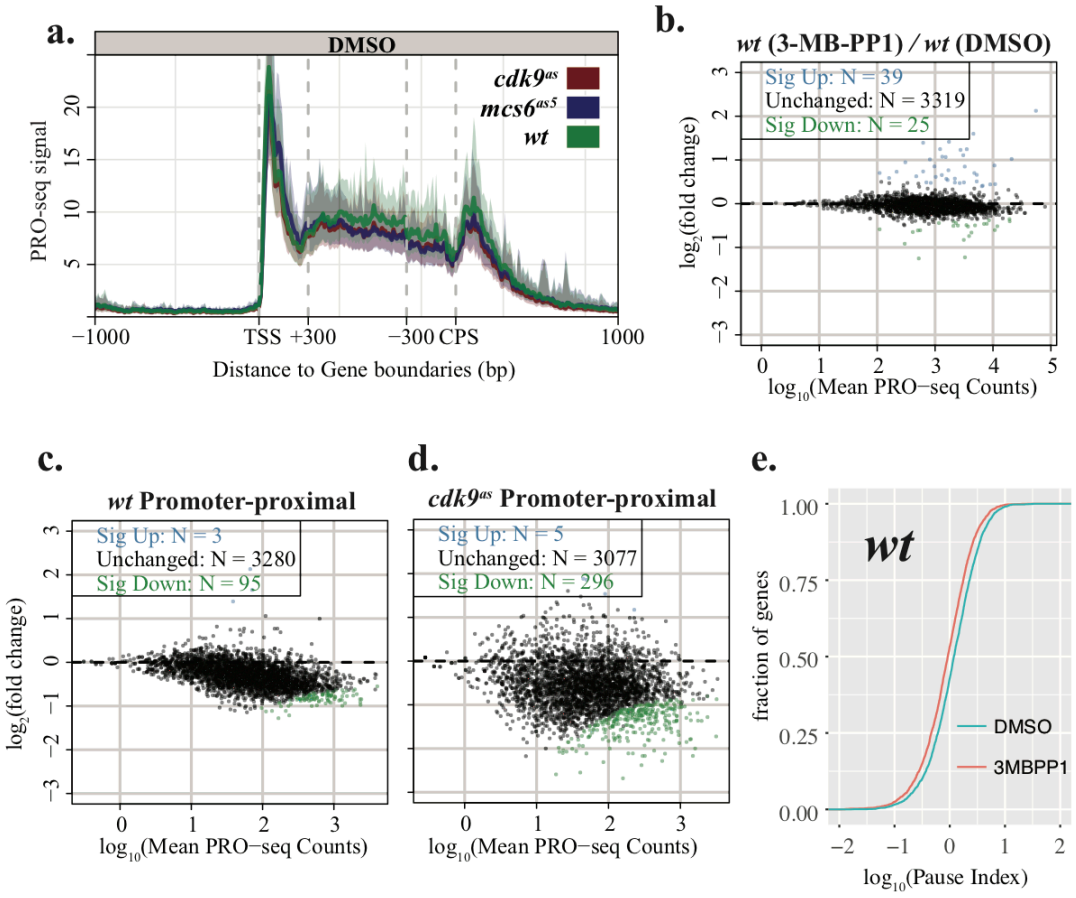

Figure 3.3 | AS kinase-dependent strains provide a highly controlled system for studying transcription. **a.** Composite profiles of combined replicate data for samples of untreated wild type, *mcs6*^{as5}, and *cdk9*^{as} strains across filtered genes separated from nearest same strand neighbors by at least 1 kb on both sides (n = 939). Composite profiles represent the median subsampled value from combining two biological replicates within each bin. Shaded regions correspond to the 12.5% and 87.5% quantiles. To scale genes to a common length, the middle gene body region of each gene was scaled to 60 bins, while the regions -1000 bp to +300 bp relative to the TSS, and -300 bp to +1000 bp relative to the CPS, are unscaled 10 bp windows ("+" and "-" indicate downstream and upstream, respectively). **b.** MA plot displaying the $\log_2$ fold change between gene body read counts from treated and untreated *wt* cells, as calculated using the DESeq2 package. **c & d.** MA plot displaying the $\log_2$ fold change between treated and untreated promoter proximal (TSS to +100 bp) read counts for wild type (**c.**), or *cdk9*^{as} (**d.**). Genes that show significantly increased or decreased PRO-seq signal (adjusted p <0.01; DESeq2: Wald test, Benjamini and Hochberg's correction) are shown in blue and green, respectively. For all MA plots, spike-in normalization was used when calculating $\log_2$ fold changes between samples. **e.** Cumulative density functions for pausing index ($\log_{10}$) of all filtered genes in treated (red) and untreated (blue) samples for the wild type strain.

***Mcs6* and *Cdk9* impact Pol II at 5' and 3' ends of genes.** CTD modification by

transcription-coupled kinases, *Mcs6* and *Lsk1*(Hsin & Manley, 2012) can modulate the association of auxiliary components with Pol II, to facilitate co-transcriptional RNA processing(Bentley, 2014) and elongation through chromatin(B. Li et al., 2007). Thus, in addition to *Cdk9*, we set out to measure the impact of *Mcs6* and *Lsk1* on global transcription. A novel AS variant of *Mcs6* (here called $mcs6^{as5}$) was as sensitive to treatment with 3-MB-PP1 as $cdk9^{as}$ (Figure 3.5a & b). Importantly, no gross transcriptional differences were observed in untreated $mcs6^{as5}$ compared with *wt* (Figure 3.3a). Within minutes of treatment, we observed measurable losses in Ser5-P and Ser7-P in the $mcs6^{as5}$ strain, while these marks in *wt* cells were unaffected (Figure 3.4a). For consistency, PRO-seq experiments with $mcs6^{as5}$ were performed using 10 μM 3-MB-PP1, which reduces Ser5-P levels to the same extent as 20 μM of the inhibitor (Figure 3.5c). The analog sensitive *Lsk1* ($lsk1^{as}$) was also inhibited by 10 μM 3-MB-PP1 (Figure 3.5c). However, despite the clear impact on Ser2 phosphorylation (Figure 3.6a), there was almost no visible change in transcription upon treatment of $lsk1^{as}$ with 10 μM 3-MB-PP1 for 5 minutes (Figure 3.6b-d). In contrast, loss of *Mcs6* or *Cdk9* activity produced abrupt changes at both ends of individual genes (Figure 3.4b). Most notably in $cdk9^{as}$, Pol II signal appeared to decrease with increasing distance from the TSS after inhibition. We also profiled distributions of Pol II in strains containing combinations of AS kinases. Combined inhibition of either $lsk1^{as}$ and $cdk9^{as}$, or $mcs6^{as5}$ and $cdk9^{as}$ produced transcriptional defects resembling those of $cdk9^{as}$ or $mcs6^{as5}$ alone, although possible basal effects of the combined mutations limited further interpretation from these strains (Figure 3.7).

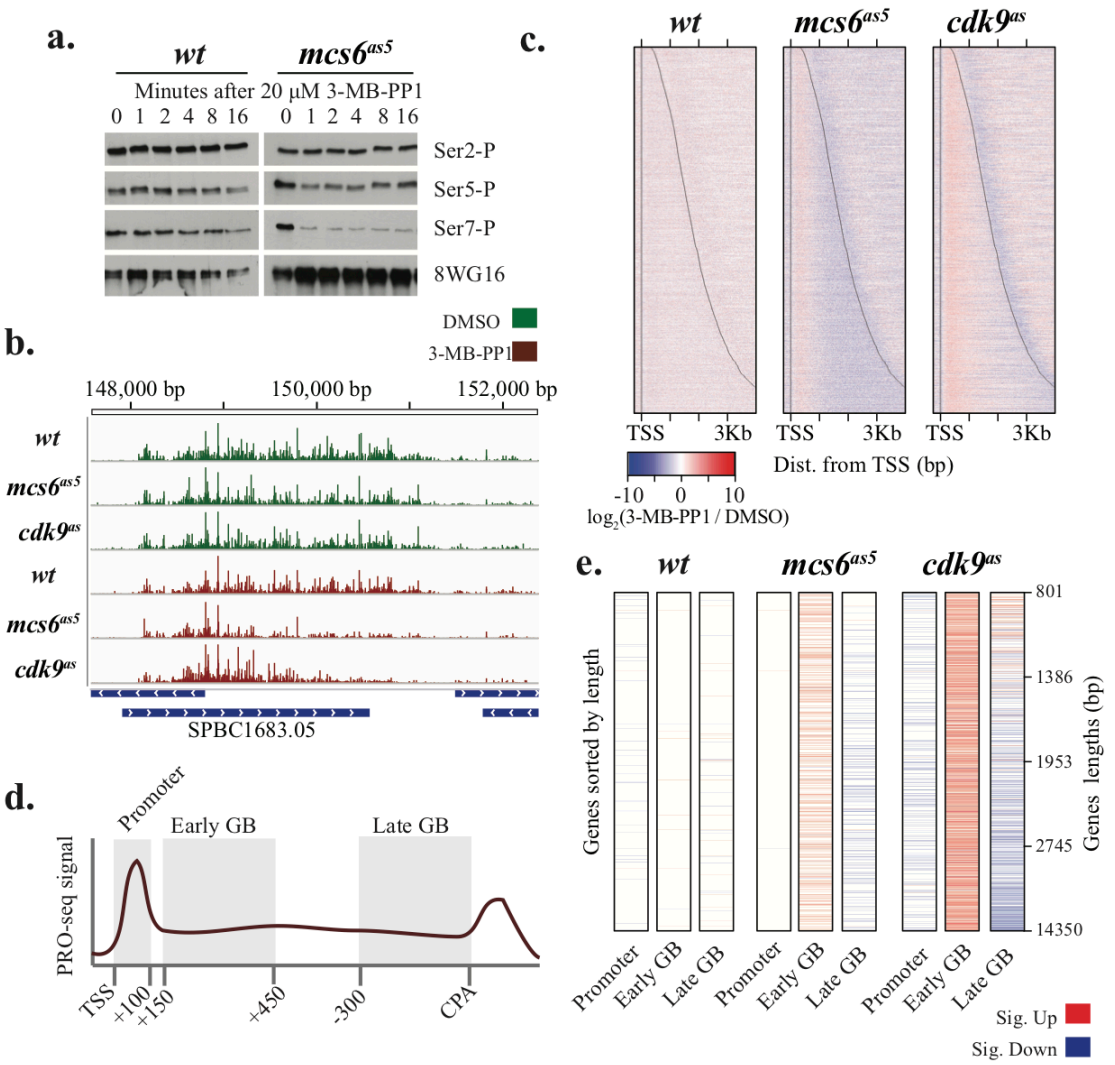

Figure 3.4 | *Mcs6* and *Cdk9* impact Pol II at 5' and 3' ends of genes. a. Western blot analysis of phosphorylated CTD residues (Ser2-P, Ser-5-P, Ser7-P) in relation to total CTD signal. Levels were measured over a time course of cultures treated with 20 µM 3-MB-PP1. **b**. Browser track image from the *SPBC1683.05* locus displaying normalized read counts from untreated (green) and treated data (red) for wild-type, *mcs6*^{as5} and *cdk9*^{as} strains from top to bottom, respectively (plus strand only). All PRO-seq samples represent treatments for 5 minutes with 10 µM 3-MB-PP1 (treated) or equivalent volume of DMSO (untreated). **c**. Heat maps depicting the $\log_2$ fold change in normalized PRO-seq signal (treated/untreated) within 10 bp bins from -250 bp to +4000 bp relative to the TSS ('+' indicates downstream and '-' indicates upstream) for wild-type, *mcs6*^{as} and *cdk9*^{as} strains from left to right, respectively. Data used in b. and c. reflect the results of combined data from two biological replicates for each treatment. Only one replicate was used for 3-MB-PP1-treated *mcs6*^{as5} (See methods). Genes within each heat map are sorted by increasing length from top to bottom, with black lines representing observed TSS and CPS (n = 3383). **d**. Hypothetical plot for the illustration of the definition of promoter (TSS to +100) early (+150 to +450 from TSS) and late (-300 to 0 from CPS) gene body regions used in E. **e**. Heat maps depicting whether each gene exhibits a

significant fold change (adjusted $p < 0.01$; treated/untreated; DESeq2: Wald test, Benjamini and Hochberg's correction) in promoter, early or late gene body regions. Genes were required to be longer than 800 nt (n = 3003) and are sorted from top to bottom by increasing gene length with length quartiles shown on the right. Significant increases and decreases in each region are shown as red and blue, respectively.

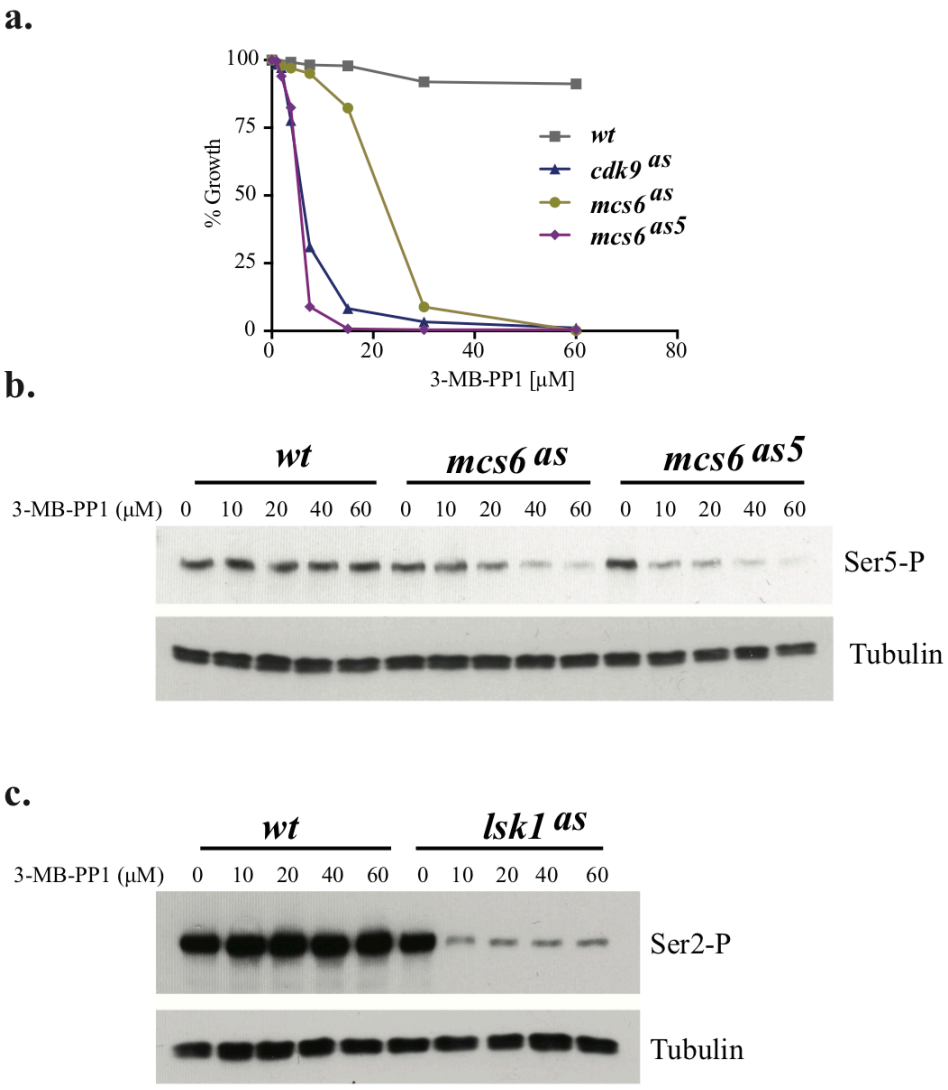

Figure 3.5 | Characterization of AS strains and the effects of 3-MB-PP1 concentration. a. The indicated strains were grown in 96-well plates in the presence of increasing amounts of 3-MB-PP1 and their growth quantified by OD_{600}. The mutant alleles and their associated mutations are: *mcs6*^{as}, L87G; *mcs6*^{as5}, N84T/L87G; *cdk9*^{as}, T120G. **b.** Western blot analysis of Ser5-P relative to tubulin (loading control) in *wt*, *mcs6*^{as}, and *mcs6*^{as5} strains after 1 hr treatment at 30 °C with the specified concentration of 3-MB-PP1. **c.** Western blot analysis of Ser2-P relative to *tubulin* in *wt*, and *lsk1*^{as} strains after 1 hr of treatment at 30 °C with the specified concentration of 3-MB-PP1.

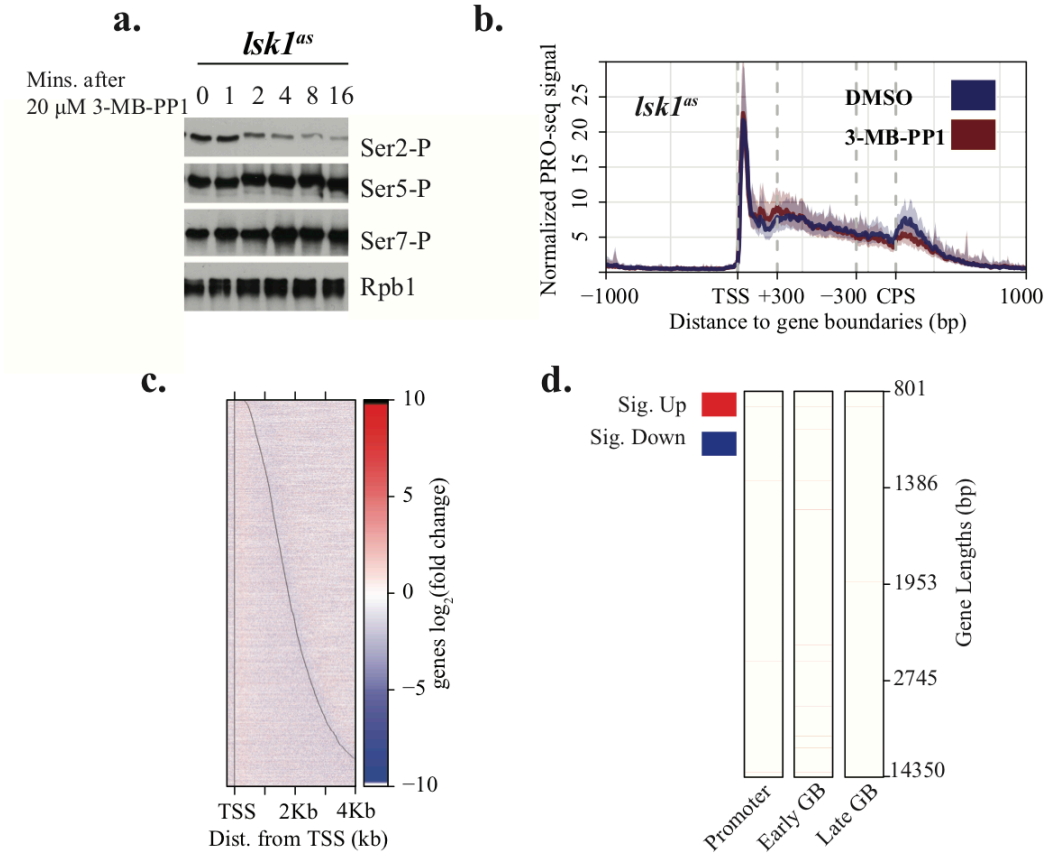

Figure 3.6 | Treatment of *lsk1*^{as} with 3-MB-PP1 has minimal impact on transcription within five minutes. a. Western blot analysis of phosphorylated CTD residues (Ser2-P, Ser-5-P, Ser7-P) in relation to total Rpb1 signal. Levels were measured over a time course of *lsk1*^{as} treated with 20 μM 3-MB-PP1. **b.** Composite PRO-seq profiles for treated (5 min 10 μM 3-MB-PP1) and untreated (5 min DMSO) *lsk1*^{as} displaying median subsampled signal across filtered genes separated from nearest same strand neighbors by at least 1 kb on both sides (n = 939). Shaded regions correspond to the 12.5% and 87.5% quantiles. The middle gene body region of each gene was scaled to 60 bins, while the regions, -1000 bp to +300 bp, relative to the TSS, and -300 bp to +1000 bp relative to the CPS, are unscaled 10 bp windows. **c.** Heat map depicting the $\log_2$ fold change in normalized PRO-seq signal (treated/untreated) in *lsk1*^{as} within 10 bp bins from -250 bp to +4000 bp relative to the TSS. Genes are sorted by increasing length, with black lines representing observed TSS and annotated CPS. Composite profiles (a.) and fold change heat maps (b.) represent combined data from two biological replicates. **d.** Heat maps depicting whether each gene exhibits a significant fold change (adjusted $p < 0.01$; treated/untreated; DESeq2: Wald test, Benjamini and Hochberg's correction) in promoter, early, or late gene body regions (defined in Figure 3.4). Genes were required to be longer than 800 nt and are sorted from top to bottom by increasing gene length (n = 3003). Gene length quartiles are shown with tick marks on the right. Significant increases and decreases in each region are shown as red and blue, respectively.

Composite PRO-seq profiles around transcription start sites (TSS) and cleavage and polyadenylation sites (CPS) revealed global kinase-dependent effects at both ends of transcription units (Figure 3.7). Surprisingly, the loss of signal at 3' ends of genes upon inhibition of *Mcs6* or *Cdk9* was only observed on longer genes (Figure 3.4c, genes near the bottom of heatmaps). Differential expression analysis(M. Love, Anders, & Huber, 2013) was used to verify differences in treated and untreated PRO-seq counts within three discrete regions of each gene: 1) the promoter-proximal region (TSS to +100), 2) the early gene body (+150 to +450 from the TSS), and 3) the late gene body (-300 to CPS) (Figure 3.4d). Most strikingly in $cdk9^{as}$, and to a lesser extent in $mcs6^{as5}$, when genes were sorted by increasing length, distance from the TSS appeared to dictate the pattern of changes within the gene body regions. Although promoter-proximal regions were only subtly affected, early gene bodies were consistently found to show increases in signal in response to *Cdk9* or *Mcs6* inhibition. However, late gene bodies displayed differing results depending on the region's distance from TSS, such that short genes exhibited increased signal and long genes had decreased signal within the last 300 bp of the transcription unit (Figure 3.4e).

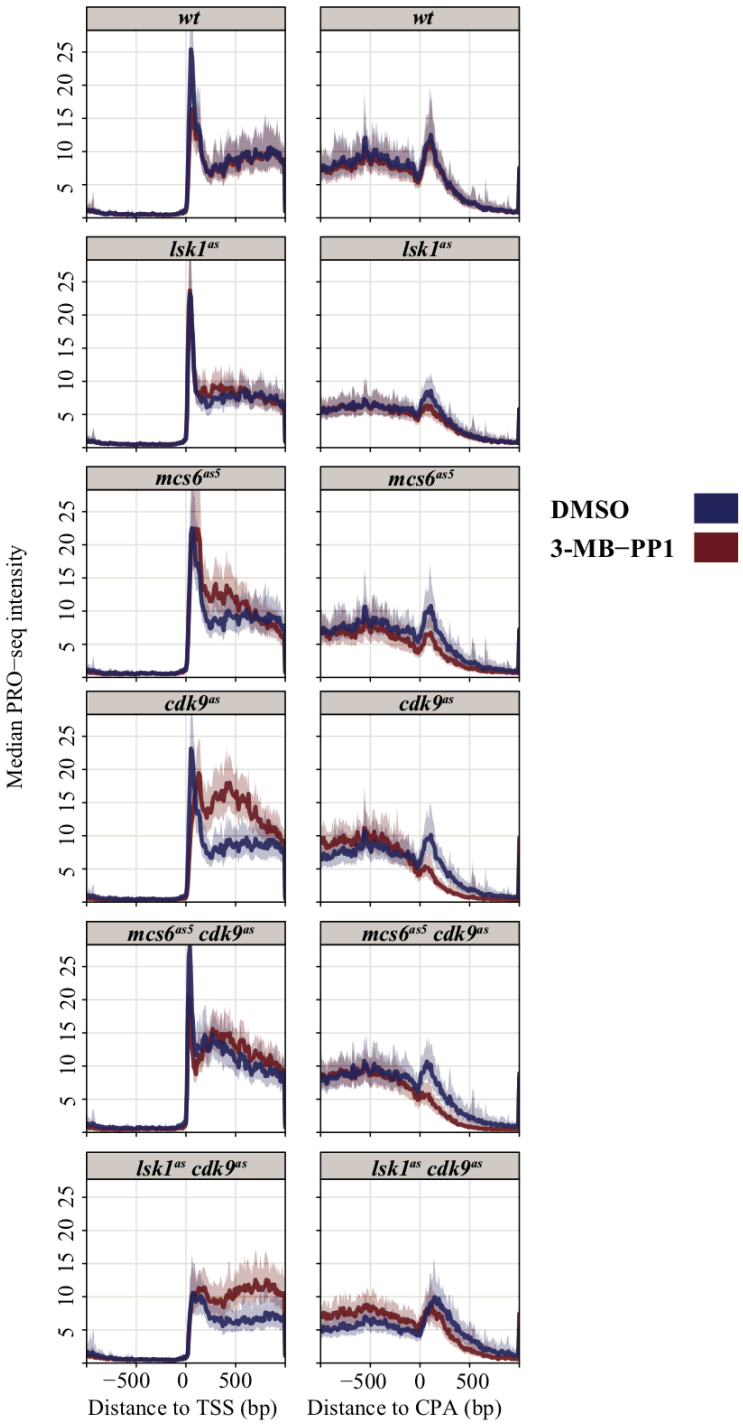

Figure 3.7 | Global effects of kinase inhibition revealed through composite profiles centered on TSS and CPS. Each panel displays centered composite PRO-seq profiles of treated (5 min 10 μM 3-MB-PP1) and untreated (5 min DMSO) samples from each strain. Dark middle lines reflect the median sub-sampled PRO-seq signal within 10 bp windows from -1 kb to +1 kb from observed TSS (left) or CPS (right) for genes separated from nearest same strand neighbors by at least 1 kb on both sides (n = 939). Shaded regions correspond to the 12.5% and 87.5% quantiles. For the combined mutants, composite

untreated PRO-seq profiles appear to deviate from those of the *wt* strain, indicating a possible synthetic effect of the mutations on transcription. Therefore, we were limited in our ability to interpret the results from these strains. With the exception of 3-MB-PP1-treated *mcs6*^{as5} (see methods), all profiles represent combined data from two biological replicates.

Several mechanisms might explain the observed gene length-dependent effects on transcription. Consistent with a loss of signal from the 3' ends of longer genes (within bottom half of heatmaps in Figure 3.4c), in the absence of *cdk9*^{as} activity, Pol II may exhibit a compromised processivity, preventing transcription elongation beyond a certain distance from the TSS(Mason & Struhl, 2005). Alternatively, the altered transcription profiles may reflect a reduced rate of early transcribing ECs incapable, or delayed in, accelerating to the natural productive speed. Assuming an unchanged initiation rate, the increased density of elongating Pol II at the 5' end would be compatible with a reduced elongation rate of Pol II, which within five minutes may not have allowed Pol II to transcribe fully across the longest genes.

A checkpoint during early elongation impacts Pol II rates. We reasoned that these proposed explanations—loss of processivity leading to premature termination versus reduced elongation rate— could be resolved by examining a time course of Pol II distribution on genes following rapid *Cdk9* inactivation. For instance, a defect in Pol II processivity would result in premature termination at roughly the same distance from the TSS, regardless of time after addition of 3-MB-PP1, whereas a rate defect might reveal a "wave front" shifting downstream with increasing time of inhibition. Thus, we performed a high-resolution time course of treatment with 3-MB-PP1 followed by PRO-seq, focusing specifically on the severe phenotype in *cdk9*^{as}.

Strikingly, individual genes reproducibly exhibited a shifting distribution of transcribing Pol II, which appeared to propagate from the TSS toward the 3' ends of genes over time (Figure 3.8a, Figure 3.9a-c). Heat maps of the fold change after each duration of 3-MB-PP1 exposure, relative to untreated (DMSO), revealed an immediate global impact of *Cdk9* inhibition on transcription that progressed as a spreading of increased signal across genes over time. Meanwhile, an early drop in signal at gene 3' ends was gradually recovered with the advancing Pol II density (Figure 3.8b).

To better understand the propagation of increased Pol II density following the initial loss in downstream signal, we prepared composite profiles at each time-point considering only the longest genes in our filtered set (at least 6 kb, n = 42). Intriguingly, average profiles for these genes reveal a rapid recession of Pol II density towards the CPS (Figure 3.8c) - a phenomenon also obvious within individual genes (Figure 3.8a; see 1-minute and 2.5-minute time-points).

