DISSERTATION

REPEATED SEQUENCES ENCODING CYS2HIS2 ZINC FINGER MOTIFS INFLUENCE

MRNA POLYADENYLATION AND LOCALIZATION

Submitted by

Aimee L. Jalkanen

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

      Advisor: Carol Wilusz
      Co-Advisor: Jeffrey Wilusz

      Susan Bailey
      Gerrit Bouma
      Douglas Thamm

ABSTRACT

REPEATED SEQUENCES ENCODING CYS2HIS2 ZINC FINGER MOTIFS INFLUENCE
MRNA POLYADENYLATION AND LOCALIZATION

The Cysteine2 Histidine2 zinc finger (C2H2-ZNF) proteins are a vast family with over
700 members in primates, many of which are transcription factors with important roles in
development, differentiation, cell cycle progression, and tumor suppression. Due to the sheer
number of C2H2-ZNF proteins and their roles in modulating expression of other genes, any
mechanism for coordinating their expression could have wide-ranging impacts on cell function
and phenotype.

Previously, a large subset of C2H2-ZNF transcripts were determined to have significant
populations with short poly(A) tails. Here, we show that multiple C2H2-ZNF mRNAs accumulate
with very short or undetectable poly(A) tails, even when newly transcribed. Furthermore, these
C2H2-ZNF mRNAs are restricted to the nucleus. Reporter mRNAs with sequences from the
ZNF12 open reading frame (ORF) and/or the 3' untranslated region (3' UTR) have short poly(A)
tails and are retained in the nucleus. Deletion analysis suggests that repeated sequence
elements in the ZNF12 mRNA that code for zinc finger protein motifs are important in controlling
both poly(A) tail length and nuclear localization. Remnants of C2H2-ZNF motif sequences found
in the ZNF12 3' UTR are also able to confer short poly(A) tails and nuclear retention. Finally, we
use RNA-fluorescence *in situ* hybridization (RNA-FISH) to reveal that ZNF12 reporter transcripts
are found in foci within the nucleus that could represent sites for storage or processing. Overall,
our findings suggest repeated sequence elements encoding C2H2-ZNF protein motifs play a
dual role as regulatory elements that may coordinate expression of the C2H2-ZNF protein family
by controlling post-transcriptional events.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1.1 Overview of the C2H2-ZNF proteins

The Cysteine2 Histidine2 zinc finger (C2H2-ZNF) genes encode transcription factors that utilize zinc finger motifs to bind DNA and influence gene expression (Brown, 2005; Gamsjaeger et al., 2007; Hall, 2005; Wolfe et al., 2000). They comprise one of the largest and most rapidly evolving gene families with over 700 members in humans (Ding et al., 2009; Emerson and Thomas, 2009). Although the functions of most C2H2-ZNF proteins are unknown, they have been proposed to protect the genome by silencing endogenous retrotransposon activity (Jacobs et al., 2014; Lukic et al., 2014; Rowe and Trono, 2011; Rowe et al., 2010) or to facilitate the coercion of transposable element sequences to regulate cellular gene expression (Ecco et al., 2016). C2H2-ZNF proteins with known functions are key regulators of stem cell maintenance, development, differentiation, cell cycle progression, and tumor suppression (Jen and Wang, 2016; Razin et al., 2012). The sheer number of these genes supports the possibility that they may have wide-ranging impacts on gene expression. In addition, the C2H2-ZNF genes possess conserved and redundant sequence elements and protein motifs that provide opportunities for shared regulatory strategies (Dovat, 2002; Huang et al., 2010; Iyengar and Farnham, 2011; Rizkallah et al., 2011; Schnall-Levin et al., 2011). The goal of this study was to uncover novel mechanisms that might control and coordinate C2H2-ZNF gene expression.

1.1.2 C2H2-ZNF protein structure

Zinc finger motifs facilitate the interaction of the C2H2-ZNF proteins with DNA and are generally present in multiple copies, ranging from one motif to over thirty (Stubbs et al., 2011). Each motif is composed of around 30 amino acids (Brown et al., 1985; Miller et al., 1985) and contains appropriately spaced cysteine and histidine residues which coordinate a zinc ion (Miller

et al., 1985). A highly conserved six amino acid linker with the consensus TGEKPY/F generally lies between each of the zinc finger motifs (Figure 1) (Stubbs et al., 2011).

In addition to the zinc finger domains, many of the C2H2-ZNF proteins also have N-terminal effector domains. The KRAB (K̲rüppel-a̲ssociated b̲ox) repressor domain is the best characterized and is found in at least a third of the C2H2-ZNF proteins (Huntley, 2006; Looman et al., 2002). After binding to the target chromatin site, C2H2-ZNF proteins containing the KRAB domain recruit KAP1 (K̲RAB-a̲ssociated p̲rotein 1 also known as TRIM28), which interacts with histone deacetylases and histone methyl transferases to modify the chromatin and repress transcription (Iyengar and Farnham, 2011; Urrutia, 2003). Notably, KAP1 also binds to and represses many of the KRAB-containing C2H2-ZNF genes to potentially coordinate C2H2-ZNF autoregulation (Groner et al., 2010; O'Geen et al., 2007).

1.1.3 Evolutionary history of the C2H2-ZNF genes

While most transcription factors are highly conserved throughout evolution, many of the C2H2-ZNF genes found in humans are primate-specific (Liu et al., 2014a). Analyses in different mammalian species indicate that the rapid expansion of this gene family occurred by tandem duplication, as many C2H2-ZNF genes are found in chromosomal clusters (Bellefroid et al., 1993; Eichler et al., 1998; Emerson and Thomas, 2009; Huntley, 2006). Evidence also suggests that specific C2H2-ZNF genes were either duplicated or lost in groups in different mammalian species after divergent points of evolution (Tadepally et al., 2008). Duplications and losses of individual ZNF domains have also been observed, as well as the appearance of "ZNF" relics in 3' UTRs following acquisition of frameshift or premature termination mutations (Looman et al., 2002).
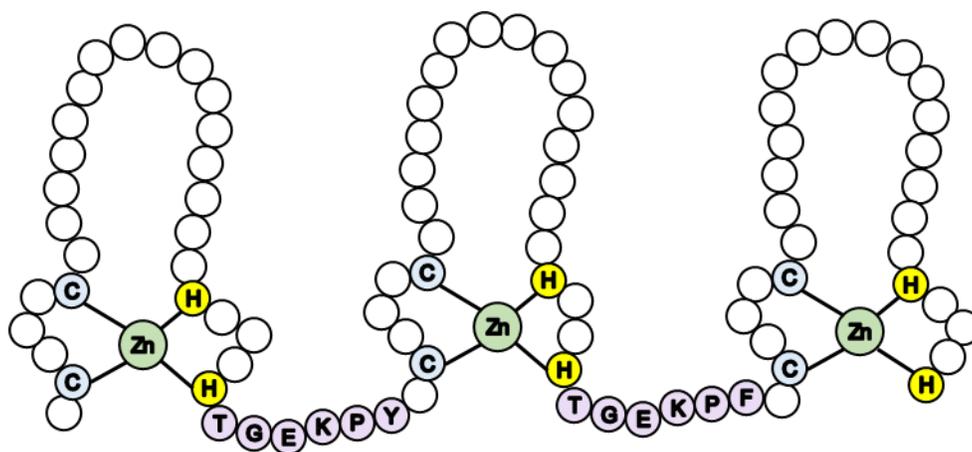
Figure 1: Structure of tandem C2H2-zinc finger motifs. Each of the zinc ions are stabilized by paired cysteine (C) and histidine residues (H). The zinc finger motifs are joined by the consensus sequence TGEKPY/F linker regions.

1.1.4 Functions of C2H2-ZNF proteins

It is proposed that many of the KRAB-domain containing C2H2-ZNF proteins evolved to repress transposable elements (TEs) as the rapid expansion of this gene family coincides with the emergence of these elements in vertebrate genomes (Thomas and Schneider, 2011). Endogenous retroviruses (ERVs) are TEs composed of remnants from a retroviral infection in which the proviral DNA becomes integrated into the host genome and is passed along to the next generation through the germline (Nelson et al., 2003; Stoye, 2012). While ERVs have played important roles in evolution, these elements can also wreak havoc on the genome by contributing to genetic instability and therefore must be silenced (Castro-Diaz et al., 2015). In mammalian cells, ZFP809 recognizes and recruits KAP1 to silence the transcription of ERVs and other retroelements (Wolf and Goff, 2009). Furthermore, depletion of KAP1 in mouse and human embryonic stem cells leads to upregulation of ERVs (Rowe et al., 2010; Turelli et al., 2014). KRAB-zinc finger proteins can also interact with other types of TEs. For example, ZNF91 and ZNF93 seemingly evolved to repress the SINE-VNTR-*Alu* (SVA) retrotransposon and long interspersed nuclear element (LINE) 1 (L1) in the human genome (Jacobs et al., 2014). Notably, the C2H2-ZNF proteins can also bind TEs in adult cells to promote or inhibit the expression of nearby genes. This coercion of TEs by the C2H2-ZNFs may be a novel way to regulate cellular gene expression (Ecco et al., 2016).

Many of the C2H2-ZNF proteins with known functions are transcription factors that regulate expression of genes involved in stem cell maintenance, erythropoiesis, cardiac remodeling, and muscle and organ development (Swamynathan, 2010). Numerous C2H2-ZNF transcription factors also influence the expression of genes involved in cancer (Jen and Wang, 2016). In some instances, C2H2-ZNF proteins enhance cancer progression. For example, ZKSCAN3 (also known as ZNF306 or ZNF309) is overexpressed in colon cancer, multiple myeloma, and prostate cancer and activates transcription of genes involved in cell proliferation, cell migration, and angiogenesis (Yang et al., 2008a, 2008b, 2011a; Zhang et al., 2012). In

prostate cancer cells, ZNF280B enhances transcription of the ubiquitin protein ligase MDM2, which promotes protein degradation of the tumor suppressor p53 (Gao et al., 2013). ZNF282 is overexpressed in many esophageal squamous cell carcinomas and serves as a co-activator of E2F1, a transcription factor involved in cell progression and cell proliferation, to promote tumorigenesis (Yeo et al., 2014). Alternatively, other C2H2-ZNF proteins function as tumor suppressors. ZNF545 is downregulated in many different tumor types and represses the activity of signaling pathways involved cell proliferation (Cheng et al., 2012). In breast cancer tissues, ZNF24 inhibits angiogenesis by repressing transcription of the vascular endothelial growth factor (VEGF) gene (Harper et al., 2007; Jia et al., 2013). ZNF668 regulates MDM2 by disrupting the activity of MDM2 to stabilize p53 (Hu et al., 2011), thus having the opposite effect on p53 compared to ZNF280B (Gao et al., 2013).

1.1.5 Coordinated control of C2H2-ZNF protein activity and expression

The massive size of the C2H2-ZNF gene family and the abundance of highly conserved and repetitive elements among these genes promote the idea of their coordinated regulation. As transcription factors, C2H2-ZNF proteins influence the expression of genes involved in numerous vital biological processes such as development, differentiation, cell-cycle progression, and apoptosis (Jen and Wang, 2016; Razin et al., 2012). Therefore, simultaneous regulation of C2H2-ZNF genes during these processes could allow for rapid changes in gene expression programs that significantly influence cell activity and behavior. Several strategies have been identified that may be designed to coordinate the activity and expression of the C2H2-ZNF proteins. These include phosphorylation of the linker region during mitosis (Dovat, 2002; Rizkallah et al., 2011), miRNA binding to repetitive mRNA sequences (Huang et al., 2010), regulated mRNA decay (Neff et al., 2012), and control of poly(A) tail length (Figure 2) (Yang et al., 2011b). Each of these are discussed in further detail below.

Phosphorylation of the protein linker region during mitosis
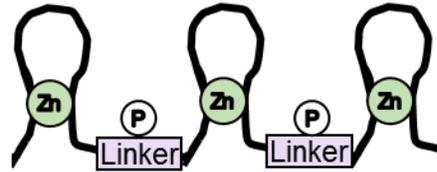
microRNA binding to conserved ORF sequences
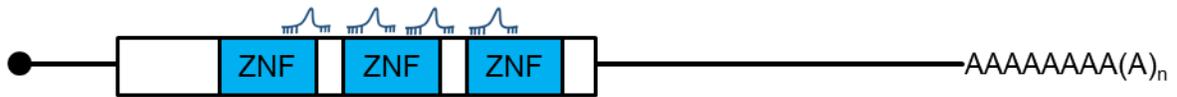
mRNA decay
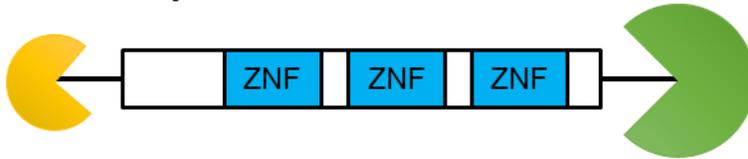
Short poly(A) tail length

Figure 2: Potential strategies to coordinate the expression and activity of the C2H2-ZNF proteins. See text for detail.

1.1.5.1 C2H2-ZNF proteins are phosphorylated during mitosis

During mitosis, the TGEKPY/F linker region of C2H2-ZNF proteins is phosphorylated to inhibit the DNA-binding ability of these proteins (Dovat, 2002). This phosphorylation event is highly synchronous and starts just after the nuclear envelope breaks down at mid-prophase. By the end of telophase (the final stage of the cell cycle), the phosphorylation is of the linker is completely reversed (Rizkallah et al., 2011). Thus, this linker region may have been conserved throughout evolution of the C2H2-ZNF genes to facilitate rapid and simultaneous inactivation of these transcription factors during mitotic progression.

1.1.5.2 MicroRNAs target C2H2-ZNF ORF sequences

The seed sequences of microRNAs bind to target mRNAs to induce translational silencing and deadenylation-mediated decay (Eulalio et al., 2008; Huntzinger and Izaurralde, 2011). Interestingly, many of the C2H2-ZNF mRNA open reading frames (ORFs) contain putative binding sites for miRNAs from four families (miR-23, miR-181, miR-188, and miR-199). These miRNA binding sites are often located near the highly conserved and repeat-rich zinc finger motif and linker sequences. Therefore, one miRNA can potentially target numerous sites in multiple C2H2-ZNF mRNAs to simultaneously downregulate their expression through translational repression and/or mRNA decay (Huang et al., 2010; Schnall-Levin et al., 2011).

1.1.5.3 C2H2-ZNF mRNAs are differentially stabilized in stem cells and differentiated cells

The amount of mRNA available for protein synthesis is influenced by the rates of both transcription and mRNA decay (Schoenberg and Maquat, 2012). In a global analysis of mRNA decay rates in human foreskin fibroblasts (HFFs) and genetically matched induced pluripotent stem (iPS) cells, the mRNAs encoding over 100 different C2H2-ZNF proteins were found to have longer half-lives in iPS cells (Neff et al., 2012). This suggests that there may be one or more common mechanisms that differentially regulate the decay of the C2H2-ZNF mRNAs in

stem cells versus differentiated cells. One such a mechanism could be miRNA-mediated decay

as the miRNAs that target C2H2-ZNF mRNAs (Schnall-Levin et al., 2011) are significantly less

abundant in iPS cells than HFFs (Neff et al., 2012).

1.1.5.4 Many C2H2-ZNF mRNAs have short poly(A) tails

A poly(A) tail is added to the 3' end of most mRNAs during transcriptional processing

and is involved in promoting nuclear export, stimulating translation, and regulating mRNA decay

(Millevoi and Vagner, 2010). Notably, multiple different C2H2-ZNF mRNAs exist with unusually

short poly(A) tails in HeLa and H9 embryonic stem cells (Yang et al., 2011b). Some C2H2-ZNF

mRNAs are primarily non-polyadenylated while others exist with both short and long poly(A)

tails and are classified as "bimorphic."

We were especially interested in determining the sequences and pathways involved in

regulating C2H2-ZNF mRNA poly(A) tail length. Conserved sequences in the C2H2-ZNF

mRNAs may recruit factors such as miRNAs and RNA binding proteins that contribute to the

short poly(A) tail. Alterations in poly(A) tail length could have impacts on splicing, nuclear

export, translation, and decay of these mRNAs to coordinate the post-transcriptional regulation

of C2H2-ZNF gene expression.

1.2.1 The mRNA poly(A) tail

The poly(A) tail is a non-templated stretch of adenosine residues added to the 3' end of

most mRNAs during processing. There are two steps to generate a poly(A) tail: cleavage at the

poly(A) site, followed by the addition of adenosine residues one at a time at the newly created 3'

end. In the first step, cleavage and polyadenylation factors recognize canonical sequences that

specify the 3' end and perform the cleavage reaction (Proudfoot, 2011). Then, poly(A)

polymerases (PAPs) work in conjunction with the nuclear poly(A) binding protein (PABPN1) to

add 150-250 adenosine residues (Eckmann et al., 2011). Another poly(A) binding protein,

8

ZC3H14 or Nab2 in yeast, also associates with the nascent poly(A) tail to control its length (Kelly et al., 2014). Both PABPN1 (Banerjee et al., 2013; Muniz et al., 2015) and ZC3H14/Nab2 influence the efficiency of splicing and export of the newly synthesized mRNA (Brockmann et al., 2012; Hurt et al., 2009; Soucek et al., 2016).

In the cytoplasm, the poly(A) tail is bound by cytoplasmic PABPs (PABPC1-4 in mammals) (Gray et al., 2015). The translation initiation factor, eIF4G interacts with PABPC1 to stimulate cap-dependent translation (Hinnebusch and Lorsch, 2012; Kahvejian et al., 2005). Poly(A) tails can be shortened via deadenylation, which generally triggers rapid mRNA decay (Chen and Shyu, 2011), but can also provide a mechanism for temporary translational silencing (Weill et al., 2012). Transcripts silenced in this manner can be re-activated through cytoplasmic polyadenylation pathways (Charlesworth et al., 2013). Thus, modulating the length of the poly(A) tail influences mRNA processing, localization, translation efficiency, and mRNA decay (Weill et al., 2012; Zhang et al., 2010). This section will discuss the mechanisms and roles of poly(A) tail length regulation in eukaryotic cells.

1.2.2 Canonical polyadenylation

During pre-mRNA 3' end processing, the cleavage/polyadenylation specificity factor (CPSF) recognizes and binds to the polyadenylation signal (PAS), generally AAUAAA or a similar variant (Proudfoot, 2011). Additional sequence elements, including an upstream element and a U/GU-rich downstream element are also required for proper poly(A) site usage (Chen and Wilusz, 1998; Tian and Graber, 2012). RNA polymerase II (Pol II), Symplekin, CPSF, the c̲leavage s̲timulation f̲actor (CstF), cleavage factors I and II (CFI and CFII), and PAP form a complex that cleaves the transcript at a CA dinucleotide about 10-30 nt downstream of the AAUAAA poly(A) signal (Figure 3) (Shi and Manley, 2015). The poly(A) tail is then added in two phases. First, PAP slowly adds an initial poly(A) tail, but this enzyme is inefficient on its own (Wahle, 1991). As soon as the tail reaches a length of 10-12 residues, PABPN1 can associate,

and in conjunction with CPSF, stimulates rapid polyadenylation by PAP (Figure 3) (Bienroth et al., 1993). PABPN1 restricts the final length of the poly(A) tail to ~250 nt (Wahle, 1995). This length disrupts the interaction between PAP and CPSF to terminate processive polyadenylation (Kuhn et al., 2009). Additional factors involved in regulating poly(A) tail length include nucleophosmin (NPM1) and ZC3H14/Nab2. NPM1 is deposited on the 3' end of transcripts following polyadenylation (Palaniswamy et al., 2006). Knockdown of NPM1 causes hyperadenylation, suggesting that this protein is involved in controlling poly(A) tail length (Sagawa et al., 2011). ZC3H14/Nab2 also keeps poly(A) tail length in check. The mechanism for this regulation is not completely understood (Fasken and Corbett, 2016), but ZC3H14/Nab2 may interact with factors involved in splicing, cleavage, and polyadenylation to restrict PAP activity or it may recruit a ribonuclease to trim the poly(A) tail to the correct length (Soucek et al., 2012).

1.2.3 Polyadenylation and splicing

Splicing is the co-transcriptional processing step that removes introns and is closely coupled to 3' end formation and polyadenylation. Multiple splicing proteins interact with 3' end formation factors to enhance the efficiency of cleavage and polyadenylation (Kyburz et al., 2006; Millevoi and Vagner, 2010; Millevoi et al., 2006). Splicing factors are also involved in defining the terminal exon and poly(A) site, which influences mRNA 3' UTR length and post-transcriptional regulation (Di Giammartino et al., 2011; Proudfoot et al., 2002). In addition, both polyadenylation and splicing facilitate the release of a processed mRNA from Pol II (Rigo and Martinson, 2009). Notably, the 3'-end processing machinery may stimulate splicing, suggesting a reciprocal relationship between these two processes (Kaida, 2016; Misra and Green, 2016). Multiple lines of evidence suggest that regulation of poly(A) tail length and splicing are linked. Splicing factors including U2 snRNP auxiliary factor (U2AF) (Gu, 2003) and the U1 snRNP-A protein (U1A) (Lutz et al., 1996) influence poly(A) tail length. PABPN1 and PAP promote
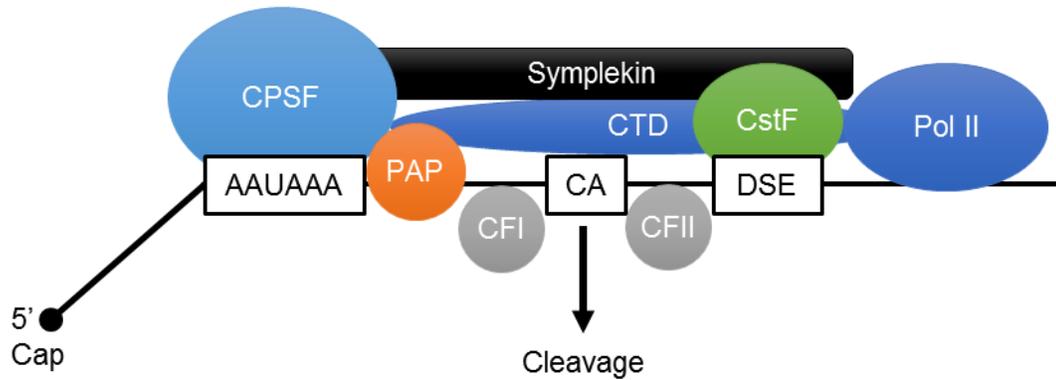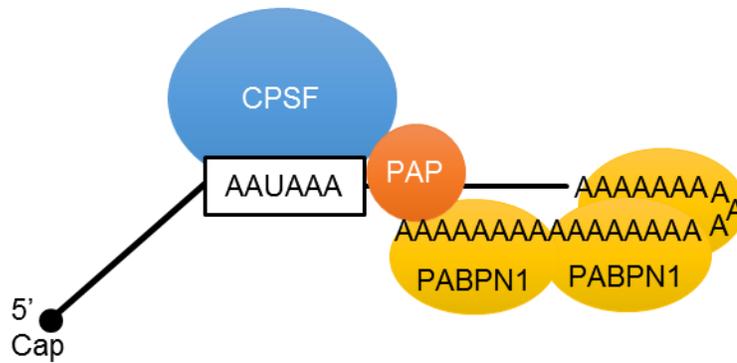
Figure 3: Complexes involved in cleavage and polyadenylation. CPSF binds the poly(A) site (AAUAAA) while CstF recognizes the GU/U-rich downstream element (DSE). These two factors, along with PAP, Symplekin, the C-terminal domain (CTD) of Pol II, CFI and CFII from the complex that cleaves the poly(A) site (often at a CA). Following the cleavage event, PAP adds the poly(A) tail. CSPF and PABPN1 enhance the processivity of poly(A) tail elongation and define its length.

splicing of terminal introns through a mechanism dependent on polyadenylation (Muniz et al.,

2015). In addition, ZC3H14/Nab interacts with the spliceosome to potentially link regulation of

splicing and polyadenylation (Soucek et al., 2016).

1.2.4 The role of the poly(A) tail in mRNA quality control

As the final stage of pre-mRNA processing, 3' end cleavage and polyadenylation may be

involved in quality control to prevent aberrant mRNAs from being exported to the cytoplasm and

translated. Proteins involved in regulating poly(A) tail length are also involved in mRNA quality

control pathways. For instance, PABPN1 and PAP will hyperadenylate nuclear mRNAs that

have not been properly spliced to stimulate their degradation by the nuclear exosome (Bresson

and Conrad, 2013; Bresson et al., 2015). ZC3H14/Nab2 interacts with splicing factors, regulates

poly(A) tail length, and is involved in nuclear export (Soucek et al., 2016) and therefore may

also play a key role in mRNA quality control (Fasken and Corbett, 2016). In yeast, the T̲rf4/5-

A̲ir1/2-M̲tr4 p̲olyadenlyation (TRAMP) complex is a key component of nuclear surveillance and

adds oligo(A) tails to non-coding and coding RNAs to promote degradation by the nuclear

exosome (Schmidt and Butler, 2013). In humans, the n̲uclear e̲x̲osome t̲argeting (NEXT)

complex mediates nuclear surveillance of RNA Pol II transcripts and is associated with the 3'

end processing complex (Di Giammartino and Manley, 2014). NEXT can recruit the exosome to

RNAs with unprotected 3' ends (Lubas et al., 2015), which could lead to degradation of mRNAs

that were not properly processed and polyadenylated (Goss and Kleiman, 2013; Liu et al.,

2014b). Thus, poly(A) tails that are excessively long or that are too short may mark aberrantly

processed mRNAs for degradation in the nucleus. However, some mRNAs are stable despite

having a short poly(A) tail (Peng et al., 2005; Subtelny et al., 2014). The 3' ends of these

mRNAs may be protected by a protein or other *trans*-acting factor or by forming an RNA

structure that stalls the nuclear exosome (Conrad, 2014).

1.2.5 The poly(A) tail and nuclear export

The poly(A) tail is an important identifier of mRNAs that are ready for nuclear export (Fuke and Ohno, 2008). Interestingly, proteins involved in regulating poly(A) tail length including PABPN1 (Bear et al., 2003), NPM1 (Crockett et al., 2004), and ZC3H14 (Hurt et al., 2009) also interact with nuclear export factors, suggesting that poly(A) tails may need to be a particular length for mRNAs to be exported. This is further supported by evidence that hyperadenylated mRNAs are retained in the nucleus, either as a quality control mechanism (Bresson and Conrad, 2013) or during disease or stress conditions (Kumar and Glaunsinger, 2010). While reduced poly(A) tail length may prevent binding of proteins necessary for nuclear export, some mRNAs with short poly(A) tails are still efficiently exported (Peng et al., 2005). These mRNAs may contain other sequence elements that interact with *trans*-acting factors to facilitate nuclear export, independent of the poly(A) tail.

1.2.6 The poly(A) tail influences translation

Once an mRNA is in the cytoplasm, the poly(A) tail continues to play an important part in post-transcriptional regulation. The cytoplasmic version of the poly(A)-binding protein (PABPC1) binds to the poly(A) tail and stabilizes the translation initiation complex composed of the cap-binding protein eIF4E, initiation factor eIF4G, the eIF4A helicase, and eIF3, which recruits the ribosome (Figure 4) (Charlesworth et al., 2013). This mRNA-protein complex forms a closed loop that enhances translation initiation and may facilitate recycling of ribosomes for continued rounds of translation (Burgess and Gray, 2010; Kahvejian et al., 2005; Wells et al., 1998). This structure may also resist mRNA decay (Weill et al., 2012). Once an mRNA is no longer needed for protein synthesis, translation termination can be coupled with deadenylation and mRNA decay (Roy and Jacobson, 2013).
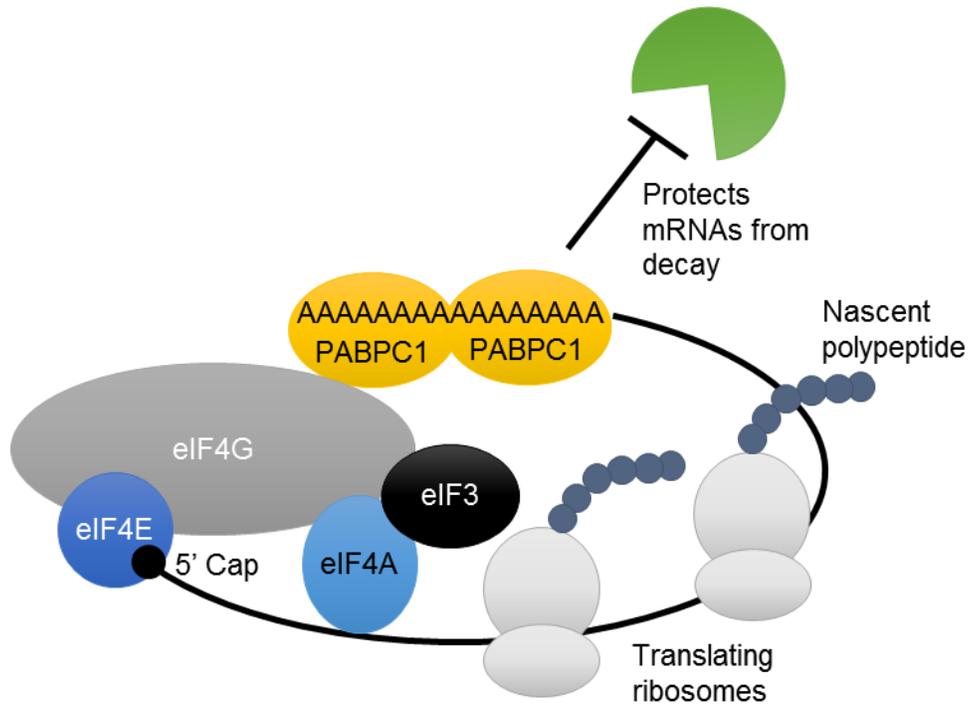
Figure 4: Closed mRNA loop enhances translation. The translation initiation complex composed of eIF4E, eIF4G, and eIF4A interact with the 5' cap and PABPC1 bound to the poly(A) tail to form a closed loop structure. This structure promotes the recruitment and recycling of ribosomes to enhance translation. The closed loop can also protect the mRNA from the decay machinery.

1.2.7 Deadenylation and mRNA decay

For most eukaryotic mRNAs, removal of the poly(A) tail initiates mRNA decay and is generally performed by the deadenylases CCR4-NOT and PAN2-PAN3 (Doidge et al., 2012). Deadenylation is a biphasic process (Figure 5). First, PAN2-PAN3 is recruited by interacting with PABPC1 and slowly shortens the poly(A) tail until it is about 110 nt. Next, CCR4-NOT further degrades the poly(A) tail until only 10 or so adenine residues remain (Chen and Shyu, 2011; Yamashita et al., 2005). Once the poly(A) tail is shortened to the critical length, decapping enzymes remove the 5' cap and the mRNA can be decayed by the 5' to 3' exonuclease XRN1 (Braun et al., 2012). Alternatively, the exosome can degrade the mRNA from 3' to 5' after the poly(A) tail has been removed (Schmid and Jensen, 2008). Thus mRNAs with short poly(A) tails generally exist for only a very short period of time.

The poly(A)-specific ribonuclease (PARN), is another important deadenylase that targets mRNAs in both the nucleus and the cytoplasm. There are several other less studied deadenylase enzymes including Nocturnin and related proteins (Godwin et al., 2013). These deadenylases may target specific transcripts, or be activated during certain cellular conditions.

