**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Kevin J. Morris Date

Defining the function of the polyadenosine RNA-binding protein ZC3H14:
Molecular insight into a disease associated RNA-binding protein

By

Kevin J. Morris
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Genetics and Molecular Biology

| | |
|---|---|
| Anita H. Corbett, Ph.D.<br>Advisor | Gary Bassell, Ph.D.<br>Committee Member |
| Yue Feng, MD, Ph.D.<br>Committee Member | Victor Faundez, MD, Ph.D.<br>Committee Member |
| Nicholas Seyfried, Ph.D.<br>Committee Member | Kenneth Moberg, Ph.D.<br>Committee Member |

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

**Defining the Neuronal Function of the Polyadenosine RNA-binding protein, ZC3H14:**

**Functional Insight into a Disease-causing RNA-binding Protein in the Brain.**

By

Kevin J. Morris

B. S. Biology, University of Mississippi, 2012

Advisor: Anita Corbett, Ph.D.

An Abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Biological and Biomedical Sciences

2018

## Abstract

Defining the neuronal function of the polyadenosine RNA-binding protein, ZC3H14:

Functional insight into a disease-causing RNA-binding protein in the brain.

One critical step for proper eukaryotic gene expression involves the successful execution of nuclear RNA processing events that are coupled to efficient export of the resulting messenger ribonucleoprotein (mRNP). Once properly processed in the nucleus the resulting mature mRNP is exported to the cytoplasm and is subject to ongoing regulation as the mRNA is targeted for transport, translation or decay. RNA-binding proteins are critical for assembling mature mRNPs and facilitating mRNA processing. In this role, RNA-binding proteins both ensure that mRNAs are properly processed and target defective RNAs for decay. The critical need for RNA-binding proteins is evidenced by the growing number of diseases that result from mutations in genes encoding these RNA-binding proteins. One key RNA-binding protein is the zinc finger polyadenosine RNA-binding protein, ZC3H14. Loss of function mutations in the *ZC3H14* gene causes a non-syndromic form of inherited, autosomal recessive intellectual disability. As ZC3H14 is ubiquitously expressed, this finding suggests that there is a critical role for ZC3H14 in the brain. However, the molecular function of the protein has yet to be determined. To probe the function of ZC3H14 in the brain, we carried out a proteomic study to identify the complete spectrum of ZC3H14 co-purifying proteins. From this approach, we identified numerous RNA processing factors that allow us to better understand the role ZC3H14 could play in the brain. By combining biochemical and molecular approaches, we assessed the importance of ZC3H14 and a variety of mRNA factors. We established links to ZC3H14 and the proper control of splicing, 3'end processing and nuclear export. To that

end, we have provided insight into how ZC3H14 and other interacting proteins influence post-transcriptional events of gene expression in the brain.

# Defining the Neuronal Function of the Polyadenosine RNA-binding protein, ZC3H14:

## Functional Insight into a Disease-causing RNA-binding Protein in the Brain.

By

Kevin J. Morris

B. S. Biology, University of Mississippi, 2012

Advisor: Anita Corbett, Ph.D.

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Biological and Biomedical Sciences

2018

**Defining the Neuronal Function of the Polyadenosine RNA-binding protein,**

**ZC3H14:**

**Functional Insight into a Disease-causing RNA-binding Protein in the Brain.**

**Table of Contents:**

**List of Figures:**

# Chapter 1: Background and Significance ~ The Critical Role of RNA-binding Proteins in Regulating Proper Gene Expression.

### 1.1 Posttranscriptional control of gene expression

Every event in a cell relies upon the successful execution of gene expression events. For example, the molecular phenotypes that distinguish a blood cell from a lung cell or the molecular pathways that allow for a stem cell to proliferate while a neuron remains post mitotic are determined through regulation of gene expression as these cells share a common genome sequence. When properly regulated, gene expression allows for a given cell to undergo processes such as differentiation, proliferation, maintenance and eventually cell death, governing the flexibility of a cell to adapt to and respond to environmental stimuli. Gene expression accomplishes these events by employing the central dogma of biology to create protein products from a DNA template, which ultimately determines cell fate and cell function. With all these processes that a cell must undergo, regulation of gene expression is a major tool that the cell employs to properly regulate critical events and functions of the cell.

The events of gene expression start with the transcription of a template DNA to generate an intermediate molecule known as a messenger RNA (mRNA). These mRNA molecules are then translated into protein. While these steps seem simple, the events of gene expression are enormously complex and require a great deal of regulatory control (1). To achieve the level of control required for gene expression to occur properly, there are regulatory control mechanisms that occur at every step of gene expression. To start, DNA is regulated at both the epigenetic and transcriptional level. The epigenetic state of DNA dictates the accessibility of the DNA to transcription factors and allows the cells to separate highly transcribed DNA from silent or inactive areas of the genome (2). Once a gene is selected for transcription, the cell continues to regulate the gene as it is transcribed and as

the nascent RNA is produced. The cell then enforces a host of regulatory mechanisms that ensure that the mRNA is produced properly. The mature mRNA is carefully regulated as it is translated into protein. This control is enforced by a number of cis-acting elements and trans-acting factors throughout the different steps of gene expression. This level of control is critical for gene expression to occur properly.

While all the steps of gene expression are extremely important in dictating the successful creation of a gene product, numerous crucial steps occur as the gene is undergoing transcription and until the mature mRNA is translated into protein. These steps are known as co- and post-transcriptional regulation and cover the entire lifecycle of an mRNA (Figure 1). An RNA is extensively processed and modified as if matures into an mRNA. The 5' end of the RNA must be capped as it emerges from RNA polymerase. The introns of the primary transcript have to be spliced out as transcription proceeds and the 3'end of the mRNA needs to be formed with a catalytic cleavage event and the addition of a 3'polyadenosine tail (Figure 1). After which an mRNA is considered mature and competent for export from the nucleus. The mature mRNA is packaged with a myriad of RNA-binding proteins to form a messenger ribonucleic particle (mRNP) and is exported from the nucleus to the cytoplasm (Figure 1). Once in the cytoplasm, the mRNA can undergo further processing and eventually be translated into protein and/or degraded (Figure 1). As RNA processing occurs both in the nucleus and the cytoplasm, one key distinction with nuclear mRNA processing is that these events occur both co- and post-transcriptionally. These steps are critical as defects in nuclear mRNA processing can affect to downstream processing events in the cytoplasm, affecting the resulting protein levels. Importantly, improperly processed mRNAs have to be retained in the nucleus and not

exported to the cytoplasm to prevent aberrant protein production. For the purposes of this review, we will focus on the regulation of the nuclear processing events that precede the export of the mature mRNA from the nucleus to the cytoplasm.

### 1.1.1 Nuclear mRNA processing events.

#### 1.1.1.1 Capping of eukaryotic mRNAs

Nuclear mRNA processing events start with the regulation of the initial capping of the RNA emerging from RNA polymerase. This event is one of the very first RNA processing steps and occurs as the first 25-30 nucleotides of the nascent mRNA emerge from RNA polymerase II (RNA Pol II) (3,4). The canonical capping step involves the addition of an evolutionarily conserved guanosine-N7 methyl cap to the 5' end of the mRNA molecule (5). The cap is added to the mRNA by capping enzymes that are recruited to the mRNA by the phosphorylation status of the C-terminal repeat domain (CTD) of RNA Pol II (6). The CTD with phospho-Ser5 residues in the heptad repeats has a crucial role in co-transcriptional capping by directly interacting with both guanine-7-methyltransferase and the mRNA-capping enzymes, stimulating the capping enzymes to transfer the cap to the RNA (7,8). The addition of the mRNA cap allows for transcription to proceed, further pre-mRNA processing, cap-dependent nuclear export, and translation of the mRNA to occur (9). Failure of the mRNA to obtain a cap will result in a severely compromised mRNA (10). This mRNA will likely be recognized by mRNA quality control factors and targeted for decay. While some research has shown that mRNAs can exist for some time without caps (11,12), the common belief is still that the cap serves a critical role in dictating further RNA processing events.

#### 1.1.1.2 Splicing of eukaryotic pre-mRNA

Almost every mRNA in higher eukaryotes is transcribed as a precursor to the mature mRNA known as pre-mRNA. Eukaryotic genes are filled with intervening sequences known as introns that need to be removed from a heterogeneous nuclear mRNA so that flanking regions known as exons can be joined together to form a mature mRNA. This process is known as splicing and is essential for mRNA processing. Splicing is made possible through a combination of cis-acting elements and trans-acting factors. The cis-acting elements are found within the mRNA and include a conserved 5′ splice site, a branch point sequence with a polypyrimidine tract, and a 3′ splice site. These cis-acting elements are recognized by a large trans-acting RNA-protein complex known as the spliceosome. The spliceosome is assembled onto the pre-mRNA and facilitates the removal of the introns and the ligation of the exons together (13). Defects in splicing can lead to a compromised RNA that may be targeted for decay or result in aberrant protein production.

#### 1.1.1.3 Alternative splicing of eukaryotic RNA transcripts

As eukaryotes have evolved, so have their genomes. The expansion of the eukaryotic genome has brought about alternative expression mechanisms that expand the coding capacity of genes. One of these mechanisms is alternative splicing which allows for differing inclusion and/or exclusion of certain gene regions to generate different mRNAs that encode distinct protein products from a single gene. For example, creating two transcripts for a single gene makes the complexity of the proteome twice as much as the genome. Recent estimates, based on analyses of expressed sequence tags (ESTs), suggest

that the transcripts from 35% of human genes are alternatively spliced (14), which is likely to be an underestimate. In fact, encoded transcripts that are alternatively spliced can produce up to tens of thousands of different mRNAs. The difficulty in detecting alternative splicing events stems from the fact that many alternative splicing events are very rare and occur only in specific tissues (15). By selectively enforcing alternative splicing events, the cell can choose to express a different splice variant at a specific time in development and/or under certain physiological conditions yielding proteins with distinct functions. With the evolution of complex organisms, the use of alternative splicing allows for a more efficient use of genetic material and the generation of extraordinary molecular diversity.

#### 1.1.1.4 3'end formation of eukaryotic mRNA

Like the 5'end of the mRNA that is modified by the addition of a methyl guanosine cap, the 3' end of the mRNA undergoes a crucial and complex formation event as well. The formation of the 3'end of an mRNA is a two-step process with an initial cleavage event followed by the addition of ~ 200 non-templated adenosines to form a 3' polyadenosine tail (16). Cleavage occurs first between a highly-conserved sequence that is upstream of the actual cleavage and polyadenylation site (16). Additional sequences upstream and downstream of the polyadenylation recruit cleavage and polyadenylation factors and enhance the efficiency of the process. Almost all 3'ends of eukaryotic mRNAs are formed with the addition of a poly(A) tail and the poly(A) tail has been shown to be required for downstream processing events of the mRNA (17). The addition of the poly(A) tail is important for nuclear transport, translation efficiency and mRNA stability as defects in 3' end formation can result in an mRNA not being efficiently transported out of the nucleus

or translated (18). As the poly(A) tail plays such a key role in the fate of the mRNA, the formation of the 3' end is a critical step in gene expression as defects in these steps can have drastic effects on the development, growth, and viability of a cell.

#### 1.1.1.5 Nuclear export of mRNAs

The complete set of mRNA processing events produces an mRNP that is competent enough to be exported from the nucleus to the cytoplasm. The cell uses an mRNP-specific export pathway that releases the mRNP from the nucleus (19). During export, the mRNP undergoes a dynamic rearrangement of associated proteins. First, the mRNP is associated with RNA regulatory proteins that generate the mRNP then the mRNP undergoes remodeling of factors that permit the translocation of the mRNP through the nuclear pore (20). Once through the nuclear pore, the mRNP is further remodeled to allow association with factors that release the mRNA into the cytoplasm (21). The completion of nuclear processing steps is all essential as the mRNA capping, splicing events, splicing-deposited proteins such as the exon junction complex (EJC), and the poly(A) tail play a role in ensuring efficient export (22). Defects in any of these steps can result in the mRNA being retained in the nucleus and targeted for decay.

### 1.1.2 Coordination of nuclear RNA processing events

The expression of protein-coding genes involves a series of highly coordinated and regulated events that process pre-mRNAs into mature mRNAs that encode for protein in the cytoplasm. The mature mRNA is then exported through the nuclear pore to the cytoplasm for further regulation. As previously discussed, the mRNA processing steps of

capping, splicing, 3'-end cleavage and polyadenylation, and nuclear export are all important processing steps for the mRNA and impairment of any of these steps can have devastating consequences for the mRNA. Although these processing steps are thought to be distinct gene-expression events, growing evidence suggests that these events are linked and influence the progression of one another (23-26). Recent advances have elucidated the mechanisms by which transcription and mRNA processing steps are coupled to dictate the fate of an mRNA (23). In this section, the discussion will focus on events in which the coupling of RNA processing steps is both beneficial and required for further processing steps of the RNA in the nucleus.

#### 1.1.2.1 RNA Polymerase II is a source for coupling mRNA processing events

As mRNA processing begins as transcription occurs, RNA Pol II serves as one of the initial factors that couples transcription with mRNA processing events. RNA Pol II is an extraordinary enzyme that uses its CTD tail to recruit factors necessary to couple transcription and mRNA processing. The CTD acts as a 'landing pad' (27) that recruits processing factors by binding directly to capping factors, splicing factors and 3′ end processing factors (28,29). The recruitment of these RNA processing factors to the CTD of RNA Pol II complex facilitates the co-transcriptional assembly of multi-subunit processing factors for events such as capping, splicing, and 3'end formation (30,31). The CTD can recruit different processing factors through differential phosphorylation patterns on the CTD as the polymerase moves from the cycles of transcription initiation, elongation, and termination (32). These reversible modifications allow for the polymerase to signal for different pre-mRNA processing factors as the polymerase proceeds (32). The CTD of RNA

Pol II directs the recruitment of mRNA processing factors making RNA Pol II an important molecule for the spatial and temporal coupling of transcription and pre-mRNA processing events.

#### 1.1.2.2 Coupling of Transcription to Splicing and Nuclear Export

Emerging studies are providing additional evidence that transcription and splicing are linked and that these two events can affect one another (33,34). Due to the fact that splicing typically occurs co-transcriptionally, these results are not terribly surprising. However, recent studies have shown that the coupling of transcription and splicing can affect the downstream event of nuclear export (35). This suggestion emerges from the finding that nuclear export components in mammals associate with splicing factors (36,37). In eukaryotes, nuclear export is initiated by a complex of factors known as the TREX complex (35). In mammals, this complex contains the six-member mRNA processing complex, THO (38) as well as the mRNA export proteins ALYREF (39) and UAP56 (40). Whether the TREX complex associates with the nascent pre-mRNA during transcription is not yet clear, recent studies show that the TREX complex functions to link both transcription and splicing to mRNA export in metazoans (41). In metazoans, introns are typically thousands to tens of thousands of nucleotides in length compared to exons that average only ~100 nucleotides. The TREX complex may be recruited during splicing to aid in distinguishing between exons that are to be exported and excised introns that are to be degraded. This model is complemented by the fact that the THO complex associates with the EJC complex (EJC) (42). By linking the processing events of transcription, splicing and nuclear export the cell can gain tighter control over these coupled mRNA

processing events and have the potential to improve the efficiency of them all. Coupling of these events is also another way to ensure that all the processing steps are complete before the RNA exits the nucleus.

#### 1.1.2.3 Coupling Splicing, 3'end formation and nuclear export

The mRNA export machinery is recruited to mRNAs as early as splicing (43,44). However, recent studies show that in addition to splicing, 3' end formation may also play a role in the recruitment of nuclear export factors (45,46). For example, work in budding yeast reveals that recruitment of the mRNA export factor Yra1 depends on proper formation of the 3'end (45). In addition, export of T7 polymerase transcripts is dependent on the 3'end formation of the transcript (47). Furthermore, deletion of several 3'end processing factors in yeast cause nuclear accumulation of mRNA (48). Taken together, these studies show that like splicing and transcription, coupling the formation of the 3'end with other processing steps provides a mechanism to the proper formation of an mRNA before it is exported from the nucleus. These events are important to prevent the export of improperly processed mRNA out to the cytoplasm.

#### 1.1.2.4 Coupling Alternative Splicing with 3'end formation

Alternative splicing mechanisms involves making decisions about which exons to splice out and which exons to retain. Many of these alternative splicing events are believed to occur co-transcriptionally (49-51), which requires quick and careful control over which factors to recruit to RNA Pol II during transcription. One of the most important splicing events is the definition of the first and last exon. These exons are unique because they only

have one flanking splice site and require a distinct mechanism of recognition as compared to internal exons(52). One way in which the cell recognizes the terminal splicing exon is through use of the cleavage and polyadenylation machinery (53,54). Furthermore, recent work shows that the poly(A) tail is important for terminal intron removal (55). Splicing factors can also influence polyadenylation by interacting with cleavage and specificity factors (56). Thus, the interactions of splicing and polyadenylation factors at terminal exons play important roles in the regulation and enhancement of both splicing and polyadenylation.

### 1.1.3 Quality Control of nuclear processing events

As previously discussed, an mRNA undergoes a series of processing steps before a mature mRNP is competent for export to the cytoplasm. Nuclear processing either yields an mRNA that is mature and thus appropriate for translation or a defective mRNA that is targeted for degradation. Nuclear pre-mRNA processing steps are coupled from the start of transcription and rely on the concurrent recruitment of mRNA processing factors to ensure that processing steps occur correctly (57,58). The benefit of coupling mRNA processing events is for an increase in the efficiency of processing and tighter control over mRNA processing events (59). An additional reason for the coupling of mRNA processing steps could be to allow early monitoring of mRNP quality as errors in processing steps can have detrimental effects on the mRNA and ultimately the cell. In the next section, we will examine mechanisms the cell employs to monitor the production of mRNA. In addition, this section will highlight ways mRNA quality control is incorporated into RNA processing steps to ensure efficient and proper processing of mRNA prior to export from the nucleus.

#### 1.1.3.1 SR Proteins as mediators of mRNA Stability

Mammalian genes contain a vast number of intronic sequences that need to be removed from the pre-mRNA to produce a mature coding mRNA. To ensure that these splicing events happen properly, cells rely on a combination of proteins that work to enhance splicing and proteins that ensure that splicing events occur properly. One set of proteins that ensures proper splcing has occurred is the exon junction complex (EJC). The EJC is deposited upstream of exon-exon boundaries and has been reported to influence mRNA export, the surveillance process known as nonsense-mediated mRNA decay (NMD), and the efficiency of mRNA translation (60). During splicing, another set of proteins termed serine arginine rich or SR proteins are loaded onto the mRNA (61). While SR proteins play a canonical role as regulators of alternative splicing (62), Recent work has revealed that some of these SR proteins are involved in mRNA quality control (63). Overexpression of some SR proteins leads to changes in splice site choice, but also significantly enhances the sensitivity of nonsense codon-containing alternatively spliced messages to degradation via the NMD pathway (64). In this case, a role for SR proteins emerges as these proteins are typically involved in alternative splicing but have now developed a new role for surveying the quality of the mRNA as it is spliced.

#### 1.1.3.2 Monitoring of mRNA processing by the Exosome and the TREX complex

The coupling of nuclear processing events also couples mRNA processing factors from different steps in mRNA biogenesis. As defective mRNAs need to be targeted for degradation, having RNA processing factors in close proximity to the RNA degradation

machinery provides an efficient mechanism. Recent work that discovered links between the transcription elongation and nuclear export complex, TREX and the nuclear RNA exosome (65,66) provides an example of such coupling. The TREX complex is typically involved with connecting the earlier mRNA processing steps to nuclear export while the nuclear exosome is a large complex with 3' to 5' exonucleases activity and involved in numerous RNA processing and degradation pathways (41,67). These studies in budding yeast show that there are genetic and physical interactions between the RNA exosome and TREX complex components. The yeast orthologue of ALYREF, Yra1 physically interacts with the RNA exosome protein Rrp45 (66). Mutants of TREX complex components generate transcripts that are truncated at their 3' ends and are sequestered near their sites of transcription (68). However, this phenotype can be suppressed by the deletion of the exosome protein Rrp6 (65). By linking the RNA exosome to TREX components the cell can quickly degrade defective mRNAs at their site of transcription before the RNA escapes the quality control checkpoint.

**1.1.3.3 mRNP surveillance and remodeling at the NPC.**

The last step in mRNA processing is the translocation of the mRNP from the nucleus to the cytoplasm via transit through the nuclear pore. As this step is the last chance for the cell to identify defective mRNAs before they reach the cytoplasm, the cell has adapted mechanisms to monitor the quality of an mRNA at the nuclear pore. In addition to the nuclear RNA exosome, several nuclear pore proteins located at the nuclear basket and at the inner face of the nuclear pore have been implicated in mRNP surveillance before translocation through the pore (69). The proteins include the nuclear pore proteins such as

Nup60 and TPR in mammals and the pre-mRNA leakage 39 (Pml39) protein as well as the nuclear basket integrity protein Esc1 in budding yeast (70). Studies with these proteins have shown that these proteins have no essential role in mRNA export but are implicated in the nuclear retention of unspliced pre-mRNAs (70). These proteins can usually detect unspliced transcripts and recruit degradation factors such as the nuclear exosome and other endonucleases such as Swt1 (71). Endonucleases such as Swt1 are typically associated with nuclear pores and have been shown to functionally interact with nuclear pore proteins such as Nup60 and mRNA export proteins such as the THO complex (72). Thus, the presence of mRNA quality control proteins at the nuclear pore showcases another mechanism by which mRNA quality control factors are integrated into mRNA processing steps to ensure only properly processed RNAs can interface with translation machinery in the cytoplasm.

**1.2 The importance of RNA-binding proteins in gene expression**

All the events of RNA-biogenesis are mediated by a class of factors known as RNA-binding proteins. RNA-binding proteins fully decorate an mRNA molecule from the time the nascent mRNA emerges from RNA polymerase to the time the mRNA is destroyed (73). The involvement of RNA binding proteins in every event of the life of an RNA makes these factors a diverse class of proteins. Current studies predict that the human genome encodes approximately 1,900 RNA binding proteins with significant diversity in their RNA binding domains (74). The defining feature of an RNA-binding protein is the ability of the protein to recognize and bind RNA. The best-characterized RNA binding domain is the RNA recognition motif, or RRM (75) but this domain is just one of the many types of domains that an RNA-binding protein can employ to bind RNA. In fact, RNA-binding

proteins can contain multiple of the same or different RNA-binding domains. However, many mechanisms by which RNA-binding proteins bind RNA have yet to be defined. As RNA-binding proteins function in every aspect of RNA biology, the use of different RNA-binding domains confers specificity to RNA-binding proteins for specific RNA sequences or structures and can allow fine tuning of the regulations of specific RNAs by designated RNA-binding proteins. In this section, we will discuss the importance of RNA-binding proteins and highlight the role of one particular RNA-binding protein in regulation of RNA processing.

### 1.2.1 RNA-Binding proteins and human disease

Defects in mRNA processing can have deleterious effects on the production of mRNA. RNA-binding proteins have a crucial role in regulating the processing of mRNA and misregulation of these factors can impair the proper production of mRNA. As mRNA processing events are coupled and mRNA quality control mechanisms are incorporated into these events, proper recruitment and functioning of RNA-binding proteins is critical for gene expression. The interplay of RNA-binding proteins in mRNA processing creates a large network of RNA-processing factors that coordinately regulate the generation of mRNA and the degradation of defective mRNAs. Proper function of these intricate networks is essential for the coordination of complex gene expression events and perturbations due to loss of a particular factor can lead to disease.

A growing number of diseases have been linked to mutations in genes encoding RNA-binding proteins (76). Pathology caused by defects in these genes encoding RNA-binding proteins is usually highly tissue-specific (77) despite the fact that many of these

proteins are ubiquitously expressed. In heritable diseases, this finding suggests that specific spatiotemporal functions of the implicated genes are disrupted and certain tissues are more susceptible to the effects of the mutation. Research has started to focus on examining single gene mutations in affected tissue (78,79). However, how mutations in ubiquitously expressed genes give rise to tissue-specific pathology is not yet known.

#### 1.2.1.1 The polyadenosine RNA-binding, ZC3H14 is associated with intellectual disability

An example of a ubiquitously expressed protein that is linked to tissue specific pathology is the RNA-binding protein ZC3H14. Two independent mutations in the gene encoding the RNA-binding protein ZC3H14 cause an autosomal recessive, non-syndromic form of intellectual disability (80), which is a disorder characterized by cognitive delays and limitations in adaptive behaviors (81). The *ZC3H14* patients display severely low IQ scores of ~30 placing them at the very severe end of the spectrum as the average IQ score of the general population is around 100 (80). The mutations in the *ZC3H14* gene severely affect the cognitive capabilities of the *ZC3H14* patients, suggesting that ZC3H14 is an important protein for brain function.

The *ZC3H14* gene is alternatively spliced to generate four distinct protein isoforms in mammals with isoforms 1-3 being the most similar to each other (Figure 1.2A) (82). These isoforms contain a proline tryptophan isoleucine (PWI)-like fold in the N-terminal domain, a nuclear localization signal (NLS), and tandem CCCH zinc finger motifs in the C-terminus (Figure 1.2A). The fourth isoform contains a unique N-terminus but retains the C-terminal zinc finger motifs (Figure 1.2A). Isoforms 1-3 are ubiquitously expressed and

are primarily found in the nucleus associated with SC35-positive nuclear speckles (82), which are sites of RNA metabolism (83). Isoform 4 which is enriched in the testis with low levels detected in other tissues, lacks an NLS and is consequently primarily localized to the cytoplasm (Figure 1.2A) (82). While the ZC3H14 protein isoforms all contain the C-terminal zinc finger motif, there are subtle differences that exist between the isoforms.

While intellectual disability has heterogeneous causes, a subset of the disease results mutations in single genes, which can provide insight to the function of the genes. In the case of the *ZC3H14* mutation, one family has a nonsense mutation that replaces the sequence encoding arginine 154 with a premature termination codon (R154Stop) in exon 6 (Figure 1.2A). The second family mutation is a deletion of 25 nucleotides in exon 17, which affects the last ammino acids of the protein (Figure 1.2A). Examining patient fibroblasts from the first family, we found that the R154 to stop mutation causes loss of expression of the nuclear isoforms of ZC3H14 (80). ZC3H14 is shown to be ubiquitously expressed but the patients display brain-specific phenotypes, suggesting that the ZC3H14 protein plays a critical role in the brain. The nature of this phenotype revealed that work needs to be performed to probe the function of ZC3H14 in the brain. At the time of this study, little was known about the function of the ZC3H14 protein. In the next section, we will explore what was previously known about the function of ZC3H14 and functional information gathered from studies in the *Drosophila* and budding yeast model systems.

### 1.2.2 The polyadenosine RNA-binding protein ZC3H14 is evolutionarily conserved

#### 1.2.2.1 Nab2, the budding yeast orthologue of ZC3H14 is important for mRNA processing.

Initial insight to the function of the ZC3H14 protein came from detailed analysis of the budding yeast orthologue of ZC3H14, Nab2. Nab2 or Nuclear poly(A)-binding protein 2 (Nab2) is essential in budding yeast and has been linked to RNA processing in the nucleus (84-87). Like ZC3H14, the Nab2 protein contains an N-terminal Proline-Tryptophan-Isoleucine (PWI)-like domain which is required for proper poly(A) RNA export from the nucleus, a nuclear targeting arginine-glycine rich (RGG) domain, and a C-terminal tandem CCCH ZnF domain which mediates binding to polyadenosine RNA (88) (Figure 1.2B). *nab2* mutant yeast cells display extended poly(A) tails and nuclear accumulation of poly(A) RNA (87). Recent analysis of the Nab2 protein has identified a role for the protein in splicing (89) and that rapid nuclear depletion of Nab2p leads to a global loss of cellular mRNA by robust nuclear mRNA decay mediated by the RNA exosome in a polyadenylation-dependent process (90). This work has helped establish Nab2 as a critical mRNA processing factor in the nucleus.

#### 1.2.2.2 dNab2, the fly orthologue of ZC3H14 has an important role in the brain

The *D. melanogaster* ortholog of ZC3H14 is dNab2 (80). dNab2 is expressed ubiquitously and has a similar domain structure to Nab2 and ZC3H14 (Figure 2B). Like Nab2 and ZC3H14, dNab2 has an N-terminal PWI-like fold, an NLS, and C-terminal tandem zinc fingers motif (Figure 2B). Ablation of the dNab2 gene results in decreased viability, with only a small percentage of flies surviving to adulthood (80). The surviving *dNab2* mutant flies can develop into adult flies likely due to a residual amount of maternal

dNab2, however these flies show a variety of organismal and molecular defects (91). Like *nab2* mutants yeast cells, *dNab2* mutant flies have extended poly(A) tails in bulk RNA (92). *dNab2* mutant flies show impaired flight and negative geotactic locomotor behavior (80). Importantly, the *Drosophila* model has provided insight into tissue specific requirements for dNab2. Studies show dNab2 is critical in neurons (80) as *dNab2* mutant flies have compromised brain function and severe brain morphology defects (91). Importantly, these defects in the fly can be rescued by the expression of either dNab2 or the mammalian ZC3H14 protein exclusively in neurons (91). Together, these data support a model where dNab2 plays a critical role in mRNA processing in the brain.

#### 1.2.2.3 ZC3H14 protein has a specialized role in regulating mRNA processing in the brain

Previous biochemical analyses of ZC3H14 protein shows that ZC3H14 binds to polyadenosine RNA with high affinity (93). Furthermore, siRNA-mediated depletion of ZC3H14 in a mammalian cell line causes extended bulk poly(A) tails (94,95). Although ZC3H14 binds with high affinity to polyadenosine RNA and in theory could bind any RNA that contains polyadenosine stretches, work from our group has shown that depletion of ZC3H14 only affects a specific subset of the transcriptome (96). Further analysis has shown that in primary hippocampal neurons, ZC3H14 protein is found both in the nucleus and the neuronal processes, suggesting a specialized role for this protein in the brain (97). Furthermore, the generation of a *Zc3h14* mutant mouse model revealed that loss of ZC3H14 is not essential and impairs higher order brain function with the mutant mice displaying defects in working memory (98). These findings coupled with the phenotype of

the patients that harbor mutations in the *ZC3H14* gene further emphasize the continued need to understand the function of ZC3H14 in the brain.

### 1.3 Using protein interaction networks to study human disease

As mRNA processing steps are not linear, RNA-binding proteins are needed to couple multiple processing steps and ensure that the mRNAs are processed properly. The coupling of RNA-processing events groups RNA-processing factors into extensive interaction networks to carry out functions. Most RNA-processing group into multimeric structural complexes that have similar functions and participate in the same biochemical pathways (99). Following this logic, one approach to understand the function of an uncharacterized protein is to define the interaction network. Several large- and medium-scale interaction studies have been undertaken using the yeast two hybrid (Y2H) system or mass spectrometry-based functional proteomics approaches (100) to build interaction networks of proteins in order to learn more about the function of a protein. These studies have provided the scientific community with predictions on protein function in a variety of model organisms. These findings indicate that identifying the interaction network of an uncharacterized protein is a useful tool to identify the complexes or pathways in which a particular protein participates. Using this approach, we can gain insight into the function of the ZC3H14 protein by defining interacting partners.

#### 1.3.1 The majority rule of determining protein function

A common approach in assigning function to an unclassified protein is to base the function of the protein on the classified functions of the identified interacting proteins. This

approach has been coined the ‘majority rule’ assignment of protein function (101,102). The majority rule is derived from the empirical observation that 70–80% of interacting protein pairs share at least one function (103). Assignment of function by majority rule relies on an unclassified protein interacting with proteins that have identified function. For example, if a protein-protein interaction map of an uncharacterized protein yields strong links to mRNA splicing, the majority rule assignment would assign the function of the protein as a RNA splicing factor. A caveat to this approach is that the analysis does not include proteins of unknown function. The reconstructed protein-protein interaction network has to enrich for

proteins with known functions otherwise this method cannot be employed. These constraints create a situation where the proposed function of a protein is determined by the characterized function of identified proteins. However, the advances to protein identification have expanded our knowledge of the function of proteins increasing the chances of an uncharacterized protein interacting with other proteins of known function. By using the majority rule, we can define the function of ZC3H14 by linking it to the function of ZC3H14 interacting partners.

