# The Role of NELF in Mediating Human Gene Expression

by

Jillian Boden Daigneault

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cancer Biology)
in The University of Michigan
2017

Doctoral Committee:
        Professor Vivian G. Cheung, Co-Chair
        Associate Professor JoAnn Sekiguchi, Co-Chair
        Professor Mark Day
        Professor Eric Fearon
        Associate Professor David O. Ferguson
        Professor Theodore Lawrence

Jillian Boden Daigneault
jmboden@umich.edu
ORCID iD: 0000-0003-0205-349

# Dedication

To my husband Chris, without your unwavering support I would not have made it.

# Acknowledgements

Thank you first and foremost to my PhD mentor, Dr. Vivian Cheung. I have learned so much from you over the past few years. You have taught me how to "pause" before answering questions and how to effectively communicate my science in any setting. Thank you for always encouraging me to pursue my career goals. I would not be where I am today without your help and guidance.

Thank you to the members of my dissertation committee (Drs. Day, Fearon, Ferguson and Lawrence) for all of your advice and support. I would also like to thank my thesis co-chair, Dr. JoAnn Sekiguchi, for helping me navigate not only thesis writing but also the entire graduate school experience. Thank you to the entire Cancer Biology Program, it has been a pleasure to be a member of such a supportive group of students and faculty. I would also like to thank the Cancer Biology Program Director, Dr. Beth Lawlor, for all of your support and guidance throughout my time in graduate school. Thank you to Zarinah Aquil for all of the wonderful coffee meetings, they made my last few years of graduate school that much more enjoyable.

I would also like to thank all the members of the Cheung lab both past and present. Thank you to Xiaorong Wang for all of your guidance and patience. You always made time for my questions (no matter how small) and I know I would not have completed this thesis without all of your help. Thank you to Zhengwei Zhu who has helped me so much throughout my thesis work even though I was a slow learner when it came to all of the bioinformatics. Thank you to Jason Watts for all of your thoughtful comments on my figures and presentations and for making ChIP

# Table of Contents

# List of Figures

# List of Tables

# Abstract

Transcription is the synthesis of RNA from a DNA template. Transcription has been shown to occur in three main steps, initiation, elongation and termination. Additional regulatory steps include a rate limiting step between initiation and elongation of RNA polymerase II (RNA Pol II) termed promoter-proximal pausing of RNA polymerase. Dysregulation at any stage of this process can lead to cancer development. There has been extensive work focusing on how drastic alterations to the transcription process contributes to cancer development; however, there has been less focus on how the fine-tuning steps of transcription regulation contribute to cancer phenotypes.  Specifically, my thesis focuses on understanding the contributions of NELF-mediated pausing to cancer development.

To begin, I focused on primary cells in order to better understand the function of NELF-mediated pausing in a normal context. My work has identified a global profile of paused genes in human cells. Following the completion of this profile, I wanted to identify pathways that were enriched for paused genes. I found that paused genes are involved in cancer associated pathways such as EGFR and TGFβ signaling as well as general transcription and gene expression. After identifying human genes with paused RNA polymerases, I focused on studying NELF.  I found that almost all genes with NELF enrichment are paused. A majority of these genes also require NELF to maintain the pause. I then wanted to investigate the biological consequences of NELF-mediated pausing. I found that the expression of more than half of NELF-mediated

paused genes are regulated by NELF. These genes include those involved in the cancer-associated pathways EGFR, TGFβ and VEGF signaling. I showed that besides localizing to gene promoters, NELF also targets other genomic regions. I found that approximately half of these sites are in enhancers, specifically those enhancers that are actively transcribed.  We showed that the abundance of these eRNAs is dependent on NELF. Finally, we found that loss of transcription of these eRNA result in changes in expression of nearby genes.

Upon identifying NELF as a key regulator of gene expression in normal cells, I investigated how this process contributed to the gene expression changes found in cancer. To begin, I identified paused genes in HT29 colorectal cancer cells. I found that many of the same pathways show pausing in both normal cells and cancer cells. I also found that most genes utilize pausing to reduce their expression in cancer cells. These genes were enriched for cellular senescence, cellular stress response and WNT signaling.

My work provides insight into the role of RNA Pol II pausing in cancer development. I have shown that cancer-associated pathways, such as EGFR, are regulated by RNA Pol II pausing.  I have also generated global profiles of pausing in both normal and cancer cells. In comparing these two profiles, I have identified differences in the function of pausing between normal and cancer cells. One of the key differences I identified is the mechanism by which pausing regulates gene expression in normal and cancer cells. Collectively, my work shows that NELF and NELF-mediated pausing are key regulators of gene expression that may be harnessed by cancer cells to promote tumorigenesis.

**Chapter 1: Introduction**

**Transcription Overview**

Transcription is a complex process in which RNA is transcribed from the corresponding DNA sequence. Transcription is mediated by three different RNA polymerases, RNA Polymerase I, II and III (Roeder and Rutter, 1969). These polymerases all share some structural similarities; however, each is responsible for transcribing different types of genes. RNA Polymerase I and III are responsible for transcribing genes that encode for transfer RNAs, ribosomal RNAs and small RNAs whereas RNA Polymerase II (RNA Pol II) is responsible for transcribing a majority of genes including protein coding genes (Weinmann et al., 1974, Winemann et al., 1974[2]). Transcription by RNA Pol II will be the focus of this thesis.

In the most simplified view, transcription begins with recruitment of RNA Pol II to genes (Roeder, 1996). RNA Pol II is directed to the start sites of genes by both sequence features as well as the general transcription factors (Roeder, 1996). The promoter region is one of the sequence features that aids in directing RNA Pol II to the start site of genes. Promoter regions are made up of an asymmetrical consensus sequence that is bound by the general transcription factors (Roeder, 1996). Upon binding, the general transcription factors recruit RNA Pol II to

form a pre-initiation complex (Figure 1.1). Following the formation of this complex, the DNA is unwound to grant RNA Pol II access to the template strand.

The next step in transcription is the initiation of the polymerase. RNA Pol II is made up of multiple subunits. The largest of which, Rpb1, contains a "tail" that is termed the C-terminal domain (CTD). The CTD contains a series 52 of heptad repeats that are able to be modified throughout transcription (Dahmus, 1981). The first of these modifications is the phosphorylation by TFIIH, a general transcription factor, of the serine 5 residues in the CTD (Roeder, 1996). This phosphorylation creates a conformational change in the polymerase and allows for the binding of RNA modifying proteins (Sims et al., 2004). Following this phosphorylation event, the polymerase is initiated and can disengage from the pre-initiation complex to begin transcribing.

Once the polymerase has left the pre-initiation complex (PIC), it is able to transcribe a short stretch of RNA, approximately 20-60 nucleotides, and then pause just downstream of the TSS (Gilmour and Lis, 1986). The polymerase is held in this paused state by the pausing factors, negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF) (Figure 1.1) (Yamaguchi et al., 1999, Yamaguchi et al., 2002, Wu et al., 2003, Wu et al., 2005). This is termed promoter-proximal pausing and has been shown to serve as the rate-limiting step between initiation and elongation (Min et al, 2011). Upon a signal, positive transcription elongation factor b (P-TEFb) is recruited to the paused complex where it phosphorylates both pausing factors as well as the serine 2 residue of RNA Pol II's CTD (Figure 1.2) (Kim and Sharp, 2001, Fujinaga et al., 2004, Yamada et al., 2006, Marshall et al., 1996). These phosphorylation events cause NELF to dissociate from the paused complex, DSIF to transform into a positive elongation factor and RNA Pol II to enter into active elongation (Figure 1.2) (Zhou et al., 2012).

Following the release from pausing, RNA Pol II enters into active elongation. Elongation is facilitated by a number of factors that aid in ensuring smooth transcription as well as performing co-transcriptional mRNA processing. RNA Pol II is able to elongate at approximately 3.8kb/min for up to 2 million base pairs (Singh and Padget, 2009). This rate is aided by the general transcription factors TFIIS and TFIIF. TFIIS aids in restarting stalled or backtracked RNA Pol II and TFIIF stimulates the elongation rate of RNA Pol II (Nechaev et al., 2010, Price et al., 1989, Flores et al., 1989). Other factors such as the Super Elongation Complex (SEC) and the Polymerase II associated factor complex (PAFc) also play a role in stimulating elongation by RNAPol II (Zhou et al, 2012). Components of the SEC aid the polymerase in overcoming temporary pauses in transcription whereas PAFc modifies histones to facilitate elongation (Zhou et al, 2012, Kim et al., 2010). These factors have also been shown to play a role in pause release and will be discussed in greater detail below.

During elongation, the mRNA is able to be processed. 5' Capping, splicing, 3' end cleavage and addition of the PolyA tail all occur co-transcriptionally. These processes are facilitated by the CTD of RNA Pol II where many of the enzymes required for these functions bind (Ho et al., 1999). The maturation of the mRNA begins with the addition of the 7-methylguanylate cap to the 5' end. Capping is performed by the cap binding complex (CBC) following phosphorylation of the serine 5 residues of the CTD (Moteki and Price, 2002). Capping typically occurs after transcription of about 25-30 nucleotides, thus placing it in the same window as RNA Pol II pausing. It has been shown that DSIF is able to stimulate the capping process and that the CBC is able to aid in the release of NELF from the paused RNA Pol II complex (Wen and Shatkin, 1999, Mandal et al., 2004).

The next step in the maturation process of the mRNA is removal of introns through splicing. Similar to the capping process, splicing is facilitated by modifications to the CTD of RNA Pol II. Phosphorylation of the serine 2 residue of the CTD, aids in the binding of splicing factors such as the Ser/Arg rich (SR) proteins (Das et al., 2007). Other components of the elongation complex have also been shown to play a role in mediating splicing, including P-TEFb (Zhou et al, 2012).

The final steps in the mRNA maturation process are the 3' cleavage of the transcript and the addition of the PolyA tail. Cleavage of the transcript occurs approximately 10-35 nucleotides downstream of the AAUAA consensus sequence (Zhou et al, 2012). Following this cleavage 200-300 adenosine nucleotides are added to the 3' end of the transcript (Zhou et al, 2012). These processes are facilitated by the CTD of RNA Pol II, specifically the serine 2 phosphorylated form (Shell et al., 2007). Other proteins in the elongation complex including, components of PAFc and SEC have been shown to facilitate this process as well (Nagaike, et al., 2011, Martinicic et al., 2009). The pausing factor, NELF, has also been shown to play a role in 3' processing RNAs. Specifically, it has been shown that NELF is responsible for the correct processing of the 3' end of histone mRNAs (Narita et al., 2007).

This thesis will focus specifically on NELF and its role in mediating RNA Pol II pausing and gene expression.

## Identification of RNA Polymerase II Pausing

Early studies of the c-myb, β-globin and hsp70 genes showed an accumulation of RNA Pol II at the 5' end. These polymerases were shown to be in a transcriptionally active state. This phenomenon was termed promoter-proximal pausing (Bender et al., 1987, Gariglio et al, 1981,

Rougvie and Lis, 1988. Further studies on *hsp70* showed that RNA Pol II pausing serves as a rate limiting step between initiation and elongation and that the expression of *hsp70* was regulated by this pause (Rougvie and Lis, 1990, O'Brien and Lis, 1991). It was shown that upon heat shock RNA Pol II released from pausing and that pausing was required to return *hsp70* expression to baseline following heat shock. They identified that, outside of *Drosophila*, a 5' accumulation of RNA Pol II, similar to that seen at *hsp70* was found at two human genes, *FOS* and *MYC* (Plet et al., 1995, Krumm et al., 1992, Strobl and Eick, 1992).

Outside of these single gene examples, it was shown that RNA Pol II tends to accumulate at the 5' end of both *Drosophila* (Muse et al., 2007, Zeitlinger et al., 2007) and human genes (Kim et al., 2005). It was originally thought that these RNA Pol II molecules represented a PIC (Kim et al., 2005). However, work in both *Drosophila* and humans showed that many of these RNA Pol II molecules had already undergone transcription initiation and entered into elongation (Muse et al., 2007, Zeitlinger et al., 2007), similar to the single gene examples identified previously (Rougvie and Lis 1988, Rougvie and Lis, 1990, Plet et al., 1995, Krumm et al., 1992, Strobl and Eick, 1992). In *Drosophila* specifically, 21% of genes were shown to have an accumulation of RNA Pol II at their 5' end (Law et al., 1998). This was identified by cloning and characterizing the DNA that was crosslinked to RNA Pol II following UV treatment. Further characterization of these regions showed that the RNA Pol II bound to them was able to resume transcription in a nuclear-run on assay, thus identifying them as transcriptionally engaged and therefore paused. (Law et al., 1998). Earlier work examining transcription elongation showed that RNA Pol II struggles in early elongation with stalling or termination occurring within the first 100 nucleotides of the gene (Kephart et al., 1992, Marshall and Price, 1992) with more fine mapping showing that the polymerase is typically found within the first 20-60 nucleotides

(Gilmour and Lis, 1986). It was not clear whether these polymerases were able to resume transcription or if they would undergo premature termination.

For a gene to truly be paused, it must be able to resume active transcription (Adelman and Lis, 2012). Core et al developed a modified nuclear-run on assay, global run-on sequencing (GRO-seq), to map the location of polymerases that were able to resume active transcription *in vitro* (Core et al., 2008). GRO-seq utilizes a nuclear-run on assay in the presence of sarkosyl to stimulate nascent transcription. During the run-on 5-bromouridine 5' triphosphate (BrUTP) is incorporated to the nascent transcripts. This allows for the specific isolation of nascent transcripts, through immunoprecipitation with an antibody specific to this nucleotide analog. Following isolation, the transcripts are made into a cDNA library which is then sequenced by the Illumina HiSeq platform (Core et al., 2008). This technique allowed for a clear distinction between stalled or terminating polymerases and paused polymerases. Application of the GRO-seq technique showed that approximately 30% of *Drosophila*, mouse and human promoters contain a paused polymerase (Core et al., 2008, Min et al., 2011, Larschan et al., 2011).

## Mechanism of RNA Polymerase II Pausing and Pause Release

RNA Pol II pausing in metazoans has been shown to be mainly mediated by two factors, DSIF and NELF (Wada et al., 1998, Yamaguchi et al., 1999, Wu et al., 2003, Wu et al., 2005). DSIF is comprised of two subunits, SPT4 and SPT5 and is conserved from yeast to human (Wada et al., 1998). It is able to function as both a pausing factor and a positive elongation factor through the direct interaction of the SPT5 subunit with RNA Pol II (Wada et al., 1998).

NELF is a 4-subunit complex, comprised of NELFA, NELFB, NELFC/D and NELFE (Figure 1.3) (Wu et al., 2005, Yamaguchi et al., 1999). *In vitro* studies show that NELF directly

6

interacts with RNA Pol II through the A subunit and the nascent RNA through the B, C and E subunits (Figure 1.3) (Yamaguchi et al., 1999, Yamaguchi et al., 2002, Narita et al 2003, Vos et al., 2016). It is also thought that NELF interacts with DSIF, though the subunit that mediates this interaction is unclear (Yamaguchi et al., 1999). Upon the loss of a single subunit, the complex becomes unstable and expression of all subunits is lost at the protein level (Yamaguchi et al., 2002, Sun et al., 2008). Complete loss of the NELF complex through the knockout of *Nelfb* in mice results in an embryonic lethal phenotype which may be due to dysregulation of the WNT pathway due to an overexpression of Lef1 upon loss of NELF (Amleh et al., 2009). Silencing of NELF in Drosophila S2 cells results in a decrease of RNA Pol II at promoters as well as the assembly of nucleosomes at a specific subset of genes (Core et al., 2012). Unlike DSIF, NELF is found only in metazoans with no homologs in Arabidopsis, yeast or *C. elegans* (Narita et al., 2003).

In *Drosophila*, RNA Pol II has been shown to utilize both DNA sequence and a third factor to establish the pause. The GAGA element, originally identified in the promoter of the *hsp70* gene, has been shown to be required for establishing RNA Pol II pausing (Lee et al., 1992). The GAGA factor, which binds the GAGA element has been found at the promoters of paused genes along with NELF in *Drosophila* (Lee et al., 2008). However, the GAGA element and the GAGA factor are not found in mammals.

Pause release is mediated by P-TEFb. P-TEFb is comprised of two subunits, Cyclin T1 and CDK9 (Marshall and Price 1995). P-TEFb is sequestered in complex with 7SK snRNA, HEXIM1 or 2, LARP7 and MEPCE (Zhou et al., 2012). Upon a signal, P-TEFb is released from this complex which allows P-TEFb to interact with BRD4 or the Super Elongation Complex (SEC) (Lu et al., 2016, Luo et al., 2012). P-TEFb has also been shown to be recruited to

promoters with paused polymerases via c-MYC, JMJD6 and TRIM28 (Rahl et al., 2010, Lu et al., 2016, Bunch et al., 2014). Because c-MYC is so frequently overexpressed in cancers, I will focus on the role of c-MYC in facilitating pause release. c-MYC has been shown to co-localize with P-TEFb at actively transcribed genes and upon loss of c-MYC, there is an increase in pausing index (PI) at its target genes (Rahl et al., 2010). There is also a decrease in the amount of serine 2 phosphorylated RNA Pol II, confirming that there is a defect in the recruitment of P-TEFb (Rahl et al., 2010). These findings provide a second mechanism by which over-expression c-MYC can contribute to the gene expression changes seen in many tumors. The ability of c-MYC to contribute to these changes by altering the initiation step is described later. Both JMJD6 and TRIM28 function to recruit P-TEFb to facilitate pause release and function in a similar manner to c-MYC. Once recruited, P-TEFb can then phosphorylate NELFE, the C-terminal domain of SPT5 and the serine-2 residue of the RNA Pol II CTD to facilitate the transition into active elongation (Kim and Sharp, 2001, Fujinaga et al., 2004, Yamada et al., 2006, Marshall et al., 1996). It has also been shown that NELFE can be ADP-ribosylated by PARP1 prior to phosphorylation by P-TEFb and that this ADP-ribosylation increases the efficiency of P-TEFb mediated pause release (Gibson et al., 2016).

Recently, it has been shown that RNA Polymerase II-associated factor 1 (PAF1) plays a role in RNA Pol II pausing and pause release. Chen et al. found that PAF1 is associated with RNA Pol II at the promoters of paused genes and that upon loss of PAF1, there is a release of RNA Pol II into the gene bodies mediated by the SEC (Chen et al., 2015). Yu et al. showed that P-TEFb recruits PAF1 to the paused RNA Pol II complex where it subsequently recruits CDK12 to facilitate pause release (Yu et al., 2015).

RNA Pol II pausing and pause release has also been shown to be regulated by epigenomic factors. Kapoor-Vazirani et al demonstrated that RNA Pol II pause release was facilitated by acetylation of the lysine 16 residue of histone H4 (H4K16ac) by the MSL complex, specifically h-MOF (2011). This histone modification aided in the recruitment of BRD4 as well as P-TEFb. This release from pausing was able to be blocked through tri-methylation on lysine 20 of histone H4 (H4K20me3) by SUV420H2. The creation of this mark inhibits the ability of h-MOF to create the H4K16ac mark and leads to a decrease in BRD4 and P-TEFb recruitment to target genes (Kapoor-Vazirani et al., 2011). Further studies on this phenomenon showed that the mono-methylation of lysine 20 of histone H4 (H4K20me1) by PR-SET7/SETD8 is responsible for the recruitment of the MSL complex and subsequent creation of the H4K16ac mark and release of paused RNA Pol II (Kapoor-Vazirani and Vertino, 2014). H4K20me1, which is created by PR-SET7, has also been shown to be necessary for the creation of the H4K20me3 mark which induces pausing (Kapoor-Vazirani and Vertino, 2014). Thus, implicating H4K20me1 as a key regulator of RNA Pol II pausing and pause release.

## RNA Polymerase II Pausing and Gene Expression

The earliest studies of pausing have shown that pausing plays a key role in regulating the expression of the *hsp70* gene in *Drosophila* as well as the human genes *FOS*, *MYC* and *JUNB* (O'Brien and Lis, 1991, Plet et al., 1992, Fivaz et al., 2000, Krumm et al., 1995, Aida et al., 2006). The expression of each of these genes have shown to be dependent on the presence of a paused polymerase at the promoter region. Min et al. showed that pausing serves as the rate limiting step between initiation and elongation for over half of all actively transcribed genes, thus implicating pausing as a regulator of global gene expression (2011). This altered the

standing theory that recruitment and initiation of the polymerase served as the most important steps in the regulation of gene expression.

In *Drosophila*, thousands of genes have been shown to have their expression regulated through RNA Pol II pausing (Gilchrist, et al., 2010). Many of these genes are involved in stimulus response and innate immunity (Glichrist, et al., 2010). Gilchrist et al. described two possible mechanisms for pausing-mediated gene expression. The first is that the polymerase is held in a paused confirmation to allow for rapid induction of the gene upon a signal. At baseline, these genes are expressed at a very low level, but upon loss of pausing, these genes increase in expression. The second is that the paused polymerase serves to clear the first nucleosome. This allows for the rapid recruitment and re-initiation of subsequent RNA Pol II. Upon loss of pausing, these genes show a decrease in expression due to the assembly of the nucleosome (Gilchrist et al., 2010). Further work in *Drosophila* has shown that pausing is able to influence development through mediating the expression of genes involved in embryo patterning (Wang et al., 2010).

In mammals, Adelman et al. found that pausing allowed for a coordinated transcriptional response of genes involved in the inflammatory response, similar to what was found in the *Drosophila* embryo (Adelman et al, 2009, Danko et al 2013). Patel et al expanded upon this study and found that BRD4 was responsible for coordinating the release from pausing by P-TEFb and subsequent expression of interferon-responsive genes (2013).

Further studies of the role of pausing in transcription and gene expression have focused almost exclusively on NELF. Because the NELFB subunit was originally characterized as a co-factor of BRCA1 (COBRA1), interest in understanding of NELF and subsequently pausing

influenced gene expression centered around the estrogen response (Ye et al., 2001). Aiyar et al found that NELF directly interacts with the estrogen receptor and represses estrogen mediated gene expression through RNA Pol II pausing (Aiyar et al 2004). Further studies on this response showed that paused RNA Pol II is commonly found at the promoters of estrogen receptor target genes and that this pause serves to create a coordinated transcriptional response to estrogen stimulation (Kininis et al, 2009). Building upon these studies, Luo et al. showed that NELF can also directly interact with the glucocorticoid receptor to mediate expression of target genes *in vitro* (2013).

Further studies into the role of pausing and gene expression have shown that NELF is able to influence pausing and gene expression independently of its interactions with nuclear hormone receptors. Specifically, genes involved in the cell cycle as well as cell growth have been shown to be regulated through NELF-mediated pausing (Sun and Li, 2010, Sun et al., 2011, Williams et al., 2015). Loss of NELF also results in defects in cell cycle progression (Sun and Li, 2010, Sun et al., 2011). NELF-mediated pausing has also been shown to regulate the expression of genes involved in signaling pathways (Williams et al., 2015). Specifically, loss of NELF causes embryonic stem cells to become resistant to FGF/ERK signaling due to dysregulation of key signaling pathway components (Williams et al., 2015). Pausing has been shown to facilitate the coordinated response to external stimuli such as estrogen stimulation, inflammation and development.

## Enhancers and Enhancer RNAs

Enhancer regions have been broadly described to alter the transcription of target genes independent of their position or orientation to the target gene (Maniatis et al., 1987). From this

initial definition, enhancers have been further characterized based on their epigenetic signatures, RNA Pol II and transcription factor binding. Several hundreds of thousands of enhancer elements have been identified in the human genome; however, not all enhancers are utilized in each cell type (ENCODE Project Consortium, 2012, Heinz, et al., 2015).

Active enhancer regions have been shown to have an acetylation mark on histone H3 lysine 27 (H3K27ac), binding by lineage determining transcription factors (LDTFs), collaborating transcription factors (CTFs), signal-dependent transcription factors (STDFs) and RNA Pol II (Creyghton et al., 2010, Heinz et al., 2015, Koch et al., 2011). Over 43,000 active enhancers have been found across hundreds of human tissues and cell-types (Andersson et al., 2014). Transcription from these enhancers was originally described in the locus control region (LCR) of the beta-globin gene cluster (Ashe et al., 1997). More recent studies have identified transcription generating from active enhancer regions throughout the genome in macrophages, breast and prostate cancer cells (Hah et al, 2011, Lam et al., 2014). The transcripts generating from these active enhancer elements are termed eRNAs (Kim et al., 2010). Genome-wide studies of eRNAs by GRO-seq and RNA sequencing (RNA-seq) have shown that eRNAs can either be polyadenylated or non-polyadenylated and also that they can be either unidirectional or bidirectional transcripts (Lam et al., 2014).

The primary function of eRNAs has been shown to be facilitating looping between active enhancer elements and protein-coding gene promoters, since many of these enhancers show high eRNA expression (Lin et al., 2012, Sanyal et al., 2012). Also, eRNAs have also been shown to interact with two components of the cohesion complex, SMC3 and RAD21, which has been shown to mediate enhancer-promoter looping (Kagey et al., 2010).

The first examples of eRNA mediated chromosomal looping directly affecting target gene expression were found at nuclear receptor-regulated genes. Specifically, knockdown of eRNAs at estrogen receptor alpha (ERα) bound enhancers resulted in a reduction in looping between the enhancer and target gene promoter as well as decreased expression of the target genes, *NR1P1* and *GREB1* (Li et al., 2013). A similar effect was seen at the MyoD enhancer and promoter. It has been shown that the loss of the eRNA transcribed from the MyoD enhancer resulted in a decrease in RNA Pol II recruitment to the promoter region resulting in a decrease in *MyoD* expression (Mousavi et al., 2013).

A second function of eRNAs has been described specifically in neurons. It has been shown that NELF is able to directly bind eRNAs and that this binding competes NELF off of paused promoters to facilitate the release from pausing (Schaukowitch et al., 2014). The loss of these eRNAs decreases the release of NELF from the paused RNA Pol II complex and results in a decrease in target gene expression (Schaukowtich et al., 2014). However, the loss of eRNAs does not affect the interaction between the enhancer and target promoter, unlike what has been described at MyoD and the ERα target genes (Schaukowitch et al., 2014). In *Drosophila*, it has also been shown that NELF is present at sites of eRNA transcription (Core et al., 2012).

## Transcription Dysregulation in Cancer

Each step of transcription is tightly regulated to ensure proper functioning of the cell. A dysregulation of a single step can lead to development of disease. Because of my interest in studying the role of NELF and pausing in regulating transcription and gene expression, I will focus on the dysregulation of transcription of RNA Pol II at the initiation and elongation phases

in cancer. However, termination as well as transcription by RNA Pol I and RNA Pol III have also been shown to play a role in the development of cancer (Bywater et al., 2013).

## Dysregulation of Initiation

The general transcription factors (GTFs) are the key players in the recruitment and initiation of RNA Pol II. The function of these factors is described earlier in this introduction. Because of their importance in coordinating the earliest step in transcription, they are rarely mutated in cancers or other disease. However, overexpression of TATA-binding protein (TBP) has been identified in colon and colorectal cancers (Johnson et al., 2003, Li et al., 2012).  It has been shown that overexpression of TBP leads to cellular transformation and tumor formation in nude mice (Johnson et al., 2003). It has also been shown that colon tumors show higher expression of TBP than normal colon tissue (Johnson et al., 2002). Because TBP is such a potent regulator of transcription, it is not surprising that it can drastically affect the transcriptional profile enough to lead to transformation and that tumors may harness this mechanism to bolster their growth. However, not all promoters contain a TATA box, so overexpression of TBP does not affect the transcription of every gene in the genome. This may be why it is the only GTF found to be mutated in cancers.

Because initiation of RNA Pol II is one of the most critical steps in ensuring proper cell functioning, it is not surprising that alterations in the GTFs are not found in tumors. However, sequence-specific transcription factors, which also can facilitate the recruitment and initiation of RNA Pol II are some of the most well described contributors to tumorigenesis. These factors regulate only the expression of specific target genes, thus making them a better target for cancer promoting mutations than the GTFs. Here, I will focus on the sequence-specific transcription

factor, *MYC*, as it is known to play a role in not only initiation, but also pausing and pause release (Rahl et al., 2010).

Increased expression of MYC has been associated with many cancer types where it functions to enhance cell growth and proliferation (Kalkat et al., 2017). This increased expression can occur in a variety of ways including gene amplification, mutation, chromosomal translocations and protein stabilization (Meyer and Penn, 2008). MYC is thought to mediate the cell cycle by regulating the transcription of MYC target genes. MYC and its partner, MAX, bind to the promoters of genes involved in cell cycle, ribosome biogenesis and metabolism (Meyer and Penn, 2008). Through this binding MYC is able to both upregulate the expression of both protein coding and non-coding RNAs to promote cell growth. Studies have also shown that MYC binding can repress gene expression by aiding in the recruitment of co-repressors as well (Meyer and Penn, 2008). It is thought that MYC is able to bind to 10-15% of the genome, thus the effects of MYC overexpression are vast (Meyer and Penn, 2008). MYC serves as the ideal target for transformation. It maintains a sequence specificity that limits the genes that are affected by overexpression while still targeting enough of the genome to create large scale changes in the transcriptional profile of the cells.

The mediator complex facilitates the initiation of RNA Pol II along with GTFs and sequence specific transcription factors. This complex was originally identified as a co-activator of nuclear-hormone receptors (Chen and Roeder, 2011). Components of the mediator complex such as CDK8, MED19 and MED1 have all been shown to be overexpressed in various tumor types (Bywater et al., 2013). MED19 has been shown to facilitate the proliferation of lung, breast, liver and bladder tumor cells (Li et al., 2011, Sun et al. 2011, Zou et al., 2011, Zhang et al., 2012). This phenotype has been shown by knockdown of MED19 only in cancer cell lines, so

it still remains unclear how an overexpression of this mediator subunit functions in a true tumor sample.

The mechanism by which an overexpression of a mediator subunit influences cancer development has been best described for MED1. MED1 overexpression leads to increased expression of both androgen and estrogen receptor target genes, thus overexpression of this component is most commonly found in breast and prostate cancers (Vijayvargia et al., 2007, Lamy et al., 2011). MED1 overexpression has been shown to aid in the development of resistance to estrogen and androgen therapies as well as induce the expression of genes that promote metastasis (Nagalingam et al., 2012, Cui et al., 2012). These studies have been completed in both cell lines and tumor tissue, thus giving a better idea for how mediator is functioning in connection with hormone driven cancer types. However, little remains known about how mediator is functioning in tumors outside of the breast and prostate.

The importance of the regulation of initiation has been well documented. With MYCs widespread influence on cellular transformation and tumorigenesis, most studies on transcriptional dysregulation in cancer have focused on sequence specific transcription factors and their ability to regulate the recruitment of RNA Pol II. These factors also serve as ideal targets for therapy since they regulate only a subset of genes. However, even MYC has been shown to have effects on processes outside of intimation. Through MYCs ability to facilitate pause release and promote elongation, the overexpression of MYC seen in so many tumors may be due to its ability to regulate multiple steps in the transcription process. Even outside of MYC, regulation of transcription following recruitment and initiation of RNA Pol II has also been shown to influence cancer development and progression and may provide novel targets for therapeutic intervention.

# Dysregulation of Elongation

Elongation of RNA Pol II is facilitated by a number of factors including chromosome remodelers, kinases and RNA processing factors. These factors have been shown to be altered in various tumor types. Some of the most well-described examples are the fusion partners of mixed linage leukemia (MLL). The most common partners of MLL are members of the super elongation complex (SEC) including ELL, AFF1, AFF4, AF9 and ENL (Bywater et al., 2013). These proteins as well as others in the complex are responsible for aiding in the release of RNA Pol II from pausing and the recruitment of histone-modifying factors and the recruitment of 3' end processing factors (Zhou et al., 2012). These fusion proteins typically exert their effects almost exclusively on the HOX cluster of genes, which leads to a dysregulation of hematopoiesis and ultimately development of leukemia (Daser and Rabbits, 2005). These fusions provided some of the first examples of the importance of the regulation of elongation in cancer development. Though the elongation factors alone have not been shown to be large contributors to tumorigenesis, these findings provided a basis for studying other regulatory steps in the transcription process.

Outside of the SEC, elongation factors such as TFIIS as well as components of the mediator complex have been shown to be overexpressed in tumors. Specifically, the overexpression of transcription elongation factor A1 (TCEA1), a form of TFIIS in humans, is found in a variety of human tumors (Shema et al., 2011). Also, overexpression of TCEA2 has been found specifically in cervical cancers (Scotto et al., 2008). The exact mechanism by which these factors facilitate tumor promoting phenotypes remain unknown; however, it is clear that dysregulated elongation even in the context of well-regulated initiation can lead to tumor

formation. This provides a novel way to examine transcription as a target for therapeutic intervention.

## Transcription as a Therapeutic Target

Transcription has been a target of therapeutic intervention even before the advent of targeted and personalized therapies. Common chemotherapeutics such as cisplatin, doxorubicin and etoposide all affect transcription (Bywater et al., 2013). In fact, most of these therapies directly affect the elongation of RNA Pol II by creating insurmountable obstacles for the polymerase.  This results in severe side-effects and high toxicity. In order to address these challenges, targeted and individualized therapeutics have come to the forefront of cancer research. Though the dysregulation of transcription leads to gene expression changes that ultimate lead to cancer, there has been little success in targeting the transcription machinery or even the general transcription factors. There has been some success in targeting the nuclear-hormone receptors; however, resistance to these therapies develops quickly. Novel findings into how transcription is regulated outside of the sequence specific transcription factors, provides new targets for therapeutic intervention.

With the challenges facing the development of targeted therapies, it is appealing to consider methods by which current therapies can be repurposed. Because elongation is the target of widely-used cancer therapies, it is necessary to gain a better understanding of how elongation works in both cancer and normal cells to create a more targeted approach when using broad cancer therapeutics.