1.2.8 The significance of poly(A) tail length

In mammalian somatic cells at steady-state conditions, median mRNA poly(A) tail length ranges from 50-100 nt (Chang et al., 2014; Subtelny et al., 2014). Poly(A) tail length is controlled at multiple stages during the mRNA lifecycle. In the nucleus, the initial poly(A) tail length is determined by nuclear polyadenylation carried out by PAP, which can be influenced by additional factors that can stimulate or inhibit its activity (Laishram, 2014). In addition, non-canonical PAPs can target specific mRNAs to control poly(A) tail length in place of canonical PAPs and may modulate gene expression profiles (Kandala et al., 2016; Kappel et al., 2015; Kuchta et al., 2016; Mohan et al., 2015). After nuclear polyadenylation is complete, poly(A) tail length can be reduced by deadenylation in the nucleus or the cytoplasm. Deadenylation can
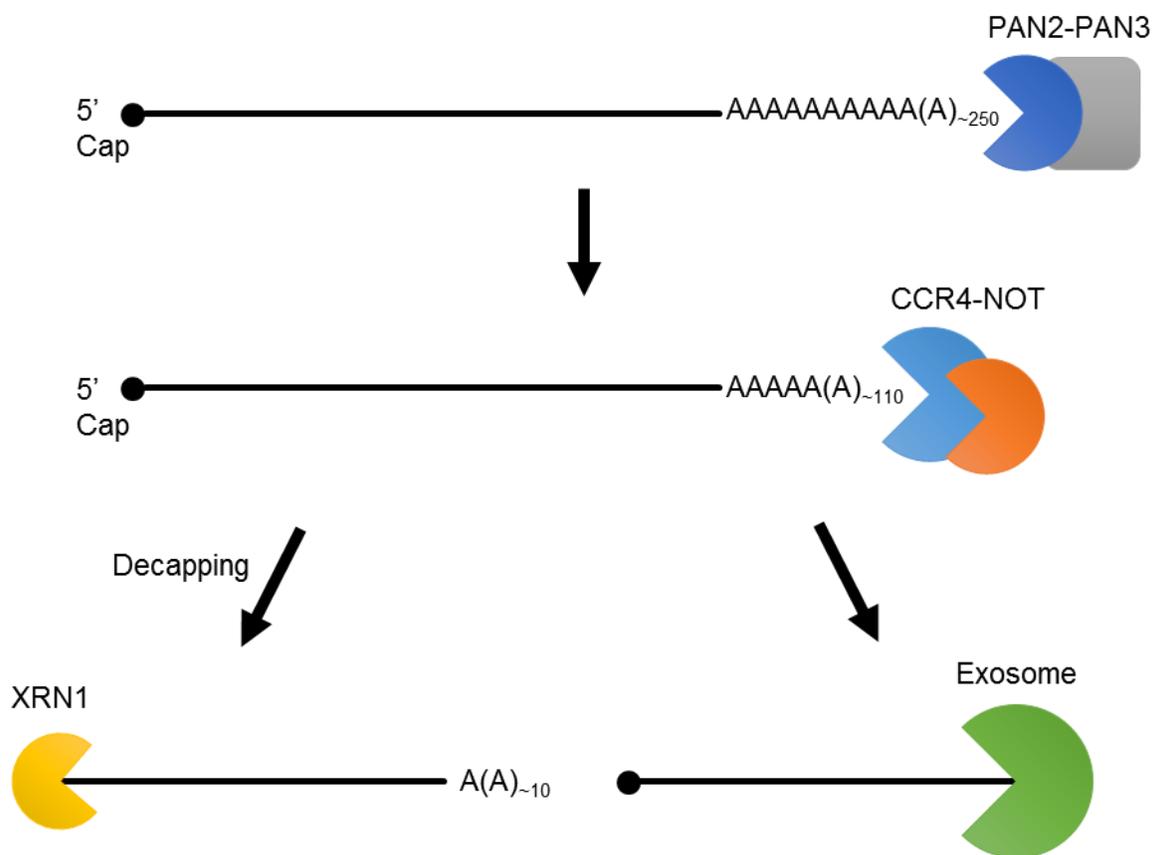
Figure 5: Deadenylation initiates mRNA decay. During the first phase of deadenylation, PAN2-PAN3 shorten the poly(A) tail. Then CCR4-NOT remove most of the remaining poly(A) tail, which can stimulate decapping and subsequent decay by XRN1 in the 5' to 3' direction. Alternatively, the RNA exosome can target deadenylated mRNAs for 3' to 5' decay.

instigate mRNA decay or temporarily repress translation (Yan, 2014). Cytoplasmic

polyadenylation can then reactivate translationally silent mRNAs by increasing mRNA poly(A)

tail length (Charlesworth et al., 2013).

While increases in poly(A) tail length enhance translation in oocytes (Reyes and Ross,

2016), embryonic cells (Subtelny et al., 2014), and neurons (Udagawa et al., 2012), this does

not appear to be the case for the majority of mRNAs in somatic cells. Recent studies of global

poly(A) tail length revealed that there is no correlation between poly(A) tail length and rates of

translation (Chang et al., 2014; Subtelny et al., 2014). This suggests that beyond the minimal

length necessary for PABPC to bind the poly(A) tail and facilitate recruitment of translation

initiation factors and prevent decay, poly(A) tail length has minimal impact on translation in

somatic cells. However, there are exceptions for certain sets of mRNAs (Burns et al., 2011;

Yamagishi et al., 2016) or under some cellular conditions (Kojima et al., 2012; Novoa et al.,

2010; Park et al., 2016). More studies are needed to clarify the overall impact poly(A) tail length

has on translation.

Transcripts with poly(A) tails that are too short for PABPs to interact with (<12 nt) are

normally rapidly targeted for decay (Gorgoni and Gray, 2004), but some mRNAs (e.g. those

encoding ribosomal proteins) are stable and translated despite having short poly(A) tails

(Subtelny et al., 2014). Such mRNAs must be protected from mRNA decay, potentially mediated

by specific RNA-binding proteins or unique 3' end structures that inhibit the degradation

machinery (Marzluff, 2012). In addition, cellular conditions could potentially inhibit efficient

decay of mRNAs with short poly(A) tails. Some mRNAs with short poly(A) tails may be

translationally silent, and are reactivated by cytoplasmic polyadenylation (Norbury, 2013).

Alternatively, mRNAs with short poly(A) tails may interact with proteins not yet identified that

promote efficient translation, independent of poly(A) tail length (Peng and Schoenberg, 2005).

1.2.9 Restriction of poly(A) tail length in the nucleus

It is assumed that the vast majority of mRNAs are polyadenylated to a default length of ~150-250 nt, but there are important examples of transcripts that have initial poly(A) tail lengths that are much shorter (Gu et al., 1999). Polyadenylation can also be inhibited during certain cellular conditions such as the cell cycle (Bond et al., 2000) and heat shock (Di Giammartino et al., 2013). In addition, deadenylation can reduce poly(A) tail length of nuclear mRNAs after dissociation of the polyadenylation machinery. While the consequences of poly(A) tail length restriction in the nucleus are not completely understood, this could impact downstream processes such as nuclear mRNA decay, export, and translation.

1.2.9.1 Poly(A) Limiting Elements

Two sequences elements have been identified that are able to restrict the length of the nascent poly(A) tail–AREs and PLEs. Tristetraprolin (TTP), an RNA-binding protein that promotes rapid cytoplasmic mRNA decay of mRNAs containing AU-rich elements (AREs) (Brooks and Blackshear, 2013), can also prevent polyadenylation in the nucleus; however, the biological significance of this regulation is not known (Su et al., 2012).

Poly(A) limiting elements (PLEs) are pyrimidine-rich sequences that inhibit the processive phase of polyadenylation to keep the poly(A) length below 20 nt. (Das Gupta et al., 1998; Gu et al., 1999; Rao, 1996; Schoenberg et al., 1989). The PLE restricts poly(A) tail length through its interaction with U2 snRNP auxiliary factor (U2AF65), a splicing factor that is predicted to recruit an additional protein to limit poly(A) tail length (Gu, 2003). It is not clear what effect the PLE has on function as mRNAs containing this element are stable and are efficiently translated (Peng and Schoenberg, 2005; Peng et al., 2005). It is possible that an unidentified *trans*-acting factor binds to the PLE to stabilize mRNAs and facilitate translation, essentially acting as a substitute for PABPC1. To date, the PLE has only been found in a small set of transcripts, therefore it is not known how many mRNAs are subject to this type of poly(A) tail

length control (Gu et al., 1999). C2H2-ZNF mRNAs may contain conserved sequences that behave like AREs or PLEs and recruit one or more factors that disrupt nuclear polyadenylation to restrict poly(A) tail length of these mRNAs.

1.2.9.2 Inhibition of polyadenylation

Restriction of PAP activity reduces mRNA poly(A) tail length to potentially downregulate the expression of certain genes. For example during the cell cycle, PAP is hyperphosphorylated by cyclins to reduce its activity and contribute to reduced poly(A) tail and protein synthesis during mitosis (Bond et al., 2000; Colgan et al., 1996). During heat shock, PAP is modified by the poly(ADP-ribose) polymerase 1 (PARP1), which adds poly(ADP-ribose) to target proteins in a reaction called PARylation. This modification reduces PAP's ability to bind and polyadenylate mRNAs that encode proteins that are not required for the heat shock response (Di Giammartino et al., 2013).

1.2.9.3 Deadenylation in the nucleus

During DNA-damaging conditions, the PARN deadenylase forms a complex with CstF to impede 3' end processing and shorten the nascent poly(A) tail (Cevher et al., 2010). This may be a mechanism to prevent further processing, export, and translation of certain mRNAs allowing the cell to prioritize expression of genes needed in response to DNA damage (Zhang et al., 2010). While other deadenylases can be found in the nucleus, it remains to be determined if they are also involved in removing poly(A) tails of nuclear mRNAs to potentially modulate gene expression (Yan, 2014). Nuclear deadenylases could target C2H2-ZNF mRNAs, either through direct recruitment to one or more sequence elements or mediated by a *trans*-acting factor, to shorten the poly(A) tails.

1.2.10 Cytoplasmic events controlling poly(A) tail length

1.2.10.1 Sequences and factors that promote cytoplasmic deadenylation

   Removal of the poly(A) tail in the cytoplasm is modulated by specific mRNA sequence

elements which recruit *trans*-acting factors. The rate of deadenylation strongly influences the

average poly(A) tail length of a specific mRNA as rapid deadenylation will result in an mRNA

population with a shorter median poly(A) tail length. Although the end result of deadenylation is

often mRNA decay, this process can also lead to translational silencing (Norbury, 2013).

PABPC1 binding to the poly(A) tail is disrupted by deadenylation, which interferes with the

closed loop structure that promotes translation (Roy and Jacobson, 2013) and leaves the 3' end

of the mRNA unprotected from the degradation machinery (Yan, 2014). Sequences that

influence deadenylation rates include AU-rich elements (AREs) (Barreau, 2005; Fabian et al.,

2013; Gherzi et al., 2004), GU-rich elements (GREs) (Moraes et al., 2006; Rattenbacher et al.,

2010), Pumilio protein binding sites (Miller and Olivas, 2011; Weidmann et al., 2014), and

miRNA seed sequences (Chekulaeva et al., 2011; Fabian et al., 2011; Huntzinger and

Izaurralde, 2011; Zhang et al., 2015). RNA-binding proteins and miRNAs bind to these sites and

directly interact with deadenylases, including CCR4-NOT and PARN, to enhance deadenylation

of targeted mRNAs (Weill et al., 2012). Multiple different factors can bind overlapping mRNA

sites or to the same RNA-binding protein to synergistically or antagonistically influence poly(A)

tail length (Ciafrè and Galardi, 2013; Dan et al., 2015; Zhang et al., 2015). Clearly, numerous

sequences and factors are involved in controlling poly(A) tail length and can ultimately influence

rates of translation and mRNA decay.

   As previously discussed, the ORFs of numerous C2H2-ZNF mRNAs contain multiple

putative miRNA binding sites (Huang et al., 2010; Schnall-Levin et al., 2011), which could be

involved in regulating poly(A) tail length of these mRNAs. It is also possible that RNA-binding

proteins could also target C2H2-ZNF mRNAs and recruit deadenylases. If C2H2-ZNF mRNA

poly(A) tail length is altered by cytoplasmic deadenylation, these mRNAs would have to possess

a mechanism to prevent rapid mRNA decay. This could be mediated by an RNA-binding protein or an mRNA structural element that inhibits the decay machinery.

1.2.10.2 Cytoplasmic polyadenylation

As noted above, not all polyadenylation occurs in the nucleus. In fact, there are a number of poly(A) polymerases that can extend the poly(A) tail in the cytoplasm. For example, in transcriptionally silent oocytes, cytoplasmic polyadenylation activates translation of specific mRNAs during meiotic progression (Piqué et al., 2008). These mRNAs undergo canonical polyadenylation in the nucleus, but the poly(A) tail is removed by deadenlyases in the cytoplasm to keep these mRNAs dormant (Weill et al., 2012). In oocytes, the cytoplasmic polyadenylation element (CPE) is a UA-rich sequence or a U-rich involved in regulating poly(A) tail length during oocyte maturation (Fox et al., 1989; Simon and Richter, 1994). The CPE is targeted by the cytoplasmic polyadenylation element binding protein 1 (CPEB1), the major factor involved in regulating cytoplasmic polyadenylation. CPEB1 forms a complex with Symplekin, CPSF, PARN, and the GLD2/PAPD4 non-canonical poly(A) polymerase. PARN continuously removes A-residues, keeping the poly(A) tail short. To stimulate translation of specific mRNAs, CPEB1 is activated by phosphorylation, which dissociates PARN from the mRNA. This allows GLD2/PAPD4 to elongate the poly(A) tail (Kim and Richter, 2006).

Only a few examples of cytoplasmic polyadenylation have been clearly demonstrated in somatic cells. In neurons, cytoplasmic polyadenylation is regulated by CPEB to stimulate localized translation at synapses, which is crucial for synaptic plasticity, learning, and memory (Darnell and Richter, 2012). During mitosis, CPEB1 and CPEB4 regulate poly(A) tail length to control translation and stimulate entry into M-phase and cell proliferation (Novoa et al., 2010). Circadian rhythm is also linked to cyclical fluctuations in poly(A) tail length, which may temporally regulate expression of specific proteins in mammals. While is some instances, increases in poly(A) tail length is attributed to nuclear polyadenylation of newly transcribed

mRNAs, other transcripts appear to be polyadenylated in the cytoplasm by CPEB1 and CPEB2 (Kojima et al., 2012). More recently, RNA-binding protein QKI-7 promotes cytoplasmic polyadenylation by recruiting GLD2/PAPD4 to target mRNAs, indicating that non-CPEB proteins can also increase poly(A) tail length to stimulate translation (Yamagishi et al., 2016). Continued studies are needed to determine the full extent of cytoplasmic polyadenylation in somatic cells. Such investigations could contribute to new insights on how remodeling poly(A) length influences key biological processes in different types of cells and in response to various cellular conditions. For example, the C2H2-ZNF mRNAs may have short poly(A) tails to temporarily silence translation. Specific cellular conditions may then stimulate cytoplasmic polyadenylation to activate synthesis of the C2H2-ZNF proteins.

1.3.1 Nuclear export of mRNAs

In eukaryotes, the nuclear membrane physically separates transcription and mRNA processing from the translation machinery. Nuclear pore complexes (NPCs), regulate the transfer of RNAs and proteins from the nucleus to the cytoplasm. The transcription/export (TREX) complex is recruited to nascent mRNA by proteins involved in transcription and processing steps including capping, splicing, and 3' end formation (Heath et al., 2016). The TREX component ALY/REF interacts with nuclear export factors that dock the mRNA on the NPC where the mRNA is translocated to the cytoplasm (Katahira, 2015). This coordination of pre-mRNA processing and nuclear export helps ensure that only properly transcribed and processed mRNAs are exported to the cytoplasm for translation, while incompletely processed or aberrant mRNAs are retained in the nucleus and degraded (Eberle and Visa, 2014).

Regulation of nuclear export can also influence gene expression. This can be mediated by specific sequence elements and *trans*-acting factors that can affect the rate of mRNA export (Wickramasinghe and Laskey, 2015). Additionally, non-membranous subnuclear compartments can concentrate proteins and other factors that facilitate nuclear retention while mRNAs

undergo further processing or as a storage site (Bahar Halpern et al., 2015; Sleeman and Trinkle-Mulcahy, 2014). Intron retention can inhibit nuclear export as a potential mechanism to regulate gene expression (Wong et al., 2016). During the course of this work, we discovered that several C2H2-ZNF transcripts are abundant in the nucleus (Chapter 3, Section 3.6) and that the ZNF12 reporter mRNA accumulates within nuclear foci (Chapter 3, Section 3.10). Therefore, here we provide an overview of mRNA nuclear export and discuss the impact this process has on post-transcriptional regulation of gene expression.

1.3.2 Overview of the mRNA nuclear export pathway

An export-competent mRNP is formed through recruitment of the TREX complex, which consists of factors involved in transcription, pre-mRNA processing, and nuclear export (Heath et al., 2016; Katahira, 2012). Components of the TREX complex including the THO complex, RNA-helicase UAP56, and ALY/REF are recruited to mRNAs through interactions with the nuclear cap-binding complex and the exon junction complex (EJC) formed during splicing (Cheng et al., 2006; Masuda, 2005). Efficient assembly of the active TREX complex is also associated with 3' end formation (Katahira, 2012). The nascent mRNA is then transferred to ALY/REF, which recruits the nuclear export factor 1 (NXF1 also known as TAP) and its co-factor NXT1/p15 to the export-ready messenger ribonucleoprotein complex (mRNP) (Heath et al., 2016). At the NPC, the NXF1-NXT1/TAP-p15 heterodimer interacts with the FG-nucleoporins to dock the mRNP (Katahira, 2015). Movement through the NPC is not completely understood, but is thought to rely on an ATP-dependent mechanism that facilitates remodeling of the mRNP at the cytoplasmic face to "ratchet" the mRNP through the pore (Stewart, 2007). In metazoans, the DEAD-box RNA helicase DDX19 along with co-factors GLE1 and inositol hexaphosphate 6 (IP$_6$) enables the remodeling of the mRNP that leads to export and release through the NPC (Folkmann et al., 2011; Sloan et al., 2016). Although most mRNAs are exported by the NXF1-NXT1 pathway, some are transported to the cytoplasm via the chromosomal region

maintenance protein 1 (CRM1) pathway, the major receptor for protein export. These mRNAs require additional adaptor proteins to mediate export through this alternative pathway (Carmody and Wente, 2009; Delaleau and Borden, 2015; Natalizio and Wente, 2013). Thus, nuclear export is dependent on the assembly of an export competent mRNP formed during pre-mRNA processing.

1.3.3 Implications of regulated mRNA export

The coupling of transcription and pre-mRNA processing with recruitment of nuclear export factors enhances the rate of mRNA transport to the cytoplasm (Oeffinger and Zenklusen, 2012; Tilgner et al., 2012). However, increasing evidence indicates that certain mRNAs accumulate primarily in the nucleus (Bahar Halpern et al., 2015; Barthelson et al., 2007; Djebali et al., 2012; Gondran et al., 1999; Prasanth et al., 2005; Solnestam et al., 2012). This can be achieved either through rapid degradation in the cytoplasm or through active retention of transcripts in the nucleus. Nuclear retention of aberrant or unprocessed mRNAs can ensure that only correctly transcribed and completely processed transcripts are exported to the cytoplasm for translation (Bonnet and Palancade, 2014; Eberle and Visa, 2014). Selective mRNA export of "normal" mRNAs can also regulate gene expression (Wickramasinghe and Laskey, 2015). In addition, nuclear retention may help overcome highly variable or "bursty" transcription to reduce variability in cytoplasmic mRNA abundances (Bahar Halpern et al., 2015; Battich et al., 2015). The implications of either promoting or inhibiting nuclear export are further discussed below.

1.3.3.1 Quality control and mRNA decay in the nucleus

Quality control measures in the nucleus prevent export and subsequent translation of aberrant mRNAs, which could lead to the synthesis of detrimental protein products. Nuclear quality control mechanisms recognize defects in the various stages of pre-mRNA processing and mRNP assembly to target defective transcripts for decay. The exonuclease DXO degrades

pre-mRNAs that have defective caps (Jiao et al., 2013). In yeast, the NPC-associated proteins

including Mlp1/2 and Pml39p can retain unspliced transcripts (Eberle and Visa, 2014). In

mammalian cells, the translocated promoter region (TPR) nuclear basket protein may be

involved in a similar pathway and prevent the export of unspliced mRNAs through NXF1-

NXT1/TAP-p15-mediated export (Coyle et al., 2011; Rajanala and Nandicoori, 2012). As

previously discussed in Section 1.2.4, 3' end processing and polyadenylation also contribute to

nuclear quality control. Improperly processed mRNAs are generally targeted for decay by the

nuclear exosome (Chlebowski et al., 2013) or the nuclear 5' to 3' exonuclease XRN2

(Nagarajan et al., 2013).

      The nuclear exosome can also degrade completely processed and polyadenylated

mRNAs. In the recently identified poly(A) tail exosome targeting (PAXT) pathway, PABPN1

directly interacts factors that recruit the nuclear exosome, leading to the decay of mature

mRNAs that have not been efficiently exported to the cytoplasm. This degradation pathway may

prevent unnecessary protein synthesis resulting from excessive transcription (Meola et al.,

2016). Notably, transcripts that stably accumulate in the nucleus must be able to avoid

PABPN1-mediated decay. Perhaps the C2H2-ZNF mRNAs have short poly(A) tails to prevent

nuclear degradation by the PAXT pathway.

1.3.3.2 Regulation of gene expression

      Selective export of mRNAs can regulate the expression of genes involved in key

biological processes including maintenance of pluripotency (Wang et al., 2013b), cell

differentiation (Tran et al., 2014), cell proliferation (Culjkovic et al., 2006), gene expression

(Wickramasinghe et al., 2014), cancer progression (Culjkovic-Kraljacic et al., 2012),

hematopoiesis (Mancini et al., 2010), and heat shock (Katahira et al., 2009). Enhanced export of

mRNAs can facilitate rapid synthesis of proteins required in response to changing cellular

conditions, while inhibited export can temporarily store transcripts that are not currently required

for future export and translation (Wickramasinghe and Laskey, 2015). The mechanisms involved in regulating nuclear export of mRNAs are not completely understood, but generally involve specific adaptor proteins that interact with export receptors (Delaleau and Borden, 2015) or NPCs (Bonnet and Palancade, 2014). These adaptor proteins can recognize specific sequence elements or other proteins that are deposited on mRNAs during splicing and other processing events (Katahira, 2015; Natalizio and Wente, 2013). Therefore, the ability of adaptor proteins to recognize and bind target mRNAs can influence nuclear export rates and alter gene expression in response to cellular stimuli. The nuclear export of the C2H2-ZNF mRNAs may be highly controlled to alter the rates of their protein synthesis during specific cellular conditions.

1.3.3.3 Nuclear retention may buffer transcription bursts

Transcription in mammalian cells frequently occurs in bursts, which can lead to increased variability in mRNA abundance. Recent evidence suggests that mRNA nuclear retention can reduce transcript variability in the cytoplasm to potentially maintain cell homeostasis despite bursts in transcription (Bahar Halpern et al., 2015; Battich et al., 2015). RNA sequencing of nuclear and cytoplasmic fractions from MIN6 mouse pancreatic beta cells and mouse liver tissue revealed multiple spliced and polyadenylated mRNAs are enriched in the nucleus. Analysis of transcription, nuclear export, and mRNA decay dynamics of two of these mRNAs, Mlxipl and Nlrp6, indicated that these mRNAs are not rapidly degraded in the cytoplasm following export, but are specifically retained in the nucleus to buffer transcriptional bursts (Bahar Halpern et al., 2015). Similarly, in HeLa cells, nuclear compartmentalization of mRNAs reduces transcriptional noise in the cytoplasm to prevent uncontrolled fluctuations in gene expression (Battich et al., 2015).

1.3.4 Regulation of nuclear export

1.3.4.1 Splicing promotes nuclear export

Splicing is closely linked with export efficiency of mRNAs as the EJC, SR splicing

factors, and other components of the splicing machinery recruit crucial components of the

export pathway to the mRNP (Björk et al., 2015; Luo and Reed, 1999; Shen, 2009). The splicing

process can even overpower the nuclear retention element of the β-globin mRNA to promote its

export (Akef et al., 2015).

1.3.4.2 Subnuclear bodies have roles in mRNA processing and storage

Specific non-membranous compartments in the nucleus or "subnuclear bodies" are

composed of proteins and RNAs involved in regulating numerous nuclear processes including

ribosomal and small nuclear RNA biogenesis, transcription, non-coding RNA and mRNA

processing, stress response, and nuclear retention (Morimoto and Boerkoel, 2013; Sleeman

and Trinkle-Mulcahy, 2014). Subnuclear bodies involved in mRNA processing and storage

include nuclear speckles and paraspeckles. It is possible that the nuclear-retained C2H2-ZNF

mRNAs could be located in these subnuclear bodies.

Nuclear speckles are dynamic structures often located near active transcription sites and

are involved in pre-mRNA splicing and 3' end processing (Spector and Lamond, 2011). Key

components of nuclear speckles include the long-noncoding RNA MALAT1, serine/arginine (SR)

splicing factors, components of the EJC, cleavage and polyadenylation factors, nuclear export

factors, and polyadenylated RNA (Huang, 1994; Hutchinson et al., 2007; Saitoh, 2004).

MALAT1 recruits splicing factors to actively transcribed chromatin sites and regulates alternative

splicing (Bernard et al., 2010; Tripathi et al., 2010). Interestingly, proteins that bind N6-

methyladenosine (m6A)-modified mRNAs are also found in nuclear speckles and regulate

splicing (Xiao et al., 2016) and nuclear export (Zheng et al., 2013). Many of the other proteins

and factors found in nuclear speckles also impact splicing, 3' end formation, and nuclear export

to regulate mRNA processing and post-transcriptionally modulate gene expression (Morimoto and Boerkoel, 2013). Notably, spliced and polyadenylated nuclear mRNAs Mlxipl and Nlrp6 (Section 1.3.3.3) co-localize with nuclear speckles, suggesting that these sites may also be involved in storing completely processed mRNAs in the nucleus (Bahar Halpern et al., 2015).

Paraspeckles are often found near speckles and are involved in regulating multiple RNA processing steps and nuclear retention of edited RNAs (Fox and Lamond, 2010). Components of paraspeckles include the long non-coding RNA NEAT1, *Drosophila* Behavior/Human Splicing (DBHS) proteins, and other RNA-binding proteins involved in RNA processing (Morimoto and Boerkoel, 2013). NEAT1 is required for paraspeckle formation and organization (Clemson et al., 2009; Sasaki et al., 2009). The DBHS proteins p54/NRB and PSF are also essential for paraspeckle formation (Sasaki et al., 2009) and are involved in transcription, splicing (Kameoka et al., 2004), and 3' formation (Rosonina et al., 2005).

1.3.5 Mechanisms of mRNA nuclear retention

1.3.5.1 A-to-I editing

One interesting mechanism for nuclear retention involves A-to-I RNA editing. Editing by ADAR1 occurs in the nucleus on duplex RNA structures, such as those formed by inverted *Alu* repeats or other highly repetitive sequences, and can contribute to nuclear retention and gene silencing of edited RNAs (Chen and Carmichael, 2008; DeCerbo and Carmichael, 2005). In mice, the polyadenylated CTN-RNA has multiple A-to-I edited sites in the 3' UTR and is retained in the nucleus. Under stress, the CTN-RNA is post-transcriptionally cleaved to remove the edited sites and produce the mCAT2 (mouse cationic amino acid transporter 2) mRNA, which is then quickly exported to the cytoplasm to encode a protein necessary for nitric oxide production as part of the inflammatory response. This interesting mechanism for storing an mRNA as a nuclear precursor RNA allows for a rapid response to a cellular stimulus to accelerate protein synthesis (Prasanth et al., 2005). The mRNAs from numerous genes, including some C2H2-

ZNFs, also possess A-to-I edited sites in *Alu*-derived sequences that regulate nuclear retention

mediated by paraspeckle components to potentially influence gene expression (Chen and

Carmichael, 2009; Chen et al., 2008; Wang et al., 2013a).

1.3.5.2 Intron retention

Emerging evidence suggests that intron retention may be a mechanism to regulate gene

expression by altering nuclear export rates of mRNAs involved in significant biological events

including cancer (Dvinge and Bradley, 2015), erythropoiesis (Pimentel et al., 2016), neuronal

development (Yap et al., 2012) and activity (Mauger et al., 2016), and the cell cycle (Heyn et al.,

2015). Several mechanisms that alter splicing to modulate intron retention have been described.

For example, RNA Pol II stalling inhibits the recruitment of the splicing machinery. This leads to

intron retention that may prevent expression of transcripts that are not relevant to the cell type

or current stage of development (Braunschweig et al., 2014). In mouse cells, the polypyrimidine

tract-binding protein (Ptbp1) represses splicing of the 3' terminal intron in multiple pre-mRNAs

involved in neurogenesis. This leads to nuclear retention and degradation of these mRNAs to

regulate gene expression during different stages of neural development (Yap et al., 2012). The

nuclear poly(A) binding protein PABPN1 similarly promotes the retention of the 3' terminal intron

of its own mRNA, which is then retained in the nucleus and decayed by the nuclear exosome as

an auto-regulatory mechanism to control PABPN1 expression (Bergeron et al., 2015).

Intron retention does not always lead to nuclear mRNA decay, but can be involved in

temporally regulating the expression of specific genes by confining polyadenylated mRNAs to

the nucleus (Boutz et al., 2015; Mauger et al., 2016). In mouse neurons, retained introns are

excised from fully transcribed and polyadenylated nuclear mRNAs in response to neuronal

activity to allow rapid export to the cytoplasm followed by translation (Mauger et al., 2016).

Therefore, intron retention can be a mechanism to facilitate rapid changes in gene expression in

response to cell stimuli.

1.3.6 Alternative roles of nuclear mRNAs

Many of the RNAs localized to the nucleus are non-coding and have a variety of functions such as serving as protein scaffolds, regulating transcription, or pre-mRNA processing (Morris and Mattick, 2014). Interestingly, mRNAs may also have functional roles in the nucleus in addition to encoding proteins in the cytoplasm (Karapetyan et al., 2013; Kloc et al., 2011). The mRNAs from multiple different genes, including some C2H2-ZNF transcripts, interact with chromatin-associated proteins (Bonasio et al., 2014; Hendrickson et al., 2016; Zhao et al., 2010). While it is not yet clear if these interactions have a specific function, it is possible that nuclear-retained mRNAs could facilitate recruitment of proteins involved in activating or repressing transcription to particular chromatin sites to regulate gene expression. Additional studies are needed to investigate the possibility that certain nuclear retained mRNAs have a dual purpose as non-coding transcripts.

1.4 Rationale and Hypothesis

C2H2-ZNF mRNAs share homologous sequences that could allow coordinated post-transcriptional regulation of these genes with potential to significantly impact cell function. In support of this, previous studies have shown that C2H2-ZNF mRNAs (i) are over-represented among non-polyadenylated transcripts in highly proliferative cells, including stem and cancer cells (Yang et al., 2011b), (ii) are more stable in stem cells than in primary fibroblasts (Neff et al., 2012), and (iii) are targeted by miRNAs that recognize repeated elements within their open reading frames. Our goals in this study were to gain insight on possible connections among these observations regarding the C2H2-ZNF mRNAs, to better characterize the short poly(A) tail phenotype, and to determine the sequence element(s) responsible.

Hypothesis: Specific elements in the C2H2-ZNF mRNA ORF and/or 3' UTR determine poly(A) tail length to impact one or more aspects of post-transcriptional regulation of gene expression.

CHAPTER 2: MATERIALS AND METHODS

2.1 Cell culture

2.1.1 HeLa cells

HeLa S3 cells adapted for adherent culture were maintained in DMEM (4.5 g/L glucose) supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, and 0.37% sodium bicarbonate without antibiotics at 37°C in 5% CO_2. HeLa Tet-Off Advanced cells (Clontech) were maintained in DMEM supplemented with 10% tetracycline-free FBS (Clontech), 2 mM L-glutamine, 0.37% sodium bicarbonate, and 100 µg/mL of G418 to maintain expression of the tet-off transactivator. For all HeLa cells, media was changed every 3 days and cells were passaged when they neared confluency, typically every 5 to 7 days. To passage cells, the media was removed, cells were washed with phosphate buffered saline (PBS), then incubated in 0.25% trypsin/EDTA (Thermo Fisher Scientific-Hyclone) at 37°C until the cells detached from the plate. Complete media was added to the cells to quench the enzymatic activity of trypsin. For maintenance, cells were diluted ~10-fold for replating. HeLa cells were cryopreserved in 50% media, 40% FBS, and 10% dimethyl sulfoxide (DMSO), cooled slowly to -80°C in a freezing chamber, and then placed in liquid nitrogen for long-term storage.

2.1.2 iPS cells

Induced pluripotent stem (iPS) cells (System Biosciences Cat # SC101A-iPSC, Lot # 110415-01) were cultured on Matrigel (BD Biosciences) in mTesR1 media (STEMCELL Technologies) at 37°C in 5% CO_2. Media was changed daily. Differentiating cells were removed from the plates regularly by marking cells that no longer showed the iPS cell morphology, manually detaching the differentiated cells from the plate using a pipette tip, and then changing the media to remove the detached cells. Cells were passaged every 5-7 days using either Dispase (STEMCELL Technologies) or manual passage by visual selection of iPS cell colonies

with the proper phenotype. iPS cells were cryopreserved in mTesR1 media with 10% DMSO, cooled slowly to -80°C in a freezing chamber, and then placed in liquid nitrogen for long-term storage.

2.1.3 Human foreskin fibroblasts

Human foreskin fibroblasts (HFFs) (System Biosciences Cat # SC101A-HFF, Lot # 110509) were maintained in DMEM (4.5 g/L glucose) supplemented with 2 mM L-glutamine, 0.1 mM non-essential amino acids, and 10% FBS without antibiotics at 37°C, 5% CO_2. Media was changed every 3 to 4 days and cells were passaged at a 1:3 ratio every 7 to 10 days. To passage cells, the media was removed, the cells were washed twice with PBS, and then incubated in a 1:1 ratio of 0.25% trypsin/EDTA (Thermo Fisher Scientific-Hyclone) and PBS at 37°C until the cells were just starting to detach from the plate. Complete media was added to the cells to quench the enzymatic activity of trypsin. The cells were collected by centrifugation at 300 x g for 5 min, then resuspended in complete media for plating.

