

Studies on RNA Regulation: From Enhancer RNAs to RBBP6 Isoform3

Yaqiong Chen

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

© 2019

Yaqiong Chen

All Rights Reserved

Abstract

Studies on RNA Regulation: From Enhancer RNAs to RBBP6 Isoform3

Yaqiong Chen

This dissertation contains two separate yet interconnected pieces of work, which shed light on the complicated RNA regulatory mechanism. The first part, as the main focus of the thesis, characterizes a large pool of human polyadenylated enhancer RNA under deficient nuclear surveillance conditions, and investigates their metabolism mechanisms. The second part elucidates the dynamic localization mechanism of RBBP6 isoform3, which inhibits pre-mRNA 3' processing by competing with RBBP6 isoform1.

Despite being composed of approximately 3 billion base pairs, only 1 to 2% of the human genome codes for proteins. The non-coding DNA regions can however function as transcription units and generate non-coding RNAs such as enhancer-derived RNAs, or eRNAs, that play crucial roles in gene expression regulation, cell differentiation, development, and diseases. Previous studies have suggested that most eRNAs are transcribed by RNA polymerase II (RNAP II), but not polyadenylated. In Chapter 3, I identify a large fraction of polyadenylated enhancer RNAs under deficient nuclear surveillance conditions via genome-wide analyses, and explore their biogenesis and degradation mechanisms. I find that the Integrator complex plays an important role in polyadenylated eRNA biogenesis, and that their exosome-dependent degradation requires two cofactor complexes containing the RNA helicase Mtr4: the PAXT/PPC complex and the NEXT complex. Additionally, the canonical poly(A) polymerases PAP- α and PAP- γ play a major role in the 3' end processing of pA⁺ eRNA. Finally, I show that under deficient nuclear surveillance

conditions, pA⁺ eRNAs accumulate in the cytoplasm and associate with polysomes, suggesting that at least some might have translation potential.

I also contributed to the discovery of two novel complexes both containing the RNA helicase Mtr4, which is a master player of the nuclear surveillance system. Mtr4 and ZFC3H1 form the PAXT/PPC complex, which facilitates the turnover of polyadenylated nuclear RNAs, including prematurely terminated RNAs (ptRNAs), upstream antisense RNAs (uaRNAs), and eRNAs (see the paper in Appendix II). Mtr4 also associates with NRDE2 to form a complex, functioning in the DNA damage response pathway (see the paper in Appendix III). These works provide additional insights into the complexity and significance of the RNA helicase Mtr4.

In the second part of the thesis, presented in Chapter 4, I studied a polyadenylation factor known as Retinoblastoma-binding protein 6 (RBBP6). RBBP6 was initially identified as a large multidomain protein, interacting with tumor suppressors p53 and Rb. Later, its diverse roles were uncovered in cell cycle progression, apoptosis, nucleic acid metabolism, differentiation, and mRNA processing. RBBP6 protein has four isoforms, among which the shortest isoform, iso3, has only one domain: the DWNN (Domain With No Name) domain. The DWNN domain displays high similarities with ubiquitin, implying its function as a novel ubiquitin-like modifier. However, I show that the DWNN domain is actually not a ubiquitin-like modifier, but is itself ubiquitinated. Moreover, the monoubiquitylation of iso3 can facilitate its localization at chromatin. Additionally, I find that the C-terminal tail of iso3 also plays a role in iso3 chromatin localization, presumably by interacting with other factors of the polyadenylation machinery. Pulldown experiments of iso3 followed by mass spectrometry identified Importin7 as an iso3-interacting factor that assists its cytoplasmic retention. Our results identified novel mechanisms for the dynamic localization of RBBP6 iso3, which shed light on the role of iso3 in mRNA 3' processing and disease.

Table of Contents

Acknowledgements.....	iii
Dedication.....	v
Preface.....	vi
Chapter 1. General Introduction	1
Part I. Enhancer RNAs.....	2
Part II. Mtr4-containing complexes.....	12
Part III. Retinoblastoma-binding protein 6.....	16
References.....	21
Chapter 2. RNA Surveillance by the Nuclear RNA Exosome: Mechanisms and Significance.....	31
Abstract.....	32
1. Introduction.....	32
2. The nuclear RNA exosome: structure and RNA degradation mechanisms.....	33
3. Molecular apparatus for RNA targeting of the exosome in yeasts and humans.....	34
4. Significance of the nuclear RNA exosome in mammalian biological processes.....	39
5. Conclusion and perspectives.....	41
References.....	42
Chapter 3. Identification of a Large Pool of Human Polyadenylated Enhancer RNAs and Characterization of Their Biogenesis and Degradation.....	53
Abstract.....	56
Introduction.....	57
Results.....	61

Discussion.....	72
Materials & Methods.....	80
References.....	88
Figure legends.....	96
Figures.....	103
Chapter 4. Dynamic Subcellular Localization of RBBP6 Isoform3 Is Regulated by Monoubiquitylation and Importin 7.....	119
Abstract.....	121
Introduction.....	122
Results.....	124
Discussion.....	133
Materials & Methods.....	138
References.....	142
Figure Legends.....	148
Figures.....	153
Chapter 5. General discussion and future directions.....	151
Appendix I. Protein sequence of RBBP6N-FLAG and Δ DWNN-FLAG.....	189
Appendix II. An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression.....	190
Appendix III. NRDE-2, the human homolog of fission yeast Nrl1, prevents DNA damage accumulation in human cells.....	229

Acknowledgements

It has been quite a journey for me. I wish I could tell others that my PhD journey was full of happiness and fulfillment, but it's only partially true. It was also a journey of struggle and desperation. I've never agreed with the opinion that you should appreciate pain and suffering, but I do always appreciate those who've enlightened the dark with their kindness, compassion, and generosity, even when they are in the dark themselves as well. Nevertheless, after six years of intense exploration in science, I feel extremely lucky to be able to witness the elegance of science.

There is a book I haven't finished, but liked very much just by its title, *The Unraveling: High Hopes and Missed Opportunities in Iraq*, by Emma Sky. The phrase "high hopes and missed opportunities" is such an accurate description for many, if not most, scientific researches. Those missed opportunities rarely come back in one scientist's life, and nonetheless will be captured by other scientists sooner or later. Six years is such a long time for me, but is just a glimpse of the long history of science.

I truly appreciate the opportunity to explore science in Columbia University with wonderful scientists. I thank Dr. James L. Manley and Dr. Carol Prives for the time they invested in me as my research advisors. I thank my committee members, Dr. Liang Tong, Dr. Richard J. Baer, Dr. Bin Tian for their advices during my PhD study. I thank my colleagues, Dr. Patricia Richard and Dr. Koichi Ogami for their guidance and advices throughout the years, my collaborator Dr. Dinghai Zheng, as well as all the lab members for their generous help and support. I thank my friends Ju Yang, Julyun Oh, Jessica Chan, Yuchao Gao, Linbi Hong, Shuang Feng, Yueh-Lin Tsai, Vitalay Fomin, Nicole Marie Benvenin, Chen Chen, and Yixuan Wu for all the happiness and sadness we've shared during the PhD journey. I also thank Women in Science at

Columbia, a student group where I've met many great people and received tremendous help. I also thank all the great mentors and teachers I met at Columbia. Some of them taught me knowledge to understand the world better, and some showed me the great virtues and moral excellences a human being should strive towards. I thank Dr. Margaret Edsall, Dr. Maya Tolstoy, Dr. Carlos J. Alonso, and Dr. Stephen Sestanovich. I hope I can pass on all the virtues I've learned from them in the future. There are so many people that have helped me and supported me in various ways. I don't think I have enough space to thank each one of them, but I'm always grateful and wish them all the best.

Finally, I thank my parents. Their incomparable love and unconditioned trust have given me enormous strength and courage in this journey. Their grit and integrity are the best gift a child can receive from her parents. Without them, I won't have made it this far.

My PhD study is coming to an end, but my life just starts. Again, I thank you all, and wish you the very best in everything, as you deserve.

Dedication

To my parents.

Preface

This thesis is divided into five chapters. Chapter 1 is a general introduction including the recent progress and the future perspectives of the enhancer RNA research, a brief summary of all Mtr4-containing complexes with a focus on the exosome adaptor complexes, as well as an overview of the functional significance of the RBBP6 protein. Chapter 2 is a published review article I co-authored, summarizes the current knowledge of the nuclear RNA surveillance system and discusses the physiological significance of the nuclear RNA exosome in mammalian biological processes. Chapter 3 is a manuscript about the identification of a large pool of human polyadenylated enhancer RNAs as well as the characterization of their metabolism mechanisms. Chapter 4 is a manuscript, elucidating the dynamic subcellular localization mechanism of RBBP6 isoform3. Chapter 5 provides a general discussion as well as future directions of the works presented in the thesis. Appendix I shows the protein sequences of two RBBP6 derivatives, whose peptide cleavage merits further study, as discussed in Chapter 5. Appendix II and III are two published articles I co-authored, about the discovery and function characterization of the Mtr4/ZFC3H1 (PAXT/PPC) complex and the Mtr4/NRDE2 complex, respectively.

Chapter 1. General Introduction

This Chapter presents a general introduction of three parts: Part I summarizes the discovery and recent progress of enhancer RNA research; Part II introduces the key Mtr4-containing complexes in the nuclear surveillance pathway; those two parts, together with Chapter 2, provide comprehensive background knowledge for the works in Chapter 3. Finally, Part III reviews the functional studies and physiological significance of the RBBP6 protein, whose isoform 3 subcellular localization mechanism is scrutinized in Chapter 4.

Part I. Enhancer RNAs

The advent of deep-sequencing techniques has largely expanded our knowledge about the transcriptome in cell. Numerous nonprotein-coding DNA elements in the genome is found to be transcribed, which create a much larger RNA repertoire than mRNAs. In this thesis, I study an abundant class of unstable RNAs transcribed by RNA polymerase II (RNAP II) from transcriptional enhancers, known as enhancer RNAs, or eRNAs. Here, I briefly summarize our current understanding of eRNAs and their biological functions.

The Characterization of Enhancers

Enhancers were originally defined as DNA elements that activate promoter transcription, characterized by a lack of orientation and distance requirements from target genes (Serfling et al., 1985). The first enhancer was discovered in SV40 viral DNA, as a tandem 72bp DNA fragment. It can enhance the transcription of a reporter gene by ~200-fold, irrespective of its orientation or distance from the gene (Banerji et al., 1981; Moreau et al., 1981). Later, with the remarkable development of next-generation sequencing technology, scientists were able to annotate putative enhancers on a genome-wide scale, based on the classic features of enhancers: DNase I

hypersensitivity; related histone modification such as higher H3K4me1/H3K4me3 ratio and histone acetylation; high binding density of transcriptional co-activators (e.g. CBP/p300) as well as multiple transcription factors (TFs) (reviewed by Calo and Wysocka, 2013; Klefogiannis et al., 2016; Shlyueva et al., 2014). The use of epigenomic markers in enhancer annotation has greatly expanded the pool of putative enhancers in humans, whose number varies between 400,000 to 1 million, almost ten-fold of the number of coding genes (Calo and Wysocka, 2013; Rivera and Ren, 2013; Shlyueva et al., 2014). However, only epigenomic markers alone were far from accurate in terms of annotating functional enhancers. For example, the DNase I hypersensitive sites also enrich in other regulatory DNA regions such as promoters; TFs bind to a variety of regulatory elements besides enhancers (Buecker and Wysocka, 2012); the H3K4me1/H3K4me3 ratio failed to identify highly transcribed enhancers (Core et al., 2014; Dao et al., 2017). I discuss the transcripts derived from enhancers regions, which can be used as another criterion to predict active enhancers in the next section.

It's noteworthy that there is a group of enhancers, named as "super enhancers" particularly interesting. They are defined by unusually high level of TF and mediator binding, and often cluster in close genomic proximity (Whyte et al., 2013). Those enhancers not only boost the expression of target genes, but also exhibit robust eRNA transcription (Hah et al., 2015; Liu et al., 2014). Many important findings about eRNA were revealed through studying super enhancer, which are elaborated below.

Enhancer as non-coding RNA transcription units

The human genome consists of approximately 3 billion base pairs, but only 1 to 2 % of it are coding genes. Nevertheless, non-coding DNA can be transcribed and generates non-coding

RNAs, including for example enhancer-derived RNAs, or eRNAs. Early studies showed sporadic examples of extragenic and intergenic transcription (Ashe et al., 1997; Masternak et al., 2003; Rogan et al., 2004; Tuan et al., 1992). Nevertheless, not until the advent of large-scale transcriptome profiling and RNA polymerase II (RNAP II) ChIP-seq analyses, were scientists able to obtain genome-wide evidence that enhancers are widely transcribed, and generate largely non-polyadenylated and bidirectional eRNAs (De Santa et al., 2010; Kim et al., 2010). Further studies using the cap analysis of gene expression (CAGE) technique, which analyze m7G capped RNA in a transcriptome, showed that eRNAs are generally capped on their 5' end, and estimated the number of eRNA-producing enhancers in humans to be ~40,000-65,000 (Andersson et al., 2014a; Arner et al., 2015). Notably, a recent pan-cancer analysis using the TCGA database found that the maximum number of eRNAs detected in any given cancer type is less than 10,000 (Chen et al., 2018), which is less than 1/4 of the total enhancers defined by CAGE. This implies that despite over 40,000 enhancers can generate transcripts, only a fraction of enhancers is actually active in a given cancer. This is not surprising, given that one typical feature of enhancer expression is its high spatio-temporal and tissue specificity.

Studies have found that eRNAs can serve as a signature of functional enhancers, given the finding that enhancers that generate eRNAs, display a higher affinity of transcriptional co-activators, greater chromatin accessibility, and higher enrichment of active histone markers (Hah et al., 2013; Kim et al., 2010; Zhu et al., 2013). It is possible that some non-transcribing enhancers might produce eRNA, whose level either is too low to be detected, or is degraded rapidly. Intriguingly, the nascent RNA sequencing analyses were able to identify enhancers merely based on their transcription level, instead of epigenetic markers (Core et al., 2014; Henriques et al., 2018).

An intriguing question is: are eRNAs just transcriptional noise after enhancer activation, or do they exert regulatory functions to instruct the activation of promoter? It's possible that at least some eRNAs might be byproducts of unspecific transcription of RNAP II: On the one hand, RNAP II has been found to constantly scans the genome randomly (Struhl, 2007); on the other hand, the high chromatic accessibility of active enhancers allows RNAP II to load conveniently. Moreover, the low expression level of eRNAs can be another evidence of non-productive “scanning” of RNAP II. Nevertheless, a genome-wide analysis of both human and mouse covering a wide range of tissues and biological stimuli, discovered that eRNAs, temporally preceding the mRNAs, respond to the stimuli. In other words, enhancer transcription represents the earliest transcription event, when stimuli or differentiation cues are given (Arner et al., 2015). This result indicates that RNAP II transcriptional machinery might first load on the enhancer regions, rather than promoters, and subsequently produces eRNAs as the earliest responses to stimuli. It is very likely that eRNAs play a role in transcriptional machinery recruiting or promoter activation.

Next, I discuss how enhancer regulates promoters, by looking into the well-established enhancer-promoter looping model.

Enhancer-promoter looping model

A number of studies have shed light on the intriguing architectural and functional similarities between enhancers and promoters, such as enrichment of general TFs (e.g., TATA-binding protein) and serine 5-phosphorylated RNAP II, DNase I hypersensitivity, related histone modification patterns, and bidirectional transcription (Core et al., 2014; Furlong and Levine, 2018; Kim et al., 2010; Kim and Shiekhattar, 2015; Whalen et al., 2016). Those similarities further backed the enhancer-promoter looping model, which has been a widely accepted mechanism for

enhancers to regulate promoters. As shown in Figure 1-1, enhancer and promoter regions are in direct contact with each other. This reflects association of TFs bound to enhancers (Levine, 2010) with factors at the promoter, with this association facilitating gene activation (Bulger and Groudine, 2011; Furlong and Levine, 2018) and allowing bidirectional exchange of components of the transcriptional machinery (Lam et al., 2014; Li et al., 2016). Moreover, a recent study found that some bidirectional enhancer can act as weak promoters, while bidirectional promoters act as strong enhancers during *Drosophila* embryogenesis (Mikhaylichenko et al., 2018). This observation further demonstrated the architectural and functional similarities between promoters and enhancers.

Despite the remarkable similarities between those two DNA elements, we also note that there are distinct sets of TFs preferentially associates with enhancers and promoters (Consortium, 2012; Core et al., 2014). For example, enhancers show preferences to recruit certain lineage-determining and signal-regulated TFs, such as oestrogen receptors (reviewed by Spitz and Furlong, 2012). This echoes the high spatio-temporal specificity of enhancer activation.

Another evidence for the enhancer-promoter interaction is made possible by the chromosome conformation capture (3C) techniques, which analyzes the contact frequencies between selected genomic sites using biochemical crosslinking (de Wit and de Laat, 2012; Denker and de Laat, 2016). Studies using 3C have confirmed the existence of long-range interactions between enhancers and promoters (Lee et al., 2015; Sanyal et al., 2012; Zhang et al., 2013). Nevertheless, we note that the results of 3C and its modified versions are not always consistent with direct observation under microscopy using fluorescence in situ hybridization (Williamson et al., 2014), which might be due to the technical limitations of different methods, and the highly dynamic interaction between enhancers and promoters (Drissen et al., 2004; Vakoc et al., 2005).

Although the mechanism of enhancer-mediated gene activation is not completely resolved, more and more evidence indicate that enhancers activate gene expression by stimulating RNAP II from pausing to productive transcriptional elongation at the target gene promoter by recruiting the positive transcription elongation factor, P-TEFb, a subunit of the super elongation complex (Bradner et al., 2017; Kim and Shiekhattar, 2015; Schaukowitch et al., 2014; Smith and Shilatifard, 2014). The enhancer-promoter looping structure would allow efficient exchange of those factors and facilitate the escape of paused RNAP II.

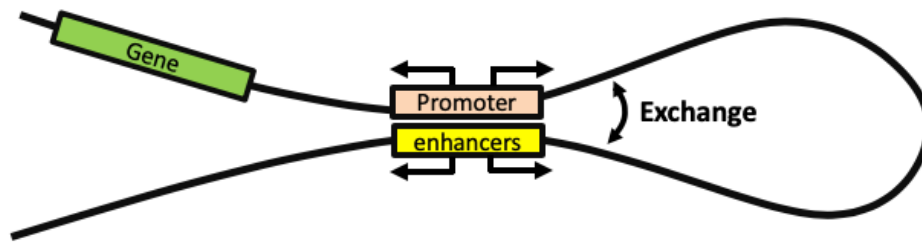


Figure 1-1. Model for enhancer-promoter looping.

Enhancers and promoters are able to engage in long-range interactions by forming “loops”, where components of the transcriptional machinery might be exchanged between enhancers and promoters.

Next, I would like to discuss the potential role of eRNAs in enhancer-promoter loop formation. Several studies have shown that after eRNA knockdown, the corresponding enhancer-promoter looping was impaired (Hsieh et al., 2014; Lai et al., 2013; Li et al., 2013; Pnueli et al., 2015). These studies indicated a potential role of eRNAs in the proper formation of enhancer-promoter loops. Nevertheless, other evidence showed that knockdown of functional eRNAs through chemical inhibition of RNAP II, didn't affect the looping (Hah et al., 2013; Schaukowitch et al., 2014). Collectively, at least some eRNAs contribute to the formation of enhancer-promoter looping. Yet, the necessity of enhancer transcription for loop formation at a larger scale is still inconclusive.

In Chapter 3, I provide additional evidence of the enhancer-promoter looping, by characterizing novel regulators shared between enhancers and promoters, as well as the similar 3' end processing mechanism between polyadenylated eRNAs and mRNAs.

Polyadenylation of eRNAs

After being transcribed from genomic DNA, RNAs that either encode proteins or execute non-coding functions undergo a series of processing steps, such as splicing, 5' capping, and 3' polyadenylation. A natural question would be whether eRNAs also undergo these processing steps? As mentioned above, results of CAGE analyses have showed that eRNAs, similar to mRNAs, are generally 5' capped (Andersson et al., 2014a; Arner et al., 2015). Nevertheless, only ~5% of eRNAs are spliced, whereas ~80% of mRNAs show splicing (Andersson et al., 2014a). This is consistent with the finding that enhancer regions lack U1 splice sites (Andersson et al., 2014b; Core et al., 2014). Lastly, although bioinformatic analyses showed that Polyadenylation Signals (PASs) are highly enriched in enhancer regions (Andersson et al., 2014a; Core et al., 2014), eRNAs have been thought to be large non-polyadenylated: one study failed to detect any eRNA in poly(A)-RNA fractions (Kim et al., 2010), whereas another estimated that approximately 90% of eRNAs were non-polyadenylated (Andersson et al., 2014a). However, all those studies were conducted under intact nuclear surveillance conditions, and thus may have failed to capture fast-turnover nuclear RNA species. Considering the low stability and abundance of eRNAs in cells, we suspected that previous studies might be biased, overlooking a fraction of eRNAs with rapid turnover and extremely short half-lives. Indeed, we identified a large pool of polyadenylated eRNAs under nuclear surveillance system in Chapter 3. Our findings shed light on the 3'

processing mechanism of eRNAs, and provide new insights on the enhancer-promoter looping model.

The biological functions of enhancer RNAs

The functional significance of enhancers as DNA elements has been extensively studied in the past. However, the genome-wide transcription of enhancers was not discovered until 2010 (De Santa et al., 2010; Kim et al., 2010). Here, I summarize the biological functions of enhancer RNAs from four different aspects: R-loop formation, diseases, development, and evolution.

- ***Enhancer RNAs and R-loops***

R-loops, three-stranded nucleic acid structures consisting of an RNA-DNA hybrid and a non-template single-stranded DNA occurs in organisms from bacteria to human. Although studies have found that R-loops are essential for many normal and physiological processes, at the same time R-loops also can cause single- or double-stranded DNA breaks, and subsequently lead to genomic instability and replication stress (Aguilera and Garcia-Muse, 2012; Santos-Pereira and Aguilera, 2015). Since R-loops are frequently formed during transcription in the genome, enhancer transcription might also serve as a source of R-loops. Indeed, a study showed that the accumulation of eRNAs after RNA exosome depletion is accompanied with increased level of R-loops at some specific enhancer regions in mouse stem cells and B cells (Pefanis et al., 2015). Consistent with this, a genome-scale computational analysis also revealed that the preferential colocalization regions for R-loop forming sequence include enhancers, promoters, and gene ends (Kuznetsov et al., 2018). Given that the genome instability caused by R-loops can be a hallmark of cancer (Macheret and Halazonetis, 2015), enhancer transcription might be also involved in tumorigenesis, which is discussed in the following section.

- ***Enhancer RNAs in disease***

As we mentioned above, enhancer transcription might cause the accumulation of R-loops, and subsequently contributes to tumorigenesis (reviewed by Rothschild and Basu, 2017). Indeed, Evidence has been provided by studying activation-induced cytidine deaminase (AID), a B cell-specific protein, which initiates both somatic hypermutations and class switch recombination of immunoglobulin genes in human B cells. Although AID is mainly specific for immunoglobulin genes, it can also damage oncogenes and consequently causes tumorigenesis. Studies have shown that transcribed intragenic super-enhancers facilitate AID mis-targeting and consequently induce genomic instability and tumorigenesis (Laffleur et al., 2017; Meng et al., 2014; Qian et al., 2014).

The dysregulation of enhancer transcription has been reported to play a crucial role in Huntington's disease (HD), a progressive neurodegenerative disease. In HD mouse striatum, a subset of eRNAs with neuronal signatures was downregulated, and some of them also lost their RNAPII binding sites, which in turn led to further downregulation of those eRNAs (Le Gras et al., 2017).

Another interesting finding was provided by computational integration of high-throughput sequencing data. Just like disease-causing variants in coding genes, a great number of genomic variants were found in enhancers, which are linked to human diseases (Epstein, 2009; Hnisz et al., 2013; Mathelier et al., 2015; Parker et al., 2013). Intriguingly, a high enrichment of autoimmune disease-associated single nucleotide polymorphisms (SNPs) was observed in a subset of transcribed enhancers (Farh et al., 2015; Vahedi et al., 2015). Additionally, another systematic genome analysis also found that a great number of disease-associated SNPs are indeed over-represented in transcribed enhancers (Andersson et al., 2014a). However, the precise function of those eRNAs, transcribed from enhancers carrying disease-associated SNPs, remain to be revealed.

- ***Enhancer RNAs in development***

Metazoan development requires precise spatio-temporal regulation of gene expression, which requires the orchestration of a mass of enhancers (Bulger and Groudine, 2010; Rickels and Shilatifard, 2018). Despite the extensive researches on enhancers, as DNA elements, in metazoan development (reviewed by Buecker and Wysocka, 2012; Rickels and Shilatifard, 2018), we have limited knowledge about the functional significance of enhancer transcription in this process. A systematic analysis of enhancer transcripts in a variety of human cell types and tissues has showed that consistent with enhancers, the majority of eRNAs are also cell-type and tissue specific (Andersson et al., 2014a). Additionally, the expression level of eRNAs correlates closely with their target gene expression (Kim et al., 2010). Nowadays, more and more studies have indicated the involvement of eRNAs during differentiation and development processes. For example, three eRNAs were found to control the expression of certain key TFs for myogenesis (Mousavi et al., 2013). In another study, a set of eRNAs was identified in multiple stages of B-cell development and activation (Brazao et al., 2016). Additionally, some erythrocyte-specific eRNAs were revealed to regulate red blood cell maturation (Alvarez-Dominguez et al., 2014). Nevertheless, the ubiquitous mechanism to elucidate the role of all eRNAs in differentiation and development remain obscure.

- ***Enhancer RNAs and evolution***

Enhancers lack sequence conservation, and undergo rapid evolution across mammalian species (Villar et al., 2015). Several studies have indicated that such rapid-evolving non-coding DNA regions help species to gain adaptive advantages by producing *de novo* genes from the perspective of evolution (Tautz and Domazet-Lošo, 2011; Young et al., 2015). Especially considering the vast number of enhancers and the widespread transcription on enhancer regions,

eRNAs can be an ideal reservoir for the translation of *de novo* micropeptide. In Chapter 3, we show that some polyadenylated eRNAs can accumulate in the cytoplasm and associate with monosome/polysomes when the nuclear surveillance pathway is compromised. Our finding implies that some polyadenylated eRNAs might have translation potential to produce *de novo* micropeptides. Nevertheless, further studies are necessary to prove the existence of *de novo* micropeptides from eRNAs.

Part II. Mtr4-containing complexes

The nuclear surveillance system guarantees transcriptional fidelity and efficient RNA turnover by regulating the maturation and degradation of a vast array of RNA species. The central role in nuclear surveillance is played by the nuclear RNA exosome, which degrades both coding and noncoding RNAs produced by all three major RNA polymerases (Mitchell, 2014; Schmid and Jensen, 2018; Zinder and Lima, 2017). Nevertheless, the exosome requires the facilitation of additional protein complexes, to recognize selective RNA substrates. Recent studies have identified a number of nuclear exosome-adaptor complexes, which are largely conserved across species. Remarkably, all of those complexes contain the RNA helicase Mtr4, responsible for unwinding RNA substrates. The published review in Chapter 2 provides a comprehensive overview of our current knowledge of the nuclear surveillance system. Nevertheless, here I briefly summarize the key Mtr4-containing complexes presented in this thesis (Figure 1-2).

NEXT complex

The RNA helicase Mtr4, together with the RNA binding protein RBM7 and the Zn-knuckle protein ZCCHC8, constitute the Nuclear Exosome Targeting (NEXT) complex (Lubas et al., 2011).

The NEXT complex has been found to be involved in the degradation of a large repertoire of RNAs, including replication-dependent histone mRNAs (Andersen et al., 2013), PROMPTs/uaRNAs (Lubas et al., 2011), eRNAs (Lubas et al., 2015). In Chapter 3, I show that the NEXT complex regulates a subset of novel polyadenylated eRNAs, half of which are also regulated by the PAXT/PPC complex mentioned below.

PAXT/PPC complex

The interaction between Mtr4 and ZFC3H1, a Zinc finger protein, was discovered by two independent studies (Meola et al., 2016; Ogami et al., 2017). The previous study in our lab showed Mtr4/ZFC3H1 facilitate the turnover of unstable nuclear RNAs. Depletion of Mtr4/ZFC3H1 led to the cytoplasmic accumulation of prematurely terminated RNAs (ptRNAs) and upstream antisense RNAs (uaRNAs), and consequently polysome association with those RNAs. Therefore, the Mtr4/ZFC3H1 was named as Polysome Protection Complex (PPC). Consistent with this, a recent study showed that ZFC3H1 indeed acts as a central nuclear retention factor for nuclear polyadenylated transcripts. Depletion of ZFC3H1 resulted in nuclear export of its polyadenylated RNA substrates (Silla et al., 2018). I contributed to the discovery of PPC complex, and the paper is available in Appendix II.

Meanwhile, another group also identified the interaction between Mtr4 and ZFC3H1. Besides, their study showed that PABPN1 also interacts with Mtr4/ZFC3H1 complex, although this interaction is more transient and likely to be RNA dependent (Meola et al., 2016). PABPN1 has been found to promote the exosome-mediated degradation of nuclear polyadenylated RNAs (Beaulieu et al., 2012; Bresson and Conrad, 2013; Bresson et al., 2015), and share a subgroup of RNA substrates with Mtr4/ZFC3H1 (Meola et al., 2016; Ogami et al., 2017). Given that this

complex preferentially regulates transcripts with long poly(A) tails, Mtr4/ZFC3H1/PABPN1 was referred as Poly(A) Exosome Targeting (PAXT) complex (Meola et al., 2016).

In Chapter 3, I show that depletion of PAXT/PPC complex results in substantial accumulation of polyadenylated eRNAs, and its “polysome protection” role also applies to those polyadenylated eRNAs. Additionally, I reveal a novel role of ZFC3H1 in RNA biogenesis, besides its known function in RNA degradation and nuclear retention.

Mtr4/NRDE2 complex

Mtr4 interacts with NRDE2, a protein mainly localizing in nuclear speckles, in human cells (Jiao et al., 2019; Lubas et al., 2011; Ogami et al., 2017; Richard et al., 2018; Wang et al., 2019). Such interaction is highly conserved, as studies showed that the Mtr4 homolog Mtl1 and the NRDE2 homolog Nrl1 in *S. pombe* also interacts with each other (Lee et al., 2013; Yamanaka et al., 2013). A published work in our lab discovered that the both Mtr4 and NRDE2 depletion causes DNA double-strand break, surprisingly, in a R-loop independent way. Nevertheless, the R-loop formation in a subset of genes was indeed regulated by the Mtr4/NRDE2 complex. These results indicate a novel role of Mtr4/NRDE2 in DNA damage response. My contribution to this work is listed in Appendix III. Additionally, a recent study showed that NRDE2 actually is a negative regulator of the nuclear surveillance system, by inhibiting the interaction between Mtr4 and the nuclear RNA exosome (Wang et al., 2019). This study has shed light on the connection between NRDE2 protein and the nuclear surveillance system.

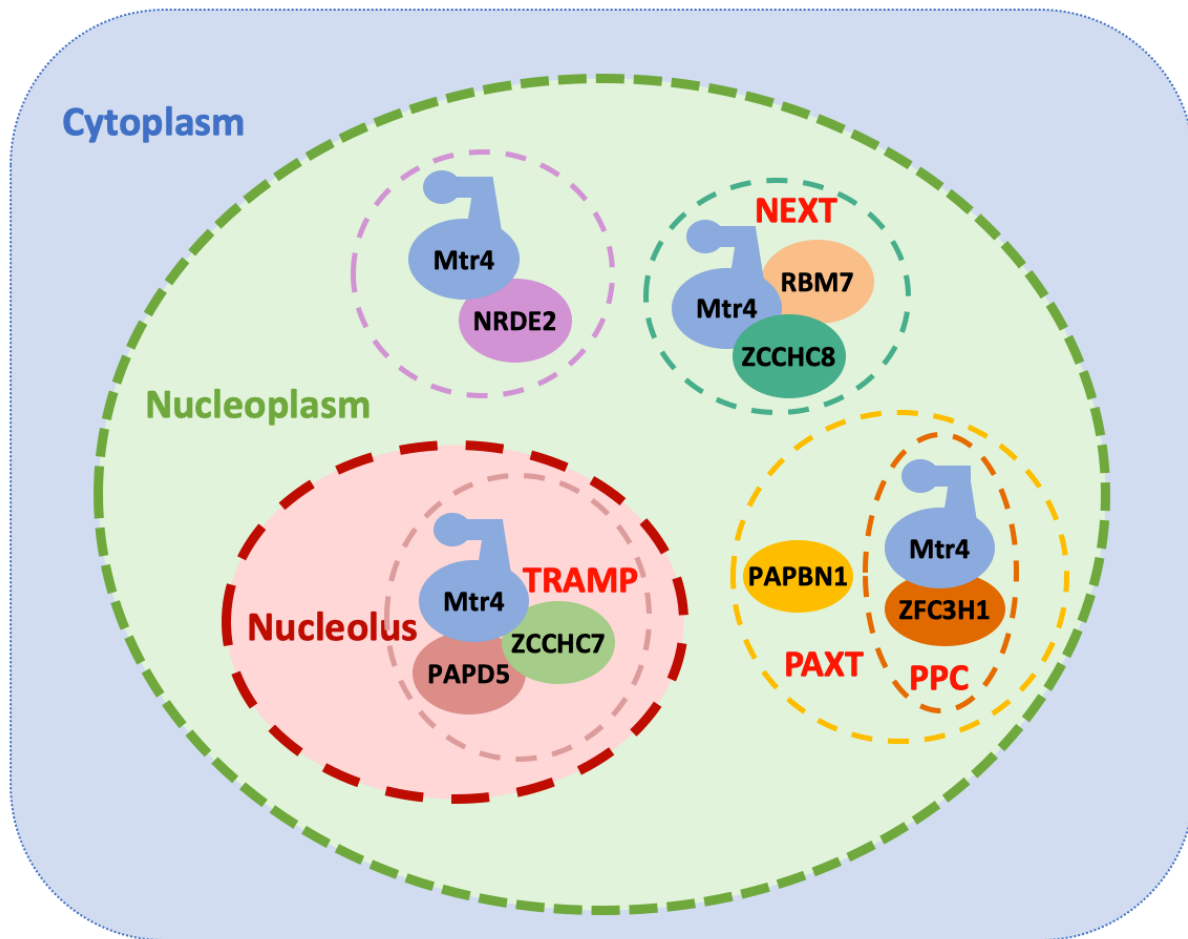


Figure 1-2. The Mtr4-containing complexes in human cells. Four Mtr4-containing complexes are listed: the NEXT complex (Mtr4/RBM7/ZCCHC8), the PPC (Mtr4/ZFC3H1) or PAXT (Mtr4/ZFC3H1/PAPBN1) complex, the Mtr4/NRDE2 complex, and the TRAMP complex (Mtr4/ZCCHC7/PAPD5). Only TRAMP complex is in the nucleolus, while other three all locate in the nucleoplasm.

TRAMP complex

The human TRAMP complex is composed of the RNA helicase Mtr4, a zinc knuckle protein ZCCHC7, and a noncanonical poly(A) polymerase PAPD5. The TRAMP complex is believed to be restricted in nucleoli, since one of its subunit ZCCHC7 can only be observed in nucleoli (Lubas et al., 2011). The other two subunits are also enriched in nucleoli (Lubas et al., 2011; Ogami et al., 2013; Rammelt et al., 2011). Consequently, the substrates of TRAMP substrates are believed to be restricted to nucleolar RNA species (Lubas et al., 2011). Since

TRAMP complex has been shown to participate eRNA degradation, I didn't knockdown any TRAMP subunits in the eRNA study presented in Chapter 3.

Part III. Retinoblastoma-binding protein 6

The retinoblastoma-binding protein 6 (RBBP6) was initially discovered as a large multidomain protein, interacting with tumor suppressor protein p53 and Rb (Saijo et al., 1995; Sakai et al., 1995; Simons et al., 1997). Later, studies found that RBBP6 participates in diverse biological processes, such as cell cycle control, apoptosis, embryonic development, and pre-mRNA processing (Di Giammartino et al., 2014; Gao and Scott, 2003; Li et al., 2007; Miotto et al., 2014). The complexity of RBBP6 isoforms echo their versatile roles in cell. Here I review the key studies of RBBP6, to provide a comprehensive background for the work in Chapter 4.

The complexity of RBBP6 isoforms

The RBBP6 gene contains 17 introns and generates four isoforms by alternative splicing (Figure 1-3) in human. All four isoforms contain a domain named DWNN (Domian With No Name) at the very N terminus. The DWNN domain, with a ubiquitin-like structure (Mather et al., 2005; Pugh et al., 2006), can be expressed as a single domain protein, RBBP6 isoform3, or iso3. Given the high similarities between DWNN and ubiquitin, an intriguing question would be whether RBBP6 iso3 itself may function as a novel ubiquitin-like modifier of other protein substrates (Pugh et al., 2006). This question is answered in Chapter 6, where I show that the DWNN domain is not a ubiquitin-like modifier, instead, itself is ubiquitinated.

The other three isoforms (isoform 1, 2 and 4) all contain a Zinc finger, a RING domain, besides the N-terminal DWNN domain. These three highly conserved domains are also shared by

all RBBP6 homologs across species. In contrast, the p53- and Rb-binding domains only present in vertebrate RBBP6 protein, suggesting that RBBP6 interacting with p53 and Rb was the result of evolution, which provides more specific or diverse regulatory mechanisms of RBBP6. We also note that the RING domain is a signature domain for E3 ubiquitin ligases. Indeed, RBBP6 possess E3 ligase activity and catalyse ubiquitination for several proteins including YB-1, ZBTB38, and p53 (Chibi et al., 2008; Li et al., 2007; Miotto et al., 2014). Most studies to date have been focusing on the longest and shortest isoforms of RBBP6, iso1 and iso3, respectively. Intriguingly, these two isoforms, though generated from a single gene, seem to exert opposing functions in mRNA processing and carcinogenesis. Consequently, the overall function of RBBP6 can be tuned by adjusting the iso1/iso3 protein ratio in cell (Chen et al., 2013; Di Giammartino et al., 2014; Moela and Motadi, 2016; Motadi et al., 2011; Motadi et al., 2018; Yoshitake et al., 2004).

The complexity of RBBP6 isoforms has been reflected in its multifaceted roles in many biological processes. Next, I summarize the key functions of RBBP6 from three aspects: mRNA processing, embryonic development, and its implication in disease.

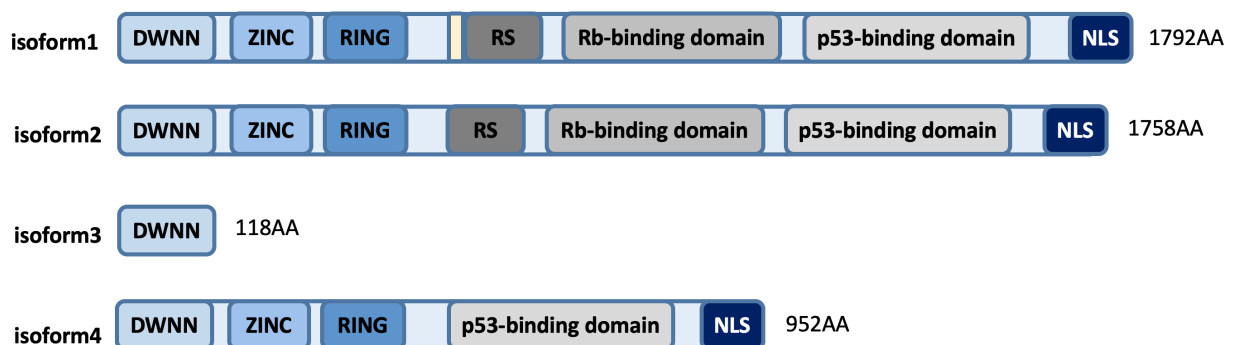


Figure 1-3. The domain organization of RBBP6 isoforms.

RBBP6 isoform1, 2 and 4 represent the long-form RBBP6, whereas isoform3 is the short-form RBBP6, only containing a DWNN domain. The only difference between isoform1 and 2 is a short coiled-coil structure (colored in light yellow) before RS domain. There are in total eight domains listed above, the DWNN domain, the Zinc finger domain, the RING domain, the coiled-coil motif, the arginine and serine (RS) domain, the p53-binding domain, the Rb-binding domain, and the nuclear localization signal (NLS). The total amino acid (AA) numbers of each isoforms are shown on the right side.

RBBP6 in pre-mRNA processing

Despite the diverse roles of RBBP6 revealed in numerous studies, mRNA 3' processing has been one of the most conserved function of RBBP6, from human (Di Giammartino et al., 2014) to yeast (Lee and Moore, 2014; Vo et al., 2001). This function can be reconstituted by the three domains at the N terminus: the DWNN domain, the Zinc finger, and the RING domain, in human, regardless of the long C-terminal fragment in RBBP6 iso1 (Di Giammartino et al., 2014). Consistent with this and as mentioned above, these three domains are highly conserved across species in RBBP6 homologs (Pugh et al., 2006), further confirming that mRNA processing is the most conserved function of RBBP6.

Studies in human cells and yeasts have shown consistent results: The Zinc finger and RING domain are required for efficient RNA binding, whereas the DWNN appears most critical for interaction with Cleavage and Polyadenylation Specificity factors, CPSFs (Di Giammartino et al., 2014; Lee and Moore, 2014). Moreover, a previous study in our lab found that RBBP6 iso3, the single DWNN domain protein, can compete with the long-form RBBP6 for binding to CstF64, a component of the core 3' processing machinery. Moreover, by adjusting the ratio between the promoting iso1 and inhibitory iso3, the alternative polyadenylation (APA) pattern and global gene expression can be modulated. As shown in Figure 1-4, low iso1/iso3 ratio leads to 3'UTR lengthening in genes with more than one APA sites, as well as defective pre-mRNA cleavage for genes with only one poly(A) site (Di Giammartino et al., 2014). High iso1/iso3 ratio, which is the case in many cancers (Chen et al., 2013; Mbita et al., 2012; Moela and Motadi, 2016; Motadi et al., 2011; Motadi et al., 2018; Yoshitake et al., 2004), corresponds with more efficient 3' processing and shorter 3' UTRs, and contributes to the activation of many oncogenes (Mayr and Bartel, 2009).

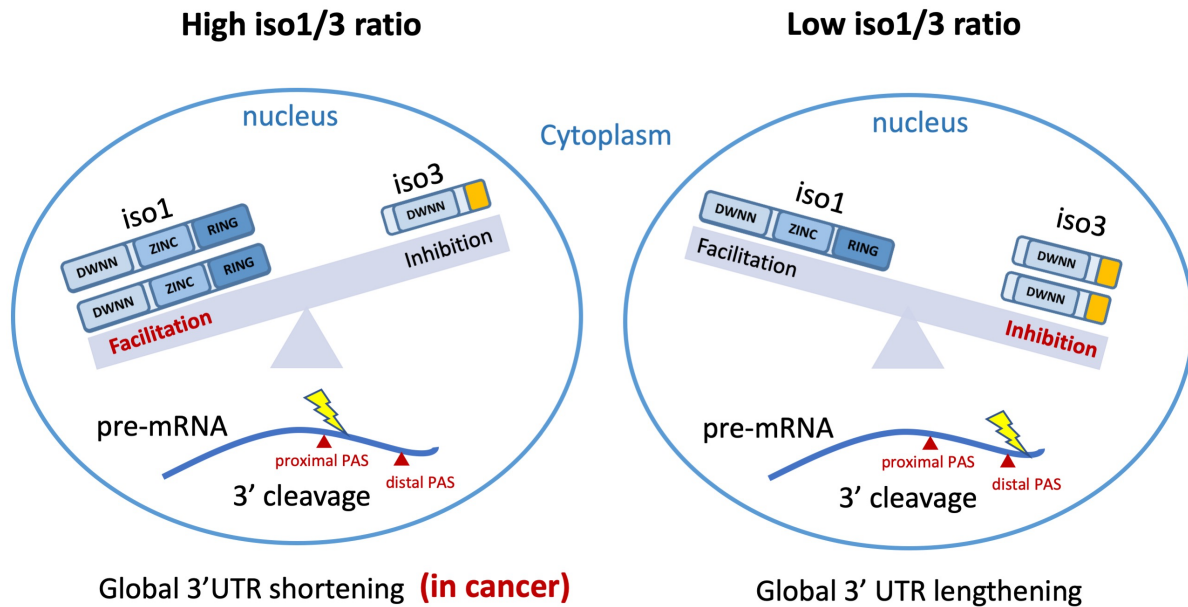


Figure 1-4. Model for pre-mRNA 3' processing mechanism modulated by RBBP6 iso1/iso3 protein ratio. RBBP6 iso1 High iso1/iso3 ratio guarantees efficient pre-mRNA 3' processing and leads to global 3'UTR shortening, which is also observed in many cancers. Low iso1/iso3 ratio inhibits pre-mRNA 3' processing, and results in the use of the strongest poly(A) sites, which are usually more distal. Eventually it will lead to global 3'UTR lengthening.

In Chapter 6, I show that the dynamic nuclear localization of RBBP6 iso3 is regulated by monoubiquitylation and Importin7. This finding provides another way to regulate the iso1/iso3 ratio in nucleus, which is by transporting iso3 in and out of the nucleus.

RBBP6 in embryonic development

In this section, I discuss the key function of long-form RBBP6 in embryonic development. Studies have found that null mutants of RBBP6 are embryonic lethal in both mice and flies (Li et al., 2007; Mather et al., 2005). Homozygous deletion of RBBP6 genes in mice resulted in early embryonic death on embryonic day 7.5 (E7.5), which can be prolonged to E11.5, by simultaneous deletion of p53. This partial rescue is likely mediated by Mdm2, a key negative regulator of p53: RBBP6 promotes p53 ubiquitination and degradation in a Mdm2 dependent manner. Moreover, knockdown of RBBP6 results in accumulated p53 and induces p53-dependent apoptosis, which

has been observed in mouse embryos (Li et al., 2007). Consistent with this, disruption of SNAMA, the RBBP6 homolog of *Drosophila*, also leads to apoptosis during embryogenesis and throughout the gastrula. Nevertheless, unlike human RBBP6, SNAMA doesn't possess p53- or Rb-binding domain, which indicates that RBBP6 might also regulate apoptosis in a p53-independent pathway. The roles of RBBP6 in embryonic development still remains to be elucidated.

RBBP6 and disease

RBBP6 has been implicated in various diseases. For example, RBBP6 iso1 has been found to be overexpressed in many types of cancer, such as breast, cervical, lung, colon, and esophagus carcinoma (Chen et al., 2013; Moela and Motadi, 2016; Motadi et al., 2011; Motadi et al., 2018; Yoshitake et al., 2004). In contrast, RBBP6 iso3 is downregulated (Mbita et al., 2012). As discussed in the previous section “RBBP6 in pre-mRNA processing”, the high iso1/iso3 ratio in cancer is consistent with the finding that cancer cells typically display global 3'UTR shortening, which often contribute to oncogene activation (Mayr and Bartel, 2009).

Besides modulating global 3'UTR, RBBP6 can also contribute to cancer development through another mechanism, as it is also found to play a role in cell cycle and apoptosis: The overexpression of near full length RBBP6 leads to cell cycle arrest at prometaphase as well as mitotic apoptosis (Gao and Scott, 2002). Lastly, we note that several studies have shown that the expression level of RBBP6 can used as a predictive or prognostic parameter for several cancers (Chen et al., 2013; Morisaki et al., 2014).

Another example is that five germline RBBP6 mutations in/adjacent its p53-binding domain display predisposition to myeloproliferative neoplasms, which suggests a p53-dependent pathway of RBBP6 in pathogenesis (Harutyunyan et al., 2016). Also, a recent study identified

RBBP6 as a negative regulator of Ebola virus replication by systematically mapping the interaction between Ebola virus proteins and human proteins (Batra et al., 2018). Moreover, a 23 amino acid region (549-571 amino acid regions of RBBP6 iso1) was characterized to bind the viral transcription regulator VP30 by X-ray crystallography, which can serve as a potential therapeutic target. Surprisingly, this region was outside of all the defined domains in RBBP6, instead, it locates between the RING domain and the RS domain. It's noteworthy that only iso1 but not iso3, binds VP30 and exhibits an inhibitory effect on viral replication.

The functional significance of RBBP6 has been backed by more and more evidence in recent years. Expectantly, my work on the nuclear localization mechanism of RBBP6 iso3, presented in Chapter 6, would contribute to a better understanding of this multifaceted protein.

References

- Aguilera, A., and Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. *Mol Cell* *46*, 115-124.
- Alvarez-Dominguez, J.R., Hu, W., Yuan, B., Shi, J., Park, S.S., Gromatzky, A.A., van Oudenaarden, A., and Lodish, H.F. (2014). Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood* *123*, 570-581.
- Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., *et al.* (2013). The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat Struct Mol Biol* *20*, 1367-1376.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., *et al.* (2014a). An atlas of active enhancers across human cell types and tissues. *Nature* *507*, 455-461.
- Andersson, R., Refsing Andersen, P., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T., and Sandelin, A. (2014b). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun* *5*, 5336.
- Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., Lennartsson, A., Ronnerblad, M., Hrydziuszko, O., Vitezic, M., *et al.* (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* *347*, 1010-1014.

- Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N.J. (1997). Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* *11*, 2494-2509.
- Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* *27*, 299-308.
- Batra, J., Hultquist, J.F., Liu, D., Shtanko, O., Von Dollen, J., Satkamp, L., Jang, G.M., Luthra, P., Schwarz, T.M., Small, G.I., *et al.* (2018). Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication. *Cell* *175*, 1917-1930 e1913.
- Beaulieu, Y.B., Kleinman, C.L., Landry-Voyer, A.M., Majewski, J., and Bachand, F. (2012). Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS Genet* *8*, e1003078.
- Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. *Cell* *168*, 629-643.
- Brazao, T.F., Johnson, J.S., Muller, J., Heger, A., Ponting, C.P., and Tybulewicz, V.L. (2016). Long noncoding RNAs in B-cell development and activation. *Blood* *128*, e10-19.
- Bresson, S.M., and Conrad, N.K. (2013). The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. *PLoS Genet* *9*, e1003893.
- Bresson, S.M., Hunter, O.V., Hunter, A.C., and Conrad, N.K. (2015). Canonical Poly(A) Polymerase Activity Promotes the Decay of a Wide Variety of Mammalian Nuclear RNAs. *PLoS Genet* *11*, e1005610.
- Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet* *28*, 276-284.
- Bulger, M., and Groudine, M. (2010). Enhancers: the abundance and function of regulatory sequences beyond promoters. *Dev Biol* *339*, 250-257.
- Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell* *144*, 327-339.
- Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Mol Cell* *49*, 825-837.
- Chen, H., Li, C., Peng, X., Zhou, Z., Weinstein, J.N., Cancer Genome Atlas Research, N., and Liang, H. (2018). A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell* *173*, 386-399 e312.
- Chen, J., Tang, H., Wu, Z., Zhou, C., Jiang, T., Xue, Y., Huang, G., Yan, D., and Peng, Z. (2013). Overexpression of RBBP6, alone or combined with mutant TP53, is predictive of poor prognosis in colon cancer. *PLoS One* *8*, e66524.

- Chibi, M., Meyer, M., Skepu, A., DJ, G.R., Moolman-Smook, J.C., and Pugh, D.J. (2008). RBBP6 interacts with multifunctional protein YB-1 through its RING finger domain, leading to ubiquitination and proteosomal degradation of YB-1. *J Mol Biol* 384, 908-916.
- ENCODE project Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74.
- Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics* 46, 1311-1320.
- Dao, L.T.M., Galindo-Albarran, A.O., Castro-Mondragon, J.A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., Stephen, T., *et al.* (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature Genetics* 49, 1073-+.
- De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* 8, e1000384.
- de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 26, 11-24.
- Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev* 30, 1357-1382.
- Di Giammartino, D.C., Li, W., Ogami, K., Yashinskies, J.J., Hoque, M., Tian, B., and Manley, J.L. (2014). RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs. *Genes Dev* 28, 2248-2260.
- Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Gene Dev* 18, 2485-2490.
- Epstein, D.J. (2009). Cis-regulatory mutations in human disease. *Brief Funct Genomic Proteomic* 8, 310-316.
- Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shores, N., Whitton, H., Ryan, R.J., Shishkin, A.A., *et al.* (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337-343.
- Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. *Science* 361, 1341-1345.
- Gao, S., and Scott, R.E. (2002). P2P-R protein overexpression restricts mitotic progression at prometaphase and promotes mitotic apoptosis. *J Cell Physiol* 193, 199-207.

- Gao, S., and Scott, R.E. (2003). Stable overexpression of specific segments of the P2P-R protein in human MCF-7 cells promotes camptothecin-induced apoptosis. *J Cell Physiol* *197*, 445-452.
- Hah, N., Benner, C., Chong, L.W., Yu, R.T., Downes, M., and Evans, R.M. (2015). Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. *Proc Natl Acad Sci U S A* *112*, E297-302.
- Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* *23*, 1210-1223.
- Harutyunyan, A.S., Giambruno, R., Krendl, C., Stukalov, A., Klampfl, T., Berg, T., Chen, D., Milosevic Feenstra, J.D., Jager, R., Gisslinger, B., *et al.* (2016). Germline RBBP6 mutations in familial myeloproliferative neoplasms. *Blood* *127*, 362-365.
- Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Gene Dev* *32*, 26-41.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* *155*, 934-947.
- Hsieh, C.L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.S., Chen, S., *et al.* (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci U S A* *111*, 7319-7324.
- Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., *et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* *465*, 182-U165.
- Kim, T.K., and Shiekhattar, R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* *162*, 948-959.
- Kleftogiannis, D., Kalnis, P., and Bajic, V.B. (2016). Progress and challenges in bioinformatics approaches for enhancer identification. *Brief Bioinform* *17*, 967-979.
- Kuznetsov, V.A., Bondarenko, V., Wongsurawat, T., Yenamandra, S.P., and Jenjaroenpun, P. (2018). Toward predictive R-loop computational biology: genome-scale prediction of R-loops reveals their association with complex promoter structures, G-quadruplexes and transcriptionally active enhancers. *Nucleic Acids Res* *46*, 8023.
- Laffleur, B., Basu, U., and Lim, J. (2017). RNA Exosome and Non-coding RNA-Coupled Mechanisms in AID-Mediated Genomic Alterations. *J Mol Biol* *429*, 3230-3241.
- Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* *494*, 497-501.

Le Gras, S., Keime, C., Anthony, A., Lotz, C., De Longprez, L., Brouillet, E., Cassel, J.C., Boutillier, A.L., and Merienne, K. (2017). Altered enhancer transcription underlies Huntington's disease striatal transcriptional signature. *Sci Rep-Uk* 7.

Lee, K., Hsiung, C.C.S., Huang, P., Raj, A., and Blobel, G.A. (2015). Dynamic enhancer-gene body contacts during transcription elongation. *Gene Dev* 29, 1992-1997.

Lee, N.N., Chalamcharla, V.R., Reyes-Turcu, F., Mehta, S., Zofall, M., Balachandran, V., Dhakshnamoorthy, J., Taneja, N., Yamanaka, S., Zhou, M., *et al.* (2013). Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell* 155, 1061-1074.

Lee, S.D., and Moore, C.L. (2014). Efficient mRNA Polyadenylation Requires a Ubiquitin-Like Domain, a Zinc Knuckle, and a RING Finger Domain, All Contained in the Mpel Protein. *Molecular and Cellular Biology* 34, 3955-3967.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Curr Biol* 20, R754-763.

Li, L., Deng, B.W., Xing, G.C., Teng, Y., Tian, C.Y., Cheng, X., Yin, X.S., Yang, J.T., Gao, X., Zhu, Y.P., *et al.* (2007). PACT is a negative regulator of p53 and essential for cell growth and embryonic development. *P Natl Acad Sci USA* 104, 7951-7956.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., *et al.* (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498, 516-520.

Liu, Z., Merkurjev, D., Yang, F., Li, W., Oh, S., Friedman, M.J., Song, X., Zhang, F., Ma, Q., Ohgi, K.A., *et al.* (2014). Enhancer activation requires trans-recruitment of a mega transcription factor complex. *Cell* 159, 358-373.

Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., and Jensen, T.H. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep* 10, 178-192.

Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A., and Jensen, T.H. (2011). Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell* 43, 624-637.

Macheret, M., and Halazonetis, T.D. (2015). DNA replication stress as a hallmark of cancer. *Annu Rev Pathol* 10, 425-448.

Masternak, K., Peyraud, N., Krawczyk, M., Barras, E., and Reith, W. (2003). Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol* 4, 132-137.

- Mathelier, A., Lefebvre, C., Zhang, A.W., Arenillas, D.J., Ding, J., Wasserman, W.W., and Shah, S.P. (2015). Cis-regulatory somatic mutations and gene-expression alteration in B-cell lymphomas. *Genome Biol* 16, 84.
- Mather, A., Rakgotho, M., and Ntwasa, M. (2005). SNAMA, a novel protein with a DWNN domain and a RING finger-like motif: a possible role in apoptosis. *Biochim Biophys Acta* 1727, 169-176.
- Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138, 673-684.
- Mbita, Z., Meyer, M., Skepu, A., Hosie, M., Rees, J., and Dlamini, Z. (2012). De-regulation of the RBBP6 isoform 3/DWNN in human cancers. *Mol Cell Biochem* 362, 249-262.
- Meng, F.L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberg, D., *et al.* (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159, 1538-1548.
- Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A., *et al.* (2016). Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol Cell* 64, 520-533.
- Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I.E., Males, M., Viales, R.R., and Furlong, E.E.M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Gene Dev* 32, 42-57.
- Miotto, B., Chibi, M., Xie, P., Koundrioukoff, S., Moolman-Smook, H., Pugh, D., Debatisse, M., He, F., Zhang, L., and Defossez, P.A. (2014). The RBBP6/ZBTB38/MCM10 axis regulates DNA replication and common fragile site stability. *Cell Rep* 7, 575-587.
- Mitchell, P. (2014). Exosome substrate targeting: the long and short of it. *Biochem Soc Trans* 42, 1129-1134.
- Moela, P., and Motadi, L.R. (2016). RBBP6: a potential biomarker of apoptosis induction in human cervical cancer cell lines. *Onco Targets Ther* 9, 4721-4735.
- Moreau, P., Hen, R., Wasyluk, B., Everett, R., Gaub, M.P., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* 9, 6047-6068.
- Morisaki, T., Yashiro, M., Kakehashi, A., Inagaki, A., Kinoshita, H., Fukuoka, T., Kasashima, H., Masuda, G., Sakurai, K., Kubo, N., *et al.* (2014). Comparative Proteomics Analysis of Gastric Cancer Stem Cells. *Plos One* 9.
- Motadi, L.R., Bhoola, K.D., and Dlamini, Z. (2011). Expression and function of retinoblastoma binding protein 6 (RBBP6) in human lung cancer. *Immunobiology* 216, 1065-1073.

- Motadi, L.R., Lekganyane, M.M., and Moela, P. (2018). RBBP6 expression effects on cell proliferation and apoptosis in breast cancer cell lines with distinct p53 statuses. *Cancer Manag Res* 10, 3357-3369.
- Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* 51, 606-617.
- Ogami, K., Cho, R., and Hoshino, S. (2013). Molecular cloning and characterization of a novel isoform of the non-canonical poly(A) polymerase PAPD7. *Biochem Biophys Res Commun* 432, 135-140.
- Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J.J., Yates, J.R., 3rd, Tian, B., and Manley, J.L. (2017). An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev*.
- Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Program, N.C.S., *et al.* (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A* 110, 17921-17926.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.P., *et al.* (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* 161, 774-789.
- Pnueli, L., Rudnizky, S., Yosefzon, Y., and Melamed, P. (2015). RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin alpha-subunit gene. *Proc Natl Acad Sci U S A* 112, 4369-4374.
- Pugh, D.J., Ab, E., Faro, A., Luty, P.T., Hoffmann, E., and Rees, D.J. (2006). DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways. *BMC Struct Biol* 6, 1.
- Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K.R., Resch, W., Liang, G., Tang, Z., Mathe, E., Benner, C., *et al.* (2014). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* 159, 1524-1537.
- Rammelt, C., Bilen, B., Zavolan, M., and Keller, W. (2011). PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *RNA* 17, 1737-1746.
- Richard, P., Ogami, K., Chen, Y., Feng, S., Moresco, J.J., Yates, J.R., 3rd, and Manley, J.L. (2018). NRDE-2, the human homolog of fission yeast Nrl1, prevents DNA damage accumulation in human cells. *RNA Biol*.
- Rickels, R., and Shilatifard, A. (2018). Enhancer Logic and Mechanics in Development and Disease. *Trends Cell Biol* 28, 608-630.

- Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. *Cell* *155*, 39-55.
- Rogan, D.F., Cousins, D.J., Santangelo, S., Ioannou, P.A., Antoniou, M., Lee, T.H., and Staynov, D.Z. (2004). Analysis of intergenic transcription in the human IL-4/IL-13 gene cluster. *Proc Natl Acad Sci U S A* *101*, 2446-2451.
- Rothschild, G., and Basu, U. (2017). Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability. *Trends in Genetics* *33*, 143-154.
- Saijo, M., Sakai, Y., Kishino, T., Niikawa, N., Matsuura, Y., Morino, K., Tama, K., and Taya, Y. (1995). Molecular-Cloning of a Human Protein That Binds to the Retinoblastoma Protein and Chromosomal Mapping. *Genomics* *27*, 511-519.
- Sakai, Y., Saijo, M., Coelho, K., Kishino, T., Niikawa, N., and Taya, Y. (1995). cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product. *Genomics* *30*, 98-101.
- Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. *Nat Rev Genet* *16*, 583-597.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* *489*, 109-113.
- Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. *Mol Cell* *56*, 29-42.
- Schmid, M., and Jensen, T.H. (2018). Controlling nuclear RNA levels. *Nat Rev Genet* *19*, 518-529.
- Serfling, E., Jasin, M., and Schaffner, W. (1985). Enhancers and Eukaryotic Gene-Transcription. *Trends in Genetics* *1*, 224-230.
- Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* *15*, 272-286.
- Silla, T., Karadoulama, E., Makosa, D., Lubas, M., and Jensen, T.H. (2018). The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts. *Cell Rep* *23*, 2199-2210.
- Simons, A., MelamedBessudo, C., Wolkowicz, R., Sperling, J., Sperling, R., Eisenbach, L., and Rotter, V. (1997). PACT: Cloning and characterization of a cellular p53 binding protein that interacts with Rb. *Oncogene* *14*, 145-155.
- Smith, E., and Shilatifard, A. (2014). Enhancer biology and enhanceropathies. *Nature Structural & Molecular Biology* *21*, 210-219.

- Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* *13*, 613-626.
- Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* *14*, 103-105.
- Tautz, D., and Domazet-Lošo, T. (2011). The evolutionary origin of orphan genes. *Nat Rev Genet* *12*, 692-702.
- Tuan, D., Kong, S., and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A* *89*, 11219-11223.
- Vahedi, G., Kanno, Y., Furumoto, Y., Jiang, K., Parker, S.C., Erdos, M.R., Davis, S.R., Roychoudhuri, R., Restifo, N.P., Gadina, M., *et al.* (2015). Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* *520*, 558-562.
- Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among Distant Regulatory Elements at the beta-Globin Locus Requires GATA-1 and FOG-1. *Molecular Cell* *17*, 453-462.
- Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deville, R., Erichsen, J.T., Jasinska, A.J., *et al.* (2015). Enhancer evolution across 20 mammalian species. *Cell* *160*, 554-566.
- Vo, L.T.A., Minet, M., Schmitter, J.M., Lacroute, F., and Wyers, F. (2001). Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. *Molecular and Cellular Biology* *21*, 8346-8356.
- Whalen, S., Truty, R.M., and Pollard, K.S. (2016). Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* *48*, 488-496.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* *153*, 307-319.
- Williamson, I., Berlivet, S., Eskeland, R., Boyle, S., Illingworth, R.S., Paquette, D., Dostie, J., and Bickmore, W.A. (2014). Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev* *28*, 2778-2791.
- Yamanaka, S., Mehta, S., Reyes-Turcu, F.E., Zhuang, F., Fuchs, R.T., Rong, Y., Robb, G.B., and Grewal, S.I. (2013). RNAi triggered by specialized machinery silences developmental genes and retrotransposons. *Nature* *493*, 557-560.
- Yoshitake, Y., Nakatsura, T., Monji, M., Senju, S., Matsuyoshi, H., Tsukamoto, H., Hosaka, S., Komori, H., Fukuma, D., Ikuta, Y., *et al.* (2004). Proliferation potential-related protein, an ideal esophageal cancer antigen for immunotherapy, identified using complementary DNA microarray

analysis. *Clinical Cancer Research* 10, 6437-6448.

Young, R.S., Hayashizaki, Y., Andersson, R., Sandelin, A., Kawaji, H., Itoh, M., Lassmann, T., Carninci, P., Consortium, F., Bickmore, W.A., *et al.* (2015). The frequent evolutionary birth and death of functional promoters in mouse and human. *Genome Res* 25, 1546-1557.

Zhang, Y., Wong, C.H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., *et al.* (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* 504, 306-310.

Zhu, Y., Sun, L., Chen, Z., Whitaker, J.W., Wang, T., and Wang, W. (2013). Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res* 41, 10032-10043.

Zinder, J.C., and Lima, C.D. (2017). Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* 31, 88-100.

Chapter 2. RNA Surveillance by the Nuclear RNA Exosome: Mechanisms and Significance

Yaqiong Chen contributed to the last part “4. Significance of the Nuclear RNA Exosome in Mammalian Biological Processes”, and wrote the 4.4 and 4.5 sections.

Review

RNA Surveillance by the Nuclear RNA Exosome: Mechanisms and Significance

Koichi Ogami ^{1,*} , Yaqiong Chen ² and James L. Manley ²

¹ Department of Biological Chemistry, Graduate School of Pharmaceutical Sciences, Nagoya City University, Nagoya 467-8603, Japan

² Department of Biological Sciences, Columbia University, New York, NY 10027, USA; yc2906@columbia.edu (Y.C.); jlm2@columbia.edu (J.L.M.)

* Correspondence: koichi_ogami@phar.nagoya-cu.ac.jp; Tel.: +81-52-836-3665

Received: 8 February 2018; Accepted: 8 March 2018; Published: 11 March 2018

Abstract: The nuclear RNA exosome is an essential and versatile machinery that regulates maturation and degradation of a huge plethora of RNA species. The past two decades have witnessed remarkable progress in understanding the whole picture of its RNA substrates and the structural basis of its functions. In addition to the exosome itself, recent studies focusing on associated co-factors have been elucidating how the exosome is directed towards specific substrates. Moreover, it has been gradually realized that loss-of-function of exosome subunits affect multiple biological processes, such as the DNA damage response, R-loop resolution, maintenance of genome integrity, RNA export, translation, and cell differentiation. In this review, we summarize the current knowledge of the mechanisms of nuclear exosome-mediated RNA metabolism and discuss their physiological significance.

Keywords: exosome; RNA surveillance; RNA processing; RNA degradation

1. Introduction

Regulation of RNA maturation and degradation is a crucial step in gene expression. The nuclear RNA exosome has a central role in monitoring nearly every type of transcript produced by RNA polymerase I, II, and III (Pol I, II, and III). The exosome guarantees fidelity of the mature 3' ends of certain stable RNA species, such as ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), telomeric RNAs, small nuclear and nucleolar RNAs (snRNAs and snoRNAs), not only by catalyzing 3' end trimming, but also by degrading transcripts containing an incomplete 3' end [1–3]. Besides, processing of messenger RNA precursors (pre-mRNAs), such as by splicing and 3' end formation, is also under the surveillance of the exosome (Figure 1) [4–13].

Intriguingly, recent advances in RNA sequencing techniques have enabled detection of novel Pol II transcripts (Figure 1), which are expressed at extremely low levels because of rapid RNA turnover by the exosome. A large fraction of these RNAs can be categorized as long non-coding RNA (lncRNA). The most well-known lncRNA substrates for the exosome are cryptic unstable transcripts (CUTs) in yeast [14–16], and their human counterparts, promoter upstream transcripts (PROMPTs) or upstream antisense RNAs (uaRNAs) [17,18], which arise due to antisense transcription from divergent gene promoters. The exosome removes sense non-coding transcripts, such as prematurely terminated RNAs (ptRNAs) [19], which are prematurely terminated and polyadenylated at a poly(A) signal (PAS) typically located in an intron of a protein-coding gene [20], and transcription start site (TSS)-associated RNAs (tssRNAs), which are infrequent short non-coding RNAs (ncRNA) (20–65 nt) generated as a result of promoter-proximal termination of sense transcription [21]. Transcriptional enhancers are also transcribed bi-directionally, and produce a class of lncRNA called enhancer RNAs (eRNAs). It was reported that exosome-sensitive eRNAs emerge from virtually all active enhancer regions, determined

by comprehensive cap analysis of gene expression (CAGE) analyses [22]. Furthermore, long intergenic RNAs (lincRNAs) are also exosome targets [23], although they are generally more stable than uaRNA and eRNA [24].

Strikingly, recent studies have been gradually revealing that the exosome is involved in multiple important biological processes. Those include the DNA damage response (DDR), R-loop resolution, maintenance of genome integrity, RNA export, translation, and cell differentiation. In this work, we review and update our current understanding regarding structural insights into RNA degradation by the exosome and its associated co-factors. We also summarize how abrogation of the functions of the exosome impacts cellular processes in mammals.

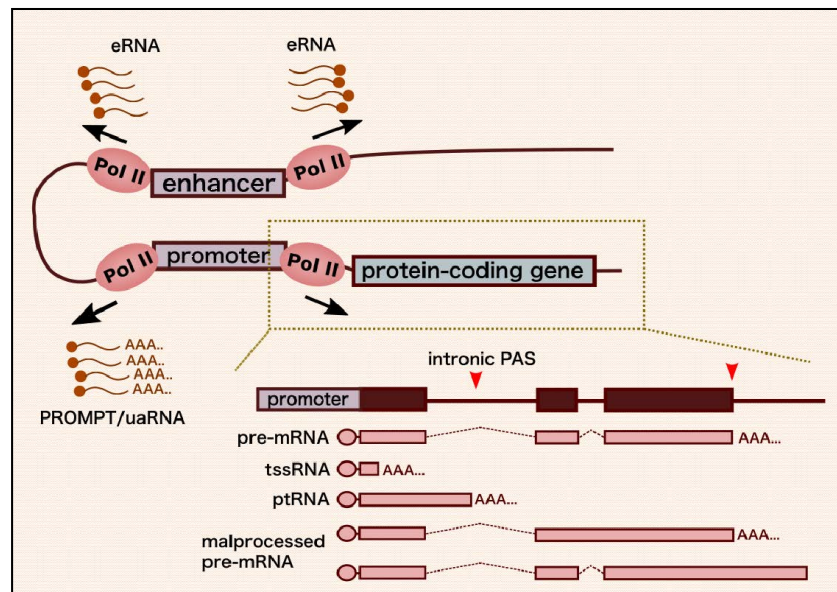


Figure 1. Schematic depiction of Polymerase II transcripts generated from enhancers and gene promoters. Both enhancers and promoters are transcribed bi-directionally and produce various types of transcripts, including messenger RNA precursors (pre-mRNA), transcription start site-associated RNA (tssRNA), prematurely terminated RNA (ptRNA), upstream antisense RNA (uaRNA) or promoter upstream transcript (PROMPT), and enhancer RNA (eRNA). The exosome functions in nuclear RNA surveillance to degrade these RNAs, as well as misprocessed messenger RNA (mRNA) precursors, such as intron-retained and poly(A) signal-mediated cleavage, and polyadenylation-defective pre-mRNAs.