## NELF in Cancer

NELF is comprised of 4 individual subunits and the expression of each subunit at the protein level is dependent on all others. Because of this, most studies of NELF in both normal tissue and tumors has focused on the consequences of a loss of NELF. However, the most common alteration of NELF in tumors is an amplification of at least one subunit (Cerami et al., 2012, Gao et al., 2013). No one subunit is amplified or mutated more than another. Amplification of a NELF subunit is found in 19 different tumor types including, breast, lung, colorectal and prostate. Interestingly, amplification of NELFA is found only in bladder cancer. However, the amplification of the other subunits is not correlated with a single tumor type. Other alterations in NELF subunits have been found in all 19 tumor types, including deletions, missense mutations, truncations and alterations in splicing. These alterations do not seem to be mutually exclusive, as one patient can have a combination of any two alterations in any two subunits within one tumor. All of these mutations occur with numerous other mutations in each tumor type. Because it is unclear how NELF is functioning in a cancer context, it is possible that these mutations may be passenger mutations that do not affect the cancer phenotype. However, with so many tumors showing an amplification of NELF, it is possible that amplification of NELF does contribute to cancer development. Because the understanding of NELFs function is limited, it remains unclear how NELF amplification can contribute to cancer phenotypes.

Though mutations in the NELF subunits have been found in many tumor types, very few studies have examined the effects of these alterations in patients. However, NELFB was originally identified through its interaction with BRCA1 (Ye et al., 2001). This sparked interest into the role of NELFB in breast cancer development. It has been shown that breast cancer progression is inversely correlated with NELFB expression (Sun et al., 2008). Overall patients with lower NELFB expression show a higher incidence of distant metastasis and local recurrence

(Sun et al, 2008). However, no deletion of NELFB is detected in any of the breast tumor samples sequenced and stored in TCGA (Cerami et al., 2012, Gao et al., 2013). Only NELFA is found to be deleted in breast tumors (Cerami et al., 2012, Gao et al., 2013).

Because expression of all NELF subunits are reliant on each other, it is thought that low expression of the entire NELF complex due to the decrease in NELFB expression may lead to poor prognosis in breast cancer (Sun et al., 2008). Follow-up studies on the findings from Sun et al., showed that a loss of pausing results in an increase in cell cycle progression and dysregulation of the estrogen response (Aiyar et al., 2004, Aiyar et al., 2006, Kinnis et al., 2009) However, these follow-up studies were performed in cell culture and mouse models rather than patient samples. Also, these phenotypes were identified following an almost complete loss of NELF. In patients, the subunits are still expressed, so it is possible that the phenotypes associated with a complete loss of NELF are not the ones responsible for the poor prognosis seen in patients. The lack of tumor types showing a deletion of any NELF subunits also indicates that loss of the NELF complex is not the main contributor to the poor prognosis seen in patients.

Though amplification of NELF subunits is the most common alteration in NELF found in tumors, only one example of amplification of a NELF subunit has been shown to be associated with cancer phenotypes. NELFE has been shown to be overexpressed at both the transcript and protein level in hepatocellular carcinoma (HCC) (Iida et al., 2012). Patients with the highest NELFE expression show portal vein invasion (PVI) which is associated with poor prognosis and metastasis (Iida et al., 2012). It has also been shown that knockdown of NELFE results in a decrease in invasion and proliferation of HCC cells (Iida et al., 2012). This does not agree with the findings in breast cancer cells. It also provides insight into which processes may be most affected by NELF overexpression. NELFE protein expression was also identified as an

20

independent risk factor for early recurrence (Iidea, et al., 2012). Outside of this single study in hepatocellular carcinoma patient samples, there are no studies examining the phenotypes associated with NELF overexpression in either normal or tumor models.

Amplification of NELFE has also been found in other liver tumors (Cerami et al., 2012, Gao et al., 2013). NELFE was originally characterized as the RD protein (RDBP) and was shown to have RNA binding activity (Levi-Strauss et al., 1988, Cheng et al., 1993). The RNA binding activity of NELFE is conferred through the RRM domain (Levi-Strauss et al., 1988, Cheng et al., 1993). As the only well characterized functional domain in a NELF subunit, it is possible that cancer associated mutations would be clustered in this area. The mutations in the RRM could affect the function of NELFE, since they may inhibit the ability of NELFE to interact with nascent RNA. This could lead to alterations in the only described function of NELF, maintaining promoter-proximal pausing. In examining NELFE mutations in the 19 tumor types shown to have mutations in a NELF subunit, only two mutations have been shown to occur in the RNA binding domain of NELFE. Both of these mutations are found in lung adenocarcinomas (Cerami et al., 2012, Gao et al., 2013). This implies that binding of nascent RNA by NELFE is not the main mechanism by which alterations in NELF subunits aid in cancer development.

Interestingly, a potential germline mutation (allele frequency = 1.0) in NELFA was identified in a single case of hepatocellular carcinoma. This mutation results in a change from a lysine to a phenylalanine at amino acid 366. There are no known functional domains in this region, so it is not clear how this mutation could be contributing to the development of the tumor. This patient also shows over 100 other mutations, so it is possible that this mutation does not have any impact on the cancer phenotype. Also, there have been no associations with NELFA and the development or progression of liver tumors (Cerami et al., 2012, Gao et al.,

2013). There are also no other potential germline mutations associated with any other tumor types.

The large number of tumors that show alterations in NELF as well as the increasing interest in understanding the role of elongation in cancer development provides a unique opportunity to investigate the role of both of these components in a cancer context. The cancer associated mutations in all subunits are depicted in Figure 1.4.

## My Thesis Summary

My work sought to expand the view of NELF and its role in facilitating human gene expression. Here, I have generated a genome-wide profile of promoter-proximal pausing in primary human dermal fibroblasts. With this profile, I have shown the importance of NELF in regulating pausing at these genes in a normal cell type. I also found that NELF is able to regulate gene expression through both pausing as well as maintenance of an open chromatin state at promoters.  Finally, I have shown that NELF regulates the transcription of eRNA. By maintaining eRNA abundance, NELF is able to influence the expression of nearby genes. These findings demonstrate novel roles for NELF in mediating transcription and global gene expression.

My work in HT29 cancer cells has shown that pausing functions as a checkpoint for key regulatory pathways. This checkpoint may serve to aid the cancer cell in surviving increased proliferation and stress by keeping genes necessary for regulating cellular senescence and stress response expressed at a stable level regardless of other genomic alterations. These findings

implicate pausing as mechanism that could be harnessed to slow cancer progression and

development.

# References

Adelman, K., Kennedy, M.A., Nechaev, S., Gilchrist., D.A., Muse, G.W., Chinenov, Y., Rogatsky, I. Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. PNAS. 2009 October 27; 106(43):18207-18212.

Adelman, K., Lis, J.T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet. 2012 Oct;13(10):720-31.

Aida, M., Chen, Y., Nakajima, K., Yamaguchi, Y., Wada, T., Handa, H. Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the *junB* gene. Mol Cell Biol. 2006 Aug;26⊗16):6094-104.

Aiyar, S.E., Sun, J., Blair, A.L., Moskaluk, C.A., Lu, Y., Ye, Q., Yamaguchi, Y., Mukherjee, A., Ren, D., Handa, H., Li., R. Attenuation of estrogen receptor a-mediated transcription through estrogen-stimulated recruitment of a negative elongation factor. Genes Dev. 2004 Sep 1; 18(17):2134-46.

Aiyar, S.E., Blair, A.L., Hopkinson, D.A., Bekiranov, S., Li, R. Regulation of clustered gene expression by cofactor of BRCA1 (COBRA1) in breast cancer cells. Oncogene. 2007 Apr 19;26(18):2543-53.

Amleh, A., Nair, S.J., Sun, J., Sutherland, A., Hasty, P., Li, R. Mouse cofactor of BRCA1 (Cobra1) is required for early embryogenesis. PLoS One. 2009 Apr 2;4(4):e5034.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, B., Otzen Bagger, F., Jorgensen, M., Refsing Anderson, P., Bertin, N., Rackham, O., Burroughs, A.M., Baillie, J.K., Ishizu, Y., Shimizu, Y., Furuhata, E., Maeda, S., Negishi, Y., Mungall, C.J., Meehan, T.F., Lassmann, T., Masayoshi, I., Kawaji, H., Kondo, N., Kawai, J., Lennartsson, A., Daub, C.O., Heutink, P., Hume, D.A., Heick Jensen, T., Suzuki, H., Hayashizaki, Y., Muller, F., The FANTOM Consortium, Forrest, A.R.R., Carninci, P., Rehli, M., Sandelin, A. An atlas of active enhancers across human cell types and tissues. Nature. 2014 Mar 27;507(7493):455-61.

Ashe, H.L., Monks, J., Wijerde, M., Fraser, P., Proudfoot, N.J. Intergenic transcription and transinduction of the human beta-globin locus. Genes Dev. 1997 Oct 1;11(19):2494-509.

Awwad, S.W., Abu-Zhayia, E.R., Guttmann-Raviv, N., Ayoub, N. NELF-E is recruited to DNA double-strand break sites to promote transcriptional repression and repair. EMBO Rep. 2017 Mar 23;pii:e201643181

Bender, T.P., Thompson, C.B., Kuehl, W.M. Differential expression of c-myb in murine B lymphomas by a block to transcription elongation. Science. 1987 Sep 18;237(4821):1473-6.

Bunch, H., Zheng, X., Burkholder, A., Dillon, S.T., Motola, S., Birrane, G., Ebmier, C., Levine, S., Fargo, D., Hu, G., Taatjes, D.J., Calderwood, S.K. TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. Nature. 2014 Aug 31;21(10):876-83.

Bywater, M.J., Pearson, R.B., McArthur, G.A., Hannan, R.D. Dysregulation of the basal RNA polymerase transcription apparatus in cancer. Nat Rev Cancer. 2013 May;13(5):299-314.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Huer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., Schultz, N. the cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012 May;2(5):401-4.

Chen, F.X., Woodfin, A.R., Gardini, A., Rickels, R.A., Marshall, S.A., Smith E.R., Sheikhattar, R., Shilatifard, A. PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. Cell. 2015 Aug 27;162(5):1003-15.

Cheng, J., Macon, K.J., Volanakis, J.E. cDNA cloning and characterization of the protein encoded by RD, a gene located in the class III region of the human major histocompatibility complex. Biochem J. 1993 Sep 1;294(Pt 2):589-93.

Chen, W., Roeder, R.G. Mediator-dependent nuclear receptor function. Semin Cell Dev Biol 2011 Sep;22(7):749-58.

Core, L.J., Waterfall, J.J., Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008 Dec 19;322(5909):1854-8.

Core, L.J., Waterfall. J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., Lis, J.T. Defining the status of RNA polymerase at promoters. Cell Rep. 2012 Oct 25;2(4):1025-35.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., Jaenisch, R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. PNAS. 2010 Dec 14;107(50):21931-6.

Cui, J., Germer, K., Wu, T., Wang, J., Luo, J., Wang, S.C., Wang, Q., Zhang, X. Cross-talk between HER2 and MED1 regulates tamoxifen resistance of human breast cancer cells. Cancer Res 2021 Nov 1;72(21):5625-34.

Dahmus, M.E. Phosphorylation of eukaryotic DNA-dependent RNA polymerase. Identification of calf thymus RNA polymerase subunits phosphorylated by two purified protein kinases, correlation with in vivo sites of phosphorylation in HeLA cell RNA polymerase II. J Biol Chem. 1981 Apr 10;256(7):3332-9.

Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L.J., Lis, J.T., Siepel, A., Kraus, W.L. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Mol Cell. 2013 Apr 25;50(2):212-22.

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Reed, R. SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. Mol Cell. 2007 Jun 22;26(6):867-81.

Daser, A., Rabbitts, T.H. The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. Semin Cancer Biol. 2005 Jun;15(3):175-88.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57-74.

Firestein, R., Bass, A.J., Kim, S.Y., Dunn, I.F., Silver, S.J., Guney, I., Freed, E., Ligon, A.H., Vena, N., Ogino, S., Chheda, M.G., Tamayo, P., Finn, S., Shrestha, Y., Boehm, J.S., Jain, S., Bojarski, E., Mermel, C., Barretina, J., Chan. J.A., Baselga, J., Tabernero, J., Root, D.E., Fuchs, C.S., Loda, M., Shivdasani, R.A., Meyerson. M., Hahn, W.C. CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. Nature. 2008 Sep 25;455(7212):547-51.

Fivaz, J., Bassi, M.C., Pinaud, S., Mirkovitch, J. RNA polymerase II promoter-proximal pausing upregulates c-fos gene expression. Gene. 2000 Sep 19;255(2):185-94.

Flores, O. Maldonado, E., Reinberg, D. Factors involved in specific transcription by mammalian RNA polymerase II: Factors IIE and IIF independently interact with RNA polymerase II. J Biol Chem. 1989 May 25;264(15)8913-21.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., Peterlin, B.M. Dynamics of Human Immunodeficiency Virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Mol Cell Bio. 2004 Jan;24(2):787-95.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., Schultz, N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013 Apr 2;6(269):pl1.

Gariglio, P., Bellard, M., Chambon, P. Clustering of RNA polymerase B molecules in the 5' moiety of the adult beta-globin gene of hen erythrocytes. Nucleic Acids Res. 1981 Jun 11;9(11):2589-98.

Gibson, B.A., Zhang, Y., Jiang, H., Hussey, K.M., Shrimp, J.H., Lin, H., Schwede, F., Yu, Y., Kraus, W.L. Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. Science 2016 Jul 1;353(6294):45-50.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K.B., Collins, J.B., Li, L., Gilmour, D.S., Adelman, K. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev. 2008 Jul 15;22(14)1921-33.

Gilmour, D.S., Lis, J.T. RNA polymerase II interacts with the promoter region of the noninduced *hsp70* gene in *Drosophila melanogaster* cells. Mol Cell Bio. 1986 Nov;6(11):3984-9.

Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., Kraus, W.L. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell. 2011 May 13;145(4):622-34.

Heinz, S., Romanoski, C.E., Benner, C., Glass, C.K. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol. 2015 Mar;16(3)144-54.

Ho, C.K., Schuman, S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. 1999 Mar;3(3):405-11.

Iida, M., Iizuka, N., Tsunedomi, R., Tsutsui, M., Yoshida, S., Maeda, Y., Tokuhisa, Y., Sakamoto, K., Yoshimura, K., Tamesa, T., Oka, M. Overexpression of the RD RNA binding protein in hepatitis C virus-related hepatocellular carcinoma. Oncol Rep. 2012 Aug;28(2):728-34.

Johnson, S.A., Dubeau, L., Kawalek, M., Dervan, A., Schonthal, A.H., Dang, C.V., Johnson, D.L. Increased expression of TATA-binding protein, the central transcription factor, can contribute to oncogenesis. Mol Cell Biol. 2003 May;23(9):3043-51.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier C.C., Goossens, J., Rahl, P.B., Levine S.S., Taatjes, D.J., Dekker, J., Young, R.A. Mediator and cohesion connect gene expression and chromatin architecture. Nature. 2010 Sep 23;467(7314):430-5.

Kalkat, M., De Melo, J., Hickman, K.A., Lourenco, C., Redel, C., Resetca, D., Tamachi, A., Tu, W.B., Penn, L.Z. MYC deregulation in primary human cancers. Genes (Basel). 2017 May 25;8(6). pii: E151.

Kapoor-Vazirani, P., Kagey, J.D., Vertino. P.M. Kapoor-Vazirani, SUV420H2-mediated H4K20 trimethylation enforces RNA polymerase II promoter-proximal pausing by blocking hMOF-dependent H4K16 acetylation. Mol Cell Biol. 2011 Apr;31(8):1594-609.

Kapoor-Vazirani, P., Vertino, P.M. A dual role for the histone methyltransferase PR-SET7/SETD8 and histone H4 lysine 20 monomethylation in the local regulation of RNA polymerase II pausing. J Biol Chem. 2014 Mar 1;289(11):7425-37,

Kephart, D.D. Marshall, N.F., Price, D.H. Stability of Drosophila RNA polymerase II elongation complexes in vitro. Mol Cell Biol. 1992 May;12(5):2067-77.

Kim, J.B., Sharp., P.A. Positive transcription elongation factor b phosphorylates hSPT5 and RNA polymerase II carboxy-terminal domain independently of cyclin-dependent kinase-activating kinase. J Biol Chem. 2001 Apr 13;276(15):12317-23.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., Ren, B. A high-resolution map of active promoters in the human genome. Nature. Aug 2005;436(11):876-80.

Kim, J. Guermah, M., Roeder, R.G. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. Cell. 2010 Feb 19;140(4):491-503.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Lapetwicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P.F., Kreiman, G., Greenberg, M.E. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010 May 13;465(7295):182-7.

Kinnis, M., Isaacs, G.D., Core, L.J., Hah, N., Kraus, W.L. Postrecruitment regulation of RNA polymerase II directs rapid signaling responses at the promoters of estrogen target genes. Mol Cell Biol. 2009 Mar;29(5):1123-33.

Koch, F., Fenouli, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicugila, S., de la Chapelle, A.L., Heidemann, M., Mintermair, C., Eick, D., Gut, I., Ferrier, P., Andrau, J.C. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat Struct Mol Biol. 2011 Jul 17;18(8):956-63.

Krumm, A., Meulia, T., Brunvand, M., Groudine, M. The block to transcriptional elongation within the human c-*myc* gene is determined in the promoter-proximal region. Genes Dev. 1992 Nov;6(11):2201-13.

Lam, M.T.Y., Li, W. Rosenfeld, M.G., Glass, C.K. Enhancer RNAs and regulated transcriptional programs. Trends Biochem Sci. 2014 Apr;39(4):170-82.

Lamy, P.J., Fina, F., Bascoul-Mollevi, C., Laberenne, A.C., Martin, P.M., Ouafik, L., Jacot, W. Quantification and clinical relevance of gene amplification at chromosome 17q12-q21 in human epidermal growth factor receptor 2-amplified breast cancers. Breast Cancer Res. 2011 Feb 2;13(1):R15.

Larschan, E., Bishop, E.P., Kharchenko, P.V., Core, L.J., Lis, J.T., Park, P.J., Kuroda, M.I. X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. Nature. 2011 Mar 3;471(7336):115-8.

Law, A., Hirayoshi, K, O'Brien, T., Lis, J.T. Direct cloning of DNA that interacts in vivo with a specific protein: application to RNA polymerase II and sites of pausing in Drosophila. Nucleic Acids Res. 1998 Feb 15;26(4):919-24.

Lee, H., Kraus, K.W., Wolfner, M.F., Lis, J.T. DNA sequence requirements for generating paused polymerase at the start of hsp70. Genes Dev. 1992 Feb;6(2):284-95.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., Gilmour, D.S. NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. Mol Cell Biol. 2008 May;28(10):3290-300.

Levi-Strauss, M., Carroll, M.C., Steinmetz, M., Meo, T. A previously undetected MHC gene with an unusual periodic structure. Science. 1988 Apr 8;240(4849):201-4.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., Oh, S., Kim, H.S., Glass, C.K., Rosenfeld, M.G. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature. 2013 Jun 27;498(7455):516-20.

Li, L.H., He, J., Hua, D., Guo, Z.J., Gao, Q. Lentivirus-mediated inhibition of Med19 suppresses growth of breast cancer cells in vitro. Cancer Chemother Pharmacol. 2011 Jul;68(1):207-15.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Chandra, V., Bossen, C., Glass, C.K., Murre, C. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat Immunol. 2012 Dec;13(12):1196-204.

Lu, X., Zhu, X., Li, Y., Liu, M., Yu, B., Wang, Y., Rao., M., Yang, H., Zhou, K., Wang, Y., Chen, Y., Chen, M., Zhuang, S., Chen, L., Lui, R., Chen, R. Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. Nucleic Acids Res. 2016 Aug 19;44(14):6853-57.

Luo, Z., Lin, C., Shilatifard, A. The super elongation complex (SEC) family in transcriptional control. Nat Rev Mol Cell Biol. 2012 Sep;13(9):543-7.

Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J., Reinberg D. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. Proc Natl Acad Sci U S A. 2004 May 18;101(20):7572-7.

Maniatis, T., Goodbourn, S., Fischer, J.A. Regulation of inducible and tissue-specific gene expression. Science. 1987 Jun 5;236(4806):1237-45.

Marshall, N.F., Price, D.H. Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol Cell Biol. 1992 May;12(5):2078-90.

Marshall, N.F., Price, D.H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J Biol Chem. 1995 May 26;270(21):12335-8.

Marshall, N.F., Peng, J., Xie, Z., Price, D.H. Control of RNA polymerase II elongation potential by a novel carboxy-terminal domain kinase. J Biol Chem. 1996 Oct 25;271(43):27176-83.

Martinicic, K., Alka, S.A., Cheatle, A., Borghesi, L., Milcarek, C. Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. Nat Immunol. 2009 Oct;10(10):1102-9.

Meyer, N., Penn, L.Z. Reflecting on 25 years with MYC. Nat Rev Cancer. 2008 Dec;8(12):976-90.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., Lis, J.T. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev. 2011 Apr 1;25(7):742-54.

Moteki, S., Price, D. Functional coupling of capping and transcription of mRNA. Mol Cell. 2002 Sep;10(3):599-609.

Mousavi, K., Zare, H., Dell'Orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., Sartorelli, V. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell. 2013 Sep 12;51(5):606-17.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grisson, S.F., Zeitlinger, J., Adelman. K. RNA polymerase is poised for activation across the genome. Nat Genet. 2007 Dec;39(12):1507-11.

Nagalingam, A., Tighiouart, M., Ryden, L., Joseph, L., Landberg, G., Saxena, N.K., Sharma, D. Med1 plays a critical role in the development of tamoxifen resistance. Carcinogenesis. 2012 Apr;33(4):918-30.

Nagake, T., Logan, C., Hotta, I., Rozenblatt-Rosen, O., Meyerson, M., Manley, J.L. Transcriptional activators enhance polyadenylation of mRNA precursors. Mol Cell. 2011 Feb 18;41(4):409-18.

Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D., Hasegawa, J., Omori, M., Inukai, N., Endoh, M., Yamada, T., Hiroshi, H. Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. Mol Cell Biol. 2003 Mar;23(6):1863-73.

Narita, T., Yung, T.M., Yamamoto, J., Tsuboi, Y., Tanabe, H., Tanaka, K., Yamaguchi, Y., Handa, H. NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. Mol Cell. 2007 May 11;26(3):349-65.

Nechaev, S. Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., Adelman, K. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science. 2010 Jan 15;327(5693):335-8.

O'Brien T., Lis, J.T. RNA polymerase II pauses at the 5' end of the transcriptionally induced *Drosophila hsp70* gene. Mol Cell Biol. 1991 Oct;11(10):5285-90.

Patel, M.C., Debrosse, M., Smith, M., Dey, A., Huyn, W., Sarai, N., Heightman, T.D., Tamura, T., Ozato, K. BRD4 coordinates recruitment of pause release factor P-TEFb and the pausing complex NELF/DSIF to regulate transcription elongation of interferon-stimulated genes. Mol Cell Biol. 2013 Jun;33(12):2497-507.

Plet, A., Eick, D., Blanchard, J.M. Elongation and premature termination of transcripts initiated from c-Fos and c-Myc promoters show dissimilar patterns. Oncogene. 1995 Jan 19;10(2):319-28.

Price, D.H., Sluder, A.E., Greenleaf, A.L. Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. Mol Cell Biol. 1989 Apr;9(4):1465-75.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., Young, R.A. c-Myc regulates transcriptional pause release. Cell. 2010 Apr 30;141(3):432-45.

Roeder, R.G., Rutter, W.J. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature. 1969 Oct 18;224(5216):234-7.

Roeder, R.G. The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci. 1996 Sep;21(9):327-35.

Rougvie, A.E., Lis, J.T. The RNA polymerase II molecule at the 5' end of the uninduced *hsp70* gene of D. melanogaster is transcriptionally engaged. Cell. 9 Sep 1988;54(6):795-804.

Sanyal, A., Lajoie, B.R., Jain, G., Dekker, J. The long-range interaction landscape of gene promoters. Nature. 2012 Sep 6;489(7414):109-13.

Schaukowitch, K., Joo, J., Liu, X., Watts, J.K., Martinez, C., Kim, T. Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell. 2014 Oct 2;56(1):29-42.

Scotto, L., Narayan, G., Nandula, S.V., Arias-Pulido, H., Subramaniyam, S., Schneider, A., Kaufmann, A.M., Wright, J.D., Pothuri, B., Mansukhani, M., Murty, V.V. Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. Genes Chromosomes Cancer. 2008 Sep;47(9):755-65.

Shell, S.A., Martincic, K., Tran, J., Milcarek, C. Increased phosphorylation of the carboxyl-terminal domain of RNA polymerase II and loading of polyadenylation and cotranscriptional

factors contribute to the regulation of the ig heavy chain mRNA in plasma cells. J Immunol. 2007 Dec 1;179(11):7663-73.

Shema, E., Kim, J., Roeder, R.G., Oren, M. RNF20 inhibits TFIIS-facilitated transcriptional elongation to suppress pro-oncogenic gene expression. Mol Cell. 2011 May 20;42(4):477-88.

Sims, R.J.3rd., Belotserkovskaya, R., Reinberg, D. Elongation by RNA polymerase II: the short and long of it. Genes Dev. 2004 Oct 15;18(20):2437-68.

Singh, J., Padgett, R.A. Rates of in situ transcription and splicing in large human genes. Nat Struct Mol Biol. 2009 Nov;16(11):1128-33.

Strobl, L.J., Eick, D. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-Myc *in vivo*. EMBO J. 1992 Sep;11(9):3307-14.

Sun, J., Watkins, G., Blair, A.L., Moskaluk, C., Ghosh, S., Jiang, W.G., Li, R. Deregulation of cofactor of BRCA1 expression in breast cancer cells. J Cell Biochem. 2008 Apr 15;103(6):1798-807.

Sun, J., Li, R. Human negative elongation factor activates transcription and regulates alternative transcription initiation. J Biol Chem. 2010 Feb 26;285(9):6443-52.

Sun. J., Pan, H., Lei, C., Yuan, B., Nair, S.J., April, C., Parmeswaran, B., Klotzle, B., Fan, J., Ruan, J., Li, R. Genetic and genomic analyses of RNA polymerase II-pausing factor in regulation of mammalian transcription and cell growth. J Biol Chem. 2011 Oct 21;286(42):36248-57.

Sun. M., Jiang, R., Li, J.D., Luo, S.L., Gao, H.W., Jin, C.Y., Shi, D.L., Wang, C.G., Wang, B., Zhang, X.Y. MED19 promotes proliferation and tumorigenesis of lung cancer. Mol Cell Biochem. 2011 Sep;355(1-2):27-33.

Vijayvargia, R., May, M.S., Fondell, J.D. A coregulatory role for the mediator complex in prostate cancer cell proliferation and gene expression. Cancer Res. 2007 May 1;67(9):4043-41.

Vos, S.M., Pollman, D., Caizzi, L., Hofmann, K.B., Rombaut, P., Zimniak, T., Herzog, F., Cramer, P. Architecture and RNA binding of the human negative elongation factor. Elife. 2016 Jun 10;5 pii:e14981.

Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., Handa, H., Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription *in vitro*. EMBO J. 1998 Dec 15;17(24):7395-403.

Wang, X., Hang, S., Prazak, Gergen, P.J. NELF potentiates gene transcription in the Drosophila embryo. PLoS One. 2010 July 9;5(7):e11498.

Weinmann, R., Roeder, R.G. Role of DNA-dependent RNA polymerase 3 in transcription of the tRNA and 5S RNA genes. Proc Natl Acad Sci U S A. 1974 May;71(5):1790-4.

Weinmann, R[2]., Raskas, H.J., Roeder, R.G. Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. Proc Natl Acad Sci U S A. 1974 Sep;71(9)3426-39.

Wen, Y, Shatkin, A.J. Transcription elongation factor hSPT5 stimulates mRNA capping. Genes Dev. 1999 Jul 15;13(14):1774-9.

Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., Adelman, K. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. Mol Cell. 2015 Apr 16;58(2):311-22.

Wu, C., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., Gilmour, D.S. NELF and DSIF cause promoter proximal pausing on the *hsp70* promoter in *Drosophila*. Genes Dev. 2003 Jun 1;17(11):1402-14.

Wu, C., Lee, C., Fan, R., Smith, M.J., Yamaguchi, Y., Handa, H., Gilmour, D.S. Molecular characterization of *Drosophila* NELF. Nucleic Acids Res. 2005 Mar 1;33(4):1269-79.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., Handa. H. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol Cell. 2006 Jan 20;21(2):227-37.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., Handa, H. NELF, a mulitisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell. 1999 Apr 2;97(1):41-51.

Yamguchi, Y., Inukai, N., Narita, T., Wada, T., Handa, H. Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol Cell Biol. 2002 May;22(9):2918-27.

Ye, Q., Hu, Y.F., Zhong, H., Nye, A.C., Belmont, A.S., Li, R. BRCA1-induced large-scale chromatin unfolding and allele-specific effects of cancer-predisposing mutations. J Cell Biol. 2001 Dec 10;155(6):911-21.

Yu, M., Yang, W., Ni, T., Tang, Z., Nakadai, T., Zhu, J., Roeder, R.G. RNA Polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science. 2015 Dec 11;350(6266):1383-6.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., Young, R.A. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet. 2007 Dec;39(12):1512-6.

Zhang, H., Jiang, H., Wang, W., Gong, J., Zhang, L., Chen, Z., Ding, Q. Expression of Med19 in bladder cancer tissues and its role on bladder cancer growth. Urol Oncol. 2012 Nov-Dec;30(6):920-7.

Zhou, Q., Li, T., Price, D.H. RNA Polymerase II elongation control. Annu Rev Biochem. 2012;81:119-43.

Zou, S.W., Ai, K.X., Wang, Z.G., Yuan, Z. Yan, Z., Yan, J., Zheng, Q. The role of Med19 in the proliferation and tumorigenesis of human hepatocellular carcinoma cells. Acta Pharmacol Sin. 2011 Mar;32(3):354-60.

**Figure 1.1**
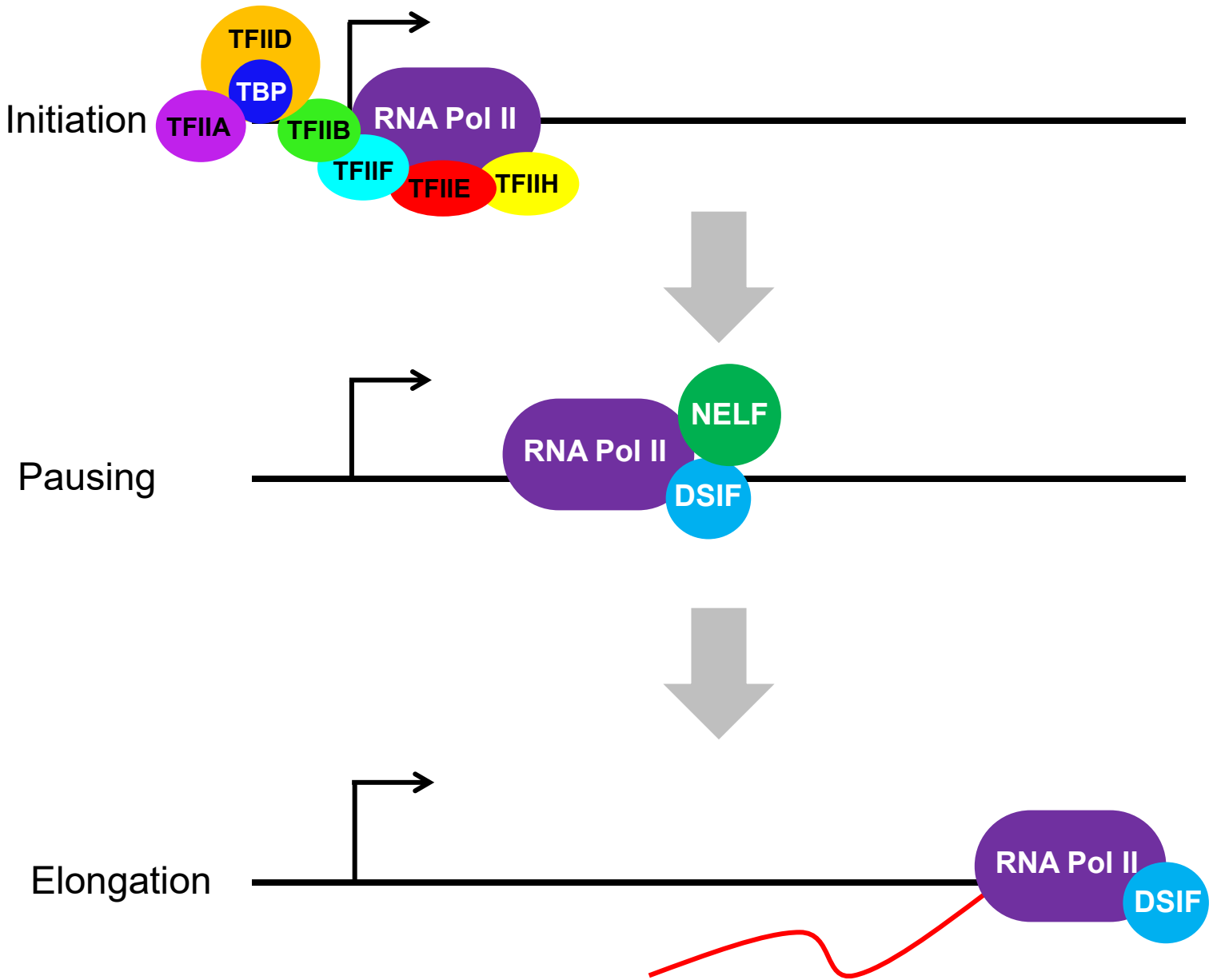
Figure 1.1 General Overview of Transcription. Transcription begins with the recruitment and initiation of RNA Pol II by the general transcription factors. Following initiation RNA Pol II can transcribe a short stretch of RNA before pausing approximately 20-60 nucleotides downstream of the TSS. This is termed promoter-proximal pausing and is mediated by the pausing factors negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF). Upon a signal, NELF is released from the paused complex and RNA Pol II enters into active elongation.
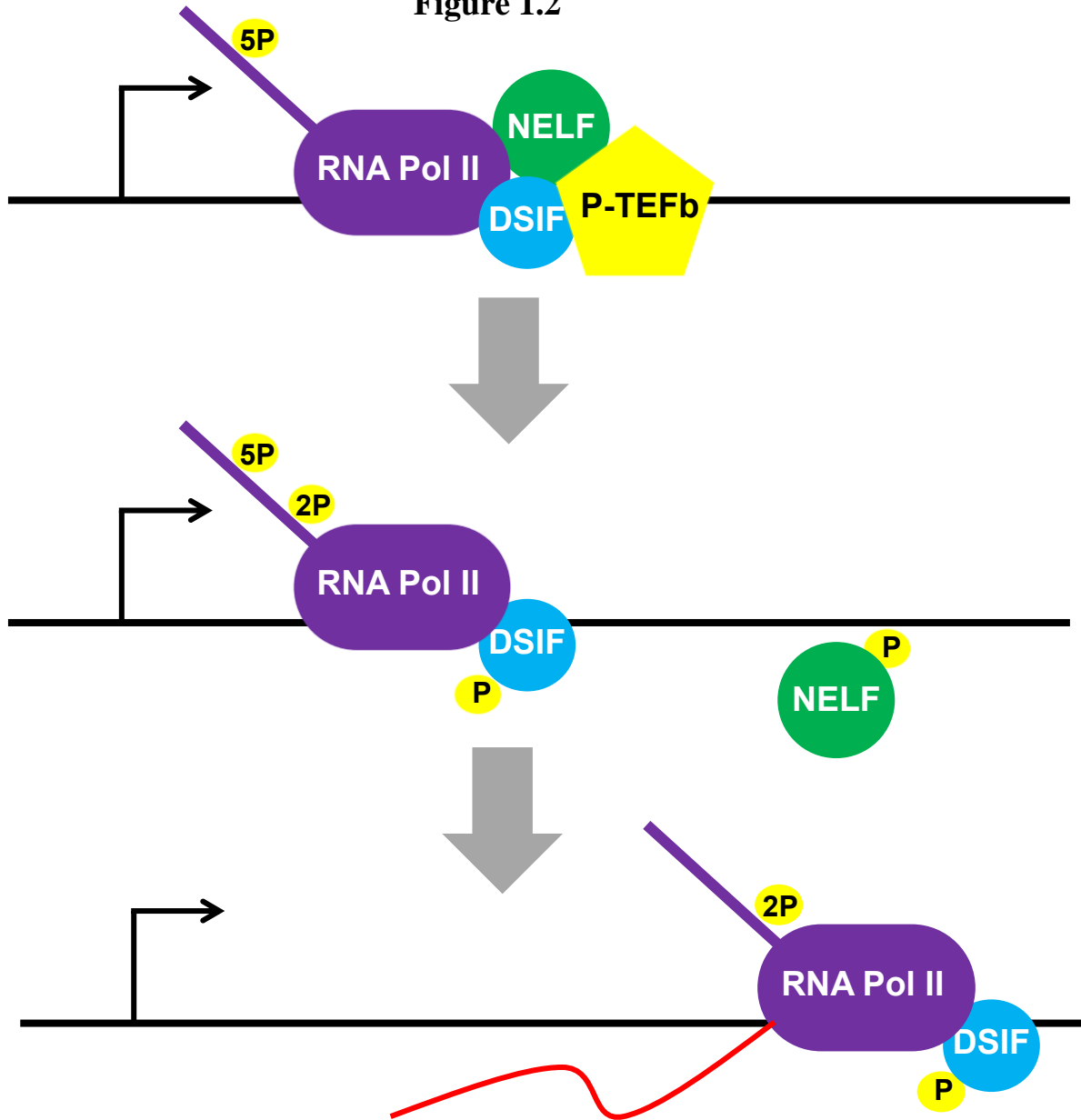
**Figure 1.2**

Figure 1.2 Mechanism of Pause Release. Positive transcription elongation factor B (P-TEFb) is recruited to paused complexes where it phosphorylates the serine 2 residue of the RNA Pol II C-terminal domain (CTD) as well as NELF and DSIF. Following these phosphorylation events, NELF dissociates from the paused complex, DSIF becomes a positive elongation factor and RNA Pol II enters into active elongation.