2.1.4 Transfections

Reporter plasmids were subjected to endotoxin removal (MiraCLEAN from Mirus Bio) and then were transfected into HeLa Tet-Off Advanced cells using the jetPRIME (Polyplus) transfection reagent according to manufacturer's instructions with the following changes: cells were transfected at the time of passage using 3 µL reagent/µg of plasmid. Cells were harvested for analysis 24-30 h after transfection.

2.1.5 Transduction and selection of stable DICER KD cell lines

DICER knockdown and control cell lines were made by Joe Russo by transducing HeLa cells with lentiviral particles containing the shRNA construct against DICER (Sigma-Aldrich: TRCN0000051261) or pLKO.1 empty vector as previously described (Stewart et al., 2003). To

select for stably transduced cells, 10 μg/mL puromycin was added to complete media 2 days after transduction. Pools of antibiotic resistant cells were maintained in 1 μg/mL puromycin for further experimentation.

2.2 Plasmids and cloning

2.2.1 Templates for in vitro transcription of polyadenylated control RNAs

Short regions derived from the ACT1, GAL1, MET3, and MET25 genes from *S. cerevisiae* were cloned into pGEM4 plasmids containing 15, 30, 45, and 149 adenosines downstream of the multiple cloning site, respectively (courtesy of Alan Godwin). *S. cerevisiae* mRNA (courtesy of the Stargell laboratory) was reverse transcribed into cDNA. The inserts were amplified using the primers listed in Appendix A Table 1 with restriction site sequences for *Bam*HI on the forward primer and *Sal*I on the reverse primer and cloned into the pGEM4 plasmids using standard molecular biology techniques. Plasmids were then purified using the PureLink MaxiPrep kit (Invitrogen).

2.2.2 ZNF12 and PPIA reporter plasmids

Luciferase reporter plasmids used the pTRE3G-BI-ZsGreen plasmid (Clontech) as a backbone. The β-globin/IgG chimeric intron was amplified from the pCI-Neo vector (Promega) by PCR using primers listed in Appendix A Table 2 with *Apa*I restriction sites and cloned intro the unique *Apa*I site of pTRE3G-BI-ZsGreen. The remainder of the cloning was performed using In-Fusion HD (Clontech) or NEBuilder HiFi (New England Biolabs) cloning kits to allow multiple fragments to be cloned simultaneously via ligation independent cloning. Renilla luciferase was amplified from pLightSwitch (SwitchGear Genomics), ZNF12 (NM_016265.3) and PPIA (NM_021130.4) ORF and 3' UTR inserts were amplified from cDNA, and the ZNF12 downstream region (NC_000007.14) was amplified from genomic DNA. Primers (see Appendix A Table 2) for each insert were designed with at least 15 bp extensions homologous to the

vector or adjacent insert sequences. For reporter constructs with the SV40 poly(A) signal left

intact, the pTRE3G-BI-ZsGreen plasmid was digested with *Not*I. To remove the SV40 region for

ZNF12 All and 3' UTR reporter constructs, plasmids were digested with *Bgl*II and *Tfi*I. The

fragments were mixed in recommended ratios and reactions were incubated at 50°C for 20-30

min before transformation into DH5α *E. coli* cells. Plasmids were then purified using the

PureLink MaxiPrep kit (Invitrogen).

2.3 In vitro transcription

The A15, A30, A45, and A149 plasmids were linearized using *Pvu*II restriction enzyme

and the empty pGEM4 plasmid was linearized with *Sma*I to make the A0 reporter. Then, 1 µg of

each plasmid was used as template for *in vitro* transcription reactions by adding 4 µL of 5X RNA

polymerase buffer, 2 µL of 5 mM NTPs, 1 µL of RiboLock RNase inhibitor (Thermo Fisher

Scientific), 2 µL of SP6 RNA polymerase (Thermo Fisher Scientific), and nuclease-free water up

to 20 µL total. Reactions were incubated overnight at 37°C and the resulting RNA products were

gel purified and extracted using phenol/chloroform. The purity and quantity of each reporter

RNA was assessed using the NanoDrop spectrophotometer (Thermo Fisher Scientific).

2.4 RNA isolation

To isolate RNA, TRIzol (Invitrogen) was adding directly to cells and the manufacturer's

protocol was followed with the following changes: after the addition of chloroform, the aqueous

phase was placed in a new tube and an additional phase separation using

phenol/chloroform/isoamyl alcohol (25:24:1) was performed prior to addition of isopropanol.

RNA from non-transfected cells was treated with 10 units of DNase I with RiboLock RNase

inhibitor (Thermo Fisher Scientific) in a 100 µL reaction at 37°C for 30 minutes. To remove

plasmid DNA from the transfected HeLa Tet-Off Advanced cell samples, the RNA was treated

with TURBO DNase (10 units/µg of RNA) with RNase inhibitor and *Dpn*I in 100 µL reactions at

37°C for 3 hours. After either DNase treatment, the RNA was purified using a phenol/chloroform/isoamyl alcohol phase separation step followed by ethanol precipitation. The purity and quantity of the RNA was assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific).

2.5 Oligo (dT) Fractionation

Oligo $(dT)_{25}$ magnetic beads (New England Biolabs) were used to separate polyadenylated RNAs from oligo- and non-adenylated RNAs. Prior to the selection, 100 µL of beads were washed in binding buffer (20 mM Tris-HCl, pH 7.5, 500 mM LiCl, 0.5% LiDS, 1 mM EDTA, 5 mM DTT) at room temperature for 2 min with agitation. A magnetic field was used to pull the beads to the side of the tube and the binding buffer was removed. Equimolar amounts (0.02 pmol) of each of the A0, A15, A45, A149 control RNAs were added to 10-25 µg of total RNA and the mixture was denatured for 2 min at 70°C in binding buffer. The washed beads were then mixed with RNA in binding buffer, and incubated for 10 min at room temperature with agitation. The magnetic field was used to isolate the RNA bound to the beads and the supernatant was removed and placed it in a separate tube for the unbound/oligo(A) fraction. The beads were washed in 100 µL of wash buffer (20 mM Tris-HCl, pH 7.5, 500 mM LiCl, 0.1% LiDS, 1 mM EDTA, 5 mM DTT) with 1 µL of Ribolock RNase inhibitor (Thermo Fisher Scientific). Using the magnetic field to isolate the RNA bound to the beads, the wash buffer supernatant was removed and added to the oligo(A) fraction. After repeating the wash step, 100 µL of low salt buffer (20 mM Tris-HCl, pH 7.5, 200 mM LiCl, 1 mM EDTA) was added to the beads and mixed with agitation for about 1 min. The magnetic field was used to isolate the beads and the supernatant was added to the oligo(A) fraction. To elute the mRNA, 50 µL of elution buffer (20 mM Tris-HCl, pH 7.5, 1 mM EDTA) and 1 µL of RNase inhibitor was added to the beads, and then incubated at 50°C for 2 min, with occasional agitation. Using the magnetic field to isolate the beads, the supernatant was put into a new tube for the bound/poly(A)+

fraction. The elution step was repeated to ensure as much poly(A)+ RNA was collected as possible. RNA in each fraction was precipitated using ethanol with ammonium acetate and glycogen. The RNA pellets were resuspended in 20 µL of nuclease-free water and 1 µL of each fraction was used to make cDNA. Following reverse transcription (Section 2.8) and quantitative PCR (Section 2.9), the relative abundance of each transcript in the unbound and bound fractions were determined. The percent of each transcript in the unbound/oligo(A) fraction of total (unbound and bound relative abundances added together) was then determined.

2.6 4-thiouridine Ultra-Short Labeling to Isolate Nascent RNAs

Newly transcribed RNAs were isolated from HeLa cells according to previously published protocols (Dolken et al., 2008; Windhager et al., 2012). Briefly, HeLa cells were grown and maintained on 150 mm cell culture dishes until near confluency then treated with 500 µM of 4-thiouridine (4sU) in the same media for 10 min at 37°C. The media was removed and cells were collected in TRIzol. Total RNA was isolated using the RNeasy kit (Qiagen) according to the manufacturer's protocol. For the labeling reaction, 75 µg of 4sU labeled RNA was combined 150 µL (2 µL/1 µg of RNA at 1 mg/mL) of EZ-Link Biotin-HPDP (Thermo Fisher Scientific), 75 µL of 10X biotinylation buffer (100 mM Tris pH 7.4, 10 mM EDTA) and nuclease-free water to bring the total reaction to 525 µL. The reactions were incubated for 2 h with rotation in the dark. After choroform/isoamyl extraction and ethanol precipitation, the RNA was resuspended in 110 µL of TE.

To isolate the 4sU labeled transcripts, 100 µL of streptavidin magnetic beads was added to the 100 µL of RNA (10 µL was reserved as Total RNA) and incubated for 15 min at room temperature while mixing. 100 µL of the equilibration buffer was added to the µMacs magnetic column (Miltenyi Biotec), then the column was washed twice with 100 µL of 1X biotinylation buffer (10 mM Tris pH 7.4, 1 mM EDTA). The collection tube was placed under the column and 100 µL of RNA-streptavidin bead solution was transferred to the equilibrated column, saving the

flow-through. The column was washed with 200 µL wash buffer (100 mM Tris-HCl pH 7.5,

10 mM EDTA, 1 M NaCl, 0.1% Tween 20, heated to 65°C) four times, saving the first two

washes with the flow-through. The column was washed twice more with room temperature wash

buffer. The bound RNA was eluted with 100 µL of 100 mM DTT and the eluate was placed into

a fresh tube. The elution step was repeated 5 min after the first elution. The flow-through and

eluted RNA fractions were then precipitated using ammonium acetate and isopropanol.

2.7 Subcellular Fractionation

Nuclear and cytoplasm cellular fractions were separated based on a previously

published protocol (Weil et al., 2000). Media was removed and cells from 60 or 100 mm plates

were collected in cold PBS using a cell scraper, then placed in 15 mL conical tubes. The tubes

were placed in a centrifuge at 500 x g for 5 min and the supernatant was discarded. 1000 µl of

NP-40 lysis buffer (0.5% vol/vol NP-40, 10 mM Tris-HCl pH 8.5, 1.5 mM $MgCl_2$, 10 mM EDTA,

140 mM NaCl) was added to each 15 mL tube. The cells were then transferred to microfuge

tubes using a 1 mL serological pipette to gently break up the cell pellets. The cells were then

incubated in the lysis buffer on ice for 10-15 min. A small amount of cell lysate was stained with

trypan blue and visualized using a microscope to ensure adequate cell lysis. The nuclei were

pelleted by centrifugation for 5 min at 500 x g at 4°C. The cytoplasmic supernatants were

removed from the nuclear pellets and placed in new tubes after each centrifugation. The nuclear

pellets were washed with 500 µL of the NP-40 lysis buffer twice, gently breaking up the nuclear

pellet each time to remove any remaining cytoplasm. The first wash was added to the

cytoplasmic fraction and the second wash was discarded. The nuclear pellet was then lysed in

500 µL of TRIzol and an equivalent volume of TRIzol was added to each tube of the cytoplasmic

fraction prior to isolating RNA as described previously (Section 2.4). Following reverse

transcription (Section 2.8) and quantitative PCR (Section 2.9), the relative abundance of each

transcript in the nuclear and cytoplasmic fractions were determined. The percent of each

transcript in the nuclear fraction of total (nuclear and cytoplasmic relative abundances added together) was then determined.

2.8 Reverse transcription

For reverse transcription (RT) reactions, 1 μg of total RNA or equal volumes (cell equivalents) of oligo (dT) fractionated or subcellular fractionated RNA samples were annealed to 0.5 μg of random hexamer primers. The reaction was brought to 10 μL total volume with nuclease-free water, incubated at 70°C for 5 min, then placed on ice for 5 min. Next, 4 μL of Improm-II 5X buffer (Promega), 3 μL of 25 mM $MgCl_2$, 1 μL of 10 mM dNTPs, 1 μL of RiboLock RNase inhibitor (Thermo Fisher Scientific), and 1 μL of Improm-II Reverse transcriptase (Promega) were added to the reaction, which was carried out at 25°C for 5 min, 42°C for 60 min, and 70°C for 15 min.

2.9 Quantitative PCR

For qPCR, 2.5 μL of cDNA was used as the template and mixed with IQ SYBR Green 2X Supermix (Bio-Rad) according to manufacturer's instructions. After an initial denaturation step of 95°C for 3 min, a two-step amplification protocol was used with an annealing temperature of 60°C for 30 s and a melting temperature of 95°C for 10 s for 40 cycles. The melt curve was generated by starting at 60°C with temperature increases of 0.5°C every 5 s until a temperature of 95°C was reached. Primer pairs (see Appendix A Table 3) were standardized by amplifying six 5-fold serial dilutions of cDNA and plotting the resulting Cq values to determine the line of best fit. Primer pairs were considered adequate if they generated a PCR efficiency between 90 and 110% (except for ZNF43 primer set, efficiency = 114.2) with a correlation coefficient (R^2 value) of >0.98.

2.10 Linker-Ligation Mediated Poly(A) Test

LLM-PAT assays were performed according to a previously published protocol (Garneau et al., 2008). Briefly, an A0 marker was made by denaturing 1 µg of RNA at 70°C for 5 min and placing it on ice to prevent refolding. The RNA was then combined with 1 µL of 10X RNase H Digestion Buffer, 2 µL RNase H, 2 µL oligo (dT)$_{18}$ (500 ng/µL), and nuclease-free water up to 10 µL total. The reaction was incubated for 45 min at 37°C. The volume of the reaction was brought up to 100 µL with water and the RNA was isolated using a phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation. The pellet was resuspended in 6 µL of nuclease-free water.

For the linker ligation, 1 µg each of RNA and RNase H-treated RNA (for the A0 marker) was mixed with: 1 µL of the 5'adenylated and 3'dideoxy C Linker-3 RNA (10 µM 5'rApp-TTTAACCGCGAATTCCAG-ddC, Integrated DNA Technologies Linker-3), 1 µL 10X Ligase buffer (New England Biolabs), 1 µL T4 ssRNA Ligase 1 (New England Biolabs), and 1 µL RiboLock RNase Inhibitor (Thermo Fisher Scientific). The reactions were incubated at 16°C overnight. The volume of the reaction was brought up to 100 µL with water, and the RNA was recovered by phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation. The pellet was resuspended in 9 µL of nuclease-free water.

The cDNA was made by adding 1 µL of the reverse transcription linker-specific primer (100 µM CTGGAATTCGCGGTT) to the A0 and PAT (non-RNase H treated) RNA. The samples were incubated at 65°C for 5 min, then placed on ice. To make cDNA, reverse transcription was performed using ImProm-II (Promega) as described previously.

For the PCR amplification step, 1 µL of A0 or PAT cDNA was combined with 5 µL of 5X OneTaq Reaction Buffer (New England Biolabs), 0.5 µL of 10 mM dNTPs, 0.125 µL of OneTaq Hot Start Polymerase (New England Biolabs), 0.5 µL of gene-specific forward primer (see Appendix A Table 4), and water to total 25 µL. Residual linker-specific primer from the reverse transcription step served as the reverse primer. The PCR conditions were: 94°C for

30 s; (94°C for 30 s, 55-59°C for 30 s, 68°C for 30 s) x 27-32 cycles; 68°C for 5 min; hold at 4°C. The PCR products were separated on 2.5% low-melting point agarose gels. The gels were stained after electrophoresis with ethidium bromide in TBE for 10-20 min, then briefly destained in water. Image Lab Software (Bio-Rad) was used to determine the most frequent poly(A) tail length of PAT PCR products. A standard curve was made using a 100 bp DNA ladder to determine the size of the A0 marker and most intense region of the PAT products based on the distance the products migrated in the gel. To determine the most common poly(A) tail length, the size of the A0 marker was subtracted from the size of the PAT product.

2.11 Western blot analysis

Protein for DICER KD validation was prepared using TRIzol according to manufacturer's instructions. Briefly, after the aqueous phase was removed during RNA extraction, DNA was isolated from the organic phase using ethanol and the supernatant was kept for protein isolation. Isopropanol was used to precipitate the protein, and the pellet was washed three times in 0.3 M guanidine hydrochloride in 95% ethanol, then in 100% ethanol, air-dried briefly, and dissolved in a 1:1 ratio of 1% SDS and 8 M urea. Protein concentration was determined using the Bradford Assay (Bio-Rad). About 50 µg of protein from the DICER KD and pLKO.1 empty vector control cells was mixed with SDS protein loading dye (0.5 M Tris-HCl pH 6.8, 20% glycerol, 2% SDS, 20% β-mercaptoethanol, and trace bromophenol blue) and denatured at 95°C for 2 min. Proteins were then resolved by SDS-PAGE using a 6% gel. The gel was transferred to the 0.45 µM PVDF Immobilon-P membrane (Millipore) charged with methanol by semi-dry transfer in transfer buffer (48 mM Tris base, 39 mM glycine, 0.1% SDS). The membrane was blocked using 5% non-fat dry milk in 1X TBST (10 mM Tris-HCl pH 8, 150 mM NaCl, 0.1% Tween-20) for 1 h then rinsed with 1X TBST. Membranes were incubated with primary antibodies (rabbit anti-DICER [H212; Santa Cruz Biotechnology; 1:50]; rat anti-HSC70 [1B5; Santa Cruz Biotechnology; 1:1000]) in 5% non-fat dry milk in 1X TBST at 4°C overnight.

Membranes were washed with TBST and incubated with secondary antibodies (goat anti-rabbit IgG HRP conjugate [Bio-Rad; 1:2,500]; goat anti-rat HRP conjugate [Santa Cruz Biotechnology; 1:5,000) for 1 h at room temperature. After washing the membranes with TBST, SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) was added to the membrane and protein bands were detected using the ChemiDoc XRS and analyzed using Image Lab Software (Bio-Rad).

2.12 RNA-fluorescence in situ hybridization procedure (RNA-FISH)

HeLa Tet-Off Advanced cells were transfected with the ZNF12 or PPIA reporter plasmids (Section 2.2.2) using jetPRIME (Polyplus) and plated onto coverslips 24-30 h prior to the RNA-fluorescence in situ hybridization procedure (RNA-FISH). Coverslips were washed with 1 mL of PBS. To fix and permeabilize the cells, the coverslips were incubated for 10 min in 1 mL of a 3:1 mixture of methanol:glacial acetic acid at room temperature, then 2 h at 4°C. The fixation and permeabilization solution was removed, then each coverslip was washed in 1 mL of wash buffer (10% deionized formamide, 2X sodium chloride/sodium citrate [SSC]) and incubated for 5 min. After removing the wash buffer, the coverslips were placed cell-side down onto hybridization buffer (10% deionized formamide in Stellaris RNA-FISH hybridization buffer [Biosearch Technologies, Inc.]) containing 125 nM of the Stellaris RNA-FISH probe recognizing Renilla luciferase labeled with Quasar 570 (VSMF-1034-5, Biosearch Technologies, Inc.) on Parafilm in a humidified chamber and incubated at 37°C in the dark for 16 hours. The hybridization buffer was removed and the coverslips were incubated in wash buffer at 37°C for 30 min. The wash buffer was removed and the coverslips were washed twice with 1X SSC. After removing the excess SSC, the coverslips were mounted onto glass microscope slides using ProLong Diamond Antifade Mountant with DAPI (Molecular Probes by Life Technologies). The slides were allowed to cure at room temperature in the dark overnight and coverslips were fixed to slides using clear nail polish. The cells were visualized using an Olympus IX71 inverted

fluorescent microscope at 100X magnification using the 31000 DAPI/Hoechst filter (EX360, EM460) to visualize DAPI and the 41002 TRITC (Rhodamine)/Cy3 filter (EX535, EM610) to visualize Quasar 570. Images were captured using a digital camera (Q Imaging Retiga 2000R).

2.13 Semi-quantitative PCR to analyze splicing

The cDNA made from HeLa Tet-Off cells transfected with PPIA or ZNF12 reporter plasmids was used as the template for PCR reactions to assess splicing efficiency of the β-globin/IgG chimeric intron. For the PCR amplification step, 1 µL of reporter cDNA was combined with 4 µL of 5X Go Taq Flexi Buffer (Promega), 0.4 µL of 10 mM dNTPs, 0.1 µL of Go Taq Flexi Polymerase (Promega), 0.2 µL of forward primer (AGGTAGCCTTGCAGAAGTTGGT), 0.2 µL of reverse primer (CTCTAGCCTTAAGAGCTGTAATTGAAC) and water to total 20 µL. The PCR conditions were: 95°C for 2 min; (95°C for 30 s, 55°C for 30 s, 72°C for 20 s) x 20-35 cycles (based on quantity of reporter mRNA in sample as determined by qPCR); 72°C for 5 min; hold at 4°C. The PCR products were separated on 2.2% agarose gels. The gels were stained with ethidium bromide in TBE for 10-20 min, and then briefly destained in water. Image Lab Software (Bio-Rad) was used to determine relative quantity of each band by measuring the intensity of the unspliced and spliced products. The relative quantities of the unspliced products were divided by 2 to adjust for molecular weight differences between spliced and unspliced bands. The percent of the intron retained in each reporter mRNA was then determined by dividing the relative quantity of the unspliced product by the total (unspliced + spliced).

CHAPTER 3: RESULTS

3.1 RNAs with poly(A) tails of <30 nucleotides bind inefficiently to oligo (dT)$_{25}$

We first wanted to determine whether we could recapitulate the findings of Yang et al. by demonstrating that C2H2-ZNF mRNAs indeed have shorter than average poly(A) tails (Yang et al., 2011b). In order to address this, we needed to establish how RNAs with poly(A) tails of different lengths behaved during oligo (dT) fractionation, which separates RNA populations by relative poly(A) tail length. Meijer et al. found that the majority of mRNAs that bound efficiently to oligo (dT) had poly(A) tails <30 nucleotides (nt) in length (Meijer et al., 2007). To see if this would also be the case for our study, we generated control RNAs by *in vitro* transcription, each containing a sequence from a different *S. cerevisiae* gene and having 15, 30, 45, or 149 A-residues at the 3' end. We also made an A0 control RNA by *in vitro* transcribing part of the pGEM-4 empty vector. Then, we tested how well each control RNA would bind to an oligo (dT)$_{25}$ matrix in order to fractionate RNAs based on relative poly(A) tail length. As RNAs require poly(A) tails to properly base pair with oligo (dT) to be retained on the matrix, we naturally expected all of the RNAs with no poly(A) tail (the 'A0' control RNA) to end up in the unbound fraction. Because of its short (A) tail and limited ability to hybridize efficiently with the oligo (dT)$_{25}$ matrix, we predicted that a portion of the RNAs with a 15 nt (A) tail might remain unbound. We expected very little, if any, of the control RNAs with longer poly(A) tails, (e.g. A30, A45, and A149) would end up in the unbound fraction, as previously demonstrated (Meijer et al., 2007).

As anticipated, all (>99%) of the A0 control ended up in the unbound fraction (Figure 6B). Approximately 30% of the A15 control RNA was in the unbound fraction, indicating that an RNA with a relatively short poly(A) tail binds inefficiently to oligo (dT)$_{25}$. The A30, A45, and A149 reporter RNAs were all highly efficient at binding oligo (dT)$_{25}$ as less than 3% of each of these mRNAs were present in the unbound fraction. Based on these results, we concluded that RNAs found in the unbound fraction will have poly(A) tails <30 nt, and probably fewer than 15 nt. In
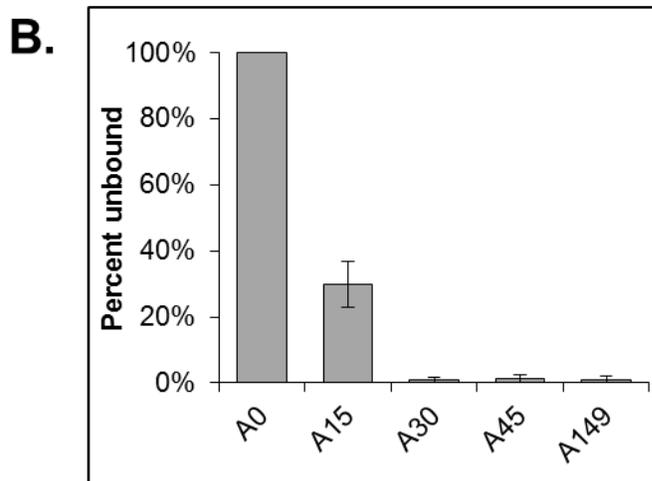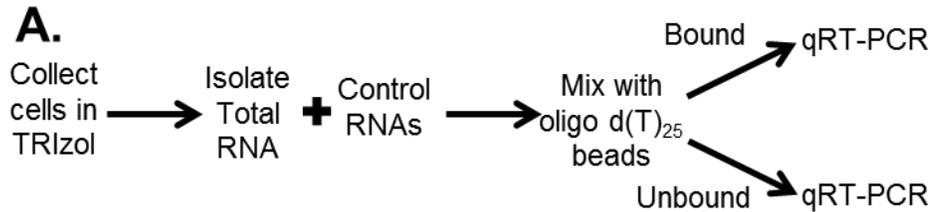
Figure 6: Characterization of the behavior of RNAs with poly(A) tails of different lengths during oligo (dT) fractionation. A. Overview of oligo (dT) fractionation assay to assess relative poly(A) tail length. Total cellular RNA was mixed with *in vitro* transcribed control RNAs with the indicated number of adenosine residues at the 3' end. Then, the RNA was added to oligo $(dT)_{25}$ magnetic beads in the presence of 500 mM LiCl. Unbound RNAs were collected and the beads were washed several times prior to the elution step, which was carried out at 50°C using a buffer containing no salt. This procedure is designed to separate RNAs with long poly(A) tails (bound to beads) from those with a shorter tail or no tail (unbound). **B.** Distribution of control *in vitro* transcribed RNAs between fractions was determined using qRT-PCR. Each RNA contains the indicated number of adenosine residues at the 3' end. Error bars represent the standard deviation of 3 independent experimental replicates.

addition, these results were consistent with the findings of Mejier et al. As the RNAs in the unbound fraction could have short poly(A) tails, we will classify this as the "oligo(A)" fraction. Thus oligo (dT)$_{25}$ fractionation followed by qRT-PCR appeared to be an appropriate method to rapidly assess relative poly(A) tail length.

3.2 Bimorphic C2H2-ZNF transcripts have large populations with short poly(A) tails

We next assessed the poly(A) status of endogenous HeLa cell mRNAs. PPIA and GAPDH mRNAs, which are expected to possess long poly(A) tails (Yang et al., 2011b), were selected to serve as endogenous positive controls for the oligo (dT) fractionation assay. We also selected seven C2H2-ZNF mRNAs for analysis, six that were projected to be bimorphic (ZNFs 2, 12, 43, 134, 280B, and 627) and one that was reported to be polyadenylated (ZNF282) (Yang et al., 2011b). As seen in Figure 7, only ~10% of the PPIA and GAPDH mRNAs had oligo(A) tails. These findings are consistent with the expectation that most mRNAs are rapidly degraded following removal of the poly(A) tail and are therefore observed primarily with long poly(A) tails (Chen and Shyu, 2011). Similarly, only ~10% of the ZNF282 mRNA possessed oligo(A) tails consistent with previous findings (Yang et al., 2011b). Interestingly, each of the bimorphic C2H2-ZNF mRNAs had a significantly larger population with oligo(A) tails (ranging from ~30 to 70% of the total RNA) compared the PPIA and GAPDH transcripts. Based on the results for the *in vitro* transcribed control RNAs (Figure 6), mRNAs ending with oligo(A) tails compose a large proportion of the population of transcripts for the six bimorphic C2H2-ZNF mRNAs we tested.

3.3 LLM-PAT in combination with oligo (dT) fractionation reveals that C2H2-ZNF mRNAs persist with unusually short poly(A) tails

We wanted to determine if each of the six bimorphic C2H2-ZNF mRNAs actually exist in a single population with relatively short poly(A) tails (<30 nt) that bind to oligo (dT) inefficiently or
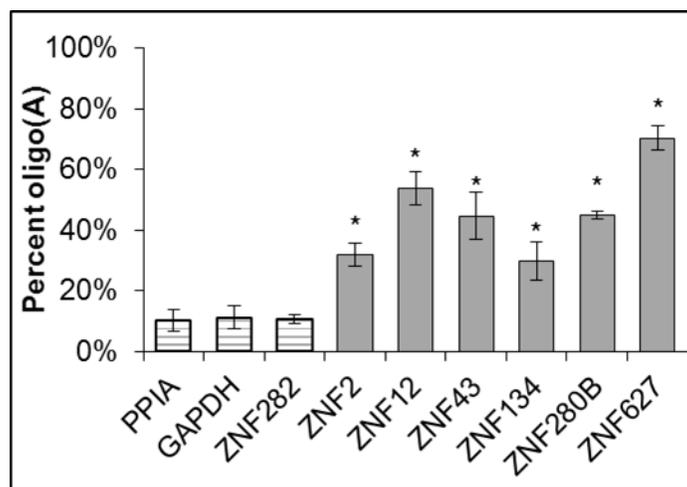
Figure 7: Multiple C2H2-ZNF mRNAs have short poly(A) tails. Total HeLa cell RNA was isolated using TRIzol and oligo (dT) fractionation was used to separate poly(A) and oligo(A) RNAs. qRT-PCR was used to measure relative abundance of the mRNA in each fraction. The percent of each mRNA in the oligo(A) fraction of total was then determined. Error bars represent standard deviation of 3 biological replicates. * $p < 0.05$ compared to PPIA by one-way ANOVA followed by Dunnett's post-hoc test for multiple comparisons to one control.

if there are indeed two distinct classes of transcripts with long and short poly(A) tails. In order to gain a more accurate picture of the poly(A) tail lengths of bimorphic C2H2-ZNF mRNAs, we performed a linker-ligation mediated poly(A) tail (LLM-PAT) assay (Garneau et al., 2008). As shown in the diagram depicting this method in Figure 8, a 5' adenylated and 3' dideoxy modified oligonucleotide was ligated to the 3' ends of HeLa RNA in an ATP-independent fashion. These modifications prevent unwanted ligation reactions. Following ligation, cDNA was made using a linker-specific primer for the reverse transcription reaction and then PCR-amplified using a linker-specific reverse primer and a gene-specific primer located approximately 100-200 bp upstream of the poly(A) site. In parallel we made A0 markers, which show where the PCR product should migrate if it was derived from an RNA without a poly(A) tail. These were made by RNase H/oligo (dT) treatment. The A0 and LLM-PAT PCR products were visualized on a native agarose gel. Because the poly(A) tails are heterogeneous, the LLM-PAT PCR product appears as a smear migrating from the A0 up to several hundred nucleotides longer. The smear produced from polyadenylated RNAs would likely obscure the products with short poly(A) tails, making it difficult to detect and thereby analyze the oligo(A) tails. Therefore, we decided to characterize the poly(A) tails after oligo (dT) fractionation.

We selected two polyadenylated mRNAs, PPIA and GAPDH, and two bimorphic C2H2-ZNF mRNAs for analysis by LLM-PAT. Based on the efficiency with which the PPIA and GAPDH mRNAs bound to the oligo $(dT)_{25}$ matrix (Figure 7), we anticipated that these mRNAs would have long poly(A) tails in the PAT assay. For the C2H2-ZNFs, there were two possibilities: (i) we would see mRNAs with relatively short tails (~15 nt) in both oligo(A) and poly(A) PCR products, or (ii) we would see a short poly(A) tail on the product from the unbound fraction and a long poly(A) tail on the PCR product from the bound fraction.

