January 2017

Molecular Mechanisms Regulating Rna Polymerase Ii Pausing During Gene Activation

Ann Sanoji Samarakkody

MOLECULAR MECHANISMS REGULATING RNA POLYMERASE II PAUSING DURING GENE ACTIVATION

by

Ann Sanoji Samarakkody
Graduateship in Chemistry, 2009

A Dissertation
Submitted to the Graduate Faculty

of the

University of North Dakota

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota
May
2017

This dissertation, submitted by Ann Sanoji Samarakkody in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is here by approved.

Sergei Y. Nechaev, Chairperson

Bryon D. Grove

Archana Dhasarathy

Joyce E. Ohm

Diane Darland

This dissertation is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Grant McGimpsey, PhD
Dean of the School of Graduate Studies

March 30, 2017
Date

PERMISSION

Title: Molecular mechanisms regulating RNA Polymerase II pausing during gene activation

Department: Biomedical Sciences, Program of Anatomy and Cell Biology

Degree: Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work, or in her absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Ann Sanoji Samarakkody
March 21, 2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There are many individuals to thank who were a tremendous influence in my life. To begin I would like to thank Sergei who has made me a better scientist over the course of 5 years. When I first joined the lab, I had very little knowledge in molecular biology but, he took me under his wing and helped me learn, grow as a professional and also to think like one. He has a true passion for science and is the kind of person who brightens up when talking about science. When my experiments had failed me, his fantastic ability to troubleshoot always kept me motivated. The countless times he has spoken with me, corrected my abstracts and write ups and listened to me has made me a proficient and confident scientist. I also like to thank Dr. Archana Dhasarathy for teaching me numerous things starting from western blots. She is a fantastic teacher and is the most organized person I've seen! I would also like to thank the members of my graduate advisory committee Dr. Bryon Grove, Dr. Joyce Ohm and Dr. Diane Darland for their guidance, advice and feedback through the course of my graduate career.

I'm very pleased to recognize all the faculty and graduate students of the previous department of Anatomy and Cell Biology. I would also like to Dr. Jane Dunlevy, the graduate advisor for the previous Department of Anatomy and Cell Biology for advising me when I was fresh on US soil about everything from

planning my course work to proper winter attire. I also thank Dr. Jon Jackson and Dr. Haris Ali for having my back through the daunting task of having to learn and teach Gross Anatomy and also for making it a fun adventure. I also acknowledge the administrative assistants especially, Bonnie Kee, Lisa Brown and Jennifer Hershey for their assistance through the years. I especially thank Bonnie for helping me through the daunting task of formatting the thesis.

I have been extremely fortunate to have worked with wonderful and brilliant colleagues: Nii Koney-Kwaku Koney; with his assistance in cell fractionations and depletions, Chuka Oscar Nnoli; for his work on the artificial bubble template, Dr. Ata Abbas and Sayantani Ghosh Dastidar. All of you have been great lab-mates through all these years. Thank you for assisting me in numerous ways and for your insights and suggestions to better my project. I also acknowledge the Dhasarathy lab; especially Jessica Warns for her guidance in data analysis and being a wonderful optimistic friend over the years. I wish all of you the very best. I would also like to thank Adam Scheidegger at the UND Epigenetics Bioinformatics Core for his constant tremendous support in analysis of genome-wide data. I also thank Oscar Nnoli for his work at the Core for performing quality controls and sequencing libraries. I would also like to thank my friends who are dear to my heart both here in the US and in Sri Lanka.

I was fortunate to be under very enthusiastic educators and exceptional mentors. I very dearly acknowledge all teachers of Good Shepherd Convent Kandy and Kandy Girls' High School and also all faculty of College of Chemical

Sciences, Institute of Chemistry Ceylon. I very dearly thank Dr. Priyani Paranagama for enkindling the passion for research in me.

My Family has always been the driving force behind my success. My parents Ann and Eric Karunarathne, you taught me the value of hard work, determination and perseverance. You are the reason I am the person I am today and I will always be grateful to you. To my loving sister Judy Karunarathne thank you for being my lifelong friend and making me strive for the best. To my grandparents Rita and Merril Dias for raising me with good values. My loving daughter Samiddi Amaara; you have been an incredibly wonderful baby since the day you were born. You have taught me to smile more and cherish every moment. Your smile and warm hugs always help me on occasional gloomy days in research. To my loving husband Nuwan; you have been a great companion and my best friend through the years. You always trusted me more than I did and never doubted that I was not capable of what I set out to do. You always have a great sense of reasoning that I always rely on. Your hard work and sacrifices have led me to success and this wouldn't have been a success without your support.

Last but not the least, I thank Almighty God for His blessings and His wondrous plan for me.

To Nuwan

ABSTRACT

RNA Polymerase II (Pol II), the enzyme that transcribes all messenger RNAs (mRNAs), has another activity by pausing at gene promoters. The paused Pol II generates short 5'-capped RNAs of about 50 nucleotides in length. Much is known about Pol II pausing including its prevalence across the genome, but the molecular mechanisms that are involved in regulating Pol II pausing and its roles in gene regulation are yet to be fully explored. In this study, I have investigated the molecular mechanisms that regulate dynamics of Pol II pausing in response to gene activation.

The main goal of this study was to determine how cells respond to stimuli by altering Pol II pausing states. To this end, we characterized changes in Pol II pausing in MCF-7 human breast cancer cells using two distinct stimuli, heat shock and Trichostatin A (TSA), a histone deacetylase inhibitor. Two genes, *SNAI1*, a master regulator of epithelial to mesenchymal transition, and *HSPA1B* (heat shock protein 70), a master regulator of heat shock response, show mRNA upregulation upon treatment with both stimuli. To determine changes in Pol II pausing and its dynamics in response to gene activation, I have used permanganate footprinting for single gene analysis, and short-capped RNA sequencing (scRNA-seq) for genome-wide analysis. I have shown that, upon activation, paused genes can retain Pol II pausing and non-paused genes can acquire Pol II pausing. Further, I observe that genes such as *HSPA1B* undergo

pause release during heat shock activation but *SNAI1* does not. In addition, I have shown that the turnover of the paused Pol II complex changes in a stimulus-specific manner, indicating that the release of paused Pol II is sensitive to the nature of the stimulus. To investigate the rate of turnover of the paused complex, I used a specific inhibitor of a general transcription factor TFIIH, Triptolide, and measured the residence time of the Pol II complex at the paused site. Genome-wide analysis of Pol II turnover demonstrates that not all genes respond the same way to heat shock. This project describes a novel mechanism for regulation of transcription during gene activation in human cells in response to environmental stresses.

CHAPTER I

INTRODUCTION

1.1 Eukaryotic Transcription

Gene expression in eukaryotes is a tightly controlled, complex process regulated by numerous transcription factors and enzymes. Chromatin-associated proteins such as histones and non-histone proteins are involved in compacting the genome, which is about 2m long in humans, into a confined space in the 5 μM nucleus. Dynamic interactions of chromatin with a variety of cellular factors such as histone modifying enzymes and histone chaperones, ATP dependent chromatin remodelers, activate and/or repress gene expression.

Eukaryotic transcription is a precisely timed process where multiple proteins come together to synthesize RNA from appropriate genes. All messenger RNA (mRNA) in eukaryotes are transcribed by the multi-subunit enzyme RNA Polymerase II (Pol II). Transcription by Pol II occurs via three major steps: transcription initiation, elongation and termination. Recruitment of Pol II to promoters (transcription initiation) was long considered as the main regulatory step of transcription in higher organisms.

1.2 Transcription Initiation

Transcription is initiated at the promoter DNA elements, which involves sequential binding of general transcription factors (GTFs), TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Matsui et al. 1980), together with Pol II, forming the

pre-initiation complex (PIC) (Figure 1). The formation of PIC is also dependent on chromatin remodeling, which involves modifications of histones in the promoter region (Kadonaga 2004). It has been suggested that histone modifications and other chromatin marks play a regulatory role in highly transcribing genes (Vakoc et al. 2005). The binding motifs of GTFs promote the assembly of the PIC by facilitating DNA-protein and protein-protein interactions. Once the PIC is established, Pol II together with other transcription factors (Pol II complex) can proceed with transcription. There are multiple events during early transcription that play regulatory roles. Pausing of Pol II and premature termination of transcription in promoter-proximal regions are two of such steps that will be discussed later in this chapter.

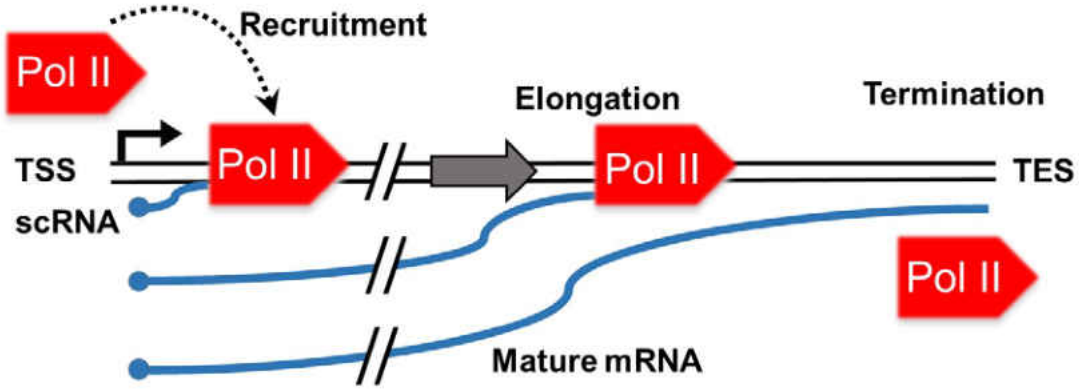

Figure 1. A schematic diagram showing an overview of transcription. Pol II gets recruited on to the promoter region along with other general transcription factors to initiate transcription. Transcription is initiated at the transcription start site (TSS), indicated by a black arrow. After transcribing about 50 nucleotides from the TSS, Pol II pauses thus, producing a short capped RNA. Once paused, Pol II can proceed into elongation. During the elongation step the mRNA molecule gets synthesized. Once the Pol II complex reaches the transcription end site (TES), it undergoes termination dissociating from the DNA template and releasing the mRNA molecule.

1.3 Elongation

Apart from initiation, elongation also plays a critical step in gene regulation. The carboxyl terminal domain (CTD) is the largest subunit of Rpb1, the largest subunit of Pol II, undergoes extensive modification during elongation. The CTD contains tandem heptad repeats of a consensus sequence of amino acids -YSPTSPS that is evolutionary conserved across eukaryotes. Yeast has 24 repeats on the CTD of Pol II whereas humans have 52 repeats. The serine residues of the CTD remain unphosphorylated during recruitment but undergo a series of modifications post recruitment or Pol II to the promoter. The serine-5 residues on CTD are phosphorylated by the cyclin dependent kinase 7 (CDK7), a subunit of TFIIH, during early elongation and pausing. Serine-5 phosphorylation is assumed to destabilize the interaction of Pol II with promoter elements and enhance the interactions with elongation factors, facilitating elongation. Transition of the Pol II complex to active elongation occurs with the recruitment of the kinases that phosphorylate the serine-2 residues on CTD. The serine-2 phosphorylated-elongating Pol II complex is capable of recruiting factors that are involved in chromatin remodeling and mRNA processing (Peterlin et al. 2006). The transition of the Pol II from paused state to the elongating state requires the kinase activity of p-TEFb (Marshall and Price 1992, Marshall and Price 1995). A second kinase comprised of cyclin K and Cdk12 or Cdk13 is also involved in phosphorylation of CTD in the 3'-end of the gene (Bartkowiak et al. 2010; Greifenberg et al. 2016). Hence, phosphorylation of the CTD plays a role in determining the status of Pol II complex during transcription.

Another feature that changes along the gene body is the Pol II elongation rate. The correlation between characteristics of the gene body and the rate at which the polymerase is able to move through the gene has been extensively studied. Elongation rates have been measured to be ranging from 1-4 kilobases per minute (Ardehali et al. 2009; Hah et al. 2011). The rate of transcription has been shown to vary depending upon the region of the gene: the elongation rate is slowest at promoter regions and is increased across the gene body (Danko et al. 2013). A recent study showed clear differences in elongation rates in response to two different stimuli, TNF-α and Estradiol (E-2), and an increased Pol II signal was measured at exons. However, this study was unable to demonstrate a common trend in Pol II elongation rate in exons (Danko et al. 2013). In mouse mesenchymal stem cells, the elongation rate was found to be negatively correlating with CpG content and exon content and positively with the active transcription mark of H3K79me2 (Jonkers et al. 2014). This evidence highlights both pausing and elongation as steps that are clearly regulated during transcription.

Pol II elongation is tightly linked to chromatin modifications. For instance, the Pol II associated factor (PAF) complex is physically associated with elongating Pol II and recruits histone H2B ubiquitinating enzymes (Pokholok et al. 2002; Wood et al. 2003). Also, the histone chaperone Facilitates Chromatin Transcription (FACT) histone chaperone complex is required for nucleosome assembly and is involved in recruitment of p-TEFb (Saunders et al. 2003; Wada et al. 2000).

The rate of Pol II elongation has been proposed to influence co-transcriptional processes such as splicing (Howe et al. 2003; de la Mata et al. 2003; Shukla & Oberdoerffer 2012), 3′ end processing (Nag et al. 2007) and termination (Hazelbaker & Buratowski 2012).

1.4 Transcription Termination

After completion of mRNA synthesis, Pol II dissociates from DNA at the transcription end site (Figure 1). This final step in transcription allows recycling of Pol II molecules for new initiation events. Most eukaryotic genes carry a highly conserved poly A signal at their 3' ends. Once the polymerase transcribes the poly A tail, the processivity of the enzyme reduces markedly. In humans, several complexes are known to facilitate transcription termination. The ser-2 hyper-phosphorylated CTD of Pol II is involved in recruiting several protein complexes that facilitate transcription termination such as cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulatory factor (CstF) and Poly (A) polymerase (Kuehner et al. 2011). Both CPSF and CstF complexes are recruited to predominantly ser-2 Phosphorylated CTD repeats of Pol II. In addition, CPSF has been shown to bind to AAUAAA, the mammalian Poly A signal, and slows down the Pol II complex to trigger termination of transcription at the 3' termini of genes. Termination factors are found proximal to transcription end sites. However, recent studies have shown occupancy of such factors close to promoter proximal regions leading to exciting discovery of premature termination in promoter regions, which will be discussed in 1.7 (Figure 2).

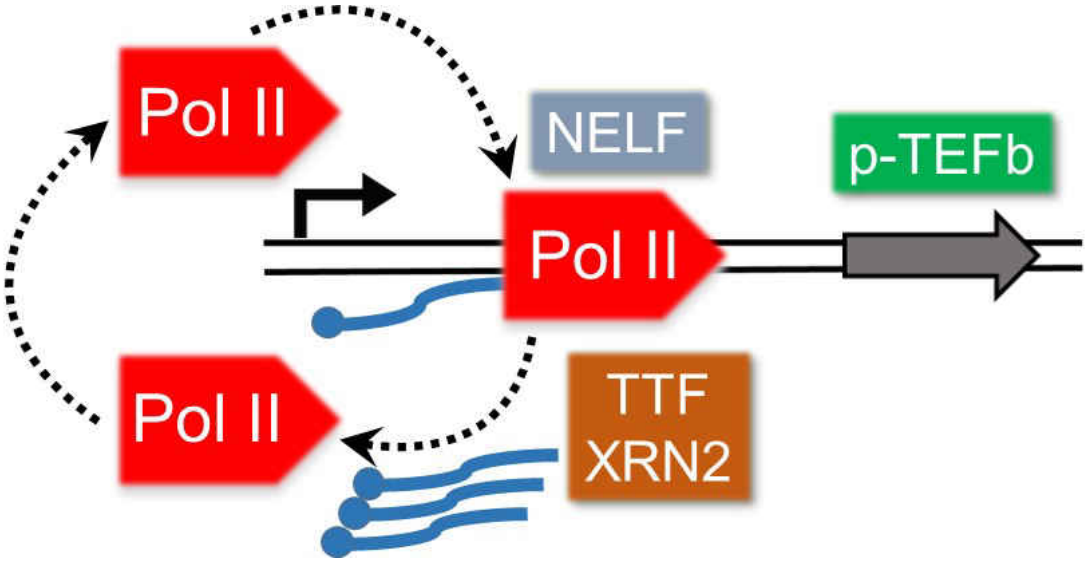

Figure 2. A scheme of pausing of RNA Polymerase II as a central mechanism in regulating gene transcription. The dashed arrows indicate the processes that occur at promoter proximal region. Once Pol II (red pentagon) is recruited and starts initiation (the solid black arrow indicates the transcription start site), it transcribed about 50 nucleotides and pauses at most mRNA gene promoters in the presence of NELF (blue). Once paused Pol II can either proceed into elongation (grey arrow) with the action of factors such as p-TEFb or dissociate from the DNA template undergoing premature termination in the presence of factors such as XRN2 (brown). The short RNA and Pol II is dissociated from the DNA template during premature termination where Pol II can initiate more cycles of transcription.

1.5 Processing of mRNA during Transcription

In eukaryotes, mRNA that is produced by Pol II undergoes several modifications. Most of these modifications occur co-transcriptionally and the three main processing steps are 5' capping, splicing of introns and addition of a poly A tail at the 3' end of the RNA.

1.5.1 Capping of Pol II Transcripts

5' capping is exclusive to transcripts of Pol II and is the first modification on the nascent mRNA. The cap structure is an essential feature of mRNA that

allows them to be recognized by translational machinery in the cytoplasm. During the early stages of transcription, the CTD domain of Pol II is hyper-phosphorylated at the ser-5 residue. A 7-methylguanosine cap (7 mG) structure is added to the nascent mRNA soon after initiation, during the transcription of the first 25-30 nucleotides (Jove & Manley 1984; Rasmussen & Lis 1993) by recruitment of capping enzymes (McCracken et al. 1997). The process of capping of the 5' end of the RNA is an intricate three step process (Mizumoto & Kaziro 1987).

In addition to 7 mG, multiple different cap structures have been observed predominantly on RNA that are much shorter, less than 200 nucleotides, than mRNA and also in virus RNAs. Some small nuclear RNA (snRNA) and small nucleolar RNA have been shown to contain a trimethylated version of the 7 mG cap structure; (2,2,7 mG) (Banerjee 1980). Some viral RNA species have been shown to contain 2,7 mG, a dimethylated cap structure (HsuChen & Dubin 1976). The capped structure has been shown to stimulate splicing of the nascent RNA in vitro (Konarska et al. 1984; Edery & Sonenberg 1985) and in vivo (Fresco & Buratowski 1996; Schwer & Shuman 1996). The capped structure recruits cap binding proteins (CBP) 20 and 80 that are involved in splicing and transport of mRNA to the cytoplasm. In addition, the capped structure can be recognized by protein synthesis machinery of the cell. 5' capped RNA binds to translation initiation factor eIF4E, promoting translation (Sonenberg et al. 1978; Edery et al. 1988; Fresco & Buratowski 1996). Further, antibodies against these specific cap structures can be used to immunoprecipitate the RNAs (Bochnig et al. 1987) that

can be subjected to microarray analysis or sequencing in downstream applications (Jia et al. 2007). The capped structure enables identifying transcripts of Pol II, which allows quantification of RNA through multiple methods including scRNA-seq (Nechaev et al. 2010; Samarakkody et al. 2015) and Pro-cap-seq (Kwak et al. 2013; Mahat et al. 2016). These genome-wide techniques allow visualizing Pol II transcription by specifically detecting RNAs that are protected at the 5' end by a cap structure.