**Figure 1.3**
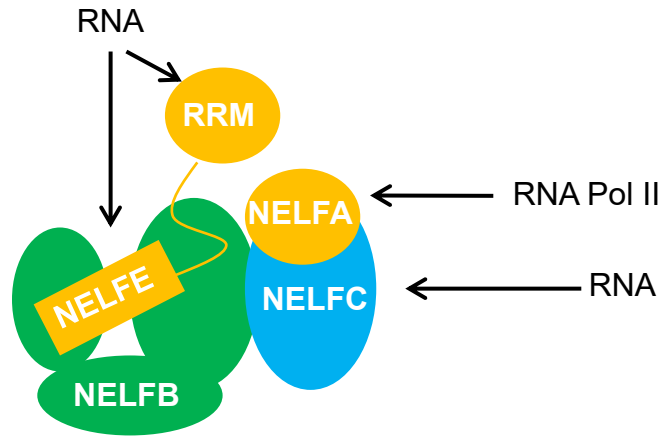


Figure 1.3 Structure of negative elongation factor (NELF). NELF is comprised of 4 individual subunits, NELFA, NELFB, NELFC/D and NELFE. The B, C and E subunits all directly interact with nascent RNA. Specifically, the NELFE subunit contains an RNA binding domain (RRM) through which it is able to bind the nascent RNA. NELFA directly interacts with RNA Pol II. This figure is adapted from Vos et al., 2016.

# Figure 1.4
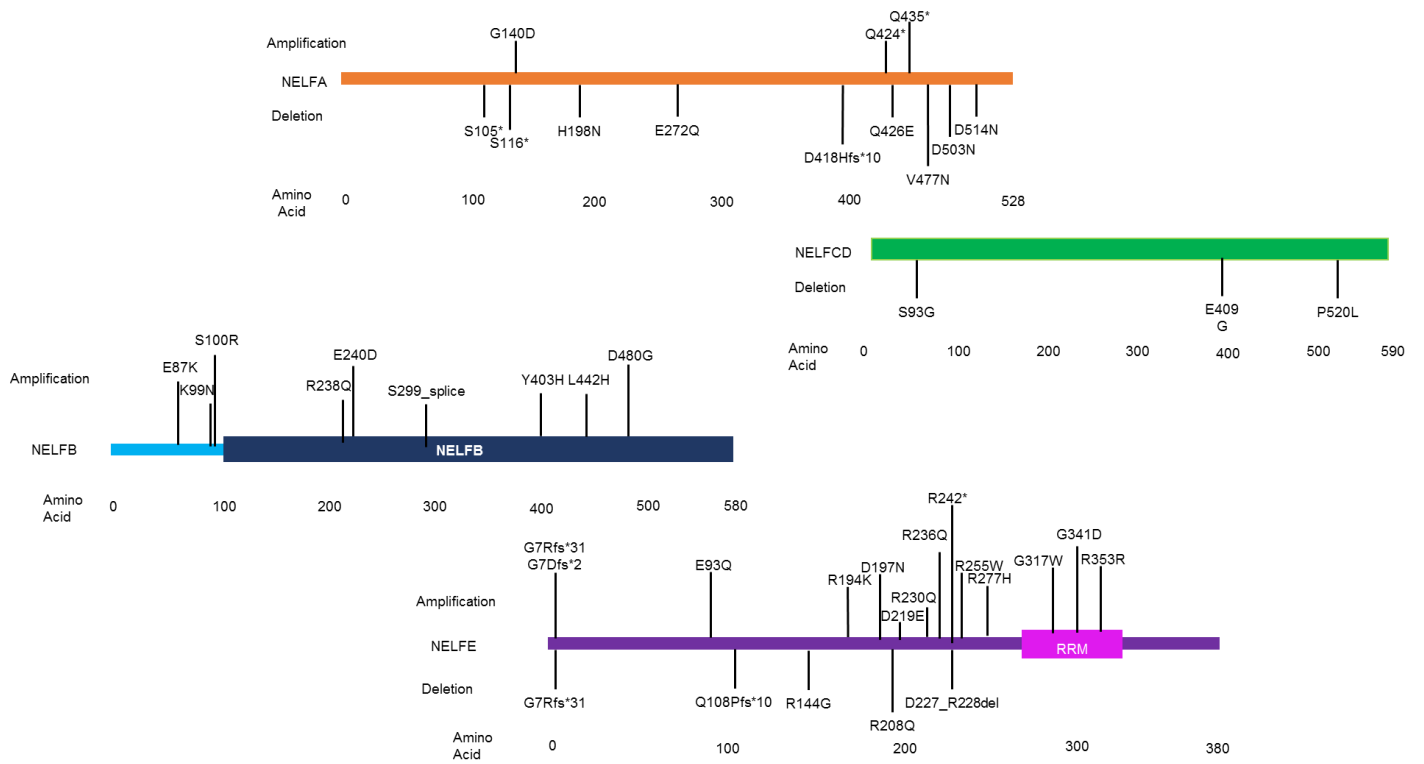


Figure 1.4 Cancer-associated mutations in NELF subunits. Mutations identified to be associated with varying tumor types. Mutations that result in an amplification or deletion are shown for each individual subunit.

# Chapter 2

## The Role of NELF in Mediating Human Gene Expression

Dr. Alan Bruzel provided technical assistance with the PRO-Seq technique used to generate the genome-wide data in Figures 2.1 and 2.4.

Dr. Jason Watts provided technical assistance with the ChIP-Seq technique used to generate the genome-wide data in Figures 2.2 and 2.6.

Zhengwei Zhu wrote the scripts for the pausing index calculations used in Figure 2.1 and Figure 2.4 as well as the scripts for quantitating active transcription from the PRO-Seq data in Figure 2.6.

Zhengwei Zhu performed the bioinformatic analysis for the genome-wide data generated from the PRO-Seq, ChIP-Seq and RNA-Seq experiments in Figures 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6.

Yooree Chung and Zhengwei Zhu wrote the script for the metagene analysis used in Figures 2.1, 2.2 and 2.4.

Dr. Joshua Burdick plotted the graph of the genome-wide data shown in Figure 2.6.

Dr. Xiaorong Wang performed the overlap analysis of the ChIP-Seq and PRO-Seq datasets in Figures 2.1, 2.2 and 2.6.

Jennifer Fox ran the Illumina HiSeq 2500 instrument for the PRO-Seq, ChIP-seq and RNA-Seq experiments in Figures 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6.

## Abstract

RNA polymerase II (RNA Pol II) pausing is a key regulatory step between transcription initiation and elongation. We identified over 7,200 paused genes in primary human dermal fibroblasts. These genes are involved in pathways such as EGFR and TGFβ signaling as well as general gene expression and transcription. The two most well characterized pausing factors are negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF). We were interested in better understanding the role of NELF in mediating pausing as well as human gene expression. As expected, we found that NELF is localized to the promoters of paused genes, and that loss of NELF results in a decrease in pausing. Upon loss of NELF, we see global gene expression changes with enrichment for genes involved in EGFR, TGFβ and VEGF signaling. We also show that NELF is able to maintain a favorable state at promoters to promote re-initiation of RNA Pol II at the genes that show a decrease in expression upon loss of NELF. Though NELF is predominantly found in promoter regions, we identified NELF enrichment in both gene bodies and intergenic regions. Many of these sites also show enrichment for H3K27ac, a mark of active enhancers and are actively transcribed suggesting that there are enhancer RNAs (eRNAs). We found that transcription of these eRNA is dependent on NELF. These eRNA are found in both known coding sequence as well as intergenic regions. Finally, we found that loss of these eRNA led to changes in expression of nearby genes. This work implicates NELF as a key regulator of human gene expression through its ability to encourage re-initiation of RNA Pol II at promoters as well as facilitate eRNA transcription. These data provide insight into the novel roles of NELF in regulating the expression of both mRNAs and eRNAs.

# Introduction

Following initiation, RNA Pol II is able to "pause" just downstream of transcription start sites. This pause, termed promoter-proximal pausing, has been shown to facilitate active elongation by maintaining the polymerase in a poised state (O'Brien and Lis, 1991, Min et al., 2011). The *hsp70* gene in *Drosophila* as well as the human genes *MYC* and *FOS* have been found to be regulated at the release from promoter-proximal pausing (Rougvie and Lis, 1988, Plet et al., 1995, Krumm et al., 1992, Strobl and Eick, 1992). Through the extensive study of these genes, promoter-proximal pausing has been defined by the presence of a transcriptionally engaged polymerase approximately 20-60 nucleotides downstream of the transcription start site as well as an enrichment of RNA Pol II in the promoter region as compared to the gene body (Gilmour and Lis, 1986, Core et al., 2008).

Genome-wide mapping of RNA Pol II has shown that RNA Pol II accumulates at the 5' end of *Drosophila* and human genes (Muse et al., 2007, Zeitlinger et al., 2007, Kim et al., 2005). These polymerases have been shown to be paused in the promoter-proximal region through nuclear-run on assays (Kim et al., 2005, Muse et al., 2007, Zeitlinger et al., 2007).

Pausing is mediated by two factors, negative elongation factor (NELF) and DRB sensitivity- inducing factor (DSIF) (Wada et al., 1998, Yamaguchi et al., 1999, Wu et al., 2003, Wu et al., 2005). DSIF is comprised of two subunits, SPT4 and SPT5 (Wada et al., 1998) and can function as both a positive and negative elongation factor (Wada et al., 1998). NELF is comprised of 4 subunits, NELFA, NELFB, NELF C/D and NELFE (Wu et al., 2005, Yamaguchi et al., 1999). In vitro studies show that NELF contacts RNA Pol II directly through the A subunit and nascent RNAs through the B, C and E subunits (Yamaguchi et al., 1999, Yamaguchi et al., 2002, Vos et al., 2016). NELF is also thought to interact directly with DSIF; however, the

subunit through which this interaction occurs is not unknown (Yamaguchi et al., 1999). It is through these direct interactions that NELF is thought to mediate RNA Pol II pausing (Yamaguchi et al., 1999, Wu et al., 2003, Wu et al., 2005).

Upon a signal, positive transcription elongation factor b (P-TEFb) is recruited to the paused complex (Marshall and Price 1992, Marshall and Price 1995, Wada et al., 1998). Recruitment of P-TEFb is mediated by a number of factors including BRD4 and the super elongation complex (SEC) (Chen et al., 2015, Lu et al., 2016, Lin et al., 2011, Rahl et al., 2010, Bunch et al., 2014). Once recruited, P-TEFb phosphorylates the serine 2 residue of the series of heptad repeats that make up the C-terminal domain of the RBP1 subunit of RNA Pol II (Marshall et al., 1996). It also phosphorylates NELFE and the C-terminal domain of SPT5 (Kim and Sharp, 2001, Fujinaga et al., 2004, Yamada et al., 2006). Upon these phosphorylation events, NELF is released from the paused complex, DSIF becomes a positive elongation factor and RNA Pol II enters into active elongation (Zhou et al., 2012).

Studies of pausing have focused on identifying and defining promoter-proximal pausing, unraveling mechanism and identifying key players. However, there are few comprehensive and in-depth studies of promoter-proximal pausing and its role in cellular function. Here, we utilized Precision Run-On Sequencing (PRO-Seq) as well as chromatin immunoprecipitation (ChIP) for the pausing factors and RNA Pol II to generate genome-wide map of pausing in primary human dermal fibroblasts. Our data show that pausing occurs throughout the human genome. Because DSIF plays a role in both pausing and elongation whereas NELF functions only in pausing (Wada et al., 1998, Yamaguchi et al., 1999). Therefore, we focused on NELF for the remainder of our studies. Loss of NELF has been shown to decrease RNA Pol II occupancy at promoters as well as alter gene expression patterns in *Drosophila*, mice and breast cancer cells (Core et al.,

2012, Gilchrist et al., 2008, Aiyar et al., 2004, Kinnis et al., 2009, Sun and Li, 2010, Sun et al., 2011, Williams et al., 2015). Our study identified three key roles of NELF. As expected, we found that NELF regulates gene expression in primary human dermal fibroblasts by maintaining promoter-proximal pausing. Second, we show that NELF maintains an open chromatin state at gene promoters to facilitate transcription by showing a decrease in the amount of active RNA Pol II at promoter regions at genes that decrease in expression. Third, we demonstrated that NELF affects gene expression by facilitating the transcription of eRNA Collectively, these findings implicate NELF as a key regulator of human gene expression.

## Results

## Promoter-Proximal Pausing Occurs Genome-Wide in Primary Human Dermal Fibroblasts

We began by identifying genes with promoter-proximal pausing in human dermal fibroblasts. These cells are easily harvested and are able to be cultured as primary cells. They also represent the most common cell type in connective tissue. For these reasons, these cells serve as a good model for human studies. To do this, we utilized Precision Run-On Sequencing (PRO-Seq) which maps the location of active RNA polymerases across the genome. PRO-Seq utilizes a modified nuclear run-on assay followed by high-throughput sequencing to map the location of active RNA polymerases at almost single base resolution (Figure S2.1) (Kwak et al., 2013). We chose PRO-Seq to study active polymerases because of its ability to map the location of the polymerase with near single nucleotide resolution through the incorporation of biotin-labeled NTPs. This provides an advantage over GRO-Seq, which maps the polymerase only

through the incorporation of a single labeled nucleotide; therefore, the location of the polymerase could be tens of base pairs away from its true location.

Promoter-proximal pausing results in an accumulation of active RNA Pol II at the 5' end of genes just downstream of the transcription start sites (TSS). We plotted the location of active RNA polymerases with our PRO-Seq data. When examining over 11,000 genes, we find an accumulation of active RNA polymerase at the 5' end (Figure 2.1A). The over 11,000 genes shown in this figure represent genes that are greater than 1kb in length and at least 1kb away from another gene. We performed these filtering steps to remove genes where the entirety of the gene fell into the promoter region and those genes that had overlapping transcripts. Both of these would inflate the peak we find at the 5' end. The script used to generate the metagene plot in Figure 2.1A is a custom script that was written by Yooree Chung and Zhengwei Zhu. This script contains the filtering steps described above. It also scaled genes into three 100 bin regions which represent the promoter, gene body and 3' end. This binning allows us to represent the density of PRO-Seq reads in each region across many genes regardless of the gene length. Our finding that there is an accumulation of RNA polymerase at the 5' end suggests that promoter-proximal pausing is occurring genome-wide.

To measure the extent of promoter-proximal pausing, for each gene, we calculated a pausing index (PI). The PI is a ratio of the number of active polymerases in a defined promoter region (+/- 1kb surrounding the TSS) as compared to the gene body (Figure S2.2) (Core et al., 2008). The number of polymerases in each region is also normalized to the length in base pairs (bp) of each region. This normalization is used to account for the differences in gene body length so that extremely long or short genes do not show an artificially high or low PI. In our analysis, we performed a number of filtering steps prior to identifying paused genes in our primary human

dermal fibroblasts. We began by removing genes that are less than 1kb in length. This eliminates genes where the entire gene falls within the promoter region. We also removed genes that were less than 1kb away from another gene. This step eliminates the possibility that reads located in nearby genes will be included in the promoter region of the gene of interest, thus leading to an artificially high PI for that gene. Next, we eliminated all genes that had less than 10 overall reads. Genes with less than 10 reads tend to give artificially high PI values, which are not representative of the true biology (Figure S2.3). We defined paused genes as having a PI $\geq$ 2. A PI of at least 2 indicates that there are more reads in the promoter region as compared to the gene body, thus implicating the presence of a paused polymerase. We chose this cutoff based on statistical analysis performed by Core et al. in their initial identification of pausing at human genes by GRO-Seq (2008). This cutoff is widely accepted in the field. We also examined genes with PIs both above and below 2 and found that those genes with a PI of greater than 2 showed a visually discernable difference in the accumulation of RNA polymerase at the promoter region as compared to those genes that had a PI of less than 2. We then identified genes that showed a significant enrichment of active RNA polymerase in the promoter region as compared to the gene body ($\chi^{2,}$ $p_c < 0.05$) (Figure S2.2). We performed this analysis to be sure that the genes that are identified as paused by PI also show a statistically significant enrichment of RNA polymerase at their promoters. This step eliminated genes that had very few reads in the promoter and gene body which lead to inflated PIs that were not representative of the biological state of the gene. Of the 11,362 genes in our analysis, 7,264 are paused based on these criteria. Figure 2.1B shows the pausing indices of these 7,264 genes. The average PI is 40 and the median PI is 16.

Among the genes that we identify as paused, we find the genes *FOS* (2.4) and *MYC* (4.8), both of which are known to have promoter-proximally paused polymerases. We also identified a large number of genes which have not been previously described to have promoter-proximally paused polymerases such as *EGFR* (PI=7.7), *FGFR1* (PI=6.9), *POLG2* (PI=257) and *MAP2K5* (PI=79) (Figure 2.1C). The paused genes we identified are enriched for signaling pathways such as EGFR and TGFβ signaling as well as general transcription and gene expression (Figure 2.1D)

We confirmed the accumulation of RNA Pol II at the 5' end of genes by chromatin immunoprecipitation (ChIP) for the serine 5 phosphorylated (ser5P) form of RNA Pol II which should localize to promoter regions (Komarnitsky et al., 2000). To identify sites of RNA Pol II (ser5P) accumulation, we called peaks using MACS2 (model-based analysis of ChIP-seq) (Zhang et al, 2008). The MACS2 program was run by Zhengwei Zhu. This method compares the ChIP samples to our input in defining peaks, thus generating a more rigorous peak definition. This allows us to be more confident in the sites of localization we identify. For this specific antibody, we examined only peaks that showed a fold enrichment of greater than 5. This fold enrichment score represents the amount of enrichment over input. We chose this cutoff based on visual inspection of >50 genes with fold enrichment scores between 7 and 3. We found the fold enrichment score of 5 to show the clearest enrichment of RNA Pol II ser5P over input as compared to lower cutoff values. We find that 64% of the paused genes identified by PRO-Seq also show an enrichment of RNA Pol II ser5P at their promoter by ChIP-Seq. We do not see a correlation between the pausing index and presence of RNA Pol II ser5P by ChIP. Because the PRO-Seq technique is a much more sensitive measure of RNA Pol II localization, we are able to identify a larger number of polymerases than by ChIP-seq. We do find that if the fold enrichment threshold is lowered, we see an increase in the number of genes that overlap between the two

data-sets; however, we believe our enrichment cutoff best represents those genes that have a true enrichment of RNA Pol II ser5P. Even with our strict enrichment cutoff and limitations of the ChIP-Seq technique, we see that all of the paused genes show a similar distribution of RNA Pol II by both techniques (Figure 2.2A).

Because PRO-Seq provides us with the ability to map the location of the polymerase with near single nucleotide resolution, we were able to map the location of the polymerase within the 2kb promoter region of our paused genes. To do this, we compared the ends of the PRO-Seq reads to the TSS. The end of the PRO-Seq reads represent the location where the polymerase stopped following the incorporation of the biotin-labeled NTP. Because the nuclear run-on is performed with an excess of biotin-labeled NTPs, the polymerase should incorporate the biotin-labeled NTP rapidly, thus stopping the polymerase within one or two nucleotides. In comparing the location of the polymerase to the TSS, we find that paused RNA Pol II is enriched between 50-200 nucleotides downstream of transcription start sites by PRO-Seq. This is similar to what has been shown at paused genes in *Drosophila* (Figure 2.2B) (Gilmour and Lis, 1986). This finding increases our confidence that our 7,264 paused genes are truly paused.

## NELF Localizes to Paused Genes

NELF and DSIF are responsible for holding RNA Pol II in a paused state just downstream of the TSS (Fraser et al., 1978, Wada et al., 1998, Yamaguchi et al., 1999, Wu et al., 2003, Wu et al., 2005). It has been shown that DSIF plays a role in both pausing and active elongation, whereas NELF has been shown to function only in pausing (Wada et al., 1998, Yamaguchi et al., 1999). For this reason, we were specifically interested in studying NELF. NELF is a 4-subunit complex comprised of NELFA, NELFB, NELFC/D and NELFE (Wu et al.,

2005, Yamaguchi et al., 1999). Because the function of this complex requires the presence of all subunits, it is possible to ascertain the location of the entire complex by examining the location of one subunit. Because NELFA interacts directly with RNA Pol II (Yamaguchi et al., 1999, Yamaguchi et al., 2002, Vos et al., 2016), we began by examining the location of the NELF complex through ChIP for the NELFA subunit. We used MACS2 to identify peaks of NELFA binding. The parameters used for peak calling for the RNA Pol II ser5P ChIP-Seq were also used for all subsequent ChIP-Seq experiments. Because we are interested in identifying paused genes with NELF localization, we searched for peaks in genes that are greater than 1kb in length and at least 1kb away from another gene. These are the same requirements we used when searching for paused genes by PRO-Seq. We also defined a NELFA peak as having a fold enrichment score of greater than 5 over input. We came to this cutoff in a similar manner as what was performed for the RNA Pol II ser5P ChIP-Seq analysis. We believe this cutoff best represents the presence of NELFA at these genes based on a visual inspection of >50 genes with a fold enrichment score between 7 and 3. We identified 8,952 NELFA peaks in our primary human dermal fibroblasts with 5,002 of these peaks found in the promoter region of genes. We found that both NELF and RNA Pol II show a similar distribution at these genes by ChIP-Seq and PRO-Seq (Figure 2.2A). The paused genes *EGFR*, *FGFR1*, *POLG2* and *MAP2K5* all show an enrichment of NELFA at their promoters (Figure 2.3A). We also confirmed the enrichment of NELF and RNA Pol II pser5 by ChIP-qPCR for the known paused genes *JUNB* and *FOS*, as well as a *TGFB1*, a gene we identified as paused (Figure 2.3B)

Next, to identify which of our paused genes also showed an enrichment of NELF in the promoter region, we compared results from PRO-seq with those from the NELF ChIP. We found that 4,433 (61%) of our 7,264 paused genes have NELF localized to their promoter. Because

48

PRO-Seq is a much more sensitive technique, it is likely that we are able to identify many more paused genes as compared to those that have NELF. This is a similar challenge to what we see with our RNA Pol II ser5P ChIP-seq. We do find that as we lower the fold enrichment threshold that we see an increase in the number of paused genes that show NELFA localization at their promoter. However, these peaks do not visually show a clear enrichment of NELFA over input and thus we are most confident in those peaks with a fold enrichment of greater than 5. We see no correlation between the PI and the presence of NELFA at the promoter. Because pausing can occur at genes with varying levels of transcription, it is reasonable that NELF would be found at genes with a wide-range of PIs. This indicates that NELF is enriched at all paused genes rather than a subset, thus implicating NELF as a key regulator of pausing in general.

Because DSIF is also known to play a role in mediating the pause, we asked if DSIF is also localized to the promoters of paused genes. We find that DSIF, NELF and RNA Pol II show similar patterns of localization (Figure 2.2A). Specifically, we find that DSIF is found at the promoters of 3,717 (51%) paused genes with over 94% of these also showing NELF localization. The steps for identifying enrichment of DSIF are the same as those used previously. We used a fold enrichment cutoff of 6 to identify what we characterize as true sites of DSIF enrichment. We used the same methods to calculate this cutoff as were described previously.

Because localization of NELF does not imply function, we knocked down NELF and look at changes in pausing. It has been shown that loss of one subunit of NELF results in the loss of all others at the protein level (Yamaguchi, et al., 2002, Sun et al., 2007), so we knocked down the NELFA subunit. We saw a greater than 80% reduction in *NELFA* transcript with a corresponding decrease in NELFA protein (Figure 2.4A and Figure S2.5). We also see a decrease in the expression of NELFB and NELFE protein following NELFA knockdown (Figure

2.4A). We see no change in cellular growth following NELF knockdown (Figure S2.6). Upon

loss of the NELF complex, we see a global decrease in active RNA Pol II in the promoter region

of paused genes as well as a significant decrease in PI (Figure 2.4B and 2.4C). We find that the

average PI before knockdown is 16.9 and the average PI following knockdown is 11.9. The

paused genes *EGFR*, *FGFR1*, *MAP2K5* and *POLG2* all show a reduction of active RNA Pol II at

their promoters upon loss of NELF (Figure 2.4D). Of the 4,433 paused genes that have NELF

localized to their promoters, 1,986 shows at least a 10% decrease in pausing index. Because

promoter-proximal pausing has been shown to be an important step in regulating elongation, a

10% decrease in pausing would have effects on the amount of elongation occurring at the target

genes. We also identify a number of genes that show a decrease in polymerase at the promoters;

however, they show no change in PI. These are not included in our list of NELF-paused genes;

however, NELF may still be playing a role in maintaining polymerase at the promoters of these

genes. We find that these genes are enriched for signaling pathways such as EGFR and TGFβ

signaling as well as lipid metabolism and general transcription (Figure 2.4E). Collectively, we

find that NELF is important for regulating promoter-proximal pausing at over 1,900 genes.

Specifically, we see an enrichment of NELF-mediated paused genes in signal transduction

pathways.

## NELF-mediated Pausing Influences Human Gene Expression

As a mediator of promoter-proximal pausing, NELF slows the transition into elongation,

which leads to decreased expression of target genes. However, it has been shown in *Drosophila*

that NELF is found at highly expressed genes where its function is to maintain an open

chromatin state to facilitate re-initiation of the polymerase and subsequence gene expression

(Gilchrist et al., 2008). To study the role of NELF in mediating human gene expression, we

examined gene expression before and after NELF knockdown in our primary human dermal

fibroblasts. We found that 11,615 genes change in expression at least 10% in either direction

following NELF knockdown. Because we are most interested in studying the role of NELF in

mediating the expression of paused genes, we focused only on the 4,433 paused genes that

showed NELF localization. Of the 4,433 paused genes that show NELF localization, 1,396 genes

show an increase in expression (>10%) following NELF knockdown (Figure 2.5A and Figure

S2.7). There are also 1,372 genes that show a decrease in expression (>10%). We believe an

expression cutoff of 10% allows us to best identify genes whose expression is regulated by

NELF. NELF serves as only one of many components that regulate pausing, thus loss of NELF

may have only a modest effect on the target genes because the other regulatory components are

able to regulate pausing to some degree. Also, pausing serves as a fine-tuning mechanism of

gene expression, so we believe a cutoff of 10% allows us to study those genes where a loss of

pausing results in only a slight increase or decrease in expression yet is still a biologically

relevant consequence of a loss of pausing. Of these 2,768 genes that change in expression at least

10%, 43% also show at least a 10% reduction in PI. Representative examples of these genes are

found in Figure 2.5B.  For genes where NELF was holding the polymerases from entering into

elongation, the silencing of NELF would facilitate elongation, and likely lead to an increase of

gene expression.  To examine how the knockdown of NELF leads to a decrease in gene

expression, we asked if it is due to a loss of active transcription. To investigate this, we

compared the amount of nascent RNA in the promoters of the genes that decreased in expression

to those genes that increased in expression following loss of NELF. We found that following

NELF knockdown, there is significantly less nascent RNA in the promoters of genes that

decrease in expression ($p=6x10^{-7}$).  This was observed for several genes in *Drosophila* (Gilchrist

et al., 2008).  Here enabled by advances in sequencing technology, we observed that for over 1,000 human genes, NELF may aid in facilitating subsequent rounds of transcription. Further studies will need to be performed to investigate the mechanism by which NELF is able to perform this function. Studies in *Drosophila* have shown that NELF is able to clear the first nucleosome to maintain an open promoter (Gilchrist et al., 2008). It is possible that NELF is performing a similar function at human promoters; however, studies investigating the location of the nucleosomes as well as the chromatin marks associated with the promoters before and after NELF knockdown need to be performed to correctly assess the chromatin state at these promoters. Metagene plots of these two classes of genes are found in Figure 2.5C. These results show that NELF influences human gene expression in two ways. The first is by holding the polymerase at the promoter region to eliminate aberrant transcription of the target genes. The second is by facilitating the re-initiation of RNA Pol II at promoter regions to increase the transcription of target genes; however, the mechanism by which NELF influences this process remains to be elucidated.

## NELF influences EGF signaling by regulating expression of the EGF receptor

Upon showing that NELF regulates the expression of *EGFR* at the transcript level. We wanted to see if this increase in transcript expression correlated with an increase in expression of EGFR at the protein level. Following NELF knockdown in two primary human dermal fibroblast lines, we see a slight increase in EGFR expression by western blot (Figure 2.6A). This is the first description of pausing leading to a change in both transcript and protein expression. Upon seeing this increase we wanted to investigate whether there was any effect on EGFR signaling. EGFR requires dimerization and autophosphorylation to activate, we hypothesized that the increase in EGFR would lead to increased response to EGF due to the increase in the amount of receptor.

We treated cells with recombinant human EGF with and without NELF. We then measured the expression of *CDKN1A*, a known target of EGFR. *CDKN1A* is not paused in our cells, so any change in *CDKN1A* expression following loss of NELF and EGF treatment should not be due to changes in pausing. We found that upon stimulation we see an increase in *CDKN1A* expression and in the absence of NELF we see an even large increase in expression of *CDKN1A* (Figure 2.6B). These findings implicate that pausing may be functioning to regulate stimulus response through a mechanism other than ensuring a coordinated gene expression response. We show that pausing can alter response to EGF stimulation by increasing the amount of receptor.

## NELF localizes to eRNA

Our chromatin immunoprecipitation study of NELF showed that NELF not only localizes to promoter regions of genes, it is also found in other parts of the human genome. Though we have focused on NELF complexes that are found in promoter regions to this point, there are sites outside of the promoter regions where NELF binds. The NELFA ChIP identified 3,945 sites that are outside of promoter regions. These sites have similar characteristics to those in the promoter regions including having a fold enrichment score of greater than 5. Of these regions, we find 3,170 in coding sequences but outside of the promoters, and 775 in intergenic regions. Since it has been shown that enhancer RNA (eRNA) can bind to NELF to facilitate the transition into active elongation (Schaukowitch et al., 2014), we asked if NELF was localizing to eRNA in these intergenic regions

Active enhancers are marked by the acetylation of lysine 27 of histone H3 (H3K27ac) (Creyghton et al., 2010). To see whether NELF found outside of the promoter region co-localize with eRNA, we asked if these sites contain the H3K27ac mark and if there is evidence for active

53

transcription. We performed ChIP-seq with an antibody specific for H3K27ac and searched for sites of H3K27ac enrichment (MACS2, fold enrichment >6). The fold enrichment cutoff analysis was performed as described previously. We believe these peaks represent active enhancers. To confirm that these active enhancers also generate transcripts, we searched for active enhancers that also show at least 10 PRO-Seq reads. We used 10 PRO-Seq reads as our cutoff since we believe at least 10 reads represents an active polymerase at both our paused genes and at these sites. We found that 2,024 of our 3,945 NELFA binding sites that are outside of the promoter region show enrichment of H3K27ac and active transcription. We define these regions as eRNA. Of these 2,024 eRNA, we found 1,635 in known coding sequence and 390 in intergenic regions. Representative examples of these eRNA are shown in Figure 2.7A.