As expected, GAPDH and PPIA were readily detected in the poly(A)+ fraction and the PCR products migrated in a smear consistent with having poly(A) tails >50 nt in length (Figure 9). We noted that the most common poly(A) tail for GAPDH in the poly(A)+ fraction was ~100 nt
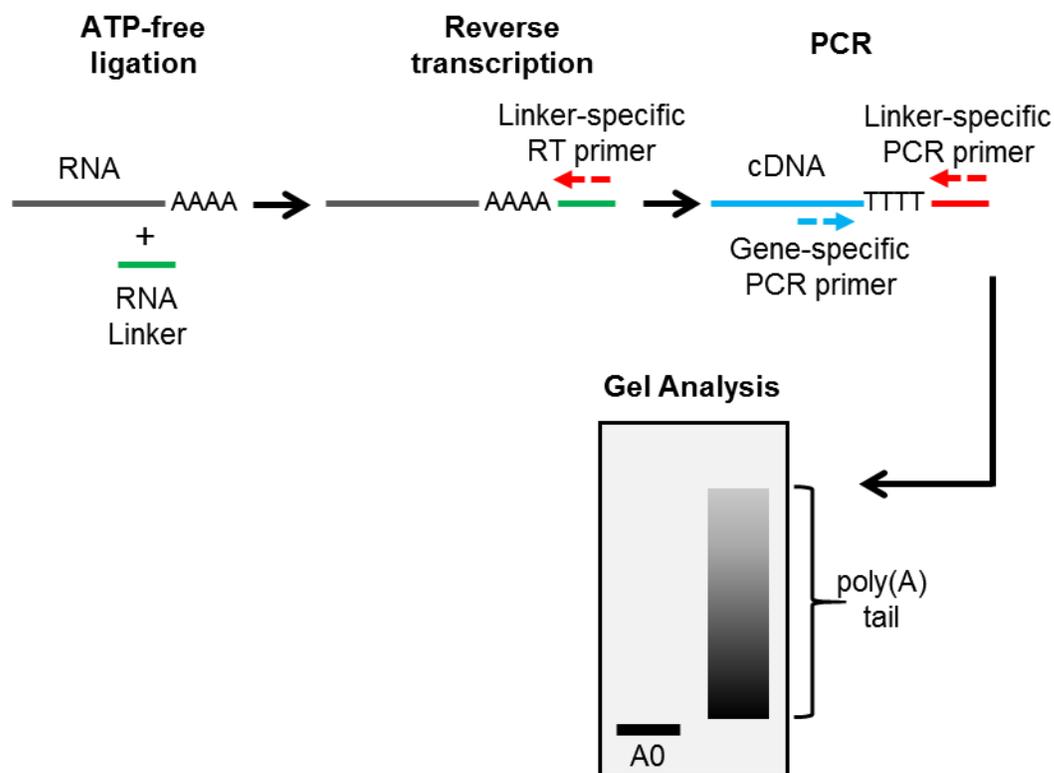
Figure 8: Overview of the linker-ligation mediated poly(A) tail (LLM-PAT) assay to assess poly(A) tail length. Briefly, an RNA linker was ligated to the 3' end of the RNAs. The RNAs were reverse transcribed using a linker-specific primer for the reaction. The cDNAs were then amplified using a gene-specific forward primer and the linker-specific primer for the reverse reaction. Products were then analyzed using agarose gel electrophoresis. The A0 marker was made using the same protocol except that the poly(A) tails were removed using oligo (dT) and RNase H prior to the ligation.

shorter compared to the results for PPIA. This difference is consistent with a recent analysis of global poly(A) tail lengths which reported that many mRNAs encoded by housekeeping genes, such as GAPDH, have shorter poly(A) tails compared to the general mRNA population (Subtelny et al., 2014). In addition, a product was detected in the unbound/oligo(A) fraction for both GAPDH and PPIA. The band was about 20-30 nt above the corresponding A0 band, suggesting that the GAPDH and PPIA mRNAs in the oligo(A) fraction possess short poly(A) tails (Figure 9). The 20-30 nt product in the oligo(A) fraction was expected given that once the poly(A) tail is sufficiently shortened, most mRNAs are rapidly degraded and no longer detectable (Chen and Shyu, 2011). These results, along with those from the reporter RNAs with artificial poly(A) tails, (Figure 6) demonstrate that mRNAs with poly(A) tails less than ~25 nt do not appreciably bind to the oligo (dT) matrix under the conditions used here.

As expected, ZNF12 and ZNF627 LLM-PAT products were also detected in both poly(A) and oligo(A) fractions (Figure 9). The ZNF12 and ZNF627 transcripts in the poly(A) fraction have poly(A) tail lengths in the same range as those of GAPDH and PPIA. However, the ZNF products detected in the oligo(A) fraction both migrate at the same molecular weight as the artificially generated A0 marker. This indicates that despite having no detectable poly(A) tail, these mRNAs are not rapidly turned over.

We noted that the relative amount of product in the bound and unbound fractions does not reflect the quantity of each mRNA found in these fractions by qRT-PCR. This is perhaps not surprising given that the steps of the LLM-PAT assay vary in efficiency, especially during the linker ligation step (Hafner et al., 2011). It is also possible that the 3' end of the A0 RNA derived from the ZNF transcripts may be inaccessible for linker ligation (Zhuang et al., 2012). Taking the results from both the oligo (dT) fractionation and LLM-PAT assays into account, we conclude that for both ZNF12 and ZNF627 at least 50% of the total mRNAs have a very short (<10 nt) poly(A) tails, while the rest of the mRNA is normally polyadenylated. These results closely reproduce and extend observations made by Yang et al. (2011), and support that in HeLa cells,

A.

GAPDH   PPIA

ZNF12   ZNF627

A0   oligo(A)   poly(A)   A0   oligo(A)   poly(A)

B.

Gene	Sample	Most common poly(A) tail length (nt)
GAPDH	oligo(A)	26.6 ± 5.7
	poly(A)	54.8 ± 7.2
PPIA	oligo(A)	26.7 ± 0
	poly(A)	151.4 ± 2.9
ZNF12	oligo(A)	4.0 ± 3.5
	poly(A)	110.4 ± 68.0
ZNF627	oligo(A)	3.2 ± 2.3
	poly(A)	80.0 ± 69.0
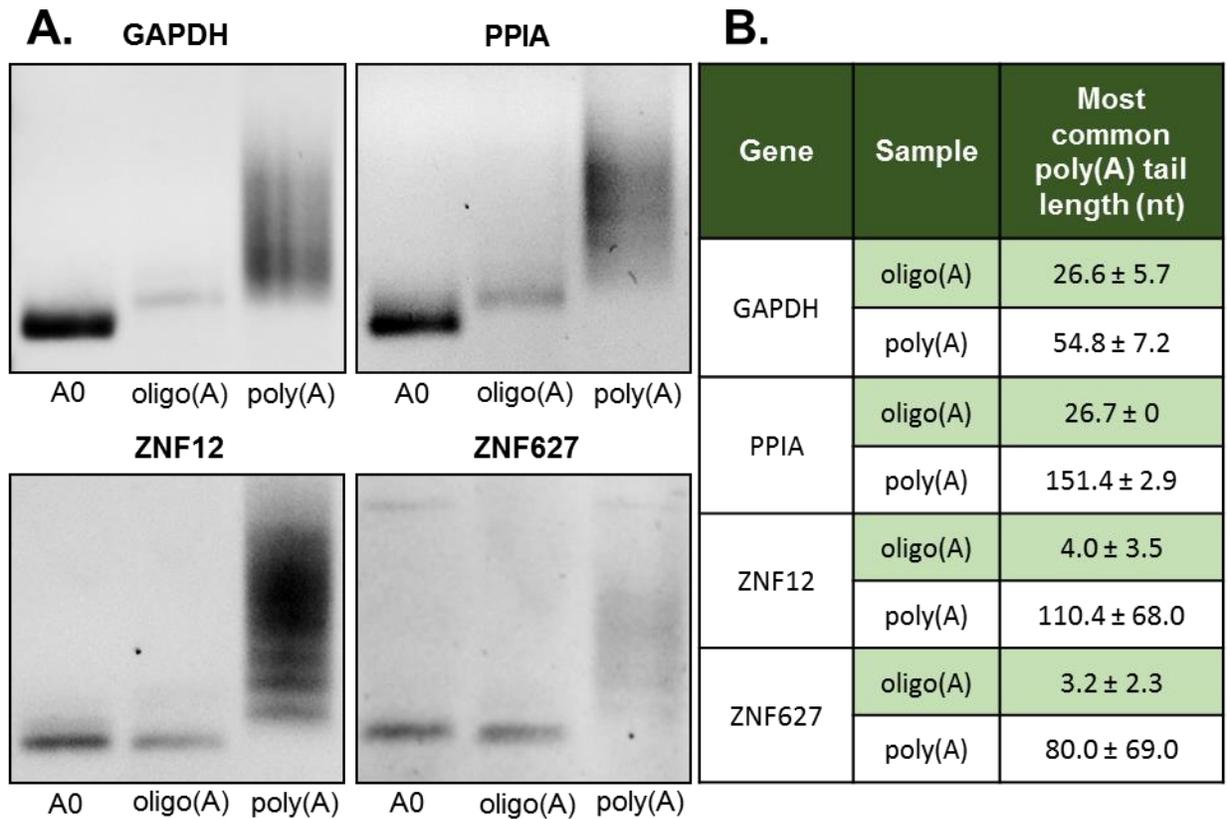
Figure 9: C2H2-ZNF mRNA poly(A) tails have a bimorphic distribution. A. LLM-PAT assays were performed on RNA after oligo (dT) selection to determine poly(A) tail length in the poly(A) fraction and oligo(A) fraction. **B.** Quantification of LLM-PAT results. A standard curve was made using a 100 bp DNA ladder to determine the molecular sizes of the A0 marker and PAT PCR products. To determine the most common poly(A) tail length, the size of the A0 marker was subtracted from the size of the most intense region of the PAT PCR product. Values for poly(A) tail length are the mean +/- SD of 2 (PPIA and ZNF627) or 3 (GAPDH and ZNF12) biological replicates.

multiple C2H2-ZNF mRNAs are truly bimorphic and have two populations with poly(A) tails of very different lengths.

3.4 Nascent bimorphic C2H2-ZNF transcripts have short poly(A) tails

There are two main ways for an mRNA to acquire a short poly(A) tail: (i) the mRNA is not polyadenylated effectively, or (ii) a long poly(A) is initially added to the mRNA and later removed by deadenylation, but the mRNA is not decayed. To distinguish these possibilities, we assessed poly(A) tail length for nascent C2H2-ZNF mRNAs. We added 4-thiouridine (4sU), a metabolic label that is incorporated into RNA during transcription, to HeLa cells in culture (Dolken et al., 2008). Consistent with previous reports (Rabani et al., 2011; Windhager et al., 2012), we used a 10 minute labeling period to label only recently made RNAs, which were then isolated via biotinylation and streptavidin pull-down as outlined in the diagram in Figure 10A.

We assessed poly(A) tail length of the nascent (4sU labeled) C2H2-ZNF mRNAs using the oligo (dT) selection assay. If the C2H2-ZNF mRNAs undergo normal polyadenylation, we would expect nascent C2H2-ZNF mRNAs to behave like the nascent PPIA mRNAs and bind efficiently to oligo (dT). Instead, the nascent C2H2-ZNF mRNAs were significantly more abundant in the oligo(A) fraction compared to the nascent PPIA mRNAs (Figure 10B). This indicates that for 40-50% of the population of these C2H2-ZNF mRNAs, poly(A) tail length may be restricted from the start. However, we cannot rule out the possibility that all of the C2H2-ZNF mRNAs are efficiently polyadenylated, and then the poly(A) tail is rapidly removed for about half of these mRNAs. Either way, the mRNAs from multiple C2H2-ZNFs have oligo(A) tails very early on in their lifespan, likely before export from the nucleus.

Compared to the steady-state mRNAs (Figure 7), the nascent PPIA and GAPDH mRNAs were approximately 5-10% more abundant in the oligo(A) fraction (Figure 10). In addition, the nascent ZNF282 mRNAs were about 20% more abundant in the oligo(A) fraction compared to steady-state. These differences were not surprising given that some of these
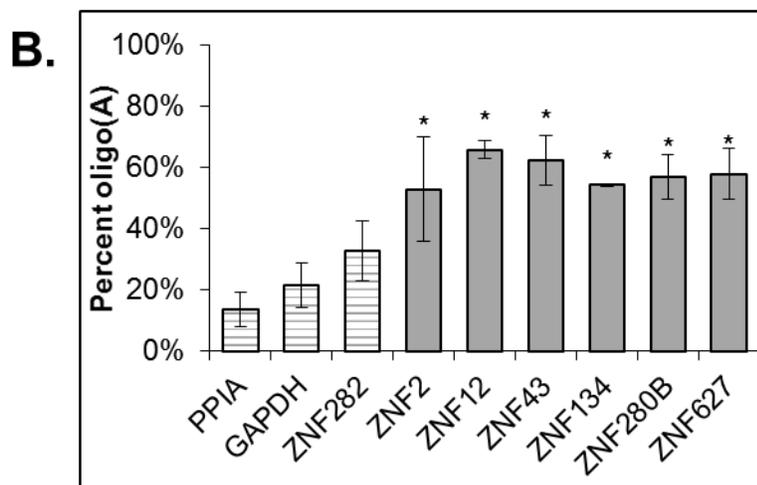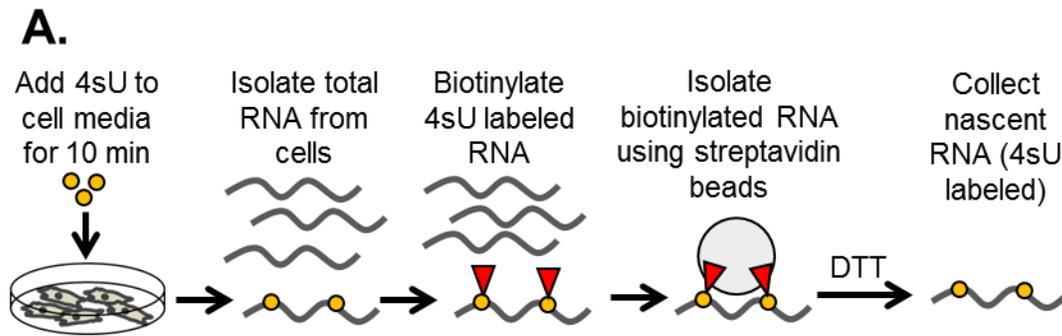
Figure 10: Nascent C2H2-ZNF mRNAs are inefficiently polyadenylated. A. HeLa cells were treated with 4-thiouridine (4sU) for 10 min to label newly made RNAs. Total HeLa RNA was collected from the cells and the 4sU labeled RNA was biotinylated. Biotin forms a disulfide bond with the 4sU. Streptavidin beads were used to bind the biotinylated RNAs (nascent fraction). Dithiothreitol (DTT) was added to break the 4sU-disulfide bond and elute the bound RNA off of the column. **B.** Oligo (dT) selection was used to separate the nascent RNA into poly(A) and oligo(A) fractions. qRT-PCR was used to measure the relative abundance of mRNA in each fraction. The percent of each mRNA in the oligo(A) fraction of total was determined. Error bars represent standard deviation of 3 biological replicates. * $p < 0.05$ compared to PPIA by one-way ANOVA followed by Dunnett's post-hoc test for multiple comparisons to one control.

mRNAs were probably not completely processed and had not yet been fully polyadenylated (Windhager et al., 2012). Nonetheless, the majority of the nascent PPIA, GAPDH, and ZNF282 mRNAs were polyadenylated.

We also noted minor differences in the abundance of nascent C2H2-ZNF mRNAs with short poly(A) tails compared to steady-state (Figure 10B). In some cases, the differences were consistent with removal of the poly(A) tail at some point after transcription, for instance by deadenylation. A specific example is the ZNF627 mRNA, which went from ~60% of the mRNA having oligo(A) tails at synthesis to ~70% at steady-state (Figure 7). In other examples, we saw the opposite trend as with ZNF134, which went from ~50% with oligo(A) tails at synthesis to ~30% at steady-state. This suggests that the poly(A) tail of ZNF134 is short early on in its lifespan and then the poly(A) tail was added by a poly(A) polymerase. It is also possible that C2H2-ZNF mRNAs with different poly(A) tail lengths have different decay rates. One might predict the mRNAs with oligo(A) tails would be degraded more rapidly because the poly(A) tail helps stabilize mRNAs (Dreyfus and Régnier, 2002; Smith et al., 2014), but it also possible that C2H2-ZNF mRNAs with short poly(A) tails are more stable given their accumulation. This is atypical, but not out of the question. Unfortunately, because both poly(A) tail remodeling (i.e. polyadenylation and deadenylation) and mRNA decay are ongoing processes, it is difficult to assess the overall mRNA stability rates of C2H2-ZNF mRNAs with short versus long poly(A) tails. In general, these C2H2-ZNF mRNAs were produced with a population possessing oligo(A) tails to start with and continued to maintain this population.

3.5 MicroRNAs do not influence the accumulation of C2H2-ZNF mRNAs with oligo(A) tails

It is likely that trans-acting factors recognize specific sequence elements within the C2H2-ZNF mRNAs to control polyadenylation. Notably, C2H2-ZNF mRNAs contain multiple miRNA target sites in their coding sequences. These sites are in regions containing highly repetitive sequence elements, primarily in the zinc finger motif and linker region encoding

sequences (Huang et al., 2010; Schnall-Levin et al., 2011). While miRNAs have established

roles in inhibiting translation and contributing to deadenylation and decay of mRNAs in the

cytoplasm (Eulalio et al., 2008; Guo et al., 2010), miRNAs also have roles in the nucleus

(Roberts, 2014). Most of the evidence thus far shows that nuclear miRNAs regulate the

biogenesis of non-coding RNAs (Liang et al., 2013), but miRNAs could potentially influence the

nuclear processing of mRNAs as well. We decided to investigate the possibility that miRNAs

bind to the C2H2-ZNF mRNAs in order to influence the poly(A) tail length.

Because multiple miRNAs are known to bind to the C2H2-ZNF mRNAs (Schnall-Levin

et al., 2011), we wanted to use a broad approach to reduce overall miRNA abundance in HeLa

cells. To do so, we used a cell line stably expressing an shRNA targeting DICER; an enzyme

required for miRNA processing. The control cell line contained the empty shRNA vector.

Western blot analysis showed noticeable reduction of the DICER protein and qRT-PCR results

showed an average of 60% reduction in DICER mRNA levels compared to the control (Figure

11). To ensure that the DICER KD had indeed reduced miRNA expression, we first assessed

the abundance of the LIN28 mRNA, a well-known target of the Let-7 microRNA family (Zhong et

al., 2010). Let-7 should be reduced in the DICER KD cells, thereby preventing miRNA mediated

decay of LIN28. Thus, we expected LIN28 to be increased in the DICER KD cells, and indeed it

was (Figure 11B).

We then assessed steady-state abundance of the C2H2-ZNF mRNAs in the DICER KD

cells. We selected three C2H2-ZNF mRNAs (ZNFs 2, 12, and 627) that were previously shown

to contain putative target sites for miRNAs 181, 188, and 199 (Schnall-Levin et al., 2011). ZNF2

contains 4 predicted miR-188 sites, ZNF12 contains 20 predicted miR-181 sites and 10

predicted miR-188 sites, and ZNF627 contains 7 predicted sites each for miR-188 and miR-199.

We used qRT-PCR to compare the relative expression of the C2H2-ZNF mRNAs in control cells

versus the DICER KD cells. We found no significant difference in steady-state abundances of
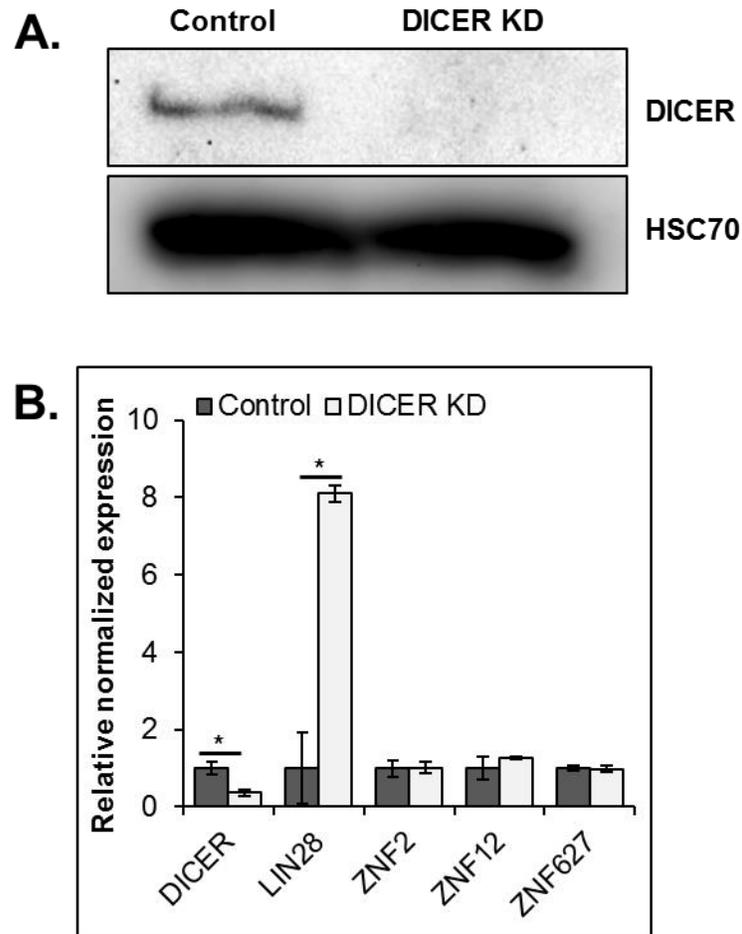
Figure 11: DICER knockdown does not alter abundance of C2H2-ZNF mRNAs.
A. Western blot analysis of the DICER protein in the DICER shRNA cells compared to the control cells. Data were normalized to HSC70. Representative of two biological replicates. **B.** The abundance of the indicated mRNAs was determined using qRT-PCR. Expression was normalized to GAPDH. Error bars represent standard deviation of three biological replicates. * $p < 0.05$ for DICER KD compared to control samples (paired, two-tailed Student's T-test).

each of the ZNF mRNAs (Figure 11B), but these findings do not rule out the possibility that

miRNAs influence poly(A) tail length.

We then used the oligo (dT) selection assay to see if the C2H2-ZNF mRNAs showed

any changes in poly(A) tail length in the DICER KD cells, but saw no differences compared to

the empty vector control cells (Figure 12). Therefore, it appears that miRNAs do not play a

significant role in determining the poly(A) tail length for C2H2-ZNF mRNAs.

3.6 C2H2-ZNF mRNAs are localized to the nucleus

In general, mRNAs require poly(A) tails of ~200 nt for efficient nuclear export, mediated

by interactions of PABPs with nuclear export factors (Di Giammartino and Manley, 2014; Fuke

and Ohno, 2008). Because a significant proportion of the C2H2-ZNF mRNAs have short poly(A)

tails, we hypothesized that these mRNAs would be inefficiently exported and therefore be more

abundant in the nucleus compared to typical mRNAs, such as PPIA and GAPDH. We used

detergent-based subcellular fractionation to isolate nuclear and cytoplasmic RNA (Weil et al.,

2000) followed by qRT-PCR to determine how much of each mRNA was in the nucleus.

We first assessed two control RNAs to ensure adequate subcellular fractionation (dark

bars in Figure 13). As tRNA is required for translation, it is generally located in the cytoplasm

(Liao et al., 2010). Only about 4% of the tRNA was in the nuclear fraction, which indicated that

this fraction was truly nuclear. The 45S ribosomal precursor RNA (45S rRNA) is processed in

the nucleus to produce the smaller ribosomal RNAs (Gondran et al., 1999), therefore we should

expect to see all of this RNA represented in the nuclear fraction. As anticipated, >99% of the

45S rRNA was in the nuclear fraction. Thus we could be confident that this subcellular

fractionation protocol adequately separated the nuclear and cytoplasmic RNAs.

We evaluated the same seven C2H2-ZNF mRNAs that we used in the oligo (dT) assay

to assess subcellular localization, expecting that the failure of polyadenylation would prevent

efficient nuclear export. Therefore, the C2H2-ZNF mRNAs, except for the polyadenylated
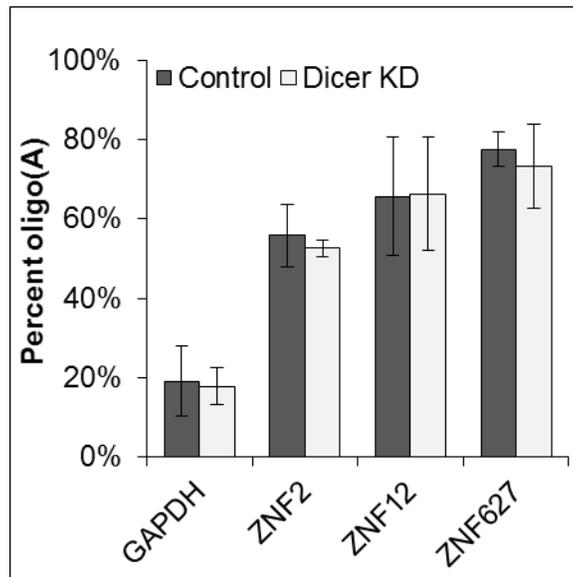
Figure 12: DICER knockdown does not influence polyadenylation of the C2H2-ZNF mRNAs. Oligo (dT) selection was used to separate poly(A) and oligo(A) RNA fractions from total RNA isolated from Control and DICER KD cells. qRT-PCR was used to measure the relative abundance of mRNA in each fraction. The percent of each mRNA in the oligo(A) fraction of total was determined. Error bars represent standard deviation of three biological replicates.
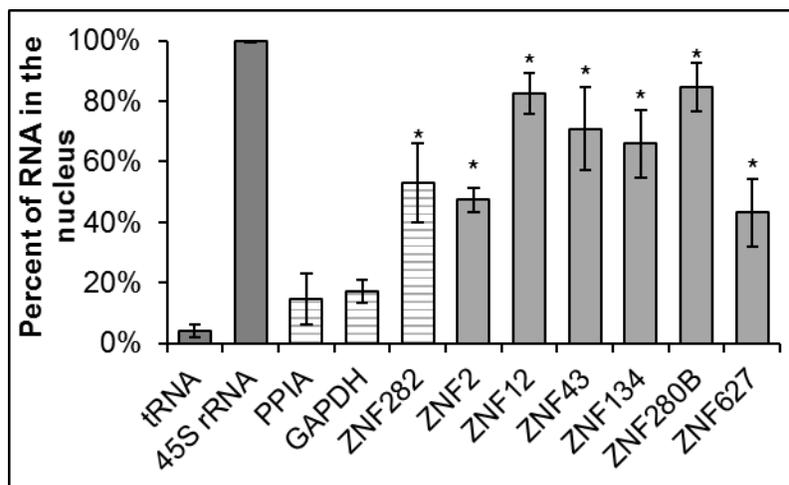
Figure 13: C2H2-ZNF mRNAs are abundant in the nucleus. Detergent-based subcellular fractionation was used to separate nuclear and cytoplasm cellular fractions. TRIzol was added to each fraction for RNA isolation. Relative abundance of the mRNAs in each fraction was determined using qRT-PCR. Fractionation controls are shown in the dark bars (tRNA=cytoplasmic control, 45S rRNA=nuclear control). Error bars represent standard deviation of 3 biological replicates. * $p < 0.05$ compared to PPIA by one-way ANOVA followed by Dunnett's post-hoc test for multiple comparisons to one control.

ZNF282 mRNA, would be more abundant in the nucleus compared to polyadenylated PPIA and GAPDH mRNAs. As shown in Figure 13, PPIA and GAPDH mRNAs were exported efficiently with less than 20% of each transcript found in the nucleus, while all of the C2H2-ZNF transcripts, including ZNF282, had over 40% of their population restricted to the nucleus. For some of the C2H2-ZNFs (ZNF12 and ZNF280B), as much as 80% of the mRNA was in the nuclear fraction.

A linear regression analysis of percent nuclear versus percent oligo(A) suggests that the nuclear localization of the bimorphic C2H2-ZNF mRNAs may be correlated with the short poly(A) tails (Figure 14). We excluded ZNF282 from the analysis because unlike the other C2H2-ZNFs, this mRNA has a long tail but is restricted to the nucleus. There are examples of mRNAs that have long poly(A) tails, but are abundant in the nucleus due to intron retention (Bhatt et al., 2012; Yap et al., 2012) or are stored in subnuclear domains, such as nuclear speckles (Bahar Halpern et al., 2015) or paraspeckles (Fox and Lamond, 2010). The ZNF282 polyadenylated mRNA may be retained in nucleus through such mechanisms. We noted that while ~70% of the ZNF627 mRNA was in the oligo(A) fraction (Figure 7), it was not as abundant in the nucleus compared to the other bimorphic C2H2-ZNF mRNAs. It is possible that the polyadenylated ZNF627 mRNAs are exported from the nucleus and deadenylated in the cytoplasm more quickly than the other ZNF mRNAs, thereby increasing the steady-state abundance of ZNF627 mRNAs possessing oligo(A) tails.

3.7 C2H2-ZNF mRNAs have oligo(A) tails and are localized to the nucleus in other cell types

HeLa cells are extensively used as a model system for molecular research; however, as cancer cells, they are highly abnormal. Therefore, we also wanted to assess the behavior of C2H2-ZNF mRNAs in other cell types. For this purpose, we selected human iPS cells as another highly proliferative cell line and HFFs as a differentiated primary cell line with limited capacity to divide. We used the oligo (dT) fractionation assay to assess the same seven C2H2-
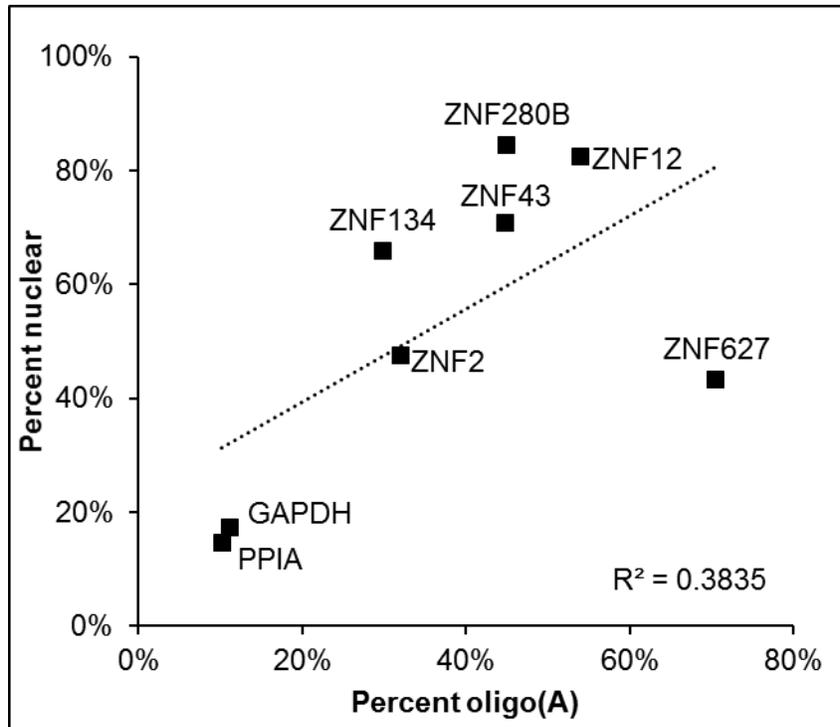
Figure 14: Percent of mRNAs in the nuclear fraction is positively correlated with the percent in the oligo(A) faction. Scatter plot of the average percent nuclear versus average percent oligo(A) for each of the mRNAs assessed by subcellular fractionation and oligo (dT) fractionation.

ZNF mRNAs previously characterized in HeLa cells. In both HFFs and iPS cells, most of the C2H2-ZNF mRNAs were more abundant in the oligo(A) fraction compared to the PPIA mRNA, except for polyadenylated ZNF282 (Figure 15). In addition, we determined that there were no significant differences in the overall distribution of the C2H2-ZNF mRNAs in the oligo(A) fraction among HeLa, HFF, and iPS cells (Figure 16). These results also suggest that the short poly(A) tails of C2H2-ZNF mRNAs do not significantly contribute to the differential stability previously identified in these mRNAs in HFFs and iPS cells (Neff et al., 2012). Thus, C2H2-ZNF mRNAs have short poly(A) tails in all human cell lines tested, suggesting this is a general property of these transcripts and not limited to cancer cells or proliferating cells.

We also used the subcellular fractionation assay followed by qRT-PCR to assess C2H2-ZNF mRNA nuclear abundance in HFFs and iPS cells. As in HeLa cells, most of the C2H2-ZNF mRNAs were enriched in the nucleus compared to the PPIA mRNA in both the HFF and iPS cells (Figure 17). We also noted that there were no significant differences in the overall nuclear localization of the C2H2-ZNF mRNAs among HeLa, HFF, and iPS cells (Figure 18). C2H2-ZNF mRNAs therefore accumulate in the nucleus in multiple human cell lines.