2. The Nuclear RNA Exosome: Structure and RNA Degradation Mechanisms

The eukaryotic nuclear RNA exosome is a 3′–5′ exonuclease complex, consisting of a 9-protein catalytically inactive core complex (EXO-9) and two catalytic subunits, Rrp6 (also known as PM/Scl-100 or EXOSC10), and Dis3 (also known as Rrp44 or EXOSC11). EXO-9 forms a double-layered barrel-like structure that comprises six ribonuclease (RNase) pleckstrin homology (PH)-like proteins (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, and Mtr3) and three S1/K homology (KH) “cap” proteins (Rrp4, Rrp40, and Csl4) [3]. The two catalytic subunits occupy opposite ends of EXO-9 to constitute EXO-11. Rrp6 is placed at the top of the S1/KH cap ring near the RNA entry pore, and Dis3 is tethered to the bottom of EXO-9 near the RNA exit pore [25–27]. Both Rrp6 and Dis3 are 3′–5′ exonucleases, but the latter also has endonucleolytic activity [28–30]. Rrp6 widens the central channel of core EXO-9 and allosterically stimulates Dis3 activity [26]. Recently, a study focused on the last 100 amino acids of Rrp6, referred to as a “lasso,” and revealed that the lasso binds RNA proximal to the EXO-9

channel and enhances RNA decay [31]. In humans, both Rrp6 and Dis3 are mostly nuclear, but Rrp6 shows significant nucleolar enrichment [32,33], whereas Dis3 is excluded from the nucleoli [33,34]. In contrast to humans, yeast Rrp6 is restricted to the nucleus, but Rrp6 and Dis3 are both present in the nucleoplasm and nucleolus [32,35].

Three additional co-factors, Mtr4 (in humans, also known as SKIV2L2 or MTREX (nomenclature recently suggested by HUGO)), Rrp47 (also known as C1D), and Mpp6, are required for maximal activity of the nuclear exosome. Rrp47 interacts with Rrp6 to provide a binding platform for Mtr4, an essential DExH-box RNA helicase [36]. Mpp6 binds to the cap subunit Rrp40, and enhances Mtr4 helicase activity [37,38]. This activity is required to unwind secondary structures formed at the 3' end of RNA substrates, so that the resultant single-stranded RNA substrates can be threaded into the central channel of the core complex in a 3' to 5' orientation [39]. Dis3 degrades RNAs threaded through the entire central channel (Figure 2a), whereas Rrp6 degrades or trims the RNA that enters into the S1/KH cap ring, and then traverses the cap to reach the Rrp6 active site (Figure 2b) [26,40,41]. In addition, there is an alternative path by which the RNA can directly access the Dis3 active site (Figure 2c) [42]. The RNA channeling, but not the direct route, induces a conformational change in Dis3 [42]. The estimated path lengths of the threading and direct access *in vitro* are ~30 nt and ~10 nt, respectively [42–45]. Recent studies in *Saccharomyces cerevisiae* have revealed that RNA substrates show preferences for a specific path to Dis3 [46,47]. Notably, identification of transcriptome-wide interactions of RNAs with individual exosome subunits using the ultraviolet (UV) crosslinking and analysis of cDNA (CRAC) technique in growing budding yeast cells showed that RNA substrates produced by all three RNA polymerases (Pol I, II, and III) exhibit preferences [47]. Interestingly, whichever the route is, Mtr4 is required for RNA degradation [47]. In addition to these two paths, a potential new route to Dis3 was recently suggested [48]: by assessing the average length of RNAs protected by the exosome in living budding yeast using CRAC analysis, it was found that there are not only ~10 nt (reflecting direct access) and 39 and 44 nt (likely reflecting RNAs threaded through the channel and also protected by co-factors) peaks, but also a ~20 nt broad peak that was not described in *in vitro* studies.

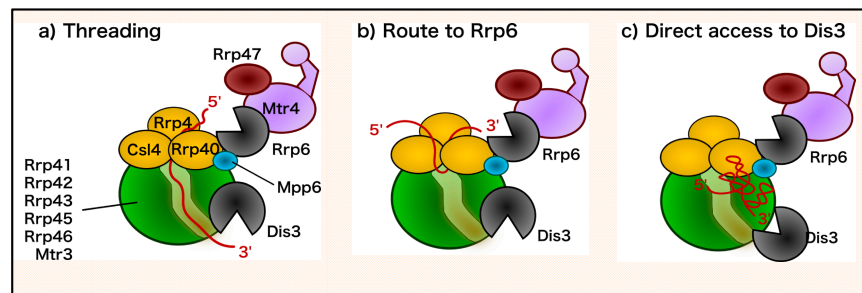


Figure 2. Structure of the RNA exosome and paths for RNA substrates to the catalytic subunits. (a) Threading route: RNA enters the central channel of the core exosome and reaches the active site of Dis3. (b) Route to Rrp6: RNA traverses the cap structure and reaches the active site of Rrp6. (c) Direct access to Dis3. RNA bypasses the central channel and directly accesses Dis3.

3. Molecular Apparatus for RNA Targeting of the Exosome in Yeasts and Humans

The fact that the exosome targets a wide variety of transcripts raises an important question: how is the exosome specifically recruited to particular RNA substrates? Recent studies have identified a number of nuclear exosome–adaptor complexes, which help the exosome load onto selective RNAs [2,3,49]. The components of the adaptors are largely conserved, especially between fission yeast and humans (Table 1). Importantly, Mtr4 is contained in all of the adaptor complexes, indicating that Mtr4 is a central and essential factor for formation of the complexes and for their functions (Figure 3).

Table 1. Conservation of exosome co-factors in yeasts and humans.

Complex	<i>Saccharomyces cerevisiae</i>	<i>Schizosaccharomyces pombe</i>	<i>Homo sapiens</i>
TRAMP	Mtr4 Air1, Air2 Trf4, Trf5	Mtr4 Air1 Cid14	Mtr4/SKIV2L2/MTREX ZCCHC7 PAPD5, PAPD7
NNS	Nrd1 Nab3 Sen1	Seb1 Nab3 Sen1	SCAF4, SCAF8 RALY, RALYL, hnRNPC, hnRNPL1, hnRNPL2, hnRNPL3, hnRNPL4 SETX
MTREC	Mtr4	Mtl1	Mtr4/SKIV2L2/MTREX
NURS	-	Red1	ZFC3H1
Mtr4/ZFC3H1	-	Iss10	-
PAXT	Pho92	Mmi1	YTHDF1, YTHDF2, YTHDF3
	Sto1	Cbc1	CBP80/NCBP1
	Cbc2	Cbc2	CBP20/NCBP2, NCBP2L
	-	Ars2/Pir2	ARS2/SRRT
	-	Red5	ZC3H3
	Sgn1/Rbp1/Rbp29	Pab2	PABPN1, PABPN1L
	-	Rmn1	RBM26, RBM27
	Pap1	Pla1	PAPOLA, PAPOLB, PAPOLG
Mtl1-Ctr1-Nrl1	Mtr4	Mtl1	Mtr4/SKIV2L2/MTREX
	-	Ctr1	CCDC174
	-	Nrl1	NRDE2
NEXT	Mtr4	Mtr4	Mtr4
	-	-	RBM7
	-	-	ZCCHC8
Other	Utp18	Utp18	UTP18
	Nop53	Rrp16	NOP53
	ISW1	-	SMARCA5
	Rix7	Rix7	NVL/NVL2
	Nsa1	Wdr74	WDR74
	-	-	DGCR8

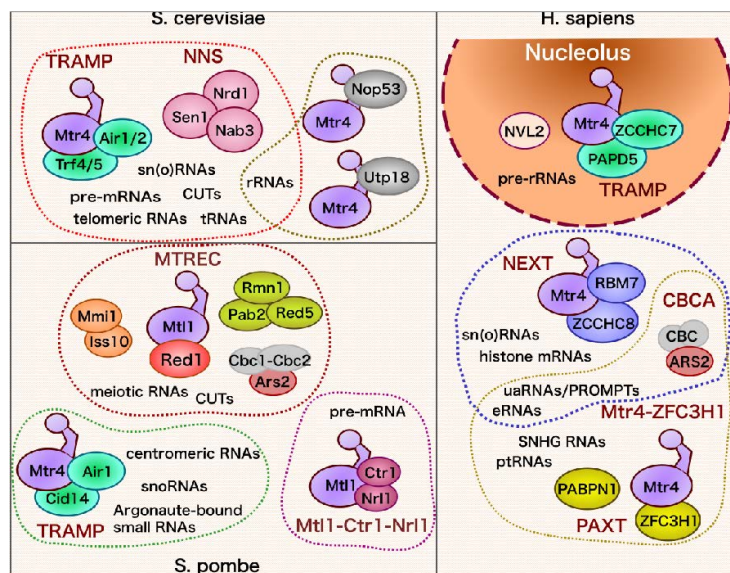


Figure 3. Overview of Mtr4-containing exosome adaptor complexes in yeasts and humans. The RNA helicase Mtr4 participates in multiple distinct exosome adaptor complexes to complete degradation and/or processing of specific RNA substrates. Mtr4-containing complexes identified in *Saccharomyces cerevisiae* (upper-left), *Schizosaccharomyces pombe* (lower-left) and *Homo sapiens* (right) are shown.

3.1. *Saccharomyces cerevisiae*

The Trf4/5-Air1/2-Mtr4 polyadenylation complex (TRAMP) was first described in *S. cerevisiae*, and now is the most well-characterized co-factor that assists exosome-mediated RNA degradation and processing in budding yeast nuclei. Soon after recognizing the importance of polyadenylation of hypomodified methionyl initiator transfer RNA (tRNA^{Met}) by the non-canonical poly(A) polymerase Trf4 for exosome-dependent tRNA quality control [50], the full composition of the responsible protein complex, TRAMP (Mtr4, Trf4, and the Zn-knuckle RNA-binding protein Air1 or Air2), was determined [14,51,52]. Later, another TRAMP complex containing Trf5, a close homolog of Trf4, was identified [53]. Air1/2 provides RNA-binding capability and is also critical for TRAMP assembly [54–56]. TRAMP recognizes a variety of transcripts [12], such as tRNAs [50,52,57–59], rRNAs [59–61], sn/snoRNAs [59,62,63], telomeric RNAs [64], CUTs [14,59,64], and pre-mRNAs [59,65–67], and these substrates are commonly polyadenylated by Trf4/5. In TRAMP, Mtr4 plays roles in RNA unwinding and modulation of poly(A)-tail length of RNA substrates [57,68–73]. Although TRAMP itself has an RNA-binding capacity, its efficient recruitment to RNA substrates is further assisted by the Nrd1–Nab3–Sen1 (NNS) complex [74]. Nrd1 and Nab3 are RNA-binding proteins that recognize specific sequence elements [59,75,76], whereas Sen1 has DNA/RNA helicase activity, which promotes dissociation of Pol II from the template DNA [77–79]. Importantly, NNS travels with a transcribing Pol II by interacting with the C-terminal domain of the Pol II largest subunit (CTD) and terminates transcription when the sequence elements emerge on the nascent RNAs [78,80–85]. NNS-dependent transcription termination is further promoted by the cleavage/polyadenylation factor Pcf11 [86]. Nrd1 interacts with the CTD-containing heptapeptide repeats (YSPTSPS) phosphorylated on Ser5 (Ser5P) through its CTD interaction domain (CID) [83,87,88]. The Nrd1 CID also binds to a CTD mimic motif in Trf4 [89]. The Nrd1 CID interacts with Trf4 and Pol II in a mutually exclusive manner, and therefore, NNS-mediated transcription termination and TRAMP/exosome-mediated RNA degradation are coordinated [89]. Notably, proteins homologous to the NNS components were found in *Schizosaccharomyces pombe* and humans (Table 1). Both *S. pombe* and humans have Sen1 homologs, Sen1 and Senataxin (SETX), respectively. *S. pombe* has the Nrd1 homolog Seb1 [90], which has Ser5P-CTD- and RNA-binding abilities [91,92]. However, although Seb1 is involved in transcription termination and alternative polyadenylation, no NNS-like function was observed [91,93,94]. Functions of the human CID-containing homolog of Nrd1, SCAF8 [95], remain unexplored, except that SCAF8 can bind to the elongating phosphorylated CTD [96,97]. Also, human RALY protein is somewhat similar to Nab3; the RNA recognition motif (RRM) in RALY shares 31% amino acid identity with the Nab3 RRM [98]. However, there is currently no evidence that these putative homologs of NNS subunits form an NNS-like complex and regulate human TRAMP functions.

Several other exosome partners exist in budding yeast. Utp18 and Nop53, an early and late associating small subunit processome factor, respectively, were shown to interact with the exosome to regulate ribosomal RNA precursor (pre-rRNA) processing [99]. Both proteins contain a conserved motif termed an arch-interacting motif (AIM), which directly dock to the arch domain of Mtr4. Recent X-ray crystallography and NMR analyses revealed the structural basis of Mtr4–Nop53 interaction and showed that the Mtr4 arch can bind Nop53 and RNA simultaneously [100]. The G-patch protein Sqs1/Pfa1 also contains a perfect AIM consensus sequence, and thus binds to the Mtr4 arch domain; however, the roles of the interaction remain elusive [99]. In addition, Babour et al. reported that the chromatin remodeling complex ISW1 physically interacts with the exosome in an RNase-insensitive manner [101]. Interestingly, this interaction is enhanced in the export-incompetent thermo-sensitive *npl3-1* mutant strain. ISW1 is required to retain export-defective poly(A)-tailed RNAs on chromatin and remove them by recruiting the exosome. This finding implies that ISW1/exosome participates in a messenger ribonucleoprotein (mRNP) nuclear export surveillance system.

3.2. *Schizosaccharomyces pombe*

The fission yeast *S. pombe* has a complex similar to *S. cerevisiae* TRAMP, consisting of Mtr4, Air1 and the Trf4/5 family of poly(A) polymerase Cid14 [102]. It functions in heterochromatic gene silencing at centromeric repeats [102–104] and polyadenylation-dependent decay of centromeric RNAs [105,106], snoRNA precursors [107], and Argonaute-bound small RNAs [108]. The precise mechanism of TRAMP recruitment to target transcripts remains unclear; however, Mlo3, the *S. pombe* homolog of mRNA export factor Yra1 or ALYREF, was shown to interact with TRAMP to silence centromeric transcripts [103,104]. Besides, the THO complex, which coordinates the steps from transcription to RNA export, is required to maintain TRAMP at snoRNA genes, and these complexes cooperate in the control of snoRNA expression, thus linking transcription and nuclear surveillance machineries [107]. Notably, Yra1 physically associates with the THO complex in *S. cerevisiae* [109,110], and therefore, it is possible that both Mlo3 and the THO complex work in the same pathway for TRAMP-mediated RNA metabolism.

S. pombe has a second Mtr4 homologue protein named Mtl1 (Mtr4-like protein 1), which is independent of TRAMP. Mtl1 interacts with the zinc-finger protein Red1 and various other proteins to form a complex called Mtl1–Red1 core (MTREC) or nuclear RNA silencing (NURS) [111,112]. MTREC interacts with the exosome, presumably through Red1 but not Mtl1 [113]. In agreement with this, Mtl1 lacks the N-terminal motif that mediates the interaction of Mtr4 with Rrp6 and Rrp41 [36]. MTREC further associates with several sub-modules such as Iss10–Mmi1, Red5–Pab2–Rmn1, Ars2–Cbc1–Cbc2, and the canonical poly(A) polymerase Pla1 [111–113]. All of these sub-modules can bind to MTREC simultaneously, forming a large 11 subunit complex [113]. However, since the sub-modules show different stoichiometry for MTREC-binding, there might be various forms of the MTREC complex. The sub-modules enable MTREC to direct specific RNA targets for exosome-mediated decay. The YTH protein Mmi1 is a well-characterized regulator of meiotic gene expression [114–116]. Mmi1 programs meiotic transcripts for co-transcriptional decay by recognizing repeats of a short nucleotide motif termed determinant of selective removal (DSR), which are found within introns in some target genes [116–119]. Iss10 is required for stable interaction between Mmi1 and Red1, and thus, involved in meiotic gene regulation [120]. Red5 and Pab2 contribute to degradation of meiotic mRNAs [121,122] and CUTs [113], whereas depletion of the associating factor Rmn1 does not affect the amount of either meiotic mRNA or CUTs [112,113]. The cap-associated complex Ars2–Cbc1–Cbc2 is responsible for efficient CUT degradation [113], which is reminiscent of the function of the human cap-binding complex (CBC)–ARS2 (CBCA) complex. The human CBCA complex is required for degradation of PROMPTs/uaRNAs [123], which are comparable to yeast CUTs [17] (discussed below).

Mtl1 also forms a Red1-independent protein complex with the *Caenorhabditis elegans* NRDE-2 homologue Nrl1 and the coiled-coil- and DUF4078 domain-containing protein Ctr1 [111,113]. The Mtl1–Ctr1–Nrl1 complex further associates with splicing factors, and is suggested to degrade unspliced pre-mRNA [111,113].

3.3. *Homo sapiens*

In addition to Mtr4, factors homologous to the yeast TRAMP subunits are present in humans; the closest orthologues of Air1/Air2 and Trf4/Trf5 are the zinc-knuckle protein ZCCHC7 and the non-canonical poly(A) polymerase PAPD5 (also known as Trf4-2), respectively. These three proteins form the TRAMP-like complex [124]. Functions of TRAMP-like are thought to be restricted to nucleoli under normal cellular conditions, due to the strict nucleolar localization of ZCCHC7 [124]. The other subunits Mtr4 and PAPD5 are restricted to the nucleus with nucleolar enrichment [124–126]. Interestingly, it was recently shown that viral infection induces cytoplasmic translocation of ZCCHC7 and Mtr4 to facilitate exosome-mediated viral RNA decay in the cytoplasm [127]. It has been shown that PAPD5 is responsible for poly- or oligo-adenylation of nucleolar RNAs, such as snoRNAs [128] and aberrant pre-rRNA species [124,129], suggesting that polyadenylation assists RNA 3' processing and/or degradation by TRAMP-like. Of note, PAPD5 has a close paralog, PAPD7

(also known as Trf4-1), that has been suggested to interact with ZCCHC7 [56]. However, roles of PAPD7 in TRAMP-like remain unclear; PAPD7 is excluded from nucleoli [126], and in agreement with this, PAPD7 is dispensable for polyadenylation of aberrant pre-rRNA species [129]. In addition, there is no evidence of an interaction between PAPD7 and Mtr4 in several independent proteomics analyses [20,124,130].

Human TRAMP-like interacts with several additional proteins. It has been shown that TRAMP-like-mediated pre-rRNA processing is modulated by the AAA-ATPase NVL2 [131,132] and its regulatory factor tryptophan-aspartic acid (WD) repeat-containing protein WDR74 [133–135]. Moreover, splicing factors such as U4/U6-U5 tri-snRNP subunits and hnRNPs are found to associate with TRAMP-like complex [20,124,130]. The function of the interaction with splicing factors awaits further investigation. The nucleolar exosome can interact with the double-stranded RNA-binding protein DGCR8, which is well known as a microprocessor subunit, to degrade mature snoRNAs and telomerase RNA (hTR) [136]. It is noteworthy that although the physical interaction between DGCR8 and TRAMP-like has not been reported, both snoRNAs and hTR are targeted by the TRAMP-like complex [136–139].

In the nucleoplasm, at least two distinct exosome adaptors are present. One is Mtr4-ZFC3H1 or poly(A) tail exosome targeting complex (PAXT), which brings the exosome to various kinds of lncRNAs, including snoRNA host gene (SNHG) transcripts, eRNAs [49], uaRNAs [20,49], and ptRNAs [20]. Another is nuclear exosome targeting complex (NEXT) [124], comprising Mtr4, the RNA binding protein RBM7 and the Zn-knuckle protein ZCCHC8, which degrades PROMPTs/uaRNAs [124], replication-dependent histone mRNAs [123], eRNAs [140], snRNAs [123,141], and snoRNAs [140]. Of note, ZFC3H1 is a close homolog of *S. pombe* Red1, and therefore, Mtr4-ZFC3H1 is the human MTREC. Although RNA substrates of Mtr4-ZFC3H1 and NEXT partly overlap, there are clear differences in their features; Mtr4-ZFC3H1 substrates are longer in RNA body size and have a long poly(A)-tail [20,49]. The precise molecular fundamentals of substrate recognition by Mtr4-ZFC3H1 await further characterization. However, Meola et al. suggested the transient and partially RNA-dependent interaction between Mtr4-ZFC3H1 and the nuclear poly(A)-binding protein PABPN1 [49]. It has been shown that PABPN1 promotes exosome-dependent decay of nuclear poly(A)-tailed transcripts [142–144]. PABPN1-mediated RNA decay is dependent on RNA polyadenylation, which requires the canonical poly(A) polymerases PAP α/γ , but not the TRAMP subunit PAPD5 [142–144], and is thus termed PABPN1- and PAP α/γ -mediated RNA decay pathway (PPD) [143]. Notably, subsets of the PABPN1 substrates overlap with those of Mtr4-ZFC3H1 [20,49]. Yet, the fact that co-depletion of Mtr4 and PABPN1 resulted in synergistic accumulation of target transcripts suggests that Mtr4-ZFC3H1 and PABPN1 may work in both the same and redundant pathways [144]. It will be interesting to investigate if and how Mtr4-ZFC3H1 participates in the PPD pathway. RNA recognition by NEXT involves the connection with the ARS2-associated cap-binding complex CBCA [123], U-rich RNA binding capacity of RBM7 [140,141], and possibly the pre-mRNA 3' processing complex [145]. CBCA and NEXT further associate with the zinc-finger CCCH domain-containing protein ZC3H18 (also known as NHN1) [123,146], and this interaction is important for cap-proximal Pol II stalling, transcription termination, 3' end formation, and RNA decay [123,146,147]. The significance of the interaction between NEXT and the pre-mRNA 3' processing complex remains undetermined.

A nucleoplasmic protein NRDE2, which is the homolog of *S. pombe* Nrl1, also interacts with Mtr4. However, in contrast to the *S. pombe* counterpart, it is unlikely that Mtr4/NRDE2 associates with the whole exosome, since analysis using size-exclusion chromatography-coupled mass spectrometry (MS) revealed that Mtr4/NRDE2 elutes around 440 kDa, which is smaller than the exosome/Mtr4 complex (>600 kDa) [20]. In agreement with this, our recent MS analysis of NRDE2-interacting proteins did not detect any exosome subunits [148].

4. Significance of the Nuclear RNA Exosome in Mammalian Biological Processes

Loss-of-function of the exosome due to mutation or depletion of its subunits and co-factors can cause alterations in various biological processes [2], and ultimately contribute to human disease, such as multiple myeloma [149–151]. Despite various interesting phenotypes in yeast, such as altered chromatin modifications in exosome-deficient cells, we restrict discussion here to evidence provided using mammalian cells.

4.1. DNA Damage Response

The activity of the nuclear exosome is altered during the cellular DDR. The change is triggered by phosphorylation of the NEXT subunit RBM7 by the stress-related kinase p38 MAPK/MK2 [152,153]. Phosphorylated RBM7 is bound by the phosphoserine-binding protein 14-3-3, and loses its RNA-binding ability, which consequently leads to stabilization and accumulation of NEXT substrates such as PROMPTs [152]. Interestingly, cells become hypersensitive to a DNA damaging reagent when RBM7 is depleted, and cells lacking RBM7 exhibit poor survival after drug treatment [152]. Although it is still largely unclear how these changes in the DDR affect cell physiology, there are interesting suggestions that a fraction of promoter-associated lncRNAs can modulate transcription of neighboring genes. For example, *cyclin D1* (CCND1) PROMPTs upregulated in response to DNA damage by ionizing irradiation provide a binding platform for the RNA-binding protein FUS/TLS. FUS/TLS recruited to the CCND1 promoter through PROMPTs represses the histone acetyltransferase activity of CBP/p300, which results in decreased CCND1 transcription [154]. However, it seems that PROMPT-mediated gene regulation is not widespread, since no correlation was observed between altered expression of the downstream gene and increased PROMPT levels in DNA damage or Rrp40 depletion [155]. This might possibly indicate that most PROMPTs lack sequence elements necessary for recruiting specific RNA-binding proteins, and the action of only a small fraction of PROMPTs may be required for the DDR.

4.2. R-Loop Resolution and Genomic Integrity

R loops are three-stranded structures composed of the nascent RNA hybridized with DNA template and the resultant displaced single-stranded DNA (ssDNA). R-loop resolution is a critical step to maintain genome integrity, since the displaced ssDNA is vulnerable to DNA damage [156–159]. Moreover, R loops are associated with human disease (reviewed in [160–163]). Intriguingly, multiple studies have reported the involvement of the exosome in R-loop resolution and genome integrity. In yeast, depletion of Rrp6 or Trf4 leads to R loop-mediated genomic instability and hyperrecombination [164,165], as well as accumulation of aberrant truncated RNA products released from an R loop [166]. These factors also promote the loading of ssDNA binding protein RPA to double-strand breaks (DSBs), and activate the checkpoint kinase Mec1/ATR, which facilitates the formation of continuous Rad51 filaments to initiate homologous recombination [167]. Strikingly, overexpression of RNase H, which removes R loops by digesting the RNA strand of RNA/DNA hybrids, dramatically rescued the rate of genome instability in TRAMP-depleted cells [168]. In human cells, the DNA/RNA helicase SETX (Senataxin), which plays a key role in R-loop resolution [169], directly interacts with the exosome subunit Rrp45 [170]. The interaction requires sumoylation of SETX, which interestingly, is blocked by certain SETX mutations in ataxia oculomotor apraxia 2 (AOA2) patients. It is speculated that SETX recruits the exosome to R loops to promote degradation of the RNA unwound and released by SETX, and thus prevents possible rehybridization and the resultant DNA damage.

Over the last decade, the concept has emerged that exosome-mediated R-loop prevention is a critical step in immunoglobulin class switch recombination (CSR) and somatic hypermutation (SHM) in B lymphocytes [171]. To initiate CSR and SHM, activation-induced cytidine deaminase (AID) deaminates cytidines on both template and non-template DNA strands of transcribing switch regions. However, the template DNA strand hybridized with a nascent transcript cannot be modified by AID

because of limited access to the template strand. Basu et al. identified the core RNA exosome EXO-9 as a key factor that promotes AID access to the template strand in the context of RNA/DNA hybrids, and thus, CSR and SHM [172]. The interaction between AID and the RNA exosome is promoted by the E3 ubiquitin ligase NEDD4, which regulates clearance of Pol II from the immunoglobulin switch region [173]. In mouse B cells and embryonic stem cells (ESCs) containing a conditional inversion allele of *Exosc3* (Rrp40) or *Exosc10* (Rrp6), which allows conditional ablation of the exosome by drug treatment, loss of the exosome results in enhanced R-loop formation and genomic instability, due to an increase of ncRNAs associated with TSS and superenhancers [174–176]. More recently, it was shown that Mtr4 has an RNA/DNA hybrid unwinding activity, and Mtr4-deficient B cells exhibited greater R-loop formation at the immunoglobulin heavy chain locus [177].

4.3. RNA Export and Translation

In addition to its role in NEXT loading to nascent transcript 5' ends, the CBC is required to initiate nuclear RNA export by recruiting various proteins. The TREX mRNA export complex is recruited to the 5' end of mRNAs through the export adaptor proteins ALYREF and THO associating with CBC [178–180]. While splicing enhances TREX recruitment [181], the interaction of ALYREF with the cap-binding protein CBP20 was shown to stimulate nuclear export of capped intronless mRNAs [180]. CBC-associating factor ZC3H18 can also enhance export of intronless mRNAs [182]. Recently, Fan et al. showed that Mtr4 competes with the export adaptor protein ALYREF for binding to ARS2, and thus inhibits nuclear export, providing an important checkpoint to prevent undesired transport of aberrant RNAs into the cytoplasm [183]. Intriguingly, CBCA (CBC–ARS2) and ZC3H18 are also found in the ZFC3H1 interactomes [49], suggesting that Mtr4–ZFC3H1 can also be recruited to CBCA assembled on the 5' cap structure. Therefore, it is possible that both NEXT and Mtr4–ZFC3H1 can antagonize ALYREF binding to CBCA. This competition, as well as the rapid RNA degradation of poly(A)-tailed lncRNAs by Mtr4–ZFC3H1, is particularly important, since normally unstable lncRNAs are exported to the cytoplasm in cells lacking Mtr4–ZFC3H1 [20,184]. Of note, there is a link between RNA 3' end cleavage/polyadenylation and export. Several 3' cleavage and polyadenylation factors interact with RNA export factors. For example, Pcf11 directly interacts with the yeast homolog of ALYREF, Yra1 [185]; CFIm68 directly binds to the mRNA export receptor NXF1 [186]; CPSF100 and CFIm proteins associate with the THO subunit THOC5 [187,188]; and Cstf64 and PABPN1 help ALYREF-binding to mRNA 3' ends [189]. Therefore, effective recruitment of RNA export complex, including ALYREF, is mediated not only by CBC, but also the 3' processing machinery and a poly(A)-tail. Recent remarkable progress in ribosome profiling technologies [190] has led to the realization that ribosome binding or even translation of lncRNAs is pervasive in mammals [191–197]. Concordantly, exported lncRNAs in Mtr4–ZFC3H1 deficient cells become ribosome-associated and likely translated. Because of the translatability of lncRNAs, as well as the more mRNA-like structures of Mtr4–ZFC3H1 substrates (presence of the cap and a poly(A)-tail) than those of NEXT substrates [49], the aberrantly exported Mtr4–ZFC3H1 substrates appear to overwhelm translation machinery and disrupt the quantitative balance between ribosomes and translatable RNAs, which leads to global reduction in heavy polysomes and translation [20,184].

Recently, Sinturel et al. reported intriguing findings that diurnal oscillations in liver mass and hepatocyte size are regulated by rhythmic changes in ribosome biogenesis, in which the nuclear exosome plays a role [198]. In this study, using mice, they demonstrated that these changes are controlled by feeding time: diurnal changes were observed only in mice fed during night and ad libitum, but not in day-fed mice. Importantly, they found that the number of ribosomes also exhibited diurnal fluctuations. In the active/dark phase, translation of ribosomal protein mRNAs was found to be significantly enhanced, and thus, protein synthesis rates increased, while in the resting/light phase, ribosomal protein synthesis was decreased, leading to an imbalance between ribosomal proteins and rRNAs. TRAMP functions to rebalance the amount of these factors by polyadenylating and degrading

excess rRNAs in incomplete ribosomal subunits. These events contribute to a daily rhythm of mouse liver protein content.

4.4. Stem Cell Self-Renewal and Differentiation

Precise regulation of the activity and maintenance of the fidelity of gene expression is vital for stem cell self-renewal, differentiation, and development. Studies have suggested that the nuclear RNA exosome is essential for maintaining progenitor cell function and preventing premature differentiation. A defective exosome pathway can lead to aberrant accumulation of RNAs, among which are mRNAs encoding differentiation-specific transcription factors, and ultimately break the balance between proliferation and differentiation. For example, the nuclear exosome directly degrades *GRHL3* transcripts, which encode a transcription factor critical for epidermal differentiation [199]. Depletion of the exosome subunit Rrp45 (EXOSC9) leads to loss of progenitor cells from the basal epidermal layer and premature differentiation. More recently, Skamagki et al. suggested that the exosome plays an important role in maintaining pluripotent stem cell redox status in mice [200]. They found that the transcription factor *ZSCAN10*, which activates transcription of *EXOSC1/2/5* genes, is expressed at a low level in induced pluripotent stem cell clones generated from aged tissue donors, and the decreased expression of RNA exosome subunits causes the accumulation of AU-rich element-containing RNAs, including glutathione peroxidase 2 (Gpx2). Overexpression of *GPX2* increases the reduced form of glutathione, thus scavenging glutathione-mediated reactive oxygen species, which consequently blunts the DDR and reduces apoptosis. Similar defects were observed following knockdown (KD) of exosome subunits EXOSC2 and/or EXOSC8 in ESCs. Mtr4 is also important in cell proliferation and differentiation. On the one hand, Mtr4 expression is highly upregulated when the self-renewal state of ESCs is induced by inhibitors of kinases, known as 2 inhibitors (2i) [201]. On the other hand, KD of Mtr4 resulted in moderate to severe mouse ESC death [202]. Additionally, depletion of Mtr4 impairs mitosis and induces cell differentiation in the murine cancer cell lines Neuro2A and P19 [203]. All the above indicates that levels of the exosome subunits correlates with cell differentiation. Indeed, Rrp4/Rrp40/Rrp42/Rrp45 (EXOSC2/3/7/9) expression is enriched in progenitor cells, but decreased upon epidermal differentiation in humans [199]. These observations strongly suggest that an abundance of the exosome is a critical prerequisite to maintain stem and progenitor cells in an undifferentiated state.

4.5. Influenza A Virus (IAV) Ribogenesis and Infectivity

A recent study revealed the significance of the exosome in influenza A virus (IAV) ribogenesis and growth [204]. In this study, Rialdi et al. analyzed the proteome of viral polymerase complex-interacting proteins, and identified the core exosome subunits. Intriguingly, they found that viral polymerase activity is attenuated in cells transfected with siRNAs against exosome subunits and in patient-derived cells harboring an *EXOSC3* (Rrp40) mutation. Importantly, viral growth was suppressed in these cells, indicating the essential role of the exosome in viral biogenesis. NEXT-assisted exosome seems to be co-opted by the viral RNA polymerase, since similar results were obtained following RBM7 KD. Moreover, synthesis of host/viral chimeric transcripts generated as a result of “cap snatching”, in which initiation of viral transcription is primed using 5' ends of host transcripts (cap with 10–20 downstream nucleotides), is decreased upon exosome-depletion. Collectively, these results suggest that the nuclear exosome coordinates with viral polymerase during the initial steps of viral transcription with Pol II at host promoters to enhance influenza A virus ribogenesis and infectivity. From the evolutionary point of view, viruses need to integrate their biological activities into hosts by recycling regulatory RNAs generated by hosts. The exosome, as the hub of RNA surveillance system, can be co-opted by viruses to facilitate the efficient formation of cellular/viral hybrid RNAs and cap-snatching.

5. Conclusions and Perspectives

The RNA exosome and its co-factors monitor the versatility and specificity of a huge variety of RNA substrates, and thus plays a crucial role in regulating the activity and maintaining the fidelity of

gene expression. Numerous studies have revealed that an impaired RNA surveillance system can break RNA homeostasis, and thus cause detrimental consequences in multiple biological processes leading to human diseases (reviewed by Morton et al. [149]). However, there are still many unanswered questions about both the fundamental and the pathological mechanisms of the nuclear exosome: how are both specificity and versatility of RNA substrates guaranteed at the same time in the RNA surveillance system? What is the comprehensive mechanism of the nuclear exosome in multiple biological processes, including maintenance of genome integrity and cell differentiation? Deeper understanding of the complexities of the RNA surveillance system has the potential to lead to novel therapeutic remedies to fight human disease.

Acknowledgments: This work was supported by a National Institute of Health grant R35 GM118136 to James L. Manley.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Mitchell, P. Exosome substrate targeting: The long and short of it. *Biochem. Soc. Trans.* **2014**, *42*, 1129–1134. [[CrossRef](#)] [[PubMed](#)]
- Kilchert, C.; Wittmann, S.; Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 227–239. [[CrossRef](#)] [[PubMed](#)]
- Zinder, J.C.; Lima, C.D. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev.* **2017**, *31*, 88–100. [[CrossRef](#)] [[PubMed](#)]
- Hilleren, P.; McCarthy, T.; Rosbash, M.; Parker, R.; Jensen, T.H. Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature* **2001**, *413*, 538–542. [[CrossRef](#)] [[PubMed](#)]
- Milligan, L.; Torchet, C.; Allmang, C.; Shipman, T.; Tollervey, D. A nuclear surveillance pathway for mRNAs with defective polyadenylation. *Mol. Cell. Biol.* **2005**, *25*, 9996–10004. [[CrossRef](#)] [[PubMed](#)]
- Torchet, C.; Bousquet-Antonelli, C.; Milligan, L.; Thompson, E.; Kufel, J.; Tollervey, D. Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs. *Mol. Cell* **2002**, *9*, 1285–1296. [[CrossRef](#)]
- Kazerouninia, A.; Ngo, B.; Martinson, H.G. Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro. *RNA* **2010**, *16*, 197–210. [[CrossRef](#)] [[PubMed](#)]
- Di Giammartino, D.C.; Li, W.; Ogami, K.; Yashinskii, J.J.; Hoque, M.; Tian, B.; Manley, J.L. RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs. *Genes Dev.* **2014**, *28*, 2248–2260. [[CrossRef](#)] [[PubMed](#)]
- Lemieux, C.; Marguerat, S.; Lafontaine, J.; Barbezier, N.; Bahler, J.; Bachand, F. A pre-mRNA degradation pathway that selectively targets intron-containing genes requires the nuclear poly(A)-binding protein. *Mol. Cell* **2011**, *44*, 108–119. [[CrossRef](#)] [[PubMed](#)]
- Bousquet-Antonelli, C.; Presutti, C.; Tollervey, D. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **2000**, *102*, 765–775. [[CrossRef](#)]
- Gudipati, R.K.; Xu, Z.; Lebreton, A.; Seraphin, B.; Steinmetz, L.M.; Jacquier, A.; Libri, D. Extensive degradation of RNA precursors by the exosome in wild-type cells. *Mol. Cell* **2012**, *48*, 409–421. [[CrossRef](#)] [[PubMed](#)]
- Schneider, C.; Kudla, G.; Wlotzka, W.; Tuck, A.; Tollervey, D. Transcriptome-wide analysis of exosome targets. *Mol. Cell* **2012**, *48*, 422–433. [[CrossRef](#)] [[PubMed](#)]
- West, S.; Gromak, N.; Norbury, C.J.; Proudfoot, N.J. Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. *Mol. Cell* **2006**, *21*, 437–443. [[CrossRef](#)] [[PubMed](#)]
- Wyers, F.; Rougemaille, M.; Badis, G.; Rousselle, J.C.; Dufour, M.E.; Boulay, J.; Regnault, B.; Devaux, F.; Namane, A.; Seraphin, B.; et al. Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **2005**, *121*, 725–737. [[CrossRef](#)] [[PubMed](#)]

15. Neil, H.; Malabat, C.; d'Aubenton-Carafa, Y.; Xu, Z.; Steinmetz, L.M.; Jacquier, A. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **2009**, *457*, 1038–1042. [[CrossRef](#)] [[PubMed](#)]
16. Xu, Z.; Wei, W.; Gagneur, J.; Perocchi, F.; Clauder-Munster, S.; Camblong, J.; Guffanti, E.; Stutz, F.; Huber, W.; Steinmetz, L.M. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **2009**, *457*, 1033–1037. [[CrossRef](#)] [[PubMed](#)]
17. Preker, P.; Nielsen, J.; Kammler, S.; Lykke-Andersen, S.; Christensen, M.S.; Mapendano, C.K.; Schierup, M.H.; Jensen, T.H. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **2008**, *322*, 1851–1854. [[CrossRef](#)] [[PubMed](#)]
18. Flynn, R.A.; Almada, A.E.; Zamudio, J.R.; Sharp, P.A. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10460–10465. [[CrossRef](#)] [[PubMed](#)]
19. Szczepinska, T.; Kalisiak, K.; Tomecki, R.; Labno, A.; Borowski, L.S.; Kulinski, T.M.; Adamska, D.; Kosinska, J.; Dziembowski, A. DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. *Genome Res.* **2015**, *25*, 1622–1633. [[CrossRef](#)] [[PubMed](#)]
20. Ogami, K.; Richard, P.; Chen, Y.; Hoque, M.; Li, W.; Moresco, J.J.; Yates, J.R., 3rd; Tian, B.; Manley, J.L. An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev.* **2017**, *31*, 1257–1271. [[CrossRef](#)] [[PubMed](#)]
21. Henriques, T.; Gilchrist, D.A.; Nechaev, S.; Bern, M.; Muse, G.W.; Burkholder, A.; Fargo, D.C.; Adelman, K. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol. Cell* **2013**, *52*, 517–528. [[CrossRef](#)] [[PubMed](#)]
22. Andersson, R.; Gebhard, C.; Miguel-Escalada, I.; Hoof, I.; Bornholdt, J.; Boyd, M.; Chen, Y.; Zhao, X.; Schmidl, C.; Suzuki, T.; et al. An atlas of active enhancers across human cell types and tissues. *Nature* **2014**, *507*, 455–461. [[CrossRef](#)] [[PubMed](#)]
23. Schlackow, M.; Nojima, T.; Gomes, T.; Dhir, A.; Carmo-Fonseca, M.; Proudfoot, N.J. Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol. Cell* **2017**, *65*, 25–38. [[CrossRef](#)] [[PubMed](#)]
24. Schwalb, B.; Michel, M.; Zacher, B.; Fruhauf, K.; Demel, C.; Tresch, A.; Gagneur, J.; Cramer, P. TT-seq maps the human transient transcriptome. *Science* **2016**, *352*, 1225–1228. [[CrossRef](#)] [[PubMed](#)]
25. Makino, D.L.; Conti, E. Structure determination of an 11-subunit exosome in complex with RNA by molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 2226–2235. [[CrossRef](#)] [[PubMed](#)]
26. Wasmuth, E.V.; Januszzyk, K.; Lima, C.D. Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. *Nature* **2014**, *511*, 435–439. [[CrossRef](#)] [[PubMed](#)]
27. Makino, D.L.; Schuch, B.; Stegmann, E.; Baumgartner, M.; Basquin, C.; Conti, E. RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature* **2015**, *524*, 54–58. [[CrossRef](#)] [[PubMed](#)]
28. Lebreton, A.; Tomecki, R.; Dziembowski, A.; Seraphin, B. Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature* **2008**, *456*, 993–996. [[CrossRef](#)] [[PubMed](#)]
29. Schaeffer, D.; Tsanova, B.; Barbas, A.; Reis, F.P.; Dastidar, E.G.; Sanchez-Rotunno, M.; Arraiano, C.M.; van Hoof, A. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat. Struct. Mol. Biol.* **2009**, *16*, 56–62. [[CrossRef](#)] [[PubMed](#)]
30. Schneider, C.; Leung, E.; Brown, J.; Tollervey, D. The N-terminal pin domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. *Nucleic Acids Res.* **2009**, *37*, 1127–1140. [[CrossRef](#)] [[PubMed](#)]
31. Wasmuth, E.V.; Lima, C.D. The Rrp6 C-terminal domain binds RNA and activates the nuclear RNA exosome. *Nucleic Acids Res.* **2017**, *45*, 846–860. [[CrossRef](#)] [[PubMed](#)]
32. Allmang, C.; Petfalski, E.; Podtelejnikov, A.; Mann, M.; Tollervey, D.; Mitchell, P. The yeast exosome and human PM-Scl are related complexes of 3'→5' exonucleases. *Genes Dev.* **1999**, *13*, 2148–2158. [[CrossRef](#)] [[PubMed](#)]
33. Tomecki, R.; Kristiansen, M.S.; Lykke-Andersen, S.; Chlebowski, A.; Larsen, K.M.; Szczesny, R.J.; Drakowska, K.; Pastula, A.; Andersen, J.S.; Stepień, P.P.; et al. The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *EMBO J.* **2010**, *29*, 2342–2357. [[CrossRef](#)] [[PubMed](#)]

34. Staals, R.H.; Bronkhorst, A.W.; Schilders, G.; Slomovic, S.; Schuster, G.; Heck, A.J.; Raijmakers, R.; Pruijn, G.J. Dis3-like 1: A novel exoribonuclease associated with the human exosome. *EMBO J.* **2010**, *29*, 2358–2367. [[CrossRef](#)] [[PubMed](#)]
35. Shiomi, T.; Fukushima, K.; Suzuki, N.; Nakashima, N.; Noguchi, E.; Nishimoto, T. Human Dis3p, which binds to either GTP- or GDP-Ran, complements *Saccharomyces cerevisiae* dis3. *J. Biochem.* **1998**, *123*, 883–890. [[CrossRef](#)] [[PubMed](#)]
36. Schuch, B.; Feigenbutz, M.; Makino, D.L.; Falk, S.; Basquin, C.; Mitchell, P.; Conti, E. The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *EMBO J.* **2014**, *33*, 2829–2846. [[CrossRef](#)] [[PubMed](#)]
37. Falk, S.; Bonneau, F.; Ebert, J.; Kogel, A.; Conti, E. Mpp6 incorporation in the nuclear exosome contributes to RNA channeling through the Mtr4 helicase. *Cell Rep.* **2017**, *20*, 2279–2286. [[CrossRef](#)] [[PubMed](#)]
38. Wasmuth, E.V.; Zinder, J.C.; Zattas, D.; Das, M.; Lima, C.D. Structure and reconstitution of yeast Mpp6-nuclear exosome complexes reveals that Mpp6 stimulates RNA decay and recruits the Mtr4 helicase. *eLife* **2017**, *6*, e29062. [[CrossRef](#)] [[PubMed](#)]
39. Schneider, C.; Tollervey, D. Threading the barrel of the RNA exosome. *Trends Biochem. Sci.* **2013**, *38*, 485–493. [[CrossRef](#)] [[PubMed](#)]
40. Wasmuth, E.V.; Lima, C.D. Structure and activities of the eukaryotic RNA exosome. *Enzymes* **2012**, *31*, 53–75. [[PubMed](#)]
41. Zinder, J.C.; Wasmuth, E.V.; Lima, C.D. Nuclear RNA exosome at 3.1 Å reveals substrate specificities, RNA paths, and allosteric inhibition of Rrp44/Dis3. *Mol. Cell* **2016**, *64*, 734–745. [[CrossRef](#)] [[PubMed](#)]
42. Liu, J.J.; Bratkowski, M.A.; Liu, X.; Niu, C.Y.; Ke, A.; Wang, H.W. Visualization of distinct substrate-recruitment pathways in the yeast exosome by electron microscopy. *Nat. Struct. Mol. Biol.* **2014**, *21*, 95–102. [[CrossRef](#)] [[PubMed](#)]
43. Bonneau, F.; Basquin, J.; Ebert, J.; Lorentzen, E.; Conti, E. The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell* **2009**, *139*, 547–559. [[CrossRef](#)] [[PubMed](#)]
44. Malet, H.; Topf, M.; Clare, D.K.; Ebert, J.; Bonneau, F.; Basquin, J.; Drazkowska, K.; Tomecki, R.; Dziembowski, A.; Conti, E.; et al. RNA channelling by the eukaryotic exosome. *EMBO Rep.* **2010**, *11*, 936–942. [[CrossRef](#)] [[PubMed](#)]
45. Wang, H.W.; Wang, J.; Ding, F.; Callahan, K.; Bratkowski, M.A.; Butler, J.S.; Nogales, E.; Ke, A. Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3' end processing. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16844–16849. [[CrossRef](#)] [[PubMed](#)]
46. Han, J.; van Hoof, A. The RNA exosome channeling and direct access conformations have distinct in vivo functions. *Cell Rep.* **2016**, *16*, 3348–3358. [[CrossRef](#)] [[PubMed](#)]
47. Delan-Forino, C.; Schneider, C.; Tollervey, D. Transcriptome-wide analysis of alternative routes for RNA substrates into the exosome complex. *PLoS Genet.* **2017**, *13*, e1006699. [[CrossRef](#)] [[PubMed](#)]
48. Delan-Forino, C.; Schneider, C.; Tollervey, D. RNA substrate length as an indicator of exosome interactions in vivo. *Wellcome Open Res.* **2017**, *2*, 34. [[CrossRef](#)] [[PubMed](#)]
49. Meola, N.; Domanski, M.; Karadoulama, E.; Chen, Y.; Gentil, C.; Pultz, D.; Vitting-Seerup, K.; Lykke-Andersen, S.; Andersen, J.S.; Sandelin, A.; et al. Identification of a nuclear exosome decay pathway for processed transcripts. *Mol. Cell* **2016**, *64*, 520–533. [[CrossRef](#)] [[PubMed](#)]
50. Kadaba, S.; Krueger, A.; Trice, T.; Krecic, A.M.; Hinnebusch, A.G.; Anderson, J. Nuclear surveillance and degradation of hypomodified initiator tRNA^{met} in *S. cerevisiae*. *Genes Dev.* **2004**, *18*, 1227–1240. [[CrossRef](#)] [[PubMed](#)]
51. LaCava, J.; Houseley, J.; Saveanu, C.; Petfalski, E.; Thompson, E.; Jacquier, A.; Tollervey, D. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **2005**, *121*, 713–724. [[CrossRef](#)] [[PubMed](#)]
52. Vanacova, S.; Wolf, J.; Martin, G.; Blank, D.; Dettwiler, S.; Friedlein, A.; Langen, H.; Keith, G.; Keller, W. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol.* **2005**, *3*, e189. [[CrossRef](#)] [[PubMed](#)]
53. Houseley, J.; Tollervey, D. Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO Rep.* **2006**, *7*, 205–211. [[CrossRef](#)] [[PubMed](#)]
54. Hamill, S.; Wolin, S.L.; Reinisch, K.M. Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15045–15050. [[CrossRef](#)] [[PubMed](#)]

55. Holub, P.; Lalakova, J.; Cerna, H.; Pasulka, J.; Sarazova, M.; Hrazdilova, K.; Arce, M.S.; Hobor, F.; Stefl, R.; Vanacova, S. Air2p is critical for the assembly and RNA-binding of the TRAMP complex and the KOW domain of Mtr4p is crucial for exosome activation. *Nucleic Acids Res.* **2012**, *40*, 5679–5693. [[CrossRef](#)] [[PubMed](#)]
56. Fasken, M.B.; Leung, S.W.; Banerjee, A.; Kodani, M.O.; Chavez, R.; Bowman, E.A.; Purohit, M.K.; Rubinson, M.E.; Rubinson, E.H.; Corbett, A.H. Air1 zinc knuckles 4 and 5 and a conserved IWRXY motif are critical for the function and integrity of the Trf4/5-Air1/2-Mtr4 polyadenylation (TRAMP) RNA quality control complex. *J. Biol. Chem.* **2011**, *286*, 37429–37445. [[CrossRef](#)] [[PubMed](#)]
57. Wang, X.; Jia, H.; Jankowsky, E.; Anderson, J.T. Degradation of hypomodified tRNA_i^{Met} in vivo involves RNA-dependent ATPase activity of the DEXH helicase Mtr4p. *RNA* **2008**, *14*, 107–116. [[CrossRef](#)] [[PubMed](#)]
58. Schneider, C.; Anderson, J.T.; Tollervey, D. The exosome subunit Rps44 plays a direct role in RNA substrate recognition. *Mol. Cell* **2007**, *27*, 324–331. [[CrossRef](#)] [[PubMed](#)]
59. Wlotzka, W.; Kudla, G.; Granneman, S.; Tollervey, D. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J.* **2011**, *30*, 1790–1803. [[CrossRef](#)] [[PubMed](#)]
60. Kadaba, S.; Wang, X.; Anderson, J.T. Nuclear RNA surveillance in *Saccharomyces cerevisiae*: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA* **2006**, *12*, 508–521. [[CrossRef](#)] [[PubMed](#)]
61. Dez, C.; Houseley, J.; Tollervey, D. Surveillance of nuclear-restricted pre-ribosomes within a subnucleolar region of *Saccharomyces cerevisiae*. *EMBO J.* **2006**, *25*, 1534–1546. [[CrossRef](#)] [[PubMed](#)]
62. Grzechnik, P.; Kufel, J. Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol. Cell* **2008**, *32*, 247–258. [[CrossRef](#)] [[PubMed](#)]
63. Losh, J.S.; King, A.K.; Bakelar, J.; Taylor, L.; Loomis, J.; Rosenzweig, J.A.; Johnson, S.J.; van Hoof, A. Interaction between the RNA-dependent ATPase and poly(A) polymerase subunits of the TRAMP complex is mediated by short peptides and important for snoRNA processing. *Nucleic Acids Res.* **2015**, *43*, 1848–1858. [[CrossRef](#)] [[PubMed](#)]
64. Houseley, J.; Kotovic, K.; El Hage, A.; Tollervey, D. Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. *EMBO J.* **2007**, *26*, 4996–5006. [[CrossRef](#)] [[PubMed](#)]
65. Ciaia, D.; Bohnsack, M.T.; Tollervey, D. The mRNA encoding the yeast ARE-binding protein Cth2 is generated by a novel 3' processing pathway. *Nucleic Acids Res.* **2008**, *36*, 3075–3084. [[CrossRef](#)] [[PubMed](#)]
66. Roth, K.M.; Byam, J.; Fang, F.; Butler, J.S. Regulation of Nab2 mRNA 3'-end formation requires the core exosome and the Trf4p component of the TRAMP complex. *RNA* **2009**, *15*, 1045–1058. [[CrossRef](#)] [[PubMed](#)]
67. Bresson, S.; Tuck, A.; Staneva, D.; Tollervey, D. Nuclear RNA decay pathways AID rapid remodeling of gene expression in yeast. *Mol. Cell* **2017**, *65*, 787–800. [[CrossRef](#)] [[PubMed](#)]
68. Bernstein, J.; Patterson, D.N.; Wilson, G.M.; Toth, E.A. Characterization of the essential activities of *Saccharomyces cerevisiae* Mtr4p, a 3'→5' helicase partner of the nuclear exosome. *J. Biol. Chem.* **2008**, *283*, 4930–4942. [[CrossRef](#)] [[PubMed](#)]
69. Jia, H.; Wang, X.; Liu, F.; Guenther, U.P.; Srinivasan, S.; Anderson, J.T.; Jankowsky, E. The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. *Cell* **2011**, *145*, 890–901. [[CrossRef](#)] [[PubMed](#)]
70. Weir, J.R.; Bonneau, F.; Hentschel, J.; Conti, E. Structural analysis reveals the characteristic features of Mtr4, a DEXH helicase involved in nuclear RNA processing and surveillance. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 12139–12144. [[CrossRef](#)] [[PubMed](#)]
71. Jia, H.; Wang, X.; Anderson, J.T.; Jankowsky, E. RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 7292–7297. [[CrossRef](#)] [[PubMed](#)]
72. Falk, S.; Weir, J.R.; Hentschel, J.; Reichelt, P.; Bonneau, F.; Conti, E. The molecular architecture of the TRAMP complex reveals the organization and interplay of its two catalytic activities. *Mol. Cell* **2014**, *55*, 856–867. [[CrossRef](#)] [[PubMed](#)]
73. Patrick, E.M.; Srinivasan, S.; Jankowsky, E.; Comstock, M.J. The RNA helicase Mtr4p is a duplex-sensing translocase. *Nat. Chem. Biol.* **2017**, *13*, 99–104. [[CrossRef](#)] [[PubMed](#)]
74. Vasiljeva, L.; Buratowski, S. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol. Cell* **2006**, *21*, 239–248. [[CrossRef](#)] [[PubMed](#)]
75. Porrua, O.; Hobor, F.; Boulay, J.; Kubicek, K.; D'Aubenton-Carafa, Y.; Gudipati, R.K.; Stefl, R.; Libri, D. In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J.* **2012**, *31*, 3935–3948. [[CrossRef](#)] [[PubMed](#)]

76. Creamer, T.J.; Darby, M.M.; Jamonnak, N.; Schaughency, P.; Hao, H.; Wheelan, S.J.; Corden, J.L. Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet.* **2011**, *7*, e1002329. [[CrossRef](#)] [[PubMed](#)]
77. Porrua, O.; Libri, D. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat. Struct. Mol. Biol.* **2013**, *20*, 884–891. [[CrossRef](#)] [[PubMed](#)]
78. Hazelbaker, D.Z.; Marquardt, S.; Wlotzka, W.; Buratowski, S. Kinetic competition between RNA polymerase II and Sen1-dependent transcription termination. *Mol. Cell* **2013**, *49*, 55–66. [[CrossRef](#)] [[PubMed](#)]
79. Han, Z.; Libri, D.; Porrua, O. Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Res.* **2017**, *45*, 1355–1370. [[CrossRef](#)] [[PubMed](#)]
80. Arigo, J.T.; Carroll, K.L.; Ames, J.M.; Corden, J.L. Regulation of yeast *NRD1* expression by premature transcription termination. *Mol. Cell* **2006**, *21*, 641–651. [[CrossRef](#)] [[PubMed](#)]
81. Arigo, J.T.; Eyler, D.E.; Carroll, K.L.; Corden, J.L. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell* **2006**, *23*, 841–851. [[CrossRef](#)] [[PubMed](#)]
82. Conrad, N.K.; Wilson, S.M.; Steinmetz, E.J.; Patturajan, M.; Brow, D.A.; Swanson, M.S.; Corden, J.L. A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* **2000**, *154*, 557–571. [[PubMed](#)]
83. Mayer, A.; Heidemann, M.; Lidschreiber, M.; Schrieck, A.; Sun, M.; Hintermair, C.; Kremmer, E.; Eick, D.; Cramer, P. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **2012**, *336*, 1723–1725. [[CrossRef](#)] [[PubMed](#)]
84. Schulz, D.; Schwalb, B.; Kiesel, A.; Baejen, C.; Torkler, P.; Gagneur, J.; Soeding, J.; Cramer, P. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* **2013**, *155*, 1075–1087. [[CrossRef](#)] [[PubMed](#)]
85. Steinmetz, E.J.; Conrad, N.K.; Brow, D.A.; Corden, J.L. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **2001**, *413*, 327–331. [[CrossRef](#)] [[PubMed](#)]
86. Grzechnik, P.; Gdula, M.R.; Proudfoot, N.J. Pcf11 orchestrates transcription termination pathways in yeast. *Genes Dev.* **2015**, *29*, 849–861. [[CrossRef](#)] [[PubMed](#)]
87. Kubicek, K.; Cerna, H.; Holub, P.; Pasulka, J.; Hrossova, D.; Loehr, F.; Hofr, C.; Vanacova, S.; Stefl, R. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.* **2012**, *26*, 1891–1896. [[CrossRef](#)] [[PubMed](#)]
88. Vasiljeva, L.; Kim, M.; Mutschler, H.; Buratowski, S.; Meinhart, A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **2008**, *15*, 795–804. [[CrossRef](#)] [[PubMed](#)]
89. Tudek, A.; Porrua, O.; Kabzinski, T.; Lidschreiber, M.; Kubicek, K.; Fortova, A.; Lacroute, F.; Vanacova, S.; Cramer, P.; Stefl, R.; et al. Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell* **2014**, *55*, 467–481. [[CrossRef](#)] [[PubMed](#)]
90. Mitsuzawa, H.; Kanda, E.; Ishihama, A. Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res.* **2003**, *31*, 4696–4701. [[CrossRef](#)] [[PubMed](#)]
91. Lemay, J.F.; Marguerat, S.; Laroche, M.; Liu, X.; van Nues, R.; Hunyadkurti, J.; Hoque, M.; Tian, B.; Granneman, S.; Bahler, J.; et al. The Nrd1-like protein Seb1 coordinates cotranscriptional 3' end processing and polyadenylation site selection. *Genes Dev.* **2016**, *30*, 1558–1572. [[CrossRef](#)] [[PubMed](#)]
92. Marina, D.B.; Shankar, S.; Natarajan, P.; Finn, K.J.; Madhani, H.D. A conserved ncRNA-binding protein recruits silencing factors to heterochromatin through an RNAi-independent mechanism. *Genes Dev.* **2013**, *27*, 1851–1856. [[CrossRef](#)] [[PubMed](#)]
93. Wittmann, S.; Renner, M.; Watts, B.R.; Adams, O.; Huseyin, M.; Baejen, C.; El Omari, K.; Kilchert, C.; Heo, D.H.; Kecman, T.; et al. The conserved protein Seb1 drives transcription termination by binding RNA polymerase II and nascent RNA. *Nat. Commun.* **2017**, *8*, 14861. [[CrossRef](#)] [[PubMed](#)]
94. Liu, X.; Hoque, M.; Laroche, M.; Lemay, J.F.; Yurko, N.; Manley, J.L.; Bachand, F.; Tian, B. Comparative analysis of alternative polyadenylation in *S. cerevisiae* and *S. pombe*. *Genome Res.* **2017**, *27*, 1685–1695. [[CrossRef](#)] [[PubMed](#)]
95. Laroche, M.; Hunyadkurti, J.; Bachand, F. Polyadenylation site selection: Linking transcription and RNA processing via a conserved carboxy-terminal domain (CTD)-interacting protein. *Curr. Genet.* **2017**, *63*, 195–199. [[CrossRef](#)] [[PubMed](#)]

96. Patturajan, M.; Wei, X.; Berezney, R.; Corden, J.L. A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II. *Mol. Cell. Biol.* **1998**, *18*, 2406–2415. [[CrossRef](#)] [[PubMed](#)]
97. Becker, R.; Loll, B.; Meinhart, A. Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **2008**, *283*, 22659–22669. [[CrossRef](#)] [[PubMed](#)]
98. Fasken, M.B.; Larabee, R.N.; Corbett, A.H. Nab3 facilitates the function of the TRAMP complex in RNA processing via recruitment of Rrp6 independent of Nrd1. *PLoS Genet.* **2015**, *11*, e1005044. [[CrossRef](#)] [[PubMed](#)]
99. Thoms, M.; Thomson, E.; Bassler, J.; Gnadig, M.; Griesel, S.; Hurt, E. The exosome is recruited to RNA substrates through specific adaptor proteins. *Cell* **2015**, *162*, 1029–1038. [[CrossRef](#)] [[PubMed](#)]
100. Falk, S.; Tants, J.N.; Basquin, J.; Thoms, M.; Hurt, E.; Sattler, M.; Conti, E. Structural insights into the interaction of the nuclear exosome helicase Mtr4 with the pre-ribosomal protein Nop53. *RNA* **2017**, *23*, 1780–1787. [[CrossRef](#)] [[PubMed](#)]
101. Babour, A.; Shen, Q.; Dos-Santos, J.; Murray, S.; Gay, A.; Challal, D.; Fasken, M.; Palancade, B.; Corbett, A.; Libri, D.; et al. The chromatin remodeler ISW1 is a quality control factor that surveys nuclear mRNP biogenesis. *Cell* **2016**, *167*, 1201–1214. [[CrossRef](#)] [[PubMed](#)]
102. Buhler, M.; Haas, W.; Gygi, S.P.; Moazed, D. RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* **2007**, *129*, 707–721. [[CrossRef](#)] [[PubMed](#)]
103. Reyes-Turcu, F.E.; Zhang, K.; Zofall, M.; Chen, E.; Grewal, S.I. Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1132–1138. [[CrossRef](#)] [[PubMed](#)]
104. Zhang, K.; Fischer, T.; Porter, R.L.; Dhakshnamoorthy, J.; Zofall, M.; Zhou, M.; Veenstra, T.; Grewal, S.I. Ctr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA. *Science* **2011**, *331*, 1624–1627. [[CrossRef](#)] [[PubMed](#)]
105. Wang, S.W.; Stevenson, A.L.; Kearsley, S.E.; Watt, S.; Bahler, J. Global role for polyadenylation-assisted nuclear RNA degradation in posttranscriptional gene silencing. *Mol. Cell. Biol.* **2008**, *28*, 656–665. [[CrossRef](#)] [[PubMed](#)]
106. Buhler, M.; Spies, N.; Bartel, D.P.; Moazed, D. TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the *Schizosaccharomyces pombe* siRNA pathway. *Nat. Struct. Mol. Biol.* **2008**, *15*, 1015–1023. [[CrossRef](#)] [[PubMed](#)]
107. Larochelle, M.; Lemay, J.F.; Bachand, F. The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. *Nucleic Acids Res.* **2012**, *40*, 10240–10253. [[CrossRef](#)] [[PubMed](#)]
108. Pisacane, P.; Halic, M. Tailing and degradation of Argonaute-bound small RNAs protect the genome from uncontrolled RNAi. *Nat. Commun.* **2017**, *8*, 15332. [[CrossRef](#)] [[PubMed](#)]
109. Strasser, K.; Masuda, S.; Mason, P.; Pfannstiel, J.; Oppizzi, M.; Rodriguez-Navarro, S.; Rondon, A.G.; Aguilera, A.; Struhl, K.; Reed, R.; et al. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **2002**, *417*, 304–308. [[CrossRef](#)] [[PubMed](#)]
110. Zenklusen, D.; Vinciguerra, P.; Wyss, J.C.; Stutz, F. Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Mol. Cell. Biol.* **2002**, *22*, 8241–8253. [[CrossRef](#)] [[PubMed](#)]
111. Lee, N.N.; Chalamcharla, V.R.; Reyes-Turcu, F.; Mehta, S.; Zofall, M.; Balachandran, V.; Dhakshnamoorthy, J.; Taneja, N.; Yamanaka, S.; Zhou, M.; et al. Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell* **2013**, *155*, 1061–1074. [[CrossRef](#)] [[PubMed](#)]
112. Egan, E.D.; Braun, C.R.; Gygi, S.P.; Moazed, D. Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. *RNA* **2014**, *20*, 867–881. [[CrossRef](#)] [[PubMed](#)]
113. Zhou, Y.; Zhu, J.; Schermann, G.; Ohle, C.; Bendrin, K.; Sugioka-Sugiyama, R.; Sugiyama, T.; Fischer, T. The fission yeast MTREC complex targets cuts and unspliced pre-mRNAs to the nuclear exosome. *Nat. Commun.* **2015**, *6*, 7050. [[CrossRef](#)] [[PubMed](#)]
114. Harigaya, Y.; Tanaka, H.; Yamanaka, S.; Tanaka, K.; Watanabe, Y.; Tsutsumi, C.; Chikashige, Y.; Hiraoka, Y.; Yamashita, A.; Yamamoto, M. Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* **2006**, *442*, 45–50. [[CrossRef](#)] [[PubMed](#)]

115. Shichino, Y.; Yamashita, A.; Yamamoto, M. Meiotic long non-coding meiRNA accumulates as a dot at its genetic locus facilitated by Mmi1 and plays as a decoy to lure Mmi1. *Open Biol.* **2014**, *4*, 140022. [[CrossRef](#)] [[PubMed](#)]
116. Kilchert, C.; Wittmann, S.; Passoni, M.; Shah, S.; Granneman, S.; Vasiljeva, L. Regulation of mRNA levels by decay-promoting introns that recruit the exosome specificity factor Mmi1. *Cell Rep.* **2015**, *13*, 2504–2515. [[CrossRef](#)] [[PubMed](#)]
117. Yamashita, A.; Shichino, Y.; Tanaka, H.; Hiriart, E.; Touat-Todeschini, L.; Vavasseur, A.; Ding, D.Q.; Hiraoka, Y.; Verdel, A.; Yamamoto, M. Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. *Open Biol.* **2012**, *2*, 120014. [[CrossRef](#)] [[PubMed](#)]
118. Chen, H.M.; Futcher, B.; Leatherwood, J. The fission yeast RNA binding protein Mmi1 regulates meiotic genes by controlling intron specific splicing and polyadenylation coupled RNA turnover. *PLoS ONE* **2011**, *6*, e26804. [[CrossRef](#)] [[PubMed](#)]
119. Hiriart, E.; Vavasseur, A.; Touat-Todeschini, L.; Yamashita, A.; Gilquin, B.; Lambert, E.; Perot, J.; Shichino, Y.; Nazaret, N.; Boyault, C.; et al. Mmi1 RNA surveillance machinery directs RNAi complex RITS to specific meiotic genes in fission yeast. *EMBO J.* **2012**, *31*, 2296–2308. [[CrossRef](#)] [[PubMed](#)]
120. Yamashita, A.; Takayama, T.; Iwata, R.; Yamamoto, M. A novel factor Iss10 regulates Mmi1-mediated selective elimination of meiotic transcripts. *Nucleic Acids Res.* **2013**, *41*, 9680–9687. [[CrossRef](#)] [[PubMed](#)]
121. St-Andre, O.; Lemieux, C.; Perreault, A.; Lackner, D.H.; Bahler, J.; Bachand, F. Negative regulation of meiotic gene expression by the nuclear poly(A)-binding protein in fission yeast. *J. Biol. Chem.* **2010**, *285*, 27859–27868. [[CrossRef](#)] [[PubMed](#)]
122. Sugiyama, T.; Wanatabe, N.; Kitahata, E.; Tani, T.; Sugioka-Sugiyama, R. Red5 and three nuclear pore components are essential for efficient suppression of specific mRNAs during vegetative growth of fission yeast. *Nucleic Acids Res.* **2013**, *41*, 6674–6686. [[CrossRef](#)] [[PubMed](#)]
123. Andersen, P.R.; Domanski, M.; Kristiansen, M.S.; Storvall, H.; Ntini, E.; Verheggen, C.; Schein, A.; Bunkenborg, J.; Poser, I.; Hallais, M.; et al. The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1367–1376. [[CrossRef](#)] [[PubMed](#)]
124. Lubas, M.; Christensen, M.S.; Kristiansen, M.S.; Domanski, M.; Falkenby, L.G.; Lykke-Andersen, S.; Andersen, J.S.; Dziembowski, A.; Jensen, T.H. Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell* **2011**, *43*, 624–637. [[CrossRef](#)] [[PubMed](#)]
125. Rammelt, C.; Bilen, B.; Zavan, M.; Keller, W. PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *RNA* **2011**, *17*, 1737–1746. [[CrossRef](#)] [[PubMed](#)]
126. Ogami, K.; Cho, R.; Hoshino, S. Molecular cloning and characterization of a novel isoform of the non-canonical poly(A) polymerase PAPD7. *Biochem. Biophys. Res. Commun.* **2013**, *432*, 135–140. [[CrossRef](#)] [[PubMed](#)]
127. Molleston, J.M.; Sabin, L.R.; Moy, R.H.; Menghani, S.V.; Rausch, K.; Gordesky-Gold, B.; Hopkins, K.C.; Zhou, R.; Jensen, T.H.; Wilusz, J.E.; et al. A conserved virus-induced cytoplasmic TRAMP-like complex recruits the exosome to target viral RNA for degradation. *Genes Dev.* **2016**, *30*, 1658–1670. [[CrossRef](#)] [[PubMed](#)]
128. Berndt, H.; Harnisch, C.; Rammelt, C.; Stohr, N.; Zirkel, A.; Dohm, J.C.; Himmelbauer, H.; Tavanez, J.P.; Huttelmaier, S.; Wahle, E. Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* **2012**, *18*, 958–972. [[CrossRef](#)] [[PubMed](#)]
129. Shcherbik, N.; Wang, M.; Lapik, Y.R.; Srivastava, L.; Pestov, D.G. Polyadenylation and degradation of incomplete RNA polymerase I transcripts in mammalian cells. *EMBO Rep.* **2010**, *11*, 106–111. [[CrossRef](#)] [[PubMed](#)]
130. Nag, A.; Steitz, J.A. Tri-snRNP-associated proteins interact with subunits of the TRAMP and nuclear exosome complexes, linking RNA decay and pre-mRNA splicing. *RNA Biol.* **2012**, *9*, 334–342. [[CrossRef](#)] [[PubMed](#)]
131. Nagahama, M.; Yamazoe, T.; Hara, Y.; Tani, K.; Tsuji, A.; Tagaya, M. The AAA-ATPase NVL2 is a component of pre-ribosomal particles that interacts with the DExD/H-box RNA helicase DOB1. *Biochem. Biophys. Res. Commun.* **2006**, *346*, 1075–1082. [[CrossRef](#)] [[PubMed](#)]
132. Sudo, H.; Nozaki, A.; Uno, H.; Ishida, Y.; Nagahama, M. Interaction properties of human TRAMP-like proteins and their role in pre-rRNA 5'ETS turnover. *FEBS Lett.* **2016**, *590*, 2963–2972. [[CrossRef](#)] [[PubMed](#)]
133. Hiraishi, N.; Ishida, Y.; Nagahama, M. AAA-ATPase NVL2 acts on Mtr4-exosome complex to dissociate the nucleolar protein WDR74. *Biochem. Biophys. Res. Commun.* **2015**, *467*, 534–540. [[CrossRef](#)] [[PubMed](#)]