Next, we investigated whether NELF plays a role in maintaining these eRNA. To study this, we knocked down NELF and examined the effect of NELF on transcription at these eRNA by PRO-Seq. We found that 1,169 of these eRNA show at least a 50% reduction in nascent transcript abundance upon loss of NELF. Because it is currently unknown how NELF is facilitating eRNA transcription, we wanted to be extremely confident in the eRNA we describe as regulated by NELF. This led us to choose a very strict cutoff of a 50% decrease in nascent transcription upon loss of NELF because these eRNA represent the eRNA that are predominantly regulated by NELF. We do find eRNA that show smaller changes in transcription level upon loss of NELF that may still represent NELF regulated eRNA; however, there may be other mechanisms regulating their transcription outside of NELF. Two representative examples of these eRNA are show in Figure 2.7A. These results implicate NELF as a regulator of eRNA transcription.

Next, we asked if loss of NELF influences gene expression through these eRNAs. To do this, we compared the expression levels of genes nearest to enhancers before and after NELF knockdown. In order to identify if the effect of NELF is due to the eRNA rather than promoter-proximal pausing, we focused only on genes that are not paused. Of the 1,169 eRNA regulated by NELF, 922 are found within the coding sequence (outside of the promoter) of genes without evidence for paused RNA polymerases. We found that of these 922 genes, the expression of 297 genes changed, either an increase or decrease, (>10%) upon loss of NELF. Because the role of eRNA in facilitating gene expression is still unclear, we wanted to include genes that show even modest changes in expression since it is possible that a 10% change is representative of the biological function of that eRNA. In addition, there are 220 eRNA that are in intergenic regions; for these, we looked at the expression of genes closest to the eRNAs. We found 97 of these genes that are eRNAs show an expression change (>10%) upon loss of NELF. Figure 2.7B shows two representative examples of genes whose expression changes following NELF knockdown. Figure 2.7C is a genome-wide map of 1,169 eRNAs whose expressions are dependent on NELF. We also plotted the 394 putative target genes of these eRNAs. We recognize that these 394 genes may not be the direct targets of the nearby eRNA since eRNA are known to act over long distances. It is also possible that we underestimated the number of genes whose expression is influenced by these eRNA by only examining nearby genes. Further investigation is needed to correctly identify the targets of these eRNA.

## Conclusions

RNA Pol II pausing is an important regulatory step between initiation and active elongation in metazoans. Here, we generated a map of human genes in primary human dermal fibroblasts with paused RNA polymerases. By PRO-Seq and chromatin immunoprecipitation

with Ser5-P RNA polymerase II, we identified over 7,000 paused genes. The RNA polymerases accumulated between 50-120 base pairs downstream of the transcription start sites (Figure 2.1B and 2.2B). These results allowed us to expand the list of genes regulated by proximal-promoter RNA polymerase pausing; these include genes involved in EGFR and TGFβ signaling. This demonstrates the importance of pausing in coordinating a transcriptional response to stimuli.

RNA Pol II is held in a paused state by the pausing factors NELF and DSIF (Wada et al., 1998, Yamaguchi et al., 1999, Wu et al., 2003, Wu et al., 2005). By ChIP for the pausing factors NELF and DSIF, we indeed found both NELF and DSIF binding to the promoters of paused genes (Figure 2.2A). However, some of the paused genes appeared not to be targets of NELF and/or DSIF. Upon closer inspection of these genes, we saw some binding of one or both proteins; however, the binding did not meet our inclusion criteria. In taking a more detailed look at NELF localization, we find that NELF is localized in the same region as the paused polymerase (Figure S2.4).

We also examined how NELF binding to genes affect their expression levels. Following NELF silencing, the expression levels of 1,396 genes increased in gene expression (by at least 10%), and a nearly number of genes (1,372) decreased in expression levels (by at least 10%). Given that NELF maintains pausing, knockdown of NELF would promote elongation and increases gene expression. For those genes that decreased in expression, instead of maintain paused polymerases, NELF maintains an open chromatin. We found a significant decrease in nascent transcription in the promoters of these following NELF knockdown (Figure 2.5C). This suggests that NELF was maintain chromatin states that promote transcription and upon its knockdown, the chromatin closes and deters subsequent rounds of transcription. These results identified a second function for NELF. We also show that NELF is able to influence the

expression of the EGF receptor at both the transcript and protein level. This indicates that NELF-mediated pausing may serve to ensure correct pathway functioning by regulating the expression of the individual components.

Our results identified yet a third function of NELF. We found that NELF is enriched at sites outside of gene promoters. Recent work has shown that NELF can localize to eRNA in *Drosophila* (Core et al., 2012); however, this phenomenon has not been seen in humans. We found that NELF and H3K27ac, a mark of active enhancers (Creyghton et al., 2010), co-localize to over 3,900 sites outside of promoter regions. We also that these sites are actively transcribed suggesting the presence of eRNA. Following NELF knockdown, there is a decrease in active transcription of these eRNAs (Table 2.4). Enhancer RNA have been shown to influence the expression of a number of genes in humans (Li et al., 2013, Mousavi et al., 2013, Schaukowitch et al., 2014). In our studies, we found that NELF regulates the abundance of eRNA. Upon loss of NELF we found that the expression of genes near the eRNA show changes in expression (Figure 2.7B). Our findings suggest that NELF influences gene expression by facilitating eRNA transcription.

Collectively, our work has generated a comprehensive view of pausing in primary human dermal fibroblasts. We have identified three ways in which NELF influences gene expression. The first is by holding the polymerase back from entering active elongation until a signal is received, the second is by facilitating re-intiation of the polymerase at promoters and the third is by regulating the transcription of eRNA. We have also shown NELF to be a key regulator of pausing and subsequent human gene expression. This is the first study showing an association of NELF with the transcription of eRNA; however, the mechanism for this remains to be elucidated. Because most enhancers are cell-type specific (Bernstein et al., 2012, Heinz, et al.,

2015), it will be important to examine other cell types to identify NELF associated eRNA and target genes.

## Materials and Methods

**Cell Culture: Primary Human Dermal Fibroblasts**

Cultured human primary newborn foreskin fibroblasts were used from 4 individuals (FB3356, FB3383, FB3385, FB3389). Fibroblasts were cultured at a density of $5 \times 10^5$ cells/mL in DMEM with 10% Fetal Bovine Serum, 100 units/mL penicillin/streptomycin and 2mM L-glutamine. Each batch was cultured separately.

**siRNA Knockdown**

Transient transfections with siRNA were performed using RNAiMAX Lipofectamine (Thermo Fisher 137708030), following the manufacturer's instructions. Non-target control (E-HUMAN-XX-0005) and NELFA (L-012156-00-0005) siRNAs were pools of 4 siRNAs and were obtained from GE Healthcare Dharmacon. RNA, protein or nuclei were harvested 96 hours post-transfection.

**RNA Isolation, reverse transcription and quantitative real-time PCR**

To measure expression of *NELFA*, *FOS*, *MYC*, *EGFR* and *SNAI1*, primary human dermal fibroblasts were transfected with non-target control or NELFA siRNAs (see above). RNA was harvested 96 hours post transfection with the Qiagen RNeasy Mini-Kit (74104). RNA (250ng) was used for cDNA synthesis using the TaqMan Reverse Transcription cDNA synthesis kit (Thermo Fisher Scientific N8080234) with oligodT. *GAPDH*, *NELFA*, *FOS*, *MYC*, *EGFR* and *SNAI1* expression levels were measured by quantitative real-time PCR using gene specific primers (Table 2.1) and Power SYBR Green Master Mix (Thermo Fisher Scientific) with three

technical replicates for each PCR. *GAPDH* was used as a housekeeping gene to normalize the amount of cDNA. Fold change of target gene mRNA levels was calculated using the $\Delta\Delta C_t$ method.

**Western Blotting**

Western blotting was performed as described previously (Gregory and Cheung, 2013). Briefly, protein lysates were collected in 1X RIPA lysis buffer (Cell Signaling 9806) following the manufacturer's protocol. Immunoblotting was performed with the following antibodies, NELFA (sc-32911), NELFB (ab48366), NELFE (Millipore ABE48), EGFR (Cell Signaling 4267S). Membranes were developed using Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific 32106).

**Precision Run-On Sequencing (PRO-seq)**

PRO-seq libraries were prepared as described previously (Kwak et al., 2013). Briefly, approximately $1.5 \times 10^6$ nuclei from each condition were added to 2 X Nuclear Run-On (NRO) reaction mixture (10 mM Tris-HCl pH 8.0, 300 mM KCl, 1% Sarkosyl, 5 mM MgCl$_2$, 1 mM DTT, 0.03 mM each of biotin-11-A/C/G/UTP (Perkin-Elmer), 0.8 u/µl Rnase inhibitor) and incubated for 3 min at 37˚C. Nascent RNA was extracted by phenol (Trizol LS)/chloroform then was fragmented by base hydrolysis in 0.2 N NaOH on ice for 15 min. The reaction was neutralized by adding 1 X volume of 1 M Tris-HCl pH 6.8. Following fragmentation, nascent RNA was purified using streptavidin beads and ligated with reverse 3' RNA adapter (5'-rCrCrUrUrGrGrCrArCrCrCrGrArGrArArUrUrCrCrA). Again, biotin-labeled products were enriched by streptavidin bead binding and extraction. For 5' end repair, the RNA products were successively treated with 5' pyrophosphatase (NEB) and polynucleotide kinase (PNK, NEB). 5' repaired RNA was ligated to reverse 5' RNA adaptor (5'-

rCrCrUrUrGrGrCrArCrCrCrGrArGrArArArUrUrCrCrA) The products were further purified by the third round of streptavidin bead binding and extraction. RNA was reverse transcribed using 25 pmol RT primer (5'AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA-3'). The product was amplified 22 cycles and products between 150-300bp (insert > 70 bp) were PAGE purified before being analyzed by Illumina HiSeq 2500 instrument. (Figure S2.1) Dr. Alan Bruzel provided technical assistance for the PRO-Seq experiments described in this chapter. Jennifer Fox ran the Illumina HiSeq 2500 Instrument.

**Chromatin Immunoprecipitation (ChIP)**

ChIP was performed as described previously (Watts et al., 2010). Briefly, cells were cross-linked with 1% formaldehyde for 15 min. Cross-linking was stopped with XM glycine for 5 min. Following cell lysis, chromatin was purified and sonicated to 1000-200bp fragments by sonication (30s on, 30s off). 10% of the supernatant was saved as input control and was processed at the cross-linking reversal step. The rest of the supernatant was pre-cleared with Protein A beads (Sigma: 11719408001) and anti-rabbit IgG. Then, the supernatant was immunoprecipitated overnight with either 5ug anti-rabbit IgG, 5ug NELFA (sc-32911), 5ug SPT5 (sc-28678X) or 5ug RNA Pol II-5P (ab5408). Following reverse cross-linking, DNA was precipitated using the Qiagen QIAquick PCR Purification Kit (28106). ChIP-seq libraries were prepared following the NuGen344 32 reaction kit (0344-32). Libraries were analyzed on the Illumina HiSeq 2500 instrument. Jennifer Fox ran the Illumina HiSeq 2500 instrument. Validation of targets was performed by real-time PCR using Power SYBR Green Master mix (Thermo Fisher Scientific) with the primers listed in Table 2.2. Dr. Jason Watts provided technical assistance with the ChIP-Seq experiments described in this chapter.

**mRNA Sequencing**

RNA was isolated from primary human dermal fibroblasts 96 hours following siRNA treatment as described above. Libraries were prepared using the Illumina TruSeq RNA Sample Preparation Protocol. Samples were analyzed on the Illumina HiSeq 2500 Instrument. Jennifer Fox ran the Illumina HiSeq 2500 Instrument.

**Sequence Alignment**

For RNA-Seq, PRO-Seq and ChIP-seq, raw sequencing files were processed by trimming the adapter sequences from the ends of reads using fastx_clipper from FASTX-Toolkit (Hannon Lab). Sequences of low-quality at the ends of reads represented by a stretch of "#" in the quality score string in FASTQ file were removed along with PolyA (more than five consecutive As at the ends of reads. Reads that were >35nt after trimming were included for downstream analysis. Reads were aligned to human reference genome (hg18) using GSNAP (Wu and Nacu, 2010) (version 2013-10-28) with the following parameters: Mismatches < [(read length +2)/12/2]; Mapping score >2-; Soft-clipping on (-trim-mismatch-score = -3). Bam files were then generated. Reads with identical sequences were counted as one read in order to remove potential PCR duplicates. Table 2.3 shows sequenced and uniquely aligned reads in each sample. This analysis was performed by Zhengwei Zhu.

**PRO-Seq Analysis**

Custom scripts were used to calculate the number of reads in the promoter and gene body regions. This script was written and run by Zhengwei Zhu. The promoter region is defined as +/-1kb surrounding the annotated transcription start sites. The read counts from each sample were normalized to the number of million mapped reads. Genes that are within 1kb of another gene or less than 1kb in length were removed. We then calculated a pausing index for the remaining genes (Core et al., 2008) (Figure S2.2).  All steps in analysis are presented in figure S2.3.

Metagene plots of paused genes were generated using custom scripts written by Yooree Chung and Zhengwei Zhu.

**ChIP-Seq Analysis**

Peaks were identified by MACS2 (Zhang et al, 2008). A fold change cutoff of greater than 5 was applied to peaks from NELFA and RNA Poll Ser5P ChIP and a fold change cutoff of greater than 6 was applied to peaks from SPT5 (DSIF) and H3K27ac ChIP. MACS2 was run by Zhengwei Zhu. Overlap of datasets was performed by Dr. Xiaorong Wang.

**RNA-Seq Analysis**

mRNA expression levels were quantitated and normalized using the RUV method (Risso et al., 2014). Bioinformatic analysis of the RNA-Seq data was performed by Zhengwei Zhu.

**EGF Time Course**

Primary human newborn dermal fibroblasts (FB3383) were plated at $5x10^5$ cells/mL and were transfected with NT or NELFA siRNA as described above. 96 hours post transfection cells were treated with human recombinant EGF (Life Technologies PHG0314) for 30m. RNA was collected following the protocol described above. qPCR was performed as described previously for *NELFA* and *CDKN1A*.

# References

Aiyar, S.E., Sun, J., Blair, A.L., Moskaluk, C.A., Lu, Y., Ye, Q., Yamaguchi, Y., Mukherjee, A., Ren, D., Handa, H., Li., R. Attenuation of estrogen receptor a-mediated transcription through estrogen-stimulated recruitment of a negative elongation factor. Genes Dev. 2004 Sep 1; 18(17):2134-46.

Amleh, A., Nair, S.J., Sun, J., Sutherland, A., Hasty, P., Li, R. Mouse cofactor of BRCA1 (Cobra1) is required for early embryogenesis. PLoS One. 2009 Apr 2;4(4):e5034.

Bunch, H., Zheng, X., Burkholder, A., Dillon, S.T., Motola, S., Birrane, G., Ebmier, C., Levine, S., Fargo, D., Hu, G., Taatjes, D.J., Calderwood, S.K. TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. Nature. 2014 Aug 31;21(10):876-83.

Chen, F.X., Woodfin, A.R., Gardini, A., Rickels, R.A., Marshall, S.A., Smith E.R., Sheikhattar, R., Shilatifard, A. PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. Cell. 2015 Aug 27;162(5):1003-15.

Core, L.J., Waterfall, J.J., Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008 Dec. 19;322(5909):1845-48.

Core, L.J., Waterfall, J.J., Glichrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., Lis, J.T., Defining that status of RNA polymerase at promoters. Cell Rep. 2012 Oct 25;2(4):1025-35.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., Jaenisch, R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. PNAS. 2010 Dec 14;107(50):21931-6.

Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L.J., Lis, J.T., Siepel, A., Kraus, W.L. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Mol Cell. 2013 Apr 25;50(2):212-22.

Fraser, N.W., Sehgal, P.B., Darnell, J.E. DRB-induced premature termination of late adenovirus transcription. Nature. 1978 Apr 13;272(5654):590-3.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., Peterlin, B.M. Dynamics of Human Immunodeficiency Virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Mol Cell Bio. 2004 Jan;24(2):787-95.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K.B., Collins, J.B., Li, L., Gilmour, D.S., Adelman, K. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev. 2008 Jul 15;22(14)1921-33.

Gilmour, D.S., Lis, J.T. RNA polymerase II interacts with the promoter region of the noninduced *hsp70* gene in *Drosophila melanogaster* cells. Mol Cell Bio. 1986 Nov;6(11):3984-9.

Gregory, BL., Cheung, V.G. Natural variation in the histone demethylase, KDM4C, influences expression levels of specific genes including those that affect cell growth. Genome Res. 2014 Jan;24(1):52-63.

Heinz, S., Romanoski, C.E., Benner, C., Glass, C.K. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol. 2015 Mar;16(3)144-54.

Kim, J.B., Sharp., P.A. Positive transcription elongation factor b phosphorylates hSPT5 and RNA polymerase II carboxy-terminal domain independently of cyclin-dependent kinase-activating kinase. J Biol Chem. 2001 Apr 13;276(15):12317-23.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., Ren, B. A high-resolution map of active promoters in the human genome. Nature. Aug 2005;436(11):876-80.

Kinnis, M., Isaacs, G.D., Core, L.J., Hah, N., Kraus, W.L. Postrecruitment regulation of RNA polymerase II directs rapid signaling responses at the promoters of estrogen target genes. Mol Cell Biol. 2009 Mar;29(5):1123-33.

Komarnitsky, P., Cho, E.J., Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. 2000 Oct 1;14(19):2452-60.

Krumm, A., Meulia, T., Brunvand, M., Groudine, M. The block to transcriptional elongation within the human c-*myc* gene is determined in the promoter-proximal region. Genes Dev. 1992 Nov;6(11):2201-13.

Kwak, H., Fuda, N.J, Core, L.J., Lis, J.T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science. 2013 Feb 22;339(6122):950-3.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., Oh, S., Kim, H., Glass, C.K., Rosenfeld, M.G. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature. 2013 Jun 27;498(7455):516-20.

Lu, X., Zhu, X., Li, Y., Liu, M., Yu, B., Wang, Y., Rao., M., Yang, H., Zhou, K., Wang, Y., Chen, Y., Chen, M., Zhuang, S., Chen, L., Lui, R., Chen, R. Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. Nucleic Acids Res. 2016 Aug 19;44(14):6853-57.

Marshall, N.F., Price, D.H. Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol Cell Biol. 1992 May;12(5):2078-90.

Marshall, N.F., Price, D.H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J Biol Chem. 1995 May 26;270(21):12335-8.

Marshall, N.F., Peng, J., Xie, Z., Price, D.H. Control of RNA polymerase II elongation potential by a novel carboxy-terminal domain kinase. J Biol Chem. 1996 Oct 25;271(43):27176-83.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., Lis, J.T. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev. 2011 Apr 1;25(7):742-54.

Mousavi, K., Zare, H., Dell'Orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., Sartorelli, V. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell. 2013 Sep 12;51(5):606-17.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grisson, S.F., Zeitlinger, J., Adelman. K. RNA polymerase is poised for activation across the genome. Nat Genet. 2007 Dec;39(12):1507-11.

Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D., Hasegawa, J., Omori, M., Inukai, N., Endoh, M., Yamada, T., Hiroshi, H. Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. Mol Cell Biol. 2003 Mar;23(6):1863-73.

O'Brien T., Lis, J.T. RNA polymerase II pauses at the 5' end of the transcriptionally induced *Drosophila hsp70* gene. Mol Cell Biol. 1991 Oct;11(10):5285-90.

Plet, A., Eick, D., Blanchard, J.M. Elongation and premature termination of transcripts initiated from c-Fos and c-Myc promoters show dissimilar patterns. Oncogene. 1995 Jan 19;10(2):319-28.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., Young, R.A. c-Myc regulates transcriptional pause release. Cell. 2010 Apr 30;141(3):432-45.

Risso, D., Nagi, J., Speed, T.P., Dudoit, S. Normalization of RNA-seq using factor analysis of control genes or samples. Nat Biotechnol. 2014 Sep;32(9):896-902.

Rougvie, A.E., Lis, J.T. The RNA polymerase II molecule at the 5' end of the uninduced *hsp70* gene of D. melanogaster is transcriptionally engaged. Cell. 9 Sep 1988;54(6):795-804.

Schaukowitch, K., Joo, J., Liu, X., Watts, J.K., Martinez, C., Kim, T. Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell. 2014 Oct 2;56(1):29-42.

Strobl, L.J., Eick, D. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-Myc *in vivo*. EMBO J. 1992 Sep;11(9):3307-14.

Sun, J., Watkins, G., Blair, A.L., Moskaluk, C., Ghosh, S., Jiang, W.G., Li, R. Deregulation of cofactor of BRCA1 expression in breast cancer cells. J Cell Biochem. 2008 Apr 15;103(6):1798-807.

Sun, J., Li, R. Human negative elongation factor activates transcription and regulates alternative transcription initiation. J Biol Chem. 2010 Feb 26;285(9):6443-52.

Sun. J., Pan, H., Lei, C., Yuan, B., Nair, S.J., April, C., Parmeswaran, B., Klotzle, B., Fan, J., Ruan, J., Li, R. Genetic and genomic analyses of RNA polymerase II-pausing factor in regulation of mammalian transcription and cell growth. J Biol Chem. 2011 Oct 21;286(42):36248-57.

Vos, S.M., Pollman, D., Caizzi, L., Hofmann, K.B., Rombaut, P., Zimniak, T., Herzog, F., Cramer, P. Architecture and RNA binding of the human negative elongation factor. Elife. 2016 Jun 10;5 pii:e14981.

Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., Handa, H., Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription *in vitro*. EMBO J. 1998 Dec 15;17(24):7395-403.

Wang, X., Hang, S., Prazak, Gergen, P.J. NELF potentiates gene transcription in the Drosophila embryo. PLoS One. 2010 July 9;5(7):e11498.

Watts, J.A., Zhang, C., Klein-Szanto, A.J., Kormish, J.D., Fu, J., Zhang, M.Q., Zaret, K.S. Study of FoxA pioneer factor at silent genes reveals Rfx-repressed enhancer at Cdx2 and a potential indicator of esophageal adenocarcinoma development. PLoS Genet. 2011 Sep;7(9):e1002277.

Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., Adelman, K. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. Mol Cell. 2015 Apr 16;58(2):311-22.

Wu, C., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., Gilmour, D.S. NELF and DSIF cause promoter proximal pausing on the *hsp70* promoter in *Drosophila*. Genes Dev. 2003 Jun 1;17(11):1402-14.

Wu, C., Lee, C., Fan, R., Smith, M.J., Yamaguchi, Y., Handa, H., Gilmour, D.S. Molecular characterization of *Drosophila* NELF. Nucleic Acids Res. 2005 Mar 1;33(4):1269-79.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., Handa. H. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol Cell. 2006 Jan 20;21(2):227-37.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., Handa, H. NELF, a mulitisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell. 1999 Apr 2;97(1):41-51.

Yamguchi, Y., Inukai, N., Narita, T., Wada, T., Handa, H. Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol Cell Biol. 2002 May;22(9):2918-27.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., Young, R.A. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet. 2007 Dec;39(12):1512-6.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
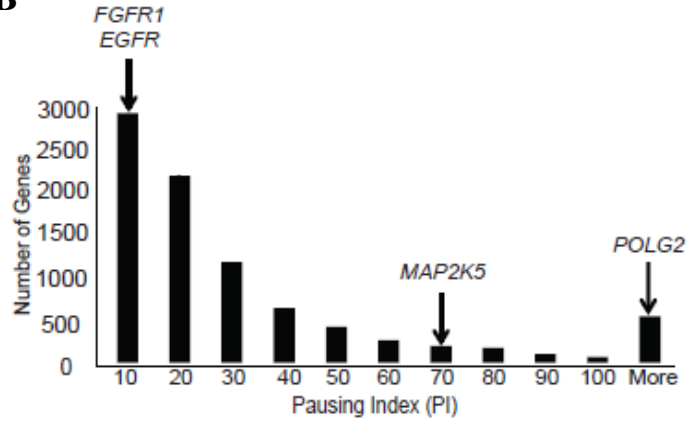
Zhou, Q., Li, T., Price, D.H. RNA Polymerase II elongation control. Annu Rev Biochem. 2012;81:119-43.
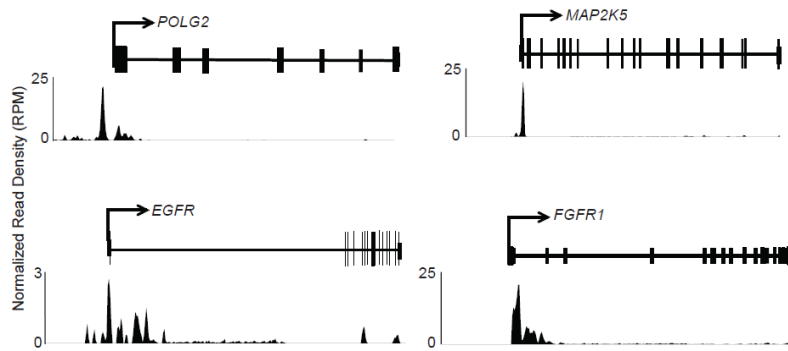
**Figure 2.1**
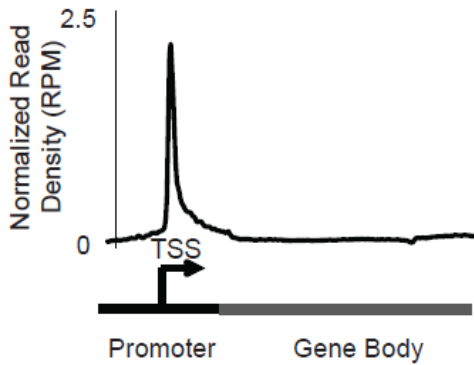


Figure 2.1 Promoter-Proximal Pausing Occurs Genome-Wide. A) Human genes show an accumulation of transcriptionally competent RNA Pol II at their 5' end. B) Histogram depicting the pausing indexes of the 7,264 paused genes (PI ≥ 2, $\chi^2 p_c < 0.05$) Script to calculate PI was written and run by Zhengwei Zhu. C) Location of active RNA polymerases on representative paused genes, *POLG2* (PI=257), *MAP2K5* (PI=79), *EGFR* (PI=7.7) and *FGFR1* (PI=6.9). The values in this plot were obtained from the custom metagene script written by Yooree Chung and Zhengwei Zhu. D) Pathway analysis of the 7,264 paused genes.
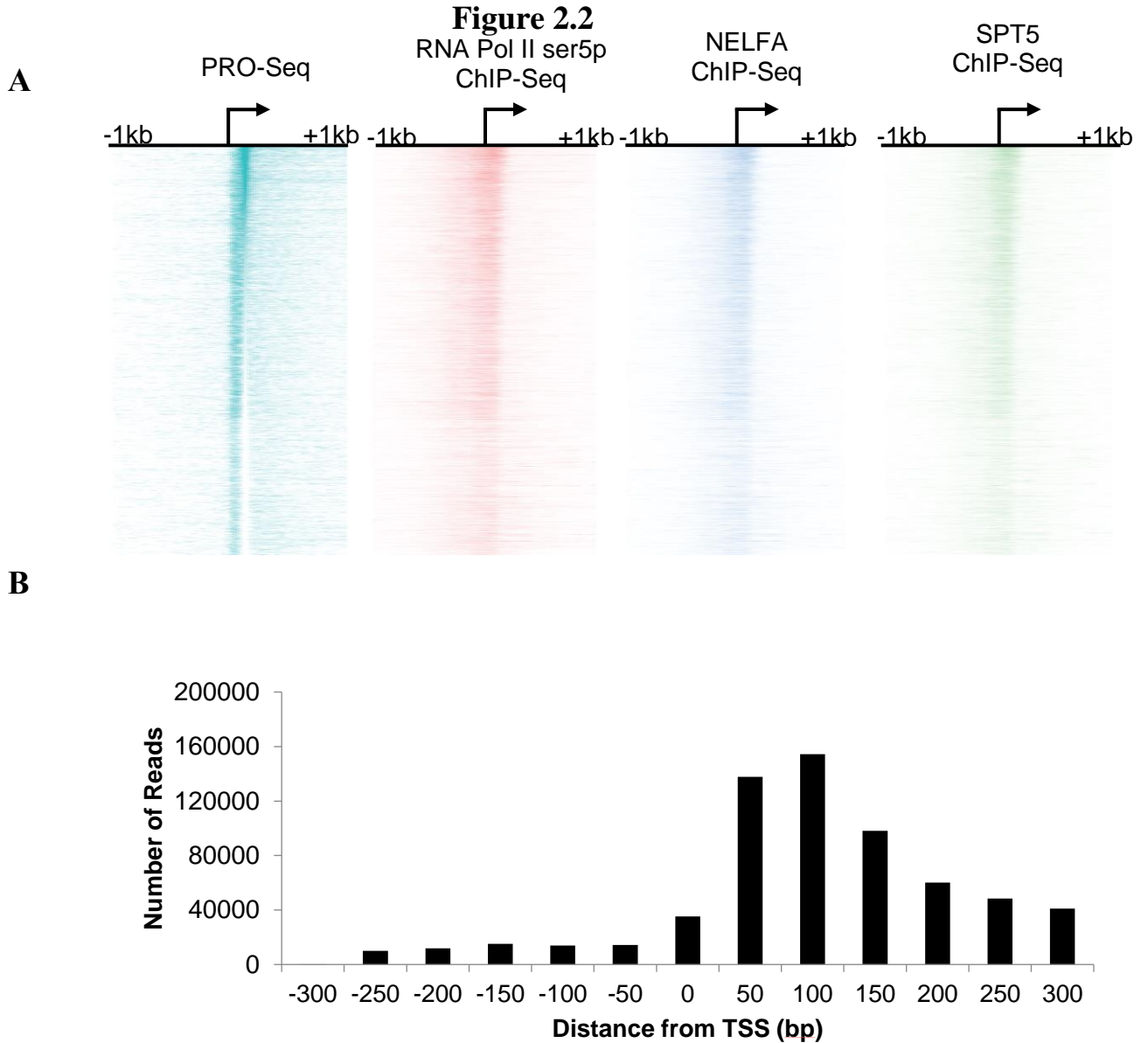
**Figure 2.2**

Figure 2.2 RNA Pol II, NELF and DSIF Localize to Paused Genes Promoters. A) Heatmaps depicting the locations of transcriptionally active RNA Pol II (PRO-Seq), RNA Pol II ser5p, NELFA and SPT5 (DSIF) in the promoter regions of 7,264 paused genes. The values represented in the heatmaps were obtained by utilizing the custom metagene script written by Yooree Chung and Zhengwei Zhu. B) Histogram depicts the location of transcriptionally active RNA Pol II identified by PRO-Seq.

**Figure 2.3**

Figure 2.3 NELF and RNA Pol II Localize to the Promoters of Paused Genes. A) PRO-seq, RNA Pol II pSer5 ChIP-seq and NELFA ChIP-seq reads across the paused genes *EGFR* and *POLG2*. The values represented in these plots were obtained by utilizing the custom metag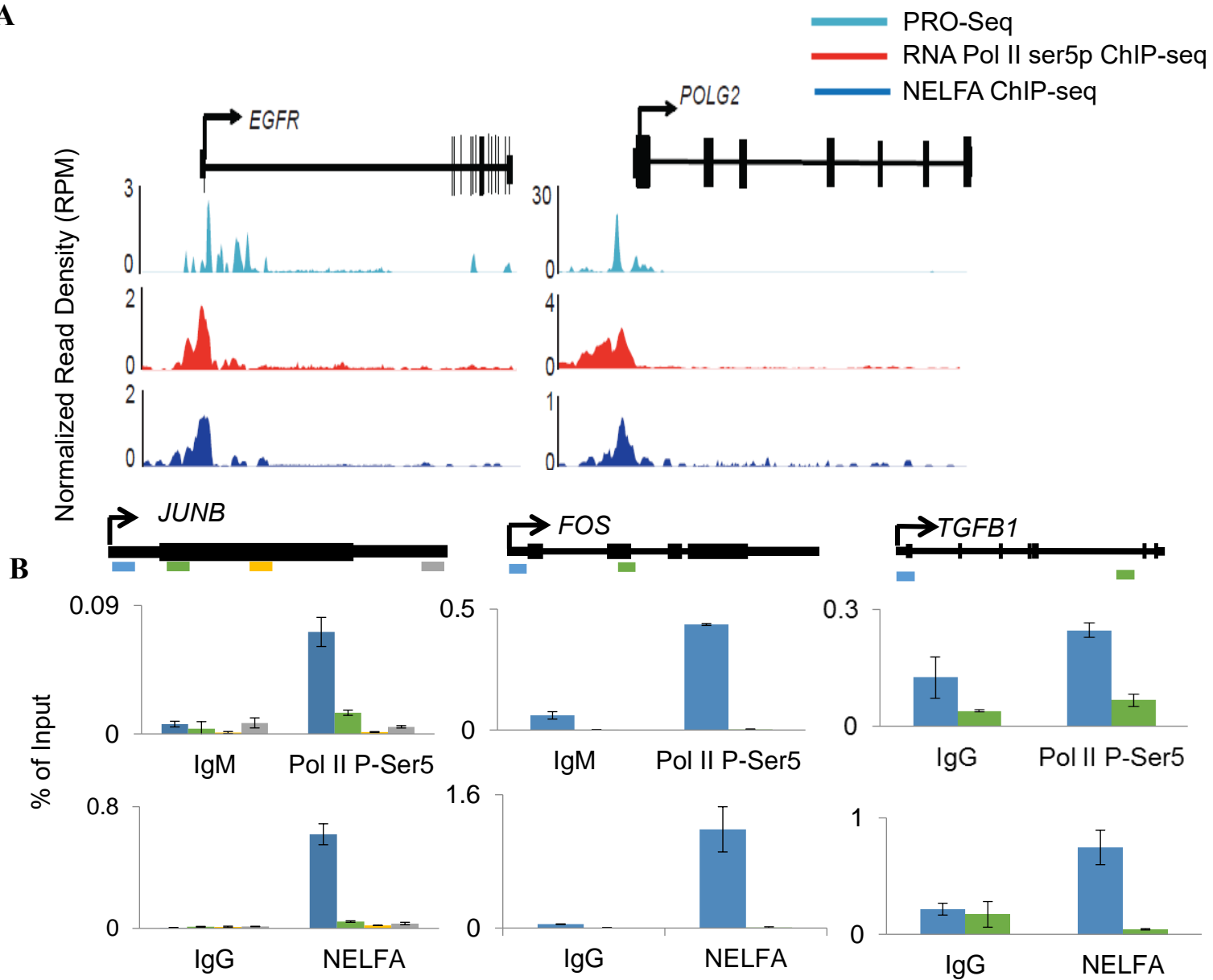ene script written by Yooree Chung and Zhengwei Zhu. B) ChIP-qPCR showing enrichment of RNA Pol II P-Ser5 at the 5' end of the paused genes *JUNB*, *FOS* and *TGFB1*.