We reason that these dynamic profiles may represent two distinct populations of transcribing Pol II captured at the moment of *cdk9*^{as} inhibition. One population had yet to transit a 5' checkpoint dependent on *Cdk9* activity and is thus impaired in elongation when *Cdk9* is inhibited. The other population, already beyond this checkpoint, may no longer require *Cdk9* to maintain its rate of transcription (Figure 3.8d), which has previously been estimated in budding yeast to be ~2 Kb/min(Mason & Struhl, 2005). Accordingly, these observations are indicative of a *Cdk9*-dependent regulatory step during the early stages of elongation in fission yeast (Figure 3.8d), ,albeit not one fixed at a specific pause position, as is the case in mammals(Adelman & Lis, 2012).

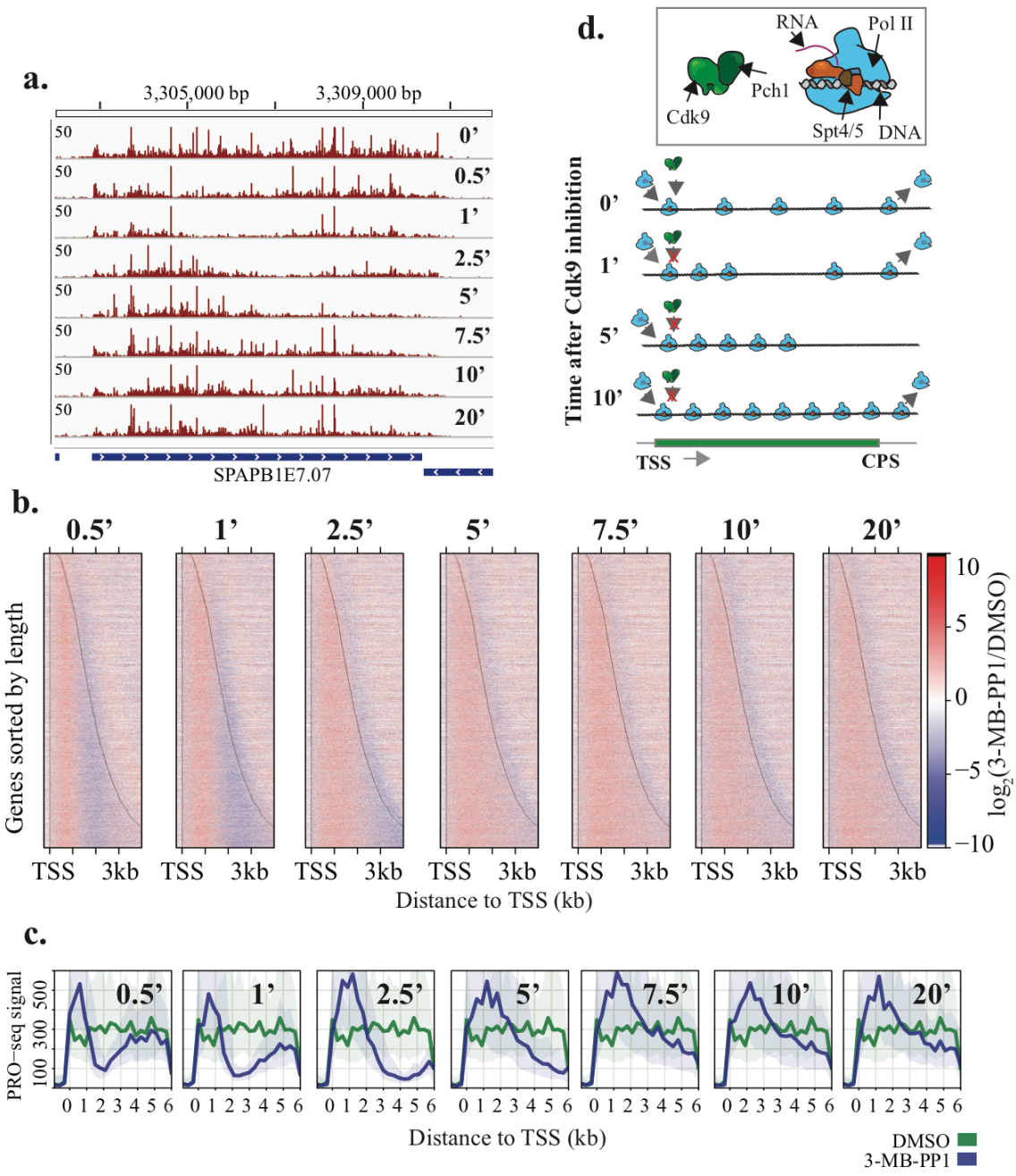

Figure 3.8 | A checkpoint during early elongation impacts Pol II rates. a. Browser track image displaying the normalized PRO-seq read count signal at the *SPBAPB1E7.07* locus for *cdk9*^{as} cells treated with 10 μM 3-MB-PP1 for an increasing duration, from top to bottom. **b**. Heat maps of $\log_2$(treated/untreated) normalized PRO-seq signal within10 bp windows from -250 bp to +4000 bp around the TSS for all filtered genes (n = 3383) ordered by increasing gene length from top to bottom. Panels show PRO-seq data from each time point of drug treatment (relative to the 0-minute sample), with increasing duration from left to right. **c**. Composite PRO-seq signal for all filtered genes at least 6 kb in length (n= 42) before and after treatment. Panels from left to right show profiles after increasing duration of treatment compared with DMSO treated cells. Data used in a-

c. reflect the results of combined data from two biological replicates for each treatment. **d**. Illustration of two populations of transcribing Pol II, which have rates differentially affected by *Cdk9* inhibition.

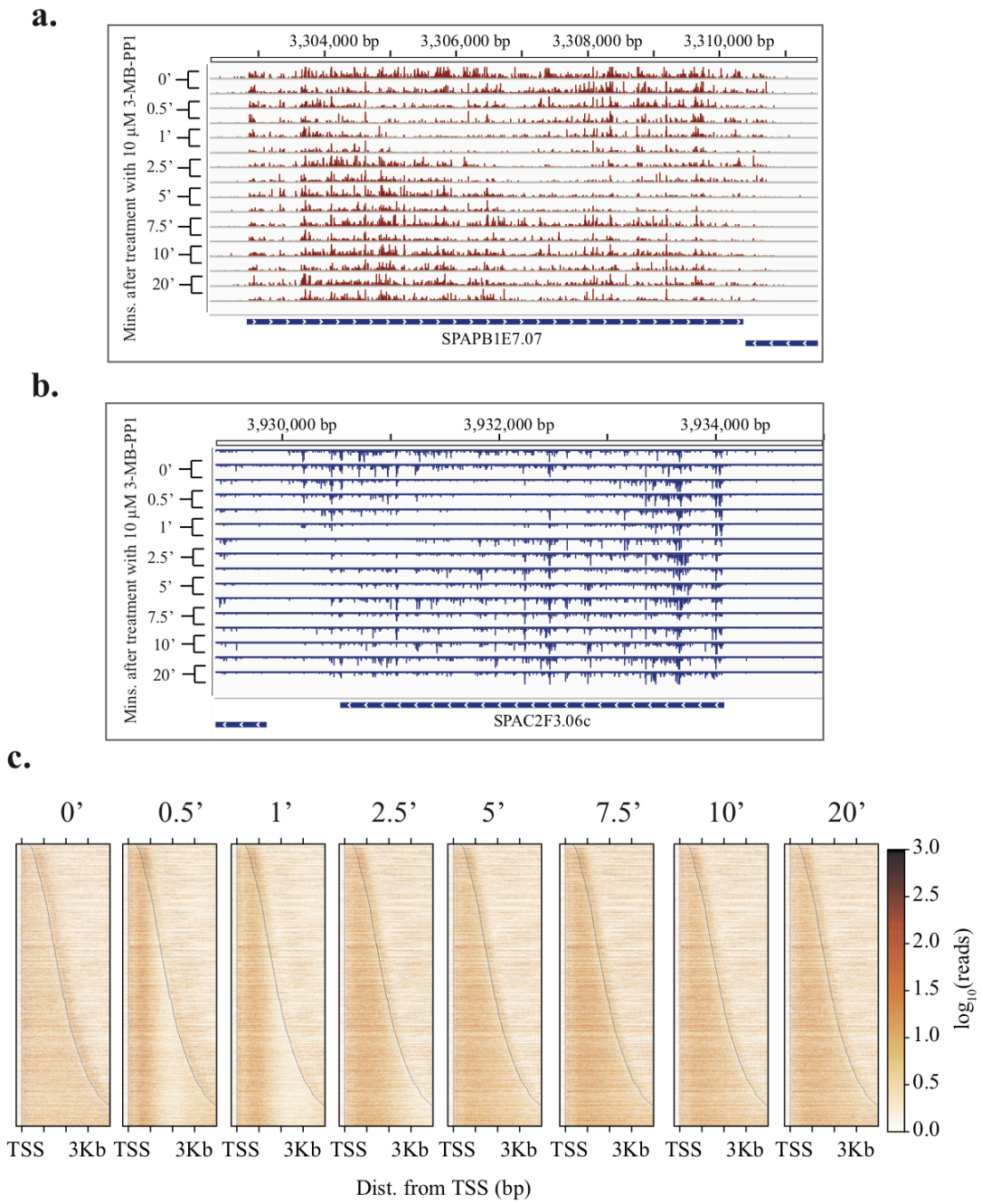

Figure 3.9 | Reproducible time-dependent impact of *Cdk9* inhibition on transcription. a & b. Browser track images from the *SPAPB1E7.07* (**a.**) and

SPAC2F3.06c (**b.**) loci. Tracks represent normalized read counts from biological replicates of *cdk9*^{as} on the plus or minus strands, respectively, treated with 10 μM 3-MB-PP1 for increasing amounts of time. **c.** Raw heatmaps of PRO-seq signal for strains before and after treatment. Heatmap signal intensity reflects normalized signal (log_{10}) within 10 bp bins from -250 to +4000 bp relative to the TSS for data from combined replicates for increasing treatment durations in *cdk9*^{as}, from left to right. Within each panel, genes are sorted by increasing length from top to bottom, with dark lines showing boundaries for each gene. All samples were treated with 10 μM 3-MB-PP1 for the specified times. The zero-minute treatment was treated with DMSO for 20 min. All heatmaps reflect combined data from two biological replicates.

Distinct effects of Cdk9 inhibition on Pol II rates. Gene activation and repression have been used to derive transcription kinetics in mammals by following changes in Pol II distributions over time(Hah et al., 2011; Jonkers et al., 2014; Mahat, Salamanca, Duarte, Danko, & Lis, 2016b). By adapting a previously described model(Danko et al., 2013), the high temporal and spatial resolution of our data allowed us to determine the distance covered by the advancing waves accurately, despite the comparatively short lengths of *S. pombe* genes (Figure 3.10a). For 43 filtered genes longer than 4 kb we were able to estimate advancing wave distance at every time point from 30 seconds to five minutes based on the differences between treated and untreated signals within tiled windows across each gene (Figure 3.11a). Consistent with a forward moving population of Pol II, distance estimates increased with time of *Cdk9* inhibition (Figure 3.10b). Moreover, by fitting a linear regression to distance travelled over time elapsed, we determined that the average rate of transcription in this population was 376 bp/minute (Figure 3.10c, Figure 3.12a), much lower than previous estimates in yeast(Mason & Struhl, 2005).

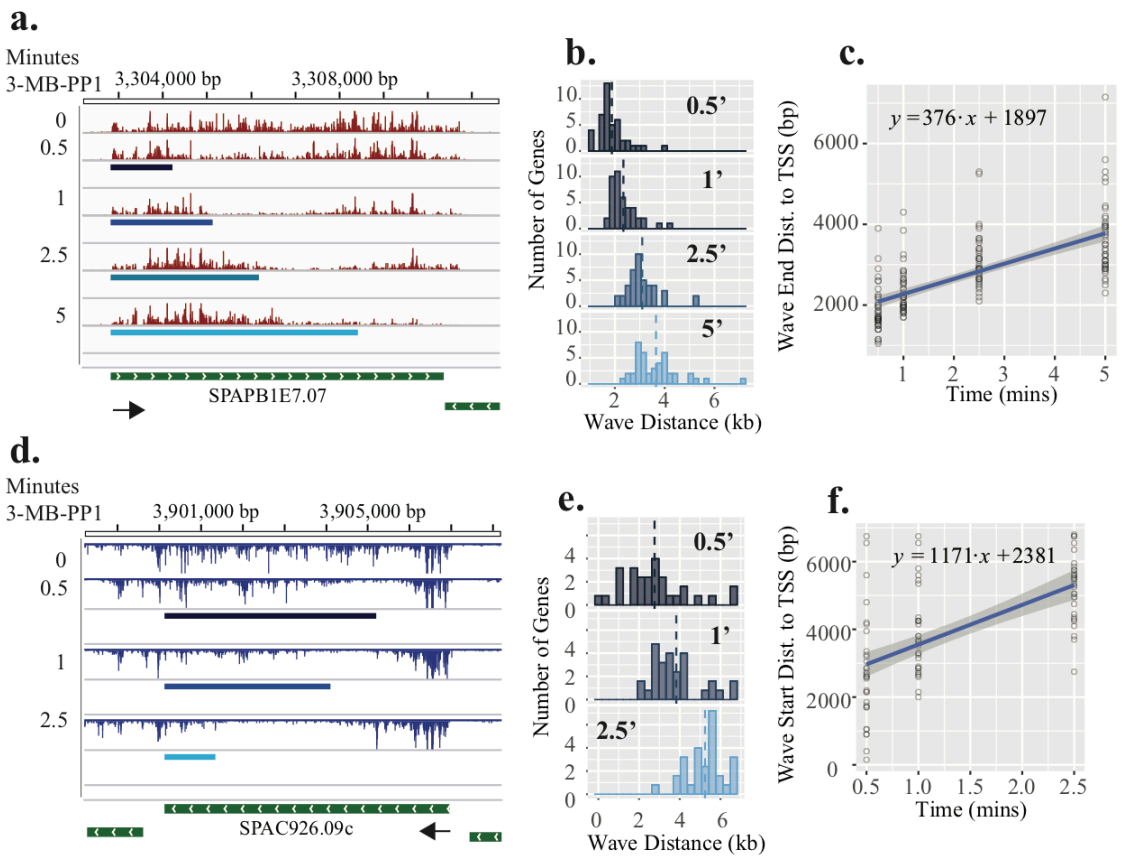

Figure 3.10 | Distinct effects of Cdk9 inhibition on Pol II rates. a. Browser image displaying normalized PRO-seq signal reflecting combined data from two biological replicates for increasing lengths of *cdk9*^{as} inhibition (top to bottom) over the *SPBAPB1E7.07* locus with advancing wave estimates (blue rectangles) for each time point (0.5 min, 1 min, 2.5 min, 5 min) plotted below the corresponding data. **b**. Histograms of estimated advancing wave end distances (relative to TSS) for filtered genes (n = 43) at each time point after addition of 10 µM 3-MB-PP1 to *cdk9*^{as} cells. Dotted, vertical lines represent the mean distance for each distribution. **c**. Linear regression of distances travelled against time elapsed gives the estimated rate of advancing wave progression (slope). **d**. Browser image showing clearing wave measurements (blue rectangles) for each time point plotted below the corresponding data at the *SPAC926.09c* locus. **e**. Histograms of estimated clearing wave tail end distances (relative to CPS) for filtered genes (n = 32) atime points (0.5 min, 1 min, 2.5 min) after *cdk9*^{as} inhibition. **f**. Linear regression of clearing wave distances travelled versus time elapsed gives the estimated rate of clearing wave progression (slope).

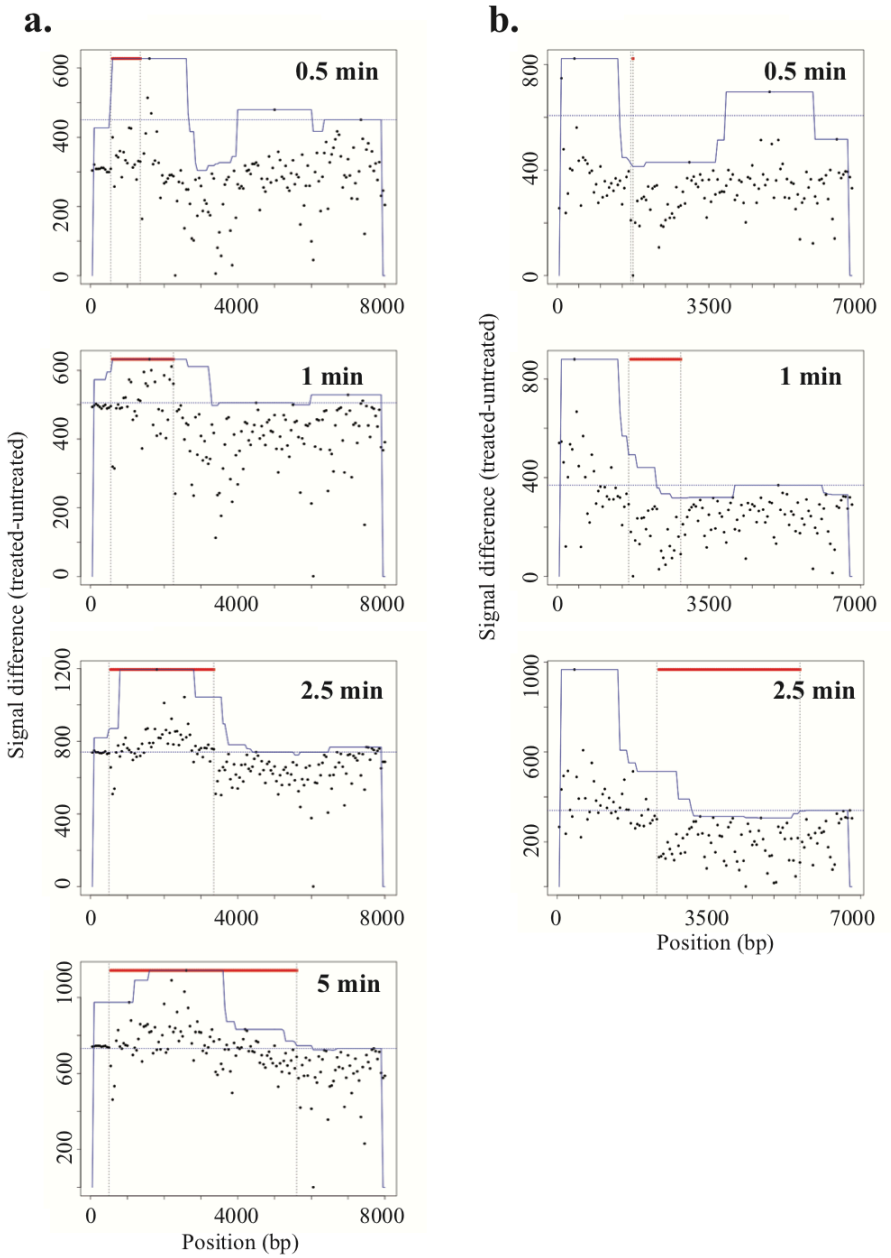

Figure 3.11| Examples of difference maps used for calling waves at each time point. **a.** Advancing wave location estimates based on difference in PRO-seq signal (treated – untreated) measured within 50 bp windows over the *SPAPB1E7.07* locus at each time point (0.5 min, 1 min, 2.5 min, and 5 min). **b.** Estimates of region between advancing and clearing waves based on difference in PRO-seq signal (treated – untreated) measured within 50 bp windows over the *SPAC926.09c* locus at each time point (0.5 min, 1 min, 2.5 min). For each plot, the y-axis displays the difference within each 50 bp window plus a constant (dotted horizontal line), which makes the minimum value equal to one. The red points indicate windows identified as being within the advancing wave (a) or between waves (b). The blue trace reflects the local maximum within a moving 1 kb window.

Slower moving Pol II is thought to be an easier target for exonuclease-mediated termination, as the window within which Pol II terminates downstream of the CPS is directly related to transcription rates(Fong et al., 2015). We therefore expected that on average, the slow population of polymerase, once downstream of the CPS would terminate within a shorter distance from the CPS. Indeed, analysis of global, post-CPS elongation after 20 minutes of *Cdk9* inhibition shows a narrowed zone of termination (Figure 3.12b) and significantly reduced PRO-seq signal within this region when compared with untreated cells (Figure 3.12c & d).

The transient expansion of the region of Pol II clearance seen on the longest genes (Figure 3.8c) suggests that suggests that Pol II further into the gene body is moving faster than the advancing wave upon loss of *Cdk9* activity. To determine the rate of this "clearing" population of Pol II, we further adapted the Danko *et al.* model(Danko et al., 2013) to identify the region between the advancing and clearing waves, which thus contains the start of the clearing wave (Figure 3.11b). Because the clearing population is fleeting and only present on the longest genes, we were forced to omit the five-minute time point and further restrict our analysis to 32 genes longer than 6 kb. Again, clearing wave positions predicted by the model agreed well with PRO-seq profiles (Figure 3.10d) and receded away from the TSS on average (Figure 3.10e). Moreover, the average transcription rate of clearing complexes was estimated to be 1171 bp/min (Figure 3.10f), significantly faster than the advancing wave (Figure 3.12a). This rapid clearing of gene bodies is reminiscent of the effects of FP in mammals, where late transcribing polymerases continue rapidly to clear off genes, apparently unaffected by the drug (Jonkers et al., 2014). Additionally, our results demonstrate disparate effects of *Cdk9*

inhibition on the elongation complex, dictated by its position relative to the TSS. Similar to the inability of metazoan Pol II to escape promoter proximal pausing in the absence of *Cdk9* activity(Ni et al., 2008), we observe a general requirement for *Cdk9* at the 5' ends of genes in *S. pombe*. These observations suggest a general, spatially localized requirement for *Cdk9* activity on the elongation complex within the promoter-proximal region that is conserved in fission yeast. In the absence of this modification, most Pol II is able to transcribe the full length of even the longest genes, but suffers a significantly reduced elongation rate.

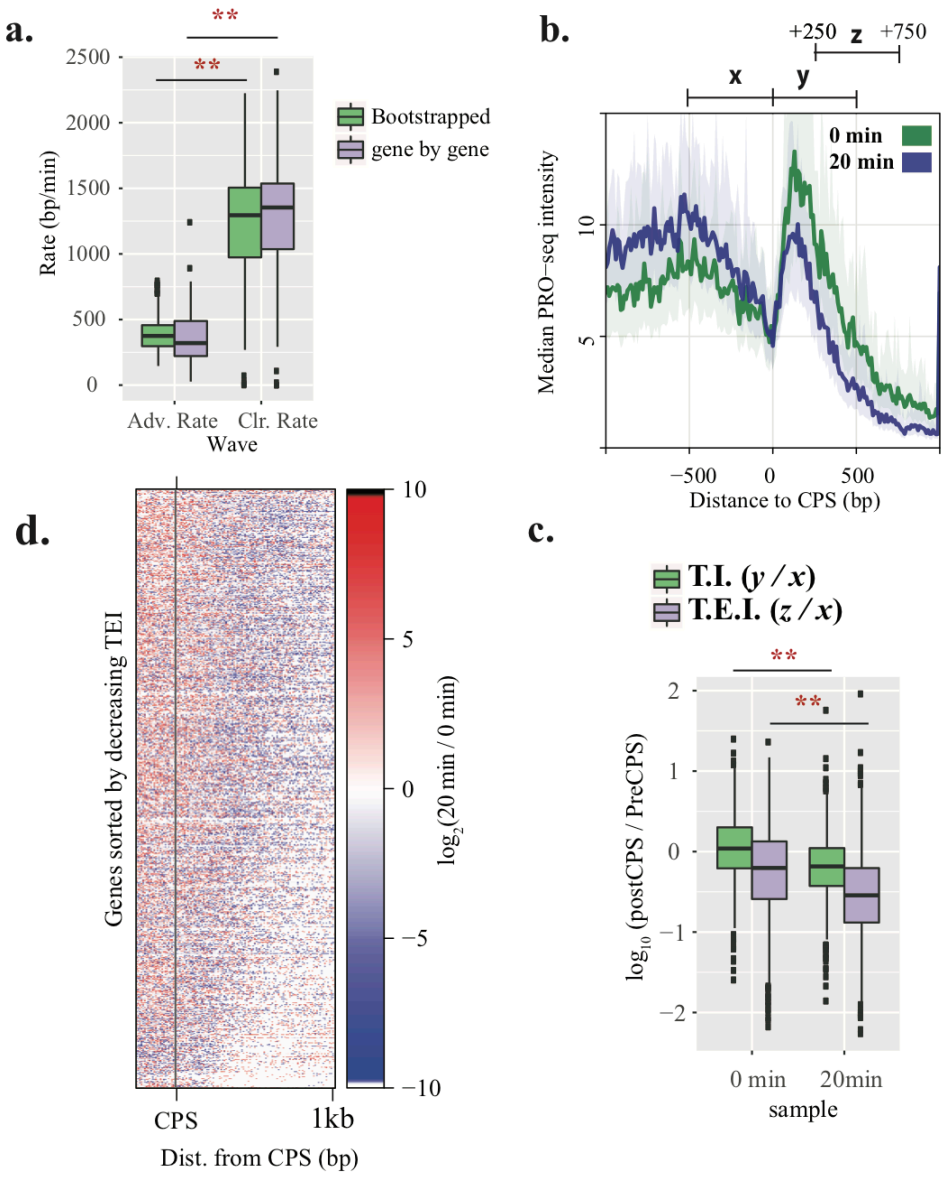

Figure 3.12 | Decreased Pol II transcription rates globally reduce post-CPS elongation upon continued *Cdk9* inhibition. a. Box plots show distributions of rate estimates for the advancing wave (left) and clearing wave (right). Significant differences were observed between advancing and clearing rates using both bootstrapping ($p < 2.2e\text{-}16$, Student's t-test) and gene-by-gene ($p = 9.619e\text{-}08$, Student's t-test) approaches. For each boxplot, the centerline represents the median, the bounds of the box represent the 1st and 3rd quartiles, and the whiskers extend 1.5 times the interquartile range. **b.** CPS centered composite PRO-seq profiles comparing *cdk9*^{as}, either untreated (green) or after 20 min of treatment with 10 μM 3-MB-PP1 (blue). Dark middle lines reflect the median subsampled PRO-seq signal within 10 bp windows from -1 kb to +1 kb of the CPS for genes separated from nearest same strand neighbors by at least 1 kb on both sides (n = 939). Shaded regions correspond to the 12.5% and 87.5% quantiles. **c.** Distributions of termination index (T.I.; green) and termination elongation index (T.E.I., purple) for all genes used in (a) before (left) and after 20 min of *Cdk9* inhibition (right). T.I. and T.E.I.

are metrics that capture elongation beyond the CPS, relative to signal in the late gene body. All 500 bp regions used for calculating T.I. (y / x) and T.E.I. (z / x) ratios are depicted above the plot in (b). **d.** Heatmap of $\log_2$ fold change (20 min/ 0 min) within 10 bp windows from -250 to +1000 bp relative to CPS for genes used in (b) and (c). Genes are sorted by decreasing T.E.I. from top to bottom as calculated from untreated *cdk9*^{as}. Data from combining two replicates for each treatment was used for all plots and calculations.

Discussion

Discerning how post-translational modifications of the elongation complex influence transcription is essential for understanding gene regulation, as these mechanisms are capable of tuning the output of gene expression(Adelman & Lis, 2012). The work presented here advances our understanding of three transcription-coupled kinases, *Cdk9*, *Mcs6*, and *Lsk1*, and their conserved roles in transcription elongation.

Unlike *Cdk9* and *Mcs6*, *Lsk1* is non-essential in *S. pombe*(Karagiannis & Balasubramanian, 2007), and loss of *lsk1*^{as} activity had little impact on transcription elongation within five minutes, despite clearly affecting Ser2-P levels. Nonetheless, subtle reductions in elongating Pol II seen within the 3' post-CPS termination region (Figure 3.6b) might be a symptom of inefficient recruitment of 3' processing machinery(Hsin & Manley, 2012). On the other hand, five-minute inhibition of *Cdk9*, *Mcs6*, or both, produced highly-related shifts in genome-wide Pol II distributions towards the 5' ends of genes. Interestingly, modification of Pol II by *Mcs6* has previously been suggested to enhance recruitment of *Cdk9* in *S. pombe*(Viladevall et al., 2009). Consequently, the inhibition of *Mcs6* might indirectly inhibit *Cdk9*, producing a *cdk9*^{as}-like effect on transcription. However, experiments in budding yeast using an irreversible analog sensitive *Kin28* mutant produced altered transcription profiles similar to what we

observe after five minutes in *S. pombe*, but a full hour after inhibition(Rodríguez-Molina et al., 2016).