134. Hiraishi, N.; Ishida, Y.I.; Sudo, H.; Nagahama, M. WDR74 participates in an early cleavage of the pre-rRNA processing pathway in cooperation with the nucleolar AAA-ATPase NVL2. *Biochem. Biophys. Res. Commun.* **2017**, *495*, 116–123. [[CrossRef](#)] [[PubMed](#)]
135. Yoshikatsu, Y.; Ishida, Y.; Sudo, H.; Yuasa, K.; Tsuji, A.; Nagahama, M. NVL2, a nucleolar AAA-ATPase, is associated with the nuclear exosome and is involved in pre-rRNA processing. *Biochem. Biophys. Res. Commun.* **2015**, *464*, 780–786. [[CrossRef](#)] [[PubMed](#)]
136. Macias, S.; Cordiner, R.A.; Gautier, P.; Plass, M.; Caceres, J.F. DGCR8 acts as an adaptor for the exosome complex to degrade double-stranded structured RNAs. *Mol. Cell* **2015**, *60*, 873–885. [[CrossRef](#)] [[PubMed](#)]
137. Tseng, C.K.; Wang, H.F.; Burns, A.M.; Schroeder, M.R.; Gaspari, M.; Baumann, P. Human telomerase RNA processing and quality control. *Cell Rep.* **2015**, *13*, 2232–2243. [[CrossRef](#)] [[PubMed](#)]
138. Nguyen, D.; Grenier St-Sauveur, V.; Bergeron, D.; Dupuis-Sandoval, F.; Scott, M.S.; Bachand, F. A polyadenylation-dependent 3' end maturation pathway is required for the synthesis of the human telomerase RNA. *Cell Rep.* **2015**, *13*, 2244–2257. [[CrossRef](#)] [[PubMed](#)]
139. Shukla, S.; Schmidt, J.C.; Goldfarb, K.C.; Cech, T.R.; Parker, R. Inhibition of telomerase RNA decay rescues telomerase deficiency caused by dyskerin or PARN defects. *Nat. Struct. Mol. Biol.* **2016**, *23*, 286–292. [[CrossRef](#)] [[PubMed](#)]
140. Lubas, M.; Andersen, P.R.; Schein, A.; Dziembowski, A.; Kudla, G.; Jensen, T.H. The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep.* **2015**, *10*, 178–192. [[CrossRef](#)] [[PubMed](#)]
141. Hrossova, D.; Sikorsky, T.; Potesil, D.; Bartosovic, M.; Pasulka, J.; Zdrahal, Z.; Stefl, R.; Vanacova, S. RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs. *Nucleic Acids Res.* **2015**, *43*, 4236–4248. [[CrossRef](#)] [[PubMed](#)]
142. Bresson, S.M.; Conrad, N.K. The human nuclear poly(A)-binding protein promotes RNA hyperadenylation and decay. *PLoS Genet.* **2013**, *9*, e1003893. [[CrossRef](#)] [[PubMed](#)]
143. Bresson, S.M.; Hunter, O.V.; Hunter, A.C.; Conrad, N.K. Canonical poly(A) polymerase activity promotes the decay of a wide variety of mammalian nuclear RNAs. *PLoS Genet.* **2015**, *11*, e1005610. [[CrossRef](#)] [[PubMed](#)]
144. Beaulieu, Y.B.; Kleinman, C.L.; Landry-Voyer, A.M.; Majewski, J.; Bachand, F. Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS Genet.* **2012**, *8*, e1003078. [[CrossRef](#)] [[PubMed](#)]
145. Shi, Y.; Di Giammartino, D.C.; Taylor, D.; Sarkeshik, A.; Rice, W.J.; Yates, J.R., 3rd; Frank, J.; Manley, J.L. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell* **2009**, *33*, 365–376. [[CrossRef](#)] [[PubMed](#)]
146. Giacometti, S.; Benbahouche, N.E.H.; Domanski, M.; Robert, M.C.; Meola, N.; Lubas, M.; Bukenborg, J.; Andersen, J.S.; Schulze, W.M.; Verheggen, C.; et al. Mutually exclusive CBC-containing complexes contribute to RNA fate. *Cell Rep.* **2017**, *18*, 2635–2650. [[CrossRef](#)] [[PubMed](#)]
147. Iasillo, C.; Schmid, M.; Yahia, Y.; Maqbool, M.A.; Descostes, N.; Karadoulama, E.; Bertrand, E.; Andrau, J.C.; Jensen, T.H. ARS2 is a general suppressor of pervasive transcription. *Nucleic Acids Res.* **2017**, *45*, 10229–10241. [[CrossRef](#)] [[PubMed](#)]
148. Richard, P.; Ogami, K.; Chen, Y.; Feng, S.; Moresco, J.J.; Yates, J.R., 3rd; Manley, J.L. NRDE-2, the human homolog of fission yeast Nrl1, prevents DNA damage accumulation in human cells. Manuscript in preparation.
149. Morton, D.J.; Kuiper, E.G.; Jones, S.K.; Leung, S.W.; Corbett, A.H.; Fasken, M.B. The RNA exosome and RNA exosome-linked disease. *RNA* **2017**, *24*, 127–142. [[CrossRef](#)] [[PubMed](#)]
150. Chapman, M.A.; Lawrence, M.S.; Keats, J.J.; Cibulskis, K.; Sougnez, C.; Schinzel, A.C.; Harview, C.L.; Brunet, J.P.; Ahmann, G.J.; Adli, M.; et al. Initial genome sequencing and analysis of multiple myeloma. *Nature* **2011**, *471*, 467–472. [[CrossRef](#)] [[PubMed](#)]
151. Tomecki, R.; Drazkowska, K.; Kucinski, I.; Stodus, K.; Szczesny, R.J.; Gruchota, J.; Owczarek, E.P.; Kalisiak, K.; Dziembowski, A. Multiple myeloma-associated hDIS3 mutations cause perturbations in cellular RNA metabolism and suggest hDIS3 PIN domain as a potential drug target. *Nucleic Acids Res.* **2014**, *42*, 1270–1290. [[CrossRef](#)] [[PubMed](#)]
152. Blasius, M.; Wagner, S.A.; Choudhary, C.; Bartek, J.; Jackson, S.P. A quantitative 14-3-3 interaction screen connects the nuclear exosome targeting complex to the DNA damage response. *Genes Dev.* **2014**, *28*, 1977–1982. [[CrossRef](#)] [[PubMed](#)]

153. Tiedje, C.; Lubas, M.; Tehrani, M.; Menon, M.B.; Ronkina, N.; Rousseau, S.; Cohen, P.; Kotlyarov, A.; Gaestel, M. p38MAPK/MK2-mediated phosphorylation of RBM7 regulates the human nuclear exosome targeting complex. *RNA* **2015**, *21*, 262–278. [[CrossRef](#)] [[PubMed](#)]
154. Wang, X.; Arai, S.; Song, X.; Reichart, D.; Du, K.; Pascual, G.; Tempst, P.; Rosenfeld, M.G.; Glass, C.K.; Kurokawa, R. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **2008**, *454*, 126–130. [[CrossRef](#)] [[PubMed](#)]
155. Lloret-Llinares, M.; Mapendano, C.K.; Martlev, L.H.; Lykke-Andersen, S.; Jensen, T.H. Relationships between PROMPT and gene expression. *RNA Biol.* **2016**, *13*, 6–14. [[CrossRef](#)] [[PubMed](#)]
156. Li, X.; Manley, J.L. Cotranscriptional processes and their influence on genome stability. *Genes Dev.* **2006**, *20*, 1838–1847. [[CrossRef](#)] [[PubMed](#)]
157. Li, X.; Manley, J.L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **2005**, *122*, 365–378. [[CrossRef](#)] [[PubMed](#)]
158. Huertas, P.; Aguilera, A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **2003**, *12*, 711–721. [[CrossRef](#)] [[PubMed](#)]
159. Gaillard, H.; Aguilera, A. Transcription as a threat to genome integrity. *Annu. Rev. Biochem.* **2016**, *85*, 291–317. [[CrossRef](#)] [[PubMed](#)]
160. Richard, P.; Manley, J.L. R loops and links to human disease. *J. Mol. Biol.* **2017**, *429*, 3168–3180. [[CrossRef](#)] [[PubMed](#)]
161. Santos-Pereira, J.M.; Aguilera, A. R loops: New modulators of genome dynamics and function. *Nat. Rev. Genet.* **2015**, *16*, 583–597. [[CrossRef](#)] [[PubMed](#)]
162. Skourti-Stathaki, K.; Proudfoot, N.J. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* **2014**, *28*, 1384–1396. [[CrossRef](#)] [[PubMed](#)]
163. Sollier, J.; Cimprich, K.A. Breaking bad: R-loops and genome integrity. *Trends Cell Biol.* **2015**, *25*, 514–522. [[CrossRef](#)] [[PubMed](#)]
164. Luna, R.; Jimeno, S.; Marin, M.; Huertas, P.; Garcia-Rubio, M.; Aguilera, A. Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. *Mol Cell* **2005**, *18*, 711–722. [[CrossRef](#)] [[PubMed](#)]
165. Gavalda, S.; Gallardo, M.; Luna, R.; Aguilera, A. R-loop mediated transcription-associated recombination in *trf4Δ* mutants reveals new links between RNA surveillance and genome integrity. *PLoS ONE* **2013**, *8*, e65541. [[CrossRef](#)] [[PubMed](#)]
166. El Hage, A.; French, S.L.; Beyer, A.L.; Tollervey, D. Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* **2010**, *24*, 1546–1558. [[CrossRef](#)] [[PubMed](#)]
167. Manfrini, N.; Trovesi, C.; Wery, M.; Martina, M.; Cesena, D.; Descrimes, M.; Morillon, A.; d'Adda di Fagnagna, F.; Longhese, M.P. RNA-processing proteins regulate Mec1/ATR activation by promoting generation of RPA-coated ssDNA. *EMBO Rep.* **2015**, *16*, 221–231. [[CrossRef](#)] [[PubMed](#)]
168. Wahba, L.; Amon, J.D.; Koshland, D.; Vuica-Ross, M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell* **2011**, *44*, 978–988. [[CrossRef](#)] [[PubMed](#)]
169. Skourti-Stathaki, K.; Proudfoot, N.J.; Gromak, N. Human Senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **2011**, *42*, 794–805. [[CrossRef](#)] [[PubMed](#)]
170. Richard, P.; Feng, S.; Manley, J.L. A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes Dev.* **2013**, *27*, 2227–2232. [[CrossRef](#)] [[PubMed](#)]
171. Laffleur, B.; Basu, U.; Lim, J. RNA exosome and non-coding RNA-coupled mechanisms in AID-mediated genomic alterations. *J. Mol. Biol.* **2017**, *429*, 3230–3241. [[CrossRef](#)] [[PubMed](#)]
172. Basu, U.; Meng, F.L.; Keim, C.; Grinstein, V.; Pefanis, E.; Eccleston, J.; Zhang, T.; Myers, D.; Wasserman, C.R.; Wesemann, D.R.; et al. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* **2011**, *144*, 353–363. [[CrossRef](#)] [[PubMed](#)]
173. Sun, J.; Keim, C.D.; Wang, J.; Kazadi, D.; Oliver, P.M.; Rabadan, R.; Basu, U. E3-ubiquitin ligase Nedd4 determines the fate of AID-associated RNA polymerase II in B cells. *Genes Dev.* **2013**, *27*, 1821–1833. [[CrossRef](#)] [[PubMed](#)]

174. Pefanis, E.; Basu, U. RNA exosome regulates AID DNA mutator activity in the B cell genome. *Adv. Immunol.* **2015**, *127*, 257–308. [[PubMed](#)]
175. Pefanis, E.; Wang, J.; Rothschild, G.; Lim, J.; Chao, J.; Rabadan, R.; Economides, A.N.; Basu, U. Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* **2014**, *514*, 389–393. [[CrossRef](#)] [[PubMed](#)]
176. Pefanis, E.; Wang, J.; Rothschild, G.; Lim, J.; Kazadi, D.; Sun, J.; Federation, A.; Chao, J.; Elliott, O.; Liu, Z.P.; et al. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* **2015**, *161*, 774–789. [[CrossRef](#)] [[PubMed](#)]
177. Lim, J.; Giri, P.K.; Kazadi, D.; Laffleur, B.; Zhang, W.; Grinstein, V.; Pefanis, E.; Brown, L.M.; Ladewig, E.; Martin, O.; et al. Nuclear proximity of Mtr4 to RNA exosome restricts DNA mutational asymmetry. *Cell* **2017**, *169*, 523–537. [[CrossRef](#)] [[PubMed](#)]
178. Chi, B.; Wang, Q.; Wu, G.; Tan, M.; Wang, L.; Shi, M.; Chang, X.; Cheng, H. Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Res.* **2013**, *41*, 1294–1306. [[CrossRef](#)] [[PubMed](#)]
179. Cheng, H.; Dufu, K.; Lee, C.S.; Hsu, J.L.; Dias, A.; Reed, R. Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **2006**, *127*, 1389–1400. [[CrossRef](#)] [[PubMed](#)]
180. Nojima, T.; Hirose, T.; Kimura, H.; Hagiwara, M. The interaction between cap-binding complex and RNA export factor is required for intronless mRNA export. *J. Biol. Chem.* **2007**, *282*, 15645–15651. [[CrossRef](#)] [[PubMed](#)]
181. Masuda, S.; Das, R.; Cheng, H.; Hurt, E.; Dorman, N.; Reed, R. Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev.* **2005**, *19*, 1512–1517. [[CrossRef](#)] [[PubMed](#)]
182. Chi, B.; Wang, K.; Du, Y.; Gui, B.; Chang, X.; Wang, L.; Fan, J.; Chen, S.; Wu, X.; Li, G.; et al. A sub-element in PRE enhances nuclear export of intronless mRNAs by recruiting the TREX complex via ZC3H18. *Nucleic Acids Res.* **2014**, *42*, 7305–7318. [[CrossRef](#)] [[PubMed](#)]
183. Fan, J.; Kuai, B.; Wu, G.; Wu, X.; Chi, B.; Wang, L.; Wang, K.; Shi, Z.; Zhang, H.; Chen, S.; et al. Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. *EMBO J.* **2017**, *36*, 2870–2886. [[CrossRef](#)] [[PubMed](#)]
184. Ogami, K.; Manley, J.L. Mtr4/ZFC3H1 protects polysomes through nuclear RNA surveillance. *Cell Cycle* **2017**, *16*, 1999–2000. [[CrossRef](#)] [[PubMed](#)]
185. Johnson, S.A.; Cubberley, G.; Bentley, D.L. Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11. *Mol. Cell* **2009**, *33*, 215–226. [[CrossRef](#)] [[PubMed](#)]
186. Ruepp, M.D.; Aringhieri, C.; Vivarelli, S.; Cardinale, S.; Paro, S.; Schumperli, D.; Barabino, S.M. Mammalian pre-mRNA 3' end processing factor CF Im68 functions in mRNA export. *Mol. Biol. Cell.* **2009**, *20*, 5211–5223. [[CrossRef](#)] [[PubMed](#)]
187. Tran, D.D.; Saran, S.; Williamson, A.J.; Pierce, A.; Dittrich-Breiholz, O.; Wiehlmann, L.; Koch, A.; Whetton, A.D.; Tamura, T. THOC5 controls 3' end-processing of immediate early genes via interaction with polyadenylation specific factor 100 (CPSF100). *Nucleic Acids Res.* **2014**, *42*, 12249–12260. [[CrossRef](#)] [[PubMed](#)]
188. Katahira, J.; Okuzaki, D.; Inoue, H.; Yoneda, Y.; Maehara, K.; Ohkawa, Y. Human TREX component THOC5 affects alternative polyadenylation site choice by recruiting mammalian cleavage factor I. *Nucleic Acids Res.* **2013**, *41*, 7060–7072. [[CrossRef](#)] [[PubMed](#)]
189. Shi, M.; Zhang, H.; Wu, X.; He, Z.; Wang, L.; Yin, S.; Tian, B.; Li, G.; Cheng, H. ALYREF mainly binds to the 5' and the 3' regions of the mRNA in vivo. *Nucleic Acids Res.* **2017**, *45*, 9640–9653. [[CrossRef](#)] [[PubMed](#)]
190. Ingolia, N.T. Ribosome footprint profiling of translation throughout the genome. *Cell* **2016**, *165*, 22–33. [[CrossRef](#)] [[PubMed](#)]
191. Chew, G.L.; Pauli, A.; Rinn, J.L.; Regev, A.; Schier, A.F.; Valen, E. Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development* **2013**, *140*, 2828–2834. [[CrossRef](#)] [[PubMed](#)]
192. Zhou, P.; Zhang, Y.; Ma, Q.; Gu, F.; Day, D.S.; He, A.; Zhou, B.; Li, J.; Stevens, S.M.; Romo, D.; et al. Interrogating translational efficiency and lineage-specific transcriptomes using ribosome affinity purification. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 15395–15400. [[CrossRef](#)] [[PubMed](#)]

193. Ingolia, N.T.; Brar, G.A.; Stern-Ginossar, N.; Harris, M.S.; Talhouarne, G.J.; Jackson, S.E.; Wills, M.R.; Weissman, J.S. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* **2014**, *8*, 1365–1379. [[CrossRef](#)] [[PubMed](#)]
194. Van Heesch, S.; van Iterson, M.; Jacobi, J.; Boymans, S.; Essers, P.B.; de Bruijn, E.; Hao, W.; MacInnes, A.W.; Cuppen, E.; Simonis, M. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* **2014**, *15*, R6. [[CrossRef](#)] [[PubMed](#)]
195. Fields, A.P.; Rodriguez, E.H.; Jovanovic, M.; Stern-Ginossar, N.; Haas, B.J.; Mertins, P.; Raychowdhury, R.; Hacohen, N.; Carr, S.A.; Ingolia, N.T.; et al. A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. *Mol. Cell* **2015**, *60*, 816–827. [[CrossRef](#)] [[PubMed](#)]
196. Ji, Z.; Song, R.; Regev, A.; Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **2015**, *4*, e08890. [[CrossRef](#)] [[PubMed](#)]
197. Calviello, L.; Mukherjee, N.; Wyler, E.; Zauber, H.; Hirsekorn, A.; Selbach, M.; Landthaler, M.; Obermayer, B.; Ohler, U. Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods* **2016**, *13*, 165–170. [[CrossRef](#)] [[PubMed](#)]
198. Sinturel, F.; Gerber, A.; Mauvoisin, D.; Wang, J.; Gatfield, D.; Stubblefield, J.J.; Green, C.B.; Gachon, F.; Schibler, U. Diurnal oscillations in liver mass and cell size accompany ribosome assembly cycles. *Cell* **2017**, *169*, 651–663. [[CrossRef](#)] [[PubMed](#)]
199. Mistry, D.S.; Chen, Y.; Sen, G.L. Progenitor function in self-renewing human epidermis is maintained by the exosome. *Cell Stem Cell* **2012**, *11*, 127–135. [[CrossRef](#)] [[PubMed](#)]
200. Skamagki, M.; Zhang, C.; Ross, C.A.; Ananthanarayanan, A.; Liu, Z.; Mu, Q.; Basu, U.; Wang, J.; Zhao, R.; Li, H.; et al. RNA exosome complex-mediated control of redox status in pluripotent stem cells. *Stem Cell Rep.* **2017**, *9*, 1053–1061. [[CrossRef](#)] [[PubMed](#)]
201. Taleahmad, S.; Mirzaei, M.; Parker, L.M.; Hassani, S.N.; Mollamohammadi, S.; Sharifi-Zarchi, A.; Haynes, P.A.; Baharvand, H.; Salekdeh, G.H. Proteome analysis of ground state pluripotency. *Sci. Rep.* **2015**, *5*, 17985. [[CrossRef](#)] [[PubMed](#)]
202. Fazzio, T.G.; Huff, J.T.; Panning, B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* **2008**, *134*, 162–174. [[CrossRef](#)] [[PubMed](#)]
203. Onderak, A.M.; Anderson, J.T. Loss of the RNA helicase SKIV2L2 impairs mitotic progression and replication-dependent histone mRNA turnover in murine cell lines. *RNA* **2017**, *23*, 910–926. [[CrossRef](#)] [[PubMed](#)]
204. Rialdi, A.; Hultquist, J.; Jimenez-Morales, D.; Peralta, Z.; Campisi, L.; Fenouil, R.; Moshkina, N.; Wang, Z.Z.; Laffleur, B.; Kaake, R.M.; et al. The RNA exosome syncs IAV-RNAPII transcription to promote viral ribogenesis and infectivity. *Cell* **2017**, *169*, 679–692. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

**Chapter 3. Identification of a Large Pool of Human Polyadenylated Enhancer RNAs and
Characterization of Their Biogenesis and Degradation**

Identification of a Large Pool of Human Polyadenylated Enhancer RNAs and Characterization of Their Biogenesis and Degradation Mechanisms

Yaqiong Chen¹, Yuch-Lin Tsai¹, Koichi Ogami^{1,2}, Dinghai Zheng³, Risa Kawaguchi^{4,5}, Bin Tian³
and James L. Manley^{1*}

¹. *Department of Biological Sciences, Columbia University, New York, New York 10027, USA.*

². *Department of Biological Chemistry, Graduate School of Pharmaceutical Sciences, Nagoya City University, Nagoya 467-8603, Japan*

³. *Department of Microbiology, Biochemistry and Molecular Genetics, Rutgers New Jersey Medical School, Newark, New Jersey 07101, USA.*

⁴. *Machine Learning Research Team, Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Koto-ku, Tokyo 135-0064, Japan*

⁵. *Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba, 277-8561, Japan*

* Corresponding author

E-mail: jl2@columbia.edu

This work was supported by NIH grant R35 GM118136.

Manuscript information:

the number of text pages: , Figures: (Supplementary Figs:)

the number of characters in the abstract:

the total # characters in the main text: (excluding supplementary figure legends)

Abstract

The human genome consists of approximately 3 billion base pairs, but only 1 to 2 % of it codes for proteins. Much of the non-coding DNA is however transcribed and generates non-coding RNAs, including for example enhancer-derived RNAs, or eRNAs. Previous studies have suggested that most eRNAs are transcribed by RNA polymerase II (RNAP II), but not polyadenylated. We now describe a genome-wide analysis of eRNAs in which the nuclear surveillance machinery has been compromised, and our analyses indicate that a large fraction of eRNAs are in fact polyadenylated. Assessing the mechanisms involved, we found that the Integrator complex plays an important role in polyadenylated (pA+) eRNA biogenesis, and that their exosome-dependent degradation requires two cofactor complexes containing the RNA helicase Mtr4. We also show that the canonical poly(A) polymerases PAP- α and PAP- γ play a major role in the 3' end processing of pA+ eRNA. Additionally, we find that under deficient nuclear surveillance conditions, pA+ eRNAs accumulate in the cytoplasm and associate with polysomes, suggesting that at least some are translated. Our findings not only identified an unexpected large pool of unstable polyadenylated eRNAs, but also revealed novel insights into their metabolism.

Introduction

Continuing improvements in deep-sequencing techniques have revealed that a large portion of the nonprotein-coding human genome is transcribed. Amongst the transcripts produced are an abundant class of unstable RNAs transcribed by RNA polymerase II (RNAP II) from transcriptional enhancers known as enhancer (e) RNAs (ENCODE Project Consortium, 2012). Enhancers were originally defined as DNA elements that activate transcription of nearby protein-coding genes, characterized by a lack of orientation and distance requirements from target promoters (Banerji et al., 1981; Bulger and Groudine, 2011; Moreau et al., 1981; Schaffner, 2015). Enhancers can be located upstream, downstream, or in the coding region of genes, but are frequently hundreds if not many thousands of base pairs distant from target promoters (Bulger and Groudine, 2011).

The current view of how enhancers function involves chromatin “looping,” such that enhancer and promoter regions are in direct contact with each other. This reflects association of transcription factors bound to enhancers (Levine, 2010) with factors at the promoter, with this association facilitating gene activation (Bulger and Groudine, 2011; (Furlong and Levine, 2018) and allowing bidirectional exchange of components of the transcriptional machinery (Lam et al., 2014; Li et al., 2016). Consistent with this, RNAP II has been shown to localize to enhancers (Heintzman et al., 2007; Koch et al., 2008; Kim et al., 2010). While this may reflect part of the gene activation mechanism, it is consistent with the fact that enhancers are frequently transcribed (De Santa et al., 2010; Kim et al., 2010; Masternak et al., 2003; Rogan et al., 2004; Tuan et al., 1992). Indeed, a number of studies have illustrated architectural and functional similarities

between enhancers and promoters: enrichment of general transcription factors (e.g. TATA-binding protein, TBP) and serine 5-phosphorylated RNAP II, DNase I hypersensitivity, related histone modification patterns, and bidirectional transcription (Core et al., 2014; Furlong and Levine, 2018; Kim et al., 2010; Kim and Shiekhattar, 2015; Whalen et al., 2016).

Active enhancers are characterized by specific chromatin modifications and by bidirectional transcription originating from the enhancer region. When an enhancer is activated, typically by binding a “pioneer” transcription factor (Schaffner, 2015), it undergoes histone modifications leading to a looser chromatin structure, enabling transcription factor binding and RNAP II recruitment to initiate bidirectional transcription (Heinz et al., 2015; Li et al., 2016). eRNA expression can thus be a signature of enhancer activation. Recent genome-wide profiling has annotated 40,000-65,000 active enhancers, based on epigenetic markers, binding of transcriptional co-activators and detection of 5' capped RNAs (Andersson et al., 2014; Djebali et al., 2012; Li et al., 2016). Another genome-wide study of nearly 9,000 cancer patient samples using TCGA RNA-seq data further validated the existence of these enhancers and detected on average approximately 4,500 active enhancers in any given cancer type (Chen et al., 2018).

Enhancer RNAs are exceptionally unstable due to the constant degradation by nuclear RNA exosome complex. The RNA helicase Mtr4 is an essential co-factor for the eukaryotic nuclear exosome, and functions by unwinding RNA substrates and delivering them to the exosome core (Johnson and Jackson, 2013; Ogami et al., 2018). Three Mtr4-containing exosome adaptor complexes have been identified so far in human nuclei, known as hTRAMP, NEXT and PAXT/PPC. The hTRAMP complex consists of the non-canonical poly(A) polymerase TRF4-

2/PAPD5, the zinc-knuckle protein ZCCHC7 and Mtr4, and is mainly responsible for rRNA processing and decay (Lubas et al., 2011). The Nuclear Exosome Targeting (NEXT) complex facilitates the turnover of various short-lived nuclear RNA substrates, including eRNAs, upstream antisense RNAs (uaRNAs, also known as promoter upstream transcripts, PROMPTs), and 3' extended products from snRNA and histone genes (Lubas et al., 2015; Lubas et al., 2011). Later, the Poly(A) tail Exosome Targeting (PAXT) or Polysome Protector Complex (PPC) was found to be involved in the degradation of more extensively polyadenylated nuclear RNAs, including a small fraction of eRNAs, prematurely terminated RNAs produced from many protein coding genes (ptRNAs), and uaRNAs (Meola et al., 2016; Ogami et al., 2017). Besides the shared subunit Mtr4, NEXT contains the Zn-finger protein ZCCHC8 and the RNA-binding protein RBM7, while the Zn-knuckle protein ZFC3H1, and perhaps the nuclear poly(A)-binding protein PABPN1, are part of the PAXT/PPC complex. The complex was dubbed the PPC because depletion of either subunit results in stabilization, cytoplasmic transport and association of normally unstable, nuclear restricted RNAs with polysomes, and inhibition of mRNA translation (Ogami et al., 2017).

While the above studies are making inroads in understanding eRNA turnover, many aspects of eRNA biogenesis remain unclear. However, one important factor is the Integrator complex. The Integrator was originally discovered to be crucial for 3' processing of certain snRNAs (Baillat et al., 2005). Additionally, Integrator is necessary for productive transcriptional elongation of EGF target genes following activation (Gardini et al., 2014). Another study on EGF-induced eRNAs found that depletion of Integrator subunits led to increased level of RNAP II on enhancer regions, and faulty termination and extended transcription of EGF-induced eRNAs,

providing strong evidence that Integrator functions in eRNA biogenesis, specifically 3' processing. Notably, this study also found a significant increase of polyadenylated eRNAs derived from the EGF-responsive enhancers under the depletion of Integrator (Lai et al., 2015). Two Integrator subunits, INTS9 and INTS11, constitute the catalytic core, and show sequence homology with the pre-mRNA 3' processing factors CPSF100 and CPSF73, respectively, while INTS4 associates with INTS9/11 as a "Symplekin-like" scaffold, to form the Integrator cleavage module (Albrecht et al., 2018). Interestingly, multiple Integrator subunits were detected in our previous proteomics analysis of the purified pre-mRNA 3' processing complex (Shi et al., 2009), suggesting their possible involvement in mRNA processing. eRNAs, however, have been thought to be largely non-polyadenylated, with previous studies suggesting that ~90% of eRNAs lack a poly(A) tail (Andersson et al., 2014; Li et al., 2016).

Here we have further investigated eRNA biogenesis, especially how their 3' ends are fashioned. By impairing the nuclear exosome degradation system, we show that a large fraction of eRNAs are indeed polyadenylated (pA⁺ eRNAs). Consistent with this, pA⁺ eRNAs typically contain AAUAAA-like and G/U-rich signal sequences near their 3' ends. However, we found that the Integrator functions in the biosynthesis of pA⁺ eRNAs as well as ptRNAs, and that both the NEXT and PAXT/PPC complexes are required for pA⁺ eRNA degradation. Moreover, we provide evidence that the canonical poly(A) polymerases, PAP- α/γ , are crucial for assuring the proper 3' cleavage of pA⁺ eRNAs. We also find that pA⁺ eRNAs, under nuclear surveillance deficient conditions, can be transported to cytoplasm and associate with polysomes. Our findings have provided new insights into eRNA biogenesis and degradation mechanism, and suggest novel

similarities between eRNAs and other RNAP II transcribed RNAs.

Results

A large fraction of eRNAs detected under nuclear surveillance deficient conditions are polyadenylated

As described above, previous work has indicated that depletion of exosome cofactors such as Mtr4 results in increased accumulation of otherwise unstable nuclear polyadenylated RNAs. To extend these findings, we set out to examine whether under such conditions there might be a significant accumulation of polyadenylated eRNAs (pA+ eRNAs). In our previous study using 3' region extraction and deep sequencing (3'READS; Hoque et al., 2013), we detected significant accumulation of polyadenylated ua- and pt- RNAs following depletion of Mtr4, but not other NEXT subunits (Ogami et al., 2017). To extend this analysis to eRNAs, we reanalyzed the data by mapping the reads to enhancer regions characterized in the ENCODE project (ENCODE Project Consortium, 2012). To reduce false detection, each pA+ eRNA was annotated only when ≥ 2 reads were assigned to an enhancer region in ≥ 2 samples. We detected 3919 pA+ eRNAs in total under NEXT subunit depletion conditions (KD of Mtr4, ZCCHC8 or RBM7), while only 1761 pA+ eRNAs under control conditions (Figure 1A). Thus, blocking the nuclear surveillance pathway allowed us to capture more pA+ eRNAs. Additionally, we found that 32% of detected pA+ eRNAs were transcribed bidirectionally, whereas 68% were unidirectional. Unidirectional eRNAs were also referred as 1d-eRNAs, whereas bidirectional as 2d-eRNAs in previous studies. Consistent with our results, 1d-eRNAs have been found to be often polyadenylated (Koch et al.,

2011; Natoli and Andrau, 2012).

Next, we classified the 3919 pA⁺ eRNAs into three groups based on their genomic locations, and dubbed them the intergenic group, the intragenic group, and the ua group (Figure S1A). The definition of these three groups is illustrated in Figure S1A: Intergenic eRNAs are eRNAs generated from enhancers located between genes; the intragenic group includes eRNAs generated from enhancers located within genes; and the ua group is defined as eRNAs produced from regions within 4 kb upstream of transcription start sites of genes. These overlap with, and can be difficult to distinguish from, uaRNAs; hence the name.

There were 1457 (37% of the total pA⁺ eRNAs detected) pA⁺ eRNAs in the intergenic group, 1843 (47%) in the intragenic group, and 619 (16%) in the ua group (Figure S1B, C). We also compared our 3'READS data with published RNA-seq data under NEXT complex depletion conditions, and found that Mtr4 depletion can indeed capture eRNAs with poly(A) tails, which were sometimes absent in these previous RNA-seq experiments (Meola et al., 2016). For example, novel poly(A) tail peaks were detected under Mtr4 KD conditions on enhancer EN515 and EN27780 RNAs in 3'READS but not RNA-seq analysis (Figure S1D). For other eRNAs, such as EN29068 and EN70250, reads were consistent between 3'READS and RNA-seq data (Figure S1D). These results raise the possibility that polyadenylation of eRNAs may be more prevalent than previously thought.

We next wished both to confirm directly that select eRNAs identified in the above analysis were polyadenylated and also analyze the length of the added poly(A) tails. To this end, we performed an RNA ligation-mediated poly(A) test (RL-PAT) assay to characterize the poly(A)

tail length of several pA⁺ eRNAs that we detected (see Figure S2 for a schematic illustration of the RL-PAT assay) (Rassa et al., 2000). siRNA knockdown (KD) of Mtr4, as well as PABPN1, a subunit of the PAXT complex (Meola et al., 2016), led to accumulation of pA⁺ eRNAs (originating from both plus (+) and minus (-) strands) with complicated polyadenylation patterns that slightly differed between Mtr4 and PABPN1 KDs (Figure 1B). All three eRNAs tested (EN11/14/17) had poly(A) tails of various lengths, from ~100 nts to, remarkably, ~1000 nts. Treatment with RNase H and oligo(dT) confirmed that these were indeed polyadenylated eRNAs. We next performed RT-qPCR (oligo(dT)-primed RT) of bidirectional pA⁺ eRNAs EN11/14/17 following Mtr4, ZFC3H1 or PABPN1 depletion. All the pA⁺ eRNAs examined accumulated after siRNA KDs compared to control (Figure 1C). Noticeably, EN11^{+/-} and EN14^{+/-} were more sensitive to Mtr4 depletion than ZFC3H1/PABPN1 KDs, indicating a minor role of PAXT/PPC complex in their degradation. In contrast, EN17^{+/-} eRNAs were very sensitive to both Mtr4 and PAXT subunit depletion, revealing a more significant role of the PAXT/PPC complex in EN17 degradation. This data indicates that pA⁺ eRNA transcripts are targeted by various exosome-targeting pathways.

PAXT/PPC and NEXT complexes share half of the pA⁺ eRNA substrates, most of which are sensitive to the depletion of the Integrator complex

The above data suggest that a substantial fraction of eRNAs are polyadenylated. Given that this was unexpected, we decided to investigate this using a more sensitive and accurate version of the 3'READS technique, 3'READS⁺ (Zheng and Tian, 2017). We performed 3'READS⁺ analysis of RNA extracted from control HeLa cells or cells depleted of the following three eRNA

regulators: ZCCHC8 (NEXT), ZFC3H1 (PAXT/PPC) or INTS4 (Integrator) (Figure 2A). INTS4 KD most likely disrupts the structural core of the Integrator complex since studies have shown that INTS4 functions as a “Symplekin-like” scaffold for the association of the catalytic module INTS9/11 (Albrecht et al., 2018). INTS4 depletion thus allowed us to examine the role of Integrator in pA⁺ eRNA biogenesis, while KD of subunits of the NEXT and PAXT/PPC complex should stabilize pA⁺ eRNAs, as above. We mapped the 3'READS⁺ signals obtained in each condition to ENCODE annotated enhancer regions, and found a total of 6453 pA⁺ eRNAs (Figure 2B). While the ENCODE project has generated large-scale annotation of ~53,000 enhancers expressed in the human genome across multiple tissues and samples (Djebali et al., 2012), a recent genome-wide study of nearly 9,000 cancer patient samples using TCGA RNA-seq data has shown that approximately 4,500 enhancers on average were active in a given cancer type (Chen et al., 2018). The numbers of active enhancers in cancer patients (~4,500) is comparable with the number of pA⁺ eRNAs detected by 3'READS⁺ (6,453), indicating that polyadenylation of eRNAs is in fact a prevalent phenomenon.

We next classified the pA⁺ eRNAs based on their genomic locations into the three categories described above, intergenic, intragenic, and ua-like (Figure 2B, C). While a subset of enhancer regions overlap with other known transcribed elements, including for example introns, ncRNAs and uaRNAs, they all have been characterized with classic enhancer-specific chromatin status, including DNase hypersensitivity, p300 binding, H3K27Ac, and H3K4Me1 in the ENCODE project, reflecting their functions as typical enhancers. However, to eliminate any potential signal contamination, further analysis of the intergenic and intragenic eRNAs was

conducted by only considering eRNAs that do not overlap with any other known transcribed elements, such as 5'UTR, 3'UTR, CDS, and ncRNAs. We analyzed the ua group as well, which was also named as ua/eRNAs. This generated 1,859 pA+ eRNAs in the intergenic group, 1,421 in the intragenic group, and 619 in the ua group. Among these 3,899 pA+ eRNAs, 2,875 were significantly upregulated (>2 fold increase as compared to siCtrl) under KD conditions (Figure 2D), while 518 were downregulated (>2 fold decrease as compared to siCtrl) (Figure S3).

To understand the target specificity of NEXT, PAXT/PPC and Integrator complexes, we also classified the pA+ eRNAs according to the identity of the factor depleted (Figure 2D). We found that regardless of genomic annotations, depletion of both NEXT and PAXT/PPC subunits led to significant accumulation of pA+ eRNAs, with substantial overlap (Figure 2D). However, ZFC3H1 depletion affected a larger number of pA+ eRNAs compared to ZCCHC8 depletion. A total of 1190 intergenic pA+ eRNAs were detected under ZCCHC8 and ZFC3H1 deficient conditions, with 587 (~49%) overlapping. Additionally, we observed ~50% overlap (486) for intragenic eRNAs, and ~55% (425) for ua/eRNAs. Thus PAXT/PPC and NEXT complexes both control pA+ eRNA accumulation, and share approximately half of the pA+ eRNAs we detected. Interestingly, INTS4 KD also resulted in enhanced pA+ eRNA accumulation. In fact, there was considerable overlap with ZCCHC8 and ZFC3H1 KDs (see the overlapping areas of the three factor KD in Figure 2D). Only a very small fraction of INTS4-depletion sensitive pA+ eRNAs were not also affected by ZCCHC8 and ZFC3H1 KD (81 (~6.8%) for intragenic eRNAs, 44 (~4.5%) for the intergenic eRNAs, and 16 (~2.1%) for ua/eRNAs). These results suggest that depletion of INTS4, and presumably inactivation of the Integrator complex, actually enhanced

synthesis of pA⁺ eRNAs, and that these eRNAs were largely the same as those subject to NEXT- or PAXT/PPC-mediated decay. We discuss the mechanistic implications of these findings below.

We validated our 3'READS⁺ analysis by performing oligo(dT) primed RT-qPCR and RL-PAT assays on six select pA⁺ eRNAs. We confirmed the bidirectional expression of the pA⁺ eRNAs tested and their sensitivities towards the KD subunit (Figure 2E). Additionally, the RL-PAT assay results were consistent with our 3'READS⁺ analysis, again revealing long poly (A) tails (from 50 nt up to 1000 nt) (Figure S4). Thus, our combined results indicate that a large fraction of eRNAs are indeed polyadenylated, and that these poly(A) tails can be unusually long.

Alternative polyadenylation is rare in eRNAs

The 3'READS⁺ analysis allowed us to thoroughly examine the global effect of ZCCHC8, ZFC3H1 and INTS4 KD on various polyadenylated RNA substrates. It is widely known that alternative polyadenylation (APA) is a very common mechanism of gene control found in a majority of mRNAs across all eukaryotic species (reviewed by Tian and Manley, 2017). Given the large number of pA⁺ eRNAs detected in our study, we were curious whether APA also occurs in those eRNAs, and if so are they regulated by any of these factors?

To investigate APA in eRNAs, we reanalyzed the 3'READS⁺ data described above. As shown in Figure 3A (upper panels), APA events are rare in pA⁺ eRNAs (less than 70 APA events detected out of 3,919 pA⁺ eRNAs). Additionally, APA of these eRNAs was not affected by ZFC3H1, INTS4, or ZCCHC8 KD. For comparison, we also examined the APA events in mRNAs, and found that ZFC3H1 and INTS4 depletion led to a global shift of poly(A) site usage, from distal

to proximal, whereas depletion of ZCCHC8 had only minimal effects (although a few affected genes also showed RNA length shortening) (Figure 3A, lower panels). Furthermore, we found that ZFC3H1 KD resulted in preferential accumulation of two types of transcripts: ptRNAs and uaRNAs, while INTS4 KD only caused ptRNA upregulation (Figure 3B).

Poly(A) signals are highly enriched in enhancer regions and actively utilized in pA+ eRNA production

Our data indicate that a large fraction of eRNAs are polyadenylated at their 3' ends. Especially in light of previous studies suggesting that most 3' ends were nonpolyadenylated and formed by the Integrator (Andersson et al., 2014; Kim et al., 2010; Lai et al., 2015), it seemed possible that the mechanism might be more analogous to that employed to form certain U snRNA 3' ends (Baillat et al., 2005). While previous work showed that poly(A) signal sequences are enriched in enhancer regions (Wang et al., 2018), we wished to extend these results and determine if such signals were located in the proximity of the 3' ends we detected in our analysis. To this end, we first searched for the most abundantly used AAUAAA canonical PAS motif and/or the AUUAAA motif, in the pA+ eRNAs detected by 3'READS+, and found a ~70% enrichment of AAUAAA and a higher enrichment (76.2%-88.6%) of both PAS motifs (Figure 4A, B). Known PAS motifs including AAUAAA and its variants were found in 79% of human mRNA (Beaudoing et al., 2000). Therefore, the abundance of PAS motifs in pA+ eRNAs is at least comparable with mRNA. Most functional PAS motifs are known to be located approximately -20 nt upstream of the cleavage and poly(A) sites (Tian et al., 2005). Importantly, we found that pA+ eRNAs displayed a

significant enrichment of PAS motifs within 20-40 nt upstream of the 3' cleavage site, regardless of their direction (+ strand vs. - strand) and genomic locations (Figure 4C). Localization of the AUUAAA motif in pA⁺ eRNAs also showed similar patterns (Figure S5). These data strongly suggest that AAUAAA motifs are utilized in generation of pA⁺ eRNA 3' ends.

Another sequence element commonly found in active polyadenylation sites is the downstream GU-rich sequences. GU-rich sequences are usually located 10-40 nt downstream of the cleavage site, and lack a clear consensus sequence (Colgan and Manley, 1997; Tian and Manley, 2017). To determine whether pA⁺ eRNAs utilize GU-rich elements similar to mRNAs, we conducted a motif search of the pA⁺ eRNAs detected in our 3'READS analysis. To simplify the algorithm, we searched for eight GU-rich elements that have been previously characterized (Perez Canadillas and Varani, 2003). As shown in Figure S6, both pA⁺ eRNAs and mRNAs showed an enrichment of these GU-rich sequences downstream of the cleavage sites. However, a GU-rich peak was also detected in the region upstream of the cleavage sites in mRNAs, while such a peak was absent in pA⁺ eRNAs. This likely corresponds to U-rich sequences in mRNAs that have been described previously (Neve et al., 2017; Tian and Manley, 2013, 2017). These U-rich elements often correspond to UGUA motifs that bind the cleavage/polyadenylation factor Cleavage Factor I (Brown and Gilmartin, 2003; Yang et al., 2011), and its absence suggests that this factor may not function in pA⁺ eRNA 3' end formation.

The absence of an upstream GU-rich peak raises the possibility that there may be differences between the polyadenylation machinery that acts to generate the 3' ends of pA⁺ eRNAs and mRNAs. However, the enrichment of upstream PAS and downstream GU-rich elements in

pA⁺ eRNAs suggests that pA⁺ eRNAs might employ a polyadenylation machinery similar to the one that functions on mRNAs. To test this hypothesis, we depleted with siRNAs two Cleavage/Polyadenylation Specificity Factor (CPSF) subunits: CPSF30 (also known as CPSF4) and CPSF73 (also known as CPSF3), together with Mtr4 (to achieve higher accumulation of eRNAs), and examined whether the cleavage efficiency of pA⁺ eRNAs was affected (Figure S7A). For this, we used RT-qPCR with primers either spanning or downstream the poly(A) site (detecting uncleaved product of pA⁺ eRNAs specifically), normalized by primers upstream of the poly(A) site, to detect possible 3' cleavage defects in both pA⁺ eRNAs after CPSF-30/-73 KD. We found that 3' cleavage of all the mRNAs tested (*gapdh*, β -actin, *fos*, *fn1*) was as expected impaired after the depletion of CPSF-30/-73, corresponding with an accumulation of uncleaved products (Figure S7B). However, 3' cleavage of pA⁺ eRNAs (EN11+/-, EN14+/-, EN17+/-) was not affected by CPSF-30/-73 KD (Figure S7C). Given the presence of canonical PASs in these pA⁺ eRNAs, these results were unexpected and suggest that pA⁺ eRNA 3' end formation might employ a distinct mechanism from that used by mRNAs.

Canonical poly(A) polymerases PAP- α / γ function in eRNA polyadenylation

The above data indicate that a large fraction of eRNAs are polyadenylated, and in a PAS-dependent manner. This suggested the involvement of the canonical 3' processing machinery, but the above data question this idea. In light of this, an intriguing question is the identity of the poly(A) polymerase(s) (PAPs) responsible for poly(A) tail synthesis. Two types of PAPs have been described, dubbed canonical and noncanonical. The canonical PAPs, PAP- α and PAP- γ , are known

to polyadenylate most mRNAs (Laishram, 2014), and polyadenylation is crucial for mRNA stabilization, export and efficient translation (Colgan and Manley, 1997). Also, more recent studies have shown that addition of a long poly(A) tail (100-500 A residues) by the canonical PAPs promotes degradation of intron-less nuclear RNAs by the nuclear exosome (Beaulieu et al., 2012; Bresson and Conrad, 2013; Bresson et al., 2015).

We therefore wanted to identify whether the canonical PAP(s) are responsible for eRNA polyadenylation. To this end, we KDed PAP- α/γ under Mtr4-depleted conditions in HeLa cells (Figure 5A). We then performed RL-PAT assays of specific eRNAs (EN11 is shown in Figure 5B, EN14 and EN17 in Figure S8) and sequenced the gel purified PCR products to identify the exact position of the RNA 3' end. Interestingly, we found that depletion of PAP- α/γ led to defective eRNA 3'-end cleavage, and consequently transcript readthrough and polyadenylation at downstream sites. Occasionally, we failed to identify the downstream cleavage sites due to extremely long readthrough (new 3'-end sites are denoted by green triangles on the genomic sequences). We also note that there was some apparent variability in precise cleavage sites and poly(A) status and tail length of individual eRNAs from experiment to experiment (e.g., EN11; Figure 5C), perhaps reflecting the unusual features of eRNA 3' processing. Nonetheless, our data indicate that the canonical poly(A) polymerases PAP- α/γ both are crucial for the proper 3' processing of pA+ eRNA.

pA+ eRNAs are transported to the cytoplasm and associate with active polysomes following Mtr4 KD

Several studies have found that certain unstable nuclear polyadenylated RNAs are

transported to the cytoplasm when Mtr4 or ZFC3H1 are depleted (Fan et al., 2017; Ogami et al., 2017; Silla et al., 2018). To determine whether pA+ eRNAs behave similarly, we analyzed their localization by subcellular fractionation followed by RT-PCR under control and Mtr4 KD conditions. Analyzing the same fractions we used in our previous study (Ogami et al., 2017), we indeed detected substantial accumulation of pA+ eRNAs (EN11/14/17 bidirectional transcripts) in the cytoplasmic fraction after Mtr4 depletion (Figure 6A, cf. lane 2) compared to control cells, which show a higher pA+ eRNA abundance in the chromatin fraction (Figure 6A, cf. lanes 1, 3, and 5). This phenomenon is consistent with our previous finding of cytoplasmic accumulation of pt/ua-RNAs following Mtr4 KD (Ogami et al., 2017).

We next asked if some of the cytoplasmic pA+ eRNAs might associate with ribosomes and be translated. Indeed, we previously observed that pt- and ua-RNAs that escaped from the nucleus following Mtr4 KD became polysome associated (Ogami et al., 2017). Interestingly, open reading frames (ORFs) of various lengths were identified in several of the eRNAs analyzed above (for example, the longest ORF for EN11+ is 465 nt, EN11- 291nt, EN14+ 204 nt, EN14- 222 nt; see Figure S9A). To examine whether these eRNAs associated with active ribosomes, we conducted polysomes fractionation to evaluate the distribution of the EN11 and EN14 bidirectional transcripts, as well as GAPDH mRNA for comparison, in both siCtrl- and siMtr4- treated cells. This analysis revealed an increased association of both EN11 and EN14 transcripts with polysomes after Mtr4 KD. Consistent with our previous analysis (Ogami et al., 2017), GAPDH mRNA polysome association was diminished (Figure 6B; quantitation in Figure S9B). We showed previously that Mtr4 KD led to a global reduction of mRNA translation, and proposed that this

reflects competition for the translation machinery by cytoplasmic accumulation of normally unstable nuclear pA⁺ RNAs (Ogami and Manley, 2017; Ogami et al., 2017). It is very likely that the stabilized pA⁺ eRNAs participate in the same process.

Discussion

In this study, we have explored degradation and biogenesis mechanisms of eRNAs. Our results revealed that a large fraction of eRNAs that accumulate under nuclear surveillance deficient conditions are polyadenylated. This contrasts with previous studies suggesting a lack of polyadenylation in eRNAs. We further identified and examined the roles of three important complexes in pA⁺ eRNA accumulation, the PAXT/PPC, NEXT Integrator complexes. We found that both the PAXT/PPC and NEXT are required for pA⁺ eRNA degradation. The Integrator, whose role has been highlighted in mRNA 3'-end cleavage and eRNA biogenesis, functions in the biosynthesis of pA⁺ eRNAs, as well as ptRNAs. Additionally, we found that the detected pA⁺ eRNAs typically contain AAUAAA-like and GU-rich signal sequences near their 3' cleavage sites. We also provide evidence that the canonical poly(A) polymerases, PAP- α/γ , function in pA⁺ eRNA cleavage and polyadenylation, and can lead to production of exceptionally long poly(A) tails. Importantly, stabilized pA⁺ eRNAs accumulate in the cytoplasm and appear to be translated. Our findings have provided new insights into eRNA metabolism, and also suggest a novel model connecting the roles of the Integrator in mRNA and eRNA synthesis.

Previous studies have identified a large pool of putative enhancers and defined bidirectional transcription from the enhancer region as a feature of enhancer activation. Although transcribed by RNAP II, eRNAs have been thought to be largely nonpolyadenylated. One study

was unable to detect any eRNAs in poly(A)- RNA fractions (Kim et al., 2010), whereas another estimated that approximately 90% of eRNAs were non-polyadenylated (Andersson et al., 2014). In sum, all studies reached the consensus that most eRNAs lack poly(A) tails (reviewed by Lam et al., 2014; Li et al., 2016). However, those studies were conducted under intact nuclear surveillance conditions, and thus may have failed to capture fast-turnover nuclear RNA species. By blocking the nuclear surveillance pathway, we found that a large number of eRNAs (~6500) are in fact polyadenylated. Notably, pan-cancer analyses using the TCGA database found that the maximum number of eRNAs detected in any given cancer type is less than 10,000, while the average number is around 4,500 (Chen et al., 2018). By comparing these numbers, we conclude that eRNA polyadenylation is a far more prevalent mechanism than previously thought. Given that highly similar molecular features and shared transcriptional machinery of enhancers and promoters have been revealed in many studies (Core et al., 2014; Furlong and Levine, 2018; Kim et al., 2010; Kim and Shiekhattar, 2015; Whalen et al., 2016), it was in fact somewhat surprising that enhancer-derived RNAs appeared to lack poly(A) tails, especially since promoter-derived non-mRNA transcripts (e.g. ptRNAs and uaRNAs) are all polyadenylated (Kaida et al., 2010; Preker et al., 2008). Our characterization of pA⁺ eRNAs provides new examples of the similarities between enhancers and promoters.

Over the past decades, increasing evidence has supported the enhancer-promoter interaction model, where the exchange of transcriptional machinery between enhancer and promoter may occur from both directions (Hatzis and Talianidis, 2002; Wang et al., 2005). Moreover, similar features between enhancers and promoters have been revealed in recent years,

such as their sequence motifs, chromatin architectures, transcription machineries, as well as regulation mechanism (Core et al., 2014; Scruggs et al., 2015; Vihervaara et al., 2017). Further extending these similarities, our results showed that both enhancer- and coding gene- derived transcripts are controlled by ZFC3H1 (a subunit of the PAXT complex) and INTS4 (a subunit of the Integrator complex).

We first consider the role of INTS4. On the one hand, depletion of INTS4 led to accumulation of ptRNAs but not uaRNAs. This can be explained by the previous finding that the Integrator complex is important for productive transcription elongation, functioning by recruiting the super elongation complex (Gardini et al., 2014). Consequently, defective transcription elongation in INTS4-depleted cells led to a decrease in full-length mRNA and an increase in ptRNAs or shorter coding transcripts. Expression of transcription elongation-independent uaRNAs was not affected by Integrator subunit depletion (Preker et al., 2011). On the other hand, we also found that Ints4 depletion caused a significant increase in pA⁺ eRNA accumulation. Based on these findings, we propose a model for the role of Integrator in mRNA/eRNA synthesis. As indicated in Figure 7A, active enhancer and promoter regions are in direct contact with each other to form a chromatin loop, with transcription machinery at both regions. When a transcription complex is in elongation status on the coding gene, depletion of the Integrator complex leads to RNAP II stalling (Gardini et al., 2014; Skaar et al., 2015; Stadelmayer et al., 2014), and consequently early termination and generation of ptRNAs. Alternatively, when RNAP II is at the promoter, we propose that depletion of Integrator causes transcription machinery switch to enhancer regions, leading to increased level of eRNA synthesis.

We note that another study found that Integrator depletion led to accumulation of extended eRNAs following EGF induction (Lai et al., 2015). However, we did not observe any significant change of pA+ eRNA length after INTS4 KD. A possible explanation is that the eRNAs in our study were all stimulus-independent, thus representing a different group of eRNAs compared with EGF-induced eRNAs. This study also observed an increased level of polyadenylated eRNAs on EGF-responsive enhancer regions under the depletion of Integrator (Lai et al., 2015), which is consistent with and expanded by our finding: the accumulation of polyadenylated eRNAs after depleting Integrator subunits, doesn't limit to EGF-responsive enhancers, but occurs without EGF stimulus.

How might ZFC3H1 function in production of mRNA? Previous studies in our lab showed that depletion of ZFC3H1 (and Mtr4) caused increased accumulation of both ptRNA and uaRNAs and transport to the cytoplasm, indicating the involvement of ZFC3H1/Mtr4 in turnover and nuclear retention of these RNAs. Later, another study confirmed the central role of ZFC3H1 as a nuclear pA+ RNA retention factor (Silla et al., 2018). In the current study, we not only detected significant accumulation of pA+ eRNAs, but also observed a significant increase of pt/ua-RNAs. Importantly, this was accompanied by a proportional decrease of full-length mRNA transcripts after ZFC3H1 depletion. In contrast, Mtr4 depletion only led to accumulation of pt/ua-RNAs, but did not affect abundance of full-length transcripts, indicating an Mtr4-independent function for ZFC3H1. Given the finding that the upregulation of ptRNAs and uaRNAs in both ZFC3H1 and INTS4 depleted cells is accompanied by a proportional downregulation of the corresponding full-length RNAs, we believe that ZFC3H1 also plays a role in RNA biogenesis, besides its known

function in RNA degradation and nuclear retention (Meola et al., 2016; Ogami et al., 2017; Silla et al., 2018). Future studies will be required to elucidate its precise role.

Our analyses of the poly(A) signals and G/U-rich regions of pA⁺ eRNAs indicate that pA⁺ eRNAs utilize AAUAAA-like and G/U-rich signal sequences to direct their polyadenylation and cleavage, just like mRNAs. It is well known that such elements are critical for 3'-end formation of most mRNAs (Gil and Proudfoot, 1987; McLauchlan et al., 1985). Therefore, we suspected that those pA⁺ eRNA might share a similar 3'-end processing mechanism with mRNAs. However, when we depleted two key cleavage and polyadenylation specificity factors, CPSF-30 and -73, which recognize PAS signals and promote mRNA cleavage (Colgan and Manley, 1997; Mandel et al., 2008; Millevoi and Vagner, 2010), unlike mRNAs, pA⁺ eRNA cleavage was not affected, indicating that a distinct mechanism of 3' end cleavage might be employed in pA⁺ eRNA. Consistent with this, the coexistence of both pA⁺ and pA⁻ eRNAs, which is not observed with mRNAs, suggests differences in mechanism. The detailed mechanism of pA⁺ eRNA 3' end cleavage remains to be elucidated.

Our analyses of individual eRNAs have discovered that the canonical poly(A) polymerases, PAP- α/γ , are crucial for the proper 3' end processing of pA⁺ eRNA. Depletion of PAP- α/γ led to read-through of normal cleavage sites. Nevertheless, those eRNAs were still polyadenylated, since they can be observed in RL-PAT assays, which only detect polyadenylated RNA species. It is possible that these eRNAs were polyadenylated by other noncanonical poly(A) polymerases under the depletion of PAP- α/γ . Consistent with this, early in vitro studies from our lab provided evidence that PAP is frequently required for cleavage as well as poly(A) tail synthesis (Ryner et al., 1989;

Takagaki et al., 1988). Confirming whether the PAP- α/γ involved polyadenylation mechanism applies globally to eRNAs will require a genome-wide study.

Addition of a poly(A) tail can be sufficient for mRNA nuclear export (Brodsky and Silver, 2000; Fuke and Ohno, 2008). Indeed, our subcellular fractionation experiments showed that pA+ eRNAs were exported into the cytoplasm when the nuclear surveillance system has been compromised. Additionally, our RL-PAT assays showed that pA+ eRNAs had poly(A) tails of various lengths, from ~100 nts to, remarkably, ~1000 nts. In contrast, mammalian mRNAs typically have poly(A) tails of ~250 nucleotides (reviewed by Weill et al., 2012). We are particularly interested in the exceptionally long poly(A) tails. Previous studies have found that PAP by itself is unable to generate poly(A) tails with defined lengths, instead, PABPN1 measures the length of poly(A) tails and guarantees the poly(A) extension no longer than ~250 nt (Bienroth et al., 1993; Wahle, 1995). It is possible that those pA+ eRNAs with exceptionally long poly(A) tails are not regulated by PABPN1. Additionally, it has been found that the longer the poly(A) tail is, the more stable the mRNA is: on the one hand, poly(A) tails of mRNAs are gradually shortened in the cytoplasm; on the other hand, mRNAs will only be degraded after removing the long poly(A) tails (reviewed by Eckmann et al., 2011; Weill et al., 2012). Therefore, the exceptionally long poly(A) tails might indicate longer half-life of pA+ eRNAs in the cytoplasm.

A previous study in our lab found that excessive polyadenylated RNA species in the nucleus “leak” into the cytoplasm under nuclear surveillance deficient conditions, thus disrupting the translational system by competing with mRNAs to bind ribosomes (Ogami et al., 2017). Consistent with this, our polysome fractionation experiments showed that nuclear pA+ eRNAs can

also accumulate in the cytoplasm following Mtr4 KD, associate with monosomes/polysomes, and we suggest facilitate the suppression of coding-gene translation. To extend this finding, we investigated whether those eRNAs, containing short open reading frames (sORFs), can be translated. While ribosome occupancy doesn't necessarily guarantee translation (Guttman et al., 2013), lncRNAs, defined as noncoding RNAs longer than 200 nucleotides, represent a source of short open reading frames (sORFs), which have the coding potential for micropeptides (Cohen, 2014; Frith et al., 2006; Pauli et al., 2015). Moreover, another study showed that a large amount of monosomes are translating sORFs actively (Heyer and Moore, 2016). It is reasonable to suspect that some of pA⁺ eRNAs, which can be transported to the cytoplasm and bind ribosomes, might also be a source to generate micropeptides, or even longer ones. In addition, enhancers lack sequence conservation, and undergo rapid evolution across mammalian species (Villar et al., 2015), thus are tend to contain de novo sORFs and produce de novo micropeptides, if translated. The generation of de novo micropeptides have been suggested to provide advantages to species during evolution (Ruiz-Orera and Alba, 2019). The concept of enhancer-derived translation merits further study.

Our findings are summarized and illustrated in Figure 7B. Both pA⁻ and pA⁺ eRNAs can be generated from the enhancer-promoter loops. All of these naturally undergo exosome mediated degradation through the NEXT and PAXT pathways. Consistent with this, a recent study found that the RNA helicase activity of Mtr4, a subunit of both PAXT/PPC and NEXT complexes, is enhanced when incorporated into the NEXT complex, whose optimal substrates actually require a polyadenylated 3' end (Puno and Lima, 2018). We have also highlighted a

critical and novel mechanism of eRNA polyadenylation, which indicates a substantially greater pool of naturally unstable polyadenylated eRNAs than previously thought. This further validates the enhancer-promoter interaction model by characterizing their shared regulators as well as shared polyadenylation mechanism. Lastly, we have provided evidence that under certain conditions eRNAs can be stabilized, transported to the cytoplasm and even translated. The significance of these unexpected findings will be the topic of future studies.

Materials & Methods

Primers and siRNAs

All siRNAs, primers and enhancer genome locations in this study are listed below.

Primers for RT-qPCR	Sense (5'-3')	Antisense (3'-5')
GAPDH	ACGACCACTTTGTCAAGCTC	TTCCTCTGTGCTCTTGCTG
NEAT1	TTGGTTCTGAGCTGCGTCTA	GTGCTGTAAAGGGGAAGAAA
EN11+	CTTACGAACTGGGGAGTCCA	TTAATAGAGACCGGGTCTGGC
EN11-	CCTACATTGTCTGCATTCAGGA	AACAGAGCCAGACCCTGTGT
EN14+	GGGCATCATTATGGATTGG	GGCTCCTTCCAGTACTAAAAAGC
EN14-	CGCTTGAACGAGTGACTGAA	GGGCATCATTATGGATTGG
EN17+	ACACAATGTCCTGTGGGTGA	GAAGAAAGCCCTTCTGCTCA
EN17-	GAGCCATGGATGGGTGATAA	GGGTCTTCCAAAGTGCAAAG
EN323+	AACCGAGATTGCTGTCCCAG	AACGGCTACTCCTAAACGGG
EN323-	TCGCTTGTGAGTCGGTAACC	TGGGGTGGCTTACTGCTTTT
EN20600+	AGCACTTTGAGCGGATCACA	CCTCCCGGGTTAAAGCGATT
EN20600-	CCGCACCGTTGACCTAGAAT	AGAAGACCAAGCAGGCAGAC
EN23771+	TGCCATACTACACCGTGCTG	TGCCCCCTTGCTGTTTTAGT
EN23771-	TGTACACCTGGCTCCTCCTT	TTCCCTTTGTGGTAGCCGAG
EN50286+	TCACTGCAGTATTGGGAGCC	CAGGCCGTCCCTAACAAATGT
EN50286-	CAGGCCGTCCCTAACAAATGT	TCACTGCAGTATTGGGAGCC

Primers for 3' cleavage RT-qPCR	Sense (5'-3')	Antisense (3'-5')
GAPDH-internal probe	ACCTGACCTGCCGTCTAGAA	ACCTGACCTGCCGTCTAGAA
GAPDH-uncleaved	CCGCACCTTGTCATGTACCA	TCTCACCTTGACACAAGCCC
b-actin-internal probe	ACTTCGAGCAAGAGATGGCC	AAGGAAGGCTGGAAGAGTGC
b-actin-uncleaved	AAGGCTTTTGGTCTCCCTGG	AAGGCTTTTGGTCTCCCTGG
fos-internal probe	GTGCCAACTTCATCCCACG	TCATGGTCTTACAACGCCA
fos-uncleaved	CTTGAGGTCTTTGACATGTGG	AAGGTTGGTCGCATTCAACT
fn1-internal probe	AAGGTTGGTCGCATTCAACT	AGGGACTTTCCTCTCTGCCA

fn1-uncleaved	GATTGCCTGCAAGGGAAATA	GCCATGCCAAAAGTGTGTCAG
EN11+ upstream of cleavage site	CTTACGAACTGGGGAGTCCA	TTAATAGAGACCGGGTCTGGC
EN11+ downstream of cleavage site	CTTACGAACTGGGGAGTCCA	AAAGGGAGGAAAAGGGGACG
EN11- upstream of cleavage site	CCTACATTGTCTGCATTCAGGA	AACAGAGCCAGACCCTGTGT
EN11- downstream of cleavage site	CCTACATTGTCTGCATTCAGGA	TCCTGAAGGGCCTGCTGTAA
EN14+ upstream of cleavage site	GGGCATCATTATGGATTTGG	GGCTCCTTCCAGTACTAAAAAGC
EN14+ downstream of cleavage site	AGGCTCAGCTGGAAAATGGA	AGGCTCAGCTGGAAAATGGA
EN14- upstream of cleavage site	CGCTTGAACGAGTACTGAA	GGGCATCATTATGGATTTGG
EN14- downstream of cleavage site	GGCCAAATCCATAATGATGCCC	GCCACAGAGTGAGAGCTTGT
EN17+ upstream of cleavage site	ACACAATGTCCTGTGGGTGA	GAAGAAAGCCCTTCTGCTCA
EN17+ downstream of cleavage site	ACACAATGTCCTGTGGGTGA	CTCTGTTTTCTGTTTCGTCAGC
EN17- upstream of cleavage site	GAGCCATGGATGGGTGATAA	AACCCATCTTGTGAGGCAGA
EN17- downstream of cleavage site	ACTCATGAGCTGCCAATCCT	CTTGCATGATGCCTGGAAGC

Primers for RL-PAT	1st Round PCR	2nd Round PCR
EN11+	ATCTCTGCACTGGCCTTACGAA	AGTCCAAGGGGTTCCAGGA
EN11-	CCTACATTGTCTGCATTCAGGA	ACACAGGGTCTGGCTCTGTT
EN14+	TGTCAGATCAGCACAAAGCA	GAGCTGGGATTGAAAAGACC
EN14-	GGATTTCTCAGGCTCCTTCC	CCAAATCCATAATGATGCCC
EN17+	ACACAATGTCCTGTGGGTGA	TGAGCAGAAGGGCTTCTTTC
EN17-	GAGCCATGGATGGGTGATAA	TCTGCCTGACAAGATGGGTT
EN323+	ACTGGAGCGTGATTGGTGAG	AACCGAGATTGCTGTCCAG
EN323-	TCGCTTGTGAGTCGGTAACC	AACGGCTACTCCTAAACGGG
EN20600+	AGCACTTTGAGCGGATCACA	CGAGACCAACCTGACCAACA
EN20600-	CCGCACCGTTGACCTAGAAT	ACTCCGTGGCTACATTGTGCG
EN23771+	TGACCACATTTCCAGGGCTC	TGCCATACTACACCGTGCTG
EN23771-	AGGTCTGACATTAGCCAGGC	TGTACACCTGGCTCCTCCTT
EN50286+	TTTGCCTTGGCCAGATAGGA	TCACTGCAGTATTGGGAGCC
EN50286-	AGCCGAATGATACAGCCCTT	CAGGCCGTCCCTAACAAATGT
5dT-anchor	GCGAGCTCCGCGGCCGCTTTTT	

siRNA	Sense (5'-3')	Antisense (3'-5')
siMtr4 #1	CAAUUAAGGCUCUGAGUAATT	UUACUCAGAGCCUAAUUGTT
siMtr4 #2	siRNA ID: s223606 (ThermoFisher)	Catalog # 4392420
siPABPN1 #1	AGUCAACCGUGUUACCAUATT	UAUGGUAACACGGUUGACUTT
siPABPN1 #2	GGCCUUAGAUGAGUCCCUATT	UAGGGACUCAUCUAAGGCCTT
siZCCHC8 #1	GAAAUACAACAGAAUAAAATT	UUUUAAUUCUGUUGUAUUUUCTT
siZCCHC8 #2	GGAAUGUACCUCAGGAUAATT	UUAUCCUGAGGUACAUCCTT
siINTS4 #1	UUGCAGUUUUGGUACUUAUTT	AUAAGUACCAAACUGCAATT
siINTS4 #2	CCUGUAAAAGUUUAUAUAATT	UUAUAUAAAACUUUUACAGGTT
siZFC3H1	GAAACAAGCUGAAGAAGAATT	UUCUUCUUCAGCUUGUUUUCTT
siPAP- α	CAAUCCAGUGCUAUUGAATT	UUCAAUAGCACUGGAUUUGTT
siPAP- γ	CAACAGAAUUCUACGUUAUATT	UAUACGUAGAAUUCUGUUGTT

eRNA list	Genomic location (UCSC: hg19)
EN11	chr11:65243603-65246000
EN14	chr14:52546966-52549000
EN17	chr17:80254353-80255909
EN323	chr1:11967731-11968976
EN20600	chr16:30644889-30646682
EN23771	chr17:41399497-41401766
EN50286	chr9:35789694-35791289

Cell culture and siRNA transfections

HeLa cells were cultured in DMEM supplemented with 10% FBS. The siRNAs were transfected using DharmaFECT 1 (GE Dharmacon) at 20nM and maintained for 48 or 72 hrs. To obtain high knockdown efficiency after 72hr treatment, the siRNA transfection was repeated 48hrs after the first transfection with half the amount of siRNA and cells were harvested 24hrs after the second transfection.

Antibodies

Mtr4 (NB100-1574), ZCCHC8 (NB100-94995), ZCCHC7 (NBP1-89175) INTS4 (NB100-60660) antibodies were from Novus Biologicals. ZFC3H1(A301-456A), PABPN1 (A303-523A), PAP- α (A301-010A), PAP/ γ (A302-427A) antibodies were from Bethyl Laboratories. CPSF-30 (sc-393316) and CPSF-73 (sc-393001) antibodies were from Santa Cruz Biotechnology. GAPDH (G8795; Sigma), U2AF65 (U4758; Sigma), histone H3 (ab1791; abcam) were also used in this study.

3'READS+

Total RNA was purified from control and siRNA-treated cells. RNA integrity was analyzed by using Agilent Bioanalyzer. Samples with an RNA integrity number (RIN) >9.0 were subjected to 3'READS+ analysis following the protocol described in Zheng et al., (2016). Briefly, after RNA fragmentation, poly(A)+ RNA fragments were captured on magnetic beads coated with a chimeric oligonucleotide (oligo CU5T45), which contained 45 thymidines at the 5' portion and five uridines at the 3' portion, and subjected to RNase H digestion, which removed the bulk of the poly(A) tail and eluted RNA from beads. Eluted RNA was ligated to 5' and 3' adapters followed by reverse transcription, PCR amplification, and deep sequencing on an Illumina platform. 3'READS+ data were analyzed as described (Zheng et al., 2016).

Poly (A) + RNA motifs analysis

Number of eRNA polyadenylation motifs were counted in the corresponding loci of the GRCh37

hg19 reference genome. Distances between pA motifs and cleavage sites were obtained by subtracting their genomic positions. Only the motifs upstream of the cleavage sites were considered as effective pA motifs.

RT-qPCR

Total RNA was isolated using TRIzol (Thermo Fischer Scientific) according to the manufacturer's instructions, followed by DNase I digestion at 37°C for 30min (Thermo Fischer Scientific). 2µg of DNaseI-treated RNA was reverse-transcribed with oligo(dT) primer using Maxima Reverse Transcriptase (Thermo Fischer Scientific). Reactions were diluted 15 times in water, and qPCR was performed with primers listed in Table. S1 using Power SYBR (Thermo Fischer Scientific). All data were normalized to GAPDH.

RNA-ligation coupled PAT assay (RL-PAT)

DNase I-treated total RNAs were ligated with KO109 P-anchor-NH₂ oligonucleotide using T4 RNA ligase 1 (New England Biolabs) in the presence of 15% PEG8000 at 14°C for 30min, and reverse-transcribed with KO108 anchor AS primer using Maxima Reverse Transcriptase (Thermo Fischer Scientific). Semi-nested PCR was performed using gene-specific forward primers and KO105 reverse primer. PCR products were resolved in 1.5-2.0% agarose gel prestained with EtBr. To remove poly(A) tails, 10µg of DNase I-treated RNAs were mixed with 50 pmol of oligo(dT)₁₈ in RNase H buffer (50mM Tris-HCl (pH 7.4), 100mM NaCl, 10mM MgCl₂) and incubated at 85°C for 5min, 42 °C for 10min, and then incubated in the presence of 2U Hybridase-Thermostable

RNase H (Epicentre) at 42°C for 30min. RNase H-treated RNAs were isolated using TRIzol.