1.5.2 Splicing

During transcription elongation, the intronic regions of the nascent RNA are removed and the exons are joined to generate a mature mRNA in a process termed splicing. Splicing allows extensive diversity of mRNA and proteins in a cell to arise from the same sequence of the gene (Wang et al. 2008; Warns et al. 2016). About 95% of protein coding genes in the cell can undergo alternative splicing (Pan et al. 2008; Wang et al. 2008). Committed factors involved in splicing are recruited to perform co-transcriptional splicing of the nascent pre-mRNA (Ardehali & Lis 2009) and this process involves extensive DNA-RNA, DNA-protein and RNA-protein interactions. General splicing factors and small nuclear ribonucleoproteins (snRNPs) interact with proteins that are specific to certain DNA sequences, to form a multisubunit complex, the spliceosome. The variations in splicing patterns are closely related to the chromatin architecture and elongation rate of Pol II along the gene body (de la Mata et al. 2003; Fong et al. 2014). Alternative splicing has been observed in many biological processes such as development, cancer (Srebrow & Kornblihtt 2006) and also during

cellular stresses (Pleiss et al. 2007). Large switching in splicing patterns during progression of cancer generates protein variants that function during changes in cell morphology and transcriptional programs (Warzecha & Carstens 2012). Among cellular stresses, splicing has been shown to be abolished in response to heat shock but not chemical stresses (Yost & Lindquist 1986; Bond 1988; Biamonti & Caceres 2009), and can be rescued by the presence of elevated levels of heat shock proteins (Yost & Lindquist 1986). Chemical stresses also induce heat shock proteins. However, the splicing patterns and mechanisms are vastly different between the two responses. The relations of variation in splicing with different stimuli to early transcriptional event are yet to be explored.

1.5.3 Polyadenylation

Most mRNAs have a tail with multiple adenosine (A) residues at their 3' termini, commonly referred to as the poly A tail. PolyA polymerase has been identified as the enzyme involved in adding A nucleotides to the 3' termini (Darnell et al. 1971; Edmonds et al. 1971; Lee et al. 1971). Most mRNAs contain a highly conserves hexanucleotide sequence, AAUAAA, approximately 10-30 bases upstream of the polyadenylation site. There's a second GU-rich region, approximately 20-40 bases downstream of the transcription end site. These two sequences and the distance between them not only determine the site of poly A addition (MacDonald et al. 1994; Takagaki & Manley 1997), but also the length of the poly A tail (Chen et al. 1995). The poly A tail stabilizes the mature mRNA as it is transported to the cytoplasm to be translated.

1.6 Promoter Proximal RNA Polymerase II Pausing

Although recruitment of Pol II and formation of the PIC are essential regulatory steps in transcription, another major step of regulation of transcription is promoter proximal Pol II pausing. After transcribing about 50 nucleotides from the transcription start site, Pol II stalls generating a small RNA with a 5' cap (Figure 2). Pausing was first characterized on heat shock proteins, mainly on HSP 70 of *Drosophila melanogaster*, when transcriptionally active polymerase was found within a narrow region at the gene promoter prior to the induction of the gene, immediately downstream of the transcription start site (Gilmour & Lis 1986). Following this work, Pol II pausing was described in other cell lines derived from higher organisms and on other genes, implying that Pol II pausing is a widespread phenomenon among metazoans. Genome-wide studies on Pol II pausing in mammals and *Drosophila* (reviewed in Adelman & Lis 2012) have indicated that pausing occurs specifically at genes involved in developmentally related pathways (Muse et al. 2007; Levine 2011), stimulus response, maintenance of pluripotency and more recently at many metabolic genes (Williams et al., 2015). Even though the exact biological function of Pol II pausing remains largely unknown, its importance in regulating gene expression is no longer questioned. Others have shown that Pol II may compete with the nucleosome at the +1 position to maintain the promoter in an open configuration (Gilchrist et al. 2008). Some alternative hypotheses that have been proposed are that pausing functions as a checkpoint for the 5' end processing of RNA and also ensures proper formation and modifications on the Pol II complex itself (Adelman

& Lis 2012). Promoter proximal pausing of Pol II has been shown to accompany histone H3K4me3 in embryonic stem cells, implying the association of pausing with active genes (Chopra et al. 2011).

A number of different factors have been identified to regulate various stages of pausing such as establishment of pausing, escape of pausing and more recently premature termination of paused Pol II complexes. The transition from the paused state to productive elongation is closely related to the changes in phosphorylation status of the CTD of the largest subunit of Pol II.

1.7 Regulation of RNA Polymerase II Pausing

Pol II pausing peaks between 35-50 nts from transcription start sites in *Drosophila* (Nechaev 2010), mouse (Core 2014) and human (Samarakkody et al. 2015) cells (Figure 2). The similarity in the location of paused Pol II in different systems implies conservation of pausing location across species. The biological systems that are being employed have vast differences in promoter sequences. Hence, an obvious question of the sequence dependency of Pol II pausing is not supported by much evidence in human systems. Transgene experiments in *Drosophila* identified both the distal and proximal regions of the *Hsp70* gene as promoting Pol II pausing and mutations in those sequences decreased levels of paused Pol II (Lee et al. 1992). Mammalian genes display prominent regions of repetitive CG nucleotides, also known as CpG islands, around their promoter regions, whereas *Drosophila* genes lacks such a feature (Hoskins et al. 2011). The presence of CpG islands brings in diverse proteins and protein complexes that are involved in DNA methylation at the cytosine nucleotides of a CG pair.

DNA methylation brings about changes in promoter architecture; deacetylation of histone H4 and methylation of histone H3 Lysine 9 (Hashimshony et al. 2003). Such changes in chromatin structure cause Pol II pausing to be different in promoter regions across species (Core et al. 2008; Rozenberg et al. 2008). The conservation of location of Pol II pausing across species is a consequence of multiple regulatory proteins and complexes that act at promoter proximal regions of genes irrespective of the DNA sequence. Among the most important problems is to understand how factors involved in Pol II pausing, elongation and premature termination integrate to produce a specific outcome; either a mature mRNA through productive elongation or short transcript through termination. Collectively, the ability of paused Pol II to take two routes makes it, potentially, a decision-making point of gene expression during early elongation. The question remains whether pausing at steady state is a rigid property of a gene or a dynamic property reflective of a cell state.

1.7.1 Establishment of Pol II Pausing through Negative Elongation Factor (NELF)

Pausing at promoter regions is enforced by two multisubunit complexes: 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor (DSIF), which contains Spt4a and Spt5, and negative elongation factor (NELF, containing subunits A-E). These two complexes bind through C/D subunits of NELF complex (Sun and Li, 2010). The five subunits of the NELF complex are thought to act synergistically since depletion of individual subunits leads to diminished levels of the others (Gilchrist et al. 2008; Sun and Li, 2010). The E subunit of the NELF complex contains arginine-aspartic acid dipeptide repeats

and has been shown to directly interact with RNA (Yamaguchi et al. 2002). A more recent study showed that the NELF complex may have at least 3 different RNA-binding domains (Vos et al. 2016), which may enable NELF to facilitate exchange of different RNA species competing either directly with each other or acting allosterically. Interaction of NELF with enhancer RNA (eRNA) has been shown to promote its release on early response genes and trigger release of paused Pol II into elongation (Schaukowitch et al. 2014). However, understanding of the functions of NELF and eRNA is limited thus far.

NELF is an essential factor and its depletion has shown to be embryonically lethal (Amleh et al. 2009). In vitro, NELF does not appear to bind to Pol II or DSIF alone but it is able to bind the Pol II-DSIF complex (Yamaguchi et al. 2002). Apart from the DRB complex, subunits of NELF has been shown to interact with multiple proteins and complexes in the nucleus such as the cap binding complex (CBC) of RNA (Narita et al. 2007) and the NELF-B subunit has been shown to associate with breast cancer susceptibility gene BRCA1 (Ye et al. 2001).

Given the role of NELF in inducing Pol II pausing, depletion of NELF is expected to increase expression of genes due to removal of pausing as a roadblock. However, studies show that depletion of NELF, contrary to expectations, leads to decreased expression of many genes in *Drosophila*, mouse and human systems (Gilchrist et al. 2008; Rahl et al. 2010; Sun & Li 2010). Moreover, NELF depletion decreases stability of paused Pol II (Henriques et al. 2013), yet these complexes do not go into productive elongation in most

genes (Core et al. 2012). NELF depletion in *Drosophila* cells was shown to affect recovery from heat shock (Ghosh et al. 2011). Thus, NELF, or pausing, clearly have roles other than gene repression, and remain to be investigated.

1.7.2 Escape of Pausing into Elongation: Positive Transcription Elongation Factor b

Positive elongation factor b (p-TEFb) alleviates the transcription halt caused by of DSIF and NELF by phosphorylating them and allowing Pol II to enter into elongation (Figure 2) (Renner et al. 2001; Yamaguchi et al. 1999). p-TEFb is a protein kinase that targets phosphorylation of ser-2 residues of Pol II CTD (Price 2000) and is comprised of cyclin dependent kinase 9 (Cdk9) and either Cyclin T1 (CycT1), Cyclin T2 (CycT2) or Cyclin K (CycK). p-TEFb is sequestered by the evolutionary conserved non-coding RNA, 7SK small nuclear RNA (snRNA) where the p-TEFb binding to the noncoding RNA is mediated by HEXIM1 or HEXIM2, and the formation of this complex inhibits the kinase activity of p-TEFb, thereby limiting the transition of promoter proximally paused Pol II into productive elongation. There are a number of factors that modulate Pol II pausing by regulating the localization of p-TEFb on a gene. Regulated release of p-TEFb has been shown to modulate global control of cell growth, proliferation and cell differentiation (He et al. 2006). A Yamanaka factor c-myc, an essential factor in embryonic stem cell renewal and proliferation(Cartwright et al. 2005), has been shown to bind p-TEFb in tumor cells (Gargano et al., 2007; Kanazawa et al., 2003), and is involved in pause release, thus acting as a transcriptional activator (Rahl et al. 2010). NF-kB is another transcriptional activator that recruits p-TEFb to genes. Bromodomain containing protein 4 (BRD4) acts as an upstream

activator by mediating the acetylated histone dependent recruitment of p-TEFb (Gregoire & Yang 2005) to gene promoters. In addition, factors such NF-κB and c-myc that act primarily on initiation, have been found to promote pause escape into productive elongation (Barboric et al. 2001; Rahl et al. 2010).

In multiple previous studies, p-TEFb has been inhibited using small molecule inhibitors including DRB (Chodosh et al. 1989) and Flavopiridol (FP). FP is a highly potent inhibitor of CDK9 (Chao & Price 2001; Ni et al. 2004) that leads to overall suppression of transcription genome-wide (Ni et al. 2008; Jonkers et al. 2014) and increase in promoter proximal Pol II in majority of genes (Henriques & Adelman 2013).

1.7.3 A Second Exit for Paused Pol II: Premature Termination

Once paused, instead of proceeding into elongation, Pol II complex can dissociate from the DNA template, terminating transcription abruptly (Figure 2). This premature termination of transcription yields a very short RNA transcript that is less than 100 nucleotides in length. With emerging evidence of occurrence of premature termination in biological systems, the molecular mechanisms and the action of candidate cofactor are still being unraveled. Recent studies have shown evidence of premature termination of transcription at the promoter proximal regions in yeast (Terzi et al. 2011), in an HIV model (Wagschal et al. 2012) and in human cell lines (Brannan et al. 2012). The mechanisms described in the studies are not similar, which suggests that premature termination may be an inherent feature of a promoter. For example, the Nrd1-Nab3-Sen1 termination pathway described in yeast indicates that H3K4 trimethylation promotes

termination. Other studies describe how termination factors such as TTF2 and a 5’ to 3’ exoribonuclease XRN-2 are recruited to promoter-proximal regions of genes to initiate termination. Consistent with a functional connection between termination and promoter proximal pausing, NELF has been shown to promote recruitment of a factor PCF11, another termination factor, to promoter-proximal regions in Human Influenza Virus (HIV (Natarajan et al. 2013). Taken together, premature termination likely involves distinct molecular mechanisms that can act through multiple proteins. How termination pathways are regulated during gene activation remains to be explored.

1.8 Pause Duration

The conservation of location of Pol II pausing implies that similar mechanisms are involved in determining the dynamics of Pol II pausing. DSIF, NELF and p-TEFb have been identified to play a role in determining the residence time of Pol II at the paused site, which is an important measure of gene or promoter activity. Recent work suggests that additional factors, such as Gdown1, the product of the *POLR2M* gene that renders Pol II responsive to Mediator, and the general transcription factor TFIIF may also influence the pause duration of the paused polymerase. These factors act by affecting the vulnerability of the paused Pol II complex to undergo premature termination (Cheng et al. 2012). Therefore, pause duration recapitulates the action of the factors that are involved in maintaining and releasing paused Pol II. Yet, the underlying molecular mechanisms remain unclear.

To determine the residence time of the paused Pol II complex, the small molecule inhibitor, Triptolide (TRP), has been used. TRP is an XPB/TFIIH inhibitor (Titov et al. 2011) that prevents the generation of newly engaged Pol II during transcription initiation by inhibiting melting of the DNA template. A time course of TRP treatment enables determination of the half-life of the paused Pol II complex. Published work has used a variety of methods such as Chromatin Immunoprecipitation (ChIP) Sequencing (ChIP-seq), Global run-on sequencing (Gro-seq), high resolution permanganate footprinting and optical imaging. The pause duration of *Hsp70* gene on *Drosophila* polytene chromosomes was estimated at 5 minutes using an optical method. The strategy involved a GFP-tagged Pol II and mCherry LacI bound *Hsp70* transgene (Buckley et al. 2014). In mouse embryonic stem cells the pause duration averaged at 7 minutes (Jonkers et al. 2014). In another study in *Drosophila* S2 cells, pause duration was estimated for 12 genes where the pause duration exceeded 15 minutes on some genes (Henriques et al., 2013). A recent study using HCT116 cells showed that some genes, when treated with low TRP concentration, had prolonged pause duration, which exceeded one hour in the case of some genes like the early response gene FOS (Chen et al. 2015). The functional significance of Pol II pausing is to allow sufficient time to integrate regulatory inputs into transcription and (Henriques et al., 2013) allowing proteins to interact and undergo post transcriptional modifications.

The above-described dedicated factors help determine how fast Pol II is transitioning through promoter regions. The activity of such factors causes

alterations in pause duration. For example, absence of NELF causes direct reduction in pause duration that enable fast promoter escape of upregulated genes (Gilchrist et al. 2008). How these factors integrate during rapid cellular responses, such as stress and other environmental stimuli, to enable survival of the cell remains to be understood.

1.9 Research Direction

1.9.1 Transcriptional Responses during Cellular Stresses

Cellular responses to stress are evolutionarily ubiquitous and play essential roles in survival of the cell and the organism (Lindquist 1986; Feder & Hofmann 1999). A cell can be exposed to many stresses from the environment such as osmotic pressure, change in pH, change in temperature, availability of nutrient content and heavy metals etc. Such suboptimal conditions trigger responses such as modifications on enzymes, alteration in metabolism, vesicular trafficking but one of the most upstream effects is change in gene expression. Gene expression changes are associated with changes in chromatin architecture; changes in nucleosome structure and positioning, regulation of specific transcription factors, and activation or repression of transcription on specific genes.

A cell responds to stress by activation of specific stress response pathways through activity of transcription factors and upregulation of specific genes. The ability to change specific biological pathways in response to stimuli enhances the chance of survival of a cell. Defects in various stress response pathways have been shown to associate with diseases such as cancer, diabetes

(Oyadomari et al. 2002), Alzheimer's disease (Terro et al. 2002), Parkinson's disease (Imai et al. 2001), etc. Defects in the heat shock response pathway compromised early development in mice (Xiao et al. 1999), and overexpressing the heat shock transcription factor in *Caenorhabditis elegans* results in increased longevity (Garigan et al. 2002; Lund et al. 2002).

1.9.2 Gap in Knowledge

The dynamics of pausing and pause release remains largely unknown. It is intriguing to look into the dynamics of pausing and pause escape of the Pol II complex specifically during rapid transcriptional responses. This will determine if the molecular mechanisms involved are common across the genome, or if they are gene specific. While multiple studies have shown pausing and pause duration to possibly be gene specific, the mechanisms involved in such regulation remain to be characterized. In this study, I have explored pausing and pause duration in MCF-7, a human breast cancer cell line, in response to two activating signals, Trichostatin A (TSA) and heat shock. TSA is a compound that has been shown to upregulate mRNA expression of the *SNAIL* gene (*SNAI1)* in MCF-7 cells. *SNAI1* is a master regulatory gene that is important in epithelial to mesenchymal transition (EMT). In addition, I have used *SLUG (SNAI2)* and E-cadherin (*CDH1)* genes which are also involved in EMT. The second activator, heat shock, has been extensively used in the field, as it was the first system where Pol II pausing was first observed. This is the first time two activators were tested on the same gene and further, the first time that dynamics of two genes are compared in the same cells.

1.9.3 EMT

EMT is a common biological process that occurs during embryonic development. During EMT cells migrate during several major milestones in development including gastrulation and neural crest formation where epithelial cells transition to a more mesenchymal morphology. During EMT, vast changes occur in the extracellular matrix, plasma membrane and cell morphology. During EMT, matrix metalloproteinases that degrade the extracellular matrix and cell surface proteins leading to release of promigratory factors making cells more migratory and invasive (Derycke & Bracke 2004). The major characteristic of EMT is loss of the cell junction protein *CDH1* on the plasma membrane. *SNAI1* has been described as a repressor of multiple target genes in human and non-human cells lines including cancer cells. In human cancer stem cells, *SNAI1* is shown to regulate the expression of a number of genes including IL-8 to induce metastasis in cells (Hwang et al. 2011). Overexpression of *SNAI1* has been associated with recurrence of breast cancer (Moody et al. 2005) and hence much attention is drawn towards *SNAI1* to understand its role in cancer metastasis.

In addition to its effects on chromatin, TSA has been reported to increase reactive oxygen species in a cell via kinase pathways (Rosato et al. 2003; Ruefli et al. 2001). Moreover, TSA is known to activate several signaling pathways that induce cellular stress such as NF-κB, MAPK and PI3K (Mayo et al. 2003; Ozaki et al. 2006; Yu et al. 2007; Ruefli et al. 2001). Taken together, TSA may act as an agent that induces cell stresses following short-term treatment.

1.9.4 Heat Shock Response

The specific protein factors that are involved in heat shock response (HSR) are broadly studied. HSR is an essential part of cell homeostasis that ensures cell survival. Upregulation of heat shock genes provides the cell with temporary enhanced tolerance to stress (Lindquist 1986; Sørensen et al. 2003). HSR is regulated by transcription factors heat shock factors (HSF). Vertebrates have 4 transcription factors HSF1-4, but HSF-1 is considered the master regulator of heat shock response. During HSR, HSF-1 trimerizes and binds to the heat shock element (HSE) in the promoter region of genes such as *Hsp70* which results in gene upregulation by increasing promoter proximal pause release (Mahat et al. 2016).

The key to transcriptional activation during HSR is the association of p-TEFb with HSF-1 (Lis et al. 2000). Not all genes are activated in a HSF-1 dependent manner during heat shock. Hence, an interesting question would be if genes that are activated by HSF-1 dependent and independent pathways show different dynamics of pausing at gene promoters. More specifically, changes in rates of premature termination and elongation are affected during gene activation will be reflected in measurements of pause duration. Therefore, using pause duration as a measurement we can determine if the mechanisms that act at the pause site are global or specific to a cohort of genes. Secondly, we will be able to determine if premature termination and escape into elongation act on the same genes in a stimulus specific manner.

1.9.5 Research Goal

The regulatory importance of Pol II pausing is not questioned in the field. Yet, the roles that paused Pol II plays during transcriptional responses remain to be explored. Many proteins are being identified as playing a role in Pol II regulation, but how these proteins contribute to produce specific transcriptional outcomes is still understudied.

In this study, I have first investigated levels of Pol II pausing on genes during basal and activated conditions in response to two stimuli, TSA and heat shock, in MCF-7 cells. Secondly, I have looked at Pol II pausing duration as a measure of dynamics of the paused Pol II complex during gene activation to determine if the molecular mechanisms that are involved during gene activation are common among genes. Finally, using sc-RNA sequencing I have investigated levels of Pol II pausing and changes in pause duration genome-wide to determine the global dynamics of Pol II pausing upon gene induction.