As previously reported, we find that the most striking effect of inhibiting *Cdk9* is a rapid reduction of *Spt5* phosphorylation(Pei et al., 2003). This near complete loss of *Spt5* phosphorylation within minutes of *cdk9*^{as} inhibition is consistent with the identification of a phosphatase opposing the activity of *Cdk9* throughout the transcription cycle(Parua et al., n.d.). Genome-wide analysis of elongating Pol II within this timeframe of *Cdk9* inhibition communicates several novel mechanistic insights into the possible role for this activity in transcription elongation. *Cdk9* activity is not essential for escape of transcribing Pol II from the promoter-proximal region in *S. pombe*, but without *Cdk9* activity these transcription complexes exhibit a severely impaired elongation rate. The progression of increased PRO-seq signal (compared with untreated cells) across the entire length of genes with increasing time of *Cdk9* inhibition excludes premature termination as the sole explanation for effects on Pol II elongation. However, we cannot rule out a combination of defects in the elongation rate and processivity of affected Pol II, which might account for non-uniform increases in Pol II density seen at later time points after *cdk9*^{as} inhibition (see Figure 3.8c).

Surprisingly, our Pol II distributions upon loss of *Spt5* phosphorylation differ notably from the depletions of total *Spt5* in *S. pombe*(Shetty et al., 2017). Shetty et al. (2017) reported general reduction of transcribing Pol II within gene bodies after two hours of *Spt5* depletion. In contrast, continued inhibition of *cdk9*^{as} (10-20 minutes) resulted in higher density across genes due to reduced elongation rate, similar to the phenotype of strains lacking the *Spt5* binding partner *Spt4*(Booth et al., 2016).

Importantly, we cannot rule out effects being a consequence of another known or unknown *Cdk9* target. Considering that *Cdk9* is capable of targeting additional transcription-coupled substrates, including the Pol II largest subunit CTD(Pei & Shuman, 2003), it is possible that the effects we observe are attributable to other targets, or combinations thereof. Alternatively, the discrepancy of effects on Pol II transcription may represent separable functions of *Spt5*. The depletion of *Spt5* might directly influence transcription elongation and processivity through loss of known interactions with Pol II, template DNA and the nascent RNA(Ehara et al., 2017; Martinez-Rucobo, Sainsbury, Cheung, & Cramer, 2011; Mason & Struhl, 2005), whereas inhibiting specific phosphorylation of the unstructured, CTR of *Spt5* might only affect transcription rate, perhaps by impairing recruitment of additional factors (Wier et al., 2013).

The loss of *Spt5* phosphorylation at the early stages of elongation may reduce association of factors, which help to negotiate nucleosomes(Lindstrom et al., 2003). Transcription elongation is known to both, modulate and be influenced by the chromatin environment(Ardehali et al., 2009; Bintu et al., 2011; K. Zhou et al., 2009). For example, inability to recruit Paf1 and the FAcilitates Chromatin Transcription (FACT) complex could reduce the ability of the elongation complex to progress through nucleosomes (Adelman et al., 2006). Alternatively, Pol II slowing may occur to ensure recruitment of 5' capping components, similar to pre-mRNA processing steps at gene 3' ends(Gromak et al., 2006; Pei & Shuman, 2002). Inhibiting phosphorylation of *Spt5* might prevent or delay the association of capping enzymes, potentially reducing elongation rates across entire transcription units. Nonetheless, further work comparing the importance of *Spt5* phosphorylation with other *Cdk9* substrates will be crucial for understanding the

mechanisms by which *Cdk9* exerts influence over the kinetics of transcription elongation in fission yeast.

The observation that Pol II farther downstream of the TSS (~1 kb) was less, or possibly unaffected by *Cdk9* inhibition appears directly comparable to clearing waves observed in metazoan systems following the inhibition of P-TEFb(Jonkers et al., 2014). But whereas most other transcribing Pol II is trapped at the promoter-proximal pause site in flies and mammals(Jonkers et al., 2014; Ni et al., 2008), these early gene body elongation complexes continue through the gene with severely impaired rates in *S. pombe*. We propose that NELF, which is a critical pausing factor in Drosophila and mammals(Fujinaga et al., 2004; Wu et al., 2003; Yamaguchi et al., 1999), may serve to exploit an early elongation checkpoint, observed here in fission yeast, creating a rate limiting, *Cdk9*-dependent, mechanism to control gene expression. Indeed, inhibiting the ability of NELF to associate with Pol II in flies has been found to reduce the effects of *Cdk9* inhibition on pause intensity(Gibson et al., 2016). Further work will be required to determine if the four-subunit NELF complex is sufficient to induce *Cdk9*-regulated pausing in other eukaryotes, and whether reduced Pol II elongation rates would result from pause escape in the absence of both NELF and *Cdk9*.

Methods

Yeast strains. Yeast strains are listed in Supplementary Table 1.

Table 3.1| List of strains used and their sources.

Strain	Allias	Genotype	Reference
JS78	*wt*	*leu1-32 ura4-D18 his3-D1 ade6-M210 h^{+}*	Saiz & Fisher 2002
LV7	*cdk9^{as}*	*cdk9^{T120G}::kanMX6 leu1-32 ura4-D18 his3-D1 ade6-M210 h^{+}*	Viladevall et al. 2009
MS340	*mcs6^{as5}*	*mcs6$^{N84T\ -L87G}$::kanMX6 leu1-32 ura4-D18 his3-D1 ade6-M210 h^{+}*	This study
CS118	*lsk1^{as}*	*lsk1^{F353G}::kanMX6 leu1-32 ura4-D18 his3-D1 ade6-M216 h^{-}*	Viladevall et al. 2009
PP20	*cdk9^{as} lsk1^{as}*	*cdk9^{T120G}::hphMX6 lsk1^{F353G}::kanMX6 leu1-32 ura4-D18 his3-D1 ade6-M210 h^{+}*	This study
PP23	*cdk9^{as} mcs6^{as5}*	*cdk9^{T120G}::hphMX6 mcs6$^{N84T-L87G}$::kanMX6 leu1-32 ura4-D18 his3-D1 ade6-M210 h^{-}*	This study
S. cerevisiae	Spike-in	W303-1a	Booth et al. 2016

Construction of *mcs6^{as5}* mutant fission yeast. We previously described *mcs6* and *cdk9* AS mutant strains, each with single substitutions of the gatekeeper residues, Leu87 and Thr120, respectively, with Gly(Viladevall et al., 2009). Although growth of both strains was sensitive to 3-MB-PP1, the *mcs6^{L87G}* strain was ~5-fold less sensitive than was the *cdk9^{T120G}* strain(Viladevall et al., 2009). By standard methods for gene replacement in *S. pombe*(Bähler et al., 1998), we introduced structure-guided second site mutations to attempt to optimize performance of *mcs6^{as}* alleles(C. Zhang et al., 2005) and obtained a double-point mutant, *mcs6$^{N84T/L87G}$* (referred to hereafter as *mcs6^{as5}*), which had a growth rate indistinguishable from that of wild-type cells in the absence of 3-MB-PP1, and was sensitive to growth inhibition by 3-MB-PP1 with an $IC_{50} \approx 5$ μM, which was nearly identical to that of *cdk9^{as}* (Figure 3.5a).

Monitoring loss of target phosphorylation upon kinase inhibition. For measuring loss of phosphorylation on *Spt5* and Rpb1 upon CDK inhibition, cell lysates were rapidly prepared using the trichloroacetic acid (TCA) lysis method(Kao & Osley, 2003). Briefly, cultures were grown to a density of ~1.2 x10^7 cells/ml in YES media at 30°C prior to treatment. After the specified duration of treatment with of 3-MB-PP1 or DMSO, ~ 6 x 10^7 cells were transferred into 500 µl 100% (w/v) TCA and collected by centrifugation. Pellets were resuspended in 20% (w/v) TCA and protein extracts were prepared using a bead beater to mechanically lyse cells with glass beads. Antibodies used in this study recognized *Spt5*-P (used at 1:2000) or total *Spt5* (used at 1:500)(Sansó et al., 2012), total Pol II (BioLegend, MMS-126R; used at 1:1000), Pol II Ser2-P (Abcam, ab5095; used at 1:1000), Pol II Ser5-P (BETHYL A304-408A; used at 1:1000), Pol II Ser7-P (EMD Millipore 04-1570; used at 1:1000), and α-tubulin (Sigma, T-5168; used at 1:5000).

Treatment of analog sensitive strains for PRO-seq. All samples were grown and treated in YES medium at 30 °C. Biological replicates were prepared by picking distinct colonies of a particular strain and growing them in separate liquid cultures. For each experiment, cultures were grown overnight and diluted to an optical density OD_{600} = 0.2. Diluted cultures were grown to mid-log phase (OD_{600} = 0.5-0.6), before treatments. Prior to treatment, all culture volumes were adjusted (based on OD_{600}) to have an equal number of *S. pombe* cells (in 10 mL volume). After normalization for cell number, a fixed amount of thawed *S. cerevisiae* (50 µL OD_{600} = 0.68) culture was spiked in to each sample. Treatments were performed by adding either 10 µM 3-MB-PP1 (stock concentration = 40 mM) or an equivalent volume of DMSO. Sample treatments were

terminated by pouring into 30 mL ice-cold water, treated with diethyl pyrocarbonate DEPC. Samples were immediately spun down and processed for cell permeabilization as described below (See PRO-seq section).

Time course experiments. The time course experiment was performed in two biological replicates, where all time-points for a given replicate was performed on cells from the same large culture. Each time-point treatment was performed as described above. In order to limit differences in time on ice before permeabilization, treatment time-points were carried out in reverse order. The 2.5, 1, and 0.5-minute treatments were performed separately to ensure accuracy of timing. The 0-minute time point was treated with DMSO for 20 minutes to account for any possible effects of adding the solvent for the maximum duration.

PRO-seq. After experimental treatments and spiking in *S. cerevisiae*, cells were spun down and washed in cold H_20. Cells were then permeabilized in 10 mL 0.5% sarkosyl at 4 °C, and incubated on ice for 20 minutes, then spun at a reduced RCF (400 x g) for 5 minutes at 4 °C. For run-on reactions, yeast pellets were resuspended in 120 µL of 2.5X run-on reaction buffer [50 mM Tris –HCl, pH 7.7, 500mM KCL, 12.5 mM $MgCl_2$] with 6 µL 0.1 M DTT and 3.75 µL of each 1 mM biotin-NTP added immediately before use. The volume of the run-on reaction mix was brought to 285 µL with DEPC-treated H_20. 15 µL 10% sarkosyl was then added to the reaction and the run-on was allowed to run on for 5 minutes at 30 °C. RNA was extracted using a hot phenol approach(Collart & Oliviero, 1993); after the run-on reaction cells were pelleted at 400 x g for 5 minutes at 4

°C and quickly resuspended in 500 µL acid phenol. An equal volume of AES buffer [50 mM NaAc pH 5.3, 10mM EDTA, 1% SDS] was added and placed at 65 °C for 5 minutes with periodic vortexing, followed by 5 minutes on ice. 200 µL chloroform was added and mixed followed by centrifugation at 14000 x g for 5 minutes (4 °C). 3 M NaOAc was added to the aqueous layer (200 mM) followed by ethanol precipitation with 3x volume of 100% ethanol. The RNA pellet was air dried before being resuspended in 20 µL DEPC-treated water. The standard PRO-seq protocol(Mahat, Kwak, Booth, Jonkers, Danko, Patel, et al., 2016a) was then followed, beginning with base-hydrolysis, through to sequencing. For base-hydrolysis of RNA, 5 µL 1N NaOH was added to the resuspended RNA and left on ice for 10 minutes before being neutralized with 25 µL 1M Tris-HCL pH 6.8. After passing hydrolyzed RNA through a P-30 column (Bio Rad), RNA was incubated with streptavidin-conjugated magnetic beads (M280, Invitrogen) for 20 minutes at room temperature, then washed with 500 µL of each of the following: high salt solution (50 mM Tris-Cl pH 7.4, 2M NaCl, 0.5% Triton X-100), bead binding solution (10 mM Tris-Cl pH 7.4, 300mM NaCl, 0.1% Triton X-100), low salt solution (5 mM Tris-Cl pH 7.4, 0.1% Triton X-100). Note that prior to incubation with sample RNA, M280 beads were washed once with a solution of 50 mM NaCl and 0.1 N NaOH, and then twice with 100 mM NaCl. After bead binding and washing, enriched nascent RNA was extracted using TRIzol (Ambion) and ethanol precipitated. An RNA adaptor was then ligated to the 3' ends of enriched RNAs (See below) using T4 RNA ligase 1 (NEB), followed by another round of bead binding, washing and TRIzol extraction, as before. RNA 5' ends were then biochemically prepared for adaptor ligation by first removing any cap structures with RNA 5' Pyrophosphohydrolase (RppH, NEB), then

restoring mono-phosphates to base hydrolyzed RNA 5' ends with T4 polynucleotide kinase (T4 PNK, NEB). T4 RNA ligase 1 was then used to ligate the 5' RNA adaptor (5'-CCUUGGCACCCGAGAAUUCCA-3') to enriched RNAs. Following another round of bead binding and washing, as before, cDNA was prepared by reverse transcription using superscript reverse transcriptase III (SSRTIII, Invitrogen) and the primer, RP1 (Illumina, TruSeq Small RNA Sample Prep oligos). Libraries were then amplified via PCR using Phusion (NEB) polymerase and Illumina oligos, RP1 and one indexed RNA PCR Primer (RPI-x; Illumina, TruSeq Small RNA Sample Prep oligos). PCR products were run on a native polyacrylamide gel, size selected (~130-400 bp), extracted, quantified and submitted to sequencing. More details regarding preparing PRO-seq libraries can be found in the published protocol for precision run-on and sequencing(Mahat, Kwak, Booth, Jonkers, Danko, Patel, et al., 2016a).

For specific batches of experiments slight modifications to the library preparation were made. Samples prepared with *wt, mcs6*^{as5}, or *cdk9*^{as}, excluding the time course experiment, were prepared according to the standard procedure described above(Mahat, Kwak, Booth, Jonkers, Danko, Patel, et al., 2016a). All libraries prepared from *lsk1*^{as} mutant and *mcs6*^{as5} *cdk9*^{as} double mutant strains received a novel 3' RNA adaptor during the first RNA ligation step. This 3' adaptor contains a random 6 nt unique molecular identifier at the 5' end that, although not used here, can be used during read processing to remove PCR duplicates (5'-/5Phos/NNNNNNGAUCGUCGGACUGUAGAACUCUGAAC/Inverted dT/-3'). For all time-course samples prepared in *cdk9*^{as}, as well as *lsk1*^{as} *cdk9*^{as} samples, a different 3' RNA adaptor was used. Here the RNA oligonucleotide possessed a known

hexanucleotide sequence preceded by a guanine at the 5' end, which differed for each library prepared (5' -/5Phos/ GNNNNNNGAUCGUCGGACUGUAGAACUCUGAAC-/Inverted dT/). The ligation of this adaptor with distinct barcodes to each library permitted the pooling of all libraries after the 3' end ligation step and thus facilitated handling. After pooling, the remainder of the library preparation was carried out as described above, but in a single tube. After sequencing, the inline barcode was used to parse reads based on their sample of origin.

Alignment and data processing. All sequencing was performed on an Illumina NextSeq 500 device. If samples were pooled during the library preparation using the novel 3' RNA adaptor described above, raw reads were parsed into their respective samples based on the inline barcode using fastx_barcode_splitter function from the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Raw reads from each sample were processed by removing any instances of partial or complete matches to the 5' adaptor sequence (5'- TGGAATTCTCGGGTGCCAAGG -3') with the fastx_clipper function. Next, if a 3' molecular, or sample barcode was included during 3' adaptor ligation, this length was removed from the beginning of each read. All reads were then trimmed to a maximum remaining length of 36 nt using the fastx_trimmer function. With the fastx_reverse_complement function we retrieved the reverse complement sequence of each read. All downstream alignments were performed using Bowtie (version 1.0.0) (Langmead et al., 2009). Reverse complemented reads derived from ribosomal RNA genes were then removed through alignment to a genome consisting of ribosomal RNA genes from *Saccharomyces cereivisiae* (spike-in) and then *Schizosaccharomyces pombe*.

To parse reads based on organism of origin, while eliminating reads of ambiguous origin, non-ribosomal DNA reads were then aligned to a combined genome, containing each chromosome from *S. cerevisiae* (sacCer3 = S288C_reference_genome_R64-1-1_20110203) and *S. pombe* (version: ASM294v2). Only uniquely mapping reads were considered for downstream analysis. Ultimately, the normalization factor for each library was calculated as: total spike-in mapped reads/10^5. Unique read alignments to the fission yeast genome were processed using the Bedtools(Quinlan & Hall, 2010) function, genomeCoverageBed, to generate bedgraph formatted files, reporting the number of read 3' ends (the last base incorporated) at each position across the genome. Bedgraph files were further converted to bigwig format for downstream analysis. Sequencing, alignment, and batch information for each sample can be found in Supplementary Table 2.

Table 3.2| Sequencing and alignment statistics from PRO-seq samples

sample	spike-in ribosomal	ribosomal	spike-in	genome	total	Norm. Factor	batch
WT_DMSOr1	10,174,782	12,285,660	46,236	6,157,670	30,465,302	0.46236	1
WT_DMSOr2	12,204,603	15,627,984	54,480	8,401,775	38,708,110	0.5448	1
WT_3MBPP1_1	12,326,132	15,585,548	80,667	10,653,445	41,242,195	0.80667	1
WT_3MBPP1_2	10,582,750	13,617,874	62,899	8,593,076	35,067,313	0.62899	1
mcs6_DMSO1	12,257,633	15,672,640	80,616	10,205,273	41,454,045	0.80616	1
mcs6_DMSO2	11,707,207	14,554,011	56,817	8,051,885	36,735,682	0.56817	1
mcs6_3MBPP1_1*	11,430,584	13,874,952	101,502	6,331,691	37,568,528	1.01502	1
mcs6_3MBPP1_2	12,052,555	14,402,612	72,480	8,466,890	37,221,280	0.7248	1
CDK9_DMSO1	12,389,079	14,034,425	54,917	6,367,375	34,927,058	0.54917	1
CDK9_DMSO2	12,778,993	16,323,383	66,519	9,358,197	41,353,473	0.66519	1
CDK9_3MBPP1_1	11,143,318	13,228,901	60,777	7,827,810	34,514,302	0.60777	1
CDK9_3MBPP1_2	12,517,108	16,637,590	88,178	12,311,216	45,190,617	0.88178	1
Isk1as_DMSOr1	8,137,437	9,062,002	73,907	6,843,886	29,144,786	0.73907	2
Isk1as_DMSOr2	10,483,655	11,431,617	67,401	8,289,801	35,769,690	0.67401	2
Isk1as_3MBPP1r1	6,935,506	7,256,537	54,000	5,481,567	23,207,739	0.54	2
Isk1as_3MBPP1r2	3,080,010	2,899,634	17,547	2,063,159	12,330,930	0.17547	2
Mcs6as_CDK9as_DMSOr1	7,733,612	8,014,461	37,501	5,750,563	26,364,431	0.37501	2
Mcs6as_CDK9as_DMSOr2	8,149,246	8,775,629	41,878	5,992,027	26,861,563	0.41878	2
Mcs6as_CDK9as_3MBPP1r1	7,206,380	7,247,941	33,500	4,536,765	23,488,806	0.335	2
Mcs6as_CDK9as_3MBPP1r2	6,378,728	6,817,084	36,155	4,582,814	20,834,265	0.36155	2
Cdk9as_0min_DMSO_rep1	941,848	1,294,433	8,443	1,240,837	6,521,490	0.08443	3
Cdk9as_0min_DMSO_rep2	1,065,358	1,504,048	13,568	2,005,311	10,353,030	0.13568	3
Cdk9as_0.5min_10uM_3MBPP1_rep1	3,205,493	3,805,818	24,773	3,807,026	15,454,733	0.24773	3
Cdk9as_0.5min_10uM_3MBPP1_rep2	1,139,637	1,662,171	16,380	1,738,590	7,779,488	0.1638	3
Cdk9as_1min_10uM_3MBPP1_rep1	1,699,913	2,207,482	16,573	2,128,619	9,584,199	0.16573	3
Cdk9as_1min_10uM_3MBPP1_rep2	2,130,628	3,151,162	30,364	3,413,440	12,760,335	0.30364	3
Cdk9as_2.5min_10uM_3MBPP1_rep1	1,753,947	2,386,420	14,135	1,973,114	8,645,482	0.14135	3
Cdk9as_2.5min_10uM_3MBPP1_rep2	2,557,168	3,377,095	23,249	3,621,442	14,120,294	0.23249	3
Cdk9as_5min_10uM_3MBPP1_rep1	2,394,714	3,307,974	20,338	2,812,736	15,961,497	0.20338	3
Cdk9as_5min_10uM_3MBPP1_rep2	2,504,115	2,672,961	26,141	3,742,474	17,870,169	0.26141	3
Cdk9as_7.5min_10uM_3MBPP1_rep1	1,998,776	2,768,515	15,073	2,612,802	16,268,203	0.15073	3
Cdk9as_7.5min_10uM_3MBPP1_rep2	2,662,662	3,131,133	33,314	5,316,423	21,503,832	0.33314	3
Cdk9as_10min_10uM_3MBPP1_rep1	2,228,285	2,681,515	17,655	3,070,909	15,621,102	0.17655	3
Cdk9as_10min_10uM_3MBPP1_rep2	1,000,575	1,441,581	14,946	1,865,992	9,003,041	0.14946	3
Cdk9as_20min_10uM_3MBPP1_rep1	1,059,830	1,275,738	11,469	1,802,836	8,859,279	0.11469	3
Cdk9as_20min_10uM_3MBPP1_rep2	2,100,685	2,635,181	22,006	3,327,234	15,608,295	0.22006	3
Lsk1as_CDK9as_DMSOr1	1,929,149	2,510,779	32,183	2,953,437	10,834,859	0.32183	3
Lsk1as_CDK9as_DMSOr2	2,795,832	3,489,519	34,863	4,400,322	16,293,462	0.34863	3
Lsk1as_CDK9as_3MBPP1r1	1,566,015	2,350,911	24,970	2,915,731	11,162,211	0.2497	3
Lsk1as_CDK9as_3MBPP1r2	2,789,279	3,492,754	30,235	4,278,326	18,766,044	0.30235	3

* Sample omitted from analysis

Combined replicates. Biological replicates correlated very well with spike-in normalization centering gene by gene scatter around the line $x=y$ (Figure 3.3), allowing us to combine the raw data from replicates (pre-alignment) for added read depth when useful. In one particular sample, *mcs6*^{as5} treated with 3-MB-PP1 for 5 minutes, we ultimately omitted biological replicate one (Supplementary Table 2) from all analysis in this work, including composite profiles, due to concerns with the library quality. Importantly, limiting differential expression analysis between treated and untreated

samples to only one replicate in the treated condition reduces power to detect differences due to overestimates of variance (M. Love et al., 2013). Although the number of real changes in the tested gene regions after inhibition of *mcs6*^{as} may be higher than listed in Figure 2e, the reported global shifts in Pol II distribution can be seen in individual replicates.

Experimental batches. Three "batches" of experiments were performed to generate the data in this work (Supplementary Table 2). In order to minimize noise introduced from across-batch comparisons, all analyses presented herein were restricted to PRO-seq libraries prepared within the same batch.

Gene sets. Using previously published PRO-cap data(Booth et al., 2016), we re-annotated transcription start sites as the position with the maximum amount of background-subtracted PRO-cap signal within the region -250 to +250 around existing annotations (version: ASM294v1.16). The selected base was required to have > 4 reads over background, while accounting for library size (annotations available at https://github.com/gregtbooth/Pombe_PROseq). All genes used in this work were required to have an observed transcription start site (n = 4654), and this observed TSS was used throughout. We then filtered genes for "activity" in our untreated, wild type data. A gene was considered active if the read density within the gene body was significantly higher ($p < 0.01$) than a set of intergenic "background" regions, based on a Poisson distribution with λ = background density ($\lambda = 0.0402$; n active = 4576). Finally, to minimize the possible influence of read-through transcription from upstream genes, we

filtered genes based on the relative amounts of PRO-seq signal directly upstream of the TSS. Thus, we only considered genes with more downstream (+250 to +550) relative to immediately upstream (-300 to observed TSS; n = 3383).

Differential expression analysis. Raw reads from the appropriate strand were counted within specified windows using custom scripts. Differential expression analysis was performed for desired regions using the DESeq2 R package(M. Love et al., 2013). Rather than using default between-sample normalization approaches, we supplied our own spike-in based normalization factors (Supplementary table 2) using the "sizeFactors" argument. Genes and regions were considered significantly changed (up or down) if they were computed to have an adjusted p-value < 0.01. To compute adjusted p-values for changes in counts DESeq2 uses the Wald test with Benjamini and Hochberg's correction(M. Love et al., 2013).

Advancing and clearing wave analysis. Both advancing and clearing waves were identified using a three state Hidden Markov Model (HMM) that was previously developed and implemented on GRO-seq data from a human cell line(Danko et al., 2013). To increase the number of time points for which we could identify advancing waves, we restricted our analysis to calling waves on filtered genes longer than 4 kb (n = 231) at each time point up to five minutes. Wave calling parameters were adjusted to accommodate the smaller genes of yeast. For calling advancing waves, the upstream region used was set to -500 to the observed TSS for each gene. Approximate wave distances, used to initialize the model, were set based on estimates derived from

inspecting the fold-change heat maps (30 sec = 1 kb; 1 min = 1.5 kb; 2.5 min = 2.5 kb; 5 min = 3.5 kb). Wave distances were determined based on the difference in signal within 50 bp windows between treated and untreated. The wave quality was determined based on the criteria used previously(Danko et al., 2013). We further required that for each gene, a wave must be called for all used time points (maximum 5 minutes) using the combined replicate data, and that the calls must not recede toward the TSS with time (n = 43).

To identify regions containing the faster population of Pol II (i.e. the clearing wave), we adjusted the HMM to identify the region between the advancing and clearing waves. For this task, we used the first 1000 nt downstream of the observed TSS (within the advancing wave) to initialize the first state of the HMM. Approximate distances of the start of the clearing wave from the TSS were used to initialize the model and were set based on estimates derived from inspecting the fold-change heat maps (30 sec = 3 kb; 1 min = 4 kb; 2.5 min = 5.5 kb). Unlike the model used to call the advancing wave, which assumed the upstream state to be normally distributed, all three states were presumed to behave according to distinct gamma distributions. Because of the fleeting nature of this population of polymerases observed in our time course, we were further limited to analysis of filtered genes longer than 6 kb (n = 42). Again, 50 nt windows were used to tile across genes, however an additional smoothing parameter in the model arguments (TSmooth) was set to 5 as a way of restricting the effect of windows with outlying differences between treated and untreated. For each gene, we required that the model be able to identify a distinct clearing wave (Kullback-Leibler divergence > 1, between states 2 and 3) at each time point and that did not move towards the TSS with time (n = 32).

We employed two methods to assess the confidence in measured average rates. For the first, bootstrapping approach, an average rate was calculated using 10% of genes for which we called waves. This process was repeated 1000 times (with replacement), giving a distribution of means. A more conservative, gene-by-gene estimate of variance was derived simply from the distribution of rates based on individual measurements for each gene.

Data Availability

The raw and processed sequencing files have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession GSE102308

Code Availability

Custom scripts and alignment pipelines have been made publicly available through the following GitHub repository; https://github.com/gregtbooth/Pombe_PROseq.

CHAPTER 4

NELF REFINES THE RESPONSE OF PAUSED RNA POLYMERASE II TO P-TEFB AND PREVENTS UNPRODUCTIVE TRANSCRIPTION

Abstract

Promoter-proximal pausing is a pervasive mechanism of transcription regulation in many metazoans, but how and why it evolved remains unknown. Negative elongation factor (NELF) is a pausing factor only found in organisms which can regulate transcription output through RNA polymerase II (Pol II) pausing, suggesting it is critical to this form of regulation. Here, we investigate the importance of NELF, as well as the conserved factor, DSIF, in regulating pause escape in *Drosophila melanogaster*. We observe that promoter-proximal pausing positively contributes to gene expression, as loss of pausing factors decreases global Pol II elongation. Surprisingly, the requirement of P-TEFb kinase activity for pause escape is largely unaffected by depletion of DSIF, NELF, or both factors, though stability of paused Pol II is reduced globally. However, in the absence of NELF, significantly more Pol II can escape promoter-proximal arrest without P-TEFb activity. Yet, transcription complexes that enter into elongation without input from P-TEFb, prematurely terminate within 3-4 kb of the transcription start site (TSS). These results suggest that NELF may refine an inherent pausing architecture of *D. melanogaster* promoters, further ensuring proper maturation of the elongation complex for productive transcription.