3.8 Overview of the reporter construct design

We established that the mRNAs of multiple different C2H2-ZNFs have short poly(A) tails and are unusually abundant in the nucleus. Our next goal was to identify potential sequences involved in regulating the short poly(A) tail and nuclear localization. We used reporter constructs to complete this task. With this type of system, we could narrow down the specific C2H2-ZNF sequences that influence the poly(A) tail length and nuclear abundance of a reporter mRNA. The ZNF12 mRNA population has a bimorphic poly(A) tail distribution, with short poly(A) tails in ~50% of the transcripts (Figure 7). In addition, the majority of the ZNF12 mRNAs (~80%) were localized to the nucleus (Figure 13). Given both of these attributes, ZNF12 was a good
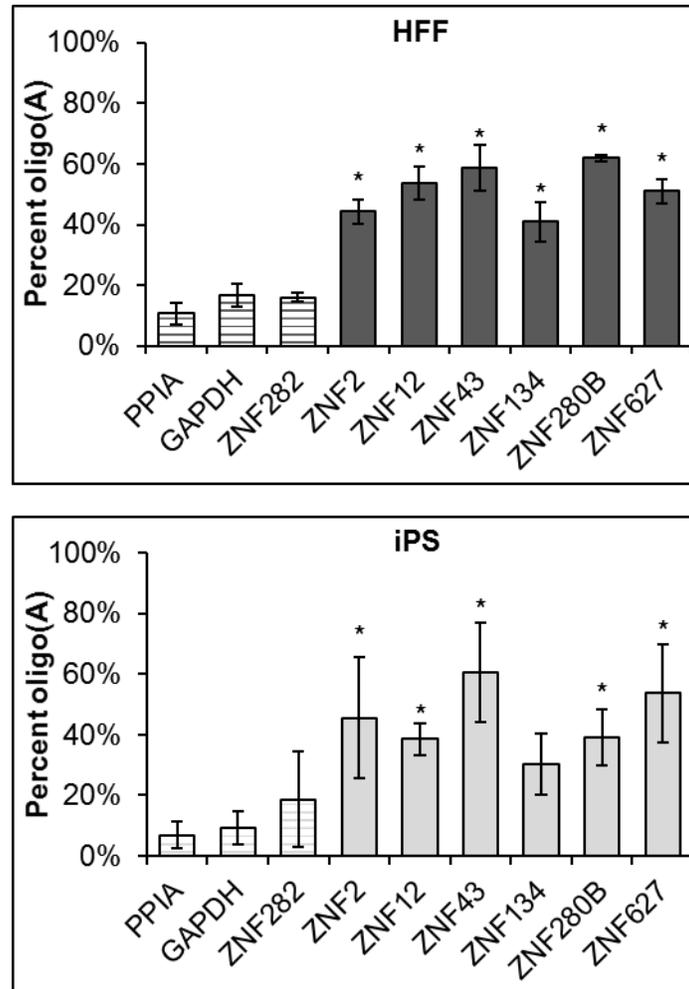
Figure 15: C2H2-ZNF mRNAs have short poly(A) tails in other human cell lines.
Total RNA was isolated from HFF and iPS cells using TRIzol. For each cell line, oligo (dT) fractionation was used to separate poly(A) and oligo(A) RNA fractions. qRT-PCR was used to measure relative abundance of the mRNA in each fraction. The percent of each mRNA in the oligo(A) fraction of total was then determined. Error bars represent standard deviation of 3 biological replicates. * $p < 0.05$ compared to PPIA by one-way ANOVA followed by Dunnett's post-hoc test for multiple comparisons to one control.
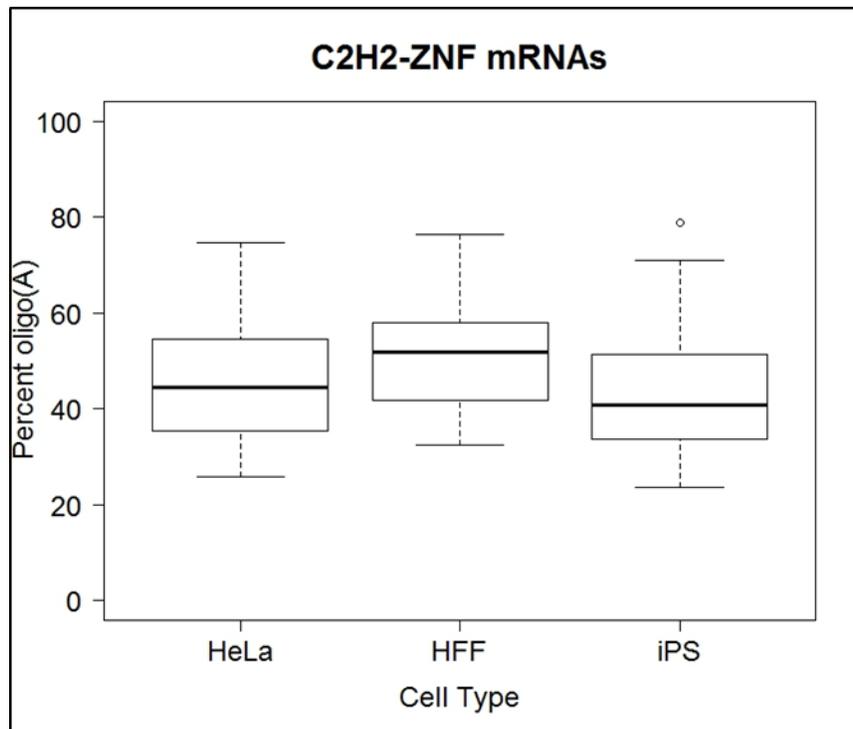
Figure 16: Box-plot comparing the proportion of C2H2-ZNF mRNAs in the oligo(A) fraction in HeLa, HFF, and iPS cells. There were no statistically significant differences among group means as determined by one-way ANOVA (F=2.43, p=0.102).
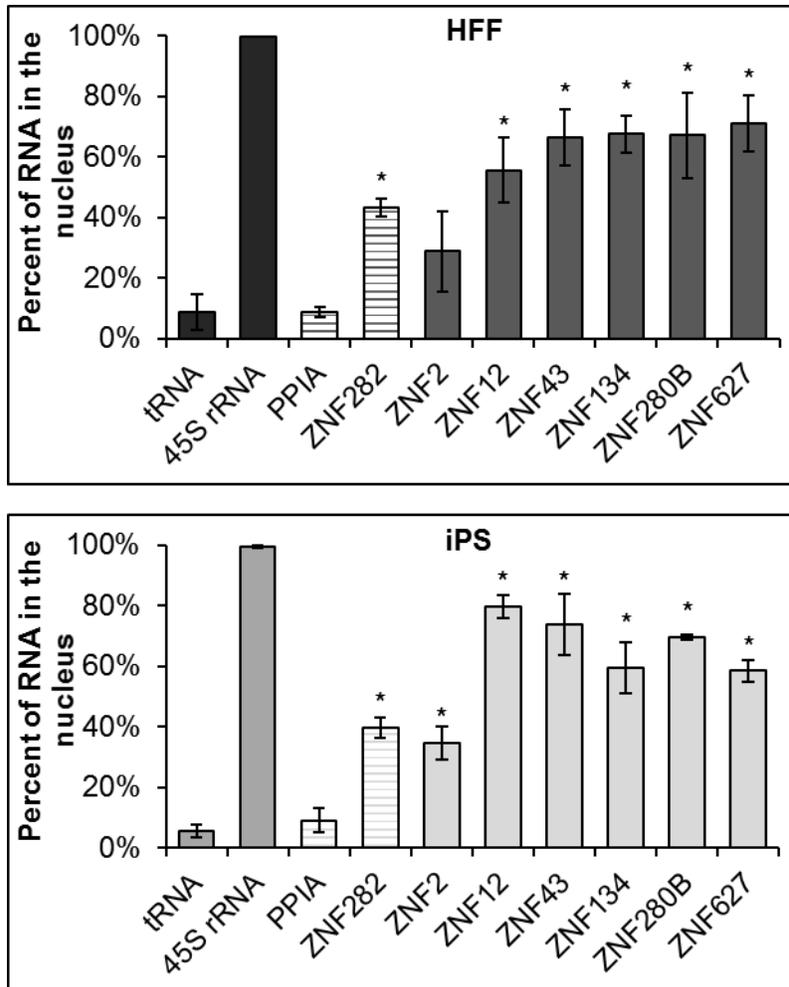
Figure 17: C2H2-ZNF mRNAs accumulate in the nucleus in other human cell lines.
For each cell line, detergent-based subcellular fractionation was used to separate nuclear and cytoplasm cellular fractions. TRIzol was added to each fraction for RNA isolation. Relative abundance of the mRNAs in each fraction was determined using qRT-PCR. The percent of each mRNA in the nucleus was determined. Fractionation controls are shown in the dark bars (tRNA=cytoplasmic control, 45S rRNA=nuclear control). Error bars represent standard deviation of 3 biological replicates. * p<0.05 compared to PPIA by one-way ANOVA followed by Dunnett's post-hoc test for multiple comparisons to one control.
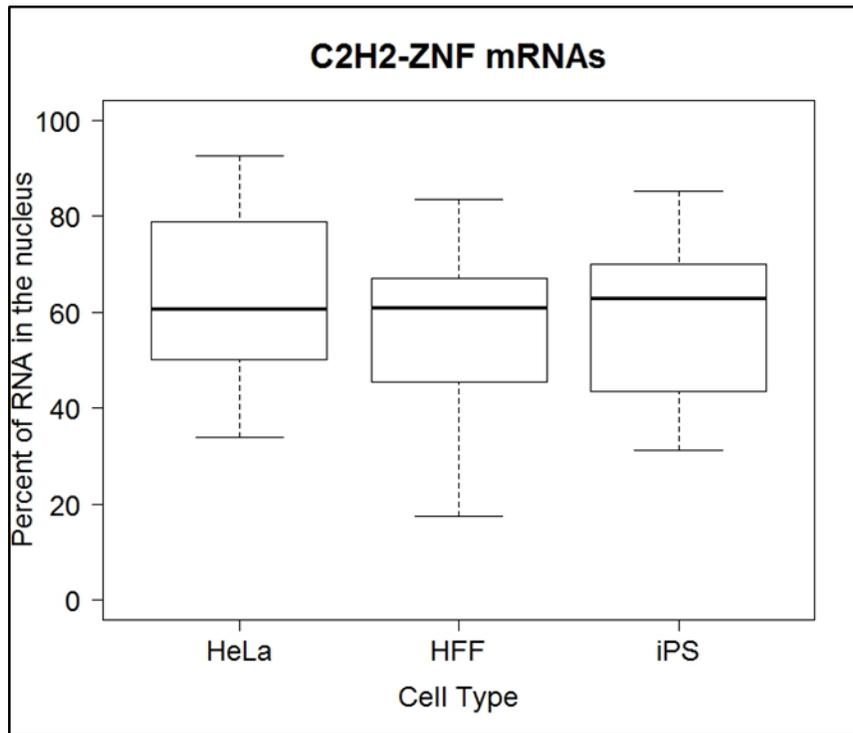
Figure 18: Box-plot comparing the proportion of C2H2-ZNF mRNAs in the nuclear fraction in HeLa, HFF, and iPS cells. There were no statistically significant differences among group means as determined by one-way ANOVA (F=3.13, p=0.054).

candidate for determining the sequences that might influence both nuclear localization and poly(A) tail length.

A schematic of the reporter design is shown in Figure 19A. The pTRE3G-BI-ZsGreen1 plasmid (Clontech) contains a bidirectional tet-off promoter to initiate transcription of both a reporter (in our case Renilla luciferase; RLuc) and ZsGreen mRNAs from the same plasmid. We fused the RLuc sequence to various regions of the from ZNF12 and PPIA transcripts while the ZsGreen mRNA was used as a control. Both RLuc-fusion and ZsGreen mRNAs were subject to the exact same experimental variation (transfection, RNA isolation, oligo (dT) selection, subcellular fractionation and qRT-PCR), making it easier to conclude that any differences were due to biological variation and not sample preparation. In addition, the plasmids contain a chimeric β-globin/IgG intron in the RLuc 5' UTR. Splicing is a key part mRNA processing (Kyburz et al., 2006; Woodward et al., 2016) and the mRNA nuclear export pathway (Björk et al., 2015). Therefore, we included an intron to allow us to assess the possibility that failure to splice is associated with short poly(A) tails and contributes to nuclear retention. The intron also increases the similarity between the reporter and the endogenous ZNF12 mRNA, which has several introns.

3.9 The majority of ZNF12 reporter mRNAs have oligo(A) tails

We cloned the entire ZNF12 ORF and 3' UTR (4507 bp), including the poly(A) signal, in frame with RLuc (Figure 19A). We also removed the plasmid SV40 poly(A) signal and replaced it with the region from the ZNF12 genomic DNA sequence containing the downstream elements (DSE), which are required for efficient polyadenylation (Chen and Wilusz, 1998; Tian and Graber, 2012). We generated a control plasmid containing the complete PPIA ORF and 3' UTR sequences (2199 bp). We used PPIA because we had established that the endogenous transcript was primarily polyadenylated and localized to the cytoplasm (Figures 7 and 13). In addition, we wanted to use a control with a relatively long mRNA sequence rather than RLuc

alone to rule out the possibility that merely placing a long sequence between RLuc sequence and the plasmid poly(A) site would affect polyadenylation. The PPIA reporter contained the SV40 poly(A) signal downstream of the 3' UTR sequence (Figure 19A).

To establish that the ZNF12 and PPIA reporter mRNAs recapitulate the behavior of the endogenous mRNAs, we used the oligo (dT) fractionation assay. For the PPIA reporter construct, ~30% of the RLuc-PPIA mRNAs had oligo (A) tails (Figure 19B). This compares favorably with the endogenous PPIA mRNAs, of which 10% were in the oligo(A) fraction (Figure 7). Furthermore, the ZsGreen mRNAs, which lack introns, were also primarily polyadenylated with <25% having oligo(A) tails. Thus both RLuc-PPIA and ZsGreen mRNAs generally have long poly(A) tails that bind to the oligo (dT)$_{25}$. The ZsGreen mRNAs derived from the ZNF12 plasmid behaved similarly to that from the PPIA construct as ~30% were in the oligo(A) fraction. In contrast, ~90% of the RLuc-ZNF12 mRNA population oligo(A) tails (Figure 19B). These results indicate that the majority of the RLuc-ZNF12 reporter mRNAs have oligo(A) tails. Interestingly, both the PPIA and ZNF12 reporter mRNAs were somewhat more abundant in the oligo(A) fraction compared to the corresponding endogenous mRNAs (Figure 7). This could be connected to the high transcription rates or other promoter-dependent regulation of plasmid-derived versus endogenous mRNAs. Regardless, it was clear that sequences contained within the ORF, 3' UTR and/or downstream region of the ZNF12 gene are sufficient to confer short poly(A) tails on the reporter transcripts.

3.10 The ZNF12 reporter mRNAs are localized in the nucleus

To demonstrate that the RLuc-PPIA and RLuc-ZNF12 reporter mRNAs recapitulate the subcellular localization of the endogenous mRNAs, we used the biochemical subcellular fractionation assay. The RLuc-PPIA reporter mRNAs were ~40% nuclear (Figure 20A). More of the RLuc-PPIA reporter mRNAs were nuclear compared to the endogenous population (~15%, Figure 13); however, the distribution was statistically similar to the ZsGreen mRNAs from the
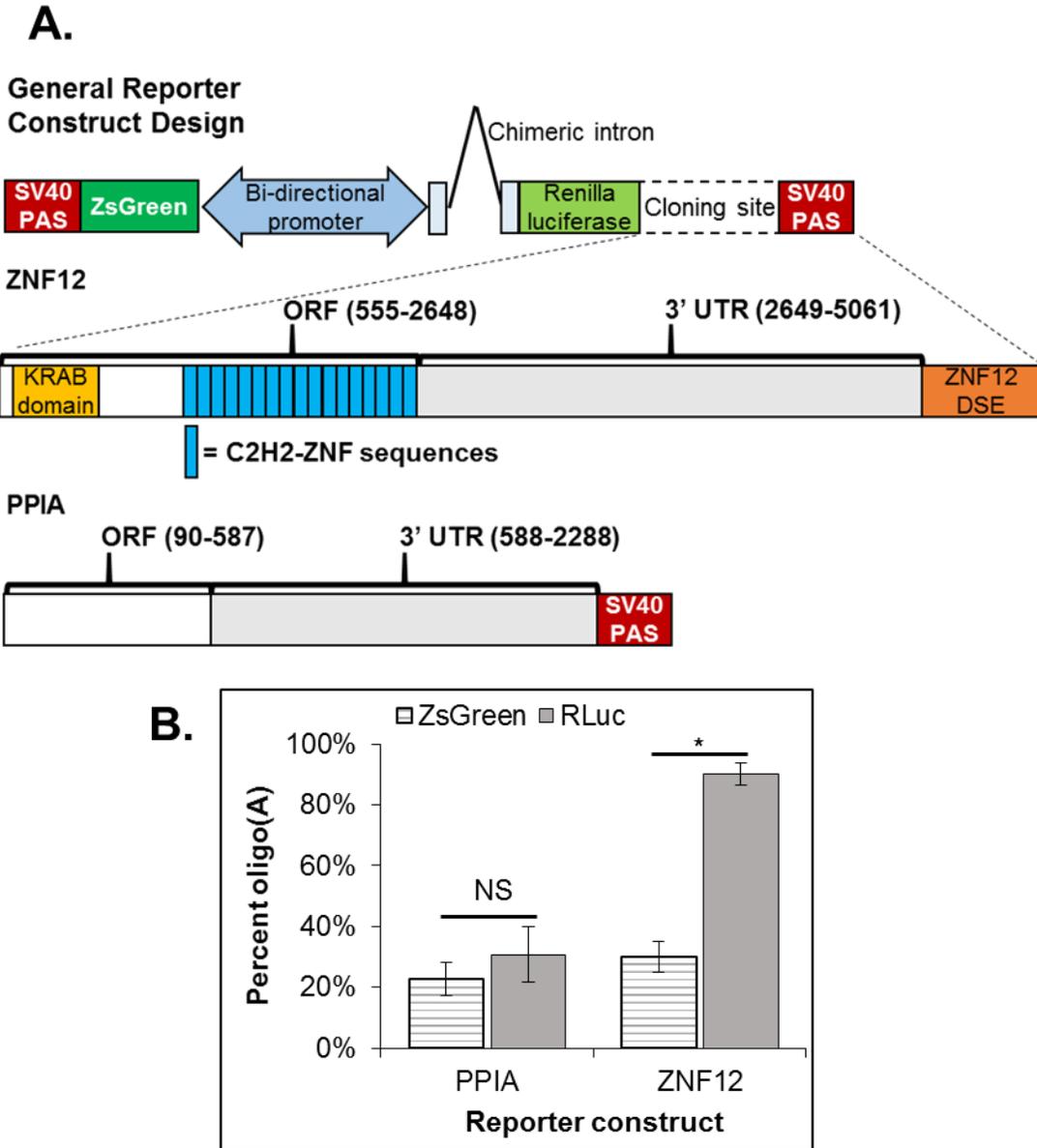
Figure 19: The ZNF12 reporter mRNAs have primarily oligo(A) tails. A. Schematic diagram of reporter constructs. Numbers indicate the nucleotides included from the mRNA sequence (NCBI RefSeq: ZNF12=NM_016265.3, PPIA=NM_021130.4). See text for additional details. **B.** 24-30 h after transfection with reporter plasmids, total RNA was isolated from Tet-Off HeLa cells using TRIzol. Oligo (dT) fractionation was used to separate poly(A) and oligo(A) RNA fractions. qRT-PCR was used to measure the relative abundance of reporter mRNAs in each fraction. The percent of mRNA in the oligo(A) fraction of total was determined. Error bars represent standard deviation of 3 biological replicates. * $p<0.05$ between ZsGreen and the Renilla luciferase (RLuc)-reporter mRNAs from the same plasmid, NS=non-significant (paired, two-tailed Student's T-test).

same plasmid, which were ~30% nuclear (Figure 20A). In contrast, >80% of the RLuc-ZNF12 reporter mRNAs were localized to the nucleus (Figure 20A). These results were highly consistent with endogenous ZNF12 mRNAs, which were ~80% nuclear (Figure 13).

We also performed RNA-fluorescence *in situ* hybridization (RNA-FISH) to verify the subcellular fractionation results, as the mRNAs in the nuclear fraction could instead be associated with the endoplasmic reticulum or other structure connected with the nucleus (Barthelson et al., 2007). In addition, RNA-FISH would reveal whether the RLuc-ZNF12 mRNA has a diffuse distribution throughout the nucleus or if it is associated with specific subnuclear domains. We used a set of probes (Stellaris) targeting RLuc to assess both of the reporter mRNAs. As anticipated, most of the RLuc-PPIA mRNAs were in the cytoplasm while the RLuc-ZNF12 reporter mRNAs were primarily in the nucleus (Figure 20B). Notably, the RLuc-ZNF12 reporter mRNAs exhibited an unusual staining pattern that could represent accumulation in subnuclear bodies such as nuclear speckles or paraspeckles. Nuclear speckles are generally associated with transcription sites and contain factors involved in pre-mRNA splicing and 3' end processing (Spector and Lamond, 2011). Paraspeckles are found near speckles and are associated with nuclear retention of edited RNAs (Sleeman and Trinkle-Mulcahy, 2014). In the future, it will be interesting to determine whether the RLuc-ZNF12 reporter mRNA colocalizes with components of nuclear speckles or paraspeckles as this may provide insights into how and why this mRNA is nuclear retained. Based on the subcellular fractionation and RNA-FISH data, sequences contained within the ORF, 3' UTR and/or downstream region of the ZNF12 gene contribute to nuclear localization of its mRNA.

3.11 The nuclear localization of the ZNF12 reporter mRNA is not due to splicing deficits

Intron retention can prevent export of mRNAs into the cytoplasm and influence polyadenylation (Akef et al., 2015; Braunschweig et al., 2014; Kyburz et al., 2006; Machinaga and Takase-Yoden, 2014; Yap et al., 2012). Therefore, we assessed splicing of the upstream
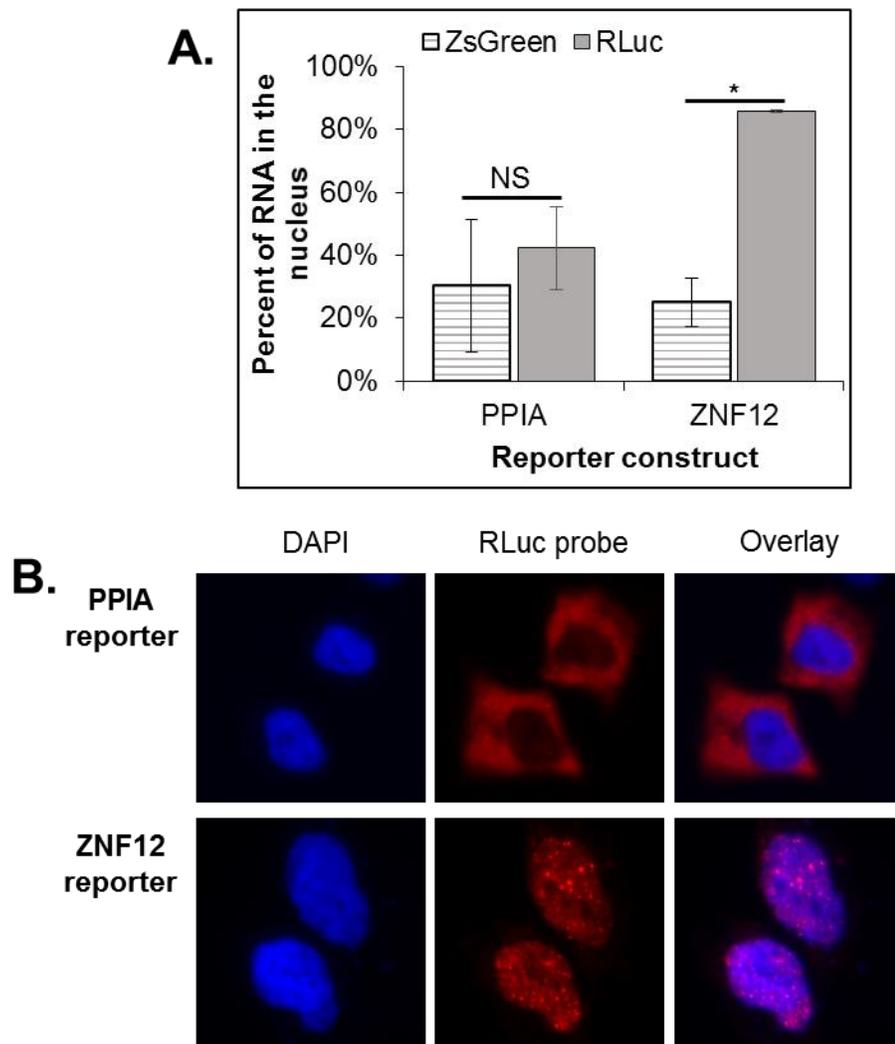
Figure 20: The ZNF12 reporter mRNAs accumulate in the nucleus. A. 24-30 h after transfection with reporter plasmids, detergent-based subcellular fractionation was used to separate nuclear and cytoplasm cellular fractions from Tet-Off HeLa cells. RNA from each fraction was isolated using TRIzol. Relative abundance of the reporter mRNAs in each fraction was determined using qRT-PCR. The percent of each mRNA in the nucleus was determined. Error bars represent standard deviation of 3 (ZNF12) or 4 (PPIA) biological replicates * $p<0.05$ between ZsGreen and the Renilla luciferase-reporter mRNAs from the same plasmid (paired, two-tailed Student's T-test). **B.** RNA-fluorescence *in situ* hybridization (RNA-FISH). Cells were fixed and stained with a set of fluorescently labeled oligo probes targeting the Renilla luciferase (RLuc) mRNA from Stellaris® (red) and nuclei were stained with DAPI (blue). Images are representative of 3 biological replicates.

intron in the RLuc reporter mRNAs using RT-PCR with primers flanking the intron. A significantly higher proportion of the RLuc-ZNF12 reporter mRNA was unspliced in the nucleus compared to the RLuc-PPIA reporter mRNA (Figure 21), suggesting that the RLuc-ZNF12 reporter mRNAs are spliced less efficiently. However, the level of intron retention (~30%) is not sufficient to explain the proportion of the transcript that is retained in the nucleus.

3.12 The ZNF12 ORF and 3' UTR sequences can influence poly(A) tail length and nuclear localization

We next wanted to determine which sequence elements were involved in restricting the poly(A) tail length of the RLuc-ZNF12 mRNA. Most of the characterized sequences that regulate polyadenylation are in the 3' UTR (Matoulkova et al., 2012) and in the region just downstream of the poly(A) site (Chen and Wilusz, 1998; Nunes et al., 2010). We therefore made a reporter plasmid containing the ZNF12 3' UTR, poly(A) signal, and DSE to determine whether one or more of these sequence elements were sufficient to restrict poly(A) tail length. We also made a reporter plasmid containing just the ZNF12 ORF sequence with the SV40 poly(A) signal (Figure 22A). To assess the relative poly(A) tail lengths of these two reporters, we used oligo (dT) fractionation followed by qRT-PCR. Unexpectedly, both the RLuc-ZNF12 ORF and the RLuc-ZNF12 3' UTR reporter mRNAs were significantly more abundant in the oligo(A) fraction than the ZsGreen mRNAs from each reporter plasmid (Figure 22B). We had expected that most of RLuc-ZNF12 3' UTR reporter mRNAs would have oligo(A) tails, but were surprised to see similar results for the Rluc-ZNF12 ORF reporter mRNA population. This was especially unexpected as the ZNF12 ORF reporter contained the SV40 poly(A) signal, suggesting that ZNF12 sequence elements found in the ORF can confer the short poly(A) tail phenotype independent of the poly(A) signal used. We can conclude that sequences in the ORF and 3' UTR independently limit the poly(A) tail length on the ZNF12 mRNA. In addition, the ZNF12
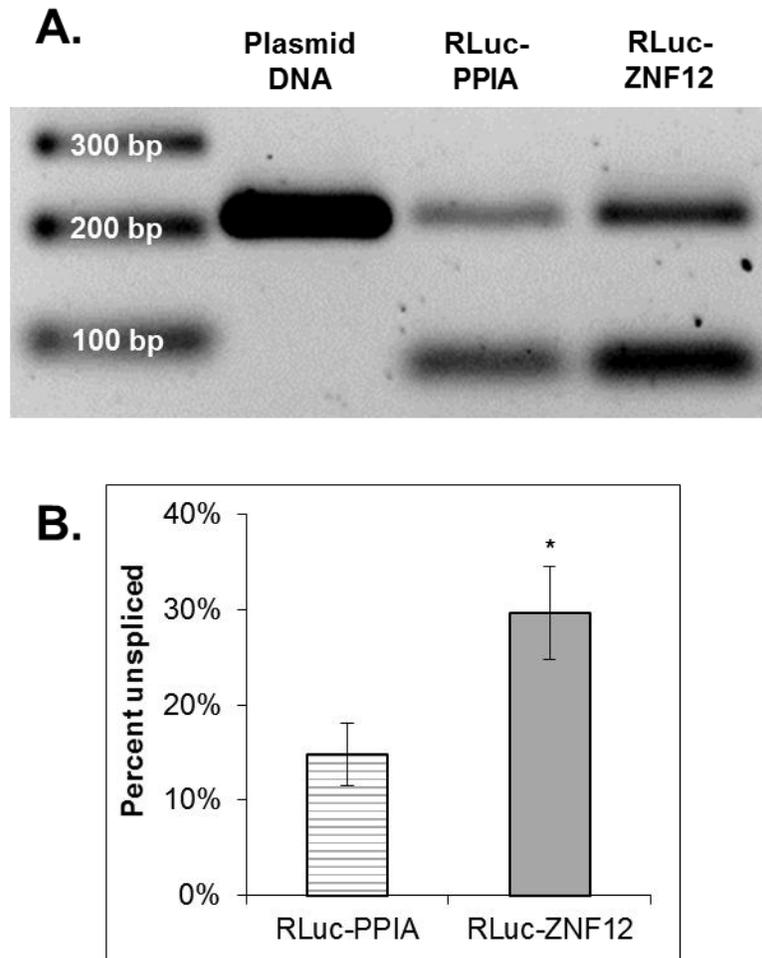
Figure 21: Nuclear localization is not due to inefficient splicing. A. Nuclear RNA from HeLa Tet-Off cells transfected with PPIA or ZNF12 reporter plasmids was analyzed by RT-PCR using primers flanking the intron followed by gel electrophoresis. The intron PCR product from the empty plasmid is shown for reference. **B.** Quantification of intron retention in the nuclear mRNA for each reporter. Relative quantity was determined by measuring the intensity of the unspliced and spliced products and adjusted for size differences. The percent of the intron retained in each reporter mRNA was then determined. Error bars represent standard deviation of 3 (ZNF12) or 4 (PPIA) biological replicates. * $p < 0.05$ compared to RLuc-PPIA (two-tailed Student's T-test).

poly(A) signal and flanking 3' UTR and DSE sequences are not required for restricting the

poly(A) tail length on the ZNF12 mRNAs.

We also assessed the subcellular localization of the RLuc-ZNF12 ORF and RLuc-3' UTR

reporter mRNAs using biochemical subcellular fractionation followed by qRT-PCR (Figure 22C).

Both of these reporter plasmids produced mRNAs that were >80% nuclear, while <30% of the

ZsGreen mRNAs from each of these plasmids were in the nucleus. Thus, sequences located in

both the ZNF12 ORF and 3' UTR can also contribute to the nuclear localization of the ZNF12

mRNAs.

3.13 Reporters mRNAs containing C2H2-ZNF motif encoding sequences have oligo(A) tails and
are localized in the nucleus

We first wanted to narrow down the potential ORF sequences involved in regulating

poly(A) tail length and nuclear localization of the ZNF12 mRNA. We divided the ORF intro three

regions, the N-terminal KRAB domain, a central domain containing 3 C2H2-ZNF motifs (3ZNF)

and the C-terminal domain containing 14 C2H2-ZNF motifs (14ZNF). The RLuc-3ZNF and

RLuc-14ZNF reporter mRNAs would also provide information as to how many C2H2-ZNF motif

encoding sequences may be necessary to confer oligo(A) tails and nuclear localization on the

reporter mRNAs.

To assess how the different ORF regions influence poly(A) tail length of the reporter

mRNAs, we used the oligo (dT) fractionation assay (Figure 23B). Only ~30% of the RLuc-KRAB

reporter mRNAs were in the oligo(A) fraction, which was statistically similar to the ZsGreen

mRNAs (~20% oligo(A)) from the same reporter. In contrast, ~60% of the RLuc-3ZNF reporter

mRNAs and ~80% of RLuc-14ZNF reporter mRNAs were in the oligo(A) fraction. This suggests

that C2H2-ZNF encoding sequences may be involved in conferring the oligo(A) tails.