CHAPTER II

MATERIALS AND METHODS

2.1 Cell Culture and Treatments

The cell line used in this study, MCF-7, is a human breast cancer cell line derived from a metastatic site by pleural effusion of a 65 year old female. MCF-7 cells were obtained from American Type Culture Collection (ATCC® HTB-22™) and were cultured in DMEM/F-12 medium (Life Technologies) with 10% fetal bovine serum (Gibco). The growth medium was replaced every three days. Cells were maintained in a humidified incubator at 37ºC and 5% CO2.

Heat shock was carried out at 42ºC for 30, 60 and 120 minutes. Cells were treated with Trichostatin A (Sigma) at a final concentration of 1μM for 3 hours and beta estradiol (Sigma) at 10 nM for 1 hour in phenol red free DMEM/F-12 supplemented with 10% charcoal dextran treated, heat inactivated FBS (Atlanta Biologicals) and human insulin, human transferrin and sodium selenite (ITS) (Sigma). Treatments with Triptolide (TRP) (Sigma) and Flavopiridol (FP) (Sigma) were carried out after the above-described activations. TRP was used at a final concentration of 2.5 μM and FP at 5 μM.

2.2 RNA Extraction

Total RNA was extracted from cells using the Qiashredder and RNeasy kit (Qiagen). 500ng of RNA was run on a 1% agarose gel with TAE buffer to check the integrity of each sample. 1 μg of total RNA was used in reverse

transcription with SSRTIII reverse transcriptase (Invitrogen) with random hexamer (Invitrogen) priming. The cDNA was then subjected to quantitativereal time PCR using gene specific primers and all samples were normalized against *GAPDH* gene. All data points are averages of three biological replicates with standard error of the mean. Once validated, the RNA was checked again for integrity and sent for RNA sequencing.

2.3 Permanganate Footprinting

Treatment of cells with potassium permanganate and all subsequent reactions were performed as described previously for mammalian cells (Samarakkody et al. 2015). All treatments with potassium permanganate were done on freshly treated cells after washing with PBS buffer. Cells pellets were resuspended in 50 µL PBS buffer, and then treated with 300 µL of ice cold 20 mM $KMnO_4$ for 1 minute and the reaction was immediately stopped with 400 µL of stop solution (20 µL of 5M NaCl, 100 µL of 1M Tris-HCl pH 7.4, 300 µL of β-mercaptoethanol (Sigma), 400 µL of 500 mM EDTA and 500 µL of 10% SDS). The cell lysate was treated with 3 µL of RNA cocktail (Ambion) at 37ºC for 30 minutes and Fungal Proteinase K (Life Technologies) overnight at 56ºC. 6 µg of purified DNA was treated with piperidine (Sigma) in 100 µL and extracted twice with chloroform. 250 ng of piperidine treated DNA was used in primer extension with a gene specific primer A using Phusion high fidelity polymerase (Thermo Scientific) and Hight Fidelity (HF) buffer. Subsequently the single stranded DNA were ligated to a double stranded linker A':B overnight at 16ºC in the presence of T4 DNA Ligase (ThermoScientific). The DNA was amplified after ethanol

precipitation with gene specific primer B and universal linker A primer for 22 cycles and an additional two cycles of PCR with ^{32}P radioactively labeled gene specific primer C. The reaction was phenol-chloroform extracted, ethanol precipitated and run on a denaturing 7% urea TBE gel.

2.4 Immunoblot Assay

Cellular extracts were prepared in lysis buffer (8M urea and 1M Tris HCl pH 8.0) with protease inhibitor cocktail (Sigma). Total cell extracts were run on a 10% SDS-PAGE resolving gel alongside full-range rainbow marker (GE Healthcare Life Sciences) and transferred onto a PVDF membrane using standard technique of wet transfer under chilled conditions. The blot was then blocked in 5% milk in TBS buffer with 0.1% tween for one hour and immunoblotted against anti-RNA Polymerase Ser-2 (Millipore), anti-Actin (Life Technologies), anti-cMyc antibody (Abcam ab32072) overnight at 4 ºC. The blots were then incubated for one hour with the horse radish peroxidase (HRP) conjugated secondary antibody (GE Healthcare Life Sciences) at a dilution of 1:10,000. The blot was then treated with Luminata Forte Western HRP substrate (EMD Millipore) for 5 minutes before imaging on the Odyssey imager.

2.5 Depletion using siRNA

siRNA sequences for XRN-2 and NELF were selected through the Silencer Select tool by ThermoFisher Scientific and those for CDK9 was used as previously published and CDK-9 antibody (A303-493A-M) from Bethyl Laboratories (Yu et al. 2015). The siRNA was diluted in a 1:1 mix of

Lipofectamine-3000 and Opti-MEM to a final concentration of 10 nM. The cells were plated at a density of 60% and grown for 72 hours.

2.6 In vitro Synthesis of Capped RNA

DNA sequences were designed based on the mouse genome with 30-75 nucleotides in length, having a G nucleotide as the start with the T7 promoter sequences at the 5' end. The designed template sequence was to ensure that it did not match to the human genome using the BLAT tool on the UCSC genome browser (Kent 2002)The whole DNA sequence together with the T7 promoter sequence was considered as the forward sequence which would serve as the sense strand during subsequent in vitro transcription. The forward and the complimentary reverse strands were mixed at a final concentration of 10 µM in TE buffer in a total of 100 µL and annealed with a program ramping down from 95^{o}C to 4^{o}C. The reaction was then precipitated with 3 volumes of ethanol and 6 µL of 5M NaCl. Using 1 µg of the annealed DNA template the in vitro transcription reaction was set up using the T7 Megascript kit (Ambion) as recommended and incubated overnight at 37^{o}C. The reaction was then treated with 1 µL of TURBO DNase as recommended and the reaction was stopped by adding the ammonium acetate solution as recommended. Then the RNA was extracted using acid-phenol/chloroform and then again with chloroform before precipitating with 3 times the volumes of ethanol. Then the RNA was then precipitated at -80^{o}C for at least 30 minutes and pelleted using a cold centrifuge at maximum speed for 20 minutes. The RNA pellet was washed with 80% ethanol, air dried to remove residual ethanol, and resuspended in 20 µL of water.

The RNA was then mixed with an equal volume of 2X gel loading dye provided with the Megascript kit and heated at 65ºC for 5 minutes. The RNA was then loaded on a 15% Urea-TBE gel that was pre-run for 20 minutes alongside the single stranded RNA marker (NEB) and run at 200 V until the dye front reaches the bottom of the gel. The band that corresponds to RNA of the expected length was extracted from the gel and extracted using the crush and soak method, where the gel pieces were placed in a gel breaker tube (Agilent) that's placed in a 1.5 mL Eppendorf tube and spun at 12000g for 5 minutes. The crushed gel was then shaken at room temperature in 400 µL of 300 mM NaCl for 4 hours. The gel homogenate was then spun through a 0.22 µM filter column (Agilent) for 5 minutes at 2000g. The flow though was then precipitated with 3 volumes of ethanol and incubated in -80ºC for 30 minutes before spinning at maximum speed in a cold centrifuge to pellet the RNA. The RNA was washed in 80% ethanol and resuspended in 20 µL of nuclease free water. A microliter each of this RNA was quantified using both the Nanodrop and the Bioanalyzer instruments.

A 7-methylaguanylate capped structure was added to the 5' end of these RNA using the Vaccinia capping system (NEB). The RNA was first heated at 65ºC for 5 minutes and immediately placed on ice. Then to 15 µL of the gel extracted RNA was combined with 2 µL of 10X capping buffer, 1 µL 10mM GTP, 1 µL of 2 mM SAM after dilution and 1 µL of Vaccinia capping enzyme was added and incubated at 37ºC for one hour. The capped RNA was then extracted using acid-phenol/chloroform as described above.

The efficiency of capping of each capped RNA was then checked using the following four reactions where 200 ng of RNA was used in each reaction:

Reaction A: Treatment with 5' phosphate dependent RNA polyphosphatase (Epicentre)

Reaction B: Treatment with 5' phosphate dependent RNA polyphosphatase (Epicentre) followed by Terminator exonuclease (Epicentre)

Reaction C: Treatment with RNA 5' Pyrophosphohydrolase (NEB)with followed by Terminator exonuclease (Epicentre)

Reaction D: Treatment with RNA 5' Pyrophosphohydrolase (NEB) alone

These four reactions were performed for each in vitro RNA and then was mixed with equal volume of gel loading buffer II (Ambion). The RNA was then loaded directly on a 15% Urea-TBE gel that was pre-run for 20 minutes and the gel was run at 200 V. A successfully capped RNA will be retained after all reactions except for reaction D, and if RNA is not reliably capped, there will be a decrease in reactions B and D. The capped RNA was diluted such that when added to the nuclei or chromatin in Trizol, the in vitro-synthesized RNA was in the range of 10-50 molecules per cell.

2.7 Short Capped RNA Library Preparation

MCF-7 cells from different treatment points were used in the library preparation using approximately 10^8 cells in two 15 cm dishes at about 90% confluency. Cells were harvested and washed with 1X PBS buffer prior to nuclei isolation.

2.7.1 Nuclei Isolation

Cells were directly resuspended in 2 mLs of the cell lysis buffer (Williams et al. 2015) containing 10 mM Tris-HCl (pH 8.0), 200 mM Sucrose, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.1% Triton X-100, SUPERase In RNase Inhibitor (Ambion), and Proteinase Inhibitor Cocktail (Sigma-Aldrich). The cell lysate was incubated on ice for 10 minutes and further lysed in a douncer with 20 strokes. The lysate was centrifuged at 1000 g in a cold centrifuge for 10 minutes to precipitate the nuclei and the nuclei was washed to remove traces of cytoplasm. 5% of the cell fractions were used in immunoblotting and qRT PCR analyses to confirm the reliability of fractionation.

2.7.2 Chromatin Isolation

Cells from two 15 cm dishes at about 80% confluency were scrapped in media and washed twice in cold PBS. The cells were resuspended in 10 mL of swelling buffer (10 mM TRIS-Cl pH 7.5, 2 mM $MgCl_2$ and 3 mM $CaCl_2$) and allowed to swell on ice for 5 minutes. Centrifuge at 2000g for 10 minutes and then remove the supernatant. Resuspend cell in 1 mL of Lysis buffer (10 mM TRIS-Cl pH 7.5, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 0.5% Igepal, 10% Glycerol, 2 units of Superase In) and gently pipette 20 times to lyse the cells. The lysate was brought up to 10 mL and the nuclei pelleted by centrifuging at 2000 g for 10 minutes. The isolated nuclei were resuspended in 125 µL of NUN1 buffer (20 mM TrisHCl pH 8.0, 75 mM NaCl, 0.5 mM EDTA, 50% Glycerol and Proteinase Inhibitor 100x (Sigma)), followed by 1.2 mL of NUN2 buffer (20 mM HEPES-KOH pH 7.6, 7.5 mM $MgCl_2$, 0.2 mM EDTA, 300 mM NaCl, 1M Urea, 1% NP40 and Proteinase

Inhibitor 100x (Sigma)). The nuclei were incubated for 15 minutes on ice for complete lysis while vortexing for 5 seconds every 4 minutes at maximum speed. The chromatin was pelleted by centrifugation at 13,000g for 10 minutes in a cold centrifuge.

2.7.3 Isolation of short RNA from Chromatin and Nuclei Fractions

Nuclei or chromatin pellets were resuspended in Trizol to extract RNA with subsequent chloroform extraction and ethanol precipitation. The RNA was resuspended in nuclease free water, treated with DNase I (NEB) and incubated at 37ºC for 30 minutes. Then the reaction mixture was mixed with equal volume of 2X gel loading buffer (Invitrogen) and heating at 65ºC for 5 minutes. The samples were loaded 20 µ g RNA per well on a 15% urea-TBE gel 10-well 1 mm thick gel (Invitrogen), alongside short range single stranded RNA ladder (NEB), after pre-running for 20 minutes at 200 V. A section with short RNAs between 20 to 80 nucleotides in length was cut from the gel and the RNA was extracted using crush and soak method (Nechaev et al. 2010). The gel pieces were crushed by placing them in gel breaker tubes (Agilent) inserted in a 1.5 mL Eppendorf tube and spinning in a bench top centrifuge at 12000g for 5 minutes. Then 400 µL of 300 mM NaCl was added to the crushed gel to extract the RNA by shaking for 4 hours at room temperature. The gel slurry was then transferred to 0.22 µ M spin filter tubes (Agilent) and spun at 2000 g for 5 minutes. The RNA was pelleted with 3 volumes of ethanol and 1 µ L of GlycoBlue co-precipitant (Ambion) by incubating in -80 ºC for at least 30 minutes before spinning down at maximum speed in a chilled centrifuge for 20 minutes. The

precipitated RNA washed with 80% ethanol and air dried prior to resuspending in 20 µ L of nuclease free water. One microliter each was used in quantification by nanodrop and bioanalyzer 2100 (Agilent).

2.7.4 Library Preparation with small capped RNA

Approximately 700 ng of nuclear RNA and 300 ng of Chromatin bound RNA was used in the library preparation protocol. 15 µL of size selected small RNA was first denatured at 65ºC and treated with 1 µL of T4 Polynecleotide Kinase- 3' phosphate minus (NEB) in the presence of 2 µ L of 10X T4 Polynucleotide Kinase buffer and 1 µL of Murine RNase Inhibitor (NEB) and 1 µL of 10 mM ATP and incubated at 37ºC for 30 minutes. This reaction would allow phosphorylation of any degradation products at their 5' end of the RNA. The reaction was made up to 100 µL with nuclease free water (Ambion) and extracted with 50 µL of acid-phenol/chloroform pH 4.5 with isoamyl alcohol, then with 50 µL of chloroform and precipitated with 3 volumes of ethanol and 6 µL of 5M NaCl and precipitated as described above. The RNA was resuspended in 17 µL of nuclease free water and denatured at 65ºC and immediately placed on ice. 2 µL of 10X reaction buffer for 5' RNA phosphate dependent polyphosphatase and 1 µL of 5' RNA phosphate dependent polyphosphatase (Epicentre) was added to the RNA and incubated at 37ºC for 30 minutes. The reaction was precipitated as described above and resuspended in 17 µ L of nuclease free water and denatured at 65ºC as described above. Then the RNA was treated with 1 µL of Terminator Exonuclease (Epicentre) in the presence of 2 µ L of 10X reaction buffer and incubated at 30ºC for an hour and there after the reaction was

precipitated as described above. The terminator exonuclease degrades RNA with a 5' monophosphate in a 5' to 3' direction.

Before the next step of ligating the 3' adapter to the RNA, the adapter sequence was preadenylated using the 5' DNA adenylation kit (NEB) as described in the NEB protocol. After the reaction was extracted with phenol/chloroform with isoamyl alcohol (pH 7.5) (Sigma) and with Chloroform (Sigma) and precipitated in the presence of 6 µL 5M NaCl with 3 volumes of ethanol. The reaction was run through Bio-Spin 6 column (Biorad) as per manufacturer's instructions to remove remaining ATP.

After the reaction with Terminator Exonuclease the RNA was resuspended in 7 µL of nuclease free water and mixed 1 µL of preadenylated 3' RNA adapter and incubated at 70°C for 2 minutes and placed on ice immediately. To the mix, 2 µL of T4 RNA Ligase buffer and 1 µL of T4 RNA Ligase 2, truncated K225Q (NEB) and incubated at 28°C for 1 hour and 20°C for two hours. Then 10 µl of 2X Gel loading buffer II was added, heated at 65°C for 5 minutes and loaded on a 15% Urea Polyacrylamide gel after pre-running for 20 minutes at 200 V, alongside low range single stranded RNA ladder (NEB). Gel selection corresponding to the RNA ligated to the 3' adapter, 100-40 nucleotides was excised and extracted from the gel using the cut and soak method as previously described. The adapter ligated RNA was then resuspended in 18 µL of nuclease free water and incubated at 37°C for 30 minutes with 1 µL of Murine RNase Inhibitor and 1 µL of Shrimp alkaline phosphatase. The RNA was extracted with acidic phenol/chloroform and ethanol precipitated as previously described. The

RNA was then resuspended in 16 µL of nuclease free water and incubated at 37ºC for 2 hours with 2 µL of 10X NEB Thermopol Buffer and 2 µL of RNA 5' Pyrophosphohydrolase and extracted with phenol/Chloroform followed by ethanol precipitation. RNA was then resuspended in 6 µL of nuclease free water and denatured at 70ºC for 2 minutes. The second ligation reaction for the RNA was set up with 1 µL 10mM ATP, 1 µL of RNA 5' adapter, 1 µL of Murine RNase Inhibitor and 1 µL T4 RNA Ligase 1 (NEB). The ligation reaction was incubated at 28ºC for an hour and 20ºC for 2 hours. After the ligation of the 5' adapter the reaction was directly used in reverse transcription. 1 µl of 50 mM TruRT primer, 5X First strand synthesis buffer (Invitrogen), 0.1 mM DTT (Invitrogen), 12.5 mM dNTPs and SSRTII (Invitrogen) was added directly to the reaction tube after ligation and incubated at 50ºC for one hour. After reverse transcription, 10 µL of 5X Phusion buffer, 2 µL of TruRTP primer and 2 µL of TruTPI indexed primer was used in PCR with high fidelity Phusion polymerase (ThermoFisher Scientific) The thermal program used was as follows; initial denaturing step of 94ºC for 30 seconds, 16 cycles of 94ºC for 15 seconds, 62ºC for 30 seconds and 70ºC for 15 seconds and a final extension at 70ºC for 5 minutes.

Each library was mixed with 10 µL of 6X purple gel loading dye (NEB) after the PCR amplification and directly loaded in two wells of a 6% non-denaturing polyacrylamide TBE 1mm thick gel (Invitrogen), alongside 25 base pair DNA ladder (NEB) and run at 100 V until the dye front reached the bottom of the gel. The library between 200-150 base pairs was cut from the gel and extracted with 400 µL of 300 mM NaCl using the crush and soak method.

The final library was resuspended in 20 µL of nuclease free water and a microliter was used in both Nanodrop and Bioanalyzer quantification. Prior to sequencing the inserts of libraries were validated by ligating about 30 ng of the gel selected purified library to pBlueScript vector, that was dephosphorylated, using T4 DNA ligase (ThermoFisher) followed by transforming the ligated DNA into DH5-α *Escherichia coli* competent cells. Plasmids from 12-15 individual clones with inserts were sequenced and the inserts were confirmed using the BLAT tool on the UCSC genome browser (Kent 2002). The scRNA libraries were sequenced on the Illumina MiSeq sequencer at the UND Epigenetics and Bioinformatics Core using the RNA LT option in 50 base pair reads in a paired end mode, using V3 150 cycle MiSeq kit.

2.8 Data Analysis

2.8.1 RNA-seq Analysis

The fastq files from the RNA-seq data were aligned to the human genome (hg19) using topthat:

tophat2 -p 5 -o <folder_name> --transcriptome-index=transcriptome_data/ -G genes.gtf genome sample.fastq.gz 2> ./ sample.stderr

The output sam file was converted to bam format, sorted, indexed and used in R to generate a list of genes with differential expression using custom scripts:

samtools sort sample.bam -o sample.BAM

samtools index sample.BAM -o sample.bai

The differential expression was determined at p value of 0.01 using 2 replicates for each time point of heat shock.