### 1.3.2 Identifying protein binding partners using mass-spectrometric analysis

Mass Spectrometry has emerged as a useful technique to identify proteins that associate with one another. This method has been termed interaction proteomics and involves a co-purification of a protein of interest with associated binding partners followed by mass-spectrometric analysis to identify all the proteins in the sample (104). One of the many benefits of this approach is that thousands of proteins can be detected in a single experiment, allowing the creation of an interaction network for a protein of interest. This type of proteomics approach has established that most gene products exert their function as members of one or more protein complexes (105,106), and that changes in different proteins participating in the same complex, such as cellular machines, rigid structures, dynamic signaling networks, and posttranslational modification systems, generally lead to similar phenotypes (107,108). Using interaction proteomics, the entire spectrum of proteins that interact with ZC3H14 can be identified in a tissue specific manner. This approach will inform the function of the ZC3H14 protein and provide insight into how the loss of ZC3H14 could cause brain dysfunction. By defining the interaction network of ZC3H14,

we can analyze the link between tissue specificity of the complexes formed with ZC3H14 and the pathological manifestations of disease.

### 1.4 Research Questions to be answered and Innovation

My thesis work seeks to define interactions between ZC3H14 and specific RNA regulatory factors in the brain to increase our understanding of the cellular functions of ZC3H14 (Figure 1.3). This study explores the hypothesis that ZC3H14 works cooperatively with specific RNA regulatory factors to ensure proper mRNA processing. The long-term goal is to provide insight into how ZC3H14, along with other interacting proteins, influences neuronal post-transcriptional events to ensure proper higher order brain function. Here, we report our work identifying the RNA regulatory factors that interact with ZC3H14 in the murine brain. We isolated mouse brain tissue and used mass spectrometry to define the complete spectrum of proteins that interact with ZC3H14 in the brain (Figure 1.3). This work has led to many discoveries about the role ZC3H14 plays in the brain and particularly provided insight about how ZC3H14 and interacting proteins cooperate to regulate mRNA processing. In the upcoming sections, I will discuss the interaction between ZC3H14 and RNA processing factors (Chapter 2). In Chapter 3, I will discuss our work elucidating the interaction and coordinated function between ZC3H14 and the mRNA processing complex, THO. Finally, I will conclude with a discussion of my finding and the future implications of my work (Chapter 4).

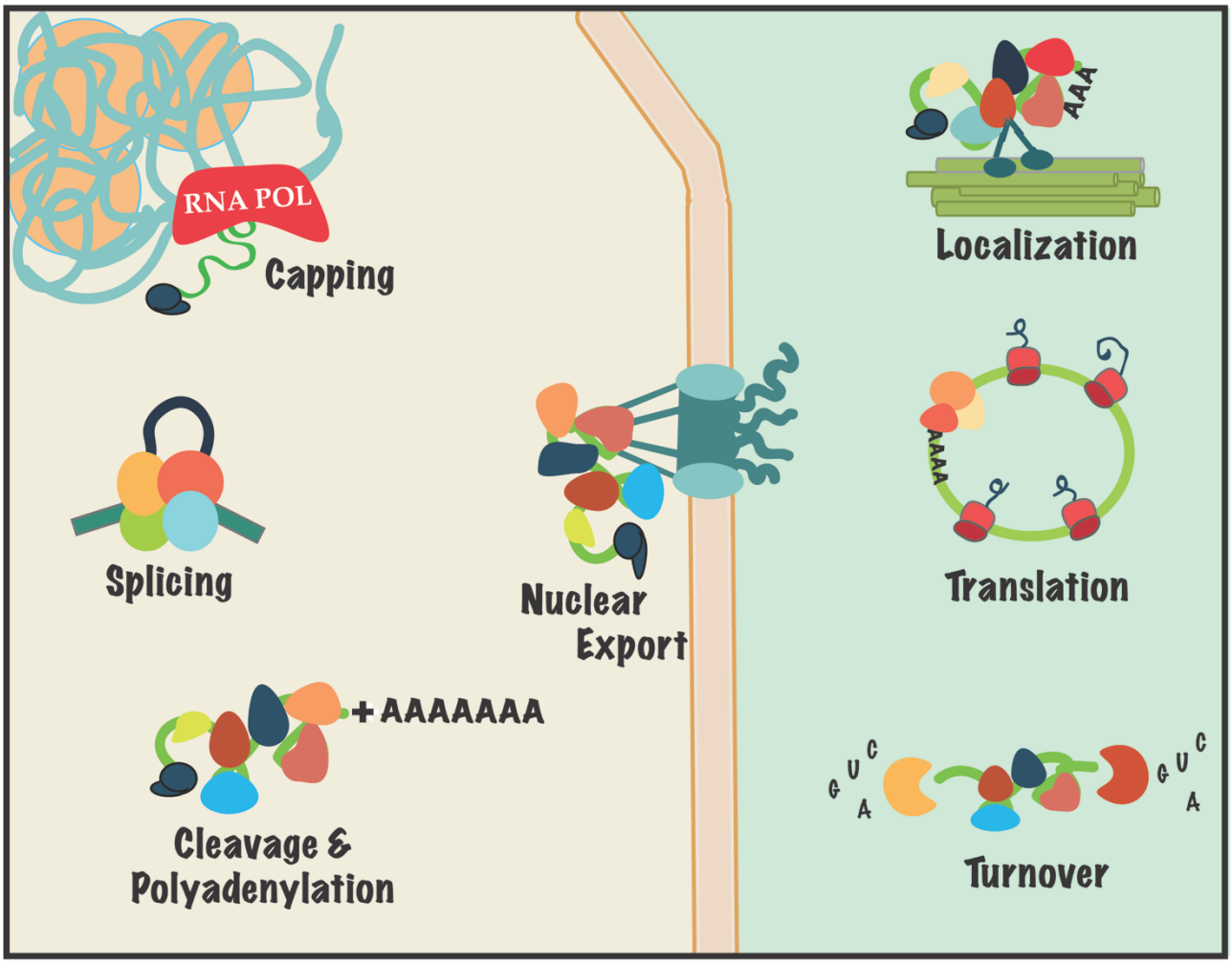



**Figure 1.1: Diagram of eukaryotic mRNA Biogenesis**
Eukaryotic mRNA processing starts as the RNA emerges from RNA polymerase II and the nascent mRNA is capped by capping proteins. The introns are then removed from the transcript and the exons are ligated together through a process known as splicing. The 3'end of the transcript is then formed with cleavage and polyadenylation. The mRNA is then handed over to the nuclear pore complex and exported from the nucleus to the cytoplasm. Once in the cytoplasm, the mRNA can undergo the processes of localization, translation and/or turnover.

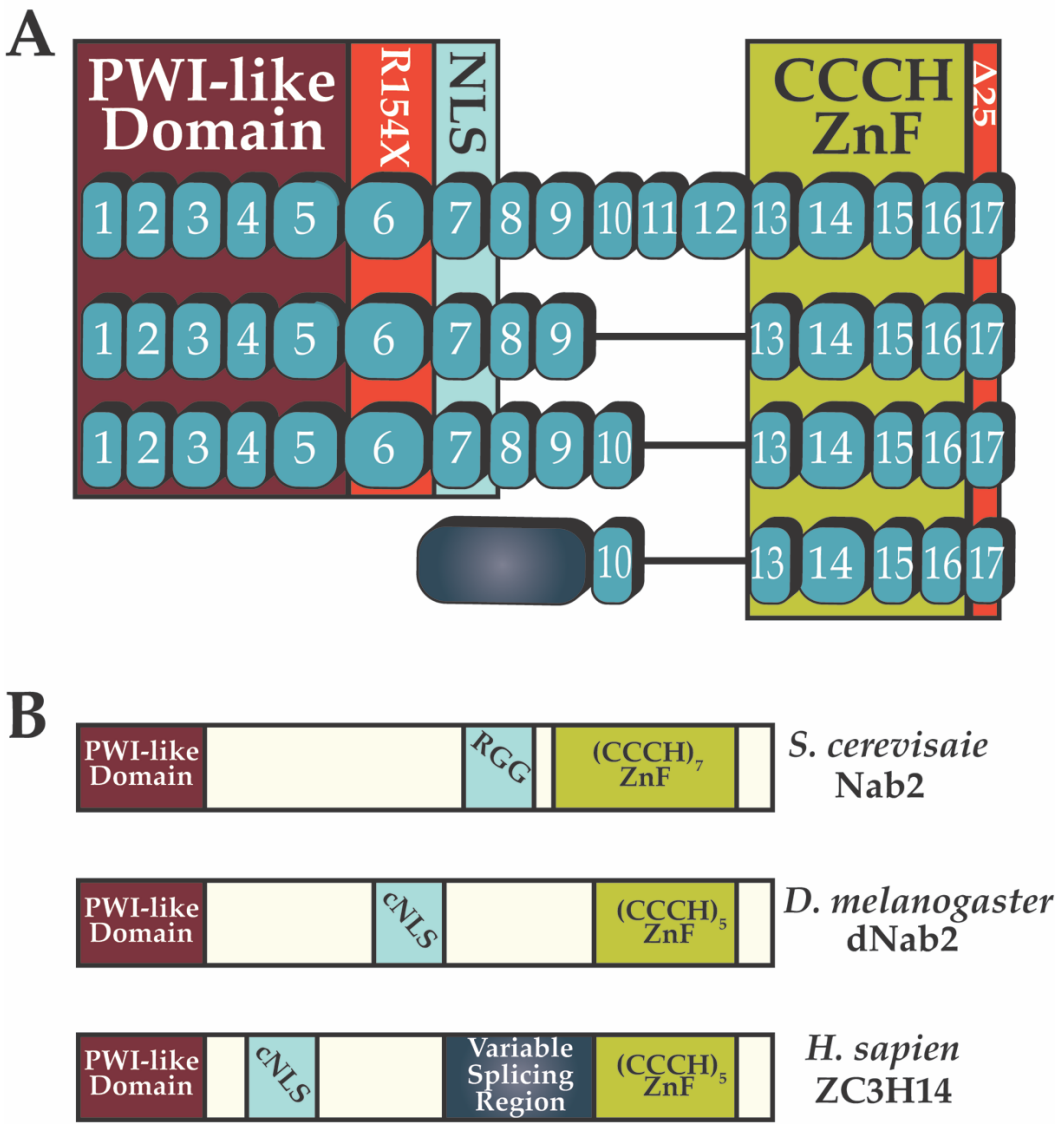



**Figure 1.2 ZC3H14 and functional orthologues in *D. melanogaster* and *S. cerevisiae*.** **A**: Human ZC3H14 encodes at least four splice variants. ZC3H14 isoforms 1-3 each encode a PWI-like domain (exons 1–5), a central cNLS (exon 7) and five tandem CCCH zinc fingers (exons 13–16). Isoforms 2 and 3 differ from isoform 1 in the exclusion of exons 10–12 or 11–12, respectively. Isoform 4 has an alternative transcriptional start site and first exon. As isoform 4 lacks exon 7, it does not contain the predicted cNLS. The identified ZC3H14 patient mutations are shown in red boxes. **B**. The Nab2/dNab2/ZC3H14 family of zinc finger polyadenosine RNA binding proteins. The Nab2/dNab2/ZC3H14 proteins share conserved domain architecture. The N-terminal domain contains a Proline-Tryptophan-Isoleucine (PWI)-like domain. Nab2 contains an RGG domain while dNab2 and ZC3H14 contain a nuclear targeting signal in the central region of the protein. ZC3H14 gene results in transcripts that encode several protein isoforms. The variation among these isoforms arises from a central alternatively, spliced region (exons 10–12) we have labeled the "variable splicing region".

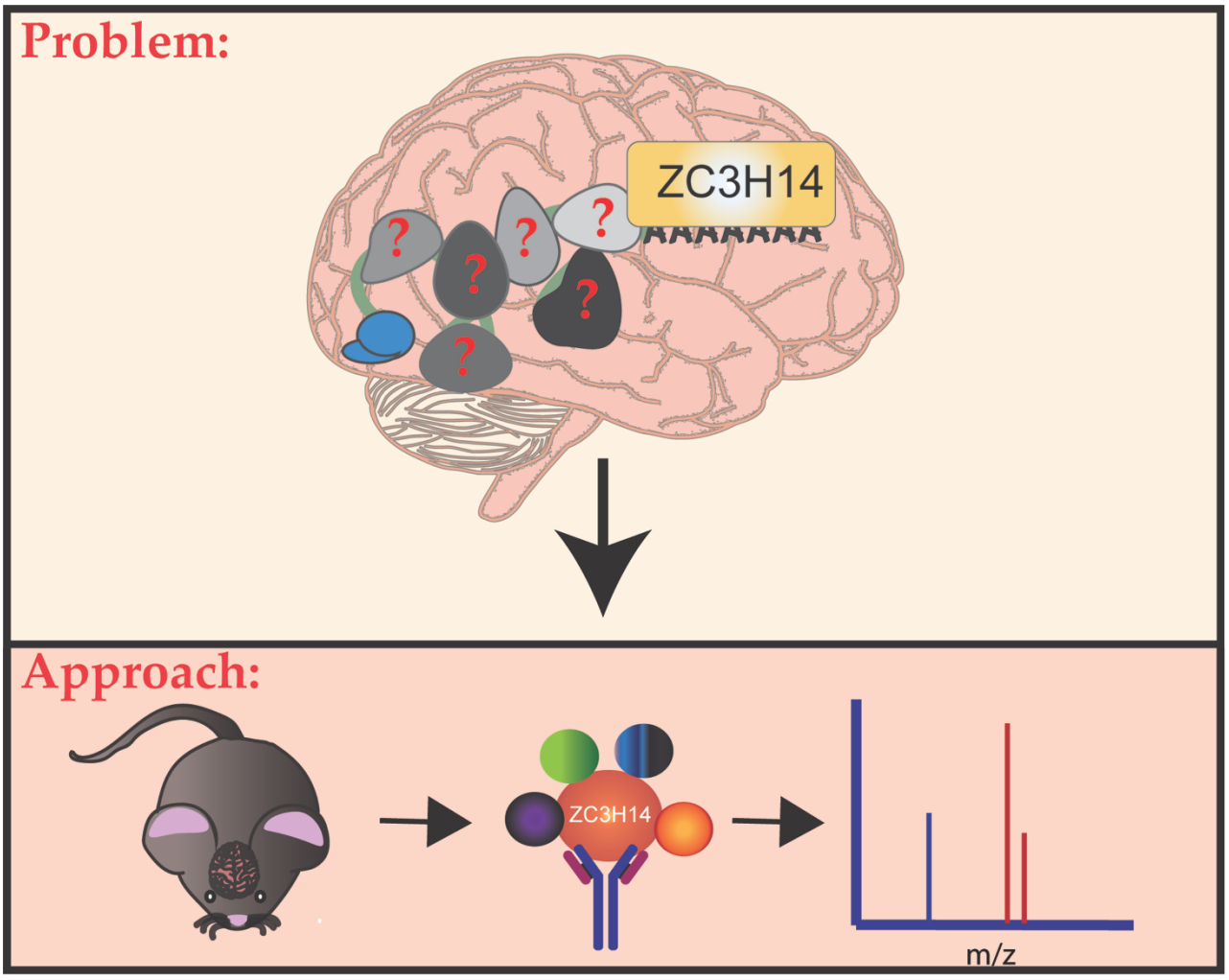



**Figure 1.3: Approach to define the function of ZC3H14 in the brain.**
Our work aims to define the function of ZC3H14 in the brain by identifying the RNA processing factors that interact with ZC3H14. Using a proteomic approach, we isolated mouse brain tissue and purified ZC3H14 protein from mouse brain lysate. Interacting proteins were identified using mass spectrometry analysis.

# Chapter 2: ZC3H14 Interacts with a Variety of RNA Regulatory Factors

## 2.1 Summary of Chapter:

Growing evidence links genes encoding RNA-binding proteins to human disease. Surprisingly, mutations in genes encoding ubiquitously expressed RNA-binding proteins often cause tissue specific diseases. Recently, our group reported inactivating mutations in the *ZC3H14* gene, which encodes an evolutionarily conserved, ubiquitously expressed polyadenosine RNA-binding protein, ZC3H14. Despite ubiquitous expression of the ZC3H14 protein, the *ZC3H14* mutant patients display non-syndromic, intellectual disability where only higher order brain functions are impaired. This finding suggests that ZC3H14 has a critical, specialized role in the brain. While the steady-state localization of ZC3H14 is nuclear, our recent studies reveal that a population of ZC3H14 is present in the cytoplasm. These results suggest that ZC3H14 could play a role in the cytoplasm, like that of other RNA-binding proteins that shuttle between the nucleus and the cytoplasm. We hypothesize that ZC3H14 works cooperatively with specific RNA regulatory factors to regulate neuronal mRNA processing. In this Chapter, we test our hypothesis by investigating the molecular functions of ZC3H14 through identifying the proteins that interact with ZC3H14. Understanding the function of ZC3H14 is a critical step to defining the complex roles of RNA-binding proteins in the regulation of gene expression.

## 2.2 Introduction:

Post-transcriptional events are orchestrated by a myriad of RNA-binding proteins and RNA regulatory factors that ensure the proper processing of RNA needed for proper gene expression (109). Aberrant gene expression can cause serious consequences and ultimately human disease (110). The timing, location, variety, and amount of each gene expressed by a given cell determines the cellular expression profile and maintenance of normal cellular function. From the start of transcription initiation, RNA-binding proteins facilitate the highly regulated post-transcriptional steps of nuclear processing, export, and turnover of mRNA transcripts (109). A given mRNA is fully decorated by a host of RNA-binding proteins that ensure every step is carried out properly (111). There are many examples of RNA-binding proteins that are dysregulated in various diseases (112), highlighting the importance of the need to functionally characterize the role of RNA-binding proteins.

One important RNA-binding protein is ZC3H14 which is a zinc finger polyadenosine RNA binding protein. While ZC3H14 is ubiquitously expressed, the inactivating mutations in the *ZC3H14* gene are linked to a non-syndromic form of intellectual disability. This finding suggests that ZC3H14 plays a critical role in the brain. While ZC3H14 is primarily localized to nuclear speckles (82), recent data shows that ZC3H14 is also present in the axons of primary cultured hippocampal neurons (97). Given that the mechanism underlying intellectual disability in the *ZC3H14* patients is unknown and the function of ZC3H14 has yet to be defined, understanding the interaction between ZC3H14 and other RNA regulatory factors is an approach to gain insight into the function of ZC3H14 in the brain.

Here, we report our work identifying the proteins that interact with ZC3H14 in the murine brain using unbiased mass spectrometry. The identified proteins that interact with ZC3H14 have allowed us to establish links between ZC3H14 and RNA regulatory pathways. This information gained is critical in understanding the role of ZC3H14 in the brain and the molecular phenotypes underlying the disease.

## 2.3 Materials and Method:

### ZC3H14 immunoprecipitation

Brain tissue was collected from wild-type Black 6 mice and homogenized in immunoprecipitation (IP) buffer (50 mM Tris-HCl [pH 7.4], 100 mM NaCl, 32 mM NaF, and 0.5% NP-40 in diethyl pyrocarbonate [DEPC]-treated water) supplemented with 1 cOmplete mini protease inhibitor tablet (Roche; 1 tablet/10 ml of buffer). Homogenate was sonicated on ice five times at 0.5% output for 10 s and then passed five times through a 27-gauge syringe. Lysates were spun at 13,000 rpm for 10 min at 4°C, and protein concentrations were determined using a standard bicinchoninic acid (BCA) assay (Pierce). Protein G-magnetic beads (Dynabeads; Invitrogen) were suspended in IP buffer and incubated with preimmune rabbit serum or an equal volume of ZC3H14 antibody for 1 h at room temperature. Bead/antibody and bead/preimmune samples were added to clarified cell lysates, followed by incubation at 4°C overnight while tumbling end over end (10% removed prior to overnight incubation for input samples). After incubation, for mass spectrometric (MS) analysis, the beads were magnetized, and unbound samples were collected (10% of input) and washed five times with ice-cold IP buffer. For immunoblotting, ZC3H14 protein complexes were eluted with reducing sample buffer (250

mM Tris-HCl, 500 mM dithiothreitol [DTT], 10% SDS, 0.5% bromophenol blue, and 50% glycerol).

**Immunoblotting**

Protein lysates were boiled in reducing sample buffer and resolved on 4-20% Criterion TGX polyacrylamide gels (Bio Rad). Proteins were transferred to nitrocellulose membranes and incubated for at least one hour in blocking buffer (5% non-fat dry milk in 0.1% TBS-Tween). This step was followed by an overnight incubation at 4°C in primary antibody diluted in blocking buffer. Primary antibodies were detected using species-specific horse radish peroxidase (HRP) conjugated secondary antibodies (Jackson ImmunoResearch) followed by incubation with enhanced chemiluminescence substrate (ECL, Sigma). Chemiluminescence was detected by exposing blots to autoradiography film (Daigger). Primary antibodies and dilutions employed as they appear: ZC3H14, (1:6000) (82), PABPN1, (1:5000) (113), (PABPC1, Abcam, ab21060), Histone, (1:10000),(Bethyl, A300-822A).

**Mass spectrometry analysis**

To identify ZC3H14 interacting proteins, bead mass spectrometry was used. After several washing steps, Sepharose beads from control and ZC3H14 immunoprecipitates were resuspended in 8 M urea–100 mM NaHPO4 (pH 8.5; 50 µl, final volume) and treated with 1 mM DTT at 25°C for 30 min, followed by 5 mM iodoacetamide at 25°C for 30 min in the dark. The samples were then diluted to 1 M urea with 50 mM ammonium bicarbonate (final volume, 400 µl) and digested with lysyl endopeptidase (Wako; 1.25 ng/µl, final concentration) at 25°C for 4 h and further digested overnight with trypsin (Promega; 1.25

ng/µl, final concentration) at 25°C. The resulting peptides were desalted with a Sep-Pak C18 column (Waters) and dried under vacuum. Each sample was resuspended in loading buffer (0.1% formic acid, 0.03% trifluoroacetic acid, 1% acetonitrile) and analyzed independently by reverses-phase liquid chromatography (LC) coupled with tandem mass spectrometry (MS/MS) as essentially previously described with slight modifications (114,115). Briefly, peptide mixtures were loaded onto a C18 nanoLC column (75 µm [inner diameter], 15-cm long, 1.9-µm resin [Maisch GmbH]) and eluted over a 5 to 0% gradient (buffer A: 0.1% formic acid; buffer B: 0.1% formic acid in 100% acetonitrile). Eluates were monitored in an MS survey scan, followed by 10 data-dependent MS/MS scans on a Q-Exactive plus Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA). The acquired MS/MS spectra were searched against a concatenated target decoy mouse reference database (v.62) of the National Center for Biotechnology Information (downloaded 14 November 2013 with 30,267 target entries) using the SEQUEST Sorcerer algorithm (version 4.3.0; SAGE-N). The search parameters included the following: fully tryptic restriction, parent ion mass tolerance of ±50 ppm, up to two missed trypsin cleavages, and dynamic modifications for oxidized Met (+15.9949 Da). The peptides were classified by charge state and first filtered by mass accuracy (10 ppm for high-resolution MS) and then dynamically by increasing XCorr and ΔCn values to reduce the protein false discovery rate to <1%. If peptides were shared by multiple members of a protein family, the matched members were clustered into a single group in which each protein identified by a unique peptide represented a subgroup. Each sample was analyzed in biological replicate, and the protein groups, the total peptide counts, and the peptide spectral counts are provided in the of supplemental material (89).

**Gene ontology (GO) enrichment and network.**

Functional enrichment of the modules was determined using the GO-Elite (v1.2.5) package (116). The set of total proteins (1,341) identified was used as the background. We analyzed the 153 proteins enriched in IP (peptide spectrum match [PSM] IP/bead fold change of >2). The Z-score determines the overrepresentation of ontologies in a module, and a permutation P value was used to assess the significance of a Z-score cutoff of 5.5 and a P value cutoff of 0.00001, with a minimum of five proteins per category used as filters in pruning the ontologies. A horizontal bar graph was plotted in R. The networks were constructed using the Circlize package in R.

**Tissue fractionation**

Brain tissue was collected from wild-type C57BL/6 mice and fractionated as first described in Guillem et al. 2005 (117). Nuclear pellets were suspended and sonicated (5 times at 10% output) in immunoprecipitation (IP) buffer (50mMTris-HCl [pH 7.4], 150 mM NaCl, and 0.5% NP-40) supplemented with 1 cOmplete mini protease inhibitor tablet (Roche). Lysates were spun at 13,000 rpm for 10 minutes (min) at 4°C and protein concentrations were determined using a standard bicinchoninic acid (BCA) assay (Pierce).

## 2.4 Results

### 2.4.1 ZC3H14 interacts with a variety of mRNA processing factors

As the function of ZC3H14 is not known, we sought to identify the factors that interact with ZC3H14 in the brain. To identify ZC3H14-interacting factors, we employed

a discovery-based approach using mass spectrometry. Mouse brain tissue was isolated from wild type mice. The tissue was homogenized and prepared for immunoprecipitation experiments as described in the Materials and Methods. Antibodies directed against the N-terminus of ZC3H14 were used to immunoprecipitate ZC3H14 isoforms 1-3. Figure 2.1A shows an immunoblot from this ZC3H14 immunoprecipitation. ZC3H14 is enriched in the ZC3H14 Bound fraction and completely absent from the control IgG Bound lane compared to the Input lane. ZC3H14 is detected as two bands with the ~100 kDa band corresponding to isoform 1 and the ~70 kDa band corresponding to a doublet band of isoforms 2 and 3. Our results confirmed that we can efficiently purify ZC3H14 protein in these immunoprecipitation experiments.

As we could immunoprecipitate ZC3H14 from mouse brain lysate, we next used unbiased mass spectrometry to identify ZC3H14-interacting proteins from this ZC3H14 immunoprecipitation. We compared results obtained from the ZC3H14 immunoprecipitation to results obtained with an IgG control antibody. These interacting proteins are presented in detail in Figures 2.1B and 2.1C (89). To explore the relationship between ZC3H14 and the interacting proteins identified, we used Gene Ontology (GO) analysis. Figure 1B shows the GO analysis of the "top" interacting candidates identified by the mass spectrometry analysis. These "top" candidates were selected based on the criteria that they were only identified in the ZC3H14 immunoprecipitation and not with the IgG control and they showed $\geq 2$ spectral counts abundance. Of the "top" proteins, 51 proteins grouped based on the three GO domains: Biological Process, Cellular Component, and Molecular Function. A Z-score cutoff of $> 4.5$, a p-value cutoff of $< 0.05$, and a 3 gene per category minimum were used as filters to refine the ontologies.

Figure 2.1B provides a circle graph indicating the specific proteins identified in each of the GO categories. The colors correspond to the GO categories, while the individual Gene terms identified are linked to their GO categories. Note that all the proteins in the circle graph are RNA processing factors and all the GO terms are related to RNA processing such as RNA splicing and mRNA export. In fact, cellular complexes such as the spliceosomal complex are enriched. These data show that ZC3H14 interacts with a variety of RNA processing factors and potentially has a role in regulating the processing of RNA in the brain.

### 2.4.2 ZC3H14 interacts with proteins in both an RNA-independent and RNA-dependent manner

As ZC3H14 is an RNA-binding protein, many of the proteins associated with ZC3H14 could interact in an RNA-dependent manner by associating with the same transcript. Figure 2.2A depicts models of how ZC3H14 could associate with proteins in the cell. In the model shown at the top, ZC3H14 and an associated protein could interact because they associate with the same transcript. In the model shown in the middle, ZC3H14 directly binds to the associated protein. In the model shown at the bottom, the interaction between ZC3H14 and another protein could be a combination of the first two models where ZC3H14 and the associated protein both bind to the RNA and to one another. These models highlight the complexity of protein interactions mediated by RNA-binding proteins.

To distinguish whether interactions between ZC3H14 and associated proteins are dependent on RNA, we immunoprecipitated ZC3H14 from control and RNase-treated lysate. Figure 2.2B shows an immunoblot of the ZC3H14 immunoprecipitation, where ZC3H14 protein is enriched in the Bound fraction and depleted in the Unbound fraction

compared to the Input fraction. The addition of RNase does not affect the immunoprecipitation of ZC3H14 as ZC3H14 protein is enriched in the bound lane of the RNase treated lysate. Importantly, ZC3H14 protein is absent from the IgG control Bound lane. To test the effect of RNase treatment, we probed for a candidate interacting protein PABPN1. PABPN1 is detected in the bound lane of ZC3H14 but partially depleted in the ZC3H14 RNase treated samples. These results show that the use of RNase distinguishes between interactions that are RNA-dependent and interactions that are RNA-independent.

As the use of RNase did not affect the precipitation of ZC3H14 protein but did alter the interaction with PABPN1, we examined ZC3H14-interacting proteins in the presence and absence of RNase treatment. We performed immunoprecipitation of ZC3H14 as described before from control and RNase treated lysate and analyzed interacting proteins with unbiased mass spectrometry. We classified proteins as RNA-independent if there was no more than a 2-spectral count difference between the control and RNase-treated sample. Partially RNA-dependent proteins had more than a 50% decrease in spectral counts between the control and RNase-treated sample. Proteins with a reduction to one or zero spectral counts upon RNase treatment were classified as RNA-dependent. In Figure 2.2C, all the ZC3H14 interacting proteins that are RNA-independent, partially RNA-dependent, and RNA-dependent are grouped in boxes. We can distinguish from the total pool of proteins that interact with ZC3H14 which protein interactions are mediated through RNA. Taken together, these results show that ZC3H14 interacts with a variety of RNA processing factors and some of these interactions are mediated through RNA.

### 2.4.3 Isoform of ZC3H14 has an isoform-specific phosphorylation site

As ZC3H14 is alternatively splice to generate multiple protein isoforms in mammals, the isoforms are likely to have specific functions. Isoform 4 of ZC3H14 has tissue specific expression in the testis, while the larger isoforms are ubiquitously expressed. Isoform 1 is the largest of the four isoforms and Isoforms 2 and 3 only differ from each other by a single small exon (see Figure 1.2). To examine isoform-specific differences in these proteins, we employed a two-pronged approach. First, as the mostly nuclear isoforms (1-3) of ZC3H14 can be detected in neuronal processes (97), we examined whether the nuclear isoforms of ZC3H14 show different nuclear and cytoplasmic distributions. Secondly, the differences in exon inclusion of ZC3H14 protein isoforms present the possibility that the protein isoforms could have isoform-specific posttranslational modifications (PTMs). Understanding the biological function of each isoform could offer insight into the multiple function of ZC3H14.

To examine the nucleocytoplasmic distribution of ZC3H14 isoforms, we isolated mouse brain tissue and fractionated the tissue into nuclear and cytoplasmic fractions. We analyzed these fractions using immunoblotting. We probed blots for ZC3H14 and nuclear and cytoplasmic markers, Histone and PABPC1, respectively. Figure 2.3A shows the resulting immunoblot with Nuclear, and Cytoplasmic (Cyto) fractions. The polyclonal antibody raised against the N-terminal domain of ZC3H14 detects multiple ZC3H14 isoforms generated by alternative splicing, but does not detect the predicted cytoplasmic Isoform 4, which lacks this N-terminal domain (82). The nuclear fraction contains the bulk of ZC3H14 protein. The cytoplasmic (Cyto) fraction contains less ZC3H14 than the nuclear fraction. The cytoplasmic fraction shows some enrichment for the ~70 kDa isoform(s) of ZC3H14 as compared to the ~100 kDa isoform. This result suggests that the ~70 kDa

isoform(s) of ZC3H14 is/are the dominant isoform in the cytoplasm and presents the possibility that there could be different functionalities for the individual ZC3H14 isoforms.