Introduction

In most well-studied metazoans, promoter-proximal Pol II pausing is accepted as a ubiquitous regulatory and often rate limiting step in gene expression(Adelman & Lis, 2012). However, contrary to the idea of negative elongation factors, or barrier to elongation, this early checkpoint has been found to positively influence transcription elongation. Rather than inducing widespread increases in gene-expression, loss of pausing, or the factors involved, often reduces mRNA production(Gilchrist et al., 2008). But why regulated stalling of the engaged transcription complex immediately downstream of the TSS would be necessary for productive downstream elongation remains poorly understood. Indeed, in many eukaryotes no such mechanism has been observed.

Unlike the large enrichments of Pol II at promoter-proximal positions of mammalian and fly genes, Pol II is relatively evenly distributed across genes in the budding yeast *Saccharomyces cerevisiae*(Booth et al., 2016; Churchman & Weissman, 2011; E. J. Steinmetz et al., 2006). The study of transcription regulation in such systems fostered the idea that transcription is controlled primarily through the recruitment of Pol II to gene promoters, rather than during active RNA synthesis(Ptashne & Gann, 1997). Even so, in another distantly related yeast, *Schizosaccharomyces pombe*, early Pol II gene body enrichments resemble pause like distributions of metazoans, despite lacking orthologs of negative elongation factor (NELF) (Booth et al., 2016; Narita et al., 2003). However, in contrast to P-TEFb regulated pause escape in metazoans, Pol II in fission yeast can transcribe across gene bodies in the absence of Cdk9 activity, although with a severely impaired elongation rate(Booth, Parua, Sansó, Fisher, & Lis, 2018). Such

results suggest that modifications performed by the conserved P-TEFb kinase, Cdk9 play a critical role in the early stages of transcription, beyond release of the paused Pol II into productive elongation.

Pol II pausing occurs almost immediately after transcription initiation, following the polymerization of just 20 to 40 nucleotides of RNA(Rougvie & Lis, 1988). Within this often-narrow window, pausing factors DSIF and NELF associate with Pol II to capture and prevent further RNA synthesis, until the complex is subject to modification by the kinase subunit of P-TEFb(Wada, 1998; Wada et al., 1998; Yamaguchi et al., 1999). This regulated capture and release of engaged Pol II is thought to enable a high degree of regulatory complexity(Peterlin & Price, 2006). Pausing has been proposed to facilitate rapid and/or synchronous gene activation or the integration of a multitude of signals for controlling gene expression(Adelman & Lis, 2012; Lagha et al., 2013). Maintaining engaged Pol II occupancy near the TSS could also help to maintain active and nucleosome-free promoters(Gilchrist et al., 2010). Moreover, capturing Pol II in the earliest stages of elongation represents an opportunity to ensure efficient and processive elongation across enormous stretches of chromatin.

The 5' ends of nascent RNA must be capped to facilitate protein translation and prevent unwanted transcript degradation and is largely coincident with promoter-proximal Pol II pausing(Nechaev et al., 2010; Rasmussen & Lis, 1993). Arresting Pol II could act to prevent wasteful rounds of transcription by assuring RNA has been capped at this early stage(Pei et al., 2003). Additionally, within these promoter-proximal regions, the elongation complex, including DISF and Pol II itself, are heavily phosphorylated within unstructured C-terminal domains(Buratowski, 2003). These modifications likely

bestow increased ability to navigate and manipulate chromatin by recruiting a battery of additional factors(Wier et al., 2013; Yamada et al., 2006; D. W. Zhang et al., 2012). Pausing could thus act to facilitate the maturation of a complex facing various co-transcriptional hurdles. In spite of this patient-maturation model, transcription across many genes in flies and mammals does not appear to slow down or pause for this input(Core et al., 2008; 2012). Thus, regardless of the modulations of the elongation complex, Pol II dwell times within the promoter-proximal regions appear to depend primarily on the recruitment and activity of P-TEFb(Buckley, Kwak, Zipfel, & Lis, 2014; Jonkers et al., 2014).

In metazoans, it is challenging to study the impact of P-TEFb on transcription elongation beyond pausing. Application of selective Cdk9 inhibitors has the effect of preventing escape of the elongation complex from the promoter-proximal position at nearly all Pol II transcribed genes. Moreover, RNA polymerases actively transcribing downstream of the pause are seemingly unaffected by the loss of Cdk9 activity, suggesting they have already received the critical input of this enzyme(Jonkers et al., 2014; Ni et al., 2008). It remains unknown exactly how P-TEFb mediates pause escape, or what effects on transcription are imparted by its kinase activity. Since NELF is ejected from the elongation complex during pause escape, it seems likely that whatever function NELF performs is critical only during promoter proximal pausing(Fujinaga et al., 2004; Rahl et al., 2010). Here, we investigate the influence of pausing factors, NELF and DSIF in creating a P-TEFb regulated barrier to transcription elongation. Using the S2 cell line derived from *Drosophila melanogaster*, we use RNA interference (RNAi) to deplete critical components of each complex before examining the effects of P-TEFb inhibition.

Precision run-on sequencing (PRO-seq) following knockdown of NELF and/or DSIF suggests precise Pol II pausing overwhelmingly affects gene activation, rather than repression. Surprisingly, even in the absence of pausing factors, inhibiting Cdk9 prevents the majority of Pol II from escaping the pause. However, without modification by Cdk9, loss of NELF results in significant leakage of Pol II into the early gene body, which are incapable of transcribing more than 3-4 kb. Our results indicate NELF refines promoter-proximal pausing to prevent unproductive elongation from occurring before P-TEFb modifies the complex.

Results

Pausing factors are required to maintain Pol II transcription elongation levels globally. Previous work has looked into the behavior of Pol II after the loss of pausing factors. In vivo studies have shown reductions in pausing, but a genome wide absolute loss of pausing, has never been achieved. To study the function of NELF and DSIF in controlling Pol II promoter proximal pausing and elongation, we used RNAi to knock down NELFe and Spt5 subunits of each complex, respectively. Our decision to target the NELFe subunit of the NELF complex was driven in part by previous work which reported that the depletion of NELFe similarly affects levels of other the other NELF subunits(J. Sun & Li, 2010). To compare the influence of these complexes on transcription, the NELF and DSIF subunits were depleted separately. Additionally, as NELF and DSIF are believed to cooperate in establishing the barrier to elongation, we included a sample in which both NELFe and Spt5 were simultaneously targeted for depletion (Figure 4.1a). To examine the impact of such perturbations on transcription by

Pol II, we prepared PRO-seq libraries from two biological replicate cultures for each knockdown condition. Importantly, prior to the precision run-on reactions and subsequent library preparations, mouse embryonic fibroblast cells were spiked into each sample to control for possible global effects of the knockdowns on global transcription (see methods). PRO-seq reads originating from spike-in were then used to normalize our experimental data from *D. melanogaster.* With the exception of the double knockdown samples (see methods), spike-in normalized Pol II density within gene bodies (TSS+200 to CPS) from biological replicates is symmetrically distributed around the line $x = y$ and highly correlated (Figure 4.2), suggesting accurate normalization as well as reproducibility of the experiments.

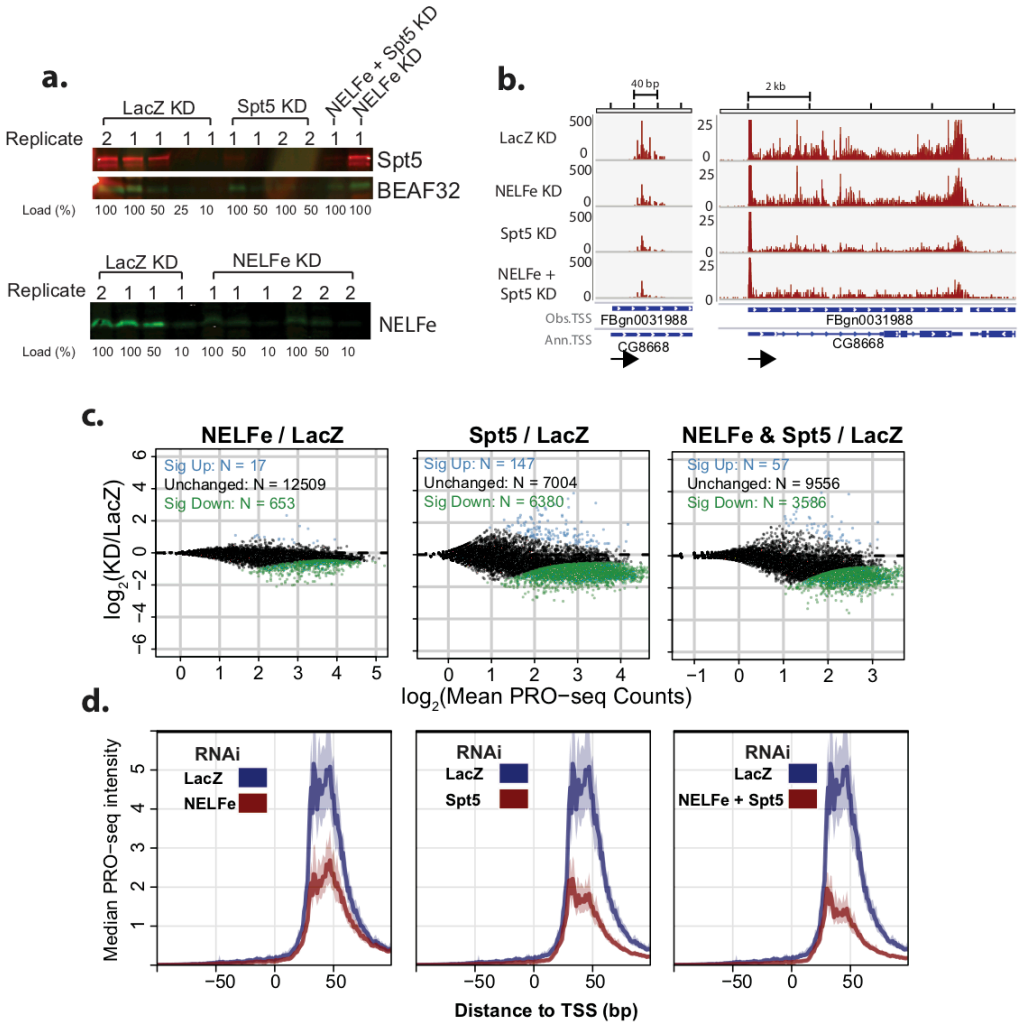

Figure 4.1| Pausing factors are required to maintain Pol II transcription elongation levels globally. **a.** Western blots validating RNAi mediated depletion of Spt5 (top) compared to a loading control (BEAF32), or NELFe (bottom). **b.** Browser tracks of normalized PRO-seq data from combined replicates (n=2) of each knockdown experiment at the *GC8668* locus (only the plus strand is shown). The left panel displays a zoomed in view of the *GC8668* promoter proximal region with an expanded y-axis (0-500) to fully capture the pause peak. The right panel shows the entire gene, with a y-axis (0-25). **c.** DESeq2(M. Love et al., 2013) was used to identify significant changes in gene body (TSS+200 to CPS) polymerase density for all genes (input n = 16164) after knocking down pausing factors (adjusted $p < 0.01$). Panels show $\log_2$ fold change in normalized gene body read density for NELFe RNAi, Spt5 RNAi, or double knockdown of NELFe and Spt5, each over LacZ RNAi treated samples, from left to right respectively. **d.** Promoter-proximal, composite PRO-seq profiles centered on the observed TSS for all filtered genes (n = 3565) for each knockdown (red) compared with LacZ RNAi (blue). From left to right, panels show NELFe RNAi, Spt5 RNAi, and NELFe +Spt5 RNAi in red and LacZ RNAi in blue. The center line displays the median

read density from all filtered genes at each base around the TSS. Shaded areas represent bootstrapped 12.5% and 87.5% confidence intervals of the median signal.

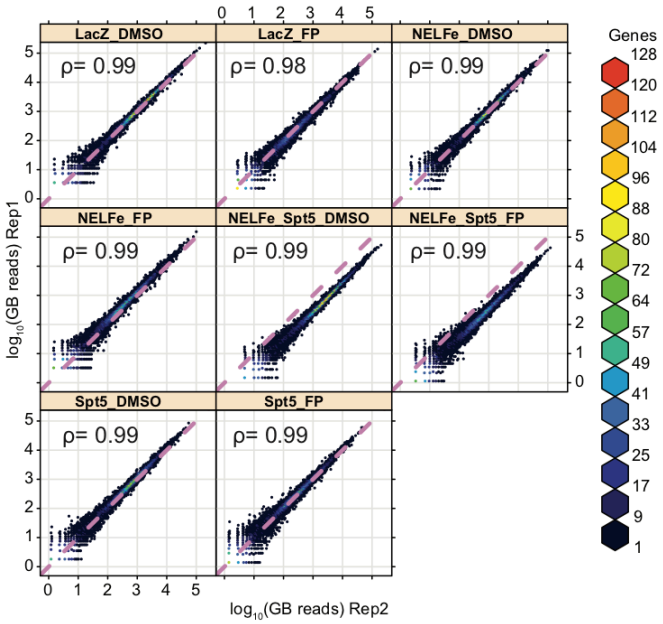

Figure 4.2| Biological replicates are highly correlated. Scatterplots display the comparison of spike-in normalized read counts within the gene body (TSS+200 to CPS) of all genes (n = 16164) from biological replicates of the same sample.

Consistent with previous observations(Gilchrist et al., 2010; Rahl et al., 2010), depletion of pausing factors dramatically reduced the level of Pol II found within promoter-proximal regions of individual genes (Figure 4.1b, left panel). Surprisingly, the amount of Pol II actively elongating within the gene body also appeared to be reduced upon loss of pausing factors, most strikingly after knockdown of Spt5 or both factors simultaneously (Figure 4.1b, right panel). To test whether levels of actively elongating Pol II were globally affected by the loss of pausing factors, we quantified the amount of PRO-seq signal within gene bodies of all genes for each experiment and compared data from each knockdown sample with the LacZ RNAi control. All knockdown conditions impacted levels of transcription significantly. In comparison with the effects of Spt5

depletion, which severely decreased the transcription of nearly half of the examined genes (6380, 47%), removing NELF had a much more modest effect on gene body Pol II density, decreasing transcription of only 653 genes (5%) (Figure 4.1c). Similar to loss of Spt5 alone, removal of both pausing factors significantly reduced transcription of thousands of genes (3586, 27%). Remarkably, in all experiments where pausing factors were removed genes were essentially exclusively unchanged or decreased, with only a handful of genes showing increased Pol II levels. These findings strongly implicate promoter-proximal pausing as a positive influence on transcription and gene expression.

As observed at the *GC8668* locus, Pol II densities within promoter proximal regions immediately downstream of the observed TSS are globally reduced in the absence of DSIF or NELF (Figure 4.1d). The most dramatic reduction in pausing occurs in the absence of both factors, and the loss of Spt5 more strikingly reduces overall Pol II density at the pause than NELF. In addition to reducing the levels of paused Pol II, NELF removal also appears to favor a downstream-shifted pause position, perhaps related to distal dispersed pausing, as described previously(H. Kwak et al., 2013). These analyses clearly indicate that the loss of either NELF or DSIF globally reduces promoter proximal pausing, likely impacting various properties of this regulatory checkpoint. As a result, rather than resulting in an increase of elongating Pol II downstream of the pause, loss of pausing often results in less Pol II transcribing within gene bodies, suggesting that pausing is required for productive elongation by Pol II.

In the absence of DSIF and NELF, Cdk9 activity remains required for the escape of most Pol II from the pause. The regulation of transcription through promoter-proximal

pausing of RNA Pol II is enabled through a release mechanism orchestrated via post-translational modification of the paused elongation complex. The conserved and essential kinase subunit of P-TEFb, Cdk9 is responsible for phosphorylation of NELF, DSIF and Pol II, and is required for releasing paused Pol II in to the gene body(Peterlin & Price, 2006). Simply inhibiting Cdk9 activity in mammals or flies prevents genome-wide pause escape of Pol II, globally shutting down transcription (Jonkers et al., 2014; Ni et al., 2008), but how it's activity triggers pause escape is not understood. To investigate the contributions of pausing factors in establishing a P-TEFb dependent barrier to elongation, we tested the ability of Cdk9 inhibition to prevent pause escape in each knockdown sample.

Inhibiting pause escape globally reduces gene expression and will gravely impact normal cell physiology within a short time frame. To distinguish changes in primary effects of Cdk9 inhibition after depletion of each pausing factor, Cdk9 was targeted for inhibition using 500 nM flavopiridol (FP) for 10 minutes in each RNAi treated sample prior to precision run-on sequencing. Surprisingly, in all knock down samples, 10-minute inhibition of Cdk9 globally induced an increase in Pol II signal within the promoter proximal regions on average (Figure 4.3a, left panels). Concomitant with the increased signal at gene 5' ends, the levels of Pol II within the gene body decreased within most genes, suggesting an inability of Pol II to escape the pause regardless of the knockdown (Figure 4.3a, right panels). Nonetheless, compared with the LacZ RNAi control, depleting pausing factors severely limited the magnitude of paused Pol II enrichment after FP treatment. Interestingly, unlike for the Spt5, or double knockdown samples, genes affected by NELFe depletion (236/261) predominantly had more Pol II

within the gene body after Cdk9 inhibition when compared with the LacZ RNAi control (Figure 4.3b, top panel). Although only a modest effect, these genes might be symptomatic of an increased ability of Pol II to escape the pause regardless of Cdk9 activity.

It is important to consider that the modest impact of pausing factor depletion on P-TEFb responsiveness and the inability to completely eliminate pausing may reflect an incomplete removal of NELF and DSIF. Nonetheless, NELF and DSIF apparently strengthen or increase the duration of the pause, perhaps by promoting stability of the complex. While NELF may act to refine Cdk9 dependence of the paused Pol II complex, these results suggest that *in vivo*, Cdk9 may be required to overcome promoter-proximal pausing, regardless of two factors initially identified as the drivers of this barrier *in vitro*(Wada et al., 1998; Yamaguchi et al., 1999).

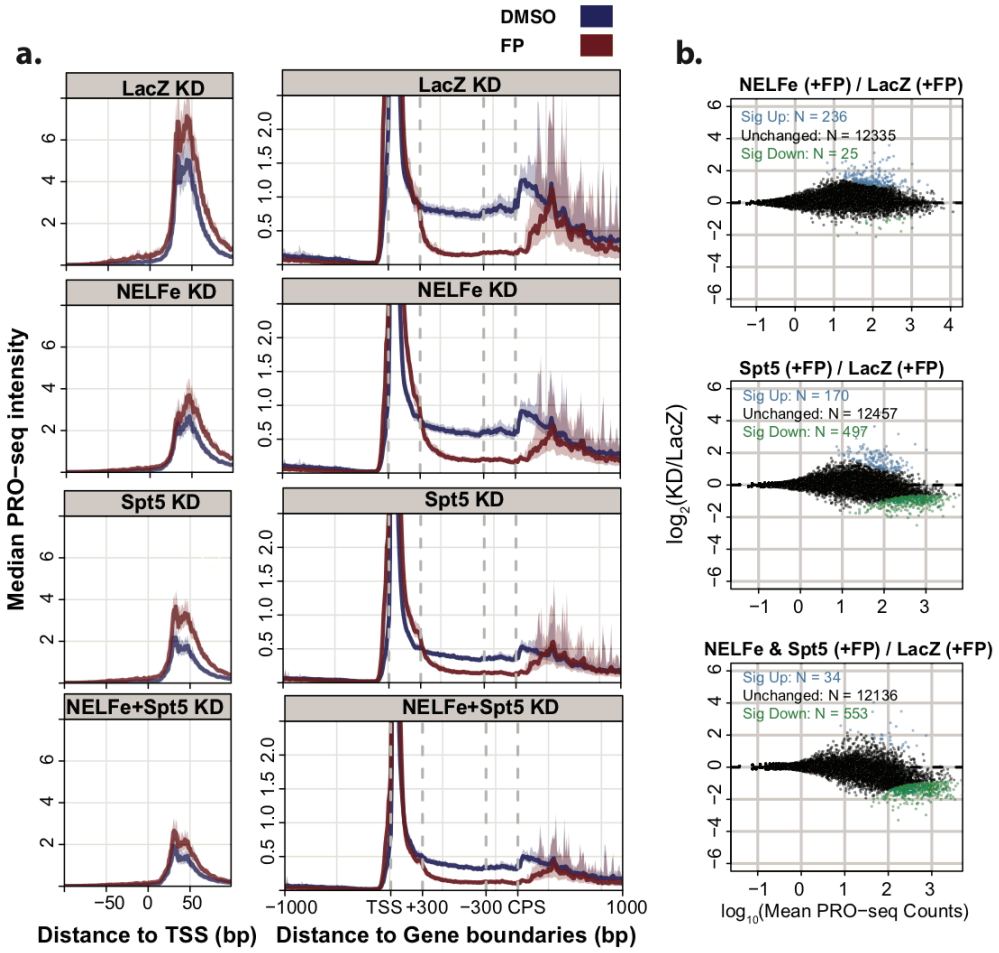

Figure 4.3| Cdk9 activity remains required for the escape of most Pol II in absence of pausing factors. **a.** Left panels display promoter-proximal composite PRO-seq profiles, while right panels display corresponding profiles covering the full length of all filtered genes (n = 3565). From top to bottom respectively, each pair of panels reflects data from LacZ RNAi, NELFe RNAi, Spt5 RNAi, and NELFe + Spt5 RNAi samples treated with either 500 nM flavopiridol (red) or an equivalent volume of DMSO (blue) for 10 minutes. Composite profiles in the left panels reflect the median signal from all filtered genes at individual bases around the TSS. Profiles on the right reflect the sum of reads within discrete windows. From -1000 to + 300 around the TSS and -300 to +1000 around the CPS, values reflect 10 bp windows. Intervening regions are composed of 30 windows, scaled based on the length of each gene. Center lines represent the median values at each position, while shaded areas represent bootstrapped 12.5% and 87.5% confidence intervals of the median signal. **b.** Panels show $\log_2$ fold change in normalized gene body read density after 10 minutes FP treatment for NELFe RNAi, Spt5 RNAi, or double knockdown of NELFe and Spt5, each over LacZ RNAi (10 min. FP treated) samples, from top to bottom respectively. DESeq2(M. Love et al., 2013) was used to determine the significance of increases (blue dots) or decreases (green dots) in Pol II density for each gene (adjusted $p < 0.01$).

NELF depletion increases frequency of premature release of Pol II from pause without P-TEFb activity. Unlike the subunits of the DSIF complex (Spt4 and Spt5), NELF is composed of ill-conserved subunits that are completely absent in many eukaryotes(Narita et al., 2003). Intriguingly, in species which lack NELF, Cdk9 is critical for efficient Pol II elongation, but not pause escape(Booth et al., 2018). Such observations support a model in which NELF functions to cooperate with conserved factors in the promoter-proximal region of genes to enable transcription regulation through P-TEFb. To investigate the principals of this model, we sought to further interrogate our results examining the impact of Cdk9 inhibition on S2 cells in which NELF had been depleted.

Looking more closely at Pol II profiles of individual genes, we identified instances in which depletion of NELFe dramatically altered the response to Cdk9 inhibition. In contrast to LacZ RNAi samples, after 10 minutes of P-TEFb inhibition in NELFe depleted cells, Pol II enrichment could be seen more than a kilobase downstream of the pause position (Figure 4.4a). Composite profiles of PRO-seq data from otherwise untreated cells recapitulated the subtle decrease in gene body elongation observed as a result of NELF knockdown (Figure 4.4b, top panel). In contrast, inhibiting Cdk9 for 10 minutes revealed a relative increase in elongating Pol II signal downstream of the pause as a result of NELFe knockdown, suggesting Pol II might be more prone to escaping the pause. Curiously, regardless of gene length, the NELF knockdown-dependent increase in escaped Pol II after Cdk9 inhibition was restricted to the first 3-4 kb of genes (Figure 4.4c).

We next investigated whether the early gene body increases in escaped Pol II constituted a statistically significant change over equivalently treated control knockdown cells. Filtered genes longer than 6.2 kb (n = 763) were partitioned into three discrete regions: promoter-proximal (TSS to +100 bp), early gene body (+200 to +3200 downstream of TSS) and late gene body (CPS-3000 to CPS) and the normalized PRO-seq signal within each region was quantified for separate biological replicates from all conditions (Figure 4.4d). Differential expression analysis performed separately on each region revealed that after Cdk9 inhibition, significant increases in Pol II density occurred in the absence of NELF, but only within the early gene body (sig. up = 254, sig. down = 2, n = 763,) (Figure 4.4, lower panels). In contrast, almost no genes exhibited changes in the late gene body, confirming the constraint of the escaped Pol II signal near the beginning of genes. While early gene body increases were only observed when pause escape was obstructed, promoter proximal regions tended to have significantly less Pol II signal in the absence of NELF, regardless of the Cdk9 activity. The regulation of transcription elongation by P-TEFb is likely enhanced by negative elongation factor, through its capacity to prevent Pol II from entering into productive elongation without modification by Ckd9.

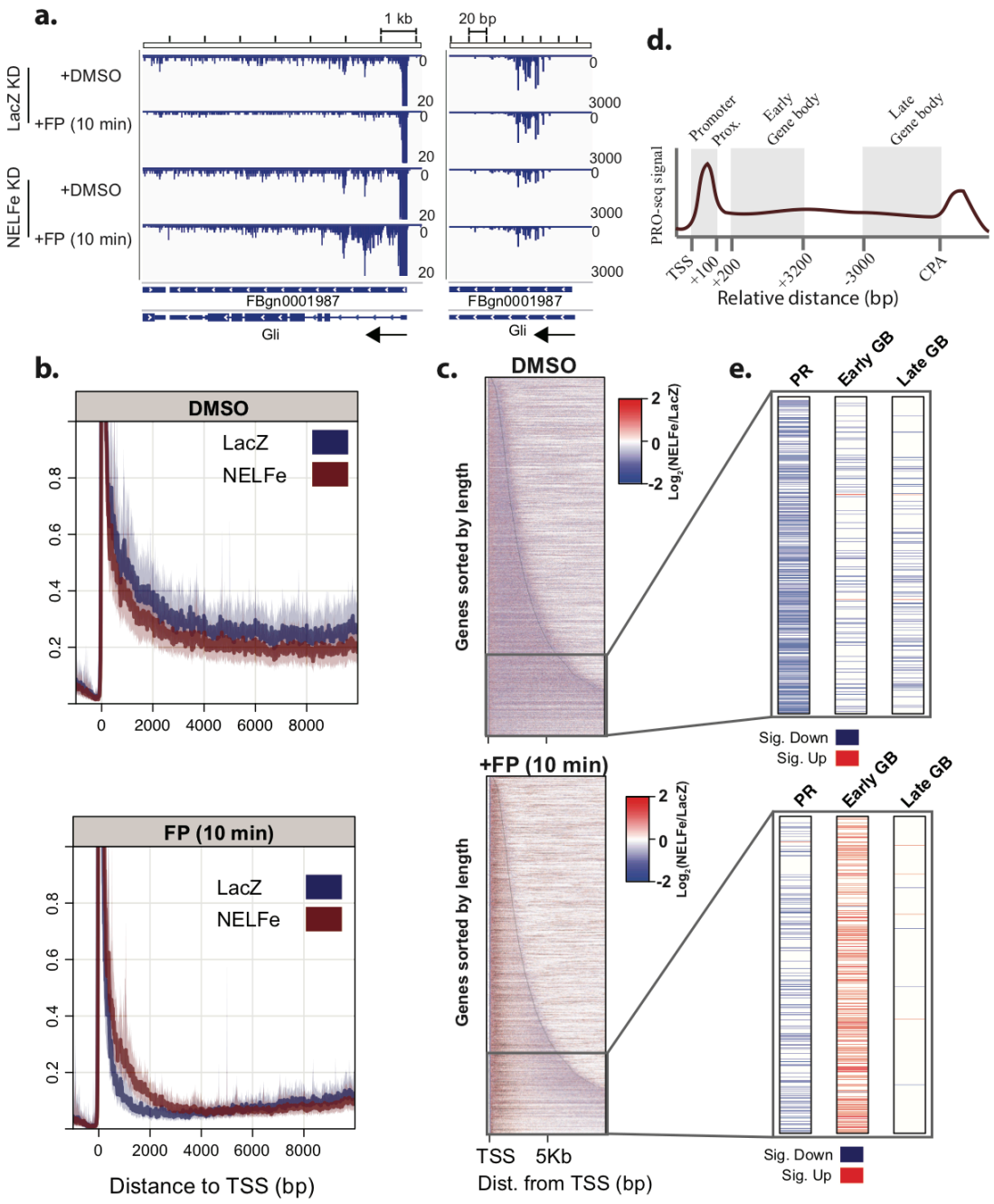

Figure 4.4| NELF depletion increases frequency of premature release of Pol II from pause. a. Browser tracks of normalized PRO-seq data from combined replicates (n=2) of LacZ RNAi (top two) and NELFe RNAi (bottom two) samples treated for 10 minutes with DMSO (upper) or FP (lower) at the *Gli* locus (only the minus strand is shown). The right panel displays a zoomed in view of the *Gli* promoter-proximal region with an expanded y-axis (0-3000) to fully capture the paused Pol II signal. **b.** Composite PRO-seq profiles around the observed TSS of filtered genes longer than 10 kb (n = 451). Each

plot displays LacZ RNAi (blue) and NELFe RNAi (red) profiles from combined replicates (n = 2) after DMSO (top panel) and FP (bottom panel) treated conditions. Position values reflect the median sum of counts within discrete 100 bp windows from -1kb to +10kb around the observed TSS. Center lines and confidence intervals were calculated as in previous figures. **c.** Heat maps display $\log_2$ fold change in PRO-seq signal (NELFe RNAi / LacZ RNAi) within 10 bp windows from -250 to +10,000 bp around the TSS for all filtered genes sorted by length (short to long from top to bottom). The top heat map reflects fold change (NELFe RNAi / LacZ RNAi) after control treatment (DMSO), while the bottom is the same comparison after 10 minutes of FP treatment. Dark lines give the positions of the observed TSS and CPS for each gene. **d.** Diagram of the partitioning of gene regions (promoter proximal, early gene body, late gene body) used in e to test for differences in Pol II density between treatments. **e.** Heat maps report whether a significant increase (red) or decrease (blue) in read density was observed within the promoter proximal, early gene body, or late gene body regions of each filtered gene from the subset of those used in c that longer than 6.2 kb (n = 763). DESeq2(M. Love et al., 2013) was used to determine whether a difference was significant (adjusted $p < 0.01$) given the same comparisons as made in c. The ordering of genes is the same as was used for c.