# Figure 2.4
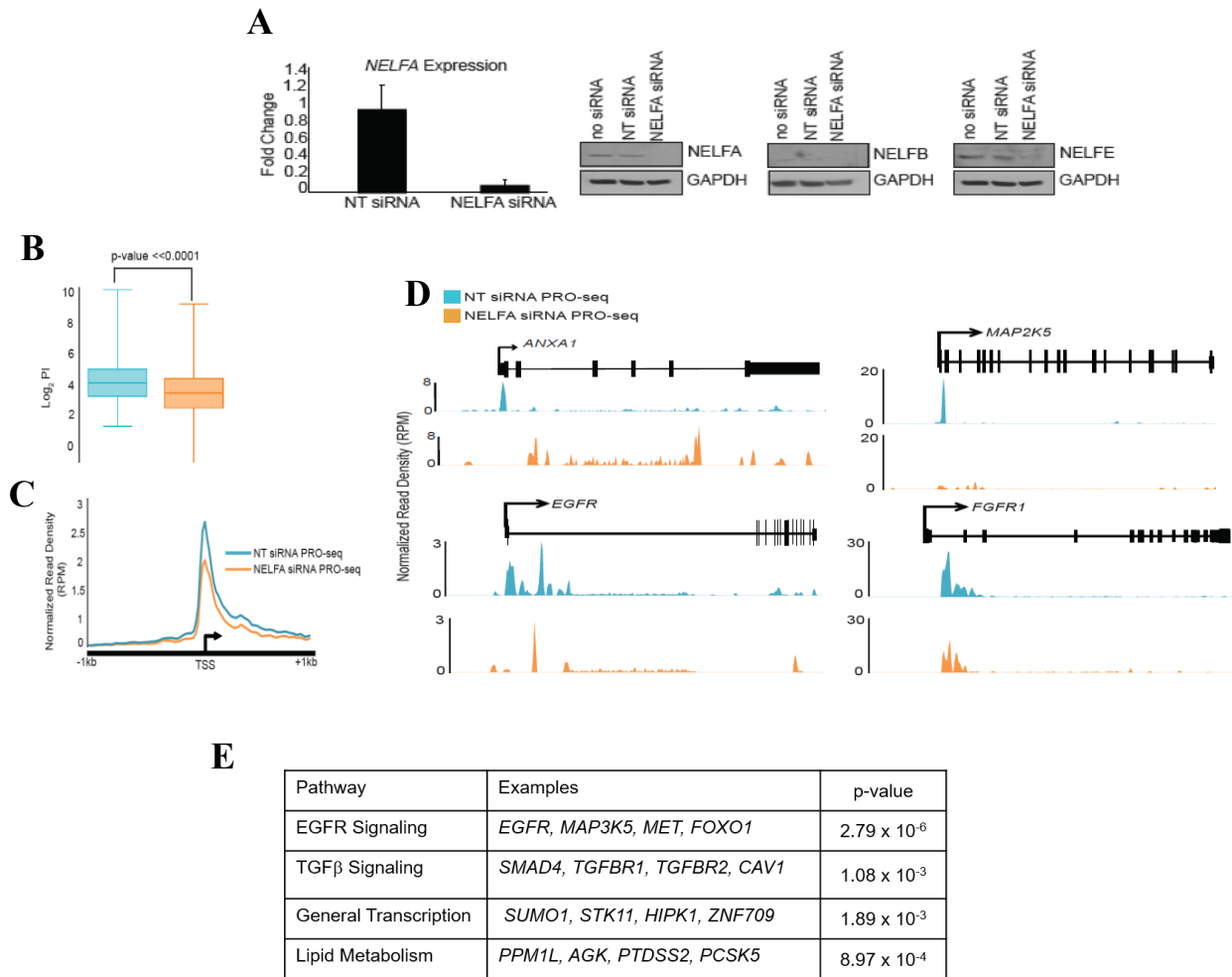


Figure 2.4 NELF Mediates Promoter-Proximal Pausing. A) Knockdown of NELFA results in loss of the entire NELF complex at the protein level. B) Box plot depicting significant decrease in pausing index (PI) following NELF knockdown (t-test, p<<0.0001). C) Metagene plot showing a decrease in the amount of transcriptionally active RNA Pol II at paused genes following NELF knockdown. D) Individual examples of PRO-seq reads across paused genes that show at >10% decrease in PI following NELF knockdown. The values shown in this figure were obtained by the custom metagene script written by Yooree Chung and Zhengwei Zhu. E) Pathway analysis of the 1,986 genes that show a >10% decrease in PI following NELF knockdown.
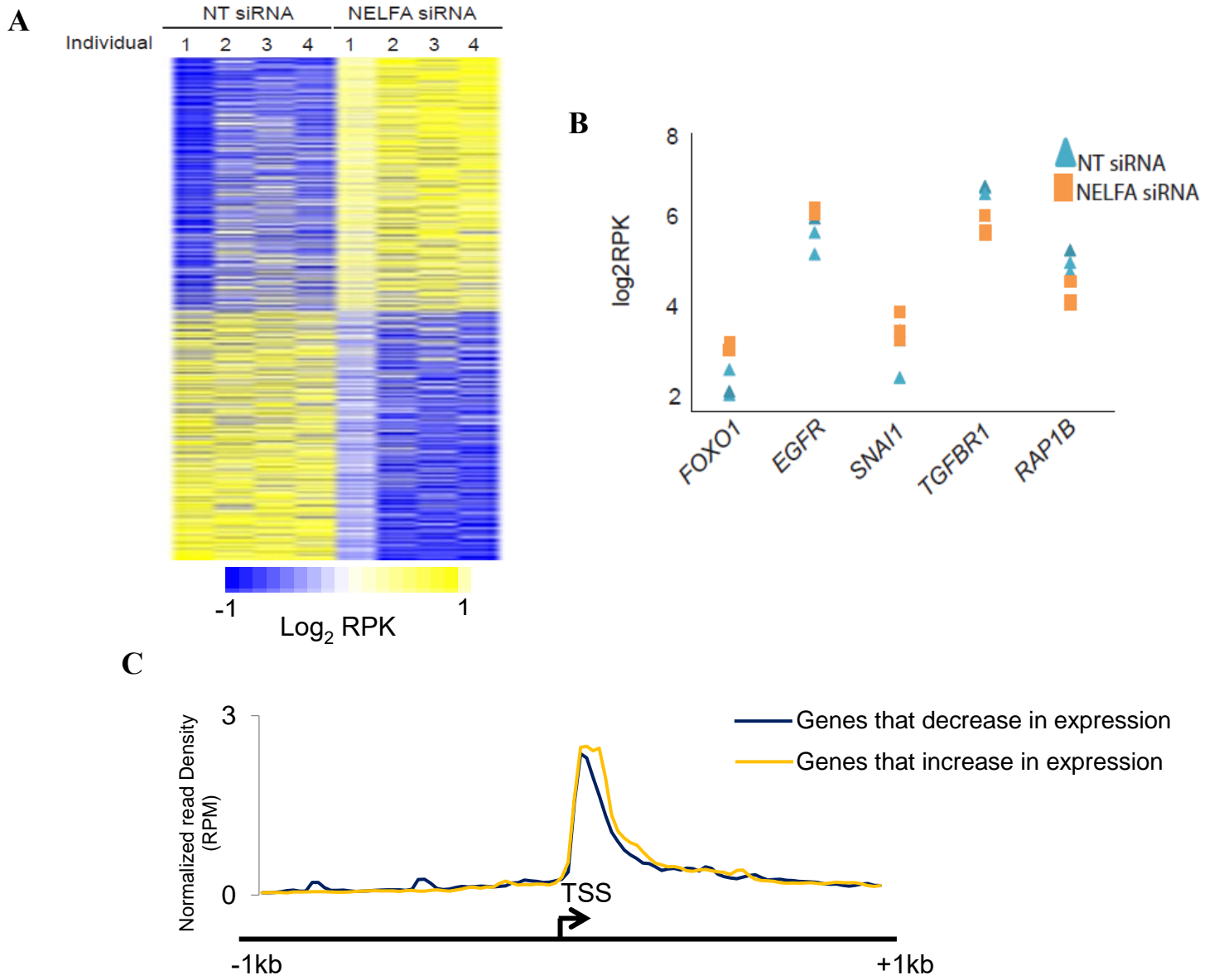
**Figure 2.5**



Figure 2.5 NELF-Mediated Pausing Influences Human Gene Expression. A) Heatmap of the 2,969 genes that change in expression >10% following NELF knockdown. Gene expression was measured by RNA-seq of four individual primary human dermal fibroblasts. B) Individual examples of genes that change in expression >10% following NELF knockdown. C) Metagene plot showing a decrease in the amount of transcriptionally active RNA Pol II at the promoter of genes whose expression decreases in expression following NELF knockdown.

# Figure 2.6

**A**



**B**
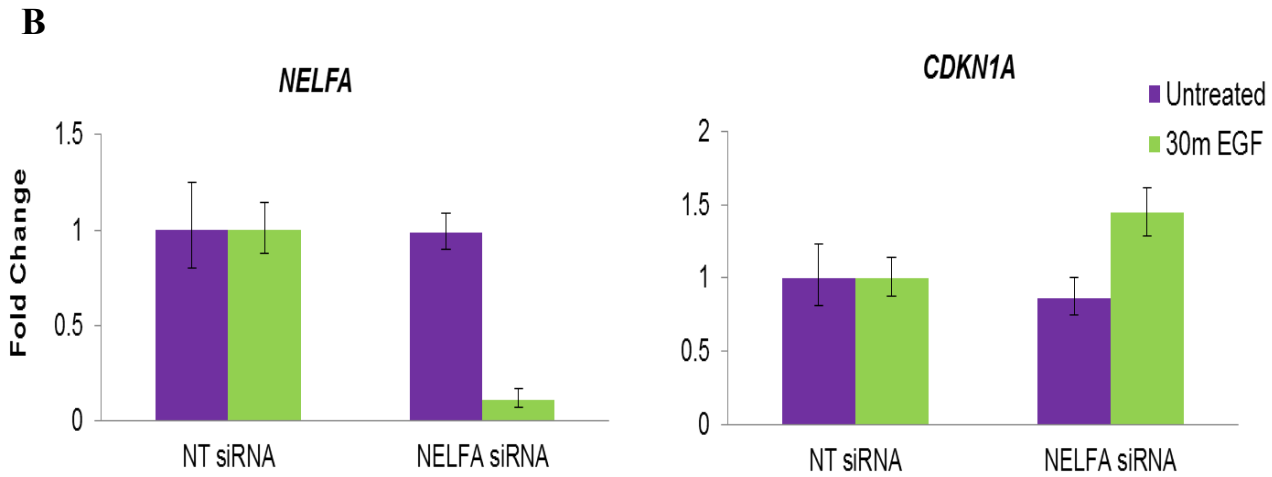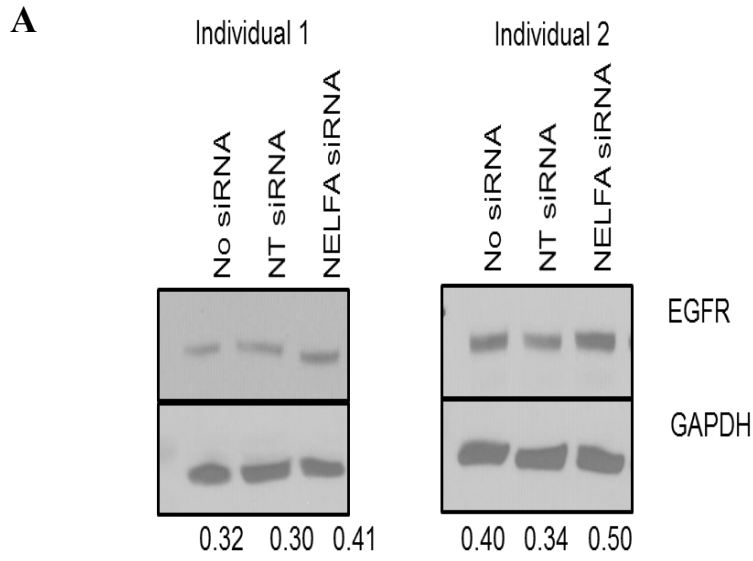


Figure 2.6 A) EGFR expression increases at the protein level following NELF knock-down. Quantitation performed in ImageJ. Intensity values are normalized to GAPDH B) *CDKN1A*, a target of EGF signaling, increases in expression 1.5 fold following 30m of EGF treatment and NELF knock-down.

# Figure 2.7
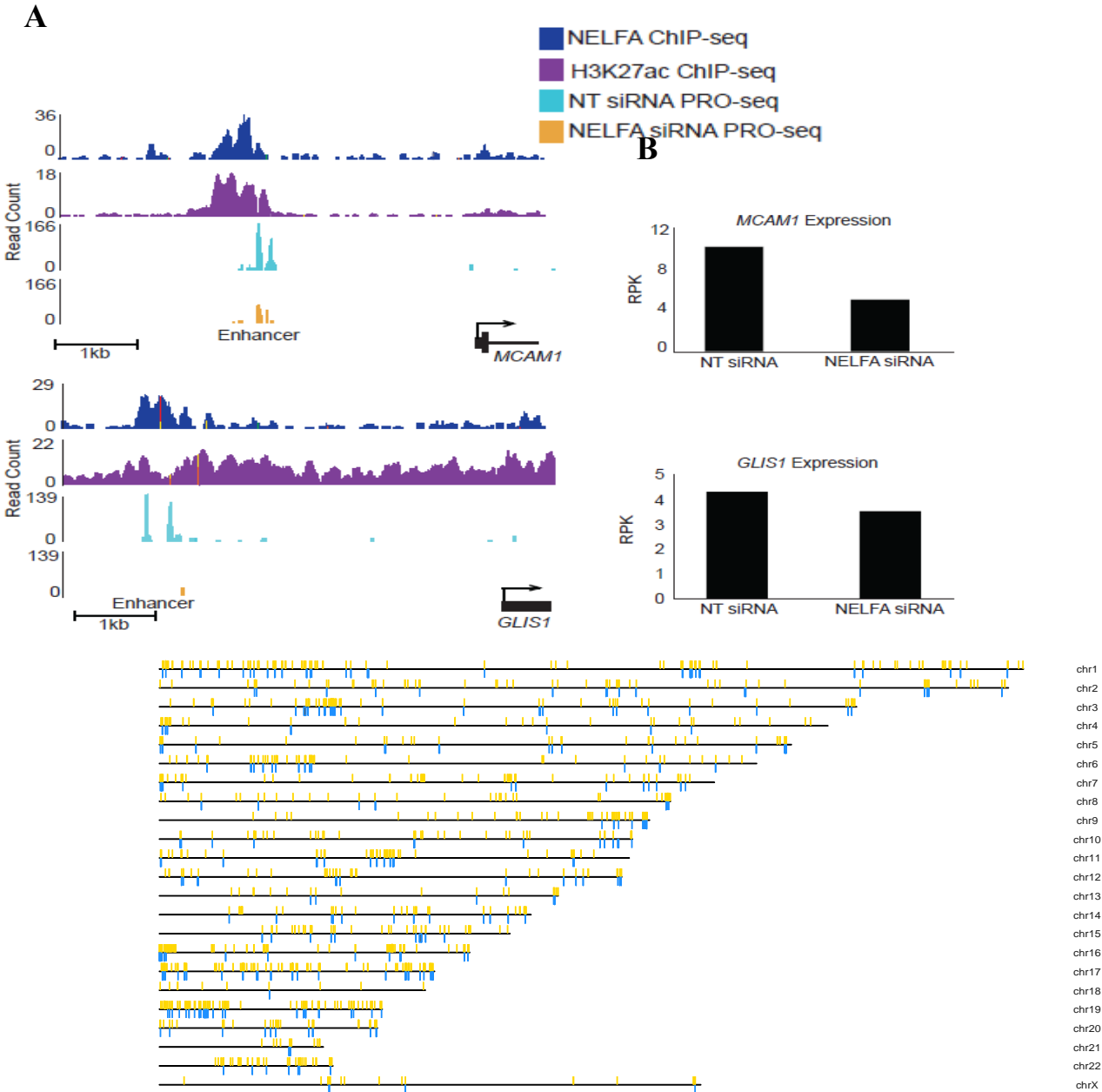


Figure 2.6 NELF Influences Gene Expression through eRNA. A) Reads from NELFA ChIP-seq, H3K27ac ChIP-seq and PRO-seq across two representative eRNA examples. B) Expression of *MCAM1* and *GLIS1* as measured by RNA-seq before and after NELF knockdown C) Global view of the locations of NELF-regulated eRNA and their putative target genes

**Figure S2.1**
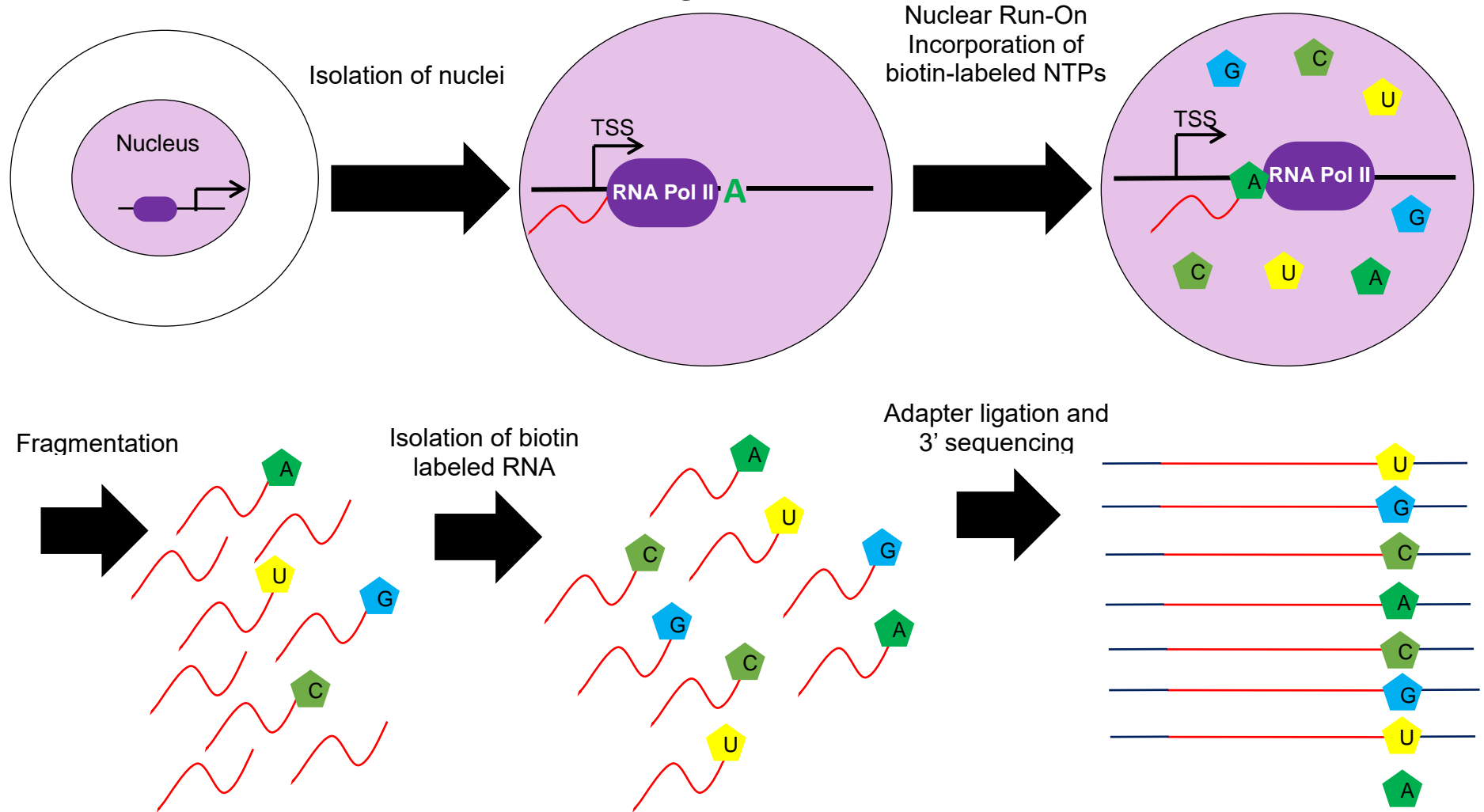
Figure S2.1 Schematic of Precision Run-On Sequencing (PRO-seq). Following nuclei isolation, biotin-labeled NTPs are incorporated by active RNA Pol II through a modified nuclear run-on assay. RNA is then harvested and fragmented. Following fragmentation, RNAs labeled with biotin are isolated thus creating a pool of strictly nascent RNA. Finally, adapters are ligated onto the nascent RNA and the nascent RNA is sequenced on the Illumina Hi-Seq platform.

**Figure S2.2**
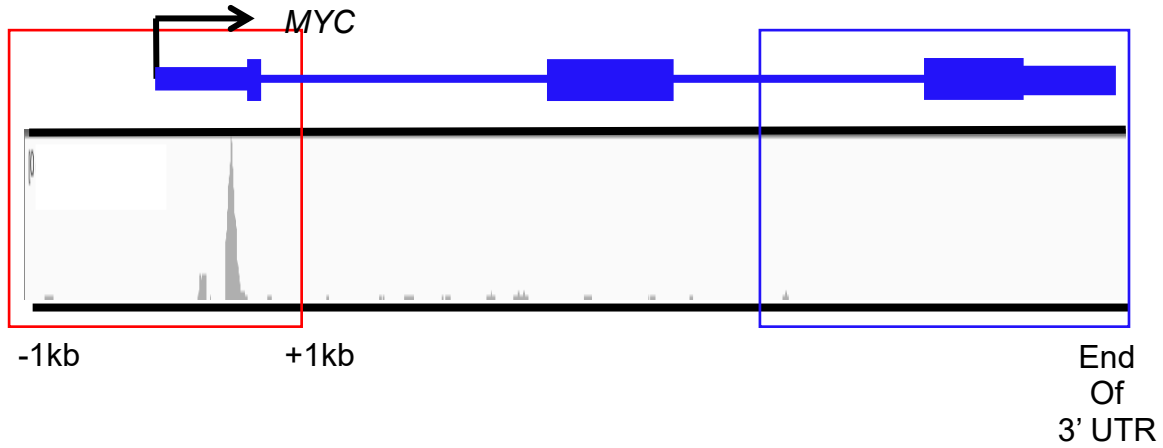


$$\text{Pausing Index (PI)} = \frac{\text{\# Normalized Reads in Promoter (-1kb, +1kb)}}{\text{\# Normalized Reads in Gene Body (+1kb, end of 3' UTR)}} = \frac{3130}{1435} = 4.8$$

Figure S2.2 Schematic of Pausing Index (PI) Calculation. The PI is a ratio of the number of reads in the promoter region as compared to the gene body. To calculate this, we first count the number of reads in the defined promoter region and the gene body. We then calculate an RPKM for each region to normalize for the size of the promoter and gene body. We then take the ratio of these two numbers to get a PI. A PI greater than 1 indicates there are more reads in the promoter as compared to the gene body.

**Figure S2.3**

Filter out genes that are less than 1kb in length and less than
1kb away from another gene

$\downarrow$

**16,333 genes**

$\downarrow$

Filter out genes that have less than 10 total reads

$\downarrow$

**11,362 genes**

$\downarrow$

Filter out genes PI<2

$\downarrow$

**8,622 genes**

$\downarrow$

Test for significant enrichment of reads in the promoter region as
compared to the gene body ($\chi^2$ test)

$\downarrow$

Filter out genes with $p_c < 0.05$

$\downarrow$

**7,264 paused genes**

Figure S2.3 Identification of Paused Genes by PRO-seq. Following alignment, we remove genes that are <1kb in length and <1kb from another gene. This allows us to avoid overlapping transcripts. We then remove genes with less than 10 total reads covering the gene. This allows us to examine only genes that are expressed. Next, we remove all genes with a PI<2 and test for a significant enrichment of reads in the promoter region as compared to the gene body by a $\chi^2$ test. This leaves us with 7,264 paused genes.

**Figure S2.4**
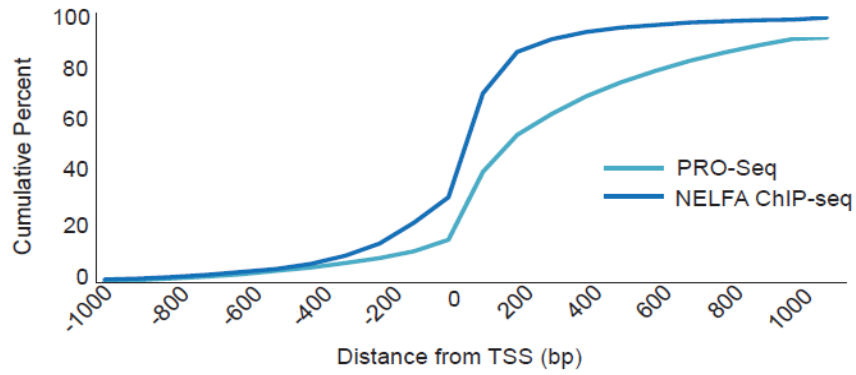


Figure S2.4 NELF Localizes Upstream of Paused RNA Pol II. A) Cumulative distribution plot depicting the locations of transcriptionally active RNA Pol II by PRO-seq and NELFA by ChIP-seq.

**Figure S2.5**
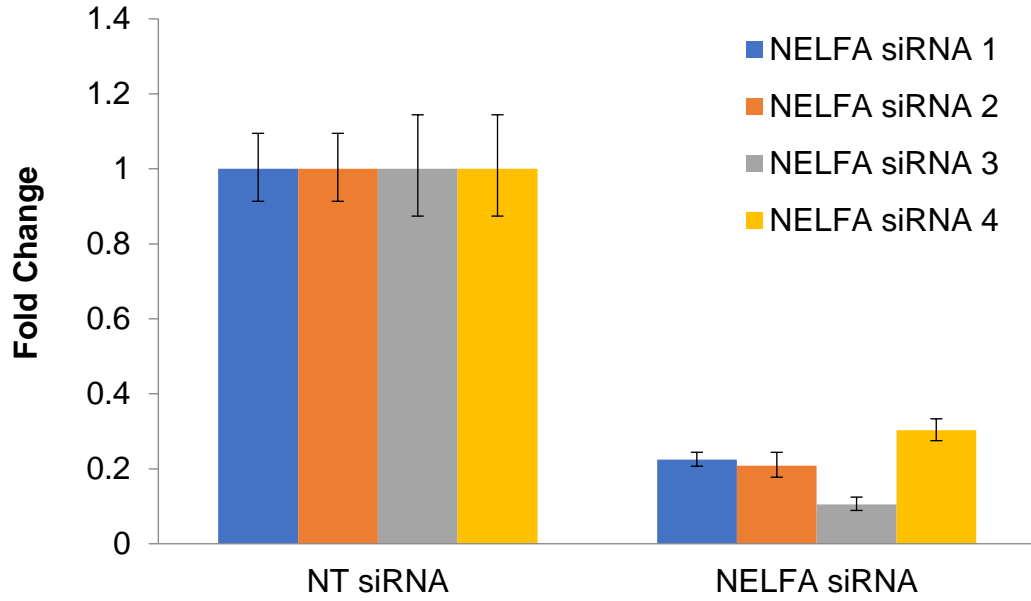
Figure S2.5 Individual NELFA siRNAs Knockdown NELF. A pool of 4 siRNAs targeting NELF was used for our NELF knockdown studies. We performed the knockdown under same conditions; however, using the individual siRNAs from the pool. We see at least a 70% knockdown of *NELFA* with each siRNA which is similar to what is seen with the pool.

**Figure S2.6**
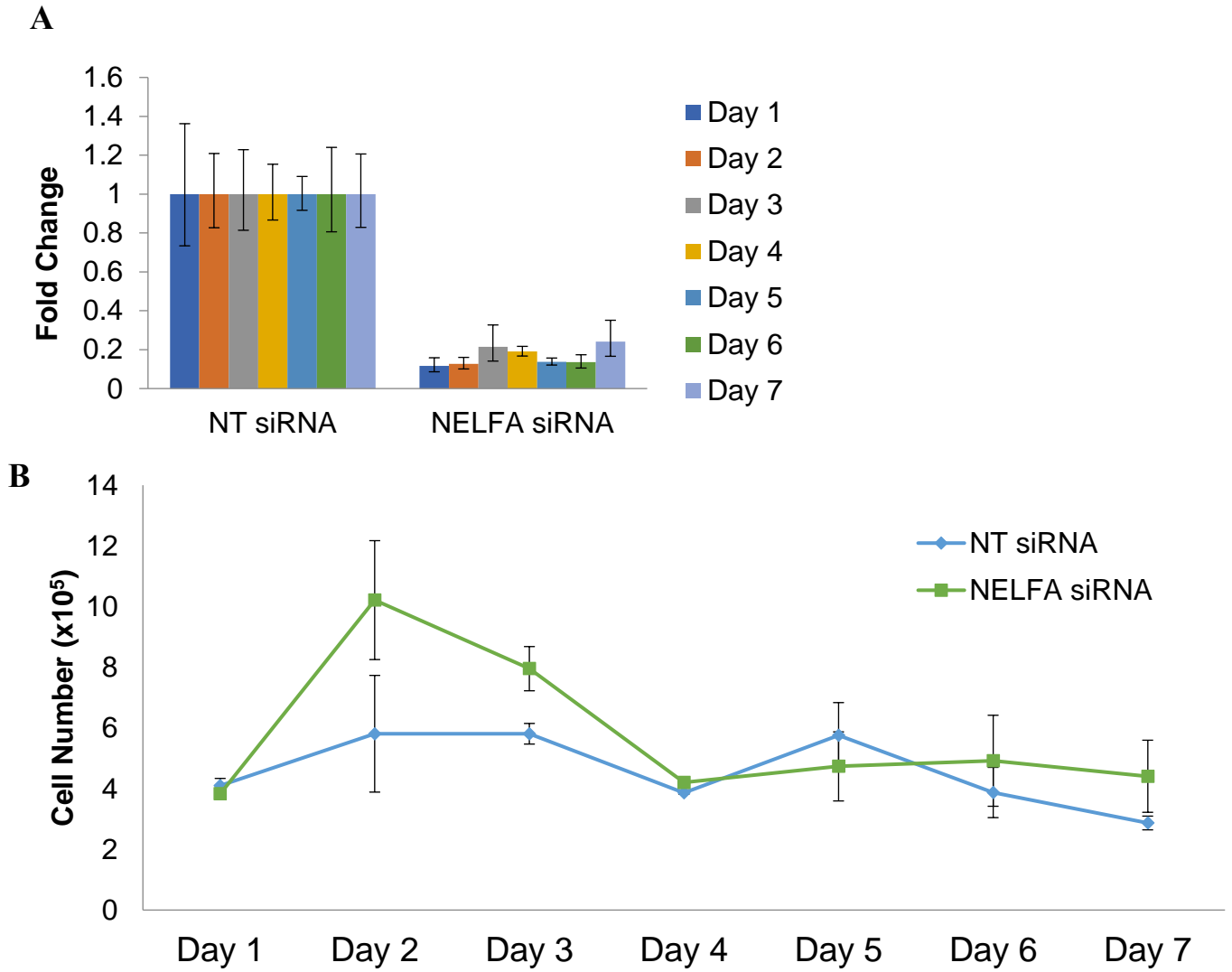


Figure S2.6 NELF Knockdown Does Not Affect Cell Growth. A) Knockdown of NELF persists over the 7-day time course used for the growth curve assay. B) Following knockdown of *NELFA*, we see no change in the growth of primary human dermal fibroblasts (FB338).

# Figure S2.7
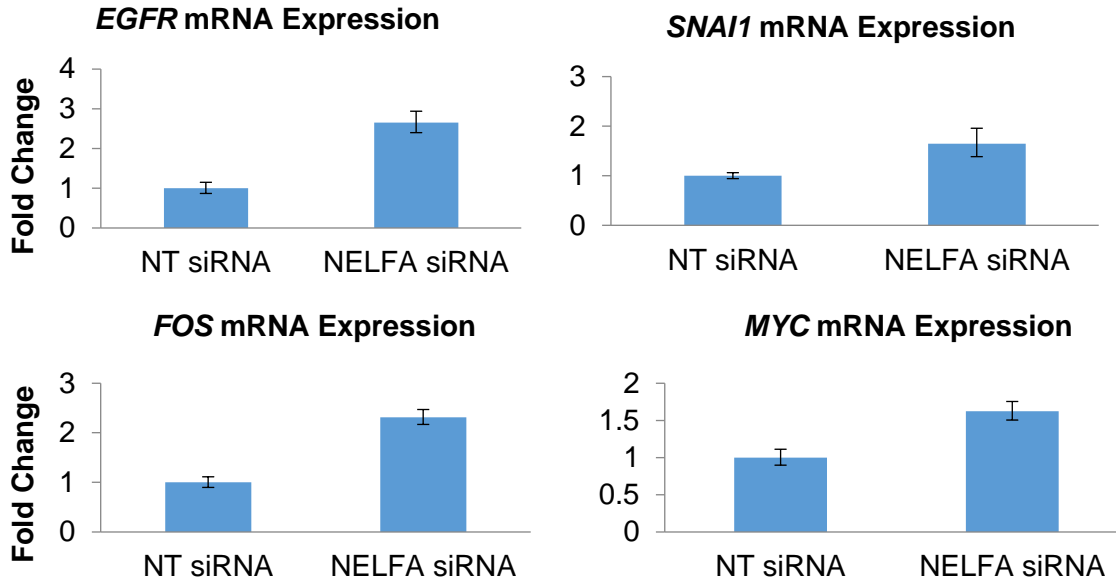


Figure S2.7 Confirmation of NELF-Mediated Expression Changes. qPCR was performed before and after NELF knockdown to validate results from RNA-seq. The expression levels of *EGFR*, *SNAI, FOS* and *MYC* increased following NELF knockdown, confirming findings from RNA-seq.