We next focused on identifying PTMs on ZC3H14 taking into consideration the different ZC3H14 protein isoforms could have distinct PTMs. We focused on phosphorylation because global proteomic studies identified phosphorylated peptides of ZC3H14 (118,119). As described previously, we isolated mouse brain tissue and fractionated into nuclear and cytoplasmic fractions. We immunoprecipitated ZC3H14 from each fraction and analyzed the bound fraction by mass spectrometry. The cytoplasmic samples were used at 4 times the concentration of the starting nuclear lysate to account for the low cytoplasmic level of ZC3H14. Figure 2.3B shows the resulting immunoblot of the nuclear and cytoplasmic immunoprecipitation of ZC3H14. ZC3H14 is enriched in the Bound lanes and depleted in the Unbound lanes compared to the Input lanes of the nuclear and cytoplasmic immunoprecipitated fractions. The IgG Bound control lanes are free of ZC3H14 protein. Through our analysis, we identified a phosphorylated peptide specific to ZC3H14 isoform 1 (Figure 2.3C). This phosphorylation site occurs at amino acid positon 516 in encoded by exon 12 of the ZC3H14 variable splicing region. This exon is not present in isoforms 2 and 3. Identification of this isoform-specific phosphorylation site suggests a mechanism that could be employed to differentially regulate distinct ZC3H14 protein isoforms.

**2.4.4 ZC3H14 forms non-canonical mRNPs with metabolic enzymes.**

The most enriched protein associated with ZC3H14 with the number of spectral counts almost equal to ZC3H14 is a metabolic protein, AMPD2 (Figure 2.4A). This result was surprising because AMPD2 is a cytoplasmic protein (120) while ZC3H14 is primarily

nuclear. To validate the interaction between ZC3H14 and AMPD2, we performed ZC3H14 immunoprecipitations from fractionated mouse brain tissue. Figure 2.4B shows the immunoprecipitation of ZC3H14 from the fractionated mouse brain lysate. ZC3H14 is enriched in the Bound lanes of the nuclear (ZC3H14-Nuc) and cytoplasmic (ZC3H14-Cyto) fractions and depleted in the Unbound lanes compared to the Input lanes. ZC3H14 protein is completely absent from the IgG control Bound lane. After confirming immunoprecipitation of ZC3H14 protein, we next examined whether AMPD2 protein co-purifies with ZC3H14. We found that AMPD2 is enriched in the bound lanes for the nuclear and cytoplasmic immunoprecipitations of ZC3H14 (Figure 4C). These results show that ZC3H14 and AMPD2 interact in both the nucleus and cytoplasm.

**2.5 Discussion**

Previous large-scale studies have been used to establish protein interacting networks for RNA-binding proteins (121,122). Here, we report our results identifying the proteins that interact with ZC3H14. Our data reveal functional links between ZC3H14 and a variety of RNA processing factors. We extended our analysis to define RNA-independent and RNA-dependent interactions. In addition, we assessed isoform-specific differences in the ZC3H14 proteins by examining subcellular distribution of the ZC3H14 isoforms and isoform-specific PTMs. Furthermore, we established that ZC3H14 forms non-canonical mRNPs as evidenced by the interaction between ZC3H14 and the adenine deaminase, AMPD2.

While ZC3H14 interacts with a variety of mRNA processing factors, how ZC3H14 cooperates with these factors to ensure proper mRNA processing remains to be determined.

One possibility is that the different isoforms of ZC3H14 protein are involved in distinct steps of mRNA processing allowing ZC3H14 to fulfil different roles in the cell. Our results support this model as Isoform 1 of ZC3H14 is primarily nuclear and has an isoform specific phosphorylation site, while Isoform 2 and/or Isoform 3 is/are the dominant isoform(s) in the cytoplasm. Differentially utilizing the different isoforms of ZC3H14 expands the function of the ZC3H14 protein. Each ZC3H14 isoform could interact with specific factors and participate in distinct mRNA processing events.

As ZC3H14 and AMPD2 interact, AMPD2 could have a role in mRNA processing. Metabolic enzymes can moonlight as RNA-binding proteins and/or be incorporated into mRNPs (123). These enzymes adopt a role in the regulation of target transcripts, while maintaining their primary function of participating in a metabolic pathway (123). The known function of AMPD2 is to maintain the cellular pool of guanine nucleotide by regulating the feedback inhibition of adenosine derivatives on de novo purine synthesis. ZC3H14 could regulate the function of AMPD2 in the brain by targeting AMPD2 to sites of mRNA processing or degradation (120,124). As adenosine deaminase deficiencies cause behavioral and neurological impairments (125), loss of ZC3H14 could result in misregulation of the AMPD2 protein, which could underlie the phenotypes observed in the *ZC3H14* mutant patients.

Based on our results, we suggest that ZC3H14 is involved in a variety of RNA processing pathways in the brain. Posttranscriptional regulation is particularly critical in neurons, which rely on sophisticated spatial and temporal control of gene expression (126,127). We present evidence that ZC3H14 is physically linked to multiple RNA processing factors in the brain suggesting that ZC3H14 could coordinate the interplay of

multiple RNA processing factors. Fully understanding the molecular function of ZC3H14 would serve as a cornerstone for elucidating the complex roles of RNA-binding proteins in regulating neuronal gene expression. Ultimately, the information gained from this study could provide insight into how ZC3H14 and other RNA-binding proteins affect proper higher order brain function.

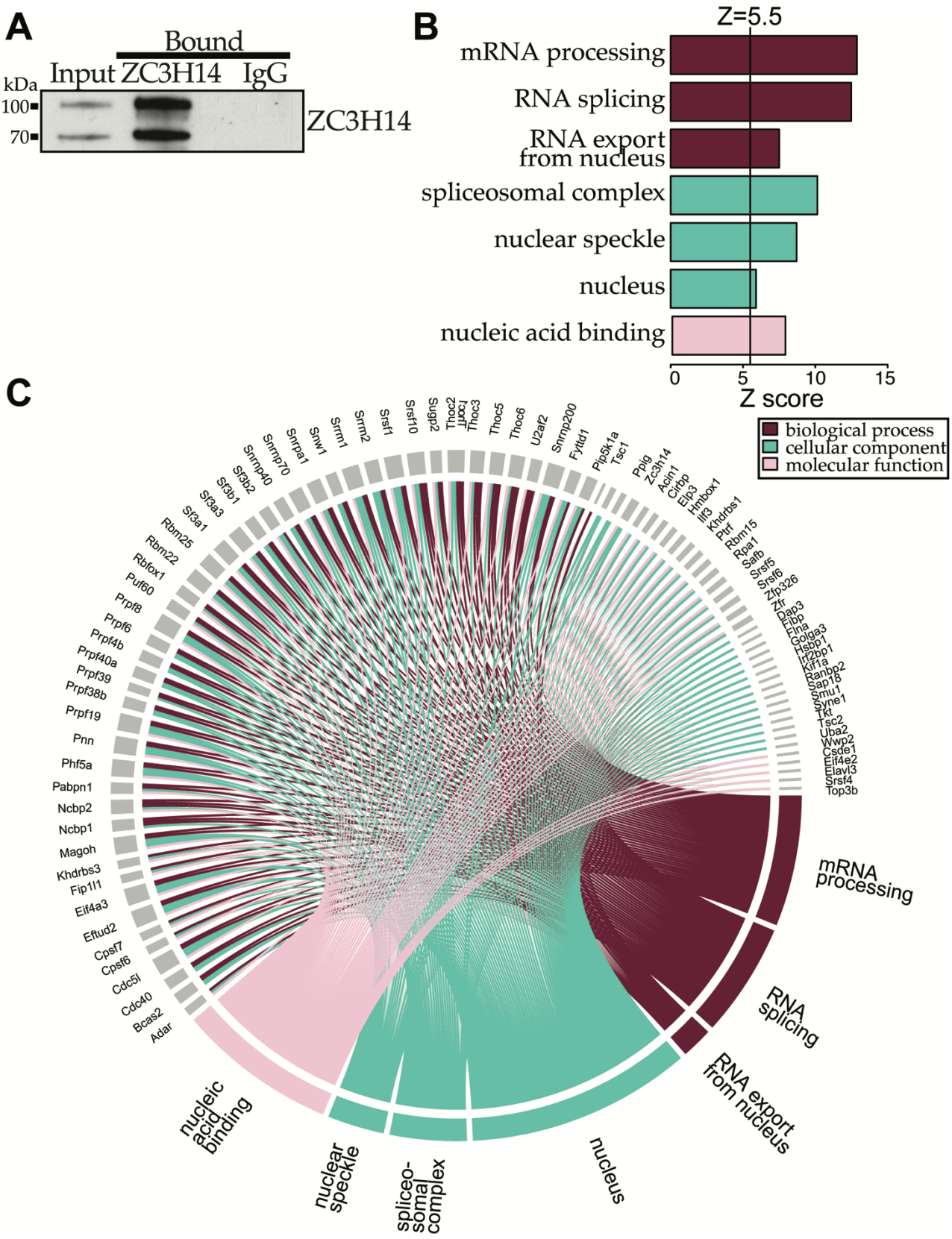
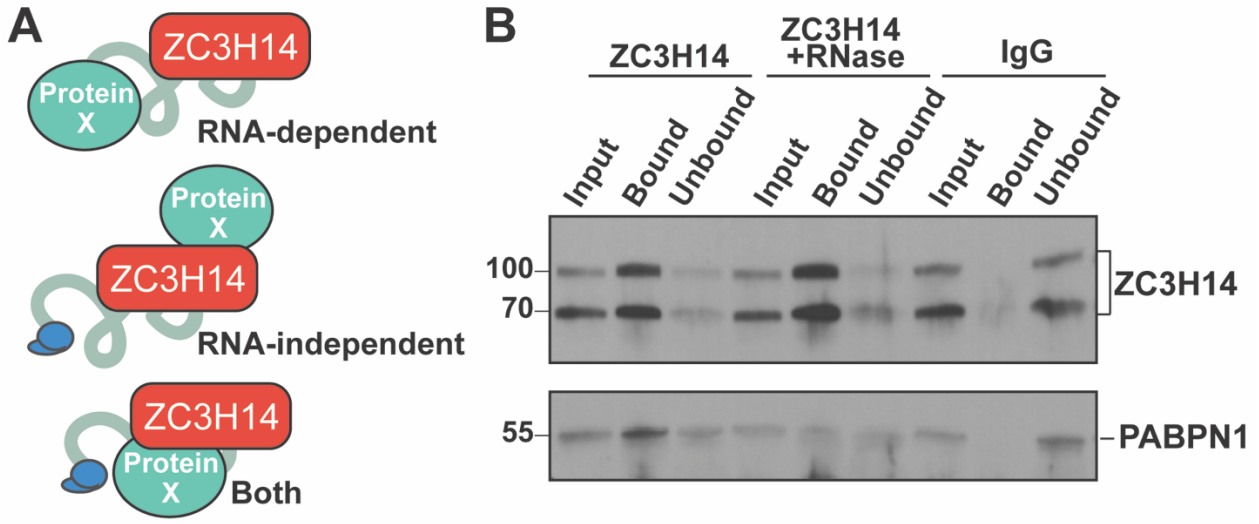

A
Bound
Input ZC3H14 IgG
kDa
100
70
ZC3H14
B
Z=5.5
mRNA processing
RNA splicing
RNA export from nucleus
spliceosomal complex
nuclear speckle
nucleus
nucleic acid binding
0 5 10 15
Z score
C
biological process
cellular component
molecular function
mRNA processing
RNA splicing
RNA export from nucleus
nucleus
spliceo-somal complex
nuclear speckle
nucleic acid binding

**Figure 2.1: ZC3H14 interacts with RNA processing factors**
**A.** Immunoblot of immunoprecipitation of ZC3H14 from mouse brain lysate. ZC3H14 is enriched in the ZC3H14 Bound lane compared to the IgG control. The ZC3H14 is detected as a two-band pattern with the top bane corresponding to the ~100 kDa of ZC3H14 and the ~70 kDa band corresponding to a doublet of the smaller isoforms of ZC3H14 **B.** Gene Ontology analysis of the "top" ZC3H14-interacting proteins. Gene terms are grouped by: biological process, cellular component, and molecular function. **C.** Graph of the gene terms that cluster into the Gene Ontology categories in B.




C

**ZC3H14-interacting Proteins**

**RNA-independent**

| | | | |
|---|---|---|---|
| Ak5 | Hnrnph3 | | Son |
| Ampd3 | Hnrnpr | Grcc10 | Srr |
| Arhgef10 | Hsbp1 | H2afy | Srrm1 |
| Asap1 | Iars | Ppig | Srsf2 |
| Bckdha | Iqsec3 | Ppm1e | Srsf5 |
| Camk2n1 | Irf2bp1 | Prpf40a | Strn3 |
| Ccdc88a | Itch | Prpf6 | Tbc1d15 |
| Cdc5l | Itih1 | Ptprs | Tcp1 |
| Cdk5 | Khdrbs1 | Ptprz1 | Thbs1 |
| Celf2 | Khdrbs3 | Ptrf | Thoc1 |
| Chtop | Kif1a | Rab5a | Thoc3 |
| Cpsf7 | Krt76 | Ranbp2 | Thoc5 |
| Cspg5 | Map1s | Rap2b | Thoc6 |
| Cul3 | Mprip | Rbm15 | Tln1 |
| Cyth3 | Myl4 | Rnps1 | Tsc1 |
| Dbt | Ncoa5 | Rogdi | Tsc2 |
| Ddx46 | Nrxn1 | Rpl36a | U2af2 |
| Dpysl5 | Pabpn1 | Rpl38 | Uba2 |
| Eif3i | Pip4k2a | Sae1 | Vgf |
| Fam171a2 | Pip5k1a | Sap18 | Vps35 |
| Fam213a | Plg | Scai | Vtn |
| Fgg | Plxna4 | Sdc4 | Wwp2 |
| Fyttd1 | Pnn | Sdpr | Zc3h14 |
| Gm5414 | Pnpla6 | Snrnp200 | |

**Partially RNA-dependent**

| | | |
|---|---|---|
| Acan | Gpc1 | Rgs7 |
| Acin1 | Hadha | Sf3a3 |
| Ampd2 | Hapln4 | Sf3b1 |
| Atp5f1 | Isca2 | Sfxn1 |
| Bckdhb | Kcnab2 | Slc9a3r1 |
| Cat | Ldhc | Snrpe |
| Cdk18 | Lmo7 | Snw1 |
| Cep170 | Lrrc47 | Srrm2 |
| Cops7a | Marcks | Srsf1 |
| Cpeb3 | Mfap1a | Srsf10 |
| Cyld | Pfkl | Tardbp |
| Dhx15 | Pgam5 | Thoc2 |
| Eif3g | Pkp1 | Tra2b |
| Eif4a3 | Prpf8 | U2surp |
| Erh | Raly | Zfp326 |
| Fam184a | Rbm22 | Zfr |
| Gm8994 | Rbm25 | |

**RNA-dependent**

| | | | |
|---|---|---|---|
| Acadvl | Dhx36 | Nptn | Sf3a1 |
| Acsl6 | Efr3b | Numbl | Sf3b2 |
| Akap8l | Eif4e2 | Pcbp1 | Slc6a17 |
| Anxa11 | Elavl3 | Pde4dip | Smap2 |
| Ap2s1 | Etfb | Pla2g4d | Smn1 |
| Apoo | Exoc4 | Pon1 | Smu1 |
| Arhgef7 | Fam195b | Ppib | Snrpa1 |
| Arvcf | Gigyf2 | Ppp2r2c | Srsf4 |
| Atl1 | Gm7698 | Ppp6c | Ssb |
| Brk1 | Golga2 | Prpf38a | Sugp2 |
| Cadm2 | Gpd1 | Prpf39 | Thoc7 |
| Ccdc175 | Gprin1 | Psma5 | Tmem65 |
| Ccdc9 | Grm2 | Puf60 | Tmx2 |
| Cct6b | Hmbox1 | Rab8b | Top3b |
| Cdc40 | Ilf3 | Rap1gds1 | Tpr |
| Cenpv | Kcnj10 | Rap2a | Tubgcp4 |
| Cirbp | Khsrp | Rap2c | Ube2d2a |
| Copa | Magoh | Rasal1 | Upf2 |
| Cox5a | Marco | Rasgrf2 | Upf3b |
| Csnk2a1 | Mars | Rpl10a | Vps26b |
| Ddx20 | Mog | Rpl22l1 | Wasf3 |
| Dfna5 | Mrps21 | Rpl35a | Zc3h11a |
| Dhx16 | Ncbp1 | | Zranb2 |

**Figure 2.2: Protein Interactions of ZC3H14 mediated through RNA**
**A.** Theoretical models of how RNA-binding proteins can interact with other proteins in RNA dependent manner, RNA-independent manner, and a combination of both. **B.** Immunoblot of ZC3H14 immunoprecipitation from control and RNase-treated lysate. ZC3H14 protein is enriched in the Bound lanes and depleted in the Unbound lanes compared to the Input lanes of the ZC3H14 and ZC3H14 + RNase fractions. IgG serves as the negative control and ZC3H14 protein is absent from the IgG Bound lane. **C.** Box grouping of ZC3H14 interaction proteins that are RNA independent, Partially RNA-dependent, and RNA-dependent.

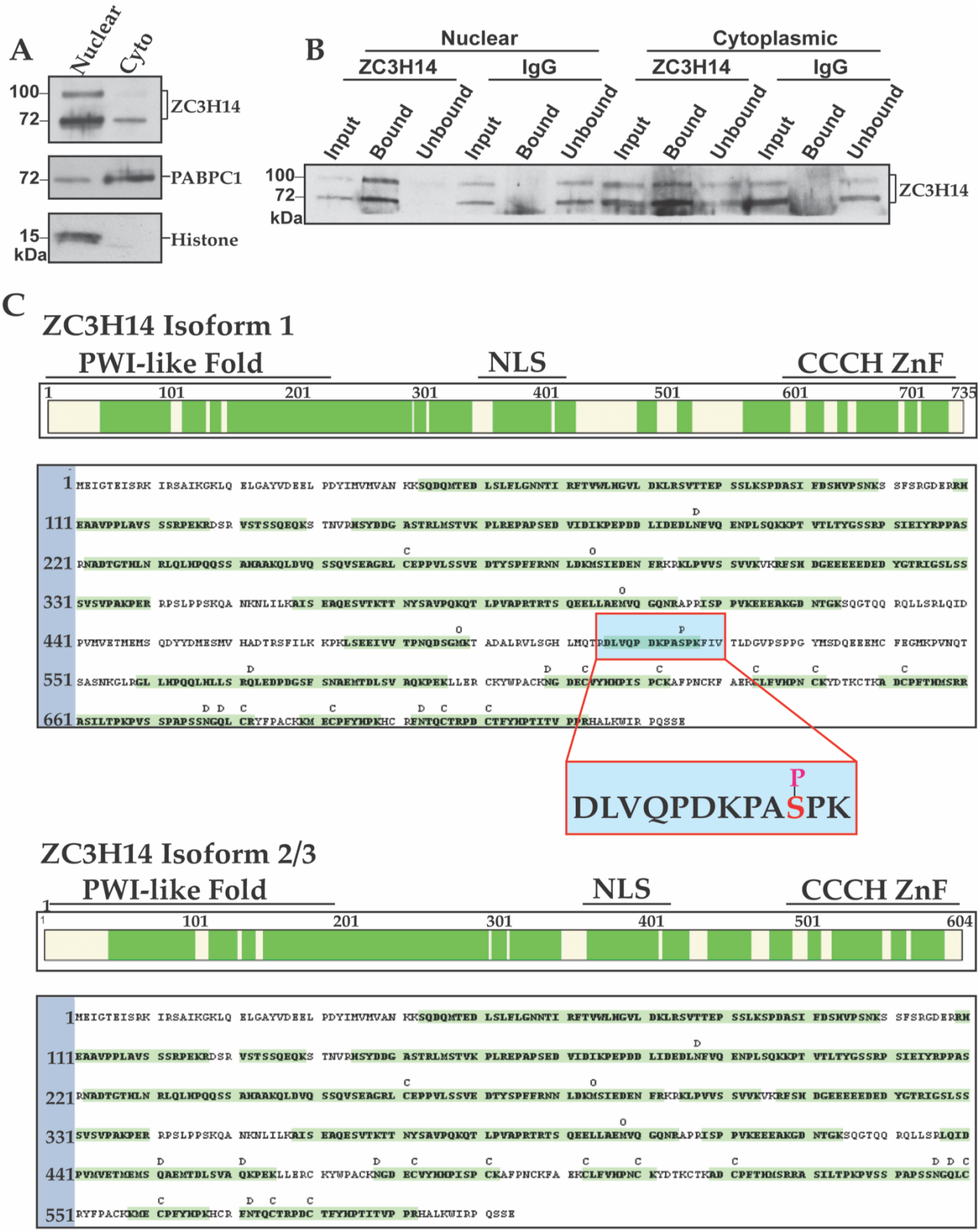

A
Nuclear
Cyto
100
72
ZC3H14
72
PABPC1
15
kDa
Histone
B
Nuclear
Cytoplasmic
ZC3H14
IgG
ZC3H14
IgG
Input
Bound
Unbound
100
72
kDa
ZC3H14
C
ZC3H14 Isoform 1
PWI-like Fold
NLS
CCCH ZnF
1
101
201
301
401
501
601
701
735
DLVQPDKPASPK
P
ZC3H14 Isoform 2/3
PWI-like Fold
NLS
CCCH ZnF
1
101
201
301
401
501
604

**Figure 2.3: ZC3H14 isoform specific phosphorylation site**
**A.** Fractionated mouse brain lysate into Nuclear and Cytoplasmic fractions. PABPC1 and Histone serve as the cytoplasmic and nuclear markers respectively. **B.** Immunoblot of ZC3H14 immunoprecipitation from Nuclear and Cytoplasmic brain lysate. Molecular weights are given in kDa to the left of each blot **C.** Identified ZC3H14 peptides from mass spectrometry analysis of ZC3H14 protein isoforms. Identified peptides are highlighted in green. Numbers correspond to sequence number of amino acids. Protein domains of ZC3H14 are indicated at the top of the protein sequence. Highlighted box indicates a phosphorylated serine at position 516.

A

| Gene Name | ZC3H14 Bound | IgG Bound |
|---|---|---|
| ZC3H14 | 18 | 0 |
| AMPD2 | 25 | 0 |

*** Values are given in Spectral Counts

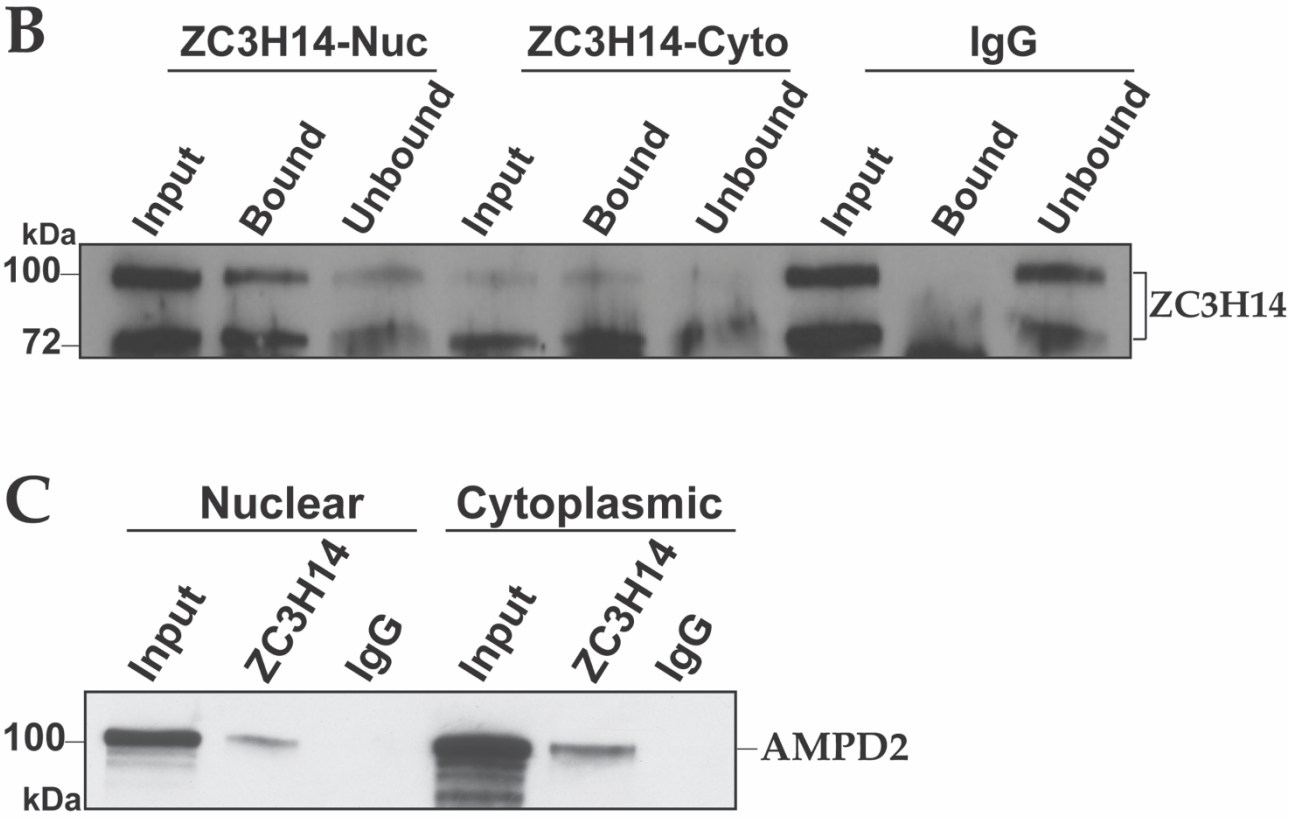



**Figure 2.4: ZC3H14 interacts with AMPD2**
**A**. Number of Spectral Counts for ZC3H14 and AMPD2 from mass spectrometry analysis of ZC3H14 immunoprecipitation. Numbers are an average from 3 experiments. **B**. Immunoblot of ZC3H14 immunoprecipitation from nuclear (ZC3H14-Nuc) and cytoplasmic (ZC3H14-Cyto) brain lysate. **C.** Immunoblot of AMPD2 from ZC3H14 immunoprecipitation experiments from Nuclear and Cytoplasmic mouse brain lysate. Molecular weights are given as kDa to the left of the blot.

# Chapter 3: The Polyadenosine RNA-Binding Protein, ZC3H14 Interacts with the THO Complex and Coordinately Regulates the Processing of Neuronal Transcripts

Kevin J. Morris

## 3.1 Summary:

The polyadenosine RNA-binding protein ZC3H14 is important for multiple steps in RNA processing. Although ZC3H14 is ubiquitously expressed, mutation of the *ZC3H14* gene causes a non-syndromic form of intellectual disability. Here, we examine the function of ZC3H14 in the brain by identifying ZC3H14-interacting proteins using unbiased mass spectrometry. Through this analysis, we identified physical interactions between ZC3H14 and multiple RNA processing factors. Notably, proteins that comprise the THO complex were amongst the most enriched proteins. We demonstrate that ZC3H14 physically interacts with THO components and that these proteins are required for proper RNA processing, as loss of ZC3H14 or THO components leads to extended bulk poly(A) tail length. Furthermore, we identified the transcripts *Atp5g1* and *Psd95* as shared RNA targets of ZC3H14 and the THO complex. Our data suggest that ZC3H14 and the THO complex are important for proper processing of *Atp5g1* and *Psd95* RNA, as depletion of ZC3H14 or THO components leads to decreased steady-state levels of the mature transcript and accumulation of *Atp5g1* and *Psd95* pre-mRNA in the cytoplasm. Taken together, this work provides the first unbiased identification of nuclear ZC3H14-interacting proteins from the

brain and further establishes a role for ZC3H14 and the THO complex in the processing of RNA.

**3.2 Introduction:**

Control of eukaryotic gene expression involves a highly ordered regulatory network of cis-acting elements and trans-acting factors to ensure that a given gene product is produced in the correct amount at the proper time and location. RNA-binding proteins are critical regulators in the co- and posttranscriptional events that control gene expression (128). These events include processing steps such as the initial capping of the nascent mRNA emerging from RNA polymerase, the splicing of the primary transcript, and the formation of the 3'end of the mRNA through cleavage and polyadenylation (129,130). Upon completion of these steps, a mature mRNA is generated that is packaged with RNA-binding proteins to form a messenger ribonucleoprotein (mRNP) complex (131). This mRNP can then be exported from the nucleus and undergo further regulation in the cytoplasm.

The importance of RNA-binding proteins is evident in both their critical role in gene expression and the growing number of diseases that are associated with the mutation of genes encoding RNA-binding proteins (112,132). Interestingly, an emerging subset of these diseases are tissue-specific diseases that result from mutation in genes encoding ubiquitously expressed proteins (132). Recently, mutations in the gene encoding the ubiquitously expressed RNA-binding protein ZC3H14 were linked to an inherited form of autosomal recessive intellectual disability (80). This finding suggests that ZC3H14 plays a critical role in the brain. However, the function of ZC3H14 has been largely studied in cultured cell lines and lower order model organisms (80,89,96).

Initial understanding of the function of ZC3H14 came from studies of the budding yeast orthologue Nab2 (88). ZC3H14 and Nab2 share conserved domain structures that

consist of an N-terminal PWI-like domain and C-terminal tandem CCCH zinc fingers (84,88). This work showed that *NAB2* is an essential yeast gene that is required for proper poly(A) tail length control and mRNA export from the nucleus (85-87). Like Nab2, ZC3H14 is primarily localized to the nucleus and both Nab2 and ZC3H14 bind with high affinity to polyadenosine RNA (82,93). Although Nab2 and ZC3H14 bind to polyadenosine RNA tracts and therefore could bind to all mRNA transcripts, previous work has shown that loss of ZC3H14 only alters the steady-state levels of a small number of transcripts (96). This finding focuses the regulation of mRNA processing by ZC3H14 to a select group of transcripts.

As patients with mutations in the *ZC3H14* gene that cause loss of the ZC3H14 protein display brain phenotypes, the role of ZC3H14 in the brain has begun to be studied (80). Studies of the ZC3H14 orthologue in *Drosophila,* dNab2, provide evidence that the function of dNab2 is critical in neurons (80). *dNab2* mutant flies have compromised brain function and severe brain morphology defects (91). Importantly, these defects in the fly can be rescued by the expression of either dNab2 or the mammalian ZC3H14 protein exclusively in neurons (91). Furthermore, the generation of a *Zc3h14* mutant mouse model revealed that loss of ZC3H14 impairs higher order brain function with the mutant mice displaying defects in working memory (98). These findings coupled with the phenotype of the *ZC3H14* patients further emphasize the continued need to understand the function of ZC3H14 in the brain.

In this study, we identify proteins that interact with ZC3H14 in the brain. We employ a proteomic approach using fractionated mouse brain lysate to identify the nuclear factors that interact with ZC3H14. Among the most enriched factors is a group of proteins

that belong to an RNA-processing complex known as the THO complex. The THO complex is a multi-subunit complex that is essential for RNA processing (41). Most studies of this complex have been performed in budding yeast and these studies have linked this complex to transcription elongation, mRNA splicing, and mRNA export (41,66,133,134). Interestingly, mutations in genes encoding components of this complex *THOC2* and *THOC6* are associated with brain disorders, similar to mutations in *ZC3H14 (135-137)*.