Release of Pol II from the pause without P-TEFb input leads to impaired transcription elongation and premature termination. Unless Pol II has already escaped the pause and entered the gene body, inhibition of Cdk9 in flies and mammals entirely prevents elongation by Pol II. Once beyond the pause, Pol II is seemingly unaffected by the same treatment, despite the fact that P-TEFb travels with the elongation complex(Ni et al., 2008; Jonkers et al., 2014). Clearly, the promoter-proximal modification of the elongation complex by Cdk9 is critical for productive transcription, but the significance of this input is poorly understood. The consequences of pause escape and Pol II elongation in the absence P-TEFb have not been studied *in vivo*, likely because it does not occur normally, or is challenging to induce. However, we found that depleting NELF had the apparent result of permitting the release of some Pol II from promoter proximal pausing in the absence of Cdk9 activity, providing an opportunity to explore the consequences of such aberrant release on Pol II transcription.

Inhibiting Cdk9 in S2 cells lacking NELF results in aberrant Pol II pause release and transcription that appears restricted within 3-4 kb downstream of the pause. However, several different fates of the leaked Pol II population could produce the observed profiles after 10 minutes. On one hand, elongation complexes which escape the pause in the absence of Cdk9 activity might be prone to premature or targeted termination within a short distance downstream. Alternatively, without NELF or Cdk9 activity, Pol II might be released with an impaired elongation rate, transcribing only 4 kb within 10 minutes. To discriminate between these possibilities, we performed a Cdk9 inhibition time course in cells with unaltered (LacZ RNAi) or severely depleted levels of NELFe (Figure 4.5a). If pause escape without Cdk9 input primarily impacts Pol II elongation rate, we surmised that, over the time-course of P-TEFb inhibition, we would observe an advancing density of Pol II signal across genes with time (Figure 4.5b). Inspection of PRO-seq data at individual loci after various time-points of FP treatment reaffirmed the increased density of Pol II downstream of the pause as a result of NELF depletion (Figure 4.5c). However, rather than advancing downstream with time after Cdk9 inhibition, NELF loss-dependent increases appeared restricted within the first few kb at all time-points (Figure 4.5c, left panel shaded area).

Consistent with our initial PRO-seq experiment following Pol II distributions 10 minutes after FP treatment, NELF depletion generally reduced promoter-proximal Pol II before and after Cdk9 inhibition when compared with LacZ RNAi (Figure 4.5d, left panels). As was observed at individual genes, loss of NELF globally resulted in increased levels of Pol II entering the gene body after 5, 10 and 20 minutes of FP treatment (Figure 4.5d, right panels). Surprisingly, the subtle global increases in

downstream transcribing Pol II resulting from NELF knock down were found to be significant at all FP time-points, but only within the early gene body as defined in Figure 4.4d (Figure 4.5e). Thus, even after 20 minutes, the released population of Pol II in NELF depleted cells was largely incapable of completing transcription of longer genes. These results suggest that modification of the elongation complex by P-TEFb during pausing is critical for processive transcription over long genes. Moreover, we report that the action of the NELF complex functions to ensure Pol II stably awaits P-TEFb signaling within the promoter-proximal region, likely to prevent wasteful rounds of transcription.

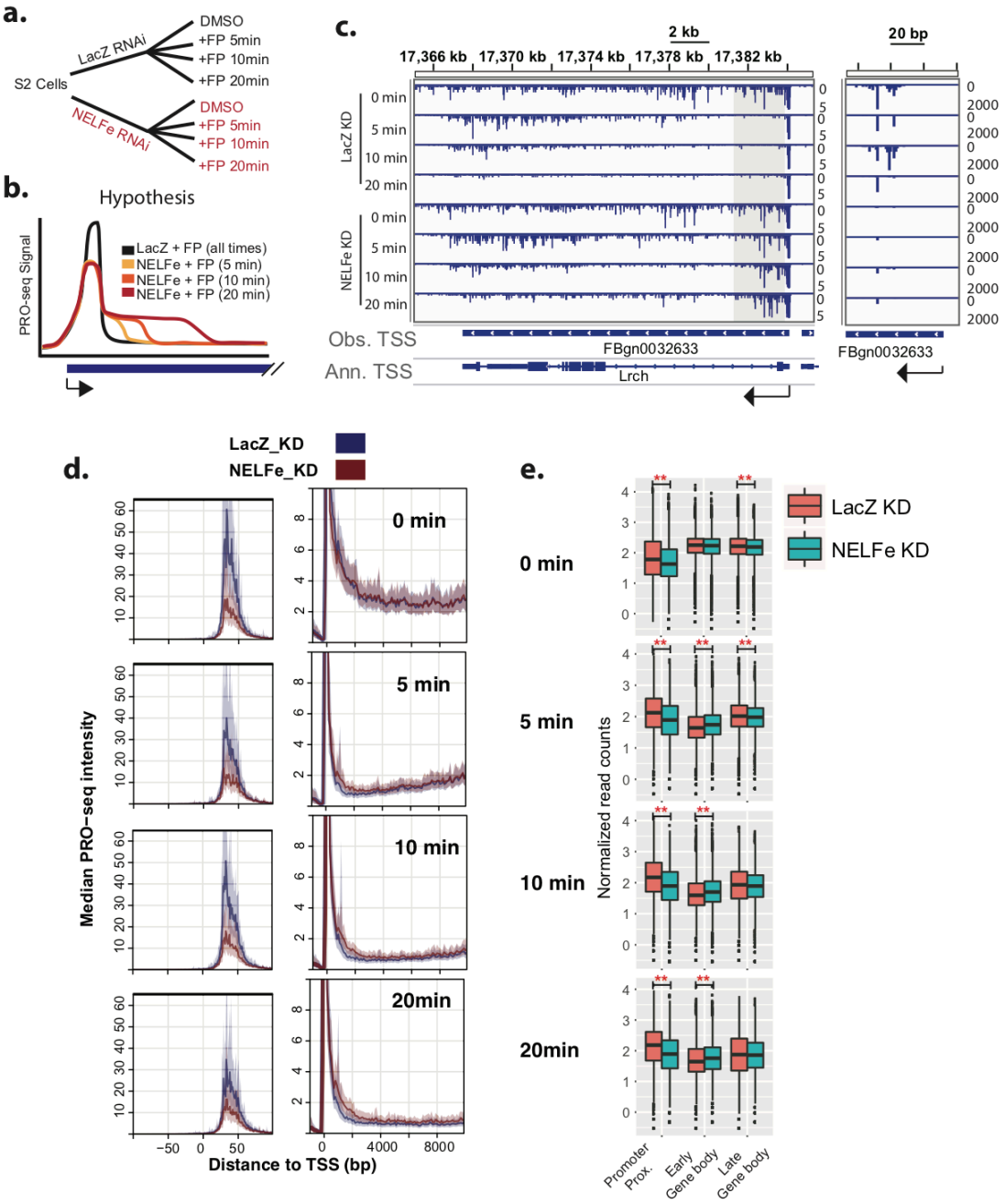

Figure 4.5| Release of Pol II from the pause without P-TEFb input leads to impaired transcription elongation and premature termination. a. Schematic of experimental conditions. **b.** Hypothetical results assuming release of slow elongating Pol II after NELFe RNAi. **c.** Browser tracks of FP timecourse in LacZ RNAi (top four tracks) and NELFe RNAi (lower 4 tracks). Tracks display PRO-seq data on the minus strand from combined replicates for each condition. Highlighted region in gray (left panel) emphasizes increased release of Pol II after FP treatment in NELFe RNAi samples. Right

panel is zoomed in on the promoter proximal region of the *Lrch* gene with an expanded y-axis (0-2000) to capture the pause signal. **d.** Composite PRO-seq profiles comparing LacZ (blue) and NELFe RNAi (red) samples at each time-point after FP treatment. Left panels display the median signal at individual bases immediately around the TSS, while right panels show signal within 100 bp windows from -1kb to +10kb around the TSS. Both panels reflect data from all filtered genes longer than 10 kb (n = 451). **e.** Box plots of counts within promoter proximal, early gene body, and late gene body regions of all genes longer than 6.2 kb (n = 763) for LacZ and NELFe RNAi cells at each time-point after FP treatment. Significant differences ($p < 0.001$) in the mean were determined using Student's t-Test between knockdowns for each gene region and are shown with red asterisks.

Discussion

Inhibition of P-TEFb globally shuts down transcription in metazoans by blocking the release of arrested Pol II captured by factors that assemble on the transcription complex shortly after initiation. But under normal conditions, many genes in metazoans do not appear rate-limited by promoter proximal pausing(Core et al., 2008; 2012). Moreover, many eukaryotes cannot regulate transcription through promoter-proximal pausing(Booth et al., 2018; Ptashne & Gann, 1997). While it is known that Cdk9 activity is critical for transcription elongation in all eukaryotes, how it became a ubiquitous mechanism for controlling promoter-proximal Pol II pausing is poorly understood. This work sought to explore how known systems of P-TEFb mediated gene regulation are established.

To dissect components involved in capturing elongating pol II downstream of the TSS we depleted two pausing factors, NELF and DSIF from S2 cells derived from *D. melanogaster*. Interestingly, although removal of NELF and DSIF drastically reduced levels of Pol II found within promoter-proximal positions, under none of our experimental conditions *in vivo* was pausing completely eliminated. Additionally, reducing the amount of Pol II found within the pause site did not lead to more downstream elongation, but instead a reduction in overall gene activity. This result is

consistent with a previous report that pausing can enhance gene expression(Gilchrist et al., 2008). The assembly of DSIF and NELF with RNA Pol II may act to stabilize various aspects of the paused complex(Narita et al., 2003; Yamaguchi et al., 1999). In their absence, rather than being released into productive elongation, Pol II might be prone to termination(Henriques et al., 2013). Alternatively, nucleosomes may be more likely to occupy promoters in the absence of frequent pausing of Pol II, leading to general repression of gene transcription(Gilchrist et al., 2010). Regarding the reduction in gene activity, the impact of depleting DSIF was far greater than the loss of NELF. However, DSIF is also known to act as a positive elongation factor that tracks with the transcription complex to the 3' ends of genes. Thus, depletion of DSIF likely also impacts Pol II elongation downstream of the pause. Indeed removing Spt5 in budding yeast, which do not regulate promoter-proximal Pol II, also globally reduces levels of transcription(Shetty et al., 2017).

NELF and DSIF are required for the sensitivity of Pol II to Cdk9 inhibitors *in vitro*(Marshall & Price, 1995; Wada et al., 1998; Yamaguchi et al., 1999). Yet, in cells depleted for either factor, we found that inhibition of Cdk9 with FP was capable of preventing the majority of Pol II from escaping into productive elongation. Thus, *in vivo*, additional factors may contribute to establishing a controlled release of Pol II through P-TEFb. For example, in *D. melanogaster*, the composition and structure of core promoter elements can strongly impact Pol II pausing, suggesting the pre-initiation complex (PIC) might restrain Pol II downstream of the TSS(H. Kwak et al., 2013). Notably, depleting NELF shifted average Pol II distributions within the pause farther downstream, both before and after FP treatment. More distal and dispersed Pol II enrichments have been

proposed to be partially enforced through contacts with first well positioned nucleosome within the gene body(H. Kwak & Lis, 2013; Mavrich et al., 2008) and the loss of NELF might promote such interactions. Importantly, trace amounts of NELFe and Spt5 were detectable after respective RNAi treatments, suggesting that the factors were not entirely eliminated from cells. Residual pausing and response to FP treatment in knockdown samples might thus represent only a partial phenotype due to small amounts of remaining functional protein. In future studies, it will be critical to explore alternative approaches for more complete and rapid removal of pausing factors.

Disentangling the influence of P-TEFb on transcription elongation from its role in Pol II pause escape has been nearly impossible in metazoans, because inhibiting Cdk9 blocks release of paused Pol II. In fission yeast, which lack NELF, inhibiting an analog sensitive Cdk9 does not prevent pause escape, but instead induces release of elongation rate-impaired promoter-proximal Pol II(Booth et al., 2018). Surprisingly, we found that by removing NELF from the *D. melanogaster* system *in* vivo, inhibition of Cdk9 resulted in Pol II "leaking" downstream of the pause. However, systematic differences between the resulting profiles exist. Unlike in fission yeast, Pol II released in the absence of Cdk9 was incapable of transcribing the full length of most genes in *D. melanogaster.* The restriction of leaked Pol II to a short range of motion is suggestive of a transcription complex prone to termination. We conclude that consistent with findings in *S. pombe*(Booth et al., 2018), Cdk9 modification during promoter-proximal pausing likely represents a checkpoint to ensure complete maturation of the elongation complex and limit wasteful downstream transcription. Ultimately, NELF may have evolutionarily reconfigured the promoter-proximal elongation complex to enable rate-limiting arrest of

Pol II, creating a mechanism for regulating gene expression via an otherwise conserved P-TEFb activity.

Future work will benefit from expanded exploration of the data presented here. Regarding the global responses to pausing factor perturbation, specific groups of genes could be examined separately depending on the characteristics of their promoters, and features of their regulation. Furthermore, non-coding distal regulatory elements, such as enhancers are also transcribed(T.-K. Kim et al., 2010b). Similar analyses should be extended to unstable transcript-producing loci to investigate the role of pausing within non-coding regions of the genome. Finally, similar pursuits in a mouse or human system could have potential medical implications by providing insight into unique properties of transcription regulation in mammals, such as long-range transcription regulation and divergent antisense transcription events. Finally, NELF is essential for proper development in mammals, possibly enhancing the ability of cells to rapidly rewire transcription programs through pausing(Williams et al., 2015). Understanding how NELF-mediated pausing influences genome response to environmental stimuli will enhance interpretation and engineering of gene circuitry.

Methods

Cell culture and RNAi knock down

Drosophila melanogaster derived S2 cells were grown in M3 + BPYE medium with 10% FBS to a cell density of ~ $5x10^6$. Cells were then diluted to a density of $1x10^6$ in M3 + BPYE media lacking FBS and 15 mL was split into separate T150 flasks for each RNAi treatment. The appropriate double stranded RNA was then added to flasks at a

concentration of 10 μg/mL. After incubating cultures with dsRNA at 26 °C for 45 minutes, an equal volume of M3 + BPYE medium with 20% FBS was then added, bringing the final FBS concentration to 10%. Cultures were then left at 26 °C for two days, then split into two new T150 flasks and a second dose of appropriate dsRNA was added directly to cultures before returning to 26 °C for two additional days. All double stranded RNAs used in this work were transcribed from dsDNA templates. DNA templates contained T7 polymerase promoters at both ends and were prepared using primers the following primers:

LacZ Forward	GAATTAATACGACTCACTATAGGGAGAGATATCCTGCTGATGAAGC
LacZ Reverse	GAATTAATACGACTCACTATAGGGAGAGCAGGAGCTCGTTATCGC
NELFe Forward	GAATTAATACGACTCACTATAGGGAAGGCACTGCAAGCGCACAAGGCGCCC
NELFe Reverse	GAATTAATACGACTCACTATAGGGACTTCATCGTATTGAACCATCTCGCGG
Spt5 Forward	GAATTAATACGACTCACTATAGGGAGAGGATACATTTACCTGGAGGCC
Spt5 Reverse	GAATTAATACGACTCACTATAGGGAGAGTTATATCGGAGACGAGCACC

All experiments were performed in biological replicate. A culture for the second replicate was set-aside during the initial seeding event for the first RNAi treatment and submitted to an additional untreated growth period and passage. This passage offsetting approach for replicates was utilized in order to capture biological and technical variation by performing all replicated knockdowns and drug treatments on different days with cells of different passage number.

Flavopiridol treatments

After four days of exposure to dsRNA, control and Flavopiridol (FP) treatments were prepared as follows: The two flasks of cultures from each knock down were pooled and the cell density was quantified in order to normalize the number of cells used in treatments across all samples. 2 mL of each pooled cell culture was set aside for assessing knock down efficiency via western blot. Pooled knockdown cultures were then split into two tubes for treatment with 500 nM FP or an equivalent volume of DMSO. Treatments were performed for 5, 10 or 20 minutes at 26 °C and abruptly stopped by placement on ice, centrifugation at 1000 x G (4 °C) for 5 min. followed by resuspension in 10 mL ice cold 1xPBS. At this stage, a small amount of thawed mouse embryonic fibroblast cells in 1xPBS were spiked into each sample (~ $4x10^4$ MEFs = 0.04% of all cells). Spike-in cells were added to serve as an exogenous control for sample normalization in the case of global changes in transcription.

Western blots for assessing knockdown efficiency

2 mL of knockdown cultures were spun down in the presence of 1% (v/v) PMSF at 1000xG. After removing media, cell pellets from each sample were resuspended in 1xSDS loading buffer such that there were $1x10^5$ cells/μL. Cells were then boiled at 95°C for 5 minutes. SDS PAGE gels and transfer reactions were performed under standard conditions and dilutions of the LacZ RNAi cell lysate were used to assess levels of protein depletion in KD samples. Laboratory stocks of rabbit anti-dNELFe (donated by David Gilmour's lab), rabbit anti-dBEAF, guinea pig anti-dChromator, and guinea pig anti-dSpt5 were used at 1:1000, 1:2000, 1:2000, 1:1000. 1 mg/mL IRDye 800CW donkey

anti-rabbit and 1 mg/mL IRDye 680LT donkey anti-guinea pig were used as the secondary antibodies and incubated with membranes at a 1:15,000 dilution. Western results were visualized using LI-COR Odyssey imaging system.

Precision run-on library preparation

After resuspending treated cultures with added spike-in cells in 1xPBS, cells were subjected to permeabilization (as opposed to nuclei isolation). Samples were washed in 5 mL wash buffer (10 mM Tris-Cl, pH 7.5; 10 mM KCl; 150 mM sucrose; 5 mM $MgCl_2$; 0.5 mM $CaCl_2$; 0.5 mM DTT; 1x Protease inhibitor cocktail (Roche); 20 units RNase inhibitor (SUPERase In, Invitrogen)), then resuspended in permeabilization buffer (10 mM Tris-Cl, pH 7.5; 10 mM KCl; 250 mM sucrose; 5 mM $MgCl_2$; 1 EGTA; 0.05% Tween-20; 0.1% NP40; 0.5 mM DTT; 1x Protease inhibitor cocktail (Roche); 20 units RNase inhibitor (SUPERase In, Invitrogen)) and left on ice for 5 minutes. Cell permeability was assessed with trypan blue staining (~99% permeable). Cells were washed twice in 5 mL wash buffer, then resuspended in 220 µL storage buffer (50mM Tris-Cl, pH 8.3; 40% glycerol; 5 mM $MgCl_2$; 0.1 mM EDTA; 0.5 mM DTT) and immediately frozen in liquid nitrogen and stored at -80 °C. 100 µL of stored, permeabilized cells were used for precision run-on reactions. PRO-seq libraries were prepared as described previously(Mahat, Kwak, Booth, Jonkers, Danko, Patel, et al., 2016a), however a modified 3' RNA adaptor was used for the purpose of molecular barcoding. The adaptor sequence used (5'-/5Phos/NNNNNNGAUCGUCGGACUGUAGAACUCUGAAC/Inverted dT/-3') contains a stretch of 6 random nucleotides at the 5' end to permit the elimination of PCR

duplicates during read processing. For time-course experiments a different 3' adaptor was used during PRO-seq library preparations. Instead of adaptors with a UMI, time-course libraries received 3' adaptors with in line barcodes (5' -/5Phos/ GNNNNNNGAUCGUCGGACUGUAGAACUCUGAAC-/Inverted dT/). The ligation of this adaptor with distinct barcodes to each library permitted the pooling of all libraries after the 3' end ligation step, facilitating handling. After pooling, the remainder of the library preparation was unchanged, but could be performed in a single tube.

PRO-cap libraries were prepared similarly to PRO-seq libraries, with several critical modifications for the selective enrichment of initiated nascent RNA 5' caps. No base hydrolysis of RNA was performed after run-on reactions. Rather than molecular barcoded adaptors, standard 3' RNA adaptors were used(Mahat, Kwak, Booth, Jonkers, Danko, Patel, et al., 2016a). Importantly, prior to enzymatic removal of the nascent RNA 5' cap, the PRO-cap sample was split into two tubes. One tube received enzymatic treatment with Tobacco acid pyrophosphatase (TAP) to remove the 5' cap structure, while the other received only reaction buffer, lacking TAP, and thus serves as a measurement of remaining background (uncapped) signal(Kruesi et al., 2013). The remainder of the library preparation was performed identically for control and experimental PRO-cap samples. Sequencing was performed using the Illumina NextSeq500 platform.

Alignment and processing

The alignment and processing of sequenced reads in this experiment were conducted largely as previously described (Booth et al., 2016; 2018). Here, since we incorporated

molecular barcodes within 3' adaptors used in PRO-seq, after removing any instances of 5' adaptor sequence (3' adaptor sequence for PRO-cap) composition, we collapsed multiple reads where the entire sequence was identical (a result of PCR duplication) using the prinseq_lite program (sourceforge.net/projects/prinseq). PRO-cap read processing did not require the duplicate removal step. After collapsing reads, the 6 nt barcodes were removed (PRO-seq only) and reads were trimmed to a maximum of 36 nt (min = 15 nt). To acquire the correct orientation of RNA origin, the reverse complement of reads were generated prior to alignment using the fastx_reverse_complement function from the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The reverse complement of PRO-cap reads was not used, as they are sequenced from 5' ends. Reads were aligned to a combined genome of *Mus musculus* (mm10) and *Drosophila melanogaster* (dm6) using Bowtie 2(Langmead & Salzberg, 2012) (version 2.2.6). Only reads that aligned uniquely were used in downstream analysis, eliminating unmappable reads and reads of ambiguous species origin. Downstream processing was performed as described previously(Booth et al., 2016; 2018).

Time-course samples did not receive a unique molecular index during library preparation and thus PCR duplicates could not be removed. For time-course samples, the first 7 nt of each read was used to separate individual samples from pooled sequencing results before being trimmed from the remaining read. Apart from this trimming step and the inability to collapse duplicate reads, remaining alignment and processing for time-course samples was carried out as described for other samples. Importantly, because of the described differences in handling of PCR duplicated reads between these batches of

experiments, we refrained from any direct comparison between the batches of PRO-seq experiments.

Normalization and experimental batches

The number of reads from each sample that uniquely aligned to the mouse genome (mm10) was divided by 100,000 and used as a spike-in normalization factor for that sample. Scatter plots of replicates for each experiment show good agreement (Figure 4.2) suggesting that the PRO-seq experiments are reproducible. Centering of scatter around the diagonal line ($x = y$) further implies that the exogenous spike-in-based normalization accurately accounts for differences in sequencing depth between samples. Notably, for the double knockdown (Spt5 RNAi + NELFe RNAi) cultures, both treated (FP) and untreated (DMSO) show scatter that is not well centered about the diagonal, implying poor normalization based on spike-in read counts. We determined that the skewedness is likely a result of replicate 2 for both treatments, as we were forced to use ~ half the number of cells for these samples due to contamination of one of the two flasks from this knockdown. Thus, for the double knockdown, rather than combining replicate data, only replicate one was used when comparing with other samples. For all other samples, combined replicates were used when useful.

Strictly for time-course experiments, after normalizing counts based on spike-in RNA, replicate comparisons failed to show adequate symmetry around the line $x = y$, indicating an inability to compare between samples and treatments with this approach. We surmised that distal genes regions of genes (> 60kb from the TSS) would remain un-affected after FP treatment after 20 minutes, given an average rate of Pol II of 2.5

kb/min(Jonkers et al., 2014). Indeed, our time course data at extremely long genes revealed a clearing wave that had yet to arrive 60 kb downstream within 20 minutes of FP treatment (Figure 4.6a). To test whether these regions would allow us to normalize between LacZ RNAi and NELFe RNAi samples, we used the original data set to assess the knockdown produced significant changes in Pol II density within the gene body regions (+60 kb from the TSS to the CPS) of all filtered genes longer than 65 kb (n = 27). After 10 minutes of FP treatment spike-in normalized counts from the original experiments showed no difference between LacZ and NELFe knockdown samples within these regions (Figure 4.6b). Thus, for each time-course sample we counted all reads within this distal gene body region of 27 genes and used this value to account for differences in sequencing depth between samples. Applying this normalization factor to individual samples produced good correlations between biological replicates that is symmetrical about the diagonal (Figure4.6c). Therefore, the "long-gene-end" normalization factor was used for all analyses of time-course experiments.

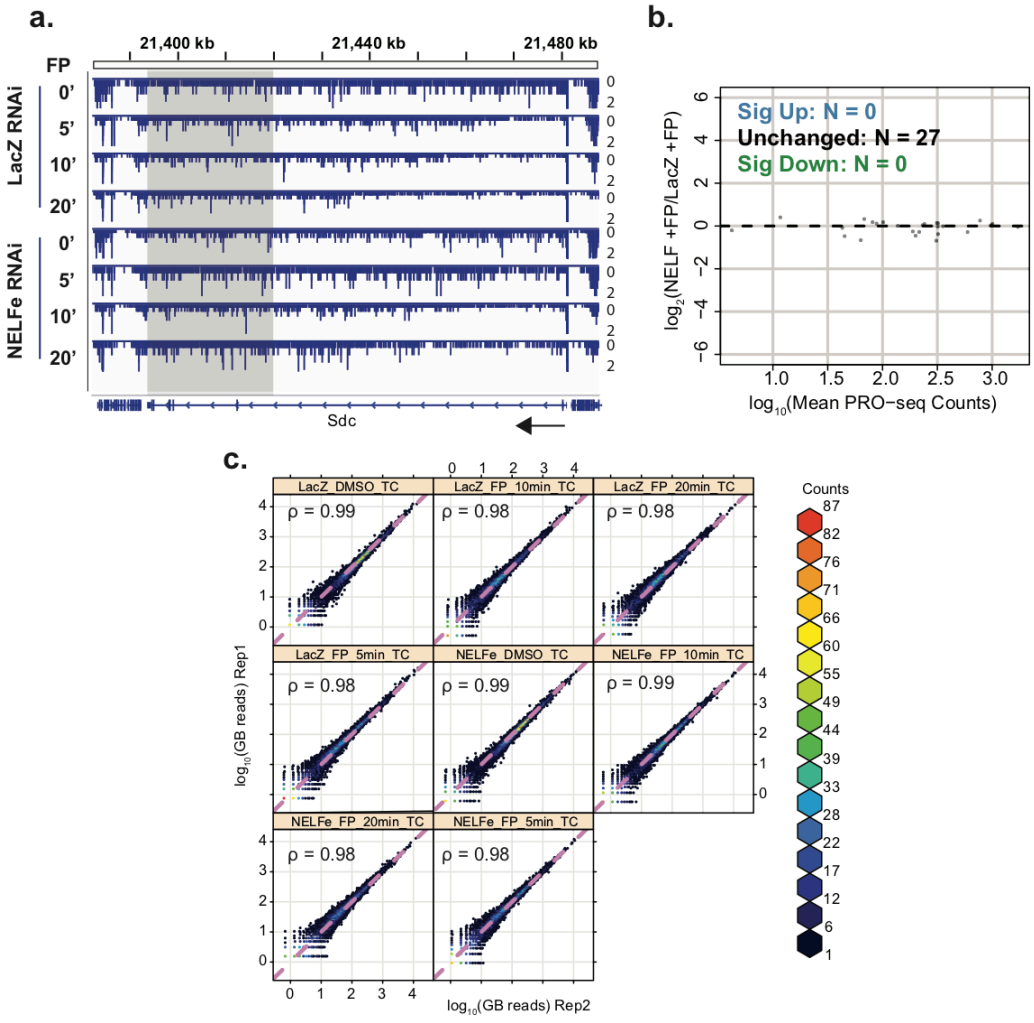

Figure 4.6| Alternative normalization approach for time course experiment data. a. Browser image of PRO-seq data from two combined replicates of either LacZ or NELFe RNAi cultures treated with 500 nM FP for the indicated amount of time. Shaded region highlights the apparently unaffected late gene body region of the *Sdc* loci. **b.** DESeq2(M. Love et al., 2013)-based differential expression analysis of spike-in normalized counts within the "long-gene-end" regions (TSS + 60kb to CPS) of all filtered genes longer than 65 kb (n = 27). Analysis compares regions from the NELFe RNAi culture treated with FP for 10 minutes with the LacZ RNAi culture treated with FP for 10 minutes (samples are not from time-course experiment). **c.** Scatterplots display the comparison of "long-gene-end"-normalized read counts within the gene body (TSS+200 to CPS) of all genes (n = 16164) from biological replicates of the same sample from time course experiment.