## Table 2.1 Primers Used for qRT-PCR

| Target Gene | FWD Primer | Reverse Primer |
|---|---|---|
| *NELFA* | cacgagagagcagatgttcg | atgaagcccaggatgagg |
| *MYC* | agcgactctgaggaggaaca | gctgtgaggaggtttgctgt |
| *FOS* | ccaagcggagacagacaac | caagggaagccacagacatc |
| *EGFR* | cgtgaagaagtgtccccgta | tcgcacttcttacacttgcg |
| *SNAI1* | gagtttaccttccagcagcc | ggtggggttgaggatctcc |

**Table 2.2 Primers Used for ChIP-qPCR**

| Target Gene | FWD Primer | Reverse Primer |
|---|---|---|
| *JUNB* Promoter | ctgggaccttgagagcgg | gcgcaaaagccctgtcag |
| *JUNB* Gene Body 1 | ccccgggacagtacttttac | ctcctgctcctcggtgac |
| *JUNB* Gene Body 2 | ccgtttacaccaacctcagc | ggaggtagctgatggtggtc |
| *JUNB* 3' End | cctctctgcaccgtactgt | cacacacaaacacaaacacgt |
| *FOS* Promoter | gttgagcccgtgacgtttac | agatgcggttggagtacgag |
| *FOS* Gene Body | ttgtgaagaccatgacaggagg | aactctagcgtactcttcctgg |
| *TGFB1* Promoter | ttgagactttccgttgccg | ccaagtcctgcctcctcg |
| *TGFB1* Gene Body | cagctcgaacactacctcct | gatctacttagggcccagg |

# Table 2.3 Sequence Reads

| Experiment | Sequenced Reads | Uniquely Aligned Reads | % of Uniquely Aligned Reads |
| --- | --- | --- | --- |
| PRO-seq | 182,098,253 | 18,295,503 | 24.57% |
| NT siRNA PRO-seq | 176,315,463 | 29,702063 | 34.43% |
| NELFA siRNA PRO-seq | 139,073,782 | 14,386,296 | 20.71% |
| NELFA ChIP-seq | 56,434,468 | 45,252,414 | 87.29% |
| RNA Pol II pSer5 ChIP-seq | 67,044,907 | 55,075,371 | 88.88% |
| SPT5 ChIP-seq | 48,988,189 | 39,516,280 | 87.92% |
| H3K27ac ChIP-seq | 16,587,050 | 14,578,638 | 95.41% |
| FB3356 NT siRNA RNA-seq | 20,848,425 | 16,744,849 | 71.72% |
| FB3356 NELFA siRNA RNA-seq | 23,760,451 | 17,354,914 | 78.36% |
| FB3383 NT siRNA RNA-seq | 8,100,232 | 5,829,497 | 77.30% |
| FB3383 NELFA siRNA RNA-seq | 8,058,689 | 6,160,127 | 81.92% |
| FB3385 NT siRNA RNA-seq | 7,073,343 | 5,410,317 | 82.13% |
| FB3385 NELFA siRNA RNA-seq | 6,232,097 | 4,734,295 | 81.25% |
| FB3389 NT siRNA RNA-seq | 4,730,306 | 4,949,487 | 78.84% |
| FB3389 NELFA siRNA RNA-seq | 5,385,887 | 4,020,459 | 80.21% |

# Chapter 3

## Preliminary Studies into RNA Polymerase II Pausing in HT29 Cells

Dr. Alan Bruzel provided technical assistance with the PRO-Seq technique used to generate the genome-wide data in Figure 3.1.

Dr. Hojoong Kwak wrote and ran the scripts for the pausing index calculations used in Figure 3.1.

Zhengwei Zhu performed the bioinformatic analysis for the genome-wide data generated from the PRO-Seq and RNA-Seq experiments in Figures 3.1 and 3.4

Jennifer Fox ran the Illumina HiSeq 2500 instrument for the PRO-Seq and RNA-Seq experiments in Figures 3.1 and 3.4. Jennifer Fox also provided the protocol and technical assistance with the IF experiments in Figures 3.7.

# Abstract

Dysregulation of transcription can lead to disease development. As such, there are many checkpoints in transcription. One of these is RNA Polymerase II (RNA Pol II) pausing. Here, we investigate the role of RNA Pol II pausing in regulating gene expression in HT29 colorectal cancer cells as well as its role in regulating R-loops. We identified over 3,900 paused genes in our HT29 cancer cells by Precision Run-On Sequencing (PRO-Seq). These genes are involved in pathways such as EGFR signaling as well as gene expression, membrane trafficking and transcription regulation by TP53. Negative elongation factor (NELF) has been shown to regulate RNA Pol II pausing at promoters. Because we were interested in understanding how pausing influences gene expression, we chose to utilize NELF to manipulate pausing in our HT29 cells. We found that over 2,900 genes change in expression following loss of NELF. These genes show enrichment for WNT signaling, cellular senescence, cellular response to stress and epigenetic regulation of gene expression. We found that most of the genes we identified show an increase in expression upon loss of NELF. We also show that these gene expression changes may also lead to an increase in cellular proliferation. Upon showing that pausing is able to regulate gene expression, we investigated the ability of pausing to regulate other transcriptional processes. We found that loss of pausing influences the formation of R-loops. Upon loss of pausing, we see an increase in R-loop formation in primary human newborn dermal fibroblasts. The findings presented below are preliminary and provide limited insight into the role of pausing in HT29 cells as well as the role of pausing in R-loop formation.  Replication of these findings is key to

drawing any formal conclusions from the data; however, we provide some preliminary

observations and hypotheses based on these single experiment findings

.

# Introduction

Colorectal cancer is the second leading cause of cancer deaths in the United States (Kuipers et al., 2015). There has also been extensive work examining the genetic components of colorectal cancer development and progression; however, many cases of colorectal cancer cannot be explained by these inherited variants. Though some work has been done to understand the genetic underpinnings of non-familial colorectal cancer, there still remains a lack of targeted therapies for both forms of this disease. Because of this, we wanted to better understand the mechanisms regulating the transcriptional program of these cells in the hopes of gaining potential new targets for therapy that can be applied to both familial and non-familial forms of the disease.

RNA Polymerase II (RNA Pol II) pauses approximately 20-60 nucleotides downstream of transcription start sites (TSS) (Gilmour and Lis, 1986). This step is thought to serve as the rate limiting step between initiation and active elongation (Min et al., 2011). This step has also been shown to influence the expression of genes in *Drosophila* and cancer models. Specifically, RNA Pol II pausing has been shown to influence the growth of breast cancer cells by upregulating genes that promote cell cycle progression (Sun and Li, 2010, Sun et al., 2011). Pausing also regulates the response to estrogen in breast cancer cells. Paused RNA Pol II serves to poise estrogen-responsive genes in order to create a coordinated response upon stimulation (Aiyar et al 2004, Kinnis et al., 2009). One of the key regulators of both of these processes is negative elongation factor (NELF). NELF is a 4-subunit complex comprised of NELFA, NELFB, NELFCD and NELFE (Wu et al., 2005, Yamaguchi et al., 1999). In both studies, the groups

focused on NELFB known also as co-factor of BRCA1 (COBRA1) because low expression of NELFB had been shown to correlate with poor prognosis in breast cancer patients (Sun et al., 2008). They showed that loss of NELFB resulted in a loss of the entire NELF complex, which subsequently led to a decrease in pausing and the resultant phenotypes (Sun et al., 2008).

Outside of breast cancer, little work has been done to investigate the role of pausing in other cancers. Only one other study has shown alterations in NELF in a tumor. NELFE has shown to be overexpressed in hepatocellular carcinoma. This overexpression has been shown to correlate with more advanced disease and poor prognosis; however, it is unclear how overexpression of NELFE relates to these phenotypes (Iida et al., 2012).

Pausing has been shown to be a key contributor to not only transcription regulation but also gene expression in model organisms as well as normal cell types. However, there is little understanding of how pausing relates to disease development. First, we wanted to investigate how pausing contributed to cancer development and progression. We chose to focus on colorectal cancer since it is a common cancer that lacks targeted therapies. By better understanding how transcription and specifically pausing is regulated in this disease, we may be able to define new targets for therapy that have not yet been investigated. Second, we wanted to investigate how pausing influences R-loop formation. Because pausing is able to limit the amount of RNA Pol II entering active elongation, we were curious to see if we could alter the level of R-loops in normal cells. In the following sections, we describe preliminary findings based off of single experiments that require additional replicates as well as in-depth follow-up studies in order to draw any firm conclusions from the data presented.

# Results

## Pausing in HT29 Colorectal Cancer Cells

## Rationale

Our initial studies into the transcriptional program of colorectal cancer began with a study of RNA Pol II pausing. We wanted to generate a global profile of transcriptionally active RNA Pol II in human cells. To do this, we chose to use Precision Run-On Sequencing (PRO-Seq) (Kwak et al. 2013). We chose this technique because of its unique ability to map the location of the polymerase with near single-nucleotide resolution as well as accurately predict paused genes. Though this technique provides us with a great deal of information, it requires a large number of cells. In order to use this technique successfully, we needed to begin our studies in the HT29 cell line since it was easily cultured and divided rapidly. These cells are also a well-established colon cancer mode and show no alterations in any of the NELF subunits (Cerami et al., 2012, Gao et al., 2013). The findings reported below are preliminary findings based off of a single experiment in HT29 cells.

## Results

We began by generating a global profile of paused genes in HT29 cells. We utilized PRO-Seq to generate this profile (Kwak et al., 2013) In order to identify paused genes, we calculated a pausing index (PI) for each gene. Pausing indexes were calculated using a custom script written by Dr. Hojoong Kwak. The general description of how pausing indexes are calculated is found in Chapter 2. We defined a paused gene as having a PI >2. We chose this cutoff based on previous work performed by the Lis Lab (Core et al., 2008). We also found that

90

genes with a PI>2 showed an enrichment of PRO-Seq reads in the promoter region that was easily visualized in a genome browser. We did not test for significant enrichment of reads in the promoter region as compared to the gene body for these samples because of differences in the output of the custom script used to calculate the PI. By using our cutoff of PI>2 we identified 3,909 paused genes in HT29 cells. Figure 3.1A shows the pausing indexes of these genes. These genes show a maximum PI of 3,143, a median PI of 20 and an average PI of 57. We identified *FOS* (4.9), a gene previously described to have a promoter-proximally paused polymerase (Plet et al., 1998). We also identified genes not previously described as paused such as *TGFBR2* (5.1), *EGFR* (9.7), *MET* (37), *MAP2K5* (426) and *POLG2* (467) (Figure 3.1B). The 3,909 paused genes we identified were enriched for genes involved in gene expression, EGFR1 signaling, membrane trafficking and transcription regulation by TP53 (Figure 3.1C).

We wanted to confirm the presence of RNA Pol II at the 5' end of some of the paused genes we identified. To do this we performed chromatin immunoprecipitation (ChIP) for the serine 5 phosphorylated form of RNA Pol II (RNA Pol II pSer5) followed by qPCR for a few of our genes of interest. We chose the serine 5 phosphorylated form of RNA Pol II since it represents the initiated form of RNA Pol II and as such should be found specifically at the 5' end of genes (K ref). We found that RNA Pol II ser5P accumulates specifically at the 5' end of the paused genes *JUNB*, *FOS* and *CDH1* (Figure 3.2). This increases our confidence in the paused genes that we identified by PRO-Seq.

RNA Pol II pausing is mediated by two factors, NELF and DSIF (Yamaguchi et al., 1999, Yamaguchi et al., 2002, Wu et al., 2003, Wu et al., 2005). Because we are interested in understanding the function of pausing in regulating transcription in these cells, we wanted to identify a way to manipulate pausing. We chose to focus on NELF in order to alter pausing since

the only described function of NELF is to mediate pausing (Narita et al., 2003). In order to confirm that NELF is at the promoters of our paused genes, we performed ChIP with an antibody to NELFA. Following ChIP for NELFA, we performed qPCR to look for enrichment of NELFA at the promoters of the paused genes *JUNB*, *FOS* and *CDH1*. We found that NELFA is specifically enriched at the promoter region of these paused genes (Figure 3.2). This again increased our confidence in the paused genes we identified by PRO-Seq as well as provided us with a target that we could use to manipulate pausing in HT29 cells.

## NELF-mediates Gene Expression in HT29 Colorectal Cancer Cells

### Rationale

Upon identifying pausing in our colorectal cancer model, we wanted to gain an understanding of the biological function of pausing in these cells. We chose to study gene expression as a readout of biological function of pausing. Though pausing serves as only one of many steps that can account for gene expression level, we believe that there will be a subset of genes that rely on pausing as a key step in regulating their expression. We are particularly interested to see if these genes are related to cancer development or progression. If the genes we identify do show a link to processes involved in disease maintenance or progression, pausing may serve as an alternate mechanism that could be targeted therapeutically for colorectal cancer. The findings below are preliminary and represent the data from only a single RNA-seq experiment. These findings require confirmation through replication studies as well as additional supporting experimental evidence.

### Results

RNA Pol II is held in a paused state just downstream of the TSS by the pausing factors

NELF and DSIF (Yamaguchi et al., 1999, Yamaguchi et al., 2002, Wu et al., 2003, Wu et al.,

2005). Loss of these factors, in particular NELF, results in a loss of pausing and subsequent

changes in gene expression in *Drosophila* (Gilchrest et al., 2010). Because we were interested in

studying the effects of pausing on gene expression, we chose to manipulate pausing through

NELF. NELF is a four-subunit complex comprised of NELFA, NELFB, NELFC/D and NELFE

(Yamaguchi et al., 1999). The expression of these subunits is interdependent, with the loss of one

subunit resulting in the loss of all others at the protein level (Yamaguchi et al., 2002, Sun et al.,

2008). To begin, we knocked down NELFA. We chose this subunit because it directly interacts

with RNA Pol II in the paused complex (Narita et al., 2003). We also identified NELFA

specifically at the promoters of paused genes (Figure 3.2). Upon knockdown of NELFA, we saw

an over 80% reduction in the amount of NELFA transcript (Figure 3.3A). We also measured the

expression of the other subunits at the transcript level following NELFA knock down (Figure

3.3A). We found that there was a slight increase in the expression of *NELFCD* following

NELFA knockdown; however, *NELFB* and *NELFE* expression remained at a similar level

(Figure 3.3A). We also examined the expression of NELFA, NELFB and NELFE at the protein

level following NELFA knockdown. We were unable to find an antibody that detected

NELFCD; therefore, we were unable to investigate whether the slight upregulation of NELFCD

following NELFA knockdown was also seen at the protein level. Upon NELFA knockdown, we

saw a reduction in the expression of each subunit (Figure 3.3B). This confirms that by knocking

down NELFA, we lose expression of the entire NELF complex at the protein level regardless of

the slight changes we see at the transcript level. We also confirmed that loss of one subunit

results in a loss of the others at the protein level by knocking down NELFE. Upon loss of

NELFE, we see an over 80% reduction in *NELFE* transcript as well as a loss of NELFE and NELFA protein (Figure 3.5A and B).

Loss of NELF leads to global changes in gene expression in *Drosophila* (Gilchrist et al, 2010). It is thought that these changes are due to a loss of RNA Pol II pausing at promoters (Gilchrist et al., 2010). Because we are interested in understanding how pausing functions to regulate gene expression in colorectal cancer, we chose to use NELF knockdown as a way to manipulate pausing and study changes in gene expression. We began by examining global gene expression by RNA-Seq before and after NELF knockdown. We found that 2,797 genes change in expression at least 1.5-fold following knockdown (Figure 3.4A). Because transcription is only a single step at which gene expression is regulated, we chose a 1.5-fold cutoff in order to identify those genes were pausing served as one of the most dominant regulators of expression. With this cutoff, we found that 2,316 genes increase in expression and 661 decrease in expression. We confirmed the expression changes of two paused genes *FOS* and *MYC* by qPCR (Figure 3.4B). We also found that upon loss of NELFE we see similar increases in the expression of *FOS* and *MYC* by qPCR (Figure 3.5C). This increased our confidence that the gene expression changes we identified were due to a loss of the NELF complex rather than a NELFA specific phenomenon. Finally, we found that these 2,797 genes were enriched for pathways such as cellular senescence, cellular response to stress, epigenetic regulation of gene expression and WNT signaling (Figure 3.4C).

NELF-mediated pausing has been shown to influence gene expression in two ways. The first is by holding the polymerase from entering into active elongation, thus resulting in lower expression of its target gene (Gilchrist et al., 2010). The second is by clearing the first nucleosome in order to facilitate subsequent rounds of transcription, thus leading to higher

expression of the target gene (Gilchrist et al., 2010). Our results show that NELF is primarily

functioning through the first mechanism in our HT29 cells.

## NELF influences growth of HT29 cells

## Rationale

Upon identifying over 2,000 genes that utilize pausing to regulate their expression, we

were interested to see if loss of NELF affected any biological processes downstream of these

genes. We identified cellular senescence as one of the key pathways to be regulated by NELF-

mediated pausing in HT29 cells. This led us to investigate if loss of NELF affected cellular

proliferation. We believe this experiment may also provide insight into how pausing plays a role

in generating cancer phenotypes.

## Results

The first phenotype we chose to examine was cell proliferation. We found that genes

involved in cellular senescence had their expression regulated by NELF-mediated pausing

(Figure 3.4C). We were interested to see if the gene expression changes induced by loss of

pausing resulted in alterations to the biological process as well. We chose to examine cell

proliferation over a 4-day time period. Though this is a short time period, we were interested in

understanding how an acute loss of pausing altered the biological phenotype. This time period

also encompasses when the HT29 cells were harvested for expression profiling by RNA-Seq,

thus providing us with a link between the expression changes we saw and the cellular phenotype.

We began by knocking down NELFA with our same pool of 4 siRNAs. We found that NELFA

expression decreases over 80% following knockdown and that this decreased level persists

across the 4-day time course (Figure 3.6A). We also confirmed that the knockdown resulted in a decreased expression of NELFA protein during the entire time course (Figure 3.6B).

We then measured cell proliferation by counting the number of HT29 cells at the same time each day for 4 days. We found that upon NELF knockdown, we see an increase in the rate of cell proliferation between days 2 and 3 as well as an increase in cell number at day 4 (Figure 3.6C). This time course indicates that the expression changes seen following loss of NELF, particularly those involved in cellular senescence pathways, may lead to increase cellular proliferation. However, the time course should be performed for a much longer time span as well as be replicated with more than one experiment to show a true change in cellular proliferation. It will also be important to measure the expression of the genes involved in cell cycle and cellular senescence to correlate their expression changes with the phenotype. These results provide a basis for future studies into a biological phenotype that may be regulated by pausing as well as a method by which pausing may be influencing cancer development and progression.

## NELF and R-loops in primary cells

### Rationale

Our finding that NELF affects the expression of over 2,000 genes as well as cell proliferation led us to investigate whether NELF also affected other transcriptional processes. We specifically chose to investigate the role of NELF and promoter-proximal pausing in mediating R-loop formation. R-loops are formed when the nascent RNA leaving the polymerase forms a hybrid with the template-strand of DNA. This causes the non-template strand to loop

out. Because R-loops are formed behind a transcriptionally active polymerase, we were interested to see if pausing altered the number or stability of R-loops. We began our studies of NELFs role in R-loop formation and maintenance in primary human newborn dermal fibroblasts. We chose this model since these cells are abundant in human connective tissue and they represent a normal human cell. Because we visualize R-loops through immunofluorescence (IF) with an antibody specific to R-loops (S.9.6) (Hu et al., 2006), we wanted a cell type that can be easily visualized using this technique. Our fibroblasts are flat with a clearly defined nucleus and cytoplasm. These traits make them an ideal candidate for IF. The findings reported below are preliminary findings and are based on a single experiment.

## Results

We began our studies by confirming that NELF knockdown was successful in primary human newborn dermal fibroblasts. Upon knockdown of NELFA with the same pool of 4 individual siRNAs, we see a greater than 80% reduction in NELFA transcript and almost complete loss of the NELFA protein (Figure 3.7A). These values correspond with what we measured in HT29 cells, so we believe this knockdown is also successful and results in loss of the entire NELF complex. Because we are using IF to measure R-loop abundance and location, we wanted to characterize NELF by IF as well. Using the same antibody as we used for western blotting, we found that NELF is localized specifically to the nucleus and is not found within nucleoli (Figure 3.7B). Next, we confirmed our NELF knockdown by IF. We found that there is a significant decrease in the intensity of NELFA staining in the nuclei of our fibroblasts (p-value=$8.6 \times 10^{-7}$) (Figure 3.7C). This confirms what we detect by western blotting and provides us with a model for examining the role of pausing in the maintenance of R-loops.

With our NELF knockdown model established in primary fibroblasts, we next wanted to examine how loss of pausing affected R-loops. We first established that knockdown in our fibroblasts was successful by measuring the amount of *NELFA* transcript before and after knockdown. We saw an 80% reduction in *NELFA* transcript which we have shown results in loss of the NELF complex in our fibroblasts previously (Figure 3.8A). Next, we wanted to identify R-loops in our fibroblasts. To do this we used an antibody specific to R-loops, S9.6. This monoclonal antibody was developed by the Leppla lab and is commercially available. It specifically recognizes RNA:DNA hybrids and does not cross react with single or double stranded RNA or DNA. This antibody has also been show to work well in the detection of R-loops via IF in our lab in experiments performed by Jennifer Fox.

Following NELF knockdown, we stained our fibroblasts with the S9.6 antibody to examine changes in R-loops. We found that S9.6 staining is most prominent in the nucleolus (Figure 3.8B). Because these regions are highly transcribed we were not surprised to find the most intense staining in this region. We did see staining throughout the nuclei of the fibroblasts as well (Figure 3.8B). We also found that the intensity of R-loop staining did not change following NELF knockdown (Figure 3.8C). However, we did see an increase in the number of S9.6 foci following NELF knockdown (Figure 3.8D). Foci were counted using ImageJ software in at least 50 nuclei.

## Preliminary Observations

Transcription is a tightly regulated process that can be harnessed by cancer to allow cells to divide rapidly, avoid death, invade and metastasize. Here, we examine one step in transcription regulation, RNA Pol II pausing, and its role in cancer development and progression.

98

RNA Pol II pausing plays a role in the transition from initiation to active elongation (Min et al., 2011). Our studies show that RNA Pol II pausing occurs at over 3,900 genes in HT29 cells. These genes are enriched for pathways that may play a role in the development of cancer such as EGFR1 signaling and transcription regulated by TP53. This implicates that pausing could be regulating the transcription of genes that promote cancer progression. These findings are based only on a single experiment. It will be important to replicate these findings through additional PRO-Seq experiments. It would also be beneficial to confirm pausing at these genes through RNA Pol II ChIP-Seq. Further work should also be conducted to confirm the enrichment and functional consequence of pausing on EGFR1 signaling and transcription regulated by TP53. These experiments may include qPCR validation as well as western blotting for downstream targets of both pathways.

The pausing factors NELF and DSIF regulate RNA Pol II pausing (Yamaguchi et al., 1999, Yamaguchi et al., 2002, Wu et al., 2003, Wu et al., 2005). We find that NELF is localized to the promoters of paused genes (Figure 3.2). Also, loss of NELF has been shown to lead to a decrease in pausing as well as subsequent changes in gene expression in *Drosophila* (Gilchrist et al., 2010). Because we were interested in how pausing was regulating transcription in our HT29 cells, we chose to utilize NELF as a way to manipulate pausing. We found that upon loss of NELF, we see over 2,900 genes that change in expression (>1.5 fold) with most of these genes showing increased expression (Figure 3.4). Gilchrist et al. as well as work in our lab (Figure 2.5), has shown that NELF is able to mediate gene expression through pausing in two ways. The first is by restraining the polymerase from entering active elongation, leading to low expression of the target gene at baseline. Many of these genes are involved in stimulus response and cell cycle (Gilchrist et al., 2008, Figure 2.5). The second is by creating a favorable environment at the

promoter for re-initiation of RNA Pol II (Gilchrist et al., 2010, Figure 2.5). In our HT29 cells, we find that most genes utilize the first mechanism. This implies that pausing may be serving as a checkpoint in regulating the expression of genes involved in key regulatory pathways such as cellular senescence and the cellular response to stress (Figure 3.4). We also see that upon loss of NELF there is an increase in cell proliferation; however, this is seen only over a short period of time (Figure 3.5). Though these initial observations provide the basis for further studies into the effects of pausing in cellular proliferation and cancer development, they are based only on a single experiment and therefore need to be replicated and confirmed.

We also see enrichment for genes involved in the WNT signaling pathway (Figure 3.4). Dysregulation of this pathway is common in colorectal cancer (Kuipers et al., 2015). HT29 cells have been shown to harbor mutations in *APC*, a regulator of the WNT pathway (https://www.atcc.org/~/media/PDFs/Culture%20Guides/Cell_Lines_by_Gene_Mutation.ashx). Our finding that genes within the WNT pathway increase in expression following loss of pausing, suggest that pausing serves as a regulatory checkpoint for components of this pathway. This may aid in lessening the transcriptional dysregulation caused by mutations in the WNT pathway in order to give the cell time to adjust to the new gene expression profile.

Following our preliminary characterization of pausing and its role in gene expression in in colorectal cancer cells, we wanted to investigate how pausing influences other components of transcription, specifically the role of pausing in facilitating R-loop formation. We found that upon loss of pausing, we see an increase in R-loop formation in primary human newborn dermal fibroblasts (Figure 3.8) This suggests that a loss of pausing affects the formation of R-loops. The mechanism by which this occurs remains to be elucidated through further experiments.

Our findings provide insight into the function of pausing in cancer. We hypothesize that pausing serves as a checkpoint to ensure that key cellular pathways maintain their function even in the context of tumor promoting mutations. In this context, pausing may be able to be targeted to restore correct functioning of tumor promoting pathways by decreasing the expression of the genes that drive cell growth, cell survival and invasion. Though these findings provide a novel and interesting line of questioning, they are based on a single experiment. Replication studies are need to confirm these results as well as functional studies to confirm the potential influence of pausing on cellular growth, survival and invasion.

## Materials and Methods

**Cell Culture: HT29, Newborn Primary Human Dermal Fibroblasts**

Cultured HT29 cells were obtained from ATCC and were cultured in RPMI 1640 with 10% Fetal Bovine Serum, 100 units/mL penicillin/streptomycin and 2mM L-gluatmine. Human primary newborn foreskin fibroblasts were used from 1 individual (FB3383). Fibroblasts were cultured at a density of 5 x $10^5$ cells/mL in DMEM with 10% Fetal Bovine Serum, 100 units/mL penicillin/streptomycin and 2mM L-glutamine. Each batch was cultured separately.

**siRNA Knockdown**

Transient transfections with siRNA were performed using RNAiMAX Lipofectamine (Thermo Fisher 137708030), following the manufacturer's instructions. Non-target control (E-HUMAN-XX-0005), NELFA (L-012156-00-0005) and NELFE (E-011761-00-0005) siRNAs were pools of 4 siRNAs and were obtained from GE Healthcare Dharmacon. RNA, protein or nuclei were harvested 96 hours post-transfection.

**RNA Isolation, reverse transcription and quantitative real-time PCR**

To measure expression of *NELFA*, *NELFB*, *NELFCD*, *NELFE*, *FOS*, *MYC* HT29 cells were transfected with non-target control or NELFA siRNAs (see above). RNA was harvested 96 hours post transfection with the Qiagen RNeasy Mini-Kit (74104). RNA (250ng) was used for cDNA synthesis using the TaqMan Reverse Transcription cDNA synthesis kit (Thermo Fisher Scientific N8080234) with oligodT. *GAPDH*, *NELFA*, *NELFB*, *NELFCD*, *NELFE*, *FOS*, *MYC* expression levels were measured by quantitative real-time PCR using gene specific primers and Power SYBR Green Master Mix (Thermo Fisher Scientific) with three technical replicates for each PCR. *GAPDH* was used as a housekeeping gene to normalize the amount of cDNA. Fold change of target gene mRNA levels was calculated using the $\Delta\Delta C_t$ method.

**Western Blotting**

Western blotting was performed as described previously (Gregory and Cheung, 2013). Briefly, protein lysates were collected in 1X RIPA lysis buffer (Cell Signaling 9806) following the manufacturer's protocol. Immunoblotting was performed with the following antibodies, NELFA (sc-32911), NELFB (ab48366), NELFE (Millipore ABE48). Membranes were developed using Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific 32106).

**Precision Run-On Sequencing (PRO-seq)**

PRO-seq libraries were prepared as described previously (Kwak et al., 2013). Briefly, approximately $1.5\times10^6$ nuclei from each condition were added to 2 X Nuclear Run-On (NRO) reaction mixture (10 mM Tris-HCl pH 8.0, 300 mM KCl, 1% Sarkosyl, 5 mM MgCl$_2$, 1 mM DTT, 0.03 mM each of biotin-11-A/C/G/UTP (Perkin-Elmer), 0.8 u/µl Rnase inhibitor) and incubated for 3 min at 37˚C. Nascent RNA was extracted by phenol (Trizol LS)/chloroform then was fragmented by base hydrolysis in 0.2 N NaOH on ice for 15 min. The reaction was neutralized by adding 1 X volume of 1 M Tris-HCl pH 6.8. Following fragmentation, nascent

RNA was purified using streptavidin beads and ligated with reverse 3' RNA adapter (5'-rCrCrUrUrGrGrCrArCrCrCrGrArGrArArUrUrCrCrA). Again, biotin-labeled products were enriched by streptavidin bead binding and extraction.  For 5' end repair, the RNA products were successively treated with 5' pyrophosphatase (NEB) and polynucleotide kinase (PNK, NEB).  5' repaired RNA was ligated to reverse 5' RNA adaptor (5'-rCrCrUrUrGrGrCrArCrCrCrGrArGrArArUrUrCrCrA) The products were further purified by the third round of streptavidin bead binding and extraction.  RNA was reverse transcribed using 25 pmol RT primer (5'AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA-3').  The product was amplified 22 cycles and products between 150-300bp (insert > 70 bp) were PAGE purified before being analyzed by Illumina HiSeq 2500 instrument. (Figure S2.1) Dr. Alan Bruzel provided technical assistance for the PRO-Seq experiments described in this chapter. Jennifer Fox ran the Illumina HiSeq 2500 Instrument.

**Chromatin Immunoprecipitation (ChIP)**

ChIP was performed as described previously (Watts et al., 2010). Briefly, cells were cross-linked with 1% formaldehyde for 15 min. Cross-linking was stopped with XM glycine for 5 min. Following cell lysis, chromatin was purified and sonicated to 1000-200bp fragments by sonication (30s on, 30s off). 10% of the supernatant was saved as input control and was processed at the cross-linking reversal step. The rest of the supernatant was pre-cleared with Protein A beads (Sigma: 11719408001) and anti-rabbit IgG (cat). Then, the supernatant was immunoprecipitated overnight with either 5ug anti-rabbit IgG, 5ug NELFA (sc-32911) or 5ug RNA Pol II-5P (ab5408). Following reverse cross-linking, DNA was precipitated using the Qiagen QIAquick PCR Purification Kit (28106). instrument. Jennifer Fox ran the Illumina HiSeq 2500 instrument. Validation of targets was performed by real-time PCR using Power SYBR. Dr.

Jason Watts provided technical assistance with the ChIP-Seq experiments described in this chapter.

**mRNA Sequencing**

RNA was isolated from HT29 cells 96 hours following siRNA treatment as described above. Libraries were prepared using the Illumina TruSeq RNA Sample Preparation Protocol. Samples were analyzed on the Illumina HiSeq 2500 Instrument. Jennifer Fox ran the Illumina HiSeq 2500 Instrument.

**Sequence Alignment**

For RNA-Seq and PRO-Seq, raw sequencing files were processed by trimming the adapter sequences from the ends of reads using fastx_clipper from FASTX-Toolkit (Hannon Lab). Sequences of low-quality at the ends of reads represented by a stretch of "#" in the quality score string in FASTQ file were removed along with PolyA (more than five consecutive As at the ends of reads. Reads that were >35nt after trimming were included for downstream analysis. Reads were aligned to human reference genome (hg18) using GSNAP (Wu and Nacu, 2010) (version 2013-10-28) with the following parameters: Mismatches < [(read length +2)/12/2]; Mapping score >2-; Soft-clipping on (-trim-mismatch-score = -3). Bam files were then generated. Reads with identical sequences were counted as one read in order to remove potential PCR duplicates. Table 2.3 shows sequenced and uniquely aligned reads in each sample. This analysis was performed by Zhengwei Zhu.

**PRO-Seq Analysis**

Custom scripts were used to calculate the number of reads in the promoter and gene body regions. This script was written and run by Hojoong Kwak. The promoter region is defined as +/- 1kb surrounding the annotated transcription start sites. The read counts from each sample were

normalized to the number of million mapped reads. Genes that are within 1kb of another gene or less than 1kb in length were removed. We then calculated a pausing index for the remaining genes (Core et al., 2008) (Figure S2.2).