Here, we demonstrate that ZC3H14 and the THO complex physically interact and coordinately regulate the processing of mRNA transcripts. As previously shown for ZC3H14, we demonstrate that THO components are required for proper control of bulk poly(A) tail length. As these factors do not affect all RNAs (96,138,139), we have identified the transcript *Atp5g1*, which encodes a component of the ATP synthase machinery (140), and the transcript *Psd95 (Dlg4),* which encodes a component of the neuronal post synaptic density (141), as shared targets of ZC3H14 and the THO complex. This work reveals that ZC3H14 and the THO complex are required to coordinate nuclear processing of *Atp5g1* and *Psd95* mRNAs with export from the nucleus. Loss of these factors leads to a decrease in the steady-state levels of the mature *Atp5g1* and *Psd95* mRNAs and accumulation of the pre-mRNAs of these transcripts in the cytoplasm. Taken together, these results suggest coordination between ZC3H14 and the THO complex to ensure proper mRNA processing prior to export from the nucleus.

## 3.3 Material and Methods

### Tissue Fractionation and Immunoprecipitations

Brain tissue was collected from wild-type C57BL/6 mice and fractionated as first described in Guillem et al. 2005 (117). Nuclear pellets were suspended and sonicated (5 times at 10% output) in immunoprecipitation (IP) buffer (50mMTris-HCl [pH 7.4], 150 mM NaCl, and 0.5% NP-40) supplemented with 1 cOmplete mini protease inhibitor tablet (Roche). Lysates were spun at 13,000 rpm for 10 minutes (min) at 4°C and protein concentrations were determined using a standard bicinchoninic acid (BCA) assay (Pierce). Protein A-magnetic beads (Dynabeads; Invitrogen) were suspended in IP buffer and incubated for 30 min with preimmune rabbit serum or an equal volume of experimental antibody (4 μg/mg of protein lysate) at room temperature. Bead/antibody complexes were added to clarified cell lysates, followed by incubation at 4°C overnight while tumbling end over end (10% removed prior to overnight incubation for input samples). After incubation, the beads were magnetized and washed five times with ice-cold IP buffer. At this point, samples were prepared for mass spectrometry analysis or for immunoblotting. For immunoblotting, protein complexes were eluted with reducing sample buffer (250 mM Tris-HCl, 500 mM dithiothreitol [DTT], 10% SDS, 0.5% bromophenol blue, and 50% glycerol).

**Mass Spectrometry Analysis**

Mass spectrometry analysis was performed as previously described in Soucek et al. 2016 (89). Magnetic beads from control and ZC3H14 immunoprecipitates were suspended in 8 M urea–100 mM NaHPO4 (pH 8.5; 50 μl, final volume) and treated with 1 mM DTT at 25°C for 30 min, followed by 5 mM iodoacetamide at 25°C for 30 min in the dark. The samples were then diluted to 1M urea with 50 mM ammonium bicarbonate (final volume, 400 μl) and digested with lysyl endopeptidase (Wako; 1.25 ng/l, final concentration) at 25°C for 4 hours and further digested overnight with trypsin (Promega; 1.25 ng/l, final

concentration) at 25°C. The resulting peptides were desalted with a Sep-Pak C18 column (Waters) and dried under vacuum. Each sample was suspended in loading buffer (0.1% formic acid, 0.03% trifluoroacetic acid, 1% acetonitrile) and analyzed independently by reverse-phase liquid chromatography (LC) coupled with tandem mass spectrometry (MS/MS) essentially as previously described with some minor modifications (115,142). Briefly, peptide mixtures were loaded onto a C18 nanoLC column (75 m [inner diameter], 15-cm long, 1.9-m resin [Maisch GmbH]) and eluted over a 5 to 0% gradient (buffer A: 0.1% formic acid; buffer B: 0.1% formic acid in 100% acetonitrile). Eluates were monitored in an MS survey scan, followed by 10 data-dependent MS/MS scans on a Q-Exactive plus Orbitrap mass spectrometer (Thermo Scientific). The acquired MS/MS spectra were searched against a concatenated target decoy mouse reference database (v.62) of the National Center for Biotechnology Information (downloaded 14 November 2013 with 30,267 target entries) using the SEQUEST Sorcerer algorithm (version 4.3.0; SAGE-N). The search parameters included the following: fully tryptic restriction, parent ion mass tolerance of 50 ppm, up to two missed trypsin cleavages, and dynamic modifications for oxidized Met (15.9949 Da). The peptides were classified by charge state and first filtered by mass accuracy (10 ppm for high-resolution MS) and then dynamically by increasing XCorr and Cn values to reduce the protein false discovery rate to 1%. If peptides were shared by multiple members of a protein family, the matched members were clustered into a single group in which each protein identified by a unique peptide represented a subgroup. The total peptide spectra match (PSM) counts are provided in Supplementary Table 1 in the Supplementary material.

**Gene Ontology (GO) Enrichment and Network**

Functional enrichment of the modules was determined using the GO-Elite (v1.2.5) package (116). The set of total proteins (1,350) identified was used as the background. We analyzed the 62 proteins enriched in IP ([PSM] IP/bead fold change of 2). The Z-score determines the overrepresentation of ontologies in a module, and a permutation p-value was used to assess the significance of a Z-score cutoff of 4.5 and a p-value cutoff of 0.05, with a minimum of three proteins per category used as filters in pruning the ontologies. A horizontal bar graph was plotted in R. The networks were constructed using the Circlize package in R.

**Immunoblotting**

Protein lysates were boiled in reducing sample buffer and resolved on 4-20% Criterion TGX polyacrylamide gels (Bio Rad). Proteins were transferred to nitrocellulose membranes and incubated for at least one hour in blocking buffer (5% non-fat dry milk in 0.1% TBS-Tween). This step was followed by an overnight incubation at 4°C in primary antibody diluted in blocking buffer. Primary antibodies were detected using species-specific horse radish peroxidase (HRP) conjugated secondary antibodies (Jackson ImmunoResearch) followed by incubation with enhanced chemiluminescence substrate (ECL, Sigma). Chemiluminescence was detected by exposing blots to autoradiography film (Daigger). Immunoblots were quantified using ImageJ software. Primary antibodies and dilutions employed as they appear: ZC3H14, (1:6000) (82), Nuclear Pore complex NUP93 and NUP62, (1:5000), (414, Abcam, ab24609), eIF5, (1:5000) (Santa Cruz, sc-282), THOC1, (1:1000), (Bethyl, A302-839A), THOC2, (1:2000), (Bethyl, A303-630A), ALYREF, (1:1000), (ALY, Santa Cruz, sc-32311), HuR, (1:1000), (Santa Cruz, sc-5261), NPM1, (1:1000), (B23, Santa Cruz, sc-6013), THOC5, (1:1000) (Bethyl, A302-119A),

HSP90, (1:1000), (Santa Cruz, sc-13119), and NXF1, (1:1000), (TAP, Santa Cruz, sc-32319).

**Glycerol Density Gradients**

Nuclei from mouse brain tissue were collected and lysed as described for tissue fractionation and immunoprecipitation. Nuclear lysate was nuclease treated using Benzonase (Sigma). Approximately 1.5 mg of nuclear lysate in a 300-µl volume was layered on top of a 10-50% glycerol gradient (10 mM HEPES pH 7.5, 2 mM $MgCl_2$, 10 mM KCl, 0.5 mM EDTA and 150 mM NaCl) in 14x89 mm tubes (Beckman Coulter), and spun in a SW41Ti rotor (Beckman Coulter) at 30,000 x RPM for 16 hours at 4ºC. Gradients were fractionated top to bottom into 500-µl fractions using the Gradient Master automated fraction collector (BioComp).

**Cell Culture and siRNA Transfections**

Neuro2a (N2a) cells are a mouse neuroblastoma cell line (143). Cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS and antibiotics and grown under standard environmental conditions. Individual siRNAs were transfected into cultured cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol in antibiotic free media. In experiments where cells were treated with two different siRNAs, cells were transfected simultaneously with double the amount of siRNA as the single transfections. siRNAs: Scrambled (Scr) control (Stealth siRNA Neg Control Medium GC, Invitrogen), *Zc3h14* (Stealth siRNA, Invitrogen, GAAGAGCCTCGATACTGACTCCAAA), *Thoc1* (MISSION esiRNA, Sigma EMU081331), *Thoc5* (MISSION esiRNA, Sigma EMU003181).

**RNA Isolation and Quantitative RT-PCR.**

Total RNA was isolated from N2a cells using the TRIzol reagent (Invitrogen) according to the manufacturer's protocol. Isolated RNA was treated with Turbo DNase (Invitrogen) to degrade contaminating DNA. For qRT-PCR analyses, cDNA was generated using the MMLV Reverse transcriptase (Invitrogen) from 1 μg of total RNA. Relative mRNA levels were measured by quantitative PCR analysis of duplicate samples of 15 ng cDNA with QuantiTect SYBR Green Master Mix using an Applied Biosystems StepOne Plus real time machine (ABI). Results were analyzed using the ΔΔCT method and normalized to 18S rRNA. Oligonucleotide sequences are provided in Supplementary Table 2.

**Bulk Poly(A) Tail Length Assays**

Bulk poly(A) tails were analyzed as described previously (89,113). Briefly, total RNA was 3'end labeled with [$^{32}$P]-pCp (cytidine 3′,5′ bisphosphate) (Perkin Elmer) using T4 RNA ligase. RNA was then digested with RNase A and T1 and resolved on TBE-Urea (90 mM Tris-borate, 2 mM EDTA, 8 M urea) 7% polyacrylamide gels. Gel images were obtained using a Typhoon phosphorimager and quantified using ImageQuant software.

**RNA Immunoprecipitation**

N2a cells were grown on 100-mm plates and UV crosslinked at 254 nm using a Stratalinker UV 2400 (Stratagene). Cells were then suspended in RNA-IP buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 10% RIPA-2 buffer in DEPC-treated water) supplemented with 1 cOmplete mini protease inhibitor tablet (Roche). Lysates were then sonicated on ice 5 times at 10% output and spun at 13,000 RPM for 10 min at 4°C. Protein concentration was

determined with a standard BCA assay (Pierce) and standard immunoprecipitation protocol followed.

**Cell Fractionation**

N2a cells were collected on ice, spun down, and suspended in ice-cold fractionation buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.5% (v/v) NP-40), supplemented with 1 mini cOmplete protease inhibitor tablet (Roche) for 10 min on ice. Cell lysates were centrifuged at 1,000 x g for 5 min and the supernatant or cytoplasmic fraction was separated from the nuclear pellet. The nuclear pellet was then washed once with fractionation buffer and centrifuged at 1,000 x g for 5 minutes. The resulting pellet was collected as the nuclear fraction. RNA was isolated from each fraction with TRIzol reagent (Invitrogen). Protein samples were suspended with RIPA-2 buffer and prepared for immunoblotting.

**Statistical Analysis**

Comparisons between experimental groups were made using Student's t-test, unless noted otherwise. All data are presented as means and standard error of the mean (SEM) (error bars) for at least three independent experiments. Asterisks (*) indicate statistical significance at *p-value < 0.05.

## 3.4 Results

### 3.4.1 ZC3H14 interacts with the THO complex.

To examine the function of ZC3H14 in the brain, we first analyzed the nucleocytoplasmic distribution of the protein in the brain. We collected and fractionated mouse brain tissue as described in Materials and Methods. Figure 3.1A shows the

distribution of the endogenous ZC3H14 protein. The polyclonal antibody raised against the N-terminal domain of ZC3H14 detects multiple ZC3H14 isoforms generated by alternative splicing, but does not detect a predicted cytoplasmic isoform that lacks the N-terminal domain (82). The nuclear fraction contains the bulk of ZC3H14 protein. The cytoplasmic (Cyto) fraction contains less ZC3H14 than the nuclear fraction with some enrichment for the ~70 kDa isoforms of ZC3H14 as compared to the ~100 kDa isoform.

As the majority of ZC3H14 is nuclear, we sought to identify factors that interact with ZC3H14 in the nucleus. To identify ZC3H14-interacting proteins, we collected mouse brain tissue and isolated the nuclear fraction. We then immunoprecipitated endogenous ZC3H14 and associated proteins from the nuclear fraction using the polyclonal ZC3H14 antibody. With unbiased mass spectrometry, we identified ZC3H14-interacting proteins and compared results from our ZC3H14 immunoprecipitation to the results obtained with control pre-immune serum. These interacting proteins are presented in detail in Supplementary Table 1 as well as Figures 3.1B and 1C.

To explore the relationship between ZC3H14 and the interacting proteins identified, we used Gene Ontology (GO) analysis. Figure 3.1B shows the GO analysis of the 62 "top" interacting candidates identified by the mass spectrometry analysis. These "top" candidates were selected based on the criteria that they were only identified in the ZC3H14 immunoprecipitation and not with the control pre-immune serum and they showed a $\geq 2$ peptide spectra matched (PSM) quantification. Of the 62 proteins, 38 proteins grouped based on the three GO domains: Biological Process, Cellular Component, and Molecular Function. A Z-score cutoff of $> 4.5$, a p-value cutoff of $< 0.05$, and a 3 gene per

category minimum were used as filters to refine the ontologies. The 62 proteins selected for GO analysis are shaded in grey in Supplementary Table 1.

From our analysis, we found that all the identified GO terms relate to mRNA processing (Figure 3.1B). This result is consistent with previous reports on the proteins that associate with ZC3H14 and the role for ZC3H14 as a regulator of mRNA processing (89). Previous studies show that ZC3H14 localizes to nuclear speckles and interacts with components of the exon junction complex (EJC) (42,82,89). Consistent with these data, the GO terms for nuclear speckle and the EJC serve as positive controls for our analysis. Figure 3.1C presents a circle graph depicting the specific genes identified in each of the GO categories. The colored sections correspond to the GO domain, where each GO term is linked to the individual gene symbol. All of the identified factors that cluster into GO terms are shown. Notably, among the most significant categories identified is the THO complex. Every member of the mammalian THO complex was identified in the ZC3H14-interacting proteome, as indicated in Figure 3.1C by the asterisks. In addition, two THO-associated factors were shown, ALYREF and CHTOP (39,144). Altogether, these data suggest a potential physical link between ZC3H14 and the THO complex.

***ZC3H14 physically associates with the THO complex***. The mammalian THO complex is a six-member RNA processing complex implicated in various RNA processing events that serve critical functions in dictating mRNA fate (145,146). Functional studies of various THO components in mammalian cells and budding yeast suggest roles in transcription elongation, mRNA splicing and nuclear mRNA export (41,66,133,134). Like ZC3H14, THO components are ubiquitously expressed and yet mutations in two THO component genes *THOC2* and *THOC6* are linked to brain disorders (135-137). This finding

strengthens the rationale to further explore the relationship between ZC3H14 and the individual THO components.

To validate the interaction between ZC3H14 and individual THO components, we immunoprecipitated ZC3H14 from mouse brain nuclear lysate as described for the mass spectrometry analysis. Utilizing SDS-PAGE and immunoblotting, we assessed the enrichment of THO components in the ZC3H14 immunoprecipitate. As shown in Figure 3.2A, ZC3H14 protein is enriched in the immunoprecipitation compared to the input and the IgG control. The THO components THOC1 and THOC2 and the THO-associated protein ALYREF are all enriched in the bound fraction with ZC3H14 but not with the IgG control (Figure 3.2A). The interaction between ZC3H14 and THOC1 is not affected by the addition of RNase, however, the interaction between ZC3H14 and THOC2 is partially reduced upon addition of RNase and the interaction with ALYREF is completely abolished. As a negative control, we used the RNA-binding protein HuR, which was not detected as an interacting protein in our mass spectrometry experiments. Consistent with our analysis, HuR shows no co-enrichment with ZC3H14. To further examine the interaction between ZC3H14 and THO components, we performed a reciprocal immunoprecipitation with THOC1. As shown in Figure 3.2B, ZC3H14 co-immunoprecipitates with THOC1 in a manner that is insensitive to RNase treatment. THOC2 serves as a positive control for THOC1 binding, as these proteins interact within the THO complex (35,147,148).

To test whether ZC3H14 and the THO complex exist in a macromolecular complex, we employed glycerol density gradients to sediment nuclear lysate. We purified nuclei from mouse brains and fractionated the nuclease-treated nuclear lysate through a 10-50% glycerol density gradient. Figure 3.2C shows the resulting immunoblot performed on the

glycerol gradient fractions. The fractions increase in density from left to right, with the lightest protein complexes sedimenting towards the left and the heaviest complexes towards the right. ZC3H14 has a wide distribution throughout the gradient. The THO components THOC1 and THOC2 sediment together and are largely restricted to fractions 7-9. Fractions 7-9 also are enriched for ZC3H14 and the THO-associated protein ALYREF. However, these THO-containing fractions do not show enrichment for all nuclear RNA-binding proteins as evidenced by the distribution of the Nucleophosmin (NPM1) protein. These data show a clear overlap between ZC3H14, ALYREF and THO component containing fractions suggesting that these proteins can exist in a complex.

### 3.4.2 Loss of ZC3H14 and THO components affect processing of mRNAs

***THO components are required for proper bulk poly(A) tail length control***. Previous studies have shown that ZC3H14 plays an evolutionarily conserved role in poly(A) tail length control (87,91,92,98). We tested whether THO components are also required for proper poly(A) tail length control. For these studies, we used a mouse neuronal cell line, Neuro2a (N2a) cells, as our experimental model (143). We treated these cells with siRNA against *Zc3h14*, *Thoc1*, or *Thoc5* and compared them to cells treated with a scrambled (Scr) control siRNA. Total RNA from these cells was then isolated and subjected to a poly(A) tail length assay as described in Materials and Methods. Figure 3.3A shows an immunoblot demonstrating efficient siRNA-mediated depletion of THOC1, THOC5, and ZC3H14. The protein eIF5 serves as a loading control. Figure 3.3B shows detection of end-labeled poly(A) tracts that were resolved on a polyacrylamide gel with the number of adenosines increasing from left to right. As shown previously for loss of ZC3H14 (92), depletion of the THO components THOC1 and THOC5 causes extended

bulk poly(A) tails when compared to the control. This result is quantified by a line scan analysis of the samples (Figure 3.3C). The signal from each lane is normalized to the beginning of the gel to correct for loading. This analysis demonstrates that loss of ZC3H14, THOC1, or THOC5 causes a significant increase in the length of bulk poly(A) tail revealing that these factors are required for proper control of poly(A) tail length.

***ZC3H14 and the THO complex regulate the processing of a shared transcript.*** To begin to define specific functions shared by ZC3H14 and the THO complex in RNA processing, processing of specific transcripts needs to be analyzed. However, the complete set of transcripts that are regulated by ZC3H14 and/or THO components in the brain is not known. To identify common transcripts between ZC3H14 and the THO complex, we performed RNA immunoprecipitations for endogenous ZC3H14, THOC5 and ALYREF and analyzed candidate target RNA transcripts. While the RNA-binding mechanism for the THO complex is not yet understood, THOC5 and ALYREF were chosen based on their suggested role as the mRNA adapters for the THO complex (39,149).

For RNA immunoprecipitations, endogenous ZC3H14, THOC5, and ALYREF were immunoprecipitated from UV crosslinked N2a cells (Figure 3.4A). Associated RNA was isolated, converted to cDNA and then subjected to qPCR for analysis (Figure 3.4B-E). Transcript levels were normalized to input and then to IgG control. Using a candidate-based approach, we tested for enrichment of RNA targets using a previously identified ZC3H14 target transcript *Atp5g1* (96) and other important neuronal transcripts with potential links to ZC3H14 (unpublished results), *Psd95* (*Dlg4*) and *Mapt*. The *Atp5g1, Psd95* and *Mapt* mRNA are all significantly enriched in ZC3H14, THOC5, and ALYREF precipitations as compared to the IgG control (Figure 3.4B, C, D). The *Rplp0* transcript is

not enriched in these precipitations, showing that not all transcripts associate with these factors (Figure 4E).

We next tested whether loss of ZC3H14 or THO components affects the steady-state levels of the *Atp5g1, Psd95,* or *Mapt* transcripts. We treated N2a cells with siRNA against *Thoc1*, *Thoc5*, *Zc3h14,* or a scrambled control. We isolated total protein and total RNA samples from these cells. Protein samples were separated by SDS-PAGE and analyzed by immunoblotting to confirm efficient siRNA-mediated depletion (Figure 3.5A). RNA from each siRNA-treated sample was analyzed by qRT-PCR to assess steady-state levels of the *Atp5g1, Psd95,* and *Mapt* transcripts. Depletion of ZC3H14, THOC1, and THOC5 all cause a statistically significant decrease (~50%) in the steady-state levels of both the *Atp5g1* and *Psd95* transcripts (Figure 3.5B, 5C). In contrast, *Mapt* mRNA is only affected by depletion of ZC3H14 with an ~50% decrease in steady-state levels (Figure 5C). Depletion of the THO components THOC1 or THOC5 does not alter the steady-state levels of *Mapt* mRNA. These results suggest that *Atp5g1* and *Psd95* are shared transcripts regulated by ZC3H14 and the THO complex but not all ZC3H14 target transcripts are shared with the THO complex.

One potential explanation for the shared phenotypes of ZC3H14 and THO depletion on poly(A) tail length and steady-state levels of RNA targets could be that depletion of ZC3H14 decreases the levels of THO components or *vice versa*. To address this possibility, we treated N2a cells with siRNA against *Zc3H14* or a scrambled control. We collected protein samples from the transfected cells and used immunoblotting to assess the effect of depletion of ZC3H14 on the protein level of THO components (Supplementary Figure 1A). We quantified protein levels comparing ZC3H14 depleted samples and control samples

using Image J software. Depletion of ZC3H14 did not cause any significant changes in THO protein levels (Supplementary Figure 1B-E). However, depletion of ZC3H14 does cause a slight decrease in THOC5 levels that is not statistically significant (Supplementary Figure 1D). We also assessed whether depletion of THO components alters ZC3H14 proteins levels. Results of this analysis confirm that depletion of THO components does not decrease ZC3H14 protein levels (Supplementary Figure 1F, 1G). These data show that the shared phenotypes detected upon depletion of ZC3H14 and THO components are not from changes of steady-state levels of the associated protein.

Previous work in budding yeast analyzed genetic interactions between the *ZC3H14* orthologue *NAB2* and *THO1*, the yeast orthologue of *THOC1*. Surprisingly, phenotypes present in *nab2* mutants were suppressed by deletion of *THO1* (68). To extend this analysis to mammalian cells, we used siRNA to simultaneously deplete either ZC3H14 and THOC1 or ZC3H14 and THOC5. As shown in Figure 3.5A, we confirmed depletion of the targeted proteins. Interestingly, in cells simultaneously treated with siRNA against *Thoc1* and *Zc3h14* or *Thoc5* and *Zc3H14,* we found that the decrease in steady-state mRNA levels observed with individual depletion of each of these individual components is rescued. Depletion of ZC3H14 and either THOC1 or THOC5 cause no statistically significant decrease in the steady-state levels of *Atp5g1, Psd95* and *Mapt* (Figure 3.5B-5D). In fact, simultaneous depletion of ZC3H14 and THOC1 causes an increase in the steady-state levels of *Psd95* compared to the scrambled control (Figure 3.5C) and depletion of ZC3H14 together with THOC1 or THOC5 causes a slight, but not statistically significant increase in the steady-state levels of *Mapt* mRNA (Figure 3.5D). These results show that

simultaneous depletion of both ZC3H14 and THO components ameliorates the individual effect of depletion of either one of these RNA processing factors.

Previous work demonstrated that ZC3H14 is critical to prevent the escape of premature or intron-containing A*tp5g1* transcript to the cytoplasm in cultured breast cancer cells (96). Furthermore, studies in budding yeast show that *THO* mutants have a pre-mRNA leakage phenotype (150). To investigate the effect of ZC3H14 and the THO complex on pre-mRNA, we depleted either ZC3H14 or THOC1 using siRNA in N2a cells and fractionated the cells into nuclear and cytoplasmic fractions. We confirmed successful knockdown and fractionation by immunoblotting (Figure 3.6A). As markers for fractionation, we used NUP62 (nuclear) and HSP90 (cytoplasmic) proteins. As illustrated in Figure 6B, we designed primers that specifically detect pre-mRNA of *Atp5g1, Psd95, Mapt*, and a control transcript *Rplp0*. The location of the qPCR primers is indicated by arrows above their targeted regions in the schematic corresponding to each pre-mRNA. We then measured the steady-state level of *Atp5g1, Psd95, and Mapt* pre-mRNA following depletion of either ZC3H14 or THOC1. We found little to no significant change in the steady-state levels of total *Atp5g1* or *Psd95* pre-mRNA upon depletion of ZC3H14 or THOC1 (Supplementary Figure 2A, 2B). In contrast, the pre-mRNA for *Mapt* is significantly increased upon depletion of THOC1 and significantly decreased upon depletion of ZC3H14 (Supplementary Figure 2C). These results suggest that *Atp5g1* and *Psd95* are likely both regulated at the posttranscriptional level by ZC3H14 and the THO complex.

As total *Atp5g1* and *Psd95* pre-mRNA levels show no statistically significant change upon depletion of ZC3H14 or THOC1, we next investigated the nuclear and

cytoplasmic levels of *Atp5g1*, *Psd95, and Mapt pre-mRNA* upon siRNA-mediated depletion of ZC3H14 or THOC1. We analyzed the nuclear pool of *Atp5g1, Psd95, Mapt*, and *Rplp0* pre-mRNAs with pre-mRNA specific primers (Figure 3.6C-F). RNA levels were normalized first to *18S* rRNA and then to the total pre-mRNA. After normalization, the fold change for each condition was compared across samples. As shown in Figures 3. 6C-6F, depletion of ZC3H14 or THOC1 did not cause a significant change in the steady-state levels of *Atp5g1, Psd95, Mapt,* or *Rplp0* pre-mRNA within the nuclear fraction.

We next assessed the cytoplasmic levels of *Atp5g1, Psd95*, *Mapt,* and *Rplp0* pre-mRNA. Cytoplasmic levels of *Atp5g1* pre-mRNA significantly increase upon depletion of either ZC3H14 or THOC1 (Figure 3.7A). The steady-state levels of *Psd95* pre-mRNA in the cytoplasmic fraction significantly increase upon depletion of THOC1 and trend toward an increase when ZC3H14 is depleted (Figure 3.7B). *Mapt* pre-mRNA cytoplasmic levels are slightly decreased upon depletion of THOC1 and are significantly increased upon depletion of ZC3H14 (Figure 3.7C). The level of the control transcript *Rplp0* pre-mRNA in the cytoplasm shows no statistically significant change under any of the conditions examined (Figure 3.7D). These data demonstrate that depletion of ZC3H14 and THOC1 can cause an accumulation of cytoplasmic pre-mRNA for multiple transcripts.

As the distribution of target pre-mRNAs were changed upon loss of ZC3H14 and THOC1, we examined whether the distribution of the mature mRNAs is affected upon loss of these factors. Using qRT-PCR and primers specific to the mature mRNA, we analyzed the nuclear cytoplasmic distribution of *Atp5g1*, *Psd95,* and *Mapt* mature mRNA upon depletion of ZC3H14 or THOC1. Depletion of THOC1 causes a significant decrease in both the nuclear and cytoplasmic level of *Atp5g1* mRNA, while depletion of ZC3H14 only

causes a slight, but not statistically significant decrease in the cytoplasmic mRNA (Supplementary Figure 3A). Nuclear and cytoplasmic steady-state levels of *Psd95* and *Mapt* mRNA are not significantly affected upon depletion of ZC3H14 (Supplementary Figure 3B-C). Depletion of THOC1 causes a slight increase in *Psd95* nuclear levels (Supplementary Figure 3B), while *Mapt* levels are significantly increased in the nucleus and slightly decreased in the cytoplasm (Supplementary Figure 3C). These data support the canonical role of THOC1 as an mRNA export factor, as slight changes are observed in the subcellular distribution of mature mRNA upon THOC1 depletion.

The accumulation of pre-mRNA in the cytoplasm upon loss of ZC3H14 or THOC1 led us to investigate how pre-mRNA could prematurely exit the nucleus. One possibility is that ZC3H14 and THOC1 modulate the interaction of transcripts with the mRNA export machinery. The ZC3H14 orthologue Nab2 mediates interactions between mRNA export factors and the nuclear pore (88,151,152). To test whether ZC3H14 modulates interactions between RNA and mRNA export components, we examined the RNA associated with the mRNA export factor NXF1 upon depletion of ZC3H14 or THOC1. Normally, NXF1 binds to processed mRNA and mediates the rapid export of these mature mRNAs through the nuclear pore (153). Thus, NXF1 should not associate with pre-mRNA that is not competent for export.

To assess whether loss of ZC3H14 or THOC1 allows premature association of NXF1 with transcripts that are not yet ready for export, we used siRNA to deplete either ZC3H14 or THOC1 in N2a cells, immunoprecipitated endogenous NXF1, and assessed the NXF1-bound RNAs by qRT-PCR. Figure 3.8A shows an immunoblot of the input levels of NXF1, THOC1, and ZC3H14 and the NXF1 immunoprecipitation from the siRNA-

treated cells. Depletion of ZC3H14 slightly decreases the steady-state level of NXF1 as compared to the control, while depletion of THOC1 has no significant effect on the levels of NXF1.

To test for the association of NXF1 with pre-mRNA, RNA was isolated from the siRNA-treated cells and subjected to qRT-PCR. RNA was normalized to the input of each condition and then compared to the scrambled control. As shown in Figure 3.8B and 3.8C, depletion of ZC3H14 or THOC1 causes a significant increase in the association of NXF1 with both *Atp5g1* (3.8B) and *Psd95* (3.8C) pre-mRNA compared to the siRNA control. Depletion of ZC3H14 but not THOC1 causes a significant increase in the association of NXF1 with *Mapt* pre-mRNA compared to scrambled control (Figure 3.8D). As a control, we measured NXF1 enrichment with *RPLP0* pre-mRNA because this transcript was unchanged upon depletion of ZC3H14 or THOC1. Neither depletion of ZC3H14 or THOC1 increases the association of NXF1 with *Rplp0* pre-mRNA (Figure 3.8E).

**3.5 Discussion:**

In this study, we employed an unbiased proteomic approach to identify factors that interact with ZC3H14 in the murine brain. From this analysis, we identified a link between ZC3H14 and proteins that comprise the THO mRNA processing complex. Detailed analysis reveals that ZC3H14 interacts with THO components in the nucleus. We present evidence that ZC3H14 and the THO complex can cooperate to ensure proper processing of mRNA. In fact, ZC3H14 and the THO components THOC1 and THOC5 are required for proper control of bulk poly(A) tail length. In addition, we identified the transcripts *Atp5g1*,

a nuclear encoded component of the mitochondria ATP synthase machinery (140) and the neuronal post synaptic density component, *Psd95 (141),* as shared targets of ZC3H14 and the THO complex. Both ZC3H14 and the THO complex are required for the proper processing of the *Atp5g1* and *Psd95* transcripts, as loss of either of these factors leads to a decrease in the steady-state levels of the mature transcripts. Furthermore, we provide evidence that the decrease in the steady-state level of the mature transcripts could be due to the improper export of the unprocessed pre-mRNAs from the nucleus via association with RNA export factors. Thus, these data suggest a model where ZC3H14 and the THO complex are critical RNA regulatory factors that cooperate to coordinate nuclear mRNA processing events with export from the nucleus.

This work highlights the requirement for ZC3H14 and the THO complex in mRNA processing and establishes ZC3H14 and the THO complex as surveillance factors for nuclear mRNA processing in vertebrate cells. These data support the model presented in Figure 3.9A, where ZC3H14 and the THO complex are recruited to the transcript as early as splicing to ensure that the proper processing of the transcript is coupled with nuclear export. Once the transcript is checked for quality, ZC3H14 and the THO complex would then facilitate subsequent association with export factors such as NXF1.