Differential expression analysis

All statistical tests for significant differences in read counts within genomic regions between samples were performed using the DESeq2 R package(M. Love et al., 2013). Rather than using default size factors to normalize data between samples, we supplied spike-in-based normalization factors (spike-in mapped reads / 10^5). Where comparisons between Spt5+NELFe RNAi samples and other knockdown samples, we simply omitted replicate 2 of the Spt5+NELFe RNAi sample, due to issues described above. However, when comparisons were made within the Spt5+NELFe RNAi samples (i.e. FP vs DMSO treatments), we were able to correct the normalization factors for the second biological replicate, by determining an "expected spike-in count" for each sample based on the relationship between spike-in counts and total library size for samples from the first replicate. Thus, all comparisons between treatments within a given knockdown take into account two replicates.

Transcript isoform and TSS selection

Selection of the preferred TSS using PRO-cap data was performed by identifying the base with the highest background-subtracted PRO-cap signal within the window -250 to +250 relative to the annotated TSS. If two bases within this region produced equal signal, the site closer to the annotation was used. Observed TSSs were additionally required to have at least 5 more reads per million mapped than background at the same position and if this criteria were not met, the annotated TSS was used. This analysis was performed for all annotated transcript isoforms (n = 52086). For genes with multiple transcript isoforms, the isoform with the highest observed TSS signal was selected and in

cases of a tie, the longest isoform was chosen. This process of isoform selection such that each gene is reported only once, resulted in a total of 17454 transcription units. Finally, transcription units were only considered if we a PRO-cap identified observed TSS was reported (n = 6970).

Gene filters and activity:

Active genes were identified by assessing whether a gene had greater PRO-seq read density than could be achieved using background signal. As described in Core et. al. (2012) Active genes were required to have a probability ($p < 0.01$) of observing the amount of reads within the gene body (+200 to CPS) based on a poison distribution with background rate defined by placing 1% of reads randomly across all mappable regions of the genome ($\lambda = (0.01*N)/ M$, where N is the number of mapped reads, and M is the number of mappable 36-mers in the dm6 genome). By these criteria, of the 6970 transcription units with a PRO-cap observed TSS, 6436 (92%) were called active. To eliminate the possible interference of upstream run through transcription on the interpretation of polymerase density within a transcription unit, we applied a filter to assess the amount of upstream versus downstream transcription around each promoter. We only considered genes with at least fourfold the amount of PRO-seq signal downstream (+300 to +500) compared with an upstream region (-300 to -100 bp), relative to the observed TSS. Using the above two filters we obtained 3969 active genes without upstream run-through signal. Finally, to prevent the possibility of downstream transcription start sites interfering with our analysis of gene body transcription, PRO-cap data was analyzed downstream of the observed TSS of the filtered genes. Genes were

used in our analyses if the selected isoform lacked any downstream 50 bp window exceeding 50 PRO-cap reads and there for unlikely to have interfering transcription initiation downstream of the called TSS.

CHAPTER 5

OUTLOOK

The regulation of gene expression through promoter-proximal pausing of engaged RNA Polymerase II is not functionally conserved across eukaryotes. Yet, much of the machinery necessary for this step in the transcription cycle is conserved from yeast to humans. Through a series of three discrete research projects presented in this thesis, I set out to explore the functional origins of this sophisticated mechanism of gene regulation.

The first project investigated transcription profiles in two highly divergent unicellular eukaryotes. Although steady state profiles of transcription indicated budding yeast, *Saccharomyces cerevisiae* and fission yeast, *Schizosaccharomyces pombe* differ greatly with respect to the kinetics of Pol II elongation along transcription units, common roles of the Spt4/Spt5 complex in Pol II regulation were gleaned from our data. In both species, we observed that deletion of the small Spt4 subunit of this complex significantly reduced the rate of Pol II elongation, suggesting it contributes to the overall ability of Pol II to translocate along DNA. Perhaps not surprisingly, this complex is conserved in archaea, even sharing homology with the NusG elongation factor in bacteria, indicating RNA polymerases have long relied on accessory factors to enhance RNA synthesis(Hartzog & Fu, 2013; Werner, 2012). However, in addition to similarities, disparate impacts of the deletion on Pol II within promoter-proximal regions of genes indicated possible divergent functions for Spt4/Spt5 between the yeast. In metazoans, the DSIF complex composed of Spt4 and Spt5 performs dual functions as a pausing and downstream elongation factor during Pol II transcription (Wada et al., 1998). Unlike in budding yeast our results suggested that fission yeast might similarly use the Spt4-spt5

complex to restrain Pol II near the transcription start sites. But whether *S. pombe* are capable of regulating Pol II transcription through this process was still unclear.

Pol II pausing in metazoans is controlled by P-TEFb, which releases the complex into productive elongation through kinase-mediated phosphorylation of Pol II, NELF and DSIF(Peterlin & Price, 2006). By contrast, yeast lack the components of negative elongation factor (NELF) involved in establishing the paused Pol II complex(Narita et al., 2003). To further understand the relationship between systems of transcription in fission yeast and metazoans, the second project sought to examine the role of the conserved P-TEFb kinase, Cdk9 in governing transcription in *S. pombe*. Using a chemical-genetic approach to inactivate enzymes, individual transcription-coupled kinases were rapidly inhibited in analog sensitive fission yeast strains. Spt5 was found to be the substrate most affected by loss of Cdk9 activity in *S. pombe*, with undetectable phosphorylation after just two minutes of inhibition. However, unlike in metazoan systems, which tightly regulate the release of paused Pol II, inhibiting Cdk9 in fission yeast did not prevent downstream elongation. Instead, promoter-proximal Pol II was severely impaired in its ability to traverse transcription units at normal speeds. Perhaps the evolution of NELF enabled a novel mechanism of gene regulation by exploiting this requisite modification of the elongation complex by Cdk9.

The final project described in this thesis endeavored to test how NELF might contribute to gene regulation through promoter-proximal pausing. Critical subunits of both DSIF and NELF pausing factors were depleted from cells derived from *Drosophila melanogaster*, which heavily utilizes promoter-proximal pausing to regulate transcription(Core et al., 2012; H. Kwak et al., 2013; Ni et al., 2008). Reduced pausing

coincided with diminished transcription, suggesting this impediment to elongation does more than regulate transcription output. As we found in fission yeast, signaling to Pol II though P-TEFb could act as a checkpoint to ensure maturation of transcription complex before pause escape. However, under normal conditions, Pol II will not escape into the gene body unless licensed by P-TEFb. By depleting NELF from cells, a significant amount of paused Pol II appeared to become resistant to Cdk9 inhibition, permitting Pol II release from arrest without modification. Indeed, when arrested Pol II is released in the absence of Cdk9 activity, it is incapable of elongating beyond 3-4 kb downstream of the pause. These findings indicate that the metazoan-specific mechanisms of gene regulation through promoter-proximal Pol II pausing may be derived from an ancestral checkpoint, which ensures functional maturity of the transcription complex. I propose that NELF may have evolutionarily exploited this checkpoint architecture, creating a novel means to regulate gene expression. Whether this added layer of gene regulation precipitated the evolution of complex and adaptable gene expression programs necessary for development is an attractive idea, but remains largely speculative. In the following section, a list of outstanding ideas, questions and experiments are presented for the continuation of inquiry into the origins and mechanisms of gene regulation.

Future directions

Mechanism by which Cdk9 substrate phosphorylation affects transcription. The findings reported in chapter 3 of this thesis strongly implicate Spt5 in the observed transcriptional defect as a result of Cdk9as inhibition in fission yeast. Appendix 1 describes the analysis of Cdk9 inhibition time courses in strains in which the known target residue, Thr1 within each Spt5 c-terminal repeat, was mutated to either an alanine

(T1A) or a glutamate (T1E). Surprisingly, neither the constitutively unphosphorylated (T1A), nor the phosphomimetic (T1E) mutants showed difference in transcriptional response to Cdk9 inhibition when compared with a strain possessing the wild type Spt5 repeats. While these results indicate that the Thr1 residue within Spt5 is disposable for Cdk9 mediated maturation of the elongating Pol II, the importance of the Spt5 CTR cannot yet be ruled out. To test the requirement of Spt5 in modulating elongation response to Cdk9 activity, it will be critical to perform the Cdk9 inhibition time course in a strain completely lacking the c-terminal repeat domain. The proposed experiment would eliminate the distinct possibility of nearby residues within the repeat ($T_1P_2A_3W_4N_5S_6G_7S_8K_9$) from receiving compensatory phosphorylation by Cdk9 in the absence of Thr1. Nonetheless, the unchanging results from strains with point mutations in Thr1 suggest that another target of Cdk9 may be responsible for the effects on transcription.

Ser2 residues within the Pol II CTD are believed to be a highly conserved substrate of Cdk9 in fission yeast as well as mammals(Phatnani & Greenleaf, 2006; Schwer & Shuman, 2011). Given the importance of this mark in coordinating co-transcriptional processes, it may very well be accountable for the observed phenotype. Interestingly, western analysis indicated the total Ser2 phosphorylation levels were not affected by *Cdk9*^{as} inhibition within the tested timeframe. However, it is possible that while lsk1/Cdk12 performs the bulk of Ser2 phosphorylation in fission yeast, Cdk9 might be responsible for the earliest appearance of this mark, and thus critical for the early elongation checkpoint observed. This model could be tested by monitoring changes in the occupancy of Ser2-phosphorylated Pol II upon Cdk9 inhibition. While resolving the

targets of Cdk9 which enable its impact on early transcription elongation is essential understanding this regulation, these events likely direct changes in Pol II transcription by altering the association and recruitment of additional factors. Future investigations should exploit our understanding of the critical Cdk9 substrates to identify candidate elongation factors, which directly manipulate the properties of Pol II transcription within chromatin.

Conserved interplay of multiple kinases to regulate transcription in fission yeast and metazoans. Transcription-coupled kinases often work in such close spatial and temporal proximity and are capable feeding back on one another. Combinatorial inhibition of various kinases targeting the elongation complex could thus yield novel mechanistic insights. For example, in fission yeast Cdk7 activity has been shown to facilitate recruitment of Cdk9 to Pol II to coordinate RNA capping and transcription elongation(Viladevall et al., 2009). In the third chapter of this thesis, using analog sensitive fission yeast strains for multiple different kinases, effects of inhibition for 5 minutes of individual and combinations of enzymes were briefly discussed. Similar to experiments performed in the *Cdk9*^{as}, time course experiments following the inhibition of Cdk7 alone or Cdk7 and Cdk9 in combination could help resolve disparate mechanistic inputs imparted by each kinase. Although after 5 minutes of inhibition of *Cdk7*^{as}, Pol II profiles in *S. pombe* reflected similar changes to that of *Cdk9*^{as} inhibition, additional time-points are needed to determine whether the phenotype shares the same dynamic properties. Indeed, irreversibly inhibiting Cdk7 in budding yeast for two hours, results in seemingly stagnant 5' enriched Pol II(Rodríguez-Molina et al., 2016).

The overarching objective of this thesis is to advance the understanding of the functional evolution and conservation of transcription regulation. Contrasting the in depth mechanistic study of transcription-coupled kinases in fission yeast with equivalent investigations in mammals would enable functional comparisons between divergent systems relating to regulation of promoter-proximal pausing. While extensive analysis of the effects of Cdk9 inhibition in various metazoan cell lines has been performed(Chao & Price, 2001; Gressel et al., 2017; Jonkers et al., 2014; Ni et al., 2008), less work has focused on the TFIIH kinase, Cdk7. Inhibitors, such as THZ1 have been effective in generally targeting wild type Cdk7 to study effects on transcription(Kwiatkowski et al., 2014), however a recent analog sensitive Cdk7 has been introduced into the human cell line HCT116, potentially allowing even tighter experimental control over its activity in the human system(Ebmeier et al., 2017; Larochelle et al., 2012). Using this cell line would enable the exploration of the role of Cdk7 in regulating general properties of Pol II elongation. Considering the wealth of literature linking this enzyme to nascent RNA capping, fine scale time courses of *Cdk7*^{as} inhibition, individually or in combination with Cdk9 inhibition, could be illuminating with regard to how capping is related to regulation of Pol II elongation in mammals. Strategies which distinguish nascent RNA produced by Pol II phopho-isoforms could be employed to investigate relationships between RNA processing and the status of Pol II. Moreover, the recent development of the coordinated PRO-seq (coPRO-seq) method enables information of both the precise position of RNA polymerase and transcription start site for every nascent RNA (unpublished). Ultimately, coupling the perturbation of co-transcriptional RNA processing and Pol II elongation

through Cdk7 and Cdk9 inhibition in mammals with such rich information on the status of Pol II and RNA would help resolve the interplay between capping and pausing.

Pausing factor influence on P-TEFb-mediated regulation. The work presented in this thesis expanded results from fission yeast, which lack P-TEFb regulated pause escape, to *D. melanogaster*, where pausing is pervasive(Ni et al., 2008). Further extending many of these experiments to a mammalian system would elaborate on the conserved roles of these pausing factors and elongation control, particularly at upstream divergent initiation sites near promoters and enhancers. Promoters and enhancers in mammals share a highly related chromatin architecture(Core et al., 2014). Moreover, divergent transcription events, including those originating from distal, intergenic regulatory regions, have also been reported to undergo promoter-proximal pausing(Henriques et al., 2018). Yet, the purpose of such elaborate regulation at non-coding sites is unknown. Manipulating the availability of pausing factors in human cells would enable direct comparison of the importance of such regulatory mechanisms for the production of protein coding RNA versus RNA which is immediately destroyed.

As mentioned in the outlook section, being able to distinguish primary or direct effects of a perturbation from downstream consequences is critical for interpreting phenotypes. One limitation of the use of RNA interference in chapter 4 to eliminate pausing factors from a cell is the delay between induction of gene knockdown and the loss of protein. With RNAi, this limitation arises due to the targeting of mRNA for degradation rather than protein directly. Protein translation is halted once all mRNA has been destroyed, but effects of the knockdown will not manifest until the pre-existing

proteins have naturally turned over. For this reason, depending on the stability of the protein, RNAi-mediated knockdown experiments often require several days to observe a phenotype. Thus, having more immediate temporal control over depletion of pausing factors would improve the detection of direct roles of factors like NELF and DSIF. One such approach is the auxin-inducible degron (AID) system. To degrade a protein of interest using the AID system, the endogenous protein must be fused with a small AID tag within a transformed cell line expressing the plant specific F-Box transport inhibitor response 1 (TIR1) protein. Simply by adding the small molecule plant hormone auxin, TIR1 and an endogenous ubiquitin ligase will target for degradation all proteins bearing the AID tag. Near complete degradation of target proteins can occur within 30 minutes of auxin treatment and can be reversed by the removal of the plant hormone(Nishimura, Fukagawa, Takisawa, Kakimoto, & Kanemaki, 2009). Using the AID system for depleting NELF and DSIF subunits would offer a significant improvement for the analysis of their roles in the regulation of promoter proximal pausing by P-TEFb. In contrast to waiting four days for protein loss, in which time secondary effects of the knockdown might accumulate, experiments inhibiting P-TEFb could be performed within half an hour of depleting each factor using the AID system.

As in the case of the cellular heat shock response, rapid, genome-wide adjustments in gene expression can be controlled via promoter-proximal pause escape(Duarte et al., 2016; Mahat, Salamanca, Duarte, Danko, & Lis, 2016b). The experiments presented in chapter 4 expand upon previous literature suggesting that the pausing factor NELF largely refines promoter-proximal Pol II pausing by enhancing the stability of Pol II and preventing its premature release into the gene body. Further

exploring the physiological requirement for NELF in executing a global transcriptional response might help to explain the emergence of NELF and gene regulation via Pol II pausing. NELF depletion resulted in a subtle but significant reduction in the ability of P-TEFb to regulate pause escape, as described in chapter 4. However, this refinement of paused Pol II could have a systematic importance in a cells ability to respond to environmental assaults. Such a physiological impact could be quantified by exposing cells, either possessing or depleted of NELF, to heat shock or other stressors and then comparing transcriptional responses.

Evolutionary coupling of NELF and P-TEFb-regulated pausing A hypothesis driving much of the work in this thesis is that NELF has enabled transcription regulation through promoter-proximal pausing. Thus far, the approach for studying the importance of NELF in transcription elongation and regulatory checkpoints has been either to interrogate species that lack orthologs for the constitutive subunits or, in species that have NELF, to remove them and observe the consequences. Two additional approaches remain untested in the investigation of NELF as sufficient for P-TEFb-regulated pause escape. First, simple sequence homology searches for eukaryotes possessing NELF subunit-like proteins has revealed unprecedented hits within simple eukaryotes, like *Dictyostelium discoidiem*, a fungal evolutionary outgroup to all animals (Charles Danko; Unpublished)(Vos et al., 2016). As P-TEFb-regulated pausing has not been demonstrated outside of metazoans, testing whether such a mechanism of regulation exists in this fungus, or other similar eukaryotes, represents an attractive functional assessment of ancestral NELF. Second, expressing all of the mammalian or *D. melanogaster* NELF

subunits in an organism with no such endogenous complex, such as *S. pombe* or *C. elegans,* would represent a direct examination of its capacity to induce regulated pausing. Admittedly, attempting to express a functional four-subunit complex in a foreign organism has its challenges, which could make negative results difficult to interpret. Nonetheless, such an ambitious undertaking could give critical insight into the mechanism of gene regulation through Pol II pausing.

Promoter-proximal pausing represents a layer of transcription regulation, which enables complex and rapidly adaptable circuitry in global gene expression. However, a complete understanding the basis of pausing and its functional impact remain elusive. Thus, continued research efforts through mechanistic, structural and evolutionary studies will be required to fully appreciate how this nexus of regulatory inputs has evolved in the context of complex cellular programming.

APPENDIX

CONSERVED TARGET RESIDUES WITHIN SPT5 ARE DISPENSABLE FOR CDK9 INFLUENCE ON THE ELONGATION COMPLEX IN FISSION YEAST

Abstract

In metazoans, transcriptionally engaged Pol II is often arrested within the promoter-proximal regions of genes and released upon specific signaling from the positive transcription elongation factor (P-TEFb) containing an essential kinase, Cdk9. In such systems the recruitment of P-TEFb to the stalled complex is widely considered to serve as mechanism for controlling the output of gene transcription(Adelman & Lis, 2012). We recently identified Pol II enrichments within promoter-proximal regions of genes in the fission yeast *Schizosaccharomyces pombe*, which resemble the profiles of engaged Pol II in organisms that regulate pause escape(Booth et al., 2016). However, unlike most well-studied metazoan systems, selective inhibition of analog-sensitive Cdk9 in fission yeast does not prevent the escape of the elongation complex from sites of Pol II enrichment immediately downstream of the transcription start site (TSS). Nonetheless, inhibiting Cdk9 dramatically reduces the elongation rate of Pol II, specifically if they have yet to transcribe beyond the early positions of a gene(Booth et al., 2018). These findings support the idea of an early checkpoint in fission yeast that might ensure maturation of the elongation complex(Pei et al., 2003). Concurrent with transcriptional defects, phosphorylation of the conserved Cdk9 substrate, Spt5 was essentially eliminated in *cdk9*^{as} within two minutes of kinase inhibition(Booth et al., 2018). However, whether

the effects of losing Cdk9 activity on transcription are a direct result of altering Spt5 phosphorylation remains unknown.

In this work, sought to directly examine whether Spt5 phosphorylation mediates the influence Cdk9 activity on the kinetics of Pol II elongation. Surprisingly, mutating all known target residues (Thr1) within nine amino-acid repeats of a truncated C-terminal domain Spt5 to either alanine or glutamate had very little effect on steady-state transcription profiles. Moreover, selectively inhibiting Cdk9 over a fine-temporal-scale time-course in the mutant Spt5 strains resulted in a nearly identical global impact on transcribing Pol II to a strain with consensus Spt5 repeats. Ultimately, despite the obvious and rapid loss of phosphorylation of Spt5 upon Cdk9 inhibition, elimination of this mark may not cause the observed transcriptional phenotypes that result from inactivation of Cdk9.

Results and discussion

Point mutations in the Cdk9 target residues of Spt5 CTR produce minimal effects on steady state transcription. Much like the unstructured C-terminal domain (CTD) of the largest subunit of Pol II, the C-terminus of Spt5 (called the CTR) contains an unstructured and repeating amino acid sequence. In fission yeast, this domain consists 18 tandem repeats of the nonapeptide $T_1P_2A_3W_4N_5S_6G_7S_8K_9$ consensus sequence(Schwer et al., 2009). As is the case in the human Spt5 counterpart, the threonine residue within this repeat (Thr1) is subject to phosphorylation by the conserved transcription coupled kinase, Cdk9 during transcription elongation(Pei & Shuman, 2003). By recruiting additional factors involved in RNA processing and chromatin modification, phosphorylation of the

Spt5 CTR has been proposed to impact transcription by Pol II(Mbogning et al., 2015; Schneider et al., 2010). To investigate the importance of Thr1 phosphorylation in transcription, we analyzed steady-state distributions of Pol II across the genome of fission yeast strains previously constructed to constitutively prevent or mimic this modification(Schneider et al., 2010). In this study, we used three strains, *Spt5-WT$_7$, Spt5-T1A$_7$* and *Spt5-T1E$_7$*, in which the CTR domain of Spt5 has been truncated to seven tandem copies of the nonapeptide sequence. While the *Spt5-WT$_7$* has all perfect matches to the consensus sequence, *Spt5-T1A$_7$* and *Spt5-T1E$_7$* strains contained substitutions of alanine or glutamate, respectively at each Thr1 position within repeats (Figure A.1a). Importantly, all three mutants occur in a background strain with the previously described analog sensitive *cdk9^{as}*(Booth et al., 2018; Viladevall et al., 2009).

To assess the impact of the Spt5 CTR mutations on transcription, we prepared precision run-on sequencing (PRO-seq) libraries from biological replicate cultures of each strain after spiking in a small amount of *Saccharomyces cerevisiae* for normalization. Correlations of gene body transcriptional activity between replicates suggested data from like samples was highly reproducible. Further, the symmetrical scatter about the line $x = y$ indicates that normalization based on spike-in derived RNA counts accurately accounted for differences in sequencing depth between samples (Figure A.1b). Surprisingly, almost no genes exhibited significant changes in transcriptional activity when comparing *Spt5-WT$_7$* with strains possessing either constitutively unphosphorylated Spt5 (*Spt5-T1A$_7$*) or phospho-mimetic Spt5 (*Spt5-T1E$_7$*) (Figure A.1c). Moreover, distributions of elongating Pol II across genes on average were largely unchanged across mutants, indicating very minimal if any impact of the CTR

substitutions on steady state Pol II elongation or its regulation (Figure A.1d). Notably, in *Spt5-T1E*$_7$ strains there is an observable drop in Pol II signal beyond the cleavage and polyadenylation signal (CPS), which could reflect a role of Spt5 Thr1 phosphorylation status in mediating transcription termination(Parua et al., 2018). While all three strains exhibit almost no growth phenotype between 25-30 °C, temperatures outside of this range appear to challenge the growth of strains with the examined mutations, particularly T1A(Schneider et al., 2010). Thus, rather than affecting steady state transcription, perhaps the Spt5 CTR enhances the ability to rapidly tune or adapt gene expression to environmental variations.

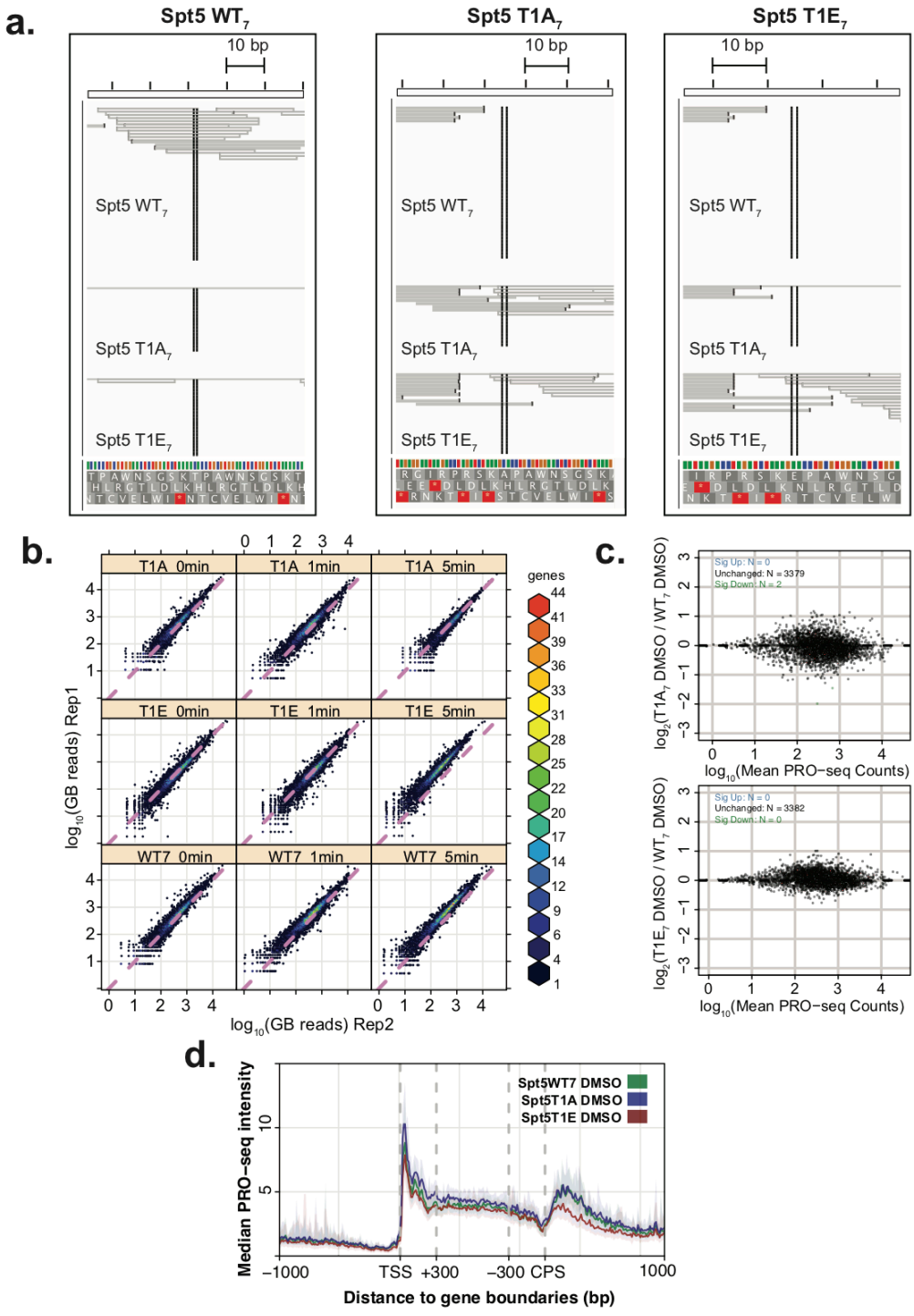

Figure A.1| Point mutations in the Cdk9 target residues of Spt5 CTR produce minimal effects on steady state transcription. a. Verification of strain mutations. Browser tracks show aligned PRO-seq data from two combined replicates for each strain within the Spt5 C-terminal repeat sequence. PRO-seq reads from each strain were aligned specifically to the Spt5 gene sequence bearing two copies of either the

WT repeat (left), T1A containing repeat (middle), or T1E containing repeat (right). Center line in each panel is placed above the Thr1 or mutated residue within the Spt5 CTR repeat. **b.** Scatter plots comparing normalized read counts within the gene body of all genes (n = 6591) between biological replicates for each sample. **c.** Differential expression analysis of gene body PRO-seq read counts from all genes (n = 3383) between untreated Spt5 $T1A_7$ and Spt5 WT_7 (top panel), or Spt5 $T1E_7$ and Spt5 WT_7 (bottom panel). **d.** Scaled composite profiles of normalized PRO-seq data from combined replicates of each untreated strain around filtered genes at least 1 kb from nearest same strand neighbors (n = 939). From -1000 to + 300 around the TSS and -300 to +1000 around the CPS, values reflect 10 bp windows. Intervening regions are composed of 30 windows, scaled based on the length of each gene. Center lines represent the median values at each position, while shaded areas represent bootstrapped 12.5% and 87.5% confidence intervals of the median signal. Normalization of individual or combined replicate data was performed by dividing all values by by spike-in RNA counts/10^5. Untreated samples were treated with DMSO for 5 minutes to control for the solvent of 3MB-PP1 at the longest treatment time-point.