**Cell Growth Assay**

HT29 cells were plated at a density of $3x10^5$ cells/mL. Transfection with NT or NELFA siRNA was performed as described above. Cells were washed 2x with PBS, trypsinized for 3 minutes and resuspended in complete RPMI 1640. 150uL of cell suspension was retained for cell counting with the Cellometer Auto T4 Automated Bright Field Cell Counter (Nexcelom). The reaming suspension was used to harvest RNA or protein (see above).
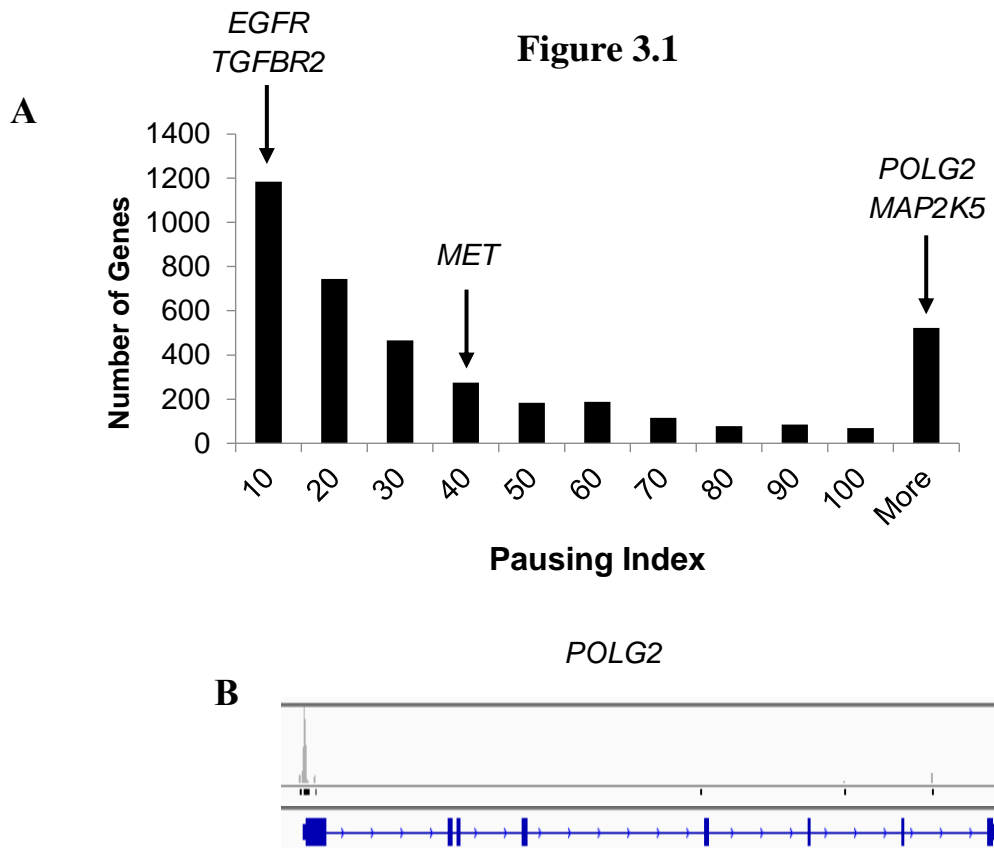
**Immunofluorescence (IF)**

Protocol and technical assistance for IF was provided by Jennifer Fox. Briefly, FB3383 were plated at $1.5x10^5$ cells/mL in a 12-well plate. Transfection of NT or NELFA siRNA was performed as described previously. 72 hours post transfection cells were plated onto glass coverslips in complete media. 96 hours post transfection cells were fixed with 4% formaldehyde for 15 minutes at RT. Coverslips were washed 3x with PBS. Cells were blocked in normal goat serum + PBS for 1h at RT. Coverslips were incubated with S9.6 antibody (1:1000) or NELFA antibody (1:50) overnight at 4$^{\circ}$C. Coverslips were washed 3x 5min with PBS and incubated with secondary antibody (S9.6 anti-mouse, NELFA anti-rabbit) for 2h at RT. Coverslips were washed with PBS 3x 2min at RT. Dapi stain (5x) was added for 5min at RT. Coverslips were washed with PBS 3x 5min at RT and ddH$_2$0 1x 5m at RT. Coverslips were mounted onto slides and left overnight at RT. Slides were imaged on Lecia DMI6000 microscope. Intensity and foci counting was performed in Image J.

# References

Aiyar, S.E., Sun, J., Blair, A.L., Moskaluk, C.A., Lu, Y., Ye, Q., Yamaguchi, Y., Mukherjee, A., Ren, D., Handa, H., Li., R. Attenuation of estrogen receptor a-mediated transcription through estrogen-stimulated recruitment of a negative elongation factor. Genes Dev. 2004 Sep 1; 18(17):2134-46.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Huer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., Schultz, N. the cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012 May;2(5):401-4.

Core, L.J., Waterfall, J.J., Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008 Dec 19;322(5909):1854-8.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., Schultz, N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013 Apr 2;6(269):pl1.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K.B., Collins, J.B., Li, L., Gilmour, D.S., Adelman, K. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev. 2008 Jul 15;22(14)1921-33.

Gilmour, D.S., Lis, J.T. RNA polymerase II interacts with the promoter region of the noninduced *hsp70* gene in *Drosophila melanogaster* cells. Mol Cell Bio. 1986 Nov;6(11):3984-9.

Hu, Z., Zhang, A., Storz, G., Gottesman, S., Leppla, S.H. An antibody-based microarray assay for small RNA detection. Nucleic Acids Res. 2006 Apr 13;34(7):e52.

Iida, M., Iizuka, N., Tsunedomi, R., Tsutsui, M., Yoshida, S., Maeda, Y., Tokuhisa, Y., Sakamoto, K., Yoshimura, K., Tamesa, T., Oka, M. Overexpression of the RD RNA binding protein in hepatitis C virus-related hepatocellular carcinoma. Oncol Rep. 2012 Aug;28(2):728-34.

Kinnis, M., Isaacs, G.D., Core, L.J., Hah, N., Kraus, W.L. Postrecruitment regulation of RNA polymerase II directs rapid signaling responses at the promoters of estrogen target genes. Mol Cell Biol. 2009 Mar;29(5):1123-33.

Kuipers, E.J., Grady, W.M., Lieberman, D., Seufferlein, T., Sung, J.J., Boelens, P.G., van de Velde, C.J., Watanbe, T. Colorectal cancer. Nat Rev Dis Primers. 2015 Nov 5;1:15065.

Kwak, H., Fuda, N.J, Core, L.J., Lis, J.T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science. 2013 Feb 22;339(6122):950-3.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., Lis, J.T. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev. 2011 Apr 1;25(7):742-54.

Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D., Hasegawa, J., Omori, M., Inukai, N., Endoh, M., Yamada, T., Hiroshi, H. Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. Mol Cell Biol. 2003 Mar;23(6):1863-73.

Sun, J., Watkins, G., Blair, A.L., Moskaluk, C., Ghosh, S., Jiang, W.G., Li, R. Deregulation of cofactor of BRCA1 expression in breast cancer cells. J Cell Biochem. 2008 Apr 15;103(6):1798-807.

Sun, J., Li, R. Human negative elongation factor activates transcription and regulates alternative transcription initiation. J Biol Chem. 2010 Feb 26;285(9):6443-52.

Sun. J., Pan, H., Lei, C., Yuan, B., Nair, S.J., April, C., Parmeswaran, B., Klotzle, B., Fan, J., Ruan, J., Li, R. Genetic and genomic analyses of RNA polymerase II-pausing factor in regulation of mammalian transcription and cell growth. J Biol Chem. 2011 Oct 21;286(42):36248-57.

Wu, C., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., Gilmour, D.S. NELF and DSIF cause promoter proximal pausing on the *hsp70* promoter in *Drosophila*. Genes Dev. 2003 Jun 1;17(11):1402-14.

Wu, C., Lee, C., Fan, R., Smith, M.J., Yamaguchi, Y., Handa, H., Gilmour, D.S. Molecular characterization of *Drosophila* NELF. Nucleic Acids Res. 2005 Mar 1;33(4):1269-79.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., Handa, H. NELF, a mulitisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell. 1999 Apr 2;97(1):41-51.

Yamguchi, Y., Inukai, N., Narita, T., Wada, T., Handa, H. Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol Cell Biol. 2002 May;22(9):2918-27.

**Figure 3.1**

**A**



**B** POLG2



**C**

| Pathway | Examples | p-value |
|---|---|---|
| Gene Expression | *SLBP; SMAD2; MAML3; EZH2* | $6.7 \times 10^{-26}$ |
| EGFR1 Signaling | *EGFR; FOXO1; MAPK9; RELA* | $7.2 \times 10^{-23}$ |
| Membrane Trafficking | *BET1; VTA1; KDELR3; VAMP3* | $2.5 \times 10^{-13}$ |
| Transcription Regulation by TP53 | *MTOR, FOS, HDAC1, CHEK1, PERP* | $2.9 \times 10^{-9}$ |

Figure 3.1 Promoter-Proximal Pausing Occurs Genome-Wide in HT29 Cells. A) Histogram depicting the pausing indexes (PI) of the 3,909 paused genes (PI>2) in HT29 cells. B) Example of the paused gene *POLG2* (PI=467). C) Pathway analysis of the 3,909 paused genes
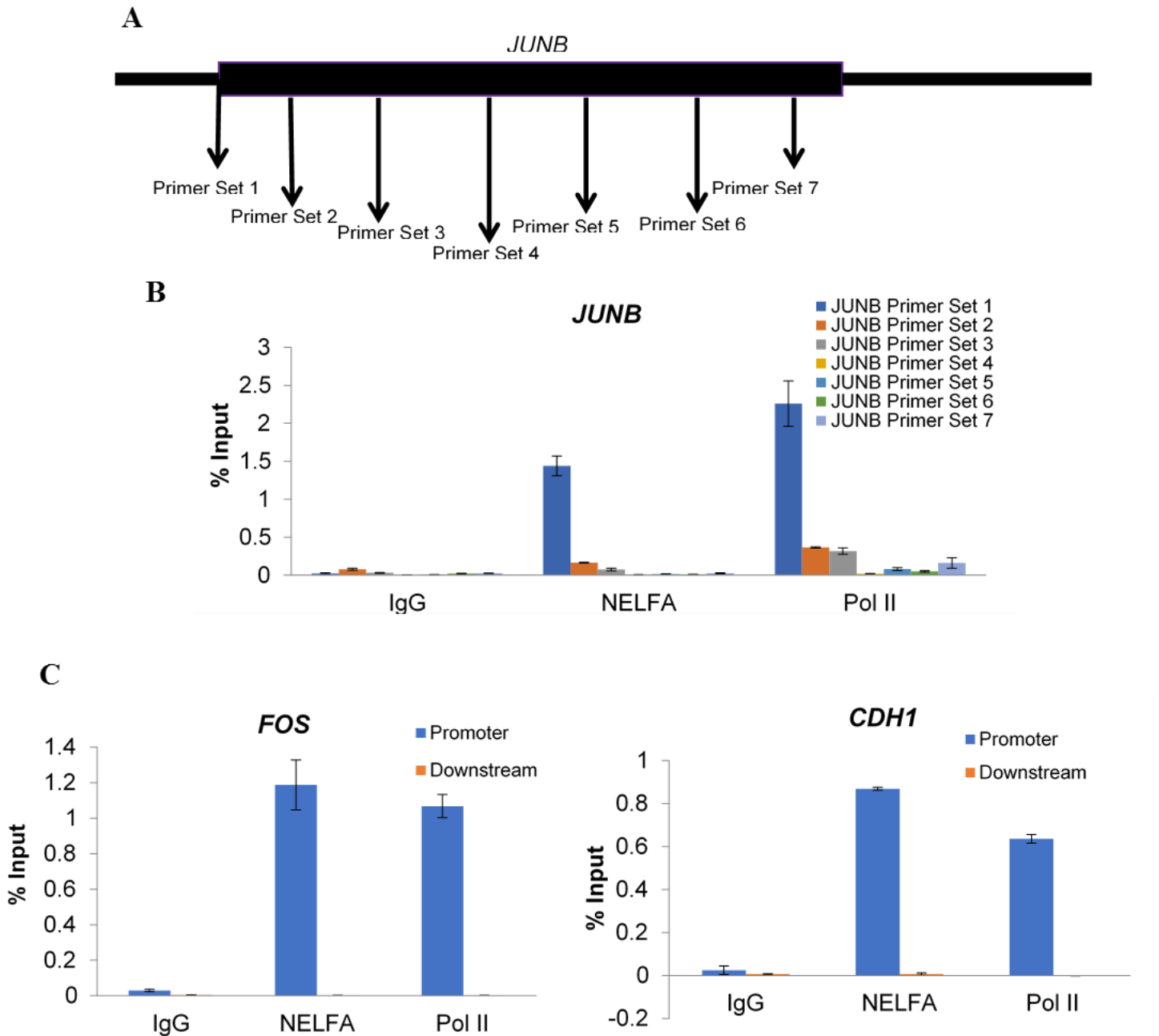
**Figure 3.2**



Figure 3.2 RNA Pol II and NELFA Localize to the Promoters of Paused Genes in HT29 Cells. A) Primer location across *JUNB*. B) ChIP-qPCR showing the enrichment of RNA Pol II and NELFA at the promoter region of the paused gene *JUNB* in HT29 cells. C) ChIP-qPCR showing the enrichment of RNA Pol II and NELFA at the promoters of the paused genes *FOS* and *CDH1*.

**Figure 3.3**

A

**NELF Subunit mRNA Expression**
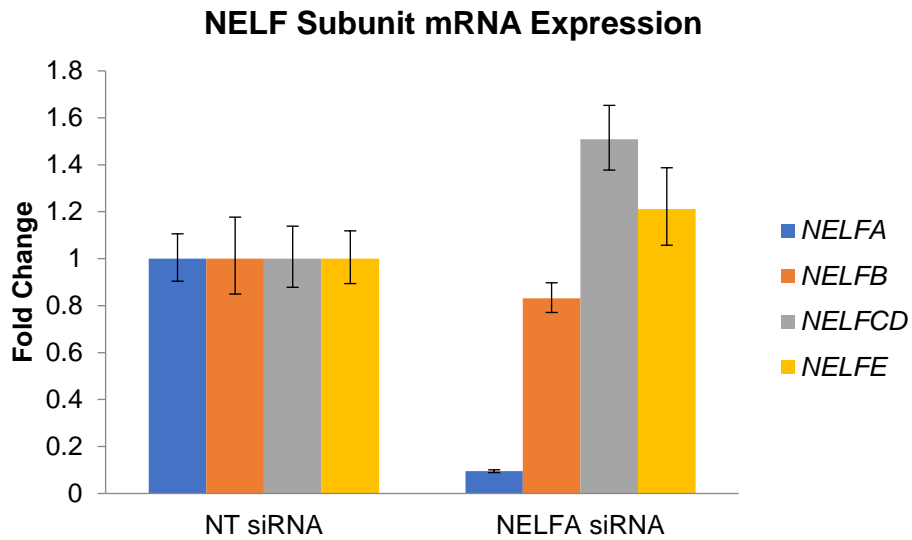


B



Figure 3.3 NELF Knockdown in HT29 Cells. A) NELFA knockdown results in a greater than 80% reduction in *NELFA* transcript; however, the expression of the other subunits at the transcript level remain unchanged. B) Loss of NELFA results in loss of NELFB and NELFE at the protein level.

# Figure 3.4

**A**



**B**



FOS

p-value: $1\times10^{-4}$

MYC

p-value: $1\times10^{-3}$

**C**
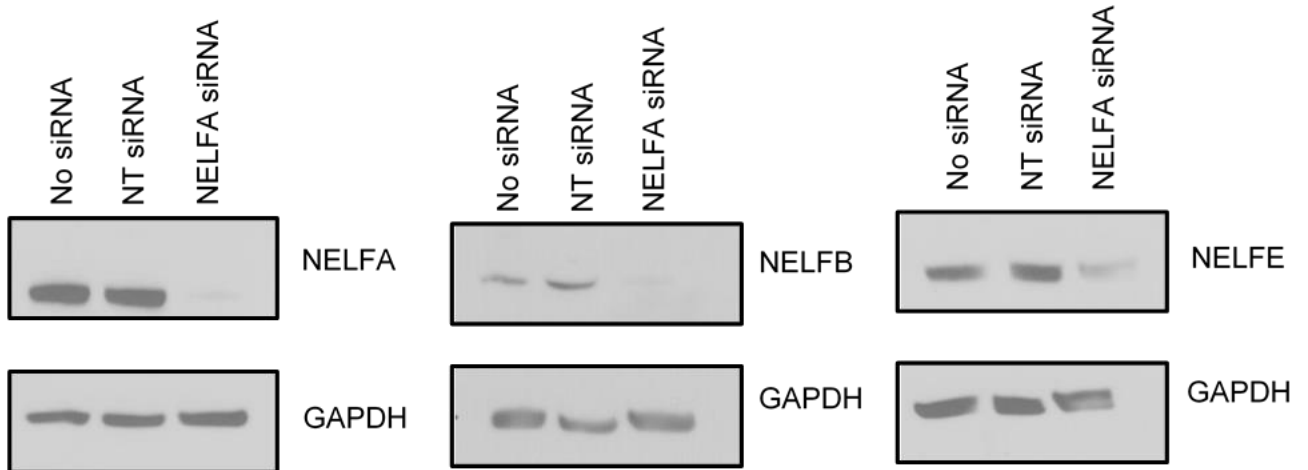
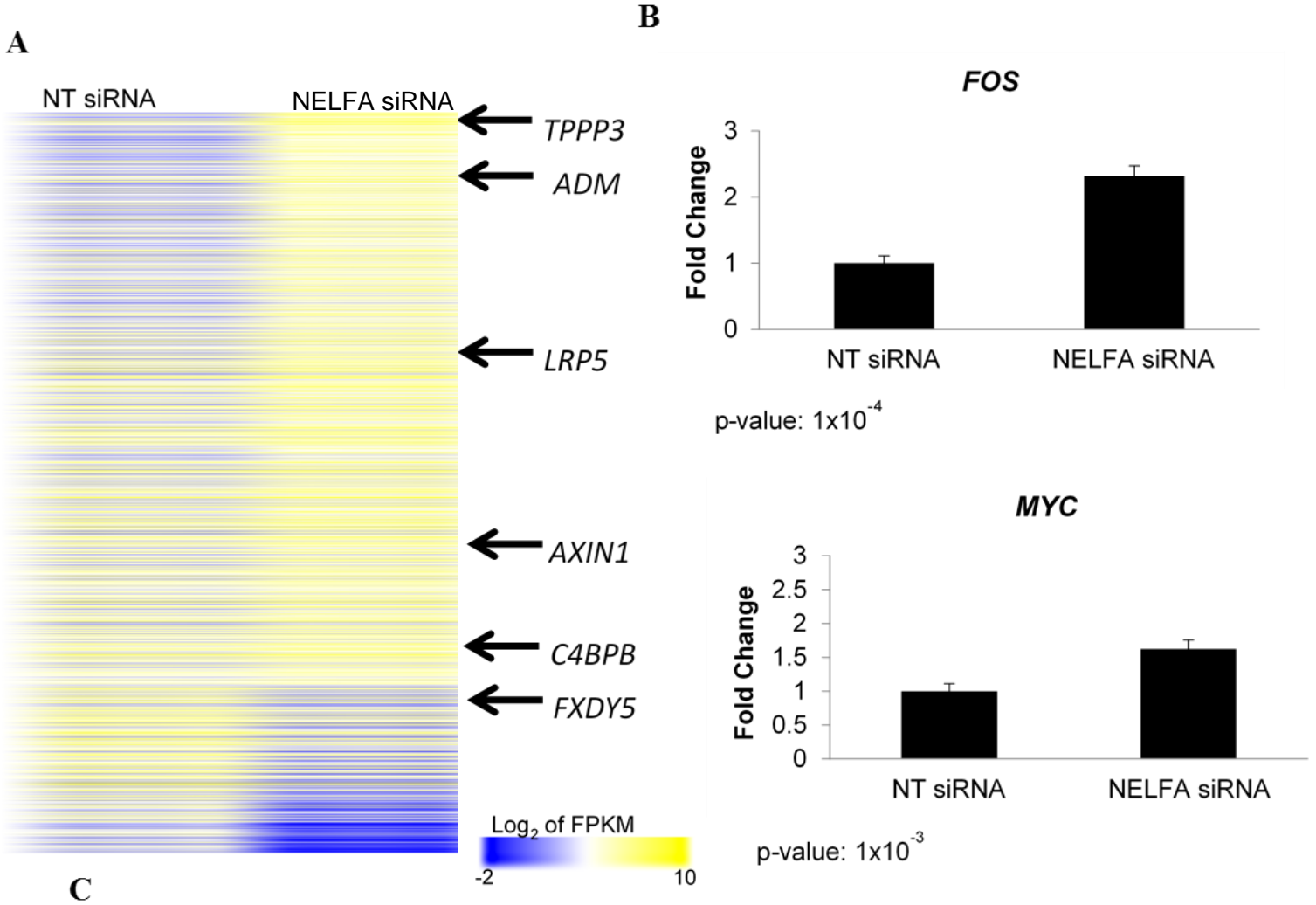| Pathway | Genes | p-value |
|---|---|---|
| Cellular Senescence | CDKN1A; ETS1; FOS; TP53; HIRA; E2F1 | $1\times10^{-9}$ |
| Cellular Response to Stress | RELA; SOD2; MAPK3; RPTOR; MAPKAPK5 | $1\times10^{-7}$ |
| Epigenetic Regulation of Gene Expression | EZH2; DNMT1; SAP18; SIN3B; SUV39H1; TET3 | $4\times10^{-6}$ |
| WNT Signaling | FZD2; AXIN1; PYGO2; PORCN; LRP5; WNT4 | $7\times10^{-5}$ |

Figure 3.4 NELF Knockdown Results in Global Gene Expression Changes in HT29 Cells. A) Heatmap depicting the 2,797 genes that change in expression (>1.5 fold) upon loss of NELF. B) The expression of the paused genes *FOS* and *MYC* decreases following NELF knockdown. C) Pathway analysis of the 2,797 genes that change in expression (>1.5 fold).

**Figure 3.5**
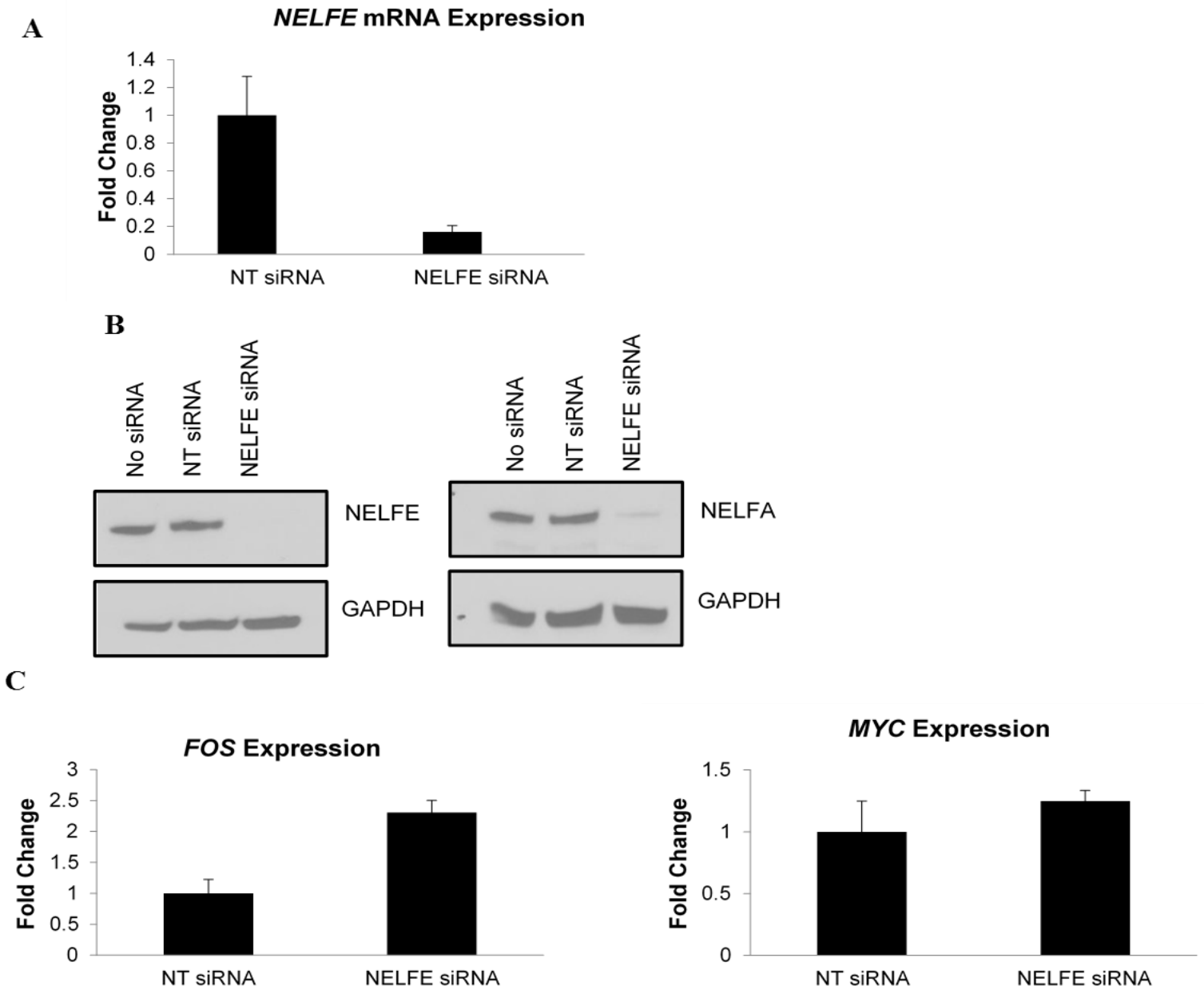


Figure 3.5 NELFE Knockdown in HT29 Cells. A) NELFE knockdown results in an >80% reduction in *NELFE* transcript. B) Knockdown of NELFE results in a loss of NELFE and NELFA at the protein level. C) NELFE knockdown results in an increase in the expression of the paused genes *FOS* and *MYC*.

**Figure 3.6**



Figure 3.6 Loss of NELFA Increases Cell Proliferation. A) NELFA knockdown results in an >80% reduction in *NELFEA* transcript throughout the time course. B) Knockdown of NELFA results in a loss of NELFA at the protein level throughout the time course. C) Cells with decreased expression of NELFA show an increase in cell proliferation from Day 3 to Day 4.

**Figure 3.7**

A  *NELFA* mRNA Expression

B  NT siRNA

Dapi          NELFA          Overlay

NELFA siRNA

Dapi          NELFA          Overlay

C  **NELFA Staining**

n= 51
p-value: $8.6 \times 10^{-7}$

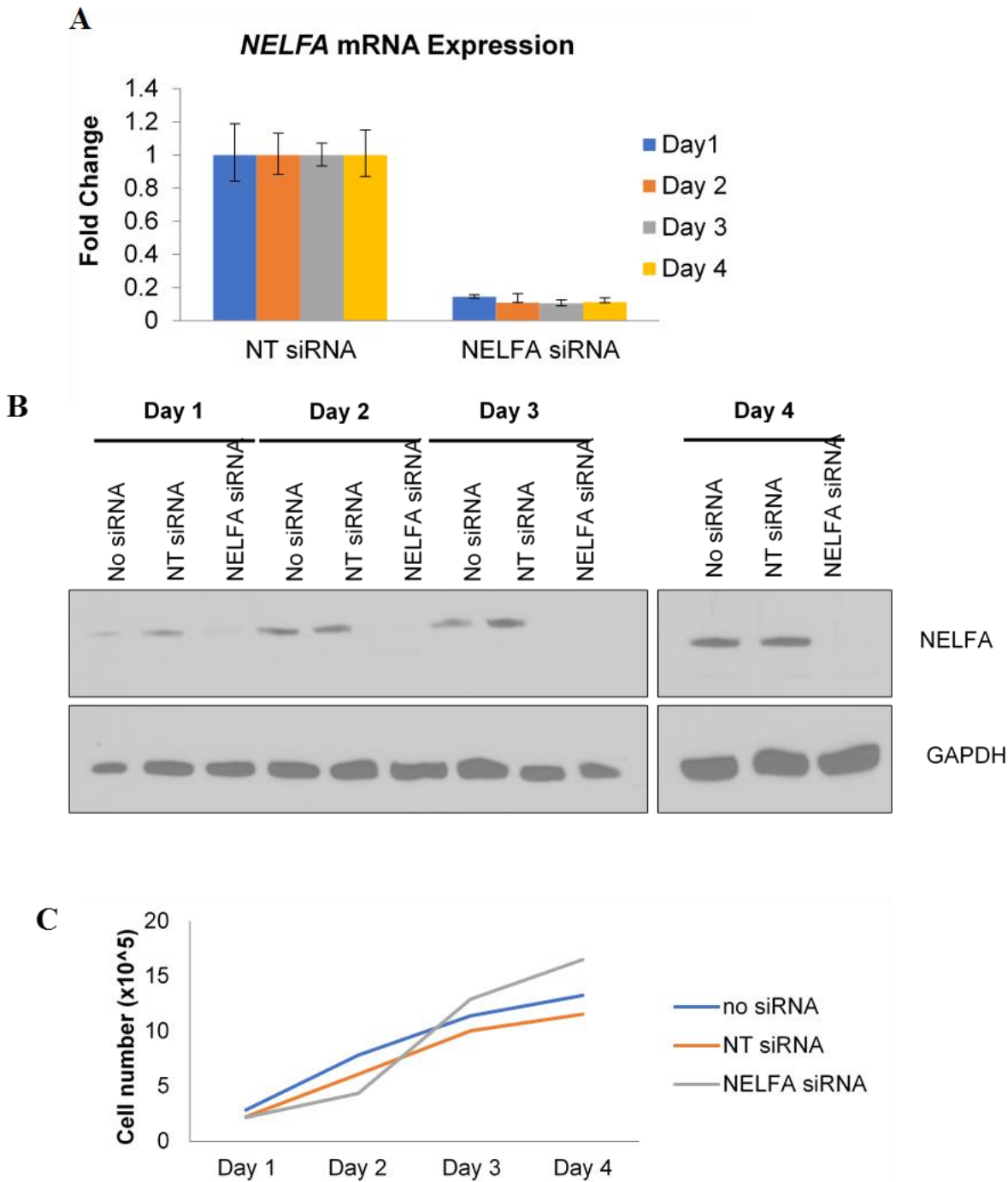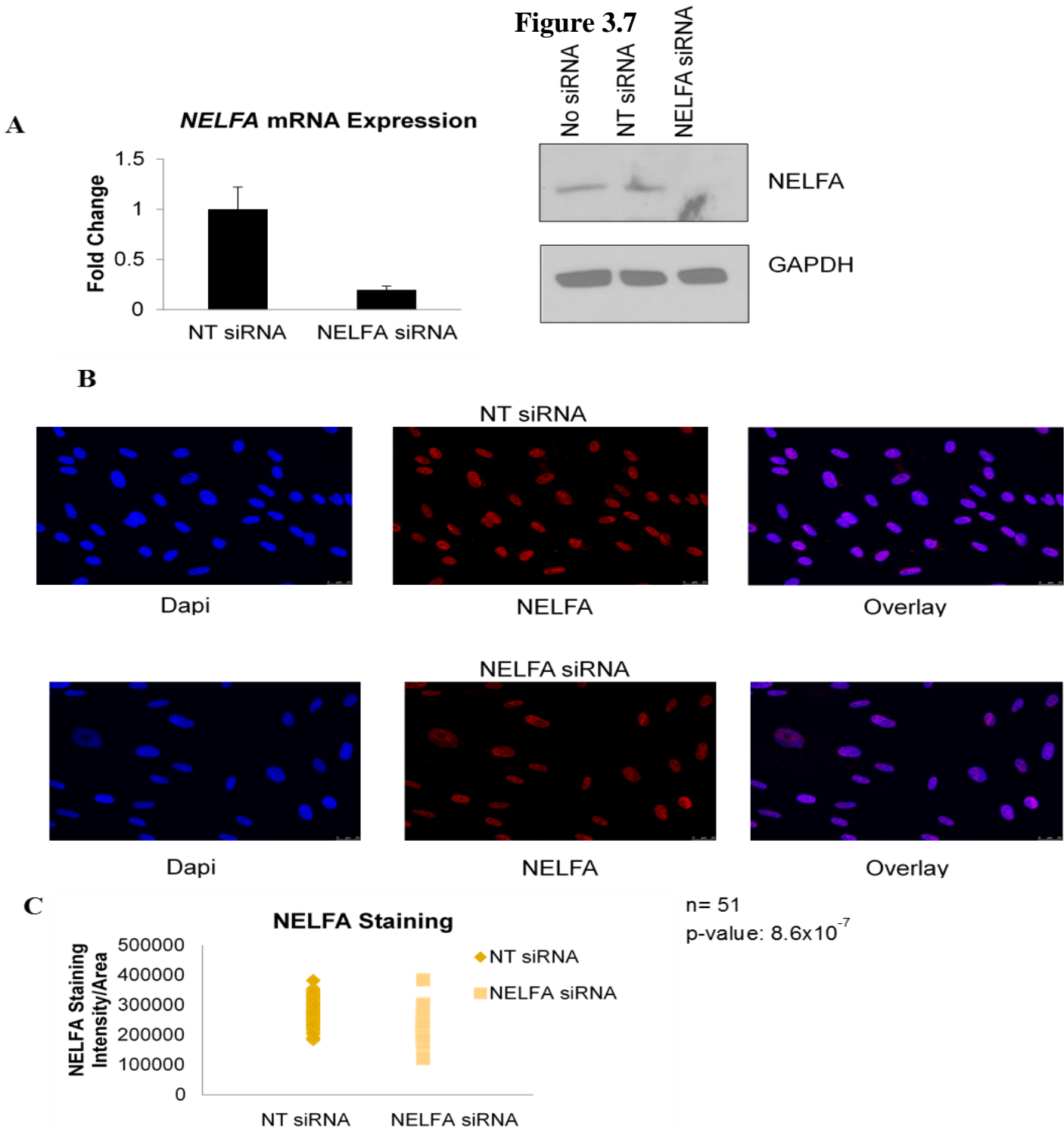Figure 3.7 NELFA knockdown can be detected by Immunofluorescence (IF). A) NELFA knockdown results in an >80% reduction in *NELFEA* transcript and loss of NELFA protein as measured by western blot. B) IF for NELFA (red) in primary dermal fibroblasts shows NELFA localizing to the nucleus. C) NELFA knockdown results in a decrease in NELFA expression at the protein level as measured by IF.