2.8.2 scRNA-seq Analysis

The paired end mode generates 2 fastq files containing the sequenced reads; read 1 (R1) and read 2 (R2). Due to the RNA insert is fairly short, during sequencing the adapters also get sequenced. Hence, the sequences were first trimmed for the adapter sequences using cutadapt (Martin 2011). The adapter trimmed files were then aligned with the hg19 human reference genome from UCSC genome browser using the version 1.1.1 of bowtie (Langmead et al. 2009). The alignment was done with 2 allowable mismatches and the reads that had non-unique alignments to the genome ware disregarded (Scheidegger et al. 2015). The unaligned sequences were output into a separate fastq file and generated a bowtie file using the following command:

```
bowtie -p 6 -q -m 1 -n 2 --un sample_un.fastq
/local_storage/annotation_db/Homo_sapiens/UCSC/hg19/Sequence/BowtieIndex/genome sample sample.bowtie
```

From the aligned file of bowtie, an unbinned bedgraph file was generated for visualization on the UCSC genome browser using the perl script bowtie2bedgraph script as published (Scheidegger et al. 2015).

```
bowtie2bedgraph.pl sample sample /local_storage/annotation_db/Homo_
sapiens/UCSC/hg19/hg19_chrominfo.txt
```

R1 reported the 5’ end of the RNA, corresponding to the transcription start site, and the R2 the 3’ end, equivalent to the pause site for short RNAs coming from

paused Pol II. For each sample of the scRNA fastq file, four bedgraph files were generated separating the forward and reverse strand for each R1 and R2 in the paired end format; R1_forward, R1_reverse, R2_forward and R2_reverse. These bedgraph files were normalized using the counts of in vitro synthesized spike-in RNAs in each sample, assuming that RNA losses of those RNAs through the processes of RNA extraction and library preparation across samples are the same. Spike-in RNA counts for each spike-in sequence was counted in each fastq file using the following command where <sequence> would be the actual sequence of the spike-in RNA:

```
grep -A 2 -B 1 <sequence> sample.fastq | sed '/--/d' > output.fastq.gz
zcat output.fastq.gz | echo $((`wc -l`/4))
```

The normalization for bedgraphs were performed by normalizing the counts for all the spike-ins across samples.

CHAPTER III

RNA POLYMERASE II PAUSING IS RETAINED DURING GENE ACTIVATION

3.1 Pol II Pausing at Basal Levels is Positively Correlated with Levels of Transcription

The levels of Pol II pausing on different genes vary greatly across the genome as shown by multiple genome-wide studies in both *Drosophila* and mammalian systems (Zeitlinger et al. 2007; Muse et al. 2007). This prompts the question whether Pol II pauses at basal conditions to allow rapid activation in response to a stimulus. Recent evidence shows that genes that lack Pol II pausing at basal levels are activated just as fast as genes that has significant levels of Pol II pausing prior to induction (Lin et al. 2011; Gilchrist et al. 2012; Chen et al. 2013). Moreover, genome-wide analysis of mammalian transcription rates revealed that induction of paused genes in response to TNF-α in fact shows a slight delay in activation (Danko et al. 2013). In *Drosophila* embryos, Pol II pausing is present on many genes that are not subsequently activated during development (Levine 2011). This evidence suggests that rapid activation is unlikely to be the main reason for Pol II pausing. Another hypothesis was that Pol II pausing may result in more synchronous activation of genes following a stimulus (Boettiger & Levine 2009) and perhaps, at basal conditions as well. To determine if pausing is coupled with productive transcription, we compared expression levels of EMT related genes to levels of Pol II pausing in a panel of

human cancer lines, selected based on their varying expression of *SNAI1* at steady state conditions; MCF-7 (human breast epithelial cancer cell line), A375 (human epithelial melanoma cell line), MDA-MB-231 (human breast epithelial adenocarcinoma cell line), T47D (human mammary ductal carcinoma cell line) and A549 (human lung carcinoma cell line). We used potassium permanganate footprinting to detect Pol II pausing on promoters of three genes, Snail (*SNAI1*), Slug (*SNAI2*) and E-Cadherin (*CDH1*) in comparison to their mRNA expression levels. Potassium permanganate footprinting is based on a previously published protocol (Gilmour & Fan 2009) that we further optimized for the analysis of mammalian cells. Noteworthy are several features of pausing: first, permanganate reactivity for each gene is present at the same location across cell lines (Figure 3), suggesting that the mechanisms that regulate Pol II pausing in these cell lines are similar. Second, Pol II pausing signal is higher with higher levels of steady state gene expression, thus highlighting a fundamental relationship between pausing and transcription in the absence of activation. The differences in levels of pausing and transcription across cell lines are, perhaps, a result of differences in chromatin architecture, which determines how accessible the promoter regions are.

3.2 Pol II Pausing levels are increased with activation of the *SNAI1* gene

SNAI1 gene is one of the highly paused genes in MCF-7 but is expressed at low levels in MCF-7 cells (Dhasarathy et al. 2007), indicating that perhaps *SNAI1* is poised for activation. Using several activating signals, we followed

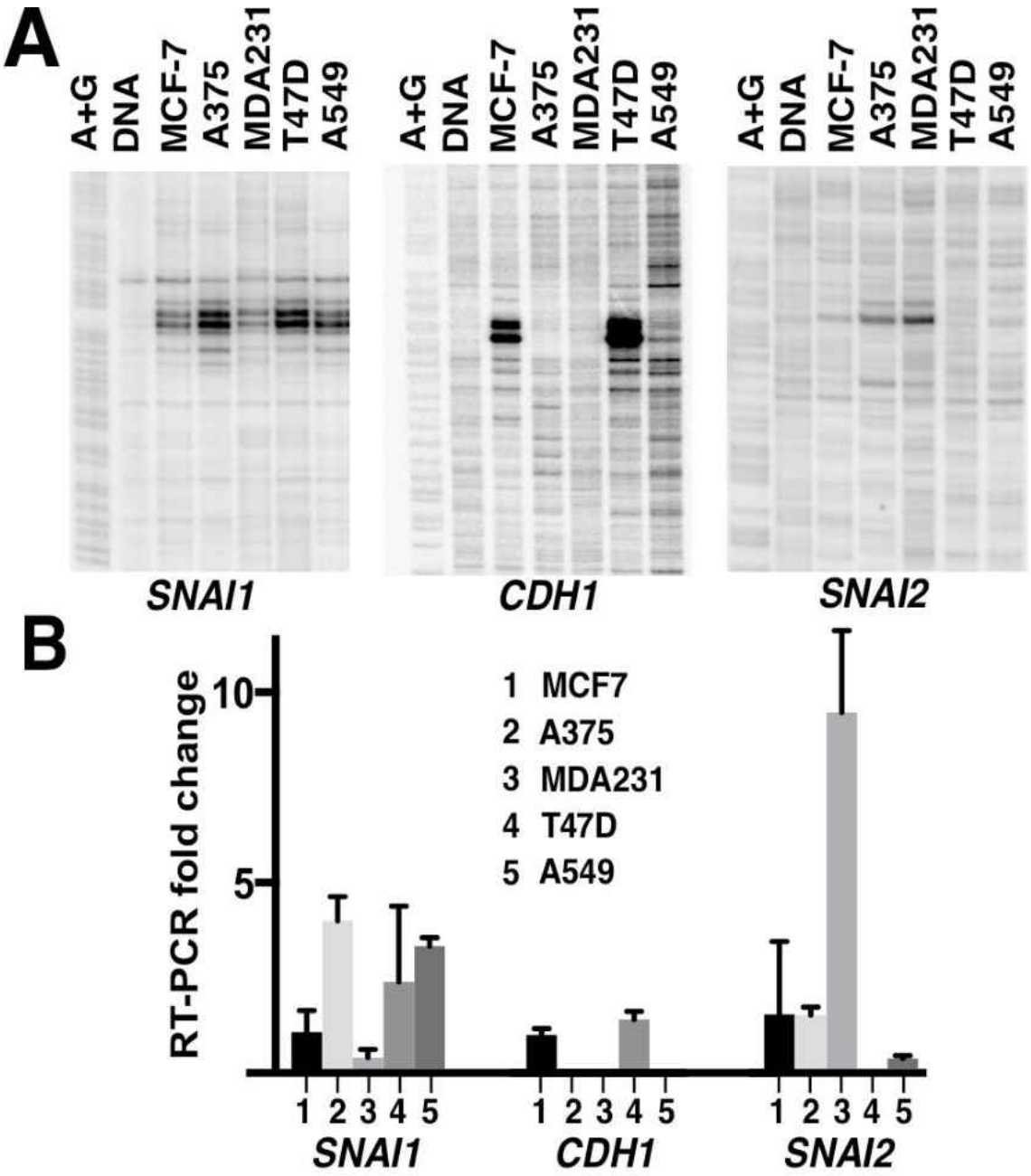

Figure 3. Levels of Pol II pausing is associated with basal levels of transcription independent of activation. (A) Permanganate footprints for *SNAI1*, *CDH1* and *SLUG* genes across different human breast cancer cell lines demonstrate that high Pol II pausing signal correlates with high levels of basal expression. (B) mRNA levels analyzed in the same cell lines as shown in A. All values derive from three independent experiments, and are plotted based on Ct difference between the indicated gene and GAPDH.

changes in Pol II pausing status on *SNAI1* upon activation. We treated MCF-7 cells with 17β-estradiol (E2), a natural ligand of estrogen receptor alpha that was previously shown to cause transient upregulation of *SNAI1* in MCF-7 cells (Hah et al. 2011), and followed a time course of transcriptional response (Figure 4A). While MCF-7 cells showed an increase in *SNAI1* RNA levels relative to *GAPDH* control gene, we did not observe a change in the location of pausing with activation, consistent with what was previously observed (Hah et al. 2011). Further, the Pol II pausing signal did not change and, specifically, did not decrease during the time course of gene activation as would be expected for pause release (Figure 4A), consistent with previous data using Global Run-on sequencing (Gro-seq) (Hah et al. 2011). However, the effect of E2 on MCF-7 cells is short-lived as demonstrated by another group (Danko et al. 2013). We tested another system, a well-studied system of EMT that has been previously shown to upregulate *SNAI1* expression. We treated NMuMg,(Namru murine mammary gland epithelial cells) with transforming growth factor beta (TGF-β). *SNAI1,* a gene that is not paused in mouse cell lines gains Pol II pausing upon treatment with TGF-β (Figure 4B). We then investigated the effects of the compound Trichostatin A (TSA), a histone deacetylase inhibitor, that activates *SNAI1*. Similar to activation with E2, activation of *SNA1* gene transcription is followed by an increase in Pol II pausing signal, which with TSA is the highest at 3 hours (Figure 4C). This response is not unique to MCF-7 cells, as *SNAI1* mRNA also gets upregulated in response to TSA treatment of MCF-10 A (non-tumorigenic human breast epithelial cell line).

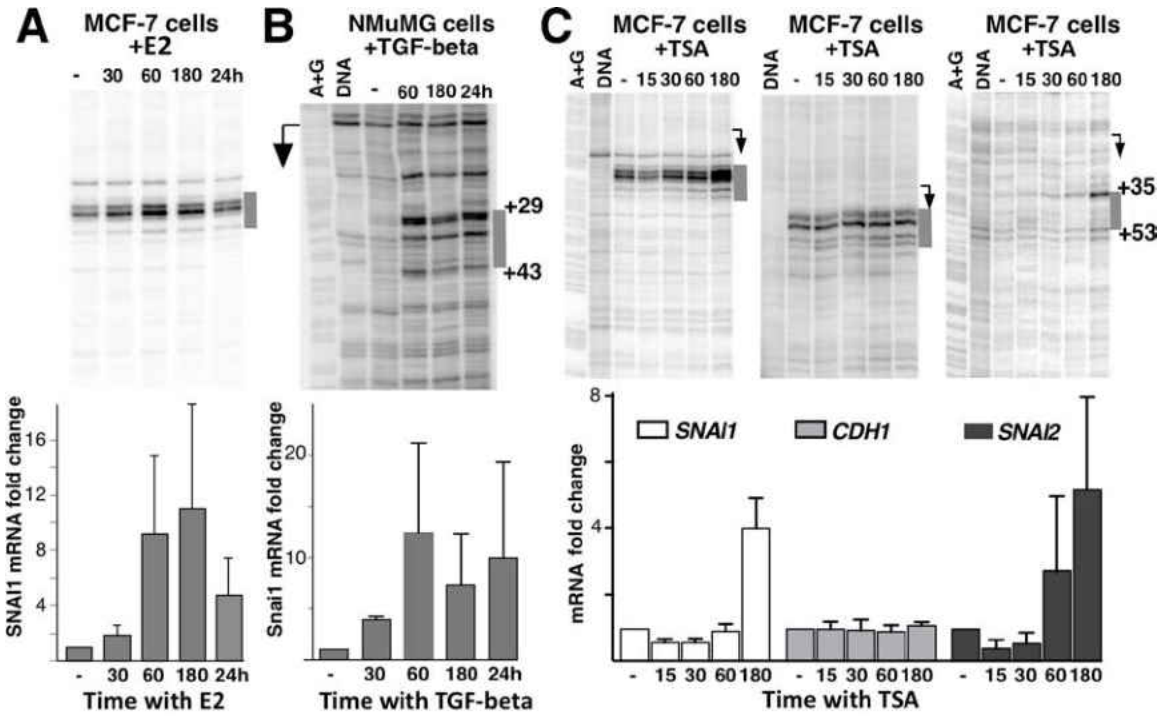

Figure 4. Pol II pausing signal is increased with gene activation. (A). Permanganate footprint on *SNAI1* gene with E2 treatment of MCF-7 cells and mRNA fold change as detected by RT-PCR normalized to *GAPDH*. RT-PCR represents an average of three independent replicates. (B). Mouse NMuMG cells treated with TGF-β RT-PCR and permanganate footprinting represents an average of three independent replicates. (C). MCF-7 cells treated with TSA. RT-PCR represents an average of four independent replicates. The regions of permanganate reactivity emerging during activation of mouse *Snai1* and human *SNAI2* genes are shown alongside the gel as a gray rectangle.

3.3 Pol II Pausing during Activation of *SNAI1* Gene

The increase in Pol II pausing during gene activation is indicative of an increase in Pol II occupancy at the promoter region. The increase in Pol II pausing can be a result of increase in recruitment of Pol II in response to gene activation, or due to a decrease in residence time leading to high turnover of Pol II at the paused site. At a given time, only a single molecule of Pol II can occupy the paused site. Therefore, the increase in pausing during gene activation is due

to more cells in the population gaining Pol II pausing on the *SNAI1* gene due to increased recruitment with activation. To investigate what fraction of cells, and gene copies, in a population was occupied by paused Pol II on *SNAI1* gene, we used artificial templates based on the human *SNAI1* gene sequence and containing an artificial transcription bubble. By measuring permanganate reactivity on artificial templates containing the known proportion of bubble, we determined that permanganate footprinting faithfully reports the relative propotion of bubbles on a gene. Based on 3 biological replicates we determined that the Pol II occupancy increases by 2-fold during TSA activation using 3 independent replicates. Given that TSA activation of MCF-7 cells leads to about 2 fold increase in permanganate signal on *SNAI1* gene, our data allowed us to postulate a 2-fold increase in the fractional occupancy of the *SNAI1* gene *in vivo*. This means that the highly *paused SNAI1* gene was not fully occupied prior to activation and the occupancy at basal levels was less than 50%.

Next we sought to find if the activation of *SNAI1* gene involves a change in residence time of paused Pol II. To determine changes in residence time of Pol II, we used TRP which specifically inhibits TFIIH (Titov et al. 2011). TRP inhibits recruitment of Pol II and thereby allows us to follow the Pol II complexes that are already paused (Henriques & Adelman 2013; Jonkers et al. 2014). We observed that the TSA activation of *SNAI1* gene (last four lanes of Figure 5A) does not speed up the turnover of pausing (first four lanes of Figure 5A). We demonstrate that the activation of *SNAI1* by TSA does not involve changes in Pol II dynamics and that the activation takes place through increase in Pol II recruitment and

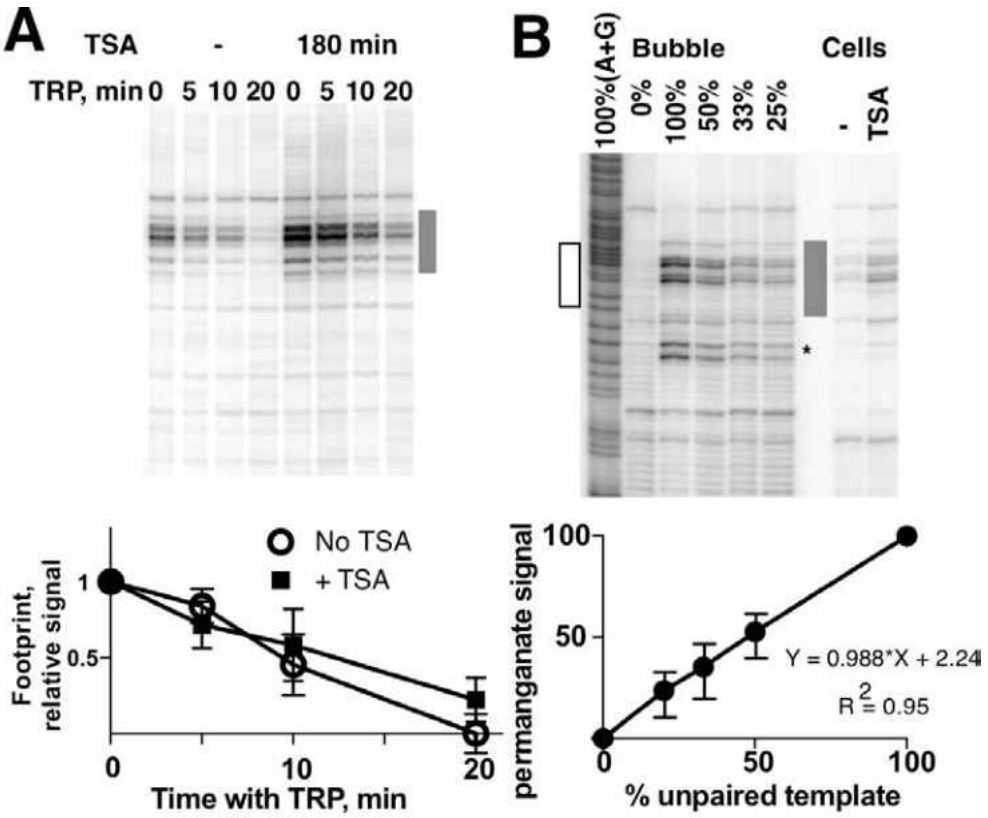

Figure 5. The *SNAI1* gene is activated through Pol II recruitment without changes in Pol II pausing duration. A. Pause signal decays in the presence of Triptolide (TRP). Quantitation of permanganate footprints (below), from three independent experiments. B. Permanganate footprinting of artificial bubble templates was done with *SNAI1* +100 primer set. A+G ladder and reactivity of MCF-7 cells, treated and untreated with TSA, is shown for comparison. Permanganate shows linear response to the proportion of transcriptional bubble. Quantitation of the experiment as in A. The values were obtained from three experiments and represent the area under the curve in the indicted part of the footprint.

transcription initiation events. This experiment demonstrates, for the first time, that a gene can be transcriptionally activated without changes in pause duration. Other genes, notably, *HSPA1B* heat shock protein 70 gene, have been shown to change pause duration during activation in both *Drosophila* and mammalian systems (Lis, 1998, Mahat et. al 2016). Our results point to a previously

unsuspected role of pausing *during* transcriptional responses. To investigate if the mechanisms involved in activation of *SNAI1* gene with TSA are distinct across cell lines, we followed activation in A375 cells, a human melanoma cell line. We note that both *SNAI1* and *SNAI2* genes get upregulated upon treatment with TSA (Figure 6A) in A375 cells just as they do in MCF-7 cells. Upon following the dynamics of the paused Pol II complex we observe that TSA activation of *SNAI1* gene in A375 cells does not involve changes in pause duration.