Subcellular fractionation

Seventy-two hours after siRNA transfection, HeLa cells grown in a 10-cm dish were washed twice with PBS and collected by scraping and centrifugation. Cell pellets were resuspended in 400 μ L of swelling buffer (10 mM Tris-HCl at pH 8.0, 1.5 mM MgCl₂, 10 mM KCl, 5 U of RNasin, 1 \times protease inhibitor cocktail) and incubated for 15 min on ice. Cells were homogenized by passing a 26-gauge needle attached to a 1-mL syringe until >90% of cells were disrupted (typically 10~20 strokes). Half of the lysate was kept in a new tube and used as whole-cell lysate. The rest of the 200 μ L of lysate was mixed with 2 μ L of 10% NP-40, gently tapped, and immediately centrifuged in an Eppendorf centrifuge 5424 at 6,000 rpm for 5 min. The supernatant was kept in a new tube and used as cytoplasmic fraction. The pellet was washed once with swelling buffer, resuspended in 100 μ L of glycerol buffer (20 mM Tris-HCl at pH 8.0, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT, 5 U of RNasin, 1 \times protease inhibitor cocktail) by pipetting, and then mixed with 100 μ L of nucleus lysis buffer (20 mM HEPES-NaOH at pH 7.6, 7.5 mM MgCl₂, 0.2 mM EDTA, 300 mM NaCl, 1 M urea, 1% NP-40, 1 mM DTT, 5 U of RNasin, 1 \times protease inhibitor cocktail). The mixture was pulse-vortexed three times, incubated for 1 min on ice, and then centrifuged in an Eppendorf centrifuge 5424 at 14,000 rpm for 2 min. The supernatant was used as nuclear-soluble fraction. The pellet was washed once with a 1:1 mixture of glycerol/nucleus lysis buffer and then resuspended in 200 μ L of water. RNA was extracted using TRIzol, DNase I-treated, and then reverse-transcribed with oligo(dT) or random primer using Maxima RT.

Polysome fractionation

HeLa cells were transfected with either siCtrl or siMtr4 and maintained in the same medium for the indicated times in five 10-cm dishes. On the day of harvest (~80% confluency), cells were treated with either 50 μ M BTdCPU or 100 μ g/mL CHX for 3 h and 5 min, respectively, at 37°C. Cells were washed once with ice-cold PBS containing either 50 μ M BTdCPU (Millipore) or 100 μ g/mL CHX and then resuspended in 400 μ L of polysome lysis buffer (20 mM HEPES-NaOH at pH 7.6, 150 mM NaCl, 15 mM MgCl₂, 0.5% NP-40, 80 U RNasin, 1 \times protease inhibitor cocktail) containing either 50 μ M BTdCPU or 100 μ g/mL CHX. After 10 min of incubation on ice, the lysates were centrifuged at 3000 rpm for 5 min, and the supernatant was centrifuged at 15,000 rpm for 5 min in a new tube. The supernatant was loaded onto 15%–45% sucrose gradients (20 mM HEPES-NaOH at pH 7.6, 150 mM NaCl, 5 mM MgCl₂, 100 μ g/mL CHX) followed by centrifugation at 39,000 rpm for 90 min using an SW41Ti rotor. Fractions (200 μ L each) were manually collected, and A₂₅₄ was determined using a Nanodrop spectrophotometer. RNA was extracted by mixing with 1 vol of TRIzol, isopropanol-precipitated in the presence of GeneElute-LPA (Sigma), and then reverse-transcribed with oligo(dT) primer using Maxima RT. To avoid efficiency differences in RT reactions, RNA amounts were equalized by adding purified yeast RNA (Thermo Fisher Scientific). No PCR products were generated when yeast RNA RT products alone were amplified with the primers used in this study.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We are grateful to Rachel Giesler (Novus Biologicals) for the Mtr4 (NB100-1574), ZCCHC8 (NB100-94995), ZCCHC7 (NBP1-89175) and INTS4 (NB100-60660) antibodies. We also thank Dr. Patricia Richard (Columbia University) for comments and editorial assistance.

Supplementary information is available at

References

Albrecht, T.R., Shevtsov, S.P., Wu, Y., Mascibroda, L.G., Peart, N.J., Huang, K.L., Sawyer, I.A., Tong, L., Dunder, M., and Wagner, E.J. (2018). Integrator subunit 4 is a 'Symplekin-like' scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic Acids Res* 46, 4241-4255.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., *et al.* (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455-461.

Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 123, 265-276.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299-308.

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Res* 10, 1001-1010.

Beaulieu, Y.B., Kleinman, C.L., Landry-Voyer, A.M., Majewski, J., and Bachand, F. (2012). Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS Genet* 8, e1003078.

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

Bresson, S.M., and Conrad, N.K. (2013). The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. *PLoS Genet* 9, e1003893.

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

- Brodsky, A.S., and Silver, P.A. (2000). Pre-mRNA processing factors are required for nuclear export. *RNA* *6*, 1737-1749.
- Brown, K.M., and Gilmartin, G.M. (2003). A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell* *12*, 1467-1476.
- Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell* *144*, 327-339.
- Chen, H., Li, C., Peng, X., Zhou, Z., Weinstein, J.N., Cancer Genome Atlas Research, N., and Liang, H. (2018). A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell* *173*, 386-399 e312.
- Cohen, S.M. (2014). Everything old is new again: (linc)RNAs make proteins! *EMBO J* *33*, 937-938.
- Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev* *11*, 2755-2766.
- ENCODE Project Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57-74.
- Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics* *46*, 1311-1320.
- De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* *8*, e1000384.
- Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., *et al.* (2012). Landscape of transcription in human cells. *Nature* *489*, 101-108.
- Eckmann, C.R., Rammelt, C., and Wahle, E. (2011). Control of poly(A) tail length. Wiley

Interdiscip Rev RNA 2, 348-361.

Fan, J., Kuai, B., Wu, G., Wu, X., Chi, B., Wang, L., Wang, K., Shi, Z., Zhang, H., Chen, S., *et al.* (2017). Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. *EMBO J* 36, 2870-2886.

Frith, M.C., Forrest, A.R., Nourbakhsh, E., Pang, K.C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Bailey, T.L., and Grimmond, S.M. (2006). The abundance of short proteins in the mammalian proteome. *PLoS Genet* 2, e52.

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

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. *Science* 361, 1341-1345.

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. *Mol Cell* 56, 128-139.

Gil, A., and Proudfoot, N.J. (1987). Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. *Cell* 49, 399-406.

Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240-251.

Hatzis, P., and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol Cell* 10, 1467-1477.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* 16, 144-154.

Heyer, E.E., and Moore, M.J. (2016). Redefining the Translational Status of 80S Monosomes. *Cell* 164, 757-769.

- Johnson, S.J., and Jackson, R.N. (2013). Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol* 10, 33-43.
- Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664-668.
- Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., *et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182-U165.
- Kim, T.K., and Shiekhattar, R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* 162, 948-959.
- Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., *et al.* (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* 18, 956-963.
- Lai, F., Gardini, A., Zhang, A., and Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525, 399-403.
- Laishram, R.S. (2014). Poly(A) polymerase (PAP) diversity in gene expression--star-PAP vs canonical PAP. *FEBS Lett* 588, 2185-2197.
- Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 39, 170-182.
- Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Curr Biol* 20, R754-763.
- Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* 17, 207-223.
- Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., and Jensen, T.H. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep* 10, 178-192.

- Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A., and Jensen, T.H. (2011). Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell* *43*, 624-637.
- Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* *65*, 1099-1122.
- Masternak, K., Peyraud, N., Krawczyk, M., Barras, E., and Reith, W. (2003). Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol* *4*, 132-137.
- McLauchlan, J., Gaffney, D., Whitton, J.L., and Clements, J.B. (1985). The consensus sequence YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res* *13*, 1347-1368.
- Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A., *et al.* (2016). Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol Cell* *64*, 520-533.
- Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* *38*, 2757-2774.
- Moreau, P., Hen, R., Wasyluk, B., Everett, R., Gaub, M.P., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* *9*, 6047-6068.
- Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet* *46*, 1-19.
- Neve, J., Patel, R., Wang, Z., Louey, A., and Furger, A.M. (2017). Cleavage and polyadenylation: Ending the message expands gene regulation. *RNA Biol* *14*, 865-890.
- Ogami, K., Chen, Y., and Manley, J.L. (2018). RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Noncoding RNA* *4*.

Ogami, K., and Manley, J.L. (2017). Mtr4/ZFC3H1 protects polysomes through nuclear RNA surveillance. *Cell Cycle* 16, 1999-2000.

Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J.J., Yates, J.R., 3rd, Tian, B., and Manley, J.L. (2017). An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev.*

Pauli, A., Valen, E., and Schier, A.F. (2015). Identifying (non-)coding RNAs and small peptides: challenges and opportunities. *Bioessays* 37, 103-112.

Perez Canadillas, J.M., and Varani, G. (2003). Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J* 22, 2821-2830.

Preker, P., Almvig, K., Christensen, M.S., Valen, E., Mapendano, C.K., Sandelin, A., and Jensen, T.H. (2011). PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res* 39, 7179-7193.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851-1854.

Puno, M.R., and Lima, C.D. (2018). Structural basis for MTR4-ZCCHC8 interactions that stimulate the MTR4 helicase in the nuclear exosome-targeting complex. *Proc Natl Acad Sci U S A*.

Rassa, J.C., Wilson, G.M., Brewer, G.A., and Parks, G.D. (2000). Spacing constraints on reinitiation of paramyxovirus transcription: the gene end U tract acts as a spacer to separate gene end from gene start sites. *Virology* 274, 438-449.

Rogan, D.F., Cousins, D.J., Santangelo, S., Ioannou, P.A., Antoniou, M., Lee, T.H., and Staynov, D.Z. (2004). Analysis of intergenic transcription in the human IL-4/IL-13 gene cluster. *Proc Natl Acad Sci U S A* 101, 2446-2451.

Ruiz-Orera, J., and Alba, M.M. (2019). Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. *Trends Genet* 35, 186-198.

Ryner, L.C., Takagaki, Y., and Manley, J.L. (1989). Multiple forms of poly(A) polymerases purified from HeLa cells function in specific mRNA 3'-end formation. *Mol Cell Biol* 9, 4229-4238.

Schaffner, W. (2015). Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol Chem* 396, 311-327.

Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Mol Cell* 58, 1101-1112.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 33, 365-376.

Silla, T., Karadoulama, E., Makosa, D., Lubas, M., and Jensen, T.H. (2018). The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts. *Cell Rep* 23, 2199-2210.

Skaar, J.R., Ferris, A.L., Wu, X., Saraf, A., Khanna, K.K., Florens, L., Washburn, M.P., Hughes, S.H., and Pagano, M. (2015). The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res* 25, 288-305.

Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S., *et al.* (2014). Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat Commun* 5, 5531.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1988). Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* 52, 731-742.

Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* 33, 201-212.

Tian, B., and Manley, J.L. (2013). Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* 38, 312-320.

Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* 18, 18-30.

Tuan, D., Kong, S., and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A* 89, 11219-11223.

Vihervaara, A., Mahat, D.B., Guertin, M.J., Chu, T.Y., Danko, C.G., Lis, J.T., and Sistonon, L. (2017). Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nature Communications* 8.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., *et al.* (2015). Enhancer evolution across 20 mammalian species. *Cell* 160, 554-566.

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

Wang, Q., Carroll, J.S., and Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol Cell* 19, 631-642.

Wang, R., Zheng, D., Yehia, G., and Tian, B. (2018). A compendium of conserved cleavage and polyadenylation events in mammalian genes. *Genome Res* 28, 1427-1441.

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

Whalen, S., Truty, R.M., and Pollard, K.S. (2016). Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* 48, 488-496.

Yang, Q., Gilmartin, G.M., and Doublet, S. (2011). The structure of human cleavage factor I(m) hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5' capping and splicing. *RNA Biol* 8, 748-753.

Zheng, D., Liu, X., and Tian, B. (2016). 3'READS+, a sensitive and accurate method for 3' end

sequencing of polyadenylated RNA. *RNA* 22, 1631-1639.

Zheng, D., and Tian, B. (2017). Polyadenylation Site-Based Analysis of Transcript Expression by 3'READS. *Methods Mol Biol* 1648, 65-77.

Figure legends

Figure 1. Polyadenylated eRNAs were detected under Mtr4 depletion via RL-PAT assay and RT-qPCR. (A) Table of pA+ eRNAs detected by reanalyzing the published 3'READS data. Each pA+ eRNA characterized here had at least 2 reads in at least 2 experiment samples. (B) RL-PAT assay analysis of pA+ eRNAs following control, Mtr4, or PABPN1 siRNA treatment of HeLa cells for 72 hours. RNase H treatment in the presence of oligo(dT) were applied to remove the poly(A) tails of RNA. The gel bands were cut out for sequencing to validate the amplification specificity. (C) RT-qPCR analysis of pA+ eRNAs after Mtr4, ZFC3H1, PABPN1 siRNA transfection. cDNAs were synthesized using oligo(dT) primers.

Figure 2. Global analysis of polyadenylated transcripts under Mtr4 depletion condition identified a large fraction of pA+ eRNAs. (A) Western blot analysis of HeLa cells extracts after 48hr of transfection of control, ZCCHC8, ZFC3H1, INTS4 siRNAs with two biological replicates. (B) Table of the classification of pA+ eRNAs detected in 3'READS+ based on their genomic locations. (C) Pie chart of the numbers of each group of pA+ eRNAs identified by 3'READS+ analysis under deficient nuclear surveillance condition. (D) Venn diagram of intergenic, intragenic and ua/eRNAs, which were upregulated at least 2 fold following indicated siRNA treatment as

compared with control siRNA. (E) RT-qPCR analysis of six bidirectional pA+ eRNAs (EN323+/-, EN16441+/-, EN20600+/-, EN23771+/-, EN28928+/-, EN50286+/-) identified in 3'READS+ experiment under the treatment of control, ZCCHC8, ZFC3H1 and INTS4 siRNAs. cDNAs were synthesized using oligo(dT) primer.

Figure 3. Alternative polyadenylation rarely exists in pA+ eRNAs, and is not affected by ZFC3H1, INTS4 or ZCCHC8 KD. (A) Regulation of alternative polyadenylation of protein-coding genes and pA+ eRNAs following KD indicated factors. mRNAs (upper panel) and pA+ eRNAs (lower panel) with significantly upregulated distal (red) and proximal (blue) pA isoforms are indicated in each graph, along with total numbers and ratios of upregulated versus downregulated. Dots and numbers in grey color indicate mRNAs and pA+ eRNAs with no significant change of APA after KD of indicated factor. Significantly regulated isoforms are those with $P < 0.05$ (Fisher's exact test) and abundance change $>5\%$. (B) Changes in relative abundance of the indicated transcript types following knockdown of ZCCHC8, ZFC3H1 or INTS4. The number of genes showing increases (UP) or decreases (DOWN) of each type of transcript are indicated. False discovery rate < 0.05 . Schematic of various transcript types analyzed is shown on the right side: transcripts using the first (F), middle (M), or last (L) potential PAS in the 3' untranslated region (UTR); the intronic PAS(I); the upstream (not 3'-most) exonic PAS (E); and the upstream antisense transcripts (UA).

Figure 4. Poly(A) signals are highly enriched in enhancers and actively utilized in pA+ eRNAs. (A) The ratio of intergenic, intragenic, and ua pA+ eRNAs with canonical PAS detected in 3'READS+ analysis. (B) The ratio of ZCCHC8 KD, ZFC3H1 KD, INTS4 KD sensitive pA+ eRNAs with canonical PAS. (C) Histogram of PAS locations in pA+ eRNAs under indicated factor knockdown conditions. Data are presented as the frequencies of PAS positions within 6 kb of the transcription start site in a strand specific way (each column is 20 nt).

Figure 5. Canonical poly(A) polymerase PAP- α/γ assure proper 3' processing of pA+ eRNAs. (A) Western blot analysis of HeLa cell extracts after 48h of transfection with control, Mtr4, PAP- α , and PAP- γ siRNAs. (B) Agarose gel of RL-PAT assay of bidirectional pA+ eRNA EN11 following PAP and Mtr4 KD as indicated. Gel bands were excised and sequenced, then mapped back to the enhancer regions. Gel bands and termination sites are boxed with different color to indicate different knockdown conditions (yellow box on gel and yellow triangle on genome sequence for control siRNA KD, red for Mtr4 KD, light green for double KD of PAP- α or PAP- γ with Mtr4, dark green for triple KD of PAP- α , PAP- γ and Mtr4). (C) The plus strand transcripts derived from EN11 detected in RL-PAT assay. Both pA+ and pA- status for EN11+ were detected simultaneously after sequencing the excised gel bands. Additionally, different poly(A) lengths and termination sites of EN11+ were observed in two independent experiments.

Figure 6. pA+ eRNAs are transported into the cytoplasm and associate with polysomes under deficient nuclear surveillance. (A) Subcellular fractionation of pA+ eRNAs, EN11, EN14 and

EN17. Subcellular fractionation was performed after 72 hr of siCtrl or siMtr4 treatment, and total RNAs were isolated from each fraction. cDNAs were synthesized using oligo(dT) primer, and the pA⁺ eRNA, EN11, EN14 and EN17 were analyzed by PCR. Gels were prestained with EtBr. (B) Polysome fractionation profile of bidirectional pA⁺ eRNA EN11^{+/-} and EN14^{+/-}. RNA extracted from polysome fractionation were used for oligo(dT)-primed cDNA synthesis, followed with RT-PCR. pA⁺ eRNA level was shown in the EtBr-prestained gel. Ribosome/polysome-associated fractions of pA⁺ eRNAs and GAPDH mRNAs are highlighted with red and green boxes, respectively. Quantification by ImageJ is in Figure S9B.

Figure 7. Model for the biogenesis and regulation mechanisms of enhancer derived transcripts. (A) Model for Integrator-mediated biogenesis of RNAs. In the scenario of transcription elongation of protein-coding genes, depletion of the Integrator complex causes early termination of transcription, and thus lead to accumulation of ptRNAs. During transcription activation, depletion of the Integrator complex causes the switch of transcription machinery from promoters to enhancer regions, and consequently generated a large amount of pA⁺ eRNAs. (B) Model for pA⁺ eRNA biogenesis and degradation mechanisms. Both pA⁻ and pA⁺ eRNAs are generated by RNAPII with the aid of the Integrator complex, and their 3' end cleavage and polyadenylation requires PAP- α/γ . The pA⁺ eRNAs will be degraded by the exosome via both NEXT and PAXT complex targeting pathways.

Supplementary Figure Legends

Figure S1. The identification of pA+ eRNA under Mtr4 deficient condition via 3'READS analysis. (A) Schematic illustration of enhancer classification based on the genomic location: intergenic (inter), intragenic (intra), upstream antisense (ua) groups, regardless the overlap of certain enhancers with other DNA elements. (B) Pie chart of the number of each group of pA+ eRNAs identified by 3'READS analysis under depletion of Mtr4. (C) Table of the classification of putative pA+ eRNAs based on the genomic location. (D) The snapshots of 3'READS and RNA-sequence reads of four pA+ eRNAs (EN515, EN29068, EN27780, EN70250) under deficient nuclear surveillance conditions. The polyA peaks are indicated with blue arrows.

Figure S2. Schematic illustration of the RL-PAT assay. RNAs were ligated with the anchor oligonucleotide and reverse-transcribed using the primer complementary to the anchor sequence. The resulting cDNA was subjected to semi-nested PCR (1st round PCR using primers indicated by red and blue arrows; 2nd round PCR using primers indicated by orange and blue arrows), and PCR products were resolved in 2.0% agarose gels prestained with EtBr. In the end, the gel bands were cut for sequencing.

Figure S3. A very small number of pA+ eRNAs were downregulated under deficient nuclear surveillance. Venn diagram of intergenic, intragenic and ua/eRNAs, which were downregulated at least 2 folds under each siRNA treatment.

Figure S4. Validation pA⁺ eRNAs detected in 3'READS analysis with RL-PAT assay. Four pA⁺ eRNAs (EN323, EN20600, EN23771, EN50286) were validated by RL-PAT assay analysis following gel sequencing. The snapshots of 3'READS+ reads are displayed below the gel images.

Figure S5. Noncanonical Poly(A) signal AUUAAA is actively utilized in pA⁺ eRNAs. Histogram of noncanonical PAS locations in pA⁺ eRNAs under indicated factor knockdown conditions. Data are presented as the frequencies of PAS positions relative to the transcription start site in a strand specific way (each column is 20 nt). Position 0 corresponds to the 3' base of the poly(A) signal.

Figure S6. G/U rich regions are enriched at the 20nt downstream of the cleavage sites of both pA⁺ eRNAs and mRNAs. Histogram of the distance between G/U rich regions and the termination sites of pA⁺ eRNAs and mRNAs. Data are presented as the density of positions relative to the transcription termination site in a strand specific way for pA⁺ eRNAs (each column is 20 nt). Position 0 corresponds to the 3' base of the poly(A) signal.

Figure S7. CPSF-30/-73 didn't regulate the 3' cleavage of pA⁺ eRNAs. (A) Western blot analysis of HeLa cells extracts after 72 hr of transfection with control, Mtr4, CPSF-30, CPSF-73 siRNAs. (B) RT-qPCR analysis of the uncleaved transcripts of four genes (gapdh, β -actin, fos, fn1). The relative amount of the indicated uncleaved transcripts was calculated by using primers spanning the last poly(A) site of each gene and normalized to siCtrl and the internal probe of the

gene. (C) RT-qPCR analysis of the uncleaved transcripts of three pA⁺ eRNAs in both transcription directions (EN11+/-, EN14+/-, EN17+/-). The amount of the uncleaved transcripts of pA⁺ eRNAs was calculated by using primers either spanning or locating downstream last poly(A) site and normalized to siCtrl and the primer upstream of the poly(A) site.

Figure S8. RL-PAT assay analysis of bidirectional pA⁺ eRNA EN14 and EN17 following knockdown of PAP- α / γ together with Mtr4. Agarose gel of RL-PAT assay of bidirectional pA⁺ eRNA EN14 and EN17 following PAP- α / γ and Mtr4 KD conditions. The gel bands were cut and sequenced, then mapped back to the enhancer regions. Gel bands and termination sites are boxed with different color to indicate different knockdown conditions (yellow box on gel and yellow triangle on genome sequence for Ctrl siRNA KD, red for Mtr4 KD, light green for double KD of PAP- α or PAP- γ with Mtr4, dark green for triple KD of PAP- α , PAP- γ and Mtr4).

Figure S9. Quantification of polysome fractionation of pA⁺ eRNA EN11 and EN14. (A) The longest putative ORFs detected in EN11 and EN14. (B) Band strength in each fraction in Figure 6B was quantitated using ImageJ. Absolute signal values after background subtraction are plotted. Ribosome-bound and -free fractions are indicated by a black dashed line.

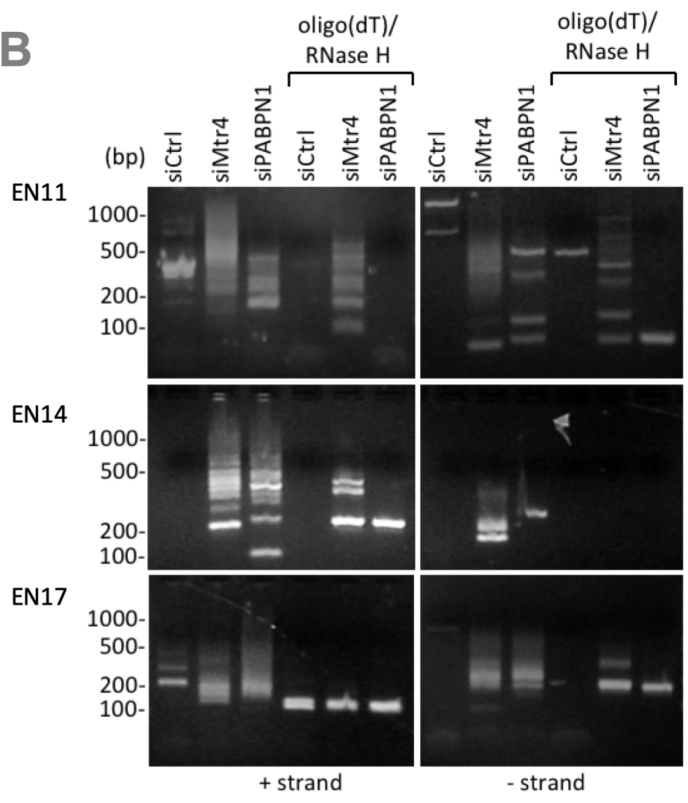
Figure 1_Chen

A

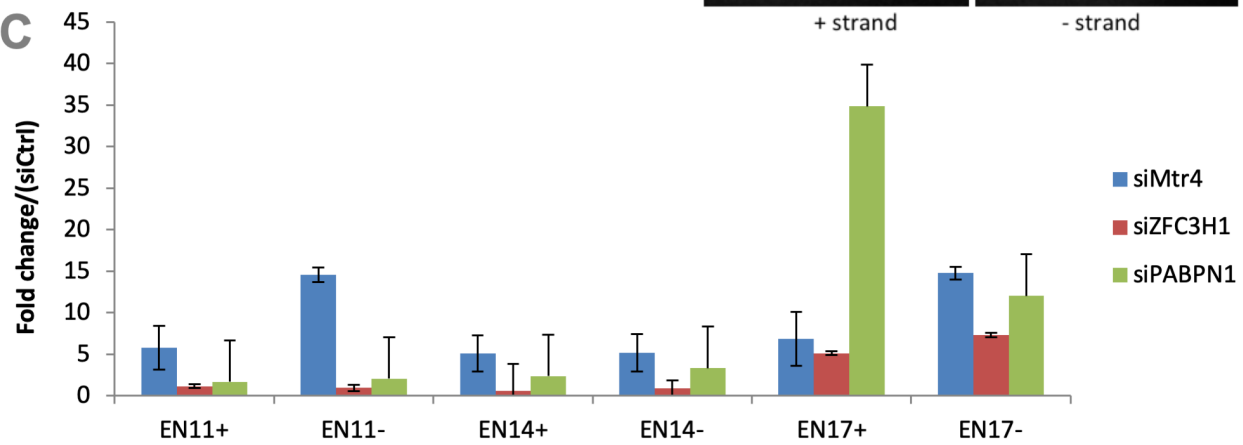
Exp condition	No. of pA+ eRNA
siCtrl	1761
siMtr4	2491
siZCCHC8	2199
siRBM7	2077
Total*	3919

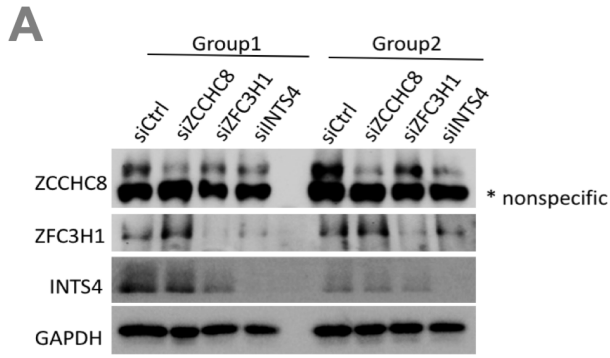
* Each pA has >= 2 reads in >= 2 samples

B



C





B

Classification	Location	Number	Sum
Intergenic Group	Intergenic	1,859	2,557
	5'UTR	54	
	ncRNA	644	
Intragenic Group	Intragenic	1,421	2,761
	CDS	42	
	3'UTR	1,298	
ua Group	ua/eRNA	1,135	1,135
		Total	6,453

*Each pA has >= 2 reads in >= 2 samples

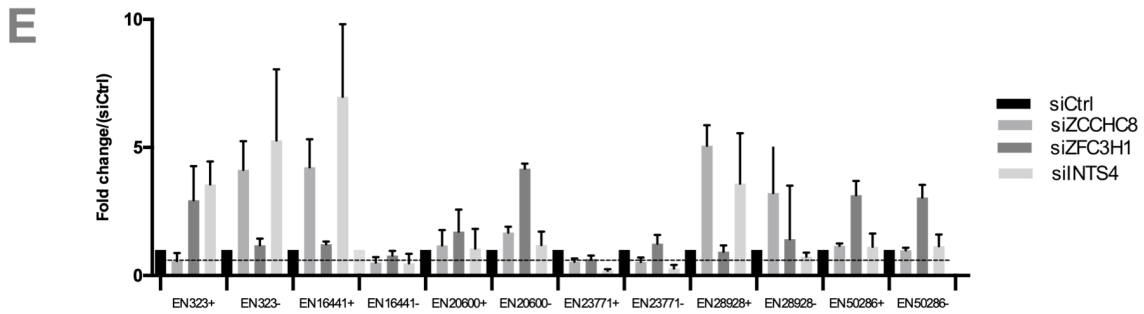
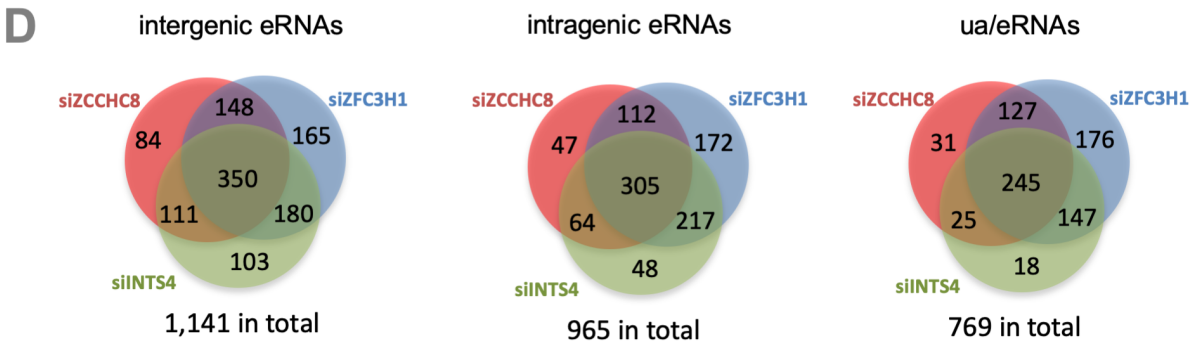
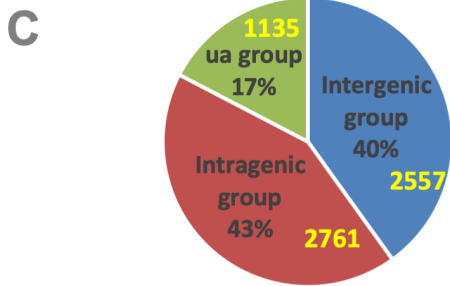
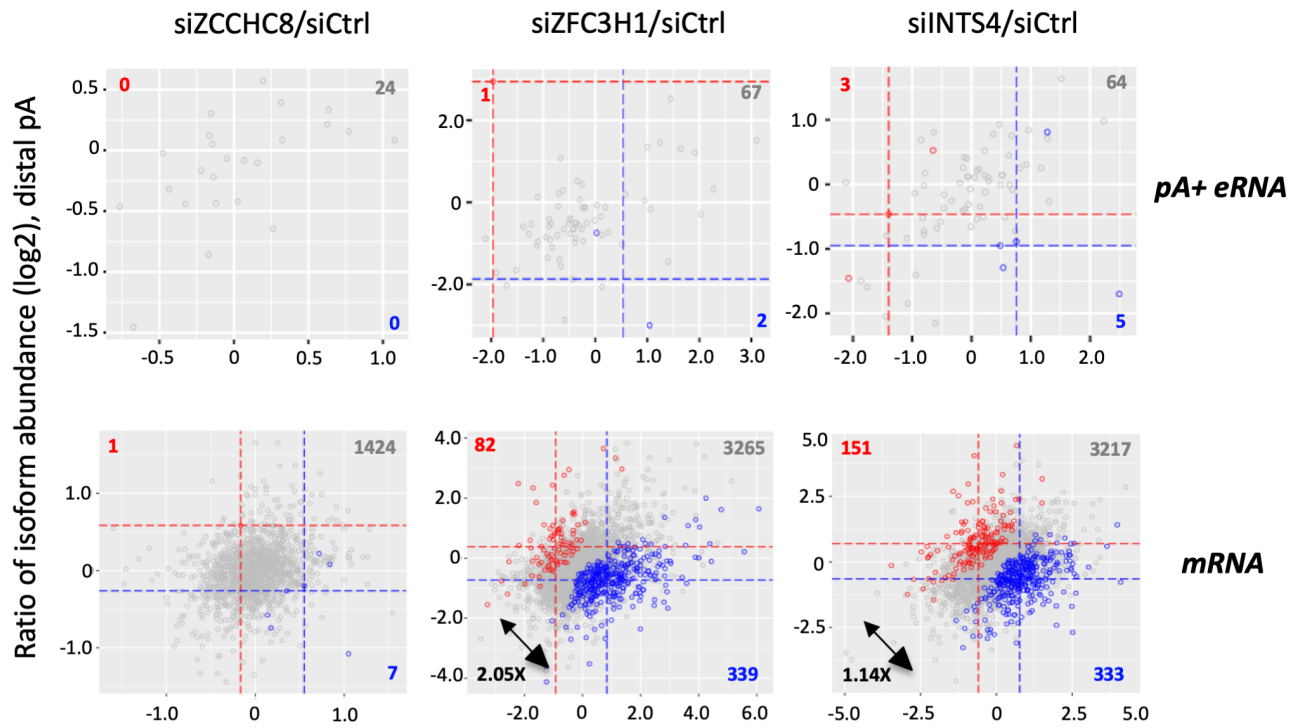


Figure 3_Chen

A



B

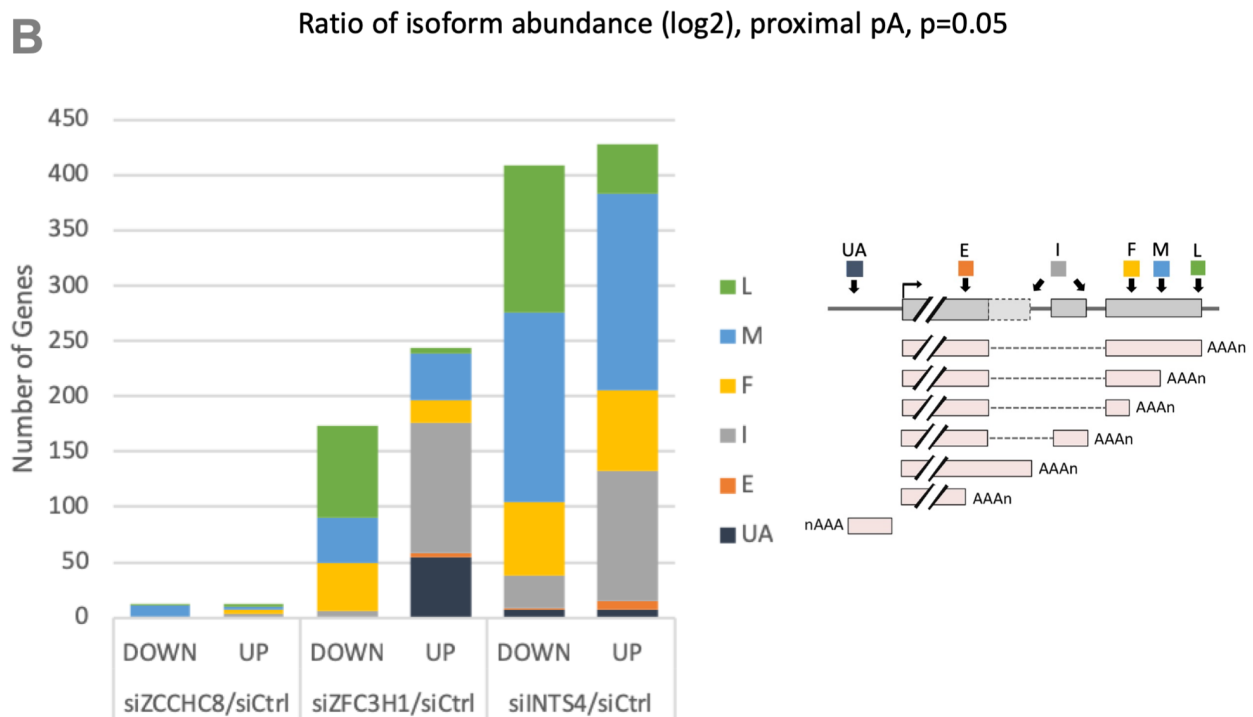


Figure 4_Chen

A

	intergenic enhancer		intragenic enhancer		ua enhancer		Strands
	AAUAAA	Total	AAUAAA	Total	AAUAAA	Total	
ENCODE enhancers (55,799 in total)	12,343	21,448	16,467	30,986	1,582	3,365	(+)
	12,419		16,602		1,567		(-)
w/ AAUAAA	79.3%		74.3%		66.0%		(+)/(-)
w/ AAUAAA or AUUAAA	88.6%		83.7%		76.2%		(+)/(-)

B

	intergenic eRNA		intragenic eRNA		ua/eRNA		Strands
	AAUAAA	Total	AAUAAA	Total	AAUAAA	Total	
pA+ eRNA in 3'READS+ (4,415 in total)	660	933	492	745	416	545	(+)
	671		443		676		449
w/ AAUAAA	71.6%		65.8%		76.2%		(+)/(-)
w/ AAUAAA or AUUAAA	91.9%		90.1%		98.3%		(+)/(-)

C

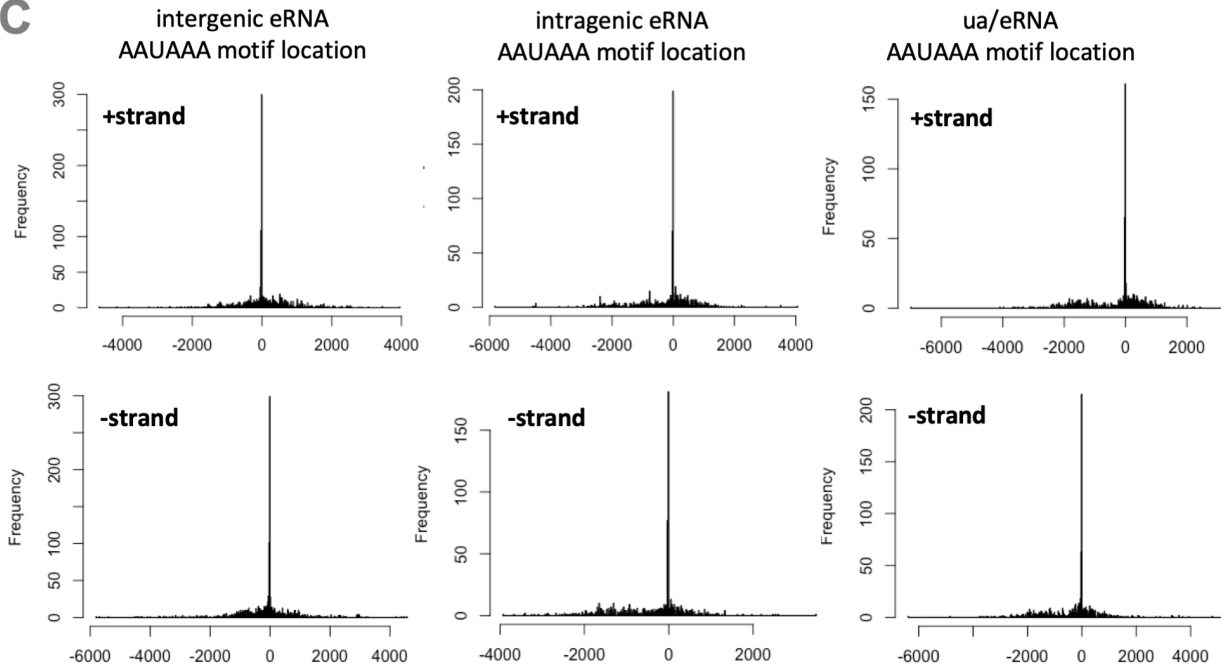


Figure 5_Chen

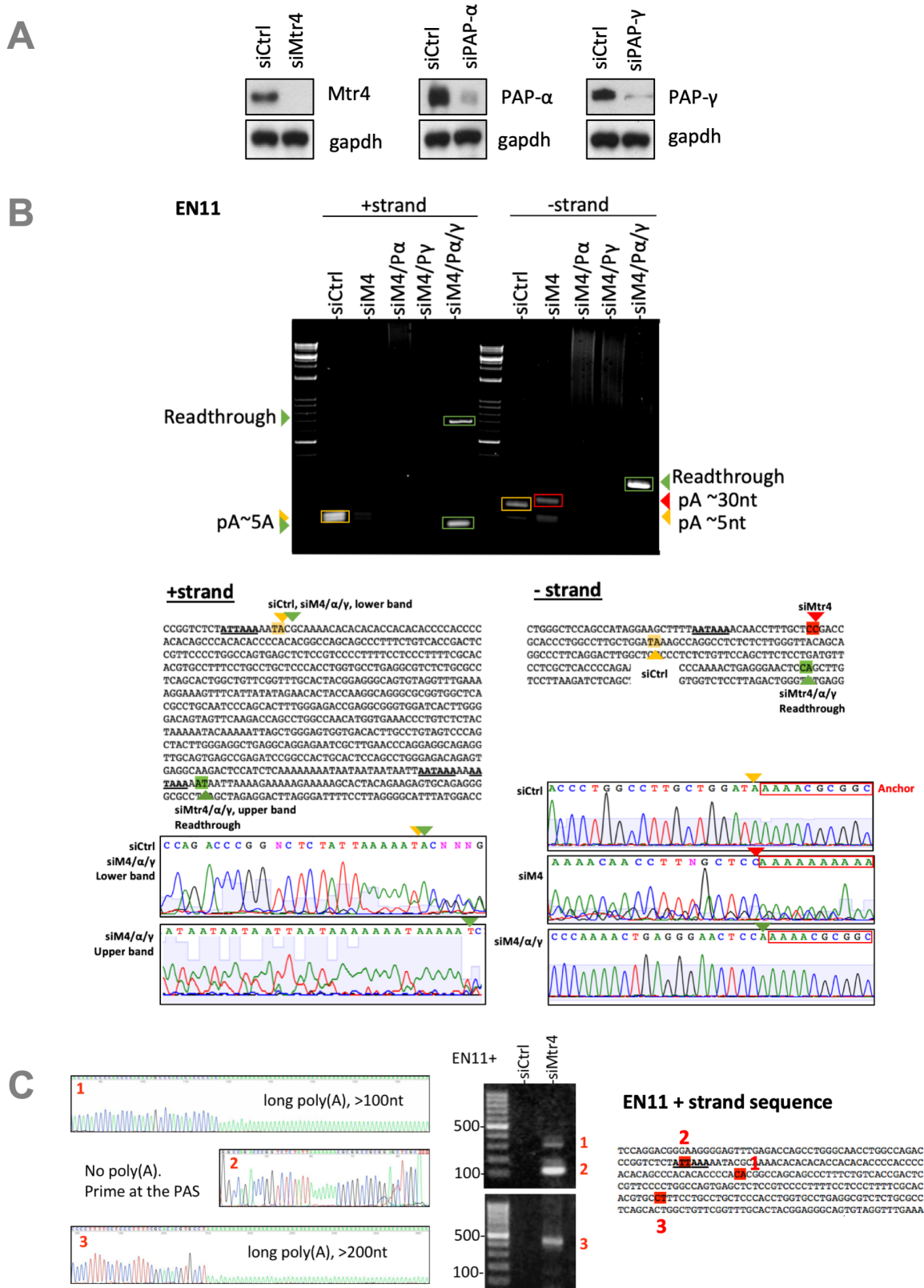


Figure 6_Chen

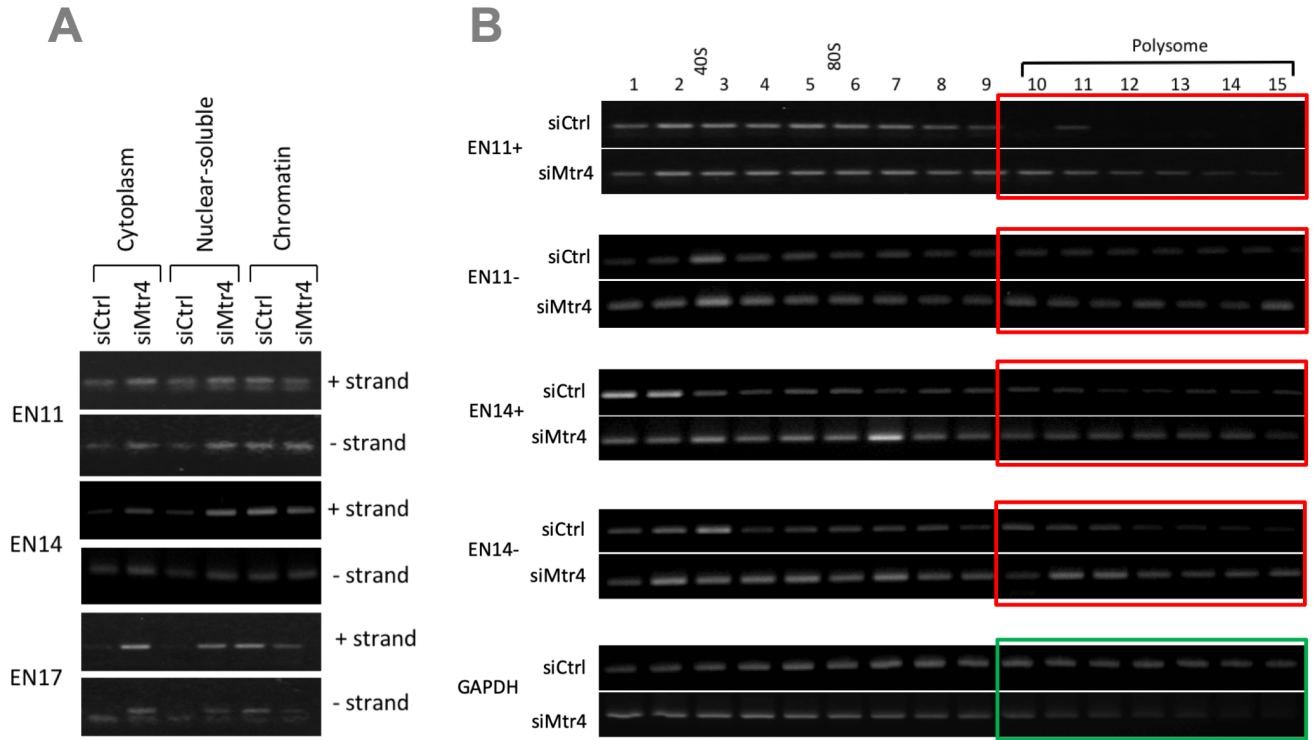


Figure 7_Chen

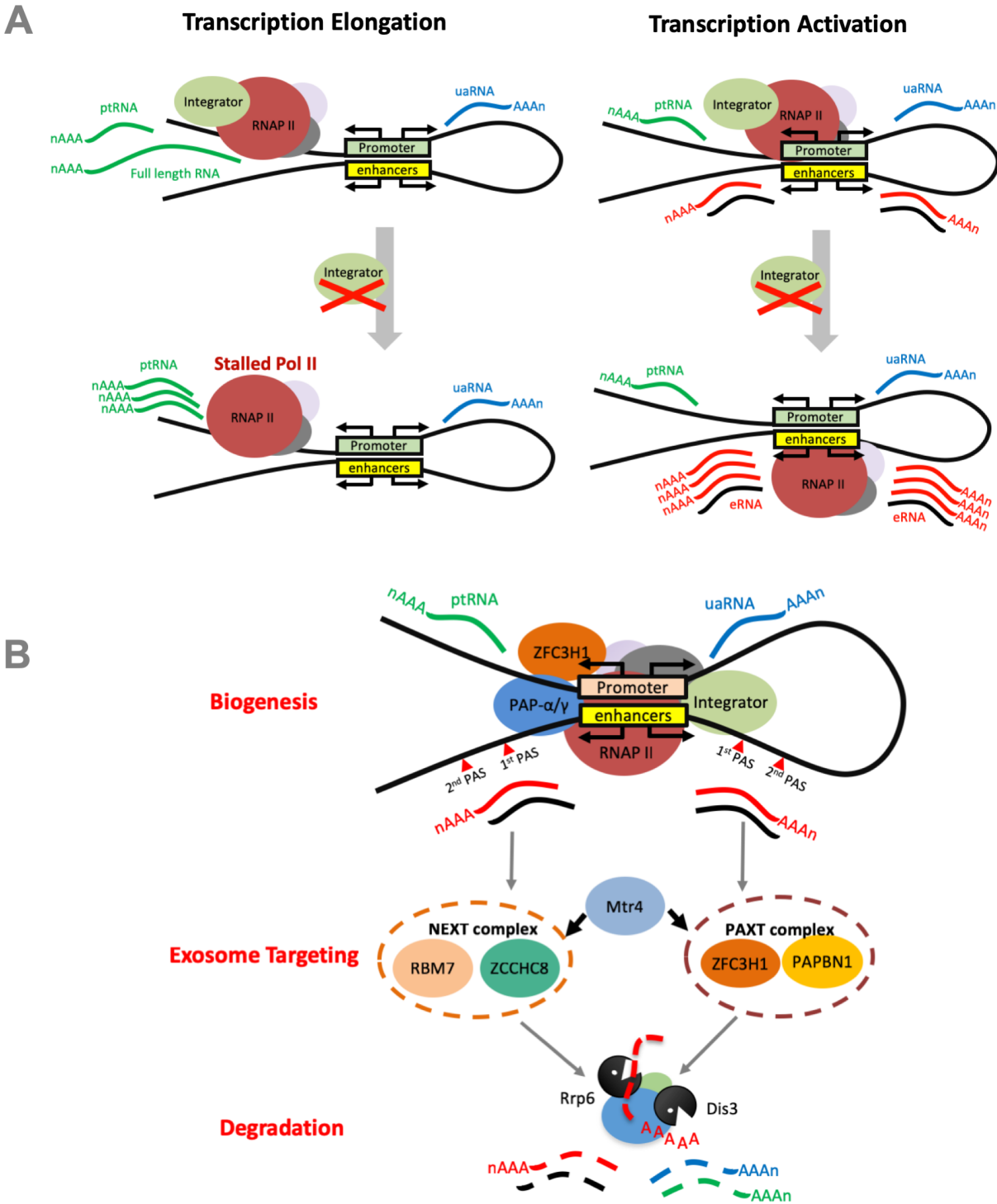
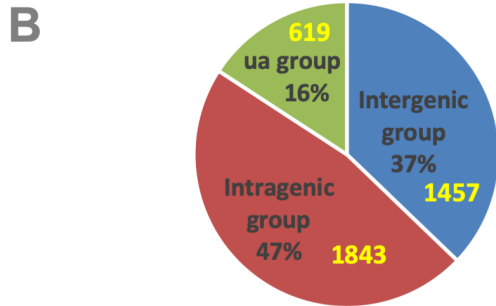
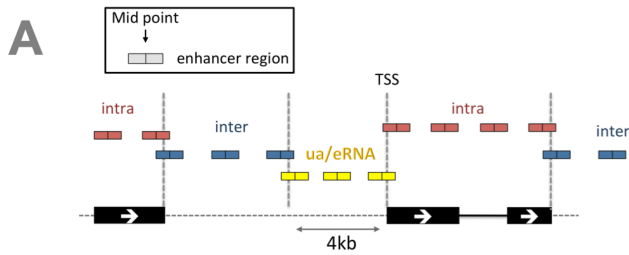


Figure S1_Chen



C

Classification	Location	Number	Sum
Intergenic Group	Intergenic	1071	1457
	5'UTR	16	
	ncRNA	370	
Intragenic Group	Intron	697	1843
	CDS	13	
	3'UTR	1069	
	3'UTR/intron	64	
ua Group	uaRNA	619	619
		Total	3919

Each pA has ≥ 2 reads in ≥ 2 samples

D

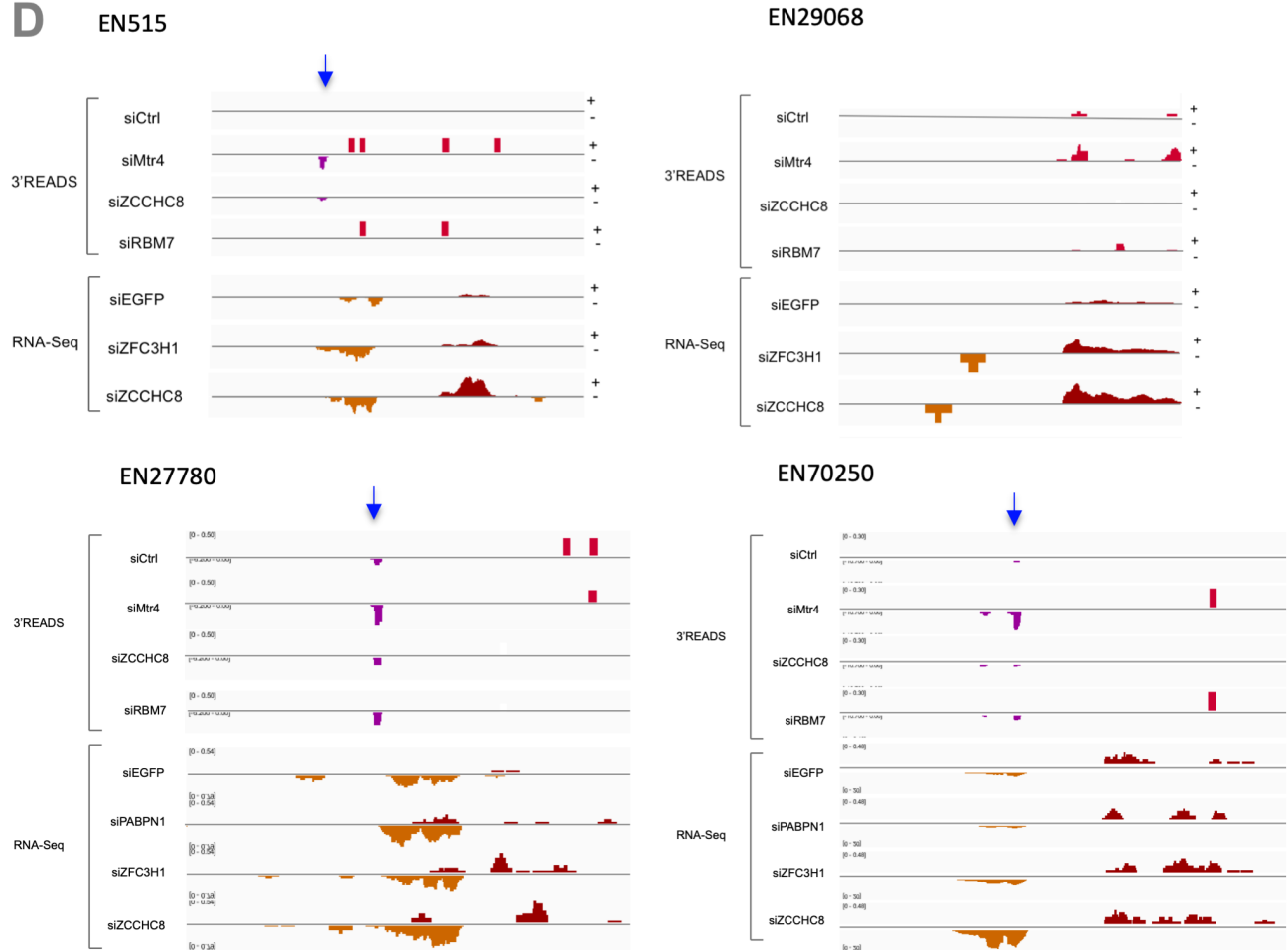
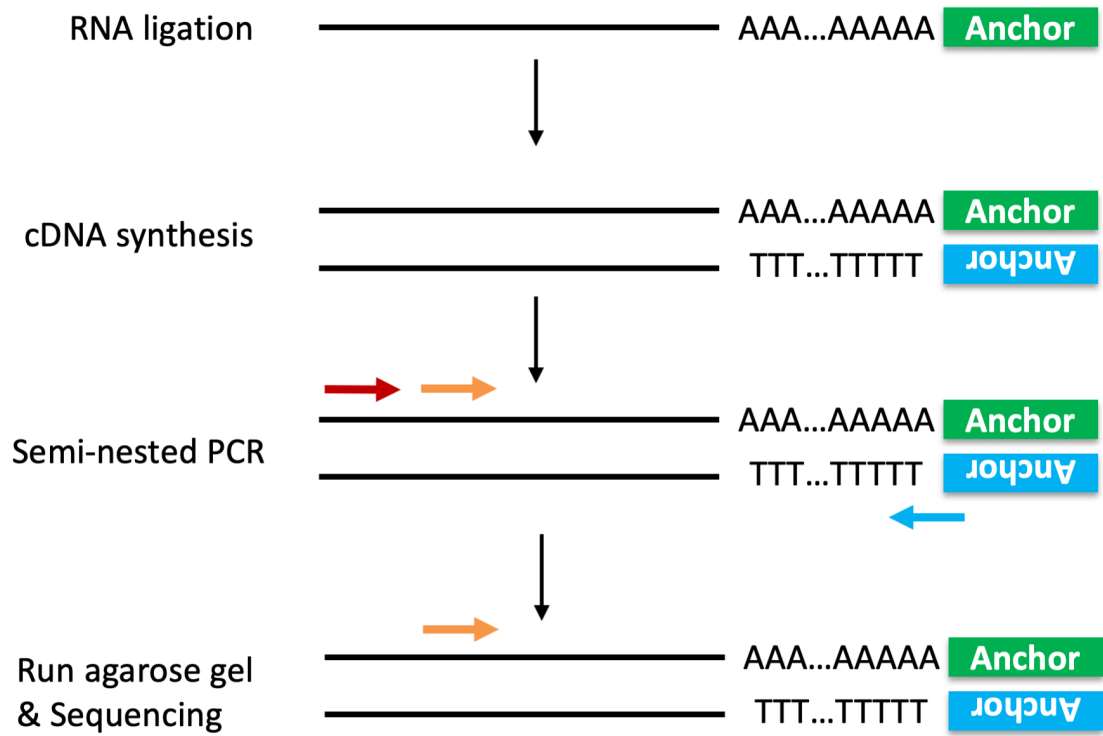


Figure S2_Chen



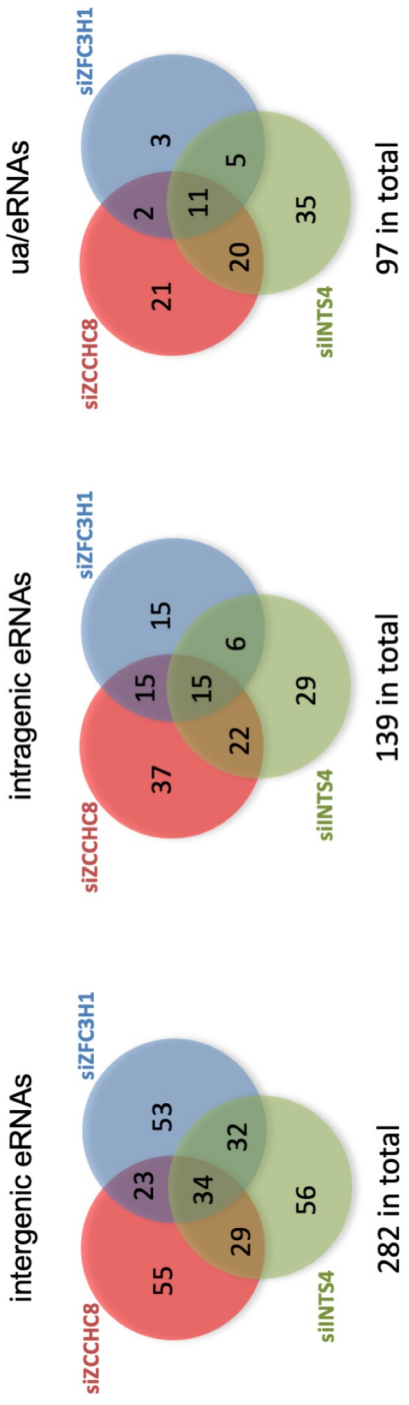
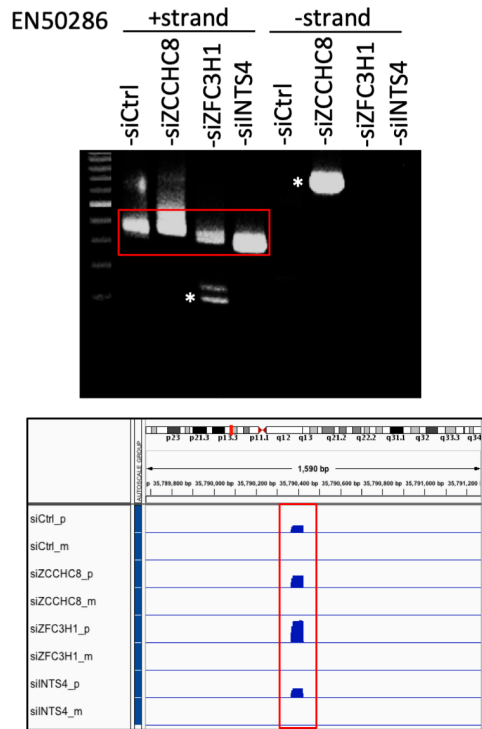
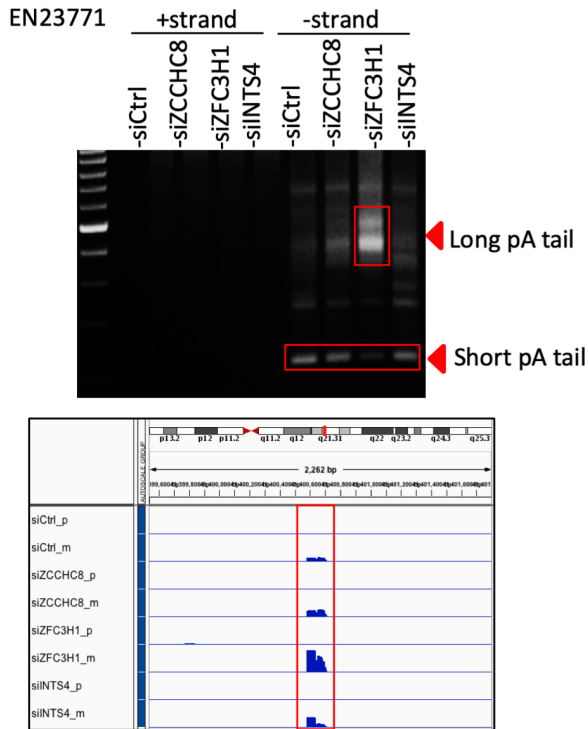
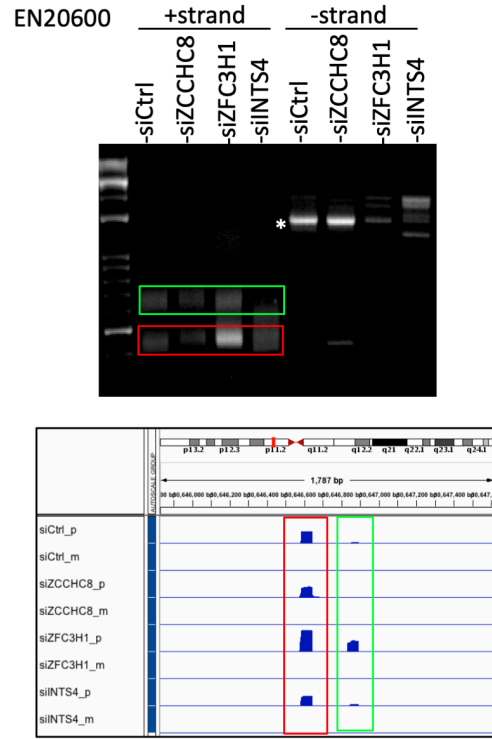
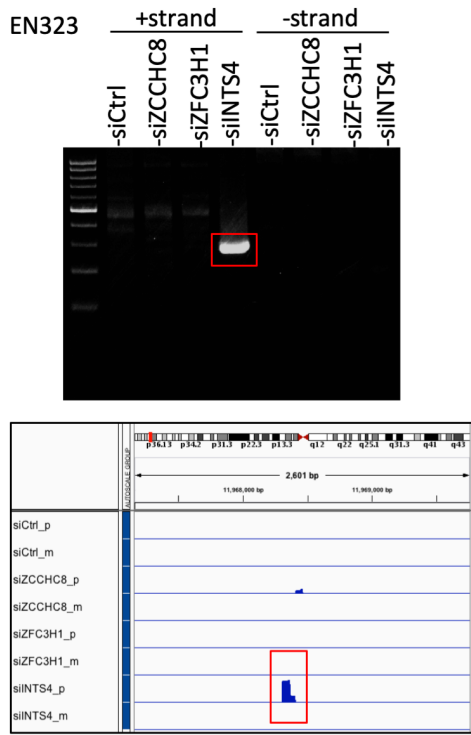
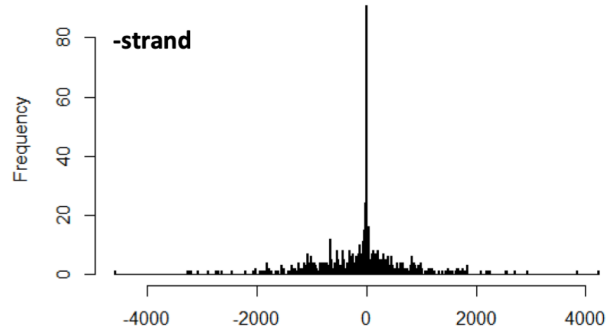
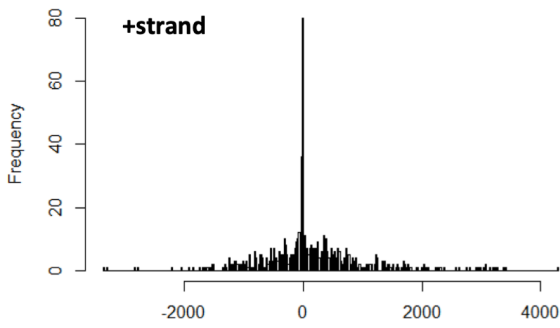


Figure S3_Chen

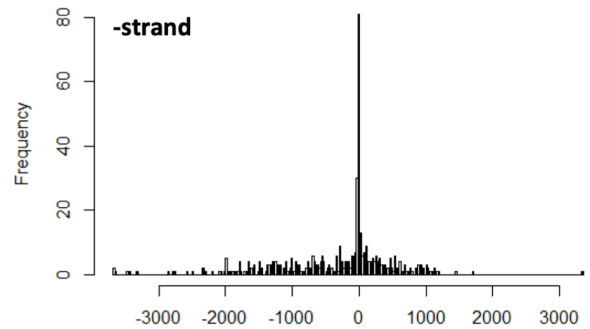
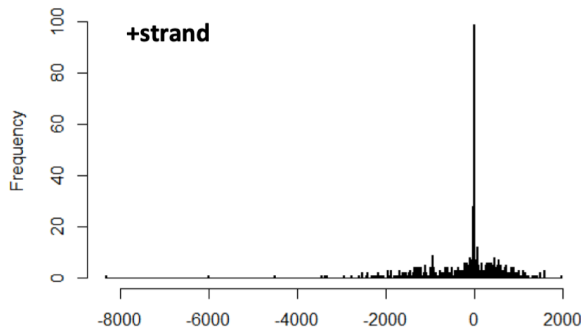
Figure S4_Chen



Histogram of AUUAAA motif locations for Intergenic eRNAs



Histogram of AUUAAA motif locations for Intragenic eRNAs



Histogram of AUUAAA motif locations for ua/eRNAs

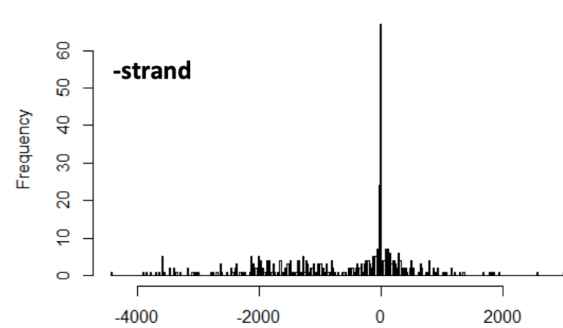
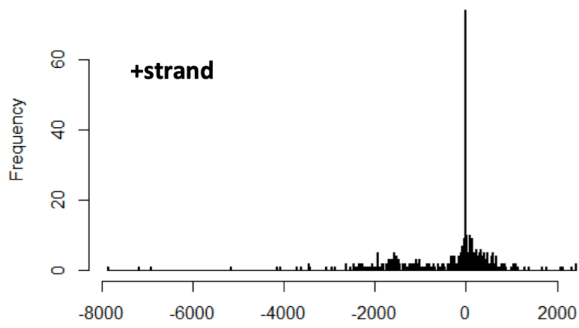
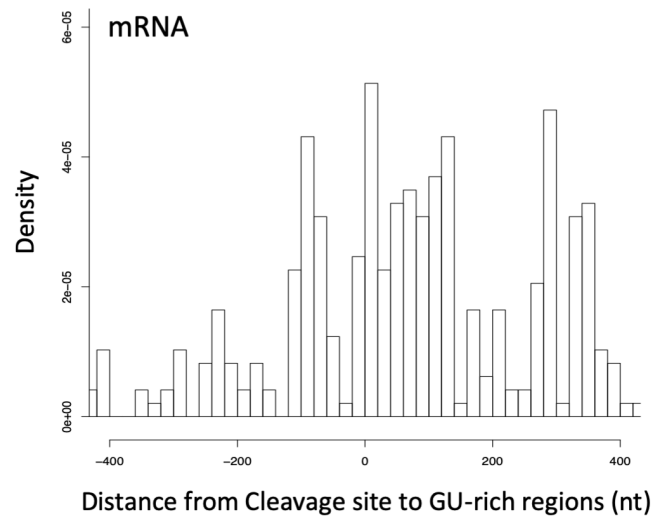
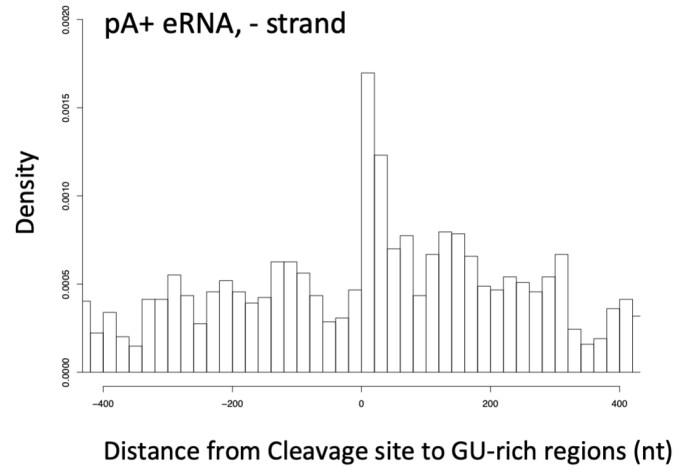
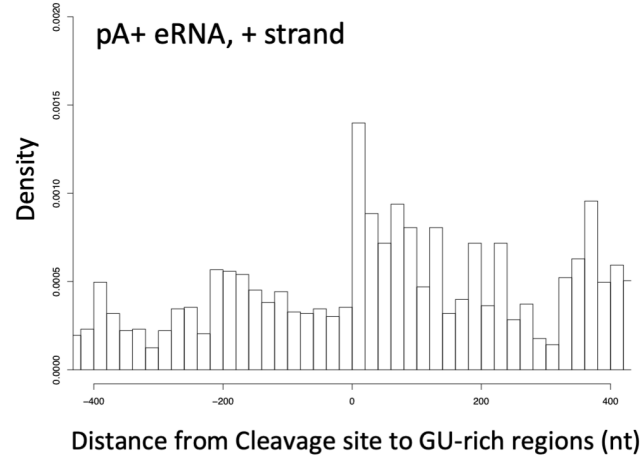