**Figure 3.8**
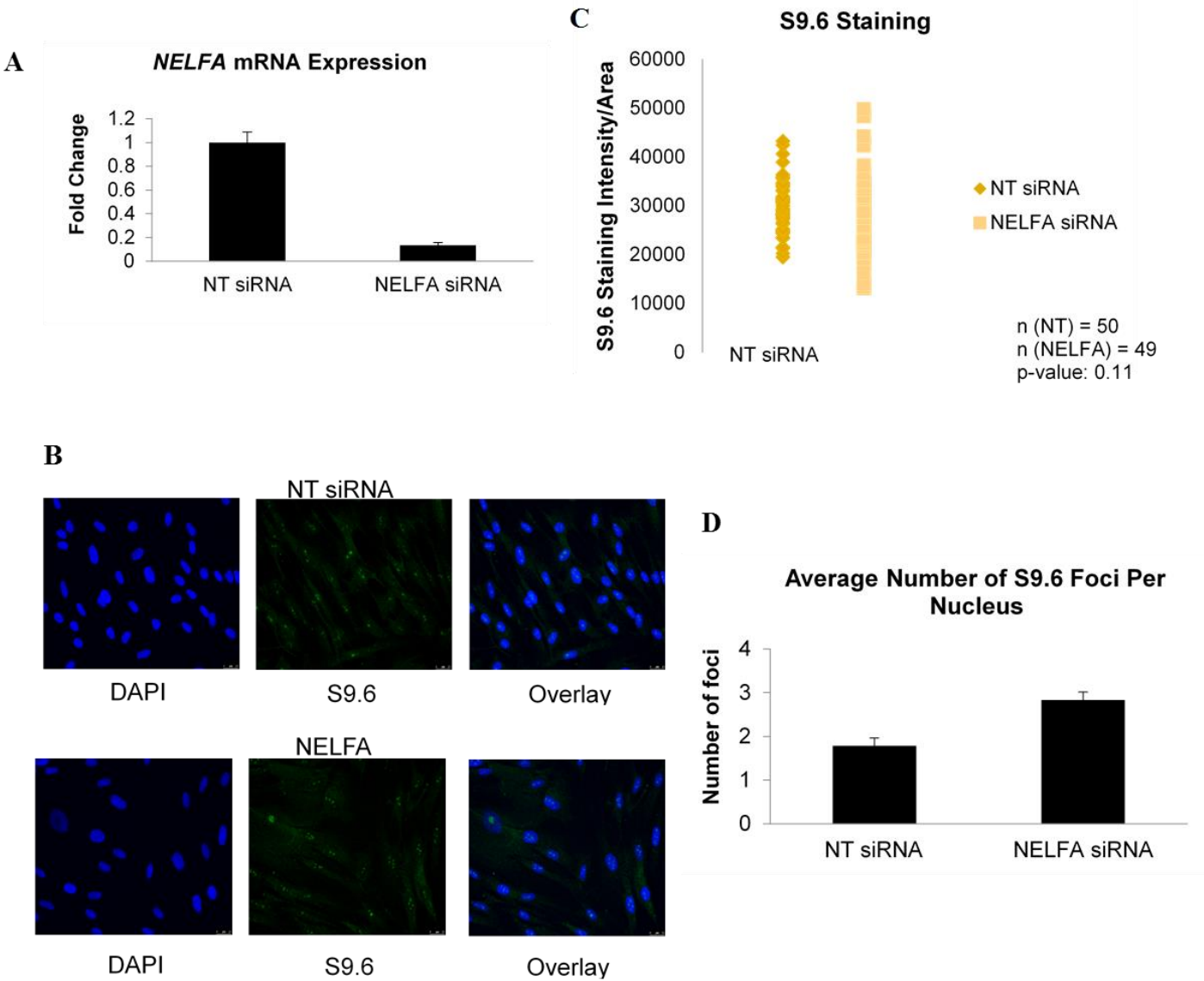


Figure 3.8 NELFA Knockdown Leads to an Increase in R-loops as measured by Immunofluorescence (IF). A) NELFA knockdown results in an >80% reduction in *NELFEA* transcript. B) IF for S9.6 (green) in primary human dermal fibroblasts shows R-loops localizing to the nucleolus. C) NELFA knockdown does not alter the intensity of S9.6 (green) staining in primary human dermal fibroblasts. D) NELFA knockdown results in an increase in the amount of R-loops in primary human dermal fibroblasts as measured by S9.6 staining.

# Chapter 4

## Conclusion

## Summary of Findings

My study of negative elongation factor (NELF) has identified a comprehensive list of paused genes in primary human dermal fibroblasts, demonstrated a role for NELF in maintaining an open chromatin state at gene promoters, and showed that NELF regulates gene expression through eRNA. We extended the findings of polymerase pausing in *Drosophila* to humans. We identified over 7,200 genes as paused in humans through both a modified nuclear run-on assay (PRO-seq) (Kwak et al., 2013) and chromatin immunoprecipitation (ChIP) for the serine 5 phosphorylated form of RNA polymerase II (RNA Pol II Ser5P). The results identified genes such as *EGFR*, *FGFR1*, *POLG2* and *MAP2K5* that were not known to have paused RNA polymerases in their promoters. We showed that for nearly 2,000 genes, the loss of NELF led to a significant decrease in pausing, and for more than 1,000 genes, it led to a decrease in transcriptionally active RNA Pol II in the promoter regions.

We next wanted to understand the influence of NELF-mediated pausing on human gene expression. In *Drosophila*, it has been shown that NELF is able to regulate gene expression through two mechanisms. The first is by inhibiting RNA polymerase II (RNA Pol II) from entering into active elongation. Thus, decreasing expression of these genes. The second, is by

clearing the first nucleosome to allow for rapid and efficient re-intiation of RNA Pol II. This

leads to high expression of these genes (Gilchrist et al., 2008). Here, we show that loss of NELF

results in the increase in expression of over 1,300 genes. Thus, demonstrating that NELF is able

to decrease the expression of target genes by decreasing the amount of RNA Pol II entering into

active elongation. We also find that a similar number of genes decrease in expression upon loss

of NELF. Next, we wanted to begin to tease apart the mechanism. In examining all genes that

change in expression, we found a significant decrease in the number of nascent transcripts at the

promoters of genes that increase versus those that decrease in expression upon loss of NELF.

This indicates that there is less active transcription at these promoters. We believe that NELF is

able to facilitate the re-initiation of RNA Pol II at the promoters of a subset of human genes,

similar to what is seen in *Drosophila*. This is the first description of this phenomenon in humans.

In our study of NELF localization, we found that over 3,900 peaks of NELF were located

outside of promoter regions. It has been shown that NELF associates with enhancer elements in

Drosophila (Core et al., 2012). More recently, NELF has been shown to bind eRNA. This

binding serves to pull NELF off of paused complexes in order to facilitate release from pausing

(Schaukowitch et al., 2014). This led us to investigate whether the just over 3,900 sites of NELF

localization outside of promoter regions occurred at eRNA. Acetylation of lysine 27 of histone

H3 as well as active transcription are marks of active enhancer regions and eRNA (Creyghton et

al., 2011, Lam et al., 2014).  We found that 2,024 of the NELF enriched regions outside of

promoters were also associated with marks of active enhancers. Upon further examination of

these sites, we found that loss of NELF decreased transcription of 1,169 eRNA. Thus,

implicating NELF as regulator of eRNA transcription. We next wanted to investigate whether

loss of these eRNA upon NELF knockdown resulted in changes in gene expression. We focused

exclusively on genes that were not paused in order to show that the effect was due to loss of

eRNA. Overall, we found that NELF regulates the expression of 394 genes through mediating-

eRNA transcription.

Following our studies in normal cells, we were interested to see how pausing influenced

gene expression in cancer cells. We identified over 3,900 paused genes in our cells by using

PRO-Seq (Kwak et al., 2013). These genes are enriched for the EGFR1 signaling pathway, gene

expression, membrane trafficking and transcription regulation by TP53. We then went on to

examine how alterations in pausing, specifically a loss of pausing, affects gene expression. We

found that over 2,700 genes change in expression upon loss of NELF. Over 80% of these genes

show an increase in expression. This implicates that the predominant role of NELF-mediated

pausing is to slow the transition from initiation to active elongation. This finding demonstrates a

difference in the function of pausing in regulating gene expression in a normal versus cancer

context. We found an enrichment for genes involved in cancer promoting pathways such as

WNT signaling. These pathways seem to be specific to our cancer cells.  We also show that there

is a potential increase in cellular proliferation following NELF knockdown which may be due to

the gene expression changes we identified.

Our finding that NELF-mediated pausing is a key regulator of gene expression, led us to

investigate whether pausing had effects on any other transcriptional processes. We chose to focus

on R-loops since they accompany an actively elongating polymerase. We found that loss of

NELF-mediated pausing results in an increase in R-loop formation in primary newborn human

dermal fibroblasts. This increase in R-loops may represent an increase in overall transcription.

This finding provides further evidence that pausing at the promoter serves as the rate limiting step between initiation and active elongation.

## Significance

Transcription must be tightly regulated in order to ensure proper functioning of the cell. Dysregulation of this process can lead to the development of many diseases including cancer. It has been shown that promoter-proximal pausing serves as the rate-limiting step between initiation and elongation of RNA polymerases and that this step is able to influence the expression of genes involved in key developmental and stimulus-responsive pathways (Adelman et al., 2009, Gilchrist et al., 2008, Amleh, et al., 2009, Wang et al., 2010, Williams et al., 2015). However, most of these discoveries were made in model organisms, such as *Drosophila* or in cancer cells. Though this has provided insight into the mechanism and function of pausing, there still remained a lack of knowledge into how polymerase pausing affects transcription in normal human cells. Here, we generated a comprehensive profile of promoter-proximal pausing in primary human dermal fibroblasts. Within this profile, we identified a number of genes not before described to be paused, including EGFR and the TGFβ receptors (TGFBR1 and TGFBR2). Both of these pathways are commonly dysregulated in cancer. Because pausing has a subtle effect on expression, the alterations typically found in these pathways may have masked the importance of pausing at these genes. This highlights the importance of using primary cells over cancer cells. Through our use of primary cells, we are able to more clearly identify the regulatory steps in these pathways that may be overridden in cancer cells to promote cellular survival and disease progression. We also showed that the expression of these genes is regulated by promoter-proximal pausing. Because these receptors are integral components in cancer

development, our study introduces a new layer of regulation onto key pathways and opens the door for novel ways to target these proteins outside of small molecule inhibitors.

Previous work has shown that NELF serves as not only a negative regulator of elongation, but also a negative regulator of gene expression. Here, my work has shown that NELF is able to serve as a positive regulator of gene expression by aiding in the re-intiation of RNA Pol II at promoter regions as well as by regulating the transcription of eRNA. Subunits of NELF have been shown to be overexpression in both hepatocellular carcinoma as well as breast cancer with the highest NELF expression found in highly metastatic tumor subtypes; however, there have been no studies examining how this overexpression leads to cancer development (Iida et al., 2012, Sun et al., 2008). Previous studies have shown that NELF is able to regulate gene expression in two ways. The first is by restraining the polymerase from entering into active elongation and the second is by maintaining an open chromatin state at promoters (Gilchrist et al., 2008). Here, we show that loss of NELF is able to decrease gene expression by lowering the amount of active transcription at promoters. Expanding upon these data, it is possible that an overexpression of NELF would facilitate re-initiation at the promoters of the target genes thus leading to increased levels of expression. We have shown that components of the TGFβ pathway, including TGFBR1, are regulated by this mechanism. The TGFβ pathway is able to act in both tumor suppressor as well as tumor promoting roles with the tumor promoting role occurring in later stage and more advanced cancers (LeBrun, 2012). It is possible that overexpression of NELF identified in late stage cancers creates an environment in which the promoters of the TGFβ pathway components are continuously held in an open conformation by NELF. This may lead to an overall increase in expression of the pathway components leading to increased TGFβ function in tumor cells thus promoting migration and invasion. Therefore,

NELF may indirectly influence cancer invasion and metastasis by facilitating expression of genes involved in tumor promoting pathways.

In our studies, we also briefly examined the role of pausing in regulating the expression of genes in HT29 colorectal cancer cells. We found that pausing functions primarily to reduce the expression of its target genes HT29 cells. These genes are enriched for pathways typically harnessed in cancer including, cellular senescence, cellular response to stress, epigenetic regulation of gene expression and WNT signaling. We do not see the same phenomenon in our primary cells. This suggests that one of the functions of NELF in cancer cells is to serve as a backstop at these genes to combat stress placed on the cell by driver mutations. It is possible that by maintaining or increasing pausing, the cell could tamper the effects of mutations in transcription factors such as *MYC*. Because *MYC* promotes uncontrolled cell growth and cell survival, pausing could function to reduce a catastrophic transcriptional response in the early transformation of the cell. Pausing may serve to reduce the transcription of target genes just enough to allow the cell to transition into a malignant phenotype gradually. This slowed transition could aid the cell in avoiding catastrophic damage due to uncontrolled division promoted by the driver. In our colorectal cancer cells, we found that pausing regulates components of the WNT pathway as well. Because mutations in the WNT pathway, particularly mutations in *APC*, drive tumorigenesis, it is possible that NELF and pausing are functioning to facilitate this process by reining in the transcriptional program under the control of the WNT pathway in the early stages of cancer development. These findings provide insight into how the earliest stages of cancer development are regulated and may also demonstrate a novel mechanism by which tumor cells are able to tolerate driver mutations.

The ability of pausing to reduce target gene expression may also be harnessed to downregulate tumor suppressor pathways. We found that loss of pausing upregulates genes involved in cellular senescence and the cellular response to stress. It is possible that cancer cells could utilize an overexpression of NELF to increase pausing at these target genes. By increasing pausing, the cells could reduce the expression of these pathways without acquiring another potentially fatal mutation. This would also serve as a gentle method to decrease gene expression, since pausing does not silence the transcription of genes completely. A controlled decrease in tumor suppressing pathways could allow a cell to still respond effectively to signals during the transition from a premalignant to a malignant state. In later stages, pausing could serve to ensure a complete loss of signaling from these pathways by creating a backstop at these genes.

Our findings provide insight into the transition of cells from a normal to a malignant state. We show that pausing regulates the expression of genes involved in signaling pathways that are commonly dysregulated in cancers such as EGFR and TGFBR1. This highlights the importance of pausing in maintaining normal cellular functioning. However, we also show that pausing may aid in the initial transition to malignancy by creating a favorable environment for driver mutations to function. Finally, we show that once transformation has occurred pausing can serve to enhance the cancer phenotype by downregulating the expression of genes involved in cellular senescence and cellular stress response. Pausing may also function in late stages of tumorigenesis. Pausing can promote invasion and metastasis by upregulating genes important for invasion and metastasis by promoting the re-initiation of RNA Pol II. Overall, our findings show that pausing can function in both tumor suppressing and tumor promoting roles throughout the development and progression of cancer.

## On-going Work

## The maintenance of open chromatin by NELF

Here, we show that loss of NELF results in changes in gene expression. In particular, we show that loss of NELF results in a decrease in expression of over 1,300 genes as well as a significant decrease in the amount of nascent transcription at these promoters. We believe that this is due to a closing of the chromatin in the absence of NELF. In Drosophila, it has been shown that the first nucleosome reassembles in the promoters of genes whose expression decreases upon loss of NELF (Gilchrist et al., 2008).

In order to gain a better picture of the chromatin state at these promoters, it will be important to determine nucleosome location as well as marks of open chromatin before and after NELF knockdown. To identify the chromatin state at these promoters, we will begin by performing chromatin-immunoprecipitation (ChIP) with an antibody targeting the tri-methylation mark on lysine 4 of histone H3 (H3K4me3). This mark indicates an open chromatin state. We will begin by examining the genes that show a decrease in expression as well as a decrease in active RNA Pol II at their promoters upon loss of NELF. We believe these genes will show alterations in their chromatin state upon loss of NELF. If we see a reduction in the amount of H3K4me3 following NELF knockdown, this will indicate a closing of the chromatin at the promoter region by histone modifications. The ability of NELF to maintain open chromatin by through histone modifications has not been previously described.

We will also perform Assay for Transposase Accessible Chromatin followed by high throughput sequencing (ATAC-Seq) to examine the location of nucleosomes before and after NELF knockdown. This assay utilizes a hyper-active transposase which will cut exposed DNA

and ligate adapters to the sequence. These fragments can then be sequenced to show areas of the genome where DNA is not wrapped around a nucleosome. This will allow us to map the location of nucleosome free DNA before and after NELF knockdown. If we see a re-assembly of the first nucleosome upon loss of NELF, this will indicate that NELF is responsible for blocking the assembly of this nucleosome at promoters. This model is similar to what has been described in *Drosophila*. Overall, these data will provide a more complete mechanism for the role of NELF-mediating pausing in regulating gene expression.

## Confirmation of NELF regulated eRNA

We identified putative enhancer RNA (eRNA) that are regulated by NELF. Though these sites show marks of active enhancers, it is unclear whether these eRNA truly promote transcription of target genes. To investigate this, we will clone a few of the putative enhancers we identified and place them near a luciferase reporter gene. Next, we will measure the level of luciferase activity in the presence and absence of the enhancer.  An increase in luciferase activity in the presence of the enhancer indicates that the enhancer is active and serves to promote gene expression.

We can also investigate the role of NELF in mediating the enhancer activity. To do this, we will utilize the same luciferase system; however, we will measure luciferase activity before and after NELF knockdown. If NELF is responsible for regulating the activity of the enhancer, we will see a decrease in luciferase expression upon loss of NELF. These data will not only confirm functionality of our putative enhancer regions, they will also confirm the ability of NELF to regulate gene expression through mediating eRNA transcription. This demonstrates an indirect effect of NELF on eRNA mediated gene expression, since we will not have measured

the abundance of the eRNA generated from these putative enhancers in this assay. There may be downstream effects outside of NELF binding to the site of eRNA that influences the change in luciferase activity.

## Identifying eRNA target genes

Upon confirming the functionality of our eRNA as well as the ability for NELF to regulate eRNA function, it will be necessary to investigate these studies in vitro. To do this, we will need to identify the targets of our eRNA. We identified 394 genes that are near our putative eRNA and also show a >10% change in expression upon loss of NELF. Using the gene nearest to the eRNA gives us a preliminary look into the ability of NELF to regulate gene expression through eRNA; however, it has been shown that eRNA and active enhancer are able to act over long distances (Maniatis et al., 1987).

To begin we will choose a few representative examples from our list of over 2,000 eRNA. We will choose these examples based on the level of expression change of the target gene. We will choose examples with large increases and decreases in expression as well as some with more subtle changes in order to achieve a representative sample. We will then knockout these active enhancers by CRISPR/Cas9. Following knockout of the enhancers, we will first check the expression of the nearby gene we identified by quantitative real-time PCR (qRT-PCR). If we see a change in expression of the nearby gene, it will confirm that the nearest gene is a target of our eRNA. If there is no change in expression then we will perform RNA-sequencing to search for genes whose expression changes upon loss of the enhancer. Once the target of the enhancer is identified, we will be able to truly examine the role of NELF in mediating the expression of the eRNA target gene. To do this, we will perform qRT-PCR for the target gene

before and after NELF knockdown. If we see a change in expression following NELF knockdown that replicates what is seen in the absence of the enhancer, it will confirm that NELF is able to regulate expression of that gene through eRNA. Overall, these data will identify the true targets of our eRNA as well as confirm the ability of NELF to regulate gene expression by mediating eRNA transcription.

## Future Directions

In my studies, I have shown that a loss of NELF results in global changes in promoter-proximal pausing as well as gene expression in both normal and tumor cells. Work by others has shown that a loss of NELF leads to poor prognosis in breast cancer patients (Sun et al., 2008). It has been shown that this phenotype may be due to an increase in cell cycle progression and dysregulation of the estrogen response in the absence of NELF (Sun et al., 2008, Aiyar et al., 2004). Sequencing of tumors has also shown that NELF subunits are deleted in a variety of tumor types. This implicates that NELF may have tumor suppressive properties. My work supports this hypothesis. I have shown that NELF is regulator of cancer promoting pathways in normal cells and that upon loss of NELF we see changes in the expression of tumor promoting pathways such as EGFR signaling. I also see that upon loss of NELF in tumor cells there is an increase in the expression of genes involved in tumor suppressive pathways such as cellular senescence and the transcriptional response to TP53. In preliminary studies, I have shown that HT29 tumor cells show an increase in cell proliferation upon loss of NELF. However, this does not correspond with an increase in the expression of cell cycle promoting genes as is seen in breast cancer cells.

Changes in cellular growth and proliferation are two of the main phenotypes currently targeted by traditional therapeutics. Because loss of NELF is associated with increased cell growth and proliferation, tumors with a loss of NELF may be more sensitive to ionizing radiation or chemotherapeutics. Recently, it has been shown that loss of NELF is associated with slight sensitivity to ionizing radiation. Specifically, NELFE has been shown to be recruited to sites of double-stranded breaks to repress transcription upon DNA damage NELFE has recently been shown to be recruited by PARP-1 to sites of double-strand breaks (DSBs) to repress transcription following DNA damage (Awwad et al., 2017).

Further insight into how NELF may be facilitating cellular growth and progression is found in Wolf-Hirschornn Syndrome. This disease is caused by a deletion of chromosome 4p16.3. This region contains *NELFA* (*WHSC2*), which was originally identified in the characterization of this disease (Mariotti et al., 2000). There have been few studies focused on the role of NELFA in the molecular underpinnings of this disease. However, one study has shown that loss of NELFA results in defects in S-phase progression (Kerzendorfer et al., 2012). Though this finding does not completely agree with the findings in breast cancer cells, it still implicates NELF as a regulator of the cell cycle.

These findings imply that tumors with deletions of NELF subunits, may be better candidates for radiation and chemotherapeutics. Alterations in NELF are found in a wide-variety of tumor types, yet these alterations do not seem to be drivers of the tumors. This eliminates NELF as a strong candidate for targeted therapeutics; however, it does make NELF a potential marker for identifying which patients may respond better to radiation and chemotherapy. With the advent of clinical Next-Gen Sequencing, sequencing the transcriptome of patients is on its way to becoming a part of diagnosis. This technique could be used to identify loss of expression

of NELF subunits without needing to characterize and verify every mutation in a NELF subunit. Because loss of one subunit leads to the loss of all of the others, detecting a decrease in a single subunit transcript would indicate the loss of the entire NELF complex. Based on my findings and those of others, patients that show this decrease in expression would be the best candidates for treatment with traditional therapeutics.

The most common alteration of NELF found in cancer is the amplification of a single or even multiple subunits (ref). However, almost all studies in both normal and cancer cells have focused on the consequences of a loss of NELF. Because of this, the current understanding of NELF and cancer progression may be lacking. Only one study has looked into the function of NELF overexpression in tumors. Iida et al., showed that NELFE is overexpressed in hepatocellular carcinomas and that a higher incidence of invasion as well as poor prognosis was correlated with this expression level (2012). This begs the question as to how an overexpression of NELF can contribute to tumor progression across many tumor types. My work provides a rationale for investigating this mechanism. In my studies, I have shown that NELF has novel functions outside of regulating pausing and gene expression which seem to be the two processes most affected by loss of NELF. Specifically, I found that NELF is able to facilitate re-initiation of RNA Pol II at promoter regions as well as regulate the transcription of eRNA. This opens up the question as to what other functions of NELF have yet to be described.

Based on the study by Iida et al., as well as my data showing that the TGFβ pathway seems to be regulated by NELF facilitating the re-initiation of RNA Pol II at promoters, it is possible that the overexpression in NELF found in tumors aids in the invasion and metastasis of tumors. We also see NELF amplification in tumors that have many other mutations, as well as those that are in a late stage or even in a recurrent tumor (Cerami et al, 2012, Gao et al., 2013).

Because of this, NELF may serve as a marker of invasion and metastasis that could be used to catch potentially invasive tumors earlier. Using transcriptome-wide sequencing or even smaller-scale methods to search for NELF overexpression at diagnosis could help to identify those patients whose tumor may have a more invasive phenotype even if no metastasis has been detected. This would allow clinicians to treat these patients more aggressively or with different therapies.

Future studies into the role of NELF in mediating the DNA damage response and cell cycle progression in a variety of tumor types will allow for NELF to be used as an even finer marker in personalized cancer therapy. Studies on the effects of NELF overexpression in both cancer and normal cells need to be undertaken in order to better understand how this genetic alteration is contributing to disease. It is possible that a deletion or an amplification in different subunits may lead to different phenotypes. Perhaps tumors that show a deletion in NELFE will have a higher sensitivity to ionizing radiation, whereas those with a deletion in NELFA will show a higher sensitivity to therapies that force the transition from S-phase to M-phase. Also, alterations in a single subunit may have different effects in different tumor types. NELF alone could serve as a marker for defining therapeutic interventions, staging and identification of the tumor type. A more in-depth characterization of the NELF complex as well as each subunit as a whole will aid in the use of NELF as a way to tailor cancer therapy on an individual basis.

# References

Adelman, K., Kennedy, M.A., Nechaev, S., Gilchrist., D.A., Muse, G.W., Chinenov, Y., Rogatsky, I. Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. PNAS. 2009 October 27; 106(43):18207-18212.

Aiyar, S.E., Sun, J., Blair, A.L., Moskaluk, C.A., Lu, Y., Ye, Q., Yamaguchi, Y., Mukherjee, A., Ren, D., Handa, H., Li., R. Attenuation of estrogen receptor a-mediated transcription through estrogen-stimulated recruitment of a negative elongation factor. Genes Dev. 2004 Sep 1; 18(17):2134-46.

Amleh, A., Nair, S.J., Sun, J., Sutherland, A., Hasty, P., Li, R. Mouse cofactor of BRCA1 (Cobra1) is required for early embryogenesis. PLoS One. 2009 Apr 2;4(4):e5034.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Huer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., Schultz, N. the cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012 May;2(5):401-4.

Core, L.J., Waterfall, J.J., Glichrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., Lis, J.T., Defining that status of RNA polymerase at promoters. Cell Rep. 2012 Oct 25;2(4):1025-35.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., Jaenisch, R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. PNAS. 2010 Dec 14;107(50):21931-6.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., Schultz, N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013 Apr 2;6(269):pl1.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K.B., Collins, J.B., Li, L., Gilmour, D.S., Adelman, K. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev. 2008 Jul 15;22(14)1921-33.

Iida, M., Iizuka, N., Tsunedomi, R., Tsutsui, M., Yoshida, S., Maeda, Y., Tokuhisa, Y., Sakamoto, K., Yoshimura, K., Tamesa, T., Oka, M. Overexpression of the RD RNA binding protein in hepatitis C virus-related hepatocellular carcinoma. Oncol Rep. 2012 Aug;28(2):728-34.

Kerzendorfer, C., Hannes, F., Colnaghi R., Abramowicz, I., Carpenter G. Vermeesch, J.R., O'Driscoll, M. Characterizing the functional consequences of the haploinsufficiency of NELF-A (WHSC2) and SLBP identifies novel cellular phenotypes in Wolf-Hirschhorn syndrome. Hum Mol Genet. 2012 May 15;21(10):2181-93.

Kwak, H., Fuda, N.J, Core, L.J., Lis, J.T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science. 2013 Feb 22;339(6122):950-3.

Lam, M.T.Y., Li, W. Rosenfeld, M.G., Glass, C.K. Enhancer RNAs and regulated transcriptional programs. Trends Biochem Sci. 2014 Apr;39(4):170-82.

LeBrun, JJ. The dual role of TGFb in human cancer: from tumor suppression to cancer metastasis. ISRN Mol Biol. 2012 Dec 24;2012:381428.

Maniatis, T., Goodbourn, S., Fischer, J.A. Regulation of inducible and tissue-specific gene expression. Science. 1987 Jun 5;236(4806):1237-45.

Mariotti, M., Manganini, M, Maier, J.A. Modulation of WHSC2 expression in human endothelial cells. FEBS Lett. 2000 Dec 29;487(2):166-70.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., Lis, J.T. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev. 2011 Apr 1;25(7):742-54.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grisson, S.F., Zeitlinger, J., Adelman. K. RNA polymerase is poised for activation across the genome. Nat Genet. 2007 Dec;39(12):1507-11.

Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D., Hasegawa, J., Omori, M., Inukai, N., Endoh, M., Yamada, T., Hiroshi, H. Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. Mol Cell Biol. 2003 Mar;23(6):1863-73.

O'Brien T., Lis, J.T. RNA polymerase II pauses at the 5' end of the transcriptionally induced *Drosophila hsp70* gene. Mol Cell Biol. 1991 Oct;11(10):5285-90.

Rougvie, A.E., Lis, J.T. The RNA polymerase II molecule at the 5' end of the uninduced *hsp70* gene of D. melanogaster is transcriptionally engaged. Cell. 9 Sep 1988;54(6):795-804.

Schaukowitch, K., Joo, J., Liu, X., Watts, J.K., Martinez, C., Kim, T. Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell. 2014 Oct 2;56(1):29-42.

Sun, J., Watkins, G., Blair, A.L., Moskaluk, C., Ghosh, S., Jiang, W.G., Li, R. Deregulation of cofactor of BRCA1 expression in breast cancer cells. J Cell Biochem. 2008 Apr 15;103(6):1798-807.

Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., Handa, H., Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription *in vitro*. EMBO J. 1998 Dec 15;17(24):7395-403.

Wang, X., Hang, S., Prazak, Gergen, P.J. NELF potentiates gene transcription in the Drosophila embryo. PLoS One. 2010 July 9;5(7):e11498.

Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., Adelman, K. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. Mol Cell. 2015 Apr 16;58(2):311-22.

Wu, C., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., Gilmour, D.S. NELF and DSIF cause promoter proximal pausing on the *hsp70* promoter in *Drosophila*. Genes Dev. 2003 Jun 1;17(11):1402-14.

Wu, C., Lee, C., Fan, R., Smith, M.J., Yamaguchi, Y., Handa, H., Gilmour, D.S. Molecular characterization of *Drosophila* NELF. Nucleic Acids Res. 2005 Mar 1;33(4):1269-79.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., Handa, H. NELF, a mulitisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell. 1999 Apr 2;97(1):41-51.

Yamguchi, Y., Inukai, N., Narita, T., Wada, T., Handa, H. Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol Cell Biol. 2002 May;22(9):2918-27.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., Young, R.A. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet. 2007 Dec;39(12):1512-6.
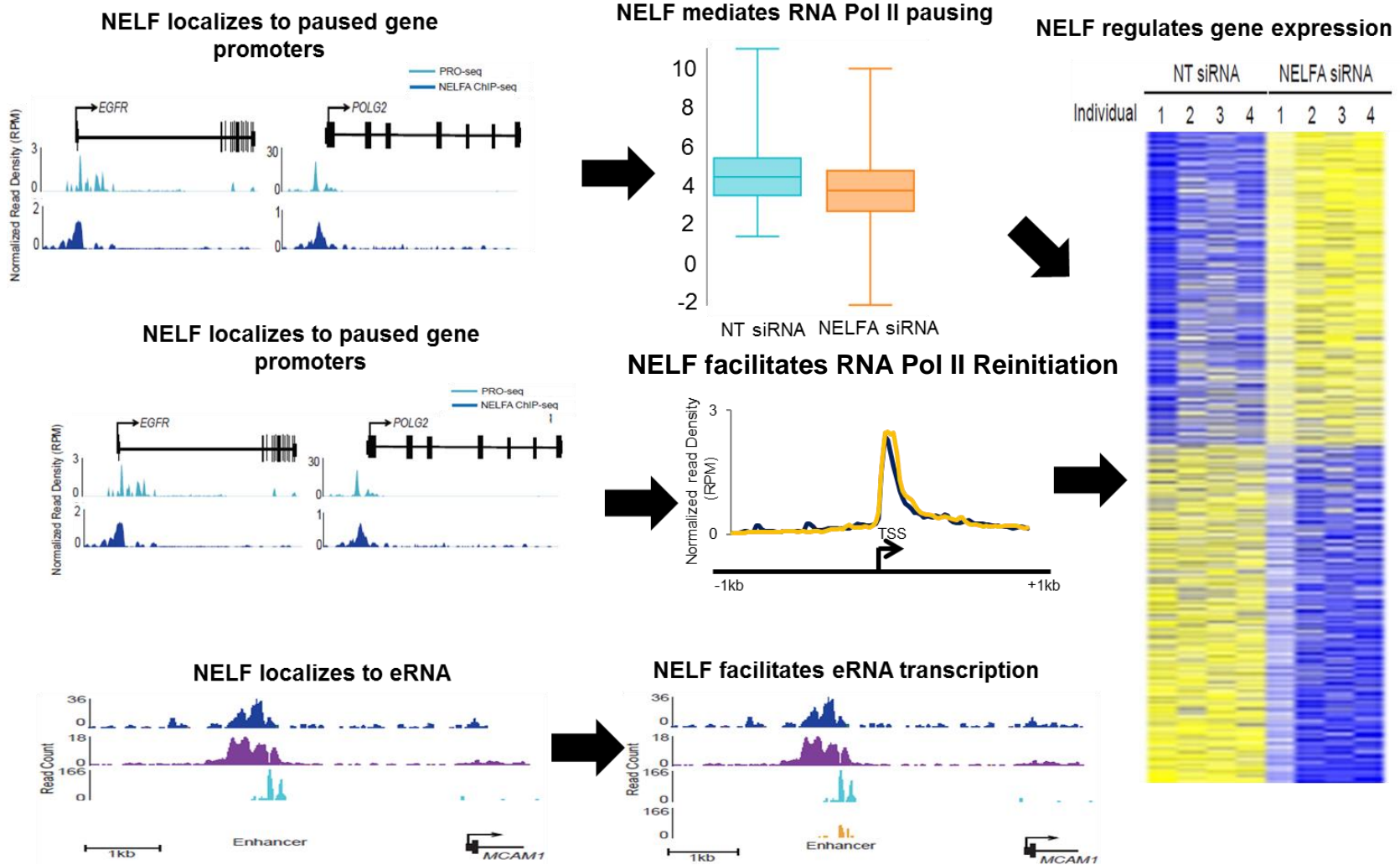
**Figure 4.1**

Figure 4.1 Thesis Summary. NELF regulates human gene expression through three mechanisms. The first is by decreasing active elongation by facilitating RNA Pol II pausing. The second is by promoting re-intiation of RNA Pol II at promoters. NELF is also able to regulate human gene expression by facilitating transcription of eRNA.