Mutations of Thr1 residues within the Spt5 CTR do not alter the impact of Cdk9 inhibition on transcription. To test whether Spt5 phosphorylation was required for the transcriptional response to Cdk9 inhibition, we exploited the fact that each Spt5-CTR mutant also contained an analog sensitive *cdk9*^{as} mutation. Thus, for each Spt5-CTR mutant strain, we conducted a time-course of Cdk9 inhibition. Performed in biological replicate (Figure A.1b), each strain was treated with 10 μM 3MB-PP1 for one or five minutes, or else an equivalent volume of DMSO for five minutes (considered the 0-min. time point). Individual genes exhibited nearly the exact same dynamic response in Pol II distributions over the time-course of *cdk9*^{as} inhibition as was previously described(Booth et al., 2018). Remarkably, this effect on Pol II was observed regardless of the status of Spt5 CTR (Figure A.2a). Moreover, in all three strains, global Pol II distributions responded instantaneously to Cdk9 inhibition and progressed in the same manner with increasing time (Figure A.2b-d). Finally, on long genes, both previously described slow advancing and rapid clearing populations of Pol II(Booth et al., 2018) could be observed

in response to addition of 3MB-PP1. Again, *cdk9^{as}Spt5-WT$_7$, cdk9^{as}Spt5-T1A$_7$,* and *cdk9^{as}Spt5-T1E$_7$,* all shared highly similar kinetic properties with respect to the two differentially affected Pol II populations. Importantly, although the T1A containing peptide is not apparently phosphorylated elsewhere by Cdk9 *in vitro*(Schneider et al., 2010), we cannot rule out that *in vivo* Cdk9 may be capable of targeting alternative residues within the examined Spt5 CTR upon substitution of the Thr1 target residue to either Ala or Glu. To convincingly exclude Spt5 phosphorylation as the consequential target of Cdk9 for impacting transcription, future experiments should investigate how cells respond to Cdk9 inhibition in mutant strains, devoid of all Spt5 C-terminal repeats.

Since kinase inhibition of Cdk9, rather than complete protein removal, produces such transcriptional consequences, it is likely that Cdk9 mediates its impact through its kinase activity. Ultimately, the results presented here make a compelling case for a target of Cdk9, other than Spt5, that likely drives its influence over transcription. Although our previous work showed minimal reduction in Pol II CTD phosphorylation after short inhibitions of Cdk9(Booth et al., 2018), Ser2 with the repeated heptad consensus sequence, $Y_1S_2P_3T_4S_5P_6S_7$, is a well characterized and conserved target of Cdk9(Ahn, Kim, & Buratowski, 2004; Czudnochowski, Bösken, & Geyer, 2012; Viladevall et al., 2009). While another transcription coupled kinase Lsk1/Cdk12 overlaps in its specificity for Ser2(Viladevall et al., 2009), Cdk9 might be specifically required for this modification of Pol II within the promoter-proximal region of genes, or during pause escape. The effects of Cdk9 inhibition could also be a consequence of loss of phosphorylation from combinations of substrates, rather than a single target. Additionally, other less characterized or as yet unidentified targets of Cdk9 may play a

role. Determining the mechanism by which Cdk9 exerts its function in transcription remains a critical problem in the field. Additional experiments are required to fully appreciate the importance of Cdk9 activity in controlling transcription by RNA Pol II.

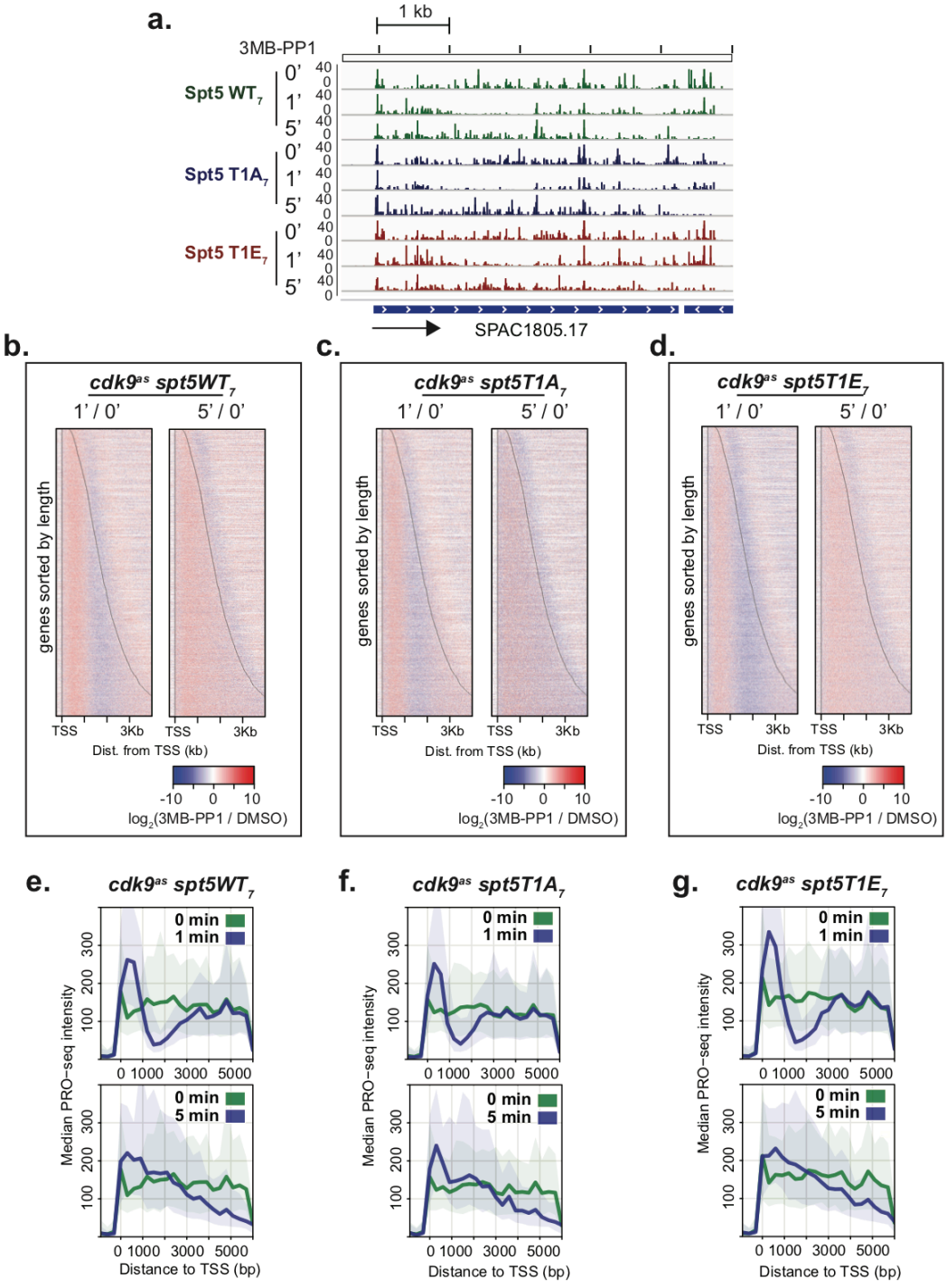

Figure A.2| Mutations of Thr1 residues within the Spt5 CTR do not alter the impact of Cdk9 inhibition on transcription. a. Browser image of normalized PRO-seq data prepared from two replicates and combined for each strain over a time-course (0 min, 1 min, 5 min) of cdk9^{as} inhibition with 10 μM 3MB-PP1. **b-d.** Heat maps of $\log_2$ (3MB-PP1 treated / DMSO treated) normalized PRO-seq counts within 10 bp windows all filtered genes (n = 3383) sorted by increasing length from top to bottom. Within each panel, the left heat map displays the fold change at each position after 1 min. treatment with 3MB-PP1 over untreated (DMSO), while the right heat map shows the fold change after 5 min. treatment with 3MB-PP1 over untreated. Equivalent comparisons are shown separately for *cdk9^{as}Spt5WT$_7$*(**b**), *cdk9^{as}Spt5T1A$_7$*(**c**), and *cdk9^{as}Spt5T1E$_7$*(**d**). **e-g.** Composite profiles of normalized PRO-seq data from combined replicates of each strain after 1 min. (top panels), or 5 min. treatment with 10 μM 3MB-PP1 compared with DMSO, around filtered genes at least 6 kb in length (n = 42), for *cdk9^{as}Spt5WT$_7$*(**e**), *cdk9^{as}Spt5T1A$_7$*(**f**), and *cdk9^{as}Spt5T1E$_7$*(**g**). Composite profiles represent the sum of normalized counts within 300 bp windows from -900 to +6000 bp upstream and downstream of the observed TSS, respecitively. Center lines represent the median values at each position, while shaded areas represent bootstrapped 12.5% and 87.5% confidence intervals of the median signal.

Methods

Methods overview. Most experimental approaches and analyses used in this work significantly overlap with those described in Chapter 3 of this thesis and will thus not be repeated here. Please see the methods section of Chapter 3 of this thesis for information regarding the following methods sections: 1) Treatment of analog sensitive strains for PRO-seq. 2) Time course experiments. 3) Alignment and data processing. 4) Gene sets. 5) Differential expression analysis. Importantly, precision run-on sequencing was also performed as described in Chapter 3, however, rather than using all four biotin-11-NTPs in the run-on reaction, only biotin-11-CTP and biotin-11-UTP were used along with unmodified ATP and GTP (substantially reduces cost per sample). This will subtly reduce the resolution of the PRO-seq

assay in cases where the downstream DNA base within the Pol II active site is not G or A.

Strains. All strains used in this work were generous gifts from Robert P. Fisher. Each of the three strains was created by crossing the previously characterized *cdk9*^{as} strain(Viladevall et al., 2009) separately with each of Spt5 CTR mutants, which were previously engineered and described by Schneider et al. (Schneider et al., 2010).

Sample barcodes for PRO-seq. All PRO-seq libraries generated for this work were prepared as described in the methods section of Chapter 3 of this thesis. However, in Chapter 3, three different 3' adaptors were described. Here all samples received a 3' RNA adaptor containing an in-line barcode for sample identification (5' -/5Phos/ GNNNNNNGAUCGUCGGACUGUAGAACUCUGAAC-/Inverted dT/). Thus, during sequencing processing, the first 7 nt of each read are used to distinguish samples and must be removed before alignment. These processing steps are described in greater detail in Chapter 3.

REFERENCES

Adelman, K., & Lis, J. T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews. Genetics*, *13*(10), 720–731. http://doi.org/10.1038/nrg3293

Adelman, K., Kennedy, M. A., Nechaev, S., Gilchrist, D. A., Muse, G. W., Chinenov, Y., & Rogatsky, I. (2009). Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(43), 18207–18212. http://doi.org/10.1073/pnas.0910177106

Adelman, K., Wei, W., Ardehali, M. B., Werner, J., Zhu, B., Reinberg, D., & Lis, J. T. (2006). Drosophila Paf1 modulates chromatin structure at actively transcribed genes. *Molecular and Cellular Biology*, *26*(1), 250–260. http://doi.org/10.1128/MCB.26.1.250-260.2006

Ahn, S. H., Kim, M., & Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Molecular Cell*, *13*(1), 67–76.

Akhtar, M. S., Heidemann, M., Tietjen, J. R., Zhang, D. W., Chapman, R. D., Eick, D., & Ansari, A. Z. (2009). TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Molecular Cell*, *34*(3), 387–393. http://doi.org/10.1016/j.molcel.2009.04.016

Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W., & Lis, J. T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. *The EMBO Journal*, *28*(8), 1067–1077. http://doi.org/10.1038/emboj.2009.56

Azuma, Y., Yamagishi, M., Ueshima, R., & Ishihama, A. (1991). Cloning and sequence determination of the Schizosaccharomyces pombe rpb1 gene encoding the largest subunit of RNA polymerase II. *Nucleic Acids Research*, *19*(3), 461–468.

Baker, K. E., & Coller, J. (2006). The many routes to regulating mRNA translation. (Vol. 7, p. 332). Presented at the Genome biology. http://doi.org/10.1186/gb-2006-7-12-332

Barboric, M., Nissen, R. M., Kanazawa, S., Jabrane-Ferrat, N., & Peterlin, B. M. (2001). NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Molecular Cell*, *8*(2), 327–337.

Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., et al. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes & Development*, *24*(20), 2303–2316. http://doi.org/10.1101/gad.1968210

Bähler, J., Wu, J. Q., Longtine, M. S., Shah, N. G., McKenzie, A., Steever, A. B., et al. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast (Chichester, England)*, *14*(10), 943–951. http://doi.org/10.1002/(SICI)1097-0061(199807)14:10<943::AID-YEA292>3.0.CO;2-Y

Beisel, C., & Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. *Nature Reviews. Genetics*, *12*(2), 123–135. http://doi.org/10.1038/nrg2932

Bensaude, O. (2014). Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity? *Transcription*, *2*(3), 103–108. http://doi.org/10.1002/(SICI)1097-4652(199607)168:1<105::AID-JCP13>3.0.CO;2-6

Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature Reviews. Genetics*, *15*(3), 163–175. http://doi.org/10.1038/nrg3662

Bentley, D. L., & Groudine, M. (1986). A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. *Nature*, *321*(6071), 702–706. http://doi.org/10.1038/321702a0

Bintu, L., Kopaczynska, M., Hodges, C., Lubkowska, L., Kashlev, M., & Bustamante, C. (2011). The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. *Nature Structural & Molecular Biology*, *18*(12), 1394–1399. http://doi.org/10.1038/nsmb.2164

Bishop, A. C., Ubersax, J. A., Petsch, D. T., Matheos, D. P., Gray, N. S., Blethrow, J., et al. (2000). A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature*, *407*(6802), 395–401. http://doi.org/10.1038/35030148

Boeger, H., Griesenbeck, J., Strattan, J. S., & Kornberg, R. D. (2003). Nucleosomes unfold completely at a transcriptionally active promoter. *Molecular Cell*, *11*(6), 1587–1598.

Booth, G. T., Parua, P. K., Sansó, M., Fisher, R. P., & Lis, J. T. (2018). Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. *Nature Communications*, *9*(1), 543.

Booth, G. T., Wang, I. X., Cheung, V. G., & Lis, J. T. (2016). Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome Research*, *26*(6), 799–811. http://doi.org/10.1101/gr.204578.116

Bösken, C. A., Farnung, L., Hintermair, C., Merzel Schachter, M., Vogel-Bachmayr, K., Blazek, D., et al. (2014). The structure and substrate specificity of human Cdk12/Cyclin K. *Nature Communications*, *5*, 115. http://doi.org/10.1038/nsb0896-696

Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., et al. (2012). mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. *Molecular Cell*, *46*(3), 311–324. http://doi.org/10.1016/j.molcel.2012.03.006

Brown, C. E., Lechner, T., Howe, L., & Workman, J. L. (2000). The many HATs of transcription coactivators. *Trends in Biochemical Sciences*, *25*(1), 15–19.

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

Buratowski, S. (2003). The CTD code. *Nature Structural & Molecular Biology*, *10*(9), 679–680. http://doi.org/10.1074/jbc.M003165200

Burova, E., Hung, S. C., Sagitov, V., Stitt, B. L., & Gottesman, M. E. (1995). Escherichia coli NusG protein stimulates transcription elongation rates in vivo and in vitro. *Journal of Bacteriology*, *177*(5), 1388–1392.

Castel, S. E., Ren, J., Bhattacharjee, S., Chang, A.-Y., Sánchez, M., Valbuena, A., et al. (2014). Dicer promotes transcription termination at sites of replication stress to maintain genome stability. *Cell*, *159*(3), 572–583. http://doi.org/10.1016/j.cell.2014.09.031

Chao, S. H., & Price, D. H. (2001). Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. *The Journal of Biological Chemistry*, *276*(34), 31793–31799. http://doi.org/10.1074/jbc.M102306200

Chen, Fei Xavier, Xie, P., Collings, C. K., Cao, K., Aoi, Y., Marshall, S. A., et al. (2017). PAF1 regulation of promoter-proximal pause release via enhancer activation. *Science (New York, N.Y.)*. http://doi.org/10.1126/science.aan3269

Chen, Yexi, Yamaguchi, Y., Tsugeno, Y., Yamamoto, J., Yamada, T., Nakamura, M., et al. (2009). DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes & Development*, *23*(23), 2765–2777. http://doi.org/10.1101/gad.1834709

Chen, Yin, Hamati, E., Lee, P.-K., Lee, W.-M., Wachi, S., Schnurr, D., et al. (2006). Rhinovirus induces airway epithelial gene expression through double-stranded RNA and IFN-dependent pathways. *American Journal of Respiratory Cell and Molecular Biology*, *34*(2), 192–203. http://doi.org/10.1165/rcmb.2004-0417OC

Cheng, B., Li, T., Rahl, P. B., Adamson, T. E., Loudas, N. B., Guo, J., et al. (2012). Functional association of Gdown1 with RNA polymerase II poised on human genes. *Molecular Cell*, *45*(1), 38–50. http://doi.org/10.1016/j.molcel.2011.10.022

Cho, E. J., Takagi, T., Moore, C. R., & Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes & Development*, *11*(24), 3319–3326.

Cho, H., Orphanides, G., Sun, X., Yang, X. J., Ogryzko, V., Lees, E., et al. (1998). A human RNA polymerase II complex containing factors that modify chromatin structure. *Molecular and Cellular Biology*, *18*(9), 5355–5363.

Churchman, L. S., & Weissman, J. S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, *469*(7330), 368–373. http://doi.org/10.1038/nature09652

Cipak, L., Zhang, C., Kovacikova, I., Rumpf, C., Miadokova, E., Shokat, K. M., & Gregan, J. (2011). Generation of a set of conditional analog-sensitive alleles of essential protein kinases in the fission yeast Schizosaccharomyces pombe. *Cell Cycle (Georgetown, Tex.)*, *10*(20), 3527–3532. http://doi.org/10.4161/cc.10.20.17792

Collart, M. A., & Oliviero, S. (1993). Preparation of yeast RNA. *Current Protocols in Molecular Biology*, 13.12.1–13.12.5.

Connelly, S., & Manley, J. L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & Development*, *2*(4), 440–452. http://doi.org/10.1101/gad.2.4.440

Corden, J. L. (1990). Tails of RNA polymerase II. *Trends in Biochemical Sciences*, *15*(10), 383–387.

Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, *46*(12), 1311–1320. http://doi.org/10.1038/ng.3142

Core, L. J., Waterfall, J. J., & Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (New York, N.Y.)*, *322*(5909), 1845–1848. http://doi.org/10.1126/science.1162228

Core, L. J., Waterfall, J. J., Gilchrist, D. A., Fargo, D. C., Kwak, H., Adelman, K., & Lis, J. T. (2012). Defining the status of RNA polymerase at promoters. *Cell Reports*, *2*(4), 1025–1035. http://doi.org/10.1016/j.celrep.2012.08.034

Coudreuse, D., van Bakel, H., Dewez, M., Soutourina, J., Parnell, T., Vandenhaute, J., et al. (2010). A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe. *Current Biology : CB*, *20*(12), 1053–1064. http://doi.org/10.1016/j.cub.2010.04.054

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: a

sequence logo generator. *Genome Research*, *14*(6), 1188–1190. http://doi.org/10.1101/gr.849004

Czudnochowski, N., Bösken, C. A., & Geyer, M. (2012). Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nature Communications*, *3*, 842. http://doi.org/10.1038/ncomms1846

Dahan, O., Gingold, H., & Pilpel, Y. (2011). Regulatory mechanisms and networks couple the different phases of gene expression. *Trends in Genetics : TIG*, *27*(8), 316–322. http://doi.org/10.1016/j.tig.2011.05.008

Danko, C. G., Hah, N., Luo, X., Martins, A. L., Core, L., Lis, J. T., et al. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Molecular Cell*, *50*(2), 212–222. http://doi.org/10.1016/j.molcel.2013.02.015

DeGennaro, C. M., Alver, B. H., Marguerat, S., Stepanova, E., Davis, C. P., Bähler, J., et al. (2013). Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. *Molecular and Cellular Biology*, *33*(24), 4779–4792. http://doi.org/10.1128/MCB.01068-13

Dekker, J. (2008). Gene regulation in the third dimension. *Science (New York, N.Y.)*, *319*(5871), 1793–1794. http://doi.org/10.1126/science.1152850

Doamekpor, S. K., Sanchez, A. M., Schwer, B., Shuman, S., & Lima, C. D. (2014). How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. *Genes & Development*, *28*(12), 1323–1336. http://doi.org/10.1101/gad.242768.114

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. http://doi.org/10.1093/bioinformatics/bts635

Duarte, F. M., Fuda, N. J., Mahat, D. B., Core, L. J., Guertin, M. J., & Lis, J. T. (2016). Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes & Development*, *30*(15), 1731–1746. http://doi.org/10.1101/gad.284430.116

Dujon, B. (2010). Yeast evolutionary genomics. *Nature Reviews. Genetics*, *11*(7), 512–524. http://doi.org/10.1038/nrg2689

Eberhardy, S. R., & Farnham, P. J. (2002). Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *The Journal of Biological Chemistry*, *277*(42), 40156–40162. http://doi.org/10.1074/jbc.M207441200

Ebmeier, C. C., Erickson, B., Allen, B. L., Allen, M. A., Kim, H., Fong, N., et al. (2017).

Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. *Cell Reports*, *20*(5), 1173–1186. http://doi.org/10.1016/j.celrep.2017.07.021

Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., & Sekine, S.-I. (2017). Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science (New York, N.Y.)*, *357*(6354), 921–924. http://doi.org/10.1126/science.aan8552

Eser, P., Wachutka, L., Maier, K. C., Demel, C., Boroni, M., Iyer, S., et al. (2016). Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. *Molecular Systems Biology*, *12*(2), 857.

Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Reviews. Genetics*, *9*(2), 102–114. http://doi.org/10.1038/nrg2290

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, *391*(6669), 806–811. http://doi.org/10.1038/35888

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M. A., & Bentley, D. L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Molecular Cell*, *60*(2), 256–267.

Fuda, N. J., Ardehali, M. B., & Lis, J. T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*, *461*(7261), 186–192. http://doi.org/10.1038/nature08449

Fuda, N. J., Guertin, M. J., Sharma, S., Danko, C. G., & Martins, A. L. (2015). GAGA Factor Maintains Nucleosome-Free Regions and Has a Role in RNA Polymerase II Recruitment to Promoters. *PLoS Genetics*.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., & Peterlin, B. M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Molecular and Cellular Biology*, *24*(2), 787–795.

Gallego, M., & Virshup, D. M. (2007). Post-translational modifications regulate the ticking of the circadian clock. *Nature Reviews. Molecular Cell Biology*, *8*(2), 139–148. http://doi.org/10.1038/nrm2106

García-Martínez, J., Aranda, A., & Pérez-Ortín, J. E. (2004). Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Molecular Cell*, *15*(2), 303–313. http://doi.org/10.1016/j.molcel.2004.06.004

Gibson, B. A., Zhang, Y., Jiang, H., Hussey, K. M., Shrimp, J. H., Lin, H., et al. (2016). Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. *Science (New York, N.Y.)*, *353*(6294), 45–50. http://doi.org/10.1126/science.aaf7865

Gilchrist, D. A., Fromm, G., Santos, Dos, G., Pham, L. N., McDaniel, I. E., Burkholder, A., et al. (2012). Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. *Genes & Development*, *26*(9), 933–944. http://doi.org/10.1101/gad.187781.112

Gilchrist, D. A., Nechaev, S., Lee, C., Ghosh, S. K. B., Collins, J. B., Li, L., et al. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes & Development*, *22*(14), 1921–1933. http://doi.org/10.1101/gad.1643208

Gilchrist, D. A., Santos, dos, G., Fargo, D. C., Xie, B., Gao, Y., Li, L., & Adelman, K. (2010). Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation. *Cell*, *143*(4), 540–551. http://doi.org/10.1016/j.cell.2010.10.004

Givens, R. M., Lai, W. K. M., Rizzo, J. M., Bard, J. E., Mieczkowski, P. A., Leatherwood, J., et al. (2012). Chromatin architectures at fission yeast transcriptional promoters and replication origins. *Nucleic Acids Research*, *40*(15), 7176–7189. http://doi.org/10.1093/nar/gks351

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R. P., & Bentley, D. L. (2009). 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*(20), 5455–5464. http://doi.org/10.1128/MCB.00637-09

Gressel, S., Schwalb, B., Decker, T. M., Qin, W., Leonhardt, H., Eick, D., & Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife*, *6*. http://doi.org/10.7554/eLife.29736

Gromak, N., West, S., & Proudfoot, N. J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Molecular and Cellular Biology*, *26*(10), 3986–3996. http://doi.org/10.1128/MCB.26.10.3986-3996.2006

Guo, J., Turek, M. E., & Price, D. H. (2014). Regulation of RNA polymerase II termination by phosphorylation of Gdown1. *The Journal of Biological Chemistry*, *289*(18), 12657–12665. http://doi.org/10.1074/jbc.M113.537662

Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G. A., et al. (2008). Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. *Structure (London, England : 1993)*, *16*(11), 1649–1658.

http://doi.org/10.1016/j.str.2008.08.013

Hah, N., Danko, C. G., Core, L., Waterfall, J. J., Siepel, A., Lis, J. T., & Kraus, W. L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*, *145*(4), 622–634. http://doi.org/10.1016/j.cell.2011.03.042

Hahn, S., & Young, E. T. (2011). Transcriptional Regulation in Saccharomyces cerevisiae: Transcription Factor Regulation and Function, Mechanisms of Initiation, and Roles of Activators and Coactivators. *Genetics*, *189*(3), 705–736. http://doi.org/10.1534/genetics.111.127019

Hall, M. A., Shundrovsky, A., Bai, L., Fulbright, R. M., Lis, J. T., & Wang, M. D. (2009). High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nature Structural & Molecular Biology*, *16*(2), 124–129. http://doi.org/10.1038/nsmb.1526

Harmston, N., & Lenhard, B. (2013). Chromatin and epigenetic features of long-range gene regulation. *Nucleic Acids Research*, *41*(15), 7185–7199. http://doi.org/10.1093/nar/gkt499

Hartzog, G. A., & Fu, J. (2013). The Spt4–Spt5 complex: A multi-faceted regulator of transcription elongation. *Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1829*(1), 105–115. http://doi.org/10.1016/j.bbagrm.2012.08.007

Hartzog, G. A., Wada, T., Handa, H., & Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes & Development*, *12*(3), 357–369.

Henriques, T., Gilchrist, D. A., Nechaev, S., Bern, M., Muse, G. W., Burkholder, A., et al. (2013). Stable Pausing by RNA Polymerase II Provides an Opportunity to Target and Integrate Regulatory Signals. *Molecular Cell*, *52*(4), 517–528. http://doi.org/10.1016/j.molcel.2013.10.001

Henriques, T., Scruggs, B. S., Inouye, M. O., Muse, G. W., Williams, L. H., Burkholder, A. B., et al. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Genes & Development*, *32*(1), 26–41. http://doi.org/10.1101/gad.309351.117

Hsin, J.-P., & Manley, J. L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & Development*, *26*(19), 2119–2137. http://doi.org/10.1101/gad.200303.112

Hu, Z., Chen, K., Xia, Z., Chavez, M., Pal, S., Seol, J.-H., et al. (2014). Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. *Genes & Development*, *28*(4), 396–408.

http://doi.org/10.1101/gad.233221.113

Ivanov, D., Kwak, Y. T., Guo, J., & Gaynor, R. B. (2000). Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Molecular and Cellular Biology*, *20*(9), 2970–2983.