Next, we assessed how the different ORF sequences would influence mRNA subcellular

localization (Figure 23C). Only ~40% of the RLuc-KRAB reporter mRNAs were nuclear,
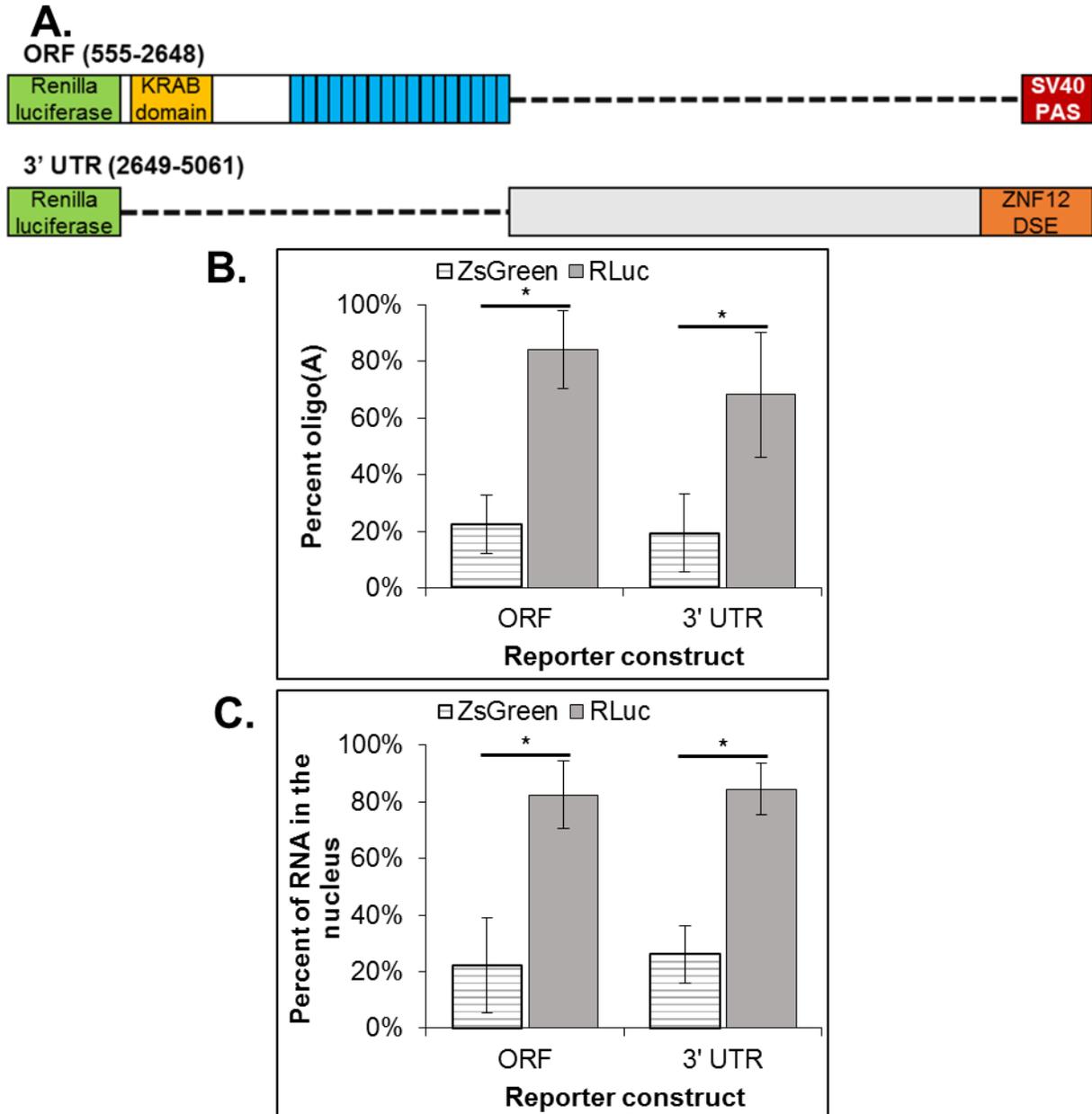
Figure 22: ZNF12 ORF and ZNF12 3'UTR reporter mRNAs have primarily oligo(A) tails and are retained in the nucleus. A. Schematic diagram of reporter constructs. **B.** 24-30 h after transfection with reporter plasmids, total RNA was isolated from Tet-Off HeLa cells using TRIzol. Oligo (dT) fractionation was used to separate poly(A) and oligo(A) RNA fractions. The percent of mRNA in the oligo(A) fraction of total was determined using qRT-PCR. **C.** 24-30 h after transfection, detergent-based subcellular fractionation was used to separate nuclear and cytoplasm cellular fractions from Tet-Off HeLa cells. TRIzol was added to each fraction for RNA isolation. The percent of each mRNA in the nucleus was determined using qRT-PCR. Error bars represent standard deviation of 3 biological replicates. * $p < 0.05$ between ZsGreen and the Renilla luciferase-reporter mRNAs from the same plasmid (paired, two-tailed Student's T-test).

suggesting that the KRAB region does not influence nuclear localization. In contrast, ~70% of the RLuc-3ZNF reporter mRNAs and ~80% of RLuc-14ZNF reporter mRNAs were nuclear. Therefore, it appears that in addition to regulating poly(A) tail length, the C2H2-ZNF motif sequences may also contribute to mRNA nuclear localization.

3.14 The ZNF12 3' UTR contains C2H2-ZNF-like sequences

The C2H2-ZNF motif sequences in the ORF appeared to confer the oligo(A) tails and contribute to nuclear retention of the ZNF12 mRNAs, but it was not clear why the 3' UTR reporter mRNAs also had oligo(A) tails and were abundant in the nucleus. Analysis using the Basic Local Alignment Search Tool (BLAST from NCBI) to compare the ZNF12 ORF and 3' UTR sequences revealed two regions (labeled E and F) in the 3' UTR that showed at least 60% sequence homology to multiple C2H2-ZNF motif sequences (labeled A-D) (Figure 24A). These 3' UTR regions may be degenerate ZNFs that are no longer functional protein elements, but are instead remnants produced during the evolution of the ZNF12 gene (Looman, 2002). In addition, the reporter mRNAs from each of the ZNF12 reporter constructs that contained one or more of the homologous sequences (ZNF12, ORF, 3' UTR, 3ZNFs, and 14ZNFs) were primarily in the oligo(A) and nuclear fractions (Figure 24B). Thus every functional region of the ZNF12 mRNA shares some sequence homology supporting the idea that C2H2-ZNF motif sequences are responsible for the oligo(A) tails and nuclear localization of the ZNF12 transcript.

3.15 Reporter mRNAs containing the ZNF-like sequence from the ZNF12 3' UTR have oligo(A) tails and are retained in the nucleus

We wanted to confirm that reporter mRNAs with the ZNF-like sequence in the 3' UTR would have oligo(A) tails and would be abundant in the nucleus. Therefore, we made a reporter containing the region of the 3' UTR with the ZNF-like sequences (3' UTRΔ). As the previous
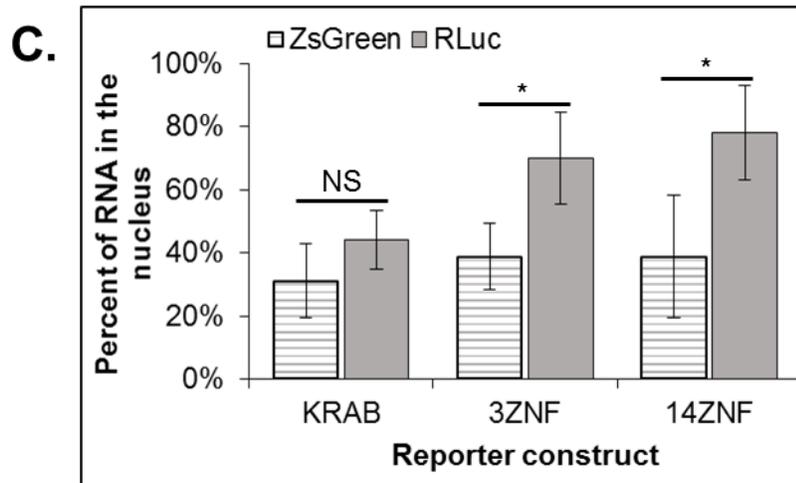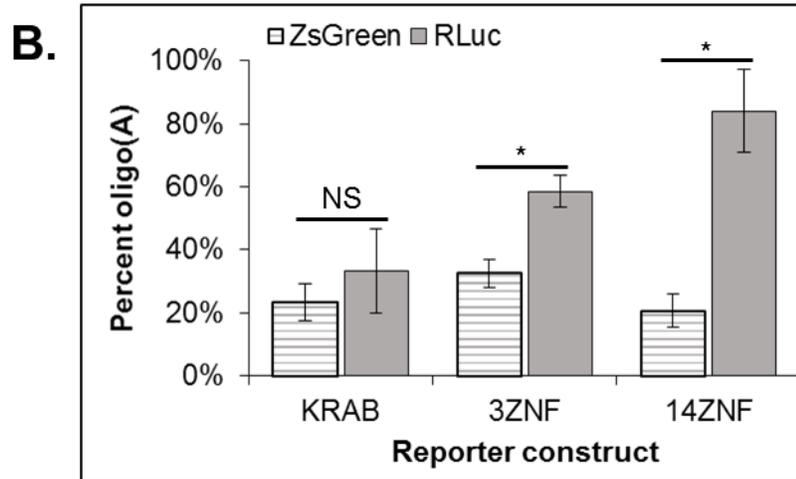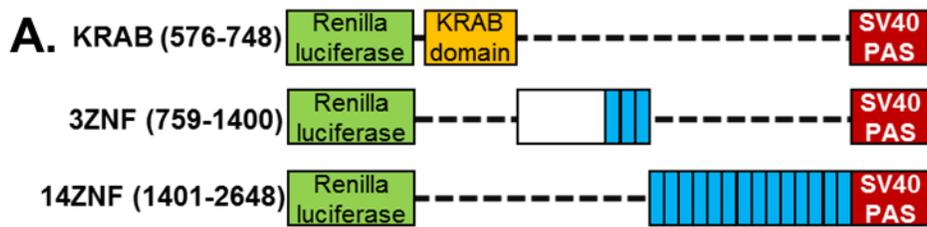
Figure 23: Reporter mRNAs with C2H2-ZNF sequences primarily have oligo(A) tails and are retained in the nucleus. A. Schematic diagram of reporter constructs. **B.** 24-30 h after transfection with reporter plasmids, total RNA was isolated from Tet-Off HeLa cells using TRIzol. Oligo (dT) selection was used to separate poly(A) and oligo(A) RNA fractions. The percent of mRNA in the oligo(A) fraction of total was determined using qRT-PCR. Error bars represent standard deviation of 3 biological replicates. **C** Tet-Off HeLa cells were transfected with reporter plasmids. 24-30 h after transfection, detergent-based subcellular fractionation was used to separate nuclear and cytoplasm cellular fractions. TRIzol was added to each fraction for RNA isolation. The percent of each mRNA in the nucleus was determined using qRT-PCR. Error bars represent standard deviation of 3 (3ZNF and 14ZNF) or 4 (KRAB) biological replicates. * $p < 0.05$ between ZsGreen and the Renilla luciferase-reporter mRNAs from the same plasmid (paired, two-tailed Student's T-test).

A.

Location of homologous
sequences in ZNF12 mRNA

ORF (555-2648)   3'UTR (2649-5061)

KRAB
domain

A B   C D   E F

A (ORF 1331-1432)   AGACATCTCATATGGAA---ATGAAGCCCTATGAATGCAGTGAATGTGGGAAATCCTTCTG
B (ORF 1923-2020)   AG--AACTCATTCAGGA---GAGAAACCCTATGAATGTAATGAATGTGGAAAAACCTTCTA
C (ORF 2091-2188)   AG--AACTCATTCAGGA--GTAAAACCCTATGAATGTAGTGAATGTGGGAAAACCTTCTA
D (ORF 2273-2356)   ---------------A--GAGAAGCCCTATGAATGTAGTGAATGTGGGAAAACCTTCTG
E (3'UTR 2652-2747) AG--ACCTCATTTTATATCAGAGAACCCTTTCAGTATAGTGAATCAGA---AA-CTCCTG

A (ORF 1331-1432)   TAAAAAGTCAAAATTTATTATACATCAGAGGACTCACACAGG
B (ORF 1923-2020)   CCTGAATTCAGCCCTCATGAGACATCAGAGAGTGCACACAGG
C (ORF 2091-2188)   CCAGAACTCAGCCCTTTGTAGACATCGGAGAATACACAAAGG
D (ORF 2273-2356)   CCAGAATTCAGCCCTTAATCGACATCAGAGAACACACACAGG
E (3'UTR 2652-2747) CCTGAAGTCAAACACCT-TGTACATCAGAGAGTTCACACAGG

B (ORF 1916-2052)   ACATTA-TAGAACTCATTCAGGAGAGAAACCCTATGAATGTAATGAATGTGGAAAAACCT
F (3'UTR 2836-2970) ACAATACTAGA-CT-ATAGACAAGTAAAAATTTATAAATATTAAGAATGT--ATATACAT

B (ORF 1916-2052)   -TCTACC-TGAATTCAGCCCTCATGAG-ACATCAGAGAGTGCACA--CAGGAGAGAAACC
F (3'UTR 2836-2970) GTC-ACCATGGATTG-GAACTGTTTTGCATATCAGG-GAAATCATAGCCAAGGGGAAATC

B (ORF 1916-2052)   TTACGAATGTAATGAATGTGGAA
F (3'UTR 2836-2970) TATC-AGTATAAGGAATGTGGAA

B.

Reporter name	>50% oligo(A)	>50% nuclear	Homologous sequences
ZNF12	Yes	Yes	A-F
ORF	Yes	Yes	A-D
KRAB	No	No	None
3ZNF	Yes	Yes	A
14ZNF	Yes	Yes	B-D
3' UTR	Yes	Yes	E,F

**Figure 24: Sequences alignment of homologous regions in the ZNF12 ORF and the 3'
UTR. A.** Homologous sequences were detected by using BLAST to compare the ZNF12
ORF and 3' UTR sequences (NM_016265.3). The sequences were aligned using the T-
Coffee webserver (Notredame et al., 2000) and final image was produced using the
BoxShade webserver. **B.** Table indicating which of the homologous sequences are in each
of the ZNF12 reporter constructs.

reporter mRNA results indicated that the ZNF12 poly(A) signal and downstream elements were not required for regulating poly(A) tail length and subcellular localization, the SV40 poly(A) signal was used for this reporter plasmid (Figure 25). Similar to the full length RLuc-3' UTR reporter mRNA, the RLuc-3' UTRΔ reporter mRNA was primarily in the oligo(A) fraction and the majority was localized in the nucleus.

3.16 Subcellular localization is strongly correlated with relative poly(A) tail length for the RLuc-ZNF12 reporter mRNAs

A linear regression analysis of percent nuclear versus percent oligo(A) for the different RLuc-ZNF12 reporter mRNAs suggests that subcellular localization is strongly correlated with relative poly(A) tail length (Figure 26). Notably, this correlation also appears to be connected to the number of C2H2-ZNF motif or ZNF-like sequences in each RLuc-ZNF12 reporter mRNA. The RLuc-ZNF12 reporter mRNAs with >14 C2H2-ZNF encoding sequences (ZNF12, ORF, and 14ZNF reporter plasmids) were more abundant in both the oligo(A) and nuclear fractions compared to the RLuc-ZNF12 reporter mRNAs with 3 or fewer C2H2-ZNF sequences (ZNFs A, 3' UTR, and 3' UTRΔ). The RLuc-KRAB mRNA did not have any C2H2-ZNF encoding sequences and was the least abundant in the oligo(A) and nuclear fractions. We noted the RLuc-3' UTR reporter mRNAs were more abundant in both the oligo(A) and nuclear fractions compared to the RLuc-mRNAs from the ZNFs A and 3' UTRΔ reporter plasmids. The complete ZNF12 3' UTR may contain additional remnants of C2H2-ZNF sequences or other sequences that further enhance nuclear localization and poly(A) tail length restriction. In addition, 3' UTR elements may be more divergent compared to the ORF sequences, but also more powerful as regulatory elements since these sequences do not need to retain protein coding abilities. Overall, the RLuc-ZNF12 reporter mRNA analysis suggests that C2H2-ZNF encoding sequences are involved in regulating ZNF12 mRNA poly(A) tail length and nuclear retention.
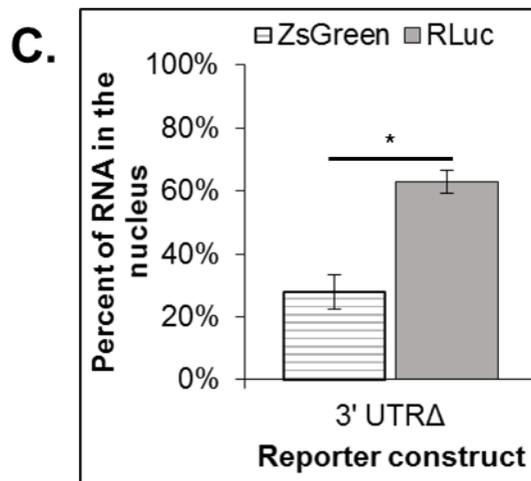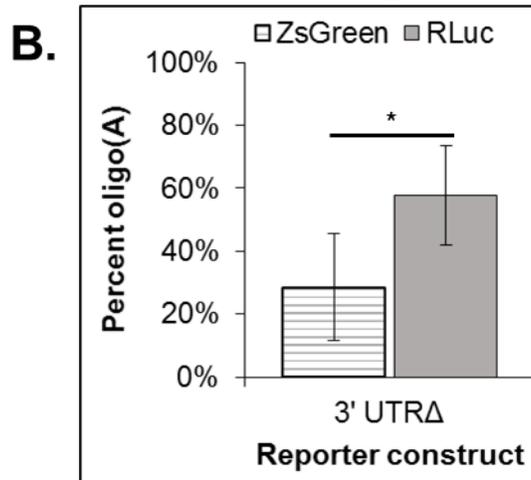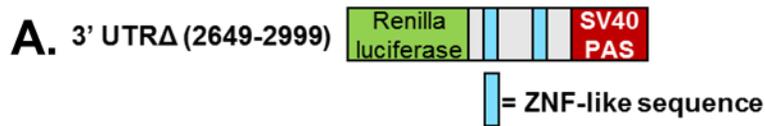
Figure 25: A C2H2-ZNF-like sequence from the ZNF12 3' UTR produces reporter mRNAs with primarily oligo(A) tails and are retained in the nucleus. A. Schematic diagram of the reporter construct. **B.** 24-30 h after transfection with reporter plasmids, total RNA was isolated from Tet-Off HeLa cells using TRIzol. Oligo (dT) selection was used to separate poly(A) and oligo(A) RNA fractions. The percent of mRNA in the oligo(A) fraction of total was determined using qRT-PCR. **C.** 24-30 h after transfection, detergent-based subcellular fractionation was used to separate nuclear and cytoplasm cellular fractions from Tet-Off HeLa cells. TRIzol was added to each fraction for RNA isolation. The percent of each mRNA in the nucleus was determined using qRT-PCR. Error bars represent standard deviation of 3 biological replicates. * $p < 0.05$ between ZsGreen and the Renilla luciferase-reporter mRNAs from the same plasmid (paired, two-tailed Student's T-test).

Figure 26: Percent of the various ZNF12 reporter mRNAs in the nuclear fraction is positively correlated with the percent in the oligo(A) faction. Scatter plot of the average percent nuclear versus average percent oligo(A) for each of the mRNAs assessed by subcellular fractionation and oligo (dT) fractionation.

CHAPTER 4: DISCUSSION

4.1 Summary of results

Consistent with previous findings (Yang et al., 2011b), here we demonstrated that large populations of six different C2H2-ZNF mRNAs persist with short poly(A) tails at steady-state conditions in HeLa, iPS cells and HFFs. The majority of nascent mRNAs for each of the six C2H2-ZNFs have short poly(A) tails, suggesting that poly(A) tail length is restricted in the nucleus around the time of transcription. We also found that C2H2-ZNF mRNAs accumulate in the nuclear fraction in HeLa, iPS cells, and HFFs, which may be related to the short poly(A) tails of these transcripts. To identify sequence elements involved in regulating poly(A) tail length and/or nuclear localization, we created a series of reporter constructs containing sequences from the ZNF12 mRNA. Reporter mRNAs with sequences from the ZNF12 ORF and/or the 3' UTR had short poly(A) tails and were primarily localized in the nucleus. Further deletion analysis indicated that repeated sequences encoding the C2H2-ZNF protein motifs are involved in poly(A) tail length restriction and nuclear retention of these mRNAs. In addition, remnants of C2H2-ZNF motif sequences from the ZNF12 3' UTR are also able to confer short poly(A) tails and contribute to nuclear retention of the reporter mRNAs. RNA-fluorescent *in situ* hybridization (RNA-FISH) revealed that ZNF12 reporter transcripts are found in foci within the nucleus that could represent sites for mRNA storage. Overall, our findings indicate that short poly(A) tails and nuclear localization are features of multiple C2H2-ZNF mRNAs in human cells. In addition, conserved C2H2-ZNF motif sequences appear to be involved in regulating poly(A) tail length and nuclear retention of the C2H2-ZNF mRNAs.

4.2 Possible connections between the short poly(A) tails and nuclear localization of the C2H2-ZNF mRNAs

We observed a positive correlation between the proportion of mRNAs in the nucleus and the proportion of mRNAs with short poly(A) tails for multiple endogenous C2H2-ZNF mRNAs (Figure 14) and for the ZNF12 reporter mRNAs (Figure 26). These findings suggest a potential connection between the regulation of poly(A) tail length and the nuclear localization of the C2H2-ZNF mRNAs. One possibility is that C2H2-ZNF mRNA poly(A) tail lengths may be restricted around the time of transcription (Figure 10) to prevent efficient binding of PABPN1, which is involved in nuclear export (Bear et al., 2003) and in facilitating the nuclear decay of polyadenylated mRNAs (Bresson and Conrad, 2013; Meola et al., 2016). Thus, restriction of poly(A) tail length could be a key aspect in enabling the nuclear accumulation of the C2H2-ZNF mRNAs. In the future, it would be interesting to determine if the underlying reason for the short poly(A) tails is indeed to prevent the binding of PABPN1. This could be investigated by analyzing the ZNF12 reporter mRNA tethered to PABPN1 or containing a synthetic long poly(A) tail (Muniz et al., 2015) to determine if this mRNA no longer efficiently accumulates in the nucleus.

While the mechanisms responsible for regulating poly(A) tail length and promoting nuclear accumulation remain to be determined, analysis of the ZNF12 reporter mRNAs indicated that the C2H2-ZNF motif encoding sequences are involved in these processes (Figure 24). An RNA-binding protein or other factor may interact with the C2H2-ZNF motif sequences to control poly(A) tail length and mediate nuclear retention. This factor could control poly(A) tail length by preventing efficient polyadenylation or by recruiting a nuclear deadenylase to remove the poly(A) tails from the C2H2-ZNF transcripts shortly after nuclear polyadenylation. In addition, this factor could also have a role in enhancing nuclear retention, possibly by facilitating localization to nuclear foci where the C2H2-ZNF mRNAs may be stored (discussed in Section 4.3). Since a proportion of the C2H2-ZNF mRNAs we analyzed had long poly(A) tails and were

located in the cytoplasm, the activity of any factor involved in these processes is limited. This

limitation could be variable accessibility of the RNA element to its binding partner. It is possible

that the binding sites in a proportion of the C2H2-ZNF transcripts are hidden by other proteins.

Changes in RNA secondary structure could also influence binding and activity of RNA-

regulatory factors (Kligun and Mandel-Gutfreund, 2015; Li et al., 2014). In addition, some factors

recognize and bind to particular RNA structural elements, which could contain the C2H2-ZNF

motif sequences. If these RNA structures are altered, it can affect the ability of proteins or other

RNAs to interact with their targets.

Identification of factors associated with the ZNF motif sequences of the nuclear C2H2-

ZNF transcripts that have short poly(A) tails will likely provide key information regarding the

post-transcriptional regulation of these mRNAs. In addition, identification of such a factor could

help uncover the full spectrum of C2H2-ZNF mRNAs that are regulated in this way. If these

mRNAs encode C2H2-ZNF proteins that are involved in the same cellular events or pathways,

this would suggest that poly(A) tail length regulation and nuclear retention are involved in

coordinately regulating C2H2-ZNF gene expression.

4.3 The nuclear foci could be paraspeckles

Using RNA-FISH, we observed that the RLuc-ZNF12 reporter mRNA localized to nuclear

foci (Figure 20B). One interesting possibility is that these sites are paraspeckles, which are non-

membranous subnuclear bodies containing the long non-coding RNA NEAT1, proteins involved

in RNA processing, and RNAs that have undergone adenosine to inosine (A-to-I) deamination

by ADAR (Morimoto and Boerkoel, 2013). Interestingly, paraspeckle formation fluctuates during

the circadian cycle to potentially regulate protein levels by controlling rhythmic nuclear retention

of mRNAs (Torres et al., 2016, 2017). Additional investigation may identify other cellular

conditions that regulate paraspeckle formation as a potential mechanism to coordinate nuclear

retention of mRNAs during significant biological events such as stress, disease, cell growth and differentiation, or the cell cycle.

Notably, many of the edited paraspeckle-associated RNAs contain *Alu* elements, which coincidentally are important features of C2H2-ZNF evolution. The exonization of *Alu* elements contributed to the formation of novel transcripts with altered translation efficiency, a process which may have influenced the primate-specific evolution of the C2H2-ZNF genes (Shen et al., 2011). While numerous C2H2-ZNFs transcripts possess *Alu* elements, these are not present in all of the nuclear retained C2H2-ZNFs mRNAs we assessed (Athanasiadis et al., 2004; Chen et al., 2008; Ramaswami et al., 2013). However, it is possible that *Alu*-containing C2H2-ZNF transcripts could mediate the localization of other C2H2-ZNF mRNAs to paraspeckles through *trans* regulation. This type of regulation allows mRNAs with a specific regulatory element to impact the behavior of other mRNAs that lack this element (Reveal et al., 2010; Wilusz and Wilusz, 2010). For example, in *Drosophila*, a translational control element in the *oskar* mRNA can regulate the translation of mRNAs lacking this element. This interaction is mediated by the polypyrimidine tract binding (PTB) RNA-binding protein, which brings the *trans*-acting mRNAs in close proximity (Macdonald et al., 2016). Perhaps an RNA-binding protein or other factor interacts with conserved C2H2-ZNF motif sequences found in C2H2-ZNF mRNAs with and without *Alu* elements. This factor may then facilitate *trans* regulation that facilitates retention of various C2H2-ZNF transcripts in paraspeckles. Disruption of this factor's interaction with the C2H2-ZNF mRNAs or dissolution of paraspeckles may then allow export of these mRNAs to the cytoplasm where they can undergo translation.

Future studies will determine whether the RLuc-ZNF12 reporter mRNA co-localizes with NEAT1 or paraspeckle proteins, such as PSF or p54/NRB (Fox and Lamond, 2010), to determine if the nuclear foci are indeed paraspeckles. Depletion of cellular NEAT1 to disrupt paraspeckle formation (Clemson et al., 2009) may prevent the nuclear localization of the C2H2-ZNF mRNAs. In rat pituitary cells, multiple C2H2-ZNF mRNAs interact with NEAT1 (Torres et

al., 2016). It would be interesting to see if this also occurs in human cells, which have a different array of C2H2-ZNF transcripts. Together, these experiments could point to potential mechanisms that contribute to nuclear retention of the C2H2-ZNF mRNAs.

4.4 Potential implications of C2H2-ZNF mRNA short poly(A) tails and nuclear localization

4.4.1 Coordinated regulation of gene expression

Many C2H2-ZNF proteins are transcription factors that influence the expression of genes involved in key biological process such as development, differentiation, cell-cycle progression, oncogenesis, and apoptosis (Jen and Wang, 2016; Razin et al., 2012). Therefore, coordinated regulation of C2H2-ZNF protein levels could promote significant changes in gene expression programs. As some of these changes may need to occur rapidly, post-transcriptional mechanisms likely have key roles in modulating C2H2-ZNF protein abundances (Liu et al., 2016). For instance, mature mRNAs that have been temporarily retained in the nucleus and are released into the cytoplasm would be available for translation more quickly than mRNAs that must be transcribed and fully processed prior to nuclear export. Perhaps a population of C2H2-ZNF mRNAs is retained in the nucleus to facilitate rapid production of these transcription factors, which would subsequently influence the expression of multiple targeted genes. Polyadenylation of the C2H2-ZNF mRNAs may be the first step in inducing their release from the nucleus. Alternatively, C2H2-ZNF mRNAs may not need to be polyadenylated, but could be exported with short poly(A) tails and interact with protein complexes or form 3' end structures that allow efficient translation.

Another possibility is that the short poly(A) tails and nuclear retention are mechanisms involved in fine-tuning C2H2-ZNF protein levels. Since many of these proteins are transcription factors that can influence the expression of numerous genes, the abundance of the C2H2-ZNF proteins must be carefully regulated. Therefore, a proportion of the C2H2-ZNF transcripts may be retained in the nucleus until additional proteins need to be synthesized. As another level of

post-transcriptional control, the short poly(A) tails of the C2H2-ZNF mRNAs may need to undergo cytoplasmic polyadenylation before these mRNAs can be translated (Reyes and Ross, 2016).

Future investigation is needed to determine if nuclear localization and/or poly(A) tail length of multiple C2H2-ZNF mRNAs changes in response to cellular events such as stress, cell differentiation or proliferation, or cancer progression. The outcomes of these studies could provide insight into how regulation of nuclear export and poly(A) tail length could be involved in coordinately controlling the expression and activity of the C2H2-ZNF transcription factors.

4.4.2 C2H2-ZNF mRNAs may have dual functions

Numerous mRNAs can have both coding and non-coding functions; therefore, it is possible that the nuclear-retained C2H2-ZNF mRNAs may have another purpose. Bifunctional RNAs can influence gene expression by regulating transcription and translation, contribute to the formation of non-membranous cellular bodies, and serve as protein scaffolds (Nam et al., 2016; Sampath and Ephrussi, 2016). In addition, mRNAs can competitively bind non-coding RNAs, such as miRNAs, and RNA-binding proteins to control the amount of these factors that are functionally available (Gardiner et al., 2015; Jens and Rajewsky, 2014; Tay et al., 2014).

There are several potential functional roles of the nuclear-retained C2H2-ZNF mRNAs. Recent analysis revealed that the C2H2-ZNF mRNAs were among the RNAs bound to chromatin-associated proteins, including those involved in gene-silencing, but the implications of such interactions have yet to be delineated (Hendrickson et al., 2016). One possibility is that the nuclear-retained C2H2-ZNF mRNAs are involved in concentrating proteins at specific chromatin sites to alter gene expression. Alternatively, these C2H2-ZNF mRNAs could sequester nuclear proteins to inhibit their functions. Continued analysis of the proteins that interact with the C2H2-ZNF mRNAs could provide insight into their possible roles in regulating protein activity. It is also

possible that the C2H2-ZNF mRNAs are competing endogenous RNAs (ceRNAs), which "sponge up" miRNAs to inhibit their regulation of other mRNAs (Figure 27) (Bak and Mikkelsen, 2014; Thomson and Dinger, 2016). Numerous C2H2-ZNF mRNAs sequences contain multiple miRNA binding sites (Schnall-Levin et al., 2011), potentially increasing their regulatory capacity as ceRNAs. While miRNA-mediated regulation of mRNAs generally occurs in the cytoplasm (Iwakawa and Tomari, 2015), some miRNAs are also present in the nucleus and may shuttle between locations (Liao et al., 2010; Solnestam et al., 2012). The C2H2-ZNF transcripts may be able to sequester these miRNAs in the nucleus to prevent their interactions with cytoplasmic targets. It would be interesting to determine if overexpression of these C2H2-ZNF transcripts would increase the stability and translation of other mRNAs that are targeted by the same miRNAs. The results of such a study could indicate that certain C2H2-ZNF mRNAs are involved in post-transcriptionally regulating gene expression by repressing miRNA activity.