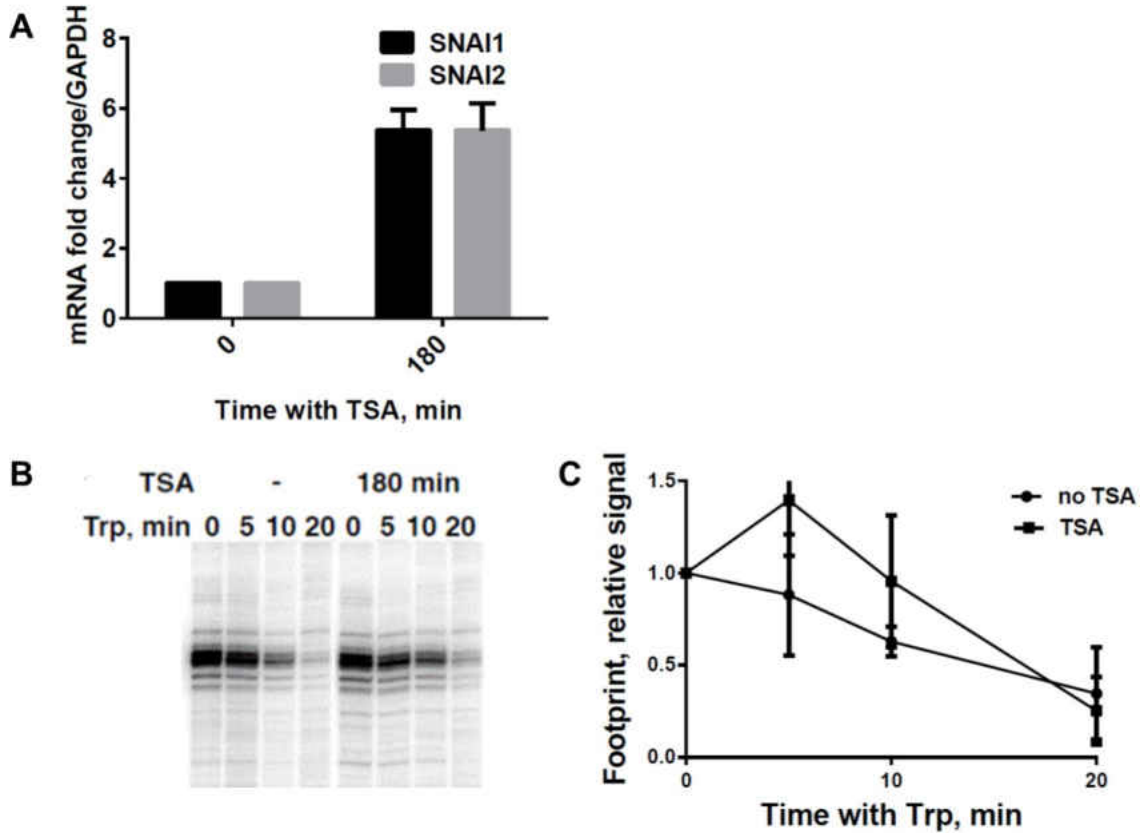

Figure 6. TSA activation of *SNAI1* gene occurs through does not involve change in Pol II pausing duration. A. mRNA fold change upon treatment with TSA in A375 cells. B. Footprint for *SNAI1* gene upon TSA activation and C. the quantification plot of the relative footprint in the presence of TRP.

3.4 Context-Dependent Release of Pol II Pausing during Gene Activation

Transcription initiation and Pol II pausing are two key processes that regulate early stages of transcription (Adelman & Lis 2012). Different signaling pathways have been in principle shown to activate transcription through either increase in initiation or release of pausing (Danko et al. 2013). However, this study employed activation in two different cell lines; MCF-7 cells stimulated with E2 and AC16 cardiomyocyte cells treated with TNF-α. Whether the same gene could be activated through two distinct mechanisms of pause release, and increase in Pol II recruitment in the same cell line was unknown. To follow activation of the same gene by different stimuli, we studied the heat shock pathway in MCF-7 cells, which upregulated the heat shock protein 70 (*HSPA1B*) gene. As previously observed in *Drosophila* (Gilmour & Lis 1986) and mouse systems (Mahat et al. 2016), we observed an increase in mRNA levels of *HSPA1B* in MCF-7 cells (Figure 7A). Previous reports have demonstrated that Pol II pausing gets completely obliterated in non-heat shock related genes in *Drosophila* (Yao et al. 2006). In contrast, the *SNAI1* gene remains paused during heat shock activation of MCF-7 cells. These results suggest that heat shock activation in mammalian cells is different than in *Drosophila*. We note that the activation of *HSPA1B* by TSA depends on an increase in Pol II recruitment and that activation by heat shock proceeds through pause release, demonstrating for the first time that the same gene can be activated by distinct mechanisms depending upon the stimuli (Figure 7B).

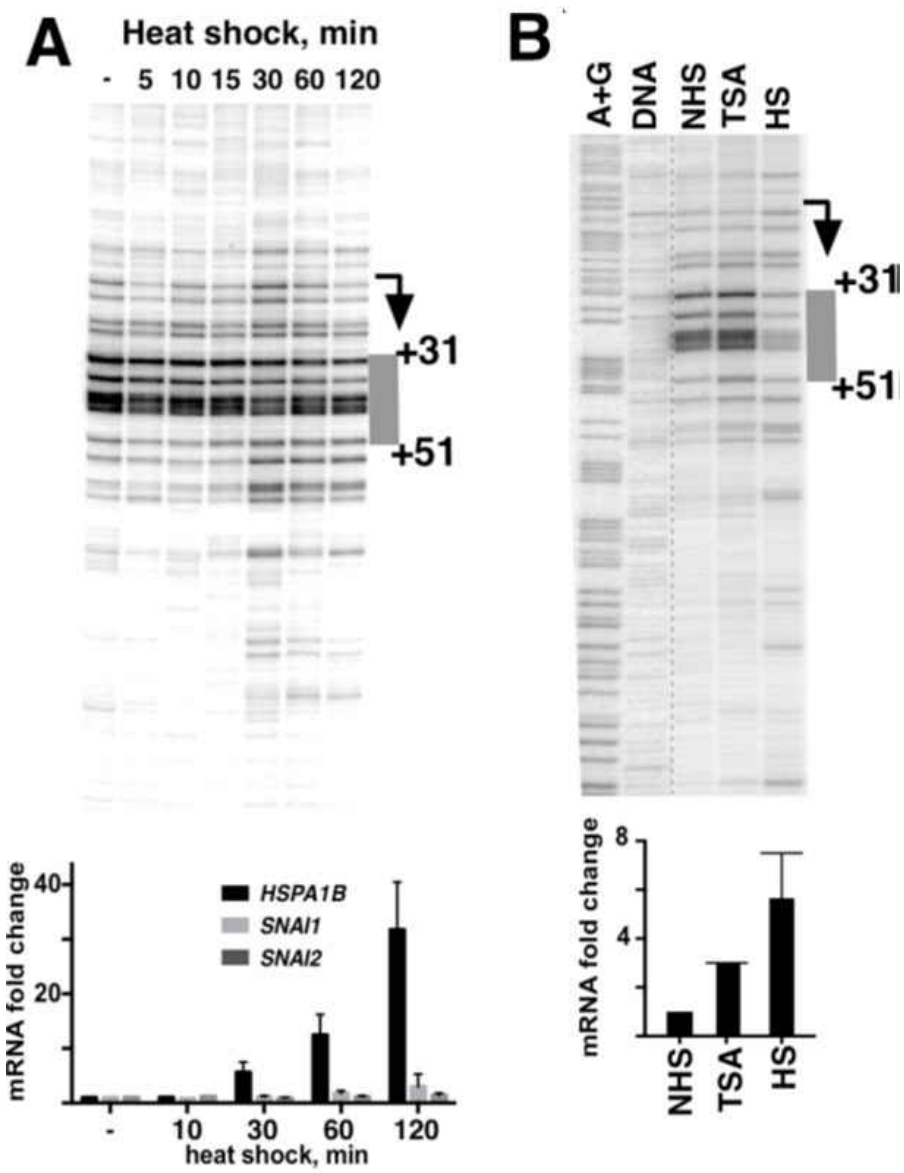

Figure 7. Pol II pausing is retained on *HSPA1B* gene during activation under the conditions of pausing release. (A) Permanganate footprinting for the time course of heat shock, shown for *HSPA1B* gene. The region of pausing is highlighted in gray. (B) Activation of *HSPA1B* gene with TSA (3 h) and heat shock (30 min), shown alongside naked DNA and the A/G ladder.

CHAPTER IV

PAUSING DYNAMICS DURING TRANSCRIPTIONAL RESPONSES REVEAL REDIRECTION OF PAUSED RNA POLYEMRASE II

4.1 Activation with TSA Results in Robust Transcriptional Response that is Common Across Cell Lines

To monitor changes in dynamics of Pol II pausing in response to stimuli, we tested global transcriptional responses to two distinct stimuli. Firstly, we chose TSA which has been previously shown to induce stress pathways. Gene activation with TSA, unlike that with heat shock, is not known to involve a specific transcription factor. To determine which genes were activated in response to two distinct stimuli TSA and heat shock, we performed mRNA-sequencing (RNA-seq) of MCF-7 cells following exposure to these stimuli. Among the time points that were published for TSA, we chose the 3 hour time point, as it gave the highest fold increase of *SNAI1* mRNA. Using a fold change cutoff of 1.3 and stringent criteria of adjusted p-value of less than 0.5, we identified genes that are differentially expressed upon TSA treatment.

Activation with TSA leads to a robust change in transcription. A higher number of genes are differentially expressed during TSA treatment of MCF-7 than in A375 cells (Figure 8). To investigate if transcriptional changes in response to stress are specific to the cell type, we performed RNA-seq analysis on A375 cells upon treatment with TSA. We observe that MCF-7 cells

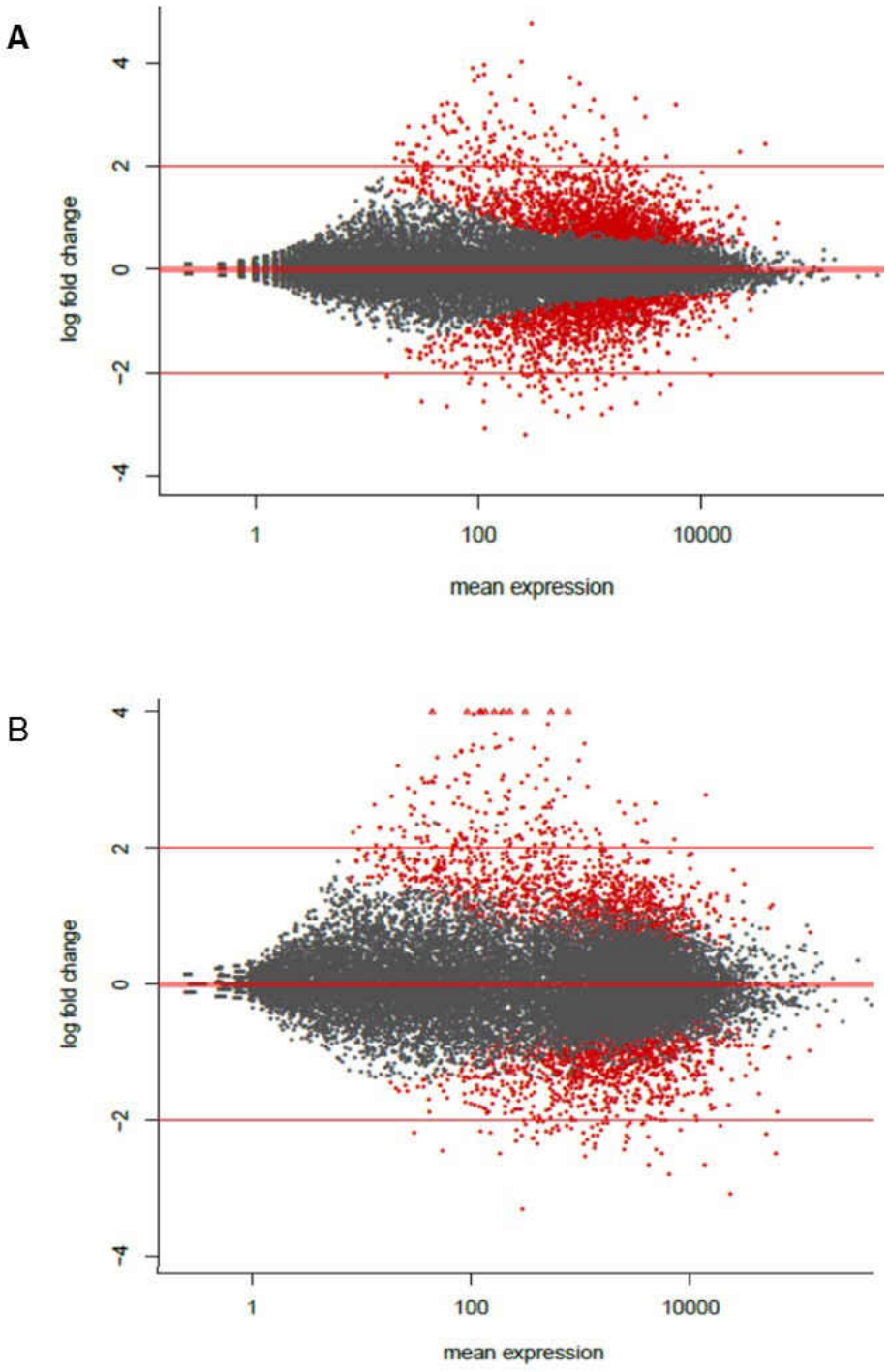

Figure 8. A large number of genes undergo changes in transcriptional levels during TSA treatment. Plot showing mean expression vs log fold change in A375 (A), and MCF-7 cells (B). in response to TSA treatment.

are more responsive to TSA in terms of the number of cells that appear to be differentially regulated upon TSA treatment relative to the control (Figure 9A and 9B). We observe that similar groups of genes are activated with TSA in both A375 and MCF-7 cells as per Gene Ontology (GO) analysis (Figure 9C and 9D). Thus, TSA treatment results in robust transcriptional responses despite the variation across the two cell lines.

4.2 Heat Shock of MCF-7 Cells Differentially Regulates Metabolism Related Genes

We next chose the heat shock response (HSR), a stress response pathway active in cancer cells (Mendillo et al. 2012), which is well studied with regards to Pol II pausing. HSR involves the recruitment of transcription factors called heat shock factors (HSF)s, especially HSF-1, to the promoter regions of some genes. During HSR not all genes are upregulated in a HSF-1 dependent manner, indicating distinct mechanisms of gene activation with the same stimuli (Mahat et al., 2016). We used 42ºC for heat shock with 30, 60 and 120 minutes as published for the *SNAI1* and *HSPA1B* genes (Samarakkody et al., 2015).

We first sought to identify early response genes of the HSR and genes that are upregulated later during heat shock. We observed a higher number of genes that are upregulated than downregulated during HSR. In addition, more genes are upregulated with longer duration of heat shock in MCF-7 cells. We note that RNA sequencing is biased against downregulated genes relative to nascent RNA analysis methods such as PRO-seq. PRO-seq is thus better suited for detecting downregulated genes (Mahat et al. 2016), and can be used in future work.

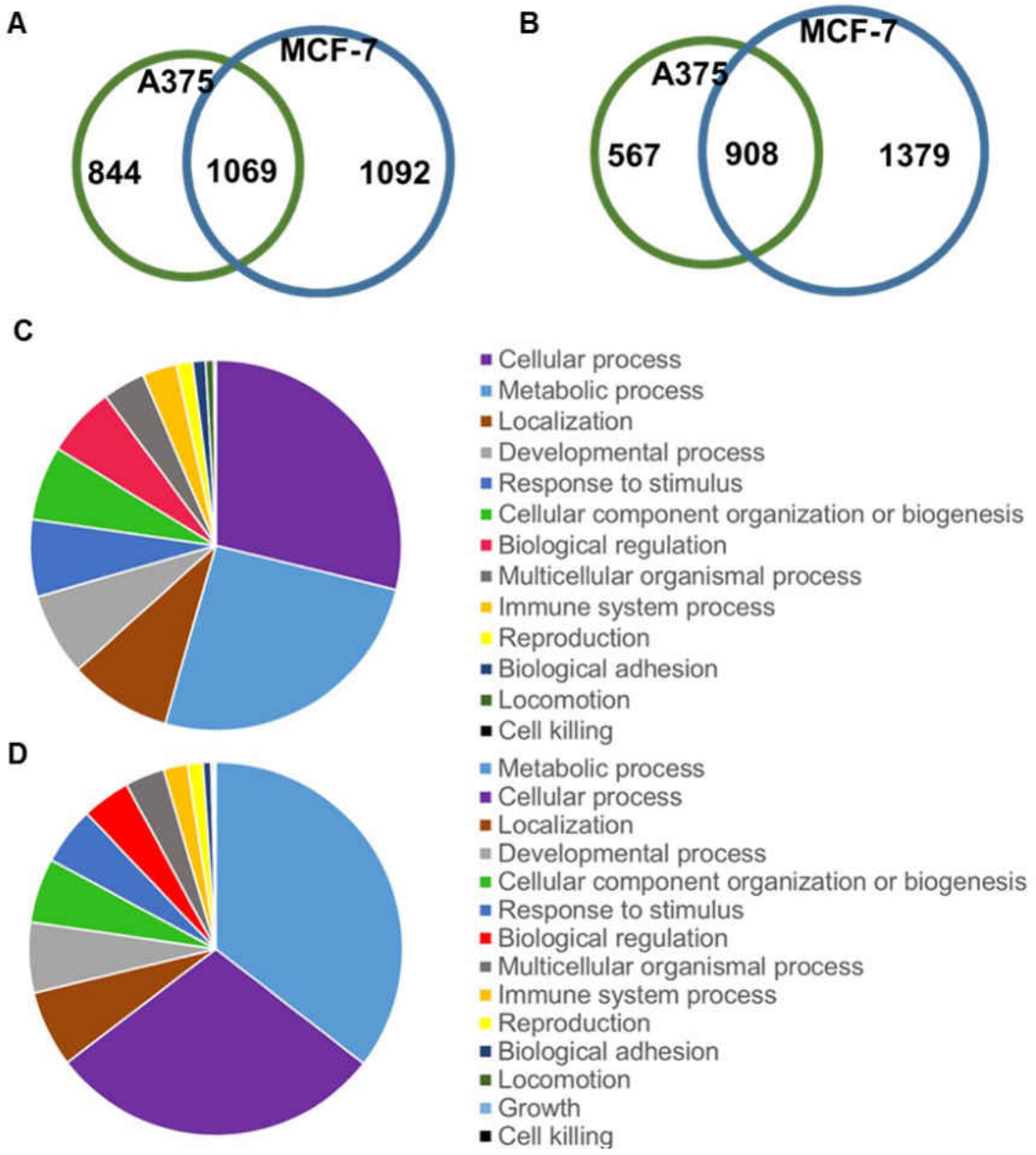

Figure 9. Metabolic processes are largely affected during TSA treatment of both MCF-7 and A375 cells. Venn diagrams showing the number of genes in MCF-7 (blue circle) and A375 (green circle) cells that are upregulated (A) and downregulated (B) during TSA treatment. Pie charts generated using the Panther program showing categories of genes that are upregulated (C) and downregulated (D) in both MCF-7 and A375 cells.

The activation of genes following the HSR that we observe here is consistent with what has been previously reported. The three genes upregulated at 30 minutes of heat shock (Figure 7B), *HSPA1B, HSPA1A* and *DNAJB1* (a heat shock protein 40), are genes in the HSR pathway (Table 1 and Figure 10A).

Table 1. The number of genes that are differentially expressed during heat shock response. The number of genes listed meet the criteria of adjusted p-value of 0.5 and have a log fold change of greater than 1.3. Each data point is representative of two independent replicates.