Loss of ZC3H14 or the THO complex also causes defects in nuclear mRNA processing, possibly by uncoupling processing events and mRNA export (Figure 3.9B). This compromised quality control pathway, resulting from loss of ZC3H14 and the THO complex, could underlie the increase in the bulk poly(A) tail length of RNA and a decrease in the steady-state levels of mature target transcripts. These defects could occur because the improperly processed transcripts can bypass further processing steps and gain access

to the export machinery. This defect would then cause an increase in the amount of pre-mRNA in the cytoplasm and possibly lead to downstream effects on gene expression.

Although our studies provide evidence that ZC3H14 and the THO complex cooperate to regulate mRNA processing, additional studies are required to define the precise nature of the interaction. Our mass spectrometry data reveals that all six members of the THO complex as well as the THO-associated proteins ALYREF (39) and CHTOP (144) co-immunoprecipitate with ZC3H14. We confirmed that ZC3H14 interacts with the THO components THOC1, THOC2, and the THO-associated protein ALYREF and that these proteins can exist in a macromolecular complex. The interaction between ZC3H14 and THOC1 is RNase insensitive, but the interaction between ZC3H14 and THOC2 is decreased by RNase treatment suggesting an RNA-independent interaction between ZC3H14 and THOC1. While the composition of the evolutionarily-conserved THO complex has been studied extensively in budding yeast (148,154-156), there is little information about the organization of the vertebrate THO complex or how the complex binds RNA. The THO complex could interface with various RNA-binding proteins to dictate the fate of different target transcripts. As our data reveals that ZC3H14 interacts with the THO complex, ZC3H14 could be an auxiliary component of the THO complex that targets the complex to certain mRNAs for specific functions. Thus, the THO complex could cooperate with distinct RNA-binding proteins to modulate different RNA processing events. Additional studies will be required to determine if ZC3H14 interacts directly with the THO complex, however; more knowledge of the overall organization of the THO complex must be acquired to lay the groundwork for these studies.

Both ZC3H14 and components of the THO complex are localized to nuclear speckles (41,82), which are nuclear domains enriched for RNA processing factors such as those required for pre-mRNA splicing (157,158). The location of ZC3H14 and the THO complex within these nuclear speckles appropriately poises these factors to surveille mRNA processing. While a direct role for ZC3H14 in mRNA quality control has not been fully established, our proteomic data provide additional evidence for a link between ZC3H14 and proteins that comprise the exon junction complex (EJC). The EJC is deposited 20-24 nucleotides upstream of the 5' end of splice junctions and then dictates further processing events including mRNA export (159,160). Together with the EJC, ZC3H14 and the THO complex could ensure that the proper processing of the mRNAs occurs prior to mRNA export.

Both our analysis of simultaneous depletion of ZC3H14 and THO components and genetics studies in budding yeast provide interesting insight into how these factors cooperate. Studies in budding yeast have shown that growth defects of *nab2* mutants can be suppressed by other mutants in the mRNA export pathway including *tho1*Δ and *pml39*Δ *(68,161)*. Deletion of *THO1* results in restoration of nuclear export and improved growth of *nab2* mutants (68), while *pml39*Δ mutants suppress the growth phenotypes of *nab2* mutants (161). In agreement with these data, our work shows that the combined loss of ZC3H14 and THO components rescues the steady-state levels of target mRNAs that are decreased upon depletion of either individual component. We suggest that this rescue could be the result of an increase in the export of defective RNA. As quality control factors, ZC3H14 and the THO complex would normally prevent export of defective or unprocessed mRNAs from the nucleus. The loss of one factor would cause an increase in mRNA

processing defects as ZC3H14 and the THO complex have roles in mRNA processing. The defective RNAs would then be retained in the nucleus and presumably degraded, resulting in the decreased steady-state levels of the target mRNA seen with individual depletion of one component. However, in our model with the simultaneous loss of both ZC3H14 and the THO complex, the defective RNAs are not retained and degraded in the nucleus but instead can escape to the cytoplasm, which accounts for the restored steady-state levels of the target mRNA. Further characterization of the defects observed upon simultaneous depletion of ZC3H14 and THO components is needed to test this model and elucidate the exact role these factors play in the mRNA quality control pathway. Nonetheless, the confluence of the results obtained with genetic studies in budding yeast and siRNA in cultured cells strongly supports functional interplay between these mRNA processing factors.

A future challenge remains in understanding precisely how ZC3H14 and the THO complex cooperate to dictate the fate of a given mRNA. The mechanisms by which ZC3H14 and the THO complex are recruited to transcripts and the exact roles they play in mRNA processing are still not understood. With ZC3H14 and the THO complex localized to nuclear speckles, one possible model is that these factors are recruited to the transcript concomitant with the processing of the pre-mRNA. As the pre-mRNA is undergoing processing, ZC3H14 and the THO complex would govern proper polyadenylation of the 3'end and then the eventual export of the mRNA. Furthermore, recent reports have found that ZC3H14 and some THO components are detected in the cytoplasm (97,162,163), suggesting these proteins could accompany the mature mRNP to the cytoplasm for further regulation. Thus, these components potentially influence gene expression both through

mediating and/or monitoring processing steps in the nucleus or through direct modulation of mRNA fate in the cytoplasm. For example, ZC3H14, which binds with high affinity to polyadenosine RNA (93) could compete with the cytoplasmic poly(A) binding protein, PABPC1, for binding to poly(A) tails to modulate translation.

As ZC3H14 and members of the THO complex are ubiquitously expressed (35,82,98), how the loss of these factors can cause brain-specific pathology is unclear. If ZC3H14 and the THO complex influence the fate of mRNA transcripts in the cytoplasm either by ensuring proper completion of nuclear processing events or by accompanying specific RNAs to the cytoplasm, this function could be more critical in neurons than in other cell types (164,165). Alternatively, ZC3H14 and the THO complex could be critical for the regulation of a specific subset of key neuronal transcripts. The loss of ZC3H14 and THO components does not affect all transcripts (96,138) raising the possibility that these factors have specificity for certain neuronal transcripts. Indeed, the *Atp5g1* and *Psd95* transcripts studied here are critical for proper neuronal function and are also trafficked to neurite extensions to modulate local mitochondrial function in the case of *Atp5g1* (166) or to maintain the integrity of the post synaptic density in the case of *Psd95* (141). Loss of either ZC3H14 or THO components would impair the processing of these transcripts, potentially interfering with proper temporal and spatial control of gene expression.

While we identified the interaction between ZC3H14 and the THO complex in the brain, this interaction is likely conserved in other tissues and cell types as these factors are all ubiquitously expressed. Thus, interplay between ZC3H14 and the THO complex could provide a fundamental mechanism to couple nuclear mRNA processing with export from the nucleus. We speculate that ZC3H14 and THO may be critical for a subset of transcripts.

Indeed, the THO complex has been implicated in the resolution of R loops (167-169). Transcripts prone to form R loops may be dependent on THO and ZC3H14 for efficient processing. In fact, recent studies report that the splicing out of introns prevents the formation of R-loops (167) and that retained introns with high GC contents are prone to the formation of R-loops leading to DNA damage (170). While the introns examined in this study have not been classified as retained introns, they do have high GC contents (> 50%). ZC3H14 and the THO complex could work alongside the splicing machinery to aid in the resolution of R-loops during the splicing out of introns with high GC contents prior to nuclear export. Depletion of ZC3H14 or THO components could lead to more R-loop formations and impact mRNA processing such that pre-mRNA accumulates, allowing for the improperly processed mRNA to gain access to the export machinery. Studies are underway to define the spectrum of direct targets for these factors in a common cell type and assess whether ZC3H14 and the THO complex are required for the proper processing of transcripts that are prone to R-loop formation.

In summary, our work identifies the THO complex as a novel protein interactor of ZC3H14. Using detailed biochemical and molecular approaches, we characterized the interaction between ZC3H14 and the THO complex as well as their regulation of target transcripts. Our work supports a model where ZC3H14 and the THO complex work together as quality control factors for nuclear mRNA processing and this work highlights the importance of both ZC3H14 and the THO complex in the mRNA quality control pathway.

**Acknowledgements:**

This study was supported in part by the Emory Proteomics Core, which is subsidized by the Emory University School of Medicine and is one of the Emory Integrated Core Facilities. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institutes of Health.

**Funding:**

This work was supported by the National Institutes of Health [5R01GM058728 to AHC and 5F31GM112418 to KJM]. Funding for open access charge: National Institutes of Health.

**Supplementary data:**

Supplementary Data are available at NAR Online.

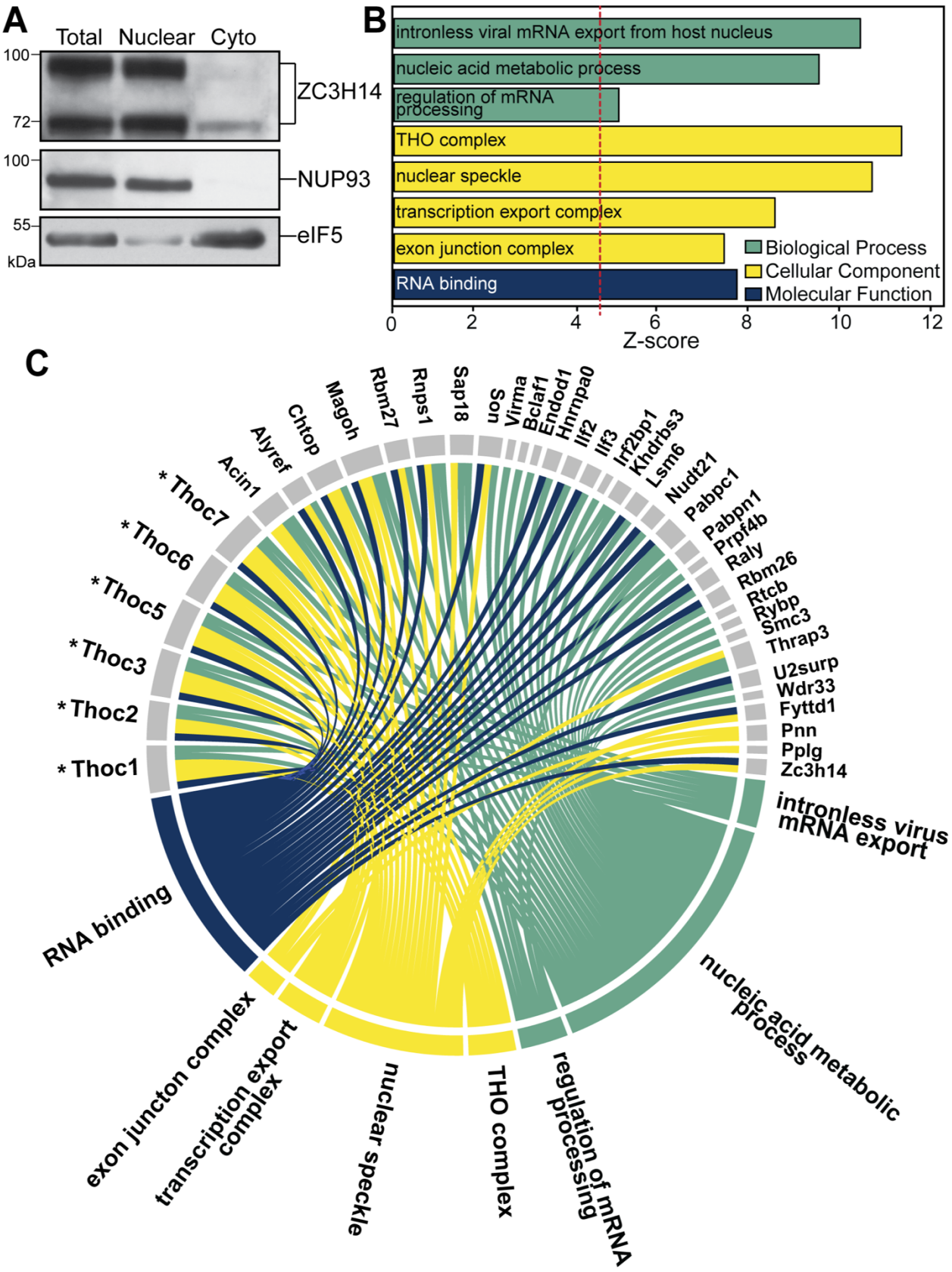

A
Total
Nuclear
Cyto
100
72
ZC3H14
100
NUP93
55
kDa
eIF5
B
intronless viral mRNA export from host nucleus
nucleic acid metabolic process
regulation of mRNA processing
THO complex
nuclear speckle
transcription export complex
exon junction complex
RNA binding
Biological Process
Cellular Component
Molecular Function
0
2
4
6
8
10
12
Z-score
C
Sap18
Son
Virma
Bclaf1
Endod1
Hnrnpa0
Ilf2
Ilf3
Irf2bp1
Khdrbs3
Lsm6
Nudt21
Pabpc1
Pabpn1
Prpf4b
Raly
Rbm26
Rtcb
Rybp
Smc3
Thrap3
U2surp
Wdr33
Fyttd1
Pnn
Pplg
Zc3h14
Rbm27
Rnps1
Magoh
Chtop
Alyref
Acin1
* Thoc7
* Thoc6
* Thoc5
* Thoc3
* Thoc2
* Thoc1
intronless virus mRNA export
nucleic acid metabolic process
regulation of mRNA processing
THO complex
nuclear speckle
transcription export complex
exon juncton complex
RNA binding

**Figure 3.1: ZC3H14 interacts with multiple mRNA processing proteins in the brain.** (A) Fractionation of mouse brain lysate. Equal amounts of protein lysate from Total, Nuclear, or Cytoplasmic (Cyto) fractions were analyzed by immunoblotting with ZC3H14 antibody. The ZC3H14 antibody recognizes multiple isoforms of ZC3H14 resulting from alternative splicing of the *ZC3H14* gene (82). To ensure proper fractionation, a nuclear marker (NUP93) and a cytoplasmic marker (eIF5) are shown. Molecular weights are indicated in kDa. (B) The gene ontology (GO) terms for the proteins that are mostly highly enriched within the immunoprecipitation of ZC3H14 compared to IgG control as identified by tandem mass spectrometry. These proteins are grouped based on the three GO domains: Biological Process, Cellular Component, and Molecular Function. The inclusion criteria for this analysis were a Z-score of $> 4.5$, a p-value of $< 0.05$, and at least three gene terms per GO term. (C) The gene terms that cluster into the listed GO terms in (B) are depicted. Colored lines connect GO categories to GO terms and individual term members. All core members of the THO complex (THOC1, THOC2, THOC3, THOC5, THOC6, THOC7) are present (indicated by an asterisk).

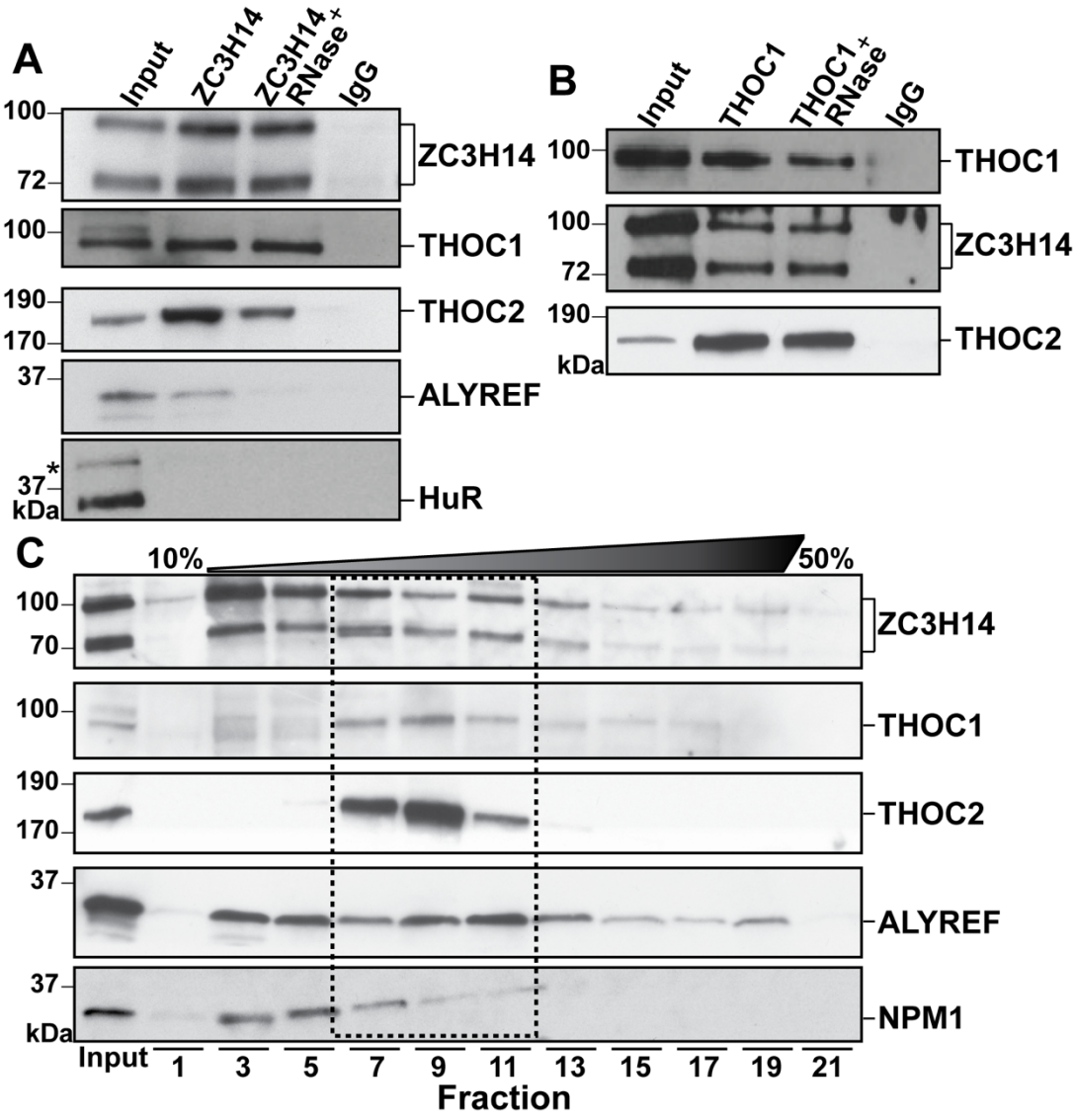



**Figure 3.2: ZC3H14 interacts with the THO complex.**
(A) Co-immunoprecipitation of ZC3H14 and THO components from mouse brain lysate. The Input and immunoprecipitated samples for ZC3H14, ZC3H14 + RNase, and control IgG are shown. THOC1 and THOC2 serve as representative members of the THO complex, together with the THO-associated protein ALYREF. HuR serves as a negative control. Asterisks indicate non-specific protein bands detected by the commercial antibodies. (B) Reciprocal co-immunoprecipitation using anti-THOC1. The Input and immunoprecipitated samples for THOC1, THOC1 + RNase and control IgG immunoprecipitations were probed to detect THOC1, ZC3H14 and THOC2 proteins. (C) Sedimentation of ZC3H14 protein complexes from the nuclear fraction of mouse brain lysate through a 10-50% glycerol gradient. Gradient fractions from top to bottom were resolved on SDS-PAGE from left to right. Fractions were probed with antibodies to detect ZC3H14, THOC1, THOC2, ALYREF, and Nucleophosmin (NPM1) as a control. The Input sample is shown on the left. Overlapping fractions are outlined. Molecular weights are shown in kDa and indicated to the left.

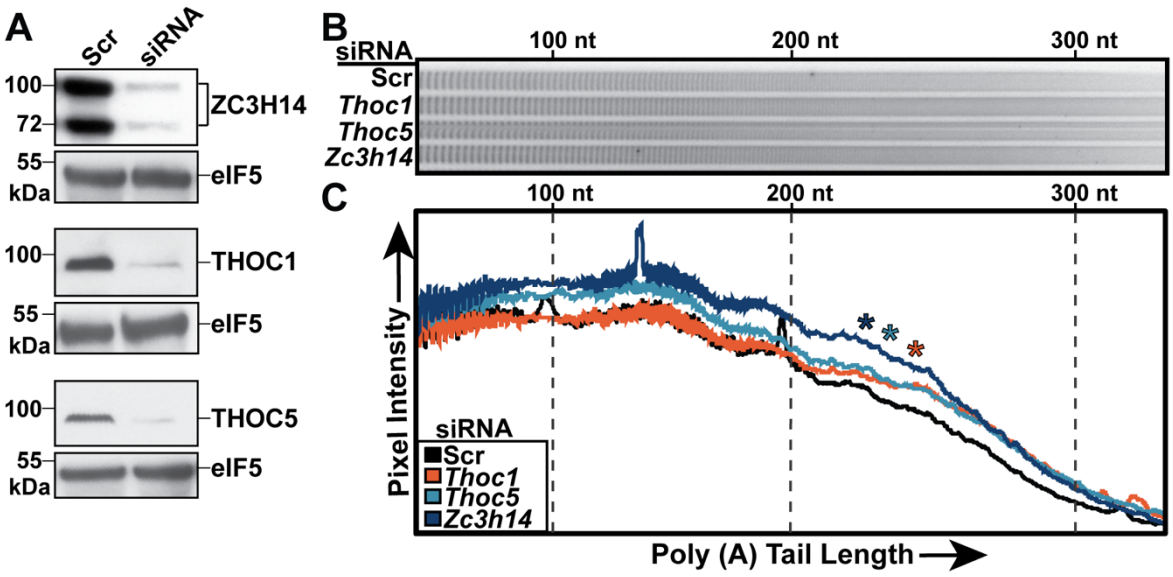



**Figure 3.3: Knockdown of ZC3H14 and THO components cause extended bulk poly(A) tails.** N2a cells were treated with scrambled (Scr), *Thoc1*, *Thoc5*, or *Zc3h14* siRNA to deplete the corresponding proteins. (A) Immunoblot of protein samples from siRNA-treated N2a cells. Antibodies against THOC1, THOC5, and ZC3H14 were used to confirm depletion of each protein. eIF5 serves as a loading control for each sample. Molecular weights are given in kDa and indicated to the left. (B) Representative image of 3'-end labeled poly(A) tracts from bulk poly(A) tails of the siRNA-treated cells that have been resolved on a denaturing polyacrylamide gel and detected by autoradiography. (C) Line scan analysis of bulk poly(A) tail length. Poly(A) tails were quantified using ImageQuant software by plotting Pixel Intensity versus poly(A) tail length and normalizing to the start of the gel. Size markers are shown for 100 nt, 200 nt, and 300 nt. Line scan data are presented as means for three independent experiments. Statistical analysis was performed using a non-parametric, one-way ANOVA Kruskal-Wallis test. Asterisks (*) indicate results that are statistically significant at *p-value < 0.05.

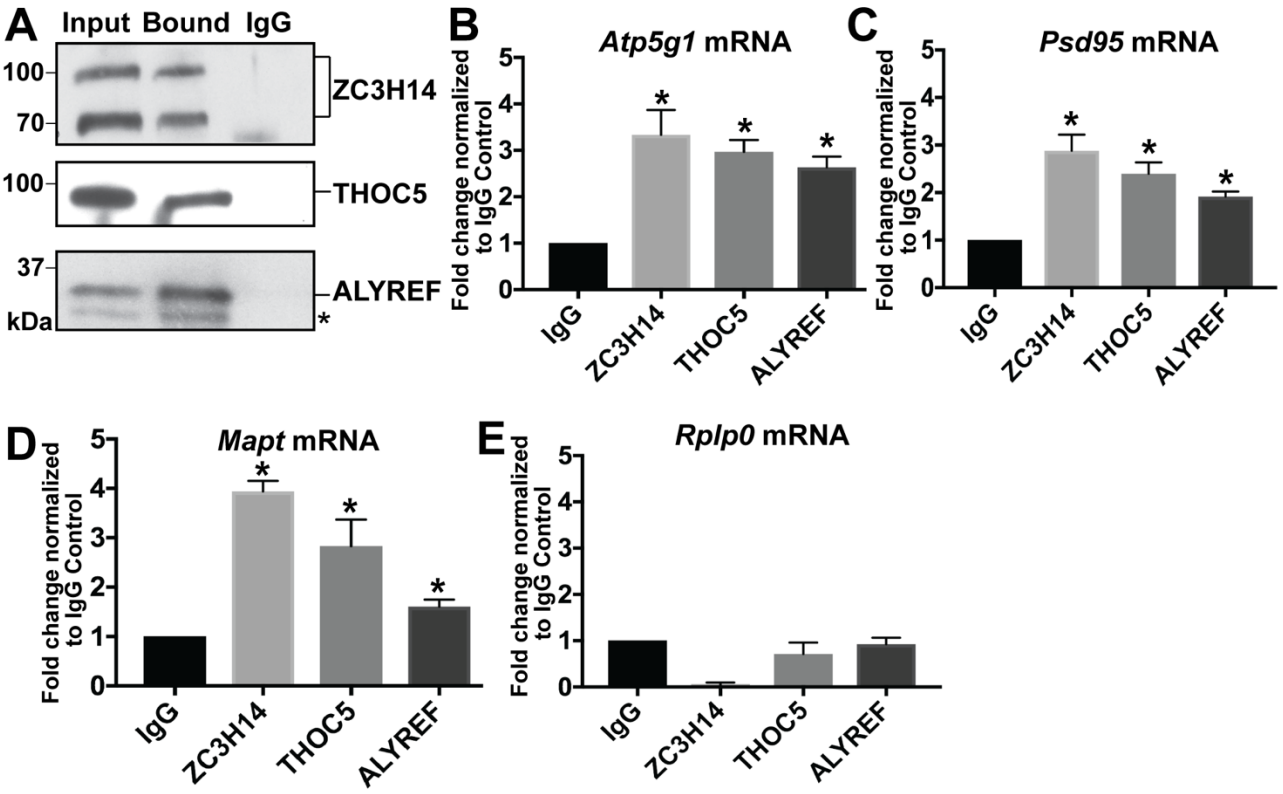



**Figure 3.4: ZC3H14 and THO components bind to common mRNA targets.** (A) Endogenous ZC3H14, THOC5, and ALYREF RNA immunoprecipitation from N2a cells. The Input fractions and the Bound immunoprecipitated fractions were compared to the IgG immunoprecipitated control to confirm efficient and specific immunoprecipitation of ZC3H14, THOC5, and ALYREF. Molecular weights are indicated in kDa to the left. Asterisk indicates a non-specific band recognized by the commercial ALYREF antibody. (B-E) Quantification of RNA IP enrichment in (A). RNA isolated from the RNA-IPs was subjected to RT-qPCR analyses to detect the (B) *Atp5g1*, (C) Psd95, (D) *Mapt*, and (E) *Rplp0* transcripts. RNA levels in the bound fractions are normalized to input levels and are presented as fold change normalized to IgG control, with the value of the IgG control set to 1.0. Values represent the mean ± SEM for n=3 independent experiments. Asterisks (*) indicate results that are statistical significance at *p-value < 0.05.

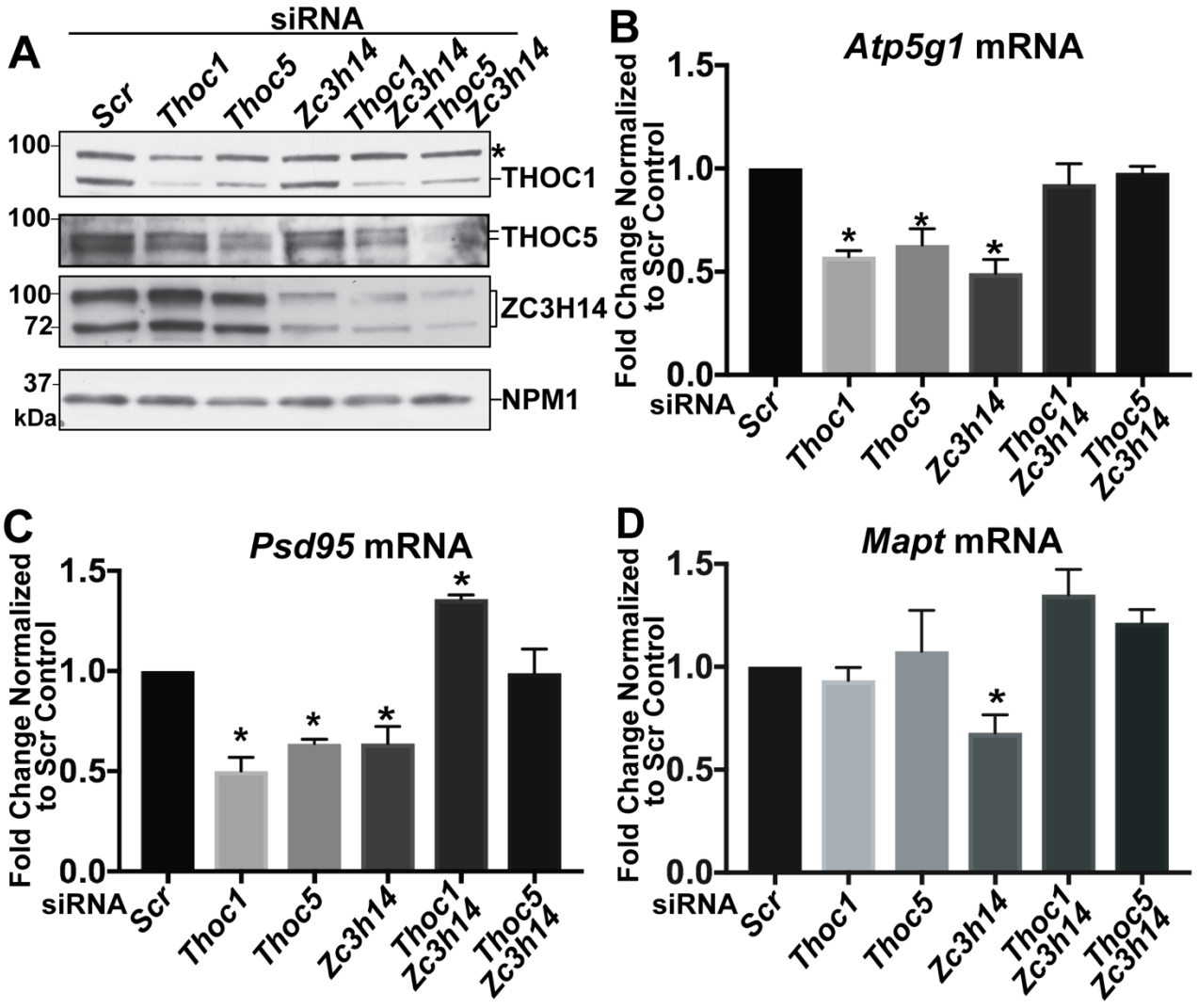



**Figure 3.5: Depletion of ZC3H14 or THO components decreases steady-state mRNA levels.** N2a cells were treated with siRNA to deplete THOC1, THOC5, or ZC3H14. Cells where either treated with individual siRNA or siRNA used in combination to deplete both THOC1 and ZC3H14 or both THOC5 and ZC3H14. Protein samples were analyzed by immunoblotting in (A) and RNA was analyzed by qRT-PCR in (B-D). (A) Antibodies against THOC1, THOC5, and ZC3H14 were used to confirm depletion of each protein with NPM1 serving as a loading control. Molecular weights are given in kDa and indicated to the left. Asterisk indicates a non-specific band recognized by the commercial THOC1 antibody. (B-D) RNA levels for (B) *Atp5g1*, (C) *Psd95*, and (D) *Mapt* mRNA are normalized to *18S* transcript levels and are presented as fold change normalized to Scr control, with the value for the Scr control set to 1.0. Values represent the mean ± SEM for n=3 independent experiments. All primers span exon-exon boundaries to specifically detect mature mRNA. Asterisks (*) indicate results that are statistically significant at *$p < 0.05$.