4.5 Conclusions

In this study, we identified highly conserved and repetitive sequence elements that not only encode key protein domains, but may also be involved in regulating post-transcriptional processes that modify the molecular behavior of the C2H2-ZNF mRNAs. In addition, this study highlights the importance of investigating alterations in mRNA poly(A) tail length and subcellular localization. Numerous mRNAs besides those encoding the C2H2-ZNF proteins have short poly(A) tails (Chang et al., 2014; Choi and Hagedorn, 2003; Meijer et al., 2007; Subtelny et al., 2014; Yang et al., 2011b) and/or are enriched in the nucleus (Bahar Halpern et al., 2015; Barthelson et al., 2007; Djebali et al., 2012; Solnestam et al., 2012; Trask et al., 2009). Most gene expression studies focus on transcripts from whole-cells without considering poly(A) tail length. Therefore, the influence that alterations in poly(A) tail status and subcellular localization could have on regulating gene expression is not well-understood. C2H2-ZNF and other mRNAs
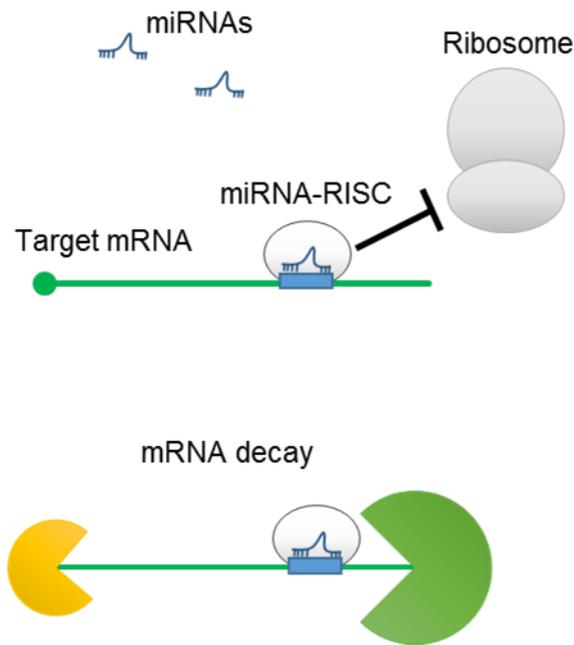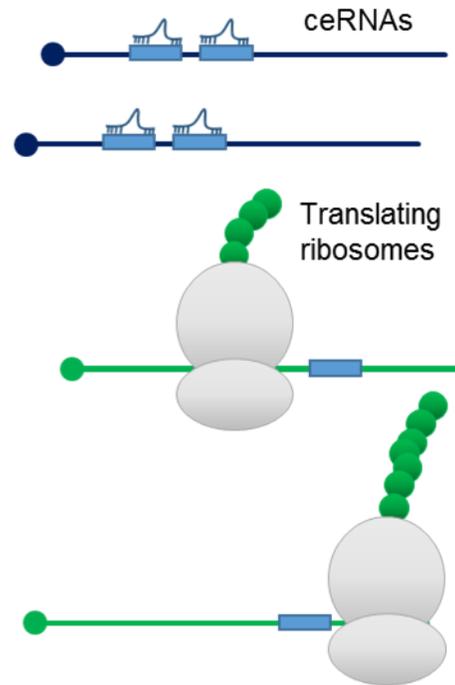
Figure 27: Mechanism of action of competing endogenous RNAs (ceRNAs).
A. Without ceRNAs present, miRNAs are readily available to interact with target mRNAs to inhibit translation and promote mRNA decay, leading to decreased protein yields. **B.** High levels of ceRNAs cause miRNA sequestration, and mRNAs are no longer targeted for miRNA-mediated repression. These mRNAs can now be translated, increasing protein synthesis.

that have atypical features could also have additional functions involved in important cellular

processes. Clearly, further investigation is needed to reveal how variations in mRNA poly(A) tail

length and subcellular localization overall impact cell biology.

REFERENCES

Akef, A., Lee, E.S., and Palazzo, A.F. (2015). Splicing promotes the nuclear export of *β-globin* mRNA by overcoming nuclear retention elements. RNA *21*, 1908–1920.

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA Editing of Alu-Containing mRNAs in the Human Transcriptome. PLoS Biol. *2*, e391.

Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., and Itzkovitz, S. (2015). Nuclear Retention of mRNA in Mammalian Tissues. Cell Rep. *13*, 2653–2662.

Bak, R.O., and Mikkelsen, J.G. (2014). miRNA sponges: soaking up miRNAs for regulation of gene expression: MicroRNA sponges. Wiley Interdiscip. Rev. RNA *5*, 317–333.

Banerjee, A., Apponi, L.H., Pavlath, G.K., and Corbett, A.H. (2013). PABPN1: molecular function and muscle disease. FEBS J. *280*, 4230–4250.

Barreau, C. (2005). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res. *33*, 7138–7150.

Barthelson, R.A., Lambert, G.M., Vanier, C., Lynch, R.M., and Galbraith, D.W. (2007). Comparison of the contributions of the nuclear and cytoplasmic compartments to global gene expression in human cells. BMC Genomics *8*, 340.

Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of Transcript Variability in Single Mammalian Cells. Cell *163*, 1596–1610.

Bear, D.G., Fomproix, N., Soop, T., Björkroth, B., Masich, S., and Daneholt, B. (2003). Nuclear poly(A)-binding protein PABPN1 is associated with RNA polymerase II during transcription and accompanies the released transcript to the nuclear pore. Exp. Cell Res. *286*, 332–344.

Bellefroid, E.J., Marine, J.C., Ried, T., Lecocq, P.J., Rivière, M., Amemiya, C., Poncelet, D.A., Coulie, P.G., de Jong, P., and Szpirer, C. (1993). Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. EMBO J. *12*, 1363–1374.

Bergeron, D., Pal, G., Beaulieu, Y.B., Chabot, B., and Bachand, F. (2015). Regulated Intron Retention and Nuclear Pre-mRNA Decay Contribute to *PABPN1* Autoregulation. Mol. Cell. Biol. *35*, 2503–2517.

Bernard, D., Prasanth, K.V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., Zhang, M.Q., Sedel, F., Jourdren, L., Coulpier, F., et al. (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J. *29*, 3082–3093.

Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions. Cell *150*, 279–290.

Bienroth, S., Keller, W., and Wahle, E. (1993). Assembly of a processive messenger RNA polyadenylation complex. EMBO J. *12*, 585–594.

Björk, P., Persson, J.-O., and Wieslander, L. (2015). Intranuclear binding in space and time of exon junction complex and NXF1 to premRNPs/mRNPs in vivo. J. Cell Biol. *211*, 63–75.

Bonasio, R., Lecona, E., Narendra, V., Voigt, P., Parisi, F., Kluger, Y., and Reinberg, D. (2014). Interactions with RNA direct the Polycomb group protein SCML2 to chromatin where it represses target genes. eLife *3*, e02637.

Bond, G.L., Prives, C., and Manley, J.L. (2000). Poly(A) polymerase phosphorylation is dependent on novel interactions with cyclins. Mol. Cell. Biol. *20*, 5310–5320.

Bonnet, A., and Palancade, B. (2014). Regulation of mRNA Trafficking by Nuclear Pore Complexes. Genes *5*, 767–791.

Boutz, P.L., Bhutkar, A., and Sharp, P.A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. Genes Dev. *29*, 63–80.

Braun, J.E., Truffault, V., Boland, A., Huntzinger, E., Chang, C.-T., Haas, G., Weichenrieder, O., Coles, M., and Izaurralde, E. (2012). A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. Nat. Struct. Mol. Biol. *19*, 1324–1331.

Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res. *24*, 1774–1786.

Bresson, S.M., and Conrad, N.K. (2013). The Human Nuclear Poly(A)-Binding Protein Promotes RNA Hyperadenylation and Decay. PLoS Genet. *9*, e1003893.

Bresson, S.M., Hunter, O.V., Hunter, A.C., and Conrad, N.K. (2015). Canonical Poly(A) Polymerase Activity Promotes the Decay of a Wide Variety of Mammalian Nuclear RNAs. PLOS Genet. *11*, e1005610.

Brockmann, C., Soucek, S., Kuhlmann, S.I., Mills-Lujan, K., Kelly, S.M., Yang, J.-C., Iglesias, N., Stutz, F., Corbett, A.H., Neuhaus, D., et al. (2012). Structural basis for polyadenosine-RNA binding by Nab2 Zn fingers and its function in mRNA nuclear export. Struct. Lond. Engl. 1993 *20*, 1007–1018.

Brooks, S.A., and Blackshear, P.J. (2013). Tristetraprolin (TTP): Interactions with mRNA and proteins, and current thoughts on mechanisms of action. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1829*, 666–679.

Brown, R.S. (2005). Zinc finger proteins: getting a grip on RNA. Curr. Opin. Struct. Biol. *15*, 94–98.

Brown, R.S., Sander, C., and Argos, P. (1985). The primary structure of transcription factor TFIIIA has 12 consecutive repeats. FEBS Lett. *186*, 271–274.

Burgess, H.M., and Gray, N.K. (2010). mRNA-specific regulation of translation by poly(A)-binding proteins. Biochem. Soc. Trans. *38*, 1517–1522.

Burns, D.M., D'Ambrogio, A., Nottrott, S., and Richter, J.D. (2011). CPEB and two poly(A) polymerases control miR-122 stability and p53 mRNA translation. Nature *473*, 105–108.

Carmody, S.R., and Wente, S.R. (2009). mRNA nuclear export at a glance. J. Cell Sci. *122*, 1933–1937.

Castro-Diaz, N., Friedli, M., and Trono, D. (2015). Drawing a fine line on endogenous retroelement activity. Mob. Genet. Elem. *5*, 1–6.

Cevher, M.A., Zhang, X., Fernandez, S., Kim, S., Baquero, J., Nilsson, P., Lee, S., Virtanen, A., and Kleiman, F.E. (2010). Nuclear deadenylation/polyadenylation factors regulate 3′ processing in response to DNA damage. EMBO J. *29*, 1674–1687.

Chang, H., Lim, J., Ha, M., and Kim, V.N. (2014). TAIL-seq: Genome-wide Determination of Poly(A) Tail Length and 3′ End Modifications. Mol. Cell *53*, 1044–1052.

Charlesworth, A., Meijer, H.A., and de Moor, C.H. (2013). Specificity factors in cytoplasmic polyadenylation: Factors in cytoplasmic polyadenylation. Wiley Interdiscip. Rev. RNA *4*, 437–461.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat. Struct. Mol. Biol. *18*, 1218–1226.

Chen, C.-Y.A., and Shyu, A.-B. (2011). Mechanisms of deadenylation-dependent decay. Wiley Interdiscip. Rev. RNA *2*, 167–183.

Chen, F., and Wilusz, J. (1998). Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. Nucleic Acids Res. *26*, 2891–2898.

Chen, L.-L., and Carmichael, G.G. (2008). Gene regulation by SINES and inosines: biological consequences of A-to-I editing of Alu element inverted repeats. Cell Cycle *7*, 3294–3301.

Chen, L.-L., and Carmichael, G.G. (2009). Altered Nuclear Retention of mRNAs Containing Inverted Repeats in Human Embryonic Stem Cells: Functional Role of a Nuclear Noncoding RNA. Mol. Cell *35*, 467–478.

Chen, L.-L., DeCerbo, J.N., and Carmichael, G.G. (2008). Alu element-mediated gene silencing. EMBO J. *27*, 1694–1705.

Cheng, H., Dufu, K., Lee, C.-S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. Cell *127*, 1389–1400.

Cheng, Y., Liang, P., Geng, H., Wang, Z., Li, L., Cheng, S.H., Ying, J., Su, X., Ng, K.M., Ng, M.H.L., et al. (2012). A Novel 19q13 Nucleolar Zinc Finger Protein Suppresses Tumor Cell Growth through Inhibiting Ribosome Biogenesis and Inducing Apoptosis but Is Frequently Silenced in Multiple Carcinomas. Mol. Cancer Res. *10*, 925–936.

Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: the exosome. Biochim. Biophys. Acta *1829*, 552–560.

Choi, Y.H., and Hagedorn, C.H. (2003). Purifying mRNAs with a high-affinity eIF4E mutant identifies the short 3' poly(A) end phenotype. Proc. Natl. Acad. Sci. *100*, 7033–7038.

Ciafrè, S.A., and Galardi, S. (2013). microRNAs and RNA-binding proteins: A complex network of interactions and reciprocal regulations in cancer. RNA Biol. *10*, 934–942.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. Mol. Cell *33*, 717–726.

Colgan, D.F., Murthy, K.G.K., Prives, C., and Manley, J.L. (1996). Cell-cycle related regulation of poly(A) polymerase by phosphorylation. Nature *384*, 282–285.

Conrad, N.K. (2014). The emerging role of triple helices in RNA biology: The emerging role of RNA triple helices. Wiley Interdiscip. Rev. RNA *5*, 15–29.

Coyle, J.H., Bor, Y.-C., Rekosh, D., and Hammarskjold, M.-L. (2011). The Tpr protein regulates export of mRNAs with retained introns that traffic through the Nxf1 pathway. RNA N. Y. N *17*, 1344–1356.

Crockett, D.K., Lin, Z., Elenitoba-Johnson, K.S., and Lim, M.S. (2004). Identification of NPM-ALK interacting proteins by tandem mass spectrometry. Oncogene *23*, 2617–2629.

Culjkovic, B., Topisirovic, I., Skrabanek, L., Ruiz-Gutierrez, M., and Borden, K.L.B. (2006). eIF4E is a central node of an RNA regulon that governs cellular proliferation. J. Cell Biol. *175*, 415–426.

Culjkovic-Kraljacic, B., Baguet, A., Volpon, L., Amri, A., and Borden, K.L.B. (2012). The Oncogene eIF4E Reprograms the Nuclear Pore Complex to Promote mRNA Export and Oncogenic Transformation. Cell Rep. *2*, 207–215.

Dan, C., Jinjun, B., Zi-Chun, H., Lin, M., Wei, C., Xu, Z., Ri, Z., Shun, C., Wen-Zhu, S., Qing-Cai, J., et al. (2015). Modulation of TNF- mRNA stability by human antigen R and miR181s in sepsis-induced immunoparalysis. EMBO Mol. Med. *7*, 140–157.

Darnell, J.C., and Richter, J.D. (2012). Cytoplasmic RNA-binding proteins and the control of complex brain function. Cold Spring Harb. Perspect. Biol. *4*, a012344.

Das Gupta, J., Gu, H., Chernokalskaya, E., Gao, X., and Schoenberg, D.R. (1998). Identification of two cis-acting elements that independently regulate the length of poly(A) on Xenopus albumin pre-mRNA. RNA N. Y. N *4*, 766–776.

DeCerbo, J., and Carmichael, G.G. (2005). Retention and repression: fates of hyperedited RNAs in the nucleus. Curr. Opin. Cell Biol. *17*, 302–308.

Delaleau, M., and Borden, K. (2015). Multiple Export Mechanisms for mRNAs. Cells *4*, 452–473.

Di Giammartino, D.C., and Manley, J.L. (2014). New Links between mRNA Polyadenylation and Diverse Nuclear Pathways. Mol. Cells *37*, 644–649.

Di Giammartino, D.C., Shi, Y., and Manley, J.L. (2013). PARP1 represses PAP and inhibits polyadenylation during heat shock. Mol. Cell *49*, 7–17.

Di Giammartino, D.C., Nishida, K., and Manley, J.L. (2011). Mechanisms and Consequences of Alternative Polyadenylation. Mol. Cell *43*, 853–866.

Ding, G., Lorenz, P., Kreutzer, M., Li, Y., and Thiesen, H.-J. (2009). SysZNF: the C2H2 zinc finger gene database. Nucleic Acids Res. *37*, D267–D273.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Doidge, R., Mittal, S., Aslam, A., and Winkler, G.S. (2012). Deadenylation of cytoplasmic mRNA by the mammalian Ccr4–Not complex. Biochem. Soc. Trans. *40*, 896–901.

Dolken, L., Ruzsics, Z., Radle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., et al. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959–1972.

Dovat, S. (2002). A common mechanism for mitotic inactivation of C2H2 zinc finger DNA-binding domains. Genes Dev. *16*, 2985–2990.

Dreyfus, M., and Régnier, P. (2002). The Poly(A) Tail of mRNAs. Cell *111*, 611–613.

Dvinge, H., and Bradley, R.K. (2015). Widespread intron retention diversifies most cancer transcriptomes. Genome Med. *7*, 45.

Eberle, A.B., and Visa, N. (2014). Quality control of mRNP biogenesis: Networking at the transcription site. Semin. Cell Dev. Biol. *32*, 37–46.

Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H.M., Turelli, P., and Trono, D. (2016). Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. Dev. Cell *36*, 611–623.

Eckmann, C.R., Rammelt, C., and Wahle, E. (2011). Control of poly(A) tail length. Wiley Interdiscip. Rev. RNA *2*, 348–361.

Eichler, E.E., Hoffman, S.M., Adamson, A.A., Gordon, L.A., McCready, P., Lamerdin, J.E., and Mohrenweiser, H.W. (1998). Complex beta-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. Genome Res. *8*, 791–808.

Emerson, R.O., and Thomas, J.H. (2009). Adaptive Evolution in Zinc Finger Transcription Factors. PLoS Genet. *5*, e1000325.

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2008). Deadenylation is a widespread effect of miRNA regulation. RNA *15*, 21–32.

Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F., et al. (2011). miRNA-mediated deadenylation is

orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. Nat. Struct. Mol. Biol. *18*, 1211–1217.

Fabian, M.R., Frank, F., Rouya, C., Siddiqui, N., Lai, W.S., Karetnikov, A., Blackshear, P.J., Nagar, B., and Sonenberg, N. (2013). Structural basis for the recruitment of the human CCR4–NOT deadenylase complex by tristetraprolin. Nat. Struct. Mol. Biol. *20*, 735–739.

Fasken, M.B., and Corbett, A.H. (2016). Links between mRNA splicing, mRNA quality control, and intellectual disability. RNA Dis. Houst. Tex *3*, e1448.

Folkmann, A.W., Noble, K.N., Cole, C.N., and Wente, S.R. (2011). Dbp5, Gle1-IP6 and Nup159: a working model for mRNP export. Nucl. Austin Tex *2*, 540–548.

Fox, A.H., and Lamond, A.I. (2010). Paraspeckles. Cold Spring Harb. Perspect. Biol. *2*, a000687.

Fox, C.A., Sheets, M.D., and Wickens, M.P. (1989). Poly(A) addition during maturation of frog oocytes: distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUAU. Genes Dev. *3*, 2151–2162.

Fuke, H., and Ohno, M. (2008). Role of poly (A) tail as an identity element for mRNA nuclear export. Nucleic Acids Res. *36*, 1037–1049.

Gamsjaeger, R., Liew, C.K., Loughlin, F.E., Crossley, M., and Mackay, J.P. (2007). Sticky fingers: zinc-fingers as protein-recognition motifs. Trends Biochem. Sci. *32*, 63–70.

Gao, S., Hsieh, C.-L., Zhou, J., and Shemshedini, L. (2013). Zinc Finger 280B Regulates sGCα1 and p53 in Prostate Cancer Cells. PLoS ONE *8*, e78766.

Gardiner, A., Twiss, J., and Perrone-Bizzozero, N. (2015). Competing Interactions of RNA-Binding Proteins, MicroRNAs, and Their Targets Control Neuronal Development and Function. Biomolecules *5*, 2903–2918.

Garneau, N.L., Sokoloski, K.J., Opyrchal, M., Neff, C.P., Wilusz, C.J., and Wilusz, J. (2008). The 3' untranslated region of sindbis virus represses deadenylation of viral transcripts in mosquito and Mammalian cells. J. Virol. *82*, 880–892.

Gherzi, R., Lee, K.-Y., Briata, P., Wegmüller, D., Moroni, C., Karin, M., and Chen, C.-Y. (2004). A KH Domain RNA Binding Protein, KSRP, Promotes ARE-Directed mRNA Turnover by Recruiting the Degradation Machinery. Mol. Cell *14*, 571–583.

Godwin, A.R., Kojima, S., Green, C.B., and Wilusz, J. (2013). Kiss your tail goodbye: the role of PARN, Nocturnin, and Angel deadenylases in mRNA biology. Biochim. Biophys. Acta *1829*, 571–579.

Gondran, P., Amiot, F., Weil, D., and Dautry, F. (1999). Accumulation of mature mRNA in the nuclear fraction of mammalian cells. FEBS Lett. *458*, 324–328.

Gorgoni, B., and Gray, N.K. (2004). The roles of cytoplasmic poly(A)-binding proteins in regulating gene expression: a developmental perspective. Brief. Funct. Genomic. Proteomic. *3*, 125–141.

Goss, D.J., and Kleiman, F.E. (2013). Poly(A) binding proteins: are they all created equal? Wiley Interdiscip. Rev. RNA *4*, 167–179.

Gray, N.K., Hrabalkova, L., Scanlon, J.P., and Smith, R.W.P. (2015). Poly(A)-binding proteins and mRNA localization: who rules the roost? Biochem. Soc. Trans. *43*, 1277–1284.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P., and Trono, D. (2010). KRAB–Zinc Finger Proteins and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading. PLoS Genet. *6*, e1000869.

Gu, H. (2003). U2AF modulates poly(A) length control by the poly(A)-limiting element. Nucleic Acids Res. *31*, 6264–6271.

Gu, H., Das Gupta, J., and Schoenberg, D.R. (1999). The poly(A)-limiting element is a conserved cis-acting sequence that regulates poly(A) tail length on nuclear pre-mRNAs. Proc. Natl. Acad. Sci. U. S. A. *96*, 8943–8948.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., Pena, J.T.G., Nusbaum, J.D., Morozov, P., Ludwig, J., et al. (2011). RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. RNA *17*, 1697–1712.

Hall, T.M.T. (2005). Multiple modes of RNA recognition by zinc finger proteins. Curr. Opin. Struct. Biol. *15*, 367–373.

Harper, J., Yan, L., Loureiro, R.M., Wu, I., Fang, J., D'Amore, P.A., and Moses, M.A. (2007). Repression of vascular endothelial growth factor expression by the zinc finger transcription factor ZNF24. Cancer Res. *67*, 8736–8741.

Heath, C.G., Viphakone, N., and Wilson, S.A. (2016). The role of TREX in gene expression and disease. Biochem. J. *473*, 2911–2935.

Hendrickson, D.G., Kelley, D.R., Tenen, D., Bernstein, B., and Rinn, J.L. (2016). Widespread RNA binding by chromatin-associated proteins. Genome Biol. *17*, 28.

Heyn, P., Kalinka, A.T., Tomancak, P., and Neugebauer, K.M. (2015). Introns and gene expression: Cellular constraints, transcriptional regulation, and evolutionary consequences: Prospects & Overviews. BioEssays *37*, 148–154.

Hinnebusch, A.G., and Lorsch, J.R. (2012). The Mechanism of Eukaryotic Translation Initiation: New Insights and Challenges. Cold Spring Harb. Perspect. Biol. *4*, a011544–a011544.

Hu, R., Peng, G., Dai, H., Breuer, E.-K., Stemke-Hale, K., Li, K., Gonzalez-Angulo, A.M., Mills, G.B., and Lin, S.-Y. (2011). ZNF668 Functions as a Tumor Suppressor by Regulating p53 Stability and Function in Breast Cancer. Cancer Res. *71*, 6524–6534.

Huang, S. (1994). In vivo analysis of the stability and transport of nuclear poly(A)+ RNA. J. Cell Biol. *126*, 877–899.

Huang, S., Wu, S., Ding, J., Lin, J., Wei, L., Gu, J., and He, X. (2010). MicroRNA-181a modulates gene expression of zinc finger family members by directly targeting their coding regions. Nucleic Acids Res. *38*, 7211–7218.

Huntley, S. (2006). A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. Genome Res. *16*, 669–677.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat. Rev. Genet. *12*, 99–110.

Hurt, J.A., Obar, R.A., Zhai, B., Farny, N.G., Gygi, S.P., and Silver, P.A. (2009). A conserved CCCH-type zinc finger protein regulates mRNA nuclear adenylation and export. J. Cell Biol. *185*, 265–277.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics *8*, 39.

Iwakawa, H., and Tomari, Y. (2015). The Functions of MicroRNAs: mRNA Decay and Translational Repression. Trends Cell Biol. *25*, 651–665.

Iyengar, S., and Farnham, P.J. (2011). KAP1 Protein: An Enigmatic Master Regulator of the Genome. J. Biol. Chem. *286*, 26267–26276.

Jacobs, F.M.J., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R., and Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature *516*, 242–245.

Jen, J., and Wang, Y.-C. (2016). Zinc finger proteins in cancer progression. J. Biomed. Sci. *23*, 53.

Jens, M., and Rajewsky, N. (2014). Competition between target sites of regulators shapes post-transcriptional gene regulation. Nat. Rev. Genet. *16*, 113–126.

Jia, D., Hasso, S.M., Chan, J., Filingeri, D., D'Amore, P.A., Rice, L., Pampo, C., Siemann, D.W., Zurakowski, D., Rodig, S.J., et al. (2013). Transcriptional repression of VEGF by ZNF24: mechanistic studies and vascular consequences in vivo. Blood *121*, 707–715.

Jiao, X., Chang, J.H., Kilic, T., Tong, L., and Kiledjian, M. (2013). A Mammalian Pre-mRNA 5′ End Capping Quality Control Mechanism and an Unexpected Link of Capping to Pre-mRNA Processing. Mol. Cell *50*, 104–115.

Kahvejian, A., Svitkin, Y.V., Sukarieh, R., M'Boutchou, M.-N., and Sonenberg, N. (2005). Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. Genes Dev. *19*, 104–113.

Kaida, D. (2016). The reciprocal regulation between splicing and 3′-end processing: The reciprocal regulation between splicing and 3′-end processing. Wiley Interdiscip. Rev. RNA *7*, 499–511.

Kameoka, S., Duque, P., and Konarska, M.M. (2004). p54nrb associates with the 5′ splice site within large transcription/splicing complexes. EMBO J. *23*, 1782–1791.

Kandala, D.T., Mohan, N., A, V., A P, S., G, R., and Laishram, R.S. (2016). CstF-64 and 3'-UTR cis-element determine Star-PAP specificity for target mRNA selection by excluding PAPα. Nucleic Acids Res. *44*, 811–823.

Kappel, C., Trost, G., Czesnick, H., Ramming, A., Kolbe, B., Vi, S.L., Bispo, C., Becker, J.D., de Moor, C., and Lenhard, M. (2015). Genome-Wide Analysis of PAPS1-Dependent Polyadenylation Identifies Novel Roles for Functionally Specialized Poly(A) Polymerases in Arabidopsis thaliana. PLOS Genet. *11*, e1005474.

Karapetyan, A., Buiting, C., Kuiper, R., and Coolen, M. (2013). Regulatory Roles for Long ncRNA and mRNA. Cancers *5*, 462–490.

Katahira, J. (2012). mRNA export and the TREX complex. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1819*, 507–513.

Katahira, J. (2015). Nuclear Export of Messenger RNA. Genes *6*, 163–184.

Katahira, J., Inoue, H., Hurt, E., and Yoneda, Y. (2009). Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. EMBO J. *28*, 556–567.

Kelly, S.M., Leung, S.W., Pak, C., Banerjee, A., Moberg, K.H., and Corbett, A.H. (2014). A conserved role for the zinc finger polyadenosine RNA binding protein, ZC3H14, in control of poly(A) tail length. RNA *20*, 681–688.

Kim, J.H., and Richter, J.D. (2006). Opposing Polymerase-Deadenylase Activities Regulate Cytoplasmic Polyadenylation. Mol. Cell *24*, 173–183.

Kligun, E., and Mandel-Gutfreund, Y. (2015). The role of RNA conformation in RNA-protein recognition. RNA Biol. *12*, 720–727.

Kloc, M., Foreman, V., and Reddy, S.A. (2011). Binary function of mRNA. Biochimie *93*, 1955–1961.

Kojima, S., Sher-Chen, E.L., and Green, C.B. (2012). Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. Genes Dev. *26*, 2724–2736.

Kuchta, K., Muszewska, A., Knizewski, L., Steczkiewicz, K., Wyrwicz, L.S., Pawlowski, K., Rychlewski, L., and Ginalski, K. (2016). FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. Nucleic Acids Res. *44*, 3534–3548.

Kuhn, U., Gundel, M., Knoth, A., Kerwitz, Y., Rudel, S., and Wahle, E. (2009). Poly(A) Tail Length Is Controlled by the Nuclear Poly(A)-binding Protein Regulating the Interaction between Poly(A) Polymerase and the Cleavage and Polyadenylation Specificity Factor. J. Biol. Chem. *284*, 22803–22814.

Kumar, G.R., and Glaunsinger, B.A. (2010). Nuclear import of cytoplasmic poly(A) binding protein restricts gene expression via hyperadenylation and nuclear retention of mRNA. Mol. Cell. Biol. *30*, 4996–5008.

Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct Interactions between Subunits of CPSF and the U2 snRNP Contribute to the Coupling of Pre-mRNA 3′ End Processing and Splicing. Mol. Cell *23*, 195–205.

Laishram, R.S. (2014). Poly(A) polymerase (PAP) diversity in gene expression - Star-PAP vs canonical PAP. FEBS Lett. *588*, 2185–2197.

Li, X., Kazan, H., Lipshitz, H.D., and Morris, Q.D. (2014). Finding the target sites of RNA-binding proteins: Finding target sites of RNA-binding proteins. Wiley Interdiscip. Rev. RNA *5*, 111–130.

Liang, H., Zhang, J., Zen, K., Zhang, C.-Y., and Chen, X. (2013). Nuclear microRNAs and their unconventional role in regulating non-coding RNAs. Protein Cell *4*, 325–330.

Liao, J.-Y., Ma, L.-M., Guo, Y.-H., Zhang, Y.-C., Zhou, H., Shao, P., Chen, Y.-Q., and Qu, L.-H. (2010). Deep Sequencing of Human Nuclear and Cytoplasmic Small RNAs Reveals an Unexpectedly Complex Subcellular Distribution of miRNAs and tRNA 3′ Trailers. PLoS ONE *5*, e10563.

Liu, H., Chang, L.-H., Sun, Y., Lu, X., and Stubbs, L. (2014a). Deep Vertebrate Roots for Mammalian Zinc Finger Transcription Factor Subfamilies. Genome Biol. Evol. *6*, 510–525.

Liu, H., Luo, M., and Wen, J. (2014b). mRNA stability in the nucleus. J. Zhejiang Univ. Sci. B *15*, 444–454.

Liu, Y., Beyer, A., and Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell *165*, 535–550.

Looman, C., Abrink, M., Mark, C., and Hellman, L. (2002). KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. Mol. Biol. Evol. *19*, 2118–2130.

Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., and Jensen, T.H. (2015). The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis. Cell Rep. *10*, 178–192.

Lukic, S., Nicolas, J.-C., and Levine, A.J. (2014). The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. Cell Death Differ. *21*, 381–387.

Luo, M.J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. Proc. Natl. Acad. Sci. U. S. A. *96*, 14937–14942.

Lutz, C.S., Murthy, K.G., Schek, N., O'Connor, J.P., Manley, J.L., and Alwine, J.C. (1996). Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. Genes Dev. *10*, 325–337.

Macdonald, P.M., Kanke, M., and Kenny, A. (2016). Community effects in regulation of translation. eLife *5*, e10965.

Machinaga, A., and Takase-Yoden, S. (2014). Polyadenylation of Friend murine leukemia virus *env* -mRNA is affected by its splicing: Splicing affects polyA tail of MLV mRNA. Microbiol. Immunol. *58*, 474–482.

Mancini, A., Niemann-Seyde, S.C., Pankow, R., El Bounkari, O., Klebba-Färber, S., Koch, A., Jaworska, E., Spooncer, E., Gruber, A.D., Whetton, A.D., et al. (2010). THOC5/FMIP, an mRNA export TREX complex protein, is essential for hematopoietic primitive cell survival in vivo. BMC Biol. *8*, 1.

Marzluff, W.F. (2012). Novel 3' ends that support translation. Genes Dev. *26*, 2457–2460.

Masuda, S. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes Dev. *19*, 1512–1517.

Matoulkova, E., Michalova, E., Vojtesek, B., and Hrstka, R. (2012). The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol. *9*, 563–576.

Mauger, O., Lemoine, F., and Scheiffele, P. (2016). Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity. Neuron *92*, 1266–1278.

Meijer, H.A., Bushell, M., Hill, K., Gant, T.W., Willis, A.E., Jones, P., and de Moor, C.H. (2007). A novel method for poly(A) fractionation reveals a large population of mRNAs with a short poly(A) tail in mammalian cells. Nucleic Acids Res. *35*, e132.

Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A., et al. (2016). Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. Mol. Cell *64*, 520–533.

Miller, M.A., and Olivas, W.M. (2011). Roles of Puf proteins in mRNA degradation and translation. Wiley Interdiscip. Rev. RNA *2*, 471–492.

Miller, J., McLachlan, A.D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. *4*, 1609–1614.

Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic Acids Res. *38*, 2757–2774.

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., and Vagner, S. (2006). An interaction between U2AF 65 and CF Im links the splicing and 3′ end processing machineries. EMBO J. *25*, 4854–4864.

Misra, A., and Green, M.R. (2016). From polyadenylation to splicing: Dual role for mRNA 3' end formation factors. RNA Biol. *13*, 259–264.

Mohan, N., Ap, S., Francis, N., Anderson, R., and Laishram, R.S. (2015). Phosphorylation regulates the Star-PAP-PIPKIα interaction and directs specificity toward mRNA targets. Nucleic Acids Res. *43*, 7005–7020.

Moraes, K.C.M., Wilusz, C.J., and Wilusz, J. (2006). CUG-BP binds to RNA substrates and recruits PARN deadenylase. RNA N. Y. N *12*, 1084–1091.

Morimoto, M., and Boerkoel, C. (2013). The Role of Nuclear Bodies in Gene Expression and Disease. Biology *2*, 976–1033.

Morris, K.V., and Mattick, J.S. (2014). The rise of regulatory RNA. Nat. Rev. Genet. *15*, 423–437.