Izban, M. G., & Luse, D. S. (1992). Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *The Journal of Biological Chemistry*, *267*(19), 13647–13655.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33 Suppl*, 245–254. http://doi.org/10.1038/ng1089

Jeronimo, C., Langelier, M.-F., Bataille, A. R., Pascal, J. M., Pugh, B. F., & Robert, F. (2016). Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. *Molecular Cell*, *64*(3), 455–466. http://doi.org/10.1016/j.molcel.2016.09.002

Jiang, C., & Pugh, B. F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews. Genetics*, *10*(3), 161–371. http://doi.org/10.1038/nrg2522

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

Jordán-Pla, A., Gupta, I., de Miguel-Jiménez, L., Steinmetz, L. M., Chávez, S., Pelechano, V., & Pérez-Ortín, J. E. (2014). Chromatin-dependent regulation of RNA polymerases II and III activity throughout the transcription cycle. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gku1349

Juven-Gershon, T., Hsu, J.-Y., Theisen, J. W., & Kadonaga, J. T. (2008). The RNA polymerase II core promoter - the gateway to transcription. *Current Opinion in Cell Biology*, *20*(3), 253–259. http://doi.org/10.1016/j.ceb.2008.03.003

Kao, C.-F., & Osley, M. A. (2003). In vivo assays to study histone ubiquitylation. *Methods (San Diego, Calif.)*, *31*(1), 59–66.

Karagiannis, J., & Balasubramanian, M. K. (2007). A cyclin-dependent kinase that promotes cytokinesis through modulating phosphorylation of the carboxy terminal domain of the RNA Pol II Rpb1p sub-unit. *PloS One*, *2*(5), e433. http://doi.org/10.1371/journal.pone.0000433

Keogh, M.-C., Podolny, V., & Buratowski, S. (2003). Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Molecular and Cellular*

Biology, *23*(19), 7005–7018.

Kephart, D. D., Marshall, N. F., & Price, D. H. (1992). Stability of Drosophila RNA polymerase II elongation complexes in vitro. *Molecular and Cellular Biology*, *12*(5), 2067–2077.

Kim, J., Guermah, M., & Roeder, R. G. (2010a). The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell*, *140*(4), 491–503. http://doi.org/10.1016/j.cell.2009.12.050

Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., et al. (2010b). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*(7295), 182–187. http://doi.org/10.1038/nature09033

Kornberg, R. D. (1977). Structure of chromatin. *Annual Review of Biochemistry*, *46*, 931–954. http://doi.org/10.1146/annurev.bi.46.070177.004435

Kornberg, R. D. (2005). Mediator and the mechanism of transcriptional activation. *Trends in Biochemical Sciences*, *30*(5), 235–239. http://doi.org/10.1016/j.tibs.2005.03.011

Kruesi, W. S., Core, L. J., Waters, C. T., Lis, J. T., & Meyer, B. J. (2013). Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife*, *2*, e00808. http://doi.org/10.7554/eLife.00808

Krumm, A., Hickey, L. B., & Groudine, M. (1995). Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. *Genes & Development*, *9*(5), 559–572.

Kwak, H., & Lis, J. T. (2013). Control of Transcriptional Elongation. *Annual Review of Genetics*. http://doi.org/10.1146/annurev-genet-110711-155440

Kwak, H., Fuda, N. J., Core, L. J., & Lis, J. T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science (New York, N.Y.)*, *339*(6122), 950–953. http://doi.org/10.1126/science.1229386

Kwiatkowski, N., Zhang, T., Rahl, P. B., Abraham, B. J., Reddy, J., Ficarro, S. B., et al. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature*, *511*(7511), 616–620. http://doi.org/10.1038/nature13393

Lagha, M., Bothma, J. P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., et al. (2013). Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. *Cell*, *153*(5), 976–987. http://doi.org/10.1016/j.cell.2013.04.045

Laitem, C., Zaborowska, J., Isa, N. F., Kufs, J., Dienstbier, M., & Murphy, S. (2015). CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-

transcribed genes. *Nature Structural & Molecular Biology*, *22*(5), 396–403. http://doi.org/10.1038/nsmb.3000

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. http://doi.org/10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. http://doi.org/10.1186/gb-2009-10-3-r25

Larochelle, S., Amat, R., Glover-Cutter, K., Sansó, M., Zhang, C., Allen, J. J., et al. (2012). Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nature Structural & Molecular Biology*, *19*(11), 1108–1115. http://doi.org/10.1038/nsmb.2399

Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature*, *424*(6945), 147–151. http://doi.org/10.1038/nature01763

Li, B., Carey, M., & Workman, J. L. (2007). The Role of Chromatin during Transcription. *Cell*, *128*(4), 707–719. http://doi.org/10.1016/j.cell.2007.01.015

Li, B., Howe, L., Anderson, S., Yates, J. R., & Workman, J. L. (2003). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *The Journal of Biological Chemistry*, *278*(11), 8897–8903. http://doi.org/10.1074/jbc.M212134200

Li, H., Hou, J., Bai, L., Hu, C., Tong, P., Kang, Y., et al. (2015). Genome-wide analysis of core promoter structures in Schizosaccharomyces pombe with DeepCAGE. *RNA Biology*, *12*(5), 525–537. http://doi.org/10.1080/15476286.2015.1022704

Li, J., Liu, Y., Rhee, H. S., Ghosh, S. K. B., Bai, L., Pugh, B. F., & Gilmour, D. S. (2013). Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Molecular Cell*, *50*(5), 711–722. http://doi.org/10.1016/j.molcel.2013.05.016

Licatalosi, D. D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J. B., & Bentley, D. L. (2002). Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Molecular Cell*, *9*(5), 1101–1111.

Lindstrom, D. L., Squazzo, S. L., Muster, N., Burckin, T. A., Wachter, K. C., Emigh, C. A., et al. (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Molecular and Cellular Biology*, *23*(4), 1368–1378.

Liu, X., Bushnell, D. A., & Kornberg, R. D. (2013). RNA polymerase II transcription: structure and mechanism. *Biochimica Et Biophysica Acta*, *1829*(1), 2–8.

http://doi.org/10.1016/j.bbagrm.2012.09.003

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W. H., et al. (2009). Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Molecular and Cellular Biology*, *29*(17), 4852–4863. http://doi.org/10.1128/MCB.00609-09

Love, J. D., Vivino, A. A., & Minton, K. W. (1985). Detection of low-level gene induction using in vitro transcription heat shock genes. *Gene Analysis Techniques*, *2*(6), 100–107. http://doi.org/10.1016/0735-0651(85)90005-6

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. http://doi.org/10.1186/s13059-014-0550-8

Love, M., Anders, S., & Huber, W. (2013). Differential analysis of count data–the DESeq2 package.

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., & Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell*, *144*(1), 16–26. http://doi.org/10.1016/j.cell.2010.11.056

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260. http://doi.org/10.1038/38444

Mahat, D. B., Kwak, H., Booth, G. T., Jonkers, I. H., Danko, C. G., Patel, R. K., et al. (2016a). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nature Protocols*, *11*(8), 1455–1476. http://doi.org/10.1038/nprot.2016.086

Mahat, D. B., Salamanca, H. H., Duarte, F. M., Danko, C. G., & Lis, J. T. (2016b). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Molecular Cell*, *62*(1), 63–78. http://doi.org/10.1016/j.molcel.2016.02.025

Marshall, N. F., & Price, D. H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Molecular and Cellular Biology*, *12*(5), 2078–2090.

Marshall, N. F., & Price, D. H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *The Journal of Biological Chemistry*, *270*(21), 12335–12338.

Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C., & Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal

transcription processivity. *The EMBO Journal*, *30*(7), 1302–1310. http://doi.org/10.1073/pnas.0806302106

Mason, P. B., & Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Molecular Cell*, *17*(6), 831–840. http://doi.org/10.1016/j.molcel.2005.02.017

Mavrich, T. N., Jiang, C., Ioshikhes, I. P., Li, X., Venters, B. J., Zanton, S. J., et al. (2008). Nucleosome organization in the Drosophila genome. *Nature*, *453*(7193), 358–362. http://doi.org/10.1038/nature06929

Maxwell, C. S., Kruesi, W. S., Core, L. J., Kurhanewicz, N., Waters, C. T., Lewarch, C. L., et al. (2014). Pol II Docking and Pausing at Growth and Stress Genes in C. elegans. *Cell Reports*, *6*(3), 455–466. http://doi.org/10.1016/j.celrep.2014.01.008

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., & Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nature Structural & Molecular Biology*, *17*(10), 1272–1278. http://doi.org/10.1038/nsmb.1903

Mbogning, J., Pagé, V., Burston, J., Schwenger, E., Fisher, R. P., Schwer, B., et al. (2015). Functional interaction of Rpb1 and Spt5 C-terminal domains in co-transcriptional histone modification. *Nucleic Acids Research*, *43*(20), 9766–9775. http://doi.org/10.1093/nar/gkv837

McGinnis, W., & Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell*, *68*(2), 283–302.

McKinlay, A., Araya, C. L., & Fields, S. (2011). Genome-Wide Analysis of Nascent Transcription in Saccharomyces cerevisiae. *G3 (Bethesda, Md.)*, *1*(7), 549–558. http://doi.org/10.1534/g3.111.000810

Meinhart, A., & Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, *430*(6996), 223–226. http://doi.org/10.1038/nature02679

Merkhofer, E. C., Hu, P., & Johnson, T. L. (2014). Introduction to cotranscriptional RNA splicing. *Methods in Molecular Biology (Clifton, N.J.)*, *1126*, 83–96. http://doi.org/10.1007/978-1-62703-980-2_6

Michels, A. A., Fraldi, A., Li, Q., Adamson, T. E., Bonnet, F., Nguyen, V. T., et al. (2004). Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *The EMBO Journal*, *23*(13), 2608–2619. http://doi.org/10.1038/sj.emboj.7600275

Min, I. M., Waterfall, J. J., Core, L. J., Munroe, R. J., Schimenti, J., & Lis, J. T. (2011).

Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes & Development*, *25*(7), 742–754. http://doi.org/10.1101/gad.2005511

Missra, A., & Gilmour, D. S. (2010). Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11301–11306. http://doi.org/10.1073/pnas.1000681107

Morimoto, R. I. (1993). Cells in stress: transcriptional activation of heat shock genes. *Science (New York, N.Y.)*, *259*(5100), 1409–1410.

Muñoz, M. J., la Mata, de, M., & Kornblihtt, A. R. (2010). The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends in Biochemical Sciences*, *35*(9), 497–504. http://doi.org/10.1016/j.tibs.2010.03.010

Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., et al. (2007). RNA polymerase is poised for activation across the genome. *Nature Genetics*, *39*(12), 1507–1511. http://doi.org/10.1038/ng.2007.21

Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., et al. (2003). Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Molecular and Cellular Biology*, *23*(6), 1863–1873.

Nechaev, S., & Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochimica Et Biophysica Acta*, *1809*(1), 34–45. http://doi.org/10.1016/j.bbagrm.2010.11.001

Nechaev, S., Fargo, D. C., Santos, dos, G., Liu, L., Gao, Y., & Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science (New York, N.Y.)*, *327*(5963), 335–338. http://doi.org/10.1126/science.1181421

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L. M., & Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, *457*(7232), 1038–1042. http://doi.org/10.1038/nature07747

Ni, Z., Saunders, A., Fuda, N. J., Yao, J., Suarez, J. R., Webb, W. W., & Lis, J. T. (2008). P-TEFb Is Critical for the Maturation of RNA Polymerase II into Productive Elongation In Vivo. *Molecular and Cellular Biology*, *28*(3), 1161–1170. http://doi.org/10.1128/MCB.01859-07

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., & Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells.

Nature Methods, *6*(12), 917–922. http://doi.org/10.1038/nmeth.1401

Parua, P. K., Booth, G. T., Sansó, M., Benjamin, B., Tanny, J. C., Lis, J. T., & Fisher, R. P. (2018). A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature*. http://doi.org/10.1038/s41586-018-0214-z

Parua, P. K., Booth, G. T., Sansó, M., Benjamin, B., Tanny, J. C., Lis, J. T., & Fisher, R. P. (n.d.). A Cdk9-PP1 kinase-phosphatase switch regulates the elongation-termination transition of RNA polymerase II. http://doi.org/10.1101/190488

Paule, M. R., & White, R. J. (2000). Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Research*, *28*(6), 1283–1298.

Pei, Y., & Shuman, S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *The Journal of Biological Chemistry*, *277*(22), 19639–19648. http://doi.org/10.1074/jbc.M200015200

Pei, Y., & Shuman, S. (2003). Characterization of the Schizosaccharomyces pombe Cdk9/Pch1 protein kinase: Spt5 phosphorylation, autophosphorylation, and mutational analysis. *The Journal of Biological Chemistry*, *278*(44), 43346–43356. http://doi.org/10.1074/jbc.M307319200

Pei, Y., Du, H., Singer, J., Stamour, C., Granitto, S., Shuman, S., & Fisher, R. P. (2006). Cyclin-dependent kinase 9 (Cdk9) of fission yeast is activated by the CDK-activating kinase Csk1, overlaps functionally with the TFIIH-associated kinase Mcs6, and associates with the mRNA cap methyltransferase Pcm1 in vivo. *Molecular and Cellular Biology*, *26*(3), 777–788. http://doi.org/10.1128/MCB.26.3.777-788.2006

Pei, Y., Schwer, B., & Shuman, S. (2003). Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. *The Journal of Biological Chemistry*, *278*(9), 7180–7188. http://doi.org/10.1074/jbc.M211713200

Pelechano, V., & Steinmetz, L. M. (2013). Gene regulation by antisense transcription. *Nature Reviews. Genetics*, *14*(12), 880–893. http://doi.org/10.1128/MCB.4.5.985

Pelechano, V., Chávez, S., & Perez-Ortin, J. E. (2014). A Complete Set of Nascent Transcription Rates for Yeast Genes (vol 5, e15442, 2010). *PloS One*, *9*(12). http://doi.org/10.1371/journal.pone.0115560

Pelechano, V., Wei, W., & Steinmetz, L. M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*, *497*(7447), 127–131. http://doi.org/10.1038/nature12121

Peng, J., Zhu, Y., Milton, J. T., & Price, D. H. (1998). Identification of multiple cyclin subunits of human P-TEFb. *Genes & Development*, *12*(5), 755–762.

Perales, R., & Bentley, D. (2009). “Cotranscriptionality”: the transcription elongation complex as a nexus for nuclear transactions. *Molecular Cell*, *36*(2), 178–191. http://doi.org/10.1016/j.molcel.2009.09.018

Peterlin, B. M., & Price, D. H. (2006). Controlling the elongation phase of transcription with P-TEFb. *Molecular Cell*, *23*(3), 297–305. http://doi.org/10.1016/j.molcel.2006.06.014

Petrenko, N., Jin, Y., Wong, K. H., & Struhl, K. (2016). Mediator Undergoes a Compositional Change during Transcriptional Activation. *Molecular Cell*, *64*(3), 443–454. http://doi.org/10.1016/j.molcel.2016.09.015

Phatnani, H. P., & Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes & Development*, *20*(21), 2922–2936. http://doi.org/10.1101/gad.1477006

Ping, Y. H. (2000). DSIF and NELF Interact with RNA Polymerase II Elongation Complex and HIV-1 Tat Stimulates P-TEFb-mediated Phosphorylation of RNA Polymerase II and DSIF during Transcription Elongation. *Journal of Biological Chemistry*, *276*(16), 12951–12958. http://doi.org/10.1074/jbc.M006130200

Ping, Y. H., & Rana, T. M. (2001). DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 Tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase II and DSIF during transcription elongation. *The Journal of Biological Chemistry*, *276*(16), 12951–12958.

Ponting, C. P. (2002). Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Research*, *30*(17), 3643–3652.

Porrua, O., & Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nature Reviews. Molecular Cell Biology*, *16*(3), 190–202. http://doi.org/10.1016/j.cell.2009.02.043

Proudfoot, N. J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. *Trends in Biochemical Sciences*, *14*(3), 105–110. http://doi.org/10.1016/0968-0004(89)90132-1

Proudfoot, N. J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science (New York, N.Y.)*, *352*(6291), aad9926. http://doi.org/10.1126/science.aad9926

Ptashne, M., & Gann, A. (1997). Transcriptional activation by recruitment. *Nature*.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842. http://doi.org/10.1093/bioinformatics/btq033

Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., & McCuine, S. (2010). c-Myc regulates transcriptional pause release. *Cell*.

Rasmussen, E. B., & Lis, J. T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(17), 7923–7927.

Rhee, H. S., & Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, *483*(7389), 295–301. http://doi.org/10.1038/nature10799

Robinson, P. J. J., Bushnell, D. A., Trnka, M. J., Burlingame, A. L., & Kornberg, R. D. (2012). Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(44), 17931–17935. http://doi.org/10.1073/pnas.1215241109

Rodriguez, C. R., Cho, E. J., Keogh, M. C., Moore, C. L., Greenleaf, A. L., & Buratowski, S. (2000). Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Molecular and Cellular Biology*, *20*(1), 104–112.

Rodríguez-Gil, A., García-Martínez, J., Pelechano, V., Muñoz-Centeno, M. de L. C., Geli, V., Pérez-Ortín, J. E., & Chávez, S. (2010). The distribution of active RNA polymerase II along the transcribed region is gene-specific and controlled by elongation factors. *Nucleic Acids Research*, *38*(14), 4651–4664. http://doi.org/10.1093/nar/gkq215

Rodríguez-Molina, J. B., Tseng, S. C., Simonett, S. P., Taunton, J., & Ansari, A. Z. (2016). Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization. *Molecular Cell*, *63*(3), 433–444. http://doi.org/10.1016/j.molcel.2016.06.036

Roeder, R. G., & Rutter, W. J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature*, *224*(5216), 234–237.

Rondón, A. G., García-Rubio, M., González-Barrera, S., & Aguilera, A. (2003). Molecular evidence for a positive role of Spt4 in transcription elongation. *The EMBO Journal*, *22*(3), 612–620. http://doi.org/10.1093/emboj/cdg047

Rosen, C. A. (1991). Regulation of HIV gene expression by RNA-protein interactions. *Trends in Genetics : TIG*, *7*(1), 9–14.

Rougvie, A. E., & Lis, J. T. (1988). The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell*, *54*(6), 795–804. http://doi.org/10.1016/S0092-8674(88)91087-2

Rougvie, A. E., & Lis, J. T. (1990). Postinitiation transcriptional control in Drosophila melanogaster. *Molecular and Cellular Biology*, *10*(11), 6041–6045.

Sansó, M., Lee, K. M., Viladevall, L., Jacques, P.-É., Pagé, V., Nagy, S., et al. (2012). A positive feedback loop links opposing functions of P-TEFb/Cdk9 and histone H2B ubiquitylation to regulate transcript elongation in fission yeast. *PLoS Genetics*, *8*(8), e1002822. http://doi.org/10.1371/journal.pgen.1002822

Saunders, A., Core, L. J., & Lis, J. T. (2006). Breaking barriers to transcription elongation. *Nature Reviews. Molecular Cell Biology*, *7*(8), 557–567. http://doi.org/10.1038/nrm1981

Schneider, S., Pei, Y., Shuman, S., & Schwer, B. (2010). Separable functions of the fission yeast Spt5 carboxyl-terminal domain (CTD) in capping enzyme binding and transcription elongation overlap with those of the RNA polymerase II CTD. *Molecular and Cellular Biology*, *30*(10), 2353–2364. http://doi.org/10.1128/MCB.00116-10

Schwer, B., & Shuman, S. (2011). Deciphering the RNA polymerase II CTD code in fission yeast. *Molecular Cell*, *43*(2), 311–318. http://doi.org/10.1016/j.molcel.2011.05.024

Schwer, B., Schneider, S., Pei, Y., Aronova, A., & Shuman, S. (2009). Characterization of the Schizosaccharomyces pombe Spt5-Spt4 complex. *RNA (New York, N.Y.)*, *15*(7), 1241–1250. http://doi.org/10.1261/rna.1572709

Scruggs, B. S., Gilchrist, D. A., Nechaev, S., & Muse, G. W. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Molecular Cell*.

Seet, B. T., Dikic, I., Zhou, M.-M., & Pawson, T. (2006). Reading protein modifications with interaction domains. *Nature Reviews. Molecular Cell Biology*, *7*(7), 473–483. http://doi.org/10.1038/nrm1960

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., et al. (2008). Divergent transcription from active promoters. *Science (New York, N.Y.)*, *322*(5909), 1849–1851. http://doi.org/10.1126/science.1162253

Shaw, P. A., Sahasrabuddhe, C. G., Hodo, H. G., & Saunders, G. F. (1978). Transcription of nucleosomes from human chromatin. *Nucleic Acids Research*, *5*(8), 2999–3012.

Shetty, A., Kallgren, S. P., Demel, C., Maier, K. C., Spatt, D., Alver, B. H., et al. (2017). Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Molecular Cell*, *66*(1), 77–88.e5. http://doi.org/10.1016/j.molcel.2017.02.023

Smale, S. T., & Kadonaga, J. T. (2003). The RNA polymerase II core promoter. *Annual Review of Biochemistry*, *72*, 449–479. http://doi.org/10.1146/annurev.biochem.72.121801.161520

Stargell, L. A., & Struhl, K. (1996). Mechanisms of transcriptional activation in vivo: two steps forward. *Trends in Genetics : TIG*, *12*(8), 311–315.

Steinmetz, E. J., Warren, C. L., Kuehner, J. N., Panbehi, B., Ansari, A. Z., & Brow, D. A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Molecular Cell*, *24*(5), 735–746. http://doi.org/10.1016/j.molcel.2006.10.023

Struhl, K., & Segal, E. (2013). Determinants of nucleosome positioning. *Nature Structural & Molecular Biology*, *20*(3), 267–273. http://doi.org/10.1016/j.cell.2012.04.036

Sun, J., & Li, R. (2010). Human negative elongation factor activates transcription and regulates alternative transcription initiation. *The Journal of Biological Chemistry*, *285*(9), 6443–6452. http://doi.org/10.1074/jbc.M109.084285

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., et al. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Research*, *22*(7), 1350–1359. http://doi.org/10.1101/gr.130161.111

Svaren, J., & Hörz, W. (1997). Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. *Trends in Biochemical Sciences*, *22*(3), 93–97.

Swanson, M. S., Malone, E. A., & Winston, F. (1991). SPT5, an essential gene important for normal transcription in Saccharomyces cerevisiae, encodes an acidic nuclear protein with a carboxy-terminal repeat. *Molecular and Cellular Biology*, *11*(8), 3009–3019.

Taube, R., Fujinaga, K., Wimmer, J., Barboric, M., & Peterlin, B. M. (1999). Tat transactivation: a model for the regulation of eukaryotic transcriptional elongation. *Virology*, *264*(2), 245–253. http://doi.org/10.1006/viro.1999.9944

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., et al. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Research*, *13*(9), 2129–2141. http://doi.org/10.1101/gr.772403

Tsankov, A. M., Gu, H., Akopian, V., Ziller, M. J., Donaghey, J., Amit, I., et al. (2015). Transcription factor binding dynamics during human ES cell differentiation. *Nature*, *518*(7539), 344–349. http://doi.org/10.1038/nature14233

Vannini, A., & Cramer, P. (2012). Conservation between the RNA polymerase I, II, and

III transcription initiation machineries. *Molecular Cell*, *45*(4), 439–446. http://doi.org/10.1016/j.molcel.2012.01.023

Venkatesh, S., & Workman, J. L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews. Molecular Cell Biology*, *16*(3), 178. http://doi.org/10.1038/nrm3941

Viladevall, L., St Amour, C. V., Rosebrock, A., Schneider, S., Zhang, C., Allen, J. J., et al. (2009). TFIIH and P-TEFb Coordinate Transcription with Capping Enzyme Recruitment at Specific Genes in Fission Yeast. *Molecular Cell*, *33*(6), 738–751. http://doi.org/10.1016/j.molcel.2009.01.029

Vojnic, E., Simon, B., Strahl, B. D., Sattler, M., & Cramer, P. (2006). Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. *The Journal of Biological Chemistry*, *281*(1), 13–15. http://doi.org/10.1074/jbc.C500423200

Vos, S. M., Pöllmann, D., Caizzi, L., Hofmann, K. B., Rombaut, P., Zimniak, T., et al. (2016). Architecture and RNA binding of the human negative elongation factor. *eLife*, *5*. http://doi.org/10.7554/eLife.14981

Wada, T. (1998). Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription invitro. *The EMBO Journal*, *17*(24), 7395–7403. http://doi.org/10.1093/emboj/17.24.7395

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & Development*, *12*(0890-9369/98), 343–356.

Weber, C. M., Ramachandran, S., & Henikoff, S. (2014). Nucleosomes Are Context-Specific, H2A.Z-Modulated Barriers to RNA Polymerase. *Molecular Cell*, *53*(5), 819–830. http://doi.org/10.1016/j.molcel.2014.02.014

Weiner, A., Hughes, A., Yassour, M., Rando, O. J., & Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Research*, *20*(1), 90–100. http://doi.org/10.1101/gr.098509.109

Wen, Y., & Shatkin, A. J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes & Development*, *13*(14), 1774–1779.

Werner, F. (2012). A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *Journal of Molecular Biology*, *417*(1-2), 13–27. http://doi.org/10.1016/j.jmb.2012.01.031

Werner, F., & Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the

three domains of life. *Nature Reviews. Microbiology*, *9*(2), 85–98. http://doi.org/10.1038/nrmicro2507

Wier, A. D., Mayekar, M. K., Héroux, A., Arndt, K. M., & VanDemark, A. P. (2013). Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(43), 17290–17295. http://doi.org/10.1073/pnas.1314754110

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., et al. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, *453*(7199), 1239–1243. http://doi.org/10.1038/nature07002

Williams, L. H., Fromm, G., Gokey, N. G., Henriques, T., Muse, G. W., Burkholder, A., et al. (2015). Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Molecular Cell*, *58*(2), 311–322. http://doi.org/10.1016/j.molcel.2015.02.003

Winston, F., Chaleff, D. T., Valent, B., & Fink, G. R. (1984). Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. *Genetics*, *107*(2), 179–197.

Wolffe, A. P., & Matzke, M. A. (1999). Epigenetics: regulation through repression. *Science (New York, N.Y.)*, *286*(5439), 481–486.

Wong, K. H., Jin, Y., & Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. *Molecular Cell*, *54*(4), 601–612. http://doi.org/10.1016/j.molcel.2014.03.024

Workman, J. L., & Kingston, R. E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annual Review of Biochemistry*, *67*, 545–579. http://doi.org/10.1146/annurev.biochem.67.1.545

Wu, C.-H., Yamaguchi, Y., Benjamin, L. R., Horvat-Gordon, M., Washinsky, J., Enerly, E., et al. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes & Development*, *17*(11), 1402–1414. http://doi.org/10.1101/gad.1091403

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., et al. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, *457*(7232), 1033–1037. http://doi.org/10.1038/nature07728

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., & Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Molecular Cell*, *21*(2), 227–237. http://doi.org/10.1016/j.molcel.2005.11.024

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., et al. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, *97*(1), 41–51.

Yu, M., Yang, W., Ni, T., Tang, Z., Nakadai, T., Zhu, J., & Roeder, R. G. (2015). RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science (New York, N.Y.)*, *350*(6266), 1383–1386. http://doi.org/10.1126/science.aad2338

Yudkovsky, N., Ranish, J. A., & Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature*, *408*(6809), 225–229. http://doi.org/10.1038/35041603

Zattas, D., & Hochstrasser, M. (2015). Ubiquitin-dependent protein degradation at the yeast endoplasmic reticulum and nuclear envelope. *Critical Reviews in Biochemistry and Molecular Biology*, *50*(1), 1–17. http://doi.org/10.3109/10409238.2014.959889

Zhang, C., Kenski, D. M., Paulson, J. L., Bonshtien, A., Sessa, G., Cross, J. V., et al. (2005). A second-site suppressor strategy for chemical genetic analysis of diverse protein kinases. *Nature Methods*, *2*(6), 435–441. http://doi.org/10.1038/nmeth764

Zhang, D. W., Rodríguez-Molina, J. B., Tietjen, J. R., Nemec, C. M., & Ansari, A. Z. (2012). Emerging Views on the CTD Code. *Genetics Research International*, *2012*(9), 1–19. http://doi.org/10.1046/j.1365-2443.2002.00522.x

Zhou, K., Kuo, W. H. W., Fillingham, J., & Greenblatt, J. F. (2009). Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(17), 6956–6961. http://doi.org/10.1073/pnas.0806302106