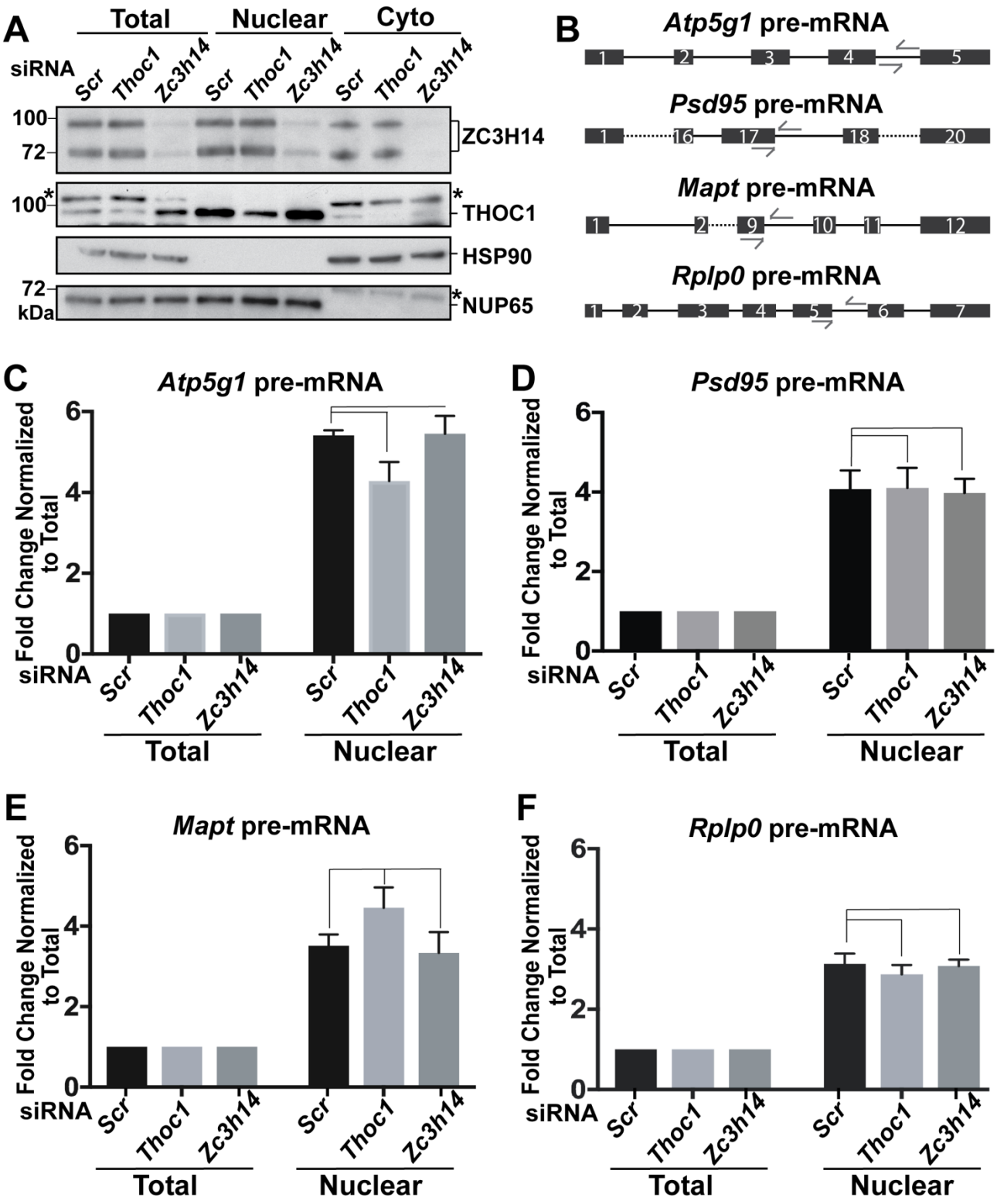

A
Total
Nuclear
Cyto
siRNA
Scr
Thoc1
Zc3h14
100
72
ZC3H14
100*
THOC1
HSP90
72
kDa
NUP65
B
Atp5g1 pre-mRNA
Psd95 pre-mRNA
Mapt pre-mRNA
Rplp0 pre-mRNA
C
Atp5g1 pre-mRNA
Fold Change Normalized to Total
D
Psd95 pre-mRNA
E
Mapt pre-mRNA
F
Rplp0 pre-mRNA
Total
Nuclear

**Figure 3.6: Depletion of ZC3H14 or THOC1 does not affect the nuclear levels of target pre-mRNA**. N2a cells were treated with scrambled (Scr), *Thoc1*, or *Zc3h14* siRNA and fractionated into Total, Nuclear, and Cytoplasmic (Cyto) fractions. Protein and RNA from these fractions were analyzed by immunoblotting (A) and qRT-PCR (C-F). (A) Fractions were probed with antibodies to detect ZC3H14 and THOC1 to confirm efficient depletion. Antibodies against the nuclear protein NUP62 and the cytoplasmic protein HSP90 were used to normalize loading and confirm proper fractionation. Molecular weights are given in kDa and indicated to the left. Asterisks indicate non-specific bands recognized by the commercial THOC1 and NUP62 antibody. (B) Schematic of pre-mRNA primers. The locations of the qPCR primers are indicated by arrows in the schematics above their targeted regions. Exons are represented as boxes and introns are represented as lines for each pre-mRNA. Dashed lines indicate regions of the transcript that are not shown. (C-F) Quantification of Nuclear levels for (C) *Atp5g1,* (D) *Psd95*, (E) *Mapt,* and (F) *Rplp0* pre-mRNA compared to Total levels. RNA levels are normalized to *18S* transcript levels and are presented as fold changes normalized to Total levels, which are set to 1.0. Fold changes of experimental groups were compared to Scr control. Values represent the mean ± SEM for n=3 independent experiments. Asterisks (*) indicate results that are statistically significant at *p-value < 0.05.

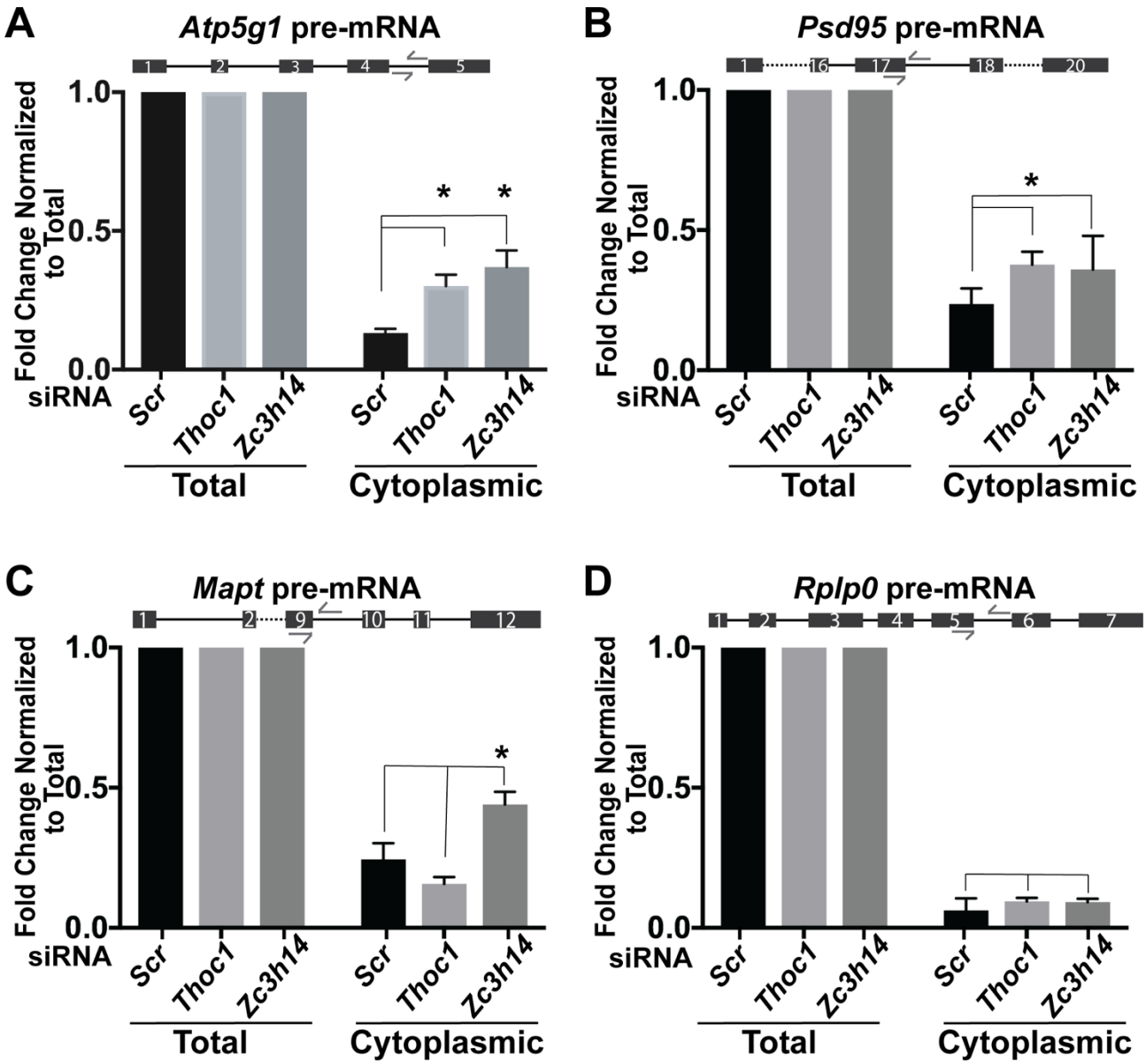



**Figure 3.7: Depletion of ZC3H14 or THOC1 leads to an increase in the cytoplasmic level of target pre-mRNA**. N2a cells were treated with scrambled (Scr), *Thoc1*, or *Zc3h14* siRNA and fractionated into Total, Nuclear, and Cytoplasmic (Cyto) fractions. RNA from the total and cytoplasmic fractions were analyzed by qRT-PCR. (A-D) Quantification of Cytoplasmic (A) *Atp5g1,* (B) *Psd95*, (C) *Mapt,* and (D) *Rplp0* pre-mRNA levels compared to Total levels. RNA levels are normalized to *18S* transcript levels and are presented as fold changes normalized to Total levels, which are set to 1.0. Fold changes for experimental groups are compared to Scr control. The locations of the qPCR primers are indicated by arrows in the schematics above each bar graph. Exons are represented as boxes and introns are represented as lines for each pre-mRNA. Dashed lines indicate regions of the transcript that are not shown. Values represent the mean ± SEM for n=3 independent experiments. Asterisks (*) indicate results that are statistically significant at *p-value < 0.05.

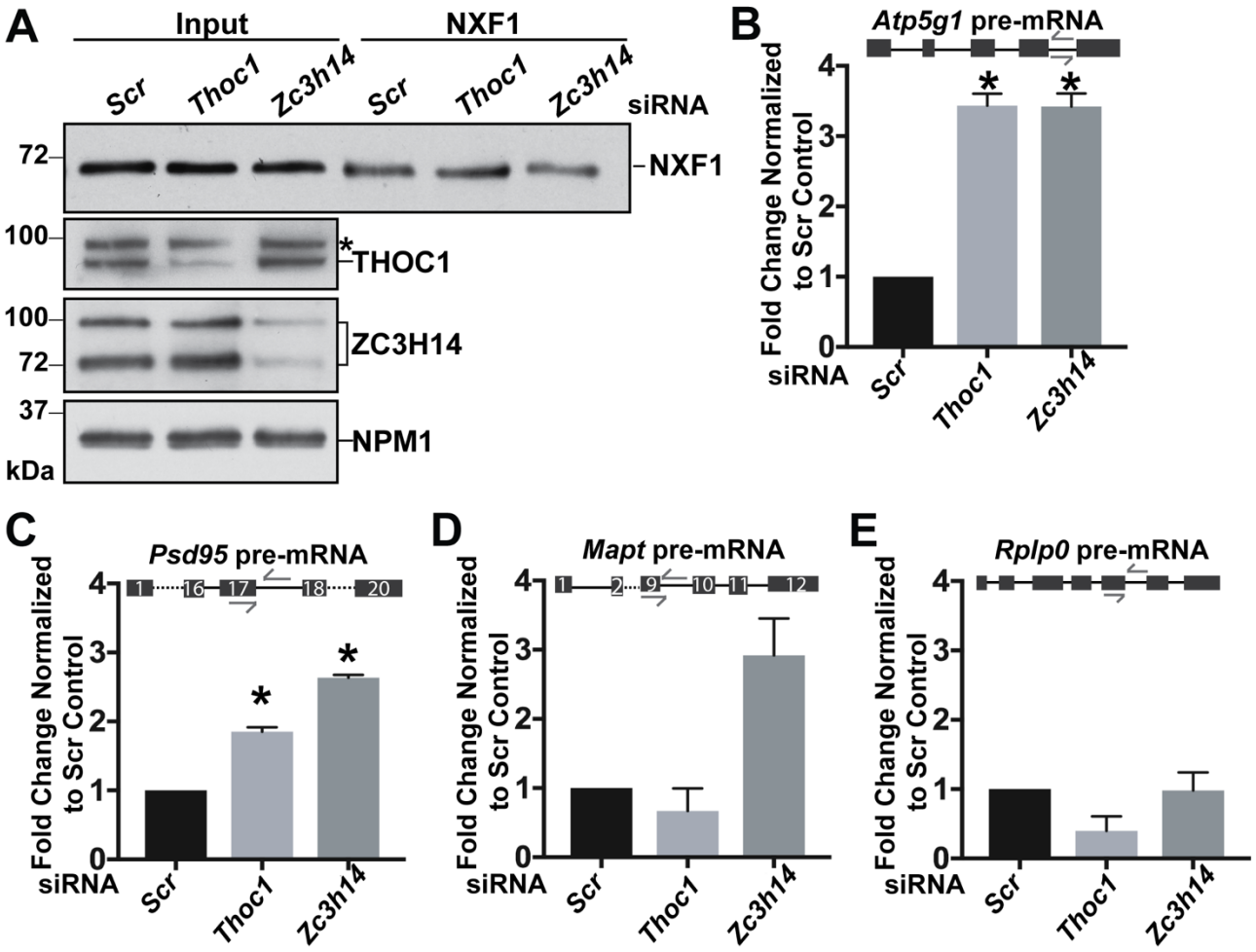



**Figure 3.8: Loss of ZC3H14 affects NXF1 target mRNA selection**
Endogenous NXF1 was immunoprecipitated from N2a cells treated with scrambled (Scr), Thoc1, or *Zc3h14* siRNA. (A) Immunoblot of NXF1 immunoprecipitation. Antibodies against THOC1 and ZC3H14 were used to confirm knockdown of ZC3H14 and THOC1. NPM1 serves as a loading control. Molecular weights are given in kDa and indicated on the left. Asterisk indicates a non-specific band recognized by the commercial THOC1 antibody. (B-E) Quantification of RNA-IP enrichment. RNA isolated from the NXF1 RNA-IP was subjected to qRT-PCR analyses to detect (B) *Atp5g1* (B), (C) Psd95, (D) Mapt, and (E) *Rplp0* pre-mRNA. RNA levels in the NXF1 bound fractions are normalized to respective input levels and are presented as fold changes normalized to Scr control, which was set to 1.0. The locations of the qPCR primers are indicated by arrows in the schematic above the graph. Exons are represented as boxes and introns are represented as lines for each pre-mRNA. Dashed lines indicate regions of the transcript that are not shown. Values represent the mean ± SEM for n=3 independent experiments. Asterisks (*) indicate that the data are statistically significant at *p-value < 0.05.

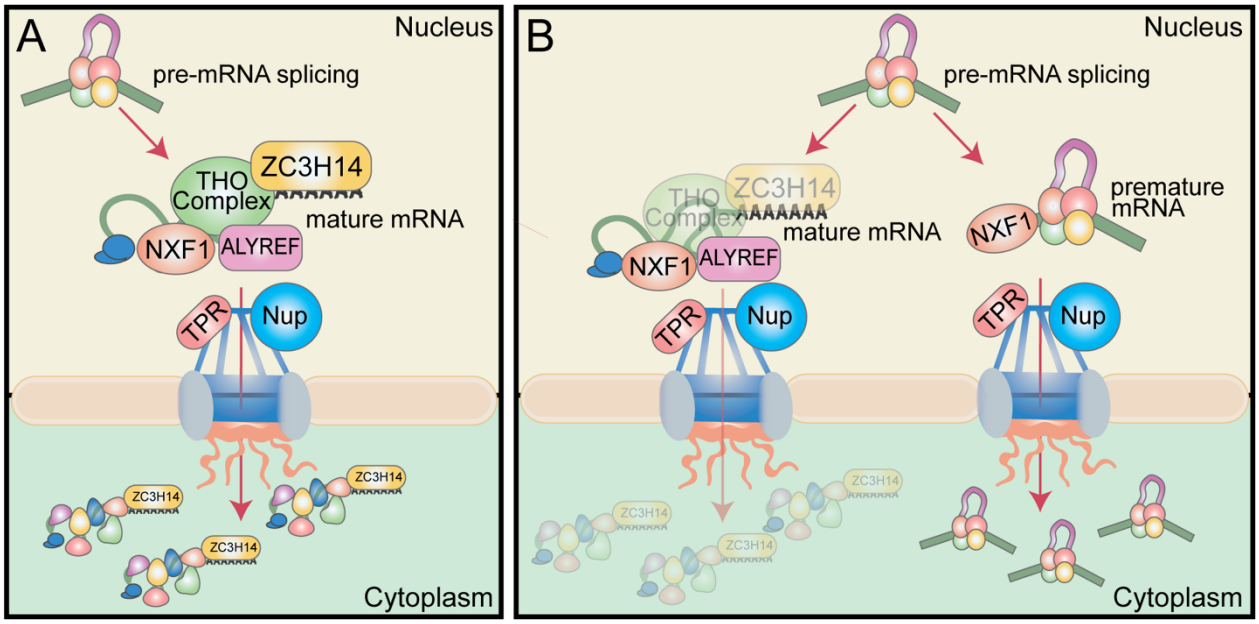



**Figure 3. 9: Model**

Nuclear surveillance model for ZC3H14 and the THO complex. (A) In our speculative model, ZC3H14 and the THO complex are recruited to mRNA upon splicing. After binding to the mRNA, ZC3H14 and the THO complex facilitate the subsequent interaction of the mRNA with export factors such as NXF1 and nuclear pore proteins (Nups), such as the nuclear basket Nup, TPR (171). (B) Loss of ZC3H14 or the THO complex causes the uncoupling of mRNA processing from nuclear export. This impairment decreases the amount of mature mRNA exported to the cytoplasm and allows for an increase in the export of intron-containing premature mRNA. Loss of ZC3H14 or the THO complex allows the pre-mRNA to bypass further processing steps and improperly gain access to the mRNA export machinery.

# Chapter 4: Discussion ~ Summary of Presented Studies and Further Perspectives About the

# Function of ZC3H14

**4. 1 Summary of presented studies:**

We began this study by using a proteomic approach to define the protein interaction network of ZC3H14 and provide insight into the function of ZC3H14. Using the rationale that proteins that interact often have similar functions, we purified ZC3H14 from mouse brain lysate and used mass spectrometry analysis to identify proteins that associate with ZC3H14. This approach led to the creation of an interaction network of the proteins that associate with ZC3H14 in the brain. Our analysis revealed that ZC3H14 interacts with a variety of mRNA processing factors, metabolic enzymes, and proteins of unknown function. We probed deeper into understanding the nature of these interactions by examining whether ZC3H14 interacts with these factors in an RNA-dependent manner. Using RNase to degrade the accessible RNA in the lysate, we determined which of the ZC3H14 protein interactions were mediated through RNA. These approaches led to a detailed interaction network of proteins that both physically interact with ZC3H14 and proteins that interact with ZC3H14 through association with RNA, allowing us to begin to define the molecular functions of ZC3H14.

One group of proteins that emerged as abundant interactors with ZC3H14 was a group of proteins that comprise the mRNA processing complex, THO. Components of the THO complex were amongst the top proteins that interact with ZC3H14 based on spectral count abundance and components of the THO complex were enriched in our gene ontology (GO) analysis of the factors that interact with ZC3H14 (Chapter 3). Interesting, like *ZC3H14*, mutations in genes encoding components of the THO complex are linked to intellectual disability and brain dysfunction despite ubiquitous expression of the proteins

(136,137) (*THOC2*, OMIM:300395) (*THOC6*, OMIM:612733). The enrichment for the THO complex in ZC3H14 immunoprecipitations as well as the links between the THO complex and brain disorders opened the possibility that ZC3H14 and the THO complex could have overlapping functions. As the function of the THO complex is to link early mRNA processing steps to nuclear export (41), ZC3H14 could also be implicated coordinating mRNA processing events. These findings strengthen the rationale to investigate the relationship between ZC3H14 and the THO complex to learn more about the function of ZC3H14.

As described in Chapter 3, we found that ZC3H14 and the THO complex physically interact in the nucleus, confirming our mass spec analysis. In examining the effect of ZC3H14 and the THO complex on mRNA processing, we found that ZC3H14 and the THO complex share two RNA targets and coordinately regulate the processing of these transcripts. Loss of either ZC3H14 or components of the THO complex causes a decrease in steady-state levels of the mature target transcripts and an accumulation of pre-mRNA of target transcripts in the cytoplasm. Furthermore, we show that this accumulation of cytoplasmic pre-mRNA is due to improper export of the pre-mRNA from the nucleus by premature association with the mRNA export factor, NXF1. Through our work, we further established a role for ZC3H14 and the THO complex in mRNA processing and proposed a model where these proteins play a role in surveilling the quality of the mRNA in the nucleus prior to nuclear export.

As we used GO ontology analysis to better understand the relationship between the proteins that interact with ZC3H14, we reexamined the individual proteins we identified co-purifying with ZC3H14 because this analysis ignores proteins that do not group into

specific categories. By prioritizing the proteins that interact with ZC3H14 based on the abundance of each protein peptide identified in the mass spectrometry analysis, we found that ZC3H14 does not just interact with RNA regulatory factors. The most abundant protein found to interact with ZC3H14 is the adenosine deaminase, AMPD2. Finding a metabolic enzyme so tightly associated with ZC3H14 was surprising but this interaction supports current models that metabolic enzymes can acts as RNA-binding proteins and/or be part of mRNPs (123). Like ZC3H14, AMPD2 is a ubiquitously expressed protein that when mutated causes a brain specific phenotype (120,124) (OMID: 102771). *AMPD2* patients display a form of pontocerebellar hypoplasia with affected patients presenting with developmental delays and neuronal defects (124). This finding supports the idea that when proteins that form complexes are dysregulated they can lead to similar pathologies (172). In Chapter 2, we demonstrate that ZC3H14 and AMPD2 physically interact in both the nucleus and the cytoplasm, suggesting that these proteins are tightly coupled as they function in both cellular compartments. This finding provides new insight into the function of ZC3H14 and the interaction between ZC3H14 and AMPD2 presents the possibility that more metabolic enzymes that are often overlooked in proteomic data sets of RNA-binding proteins have biological relevance in RNA processing.

To fully understand the function of ZC3H14, we considered that different ZC3H14 isoforms could have distinct functions and interactions of the ZC3H14 protein isoforms. As *ZC3H14* is alternatively spliced to produce at least four distinct protein isoforms (82), there could be specific functions for each isoform. Having the high throughput analyzing power of mass spectrometry allowed us to expand our analysis of the ZC3H14 protein and study isoform-specific differences in ZC3H14. We found that while the majority of the

protein is nuclear, a population of ZC3H14 exists in the cytoplasm (Chapter 2) (97). In fact, the ~70 kDa isoform(s) (Isoform 2/3) of ZC3H14 is/are the primary ZC3H14 isoform(s) in the cytoplasm. In addition, we identified an isoform-specific PTM on the ~100 kDa isoform (Isoform 1) of ZC3H14 (Chapter 2). Isoform 1 contains a phosphorylated serine at amino acid position 516. Interestingly, this region is encoded by exon 12 which is located in the variable splicing region of ZC3H14 and is specific to Isoform 1. As Isoform 1 is not detected in the cytoplasm, the phosphorylation site could be a mechanism by which Isoform 1 is restricted to the nucleus, while Isoforms 2/3 can travel to the cytoplasm. This possibility presents a model where Isoform 1 is the dominant isoform in early mRNA processing steps. As the mRNP undergoes remodeling, Isoform 1 is phosphorylated and released from the transcript. Isoforms 2/3 bind the transcript and facilitate the translocation of the mRNA through the nuclear pore. These analyses have provided interesting observations about the isoform-specific differences between the isoforms of ZC3H14, demonstrating the need to study the isoforms independently to fully understand the multifaceted functions of ZC3H14.

By defining the network of ZC3H14 interacting proteins, we have gained insight into how ZC3H14 functions in the cell. We found that ZC3H14 interacts with a variety of mRNA processing factors, which supports a role for ZC3H14 in the coupling of early nuclear processing steps. The information gained from this study has laid the foundation for future studies that will probe deeper into understanding the relationship between ZC3H14 and splicing factors, components of the THO complex, and other interacting partners. By understanding the relationship between ZC3H14 and associated proteins, we can increase our knowledge about the function of ZC3H14. As more mutations in genes

that encode proteins of unknown function are being found (173), this study serves as a foundational study in seeking to elucidate the function of a protein by defining the complete spectrum of interacting proteins.

**4.2 Conclusions from our presented studies**

Our work sought to understand the function of ZC3H14 in the brain by identifying the regulatory factors that interact with ZC3H14 in murine lysate. Through this study, we have provided information that has improved our understanding about the cellular function of ZC3H14. The identified protein interaction network of ZC3H14 revealed interesting protein factors that links the major function of ZC3H14 to RNA processing. Like ZC3H14, the majority of identified interacting proteins are ubiquitously expressed, making it likely that these interactions occur in various tissues. The enrichment of factors associated with pre-mRNA splicing and mRNA export from the nucleus links the function of ZC3H14 to these processes. Through our work, we demonstrated that ZC3H14 is important for proper mRNA processing in the nucleus, as loss of ZC3H14 leads to extended bulk poly(A) tails, decreased steady-state levels of target transcripts, and escape of pre-mRNA from the nucleus to the cytoplasm (Chapter 3). ZC3H14 likely works with the identified factors in various cells to confer proper mRNA processing. However, ZC3H14 has been shown to only affect a subset of transcripts (96) and the transcripts regulated by ZC3H14 could differ within different cell types. The RNA processing defects that occur due to loss of ZC3H14 could determine which cells are more sensitive to loss of ZC3H14, as the expression of critical genes could be compromised in select cells. Understanding cell-specific requirements for ZC3H14 will require further analysis of the exact processing defects in specific transcripts within different cell types.

By examining RNA processing defects that result upon depletion of ZC3H14 protein, we can begin to model the general role of ZC3H14. Recent work with ZC3H14 has demonstrated that this protein is important in the brain, particularly in neurons. Analysis of the expression of the ZC3H14 protein in the brain revealed that ZC3H14 protein levels are higher in neurons compared to glia cells of the hippocampus (80), possibly indicating a more critical requirement for the protein. In agreement with this data, work in *Drosophila* demonstrated that *dNab2,* the *ZC3H14* orthologue is important for neuronal functions as *dNab2* mutant flies have neuronal morphological defects. Furthermore, the reduced vitality of the *Nab2* mutant flies can be rescued by the expression of dNab2 or the 100 kDa isoform of the human ZC3H14 protein specifically in neurons. Moreover, the generation of the *Zc3h14* mutant mouse model allowed us to test whether ZC3H14 is required for proper brain function, revealing that *Zc3h14* mutant mice display impairments in learning and memory. These findings support the argument that ZC3H14 is important for neuronal function and development. Whether ZC3H14 plays a critical role in neurons by regulating the processing of specific transcripts or some other function is yet to be determined. Continuing research is needed to elucidate the exact functional requirements for ZC3H14 in neurons.

### 4.2.1 Linking mutations in *ZC3H14* to intellectual disability

To understand why *ZC3H14* mutant patients are affected by intellectual disability, we need to understand more about how inherited genetic mutations give rise to intellectual disability. Recent large-scale sequencing studies performed on various neurodevelopment diseases have revealed an enrichment of *de novo* mutations in certain genes (174). These analyses have revealed patterns of overlapping affected genes in a variety of

neurodevelopmental diseases. For instance, intellectual disability and Autism Spectrum Disorder (ASD) show a large overlap of loss-of-function *de novo* mutations (175), suggesting that affected genes shared between disorders possibly overlap in biological pathways. Another interesting finding is that mutations in some genes give rise to a wide variety of neurodevelopmental phenotypes. For example, mutations in sodium channel, voltage gated, type II α-subunit (*SCN2A*) have been identified in individuals with intellectual disability, ASD, epilepsy and ataxia (176-178). The different manifestations of neurological phenotypes for mutations within the same gene could be attributed to stochastic processes during development, the difference in genetic background between patients and/or the effects of different mutations (179). Taken together, these findings demonstrate that some basic molecular pathways are shared between all neurodevelopmental disorders, indicating why some genetic mutations can give rise to multiple neurodevelopmental phenotypes while some are more exclusively linked to a particular neurological disorder.

To fully understand how mutations in *ZC3H14* give rise to intellectual disability, we need to establish a better framework that assesses shared characteristics of *de novo* mutations in the *ZC3H14* gene. Recent studies have developed assessment criteria to determine the role of *de novo* mutations in different brain diseases (180,181). For example, a study examining the *de novo* mutations of ASD, found a significant role for loss-of-function mutations in the neurodevelopmental process (181). As of now, the number of individuals with loss-of-function mutations in *ZC3H14* is limited to 6 known individuals (80). To extend this type of assessment to *ZC3H14* mutations, we would need a greater number of individuals to determine whether mutations in *ZC3H14* are

linked to neuronal developmental processes. As a lot of cases of inherited intellectual disability are not genetically characterized (182), the number of individuals with mutations in *ZC3H14* could be a lot greater than the reported cases. Alternatively, the complete loss of ZC3H14 could be lethal, explaining why no other ZC3H14 mutant patients have emerged. This event is likely based on evidence that the functional orthologues of *ZC3H14* in budding yeast (86) and *Drosophila* are essential (80). With the discovery of more patients, we can start to assess the contribution of *de novo* mutations in *ZC3H14* to the intellectual disability phenotypes.

## 4.3 Further perspectives on the function of ZC3H14

### 4.3.1 Role of ZC3H14 in nuclear mRNA processing

As the ZC3H14 protein is primarily localized to the nucleus at steady-state, understanding the nuclear function of the protein is critical. Our work reveals that ZC3H14 interacts with a variety of mRNA processing factors and works with the THO complex to coordinately regulate the processing of transcripts. We demonstrate that ZC3H14 and THO components are needed for proper poly(A) tail length control. We found that ZC3H14 and the THO complex regulate the processing of two transcripts, *Atp5g1* and *Psd95*. Interestingly, loss of ZC3H14 or the THO component THOC1 causes an accumulation of the pre-mRNAs of these transcripts in the cytoplasm due to premature association with the mRNA export factor, NXF1. The defects observed in mRNA processing could stem from defects in splicing that result from loss of ZC3H14 or THOC1. The THO complex interacts with splicing factors (41), and our work reveals that ZC3H14 interacts with splicing factors, strengthening the possibility that these factors are involved in splicing (89) (Chapter 2). The defects observed in the processing of the pre-mRNA upon loss of ZC3H14 or THOC1

could be indicate a requirement for ZC3H14 and the THO complex in splicing. Indeed, recent studies of the ZC3H14 orthologue Nab2 in buddy yeast support a role for this protein in splicing (89) (Chapter 2).