Figure S6_Chen



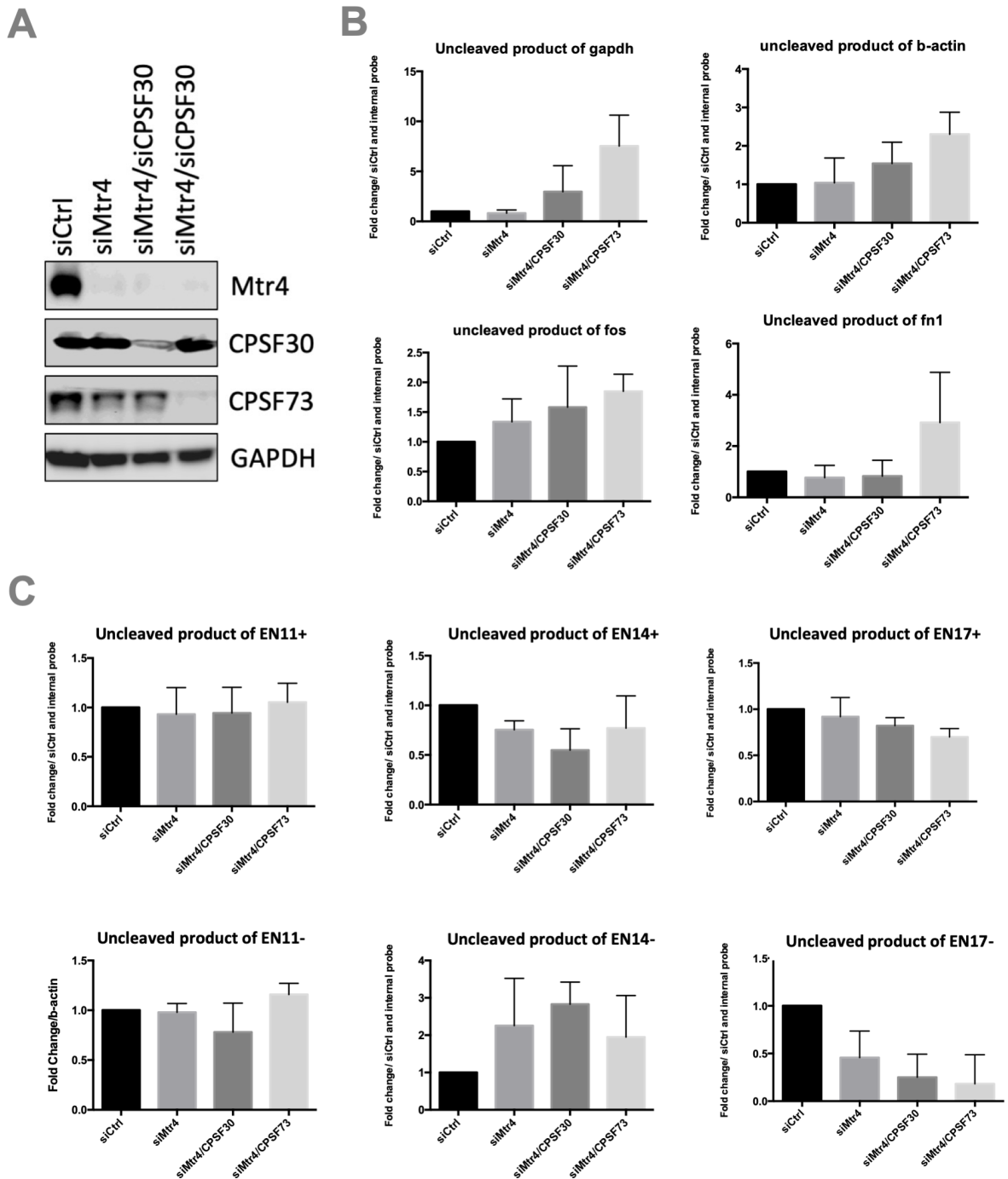


Figure S9_Chen

A EN11 (hg19_dna range=chr11:65243603-65246000)

+ strand longest putative ORF (465nt)

```

M G L T K P G W Q P V P P H
GAAGGCGAGAAGGGGGCATGGGGCTGACAAAGCCGGGGTGGCAGCCCGTCCCACCCAC
S V W V P A V C S A Q P Q G P P T T L A
TCAGTGTGGGTCCCGCGGTCTGTTCGGCGCAGCCACAGGGTCTCCACCCACCTCGCG
A L R Q P R Y R F P D F G T E T T T P A
GCCCTTCGCCAGCCAGATACCGTTTCCCGACTTTGGAACGGAGACCACAACACCCGCC
P E L L K A Q V R S T R Q E T G S Q I P
CCGGAACACTGAAGGCTCAAGTTAGGTCAACCCGGCAGGAACCGGTTCCCAAATCCCA
R R T Q A R G S L P Q G W N S A Q W P P
AGGGCGACCCAGGCCCGGTTCCTCCCTCAAGCTGGAACAGCGCCAGTGGCCACCG
P P G L Q P P G A R P P P G L S P M R S
CCGCCGGTCTCCAGCCCCAGGAGCGCGGCTCTCCCGACTCTCTCCGATGCGGTCC
C P G D T A G R R D L A G G R G P G A G
TGCCCTGGGACACGGCAGGAAGAAGGACTTGGTGGGGACGTGGCCCCGGCGGGGG
A G E A A R D S Q R P L R G C P P A V T
GCAGGGGAGGCAGCGGGGACTCCAGCGGCCCTGCGCGGGTGCCACCCCGCGCTCACG
*
TGACACCGCGGGCGCTGAGTCACCGGAAGCATCTGGCTGGGCGCGGGATTGGCCAAT
    
```

- strand longest putative ORF (291nt)

```

M E S C L T L S P R L E C S G R I S A
GAGATGGAGTCTTGCCCTACTCTGTCTCCAGGCTGGAGTGCAGTGGCCGGATCTCGGCT
H C N L C L L G S S D S P A S A S Q V A
CACTGCAACCTCTGCCTCCGGTTCAAGCGATTCTCTGCCTCAGCTCCCAAGTAGCT
G T T G K C H H S Q L I F V F L V E T G
GGGACTACAGGCAAGTGTCAACACTCCAGCTAATTTTGTATTTTAGTAGAGACAGGG
F H H V G Q A G L E L L S P S D P P A S
TTTACCATGTTGGCCAGGCTGGTCTTGAACTACTGTCCCAAGTATCCACCCCGCTCG
V S Q S A G I A G V S H R A L P W *
GTCTCCAAAGTGCTGGGATTGACAGGCTGAGCCACCGCGCCCTGCCTTGGTAGTGTCT
    
```

EN14 (hg19_dna range=chr14:52546966-52549000)

+ strand longest putative ORF (204nt)

```

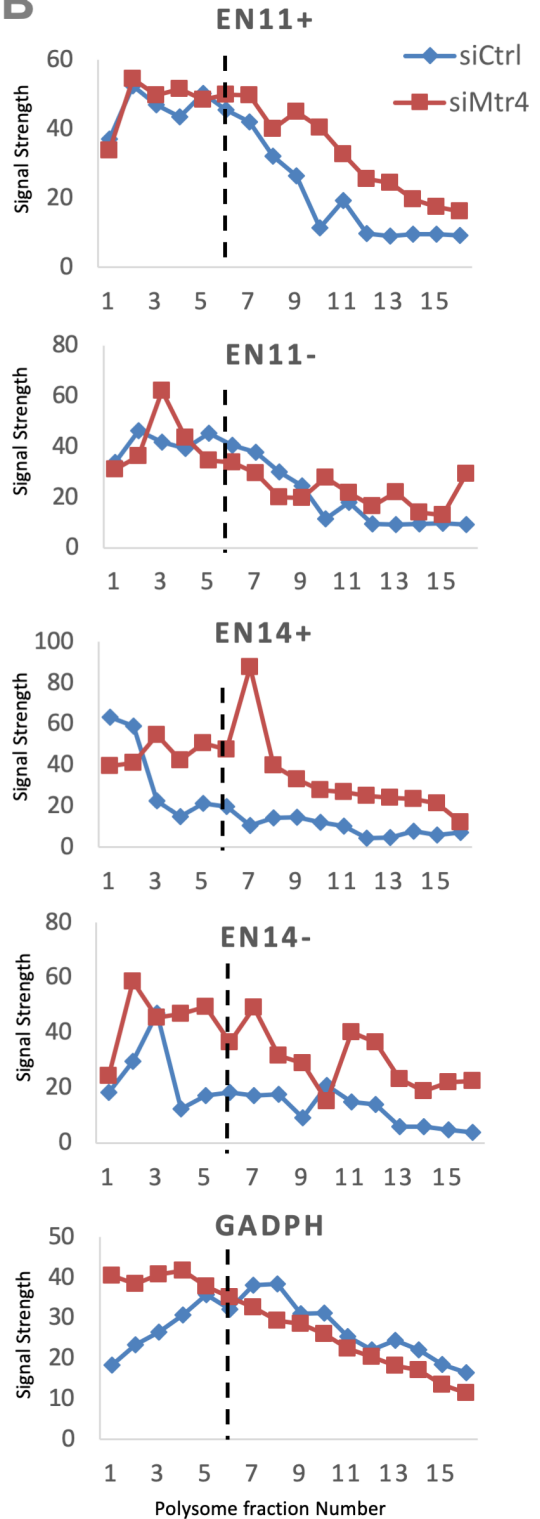
M D L T K L S S A F Q P W A
TTTTTAGGCTCAGCTGGAATAATGGACTTGACAAAGCTCAGCTCAGCTTCCAGCCCTGGG
Q E I L L C C Y F C V C V C V F S I S S
CTCAGGAGATTCTGTGTGTTGTTATTTTGTGTGTGTGTGTGTGTGTCTTCTATCTCCT
S F F S S F V L F C F Y C C K V K Y N W
CCTCTTTCTTCTCTTTGTTTGTGTTTGTGTTTACTGCTGCAAGGTAATAATAAAT
E Q G G Y F A K S D A N C *
GGGAACAGGGTGGCTATTTTGCCAAATCCGATGCTAAGTCTAAAGAGGTGATTCTCATG
    
```

- strand longest putative ORF (222nt)

```

M T G G C G W
AAGTGTGTAGTGGAGTTCACGCTCAGTTACTACCATGCATGACAGGGGTTGGGATGG
C P S R L E H Q T A P P S Q V R A P F P
TGTCATCAAGGCTTGAACACCAGACAGCACCCTTACAGGTGAGGGTCCCTTTCCC
Y P K D G F I M F L L N S S S S S S F Y
TACCCGAAGGATGGATTCAATGTTTCTCTTAAATCCTCAAGCTCCTCTTCTCTCTAT
R N C E L K R H L K G R T R V S L L F Q
CGAACTGTGAATAAAAAGGCACCTTGAAGGGCAGAACAGGGTCTCTCTCTTCCAG
F G D P N A *
TTTGGGACCCAAATGCTTAAGCTCCACAAGTAACTGCAACGCTCATGGACAGCATGA
    
```

B



**Chapter 4. Dynamic Subcellular Localization of RBBP6 Isoform3 Is Regulated by
Monoubiquitylation and Importin 7**

Dynamic Subcellular Localization of RBBP6 Isoform3 Is Regulated by Monoubiquitylation and Importin 7

Yaqiong Chen¹, Dafne Campigli Di Giammartino^{1,2}, James J. Moresco³, John R. Yates III³, Carol Prives¹ and James L. Manley^{1*}

¹. *Department of Biological Sciences, Columbia University, New York, New York 10027, USA.*

². *Joan & Sanford I. Weill Department of Medicine, Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10021, USA.*

³. *Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA*

* Corresponding author

E-mail: jlm2@columbia.edu

Abstract

Retinoblastoma-binding protein 6 (RBBP6) was initially identified as a large multidomain protein, interacting with p53 and Rb. Later, its diverse roles were uncovered in cell cycle progression, apoptosis, nucleic acid metabolism, differentiation, and mRNA processing. RBBP6 protein has four isoforms, among which the shortest isoform, iso3, has only one domain: the DWNN (domain with no name) domain. The DWNN domain displays high similarities with ubiquitin, implying its function as a novel ubiquitin-like modifier. However, we show that the DWNN domain is actually not a ubiquitin-like modifier, but is itself ubiquitinated. Moreover, the monoubiquitylation of iso3 can facilitate its localization at chromatin. Additionally, we found that the C-terminal tail of iso3 also plays a role in iso3 chromatin localization, presumably by interacting with other factors of the polyadenylation machinery. Pulldown experiments of iso3 followed by mass spectrometry identified Importin 7 as an iso3-interacting factor that assists its cytoplasmic localization. Our results identified novel mechanisms for the dynamic localization of RBBP6 iso3, which shed light on the role of iso3 in mRNA 3' processing and disease.

Introduction

The retinoblastoma-binding protein 6 (RBBP6) was initially discovered as a large multidomain protein (~ 250 KDa), interacting with both p53 and Rb (Saijo et al., 1995; Sakai et al., 1995; Simons et al., 1997). While the human *RBBP6* gene codes for four protein isoforms, a majority of the early studies focused on the long-form RBBP6 protein, and revealed its crucial roles in various biological processes, including cell-cycle progression, apoptosis, nucleic acid metabolism, differentiation and mRNA processing (Di Giammartino et al., 2014; Gao and Scott, 2003; Li et al., 2007; Miotto et al., 2014). RBBP6 was also found to be involved in various diseases: for example, the protein is overexpressed in many types of cancer (Chen et al., 2013; Moela and Motadi, 2016; Motadi et al., 2011; Motadi et al., 2018; Yoshitake et al., 2004). This is consistent with the finding that the long isoform of RBBP6 is a negative regulator of p53, by functioning as a scaffold protein to promote MDM2-mediated ubiquitination and degradation of p53/TP53 (Li et al., 2007). Another example is that RBBP6 mutation in its p53-binding domain leads to predisposition to myeloproliferative neoplasms, which suggests a p53-dependent pathway of RBBP6 in pathogenesis (Harutyunyan et al., 2016). Additionally, a more recent study identified RBBP6 as a negative regulator of Ebola virus replication, and thus has the potential to be a therapeutic target (Batra et al., 2018).

Despite the diverse roles of RBBP6 revealed in previous studies, mRNA 3' processing has been one of the most conserved function of RBBP6, from human (Di Giammartino et al., 2014) to yeast (Lee and Moore, 2014; Vo et al., 2001). Interestingly, RBBP6 is expressed in all eukaryotic organisms but its p53- and Rb-binding domains are only present in vertebrates (Figure 1A),

indicating that the ability to modulate mRNA processing might be a more conserved functions of RBBP6 in all organisms, while the p53 and Rb binding domains in vertebrates allows for more specific regulatory mechanisms.

RBBP6 homologs across species share three well-conserved domains at their N-terminus: the “Domain With No Name”, i.e. DWNN, the Zinc knuckle, and the RING finger domains (Figure 1A). It has been found that the DWNN domain itself is expressed in vertebrates with a short C-terminal tail, which is identified as RBBP6 isoform3 (iso3). A previous study in our lab has revealed the inhibitory function of RBBP6 iso3 by competing with long-form RBBP6 for the polyadenylation machinery binding (Di Giammartino et al., 2014). Remarkably, the structure of the DWNN domain displays high similarities with ubiquitin, which raised the intriguing possibility that RBBP6 iso3 itself may function as a novel ubiquitin-like modifier of other protein substrates (Pugh et al., 2006). Furthermore, the RING domain of the long-form RBBP6 is a signature domain for E3 ubiquitin ligases, and RBBP6 has been found to be an E3 ligase for several proteins including YB-1, ZBTB38, and TP53 (Chibi et al., 2008; Li et al., 2007; Miotto et al., 2014). We consequently speculated that the long-form RBBP6 might serve as the E3 ligase for iso3 modification, if such modification exists.

Here we describe experiments that explore whether iso3 is a ubiquitin-like modifier, and whether long-form RBBP6 functions as an E3 ligase for such a modification mediated by iso3. Surprisingly, we found that the DWNN domain is not a ubiquitin-like modifier, but instead is itself ubiquitinated. We investigated further the ubiquitylation process of iso3, and found that different ubiquitylation status, especially monoubiquitylation of iso3, can modulate its subcellular

localization. We also found that the C-terminal tail of iso3 facilitates its nuclear localization. We also showed that the long form of RBBP6 is not the E3 ligase of iso3 ubiquitination. Furthermore, we identified by his-tag pulldown and mass spectrometry analysis an import factor, Importin 7 (Imp7), that physically interacts with iso3, and revealed its crucial function in facilitating iso3 cytoplasmic retention. Finally, we discuss the functional significance of such dynamic subcellular localization of RBBP6 iso3 protein, using cancer development as an example.

Results

Lysine-dependent protein modification is observed in RBBP6 iso3 protein

The DWNN domain at the N-terminus of all four isoforms of the RBBP6 protein has been found to have a ubiquitin-like structure, and lysine residues that are likely to be equivalent to ubiquitin lysines (Pugh et al., 2006). Consequently, the initial aim of our experiments was to investigate whether the DWNN domain is a ubiquitin-like modifier as previously suggested. As shown in Figure 1A, the DWNN is expressed as a small single-domain protein, RBBP6 isoform3 (iso3), in human. We constructed an N-terminally HA- and his-tagged iso3 (HA-his-iso3) vector, and overexpressed iso3 protein in HeLa cells. Exogenous iso3 protein accumulated in a time-dependent manner, and, intriguingly, multiple bands appeared after 36hr transfection (Figure 1B). The molecular weight of HA-his-iso3 protein is approximately 17KDa, thus bands with molecular weights above 17KDa suggested that either RBBP6 iso3 protein can act as a modifier that can be attached to other proteins, or is itself modified. We further validated the existence of such protein modification by Western blot analysis with 3 different RBBP6 antibodies, which recognize distinct

epitopes of the DWNN domain (Figure S1). Consistent with this, we obtained similar results for RBBP6 iso3 specific antibodies, showing that iso3 is indeed itself modified.

We next wanted to investigate whether this iso3-involved protein modification is also lysine-dependent, just as ubiquitylation. We did a side-by-side comparison of the protein sequences of ubiquitin and the DWNN domain (Figure 2A). Lysine residues are highlighted in blue and labeled with amino acid numbers while the di-Glycine (di-Gly) residues are highlighted in pink. Mutation of lysine/di-Gly residues in ubiquitin has been widely used to abolish ubiquitination. The first ubiquitin molecule is covalently bound through its C-terminal di-Gly motif to lysine residues (sometimes cysteine, serine, threonine etc.) on protein substrates, which can be eliminated by mutating di-Gly residues. The second ubiquitin molecule in polyubiquitin chains are always bound to one of the seven lysine residues or the N-terminal methionine of the previous ubiquitin molecule. Therefore, mutating lysine residues on ubiquitin abolishes the formation of polyubiquitin chains (Hershko and Ciechanover, 1998; Varshavsky, 2012). For this reason, we mutated the lysine/di-Gly residues of the DWNN domain into arginine and alanine, respectively, to check the effect on protein modification. Surprisingly, even after mutating all the lysine residues and di-Gly residues in the DWNN domain (the last two lanes: K8-66 and diGly in Figure 2B), we still observed protein modification by Western blot.

RBBP6 iso3 actually has four additional lysine residues outside of the DWNN domain (Figure 2C). RBBP6 iso3 consists of the DWNN domain (76 amino acids), and a short C-terminal tail (32 amino acids). To examine the effect of the lysine residues in the C-terminal tail on protein modification, we next mutated these residues to arginine. Strikingly, only mutation of every single

lysine residue in iso3 can fully eliminate the high molecular weight protein modifications (last lane “D+C1+C2” in Figure D2). This result provided evidence that the protein modification observed in RBBP6 iso3 is lysine-dependent. Especially, this modification doesn’t appear to rely on any specific lysine of iso3, since one lysine mutation can always be compensated by other remaining lysine residues without affecting the pattern of the protein modification.

DWNN is not a ubiquitin-like modifier, but is itself ubiquitinated

We next considered two possibilities that can explain what we have observed so far: The observed protein modification is due to ubiquitylation of RBBP6 iso3, or to a novel modification mediated by the DWNN domain. Given the fact that the modification is lysine-dependent, it is more likely to be ubiquitylation of RBBP6 iso3 protein. To test this, we transfected HeLa cells with HA-his-tagged iso3/DWNN vector and FLAG-ubiquitin vector together and performed his-tag purification under denaturing conditions, in which only covalent protein-protein interaction could be detected. FLAG-Ub bound to both iso3 (Figure 3A) and the DWNN (Figure 3B) covalently. Consistently, overexpression of FLAG-Ub enhanced iso3 ubiquitylation significantly. It is noteworthy that the molecular weights of the bands detected by the FLAG antibody (i.e. ubiquitylated protein) were also consistent with those detected by HA antibody (i.e. modified iso3). Together, these results indicate that the DWNN domain is not a ubiquitin-like modifier, but is instead itself ubiquitylated.

Finally, to validate iso3 ubiquitylation, we determined whether ubiquitylated iso3 is regulated by proteasome-mediated degradation. Iso3 transfected HeLa cells were cultured in the

presence or absence of the proteasome inhibitors MG132, MG115 or lactacystin in a time-course manner (0hr, 6hr, 12hr, 24hr treatment, respectively). MG132 and MG115 can block the ubiquitin-dependent 26S proteasome reversibly (Lee and Goldberg, 1998), whereas lactacystin inhibits the proteolytic activity of 20S proteasome, which is ubiquitin-and ATP-independent (Fenteany and Schreiber, 1998). Interestingly, we found iso3 protein modification only responded to MG132 and MG115 treatment, but not lactacystin (Figure 3C), confirming that this lysine-dependent and MG132/MG115-inducible protein modification of RBBP6 iso3 is indeed ubiquitylation.

Given the fact that the long forms of RBBP6 are known to be a E3 ubiquitin-protein ligases, an intriguing possibility was that a long-form RBBP6 might function as the E3 ligase for iso3 ubiquitylation. As in Figure 1A, all RBBP6 homologs share three well-conserved domains at their N-terminus: the DWNN, the RING, and the Zinc knuckle domains. The RING domain is known to be a key domain to carry E3 ligase activity by binding to E2-ubiquitin thioester and activating discharge of ubiquitin cargo (Deshaies and Joazeiro, 2009). To test the possibility that RBBP6 functions as the iso3 E3 ligase, we constructed a 3xFLAG-tagged RBBP6 N-terminal derivative (RBBP6N), containing the key RING domain, Zinc knuckle, the DWNN domain, as well as a RBBP6N derivative that only contained the RING domain and the Zinc knuckle (Δ DWNN). We additionally made two mutant RBBP6N vectors (“RING mut” and “Zinc mut” vectors), in which the conserved cysteine residues of either the RING or the Zinc knuckle domains are mutated into alanine. The mutated cysteines are highlighted in red in Figure 4A. The FLAG-tagged RBBP6N plasmid and its mutant derivatives were co-transfected with HA-his-iso3 into HeLa cells for 48hrs. Overexpression of neither RBBP6N nor its mutants affected iso3 ubiquitination, even when

transfecting increasing amounts of the vectors into cells. Taken together, we concluded that RBBP6N is not the E3 ligase for iso3 ubiquitylation (Figure 4B).

Chromatin localization of RBBP6 iso3 is facilitated by monoubiquitylation and the C-terminal tail of iso3

We next wished to look into what role ubiquitylation plays in RBBP6 iso3 regulation. Ubiquitin is a small protein of about 8.6 kDa that can be linked to proteins as a monomer on single (monoubiquitylation) or multiple substrate lysine residues (multi-monoubiquitylation) or as a polymer (polyubiquitylation) by the sequential addition of ubiquitin to each other through ubiquitin lysines (reviewed by Sadowski et al., 2012) (Figure 3D). Iso3 appeared to be both monoubiquitylated (MW ~25 kDa) and multi-monoubiquitylated or poly-ubiquitylated (Figure 3C).

Our next question was whether and how the various ubiquitylation states help modulate the function of iso3 protein. Studies have found that monoubiquitylation regulates DNA repair, histone function, gene expression and intracellular localization (Bergink and Jentsch, 2009; Haglund et al., 2003b; Hicke, 2001; Passmore and Barford, 2004; Trotman et al., 2007). Multi-monoubiquitylation plays crucial roles in receptor endocytosis (Haglund et al., 2003b). Polyubiquitylation is widely-known to target protein substrates mainly for proteasomal degradation (Thrower et al., 2000).

Since monoubiquitylation of iso3 was consistently observed in our experiments, and because it can be difficult to distinguish multi-monoubiquitylation from polyubiquitylation, we

decided to focus on investigating whether monoubiquitylation affects the nuclear localization of the RBBP6 iso3 protein. As described above, ubiquitylation of iso3 can only be abolished by mutating every single lysine in the protein, and this mutant iso3 protein “D+C1+C2” was denoted as “iso3-All-K”. We first overexpressed iso3 and iso3-All-K proteins by transfecting HeLa cells with specific vectors, then separated cell compartments into cytoplasmic, nuclear-soluble, and chromatin fractions, analyzed different fractions by Western blotting, and quantified images using ImageJ. The effectiveness of the fractionation protocol was also verified by Western blotting: GAPDH, U2AF65, and histone H3 were detected predominantly in the cytoplasmic, nuclear-soluble, and chromatin fractions, respectively (Figure 5A). As shown in Figure 5A, both iso3 and iso3-All-K were predominantly enriched in the chromatin fractions. Yet, more iso3-All-K, as compared with wild-type iso3 protein, was observed in the cytoplasmic and nuclear-soluble fractions, indicating that chromatin localization of iso3 is facilitated by monoubiquitylation. Furthermore, multiple bands above 25 KDa (molecular weight of monoubiquitinated iso3) were also detected in the chromatin fractions. We suspected that those species were more likely to be multi-monoubiquitinated rather than polyubiquitinated, given their enrichment in the chromatin fraction and absence in the cytoplasmic fraction. In other words, polyubiquitinated substrates should be degraded in the cytoplasm, instead of being transported into the nucleus. We note that, although there are numerous studies showing that polyubiquitination can also serve for non-degradation purposes, all the non-proteolytic functions of polyubiquitin described appear to be carried out in the cytoplasm (Husnjak and Dikic, 2012; Li and Ye, 2008; Swatek and Komander, 2016). Therefore, it's possible that multi-monoubiquitylation of iso3 also facilitates its chromatin

localization.

Next, we determined the subcellular localization of the DWNN domain itself by mutating every lysine to arginine, i.e. “DWNN-All-K”. In contrast to the localization of iso3 and its All-K mutant, both the DWNN and the DWNN-All-K mutant were almost evenly distributed in the cytoplasmic and chromatin fractions (compare lane “cyto” with lane “chromatin” in Figure 5A and B). Since the only difference between iso3 and the DWNN is a short C-terminal tail, our data suggested that the C-terminal tail facilitates chromatin localization of iso3. Indeed, given that the RBBP6 iso3 is a small protein lacking a nuclear localization signal, and that iso3 competes with the long forms of RBBP6 for binding to the core 3' processing machinery, it is very possible that the C-terminal tail helps iso3 to localize to chromatin by interacting with other nuclear proteins.

To characterize the role of the C-terminal tail in iso3 nuclear localization, we performed immunofluorescence (IF) in HeLa cells transfected with HA-his-tagged iso3/DWNN using a HA antibody as well as RBBP6 Ab1, which recognizes an epitope of the DWNN domain (RBBP6 Ab1 shown in Figure S1) (Figure 5C). Consistent with our subcellular fractionation data, the iso3 protein was almost entirely enriched in the nucleus, whereas the DWNN protein was located in both cytoplasm and nucleus. Altogether these data suggest that the chromatin localization of iso3 is facilitated by monoubiquitylation (and possibly multi-monoubiquitylation) as well as the C-terminal tail of iso3.

Importin7 interacts with RBBP6 iso3 and facilitates its cytoplasmic retention

We next wanted to characterize proteins interacting with ubiquitylated iso3, such as its E3

ubiquitin ligase. To investigate this, we prepared extracts from HEK293 cells overexpressing HA-his-tagged iso3, performed his-tag pull down under stringent conditions (Figure 6A and B) followed by Mass Spectrometry (MS) analysis of mock and iso3 transfected cells (Figure 6C). Two independent purifications (i.e. Group1 and Group2, indicated in Figure 6A and B) identified 12 overlapping proteins interacting with iso3. As expected, the top 4 most enriched proteins identified in MS were all ubiquitin. The other 8 proteins were classified by their spectral count numbers into two groups. In the high spectral counts group, which reflects higher protein abundance in the purified samples, two proteins were notable. The first was E3 ubiquitin-protein ligase CBL, which might function as the E3 ligase for iso3. However, there is a natural 6xHis sequence in the protein sequence of CBL, which might be why it was present in the his-tag pulldown samples. The second protein, Importin 7 (Imp7), is an import factor that shuttles between nucleus and cytoplasm to transport several nucleic acids-binding proteins, such as ribosomal proteins (Jakel and Gorlich, 1998) and the HIV-1 integrase protein (Ao et al., 2007; Fassati et al., 2003).

We next performed Metascape analysis (<http://metascape.org/>) of all the proteins detected in the MS analysis to obtain a better idea of the enrichment networks of iso3 interacting proteins. As shown in Figure S2, the 78 proteins detected in Group1, and the 165 proteins in Group2, were bundled into biological process nodes, which implied the potential participation of iso3 protein in those pathways. It is noteworthy that the Western blot analysis of Group1 and Group2 pulldown samples displayed distinctive ubiquitylation statuses (Figure 6A). Indeed, Group2 showed much less monoubiquitylated and more multi-/poly-ubiquitylated iso3 compared to Group1 samples,

corresponding with more cell death in Group2 than Group1, as observed during sample collection and preparation. The differences of the enrichment networks between the two groups might imply the different functions of iso3 under mono-/multi-mono-/poly-ubiquitylation.

We next investigated whether Imp7 plays a role in the nuclear trafficking of iso3. We first verified the physical interaction between Imp7 and iso3 in the his-tag pulldown experiment. As indicated in Figure 7A, Imp7 indeed physically interacts with the iso3. We next determined whether depletion of Imp7 affects the localization of iso3. We depleted Imp7 by siRNA transfection, simultaneously overexpressed iso3 into HeLa cells, and performed subcellular fractionation as described earlier (Figure 7B). Imp7 is predominantly located in the cytoplasmic fraction, which is consistent with a previous report (Gorlich et al., 1997). Strikingly, we found that depletion of Imp7 led to decreased level of the iso3 protein in the cytoplasmic fraction. Quantification of three independent replicates is indicated in Figure7C. This data strongly indicates that Imp7 might facilitate the cytoplasmic transport of iso3 protein.

To extend our findings, we wished to investigate the possible physiological significance of the RBBP6-IMP7 interaction. To this end, we examined the cross-cancer alteration of *RBBP6* and *IPO7* genes using cBioPortal for Cancer Genomics (<http://www.cbioportal.org/>). Remarkably, *RBBP6* and *IPO7* alterations are found in the same top five most altered cancer types: prostate cancer, endometrial cancer, cutaneous melanoma, colorectal adenocarcinoma and bladder cancer (highlighted in red boxes) (Figure S3A). Significant co-occurrence of amplifications is associated with neuroendocrine prostate cancer, which has the highest frequency of alteration for *RBBP6* and *IPO7* (Trento/Cornell/Broad 2016) (Figure S3B). mRNA expression level of *RBBP6* and *IPO7*

was also found to be significantly positively correlated (Pearson correlation = 0.62 and Spearman correlation = 0.75) (Figure S3C). A number of studies have observed a decreased level of iso3 and an increased level of long-form RBBP6 in cancer (Chen et al., 2013; Moela and Motadi, 2016; Motadi et al., 2011; Motadi et al., 2018; Yoshitake et al., 2004). Consistent with this, our results suggest that upregulation of Imp7 in cancer would retain more iso3 in the cytoplasm, so that less iso3 would compete with long-form RBBP6 in the nucleus.

Discussion

In this study, we investigated whether the RBBP6 iso3 protein is a ubiquitin-like modifier, and how its subcellular localization is regulated in the cell. Our results showed that the DWNN, previously known as a ubiquitin-like domain (Pugh et al., 2006), doesn't exert ubiquitin-like modification function. We however showed that RBBP6 iso3 is itself ubiquitinated. Next, we demonstrated that the nuclear localization of iso3 is facilitated by its monoubiquitylation as well as the C-terminal tail. In addition, our MS analysis confirmed the ubiquitylation of iso3, and further identified a novel interaction between Imp7 and iso3. Finally, we showed that Imp7 can facilitate the cytoplasmic retention of iso3. Below we discuss how the dynamic regulation of iso3 subcellular localization contributes to the functions of RBBP6 in a variety of biological processes, with an emphasis in mRNA processing and cancer.

Previous studies have found that the DWNN domain adopts a ubiquitin-like structure, with characteristic ubiquitin di-Glycine and lysine residues (Pugh et al., 2006). Ubiquitin is a 76-residues protein with seven lysine residues and di-Glycine residues at the C-terminus. The DWNN

domain contains 81 residues with eight lysine residues and C-terminal di-Glycine motif. It is expressed as a single domain protein with an additional C-terminal tail in human, also known as RBBP6 isoform3. The similarities between ubiquitin and DWNN raised the intriguing possibility that the RBBP6 iso3 protein might be able to modify protein substrates, as a novel ubiquitin-like modifier. Indeed, we observed multiple high molecular weight bands above the RBBP6 iso3 protein in Western blot analysis of cell lysate overexpressing tagged iso3, and such modification disappeared when all lysine residues were mutated. Yet, we showed evidence indicating that the observed lysine-dependent modification in RBBP6 iso3 is most likely ubiquitylation

Although we have shown that iso3 is extensively ubiquitinated, we cannot exclude the possibility that iso3 can modify ubiquitin and even form “iso3 chain” on ubiquitin. Indeed ubiquitin-like molecules such as SUMO and NEDD8 are able to modify ubiquitin to generate NEDDylated ubiquitin and SUMOylated ubiquitin, respectively (Galisson et al., 2011; Hendriks et al., 2014; Lamoliatte et al., 2013; Singh et al., 2012). Nevertheless, this is made unlikely by comparing the molecular weight of the iso3 modification bands in Western blot images: the iso3 protein is around 18 KDa, whereas its modification bands are approximately 25 KDa, 37 KDa, 50 KDa, and above 75 KDa. The molecular weight difference between modification bands is around 8 KDa, which matches the molecular weight of ubiquitin, 8.6KDa. If iso3 could modify ubiquitin and form “iso3 chain” on ubiquitin, bands with much higher molecular weight should be observed in Western blot analysis. Also, the di-glycine motif, which is essential for attaching ubiquitin-like proteins to substrates, can be substituted with alanine residues without affecting the modification of iso3. Taken together, we conclude that iso3 is not a ubiquitin-like modifier but is itself

ubiquitylated. Additionally, although long-form RBBP6 is known to be an E3 ligase for various proteins, we found that overexpression of recombinant RBBP6 N-terminal derivative containing the DWNN, Zinc knuckle, and RING domains, had no impact on iso3 ubiquitination. We cannot exclude the possibility that full-length RBBP6 might facilitate iso3 ubiquitination. Nevertheless, we were unfortunately unable to construct a full-length RBBP6 plasmid to test whether the protein is the E3 ligase for iso3 ubiquitination. To date, the E3 ubiquitin ligase responsible for iso3 ubiquitination remains unknown.

Predominant chromatin localization of iso3 is consistent with our previous finding that iso3 can bind to the 3' pre-mRNA processing core machinery (Di Giammartino et al., 2014). Intriguingly, we also observed enrichment of monoubiquitylated iso3 at chromatin, and a higher level of all-lysine mutated iso3 in the cytoplasm, as compared with wild-type iso3. Those results indicate that monoubiquitylation might facilitate the chromatin localization of iso3 protein. Monoubiquitylation has been widely implicated as a signal for intracellular trafficking (reviewed by Komander and Rape, 2012; Sadowski et al., 2012; Sigismund et al., 2004). For example, several receptors such as epidermal growth factor receptor and receptor tyrosine kinase that undergo endocytosis require monoubiquitylation (reviewed by Haglund et al., 2003a; Hicke, 2001). Since iso3 is not a receptor, monoubiquitylation might affect iso3 localization in an indirect way, possibly by affecting the accessibility of iso3 protein to its binding partners, rather than being an export signal itself. For example, monoubiquitylation of the transcription factor p53 will lead to its cytoplasmic accumulation, possibly by altering the accessibility of p53's nuclear export sequence to the export machinery (Carter et al., 2007; Li et al., 2003). In our analysis of iso3 subcellular

fractionation, we also noticed two modified iso3 species (approximately 37 KDa and 50 KDa, respectively) larger than the monoubiquitinated iso3 species (~ 25 KDa), which can be either polyubiquitylated or multi-monoubiquitylated iso3. Given their enrichment in the chromatin fraction, we suspect these two isoforms correspond to multi-monoubiquitylated iso3 proteins and facilitate iso3 chromatin localization. However, further experiments are needed to prove the existence of such modifications in iso3.

We also discovered a novel interaction between Importin 7 (Imp7) and iso3, which further extended our understanding about the dynamic subcellular localization of iso3. Imp7, encoded by the gene *IPO7*, is a nucleocytoplasmic transport protein related to the importin- β family, which can cross the nuclear envelope rapidly in both directions (Gorlich et al., 1997). Studies have shown that imp7 acts as an import factor for various nucleic acid-binding proteins, such as ribosomal proteins (Jakel and Gorlich, 1998), the glucocorticoid receptor, the HIV-1 integrase protein (Ao et al., 2007; Fassati et al., 2003), as well as histone H1, by forming a heterodimer with importin- β (Jakel et al., 1999). Additionally, Imp7 can also mediate the nuclear trafficking of both endogenous and exogenous DNA in mammalian cells, such as mitochondrial DNA and adenovirus DNA (Dhanoya et al., 2013; Trotman et al., 2001). Besides acting as an import factor, Imp7 can also facilitate cytoplasmic retention. For example, Imp7 can bind to the androgen receptor and block its nuclear localization signal, thus causing its cytoplasmic retention (Ni et al., 2013). It is very likely that Imp7 facilitates the cytoplasmic retention of iso3 by blocking the region responsible for its nuclear localization.

Previous studies of RBBP6 suggested that its function in various cellular processes is

modulated by the ratio between the long and short isoforms of RBBP6 (i.e. long-form RBBP6 and iso3). For example, long-form RBBP6 has been discovered to be overexpressed in a number of tumors (Chen et al., 2013; Moela and Motadi, 2016; Motadi et al., 2011; Motadi et al., 2018; Yoshitake et al., 2004) while iso3 has been found to be down-regulated in cancers (Mbita et al., 2012). Besides cell proliferation, the ratio change between RBBP6 isoforms also affects 3' pre-mRNA processing efficiency. We showed previously that iso3 can compete with RBBP6 long-form for the binding to the core factors of the 3' pre-mRNA processing machinery, and therefore play an inhibitory role in 3' processing. In other words, pre-mRNA 3' processing efficiency of RBBP6-sensitive transcripts can be regulated by changing the ratio between active long-form RBBP6 and inhibitory iso3 (Di Giammartino et al., 2014). Here, we propose a model that explains the dynamic subcellular localization of RBBP6 iso3 protein (Figure 7D). Monoubiquitylation and the C-terminal tail of iso3 facilitate the nuclear localization of iso3, whereas Imp7 participates in the cytoplasmic retention of iso3, possibly by blocking its interaction with other nuclear proteins. By shuttling between cytoplasm and nucleus, iso3 can either participate in cell proliferation, mRNA processing, differentiation, or be degraded. Additionally, we observed a striking correlation between *RBBP6* and *IPO7* in cancer types as well as in their mRNA expression levels, by analyzing cancer genomic data. This is consistent with previous findings as well as our model: On the one hand, long-form RBBP6 is overexpressed in cancer, whereas the level of iso3 is decreased. On the other hand, upregulation of Imp7 would retain more iso3 in the cytoplasm so that less iso3 would compete with long-form RBBP6 in the nucleus.

In conclusion, we have shown that RBBP6 iso3 is not a ubiquitin-like modifier but is a

ubiquitylated protein. The monoubiquitylation of iso3 and its C-terminal tail facilitate its chromatin localization, whereas the import factor Imp7 participates in its cytoplasmic retention. Our study highlighted a novel mechanism by which iso3 subcellular localization can modulate the ratio between long-form RBBP6 and iso3, and possibly regulate mRNA 3' end processing.

Material & Methods

Cell culture and siRNA transfections

HeLa and HEK293 cells were cultured in DMEM supplemented with 10% FBS. The siRNAs against IPO7 were transfected using DharmaFECT1 (GE Dharmacon) (#1, GAUGGAGCCCUGCAUAUGA; #2, UGAUGACCUUACCAAUGUA) at 20nM and maintained for 72 hrs. To obtain high knockdown efficiency after 72hr treatment, the siRNA transfection was repeated 48hrs after the first transfection with half the amount of siRNA and cells were harvested 24hrs after the second transfection.

Antibodies

RBBP6 Ab1, RBBP6 Ab2, RBBP6 Ab3 antibodies were made by courtesy of Bethyl Laboratories. RBBP6 C-terminus were from Santa Cruz (sc-6359). GAPDH (G8795; Sigma), U2AF65 (U4758; Sigma), histone H3 (ab1791; abcam), HA (901501; Biolegend), FLAG (F1804; Sigma) were also used in this study.

Subcellular fractionation

Seventy-two hours after siRNA/plasmid transfection, HeLa cells grown in a 10-cm dish were washed twice with PBS and collected by scrapping and centrifugation. Cell pellets were resuspended in 400 μ L of swelling buffer (10 mM Tris-HCl at pH 8.0, 1.5 mM MgCl₂, 10 mM KCl, 5 U of RNasin, 1 \times protease inhibitor cocktail) and incubated for 15 min on ice. Cells were homogenized by passing a 26-gauge needle attached to a 1 mL syringe until >90% of cells were disrupted (typically 10~20 strokes). Half of the lysate was kept in a new tube and used as whole-cell lysate. The rest of the 200 μ L of lysate was mixed with 2 μ L of 10% NP-40, gently tapped, and immediately centrifuged in an Eppendorf centrifuge 5424 at 6,000 rpm for 5 min. The supernatant was kept in a new tube and used as cytoplasmic fraction. The pellet was washed once with swelling buffer, resuspended in 100 μ L of glycerol buffer (20 mM Tris-HCl at pH 8.0, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT, 5 U of RNasin, 1 \times protease inhibitor cocktail) by pipetting, and then mixed with 100 μ L of nucleus lysis buffer (20 mM HEPES-NaOH at pH 7.6, 7.5 mM MgCl₂, 0.2 mM EDTA, 300 mM NaCl, 1 M urea, 1% NP-40, 1 mM DTT, 5 U of RNasin, 1 \times protease inhibitor cocktail). The mixture was pulse-vortexed three times, incubated for 1 min on ice, and then centrifuged in an Eppendorf centrifuge 5424 at 14,000 rpm for 2 min. The supernatant was used as nuclear-soluble fraction. The pellet was washed once with a 1:1 mixture of glycerol/nucleus lysis buffer and then resuspended in 200 μ L of water.

His-Tag Pulldown

HeLa or HEK293 cells grown in 10-cm dishes were washed twice with PBS and lysed in lysis buffer (6 M guanidinium-HCl, 0.1 M Na₂HPO₄/NaH₂PO₄, 10 mM Tris-HCl (pH 8), 0.005

M imidazole, 5mM dithiothreitol, 1× protease inhibitor cocktail) for 5 min on ice. The lysates were sonicated and then centrifuged in an Eppendorf centrifuge 5424 at 6000 rpm for 15 min at 4°C, and supernatants were rotated with NTA agarose beads (Sigma) for 1hr. The beads were then washed three times with washing buffer (8 M urea, 0.1 M Na₂HPO₄/NaH₂PO₄, 10 mM Tris-HCl (pH 6.8), 0.005 M imidazole, 5mM dithiothreitol), and eluted by the elution buffer (0.5 M imidazole, 0.15 M Tris-HCl (pH 6.8), 30% glycerol, 5mM dithiothreitol, 5% SDS) for Western blot analysis or Mass spectrometry analysis.

Mass Spectrometry Analysis

Protein samples were digested for 18 h at 37°C in 2 M urea, 100 mM Tris (pH 8.5), and 1 mM CaCl₂ with 2µg of trypsin (Promega). Multidimensional protein identification technology (MudPIT) analysis was performed using an Eksigent nanoLC pump and a Thermo LTQ-Orbitrap using an in-house built electrospray stage (Wolters et al. 2001). Protein and peptide identification and protein quantitation were done with Integrated Proteomics Pipeline (IP2; Integrated Proteomics Applications, Inc., <http://www.integratedproteomics.com>). Tandem mass spectra were extracted from raw files using RawConverter (He et al. 2015) and were searched against a UniProt human database with reversed sequences using ProLuCID (Peng et al. 2003; Xu et al. 2015). The search space included all fully tryptic and half-tryptic peptide candidates. Peptide candidates were filtered using DTASelect with the following parameters: -p 2 -y 1 --trypstat --extra --pI -DM 10 -DB --dm -in --brief --quiet (Tabb et al. 2002).

Immunofluorescence Staining

Cells were washed twice with PBS and fixed with 4% paraformaldehyde for 10 min. Following fixation, cells were permeabilized with 1% Triton X-100 in PBS for 10 min, followed by 30 min incubation with 3% BSA in PBS. Cells were next incubated with rabbit anti-RBBP6 Ab1 (Bethyl Lab) and mouse anti-HA (901501; Biolegend) overnight at 4°C. Both antibodies were used at a dilution of 1:500. After three washes with PBS, cells were incubated with an appropriate secondary antibody (goat anti-rabbit Alexa Fluor 488 or rabbit anti- mouse Alexa Fluor 564) for 2hrs at room temperature. Nuclei were counterstained with DAPI. Images were taken with a Zeiss LSM700 confocal microscope.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We are grateful to Bethyl lab for making RBBP6 antibodies. We also thank Dr. Patricia Richard (Columbia University) for comments and editorial assistance. This work was supported by NIH grant R35GM118136 to JLM.

Supplementary information is available at

References

- Ao, Z., Huang, G., Yao, H., Xu, Z., Labine, M., Cochrane, A.W., and Yao, X. (2007). Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. *J Biol Chem* 282, 13456-13467.
- Batra, J., Hultquist, J.F., Liu, D., Shtanko, O., Von Dollen, J., Satkamp, L., Jang, G.M., Luthra, P., Schwarz, T.M., Small, G.I., *et al.* (2018). Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication. *Cell* 175, 1917-1930 e1913.
- Bergink, S., and Jentsch, S. (2009). Principles of ubiquitin and SUMO modifications in DNA repair. *Nature* 458, 461-467.
- Carter, S., Bischof, O., Dejean, A., and Vousden, K.H. (2007). C-terminal modifications regulate MDM2 dissociation and nuclear export of p53. *Nature Cell Biology* 9, 428-U111.
- Chen, J., Tang, H., Wu, Z., Zhou, C., Jiang, T., Xue, Y., Huang, G., Yan, D., and Peng, Z. (2013). Overexpression of RBBP6, alone or combined with mutant TP53, is predictive of poor prognosis in colon cancer. *PLoS One* 8, e66524.

Chibi, M., Meyer, M., Skepu, A., DJ, G.R., Moolman-Smook, J.C., and Pugh, D.J. (2008). RBBP6 interacts with multifunctional protein YB-1 through its RING finger domain, leading to ubiquitination and proteosomal degradation of YB-1. *J Mol Biol* 384, 908-916.

Deshaies, R.J., and Joazeiro, C.A.P. (2009). RING Domain E3 Ubiquitin Ligases. *Annu Rev Biochem* 78, 399-434.

Dhanoya, A., Wang, T., Keshavarz-Moore, E., Fassati, A., and Chain, B.M. (2013). Importin-7 mediates nuclear trafficking of DNA in mammalian cells. *Traffic* 14, 165-175.

Di Giammartino, D.C., Li, W., Ogami, K., Yashinski, J.J., Hoque, M., Tian, B., and Manley, J.L. (2014). RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs. *Genes Dev* 28, 2248-2260.

Fassati, A., Gorlich, D., Harrison, I., Zaytseva, L., and Mingot, J.M. (2003). Nuclear import of HIV-1 intracellular reverse transcription complexes is mediated by importin 7. *EMBO J* 22, 3675-3685.

Fenteany, G., and Schreiber, S.L. (1998). Lactacystin, proteasome function, and cell fate. *J Biol Chem* 273, 8545-8548.

Galisson, F., Mahrouche, L., Courcelles, M., Bonneil, E., Meloche, S., Chelbi-Alix, M.K., and Thibault, P. (2011). A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol Cell Proteomics* 10, M110 004796.

Gao, S., and Scott, R.E. (2003). Stable overexpression of specific segments of the P2P-R protein in human MCF-7 cells promotes camptothecin-induced apoptosis. *J Cell Physiol* 197, 445-452.

Gorlich, D., Dabrowski, M., Bischoff, F.R., Kutay, U., Bork, P., Hartmann, E., Prehn, S., and Izaurralde, E. (1997). A novel class of RanGTP binding proteins. *J Cell Biol* 138, 65-80.

Haglund, K., Di Fiore, P.P., and Dikic, I. (2003a). Distinct monoubiquitin signals in receptor endocytosis. *Trends in Biochemical Sciences* 28, 598-603.

Haglund, K., Sigismund, S., Polo, S., Szymkiewicz, I., Di Fiore, P.P., and Dikic, I. (2003b).

Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. *Nat Cell Biol* 5, 461-466.

Harutyunyan, A.S., Giambruno, R., Krendl, C., Stukalov, A., Klampfl, T., Berg, T., Chen, D., Milosevic Feenstra, J.D., Jager, R., Gisslinger, B., *et al.* (2016). Germline RBBP6 mutations in familial myeloproliferative neoplasms. *Blood* 127, 362-365.

Hendriks, I.A., D'Souza, R.C., Yang, B., Verlaan-de Vries, M., Mann, M., and Vertegaal, A.C. (2014). Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat Struct Mol Biol* 21, 927-936.

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. *Annu Rev Biochem* 67, 425-479.

Hicke, L. (2001). Protein regulation by monoubiquitin. *Nat Rev Mol Cell Biol* 2, 195-201.

Husnjak, K., and Dikic, I. (2012). Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu Rev Biochem* 81, 291-322.

Jakel, S., Albig, W., Kutay, U., Bischoff, F.R., Schwamborn, K., Doenecke, D., and Gorlich, D. (1999). The importin beta/importin 7 heterodimer is a functional nuclear import receptor for histone H1. *EMBO J* 18, 2411-2423.

Jakel, S., and Gorlich, D. (1998). Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO J* 17, 4491-4502.

Komander, D., and Rape, M. (2012). The Ubiquitin Code. *Annu Rev Biochem* 81, 203-229.

Lamoliatte, F., Bonneil, E., Durette, C., Caron-Lizotte, O., Wildemann, D., Zerweck, J., Wenshuk, H., and Thibault, P. (2013). Targeted identification of SUMOylation sites in human proteins using affinity enrichment and paralog-specific reporter ions. *Mol Cell Proteomics* 12, 2536-2550.

Lee, D.H., and Goldberg, A.L. (1998). Proteasome inhibitors: valuable new tools for cell biologists. *Trends Cell Biol* 8, 397-403.

Lee, S.D., and Moore, C.L. (2014). Efficient mRNA Polyadenylation Requires a Ubiquitin-Like

Domain, a Zinc Knuckle, and a RING Finger Domain, All Contained in the Mpel Protein. *Molecular and Cellular Biology* 34, 3955-3967.

Li, L., Deng, B.W., Xing, G.C., Teng, Y., Tian, C.Y., Cheng, X., Yin, X.S., Yang, J.T., Gao, X., Zhu, Y.P., *et al.* (2007). PACT is a negative regulator of p53 and essential for cell growth and embryonic development. *P Natl Acad Sci USA* 104, 7951-7956.

Li, M.Y., Brooks, C.L., Wu-Baer, F., Chen, D.L., Baer, R., and Gu, W. (2003). Mono-versus polyubiquitination: Differential control of p53 fate by Mdm2. *Science* 302, 1972-1975.

Li, W., and Ye, Y. (2008). Polyubiquitin chains: functions, structures, and mechanisms. *Cell Mol Life Sci* 65, 2397-2406.

Mbita, Z., Meyer, M., Skepu, A., Hosie, M., Rees, J., and Dlamini, Z. (2012). De-regulation of the RBBP6 isoform 3/DWNN in human cancers. *Mol Cell Biochem* 362, 249-262.

Miotto, B., Chibi, M., Xie, P., Koundrioukoff, S., Moolman-Smook, H., Pugh, D., Debatisse, M., He, F., Zhang, L., and Defosse, P.A. (2014). The RBBP6/ZBTB38/MCM10 axis regulates DNA replication and common fragile site stability. *Cell Rep* 7, 575-587.

Moela, P., and Motadi, L.R. (2016). RBBP6: a potential biomarker of apoptosis induction in human cervical cancer cell lines. *Onco Targets Ther* 9, 4721-4735.

Motadi, L.R., Bhoola, K.D., and Dlamini, Z. (2011). Expression and function of retinoblastoma binding protein 6 (RBBP6) in human lung cancer. *Immunobiology* 216, 1065-1073.

Motadi, L.R., Lekganyane, M.M., and Moela, P. (2018). RBBP6 expressional effects on cell proliferation and apoptosis in breast cancer cell lines with distinct p53 statuses. *Cancer Manag Res* 10, 3357-3369.

Ni, L., Llewellyn, R., Kesler, C.T., Kelley, J.B., Spencer, A., Snow, C.J., Shank, L., and Paschal, B.M. (2013). Androgen induces a switch from cytoplasmic retention to nuclear import of the androgen receptor. *Mol Cell Biol* 33, 4766-4778.

Passmore, L.A., and Barford, D. (2004). Getting into position: the catalytic mechanisms of protein

ubiquitylation. *Biochem J* 379, 513-525.

Pugh, D.J., Ab, E., Faro, A., Lulya, P.T., Hoffmann, E., and Rees, D.J. (2006). DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways. *BMC Struct Biol* 6, 1.

Sadowski, M., Suryadinata, R., Tan, A.R., Roesley, S.N.A., and Sarcevic, B. (2012). Protein monoubiquitination and polyubiquitination generate structural diversity to control distinct biological processes. *Iubmb Life* 64, 136-142.

Saijo, M., Sakai, Y., Kishino, T., Niikawa, N., Matsuura, Y., Morino, K., Tama, K., and Taya, Y. (1995). Molecular-Cloning of a Human Protein That Binds to the Retinoblastoma Protein and Chromosomal Mapping. *Genomics* 27, 511-519.

Sakai, Y., Saijo, M., Coelho, K., Kishino, T., Niikawa, N., and Taya, Y. (1995). cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product. *Genomics* 30, 98-101.

Sigismund, S., Polo, S., and Di Fiore, P.P. (2004). Signaling through monoubiquitination. *Curr Top Microbiol* 286, 149-185.

Simons, A., MelamedBessudo, C., Wolkowicz, R., Sperling, J., Sperling, R., Eisenbach, L., and Rotter, V. (1997). PACT: Cloning and characterization of a cellular p53 binding protein that interacts with Rb. *Oncogene* 14, 145-155.

Singh, R.K., Zerath, S., Kleifeld, O., Scheffner, M., Glickman, M.H., and Fushman, D. (2012). Recognition and cleavage of related to ubiquitin 1 (Rub1) and Rub1-ubiquitin chains by components of the ubiquitin-proteasome system. *Mol Cell Proteomics* 11, 1595-1611.

Swatek, K.N., and Komander, D. (2016). Ubiquitin modifications. *Cell Res* 26, 399-422.

Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. *Embo Journal* 19, 94-102.

Trotman, L.C., Mosberger, N., Fornerod, M., Stidwill, R.P., and Greber, U.F. (2001). Import of

adenovirus DNA involves the nuclear pore complex receptor CAN/Nup214 and histone H1. *Nat Cell Biol* 3, 1092-1100.

Trotman, L.C., Wang, X., Alimonti, A., Chen, Z., Teruya-Feldstein, J., Yang, H., Pavletich, N.P., Carver, B.S., Cordon-Cardo, C., Erdjument-Bromage, H., *et al.* (2007). Ubiquitination regulates PTEN nuclear import and tumor suppression. *Cell* 128, 141-156.

Varshavsky, A. (2012). The ubiquitin system, an immense realm. *Annu Rev Biochem* 81, 167-176.

Vo, L.T.A., Minet, M., Schmitter, J.M., Lacroute, F., and Wyers, F. (2001). Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. *Molecular and Cellular Biology* 21, 8346-8356.

Yoshitake, Y., Nakatsura, T., Monji, M., Senju, S., Matsuyoshi, H., Tsukamoto, H., Hosaka, S., Komori, H., Fukuma, D., Ikuta, Y., *et al.* (2004). Proliferation potential-related protein, an ideal esophageal cancer antigen for immunotherapy, identified using complementary DNA microarray analysis. *Clinical Cancer Research* 10, 6437-6448.

Figure legends

Figure 1. RBBP6 iso3 is extensively modified. (A) Schematic of RBBP6 isoforms identified in different species. RBBP6 isoform3 was only identified in human, while the longer isoforms were characterized in various species, containing not only the DWNN domain, but also the Zinc knuckle and RING domains. (B) Western blot analysis of RBBP6 iso3 protein modification. HA-his-iso3 in pCMV plasmid was overexpressed in HeLa cells in a time course of transfection from 0hr to 48hr. (C) Validation of RBBP6 iso3 protein modification with RBBP6 specific antibodies. The epitopes of each antibody were shown in Fig. S1.

Figure 2. RBBP6 isoform3 protein is modified in a lysine-dependent manner. (A) Comparison of the amino acid sequences between ubiquitin and DWNN domain. Lysine residues were highlighted in blue and labeled with amino acid numbers, while the di-Glycine residues in pink. (B) Western blot analysis of RBBP6 iso3 mutants. Lysine8 and 11 were mutated to arginine in “K8/K11” mutant. Lysine 30, 31, 38, 40 were mutated to arginine in “K30-40” mutant. Every Lysine (from K8 to K66) in the DWNN domain of RBBP6 iso3 was mutated to arginine in “K8-66” mutant. Di-Glycine of iso3 were mutated to Alanine in “diGly” mutant. The upper panel for long exposure, lower panel for short exposure. (C) Schematic of the DWNN domain and the C-terminal tail in RBBP6 iso3 protein. The first 76 amino acids of the iso3 protein are the DWNN domain, whereas the last 43 amino acids are the C-terminal tail of iso3. The abbreviations of various mutations in iso3 plasmid are listed on the right side: we abbreviated the first two lysine mutations in the C-terminal tail as “C1”, while the last two lysine mutations as “C2”. Also,

mutation of all lysine residues (K8-66) in the DWNN domain was indicated as “D”. (D) Abolishment of protein modification by mutating every single lysine of RBBP6 iso3. HA-iso3 vector with four lysine residues mutated in the C-terminal tail, was indicated as “C1+C2”, whereas HA-iso3 with every single lysine mutated was denoted as “D+C1+C2”. HA-iso3 vectors carrying various combinations of lysine mutations in the C-terminal tail were transfected into HeLa cells, and protein extracts were analyzed by Western blot after 48hr of transfection.

Figure 3. The protein modification of RBBP6 isoform3 is ubiquitylation. (A) His-tag pulldown of HA-his-iso3 under denaturing conditions. Cell extracts prepared from HeLa cells expressing HA-his-iso3 or/and FLAG-Ubiquitin were used for the denaturing his-tag pulldown experiments. (B) His-tag pulldown of HA-his-DWNN under denaturing conditions. Cell extracts prepared from HeLa cells expressing HA-his-DWNN or/and FLAG-Ubiquitin were used for denaturing his-tag pulldown experiment. (C) MG132, MG115 and lactacystin treatment of iso3 overexpressed HeLa cells. HeLa cells were first transfected with HA-his-iso3 vectors for 24hrs, and then treated with DMSO, 10uM MG132, 10uM MG115, and 10uM lactacystin, respectively. Four time points of drug treatment were taken in the experiment: 0hr, 6hr, 12hr, and 24hr. The overall plasmid transfection time across all the samples is 48hrs. Cell lysates were analyzed by Western blotting. The signal intensity of protein modification was quantified by ImageJ, and normalized to GAPDH. (E) Illustration of poly-/multi-mono-/mono-ubiquitylation on protein substrates.

Figure 4. Overexpression of the E3 ligase RBBP6N does not affect iso3 ubiquitylation. (A)

Western blot analysis of His-tag pulldown of HeLa cells co-expressing HA-his-iso3 and RBBP6N-FLAG/ Δ DWNN-FLAG under denaturing conditions. (B) Western blot analysis of HeLa cell extracts co-expressing of RBBP6N-FLAG and its mutants (RING mut and Zinc mut) together with HA-his-iso3. Four Cystines in the RING domain of RBBP6N were mutated to alanine in the “RING mut”. 2 Cystines in the ZINC domain of RBBP6N were mutated to alanine in the “Zinc mut”.

Figure 5. Chromatin localization of RBBP6 iso3 is facilitated by monoubiquitylation and the C-terminal tail of iso3. (A) Subcellular fractionation of wild-type iso3 and its all Lysine mutant D+C1+C2 (i.e. “iso3-All-K”). Western blot analysis of iso3 subcellular localization was shown in the left panel. The quantification by ImageJ was shown in the right panel. (B) Subcellular fractionation of wild-type DWNN and its all Lysine mutant (i.e. “DWNN-All-K”). (C) Immunofluorescence of HeLa cells expressing HA-his-iso3/-DWNN. HeLa cells were transfected with iso3/DWNN plasmids for 48hrs, then fixed and immunostained with antibodies against HA (green) and RBBP6 Ab1 (red). Nuclei were counterstained with DAPI (blue). The fluorescence intensity in the nucleus and cytoplasm was quantified by ImageJ, and compared using an unpaired Student’s *t*-test ($p < 0.05$).

Figure 6. Mass Spectrometry analysis of his-tag pull down of HA-his-iso3 under denaturing conditions. (A) Western blot analysis of denaturing his-tag samples of HA-his-iso3 overexpressed HeLa cells. Two biological replicates were shown as Group1 and Group2, respectively. (B) Silver

staining of Group1 and Group2 input and pulldown samples. (C) Venn diagram of proteins detected in Group1 and Group2 via Mass Spectrometry analysis. (D) List of accession numbers and spectral counts of the 12 proteins detected in both groups. The top 4 proteins were all ubiquitin protein fragments. The next 4 proteins were classified as “high spectral counts”, while the last 4 proteins “low spectral counts” were based on the number of their spectral counts.

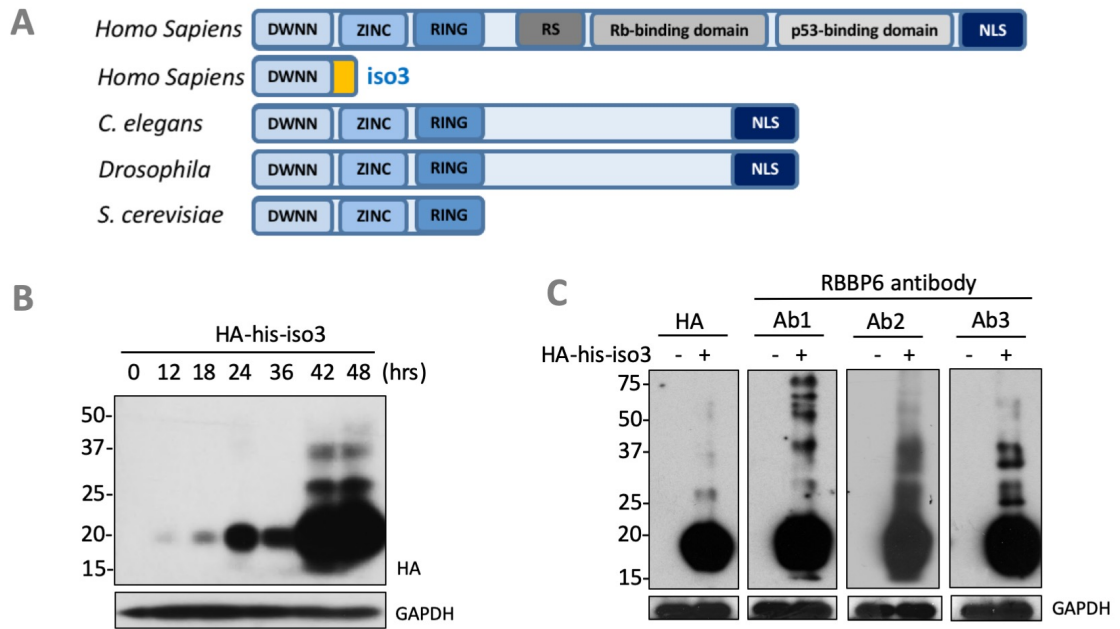
Figure 7. Importin 7 physically interacts with RBBP6 iso3 and facilitates its cytoplasmic retention. (A) Western blot analysis of the his-tag pulldown of HeLa cells overexpressing HA-his-iso3 using HA antibody and Imp7 antibody. (B) Subcellular fractionation of iso3 under control siRNA and siImp7 KD conditions. (C) ImageJ quantification of the subcellular fractionation western blotting results. Values were normalized to the Whole cell signals and set as percentage. Bar represent mean \pm SD. n=3. Asterisk denotes significant difference of the cytoplasmic signals between siCtrl and siImp7 using a paired Student’s t-test. (D) Schematic illustration of dynamic nuclear localization of RBBP6 iso3. Imp7 facilitates the cytoplasmic retention of iso3, while monoubiquitylation and iso3 C-terminal tail enhance its nuclear localization.

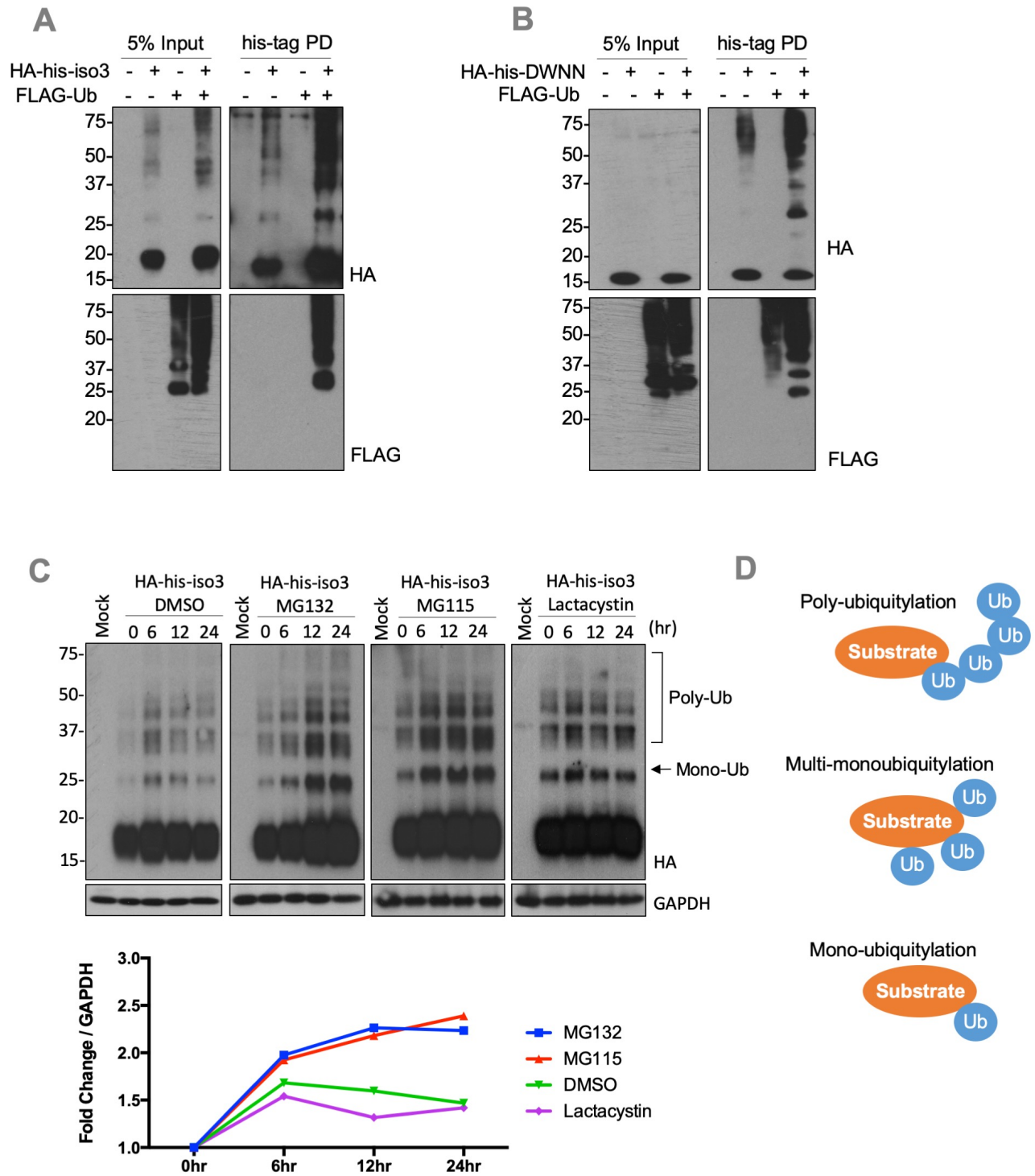
Supplementary Figure legends

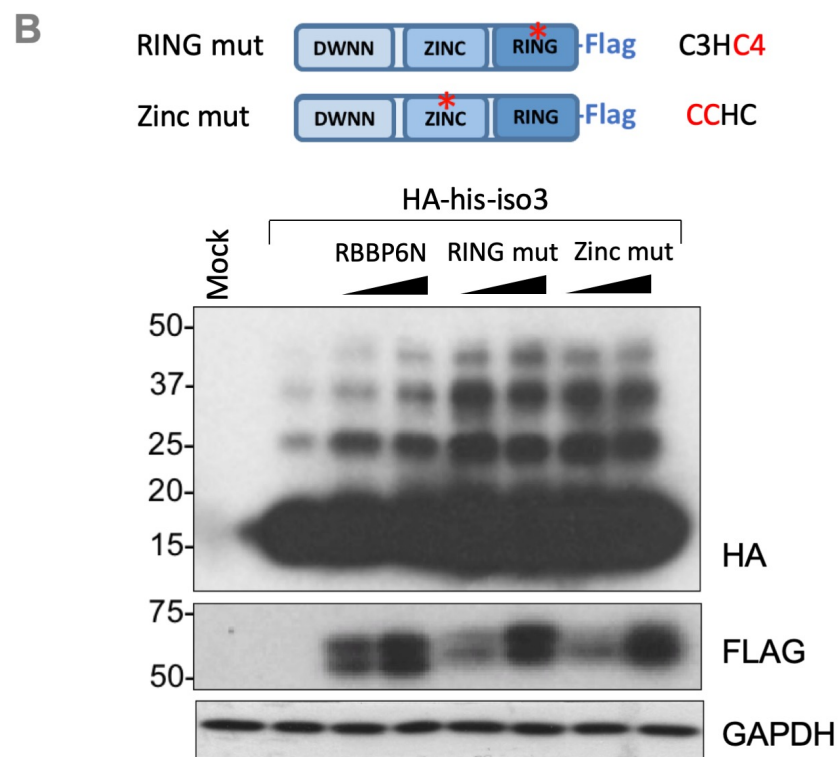
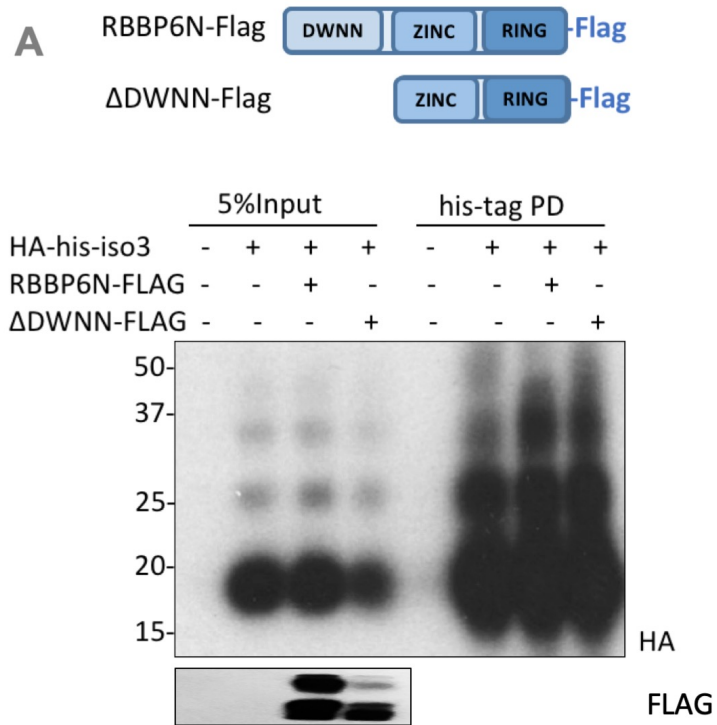
Figure S1. The epitopes of RBBP6 antibodies in iso3 and RBBP6N proteins. The protein sequences of RBBP6 iso3 (upper) and RBBP6N (lower) were listed. The epitope of RBBP6 Ab1, Ab2, Ab3 antibodies are highlighted in orange, green, and red, respectively. *Only partial amino acid sequence on the N-terminus is shown here.