	Upregulated genes	Downregulated genes
Heat shock 30 minutes	3	0
Heat shock 60 minutes	108	2
Heat shock 120 minutes	213	5

When cells were heat shocked for a longer period of 60 minutes, 108 genes, including *HSPA1B, HSPA1A* and *DNAJB1*, were upregulated. While there were a larger number of genes that were significantly upregulated, 2 genes (red dots in Figure 10B) appeared to downregulated. Among the upregulated genes most genes are related to metabolism and other cellular processes (Figure 10C) as expected. At a longer time point with heat shock, more genes get upregulated (Figure 10D). We also note that some genes get transiently upregulated during heat shock. We note that 17 genes appear as differentially expressed only at 60 minutes (Table 2) and 125 genes only at 120 minutes. Well-known early response genes such as *JUN* and *FOS* are two examples of genes that are significantly upregulated at 60 minutes but not at 120 minutes of heat

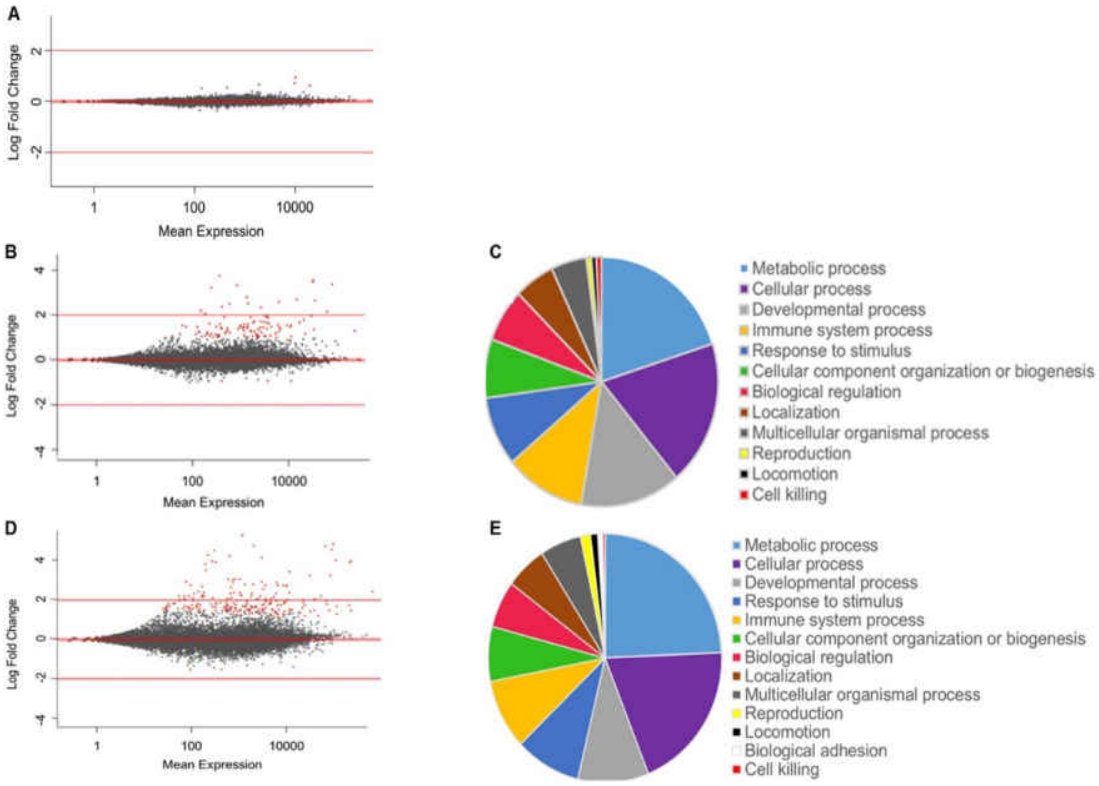

Figure 10. Genes involved in metabolism are upregulated during heat shock response. Plot showing the average expression and fold change at (A) 30 minutes, (B) 60 minutes and (D) 120 minutes Gene ontology analysis of genes that are identified as differentially expressed upon heat shock at (C) 60 minutes and (D) 120 minutes.

shock. These data provide evidence that some genes are activated early during the stress response, and some genes are activated in a transient manner.

Table 2. List of genes that are differentially expressed only at 60 minutes. The list of genes is sorted by the fold change and the p value is adjusted to disregard genes that have very low level of expression at basal stage.

Gene	Fold Change	p value
JUN	2.79	6.63E-06
PPP1R3C	2.78	1.01E-05
AP4B1-AS1	2.71	1.13E-05
RAET1K	2.41	0.000283347
PPP1R15A	2.38	0.000449711
ZC3H12A	2.38	0.000324101
FAM43A	2.28	0.000459722
FOS	2.26	0.00101959
ARC	2.23	0.001200184
DNAJB5	2.20	0.001553058
PSMD6-AS2	2.10	0.003984572
CIART	2.09	0.008118012
MYLIP	2.07	0.000161757
CCDC117	1.99	0.001246679
FRMD6	1.88	0.00493782
NUFIP2	1.77	0.009887267
DKK1	0.52	0.007441079

CHAPTER V

POL II PAUSING MAINTAINS A BALANCE BETWEEN PRODUCTIVE ELONGATION AND PREMATURE TERMINATION DURING TRANSCRIPTIONAL RESPONSES

5.1 Turnover of the Paused Pol II Complex is Determined by the Stimulus

Escape of paused Pol II into productive elongation is essential in producing mRNAs from a gene. The kinase activity of the p-TEFb complex, which is composed of the cyclin-dependent kinase CDK9 and Cyclin-T, phosphorylates the CTD of the large subunit of Pol II and facilitates the transition to productive elongation. During activation of genes, two distinct mechanisms come together to increase transcription: increased recruitment of Pol II to promoter regions and increased activity of p-TEFb facilitating escape of paused Pol II into elongation. Gene activation has been proposed to follow two distinct mechanisms in response to E2 and TNF-α in two different cell lines (Danko et al. 2013). We have previously shown (Samarakkody et al. 2015), as demonstrated in chapter 1, that genes can be activated by an increase in recruitment of Pol II onto promoter regions without changes in pausing duration. This study demonstrated that E2 increases Pol II initiation and TNF-α increases the escape of paused Pol II into productive elongation. Although transcriptional activation of E2 and TNF-α follows two distinct mechanisms, the inherent differences between the two cell lines may have also contributed to the different responses, leaving

the question of whether the same gene can be activated through the same mechanism open. We next determined whether the response of paused Pol II to gene activation is determined by the gene or by the stimulus. To do so, we selected genes that are upregulated by both TSA and heat shock in MCF-7 cells. If the mechanism of activation is determined by the gene, dynamics of paused Pol II and hence the duration of pausing at a given gene will be the same with different stimuli. However, if activation of a gene is determined by the stimulus, the pause duration of a gene may be different depending upon the stimulus. To determine pausing duration, we used Triptolide (TRP), as described in the previous chapter, which has been used by us and others to measure pause duration (Samarakkody et al. 2015; Henriques et al. 2013) at a final concentration of 2.5 μM. We chose to investigate *SNAI1* and *HSPA1B* genes that are upregulated both with TSA and heat shock. We first compared how TSA treatment affects activation of both *SNAI1* and *HSPA1B* (Figure 11A-D). We noted that the pause duration does not change with TSA treatment on both tested genes. However, the pause duration is drastically reduced upon heat shock treatment on both *SNAI1* and *HSPA1B* (Figure 11E-H) indicating that dynamics of the paused complex changes in a stimulus specific manner, and not in a gene specific manner. During the heat shock response, the transcription factor HSF-1 is known to be recruited to some promoter regions. HSF-1 can recruit p-TEFb, a pause release factor. The *HSPA1B* gene is activated in an HSF-1 dependent manner, but *SNAI1* is not. The fast decay of Pol II at the pause site in *HSPA1B* can be attributed to pause release that may occur in an HSF-1

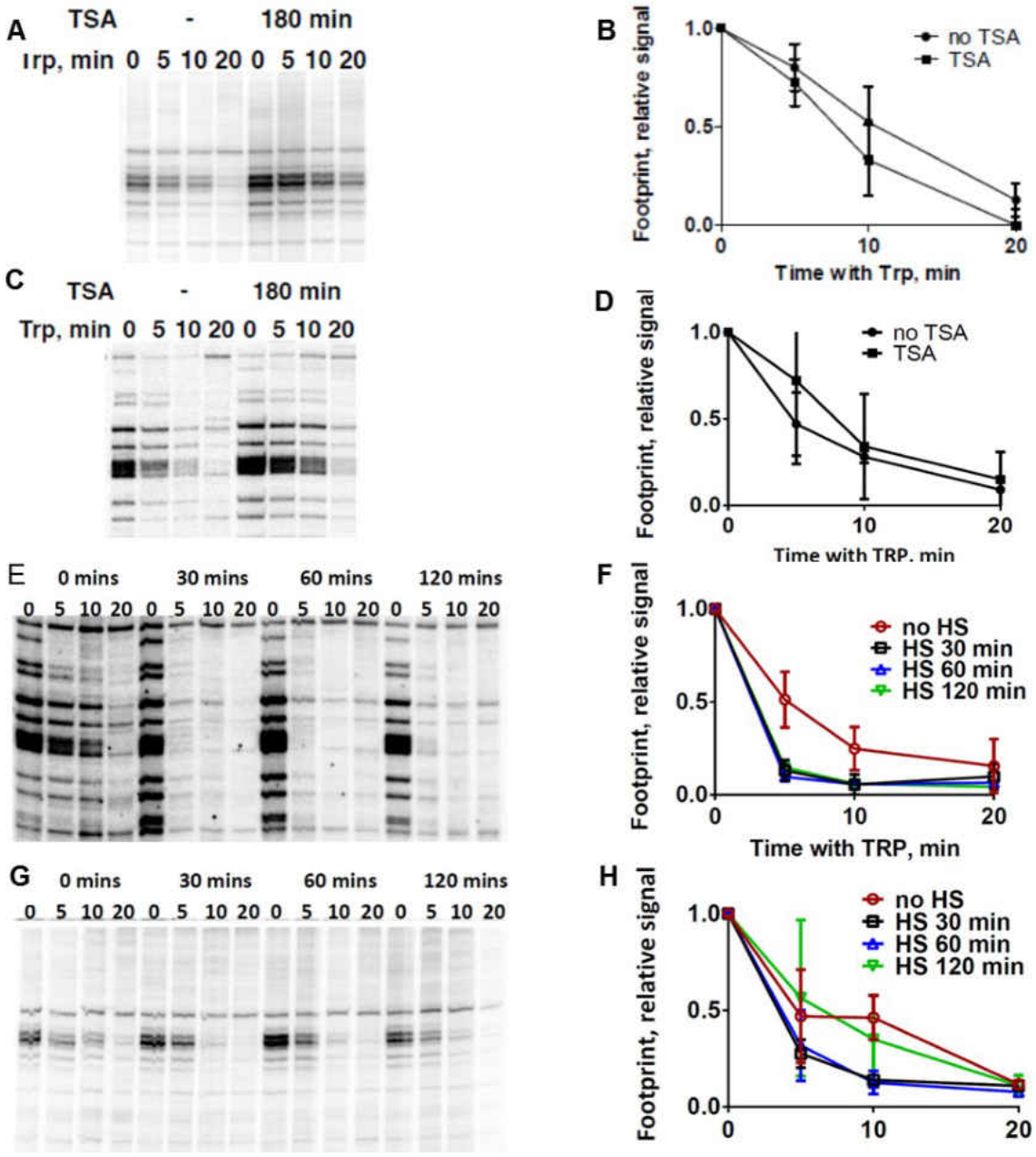

Figure 11. Pause duration changes with heat shock and not with TSA treatment. Permanganate footprints and quantification plots (right) for *SNAI1* (A) and (G) and *HSPA1B* (C) and (E) with TSA and Heat shock in the presence of TRP. Each time point is an average of 3 independent replicates.

dependent manner. However, it is yet to be explored whether the rapid decay of the paused complex on the *SNAI1* gene is due to pause release into elongation.

5.2 Pol II Turnover Increases Genome wide during Heat Shock Response

Knowing that two genes change dynamics of the paused Pol II in the presence of heat shock, we next sought to address if it is a characteristic of a cohort of upregulated genes, or is common across the genome. To investigate Pol II turnover on genes across the genome, we used short capped RNA sequencing. With modifications to the published protocol, we selected for short capped RNA from the nuclei of MCF-7 cells to investigate genome-wide Pol II pausing (Figure 12). Two genes that are known to be paused in MCF-7 cells, *SNAI1* and *CDH1* genes, showed a signal with short-capped RNA sequencing (scRNA-seq) while *SNAI2,* a gene that is not paused at delectable levels in MCF-7 cells does not (Figure 12). Consistent with what has been previously observed in *Drosophila* and mouse, we observe that Pol II pauses about 35 nucleotides downstream of the transcription start site in humans (Figure 13).

To select for RNA that results from Pol II pausing, we fractionated cells to obtain RNA that is bound to chromatin. Using a previously published protocol, (Samarakkody et al., 2015) with modifications, we were able to reliably detect Pol II pausing genome-wide using chromatin bound RNA. During the time course with TRP the amount of chromatin bound RNA at the 20-minute time point is expected to be low to none due to the decay of complexes from the paused site (Figure 14). Hence, the libraries from these time points cannot be normalized

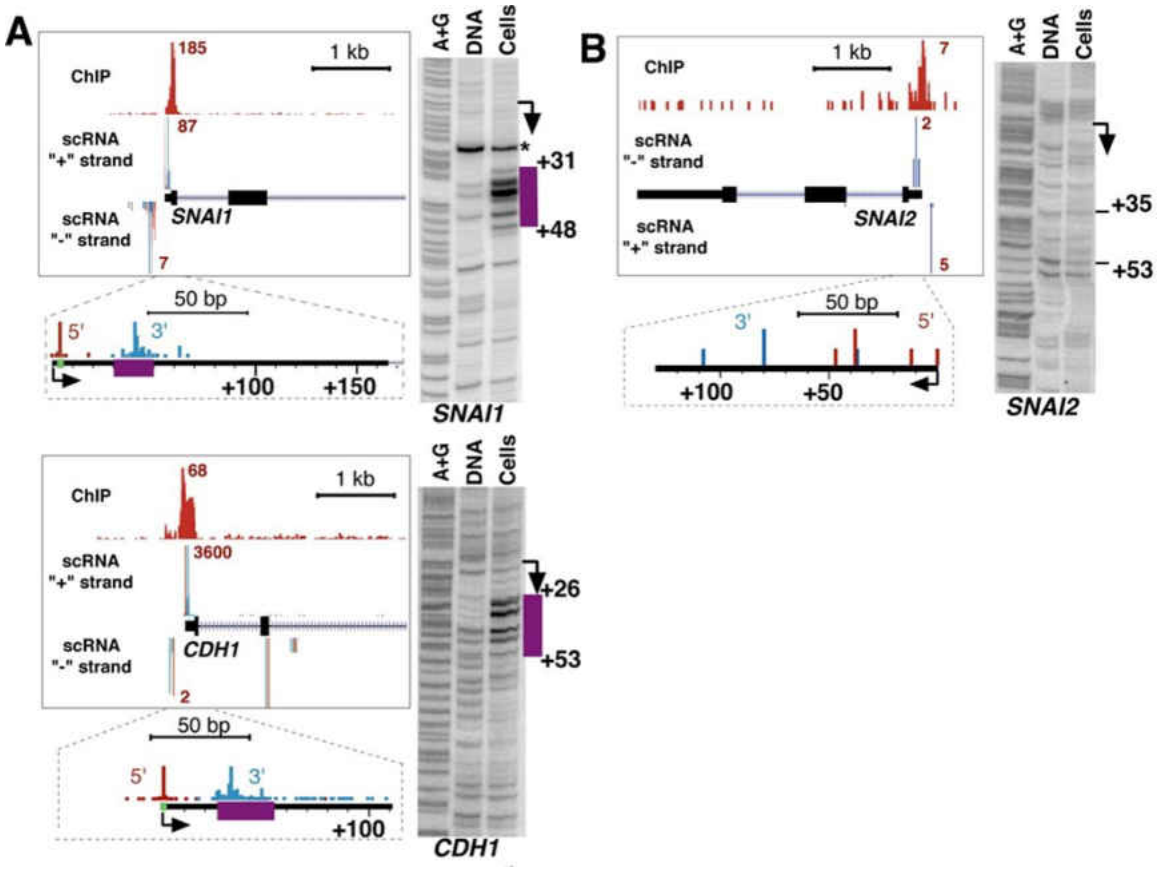

Figure 12. Pol II pausing on individual genes. (A). *SNAI1* and *CDH1* genes show evidence of Pol II pausing but (B). *SNAI2* gene does not. For each gene, a UCSC genome browser view is shown, with positions of Pol II ChIP-seq bins and the positions of 5′ and 3′-RNAs indicated. For scRNAs, plus and minus strand matches are shown as separate tracks. Numbers near each peak represent the maximum number of sequence reads within the displayed interval. An inset below each gene shows a zoomed-in view of scRNA tracks around the TSS. Permanganate footprinting of the same gene is shown on the right, alongside naked DNA and A/G ladder. Annotated TSSs are shown with arrows and TSS start sites are shown in green. The region of permanganate reactivity is shown alongside the gel and on the gene scheme in purple. An asterisk at the *SNAI1* gene indicates a non-specific band present both in cells and naked DNA.

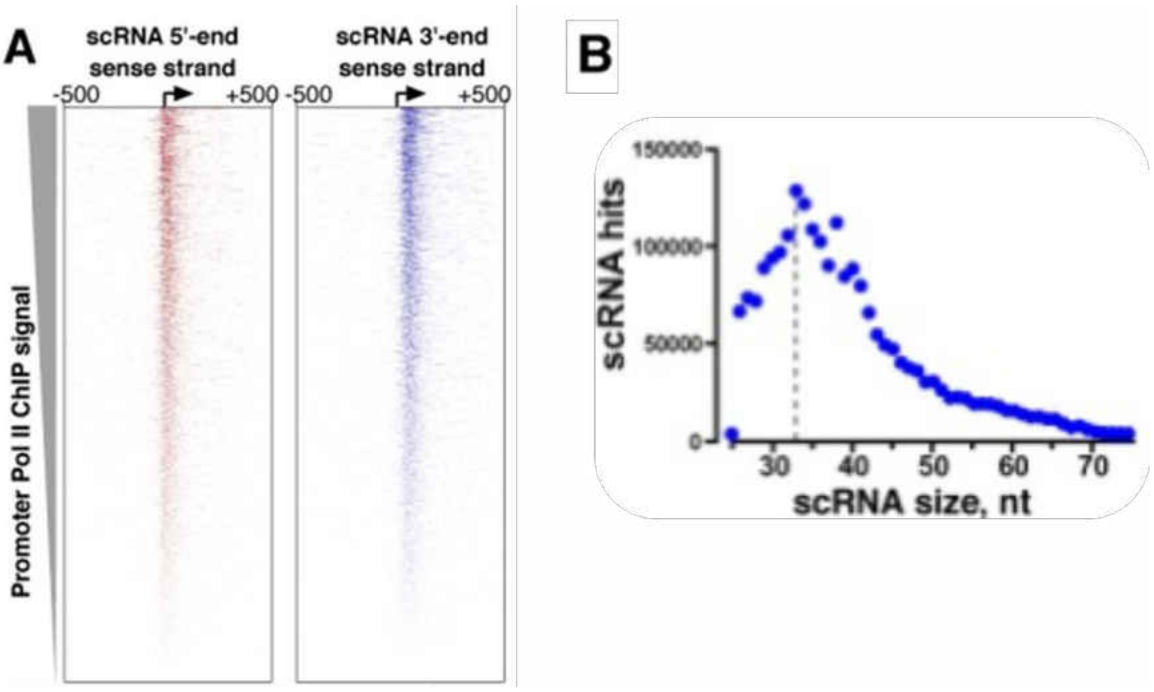

Figure 13. (A) Heatmaps showing scRNAs for 7302 genes with the highest promoter Pol II enrichment, in the region of ±500 nt around their RefSeq (annotated) TSSs. The position of the TSS is shown by an arrow. Only the sense strand scRNAs are shown. (B) Metagene plot showing the frequency distribution of lengths of scRNAs based on distances between R1 and R2 paired-end sequencing read positions.

based on the total number of reads. To overcome this, we have used known amounts of artificially synthesized short capped RNA, with known lengths and sequences, which we added in to Trizol at the step of RNA extraction which are then used in normalization.