#### 4.3.1.1 How ZC3H14 and THO components can affect poly(A) tail length:

In Chapter 3, we demonstrate that ZC3H14 and THO components are important for proper poly(A) tail length control. ZC3H14 has been shown to bind to polyadenosine RNA tracts (95), making it likely that ZC3H14 is associated with the poly(A) tails of RNAs. The mechanism by which the THO complex binds RNA has not been determined. One potential mechanism is that the THO complex binds to RNAs through an adapter protein, which could be ZC3H14. With ZC3H14 recruiting the THO complex to the poly(A) tail, these proteins could ensure that the poly(A) tail is formed properly before the mRNA is exported from the nucleus. Based on a model of Nab2 function in budding yeast (87,89), ZC3H14 could serve as a molecular ruler for the poly(A) tail by recruiting RNA processing factors such as the RNA exosome to trim the poly(A) tail. In fact, the THO complex has been shown to interact with the RNA exosome in yeast to degrade improperly processed RNAs (66). Together, ZC3H14 and the THO complex could work to recruit the RNA exosome to ensure that the poly(A) tails are the proper length needed for export. Alternatively, ZC3H14 and the THO complex could signal for the RNA exosome to degrade RNAs with extended poly(A) tails. With either model, the loss of ZC3H14 or the THO complex would result in the accumulation of RNAs with extended poly(A) tails.

#### 4.3.1.2 ZC3H14 ensures proper processing of target transcripts:

Our analysis of the effect of ZC3H14 and the THO complex on mRNAs revealed that depletion of ZC3H14 or THO components causes a decrease in the steady-state levels of mature target transcripts. Recent work in budding yeast shows that the depletion of Nab2 leads to the rapid depletion of RNAs mediated by the RNA exosome (90). This finding suggests that loss of Nab2 would either lead to an increase in the misprocessing of RNAs that would be targeted for decay or a failure of Nab2 to protect processed RNAs from degradation. Similarly, the loss of ZC3H14 could cause an increase in mRNA processing defects or a failure to protect the RNA from degradation factors. Our work in Chapter 3 would support the first model where loss of ZC3H14 leads to misprocessed target transcripts. The defective RNAs would then be retained in the nucleus and presumably degraded, resulting in the decreased steady-state levels of the target mRNA seen with depletion of ZC3H14 (Chapter 3). Further characterization of the defects observed upon depletion of ZC3H14 is needed to test this model and elucidate the exact role ZC3H14 plays in ensuring proper mRNA processing.

#### 4.3.1.3 How ZC3H14 prevents the escape of pre-mRNA to the cytoplasm

Another interesting observation is how depletion of ZC3H14 or THOC1 leads to the escape of pre-mRNA out to the cytoplasm. Recent work shows that the THO complex is important for the resolution of R-loops during splicing (168). R-loops are structures that form from base pairing of the nascent RNA emerging from RNA polymerase and the transcribing DNA template, interfering with mRNA processing and possibly causing to DNA damage (170,183). Generally, the splicing out of introns prevents the formation of R-loops but introns with high GC contents are prone to R-loop formation and pose a challenge for the splicing machinery (150,170). Like the THO complex, ZC3H14 could

have a role in the resolution of R-loops during splicing. The pre-mRNAs of the two identified transcripts of ZC3H14 and THO complex, *Atp5g1* and *Psd95,* contain introns with high GC contents (greater than 50%). We analyzed the high GC-content introns of these transcripts by qRT-PCR in Chapter 3 and found that they were retained in the cytoplasmic pre-mRNA. These findings present the possibility that with loss of ZC3H14 or THOC1, the high GC-content introns remain in the transcript and are exported to the cytoplasm.

In Figure 4.1A, we present a model where ZC3H14 and the THO complex contributes to pre-mRNA splicing by preventing the formation of R-loops for introns that have high GC contents. Through the actions of ZC3H14 and the THO complex, the splicing machinery is not inhibited by the formation of R-loops and can excise the intron allowing downstream processing of the mRNA to occur. ZC3H14 and the THO complex then facilitate the transfer of the transcript to the mRNA export machinery and the mature mRNA is exported to the cytoplasm. Conversely, the loss of ZC3H14 or the THO complex allows R-loops of GC-rich introns to form during splicing, preventing the splicing machinery to excise the intron (Figure 4.1B). The formation of the R-loops causes incomplete splicing of the transcript but the intron-containing RNA is still recognized as spliced and gains access to the RNA export machinery. The requirement for ZC3H14 and the THO complex for the resolution of R-loops during splicing makes these factors critical for nuclear processing and prevents the export of improperly processed mRNA.

### 4.3.2 Role of ZC3H14 in the cytoplasm

As a pool of ZC3H14 exists in the cytoplasm (97) (Chapter 3), work has begun to elucidate the function of ZC3H14 in this compartment. To better understand the

cytoplasmic function of ZC3H14, we performed mass spectrometry analysis on the cytoplasmic fraction of ZC3H14 in murine lysate to identify the factors that interact with ZC3H14 (Appendix 1). This work revealed that in the cytoplasm ZC3H14 interacts with a variety of factors including translation machinery and cytoskeletal components. Additional work with ZC3H14 demonstrates that the protein associates with ribosomes in the polysome profile (97). Furthermore, work in *Drosophila* showed that dNab2 interacts with a known regulator of translation dFMR (184) and regulates the expression of a translational reporter (97) Taken together, these data present the possibility that ZC3H14 works with other RNA-binding proteins and the translation machinery to regulate translation. One role of ZC3H14 could be to ensure proper nuclear processing and then regulate the translation of the transcript in the cytoplasm. As dNab2 interacts with dFMR, ZC3H14 could exhibit functions similar to FMRP, where ZC3H14 would negatively regulate the translation of the transcript until it is properly localized in the cytoplasm.

My work has shown that both *Psd94* and *Atp5g1* are transcripts of ZC3H14 and these transcripts are locally translated in neurons (166,185). Taken together, the fact that ZC3H14 interacts with translational machinery and cytoskeletal components in the cytoplasm, associates with ribosomes in the polysomes, and binds to transcripts that require localized translations presents the possibility that ZC3H14 is involved in local translation. As ZC3H14 binds to polyadenosine tracts (93), the protein is likely associated with the poly(A) tails of mRNAs. ZC3H14 is poised to prevent access of the poly(A) tail from the translation machinery. One model for the role of ZC3H14 in the cytoplasm is that ZC3H14 travels with the mRNA from the nucleus and safe guards the poly(A) tail as the mRNA is trafficked to either the axon in the case of *Atp5g1* (166) or the dendrite in the case of *Psd95*

(185). Upon loss of ZC3H14, the poly(A) tail is not protected as the mRNA exits the nucleus and the translation initiation machinery can gain access to the transcript. The failure of the transcripts to be targeted to the correct region of the neuron prevents the proteins from being expressed where they are needed, resulting in neuronal defects. As ATP5G1 is important for axonal maintenance (166) and PSD95 is important for the integrity of the post synaptic density (185), loss of the expression of these proteins in the proper areas would have devastating effects on the function of the neuron and could underlie the molecular phenotypes of the *ZC3H14* patients. Thus, this model prevents one possibility for how loss of ZC3H14 preferentially affects the brain.

### 4.3.3 The role of ZC3H14 in non-canonical mRNPs

The association between ZC3H14 and AMPD2 presents the possibility that these proteins function together. AMPD2 functions in purine metabolism by maintaining guanosine pools and loss of AMPD2 leads to increased levels of adenosine (120). As ZC3H14 is involved in nuclear mRNA processing, AMPD2 could associate with ZC3H14 to regulate nucleotide levels at sites of mRNA processing. Together as an mRNP, ZC3H14 could target AMPD2 to sites of RNA processing where maintenance of nucleotide pools is needed, such as sites of transcription, polyadenylation or RNA degradation. As loss of ZC3H14 causes extended bulk poly(A) tails (94), ZC3H14 could normally direct AMPD2 to sites of polyadenylation to control and regulate the pool of adenosines for poly(A) polymerase. As growing evidence suggests that elevated adenosine levels in the cell are toxic and can lead to neurodegeneration (186), misregulation of AMPD2 due to loss of ZC3H14 could underlie the phenotypes of the *ZC3H14* patients. Further examination of

the relationship between ZC3H14 and AMPD2 is needed to uncover how these factors could work together to influence mRNA processing.

### 4.4 Future research in examining the function of ZC3H14

The advancements in the understanding of ZC3H14 from the time of this study has positioned us to carry out more detailed analysis to probe the function of ZC3H14. My work has defined the protein interaction network of ZC3H14 in the brain. Characterization of the *Zc3h14* mutant mouse model has revealed that the mutant mice have defects in working memory and changes in the steady-state levels of hippocampal proteins. Furthermore, work in *Drosophila* with dNab2 and follow up studies with ZC3H14 have revealed that in addition to RNA processing, these factors are possibly involved with regulating the translation of target transcripts (97). In addition to the established role in regulation of poly(A) tail length, my work has shown that ZC3H14 interacts with splicing factors and works with mRNA processing complexes such as the THO complex to coordinate nuclear mRNA processing. Taken together, we have stronger evidence that ZC3H14 plays a critical role in mRNA processing, a well characterized list of proteins that interact with ZC3H14, and evidence that ZC3H14 is involved in regulating the translation of target transcripts. More work is needed to further characterize the exact RNA processing defects that arise from loss of ZC3H14 and the function of ZC3H14 both in the nucleus and the cytoplasm. In the next section, we will discuss ways in which we can continue to assess the function of ZC3H14.

#### 4.4.1 Characterizing the mRNA processing defects resulting from loss of ZC3H14

My work provides insight into the function of ZC3H14 by discovering RNA processing defects that result from loss of ZC3H14. Previous work showed that loss of ZC3H14 causes extended bulk poly(A) tails (94) and that ZC3H14 affects the steady-state levels of a target transcript, *Atp5g1 (96)*. My work shows that loss of ZC3H14 causes a decrease in the steady-state levels of two additional target transcripts *Psd95* and *Mapt* and that loss of ZC3H14 causes an accumulation of the pre-mRNA of these transcripts in the nucleus due to premature export by NXF1. Establishing the mRNA processing defects that occur due to loss of ZC3H14 positions us to examine these defects globally by performing targeted RNA-seq experiments on ZC3H14-depleted neuronal cells or ZC3H14 knockout mouse tissue. As loss of ZC3H14 causes extended poly(A) tails, we can use new advances in RNA sequencing to sequencing to assessing changes in poly(A) tail length for specific transcripts (187). To complement this approach, we can examine changes in RNA steady-state levels and changes in splicing of mRNAs upon loss of ZC3H14. As recent work shows that polyadenylation factors can influence 3-end formation (188), we can examine whether changes in the 3'end of transcripts occur upon ZC3H14 depletion. Lastly, we can expand upon our finding of improper nuclear export in Chapter 3 by performing RNA-seq on fractionated samples. We can examine whether cytoplasmic accumulation of more pre-mRNAs occurs upon loss of ZC3H14. By using these high-throughput sequencing techniques, we can gain a greater understanding of the exact defects that occur on an RNA transcript with the loss of ZC3H14.

### 4.4.2 Testing whether ZC and R- loops

In Chapter 3, we postulated that ZC3H14 and the THO complex could be involved in the resolution of DNA:RNA hybrids or R-loops during splicing. To test this model, we

can examine whether transcripts are more prone to R-loop formation upon loss of ZC3H14 by using a newly developed monoclonal antibody, S9.6 (189) to detect R-loops. Using the antibody, we can perform immunostaining on cells that have been depleted of ZC3H14 or primary cultures from the *Zc3h14* knockout mouse. The S9.6 antibody has also been used in immunoprecipitation experiments to purifying and sequence R-loops in yeast (190). We can take advantage of this method and purify R-loops from ZC3H14-depleted cells and assess whether any of the ZC3H14 target transcripts (*Atp5g1*, *Mapt* or *Psd95*) are more enriched with the R-loop purification from ZC3H14-depleted cells. These approaches will allow for us to determine if ZC3H14 contributes to R-loop resolution and identify a pool of transcripts that are prone to R-loop formation with loss ZC3H14.

### 4.4.3 Examining the requirement for ZC3H14 in local translation

We can test the requirement of ZC3H14 in local translation by using a protein tagging method to measure the effect of ZC3H14 on the rate of protein synthesis. In neurons, RNA-binding proteins regulate spatial and temporal control of gene expression to facilitate local translation (191). As ZC3H14 is localized to the processes of primary hippocampal neurons (97), we can assess whether ZC3H14 mediates translation in these localized areas. We will examine compartmentalized protein synthesis in cultured primary hippocampal neurons obtained from wild type and *Zc3h14* knockout mice, using Click-it Detection technology (192). Click-it Detection measures protein synthesis by using the chemistry of azide and alkyne groups to metabolically label newly synthesized proteins. Fluorescent intensity of the incorporated biomolecule is used as a direct measure of the rate of protein synthesis, which is detected by microscopy to provide spatial information. With

this approach, we will be able to determine if the loss of ZC3H14 influences the rate of protein expression in the processes of hippocampal neurons.

### 4.4.4 Future of the ZC3H14 mutant mouse model

The generation of the *Zc3h14* knockout mouse has been a great tool to assess the requirement of ZC3H14 in various tissues. Initial characterization of the *Zc3h14* knockout mouse yielded that the mice have working memory impairments (98) which agrees with the brain defects present in the *ZC3H14* mutant patients. Starting with the neuronal defects of the *Zc3h14* mouse, we can examine the molecular phenotypes underlying the memory disorder in the mutant mice. First, we can determine whether the neurons of the *Zc3h14* mutant mouse develop properly by observing the morphology of the neurons in the mutant mice. Using Golgi staining, we can examine whether neurons have gross morphological defects. We will prioritize neurons of the hippocampus as the mice have memory defects. In addition to our in vivo approaches, we can take advantage of the ability to culture and isolate primary hippocampal neurons. We can assess whether neurons from the mutant mice have any growth and developmental defects. To support our claim that ZC3H14 is important for neuronal development, we can express ZC3H14 protein in the neurons from the knockout mouse and assess whether there is rescue of the morphological phenotypes. These studies will allow for us to test the requirement for ZC3H14 in neuronal development and maintenance.

#### 4.4.4.1 Examining the *Zc3h14* mutant mouse along with the *Thoc1* mutant mouse

As our work in Chapter 3 demonstrated a physical and functional interaction between ZC3H14 and the THO complex, we can exploit mouse genetics to further examine

the relationship between ZC3H14 and the THO complex. Recently, a *Thoc1* mutant mouse was generated that has reduced expression of the *Thoc1* gene (193). The initial study of this mouse has no observable phenotypes; however, the studies of this mouse did not extend to analysis of molecular phenotypes. As loss of ZC3H14 and THOC1 share phenotypes in N2a cells, we can observe the *Thoc1* mutant mouse for neuronal or learning and memory defects. We can expand upon or work in N2a cells and test whether the *Zc3h14* and *Thoc1* mutant mice have defects in RNA processing. As simultaneous depletion of ZC3H14 and THO resulted in a rescue of RNA steady-state levels (Chapter 3), we can observe whether crossing the mutant *Zc3h14* mouse and *Thoc1* rescues the behavioral and molecular defects in the *Zc3h14* mice. The combined use of the mouse models will allow for us to further increase our understanding of the functional relationship between ZC3H14 and the THO complex.

#### 4.4.4.2 ZC3H14 and FMRP mouse models

Recent work in *Drosophila* established a network of interactions that link dNab2 to the cytoplasmic control of neuronal mRNAs in conjunction with the fragile X protein ortholog dFMR. dNab2 and dfmr1 genetically interact in control of neurodevelopment and olfactory memory, co-localize in puncta within neuronal processes, and share regulation of target RNAs (97). These findings make the FMRP mouse model an interesting mouse model to study in conjunction with the *Zc3h14* mutant mice. The use of these two mouse models affords us the opportunity to build upon our fly studies and observe the relationship between ZC3H14 and FMRP in the mammalian system and assessed more shared defects between the two mouse models. By crossing the two mouse models, we can gain better insight into the functional relationship between these two factors. We can assess whether

the combined loss of ZC3H14 and FMRP ameliorate or worsen the observed phenotypes in each individual mouse model. The characterization of the double mutant mouse will provide a greater understanding about the relationship between ZC3H14 and FMRP and highlight shared pathways that these proteins participate in.

**4.5 Conclusion and final thoughts.**

Our work sought to define the protein interaction network of ZC3H14 and assess the requirement for ZC3H14 in neuronal mRNA processing to increase our understanding of the cellular functions of ZC3H14 in the brain (Figure 4.2A). We hypothesized that ZC3H14 works cooperatively with specific RNA regulatory factors to ensure proper mRNA processing. Critically, we identified proteins that interact with ZC3H14 and started to elucidate the mechanisms by which ZC3H14 works with these factors to coordinate mRNA processing. We have now established novel connections between ZC3H14, other RNA-processing factors, and metabolic enzymes that uniquely position us to pursue our goal of understanding the molecular function of ZC3H14 and gaining insight into the complex roles of RNA-binding proteins in regulating gene expression (Figure 4.2B).

By understanding the function of ZC3H14, we enhanced the current knowledge about the role of RNA-binding proteins in mRNA processing. Our work suggests that ZC3H14 is involved in mRNA splicing, mRNA export, and mRNA quality control in the nucleus. In defining the function of ZC3H14, we provided insight into all of these processes and increased our understanding of how mRNA processing events are coordinated. In addition, identifying the protein partners of ZC3H14 not only provided information about the function of ZC3H14 but also provided insight into the function of ZC3H14 protein partners. For example, our work in Chapter 3, established a role for the THO complex in

mRNA processing and identified novel mRNA processing defects that result from loss of THO components. Ultimately, through understanding the function of ZC3H14, we gained insight into how RNA-binding proteins interact with one another in the brain and work together to coordinately regulate mRNA processing.

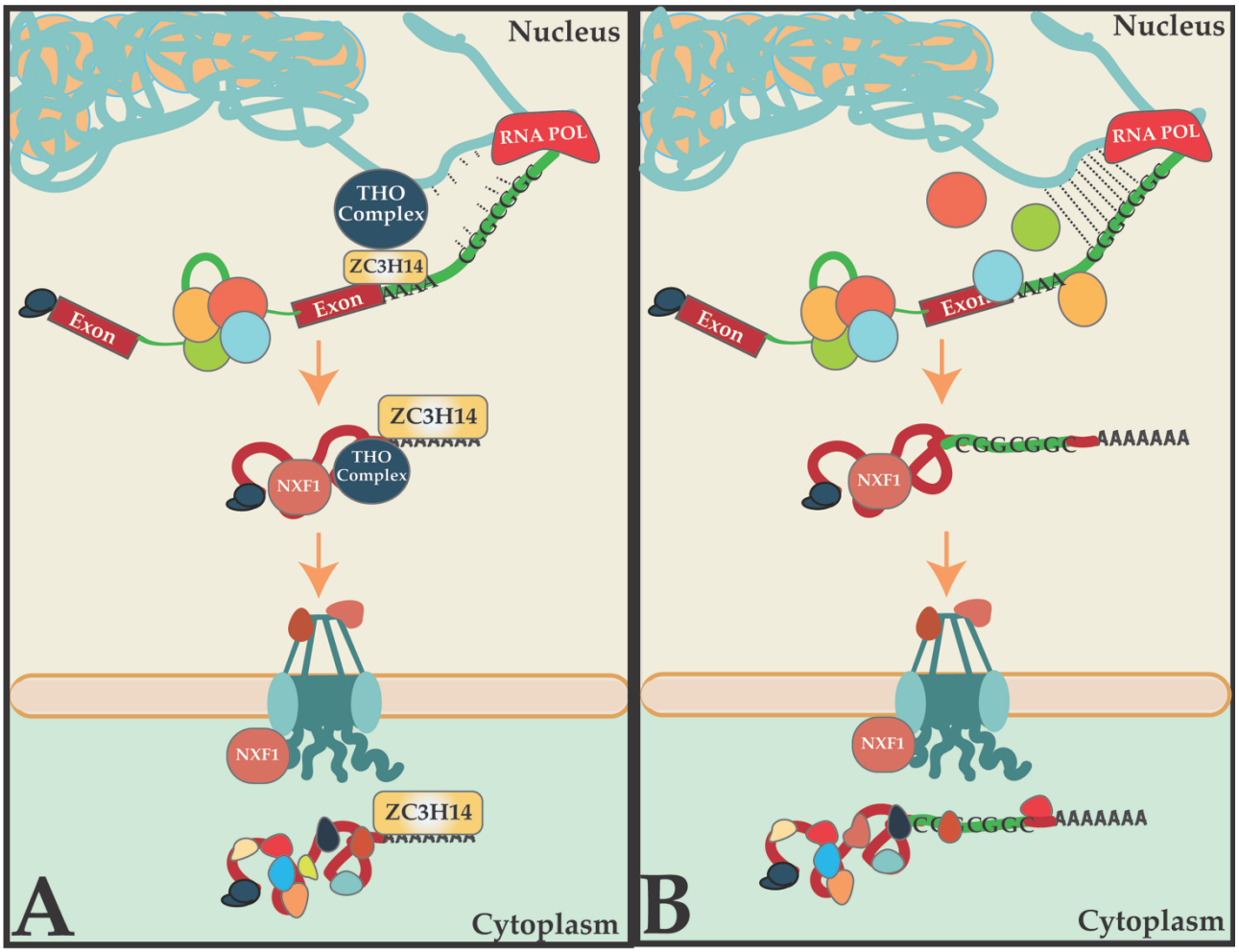



**Figure 4.1: ZC3H14 and the THO complex are responsible for the resolution of R-loops during splicing.**

**A.** ZC3H14 and the THO complex contribute to pre-mRNA splicing by preventing the formation of R-loops of introns that have high GC contents. Through the actions of ZC3H14 and the THO complex, the splicing machinery is not inhibited by the formation of R-loops and can excise the intron allowing for downstream processing of the mRNA to occur. ZC3H14 and the THO complex then facilitate the transfer of the transcript to the mRNA export machinery and the mature mRNA is then exported to the cytoplasm. **B.** The loss of ZC3H14 or the THO complex allows R-loops of GC-rich introns to form during splicing, preventing the splicing machinery to excise the intron. The formation of the R-loops causes incomplete splicing of the transcript but the intron-containing RNA is still recognized as spliced and gains access to the RNA export machinery.

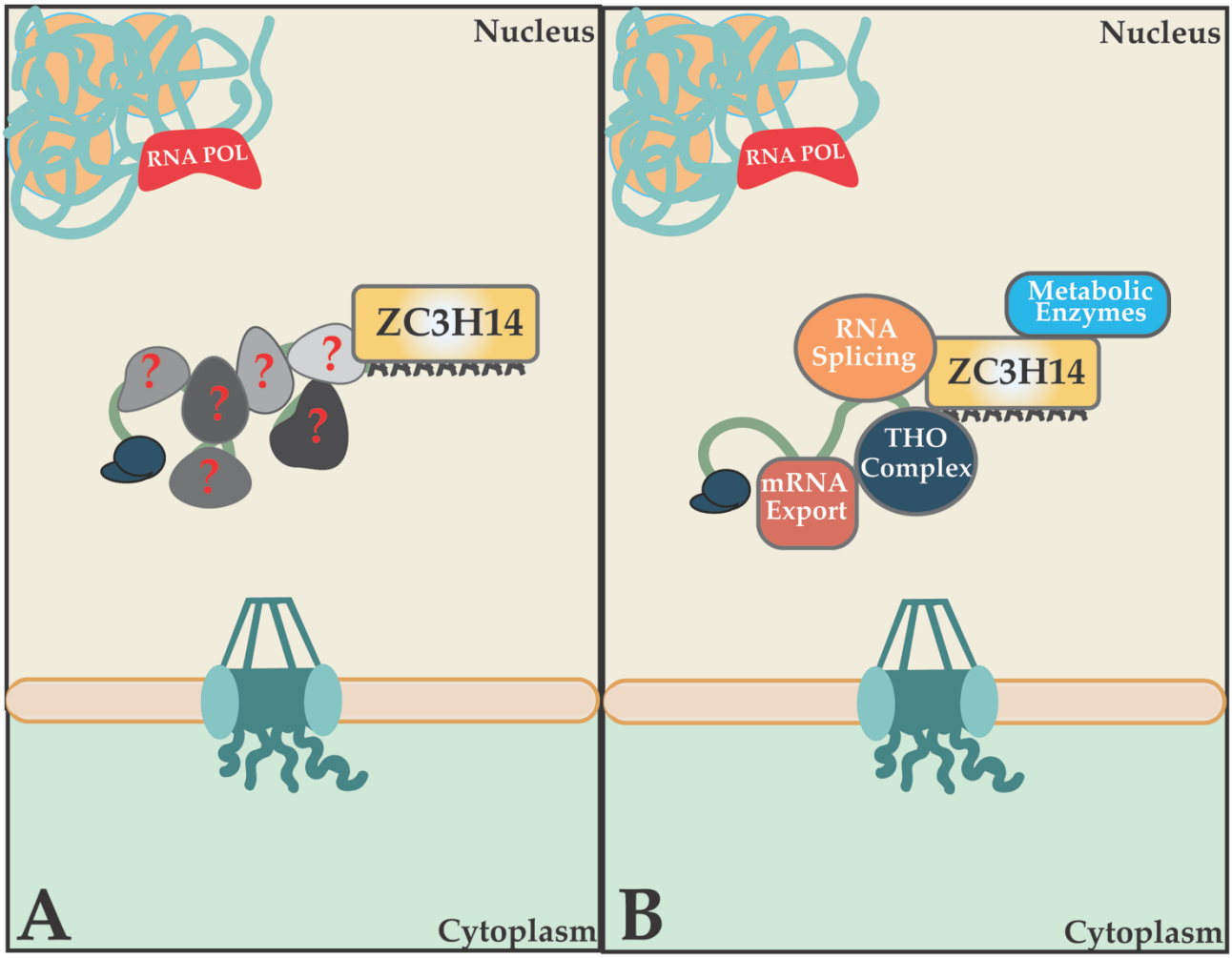



**Figure 4.2: ZC3H14 interacts with splicing factors, mRNA export factors, the THO complex and metabolic enzymes.**
**A.** Our work aimed to define the protein interaction network of ZC3H14 and assess the requirement for ZC3H14 in neuronal mRNA processing to increase our understanding of the cellular functions of ZC3H14 in the brain. **B.** We identified proteins that interact with ZC3H14 and started to elucidate the mechanisms by which ZC3H14 works with these factors to coordinate mRNA processing. We have now established novel connections between ZC3H14, other RNA-processing factors, and metabolic enzymes.

**5.0 References:**


1. Bauer, A.J. and Martin, K.A. (2017) Coordinating Regulation of Gene Expression in Cardiovascular Disease: Interactions between Chromatin Modifiers and Transcription Factors. *Front Cardiovasc Med*, **4**, 19.
2. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev*, **16**, 6-21.
3. Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S. and Darnell, J.E., Jr. (1980) The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell*, **19**, 69-78.
4. Rasmussen, E.B. and Lis, J.T. (1993) In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc Natl Acad Sci U S A*, **90**, 7923-7927.
5. Moteki, S. and Price, D. (2002) Functional coupling of capping and transcription of mRNA. *Mol Cell*, **10**, 599-609.
6. Ghosh, A. and Lima, C.D. (2010) Enzymology of RNA cap synthesis. *Wiley Interdiscip Rev RNA*, **1**, 152-172.
7. Cho, E.J., Takagi, T., Moore, C.R. and Buratowski, S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev*, **11**, 3319-3326.
8. McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S. and Bentley, D.L. (1997) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev*, **11**, 3306-3318.
9. Martinez-Rucobo, F.W., Kohler, R., van de Waterbeemd, M., Heck, A.J., Hemann, M., Herzog, F., Stark, H. and Cramer, P. (2015) Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Mol Cell*, **58**, 1079-1089.
10. Decker, C.J. and Parker, R. (1994) Mechanisms of mRNA degradation in eukaryotes. *Trends Biochem Sci*, **19**, 336-340.
11. van Dijk, E., Cougot, N., Meyer, S., Babajko, S., Wahle, E. and Seraphin, B. (2002) Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J*, **21**, 6915-6924.
12. Brengues, M., Teixeira, D. and Parker, R. (2005) Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science*, **310**, 486-489.
13. Collins, C.A. and Guthrie, C. (2000) The question remains: is the spliceosome a ribozyme? *Nat Struct Biol*, **7**, 850-854.
14. Mironov, A.A., Fickett, J.W. and Gelfand, M.S. (1999) Frequent alternative splicing of human genes. *Genome Res*, **9**, 1288-1293.
15. Pan, S.H., Chao, Y.C., Hung, P.F., Chen, H.Y., Yang, S.C., Chang, Y.L., Wu, C.T., Chang, C.C., Wang, W.L., Chan, W.K. *et al.* (2011) The ability of LCRMP-1 to promote cancer invasion by enhancing filopodia formation is antagonized by CRMP-1. *J Clin Invest*, **121**, 3189-3205.
16. Zhao, J., Hyman, L. and Moore, C. (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev*, **63**, 405-445.