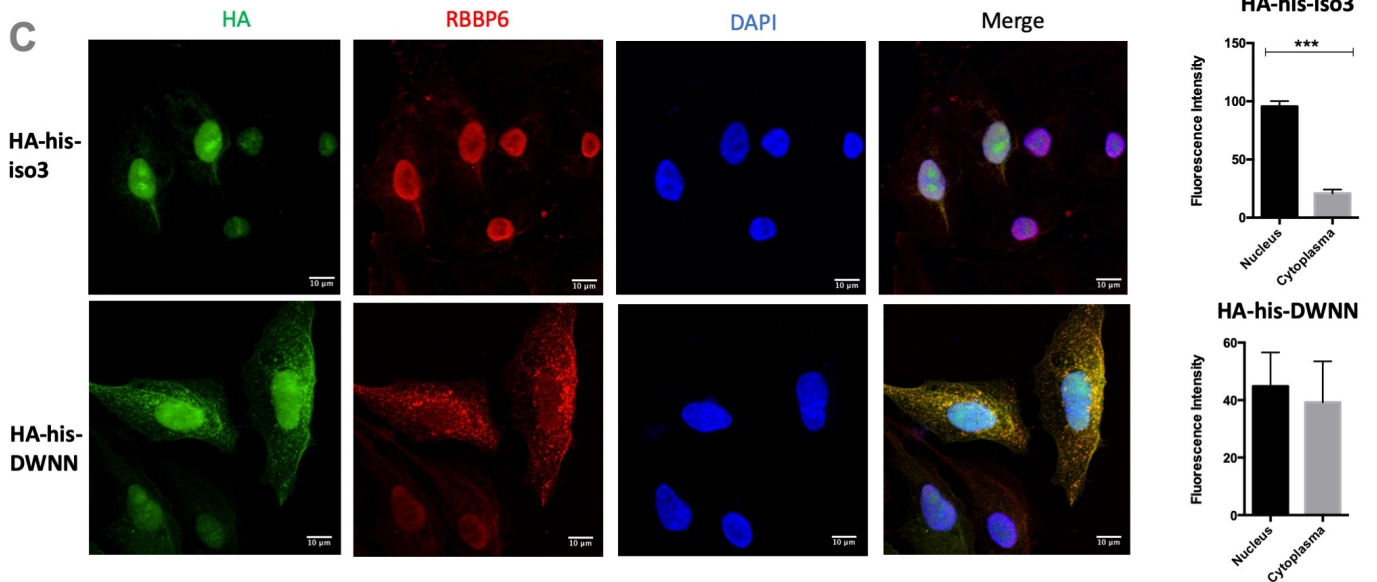
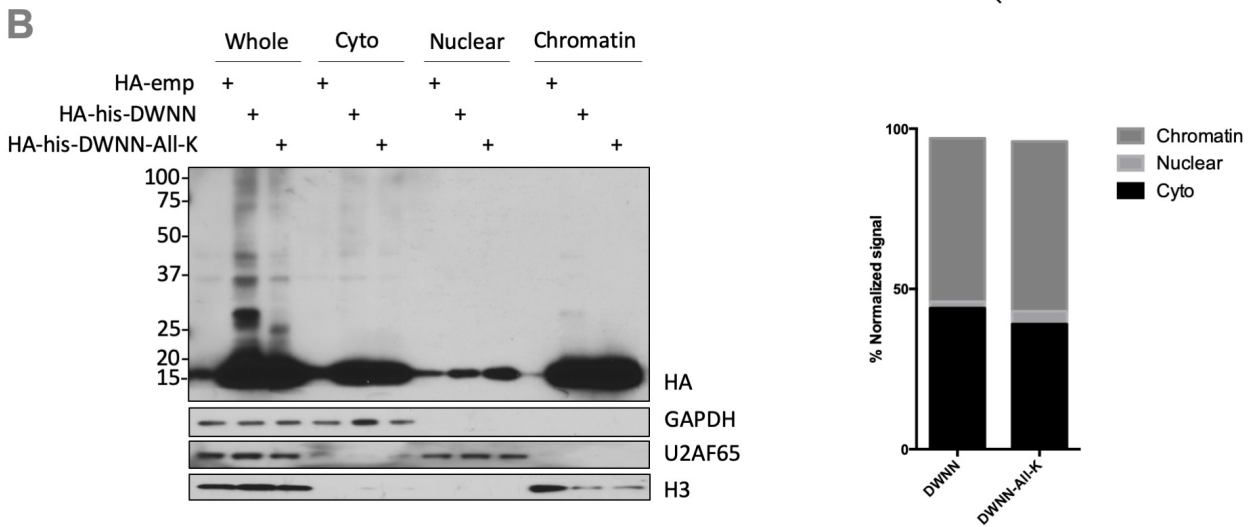
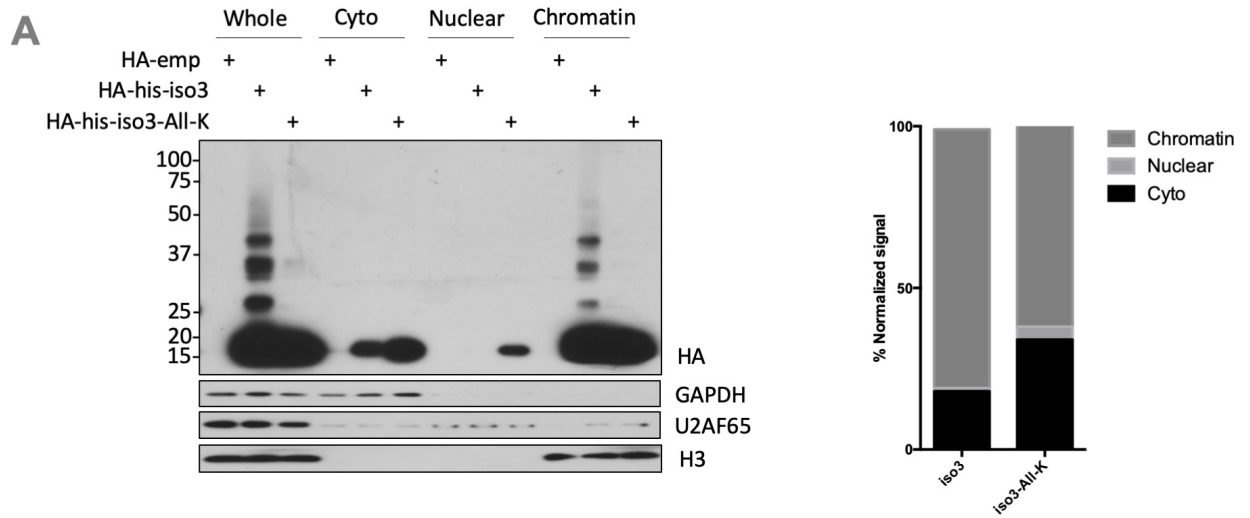
Figure S2. Metascape analysis of proteins detected in mass spectrometry of denaturing HA-his-iso3 pulldown samples. 78 proteins detected in group1, and 165 proteins detected in group2 were annotated using Metascape Analysis (<http://metascape.org/>), respectively.

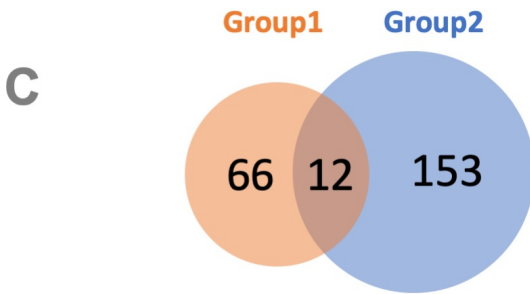
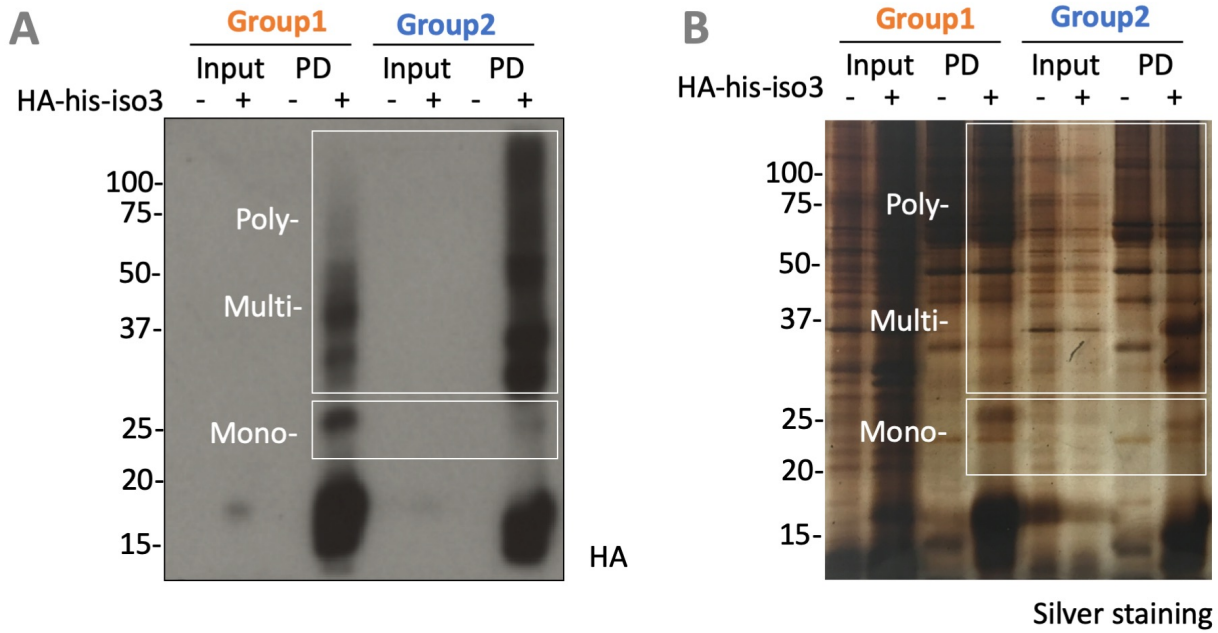
Figure S3. Significant co-occurrence of RBBP6 and IPO7 was observed in cancer. (A) Cross-cancer alteration summary for RBBP6 and IPO7 by cBioPortal for Cancer Genomics. RBBP6 and IPO7 share the top 5 most altered cancer types: prostate cancer, endometrial cancer, cutaneous melanoma, colorectal adenocarcinoma, and bladder cancer. (B) Significant co-occurrence of RBBP6 and IPO7 amplification was observed in neuroendocrine prostate cancer (Trento/Cornell/Broad 2016) by cBioPortal for Cancer Genomics (p -value <0.001). (C) Positive correlation was detected in RBBP6 and IPO7 mRNA expression levels (Pearson correlation = 0.62 and Spearman correlation = 0.75) in neuroendocrine prostate cancer (Trento/Cornell/Broad 2016).





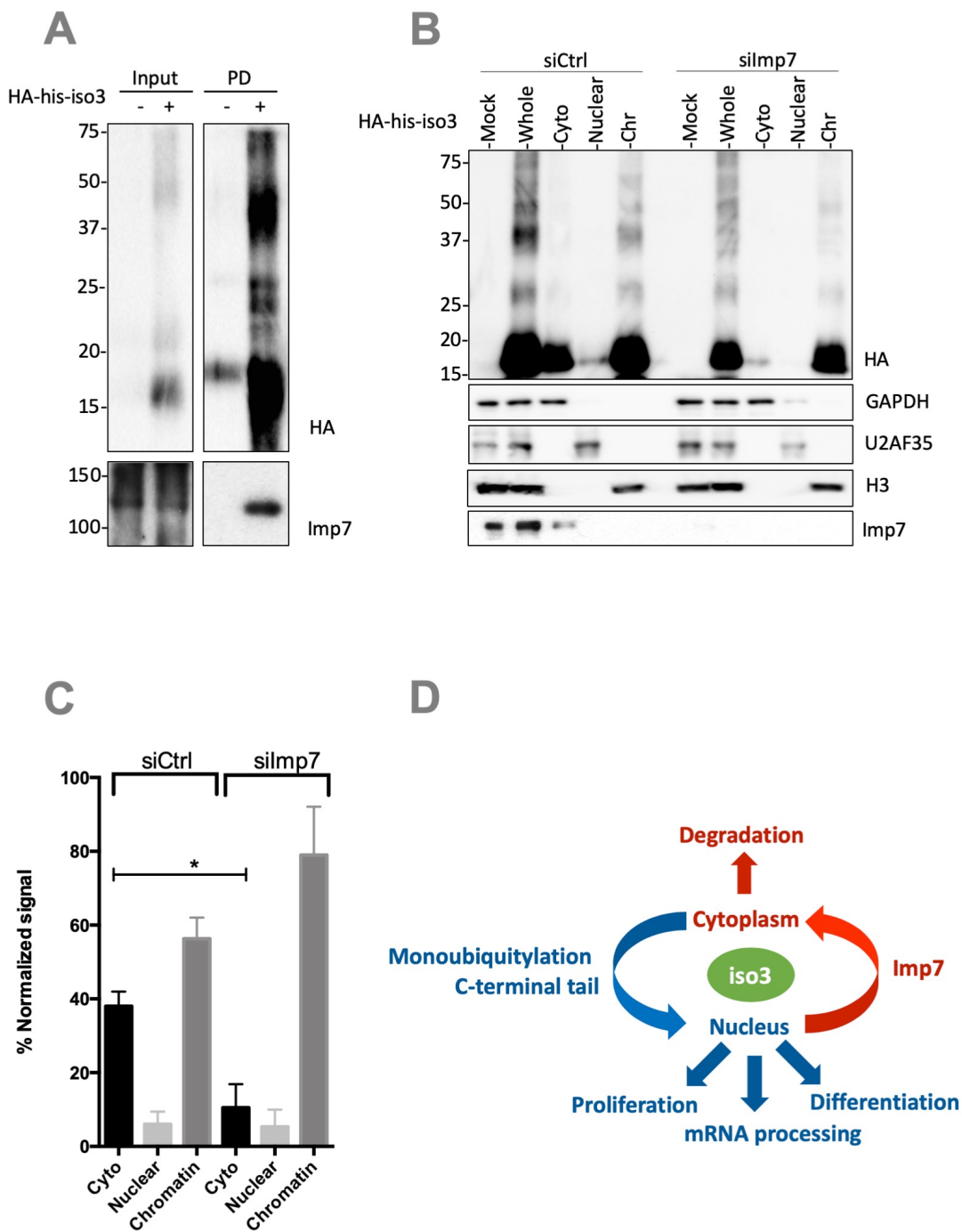






D

Accession No.	Protein Name	Spectral Counts				
		Group1		Group2		
		Ctrl	iso3	Ctrl	iso3	
P0CG47	Polyubiquitin-B	42	126	22	327	Ubiquitination
P0CG48	Polyubiquitin-C	43	131	22	334	
P62979	Ubiquitin-40S ribosomal protein S27a	41	119	22	318	
P62987	Ubiquitin-60S ribosomal protein L40	41	119	22	318	
Q9NNW5	WD repeat-containing protein 6	4	11	31	85	High spectral counts
P22681	E3 ubiquitin-protein ligase CBL	5	13	40	72	
Q8WV24	Pleckstrin homology-like domain family A member 1	14	28	9	40	
O95373	Importin-7 (IPO7)	0	5	15	26	
P04792	Heat shock protein beta-1	0	8	0	7	Low spectral counts
Q8TB05	UBA-like domain-containing protein 1	0	7	0	6	
Q8IYN6	UBA-like domain-containing protein 2	0	3	0	6	
P35232	Prohibitin	2	6	0	5	



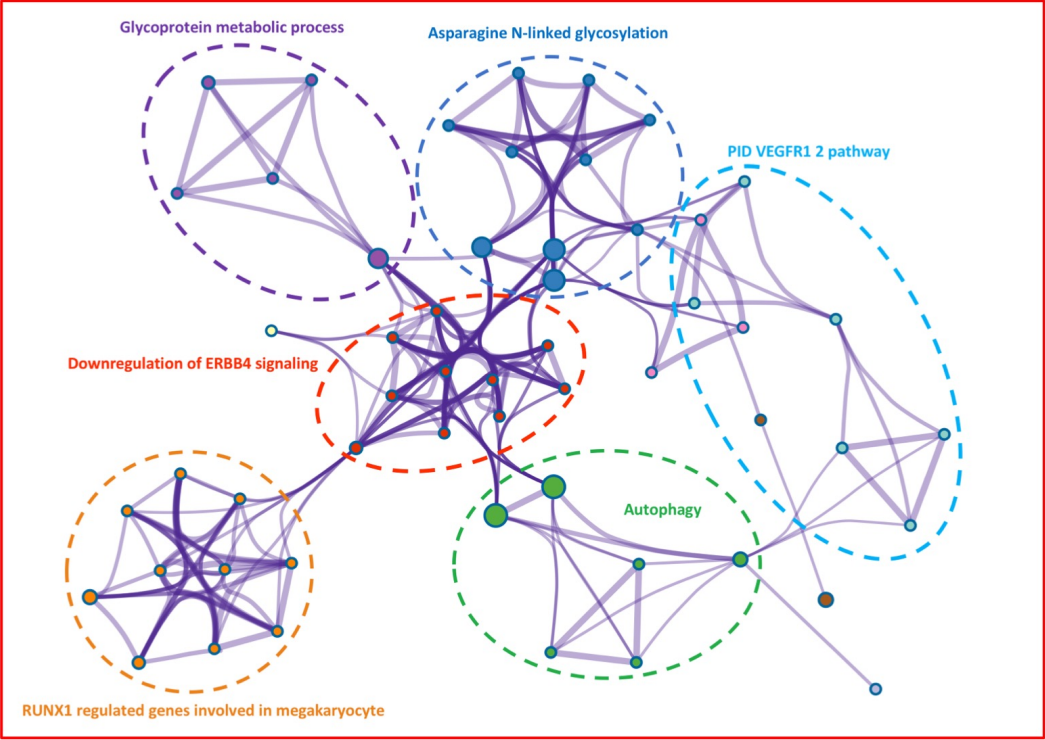
Protein sequence of RBBP6 iso3

MSCVHYKFSSKLNVDTVTFDGLHISLCDLKKQIMGREKLKAADCDLQITNAQTKEEYTDNALIPKNSS
VIVRRIPIGGVKSTSKTYVISRTEPAMATTKAVCKNTISHFFYTLPL*
RBBP6 Ab1 RBBP6 Ab2
RBBP6 Ab3

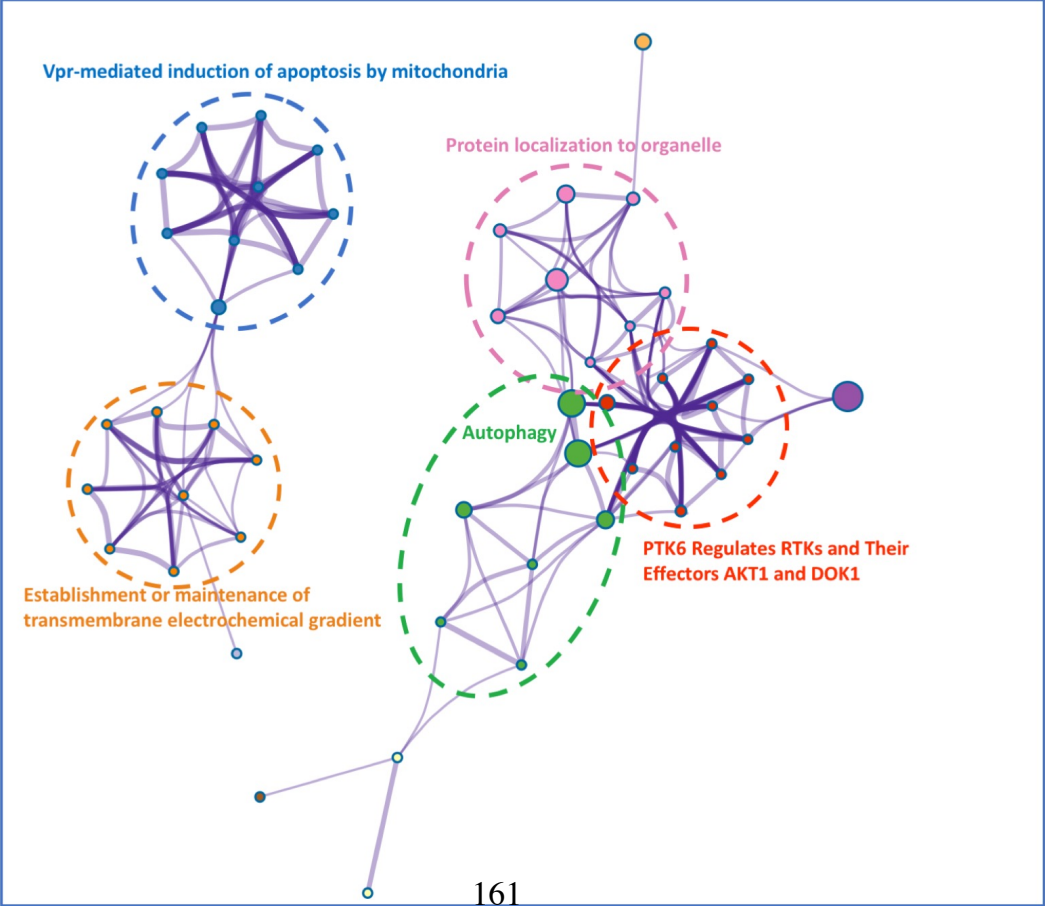
Protein sequence of RBBP6N *

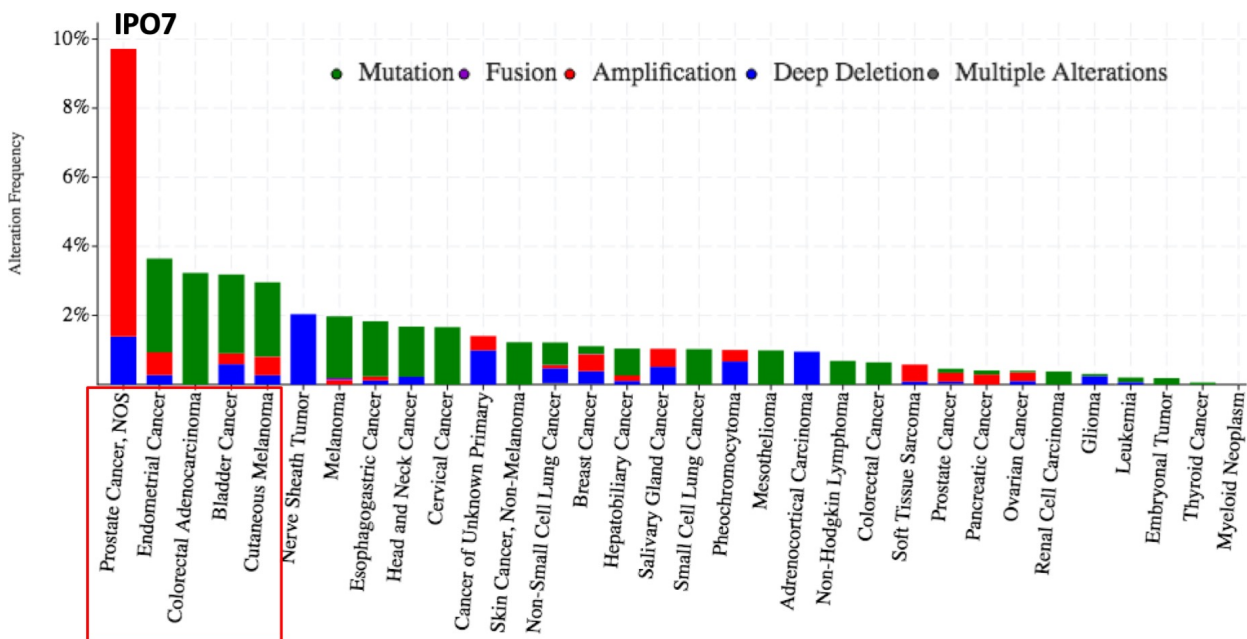
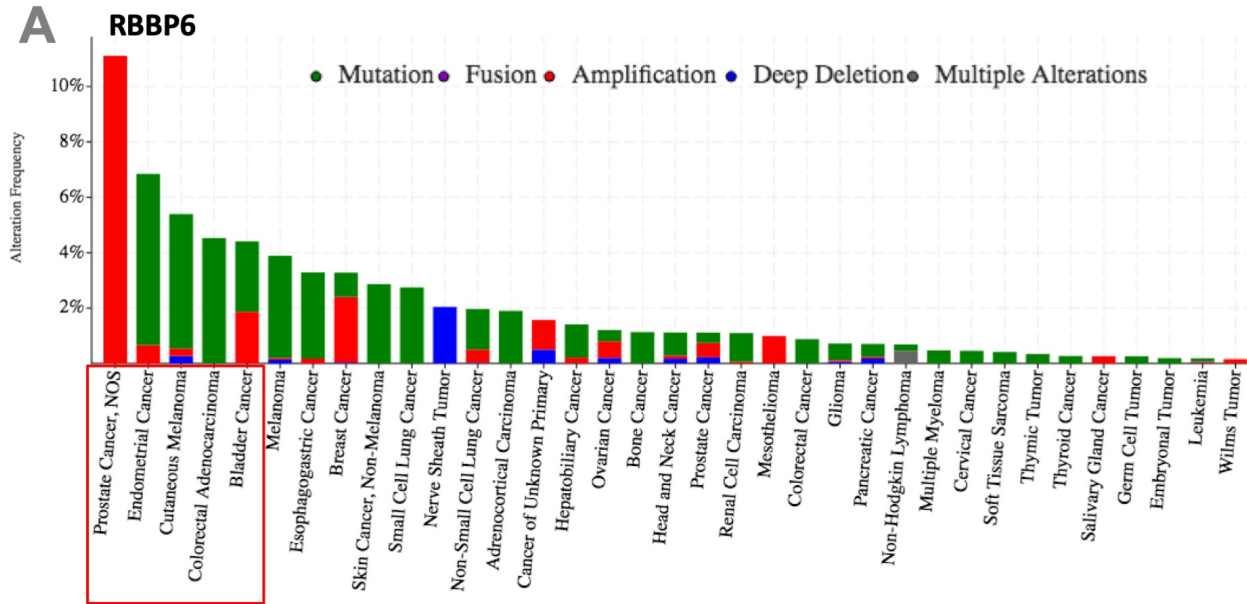
MSCVHYKFSSKLNVDTVTFDGLHISLCDLKKQIMGREKLKAADCDLQITNAQTKEEYTDNALIPKNSS
VIVRRIPIGGVKSTSKTYVISRTEPAMATTKAIDDSSASISLAQLTKTANLAEANASEEDKIKAMMSQSG
HEYDPINYMKKPLGPPPSYTCFRGKPGHYIKNCPTNGDKNFESGPRIKKSTGIPRSFMMMEVK.....
RBBP6 Ab1 RBBP6 Ab2
RBBP6 Ab3

**Mass Spec
Group1**



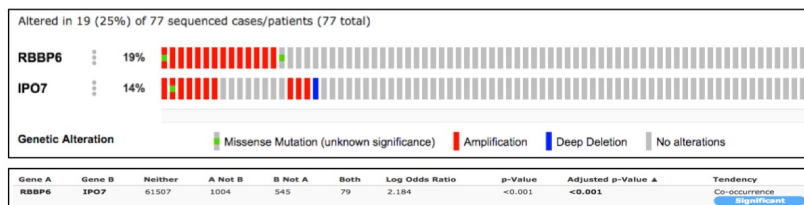
**Mass Spec
Group2**





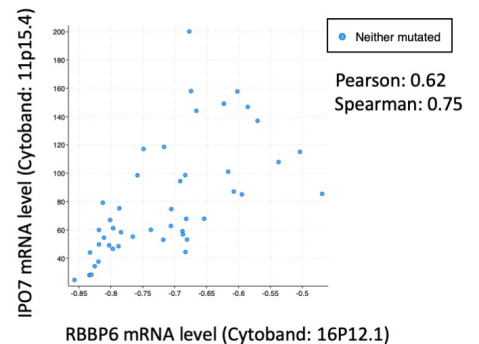
B

Neuroendocrine Prostate Cancer (Trento/Cornell/Broad 2016), Tumor Samples with sequencing and CNA data (107 samples)



C

RBBP6 and IPO7 mRNA co-expression in Neuroendocrine Prostate Cancer



Chapter 5. General discussion and future directions

In the first part of the thesis, I explored the biogenesis and degradation mechanisms of eRNAs. Our genome-wide analyses characterized a large pool of human polyadenylated enhancer RNAs under nuclear surveillance deficient conditions. This result expanded our knowledge about eRNAs, which are typically considered to be nonpolyadenylated. I further investigated the roles of the three complexes, PAXT/PPC, NEXT, and the Integrator, in regulating pA⁺ eRNA metabolism, and proposed a model of the Integrator-mediated RNAP II switch between enhancers and promoters. Additionally, I found an enrichment of PAS and GU-rich sequences near the 3' cleavage sites of those pA⁺ eRNAs, and provided evidence of the involvement of canonical poly(A) polymerase in their 3' processing. Lastly, I showed that stabilized nuclear pA⁺ eRNAs were transported into the cytoplasm and associated with polysomes, implying translational potential.

Nevertheless, several questions remain to be answered. For example, do all enhancers produce pA⁺ and pA⁻ eRNA simultaneously? If yes, then what is the ratio of pA⁺ and pA⁻ eRNA levels for any given enhancer under deficient nuclear surveillance conditions? My RL-PAT assay indeed detected the coexistence of pA⁺ and pA⁻ eRNAs derived from individual enhancers. Nevertheless, a genome-wide analyses would provide a more comprehensive picture. Since the 3'READS⁺ technique only analyzes polyadenylated RNA species, it would be nice to conduct both RNA-seq and 3'READS⁺ side by side, and analyze eRNAs with different polyadenylation statuses under the same experiment conditions.

Another question is whether the canonical poly(A) polymerases PAP- α/γ regulate the 3' processing of all pA⁺ eRNAs? My analyses of individual eRNAs showed that depletion of PAP- α/γ led to transcript readthrough and polyadenylation at much more downstream sites. To find out whether the involvement of PAP- α/γ in pA⁺ eRNA 3' processing applies globally, a genome-wide analysis is necessary.

The polysome fractionation experiments in Chapter 3 indicated that stabilized pA+ eRNAs accumulated in the cytoplasm and associated with polysomes, implying the possibility of producing *de novo* micropeptides. The genome editing technique CRISPR/Cas9 would be an ideal tool to help identify those micropeptides, if there are any, by adding specific tags such as GFP on them (Lackner et al., 2015). One big technical challenge would be detecting micropeptides, especially those with very low abundance (Makarewich and Olson, 2017). Nevertheless, this topic might be key evidence for the evolutionary significance of enhancers, and merits further study.

Lastly, I'd like to discuss some preliminary data about the role of Mtr4 in stem cell differentiation. Chapter 2 has reviewed the importance of the nuclear surveillance system in stem cell self-renewal and differentiation, which requires precise regulation and fidelity maintenance of gene expression. Aberrant accumulation of RNAs under deficient nuclear surveillance conditions would lead to the imbalance between proliferation and differentiation (Ogami et al., 2018). It has been reported that Mtr4 depletion in murine cancer cell lines Neuro2A and P19 cells resulted in decreased cellular proliferation and increased expression of genes indicative of cell differentiation (Onderak and Anderson, 2017). Consistent with this, I also found the involvement of Mtr4 in human embryonic stem cell (ESC) differentiation, by collaborating with Dr. Takashi Yamazaki, who performed the embryonic body differentiation assay presented below (Figure 5-1).

A human H9 ESC line (WA09) was differentiated using a 20-day embryoid body (EB) differentiation assay (protocol from Yamazaki et al., 2018), and then harvested at various time points (0, 5, 10, 20 days of differentiation, respectively) for qRT-PCR and Western blotting analyses. Downregulation of the pluripotency markers OCT4 and NANOG, accompanied with upregulation of differentiation markers IGBP7, RUNX1, FOXA, and MSX1 validated the effectiveness of the EB differentiation assay (Figure 5-1A). Remarkably, the protein levels of Mtr4

and ZCCHC8, both of which are subunits of the NEXT complex, were drastically decreased after 10 days of differentiation. In contrast, the protein level of PABPN1, a subunit of PAXT complex, remained unchanged (Figure 5-1B). This result indicates that different nuclear exosome adaptor complexes might play distinct roles in ESC differentiation. I further examined the expression level of two eRNAs (eRNA11+ and eRNA17+) and two well-known lncRNAs (MALAT1 and NEAT1). One eRNA and both lncRNAs (eRNA17+, MALAT1, NEAT1) showed tens of folds increase after hESC differentiation. Additionally, I transfected hESCs with control or Mtr4 siRNAs, and noticed that KD of Mtr4 induced a flattened epithelial-like appearance (Figure 5-1D), which is typically observed in differentiating ESCs (Ji et al., 2012; Tay et al., 2008). Nevertheless, it is noteworthy that transfecting ESCs with siRNAs is rather difficult. Based on the experiences in our lab, the maximum siRNA transfection efficiency in ESC is 10%. Thus, we wondered whether downregulation of Mtr4 in a small number of cells is sufficient to induce differentiation. Viral transfection and differentiation marker staining would be necessary to test this hypothesis. Nevertheless, these preliminary observations strongly suggested the involvement of Mtr4 in ESC differentiation, which merits further study in the future.

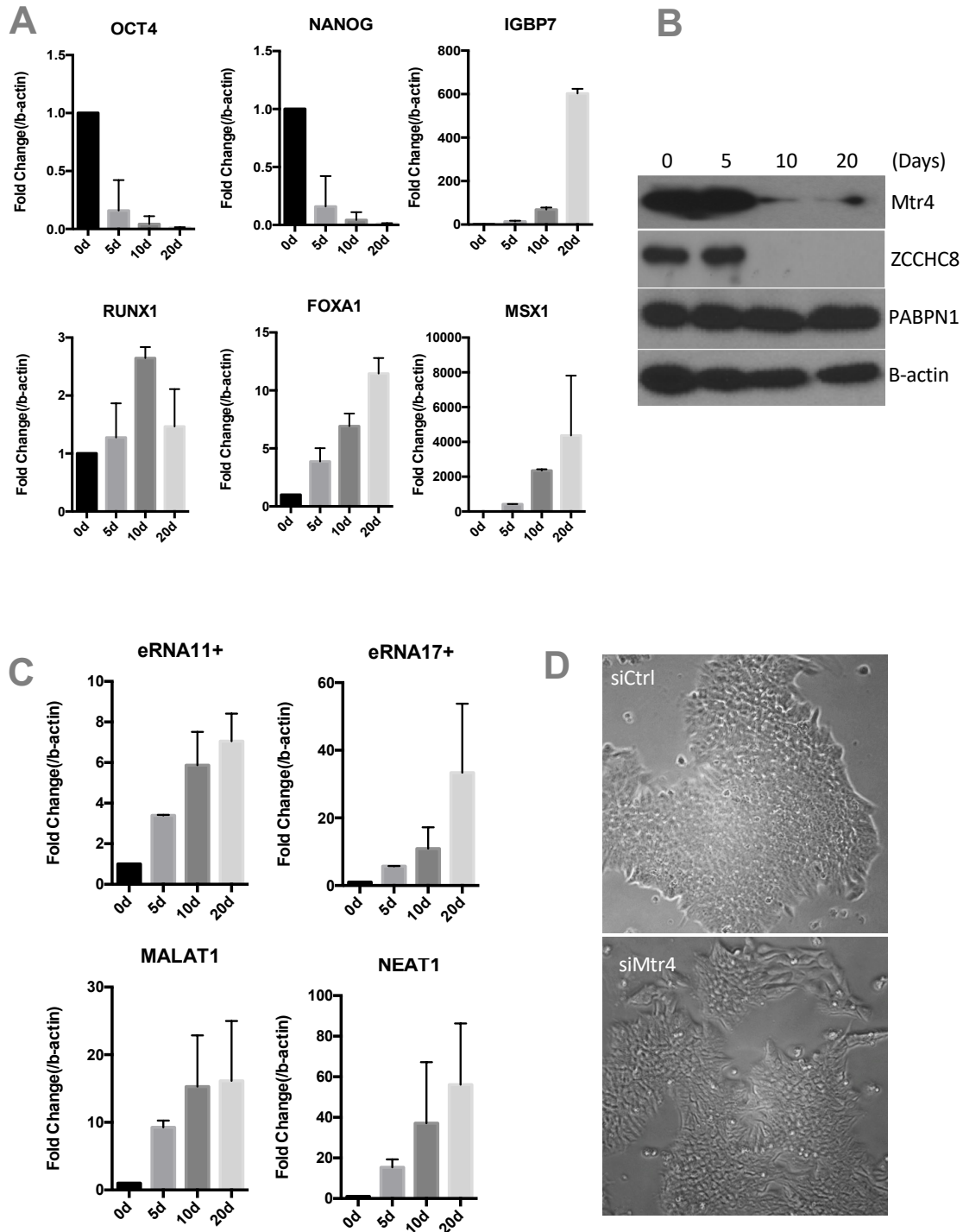


Figure 5-1. Mtr4 decreased drastically after ESC differentiation.

A. RT-qPCR analysis of ESC pluripotency and differentiation markers in a 20 day embryoid body differentiation assay: OCT4 and NANOG (pluripotency markers, IGBP7 (differentiation marker), RUNX1 (mesoderm marker), FOXA (endoderm marker), MSX1 (ectoderm marker). B. Western blot analysis of embryoid body after 20 days of differentiation with antibodies listed on the right. C. RT-qPCR result of eRNA17+, eRNA11+, MALAT1, and NEAT1. D. siMtr4 treatment of ESC cells induced a differentiation morphology 72hrs after siRNA transfection.

In the second part of the thesis, I studied RBBP6, a polyadenylation factor. The RBBP6 isoform3 protein displays high similarities with ubiquitin, indicating its potential role as a ubiquitin-like modifier. Nevertheless, I found that RBBP6 iso3 is not a ubiquitin-like modifier, instead is itself ubiquitylated by both monoubiquitin and polyubiquitin chains. I also showed that the nuclear localization of iso3 is facilitated by monoubiquitylation and its C-terminal tail. Finally, I identified a novel interaction between iso3 and Importin7, which assists the cytoplasmic retention of iso3. My work revealed the dynamic subcellular localization mechanism of RBBP6 iso3, which can further modulate the ratio between long-form RBBP6 and iso3 in a variety of biological processes.

I also examined the cross-cancer alteration of RBBP6 and Importin7 genes in the Cancer Genomics database, and observed a significant co-occurrence of amplification of the two genes. All the findings combined together show that Importin7 and RBBP6 iso1 are often upregulated in cancer, while RBBP6 iso3 is downregulated. Therefore, I proposed a model: upregulation of Importin7 would retain more iso3 protein in the cytoplasm, so that less iso3 competes with RBBP6 iso1 in the nucleus. However, this cause-effect relationship between Importin7 and RBBP6 iso3 in cancer still needs more evidence to back it up, which can be a direction for future study.

Another open question is about the unexpected peptide cleavage of the RBBP6N protein. As shown in Figure 4 of Chapter 4, two RBBP6 N-terminal derivative plasmids were constructed: a 3xFLAG-tagged RBBP6N containing the key RING domain, Zinc knuckle, the DWNN domain, as well as a 3xFLAG-tagged Δ DWNN only containing the RING domain and Zinc knuckle. Both RBBP6N and Δ DWNN were tagged on the C-terminus. Surprisingly, I noticed that the western blot analyses of RBBP6N-FLAG always detected two bands, with molecular weights of 75 KDa and 50 KDa, respectively, whereas Δ DWNN-FLAG showed one band (approximately 50KD)

(Figure 5-2A). Given that the molecular weight of RBBP6N-FLAG is 75 KDa, it's very possible that the 50 KDa band is generated by peptide cleavage of RBBP6N protein, by cleaving off the DWNN domain (approximately 15 KDa). To test this possibility, I constructed another RBBP6N plasmid, with 3xFLAG tag on the C-terminus (FLAG-RBBP6N), and transfected it side by side with RBBP6N-FLAG into HeLa cells, with MG132 treatment to block potential protein degradation. As shown in Figure 5-2B, FLAG-RBBP6N only displayed one band at 75 KDa, while RBBP6N-FLAG displayed two bands. It is very likely that the N-terminus of RBBP6N is cleaved off, so that the cleaved RBBP6N derivative of FLAG-RBBP6N didn't contain the FLAG tag anymore, and therefore couldn't be detected via Western blot analysis.

In order to determine the peptide cleavage site, I performed immunoprecipitation of HeLa cells overexpressing RBBP6N-FLAG proteins with an anti-FLAG antibody. As shown in Figure 5-2C, both RBBP6N and cleaved RBBP6N were pulled down. The gel band of the cleaved RBBP6N protein was analyzed by mass spectrometry, and the peptide closest to the N-terminus of RBBP6N was shown in Figure 5-2D. Thus, the peptide cleavage site of RBBP6N should be at least before this peptide sequence detected by mass spectrometry.

Next, I used three RBBP6 antibodies recognizing different N-terminal epitopes of RBBP6 protein, to further pinpoint the peptide cleavage site (Figure 5-2E). I found that all three RBBP6 antibodies could only detect the uncleaved RBBP6N, which implied that the cleavage occurs after their epitopes. Taken together, as shown in Figure 5-2F, we concluded that the peptide cleavage site of RBBP6N is after the DWNN domain, and is very likely within the epitope recognized by RBBP6 Ab3.

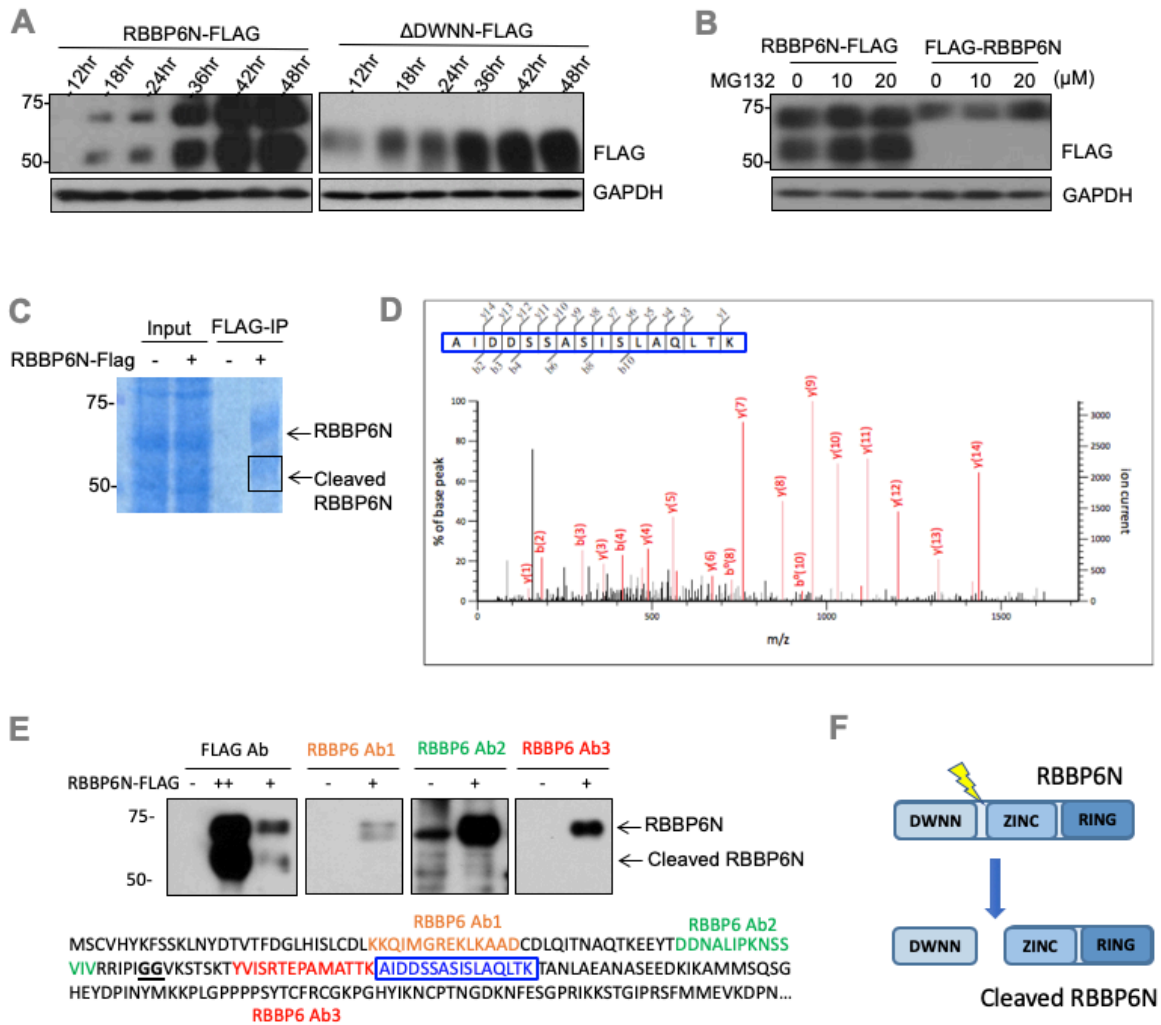


Figure 5-2. RBBP6N underwent peptide cleavage after the N-terminal DWNN domain.

(A) Western blot analysis of the overexpression of RBBP6N-FLAG and ΔDWNN-FLAG in HeLa cells from 12hrs to 48hrs after plasmid transfection. (B) Western blot analysis of C-terminal FLAG tagged RBBP6N (RBBP6N-FLAG) and N-terminal FLAG tagged RBBP6N (FLAG-RBBP6N) under 20uM MG132 treatment. (C) Coomassie blue staining of FLAG-IP samples of RBBP6N-FLAG transfected HeLa cells. The upper band is the RBBP6N protein, while the lower band is cleaved C-terminal RBBP6N. (D) Mass spectrometry analysis of the cleaved RBBP6N gel band. The peptide detected closest to the N-terminus of RBBP6N was boxed in blue. (E) Western blot analysis of RBBP6N peptide cleavage with the FLAG antibody and RBBP6 specific antibodies. The epitopes of RBBP6 specific antibodies were highlighted in orange (Ab1), green (Ab2), and red (Ab3), respectively. The peptide fragment detected by mass spectrometry was highlighted in blue. (F) Schematic illustration of RBBP6N peptide cleavage. The peptide cleavage of RBBP6N occurred right after the DWNN domain.

It is known that the DWNN domain itself has opposite functions from the full-length RBBP6 in various biological process, such as mRNA 3' processing and cell proliferation. Consequently, different ratios between isoforms could help regulate those processes (Di

Giammartino et al., 2014; Mbita et al., 2012). Therefore, the peptide cleavage of the N-terminus DWNN domain from the RBBP6N protein might facilitate the ratio change between the isoforms. For example, enhanced peptide cleavage of RBBP6N would generate more DWNN proteins and less functional RBBP6N, and vice versa. However, we didn't observe such peptide cleavage in full-length RBBP6 (Figure 5-3A). We treated HeLa cells with control siRNAs, siRNAs against only isoform1 and 2, as well as all three isoforms of RBBP6, and performed Western blot analysis using two different antibodies recognizing the C-terminus of the RBBP6 protein. If such peptide cleavage exists, we should have been able to see the cleaved product (approximately 230 KDa) just below full-length RBBP6 (250 KDa).

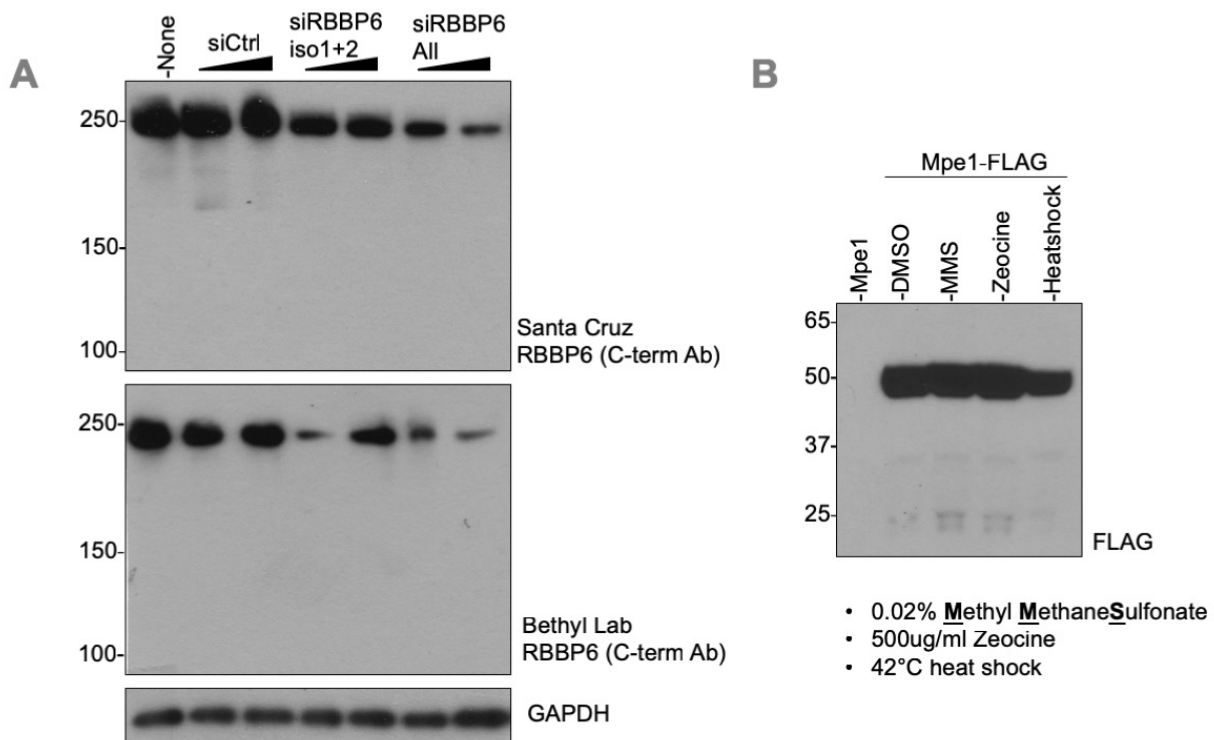


Figure 5-3. The peptide cleavage of RBBP6N was not found in full-length RBBP6 or its yeast homolog Mpe1. (A) Western blot analysis of full-length RBBP6 under siRNA knockdown. (B) Western blot analysis of RBBP6 yeast homolog Mpe1-FLAG under DMSO and DNA damage treatment conditions. Yeasts expressing Mpe1-FLAG were treated with DMSO, 0.02% Methyl Methane Sulfonate (MMS), 500ug/ml Zeocine, and 42 °C degree heat shock for 1hr, respectively.

Given that full-length human RBBP6 contains a very long C-terminus compared with RBBP6N, we decided to look into the yeast homolog of RBBP6, Mpe1, which only contains the DWNN domain, Zinc Knuckle, and RING domain, similar to RBBP6N. We collaborated with Dr. Julyun Oh from Dr. Lorraine S. Symington lab, and tagged the endogenous Mpe1 with 3xFLAG on the C-terminus in yeast. Still, no peptide cleavage of Mpe1 was detected, with or without DNA damage treatment, as indicated in Figure 5-3B.

Despite the fact that N-terminal peptide cleavage only exists in RBBP6N but not endogenous full-length RBBP6, this finding provides crucial information about the features of RBBP6N. If this plasmid construct will be used for other studies, researchers should be aware that the N-terminus of RBBP6N undergoes peptide cleavage.

It is still an open question, how and why the DWNN domain is cleaved off in RBBP6N proteins. If we can find out the underlying mechanism and reproduce the cleavage in full-length RBBP6, it might provide a novel way to modulate the function of RBBP6.

References

Aguilera, A., and Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. *Mol Cell* *46*, 115-124.

Albrecht, T.R., Shevtsov, S.P., Wu, Y., Mascibroda, L.G., Peart, N.J., Huang, K.L., Sawyer, I.A., Tong, L., Dundr, M., and Wagner, E.J. (2018). Integrator subunit 4 is a 'Symplekin-like' scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic Acids Res* *46*, 4241-4255.

Alvarez-Dominguez, J.R., Hu, W., Yuan, B., Shi, J., Park, S.S., Gromatzky, A.A., van Oudenaarden, A., and Lodish, H.F. (2014). Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood* *123*, 570-581.

Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., *et al.* (2013). The human cap-binding complex is

functionally connected to the nuclear RNA exosome. *Nat Struct Mol Biol* 20, 1367-1376.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., *et al.* (2014a). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455-461.

Andersson, R., Refsing Andersen, P., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T., and Sandelin, A. (2014b). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun* 5, 5336.

Ao, Z., Huang, G., Yao, H., Xu, Z., Labine, M., Cochrane, A.W., and Yao, X. (2007). Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. *J Biol Chem* 282, 13456-13467.

Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., Lennartsson, A., Ronnerblad, M., Hrydziuszko, O., Vitezic, M., *et al.* (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* 347, 1010-1014.

Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N.J. (1997). Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* 11, 2494-2509.

Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 123, 265-276.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299-308.

Batra, J., Hultquist, J.F., Liu, D., Shtanko, O., Von Dollen, J., Satkamp, L., Jang, G.M., Luthra, P., Schwarz, T.M., Small, G.I., *et al.* (2018). Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication. *Cell* 175, 1917-1930 e1913.

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Res* 10, 1001-1010.

Beaulieu, Y.B., Kleinman, C.L., Landry-Voyer, A.M., Majewski, J., and Bachand, F. (2012). Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS Genet* 8, e1003078.

Bergink, S., and Jentsch, S. (2009). Principles of ubiquitin and SUMO modifications in DNA repair. *Nature* 458, 461-467.

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

Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. *Cell* 168,

629-643.

Brazao, T.F., Johnson, J.S., Muller, J., Heger, A., Ponting, C.P., and Tybulewicz, V.L. (2016). Long noncoding RNAs in B-cell development and activation. *Blood* *128*, e10-19.

Bresson, S.M., and Conrad, N.K. (2013). The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. *PLoS Genet* *9*, e1003893.

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

Brodsky, A.S., and Silver, P.A. (2000). Pre-mRNA processing factors are required for nuclear export. *RNA* *6*, 1737-1749.

Brown, K.M., and Gilmartin, G.M. (2003). A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell* *12*, 1467-1476.

Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet* *28*, 276-284.

Bulger, M., and Groudine, M. (2010). Enhancers: the abundance and function of regulatory sequences beyond promoters. *Dev Biol* *339*, 250-257.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell* *144*, 327-339.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Mol Cell* *49*, 825-837.

Carter, S., Bischof, O., Dejean, A., and Vousden, K.H. (2007). C-terminal modifications regulate MDM2 dissociation and nuclear export of p53. *Nature Cell Biology* *9*, 428-U111.

Chen, H., Li, C., Peng, X., Zhou, Z., Weinstein, J.N., Cancer Genome Atlas Research, N., and Liang, H. (2018). A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell* *173*, 386-399 e312.

Chen, J., Tang, H., Wu, Z., Zhou, C., Jiang, T., Xue, Y., Huang, G., Yan, D., and Peng, Z. (2013). Overexpression of RBBP6, alone or combined with mutant TP53, is predictive of poor prognosis in colon cancer. *PLoS One* *8*, e66524.

Chibi, M., Meyer, M., Skepu, A., DJ, G.R., Moolman-Smook, J.C., and Pugh, D.J. (2008). RBBP6 interacts with multifunctional protein YB-1 through its RING finger domain, leading to ubiquitination and proteosomal degradation of YB-1. *J Mol Biol* *384*, 908-916.

Cohen, S.M. (2014). Everything old is new again: (linc)RNAs make proteins! *EMBO J* *33*, 937-

938.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev* *11*, 2755-2766.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57-74.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics* *46*, 1311-1320.

Dao, L.T.M., Galindo-Albarran, A.O., Castro-Mondragon, J.A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., Stephen, T., *et al.* (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature Genetics* *49*, 1073-+.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* *8*, e1000384.

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev* *26*, 11-24.

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev* *30*, 1357-1382.

Deshai, R.J., and Joazeiro, C.A.P. (2009). RING Domain E3 Ubiquitin Ligases. *Annu Rev Biochem* *78*, 399-434.

Dhanoya, A., Wang, T., Keshavarz-Moore, E., Fassati, A., and Chain, B.M. (2013). Importin-7 mediates nuclear trafficking of DNA in mammalian cells. *Traffic* *14*, 165-175.

Di Giammartino, D.C., Li, W., Ogami, K., Yashinski, J.J., Hoque, M., Tian, B., and Manley, J.L. (2014). RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs. *Genes Dev* *28*, 2248-2260.

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

Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Gene Dev* *18*, 2485-2490.

Eckmann, C.R., Rammelt, C., and Wahle, E. (2011). Control of poly(A) tail length. Wiley

Interdiscip Rev RNA 2, 348-361.

Epstein, D.J. (2009). Cis-regulatory mutations in human disease. *Brief Funct Genomic Proteomic* 8, 310-316.

Fan, J., Kuai, B., Wu, G., Wu, X., Chi, B., Wang, L., Wang, K., Shi, Z., Zhang, H., Chen, S., *et al.* (2017). Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. *EMBO J* 36, 2870-2886.

Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shores, N., Whitton, H., Ryan, R.J., Shishkin, A.A., *et al.* (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337-343.

Fassati, A., Gorlich, D., Harrison, I., Zaytseva, L., and Mingot, J.M. (2003). Nuclear import of HIV-1 intracellular reverse transcription complexes is mediated by importin 7. *EMBO J* 22, 3675-3685.

Fenteany, G., and Schreiber, S.L. (1998). Lactacystin, proteasome function, and cell fate. *J Biol Chem* 273, 8545-8548.

Frith, M.C., Forrest, A.R., Nourbakhsh, E., Pang, K.C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Bailey, T.L., and Grimmond, S.M. (2006). The abundance of short proteins in the mammalian proteome. *PLoS Genet* 2, e52.

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

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. *Science* 361, 1341-1345.

Galisson, F., Mahrouche, L., Courcelles, M., Bonneil, E., Meloche, S., Chelbi-Alix, M.K., and Thibault, P. (2011). A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol Cell Proteomics* 10, M110 004796.

Gao, S., and Scott, R.E. (2002). P2P-R protein overexpression restricts mitotic progression at prometaphase and promotes mitotic apoptosis. *J Cell Physiol* 193, 199-207.

Gao, S., and Scott, R.E. (2003). Stable overexpression of specific segments of the P2P-R protein in human MCF-7 cells promotes camptothecin-induced apoptosis. *J Cell Physiol* 197, 445-452.

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. *Mol Cell* 56, 128-139.

Gil, A., and Proudfoot, N.J. (1987). Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. *Cell* 49, 399-406.

Gorlich, D., Dabrowski, M., Bischoff, F.R., Kutay, U., Bork, P., Hartmann, E., Prehn, S., and Izaurralde, E. (1997). A novel class of RanGTP binding proteins. *J Cell Biol* *138*, 65-80.

Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* *154*, 240-251.

Haglund, K., Di Fiore, P.P., and Dikic, I. (2003a). Distinct monoubiquitin signals in receptor endocytosis. *Trends in Biochemical Sciences* *28*, 598-603.

Haglund, K., Sigismund, S., Polo, S., Szymkiewicz, I., Di Fiore, P.P., and Dikic, I. (2003b). Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. *Nat Cell Biol* *5*, 461-466.

Hah, N., Benner, C., Chong, L.W., Yu, R.T., Downes, M., and Evans, R.M. (2015). Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. *Proc Natl Acad Sci U S A* *112*, E297-302.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* *23*, 1210-1223.

Harutyunyan, A.S., Giambruno, R., Krendl, C., Stukalov, A., Klampfl, T., Berg, T., Chen, D., Milosevic Feenstra, J.D., Jager, R., Gisslinger, B., *et al.* (2016). Germline RBBP6 mutations in familial myeloproliferative neoplasms. *Blood* *127*, 362-365.

Hatzis, P., and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol Cell* *10*, 1467-1477.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* *16*, 144-154.

Hendriks, I.A., D'Souza, R.C., Yang, B., Verlaan-de Vries, M., Mann, M., and Vertegaal, A.C. (2014). Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat Struct Mol Biol* *21*, 927-936.

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Gene Dev* *32*, 26-41.

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. *Annu Rev Biochem* *67*, 425-479.

Heyer, E.E., and Moore, M.J. (2016). Redefining the Translational Status of 80S Monosomes. *Cell* *164*, 757-769.

Hicke, L. (2001). Protein regulation by monoubiquitin. *Nat Rev Mol Cell Biol* *2*, 195-201.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young,

- R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934-947.
- Hsieh, C.L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.S., Chen, S., *et al.* (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci U S A* 111, 7319-7324.
- Husnjak, K., and Dikic, I. (2012). Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu Rev Biochem* 81, 291-322.
- Jakel, S., Albig, W., Kutay, U., Bischoff, F.R., Schwamborn, K., Doenecke, D., and Gorlich, D. (1999). The importin beta/importin 7 heterodimer is a functional nuclear import receptor for histone H1. *EMBO J* 18, 2411-2423.
- Jakel, S., and Gorlich, D. (1998). Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO J* 17, 4491-4502.
- Ji, L., LaPointe, V.L., Evans, N.D., and Stevens, M.M. (2012). Changes in embryonic stem cell colony morphology and early differentiation markers driven by colloidal crystal topographical cues. *Eur Cell Mater* 23, 135-146.
- Jiao, A.L., Perales, R., Umbreit, N.T., Haswell, J.R., Piper, M.E., Adams, B.D., Pellman, D., Kennedy, S., and Slack, F.J. (2019). Human nuclear RNAi-defective 2 (NRDE2) is an essential RNA splicing factor. *RNA* 25, 352-363.
- Johnson, S.J., and Jackson, R.N. (2013). Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol* 10, 33-43.
- Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664-668.
- Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., *et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182-U165.
- Kim, T.K., and Shiekhattar, R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* 162, 948-959.
- Kleftogiannis, D., Kalnis, P., and Bajic, V.B. (2016). Progress and challenges in bioinformatics approaches for enhancer identification. *Brief Bioinform* 17, 967-979.
- Komander, D., and Rape, M. (2012). The Ubiquitin Code. *Annu Rev Biochem* 81, 203-229.
- Kuznetsov, V.A., Bondarenko, V., Wongsurawat, T., Yenamandra, S.P., and Jenjaroenpun, P. (2018). Toward predictive R-loop computational biology: genome-scale prediction of R-loops reveals their association with complex promoter structures, G-quadruplexes and transcriptionally active enhancers. *Nucleic Acids Res* 46, 8023.

- Lackner, D.H., Carre, A., Guzzardo, P.M., Banning, C., Mangena, R., Henley, T., Oberndorfer, S., Gapp, B.V., Nijman, S.M., Brummelkamp, T.R., *et al.* (2015). A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nat Commun* 6, 10237.
- Laffleur, B., Basu, U., and Lim, J. (2017). RNA Exosome and Non-coding RNA-Coupled Mechanisms in AID-Mediated Genomic Alterations. *J Mol Biol* 429, 3230-3241.
- Lai, F., Gardini, A., Zhang, A., and Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525, 399-403.
- Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494, 497-501.
- Laishram, R.S. (2014). Poly(A) polymerase (PAP) diversity in gene expression--star-PAP vs canonical PAP. *FEBS Lett* 588, 2185-2197.
- Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 39, 170-182.
- Lamoliatte, F., Bonneil, E., Durette, C., Caron-Lizotte, O., Wildemann, D., Zerweck, J., Wenshuk, H., and Thibault, P. (2013). Targeted identification of SUMOylation sites in human proteins using affinity enrichment and paralog-specific reporter ions. *Mol Cell Proteomics* 12, 2536-2550.
- Le Gras, S., Keime, C., Anthony, A., Lotz, C., De Longprez, L., Brouillet, E., Cassel, J.C., Boutillier, A.L., and Merienne, K. (2017). Altered enhancer transcription underlies Huntington's disease striatal transcriptional signature. *Sci Rep-Uk* 7.
- Lee, D.H., and Goldberg, A.L. (1998). Proteasome inhibitors: valuable new tools for cell biologists. *Trends Cell Biol* 8, 397-403.
- Lee, K., Hsiung, C.C.S., Huang, P., Raj, A., and Blobel, G.A. (2015). Dynamic enhancer-gene body contacts during transcription elongation. *Gene Dev* 29, 1992-1997.
- Lee, N.N., Chalamcharla, V.R., Reyes-Turcu, F., Mehta, S., Zofall, M., Balachandran, V., Dhakshnamoorthy, J., Taneja, N., Yamanaka, S., Zhou, M., *et al.* (2013). Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell* 155, 1061-1074.
- Lee, S.D., and Moore, C.L. (2014). Efficient mRNA Polyadenylation Requires a Ubiquitin-Like Domain, a Zinc Knuckle, and a RING Finger Domain, All Contained in the Mpel Protein. *Molecular and Cellular Biology* 34, 3955-3967.
- Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Curr Biol* 20, R754-763.

- Li, L., Deng, B.W., Xing, G.C., Teng, Y., Tian, C.Y., Cheng, X., Yin, X.S., Yang, J.T., Gao, X., Zhu, Y.P., *et al.* (2007). PACT is a negative regulator of p53 and essential for cell growth and embryonic development. *P Natl Acad Sci USA* *104*, 7951-7956.
- Li, M.Y., Brooks, C.L., Wu-Baer, F., Chen, D.L., Baer, R., and Gu, W. (2003). Mono-versus polyubiquitination: Differential control of p53 fate by Mdm2. *Science* *302*, 1972-1975.
- Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., *et al.* (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* *498*, 516-520.
- Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* *17*, 207-223.
- Li, W., and Ye, Y. (2008). Polyubiquitin chains: functions, structures, and mechanisms. *Cell Mol Life Sci* *65*, 2397-2406.
- Liu, Z., Merkurjev, D., Yang, F., Li, W., Oh, S., Friedman, M.J., Song, X., Zhang, F., Ma, Q., Ohgi, K.A., *et al.* (2014). Enhancer activation requires trans-recruitment of a mega transcription factor complex. *Cell* *159*, 358-373.
- Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., and Jensen, T.H. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep* *10*, 178-192.
- Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A., and Jensen, T.H. (2011). Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell* *43*, 624-637.
- Macheret, M., and Halazonetis, T.D. (2015). DNA replication stress as a hallmark of cancer. *Annu Rev Pathol* *10*, 425-448.
- Makarewich, C.A., and Olson, E.N. (2017). Mining for Micropeptides. *Trends Cell Biol* *27*, 685-696.
- Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* *65*, 1099-1122.
- Masternak, K., Peyraud, N., Krawczyk, M., Barras, E., and Reith, W. (2003). Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol* *4*, 132-137.
- Mathelier, A., Lefebvre, C., Zhang, A.W., Arenillas, D.J., Ding, J., Wasserman, W.W., and Shah, S.P. (2015). Cis-regulatory somatic mutations and gene-expression alteration in B-cell lymphomas. *Genome Biol* *16*, 84.

- Mather, A., Rakgotho, M., and Ntwasa, M. (2005). SNAMA, a novel protein with a DWNN domain and a RING finger-like motif: a possible role in apoptosis. *Biochim Biophys Acta* 1727, 169-176.
- Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138, 673-684.
- Mbita, Z., Meyer, M., Skepu, A., Hosie, M., Rees, J., and Dlamini, Z. (2012). De-regulation of the RBBP6 isoform 3/DWNN in human cancers. *Mol Cell Biochem* 362, 249-262.
- McLauchlan, J., Gaffney, D., Whitton, J.L., and Clements, J.B. (1985). The consensus sequence YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res* 13, 1347-1368.
- Meng, F.L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberg, D., *et al.* (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159, 1538-1548.
- Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A., *et al.* (2016). Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol Cell* 64, 520-533.
- Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I.E., Males, M., Viales, R.R., and Furlong, E.E.M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Gene Dev* 32, 42-57.
- Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* 38, 2757-2774.
- Miotto, B., Chibi, M., Xie, P., Koundrioukoff, S., Moolman-Smook, H., Pugh, D., Debatisse, M., He, F., Zhang, L., and Defossez, P.A. (2014). The RBBP6/ZBTB38/MCM10 axis regulates DNA replication and common fragile site stability. *Cell Rep* 7, 575-587.
- Mitchell, P. (2014). Exosome substrate targeting: the long and short of it. *Biochem Soc Trans* 42, 1129-1134.
- Moela, P., and Motadi, L.R. (2016). RBBP6: a potential biomarker of apoptosis induction in human cervical cancer cell lines. *Onco Targets Ther* 9, 4721-4735.
- Moreau, P., Hen, R., Wasyluk, B., Everett, R., Gaub, M.P., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* 9, 6047-6068.
- Morisaki, T., Yashiro, M., Kakehashi, A., Inagaki, A., Kinoshita, H., Fukuoka, T., Kasashima, H., Masuda, G., Sakurai, K., Kubo, N., *et al.* (2014). Comparative Proteomics Analysis of Gastric Cancer Stem Cells. *Plos One* 9.

Motadi, L.R., Bhoola, K.D., and Dlamini, Z. (2011). Expression and function of retinoblastoma binding protein 6 (RBBP6) in human lung cancer. *Immunobiology* 216, 1065-1073.

Motadi, L.R., Lekganyane, M.M., and Moela, P. (2018). RBBP6 expressional effects on cell proliferation and apoptosis in breast cancer cell lines with distinct p53 statuses. *Cancer Manag Res* 10, 3357-3369.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* 51, 606-617.

Neve, J., Patel, R., Wang, Z., Louey, A., and Furger, A.M. (2017). Cleavage and polyadenylation: Ending the message expands gene regulation. *RNA Biol* 14, 865-890.

Ni, L., Llewellyn, R., Kesler, C.T., Kelley, J.B., Spencer, A., Snow, C.J., Shank, L., and Paschal, B.M. (2013). Androgen induces a switch from cytoplasmic retention to nuclear import of the androgen receptor. *Mol Cell Biol* 33, 4766-4778.

Ogami, K., Chen, Y., and Manley, J.L. (2018). RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Noncoding RNA* 4.

Ogami, K., Cho, R., and Hoshino, S. (2013). Molecular cloning and characterization of a novel isoform of the non-canonical poly(A) polymerase PAPD7. *Biochem Biophys Res Commun* 432, 135-140.

Ogami, K., and Manley, J.L. (2017). Mtr4/ZFC3H1 protects polysomes through nuclear RNA surveillance. *Cell Cycle* 16, 1999-2000.

Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J.J., Yates, J.R., 3rd, Tian, B., and Manley, J.L. (2017). An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev*.

Onderak, A.M., and Anderson, J.T. (2017). Loss of the RNA helicase SKIV2L2 impairs mitotic progression and replication-dependent histone mRNA turnover in murine cell lines. *RNA* 23, 910-926.

Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Program, N.C.S., *et al.* (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A* 110, 17921-17926.

Passmore, L.A., and Barford, D. (2004). Getting into position: the catalytic mechanisms of protein ubiquitylation. *Biochem J* 379, 513-525.

Pauli, A., Valen, E., and Schier, A.F. (2015). Identifying (non-)coding RNAs and small peptides: challenges and opportunities. *Bioessays* 37, 103-112.

- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.P., *et al.* (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* *161*, 774-789.
- Perez Canadillas, J.M., and Varani, G. (2003). Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J* *22*, 2821-2830.
- Pnueli, L., Rudnizky, S., Yosefzon, Y., and Melamed, P. (2015). RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin alpha-subunit gene. *Proc Natl Acad Sci U S A* *112*, 4369-4374.
- Preker, P., Almvig, K., Christensen, M.S., Valen, E., Mapendano, C.K., Sandelin, A., and Jensen, T.H. (2011). PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res* *39*, 7179-7193.
- Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science* *322*, 1851-1854.
- Pugh, D.J., Ab, E., Faro, A., Luty, P.T., Hoffmann, E., and Rees, D.J. (2006). DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways. *BMC Struct Biol* *6*, 1.
- Puno, M.R., and Lima, C.D. (2018). Structural basis for MTR4-ZCCHC8 interactions that stimulate the MTR4 helicase in the nuclear exosome-targeting complex. *Proc Natl Acad Sci U S A*.
- Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K.R., Resch, W., Liang, G., Tang, Z., Mathe, E., Benner, C., *et al.* (2014). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* *159*, 1524-1537.
- Rammelt, C., Bilen, B., Zavolan, M., and Keller, W. (2011). PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *RNA* *17*, 1737-1746.
- Rassa, J.C., Wilson, G.M., Brewer, G.A., and Parks, G.D. (2000). Spacing constraints on reinitiation of paramyxovirus transcription: the gene end U tract acts as a spacer to separate gene end from gene start sites. *Virology* *274*, 438-449.
- Richard, P., Ogami, K., Chen, Y., Feng, S., Moresco, J.J., Yates, J.R., 3rd, and Manley, J.L. (2018). NRDE-2, the human homolog of fission yeast Nrl1, prevents DNA damage accumulation in human cells. *RNA Biol*.
- Rickels, R., and Shilatifard, A. (2018). Enhancer Logic and Mechanics in Development and Disease. *Trends Cell Biol* *28*, 608-630.
- Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. *Cell* *155*, 39-55.