We show that paused Pol II complexes decay in the presence of TRP genome-wide as observed for *SNAI1* and *HSPA1B* genes. Secondly, we note that most complexes decay faster in the presence of TRP with heat shock activation (Figure 15). For genes that are upregulated, the increase in turnover of

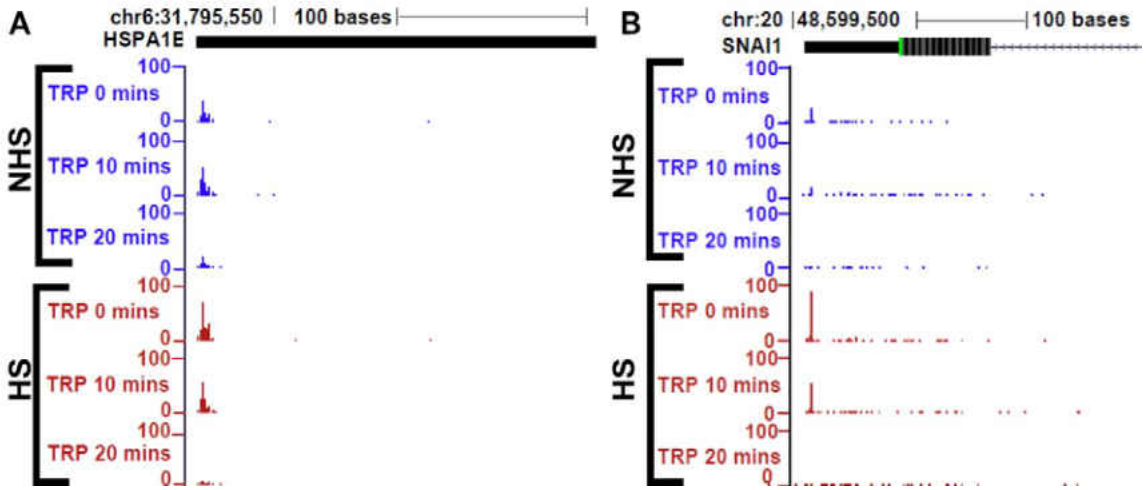

Figure 14. Chromatin bound RNA allows to detect decay of paused Pol II complex genome wide. A *HSPA1B* and B *SNAI1 I* genes showing decay of signal from R1 reads from scRNA-seq in the presence of TRP.

paused complexes can be a result of more pause release in to productive elongation to produce more mRNA. Alternatively, the high turnover can also be a consequence of increased premature termination which could be especially true for genes that are downregulated or not significantly upregulated during activation. We also observe that a few genes such as *FTL*, *SNHG1* and *RPLP1* that show increased paused duration upon heat shock.

5.3 Pause Release by p-TEFb affects Pol II Turnover

Having observed that the fast release of paused Pol II is stimulus dependent, we sought to answer if pause release completely relies on the activity of p-TEFb. We used Flavopiridol (FP), a potent inhibitor of p-TEFb, to investigate if paused Pol II complexes are released slower when pause escape is blocked by FP. In fact, what we observed was a faster release of paused Pol II at basal levels (Figure 16), indicating that the paused complexes may be actively

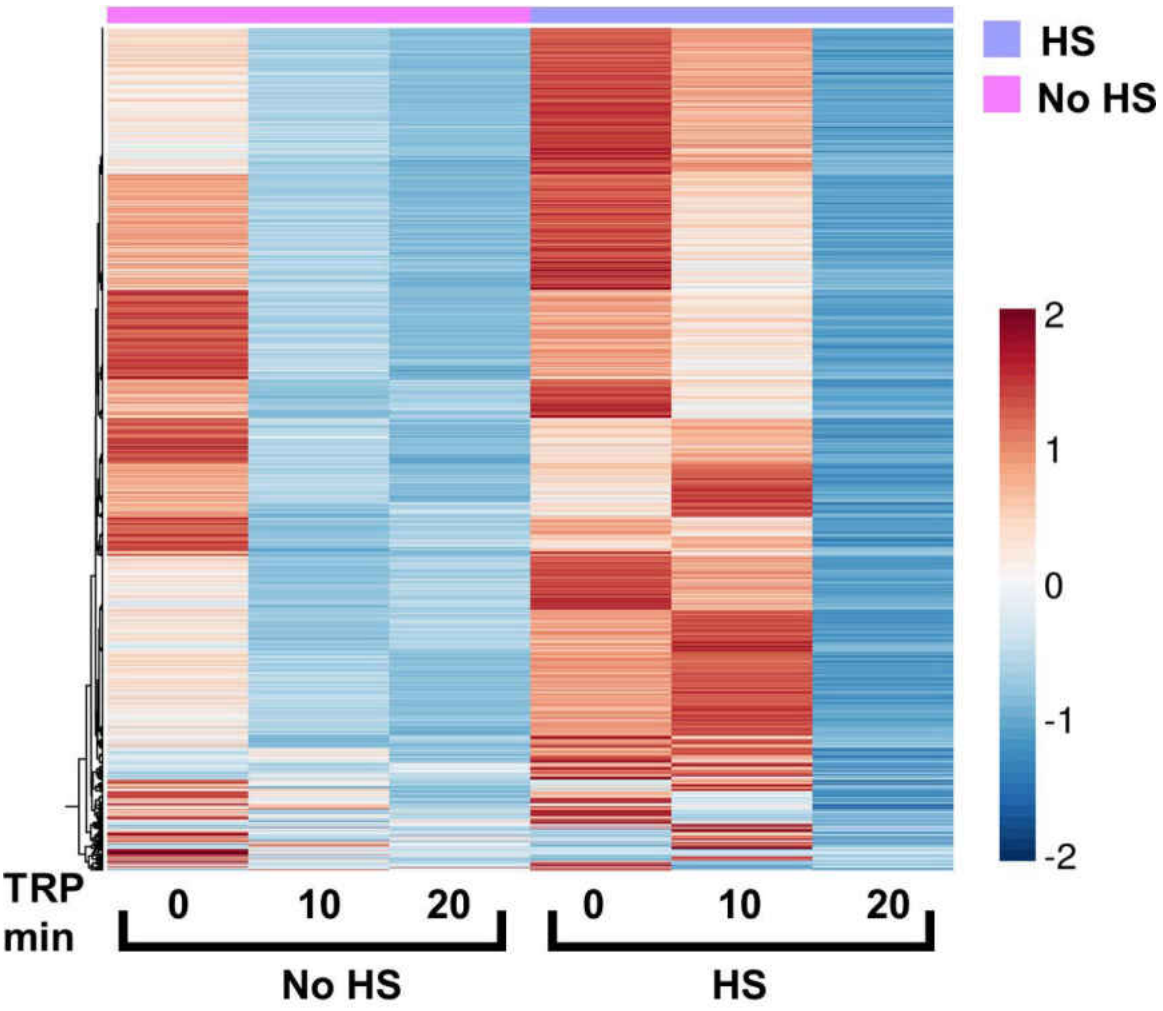

Figure 15. Heat map showing faster decay of complexes genome-wide in the presence of TRP. The scale on the left shows the fold change of signal and the genes are being ranked after unsupervised clustering.

redirected towards premature termination when paused Pol II cannot proceed into elongation. Next, we investigated how the dynamics of paused Pol II may change during gene activation. Using TSA and heat shock as activators, we followed paused Pol II in the presence of FP. We observe that with activation of *SNAI1* gene with TSA, the complexes showed faster turnover. We compared this observation with heat shock activation and noted that the turnover rate of the

paused complexes increased in the presence of FP. Hence, for the first time, we show that the balance between productive elongation and premature termination changes in a stimulus dependent manner.

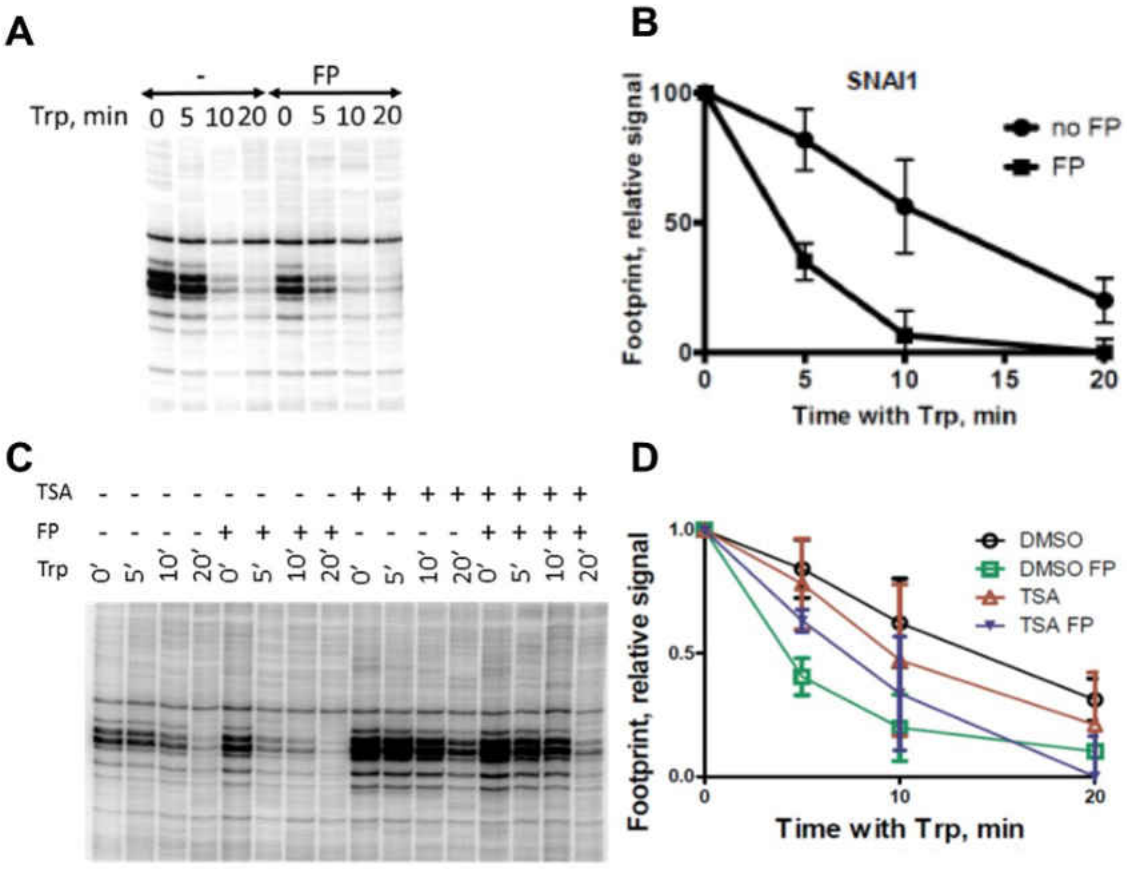

Figure 16. Inhibition of p-TEFb with FP shows a stimulus specific effect on pausing duration. A. *SNAI1* gene showing faster decay in the presence of FP along with the quantification plot on the right B. C. Gel image for *SNAI1* gene in the presence of FP with TSA activation with the quantification plot on the right D.

CHAPTER VI

DISCUSSION

The main objective of this study was to determine how Pol II pausing could regulate transcriptional programs in cells in response to stimuli. Pausing has been previously suggested to serve as a platform to integrate regulatory inputs (Henriques et al. 2013) into a transcriptional outcome. However, the mechanisms that involve multiple transcription factors and dedicated proteins in regulating transcriptional outcome are not well understood. We demonstrate that, firstly, Pol II pausing is retained during activation on genes with both TSA and heat shock activation. Secondly, we show that genes can be activated through distinct molecular mechanisms primarily through either pause retention or pause release. Here, we show that the same gene can be activated by either increase in Pol II pausing or by release of pausing depending upon the stimulus. This demonstrates that genes are flexible in their response to stimuli and that activation of the same gene can occur through distinct molecular mechanisms as measured by pause dynamics. Regardless of the mechanism of activation, Pol II pausing plays an important role in regulating the transcriptional outcomes.

Using global sequencing of scRNA to map Pol II pausing across the genome in human MCF-7 breast cancer cells, we show that (i) transcription start sites can be reliably detected by analysis of the 5’ ends of scRNA reads and

(ii) the location of Pol II pausing can be detected by the 3' ends. These studies show that pausing location is highly conserved across the genomes of different classes, *Drosophila*, mice, and as we show, humans. We note that, consistent with what has been observed in other systems, the predominant distance of 33–35 nucleotides downstream of the TSS was observed across the genome on MCF-7 cells, indicating that the underlying molecular mechanisms govern pausing of Pol II regardless of the promoter sequence. Furthermore, using the scRNA sequencing on RNA from chromatin-bound fraction, we were able to detect pause duration on a global scale. We note that most genes show an increased turnover at the paused site upon heat shock activation. Future studies will determine if genes that demonstrate increased turnover also show increased rates of premature termination and pause release upon heat shock as well as other stimuli. Monitoring the dynamics of Pol II pausing in response to stimuli will be used in future work to for profiling of cells perhaps during differentiation, development and also during progression of diseases such as cancers.

The work characterizes several novel properties of Pol II pausing. First, we demonstrate the plasticity that genes exhibit in response to different stimuli at the level of pausing. We show that Pol II pausing continues to provide a stable platform not only prior to, but also during transcriptional responses to stimuli. We report for the first time that the dynamics of paused Pol II in response to activation is not hardwired in the gene, but is dynamically determined by the stimulus. Using model genes *SNAI1* and HSPA1B, we show that the same gene can respond to different stimuli differently and specifically, leading to faster

release of paused Pol II with heat shock, yet retention of pausing with TSA. At the conditions of fast turnover of Pol II during activation, the location of pausing did not change, suggesting that all factors necessary for pause establishment and release remain available at the promoter, even at conditions of high-level activation. The mechanisms that involve pause release rely on multiple steps of p-TEFb recruitment, phosphorylation of NELF, and the interaction or recruitment of multiple functions through protein complexes such as the Super Elongation Complex. While the interplay between pausing establishment and release across the genome remains to be fully understood, our data show that Pol II pausing remains the slowest step on heat shock gene in human cells even at the conditions of high-level activation. Further analysis of early transcription combining nuclear run-on and free RNA-based methods will shed light on the mechanisms and global dynamics of pausing during stimulus response.

Using several EMT related genes such as *SNAI1, CDH1* and *SNAI2,* we show that gene activation can take place without changes in pausing location and dynamics as measured by pause duration. Among these genes, *SNAI2* is considered a non-paused gene due to low levels of promoter Pol II observed in cells prior to activation. In addition, *SNAI1* and *SNAI2* genes are transcribed at very low levels in MCF-7 cells. The human *SNAI1* gene upon activation with TSA acquires pausing signature revealed by increased recruitment of Pol II to the promoter. From a perspective of mechanisms, the availability of the negative elongation factor (NELF) during each transcription initiation determines the establishment of pausing (Gilchrist et al. 2008; Yamaguchi et al. 1999). During

differentiation and development, the availability of factors such as NELF may change, thus altering pausing status on genes. Overall, we propose that the increase of the pausing signal represents the default outcome of gene activation that enables context-specific regulatory inputs, one of which is pause release, to affect the rate of mRNA synthesis. Further analysis of pausing in different systems will uncover its role both in maintaining genes in a “poised” st

In *Drosophila* genes with high levels of Pol II pausing contain initially transcribed sequences that favor nucleosomes and promote chromatin assembly. In contrast, genes that have low levels of Pol II pausing on promoter sequences disfavor nucleosome assembly (Gilchrist et al. 2010). In mammals, Pol II pausing is enriched at genes that contain CpG islands at promoter regions (Core et al. 2008). Although much debated, there is evidence supporting the idea that, as in *Drosophila,* CG rich mammalian promoters favor nucleosome formation, suggesting that pausing at such promoters helps maintain the promoter in an open configuration to be readily accessed by the transcription machinery. We note that human *SNAI2* gene, which is not paused, does not contain a CpG island overlapping its promoter, while the paused *SNAI1* gene does. Although *SNAI2* does not contain a CG rich promoter sequence, it is able to acquire pausing with increase in recruitment of Pol II. Hence, the presence of an overlapping CpG island is not strictly required for pausing, suggesting that mechanisms other than DNA methylation, which occurs at CpG islands, may control pausing. Further, our finding that TSA upregulates the heat shock *HSP70* gene independent of the HSF-1 factor suggests that TSA, and

HDAC inhibitors in general, can affect genes through bypassing their default activation mechanisms including requirement for specific transcription factors. In addition, using *SNAI1* as an example, we show that genes do not get fully occupied by Pol II upon activation. With the use of an artificially synthesized bubble template for *SNAI1* gene we have determined the relative level of pausing in a population of cells. We first note that about 20% of cells contain Pol II pausing at basal state on *SNAIL* gene and that Pol II pausing signal is increased by about 2-fold upon activation with TSA. It is likely that most genes, even those that are highly paused, are not fully occupied by Pol II, indicating that initiation of transcription, and not pausing, may be a rate limiting step in early events of transcription of most if not all genes.

We note that upon heat shock, *HSPA1B* is activated through pause release in MCF-7 cells. A large fraction of P-TEFb is bound to the 7SK snRNP that sequesters the kinase in an inactive state, and another fraction of P-TEFb is found in a protein complex known as the Super Elongation Complex (SEC) (reviewed in Luo et al 2012). The multiprotein complex of SEC appears to be particularly important for rapid induction of transcription in stem cells (Lin et al. 2011). Although the mechanisms of recruitment of p-TEFb and SEC to gene promoters are not completely understood, the mediator subunit MED26 has been recently shown to be involved (Takahashi et al. 2011).

Most gene promoters contain multiple motifs for binding of transcription factors (TF). Among such promoter binding TFs, some assist in recruiting p-TEFb to the promoter. BRD4 and cMyc are two of the transcription factors that are

largely studied in the context of pause escape. Similarly, HSF-1 has been shown to recruit p-TEFb in *Drosophila* to promote pause release on genes upon heat shock activation. Recently, HSF-1 was also shown to regulate pause release in mouse cells during heat shock response (Mahat et al. 2016). While the interplay between pausing establishment and release remains to be fully understood, our data show that Pol II pausing remains the slowest step on heat shock genes in human cells even at conditions of high level of gene activation.

Recent work showed that Pol II pausing, not the pre-initiation complex, is likely the main form of Pol II near promoters in metazoans (Pugh and Venters, 2016) and our recent analysis of several individual genes supports this notion (Samarakkody et al., 2015). Promoters recruit Pol II into a paused site that gets distributed into distinct fates of termination versus elongation. While the exact factors, and the interplay among factors, remain to be characterized, we suggest that termination and pause release into productive elongationcompete during transcriptional responses. We suggest that termination does not require P-TEFb function (Figure 2). Thus, the fate of the paused complex is not pre-determined during transcription initiation, but rather, determined at the paused complex. Multiple other factors are involved in pausing and pause release. Some of these factors are direct and some are indirect in terms of their effects on Pol II pausing. For example, Paf1 may or may not be acting directly. It is clear, however, that Pol II pausing emerges as a complex platform for regulatory inputs that enables transcriptional signals through multiple mechanisms that remain to be investigated.