17. Proudfoot, N. and O'Sullivan, J. (2002) Polyadenylation: a tail of two complexes. *Curr Biol*, **12**, R855-857.
18. Sawazaki, R., Imai, S., Yokogawa, M., Hosoda, N., Hoshino, S.I., Mio, M., Mio, K., Shimada, I. and Osawa, M. (2018) Characterization of the multimeric structure of poly(A)-binding protein on a poly(A) tail. *Sci Rep*, **8**, 1455.
19. Jarmolowski, A., Boelens, W.C., Izaurralde, E. and Mattaj, I.W. (1994) Nuclear export of different classes of RNA is mediated by specific factors. *J Cell Biol*, **124**, 627-635.
20. Brodsky, A.S. and Silver, P.A. (2000) Pre-mRNA processing factors are required for nuclear export. *RNA*, **6**, 1737-1749.
21. Terry, L.J. and Wente, S.R. (2007) Nuclear mRNA export requires specific FG nucleoporins for translocation through the nuclear pore complex. *J Cell Biol*, **178**, 1121-1132.
22. Elbarbary, R.A. and Maquat, L.E. (2016) Coupling pre-mRNA splicing and 3' end formation to mRNA export: alternative ways to punch the nuclear export clock. *Genes Dev*, **30**, 487-488.
23. Lee, K.M. and Tarn, W.Y. (2013) Coupling pre-mRNA processing to transcription on the RNA factory assembly line. *RNA Biol*, **10**, 380-390.
24. Manley, J.L. (2002) Nuclear coupling: RNA processing reaches back to transcription. *Nat Struct Biol*, **9**, 790-791.
25. Minvielle-Sebastia, L. and Keller, W. (1999) mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr Opin Cell Biol*, **11**, 352-357.
26. Bentley, D. (1999) Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr Opin Cell Biol*, **11**, 347-351.
27. Greenleaf, A.L. (1993) Positive patches and negative noodles: linking RNA processing to transcription? *Trends Biochem Sci*, **18**, 117-119.
28. Perales, R. and Bentley, D. (2009) "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol Cell*, **36**, 178-191.
29. Egloff, S., Dienstbier, M. and Murphy, S. (2012) Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet*, **28**, 333-341.
30. Johnson, S.A., Kim, H., Erickson, B. and Bentley, D.L. (2011) The export factor Yra1 modulates mRNA 3' end processing. *Nat Struct Mol Biol*, **18**, 1164-1171.
31. Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A. and Varani, G. (2010) Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol*, **17**, 1195-1201.
32. Buratowski, S. (2009) Progression through the RNA polymerase II CTD cycle. *Mol Cell*, **36**, 541-546.
33. Auboeuf, D., Honig, A., Berget, S.M. and O'Malley, B.W. (2002) Coordinate regulation of transcription and splicing by steroid receptor coregulators. *Science*, **298**, 416-419.
34. Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A. and Proudfoot, N.J. (2002) Promoter proximal splice sites enhance transcription. *Genes Dev*, **16**, 2792-2799.
35. Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R. *et al.* (2002) TREX

is a conserved complex coupling transcription with messenger RNA export. *Nature*, **417**, 304-308.
36. Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N. and Moore, M.J. (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, **8**, 426-439.
37. Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M. and Luhrmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci U S A*, **99**, 16719-16724.
38. Chi, B., Wang, Q., Wu, G., Tan, M., Wang, L., Shi, M., Chang, X. and Cheng, H. (2013) Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Res*, **41**, 1294-1306.
39. Zhou, Z., Luo, M.J., Straesser, K., Katahira, J., Hurt, E. and Reed, R. (2000) The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature*, **407**, 401-405.
40. Gatfield, D., Le Hir, H., Schmitt, C., Braun, I.C., Kocher, T., Wilm, M. and Izaurralde, E. (2001) The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. *Curr Biol*, **11**, 1716-1721.
41. Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N. and Reed, R. (2005) Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev*, **19**, 1512-1517.
42. Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z. and Moore, M.J. (2012) The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell*, **151**, 750-764.
43. Reed, R. and Hurt, E. (2002) A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell*, **108**, 523-531.
44. Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499-506.
45. Lei, E.P. and Silver, P.A. (2002) Protein and RNA export from the nucleus. *Dev Cell*, **2**, 261-272.
46. Jensen, T.H. and Rosbash, M. (2003) Co-transcriptional monitoring of mRNP formation. *Nat Struct Biol*, **10**, 10-12.
47. Dower, K. and Rosbash, M. (2002) T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export. *RNA*, **8**, 686-697.
48. Hammell, C.M., Gross, S., Zenklusen, D., Heath, C.V., Stutz, F., Moore, C. and Cole, C.N. (2002) Coupling of termination, 3' processing, and mRNA export. *Mol Cell Biol*, **22**, 6441-6457.
49. Listerman, I., Sapra, A.K. and Neugebauer, K.M. (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol*, **13**, 815-822.
50. Kornblihtt, A.R. (2006) Chromatin, transcript elongation and alternative splicing. *Nat Struct Mol Biol*, **13**, 5-7.
51. Brugiolo, M., Herzel, L. and Neugebauer, K.M. (2013) Counting on co-transcriptional splicing. *F1000Prime Rep*, **5**, 9.

52. Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J Biol Chem*, **270**, 2411-2414.
53. Cooke, C., Hans, H. and Alwine, J.C. (1999) Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol Cell Biol*, **19**, 4971-4979.
54. Niwa, M., Rose, S.D. and Berget, S.M. (1990) In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev*, **4**, 1552-1559.
55. 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.
56. 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.
57. Neugebauer, K.M. (2002) On the importance of being co-transcriptional. *J Cell Sci*, **115**, 3865-3871.
58. Proudfoot, N.J., Furger, A. and Dye, M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501-512.
59. Moore, M.J. and Proudfoot, N.J. (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*, **136**, 688-700.
60. Maquat, L.E. (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol*, **5**, 89-99.
61. Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P. and Reed, R. (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell*, **26**, 867-881.
62. Bourgeois, C.F., Lejeune, F. and Stevenin, J. (2004) Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. *Prog Nucleic Acid Res Mol Biol*, **78**, 37-88.
63. Santamaria, R., Vilageliu, L. and Grinberg, D. (2008) SR proteins and the nonsense-mediated decay mechanism are involved in human GLB1 gene alternative splicing. *BMC Res Notes*, **1**, 137.
64. Zhang, Z. and Krainer, A.R. (2004) Involvement of SR proteins in mRNA surveillance. *Mol Cell*, **16**, 597-607.
65. Libri, D., Dower, K., Boulay, J., Thomsen, R., Rosbash, M. and Jensen, T.H. (2002) Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. *Mol Cell Biol*, **22**, 8254-8266.
66. Zenklusen, D., Vinciguerra, P., Wyss, J.C. and Stutz, F. (2002) Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Molecular and cellular biology*, **22**, 8241-8253.
67. Losh, J.S. and van Hoof, A. (2015) Gateway Arch to the RNA Exosome. *Cell*, **162**, 940-941.
68. Jimeno, S., Luna, R., Garcia-Rubio, M. and Aguilera, A. (2006) Tho1, a novel hnRNP, and Sub2 provide alternative pathways for mRNP biogenesis in yeast THO mutants. *Mol Cell Biol*, **26**, 4387-4398.

69. Fasken, M.B. and Corbett, A.H. (2009) Mechanisms of nuclear mRNA quality control. *RNA Biol*, **6**, 237-241.
70. Palancade, B. and Doye, V. (2008) Sumoylating and desumoylating enzymes at nuclear pores: underpinning their unexpected duties? *Trends Cell Biol*, **18**, 174-183.
71. Peng, S., Zhou, K., Wang, W., Gao, Z., Dong, Y. and Liu, Q. (2014) High-resolution crystal structure reveals a HEPN domain at the C-terminal region of S. cerevisiae RNA endonuclease Swt1. *Biochem Biophys Res Commun*, **453**, 826-832.
72. Skruzny, M., Schneider, C., Racz, A., Weng, J., Tollervey, D. and Hurt, E. (2009) An endoribonuclease functionally linked to perinuclear mRNP quality control associates with the nuclear pore complexes. *PLoS Biol*, **7**, e8.
73. Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D. and Brown, P.O. (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol*, **6**, e255.
74. Gerstberger, S., Hafner, M. and Tuschl, T. (2014) A census of human RNA-binding proteins. *Nat Rev Genet*, **15**, 829-845.
75. Maris, C., Dominguez, C. and Allain, F.H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J*, **272**, 2118-2131.
76. Hashemikhabir, S., Neelamraju, Y. and Janga, S.C. (2015) Database of RNA binding protein expression and disease dynamics (READ DB). *Database (Oxford)*, **2015**, bav072.
77. Gerstberger, S., Hafner, M., Ascano, M. and Tuschl, T. (2014) Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. *Adv Exp Med Biol*, **825**, 1-55.
78. Hayward, J. and Chitty, L.S. (2018) Beyond screening for chromosomal abnormalities: Advances in non-invasive diagnosis of single gene disorders and fetal exome sequencing. *Semin Fetal Neonatal Med*.
79. Gerson, K.D. and O'Brien, B.M. (2018) Cell-Free DNA: Screening for Single-Gene Disorders and Determination of Fetal Rhesus D Genotype. *Obstet Gynecol Clin North Am*, **45**, 27-39.
80. Pak, C., Garshasbi, M., Kahrizi, K., Gross, C., Apponi, L.H., Noto, J.J., Kelly, S.M., Leung, S.W., Tzschach, A., Behjati, F. *et al.* (2011) Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in Drosophila and humans. *Proc Natl Acad Sci U S A*, **108**, 12390-12395.
81. Devapriam, J., Fosker, H., Chester, V., Gangadharan, S., Hiremath, A. and Alexander, R.T. (2018) Characteristics and outcomes of patients with intellectual disability admitted to a specialist inpatient rehabilitation service. *J Intellect Disabil*, 1744629518756698.
82. Leung, S.W., Apponi, L.H., Cornejo, O.E., Kitchen, C.M., Valentini, S.R., Pavlath, G.K., Dunham, C.M. and Corbett, A.H. (2009) Splice variants of the human ZC3H14 gene generate multiple isoforms of a zinc finger polyadenosine RNA binding protein. *Gene*, **439**, 71-78.
83. Villagra, N.T., Bengoechea, R., Vaque, J.P., Llorca, J., Berciano, M.T. and Lafarga, M. (2008) Nuclear compartmentalization and dynamics of the poly(A)-

binding protein nuclear 1 (PABPN1) inclusions in supraoptic neurons under physiological and osmotic stress conditions. *Mol Cell Neurosci*, **37**, 622-633.

84. Marfatia, K.A., Crafton, E.B., Green, D.M. and Corbett, A.H. (2003) Domain analysis of the Saccharomyces cerevisiae heterogeneous nuclear ribonucleoprotein, Nab2p. Dissecting the requirements for Nab2p-facilitated poly(A) RNA export. *J Biol Chem*, **278**, 6731-6740.
85. Green, D.M., Marfatia, K.A., Crafton, E.B., Zhang, X., Cheng, X. and Corbett, A.H. (2002) Nab2p is required for poly(A) RNA export in Saccharomyces cerevisiae and is regulated by arginine methylation via Hmt1p. *J Biol Chem*, **277**, 7752-7760.
86. Anderson, J.T., Wilson, S.M., Datar, K.V. and Swanson, M.S. (1993) NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. *Mol Cell Biol*, **13**, 2730-2741.
87. Hector, R.E., Nykamp, K.R., Dheur, S., Anderson, J.T., Non, P.J., Urbinati, C.R., Wilson, S.M., Minvielle-Sebastia, L. and Swanson, M.S. (2002) Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *EMBO J*, **21**, 1800-1810.
88. Grant, R.P., Marshall, N.J., Yang, J.C., Fasken, M.B., Kelly, S.M., Harreman, M.T., Neuhaus, D., Corbett, A.H. and Stewart, M. (2008) Structure of the N-terminal Mlp1-binding domain of the Saccharomyces cerevisiae mRNA-binding protein, Nab2. *J Mol Biol*, **376**, 1048-1059.
89. 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*.
90. Schmid, M., Olszewski, P., Pelechano, V., Gupta, I., Steinmetz, L.M. and Jensen, T.H. (2015) The Nuclear PolyA-Binding Protein Nab2p Is Essential for mRNA Production. *Cell Rep*, **12**, 128-139.
91. Kelly, S.M., Bienkowski, R., Banerjee, A., Melicharek, D.J., Brewer, Z.A., Marenda, D.R., Corbett, A.H. and Moberg, K.H. (2016) The Drosophila ortholog of the Zc3h14 RNA binding protein acts within neurons to pattern axon projection in the developing brain. *Dev Neurobiol*, **76**, 93-106.
92. Kelly, S., Leung, S.W., Pak, C., Banerjee, A., Moberg, K. and Corbett, A.H. (2013) A Conserved Role for the Zinc Finger Polyadenosine RNA Binding Protien, ZC3H14, in Control of RNA Poly(A) Tail Length *RNA Biol*, **Submitted**.
93. Kelly, S.M., Pabit, S.A., Kitchen, C.M., Guo, P., Marfatia, K.A., Murphy, T.J., Corbett, A.H. and Berland, K.M. (2007) Recognition of polyadenosine RNA by zinc finger proteins. *Proc Natl Acad Sci U S A*, **104**, 12306-12311.
94. 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.
95. Kelly, S.M., Leung, S.W., Apponi, L.H., Bramley, A.M., Tran, E.J., Chekanova, J.A., Wente, S.R. and Corbett, A.H. (2010) Recognition of polyadenosine RNA by the zinc finger domain of nuclear poly(A) RNA-binding protein 2 (Nab2) is required for correct mRNA 3'-end formation. *The Journal of biological chemistry*, **285**, 26022-26032.

96. Wigington, C.P., Morris, K.J., Newman, L.E. and Corbett, A.H. (2016) The Polyadenosine RNA-binding Protein, Zinc Finger Cys3His Protein 14 (ZC3H14), Regulates the Pre-mRNA Processing of a Key ATP Synthase Subunit mRNA. *J Biol Chem*, **291**, 22442-22459.
97. Bienkowski, R.S., Banerjee, A., Rounds, J.C., Rha, J., Omotade, O.F., Gross, C., Morris, K.J., Leung, S.W., Pak, C., Jones, S.K. *et al.* (2017) The Conserved, Disease-Associated RNA Binding Protein dNab2 Interacts with the Fragile X Protein Ortholog in Drosophila Neurons. *Cell Rep*, **20**, 1372-1384.
98. Rha, J., Jones, S.K., Fidler, J., Banerjee, A., Leung, S.W., Morris, K.J., Wong, J.C., Inglis, G.A.S., Shapiro, L., Deng, Q. *et al.* (2017) The RNA-binding Protein, ZC3H14, is Required for Proper Poly(A) Tail Length Control, Expression of Synaptic Proteins, and Brain Function in Mice. *Hum Mol Genet*.
99. Vazquez, A., Flammini, A., Maritan, A. and Vespignani, A. (2003) Global protein function prediction from protein-protein interaction networks. *Nat Biotechnol*, **21**, 697-700.
100. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, **417**, 399-403.
101. Schwikowski, B., Uetz, P. and Fields, S. (2000) A network of protein-protein interactions in yeast. *Nat Biotechnol*, **18**, 1257-1261.
102. Hishigaki, H., Nakai, K., Ono, T., Tanigami, A. and Takagi, T. (2001) Assessment of prediction accuracy of protein function from protein--protein interaction data. *Yeast*, **18**, 523-531.
103. Mayer, M.L. and Hieter, P. (2000) Protein networks-built by association. *Nat Biotechnol*, **18**, 1242-1243.
104. Dunham, W.H., Mullin, M. and Gingras, A.C. (2012) Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics*, **12**, 1576-1590.
105. Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J. and Aebersold, R. (2011) The quantitative proteome of a human cell line. *Mol Syst Biol*, **7**, 549.
106. Mann, M., Kulak, N.A., Nagaraj, N. and Cox, J. (2013) The coming age of complete, accurate, and ubiquitous proteomes. *Mol Cell*, **49**, 583-590.
107. Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S. and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol*, **7**, 548.
108. Azimifar, S.B., Nagaraj, N., Cox, J. and Mann, M. (2014) Cell-type-resolved quantitative proteomics of murine liver. *Cell Metab*, **20**, 1076-1087.
109. Glisovic, T., Bachorik, J.L., Yong, J. and Dreyfuss, G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett*, **582**, 1977-1986.
110. Jiang, Q., Hu, Y. and Liu, G. (2016) Association of Alzheimer Disease Susceptibility Variants and Gene Expression in the Human Brain. *JAMA Neurol*, **73**, 1255.
111. Swinburne, I.A., Meyer, C.A., Liu, X.S., Silver, P.A. and Brodsky, A.S. (2006) Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription. *Genome Res*, **16**, 912-921.

112. Lukong, K.E., Chang, K.W., Khandjian, E.W. and Richard, S. (2008) RNA-binding proteins in human genetic disease. *Trends Genet*, **24**, 416-425.
113. Apponi, L.H., Leung, S.W., Williams, K.R., Valentini, S.R., Corbett, A.H. and Pavlath, G.K. (2010) Loss of nuclear poly(A)-binding protein 1 causes defects in myogenesis and mRNA biogenesis. *Hum Mol Genet*, **19**, 1058-1065.
114. Dammer, E.B., Duong, D.M., Diner, I., Gearing, M., Feng, Y., Lah, J.J., Levey, A.I. and Seyfried, N.T. (2013) Neuron enriched nuclear proteome isolated from human brain. *J Proteome Res*, **12**, 3193-3206.
115. Herskowitz, J.H., Seyfried, N.T., Gearing, M., Kahn, R.A., Peng, J., Levey, A.I. and Lah, J.J. (2011) Rho kinase II phosphorylation of the lipoprotein receptor LR11/SORLA alters amyloid-beta production. *J Biol Chem*, **286**, 6117-6127.
116. Zambon, A.C., Gaj, S., Ho, I., Hanspers, K., Vranizan, K., Evelo, C.T., Conklin, B.R., Pico, A.R. and Salomonis, N. (2012) GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics*, **28**, 2209-2210.
117. Guillemin, I., Becker, M., Ociepka, K., Friauf, E. and Nothwang, H.G. (2005) A subcellular prefractionation protocol for minute amounts of mammalian cell cultures and tissue. *Proteomics*, **5**, 35-45.
118. Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A.N., Heck, A.J. and Mohammed, S. (2013) Toward a comprehensive characterization of a human cancer cell phosphoproteome. *J Proteome Res*, **12**, 260-271.
119. Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P. and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, **127**, 635-648.
120. Akizu, N., Cantagrel, V., Schroth, J., Cai, N., Vaux, K., McCloskey, D., Naviaux, R.K., Van Vleet, J., Fenstermaker, A.G., Silhavy, J.L. *et al.* (2013) AMPD2 regulates GTP synthesis and is mutated in a potentially treatable neurodegenerative brainstem disorder. *Cell*, **154**, 505-517.
121. Pichon, X., Wilson, L.A., Stoneley, M., Bastide, A., King, H.A., Somers, J. and Willis, A.E. (2012) RNA binding protein/RNA element interactions and the control of translation. *Curr Protein Pept Sci*, **13**, 294-304.
122. Klass, D.M., Scheibe, M., Butter, F., Hogan, G.J., Mann, M. and Brown, P.O. (2013) Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in Saccharomyces cerevisiae. *Genome Res*, **23**, 1028-1038.
123. Castello, A., Hentze, M.W. and Preiss, T. (2015) Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. *Trends Endocrinol Metab*, **26**, 746-757.
124. Accogli, A., Iacomino, M., Pinto, F., Orsini, A., Vari, M.S., Selmi, R., Torella, A., Nigro, V., Minetti, C., Severino, M. *et al.* (2017) Novel AMPD2 mutation in pontocerebellar hypoplasia, dysmorphisms, and teeth abnormalities. *Neurol Genet*, **3**, e179.
125. Stone, T.W., Ceruti, S. and Abbracchio, M.P. (2009) Adenosine receptors and neurological disease: neuroprotection and neurodegeneration. *Handb Exp Pharmacol*, 535-587.

126. Liu-Yesucevitz, L., Bassell, G.J., Gitler, A.D., Hart, A.C., Klann, E., Richter, J.D., Warren, S.T. and Wolozin, B. (2011) Local RNA translation at the synapse and in disease. *J Neurosci*, **31**, 16086-16093.
127. Iacoangeli, A. and Tiedge, H. (2013) Translational control at the synapse: role of RNA regulators. *Trends Biochem Sci*, **38**, 47-55.
128. Siomi, H. and Dreyfuss, G. (1997) RNA-binding proteins as regulators of gene expression. *Curr Opin Genet Dev*, **7**, 345-353.
129. Moore, M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514-1518.
130. Singh, G., Pratt, G., Yeo, G.W. and Moore, M.J. (2015) The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu Rev Biochem*, **84**, 325-354.
131. Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol*, **3**, 195-205.
132. Zhou, H., Mangelsdorf, M., Liu, J., Zhu, L. and Wu, J.Y. (2014) RNA-binding proteins in neurological diseases. *Sci China Life Sci*, **57**, 432-444.
133. Huertas, P., Garcia-Rubio, M.L., Wellinger, R.E., Luna, R. and Aguilera, A. (2006) An hpr1 point mutation that impairs transcription and mRNP biogenesis without increasing recombination. *Mol Cell Biol*, **26**, 7451-7465.
134. Rougemaille, M., Dieppois, G., Kisseleva-Romanova, E., Gudipati, R.K., Lemoine, S., Blugeon, C., Boulay, J., Jensen, T.H., Stutz, F., Devaux, F. *et al.* (2008) THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association. *Cell*, **135**, 308-321.
135. Kumar, R., Corbett, M.A., van Bon, B.W., Woenig, J.A., Weir, L., Douglas, E., Friend, K.L., Gardner, A., Shaw, M., Jolly, L.A. *et al.* (2015) THOC2 Mutations Implicate mRNA-Export Pathway in X-Linked Intellectual Disability. *Am J Hum Genet*, **97**, 302-310.
136. Di Gregorio, E., Bianchi, F.T., Schiavi, A., Chiotto, A.M., Rolando, M., Verdun di Cantogno, L., Grosso, E., Cavalieri, S., Calcia, A., Lacerenza, D. *et al.* (2013) A de novo X;8 translocation creates a PTK2-THOC2 gene fusion with THOC2 expression knockdown in a patient with psychomotor retardation and congenital cerebellar hypoplasia. *J Med Genet*, **50**, 543-551.
137. Beaulieu, C.L., Huang, L., Innes, A.M., Akimenko, M.A., Puffenberger, E.G., Schwartz, C., Jerry, P., Ober, C., Hegele, R.A., McLeod, D.R. *et al.* (2013) Intellectual disability associated with a homozygous missense mutation in THOC6. *Orphanet J Rare Dis*, **8**, 62.
138. Guria, A., Tran, D.D., Ramachandran, S., Koch, A., El Bounkari, O., Dutta, P., Hauser, H. and Tamura, T. (2011) Identification of mRNAs that are spliced but not exported to the cytoplasm in the absence of THOC5 in mouse embryo fibroblasts. *RNA*, **17**, 1048-1056.
139. Wang, L., Miao, Y.L., Zheng, X., Lackford, B., Zhou, B., Han, L., Yao, C., Ward, J.M., Burkholder, A., Lipchina, I. *et al.* (2013) 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.
140. Bonora, M., Bononi, A., De Marchi, E., Giorgi, C., Lebiedzinska, M., Marchi, S., Patergnani, S., Rimessi, A., Suski, J.M., Wojtala, A. *et al.* (2013) Role of the c

subunit of the FO ATP synthase in mitochondrial permeability transition. *Cell Cycle*, **12**, 674-683.
141. Kornau, H.C., Schenker, L.T., Kennedy, M.B. and Seeburg, P.H. (1995) Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. *Science*, **269**, 1737-1740.
142. Donovan, L.E., Dammer, E.B., Duong, D.M., Hanfelt, J.J., Levey, A.I., Seyfried, N.T. and Lah, J.J. (2013) Exploring the potential of the platelet membrane proteome as a source of peripheral biomarkers for Alzheimer's disease. *Alzheimers Res Ther*, **5**, 32.
143. Mao, A.J., Bechberger, J., Lidington, D., Galipeau, J., Laird, D.W. and Naus, C.C. (2000) Neuronal differentiation and growth control of neuro-2a cells after retroviral gene delivery of connexin43. *The Journal of biological chemistry*, **275**, 34407-34414.
144. Chang, C.T., Hautbergue, G.M., Walsh, M.J., Viphakone, N., van Dijk, T.B., Philipsen, S. and Wilson, S.A. (2013) Chtop is a component of the dynamic TREX mRNA export complex. *EMBO J*, **32**, 473-486.
145. Luna, R., Rondon, A.G. and Aguilera, A. (2012) New clues to understand the role of THO and other functionally related factors in mRNP biogenesis. *Biochim Biophys Acta*, **1819**, 514-520.
146. Rondon, A.G., Jimeno, S. and Aguilera, A. (2010) The interface between transcription and mRNP export: from THO to THSC/TREX-2. *Biochim Biophys Acta*, **1799**, 533-538.
147. Rehwinkel, J., Herold, A., Gari, K., Kocher, T., Rode, M., Ciccarelli, F.L., Wilm, M. and Izaurralde, E. (2004) Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. *Nat Struct Mol Biol*, **11**, 558-566.
148. Ren, Y., Schmiege, P. and Blobel, G. (2017) Structural and biochemical analyses of the DEAD-box ATPase Sub2 in association with THO or Yra1. *Elife*, **6**.
149. 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.
150. Bonnet, A., Bretes, H. and Palancade, B. (2015) Nuclear pore components affect distinct stages of intron-containing gene expression. *Nucleic Acids Res*, **43**, 4249-4261.
151. Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A.H., Dargemont, C. and Stutz, F. (2010) Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. *Genes Dev*, **24**, 1927-1938.
152. Saroufim, M.A., Bensidoun, P., Raymond, P., Rahman, S., Krause, M.R., Oeffinger, M. and Zenklusen, D. (2015) The nuclear basket mediates perinuclear mRNA scanning in budding yeast. *J Cell Biol*, **211**, 1131-1140.
153. Viphakone, N., Hautbergue, G.M., Walsh, M., Chang, C.T., Holland, A., Folco, E.G., Reed, R. and Wilson, S.A. (2012) TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nat Commun*, **3**, 1006.
154. Gewartowski, K., Cuellar, J., Dziembowski, A. and Valpuesta, J.M. (2012) The yeast THO complex forms a 5-subunit assembly that directly interacts with active chromatin. *Bioarchitecture*, **2**, 134-137.

155. Poulsen, J.B., Sanderson, L.E., Agerschou, E.D., Dedic, E., Boesen, T. and Brodersen, D.E. (2014) Structural characterization of the Saccharomyces cerevisiae THO complex by small-angle X-ray scattering. *PloS one*, **9**, e103470.
156. Pena, A., Gewartowski, K., Mroczek, S., Cuellar, J., Szykowska, A., Prokop, A., Czarnocki-Cieciura, M., Piwowarski, J., Tous, C., Aguilera, A. *et al.* (2012) Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. *EMBO J*, **31**, 1605-1616.
157. Wansink, D.G., Schul, W., van der Kraan, I., van Steensel, B., van Driel, R. and de Jong, L. (1993) Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J Cell Biol*, **122**, 283-293.
158. Phair, R.D. and Misteli, T. (2000) High mobility of proteins in the mammalian cell nucleus. *Nature*, **404**, 604-609.
159. Le Hir, H., Gatfield, D., Izaurralde, E. and Moore, M.J. (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *The EMBO journal*, **20**, 4987-4997.
160. Le Hir, H., Moore, M.J. and Maquat, L.E. (2000) Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev*, **14**, 1098-1108.
161. Palancade, B., Zuccolo, M., Loeillet, S., Nicolas, A. and Doye, V. (2005) Pml39, a novel protein of the nuclear periphery required for nuclear retention of improper messenger ribonucleoparticles. *Mol Biol Cell*, **16**, 5258-5268.
162. Tascou, S., Kang, T.W., Trappe, R., Engel, W. and Burfeind, P. (2003) Identification and characterization of NIF3L1 BP1, a novel cytoplasmic interaction partner of the NIF3L1 protein. *Biochemical and biophysical research communications*, **309**, 440-448.
163. El Bounkari, O., Guria, A., Klebba-Faerber, S., Claussen, M., Pieler, T., Griffiths, J.R., Whetton, A.D., Koch, A. and Tamura, T. (2009) Nuclear localization of the pre-mRNA associating protein THOC7 depends upon its direct interaction with Fms tyrosine kinase interacting protein (FMIP). *FEBS Lett*, **583**, 13-18.
164. Holt, C.E. and Schuman, E.M. (2013) The central dogma decentralized: new perspectives on RNA function and local translation in neurons. *Neuron*, **80**, 648-657.
165. Davis, H.P. and Squire, L.R. (1984) Protein synthesis and memory: a review. *Psychol Bull*, **96**, 518-559.
166. Natera-Naranjo, O., Kar, A.N., Aschrafi, A., Gervasi, N.M., Macgibeny, M.A., Gioio, A.E. and Kaplan, B.B. (2012) Local translation of ATP synthase subunit 9 mRNA alters ATP levels and the production of ROS in the axon. *Molecular and cellular neurosciences*, **49**, 263-270.
167. Bonnet, A., Grosso, A.R., Elkaoutari, A., Coleno, E., Presle, A., Sridhara, S.C., Janbon, G., Geli, V., de Almeida, S.F. and Palancade, B. (2017) Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability. *Mol Cell*, **67**, 608-621 e606.
168. Gomez-Gonzalez, B., Garcia-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marin, A., Foiani, M. and Aguilera, A. (2011) Genome-wide function of

THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J*, **30**, 3106-3119.
169. Pfeiffer, V., Crittin, J., Grolimund, L. and Lingner, J. (2013) The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *The EMBO journal*, **32**, 2861-2871.
170. Jangi, M., Fleet, C., Cullen, P., Gupta, S.V., Mekhoubad, S., Chiao, E., Allaire, N., Bennett, C.F., Rigo, F., Krainer, A.R. *et al.* (2017) SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. *Proc Natl Acad Sci U S A*, **114**, E2347-E2356.
171. Cordes, V.C., Reidenbach, S., Rackwitz, H.-R. and Franke, W.W. (1997) Identification of protein p270/Tpr as a constitutive component of the nuclear pore complex–attached intranuclear filaments. *The Journal of cell biology*, **136**, 515-529.
172. Aebersold, R. and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature*, **537**, 347-355.
173. Sauna, Z.E. and Kimchi-Sarfaty, C. (2011) Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet*, **12**, 683-691.
174. Hoischen, A., Krumm, N. and Eichler, E.E. (2014) Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nature neuroscience*, **17**, 764-772.
175. Fombonne, E. (2003) Epidemiological surveys of autism and other pervasive developmental disorders: an update. *J Autism Dev Disord*, **33**, 365-382.
176. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N. *et al.* (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*, **380**, 1674-1682.
177. Jiang, Y.H., Yuen, R.K., Jin, X., Wang, M., Chen, N., Wu, X., Ju, J., Mei, J., Shi, Y., He, M. *et al.* (2013) Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *American journal of human genetics*, **93**, 249-263.
178. Liao, Y., Anttonen, A.K., Liukkonen, E., Gaily, E., Maljevic, S., Schubert, S., Bellan-Koch, A., Petrou, S., Ahonen, V.E., Lerche, H. *et al.* (2010) SCN2A mutation associated with neonatal epilepsy, late-onset episodic ataxia, myoclonus, and pain. *Neurology*, **75**, 1454-1458.
179. Vissers, L.E., Gilissen, C. and Veltman, J.A. (2016) Genetic studies in intellectual disability and related disorders. *Nat Rev Genet*, **17**, 9-18.
180. Gratten, J., Visscher, P.M., Mowry, B.J. and Wray, N.R. (2013) Interpreting the role of de novo protein-coding mutations in neuropsychiatric disease. *Nature genetics*, **45**, 234-238.
181. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnstrom, K., Mallick, S., Kirby, A. *et al.* (2014) A framework for the interpretation of de novo mutation in human disease. *Nature genetics*, **46**, 944-950.
182. Mefford, H.C., Batshaw, M.L. and Hoffman, E.P. (2012) Genomics, intellectual disability, and autism. *N Engl J Med*, **366**, 733-743.

183. Bhatia, V., Herrera-Moyano, E., Aguilera, A. and Gomez-Gonzalez, B. (2017) The Role of Replication-Associated Repair Factors on R-Loops. *Genes (Basel)*, **8**.
184. Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W. *et al.* (2011) FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*, **146**, 247-261.
185. Bustos, F.J., Varela-Nallar, L., Campos, M., Henriquez, B., Phillips, M., Opazo, C., Aguayo, L.G., Montecino, M., Constantine-Paton, M., Inestrosa, N.C. *et al.* (2014) PSD95 suppresses dendritic arbor development in mature hippocampal neurons by occluding the clustering of NR2B-NMDA receptors. *PLoS One*, **9**, e94037.
186. Mojsilovic-Petrovic, J., Jeong, G.B., Crocker, A., Arneja, A., David, S., Russell, D.S. and Kalb, R.G. (2006) Protecting motor neurons from toxic insult by antagonism of adenosine A2a and Trk receptors. *J Neurosci*, **26**, 9250-9263.
187. 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. *Molecular cell*, **53**, 1044-1052.
188. Tian, B. and Manley, J.L. (2013) Alternative cleavage and polyadenylation: the long and short of it. *Trends in biochemical sciences*, **38**, 312-320.
189. Konig, F., Schubert, T. and Langst, G. (2017) The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences. *PLoS One*, **12**, e0178875.
190. Wahba, L., Costantino, L., Tan, F.J., Zimmer, A. and Koshland, D. (2016) S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev*, **30**, 1327-1338.
191. 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.
192. Roche, F.K., Marsick, B.M. and Letourneau, P.C. (2009) Protein synthesis in distal axons is not required for growth cone responses to guidance cues. *J Neurosci*, **29**, 638-652.
193. Wang, X., Li, Y., Zhang, X. and Goodrich, D.W. (2007) An allelic series for studying the mouse Thoc1 gene. *Genesis*, **45**, 32-37.