- Rogan, D.F., Cousins, D.J., Santangelo, S., Ioannou, P.A., Antoniou, M., Lee, T.H., and Staynov, D.Z. (2004). Analysis of intergenic transcription in the human IL-4/IL-13 gene cluster. *Proc Natl Acad Sci U S A* *101*, 2446-2451.
- Rothschild, G., and Basu, U. (2017). Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability. *Trends in Genetics* *33*, 143-154.
- Ruiz-Orera, J., and Alba, M.M. (2019). Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. *Trends Genet* *35*, 186-198.
- Ryner, L.C., Takagaki, Y., and Manley, J.L. (1989). Multiple forms of poly(A) polymerases purified from HeLa cells function in specific mRNA 3'-end formation. *Mol Cell Biol* *9*, 4229-4238.
- Sadowski, M., Suryadinata, R., Tan, A.R., Roesley, S.N.A., and Sarcevic, B. (2012). Protein monoubiquitination and polyubiquitination generate structural diversity to control distinct biological processes. *Iubmb Life* *64*, 136-142.
- Saijo, M., Sakai, Y., Kishino, T., Niikawa, N., Matsuura, Y., Morino, K., Tama, K., and Taya, Y. (1995). Molecular-Cloning of a Human Protein That Binds to the Retinoblastoma Protein and Chromosomal Mapping. *Genomics* *27*, 511-519.
- Sakai, Y., Saijo, M., Coelho, K., Kishino, T., Niikawa, N., and Taya, Y. (1995). cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product. *Genomics* *30*, 98-101.
- Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. *Nat Rev Genet* *16*, 583-597.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* *489*, 109-113.
- Schaffner, W. (2015). Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol Chem* *396*, 311-327.
- Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. *Mol Cell* *56*, 29-42.
- Schmid, M., and Jensen, T.H. (2018). Controlling nuclear RNA levels. *Nat Rev Genet* *19*, 518-529.
- Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Mol Cell* *58*, 1101-1112.
- Serfling, E., Jasin, M., and Schaffner, W. (1985). Enhancers and Eukaryotic Gene-Transcription. *Trends in Genetics* *1*, 224-230.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* *33*, 365-376.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* *15*, 272-286.

Sigismund, S., Polo, S., and Di Fiore, P.P. (2004). Signaling through monoubiquitination. *Curr Top Microbiol* *286*, 149-185.

Silla, T., Karadoulama, E., Makosa, D., Lubas, M., and Jensen, T.H. (2018). The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts. *Cell Rep* *23*, 2199-2210.

Simons, A., MelamedBessudo, C., Wolkowicz, R., Sperling, J., Sperling, R., Eisenbach, L., and Rotter, V. (1997). PACT: Cloning and characterization of a cellular p53 binding protein that interacts with Rb. *Oncogene* *14*, 145-155.

Singh, R.K., Zerath, S., Kleifeld, O., Scheffner, M., Glickman, M.H., and Fushman, D. (2012). Recognition and cleavage of related to ubiquitin 1 (Rub1) and Rub1-ubiquitin chains by components of the ubiquitin-proteasome system. *Mol Cell Proteomics* *11*, 1595-1611.

Skaar, J.R., Ferris, A.L., Wu, X., Saraf, A., Khanna, K.K., Florens, L., Washburn, M.P., Hughes, S.H., and Pagano, M. (2015). The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res* *25*, 288-305.

Smith, E., and Shilatifard, A. (2014). Enhancer biology and enhanceropathies. *Nature Structural & Molecular Biology* *21*, 210-219.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* *13*, 613-626.

Stadlmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S., *et al.* (2014). Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat Commun* *5*, 5531.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* *14*, 103-105.

Swatek, K.N., and Komander, D. (2016). Ubiquitin modifications. *Cell Res* *26*, 399-422.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1988). Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* *52*, 731-742.

Tautz, D., and Domazet-Lošo, T. (2011). The evolutionary origin of orphan genes. *Nat Rev Genet*

12, 692-702.

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 455, 1124-1128.

Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. *Embo Journal* 19, 94-102.

Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* 33, 201-212.

Tian, B., and Manley, J.L. (2013). Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* 38, 312-320.

Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* 18, 18-30.

Trotman, L.C., Mosberger, N., Fornerod, M., Stidwill, R.P., and Greber, U.F. (2001). Import of adenovirus DNA involves the nuclear pore complex receptor CAN/Nup214 and histone H1. *Nat Cell Biol* 3, 1092-1100.

Trotman, L.C., Wang, X., Alimonti, A., Chen, Z., Teruya-Feldstein, J., Yang, H., Pavletich, N.P., Carver, B.S., Cordon-Cardo, C., Erdjument-Bromage, H., *et al.* (2007). Ubiquitination regulates PTEN nuclear import and tumor suppression. *Cell* 128, 141-156.

Tuan, D., Kong, S., and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A* 89, 11219-11223.

Vahedi, G., Kanno, Y., Furumoto, Y., Jiang, K., Parker, S.C., Erdos, M.R., Davis, S.R., Roychoudhuri, R., Restifo, N.P., Gadina, M., *et al.* (2015). Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* 520, 558-562.

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among Distant Regulatory Elements at the beta-Globin Locus Requires GATA-1 and FOG-1. *Molecular Cell* 17, 453-462.

Varshavsky, A. (2012). The ubiquitin system, an immense realm. *Annu Rev Biochem* 81, 167-176.

Vihervaara, A., Mahat, D.B., Guertin, M.J., Chu, T.Y., Danko, C.G., Lis, J.T., and Sistonon, L. (2017). Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nature Communications* 8.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., *et al.* (2015). Enhancer evolution across 20 mammalian species. *Cell* 160, 554-566.

- Vo, L.T.A., Minet, M., Schmitter, J.M., Lacroute, F., and Wyers, F. (2001). Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. *Molecular and Cellular Biology* 21, 8346-8356.
- Wahle, E. (1995). Poly(A) tail length control is caused by termination of processive synthesis. *J Biol Chem* 270, 2800-2808.
- Wang, J., Chen, J., Wu, G., Zhang, H., Du, X., Chen, S., Zhang, L., Wang, K., Fan, J., Gao, S., *et al.* (2019). NRDE2 negatively regulates exosome functions by inhibiting MTR4 recruitment and exosome interaction. *Genes Dev.*
- Wang, Q., Carroll, J.S., and Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol Cell* 19, 631-642.
- Wang, R., Zheng, D., Yehia, G., and Tian, B. (2018). A compendium of conserved cleavage and polyadenylation events in mammalian genes. *Genome Res* 28, 1427-1441.
- Weill, L., Belloc, E., Bava, F.A., and Mendez, R. (2012). Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat Struct Mol Biol* 19, 577-585.
- Whalen, S., Truty, R.M., and Pollard, K.S. (2016). Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* 48, 488-496.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307-319.
- Williamson, I., Berlivet, S., Eskeland, R., Boyle, S., Illingworth, R.S., Paquette, D., Dostie, J., and Bickmore, W.A. (2014). Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev* 28, 2778-2791.
- Yamanaka, S., Mehta, S., Reyes-Turcu, F.E., Zhuang, F., Fuchs, R.T., Rong, Y., Robb, G.B., and Grewal, S.I. (2013). RNAi triggered by specialized machinery silences developmental genes and retrotransposons. *Nature* 493, 557-560.
- Yamazaki, T., Liu, L., Lazarev, D., Al-Zain, A., Fomin, V., Yeung, P.L., Chambers, S.M., Lu, C.W., Studer, L., and Manley, J.L. (2018). TCF3 alternative splicing controlled by hnRNP H/F regulates E-cadherin expression and hESC pluripotency. *Genes Dev* 32, 1161-1174.
- Yang, Q., Gilmartin, G.M., and Doublet, S. (2011). The structure of human cleavage factor I(m) hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5' capping and splicing. *RNA Biol* 8, 748-753.
- Yoshitake, Y., Nakatsura, T., Monji, M., Senju, S., Matsuyoshi, H., Tsukamoto, H., Hosaka, S., Komori, H., Fukuma, D., Ikuta, Y., *et al.* (2004). Proliferation potential-related protein, an ideal

esophageal cancer antigen for immunotherapy, identified using complementary DNA microarray analysis. *Clinical Cancer Research* *10*, 6437-6448.

Young, R.S., Hayashizaki, Y., Andersson, R., Sandelin, A., Kawaji, H., Itoh, M., Lassmann, T., Carninci, P., Consortium, F., Bickmore, W.A., *et al.* (2015). The frequent evolutionary birth and death of functional promoters in mouse and human. *Genome Res* *25*, 1546-1557.

Zhang, Y., Wong, C.H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., *et al.* (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* *504*, 306-310.

Zheng, D., Liu, X., and Tian, B. (2016). 3'READS+, a sensitive and accurate method for 3' end sequencing of polyadenylated RNA. *RNA* *22*, 1631-1639.

Zheng, D., and Tian, B. (2017). Polyadenylation Site-Based Analysis of Transcript Expression by 3'READS. *Methods Mol Biol* *1648*, 65-77.

Zhu, Y., Sun, L., Chen, Z., Whitaker, J.W., Wang, T., and Wang, W. (2013). Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res* *41*, 10032-10043.

Zinder, J.C., and Lima, C.D. (2017). Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* *31*, 88-100.

Appendix I. Protein sequence of RBBP6N-FLAG and ΔDWNN-FLAG

Protein sequence of RBBP6N-FLAG

MSCVHYKFSSKLNVDYTFDGLHISLCDLKKQIMGREKLKAADCDLQITNAQTKEEYTD
DNALIPKNSSVIVRRIPIGGVKSTSKTYVISRTEPAMATTKAIDDSSASISLAQLTKTANLA
EANASEEDKIKAMMSQSGHEYDPINYMKKPLGPPPPSYTCFRCGKPGHYIKNCPTNGDK
NFESGPRIKKSTGIPRSFMMEVKDPNMKGAMLTNTGKYAIPTIDAEAYAIGKKEKPPFLP
EEPSSSSEEDDPIPDELLCLICKDIMTDAVVIPCCGNSYCDECIRTALLESEHTCPTCHQN
DVSPDALIANKFLRQAVNNFKNETGYTKRLRKQLPPPPPIPPPRPLIQRNLQPLMRSPISR
QQDPLMIPVTSSSTHPAPSISSLTSNQSSLAPPVSGNPSSAPAPVPDITATVSISVHSEKSDG
PFRSDNKILPAAALASEHSGTSSIAITALMEEKGYQVPVLGTPSLLGQSLLHGQLIPTT
GPVRINTARPGGGRPGWEHSNKLGYLVSPQQIRRSRGSRA**DYKDHDGDYKDHDIDY**

KDDDDK (FLAG tag is highlighted in red)

Protein sequence of ΔDWNN-FLAG

IGGVKSTSKTYVISRTEPAMATTKAIDDSSASISLAQLTKTANLAEANASEEDKIKAMMS
QSGHEYDPINYMKKPLGPPPPSYTCFRCGKPGHYIKNCPTNGDKNFESGPRIKKSTGIPRS
FMMEVKDPNMKGAMLTNTGKYAIPTIDAEAYAIGKKEKPPFLPEEPSSSSEEDDPIPDEL
LCLICKDIMTDAVVIPCCGNSYCDECIRTALLESEHTCPTCHQNDVSPDALIANKFLRQ
AVNNFKNETGYTKRLRKQLPPPPPIPPPRPLIQRNLQPLMRSPISRQQDPLMIPVTSSSTH
PAPSISSLTSNQSSLAPPVSGNPSSAPAPVPDITATVSISVHSEKSDGPFRSDNKILPAAAL
ASEHSGTSSIAITALMEEKGYQVPVLGTPSLLGQSLLHGQLIPTTGPVRINTARPGGGRP
GWEHSNKLGYLVSPQQIRRSRGSRA**DYKDHDGDYKDHDIDYKDDDDK**

**Appendix II. An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs
to prevent their cytoplasmic transport and global translational repression**

Yaqiong Chen performed experiments and analyzed data in Figure 5 and Figure S5

An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression

Koichi Ogami,^{1,4} Patricia Richard,¹ Yaqiong Chen,¹ Mainul Hoque,² Wencheng Li,² James J. Moresco,³ John R. Yates III,³ Bin Tian,² and James L. Manley¹

¹Department of Biological Sciences, Columbia University, New York, New York 10027, USA; ²Department of Microbiology, Biochemistry, and Molecular Genetics, Rutgers New Jersey Medical School, Newark, New Jersey 07103, USA; ³Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA

Many long noncoding RNAs (lncRNAs) are unstable and rapidly degraded in the nucleus by the nuclear exosome. An exosome adaptor complex called NEXT (nuclear exosome targeting) functions to facilitate turnover of some of these lncRNAs. Here we show that knockdown of one NEXT subunit, Mtr4, but neither of the other two subunits, resulted in accumulation of two types of lncRNAs: prematurely terminated RNAs (ptRNAs) and upstream antisense RNAs (uaRNAs). This suggested a NEXT-independent Mtr4 function, and, consistent with this, we isolated a distinct complex containing Mtr4 and the zinc finger protein ZFC3H1. Strikingly, knockdown of either protein not only increased pt/uaRNA levels but also led to their accumulation in the cytoplasm. Furthermore, all pt/uaRNAs examined associated with active ribosomes, but, paradoxically, this correlated with a global reduction in heavy polysomes and overall repression of translation. Our findings highlight a critical role for Mtr4/ZFC3H1 in nuclear surveillance of naturally unstable lncRNAs to prevent their accumulation, transport to the cytoplasm, and resultant disruption of protein synthesis.

[*Keywords:* Mtr4; ZFC3H1; exosome; lncRNA; polyadenylation]

Supplemental material is available for this article.

Received May 25, 2017; revised version accepted June 22, 2017.

RNA polymerase II is responsible for production of a large repertoire of RNAs. In addition to mRNAs, these include a variety of functional, relatively stable RNAs, such as small nuclear RNAs (snRNAs) and microRNAs (miRNAs), and a diverse set of long noncoding RNAs (lncRNAs). Many of these lncRNAs have well-documented functions in either the cytoplasm or nucleus and are also often quite stable (Chen 2016). However, a large number of lncRNAs have no known functions and can be very unstable and rapidly degraded in the nucleus. These include, for example, RNAs that are transcribed upstream of protein-coding gene promoters (Preker et al. 2008; Flynn et al. 2011) as well as RNAs transcribed, frequently bidirectionally, from transcriptional enhancers (Djebali et al. 2012; Andersson et al. 2014a). Synthesis of these RNAs often involves multiple RNA processing reactions, which are typically closely linked to their transcription.

Such processing invariably involves 5' capping, frequently splicing as well as 3' end formation, and often cleavage and polyadenylation. It is noteworthy that polyadenylation can have either a stabilizing effect, as typically observed with mRNAs, or a destabilizing effect, as found with many nuclear lncRNAs subject to rapid decay (Beaulieu et al. 2012; Ntini et al. 2013; Bresson et al. 2015).

Two interesting classes of lncRNAs are the promoter upstream transcripts (PROMPTs) and prematurely terminated RNAs (ptRNAs). PROMPTs are transcribed in both sense and antisense directions relative to transcription start sites of protein-coding genes and are polyadenylated (Preker et al. 2008). Notably, PROMPTs that are transcribed in the antisense direction, which we specifically refer to as upstream antisense RNAs (uaRNAs) (Flynn et al. 2011), appear to be processed by mechanisms of pre-mRNA 3' end formation that are the same as or

⁴Present address: RIKEN Center for Life Science Technologies, Post-Transcriptional Control Research Unit, Tsurumi, Yokohama 230-0045, Japan. Corresponding author: jlm2@columbia.edu
Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.302604.117>.

© 2017 Ogami et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License [Attribution-NonCommercial 4.0 International], as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Ogami et al.

similar to those of mRNAs (Almada et al. 2013; Ntini et al. 2013; for review, see Richard and Manley 2013). Motifs similar to those that constitute poly(A) sites (PASs) in pre-mRNAs are found at the 3' ends of uaRNAs, and much of the same complex protein machinery that is responsible for mRNA polyadenylation (Tian and Manley 2017) functions in uaRNA 3' end formation (Ntini et al. 2013). Notably, PASs are more enriched in the upstream antisense region compared with the downstream region, whereas U1 snRNA-binding sites (which, when bound by U1 snRNP, prevent polyadenylation at nearby PASs) (Almada et al. 2013; Ntini et al. 2013) are more frequent in the sense-coding direction. Failure of the suppression of proximal PASs results in premature cleavage and polyadenylation, giving rise to ptRNAs (Kaida et al. 2010; Berg et al. 2012). PAS-driven early termination of pt/uaRNAs is linked to rapid degradation of these RNAs by the nuclear exosome despite the fact that the RNA signals and protein factors are very similar to those used for relatively stable mRNAs (Ntini et al. 2013).

What targets unstable nuclear RNAs for rapid turnover? An interesting candidate that might link PAS-mediated 3' processing of lncRNAs to degradation is the exosome adaptor complex NEXT (nuclear exosome targeting), which consists of the RNA helicase Mtr4, the RNA-binding protein RBM7, and the zinc knuckle protein ZCCHC8 (Lubas et al. 2011). NEXT physically associates with the nuclear exosome to facilitate turnover of various types of RNA substrates, including PROMPTs, enhancer RNAs (eRNAs), 3' extended snRNAs/small nucleolar RNAs (snoRNAs), and replication-dependent histone-encoding transcripts (Lubas et al. 2011; Andersen et al. 2013). Mtr4, which shows a higher affinity for poly(A) relative to non-poly(A) RNA (Bernstein et al. 2008), plays an essential role in exosome activation, presumably by RNA unwinding and/or feeding RNA substrates into the exosome (Wang et al. 2008; Houseley and Tollervey 2009). The substrate recognition activity of NEXT is conferred by RBM7, which shows some preference for U-rich sequences (Andersen et al. 2013). Intriguingly, all three NEXT subunits were identified in a proteomics analysis of the purified human pre-mRNA polyadenylation complex (Shi et al. 2009), suggesting a possible role for NEXT coupled to 3' processing. Human cells possess another exosome adaptor complex, TRAMP (Trf4–Air–Mtr4 polyadenylation complex), comprising Mtr4, the noncanonical poly(A) polymerase PAPD5, and the zinc knuckle protein ZCCHC7. TRAMP substrates, however, are thought to be restricted to nucleolar RNA targets such as ribosomal RNA (rRNA) precursors (Lubas et al. 2011).

In this study, we analyzed the polyadenylated transcriptomes of cells depleted of the individual NEXT subunits to investigate whether NEXT might affect polyadenylation generally. Unexpectedly, we found that siRNA-mediated knockdown of Mtr4, but neither of the other two NEXT subunits or a TRAMP subunit, resulted in strong and specific accumulation of ptRNAs and uaRNAs. We then examined Mtr4 complexes and interacting proteins by gel filtration followed by mass spectrometry (MS) and identified the zinc finger protein ZFC3H1 as an Mtr4 part-

ner (see also Meola et al. 2016) also necessary for degradation of ptRNAs and uaRNAs. Unexpectedly, knockdown of Mtr4 or ZFC3H1 resulted in accumulation of polyadenylated Mtr4 targets in the cytoplasm, and the exported RNAs were bound by ribosomes. Mtr4/ZFC3H1-deficient cells also showed a surprising global reduction in heavier polysomes, suggesting that ribosomes naturally bound to mRNAs were occupied by the short ORF-containing pt/uaRNAs. Consistent with this, the Mtr4/ZFC3H1-depleted cells showed a significant specific inhibition of translation. Our findings illustrate the importance of nuclear surveillance of polyadenylated lncRNAs by Mtr4/ZFC3H1 to prevent the unwanted and deleterious transport to and accumulation of these RNAs in the cytoplasm.

Results

Mtr4 prevents accumulation of ptRNAs and uaRNAs

The initial aim of our experiments was to investigate the significance of the previously observed association of NEXT with the pre-mRNA polyadenylation machinery. To this end, we depleted each of the three NEXT subunits from HeLa cells with siRNAs (Fig. 1A,B) and performed 3' region extraction and deep sequencing (3'READS) (Hoque et al. 2013) to analyze the global effects on accumulation of polyadenylated RNAs. Unexpectedly, depletion of Mtr4, but neither of the other two NEXT subunits (RBM7 or ZCCHC8), resulted in preferential accumulation of two types of transcripts: uaRNAs and ptRNAs (Fig. 1C,D). Consistent with this, metagene plots revealed that the Mtr4-depleted cells showed a sharp increase of promoter-proximal PAS usage in both sense and antisense directions, whereas depletion of RBM7 and ZCCHC8 had only minimal effects (Fig. 1E). An increase of reads corresponding to intronic PASs was observed in Mtr4 knockdown cells, while those corresponding to the 3'-most PASs, reflecting full-length mRNAs, were not affected by knockdown (Supplemental Fig. S1A), suggesting that the change was due to ptRNA stabilization as opposed to increased 3' processing at the intronic PAS.

We next performed RT-qPCR [oligo(dT)-primed RT and quantitative PCR (qPCR)] to confirm and extend the results of 3'READS. To distinguish ptRNAs and full-length mRNAs, primers were designed as shown in Figure 1G. Consistent with the 3'READS data, levels of both ptRNAs and uaRNAs were increased by Mtr4 knockdown but not by knockdown of RBM7 or ZCCHC8 (Fig. 1F,H). In contrast to the uaRNAs, increased levels of two known NEXT substrates—*RBM39* and *FBXO7* PROMPTs (proRBM39 and proFBXO7, respectively) (Lubas et al. 2011)—were observed in all knockdown cells (Fig. 1H). Similar results were obtained using a second Mtr4 siRNA (Supplemental Fig. S1B,C). RT-qPCR also confirmed that there were no changes in full-length mRNA levels for genes that displayed elevated ptRNA levels (Fig. 1F, FL mRNA). The other known Mtr4-containing complex, TRAMP, appears not to be involved in degradation of ptRNAs and uaRNAs, as there were no significant changes of the Mtr4 target RNAs after ZCCHC7 knockdown

Mtr4/ZCCH3H1 protects polysomes

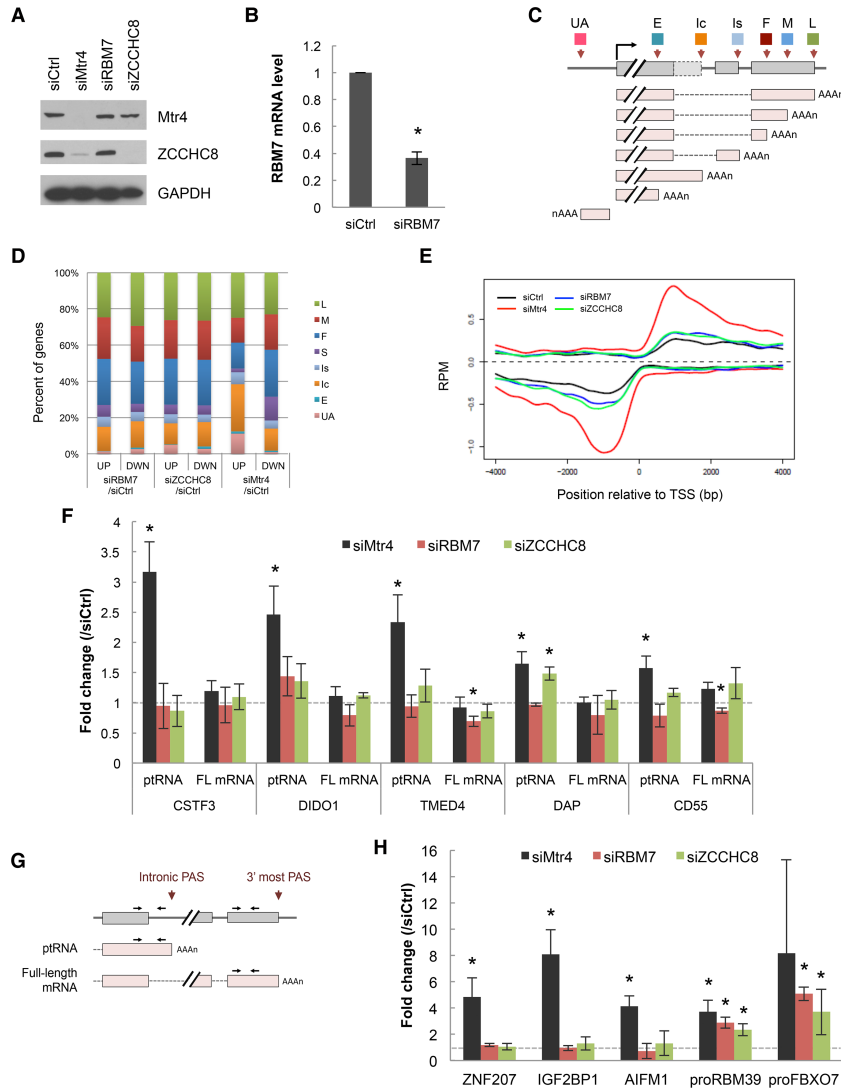


Figure 1. Global analysis of poly(A)⁺ transcript levels following depletion of individual NEXT subunits. *(A)* Western blot analysis of HeLa cell extracts after 48 h of transfection of control (Ctrl), Mtr4, RBM7, or ZCCHC8 siRNA. *(B)* RBM7 mRNA level after 48 h of siRBM7 treatment. RBM7 mRNA was normalized to GAPDH mRNA, and the normalized levels in siCtrl-treated cells were set to 1. Bars represent mean \pm SD. $n = 3$. An asterisk denotes significant difference from siCtrl ($P < 0.05$) using an unpaired Student's t -test. *(C)* Schematic of different transcript types analyzed: transcripts using the first (F), middle (M), or last (L) potential PAS in the 3' untranslated region (UTR); the single (S); no 3' UTR APA PAS in the 3' UTR; the intronic PAS in the composite terminal exon (Ic); the intronic PAS in the skipped terminal exon (Is); the upstream (not 3'-most) exonic PAS (E); and the upstream antisense transcripts (UA). *(D)* Changes in relative abundance of the indicated transcript types following knockdown of indicated NEXT subunits. The percentage of genes showing increases (UP) or decreases (DWN) of each type of transcript are indicated. False discovery rate < 0.05 . *(E)* Metagene plots of ptRNAs and uARNAs. Data are presented as strand-specific reads per million (RPM) at PAS positions within 4 kb upstream of or downstream from the transcription start site. *(F, H)* RT-qPCR [oligo(dT)-primed RT and quantitative PCR (qPCR)] analysis of select ptRNAs and corresponding full-length (FL) mRNAs (*F*) and uARNAs (*H*) after knockdown of the individual NEXT subunits. Analysis of two representative PROMPTs—proRBM39 and proFBXO7—is also shown in *H*. Values were normalized to GAPDH mRNA, and the normalized levels in siCtrl-treated cells were set to 1. Bars represent mean \pm SD. $n = 3$. Asterisks denote significant difference from siCtrl ($P < 0.05$) using an unpaired Student's t -test. *(G)* Diagram of a ptrRNA-producing gene and primers used for RT-qPCR. Arrows indicate the positions of primer targeting sites to analyze ptrRNA and full-length mRNA.

Ogami et al.

(Supplemental Fig. S1D,E). In contrast and as expected, the nuclear exosome is required for ptRNA and uaRNA degradation, as all tested pt/uaRNAs accumulated following codepletion of the two catalytic subunits Rrp6 and Dis3 (Supplemental Fig. S1F,G). Together, these results indicate that the exosome degrades these RNAs in an Mtr4-dependent, but NEXT- and TRAMP-independent, manner.

Identification of Mtr4-interacting proteins

The above results suggested the possible existence of an additional Mtr4-containing protein complex that functions in ptRNA and/or uaRNA turnover. To investigate this, we prepared extracts from HEK293 cells stably expressing N-terminally 3xFlag-tagged Mtr4 (Flag-Mtr4, with expression equivalent to endogenous Mtr4) (Fig.

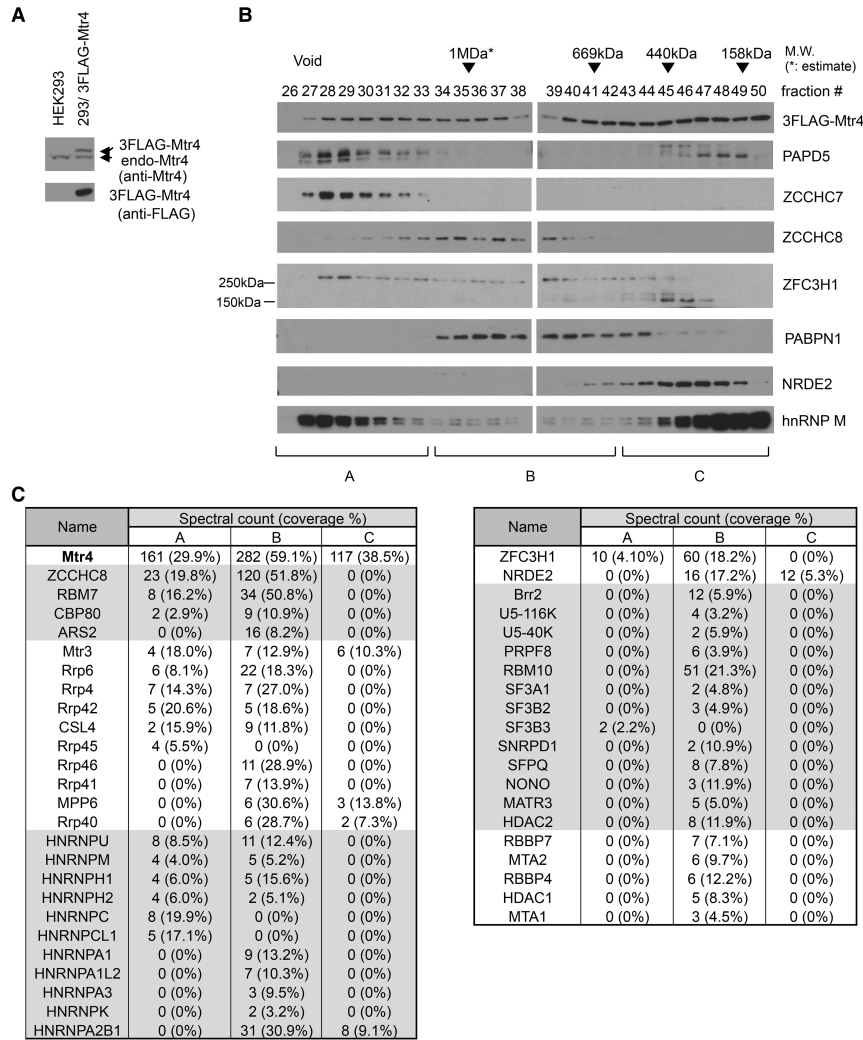


Figure 2. Identification of Mtr4-interacting proteins by cofractionation and MS. (A) Western blotting analysis of HEK293 cells and HEK293 cells stably expressing 3Flag-Mtr4. (Top panel) Blotted with anti-Mtr4 antibodies. (Bottom panel) Blotted with anti-Flag antibodies. (B) Fractions from Superose 6 gel filtration 3Flag-Mtr4-expressing HEK293 cells were analyzed by Western blotting using antibodies against proteins shown at the right. Approximate molecular sizes are indicated at the top, and fractions pooled are indicated at the bottom. (C) Selected proteins copurified with 3Flag-Mtr4 in the indicated pools. Spectral counts and sequence coverage of known Mtr4-interacting partners [NEXT, exosome, and NRDE2], proteins detected as complexes (e.g., NuRD and spliceosome), and RNA processing or RNA-binding proteins are shown. A full protein list is in Supplemental Table S1.

2A) in the presence of RNase A and performed size fractionation using Superose 6 gel filtration chromatography followed by Western blotting (Fig. 2B). Mtr4 was detected in all fractions from the void to <158 kDa in a bimodal distribution with peaks at fraction 31 and at <158 kDa, likely reflecting the existence of multiple Mtr4-containing complexes. The NEXT subunit ZCCHC8 eluted at ~1 MDa, whereas TRAMP subunits ZCCHC7 and PAPD5 appeared mainly in the void fractions.

To identify additional Flag-Mtr4-interacting proteins, we collected three Flag-Mtr4-containing pools (A–C) according to the distribution of TRAMP (ZCCHC7 and PAPD5) and NEXT (ZCCHC8). The pools were subjected to Flag immunoprecipitation, and the coimmunoprecipitated proteins were identified by MS (Fig. 2C; Supplemental Table S1). The subunits of NEXT, the exosome, and cap-binding complex (CBC), which were shown previously to associate with NEXT (Andersen et al. 2013), were detected in pools A and B. Consistent with a previous report (Nag and Steitz 2012), various splicing factors also associated with Flag-Mtr4 in pool B. Pool B also contained NuRD complex subunits, which function in histone

modification (Xue et al. 1998; Zhang et al. 1998). We also found several heterogeneous nuclear ribonucleoproteins (hnRNPs)—PSF/SFPQ and p54nrb/NONO—in pool B as well as two other possibly relevant proteins: NRDE2 and ZFC3H1. Although the role of NRDE2 in human cells is unknown, the fission yeast homolog Nrl1 interacts physically with an Mtr4-like protein, Mtl1 (Lee et al. 2013; Aronica et al. 2016), and is involved in suppression of R-loop formation (Aronica et al. 2016). ZFC3H1 is a large (~230-kDa) protein localized in the nucleus and shown to modulate *IL-8* transcription (Tomita et al. 2014). ZFC3H1 is the apparent homolog of the *Schizosaccharomyces pombe* protein Red1, which functions in a distinct Mtl1-containing complex and, significantly, plays essential roles in the degradation of various unstable RNAs (Lee et al. 2013; Egan et al. 2014; Zhou et al. 2015).

We next performed coimmunoprecipitation (co-IP) experiments to verify several of the interactions suggested by the above data. We validated RNase-resistant interactions of Mtr4 with ZFC3H1 (Fig. 3A,B), NRDE2, U5-40K, and hnRNP M (Supplemental Fig. S2). (Note that ZFC3H1 appears as two major bands of ~250 and 150

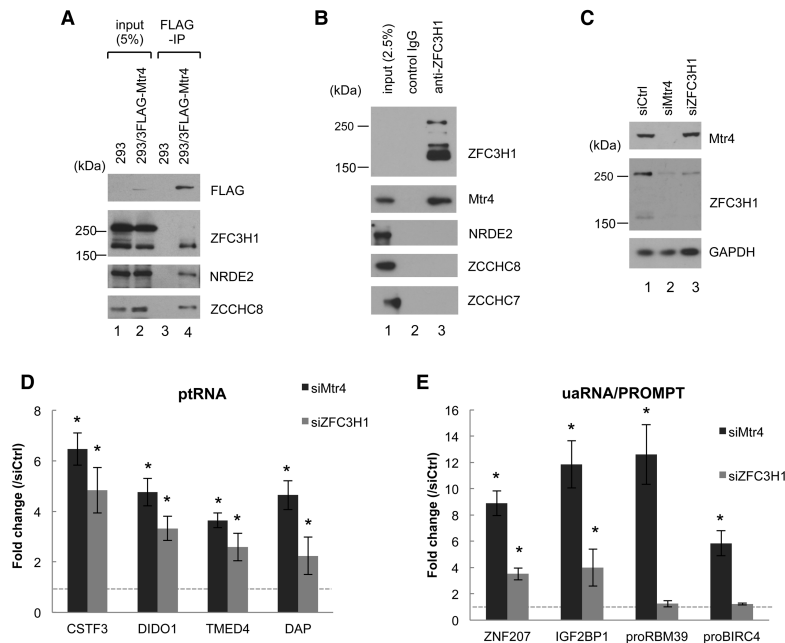


Figure 3. Mtr4-associated ZFC3H1 is required for down-regulation of ptRNAs and uRNAs but not NEXT substrates. (A) Cell extracts prepared from HEK293 cells and HEK293 cells stably expressing 3Flag-Mtr4 were used for immunoprecipitation with anti-Flag antibodies in the presence of benzonase and RNase A followed by Western blotting with the indicated antibodies. (B) Cell extracts prepared from HEK293 cells were used for co-IP experiments with anti-ZFC3H1 in the presence of benzonase and RNase A followed by Western blotting with antibodies against the proteins indicated at the right. (C) Western blot analysis of HeLa cell extracts after 72 h of knockdown treatment with the siRNAs indicated at the top; antibodies against the proteins are indicated at the right. (D,E) RT-qPCR analysis of the indicated ptRNAs (D) and the indicated uRNAs and NEXT substrates proRBM39 and proBIRC4 (E) after the indicated siRNA transfections. Transcript levels were normalized to GAPDH mRNA, and the normalized levels in siCtrl-treated cells were set to 1. Bars represent mean \pm SD. $n = 3$. Asterisks denote significant difference from siCtrl ($P < 0.05$) using an unpaired Student's *t*-test.

Ogami et al.

kDa, both of which were decreased by ZFC3H1 knockdown [Figs. 3C, 6A]. The 150-kDa isoform was more efficiently immunoprecipitated with Mtr4 or ZFC3H1 antibodies [Fig. 3A], which may reflect limited epitope accessibility in the 250-kDa ZFC3H1–Mtr4 complex. The existence of a 250-kDa ZFC3H1–Mtr4 complex is supported by the observation that Mtr4 knockdown caused decreases in both the 250- and 150-kDa isoforms, likely reflecting protein destabilization [Figs. 3C, 6A]. The origin of the smaller species remains to be determined.) Importantly, neither TRAMP nor NEXT subunits were coimmunoprecipitated with ZFC3H1 or NRDE2 (Fig. 3B; Supplemental Fig. S2B,C), indicating that these proteins form complexes that are distinct from NEXT and TRAMP. In contrast, hnRNP M and U5-40K coimmunoprecipitated with ZCCHC7 and PAPP5 but not with ZCCHC8, suggesting that Mtr4 interacts with these proteins in the context of TRAMP (Supplemental Fig. S2A).

ZFC3H1 is required for repression of ptRNAs and uaRNAs but not NEXT substrates

Among the verified Mtr4-interacting partners described above, we decided to focus on ZFC3H1. (While this work was in progress, Meola et al. [2016] also identified ZFC3H1 as an Mtr4-interacting protein.) To determine whether ZFC3H1, like Mtr4, is required for ptRNA and uaRNA turnover, we depleted ZFC3H1 and evaluated the accumulation of these RNAs by RT-qPCR (Fig. 3C–E). Importantly, as with Mtr4 knockdown, both ptRNA and uaRNA levels were increased by ZFC3H1 knockdown (Fig. 3D,E). In contrast, ZFC3H1 knockdown, unlike NEXT knockdown, had no significant effect on the levels of two PROMPTs, proRBM39 and proBIRC4, indicating that ZFC3H1 and NEXT target distinct sets of RNA substrates (Fig. 3E).

Another protein that might be involved in pt/uaRNA degradation is the nuclear poly(A)-binding protein (PABPN1). PABPN1 interacts physically with the nuclear exosome to degrade subsets of polyadenylated lncRNA species (Beaulieu et al. 2012), and these targets include both ptRNAs (Li et al. 2015) and uaRNAs (Bresson et al. 2015; Li et al. 2015). Despite the absence of PABPN1 in our Flag-Mtr4 co-IP/MS (Fig. 2; Supplemental Table S1) and co-IP/Western blot analyses (Supplemental Fig. S3A, B), we next investigated whether PABPN1 functions similarly to Mtr4/ZFC3H1. RT-qPCR results indicate that Mtr4/ZFC3H1 and PABPN1 share RNA substrates (Fig. 3C–E; Supplemental Fig. S3C–E), although an exception was uaGF2BP1, which was sensitive only to Mtr4/ZFC3H1 knockdown (Fig. 3E; Supplemental Fig. S3E). As with Mtr4/ZFC3H1 knockdown, PABPN1 knockdown had only minimal effect on proRBM39 and proBIRC4 levels (Supplemental Fig. S3E).

Depletion of Mtr4/ZFC3H1 causes cytoplasmic accumulation of ptRNAs and uaRNAs

We next investigated the consequences of the increased accumulation of ptRNAs and uaRNAs caused by Mtr4/

ZFC3H1 knockdown. Given that addition of a poly(A) tail can be sufficient to target mRNAs for nuclear export (Brodsky and Silver 2000; Fuke and Ohno 2008), we first investigated whether there were any changes in subcellular localization of these RNAs in knockdown cells. To this end, we separated cell compartments into cytoplasmic, nuclear-soluble, and chromatin fractions. The effectiveness of our fractionation protocol was verified by Western blotting: GAPDH, U2AF65, and histone H3 were detected predominantly in the cytoplasmic, nuclear-soluble, and chromatin fractions, respectively (Fig. 4A–C). Mtr4 was predominantly nuclear and evenly distributed in the nuclear-soluble and chromatin fractions (Fig. 4A,B), while ZFC3H1 was almost entirely in the chromatin fraction (Fig. 4C). NEXT subunit ZCCHC8 and TRAMP subunits ZCCHC7 and PAPP5 displayed different distributions; ZCCHC8 was largely nuclear-soluble, whereas ZCCHC7 and PAPP5 were exclusively in the chromatin fraction (Fig. 4A). PABPN1 was predominantly in the two nuclear fractions (Fig. 4A).

To investigate ptRNA and uaRNA localization, we analyzed by RT-PCR RNA from fractions prepared as above from cells treated with Ctrl, Mtr4, or ZFC3H1 siRNAs (Fig. 4B,C). Specifically, RNA from the cytoplasmic (Fig. 4B,C, lanes 1,2), nuclear-soluble (Fig. 4B,C, lanes 3,4), and chromatin (Fig. 4B,C, lanes 5,6) fractions of knockdown cells was analyzed. Effective fractionation was verified by Western blotting of GAPDH, U2AF65, and histone H3 as above as well as by RT-PCR of a cytoplasmic lncRNA (RPPH1) and a nuclear-insoluble lncRNA (NEAT1). Moreover, unspliced forms of multiexonic ptRNAs (CSTF3 and TMED4) were enriched in the chromatin fraction, further demonstrating the validity of the fractionation. In siCtrl cells, all types of Mtr4/ZFC3H1 targets were most abundant in the chromatin fraction (Fig. 4B,C, cf. lanes 1, 3, and 5). This trend was most pronounced with the single-exonic DAP ptRNA as well as all uaRNAs, which were almost exclusively in the chromatin fraction, whereas multiexonic CSTF3 and TMED4 ptRNAs were detected in all three fractions. Most importantly, however, increased accumulation of all of the ptRNAs and uaRNAs was detected in cytoplasmic and nuclear-soluble fractions after Mtr4 or ZFC3H1 knockdown, while levels in the chromatin fraction were unchanged. These results imply that Mtr4/ZFC3H1 target transcripts, especially those with single exons, are degraded immediately after release from chromatin, and failure of this surveillance system results in significant accumulation of these RNAs in the cytoplasm.

Exported ptRNA and uaRNA associate with ribosomes

We next investigated the fate of the ua/ptRNAs that accumulate in the cytoplasm in the knockdown cells. Notably, all of these RNAs contain putative ORFs, and one possibility therefore is that they are bound by ribosomes and translated. This possibility is supported by the fact that ptRNAs and uaRNAs have very long, ~300-nucleotide poly(A) tails, as determined by RL-PAT assays (Supplemental Fig. S4). To investigate the association of ptRNAs

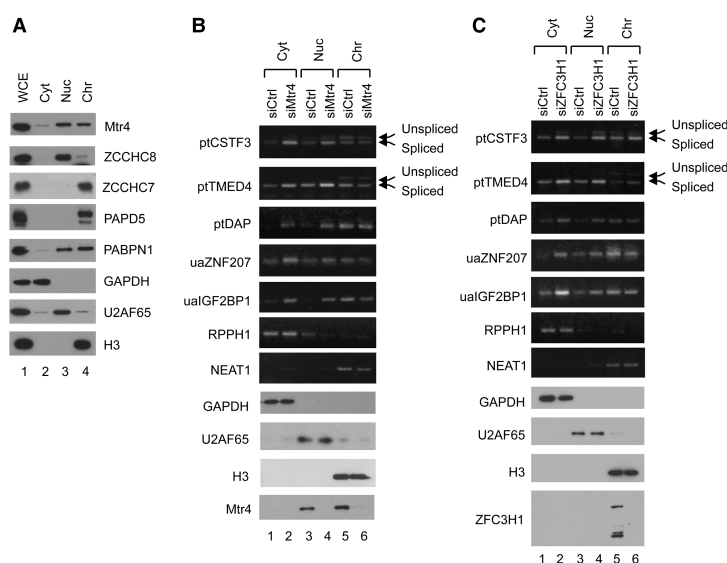


Figure 4. Mtr4 knockdown causes cytoplasmic accumulation of stabilized ptRNAs and uaRNAs. (A) Western blotting of subcellular fractions prepared from HeLa cells. Proteins from whole-cell extract (WCE), cytoplasm (Cyt), nuclear-soluble chromatin (Nuc), and nuclear-insoluble chromatin (Chr) fractions were analyzed using antibodies directed against the proteins listed on the right. (B, C) Subcellular fractionation was performed after 72 h of siMtr4 (B) or siZFC3H1 (C) treatment, and total RNAs were isolated from each fraction, as indicated at the top. cDNA was synthesized using random or oligo(dT) primer, and the indicated transcripts (shown at the left) were analyzed by PCR. Gels were pre-stained with ethidium bromide (EtBr). RPPH1 and NEAT1 RNAs were amplified using random-primed RT products and served as cytoplasmic and nuclear-insoluble markers. Other RNAs were amplified using oligo(dT)-primed RT products.

and uaRNAs with ribosomes, we first performed polysome fractionation by sedimenting cytoplasmic extracts prepared from siRNA-treated cells through 15%–45% sucrose gradients (Fig. 5A) and then evaluated the distribution of individual Mtr4 target transcripts by RT-PCR (Fig. 5B–D; quantitation in Supplemental Fig. S5). In Ctrl siRNA-treated cells, CSTF3 and TMED4 ptRNAs were detected mainly in polysomes, while the DAP ptRNA was found in monosomes to light polysomes (Fig. 5B). All four uaRNAs analyzed (ZNF207, IGF2BP1, IGF2BP3, and MANIA2), although present at only low levels in the cytoplasm, were distributed between free cytosolic to light polysome fractions (Fig. 5C). While Mtr4 knockdown in general did not cause significant changes in the distribution of these RNAs, all of the ptRNAs and uaRNAs that accumulated in the cytoplasm associated with ribosomes, and the absolute amount of ribosome-bound RNA thus increased in all cases. We also analyzed the polysome profile of two full-length mRNAs (GAPDH and CSTF3), and, in contrast to the ptRNAs and uaRNAs, both shifted to lighter fractions following Mtr4 knockdown (Fig. 5D).

We next investigated whether the ptRNAs and uaRNAs were indeed associated with active ribosomes. To this end, we treated cells with the eIF2-GTP-tRNA^{iMet} ternary complex inhibitor BTdCPU, which blocks formation of the preinitiation complex (Chen et al. 2011), and subjected cell extracts to sucrose gradient analysis as above. UV absorption profiles showed a sharp inhibition of translation after 3 h of BTdCPU treatment (Fig. 5A; Supplemental Fig. S6A, +BTdCPU). RT-PCR analysis of the individual Mtr4 target RNAs revealed in all cases a shift of peak positions from heavy to light fractions following BTdCPU treatment (Fig. 5B–D), providing evidence that the RNAs were associated with active polysomes. Together,

our results indicate that Mtr4 target transcripts that escaped RNA surveillance in the nucleus and were exported to the cytoplasm were then bound by ribosomes and likely translated.

Despite the increased association of uaRNAs and ptRNAs with ribosomes in the Mtr4 knockdown cells, we detected an unexpected decrease in polysomes. Specifically, analysis of the UV absorption profiles revealed that 48 h of Mtr4 siRNA treatment caused a reduction in polysomes, especially in heavier polysomes (Fig. 5A; Supplemental Fig. S6A). Notably, this is consistent with the behavior of the two full-length mRNAs analyzed (see above). An even more robust reduction in polysomes was evident after 72 h of Mtr4 knockdown, again especially notable in the heavy polysome fractions (Supplemental Fig. S6B). A second Mtr4 siRNA gave similar results (Supplemental Fig. S6C). Also of note, although Mtr4-depleted cells showed reduced growth and morphological changes after 72 h, ptRNA and uaRNA levels continued to increase even after 96 h of knockdown (Supplemental Fig. S6D). These findings together indicate that normally unstable, nuclear, and hence untranslated lncRNAs associate with active ribosomes following Mtr4 knockdown, but, paradoxically, this correlates with an overall reduction in polysomes and hence, very likely, translation.

To verify that Mtr4 knockdown indeed led to a global reduction of translation and determine whether ZFC3H1 knockdown might have similar effects, we performed puromycin incorporation assays. Puromycin is a chain terminator that is incorporated into growing nascent polypeptide chains and thus can be used to label nascent polypeptides (Schmidt et al. 2009). We treated cells with low concentrations of puromycin (1 μ g/mL), and translation efficiency was evaluated by detecting puromylated nascent chains using anti-puromycin antibodies

Ogami et al.

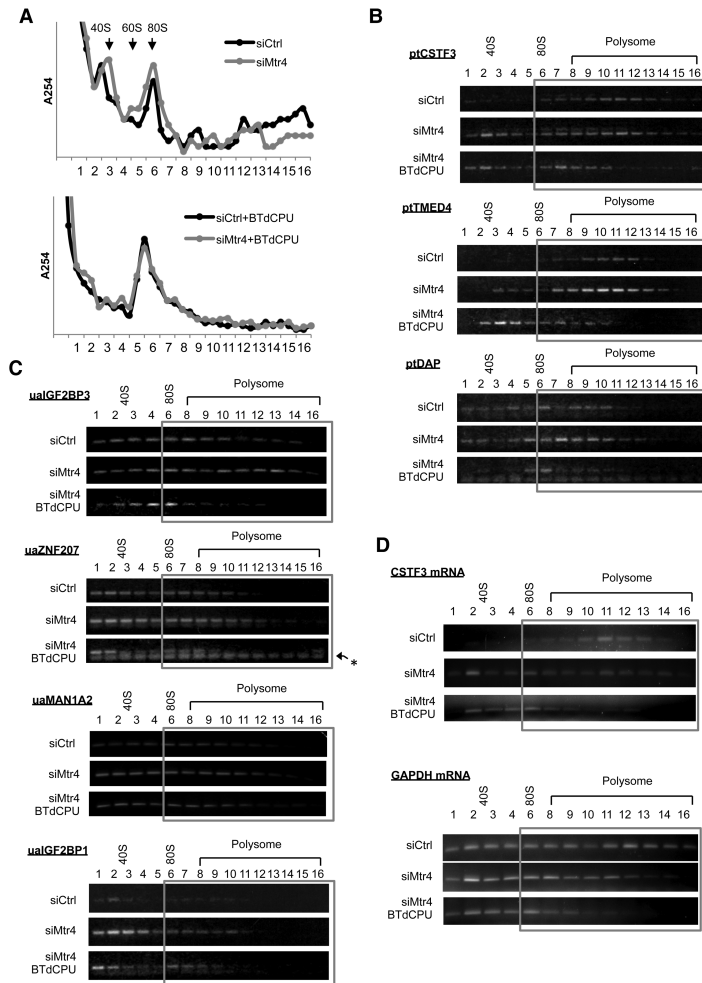


Figure 5. Cytoplasmic ptRNAs and uRNAs associate with active ribosomes but lead to reduced global translation. (A) UV absorption profiles at 254 nm of 15%–45% sucrose gradients. HeLa cells were transfected with either control siRNA (siCtrl) or Mtr4 siRNA (siMtr4) for 48 h, and cytoplasmic extracts were prepared from cells with or without 50 μ M/mL BTdCPU for 3 h. (B–D) RNAs extracted from each fraction as in A were used for oligo(dT)-primed cDNA synthesis, and the indicated transcripts were analyzed by RT-PCR. Gels were pre-stained with EtBr. Ribosome/polysome-associated fractions are highlighted with a gray box. An asterisk marks primer dimers.

for Western blotting of whole-cell lysates (Fig. 6A, quantitation in B). Puromycin incorporation was completely blocked by pretreatment with cycloheximide (CHX) (Fig. 6A, lane 3), confirming that this method allowed us to analyze newly synthesized proteins. In agreement with the polysome fractionation data, efficiency of puromycin incorporation was substantially lower after Mtr4 or ZFC3H1 knockdown (Fig. 6A, cf. lanes 2 and 4,8). Note that siZFC3H1 treatment was less toxic than siMtr4: We observed only a small reduction in cell proliferation and slight morphological changes. Importantly, depletion of NEXT and TRAMP subunits RBM7, ZCCHC8, and ZCCHC7 did not cause a significant decrease in puromycin incorporation (Fig. 6A, lanes 5–7).

Next, we wished to address the possibility that the reduction in polysomes and translation in the Mtr4/ZFC3H1 knockdown cells might reflect another function of the proteins. Specifically, aberrant unprocessed pre-

rRNAs have been observed to accumulate in Mtr4-compromised cells (Schilders et al. 2007; Tafforeau et al. 2013). We also detected pre-rRNAs by Northern blotting using probes hybridizing to 3' extended sequences of 5.8S (ITS2) and 18S (ITS1) rRNAs in Mtr4-depleted cells (Supplemental Fig. S7A). However, Northern blotting using probes that detect both mature and unprocessed rRNAs revealed that the amount of unprocessed rRNAs relative to the mature species was extremely low: The unprocessed pre-rRNAs were detectable only when blots were overexposed such that signals for mature rRNAs were saturated and no longer in a quantitative range (Supplemental Fig. S7B, 5.8S and 18S). In addition, there were no detectable effects on the levels of mature rRNAs. Moreover, ZFC3H1 knockdown, which also caused reduced translation, did not lead to accumulation of aberrant pre-rRNAs or decrease of mature rRNAs (Supplemental Fig. S7B). These findings and other results discussed

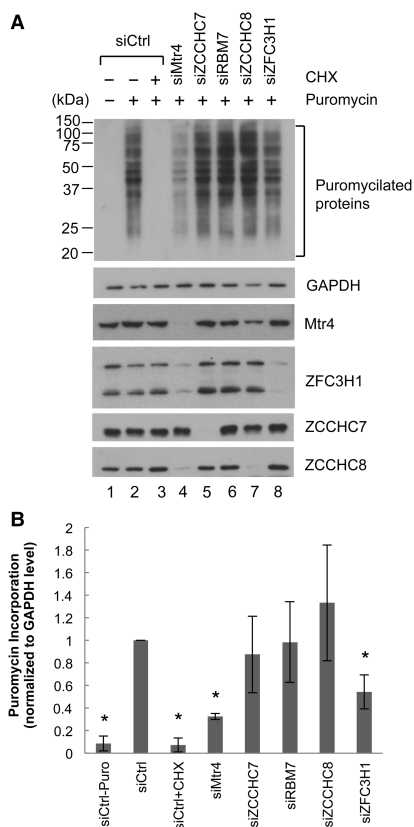


Figure 6. Mtr4/ZFC3H1 depletion causes global reduction of translation. (A) Puromycin incorporation assay. HeLa cells transfected with the indicated siRNAs for 48 h were treated with 1 μ g/mL puromycin for 30 min. (Lane 3) CHX treatment was performed 10 min prior to puromycin addition. Cell lysates were resolved by SDS-PAGE, and puromylated proteins were detected using an anti-puromycin antibody. (B) Puromycin-incorporated protein levels as in A were quantitated using LI-COR Image Studio software and normalized by GAPDH levels. The normalized levels in lane 2 were set to 1. Bars represent mean \pm SD. $n = 3$. Asterisks denote significant difference from lane 2 ($P < 0.05$) using an unpaired Student's t -test.

below argue strongly that the effects of Mtr4/ZFC3H1 knockdown on translation are not due to defects in ribosome assembly.

Together, our experiments revealed an unexpected role for Mtr4-ZFC3H1 in preventing global disruption of mRNA association with polysomes and subsequent translation.

Discussion

It is becoming increasingly clear that a much larger fraction of the human genome is transcribed than previously

appreciated. Much of this “non-mRNA” transcription is by RNA polymerase II, and the RNAs produced are frequently subject to the processing reactions that typically give rise to mRNAs, such as splicing and polyadenylation (Jensen et al. 2013). However, unlike mRNAs, these RNAs are often retained in the nucleus and rapidly turned over. While some of the key factors in this process are known, such as the nuclear exosome and NEXT complex (Lubas et al. 2011; Meola et al. 2016; for review, see Zinder and Lima 2017), it is not well understood how and whether the degradation of these RNAs is coordinated and what the consequences might be if their turnover is prevented. In this study, we identified Mtr4 together with ZFC3H1 as a potential “master regulator” of polyadenylated lncRNA metabolism and showed that when its activity is reduced, normally unstable lncRNAs accumulate and are transported to the cytoplasm, where they appear to “swamp” ribosomes and thereby inhibit translation globally (Fig. 7). Based on these properties, we refer to the Mtr4/ZFC3H1 complex as the “polysome protector complex” (PPC).

ptRNAs and uaRNAs are targeted for turnover by the PPC and not by either of the other characterized Mtr4-containing complexes—NEXT or TRAMP. A large majority of these RNAs contain a PAS at their 3' end, implying that the canonical or very similar pre-mRNA 3' processing machinery is used for their polyadenylation. However, in contrast to mRNAs (which are generally more stable, efficiently exported to the cytoplasm, and translated), uaRNAs and ptRNAs as well as many other lncRNAs are typically rapidly degraded in the nucleus (Andersson et al. 2014b; Li et al. 2015; Schlackow et al. 2017). While neither NEXT nor TRAMP is required for ptRNA and uaRNA turnover, the nuclear exosome is. Thus, the PPC is distinct from NEXT and TRAMP, suggesting that these complexes target distinct sets of RNA substrates for degradation by the exosome. Importantly, though, only the PPC prevents accumulation and transport of lncRNAs from the nucleus that is sufficient to disrupt normal translation, as knockdown of NEXT/TRAMP subunits did not detectably affect translation.

While this work was in preparation, Meola et al. (2016) reported that Mtr4 and ZFC3H1 together with PABPN1 form a complex that preferentially degrades polyadenylated lncRNAs such as snoRNA host gene (SNHG) transcripts. They also showed that ZFC3H1 and PABPN1 knockdown resulted in the accumulation of subsets of uaRNAs and eRNAs. There are similarities as well as differences between our results and those of Meola et al. (2016). For example, in addition to uaRNAs, we identified ptRNAs as Mtr4/ZFC3H1 substrates and, notably, found that these RNAs had very long poly(A) tails and were not substrates for NEXT-mediated degradation. Also, while we did not detect an interaction between PABPN1 and Mtr4/ZFC3H1, we did observe the accumulation of most, but not all, ptRNAs and uaRNAs tested following PABPN1 knockdown. Finally and most importantly, we demonstrated that an important function of Mtr4/ZFC3H1 involves the maintenance of polysome integrity by preventing the accumulation of polyadenylated lncRNAs in the cytoplasm.

Ogami et al.

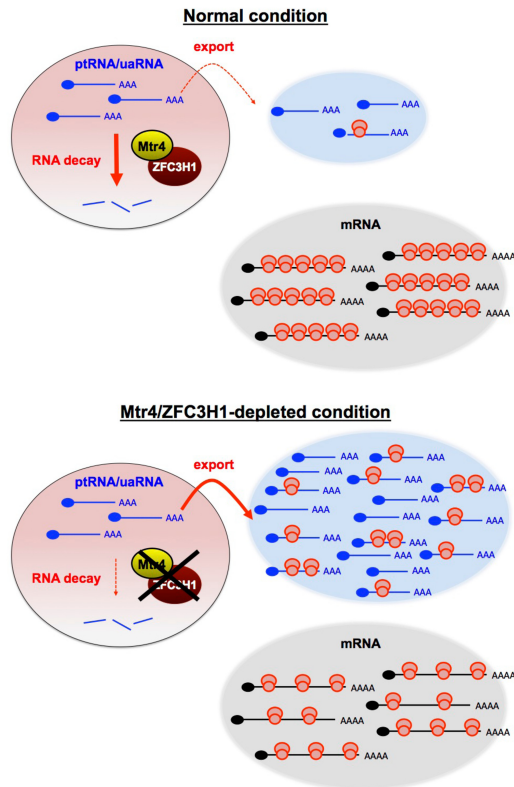


Figure 7. Model for the role of the Mtr4/ZFC3H1 complex in the turnover of nuclear polyadenylated transcripts and how its loss affects translation. Model depicting the impact of PPC deficiency on polyadenylated transcriptomes and global translation. Loss of the PPC results in stabilization of ptRNAs and uaRNAs, which are normally rapidly degraded in the nucleus, and these RNAs are then transported to the cytoplasm. The exported RNAs become ribosome-associated and overwhelm the translational machinery, which leads to disruption of the quantitative balance between available ribosomes and translatable RNAs. See the text for details.

The fact that Mtr4 participates in multiple distinct complexes is reminiscent of Mtr4 proteins in fission yeast. *S. pombe* has two Mtr4 paralogs: Mtr4 and Mtl1. Mtr4 is a TRAMP component (Zhang et al. 2011), whereas Mtl1 forms a core complex with Red1, called MTREC (Mtl1-Red1 core) or NURS (nuclear RNA silencing) (Lee et al. 2013; Egan et al. 2014; Zhou et al. 2015). As mentioned above, Red1 is the fission yeast homolog of ZFC3H1 and is essential for exosomal decay of various RNAs, including CUTs (cryptic unstable transcripts), which are similar to uaRNAs and PROMPTs in mammals; meiotic mRNAs; and unspliced pre-mRNAs (Lee et al. 2013; Egan et al. 2014; Zhou et al. 2015). Specificity for RNA targeting by MTREC/NURS is determined by at least three distinct submodule complexes (Zhou et al. 2015). Further studies

are required to determine whether the PPC also uses associated proteins to distinguish specific RNA substrates.

The global reduction of translation that we observed in PPC-depleted cells highlights the importance of rapidly degrading naturally unstable nuclear lncRNAs. Since failure of RNA surveillance by PPC depletion leads to increased accumulation of cytoplasmic polyadenylated RNAs that have the ability to recruit ribosomes, we propose that heavy polysome formation on mRNAs is hampered by “dilution of ribosomes” by the accumulated normally unstable nuclear “noncoding” RNAs in the cytoplasm (Fig. 7). Translation or ribosome binding of lncRNAs, which contain small ORFs (sORFs) and thus possibly produce micropeptides (Slavoff et al. 2013; Ruiz-Orera et al. 2014; Raj et al. 2016) has been reported in multiple species, including yeast (Ingolia et al. 2014; Smith et al. 2014), fruit flies (Dunn et al. 2013; Aspden et al. 2014), zebrafish (Chew et al. 2013; Bazzini et al. 2014), and mammals (Chew et al. 2013; Zhou et al. 2013; Ingolia et al. 2014; van Heesch et al. 2014; Fields et al. 2015; Ji et al. 2015; Calviello et al. 2016; for review, see Ingolia 2016). Moreover, a recent study provided evidence that a vast majority of monosomes is actively elongating and translating sORFs (Heyer and Moore 2016). Thus, ribosome binding/translation of some lncRNAs occurs and appears compatible with efficient cellular protein synthesis. However, our data provide evidence that when such RNAs are globally stabilized and accumulate in the cytoplasm, they “overwhelm” the system, and translation of mRNAs is repressed (Fig. 7). The fact that Mtr4/ZFC3H1 substrates have long poly(A) tails, which is in contrast to NEXT targets that largely lack poly(A) tails (Meola et al. 2016), may help explain their efficient ribosome association (Peng et al. 2008; Park et al. 2016). It is noteworthy that although eRNAs are largely nonpolyadenylated (Djebali et al. 2012; Andersson et al. 2014a), we recently identified a class of poly(A)⁺ eRNAs that are stabilized by Mtr4/ZFC3H1 knockdown. These eRNAs can also be transported to the cytoplasm and associate with ribosomes (K Ogami, Y Chen, and JL Manley, unpubl.), increasing the pool of lncRNAs that require surveillance by the PPC. Our results thus highlight how critical it is that such lncRNAs be degraded rapidly in the nucleus because, if they survive surveillance by the PPC, they become toxic.

Could another function of Mtr4 or ZFC3H1 be responsible for the disruption of translation that we observed? As noted above, Mtr4 is known to function in the maturation of 5.8S and 18S rRNA from cleaved rRNA precursors (de la Cruz et al. 1998; Schilders et al. 2007; Tafforeau et al. 2013). Might defects in rRNA processing contribute to reduced translation in Mtr4-deficient cells? We consider this unlikely for several reasons: First, since mature rRNAs are abundant and very stable, with half-lives that are days long (Yi et al. 1999; Defoiche et al. 2009), it would be unlikely that mature rRNA levels decrease sufficiently to affect ribosome levels and perturb translation. Our finding that the amount of unprocessed pre-rRNA that accumulated following Mtr4 knockdown was extremely small and that levels of 5.8S and 18S rRNAs were

essentially unaffected is consistent with this. Second, while ZFC3H1 knockdown was shown previously to result in reductions in 47S and 45S pre-rRNAs (Tafforeau et al. 2013), our data showed no changes in downstream pre-rRNA and mature rRNA levels following ZFC3H1 knockdown. Third, unprocessed pre-5.8S rRNA in fact assembles into 60S ribosomes (Briggs et al. 1998), and the resulting immature 60S particles engage in apparently normal translation (Rodriguez-Galan et al. 2015). Indeed, polysome disassembly has not been observed in yeast or mammals under conditions that allow accumulation of aberrant pre-rRNA (Briggs et al. 1998; Strezoska et al. 2000). It is thus unlikely that defective pre-rRNA processing is responsible for the impaired polysome formation/translation that we observed in Mtr4/ZFC3H1 knockdown cells. Indeed, the fact that the excess uRNAs and ptRNAs that accumulated in Mtr4 knockdown cells associated with active ribosomes further argues against this possibility. Mtr4-depleted cells can also accumulate mature and 3' extended snRNA (Hrossova et al. 2015; Lubas et al. 2015) and pri-miRNA 5' by-products (Dorweiler et al. 2014; Lubas et al. 2015). However, accumulation of these transcripts does not contribute to the decreased translation that we observed, since these RNAs are NEXT substrates (Hrossova et al. 2015; Lubas et al. 2015; K Ogami, Y Chen, M Hoque, W Li, B Tian, and JL Manley, unpubl.), and NEXT subunit knockdown had no effect on protein synthesis.

Our model implies that ribosomes must not be present in significant excess or they would otherwise be able to handle the increase in substrates produced when the PPC is depleted. Indeed, studies in yeast have suggested that ribosomes in fact are limiting for translation (Chu and von der Haar 2012; Shah et al. 2013). A similar situation likely exists in human cells. A good example is virus infection. In infected cells, there is often a competition between viral and cellular RNAs for limiting translation components. To overcome this, some viruses alter the balance of viral and cellular mRNA availability for translation by decreasing cytoplasmic cellular mRNA levels by stimulation of mRNA turnover or inhibition of mRNA export (for review, see Walsh and Mohr 2011). For example, herpes simplex virus 1 expresses the endonuclease virion host shutoff (vhs) to accelerate cellular mRNA decay, thereby preventing mRNA overload in infected cells (Dauber et al. 2014). Viral mRNAs associated with polysomes dramatically decrease in the absence of vhs, indicating that the total amount of translatable RNA needs to be regulated to ensure optimal translation of viral mRNAs.

In conclusion, we identified a complex containing the RNA helicase Mtr4 and the zinc finger protein ZFC3H1. This complex, dubbed the PPC, functions in nuclear surveillance of certain unstable polyadenylated lncRNAs to prevent their accumulation, export to the cytoplasm, and consequent disruption of protein synthesis. Our findings are significant because they provide an explanation of why so many lncRNAs are degraded in the nucleus essentially as soon as they are synthesized, as they otherwise have the potential to escape from the nucleus and over-

whelm the cell's translational capacity. While it will be important in the future to learn more about the PPC (e.g., precisely how it functions and whether it can be regulated), our results have uncovered a new and unexpected function for nuclear RNA surveillance.

Materials and methods

Primers and siRNAs

All primers and siRNAs used in this study are listed in Supplemental Table S2. siRNAs against Rrp6 and Dis3 were described previously (Richard et al. 2013; Di Giammartino et al. 2014). siMtr4 s223606 was obtained from Thermo Fischer Scientific.

Cell culture and siRNA transfections

HeLa cells were cultured in DMEM supplemented with 10% FBS. The siRNAs were transfected using DharmaFECT 1 (GE Dharmacon) at 20 nM and maintained for 48, 72, or 96 h as indicated. To maintain high knockdown efficiency after 96 h, the siRNA transfection was repeated 48 h after the first transfection with half the amount of siRNA, and cells were harvested 48 h after the second transfection.

Antibodies

Mtr4 (NB100-1574), ZCCHC8 (NB100-94995), ZCCHC7 (NBP1-89175), Dis3 (H00022894-B01P), and Rrp6 (NBP1-32870) antibodies were from Novus Biologicals. ZFC3H1 (Bethyl Laboratories, A301-456A), hnRNP M (Bethyl Laboratories, A303-910A), NRDE2 (Proteintech, 24968-1-AP), U2AF65 (Sigma, U4758), histone H3 (Abcam, ab1791), and puromycin (Kerafast, 3RH11) were also used in this study. PAPD5 antibody was a generous gift from Dr. Shin-ichi Hoshino (Ogami et al. 2013).

RT-qPCR

Total RNA was isolated using TRIzol according to the manufacturer's instructions followed by DNase I digestion for 30 min at 37°C. Two micrograms of DNase I-treated RNA was reverse-transcribed with oligo(dT) primer using Maxima RT. Reactions were diluted 15 times in water, and qPCR was performed with the primers listed in Supplemental Table S1 and Power SYBR using StepOnePlus (Applied Biosystems). All data were normalized to GAPDH mRNA levels. Reagents for RT-qPCR were all from Thermo Fischer Scientific.

3'READS

Total RNA was purified from control and siRNA-treated cells. RNA integrity was analyzed by using Agilent Bioanalyzer. Samples with an RNA integrity number (RIN) >9.0 were subjected to 3'READS analysis following the protocol described in Hoque et al. (2014). Briefly, after RNA fragmentation, poly(A)⁺ RNA fragments were captured on magnetic beads coated with a chimeric oligonucleotide (oligo CU₅T₄₅), which contained 45 thymidines at the 5' portion and five uridines at the 3' portion, and subjected to RNase H digestion, which removed the bulk of the poly(A) tail and eluted RNA from beads. Eluted RNA was ligated to 5' and 3' adapters followed by reverse transcription, PCR amplification, and deep sequencing on an Illumina platform. 3'READS data were analyzed as described (Li et al. 2015).

Ogami et al.

Gel filtration

HEK293 or HEK293/3Flag-Mtr4 cells from 10 10-cm dishes were washed twice with PBS and lysed in lysis buffer (20 mM Tris-HCl at pH 8.0, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40, 1× protease inhibitor cocktail [Biotools], 100 µg/mL RNase A) for 10 min at room temperature and 15 min on ice. Lysates were sonicated and cleared by centrifugation at 15,000 rpm in an Eppendorf centrifuge 5424 for 20 min at 4°C. The supernatants were filtered using Spin-X 0.45 µm cellulose acetate membrane (Sigma) at 16,000g for 10 min at 4°C prior to applying to a Superose 6 column. Gel filtration was performed in FPLC buffer (50 mM Tris-HCl at pH 8.0, 150 mM NaCl, 0.5% NP-40) with a flow rate of 0.15 mL/min. Eluates were collected every 5 min.