Cells orchestrate transcriptional responses by rapidly and precisely coordinating expression of thousands of genes in a cell type specific manner. Unlike single-celled organisms, metazoans are not driven directly by the environment, but must follow a predetermined epigenetic program in different environmental conditions. Metazoan cells are specialized and must undergo responses not only respond to the signal, but also must stay within their epigenetically specified programs. Evolutionary advantages have given metazoans numerous means of retaining their epigenetic identity in the presence of stress conditions, which is essential in survival of the organism. Because only a small number of genes respond directly to the stimulus, additional mechanisms must be involved in global transcriptional responses. How transcriptional machinery is “primed” for transcriptional responses remains largely unclear. The phenomenon of transcriptional pausing is a key regulatory mechanism that has received much attention due to its prevalence in early response genes and its association with important regulatory process such as cell differentiation, development, and responses to stimuli. Although the duration of the paused complex is known to change during responses to stimuli, how Pol II pausing changes during transcriptional responses remains to be investigated in the future. It is not clear whether activation of genes must require release of paused Pol II, or if transcriptional events lead to a decrease in pausing duration. Here we demonstrate that not only levels of Pol II pausing, but pause duration are measures of the state of a gene. The availability of factors that regulate Pol II pausing in promoter proximal regions may play a determining role in how genes

would respond to stimuli. For instance, cMyc that plays a role in promoter proximal regions has been shown to be overexpressed in cancer cells. Perhaps the dysregulation of genes during disease conditions may occur at the levels of transcription at the paused site.

REFERENCES

Adelman, K. & Lis, J.T., 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews. Genetics*, 13(10), pp.720–31.

Ardehali, M.B. et al., 2009. Spt6 enhances the elongation rate of RNA polymerase II in vivo. *The EMBO journal*, 28(8), pp.1067–77.

Ardehali, M.B. & Lis, J.T., 2009. Tracking rates of transcription and splicing in vivo. *Nature Structural & Molecular Biology*, 16(11), pp.1123–1124.

Banerjee, A.K., 1980. 5'-terminal cap structure in eucaryotic messenger ribonucleic acids. *Microbiological reviews*, 44(2), pp.175–205.

Barboric, M. et al., 2001. NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Molecular cell*, 8(2), pp.327–37.

Bartkowiak, B. et al., 2010. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes & Development*, 24(20), pp.2303–2316.

Biamonti, G. & Caceres, J.F., 2009. Cellular stress and RNA splicing. *Trends in Biochemical Sciences*, 34(3), pp.146–153.

Bochnig, P. et al., 1987. A monoclonal antibody against 2,2,7-trimethylguanosine that reacts with intact, class U, small nuclear ribonucleoproteins as well as with 7-methylguanosine-capped RNAs. *European journal of biochemistry*, 168(2), pp.461–7.

Bond, U., 1988. Heat shock but not other stress inducers leads to the disruption of a sub-set of snRNPs and inhibition of in vitro splicing in HeLa cells. *The EMBO journal*, 7(11), pp.3509–18.

Brannan, K. et al., 2012. mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. *Molecular Cell*, 46(3), pp.311–324.

Buckley, M.S. et al., 2014. Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation. *Genes & development*, 28(1), pp.14–9.

Cartwright, P. et al., 2005. LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development*, 132(5), pp.885–896.

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), pp.31793–9.

Chen, F., Gao, X. & Shilatifard, A., 2015. Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. *Genes & Development*, 29(1), pp.39–47.

Chen, F., MacDonald, C.C. & Wilusz, J., 1995. Cleavage site determinants in the mammalian polyadenylation signal. *Nucleic acids research*, 23(14), pp.2614–20.

Cheng, B. et al., 2012. Functional Association of Gdown1 with RNA Polymerase II Poised on Human Genes. *Molecular Cell*, 45(1), pp.38–50.

Chodosh, L.A. et al., 1989. 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. *The Journal of biological chemistry*, 264(4), pp.2250–7.

Chopra, V.S. et al., 2011. The polycomb group mutant esc leads to augmented levels of paused Pol II in the Drosophila embryo. *Molecular cell*, 42(6), pp.837–44.

Core, L.J. et al., 2012. Defining the status of RNA polymerase at promoters. *Cell reports*, 2(4), pp.1025–35.

Core, L.J., Waterfall, J.J. & Lis, J.T., 2008. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science*, 322(5909), pp.1845–1848.

Danko, C.G. et al., 2013. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Molecular cell*, 50(2), pp.212–22.

Derycke, L.D.M. & Bracke, M.E., 2004. N-cadherin in the spotlight of cell-cell adhesion, differentiation, embryogenesis, invasion and signalling. *The International journal of developmental biology*, 48(5–6), pp.463–76.

Edery, I., Altmann, M. & Sonenberg, N., 1988. High-level synthesis in Escherichia coli of functional cap-binding eukaryotic initiation factor eIF-4E and affinity purification using a simplified cap-analog resin. *Gene*, 74(2), pp.517–25.

Edery, I. & Sonenberg, N., 1985. Cap-dependent RNA splicing in a HeLa nuclear extract. *Proceedings of the National Academy of Sciences of the United States of America*, 82(22), pp.7590–4.

Feder, M.E. & Hofmann, G.E., 1999. HEAT-SHOCK PROTEINS, MOLECULAR CHAPERONES, AND THE STRESS RESPONSE: Evolutionary and Ecological Physiology. *Annual Review of Physiology*, 61(1), pp.243–282.

Fong, N. et al., 2014. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes & development*, 28(23), pp.2663–76.

Fresco, L.D. & Buratowski, S., 1996. Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing. *RNA (New York, N.Y.)*, 2(6), pp.584–96.

Garigan, D. et al., 2002. Genetic analysis of tissue aging in Caenorhabditis elegans: a role for heat-shock factor and bacterial proliferation. *Genetics*, 161(3), pp.1101–12.

Ghosh, S.K.B., Missra, A. & Gilmour, D.S., 2011. Negative Elongation Factor Accelerates the Rate at Which Heat Shock Genes Are Shut off by Facilitating Dissociation of Heat Shock Factor. *Molecular and Cellular Biology*, 31(20), pp.4232–4243.

Gilchrist, D.A. et al., 2008. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes & Development*, 22(14), pp.1921–1933.

Gilmour, D.S. & Lis, J.T., 1986. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. , 6(11).

Gregoire, S. & Yang, X.-J., 2005. Association with Class IIa Histone Deacetylases Upregulates the Sumoylation of MEF2 Transcription Factors. *Molecular and Cellular Biology*, 25(6), pp.2273–2287.

Greifenberg, A.K. et al., 2016. Structural and Functional Analysis of the Cdk13/ Cyclin K Complex.

Hah, N. et al., 2011. A Rapid, Extensive, and Transient Transcriptional Response to Estrogen Signaling in Breast Cancer Cells. *Cell*, 145(4), pp.622–634.

Hashimshony, T. et al., 2003. The role of DNA methylation in setting up chromatin structure during development. *Nature Genetics*, 34(2), pp.187–192.

Hazelbaker, D.Z. & Buratowski, S., 2012. Transcription: Base J Blocks the Way. *Current Biology*, 22(22), pp.R960–R962.

He, N., Pezda, A.C. & Zhou, Q., 2006. Modulation of a P-TEFb functional equilibrium for the global control of cell growth and differentiation. *Molecular and cellular biology*, 26(19), pp.7068–76.

Henriques, T. et al., 2013. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Molecular cell*, 52(4), pp.517–28.

Henriques, T. & Adelman, K., 2013. Catching the Waves: Following the Leading Edge of Elongating RNA Polymerase II. *Molecular Cell*, 50(2), pp.159–160.

Hoskins, R.A. et al., 2011. Genome-wide analysis of promoter architecture in Drosophila melanogaster. *Genome research*, 21(2), pp.182–92.

Howe, K.J., Kane, C.M. & Ares, M., 2003. Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. *RNA (New York, N.Y.)*, 9(8), pp.993–1006.

HsuChen, C.C. & Dubin, D.T., 1976. Di-and trimethylated congeners of 7-methylguanine in Sindbis virus mRNA. *Nature*, 264(5582), pp.190–1.

Imai, Y. et al., 2001. An unfolded putative transmembrane polypeptide, which can lead to endoplasmic reticulum stress, is a substrate of Parkin. *Cell*, 105(7), pp.891–902.

Jia, D. et al., 2007. Systematic identification of non-coding RNA 2,2,7-trimethylguanosine cap structures in Caenorhabditis elegans. *BMC molecular biology*, 8, p.86.

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, p.e02407.

Jove, R. & Manley, J.L., 1984. In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites. *The Journal of biological chemistry*, 259(13), pp.8513–21.

Kadonaga, J.T., 2004. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, 116(2), pp.247–57.

Konarska, M.M., Padgett, R.A. & Sharp, P.A., 1984. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell*, 38(3), pp.731–6.

Kuehner, J.N., Pearson, E.L. & Moore, C., 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews. Molecular cell biology*, 12(5), pp.283–94.

Kwak, H. et al., 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science (New York, N.Y.)*, 339(6122), pp.950–3.

de la Mata, M. et al., 2003. A slow RNA polymerase II affects alternative splicing in vivo. *Molecular cell*, 12(2), pp.525–32.

Lee, H. et al., 1992. DNA sequence requirements for generating paused polymerase at the start of hsp70. *Genes & development*, 6(2), pp.284–95.

Levine, M., 2011. Paused RNA polymerase II as a developmental checkpoint. *Cell*, 145(4), pp.502–11.

Lindquist, S., 1986. The Heat-Shock Response. *Annual Review of Biochemistry*, 55(1), pp.1151–1191.

Lis, J.T. et al., 2000. P-TEFb kinase recruitment and function at heat shock loci. *Genes & development*, 14(7), pp.792–803.

Lund, J. et al., 2002. Transcriptional profile of aging in C. elegans. *Current biology : CB*, 12(18), pp.1566–73.

MacDonald, C.C., Wilusz, J. & Shenk, T., 1994. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Molecular and cellular biology*, 14(10), pp.6647–54.

Mahat, D.B. et al., 2016. Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Molecular Cell*, 62(1), pp.63–78.

Matsui, T. et al., 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *The Journal of biological chemistry*, 255(24), pp.11992–6.

Mayo, M.W. et al., 2003. Ineffectiveness of Histone Deacetylase Inhibitors to Induce Apoptosis Involves the Transcriptional Activation of NF- B through the Akt Pathway. *Journal of Biological Chemistry*, 278(21), pp.18980–18989.

McCracken, S. et al., 1997. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes & development*, 11(24), pp.3306–18.

Mendillo, M.L. et al., 2012. HSF1 drives a transcriptional program distinct from heat shock to support highly malignant human cancers. *Cell*, 150(3), pp.549–62.

Mizumoto, K. & Kaziro, Y., 1987. Messenger RNA capping enzymes from eukaryotic cells. *Progress in nucleic acid research and molecular biology*, 34, pp.1–28.

Moody, S.E. et al., 2005. The transcriptional repressor Snail promotes mammary tumor recurrence. *Cancer Cell*, 8(3), pp.197–209.

Muse, G.W. et al., 2007. RNA polymerase is poised for activation across the genome. *Nature genetics*, 39(12), pp.1507–11.

Nag, A., Narsinh, K. & Martinson, H.G., 2007. The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nature structural & molecular biology*, 14(7), pp.662–9.

Natarajan, M. et al., 2013. Negative Elongation Factor (NELF) Coordinates RNA Polymerase II Pausing, Premature Termination, and Chromatin Remodeling to Regulate HIV Transcription. *Journal of Biological Chemistry*, 288(36), pp.25995–26003.

Nechaev, S. et al., 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), pp.335–8.

Ni, Z. et al., 2004. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Molecular cell*, 13(1), pp.55–65.

Ni, Z. et al., 2008. P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo. *Molecular and cellular biology*, 28(3), pp.1161–70.

Oyadomari, S., Araki, E. & Mori, M., 2002. Endoplasmic reticulum stress-mediated apoptosis in pancreatic beta-cells. *Apoptosis : an international journal on programmed cell death*, 7(4), pp.335–45.

Ozaki, K. et al., 2006. Blockade of the ERK pathway markedly sensitizes tumor cells to HDAC inhibitor-induced cell death. *Biochemical and Biophysical Research Communications*, 339(4), pp.1171–1177.

Peterlin, B.M. et al., 2006. Controlling the elongation phase of transcription with P-TEFb. *Molecular cell*, 23(3), pp.297–305.

Pleiss, J.A. et al., 2007. Rapid, transcript-specific changes in splicing in response to environmental stress. *Molecular cell*, 27(6), pp.928–37.

Pokholok, D.K., Hannett, N.M. & Young, R.A., 2002. Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Molecular cell*, 9(4), pp.799–809.

Price, D.H., 2000. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Molecular and cellular biology*, 20(8), pp.2629–34.

Rahl, P.B. et al., 2010. c-Myc Regulates Transcriptional Pause Release. *Cell*, 141(3), pp.432–445.

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*, 90(17), pp.7923–7927.

Renner, D.B. et al., 2001. A Highly Purified RNA Polymerase II Elongation Control System. *Journal of Biological Chemistry*, 276(45), pp.42601–42609.

Rosato, R.R., Almenara, J.A. & Grant, S., 2003. The histone deacetylase inhibitor MS-275 promotes differentiation or apoptosis in human leukemia cells through a process regulated by generation of reactive oxygen species and induction of p21CIP1/WAF1 1. *Cancer research*, 63(13), pp.3637–45.

Rozenberg, J.M. et al., 2008. All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. *BMC Genomics*, 9(1), p.67.

Ruefli, A.A. et al., 2001. The histone deacetylase inhibitor and chemotherapeutic agent suberoylanilide hydroxamic acid (SAHA) induces a cell-death pathway characterized by cleavage of Bid and production of reactive oxygen species. *Proceedings of the National Academy of Sciences*, 98(19), pp.10833–10838.

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

Saunders, A. et al., 2003. Tracking FACT and the RNA Polymerase II Elongation Complex Through Chromatin in Vivo. *Science*, 301(5636), pp.1094–1096.

Schwer, B. & Shuman, S., 1996. Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing in vivo. *RNA (New York, N.Y.)*, 2(6), pp.574–83.

Shukla, S. & Oberdoerffer, S., 2012. Co-transcriptional regulation of alternative pre-mRNA splicing. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(7), pp.673–683.

Sonenberg, N. et al., 1978. A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 75(10), pp.4843–7.

Sørensen, J.G., Kristensen, T.N. & Loeschcke, V., 2003. The evolutionary and ecological role of heat shock proteins. *Ecology Letters*, 6(11), pp.1025–1037.

Srebrow, A. & Kornblihtt, A.R., 2006. The connection between splicing and cancer. *Journal of Cell Science*, 119(13), pp.2635–2641.

Sun, J. & Li, R., 2010. Human negative elongation factor activates transcription and regulates alternative transcription initiation. *The Journal of biological chemistry*, 285(9), pp.6443–52.

Takagaki, Y. & Manley, J.L., 1997. RNA recognition by the human polyadenylation factor CstF. *Molecular and cellular biology*, 17(7), pp.3907–14.

Terro, F. et al., 2002. Neurons overexpressing mutant presenilin-1 are more sensitive to apoptosis induced by endoplasmic reticulum-golgi stress. *Journal of Neuroscience Research*, 69(4), pp.530–539.

Terzi, N. et al., 2011. H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. *Molecular and cellular biology*, 31(17), pp.3569–83.

Titov, D. V et al., 2011. XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nature Chemical Biology*, 7(3), pp.182–188.

Vakoc, C.R. et al., 2005. Histone H3 Lysine 9 Methylation and HP1γ Are Associated with Transcription Elongation through Mammalian Chromatin. *Molecular Cell*, 19(3), pp.381–391.

Vos, S.M. et al., 2016. Architecture and RNA binding of the human negative elongation factor. *eLife*, 5.

Wada, T. et al., 2000. FACT relieves DSIF/NELF-mediated inhibition of transcriptional elongation and reveals functional differences between P-TEFb and TFIIH. *Molecular cell*, 5(6), pp.1067–72.

Wagschal, A. et al., 2012. Microprocessor, Setx, Xrn2, and Rrp6 Co-operate to Induce Premature Termination of Transcription by RNAPII. *Cell*, 150(6), pp.1147–1157.

Wang, E.T. et al., 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221), pp.470–476.

Warns, J.A., Davie, J.R. & Dhasarathy, A., 2016. Connecting the dots: chromatin and alternative splicing in EMT. *Biochemistry and Cell Biology*, 94(1), pp.12–25.

Warzecha, C.C. & Carstens, R.P., 2012. Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal transition (EMT). *Seminars in Cancer Biology*, 22(5–6), pp.417–427.

Wood, A. et al., 2003. The Paf1 Complex Is Essential for Histone Monoubiquitination by the Rad6-Bre1 Complex, Which Signals for Histone Methylation by COMPASS and Dot1p. *Journal of Biological Chemistry*, 278(37), pp.34739–34742.

Xiao, X. et al., 1999. HSF1 is required for extra-embryonic development, postnatal growth and protection during inflammatory responses in mice. *The EMBO journal*, 18(21), pp.5943–52.

Yamaguchi, Y. et al., 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, 97(1), pp.41–51.

Ye, Q. et al., 2001. BRCA1-induced large-scale chromatin unfolding and allele-specific effects of cancer-predisposing mutations. *The Journal of cell biology*, 155(6), pp.911–21.

Yost, H.J. & Lindquist, S., 1986. RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. *Cell*, 45(2), pp.185–93.

Yu, C. et al., 2007. Abrogation of MAPK and Akt Signaling by AEE788 Synergistically Potentiates Histone Deacetylase Inhibitor-Induced Apoptosis through Reactive Oxygen Species Generation. *Clinical Cancer Research*, 13(4), pp.1140–1148.

APPENDIX A

ABBREVIATIONS

7-mG cap	7-methylguanosine cap
ATP	Adenosine triphosphate
CBC	Cap Binding Complex
CBP	Cap Binding Protein
CDK	Cyclin-dependent Kinase
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation sequencing
CPSF	cleavage and polyadenylation specificity factor
CTD	Carboxyl-Terminal Domain
DNA	Deoxyribonucleic Acid
DRB	5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole
DSIF	(DRB) Sensitivity-Inducing Factor
E2	17β-Estradiol
EMT	Epethelial to Mesenchymal Transition
FACT	Facilitates Chromatin Transcription
FP	Flavapiridol
GRO	Global Run-on
GRO-seq	Global Run-on-sequencing
GTF	General Transcription Factors
GTP	Guanosine triphosphate
HIV	Human Influenza Virus
HSE	Heat Shock Element
HSF	Heat Shock Facor
HSP70	Heat Shock Protein 70
HSR	Heat Shock Response
IL-8	Interleukin-8

MAPK	Mitogen-Activated Protein Kinase
mG cap	methylguanosine
min	minutes
mRNA	messenger RNA
NELF	Negative Elongation Factor
NF-κB	Neclear Factor Kappa Beta
nt	nucleotide
PAF	Polymerase II Asssociated Factor
PCF11	Cleavage and Polyadenylation Factor subunit
PCR	Polymerase Chain Reaction
PI3K	Phosphoinositide 3-kinase
PIC	Pre-initiation Complex
Pol II	RNA Polymerase II
Pro-seq	Precision nuclear run-on sequencing
p-TEFb	Positive Transcription Elongation Factor
RNA	Ribonucleic Acid
RT	Reverse Transcription
RT-PCR	Reverse Transcription- Polymerase Chain Reaction
SAM	S-adenosylmethionine
scRNA	Short capped RNA
scRNA-seq	Short capped RNA sequencing
snRNP	small nuclear ribonucleic proteins
TES	Transcription End Site
TF	Transcription Factor
TNF-α	Tumor Necrosis Factor alpha
TRP	Triptolide
TSA	Trichostatin A
TSS	Transcription Start Site
TTF	Transcription termination Factor
XRN-2	5'-3' exoribonuclease 2

APPENDIX B

LIST OF PRIMERS

hCDH1+272A	tttcttggaagaagggaag
hCDH1+272B	ttcttggaagaagggaagc
hCDH1+272C	agaagggaagcggtgacgac
hSnail+210A	caatggtccacaaaacatc
hSnail+210B	acatcctgtgactcgatcct
hSnail+210C	tgtctcccccaaacctcctg
hSlug+225A	tttgcaaagctctagatacg
hSlug+225B	tgtccagttcgctgtagttt
hSlug+225C	ggctttttggaggcgttgaa
mSnail+259A	agctccagatccacctgt
mSnail+259B	gctccagatccacctgtc
mSnail+259C	gcagagcccgattctcacct