Muniz, L., Davidson, L., and West, S. (2015). Poly(A) Polymerase and the Nuclear Poly(A) Binding Protein, PABPN1, Coordinate the Splicing and Degradation of a Subset of Human Pre-mRNAs. Mol. Cell. Biol. *35*, 2218–2230.

Nagarajan, V.K., Jones, C.I., Newbury, S.F., and Green, P.J. (2013). XRN 5'→3' exoribonucleases: structure, mechanisms and functions. Biochim. Biophys. Acta *1829*, 590–603.

Nam, J.-W., Choi, S.-W., and You, B.-H. (2016). Incredible RNA: Dual Functions of Coding and Noncoding. Mol. Cells *39*, 367–374.

Natalizio, B.J., and Wente, S.R. (2013). Postage for the messenger: designating routes for nuclear mRNA export. Trends Cell Biol. *23*, 365–373.

Neff, A.T., Lee, J.Y., Wilusz, J., Tian, B., and Wilusz, C.J. (2012). Global analysis reveals multiple pathways for unique regulation of mRNA decay in induced pluripotent stem cells. Genome Res. *22*, 1457–1467.

Nelson, P.N., Carnegie, P.R., Martin, J., Davari Ejtehadi, H., Hooley, P., Roden, D., Rowland-Jones, S., Warren, P., Astley, J., and Murray, P.G. (2003). Demystified. Human endogenous retroviruses. Mol. Pathol. MP *56*, 11–18.

Norbury, C.J. (2013). Cytoplasmic RNA: a case of the tail wagging the dog. Nat. Rev. Mol. Cell Biol. *14*, 643–653.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. *302*, 205–217.

Novoa, I., Gallego, J., Ferreira, P.G., and Mendez, R. (2010). Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. Nat. Cell Biol. *12*, 447–456.

Nunes, N.M., Li, W., Tian, B., and Furger, A. (2010). A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. EMBO J. *29*, 1523–1536.

Oeffinger, M., and Zenklusen, D. (2012). To the pore and through the pore: A story of mRNA export kinetics. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1819*, 494–506.

O'Geen, H., Squazzo, S.L., Iyengar, S., Blahnik, K., Rinn, J.L., Chang, H.Y., Green, R., and Farnham, P.J. (2007). Genome-Wide Analysis of KAP1 Binding Suggests Autoregulation of KRAB-ZNFs. PLoS Genet. *3*, e89.

Palaniswamy, V., Moraes, K.C.M., Wilusz, C.J., and Wilusz, J. (2006). Nucleophosmin is selectively deposited on mRNA during polyadenylation. Nat. Struct. Mol. Biol. *13*, 429–435.

Park, J.-E., Yi, H., Kim, Y., Chang, H., and Kim, V.N. (2016). Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. Mol. Cell *62*, 462–471.

Peng, J., and Schoenberg, D.R. (2005). mRNA with a <20-nt poly(A) tail imparted by the poly(A)-limiting element is translated as efficiently in vivo as long poly(A) mRNA. RNA N. Y. N *11*, 1131–1140.

Peng, J., Murray, E.L., and Schoenberg, D.R. (2005). The poly(A)-limiting element enhances mRNA accumulation by increasing the efficiency of pre-mRNA 3' processing. RNA N. Y. N *11*, 958–965.

Pimentel, H., Parra, M., Gee, S.L., Mohandas, N., Pachter, L., and Conboy, J.G. (2016). A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. Nucleic Acids Res. *44*, 838–851.

Piqué, M., López, J.M., Foissac, S., Guigó, R., and Méndez, R. (2008). A Combinatorial Code for CPE-Mediated Translational Control. Cell *132*, 434–448.

Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating Gene Expression through RNA Nuclear Retention. Cell *123*, 249–263.

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev. *25*, 1770–1782.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA Processing with Transcription. Cell *108*, 501–512.

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat. Biotechnol. *29*, 436–442.

Rajanala, K., and Nandicoori, V.K. (2012). Localization of nucleoporin Tpr to the nuclear pore complex is essential for Tpr mediated regulation of the export of unspliced RNA. PloS One *7*, e29921.

Ramaswami, G., Zhang, R., Piskol, R., Keegan, L.P., Deng, P., O'Connell, M.A., and Li, J.B. (2013). Identifying RNA editing sites using RNA sequencing data alone. Nat. Methods *10*, 128–132.

Rao, M. (1996). Regulated nuclear polyadenylation of Xenopus albumin pre-mRNA. Nucleic Acids Res. *24*, 4078–4083.

Rattenbacher, B., Beisang, D., Wiesner, D.L., Jeschke, J.C., von Hohenberg, M., St. Louis-Vlasova, I.A., and Bohjanen, P.R. (2010). Analysis of CUGBP1 Targets Identifies GU-Repeat Sequences That Mediate Rapid mRNA Decay. Mol. Cell. Biol. *30*, 3970–3980.

Razin, S.V., Borunova, V.V., Maksimenko, O.G., and Kantidze, O.L. (2012). Cys2His2 zinc finger protein family: classification, functions, and major members. Biochem. Biokhimii͡a *77*, 217–226.

Reveal, B., Yan, N., Snee, M.J., Pai, C.-I., Gim, Y., and Macdonald, P.M. (2010). BREs mediate both repression and activation of oskar mRNA translation and act in trans. Dev. Cell *18*, 496–502.

Reyes, J.M., and Ross, P.J. (2016). Cytoplasmic polyadenylation in mammalian oocyte maturation. Wiley Interdiscip. Rev. RNA *7*, 71–89.

Rigo, F., and Martinson, H.G. (2009). Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. RNA *15*, 823–836.

Rizkallah, R., Alexander, K.E., and Hurt, M.M. (2011). Global mitotic phosphorylation of C2H2 zinc finger protein linker peptides. Cell Cycle Georget. Tex *10*, 3327–3336.

Roberts, T.C. (2014). The MicroRNA Biology of the Mammalian Nucleus. Mol. Ther. Acids *3*, e188.

Rosonina, E., Ip, J.Y.Y., Calarco, J.A., Bakowski, M.A., Emili, A., McCracken, S., Tucker, P., Ingles, C.J., and Blencowe, B.J. (2005). Role for PSF in Mediating Transcriptional Activator-Dependent Stimulation of Pre-mRNA Processing In Vivo. Mol. Cell. Biol. *25*, 6734–6746.

Rowe, H.M., and Trono, D. (2011). Dynamic control of endogenous retroviruses during development. Virology *411*, 273–287.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237–240.

Roy, B., and Jacobson, A. (2013). The intimate relationships of mRNA decay and translation. Trends Genet. *29*, 691–699.

Sagawa, F., Ibrahim, H., Morrison, A.L., Wilusz, C.J., and Wilusz, J. (2011). Nucleophosmin deposition during mRNA 3′ end processing influences poly(A) tail length: NPM1 affects mRNA poly(A) tail length. EMBO J. *30*, 3994–4005.

Saitoh, N. (2004). Proteomic Analysis of Interchromatin Granule Clusters. Mol. Biol. Cell *15*, 3876–3890.

Sampath, K., and Ephrussi, A. (2016). CncRNAs: RNAs with both coding and non-coding roles in development. Development *143*, 1234–1241.

Sasaki, Y.T.F., Ideue, T., Sano, M., Mituyama, T., and Hirose, T. (2009). MEN / noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc. Natl. Acad. Sci. *106*, 2525–2530.

Schmid, M., and Jensen, T.H. (2008). The exosome: a multipurpose RNA-decay machine. Trends Biochem. Sci. *33*, 501–510.

Schmidt, K., and Butler, J.S. (2013). Nuclear RNA surveillance: role of TRAMP in controlling exosome specificity. Wiley Interdiscip. Rev. RNA *4*, 217–231.

Schnall-Levin, M., Rissland, O.S., Johnston, W.K., Perrimon, N., Bartel, D.P., and Berger, B. (2011). Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs. Genome Res. *21*, 1395–1403.

Schoenberg, D.R., and Maquat, L.E. (2012). Regulation of cytoplasmic mRNA decay. Nat. Rev. Genet. *13*, 246–259.

Schoenberg, D.R., Moskaitis, J.E., Smith, L.H., Jr, and Pastori, R.L. (1989). Extranuclear estrogen-regulated destabilization of Xenopus laevis serum albumin mRNA. Mol. Endocrinol. Baltim. Md *3*, 805–814.

Shen, H.-H. (2009). UAP56- a key player with surprisingly diverse roles in pre-mRNA splicing and nuclear export. BMB Rep. *42*, 185–188.

Shen, S., Lin, L., Cai, J.J., Jiang, P., Kenkel, E.J., Stroik, M.R., Sato, S., Davidson, B.L., and Xing, Y. (2011). Widespread establishment and regulatory impact of Alu exons in human genes. Proc. Natl. Acad. Sci. *108*, 2837–2842.

Shi, Y., and Manley, J.L. (2015). The end of the message: multiple protein–RNA interactions define the mRNA polyadenylation site. Genes Dev. *29*, 889–897.

Simon, R., and Richter, J.D. (1994). Further analysis of cytoplasmic polyadenylation in Xenopus embryos and identification of embryonic cytoplasmic polyadenylation element-binding proteins. Mol. Cell. Biol. *14*, 7867–7875.

Sleeman, J.E., and Trinkle-Mulcahy, L. (2014). Nuclear bodies: new insights into assembly/dynamics and disease relevance. Curr. Opin. Cell Biol. *28*, 76–83.

Sloan, K.E., Gleizes, P.-E., and Bohnsack, M.T. (2016). Nucleocytoplasmic Transport of RNAs and RNA–Protein Complexes. J. Mol. Biol. *428*, 2040–2059.

Smith, R.W.P., Blee, T.K.P., and Gray, N.K. (2014). Poly(A)-binding proteins are required for diverse biological processes in metazoans. Biochem. Soc. Trans. *42*, 1229–1237.

Solnestam, B., Stranneheim, H., Hällman, J., Käller, M., Lundberg, E., Lundeberg, J., and Akan, P. (2012). Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs. BMC Genomics *13*, 574.

Soucek, S., Corbett, A.H., and Fasken, M.B. (2012). The long and the short of it: The role of the zinc finger polyadenosine RNA binding protein, Nab2, in control of poly(A) tail length. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1819*, 546–554.

Soucek, S., Zeng, Y., Bellur, D.L., Bergkessel, M., Morris, K.J., Deng, Q., Duong, D., Seyfried, N.T., Guthrie, C., Staley, J.P., et al. (2016). The Evolutionarily-conserved Polyadenosine RNA Binding Protein, Nab2, Cooperates with Splicing Machinery to Regulate the Fate of pre-mRNA. Mol. Cell. Biol. *36*, 2697–2714.

Spector, D.L., and Lamond, A.I. (2011). Nuclear Speckles. Cold Spring Harb. Perspect. Biol. *3*, a000646.

Stewart, M. (2007). Ratcheting mRNA out of the Nucleus. Mol. Cell *25*, 327–330.

Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S.Y., Hahn, W.C., Sharp, P.A., et al. (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA N. Y. N *9*, 493–501.

Stoye, J.P. (2012). Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nat. Rev. Microbiol. *10*, 395–406.

Stubbs, L., Sun, Y., and Caetano-Anolles, D. (2011). Function and Evolution of C2H2 Zinc Finger Arrays. Subcell. Biochem. *52*, 75–94.

Su, Y.-L., Wang, S.-C., Chiang, P.-Y., Lin, N.-Y., Shen, Y.-F., Chang, G.-D., and Chang, C.-J. (2012). Tristetraprolin Inhibits Poly(A)-Tail Synthesis in Nuclear mRNA that Contains AU-Rich Elements by Interacting with Poly(A)-Binding Protein Nuclear 1. PLoS ONE *7*, e41313.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66–71.

Swamynathan, S.K. (2010). Krüppel-like factors: three fingers in control. Hum. Genomics *4*, 263–270.

Tadepally, H.D., Burger, G., and Aubry, M. (2008). Evolution of C2H2-zinc finger genes and subfamilies in mammals: Species-specific duplication and loss of clusters, genes and effector domains. BMC Evol. Biol. *8*, 176.

Tay, Y., Rinn, J., and Pandolfi, P.P. (2014). The multilayered complexity of ceRNA crosstalk and competition. Nature *505*, 344–352.

Thomas, J.H., and Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. Genome Res. *21*, 1800–1812.

Thomson, D.W., and Dinger, M.E. (2016). Endogenous microRNA sponges: evidence and controversy. Nat. Rev. Genet. *17*, 272–283.

Tian, B., and Graber, J.H. (2012). Signals for pre-mRNA cleavage and polyadenylation. Wiley Interdiscip. Rev. RNA *3*, 385–396.

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigo, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome Res. *22*, 1616–1625.

Torres, M., Becquet, D., Blanchard, M.-P., Guillen, S., Boyer, B., Moreno, M., Franc, J.-L., and François-Bellan, A.-M. (2016). Circadian RNA expression elicited by 3'-UTR IRAlu-paraspeckle associated elements. eLife *5*, e14837.

Torres, M., Becquet, D., Blanchard, M.-P., Guillen, S., Boyer, B., Moreno, M., Franc, J.-L., and François-Bellan, A.-M. (2017). Paraspeckles as rhythmic nuclear mRNA anchorages responsible for circadian gene expression. Nucleus 1–6.

Tran, D.D.H., Koch, A., and Tamura, T. (2014). THOC5, a member of the mRNA export complex: a novel link between mRNA export machinery and signal transduction pathways in cell proliferation and differentiation. Cell Commun. Signal. *12*, 3.

Trask, H.W., Cowper-Sal-lari, R., Sartor, M.A., Gui, J., Heath, C.V., Renuka, J., Higgins, A.-J., Andrews, P., Korc, M., Moore, J.H., et al. (2009). Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs. RNA *15*, 1917–1928.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. Mol. Cell *39*, 925–938.

Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., and Trono, D. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res. *24*, 1260–1270.

Udagawa, T., Swanger, S.A., Takeuchi, K., Kim, J.H., Nalavadi, V., Shin, J., Lorenz, L.J., Zukin, R.S., Bassell, G.J., and Richter, J.D. (2012). Bidirectional control of mRNA translation and synaptic plasticity by the cytoplasmic polyadenylation complex. Mol. Cell *47*, 253–266.

Urrutia, R. (2003). KRAB-containing zinc-finger repressor proteins. Genome Biol. *4*, 231.

Wahle, E. (1991). Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. J. Biol. Chem. *266*, 3131–3139.

Wahle, E. (1995). Poly(A) tail length control is caused by termination of processive synthesis. J. Biol. Chem. *270*, 2800–2808.

Wang, I.X., So, E., Devlin, J.L., Zhao, Y., Wu, M., and Cheung, V.G. (2013a). ADAR Regulates RNA Editing, Transcript Stability, and Gene Expression. Cell Rep. *5*, 849–860.

Wang, L., Miao, Y.-L., Zheng, X., Lackford, B., Zhou, B., Han, L., Yao, C., Ward, J.M., Burkholder, A., Lipchina, I., et al. (2013b). The THO Complex Regulates Pluripotency Gene mRNA Export and Controls Embryonic Stem Cell Self-Renewal and Somatic Cell Reprogramming. Cell Stem Cell *13*, 676–690.

Weidmann, C.A., Raynard, N.A., Blewett, N.H., Van Etten, J., and Goldstrohm, A.C. (2014). The RNA binding domain of Pumilio antagonizes poly-adenosine binding protein and accelerates deadenylation. RNA *20*, 1298–1319.

Weil, D., Boutain, S., Audibert, A., and Dautry, F. (2000). Mature mRNAs accumulated in the nucleus are neither the molecules in transit to the cytoplasm nor constitute a stockpile for gene expression. RNA N. Y. N *6*, 962–975.

Weill, L., Belloc, E., Bava, F.-A., and Méndez, R. (2012). Translational control by changes in poly(A) tail length: recycling mRNAs. Nat. Struct. Mol. Biol. *19*, 577–585.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Mol. Cell *2*, 135–140.

Wickramasinghe, V.O., and Laskey, R.A. (2015). Control of mammalian gene expression by selective mRNA export. Nat. Rev. Mol. Cell Biol. *16*, 431–442.

Wickramasinghe, V.O., Andrews, R., Ellis, P., Langford, C., Gurdon, J.B., Stewart, M., Venkitaraman, A.R., and Laskey, R.A. (2014). Selective nuclear export of specific classes of mRNA from mammalian nuclei is promoted by GANP. Nucleic Acids Res. *42*, 5059–5071.

Wilusz, C.J., and Wilusz, J. (2010). trans Regulation: Do mRNAs Have a Herd Mentality? Dev. Cell *18*, 333–334.

Windhager, L., Bonfert, T., Burger, K., Ruzsics, Z., Krebs, S., Kaufmann, S., Malterer, G., L'Hernault, A., Schilhabel, M., Schreiber, S., et al. (2012). Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. Genome Res. *22*, 2031–2042.

Wolf, D., and Goff, S.P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature *458*, 1201–1204.

Wolfe, S.A., Nekludova, L., and Pabo, C.O. (2000). DNA recognition by Cys2His2 zinc finger proteins. Annu. Rev. Biophys. Biomol. Struct. *29*, 183–212.

Wong, J.J.-L., Au, A.Y.M., Ritchie, W., and Rasko, J.E.J. (2016). Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. BioEssays *38*, 41–49.

Woodward, L.A., Mabin, J.W., Gangras, P., and Singh, G. (2016). The exon junction complex: a lifelong guardian of mRNA fate. Wiley Interdiscip. Rev. RNA *8*, e1411.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., et al. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Mol. Cell *61*, 507–519.

Yamagishi, R., Tsusaka, T., Mitsunaga, H., Maehata, T., and Hoshino, S. (2016). The STAR protein QKI-7 recruits PAPD4 to regulate post-transcriptional polyadenylation of target mRNAs. Nucleic Acids Res. *44*, 2475–2490.

Yamashita, A., Chang, T.-C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.-Y.A., and Shyu, A.-B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat. Struct. Mol. Biol. *12*, 1054–1063.

Yan, Y.-B. (2014). Deadenylation: enzymes, regulation, and functional implications: Regulation of deadenylation. Wiley Interdiscip. Rev. RNA *5*, 421–443.

Yang, L., Hamilton, S.R., Sood, A., Kuwai, T., Ellis, L., Sanguino, A., Lopez-Berestein, G., and Boyd, D.D. (2008a). The previously undescribed ZKSCAN3 (ZNF306) is a novel "driver" of colorectal cancer progression. Cancer Res. *68*, 4321–4330.

Yang, L., Zhang, L., Wu, Q., and Boyd, D.D. (2008b). Unbiased Screening for Transcriptional Targets of ZKSCAN3 Identifies Integrin 4 and Vascular Endothelial Growth Factor as Downstream Targets. J. Biol. Chem. *283*, 35295–35304.

Yang, L., Wang, H., Kornblau, S.M., Graber, D.A., Zhang, N., Matthews, J.A., Wang, M., Weber, D.M., Thomas, S.K., Shah, J.J., et al. (2011a). Evidence of a role for the novel zinc-finger transcription factor ZKSCAN3 in modulating Cyclin D2 expression in multiple myeloma. Oncogene *30*, 1329–1340.

Yang, L., Duff, M.O., Graveley, B.R., Carmichael, G.G., and Chen, L.-L. (2011b). Genomewide characterization of non-polyadenylated RNAs. Genome Biol. *12*, R16.

Yap, K., Lim, Z.Q., Khandelia, P., Friedman, B., and Makeyev, E.V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. Genes Dev. *26*, 1209–1223.

Yeo, S.-Y., Ha, S.Y., Yu, E.J., Lee, K.-W., Kim, J.H., and Kim, S.-H. (2014). ZNF282 (Zinc finger protein 282), a novel E2F1 co-activator, promotes esophageal squamous cell carcinoma. Oncotarget *5*, 12260–12272.

Zhang, X., Virtanen, A., and Kleiman, F.E. (2010). To polyadenylate or to deadenylate: that is the question. Cell Cycle Georget. Tex *9*, 4437–4449.

Zhang, X., Jing, Y., Qin, Y., Hunsucker, S., Meng, H., Sui, J., Jiang, Y., Gao, L., An, G., Yang, N., et al. (2012). The zinc finger transcription factor ZKSCAN3 promotes prostate cancer cell migration. Int. J. Biochem. Cell Biol. *44*, 1166–1173.

Zhang, X., Devany, E., Murphy, M.R., Glazman, G., Persaud, M., and Kleiman, F.E. (2015). PARN deadenylase is involved in miRNA-dependent degradation of TP53 mRNA in mammalian cells. Nucleic Acids Res. *43*, 10925–10938.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. Mol. Cell *40*, 939–953.

Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.-M., Li, C.J., Vågbø, C.B., Shi, Y., Wang, W.-L., Song, S.-H., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol. Cell *49*, 18–29.

Zhong, X., Li, N., Liang, S., Huang, Q., Coukos, G., and Zhang, L. (2010). Identification of MicroRNAs Regulating Reprogramming Factor LIN28 in Embryonic Stem Cells and Cancer Cells. J. Biol. Chem. *285*, 41961–41971.

Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y., and Robb, G.B. (2012). Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. Nucleic Acids Res. *40*, e54–e54.

APPENDICES

Appendix A: Primer sequences

Table 1: Primers used to generate plasmid templates for *in vitro* transcribed polyadenylated control RNAs

Plasmid	*S. cerevisiae* Gene name	Gene ID	Forward Primer	Reverse Primer
A15	ACT1	850504	CGCGGATCCCCAAGAGAGGTATCTTGACTTTAC	CACGTCGACAGAAGGTATGATGCCAGATCTTTT
A35	GAL1	852308	CGCGGATCCTGCGTATTACGGTCGTTGCA	CACGTCGACAACAGAGGCAGCCTGATCCA
A45	MET3	853466	CGCGGATCCACGTTGCCGGTGATTATTACGTCG	CACGTCGACAGCTTCTCTGGCGGCTCTCACA
A149	MET25	851010	CGCGGATCCGCTCACAAACACGGTATTCCAGTTGTC	CACGTCGACGGCAGGTTGAGAGAATTGAGGGAACT

Table 2: Primers used to generate PPIA and ZNF12 reporter plasmids

Reporter plasmid	Insert name	Forward Primer	Reverse Primer
All plasmids	Chimeric intron	GACGGGCCCAGGTAGCCTTGCAGAAGTTG	GACGGGCCCCTCTAGCCTTAAGAGCTGTAAT
All plasmids	Renilla luciferase	AGAGGGGCCCAGATCTATGGCTTCCAAGGTGTACGA	ATATCGCCGGCGGCCGCGACGTTGATCCTGGCGCT
PPIA	PPIA ORF + 3' UTR	CAGGATCAACGTCGCGATGGTCAACCCCACCGTG	GGAGATATCGCCGGCGGCCGCTAGGGATAATGAAAAGATGGTTGATGTTTATTTC
ZNF12	Z12 ORF	ATGAATAAATCCCTGGGGCCAG	TGAGGTCTGACTTCAGAGAAGCCTTC
ZNF12	Z12 3' UTR	GAAGGCTTCTCTGAAGTCAGACCTCA	AAGCCATTTACAGTTTTCAAATATTTTACTGAAAATG
ZNF12	Z12 DSE	GAAAACTGTAAATGGCTTGGCATTTTCTTCTTATC	CGTTATCCCCTGATTCTGTTGTAACAAGCCAGCAGCCCA
ORF	Z12 ORF only	ATGAATAAATCCCTGGGGCCAG	ATATCGCCGGCGGCCGCTCAGAGAAGCCTTCCCAC
3' UTR	Z12 3' UTR + DSE only	ATCAACGTCGCGGCCGCCTGAAGTCAGACCTCA	CGTTATCCCCTGATTCTGTGTAACAAGCCAGCAGCCCA
KRAB	Z12 KRAB	CAGGATCAACGTCGCGGTGTCATTCAAGGACGTG	GGAGATATCGCCGGCGGCCGCTCATATCCATGGCTCTTCTCC
3ZNF	3ZNF	AGCGCCAGGATCAACGTCGCGGTAGAAGGAGAATTCCTACTTC	GGAGATATCGCCGGCGGCCGCTCATGACTTTTTACAGAAGGATTTC
14ZNF	14ZNF	GCCAGGATCAACGTCGCGAAATTTATTATACATCAGAGGACTCAC	GGAGATATCGCCGGCGGCCGCTCAGAGAAGCCTTCCCAC
3' UTRΔ	Z12 3' UTRΔ	ATCAACGTCGCGGCCGCCTGAAGTCAGACCTCA	CCTCTGGAGATATCGCCGGCAGTATTAAACAGTTTCCAAAGGATTATG

Table 3: Primers used for qPCR analysis

Target	Gene ID	Forward Primer	Reverse Primer	Efficiency
A0 (pGEM4)	N/A	GAATACACGGAATTCGAGCTC	TCATTAATGCAGCTGGCTTAT	98.0
A15 (ACT1)	850504	CCAAGAGAGGTATCTTGACTTTAC	AGAAGGTATGATGCCAGATCTTTT	93.3
A35 (GAL1)	852308	TGCGTATTACGGTCGTTGCA	AACAGAGGCAGCCTGATCCA	101.8
A45 (MET3)	853466	ACGTTGCCGGTGATTATTACGTCG	AGCTTCTCTGGCGGCTCTCACA	102.2
A149 (MET25)	851010	GCTCACAAACACGGTATTCCAGTTGTC	GGCAGGTTGAGAGAATTGAGGGAACT	90.6
DICER	23405	TTGGCTTCCTCCTGGTTA	CACATCAGGCTCTCCTCCTC	98.8
GAPDH	2597	AAGGTGAAGGTCGGAGTCAA	AATGAAGGGGTCATTGATGG	93.0
PPIA	5478	GTGGTATAAAAGGGGCGGGAG	GTCTGCAAACAGCTCAAAGGA	107.3
LIN28A	79727	GAAGCGCAGATCAAAAGGAG	GCTGATGCTCTGGCAGAAGT	108.5
45S rRNA	100861532	GAACGGTGGTGTGTCGTT	GCGTCTCGTCTCGTCTCACT	91.0
tRNATyr	100009601	CCTTCGATAGCTCAGCTGGTAGAGCGGAGG	CGGAATCGGAACCAGCGACCTAAGGATGTCC	102.7
ZNF2	7549	GTTTCCCGACCTGAAGAGG	TCCTGTGTGCTCTCCTTGTG	103.7
ZNF12	7559	ATCAGCAAGTTGGAGCAAGG	TTAGGTCATCAGTTTGCCAGA	91.3
ZNF43	7594	TGATCACCTGTCTGGAGCAA	TGGCCAAAAGTCTTGGGTAA	114.2
ZNF134	7693	GAGTCATGCCCTCTCAGGC	CTACGAACTGTGATGGCGG	108.2
ZNF627	199692	TTGAGGATGTGGCTGTGAAC	CAGAAGCCAGGTTCCTGAAG	97.8
ZNF280B	140883	GAGCACAAGACCCAGTGTCA	TTAGACCTGGGGAGTGATGG	103.5
ZNF282	8427	GAATCCATTACCGACTCCCCA	TGTCTGCCAAATCCTGCTGA	104.2
ZsGreen	N/A	GTCAGCTTGTGCTGGATGAA	CCCCGTGATGAAGAAGATGA	99.0
Renilla luciferase	N/A	CTATCGTGCATGCTGAGAGC	CAGGTAAGCAGCGAACTCCT	99.2

Table 4: LLM-PAT primers

Gene name	Gene ID	Primer
GAPDH	2597	AGGAGTAAGACCCCTGGACCACC
PPIA	5478	CACCACCACTTGTGCATCTC
ZNF12	7559	GAGTATGGTACGTAACAGGAATTACATCTG
ZNF627	199692	CACTGTGAAGTCAGCGTTAACC
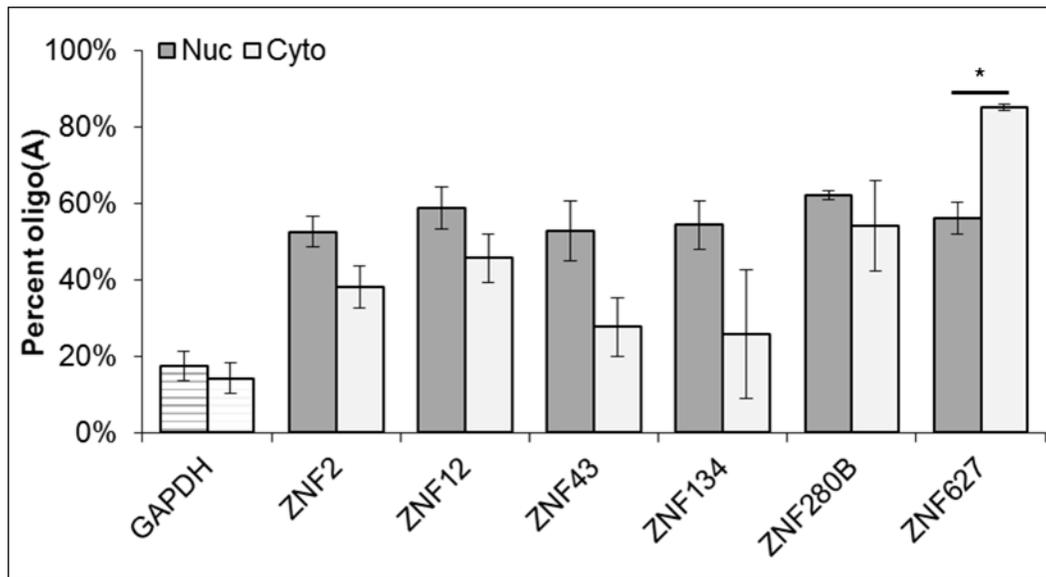
Figure 28: C2H2-ZNF mRNAs have short poly(A) tails in the nucleus and the cytoplasm. Detergent-based subcellular fractionation was used to separate nuclear and cytoplasm cellular fractions prior to RNA isolation using TRIzol. Oligo (dT) fractionation was used to separate poly(A) and oligo(A) nuclear and cytoplasmic RNAs. qRT-PCR was used to measure relative abundance of the mRNA in each fraction. The percent of each mRNA in the oligo(A) fraction of total was then determined. Error bars represent standard deviation of 3 biological replicates. * $p < 0.05$ Nuc compared to Cyto (paired, two-tailed Student's T-test).

LIST OF ABBREVIATIONS

ARE   AU-rich element

C2H2-ZNF   Cysteine2 Histidine2 zinc finger

ceRNA   competing endogenous ribonucleic acid

CF   cleavage factor

CPSF   cleavage/polyadenylation specificity factor

CPE   cytoplasmic polyadenylation element

CPEB   cytoplasmic polyadenylation element binding protein

CRM1   chromosomal region maintenance protein 1

CstF   cleavage stimulation factor

DBHS   *Drosophila* Behavior/Human Splicing

DMSO   dimethyl sulfoxide

DNA   deoxyribonucleic acid

DSE   downstream element

EJC   exon junction complex

ERV   endogenous retrovirus

FBS   fetal bovine serum

FISH   fluorescence *in situ* hybridization

GRE   GU-rich element

HFF   human foreskin fibroblast

IP$_6$ inositol hexaphosphate 6

iPS   induced pluripotent stem [cell]

KAP1   KRAB-associated protein 1

KRAB   Krüppel-associated box

LINE   long interspersed nuclear element

LLM-PAT   linker-ligation mediated poly(A) test

MALAT1	metastasis associated lung adenocarcinoma transcript 1
mCAT2	mouse cationic amino acid transporter 2
miRNA	micro ribonucleic acid
Mlxipl	Max-like protein X interacting protein-like
mRNP	messenger ribonucleoprotein complex
m6A	N6-methyladenosine
NEAT1	nuclear-enriched abundant transcript 1
NEXT	nuclear exosome targeting
Nlrp6	NOD-like receptor family pyrin domain containing 6
NPC	nuclear pore complex
NPM1	nucleophosmin
NS	non-significant
nt	nucleotide
NXF1	nuclear export factor 1
ORF	open reading frame
PABPC	cytoplasmic poly(A) binding protein
PABPN1	nuclear poly(A) binding protein
PAP	poly(A) polymerase
PARN	poly(A)-specific ribonuclease
PARP1	poly(ADP-ribose) polymerase 1
PAS	poly(A) signal
PAXT	poly(A) tail exosome targeting
PBS	phosphate buffered saline
PLE	poly(A) limiting element
Ptbp1	polypyrimidine tract-binding protein
Pol II	RNA polymerase II

qPCR	quantitative polymerase chain reaction
RLuc	Renilla luciferase
RNA	ribonucleic acid
RT	reverse transcription
SR	serine/arginine
TE	transposable element
TPR	translocated promoter region
TRAMP	Trf4/5-Air1/2-Mtr4 polyadenlyation
TREX	transcription/export
UTR	untranslated region
VEGF	vascular endothelial growth factor
XRN	5'-3' exoribonuclease
ZC3H14	zinc finger CCCH-type containing 14
4sU	4-thiouridine