MS analysis

Pooled FPLC fractions 26–33 (pool A), 34–42 (pool B), and 43–50 (pool C) were mixed with anti-Flag M2 magnetic beads for 2 h. Beads were then washed three times with wash buffer (20 mM Tris-HCl at pH 8.0, 300 mM NaCl, 0.5% NP-40, 0.5 mM EDTA). Proteins remaining on the resin were eluted using 100 µL of 150 µg/mL 3Flag peptide (ApexBio) three times and precipitated in 23% TCA and washed with cold acetone. Proteins were reduced with 5 mM Tris (2-carboxyethyl) phosphine hydrochloride (Sigma) and alkylated with 55 mM 2-chloroacetamide (Fluka Analytical). Proteins were digested for 18 h at 37°C in 2 M urea, 100 mM Tris (pH 8.5), and 1 mM CaCl₂ with 2 µg of trypsin (Promega). Multidimensional protein identification technology (MudPIT) analysis was performed using an Eksigent nanoLC pump and a Thermo LTQ-Orbitrap using an in-house built electrospray stage (Wolters et al. 2001).

Protein and peptide identification and protein quantitation were done with Integrated Proteomics Pipeline (IP2; Integrated Proteomics Applications, Inc., <http://www.integratedproteomics.com>). Tandem mass spectra were extracted from raw files using RawConverter (He et al. 2015) and were searched against a UniProt human database with reversed sequences using ProLuCID (Peng et al. 2003; Xu et al. 2015). The search space included all fully tryptic and half-tryptic peptide candidates. Peptide candidates were filtered using DTASelect with the following parameters: -p 2 -y 1 --trypstat --extra --pf -DM 10 --dm --in --brief --quiet (Tabb et al. 2002).

Co-IP

HEK293 or HEK293/3Flag-Mtr4 cells grown in two 10-cm dishes were washed twice with PBS and lysed in lysis buffer (20 mM Tris-HCl at pH 8.0, 300 mM NaCl, 1.5 mM MgCl₂, 0.5 mM EDTA, 0.5% NP-40, 1× protease inhibitor cocktail, 10 µg/mL RNase A, >250 U/mL benzonase) for 20 min on ice. The lysates were sonicated and then centrifuged in an Eppendorf centrifuge 5424 at 15,000 rpm for 20 min at 4°C, and supernatants were rotated with either anti-Flag M2 magnetic beads (Sigma) for 1 h or protein G beads (GE healthcare) in the presence of antibodies for 4 h. The beads were then washed three times with lysis buffer, and proteins retained on the resin were subjected to Western blot analysis.

Subcellular fractionation

Seventy-two hours after siRNA transfection, HeLa cells grown in a 10-cm dish were washed twice with PBS and collected by scraping and centrifugation. Cell pellets were resuspended in 400 µL of swelling buffer (10 mM Tris-HCl at pH 8.0, 1.5 mM MgCl₂, 10 mM KCl, 5 U of RNasin, 1× protease inhibitor cocktail) and incubated for 15 min on ice. Cells were homogenized by passing a 26-gauge needle attached to a 1-mL syringe until >90% of cells were disrupted (typically 10–20 strokes). Half of the lysate was kept in a new tube and used as whole-cell lysate. The rest of the 200 µL of lysate was mixed with 2 µL of 10% NP-40, gently tapped, and immediately centrifuged in an Eppendorf centrifuge 5424 at 6,000 rpm for 5 min. The supernatant was kept in a new tube and used as cytoplasmic fraction. The pellet was washed once with swelling buffer, resuspended in 100 µL of glycerol buffer (20 mM Tris-HCl at pH 8.0, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT, 5 U of RNasin, 1× protease inhibitor cocktail) by pipetting, and then mixed with 100 µL of nucleus lysis buffer (20 mM HEPES-NaOH at pH 7.6, 7.5 mM MgCl₂, 0.2 mM EDTA, 300 mM NaCl, 1 M urea, 1% NP-40, 1 mM DTT, 5 U of RNasin, 1× protease inhibitor cocktail). The mixture was pulse-vortexed three times, incubated for 1 min on ice, and then centrifuged in an Eppendorf centrifuge 5424 at 14,000 rpm for 2 min. The supernatant was used as nuclear-soluble fraction. The pellet was washed once with a 1:1 mixture of glycerol/nucleus lysis buffer and then resuspended in 200 µL of water. RNA was extracted using TRIzol, DNase I-treated, and then reverse-transcribed with oligo(dT) or random primer using Maxima RT.

Polysome fractionation

Polysome fractionation

HeLa cells were transfected with either siCtrl or siMtr4 and maintained in the same medium for the indicated times in five 10-cm dishes. On the day of harvest (~80% confluency), cells were treated with either 50 µM BTdCPU or 100 µg/mL CHX for 3 h and 5 min, respectively, at 37°C. Cells were washed once with ice-cold PBS containing either 50 µM BTdCPU (Millipore) or 100 µg/mL CHX and then resuspended in 400 µL of polysome lysis buffer (20 mM HEPES-NaOH at pH 7.6, 150 mM NaCl, 15 mM MgCl₂, 0.5% NP-40, 80 U RNasin, 1× protease inhibitor cocktail) containing either 50 µM BTdCPU or 100 µg/mL CHX. After 10 min of incubation on ice, the lysates were centrifuged at 3000 rpm for 5 min, and the supernatant was centrifuged at 15,000 rpm for 5 min in a new tube. The supernatant was loaded onto 15%–45% sucrose gradients (20 mM HEPES-NaOH at pH 7.6, 150 mM NaCl, 5 mM MgCl₂, 100 µg/mL CHX) followed by centrifugation at 39,000 rpm for 90 min using an SW41Ti rotor. Fractions (200 µL each) were manually collected, and A254 was determined using a Nanodrop spectrophotometer. RNA was extracted by mixing with 1 vol of TRIzol, isopropanol-precipitated in the presence of GeneElute-LPA (Sigma), and then reverse-transcribed with oligo(dT) primer using Maxima RT. To avoid efficiency differences in RT reactions, RNA amounts were equalized by adding purified yeast RNA (Thermo Fisher Scientific). No PCR products were generated when yeast RNA RT products alone were amplified with the primers used in this study.

Puromycin incorporation assay

siRNA-transfected HeLa cells (~50% confluent) were treated with 1 µg/mL puromycin for 30 min to label nascent polypeptides. CHX treatment was done at 10 µg/mL 10 min prior to adding puromycin. Puromycylated proteins were resolved by SDS-PAGE and detected by Western blotting using anti-puromycin antibody.

Acknowledgments

We are grateful to Shin-ichi Hoshino (Nagoya City University) for sharing the PAPD5 antibody. We also thank Rachel Giesler

- [Novus Biologicals) for the Mtr4, ZCCHC8, and ZCCHC7 antibodies. This work was supported by National Institutes of Health grants R01 GM28983 and R35 GM118136 to J.L.M., and R01 GM084089 to B.T. J.J.M. and J.R.Y. were supported by the National Center for Research Resources (5P41RR011823).
- ## References
- Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA. 2013. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**: 360–363.
- Andersen PR, Domanski M, Kristiansen MS, Storrval H, Ntini E, Verheggen C, Schein A, Bunkenborg J, Poser I, Hallais M, et al. 2013. The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat Struct Mol Biol* **20**: 1367–1376.
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al. 2014a. An atlas of active enhancers across human cell types and tissues. *Nature* **507**: 455–461.
- Andersson R, Refsing Andersen P, Valen E, Core LJ, Bornholdt J, Boyd M, Jensen TH, Sandelin A. 2014b. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun* **5**: 5336.
- Aronica L, Kasparek T, Ruchman D, Marquez Y, Cipak L, Cipakova I, Anrather D, Mikolaskova B, Radtke M, Sarkar S, et al. 2016. The spliceosome-associated protein Nrl1 suppresses homologous recombination-dependent R-loop formation in fission yeast. *Nucleic Acids Res* **44**: 1703–1717.
- Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MA, Brocard M, Couso JP. 2014. Extensive translation of small open reading frames revealed by Poly-Ribo-Seq. *Elife* **3**: e03528.
- Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, et al. 2014. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J* **33**: 981–993.
- Beaulieu YB, Kleinman CL, Landry-Voyer AM, Majewski J, Bachand F. 2012. Polyadenylation-dependent control of long non-coding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS Genet* **8**: e1003078.
- Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L, et al. 2012. U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**: 53–64.
- Bernstein J, Patterson DN, Wilson GM, Toth EA. 2008. Characterization of the essential activities of *Saccharomyces cerevisiae* Mtr4p, a 3' → 5' helicase partner of the nuclear exosome. *J Biol Chem* **283**: 4930–4942.
- Bresson SM, Hunter OV, Hunter AC, Conrad NK. 2015. Canonical poly(A) polymerase activity promotes the decay of a wide variety of mammalian nuclear RNAs. *PLoS Genet* **11**: e1005610.
- Briggs MW, Burkard KT, Butler JS. 1998. Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. *J Biol Chem* **273**: 13255–13263.
- Brodsky AS, Silver PA. 2000. Pre-mRNA processing factors are required for nuclear export. *RNA* **6**: 1737–1749.
- Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M, Obermayer B, Ohler U. 2016. Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods* **13**: 165–170.
- Chen LL. 2016. Linking long noncoding RNA localization and function. *Trends Biochem Sci* **41**: 761–772.
- Chen T, Ozel D, Qiao Y, Harbinski F, Chen L, Denoyelle S, He X, Zvereva N, Supko JG, Chorev M, et al. 2011. Chemical genetics identify eIF2α kinase heme-regulated inhibitor as an anti-cancer target. *Nat Chem Biol* **7**: 610–616.
- Chew GL, Pauli A, Rinn JL, Regev A, Schier AF, Valen E. 2013. Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development* **140**: 2828–2834.
- Chu D, von der Haar T. 2012. The architecture of eukaryotic translation. *Nucleic Acids Res* **40**: 10098–10106.
- Dauber B, Saffran HA, Smiley JR. 2014. The herpes simplex virus 1 virion host shutoff protein enhances translation of viral late mRNAs by preventing mRNA overload. *J Virol* **88**: 9624–9632.
- Defoiche J, Zhang Y, Lagneaux L, Pettengell R, Hegedus A, Willem L, Macallan DC. 2009. Measurement of ribosomal RNA turnover in vivo by use of deuterium-labeled glucose. *Clin Chem* **55**: 1824–1833.
- de la Cruz J, Kressler D, Tollervey D, Linder P. 1998. Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in *Saccharomyces cerevisiae*. *EMBO J* **17**: 1128–1140.
- Di Giammartino DC, Li W, Ogami K, Yashinskii JJ, Hoque M, Tian B, Manley JL. 2014. RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs. *Genes Dev* **28**: 2248–2260.
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. 2012. Landscape of transcription in human cells. *Nature* **489**: 101–108.
- Dorweiler JE, Ni T, Zhu J, Munroe SH, Anderson JT. 2014. Certain adenylated non-coding RNAs, including 5' leader sequences of primary microRNA transcripts, accumulate in mouse cells following depletion of the RNA helicase MTR4. *PLoS One* **9**: e99430.
- Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS. 2013. Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *Elife* **2**: e01179.
- Egan ED, Braun CR, Gygi SP, Moazed D. 2014. Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. *RNA* **20**: 867–881.
- Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, Raychowdhury R, Hacohen N, Carr SA, Ingolia NT, et al. 2015. A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. *Mol Cell* **60**: 816–827.
- Flynn RA, Almada AE, Zamudio JR, Sharp PA. 2011. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc Natl Acad Sci* **108**: 10460–10465.
- Fuke H, Ohno M. 2008. Role of poly(A) tail as an identity element for mRNA nuclear export. *Nucleic Acids Res* **36**: 1037–1049.
- He L, Diedrich J, Chu YY, Yates JR III. 2015. Extracting accurate precursor information for tandem mass spectra by Raw-Converter. *Anal Chem* **87**: 11361–11367.
- Heyer EE, Moore MJ. 2016. Redefining the translational status of 80S monosomes. *Cell* **164**: 757–769.
- Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods* **10**: 133–139.
- Hoque M, Li W, Tian B. 2014. Accurate mapping of cleavage and polyadenylation sites by 3' region extraction and deep sequencing. *Methods Mol Biol* **1125**: 119–129.

Ogami et al.

- Houseley J, Tollervey D. 2009. The many pathways of RNA degradation. *Cell* **136**: 763–776.
- Hrossova D, Sikorsky T, Potesil D, Bartosovic M, Pasulka J, Zdrahal Z, Stefl R, Vanacova S. 2015. RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs. *Nucleic Acids Res* **43**: 4236–4248.
- Ingolia NT. 2016. Ribosome footprint profiling of translation throughout the genome. *Cell* **165**: 22–33.
- Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJ, Jackson SE, Wills MR, Weissman JS. 2014. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep* **8**: 1365–1379.
- Jensen TH, Jacquier A, Libri D. 2013. Dealing with pervasive transcription. *Mol Cell* **52**: 473–484.
- Ji Z, Song R, Regev A, Struhl K. 2015. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife* **4**: e08890.
- Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**: 664–668.
- Lee NN, Chalamcharla VR, Reyes-Turcu F, Mehta S, Zofall M, Balachandran V, Dhakshnamoorthy J, Taneja N, Yamanaka S, Zhou M, et al. 2013. Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell* **155**: 1061–1074.
- Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park JY, Gunderson SI, Kalsotra A, Manley JL, et al. 2015. Systematic profiling of poly(A)⁺ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet* **11**: e1005166.
- Lubas M, Christensen MS, Kristiansen MS, Domanski M, Falckenby LG, Lykke-Andersen S, Andersen JS, Dziembowski A, Jensen TH. 2011. Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell* **43**: 624–637.
- Lubas M, Andersen PR, Schein A, Dziembowski A, Kudla G, Jensen TH. 2015. The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep* **10**: 178–192.
- Meola N, Domanski M, Karadoulama E, Chen Y, Gentil C, Pultz D, Vitting-Seerup K, Lykke-Andersen S, Andersen Jens S, Sandelin A, et al. 2016. Identification of a nuclear exosome decay pathway for processed transcripts. *Mol Cell* **64**: 520–533.
- Nag A, Steitz JA. 2012. Tri-snRNP-associated proteins interact with subunits of the TRAMP and nuclear exosome complexes, linking RNA decay and pre-mRNA splicing. *RNA Biol* **9**: 334–342.
- Ntini E, Jarvelin AI, Bornholdt J, Chen Y, Boyd M, Jorgensen M, Andersson R, Hoof I, Schein A, Andersen PR, et al. 2013. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol* **20**: 923–928.
- Ogami K, Cho R, Hoshino S. 2013. Molecular cloning and characterization of a novel isoform of the non-canonical poly(A) polymerase PAPD7. *Biochem Biophys Res Commun* **432**: 135–140.
- Park JE, Yi H, Kim Y, Chang H, Kim VN. 2016. Regulation of poly(A) tail and translation during the somatic cell cycle. *Mol Cell* **62**: 462–471.
- Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2**: 43–50.
- Peng J, Murray EL, Schoenberg DR. 2008. In vivo and in vitro analysis of poly(A) length effects on mRNA translation. *Methods Mol Biol* **419**: 215–230.
- Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**: 1851–1854.
- Raj A, Wang SH, Shim H, Harpak A, Li YI, Engelmann B, Stephens M, Gilad Y, Pritchard JK. 2016. Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *Elife* **5**: e13328.
- Richard P, Manley JL. 2013. How bidirectional becomes unidirectional. *Nat Struct Mol Biol* **20**: 1022–1024.
- Richard P, Feng S, Manley JL. 2013. A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes Dev* **27**: 2227–2232.
- Rodriguez-Galan O, Garcia-Gomez JJ, Kressler D, de la Cruz J. 2015. Immature large ribosomal subunits containing the 7S pre-rRNA can engage in translation in *Saccharomyces cerevisiae*. *RNA Biol* **12**: 838–846.
- Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM. 2014. Long non-coding RNAs as a source of new peptides. *Elife* **3**: e03523.
- Schilders G, van Dijk E, Pruijn GJ. 2007. C1D and hMtr4p associate with the human exosome subunit PM/Scf-100 and are involved in pre-rRNA processing. *Nucleic Acids Res* **35**: 2564–2572.
- Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ. 2017. Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol Cell* **65**: 25–38.
- Schmidt EK, Clavarino G, Ceppi M, Pierre P. 2009. SUnSET, a nonradioactive method to monitor protein synthesis. *Nat Methods* **6**: 275–277.
- Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. 2013. Rate-limiting steps in yeast protein translation. *Cell* **153**: 1589–1601.
- Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR III, Frank J, Manley JL. 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* **33**: 365–376.
- Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A. 2013. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol* **9**: 59–64.
- Smith JE, Alvarez-Dominguez JR, Kline N, Huynh NJ, Geisler S, Hu W, Collier J, Baker KE. 2014. Translation of small open reading frames within unannotated RNA transcripts in *Saccharomyces cerevisiae*. *Cell Rep* **7**: 1858–1866.
- Strezoska Z, Pestov DG, Lau LF. 2000. Bop1 is a mouse WD40 repeat nucleolar protein involved in 28S and 5.8S rRNA processing and 60S ribosome biogenesis. *Mol Cell Biol* **20**: 5516–5528.
- Tabb DL, McDonald WH, Yates JR III. 2002. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **1**: 21–26.
- Tafforeau L, Zorbas C, Langhendries JL, Mullineux ST, Stamatopoulou V, Mullier R, Wacheul L, Lafontaine DL. 2013. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of pre-rRNA processing factors. *Mol Cell* **51**: 539–551.
- Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18**: 18–30.
- Tomita T, Ieguchi K, Coin F, Kato Y, Kikuchi H, Oshima Y, Kurata S, Maru Y. 2014. ZFC3H1, a zinc finger protein, modulates IL-8 transcription by binding with celastrol A, a potential immune suppressor. *PLoS One* **9**: e108957.

- van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, MacInnes AW, Cuppen E, Simonis M. 2014. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol* **15**: R6.
- Walsh D, Mohr I. 2011. Viral subversion of the host protein synthesis machinery. *Nat Rev Microbiol* **9**: 860–875.
- Wang X, Jia H, Jankowsky E, Anderson JT. 2008. Degradation of hypomodified tRNA^(iMet) in vivo involves RNA-dependent ATPase activity of the DExH helicase Mtr4p. *RNA* **14**: 107–116.
- Wolters DA, Washburn MP, Yates JR III. 2001. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73**: 5683–5690.
- Xu T, Park SK, Venable JD, Wohlschlegel JA, Diedrich JK, Cociorva D, Lu B, Liao L, Hewel J, Han X, et al. 2015. ProLuCID: an improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J Proteomics* **129**: 16–24.
- Xue Y, Wong J, Moreno GT, Young MK, Cote J, Wang W. 1998. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell* **2**: 851–861.
- Yi X, Tesmer VM, Savre-Train I, Shay JW, Wright WE. 1999. Both transcriptional and posttranscriptional mechanisms regulate human telomerase template RNA levels. *Mol Cell Biol* **19**: 3989–3997.
- Zhang Y, LeRoy G, Seelig HP, Lane WS, Reinberg D. 1998. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**: 279–289.
- Zhang K, Fischer T, Porter RL, Dhakshnamoorthy J, Zofall M, Zhou M, Veenstra T, Grewal SI. 2011. Ctr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA. *Science* **331**: 1624–1627.
- Zhou P, Zhang Y, Ma Q, Gu F, Day DS, He A, Zhou B, Li J, Stevens SM, Romo D, et al. 2013. Interrogating translational efficiency and lineage-specific transcriptomes using ribosome affinity purification. *Proc Natl Acad Sci* **110**: 15395–15400.
- Zhou Y, Zhu J, Schermann G, Ohle C, Bendrin K, Sugioka-Sugiyama R, Sugiyama T, Fischer T. 2015. The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat Commun* **6**: 7050.
- Zinder JC, Lima CD. 2017. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* **31**: 88–100.

SUPPLEMENTAL INFORMATION

FIGURE LEGENDS

Figure S1. Mtr4 controls pt- and ua-RNA expression via a NEXT and TRAMP-independent mechanism.

(A) UCSC genome browser screenshot of a representative example of RNA-Seq data at *CSTF3* locus. (B) Western blot analysis of HeLa cell extracts after 48 hours KD of Mtr4 using a second Mtr4 siRNA, s223606. (C) RT-qPCR analysis of ptRNA and uaRNA after Mtr4 KD using s223606. proRBM39 was included as a positive control of RNA accumulation after Mtr4 depletion. Transcript levels were normalized to GAPDH mRNA, and normalized levels in siCtrl-treated cells were set to 1. Bars represent mean \pm SD (n=3), and asterisks denotes significant difference from siCtrl ($P<0.05$) using an unpaired Student's t-test. (D) Western blot analysis of HeLa cell extracts after 96 hours KD of ZCCHC7. (E) RT-qPCR analysis of ptRNA, full-length mRNA (FL mRNA) and uaRNAs. RNA levels were normalized to GAPDH mRNA, and the normalized levels in siCtrl-treated cells were set to 1. Bars represent mean \pm SD (n=3), and asterisks denote significant difference from siCtrl ($P<0.05$) using an unpaired Student's t-test. (F) Western blot analysis of HeLa cell extracts after 96 hours KD of the nuclear exosome catalytic subunits Dis3 and Rrp6. (G) RT-qPCR analysis of ptRNA and uaRNA after depletion of Dis3/Rrp6. Transcript levels were normalized to GAPDH mRNA, and the normalized levels in siCtrl-treated cells were set to 1. Bars represent mean \pm SD (n=3), and asterisks denote significant difference from siCtrl ($P<0.05$) using an unpaired Student's t-test.

Figure S2. Co-IP analysis of Mtr4 complexes.

(A, B) HeLa cells were transfected with the indicated plasmids for 30 hours, and cell extracts were used for IP with anti-Flag in the presence of Benzonase and RNase A followed by western blotting with antibodies against the proteins indicated on the right. Note that the interaction of TRAMP with U5-40K and hnRNP M was observed only when nuclear extracts prepared using high salt buffer were used. (C) Cell extracts prepared from HEK293 cells were used for co-IP with anti-ZCCHC8 antibodies in the presence of Benzonase and RNase A followed by western blotting with antibodies against the proteins indicated on the right.

Figure S3. Effects of PABPN1 knockdown on pt- and ua-RNA abundances.

(A) Cell extracts prepared from HEK293 and HEK293 cells stably expressing 3FLAG-Mtr4 were used for IP with anti-Flag antibodies in the presence of Benzonase and RNase A followed by western blotting with the indicated antibodies. (B) Cell extracts prepared from HEK293 cells were used for co-IP experiment with anti-ZFC3H1 in the presence of Benzonase and RNase A followed by western blotting with antibodies against the proteins indicated on the right. (C) Western blot analysis of HeLa cell extracts after 72 hours KD treatment with the siRNAs indicated at the top, with antibodies against the proteins indicated on the right. (D, E) RT-qPCR analysis of indicated ptRNAs (D) and indicated uaRNAs and NEXT substrates proRBM39 and proBIRC4 (E) after the indicated siRNA transfections. Transcript levels were normalized to GAPDH mRNA, and the normalized levels in siCtrl-treated cells were set to 1. Bars represent mean \pm SD (n=3), and asterisks denote significant difference from siCtrl ($P < 0.05$) using an unpaired Student's t-test.

Figure S4. Poly(A) tail length analysis of pt- and ua-RNAs by RNA ligation coupled poly(A) test (RL-PAT) assay.

(A) Schematic illustration of the RL-PAT assay. RNAs were ligated with the anchor oligonucleotide and reverse-transcribed using the primer complementary to the anchor sequence. The resulting cDNA was subjected to semi-nested PCR, and PCR products were resolved in 2.0% agarose gels prestained with EtBr. (B) RL-PAT assay analysis of pt- and ua-RNAs following either control siRNA (siCtrl) or Mtr4 siRNA (siMtr4) treatment of HeLa cells for 72 hours. RNAs digested with RNase H in the presence of oligo(dT) were analyzed to mark positions of deadenylated RNAs (lanes 3 and 4).

Figure S5. Quantitation of polysome fractionation after Mtr4 knockdown.

Band strength in each fraction in Figure 6B-6D was quantitated using Image-studio software LICOR. Absolute signal values after background subtraction are plotted. Ribosome-bound and -free fractions are indicated by a red dashed line.

Figure S6. Polysome fractionation after Mtr4 knockdown.

(A) Agarose gel electrophoresis-EtBr staining analysis of extracted RNAs from sucrose gradient fractions, corresponding to Figure 6A. (B) UV absorption profiles at 254 nm (upper) and agarose gel electrophoresis-EtBr staining analysis (below) of sucrose gradients. HeLa cells were transfected with either control (Ctrl) or Mtr4 siRNA for 72 hours, and cytoplasmic extracts were separated using 15-45% sucrose gradients. (C) UV absorption profiles at 254 nm of sucrose gradients of samples from Mtr4 KD using a second siRNA, s223606. HeLa cells were transfected with either siCtrl or siMtr4 s223606 for 72 hours, and cytoplasmic

extracts were resolved using 15-45% sucrose gradients. **(D)** RT-qPCR analysis of ptRNA and uaRNA after the indicated hours of Mtr4 KD. Transcript levels were normalized to GAPDH mRNA, and the normalized levels in siCtrl-treated cells were set to 1.

Figure S7. Northern blotting analysis of pre-rRNA processing.

(A) Schematic of the human pre-rRNA showing the positions of the oligonucleotide probes.

(B) Northern blot analysis of HeLa RNA extracts after 72 hours treatment of the indicated siRNAs. Unprocessed and mature rRNAs were detected using the indicated probes. Arrows and asterisks indicate accumulated precursors and mature rRNAs, respectively.

MATERIALS AND METHODS

Northern blotting

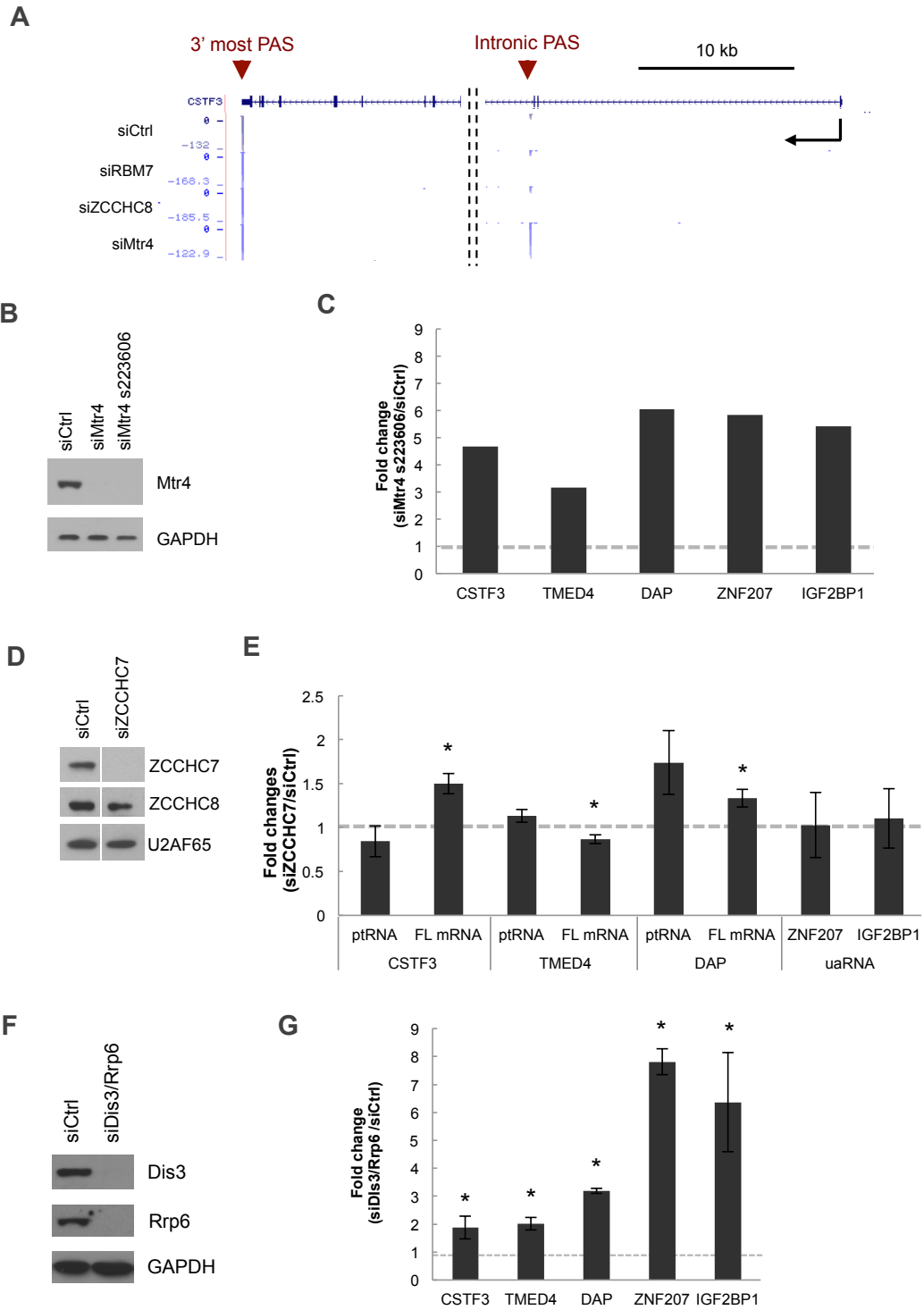
Northern blotting conditions were adapted from Wang and Pestov (2016). In brief, total RNA was extracted with TRIzol (Invitrogen) from untransfected HeLa cells or 72h after transfection with siRNAs. One (for 18S), 2 (for ITS1) or 8 (for 5.8S and ITS2) μ g of total RNAs were separated on a 1% (for 18S and ITS1) or 1.8% (for 5.8S and ITS2) formaldehyde agarose gel using a Tricine-triethanolamine buffer. After O/N passive transfer to Hybond N+ membranes, blots were hybridized with 32 P-labeled oligo probes in hybridization solution (5 \times SSC, 5 \times Denhardt's solution, 0.5% SDS). After O/N incubation at 37°C, blots were washed in 2XSSC/0.1% SDS at RT for 10 min, in 2XSSC/0.1% SDS at 42°C for 10 min, in 1XSSC/0.1% SDS at 42°C for 10 min followed by 1XSSC/0.1% SDS at 42°C for 20 min and autoradiography.

RNA-ligation coupled PAT assay (RL-PAT)

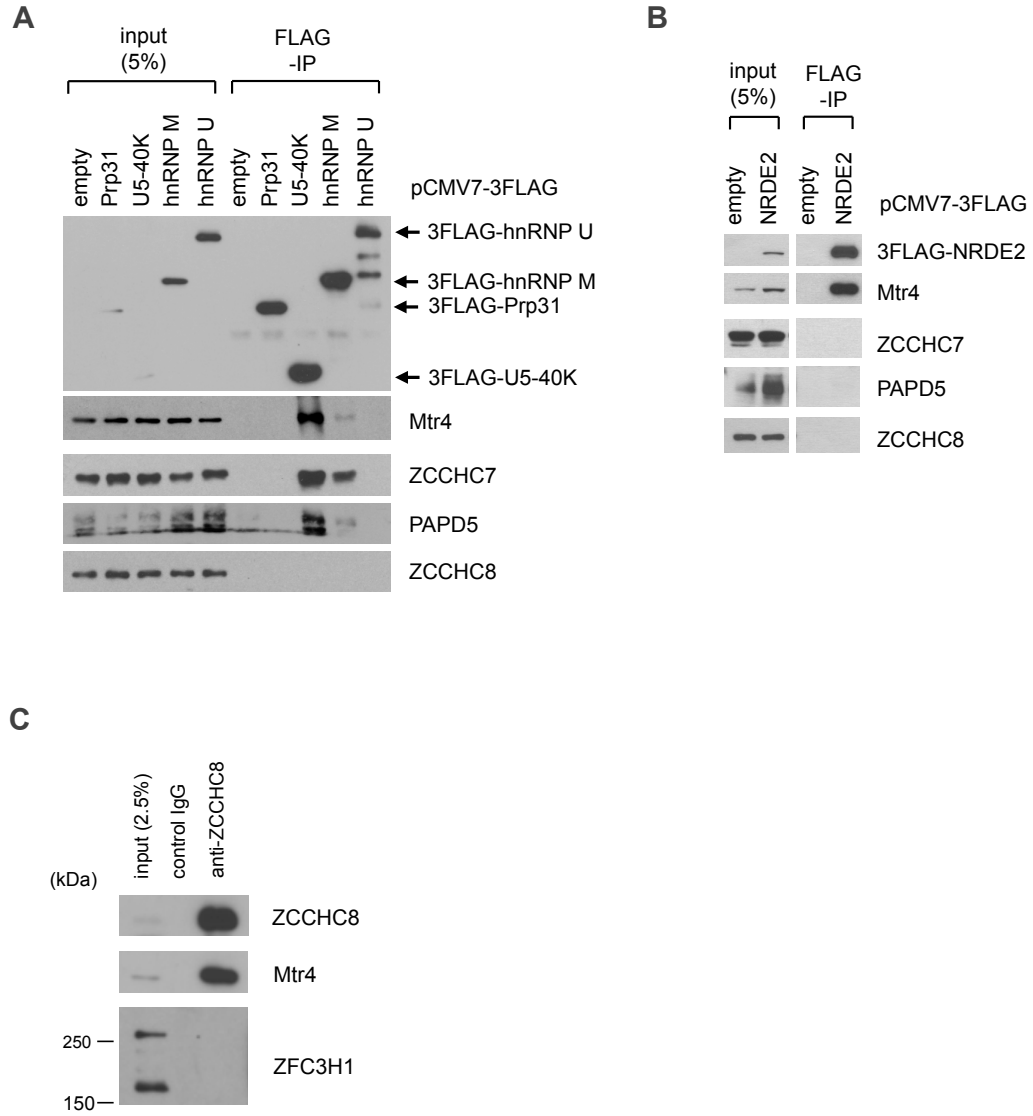
DNase I-treated total RNAs were ligated with KO109 5'-Phospho-CGC GGC CGC GGA GCT CGC- NH₂-3' oligonucleotide using T4 RNA ligase 1 (New England Biolabs) in the presence of 15% PEG8000 at 14°C for 30 min, and reverse-transcribed with KO108 5'-GCG AGC TCC GCG GCC GCG-3' using Maxima Reverse Transcriptase (Thermo Fischer Scientific). Semi-nested PCR was performed using gene-specific forward primers (Table S2) and KO105 5'-GCG AGC TCC GCG GCC GCG TTT TT-3' reverse primer. PCR products were resolved in 2.0% agarose gel prestained with EtBr. To remove poly(A) tails, 10 µg of DNase I-treated RNAs were mixed with 50 pmol of oligo(dT)₁₈ in RNase H buffer (50 mM Tris-HCl (pH 7.4), 100 mM NaCl, 10 mM MgCl₂) and incubated at 85°C for 5 min, 42 °C for 10 min, and then incubated in the presence of 2U Hybridase-Thermostable RNase H (Epicentre) at 42°C for 30 min. RNase H-treated RNAs were isolated using TRIzol.

SUPPLEMENTAL REFERENCE

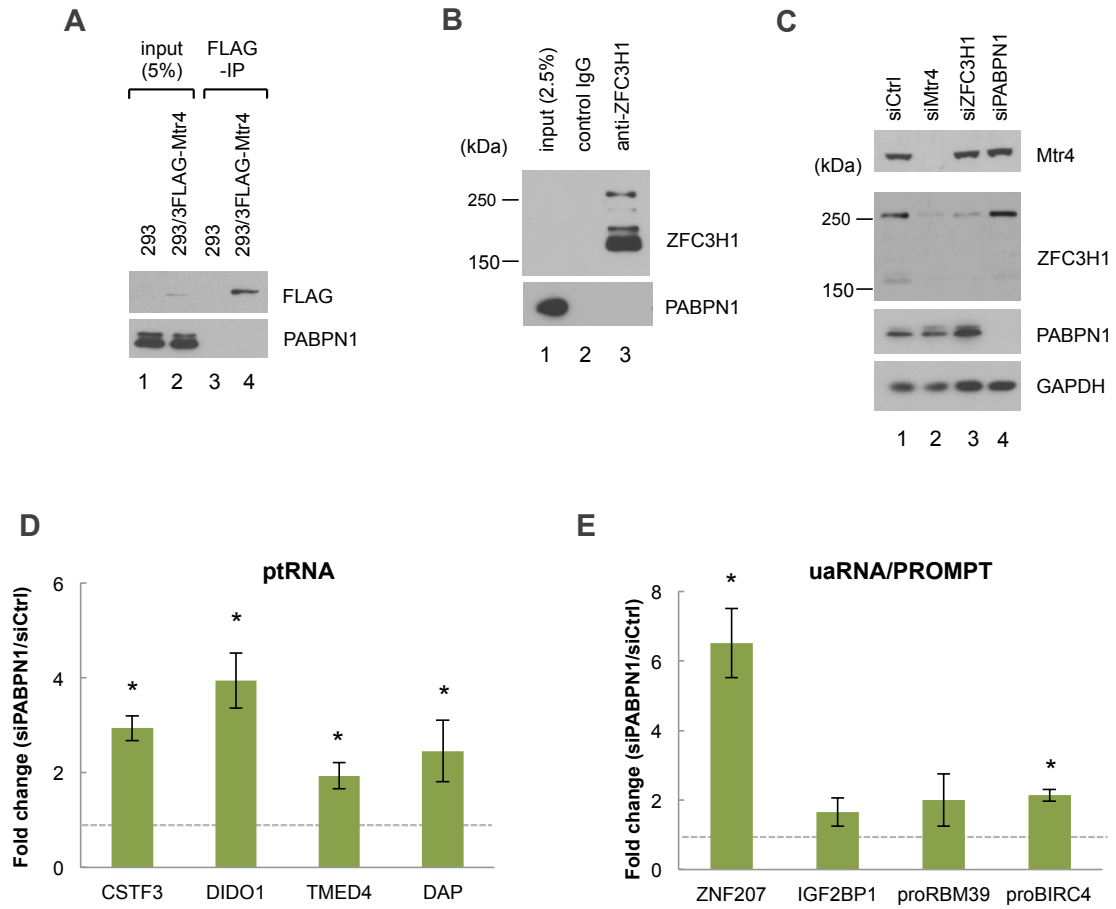
Wang M, Pestov DG 2016. Quantitative northern blot analysis of mammalian rRNA processing. *Methods Mol Biol* **1455**:147-157.



Supplementary Figure 2_K. Ogami

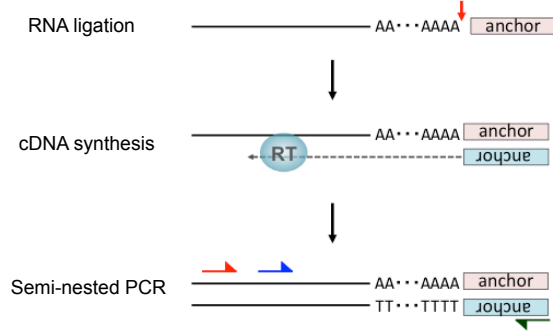


Supplementary Figure 3_K. Ogami

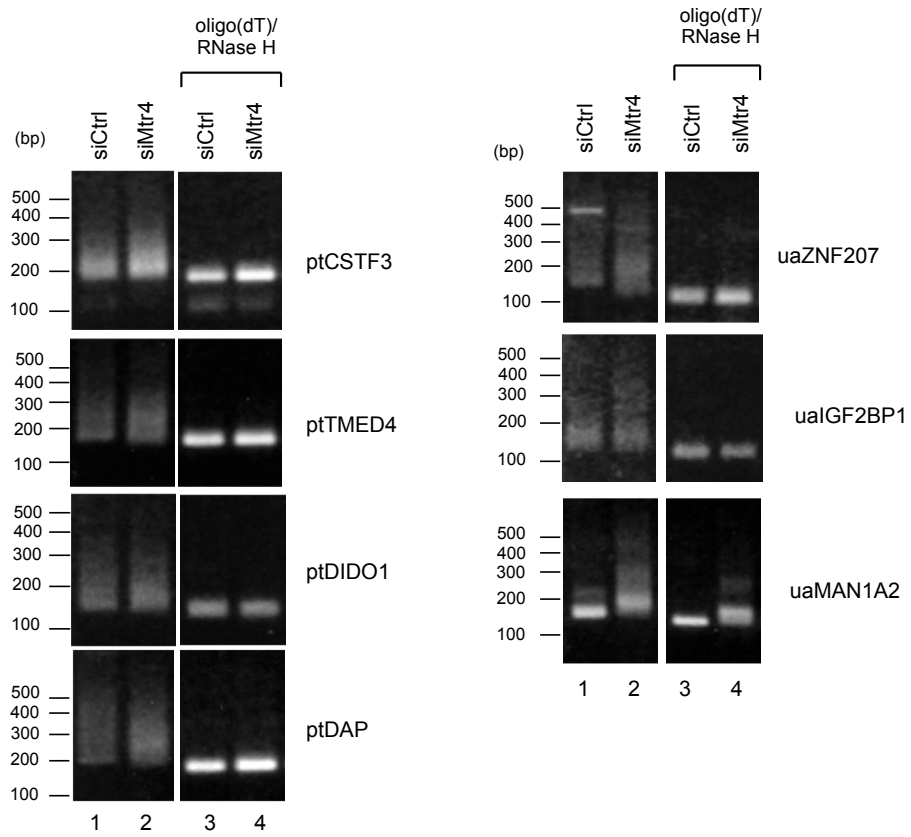


Supplementary Figure 4_K. Ogami

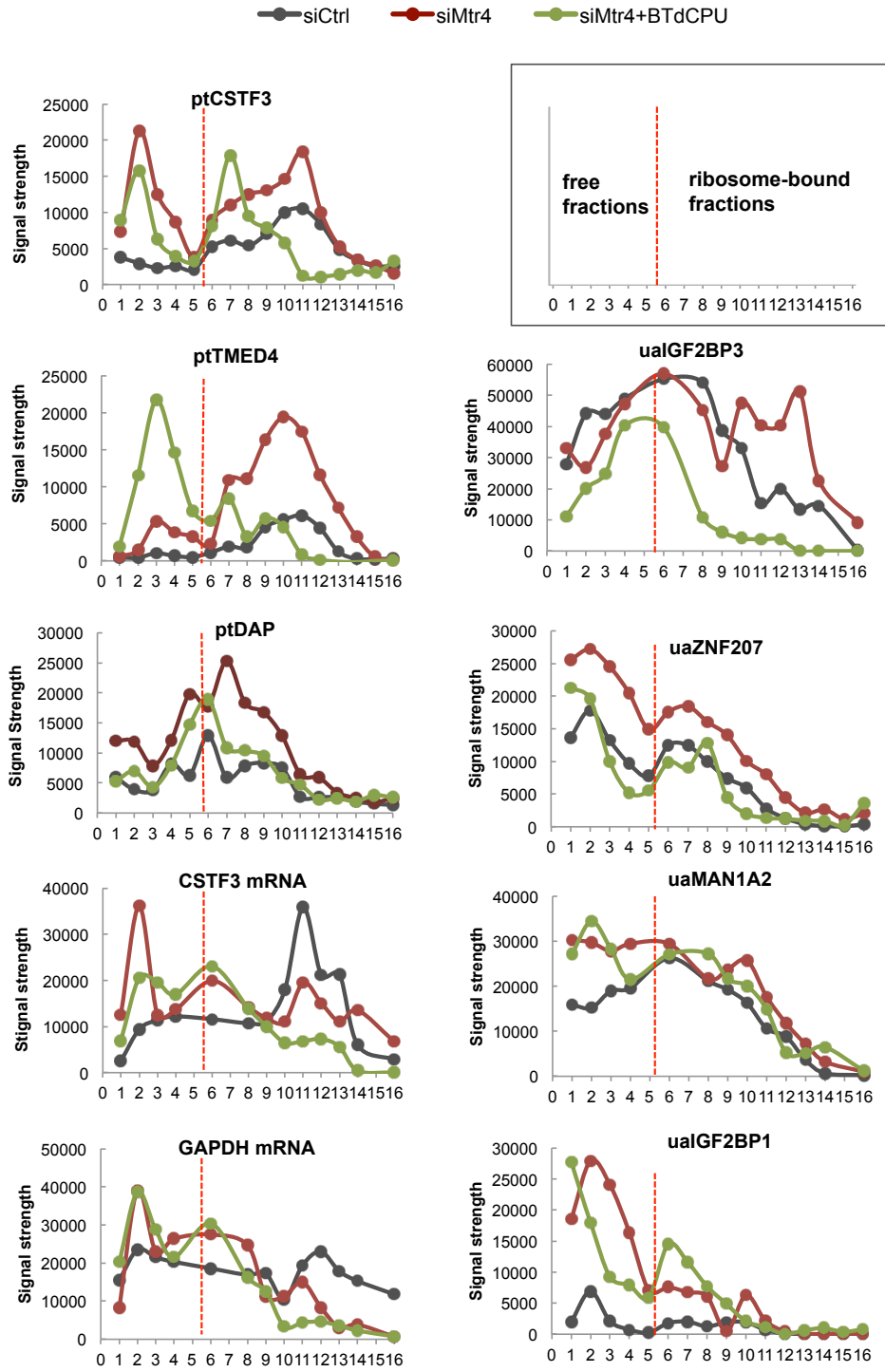
A



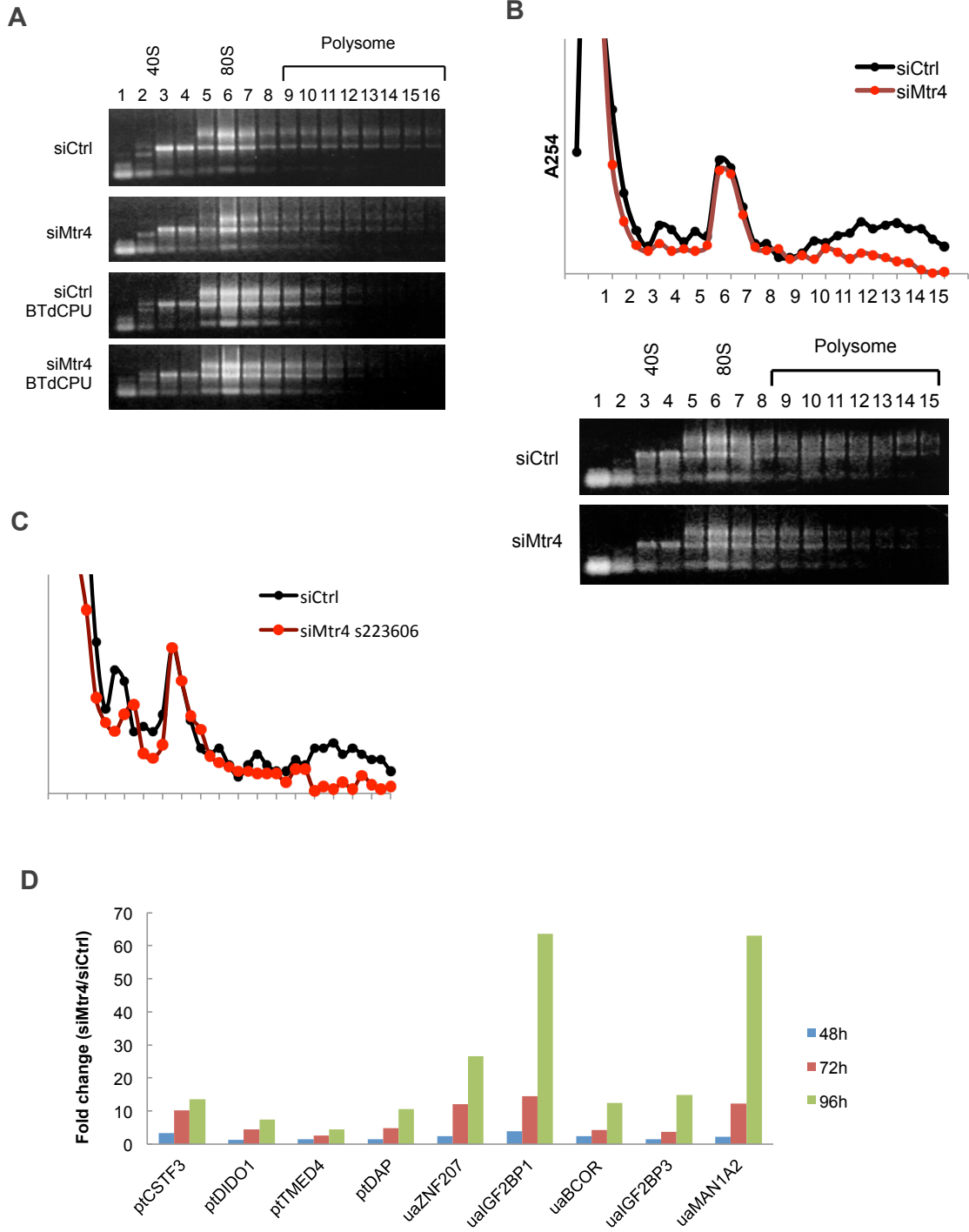
B



Supplementary Figure 5_K. Ogami



Supplementary Figure 6_K. Ogami



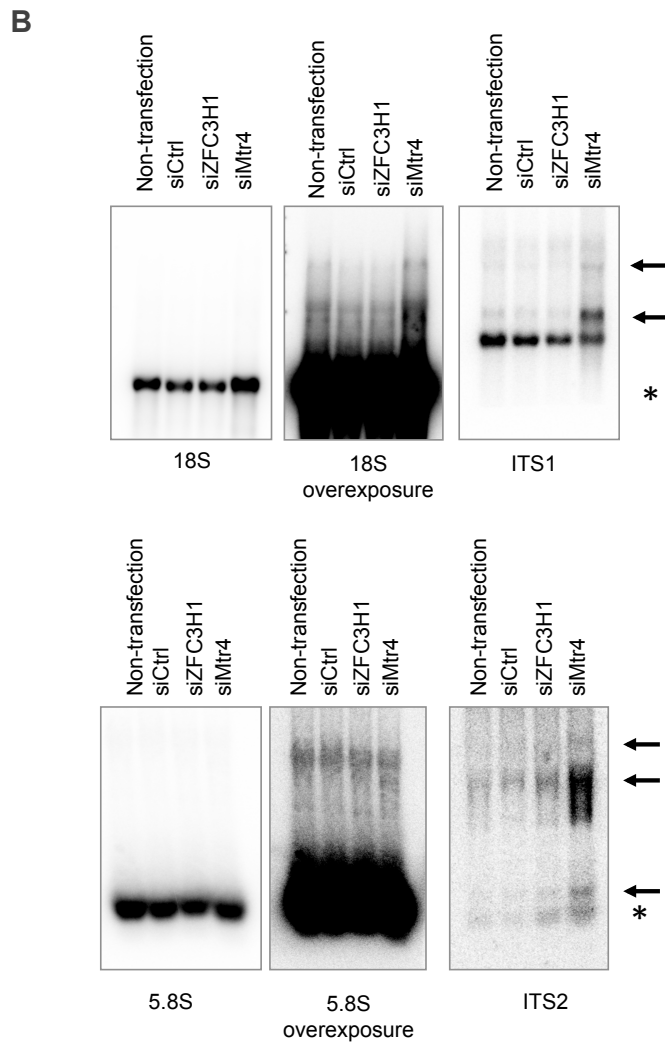
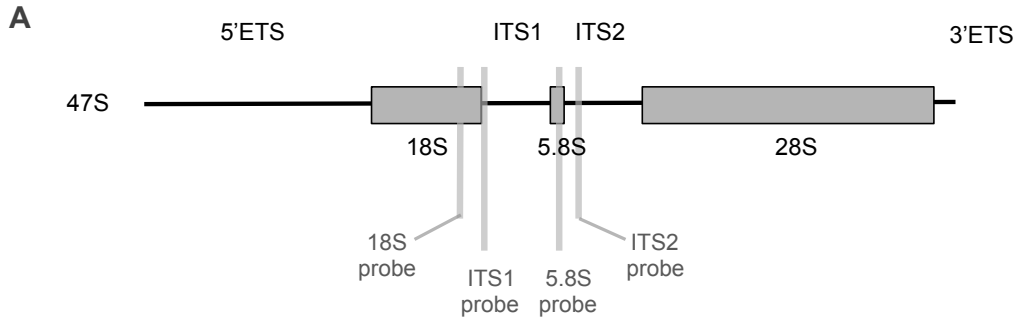


Table S1. Proteins identified in gel filtration-coupled co-IP and MS analysis

Accession	Description	A		B		C	
		293	Mtr4	293	Mtr4	293	Mtr4
P42285	SKIV2L2	0(0%)	161(29.9%)	0(0%)	282(59.1%)	0(0%)	117(38.5%)
Q6NZY4	ZCCHC8	0(0%)	23(19.8%)	0(0%)	120(51.8%)	0(0%)	0(0%)
O60293	ZFC3H1	0(0%)	10(4.1%)	0(0%)	60(18.2%)	0(0%)	0(0%)
P21333	FLNA	0(0%)	4(1.6%)	0(0%)	60(17.7%)	0(0%)	0(0%)
Q14152	EIF3A	0(0%)	0(0%)	0(0%)	45(17.6%)	0(0%)	0(0%)
Q9P2R3	ANKFY1	0(0%)	0(0%)	0(0%)	44(16.6%)	0(0%)	0(0%)
Q9Y580	RBM7	0(0%)	8(16.2%)	0(0%)	34(51.8%)	0(0%)	0(0%)
P78371	CCT2	0(0%)	0(0%)	0(0%)	42(40.2%)	0(0%)	0(0%)
P55884	EIF3B	0(0%)	0(0%)	0(0%)	41(29.6%)	0(0%)	0(0%)
Q9H7Z3	NRDE2	0(0%)	0(0%)	0(0%)	16(17.2%)	0(0%)	12(5.3%)
Q01780	EXOSC10	0(0%)	6(8.1%)	0(0%)	22(18.3%)	0(0%)	0(0%)
P55072	VCP	0(0%)	0(0%)	0(0%)	26(24.8%)	0(0%)	0(0%)
P49368	CCT3	0(0%)	0(0%)	0(0%)	24(29.2%)	0(0%)	2(3.3%)
P68363	TUBA1B	0(0%)	4(11.8%)	0(0%)	22(19.3%)	0(0%)	0(0%)
Q71U36	TUBA1A	0(0%)	4(11.8%)	0(0%)	21(16.2%)	0(0%)	0(0%)
Q9BQE3	TUBA1C	0(0%)	4(11.8%)	0(0%)	21(14%)	0(0%)	0(0%)
P17987	TCP1	0(0%)	0(0%)	0(0%)	24(18.9%)	0(0%)	0(0%)
Q6PEY2	TUBA3E	0(0%)	2(8.4%)	0(0%)	21(16.2%)	0(0%)	0(0%)
Q13748	TUBA3C	0(0%)	2(8.4%)	0(0%)	21(16.2%)	0(0%)	0(0%)
Q13347	EIF3I	0(0%)	0(0%)	0(0%)	23(46.8%)	0(0%)	0(0%)
P50991	CCT4	0(0%)	0(0%)	0(0%)	22(26.3%)	0(0%)	0(0%)
P68366	TUBA4A	0(0%)	0(0%)	0(0%)	22(19.4%)	0(0%)	0(0%)
Q99832	CCT7	0(0%)	0(0%)	0(0%)	21(27.4%)	0(0%)	0(0%)
B5ME19	EIF3CL	0(0%)	0(0%)	0(0%)	20(17.3%)	0(0%)	0(0%)
Q99613	EIF3C	0(0%)	0(0%)	0(0%)	20(17.3%)	0(0%)	0(0%)
Q00839	HNRNPU	0(0%)	8(8.5%)	0(0%)	11(12.4%)	0(0%)	0(0%)
Q07020	RPL18	0(0%)	16(26.6%)	0(0%)	3(11.7%)	0(0%)	0(0%)
Q02878	RPL6	0(0%)	12(14.2%)	0(0%)	7(16.7%)	0(0%)	0(0%)
P62913	RPL11	0(0%)	14(17.4%)	0(0%)	5(16.9%)	0(0%)	0(0%)
Q13200	PSMD2	0(0%)	0(0%)	0(0%)	19(16.5%)	0(0%)	0(0%)
P05387	RPLP2	0(0%)	14(691.6%)	0(0%)	3(46.1%)	0(0%)	0(0%)

P62266	RPS23	0(0%)	14(23.1%)	0(0%)	3(13.3%)	0(0%)	0(0%)
P62241	RPS8	0(0%)	11(31.7%)	0(0%)	6(13%)	0(0%)	0(0%)
Q5RKV6	EXOSC6	0(0%)	4(18%)	0(0%)	7(12.9%)	0(0%)	6(10.3%)
Q5VTE0	EEF1A1P5	0(0%)	4(5%)	0(0%)	13(16.5%)	0(0%)	0(0%)
P68104	EEF1A1	0(0%)	4(5%)	0(0%)	13(16.5%)	0(0%)	0(0%)
Q9BXP5	SRRT	0(0%)	0(0%)	0(0%)	16(8.2%)	0(0%)	0(0%)
P62424	RPL7A	0(0%)	12(23.7%)	0(0%)	4(7.1%)	0(0%)	0(0%)
Q13162	PRDX4	0(0%)	0(0%)	0(0%)	11(38.4%)	0(0%)	5(7.7%)
P60228	EIF3E	0(0%)	0(0%)	0(0%)	15(21.1%)	0(0%)	0(0%)
P11940	PABPC1	0(0%)	0(0%)	0(0%)	15(14.8%)	0(0%)	0(0%)
Q13310	PABPC4	0(0%)	0(0%)	0(0%)	15(14.1%)	0(0%)	0(0%)
O75821	EIF3G	0(0%)	0(0%)	0(0%)	15(37.2%)	0(0%)	0(0%)
O00231	PSMD11	0(0%)	0(0%)	0(0%)	14(19.4%)	0(0%)	0(0%)
Q13868	EXOSC2	0(0%)	7(14.3%)	0(0%)	7(27%)	0(0%)	0(0%)
P40227	CCT6A	0(0%)	0(0%)	0(0%)	14(18.5%)	0(0%)	0(0%)
P06576	ATP5B	0(0%)	0(0%)	0(0%)	14(14.2%)	0(0%)	0(0%)
P26641	EEF1G	0(0%)	0(0%)	0(0%)	14(16.9%)	0(0%)	0(0%)
P78527	PRKDC	0(0%)	0(0%)	0(0%)	13(3.5%)	0(0%)	0(0%)
O43242	PSMD3	0(0%)	0(0%)	0(0%)	12(18.9%)	0(0%)	0(0%)
P39023	RPL3	0(0%)	0(0%)	0(0%)	12(22.8%)	0(0%)	0(0%)
P61353	RPL27	0(0%)	7(20.6%)	0(0%)	5(23.5%)	0(0%)	0(0%)
P46781	RPS9	0(0%)	5(16%)	0(0%)	7(13.9%)	0(0%)	0(0%)
O15371	EIF3D	0(0%)	0(0%)	0(0%)	12(18.1%)	0(0%)	0(0%)
O75643	SNRNP200	0(0%)	0(0%)	0(0%)	12(5.9%)	0(0%)	0(0%)
P36578	RPL4	0(0%)	7(6.6%)	0(0%)	5(13.1%)	0(0%)	0(0%)
Q9Y262	EIF3L	0(0%)	0(0%)	0(0%)	12(8.7%)	0(0%)	0(0%)
Q9Y3B2	EXOSC1	0(0%)	2(15.9%)	0(0%)	9(11.8%)	0(0%)	0(0%)
Q9NQ4	EXOSC5	0(0%)	0(0%)	0(0%)	11(28.9%)	0(0%)	0(0%)
Q09161	NCBP1	0(0%)	2(2.9%)	0(0%)	9(10.9%)	0(0%)	0(0%)
P05388	RPLP0	0(0%)	11(30.6%)	0(0%)	0(0%)	0(0%)	0(0%)
P62249	RPS16	0(0%)	0(0%)	0(0%)	11(37.7%)	0(0%)	0(0%)
P62753	RPS6	0(0%)	7(20.1%)	0(0%)	4(16.9%)	0(0%)	0(0%)
P33176	KIF5B	0(0%)	0(0%)	0(0%)	10(10%)	0(0%)	0(0%)
P50914	RPL14	0(0%)	6(8.8%)	0(0%)	4(5.6%)	0(0%)	0(0%)

P38646	HSPA9	0(0%)	0(0%)	0(0%)	10(10.6%)	0(0%)	0(0%)
Q13263	TRIM28	0(0%)	0(0%)	0(0%)	10(17.4%)	0(0%)	0(0%)
Q15024	EXOSC7	0(0%)	5(20.6%)	0(0%)	5(18.6%)	0(0%)	0(0%)
P52272	HNRNPM	0(0%)	4(4%)	0(0%)	5(5.2%)	0(0%)	0(0%)
P60866	RPS20	0(0%)	2(19.3%)	0(0%)	7(22.7%)	0(0%)	0(0%)
P10809	HSPD1	0(0%)	0(0%)	0(0%)	9(14%)	0(0%)	0(0%)
Q99547	MPHOSPH6	0(0%)	0(0%)	0(0%)	6(30.6%)	0(0%)	3(13.8%)
Q8NHW5	RPLP0P6	0(0%)	9(18.3%)	0(0%)	0(0%)	0(0%)	0(0%)
P31943	HNRNPH1	0(0%)	4(6%)	0(0%)	5(15.6%)	0(0%)	0(0%)
Q9NQ75	EXOSC3	0(0%)	0(0%)	0(0%)	6(28.7%)	0(0%)	2(7.3%)
P07910	HNRNPC	0(0%)	8(19.9%)	0(0%)	0(0%)	0(0%)	0(0%)
P11021	HSPA5	0(0%)	0(0%)	0(0%)	8(14.1%)	0(0%)	0(0%)
P23246	SFPQ	0(0%)	0(0%)	0(0%)	8(7.8%)	0(0%)	0(0%)
P15880	RPS2	0(0%)	3(7.5%)	0(0%)	5(12.6%)	0(0%)	0(0%)
P29692	EEF1D	0(0%)	0(0%)	0(0%)	8(23.5%)	0(0%)	0(0%)
O15397	IPO8	0(0%)	0(0%)	0(0%)	8(9.8%)	0(0%)	0(0%)
Q99460	PSMD1	0(0%)	0(0%)	0(0%)	8(11.9%)	0(0%)	0(0%)
Q92769	HDAC2	0(0%)	0(0%)	0(0%)	8(11.9%)	0(0%)	0(0%)
P51665	PSMD7	0(0%)	0(0%)	0(0%)	8(14.2%)	0(0%)	0(0%)
Q9Y4E8	USP15	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	7(6.8%)
Q9NPD3	EXOSC4	0(0%)	0(0%)	0(0%)	7(13.9%)	0(0%)	0(0%)
Q9H361	PABPC3	0(0%)	0(0%)	0(0%)	7(2.2%)	0(0%)	0(0%)
P62195	PSMC5	0(0%)	0(0%)	0(0%)	7(21.2%)	0(0%)	0(0%)
P62280	RPS11	0(0%)	5(21.5%)	0(0%)	2(18.4%)	0(0%)	0(0%)
Q16576	RBBP7	0(0%)	0(0%)	0(0%)	7(7.1%)	0(0%)	0(0%)
Q32P51	HNRNPA1L2	0(0%)	0(0%)	0(0%)	7(10.3%)	0(0%)	0(0%)
P06702	S100A9	0(0%)	0(0%)	0(0%)	7(34.2%)	0(0%)	0(0%)
P18124	RPL7	0(0%)	7(19.4%)	0(0%)	0(0%)	0(0%)	0(0%)
P24534	EEF1B2	0(0%)	0(0%)	0(0%)	7(34.2%)	0(0%)	0(0%)
Q08211	DHX9	0(0%)	0(0%)	0(0%)	7(4.2%)	0(0%)	0(0%)
O94776	MTA2	0(0%)	0(0%)	0(0%)	6(9.7%)	0(0%)	0(0%)
Q6P2Q9	PRPF8	0(0%)	0(0%)	0(0%)	6(3.9%)	0(0%)	0(0%)
P43686	PSMC4	0(0%)	0(0%)	0(0%)	6(22.7%)	0(0%)	0(0%)
P55795	HNRNPH2	0(0%)	4(6%)	0(0%)	2(5.1%)	0(0%)	0(0%)

P05141	SLC25A5	0(0%)	0(0%)	0(0%)	6(11.7%)	0(0%)	0(0%)
P49327	FASN	0(0%)	0(0%)	0(0%)	6(4.1%)	0(0%)	0(0%)
P07437	TUBB	0(0%)	0(0%)	0(0%)	6(28.2%)	0(0%)	0(0%)
Q09028	RBBP4	0(0%)	0(0%)	0(0%)	6(12.2%)	0(0%)	0(0%)
O00303	EIF3F	0(0%)	0(0%)	0(0%)	6(9.2%)	0(0%)	0(0%)
P27635	RPL10	0(0%)	2(11.7%)	0(0%)	4(28%)	0(0%)	0(0%)
P62191	PSMC1	0(0%)	0(0%)	0(0%)	6(11.6%)	0(0%)	0(0%)
P62333	PSMC6	0(0%)	0(0%)	0(0%)	6(20.8%)	0(0%)	0(0%)
P27708	CAD	0(0%)	0(0%)	0(0%)	6(3.9%)	0(0%)	0(0%)
P46777	RPL5	0(0%)	6(16.8%)	0(0%)	0(0%)	0(0%)	0(0%)
Q15008	PSMD6	0(0%)	0(0%)	0(0%)	6(12.9%)	0(0%)	0(0%)
P54105	CLNS1A	0(0%)	0(0%)	0(0%)	6(24.1%)	0(0%)	0(0%)
P53621	COPA	0(0%)	0(0%)	0(0%)	6(5.4%)	0(0%)	0(0%)
P08238	HSP90AB1	0(0%)	0(0%)	0(0%)	4(7.9%)	0(0%)	2(2.9%)
P08670	VIM	0(0%)	0(0%)	0(0%)	6(16.5%)	0(0%)	0(0%)
P35580	MYH10	0(0%)	0(0%)	0(0%)	5(3.4%)	0(0%)	0(0%)
Q9BYE4	SPRR2G	0(0%)	0(0%)	0(0%)	5(30.1%)	0(0%)	0(0%)
P47914	RPL29	0(0%)	5(14.5%)	0(0%)	0(0%)	0(0%)	0(0%)
O60812	HNRNPCL1	0(0%)	5(17.1%)	0(0%)	0(0%)	0(0%)	0(0%)
P43243	MATR3	0(0%)	0(0%)	0(0%)	5(5%)	0(0%)	0(0%)
P62263	RPS14	0(0%)	0(0%)	0(0%)	5(29.1%)	0(0%)	0(0%)
P25398	RPS12	0(0%)	0(0%)	0(0%)	5(40.9%)	0(0%)	0(0%)
Q13547	HDAC1	0(0%)	0(0%)	0(0%)	5(8.3%)	0(0%)	0(0%)
P62701	RPS4X	0(0%)	5(14.8%)	0(0%)	0(0%)	0(0%)	0(0%)
P35325	SPRR2B	0(0%)	0(0%)	0(0%)	5(30.6%)	0(0%)	0(0%)
P35326	SPRR2A	0(0%)	0(0%)	0(0%)	5(30.6%)	0(0%)	0(0%)
P17980	PSMC3	0(0%)	0(0%)	0(0%)	5(15.9%)	0(0%)	0(0%)
P68371	TUBB4B	0(0%)	0(0%)	0(0%)	5(24%)	0(0%)	0(0%)
P46779	RPL28	0(0%)	5(17.5%)	0(0%)	0(0%)	0(0%)	0(0%)
P40429	RPL13A	0(0%)	5(9.9%)	0(0%)	0(0%)	0(0%)	0(0%)
P12236	SLC25A6	0(0%)	0(0%)	0(0%)	5(8.4%)	0(0%)	0(0%)
P12235	SLC25A4	0(0%)	0(0%)	0(0%)	5(8.4%)	0(0%)	0(0%)
P20930	FLG	0(0%)	0(0%)	0(0%)	5(1%)	0(0%)	0(0%)
P22532	SPRR2D	0(0%)	0(0%)	0(0%)	5(30.6%)	0(0%)	0(0%)

P22531	SPRR2E	0(0%)	0(0%)	0(0%)	5(30.6%)	0(0%)	0(0%)
Q9BRS2	RIOK1	0(0%)	0(0%)	0(0%)	4(12.3%)	0(0%)	0(0%)
P31151	S100A7	0(0%)	0(0%)	0(0%)	4(35.6%)	0(0%)	0(0%)
Q14683	SMC1A	0(0%)	0(0%)	0(0%)	4(2.4%)	0(0%)	0(0%)
Q58FF7	HSP90AB3P	0(0%)	0(0%)	0(0%)	2(4.2%)	0(0%)	2(3.5%)
Q07866	KLC1	0(0%)	0(0%)	0(0%)	4(2.4%)	0(0%)	0(0%)
Q00610	CLTC	0(0%)	0(0%)	0(0%)	4(3.8%)	0(0%)	0(0%)
P35998	PSMC2	0(0%)	0(0%)	0(0%)	4(12.2%)	0(0%)	0(0%)
Q9BVA1	TUBB2B	0(0%)	0(0%)	0(0%)	4(18%)	0(0%)	0(0%)
O43852	CALU	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	4(14.3%)
P23458	JAK1	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	4(2.9%)
P04350	TUBB4A	0(0%)	0(0%)	0(0%)	4(20%)	0(0%)	0(0%)
O43143	DHX15	0(0%)	2(3.3%)	0(0%)	2(3%)	0(0%)	0(0%)
A6NHT5	HMX3	0(0%)	4(4.5%)	0(0%)	0(0%)	0(0%)	0(0%)
P05109	S100A8	0(0%)	0(0%)	0(0%)	4(20.4%)	0(0%)	0(0%)
Q8WXI9	GATAD2B	0(0%)	0(0%)	0(0%)	4(6.7%)	0(0%)	0(0%)
Q7L2H7	EIF3M	0(0%)	0(0%)	0(0%)	4(15.5%)	0(0%)	0(0%)
Q9H0B6	KLC2	0(0%)	0(0%)	0(0%)	4(2.3%)	0(0%)	0(0%)
O00487	PSMD14	0(0%)	0(0%)	0(0%)	4(10%)	0(0%)	0(0%)
P62081	RPS7	0(0%)	0(0%)	0(0%)	4(13.4%)	0(0%)	0(0%)
Q13885	TUBB2A	0(0%)	0(0%)	0(0%)	4(18%)	0(0%)	0(0%)
P25787	PSMA2	0(0%)	0(0%)	0(0%)	4(23.5%)	0(0%)	0(0%)
Q6PKG0	LARP1	0(0%)	0(0%)	0(0%)	4(2.8%)	0(0%)	0(0%)
Q15750	TAB1	0(0%)	0(0%)	0(0%)	4(9.3%)	0(0%)	0(0%)
Q06265	EXOSC9	0(0%)	4(5.5%)	0(0%)	0(0%)	0(0%)	0(0%)
Q8NC51	SERBP1	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	4(7.1%)
Q15029	EFTUD2	0(0%)	0(0%)	0(0%)	4(3.2%)	0(0%)	0(0%)
P26640	VARS	0(0%)	0(0%)	0(0%)	4(3%)	0(0%)	0(0%)
P14625	HSP90B1	0(0%)	0(0%)	0(0%)	2(3.5%)	0(0%)	2(5.1%)
Q14315	FLNC	0(0%)	0(0%)	0(0%)	3(1.2%)	0(0%)	0(0%)
Q14697	GANAB	0(0%)	0(0%)	0(0%)	3(1.8%)	0(0%)	0(0%)
Q5T750	XP32	0(0%)	0(0%)	0(0%)	3(6.4%)	0(0%)	0(0%)
P05023	ATP1A1	0(0%)	0(0%)	0(0%)	3(3%)	0(0%)	0(0%)
Q99623	PHB2	0(0%)	0(0%)	0(0%)	3(11%)	0(0%)	0(0%)

P07814	EPRS	0(0%)	0(0%)	0(0%)	3(3.2%)	0(0%)	0(0%)
Q99873	PRMT1	0(0%)	0(0%)	0(0%)	3(7.5%)	0(0%)	0(0%)
Q16531	DDB1	0(0%)	0(0%)	0(0%)	3(2.8%)	0(0%)	0(0%)
P60900	PSMA6	0(0%)	0(0%)	0(0%)	3(8.9%)	0(0%)	0(0%)
Q13435	SF3B2	0(0%)	0(0%)	0(0%)	3(4.9%)	0(0%)	0(0%)
P25705	ATP5A1	0(0%)	0(0%)	0(0%)	3(4.3%)	0(0%)	0(0%)
Q04637	EIF4G1	0(0%)	0(0%)	0(0%)	3(1.8%)	0(0%)	0(0%)
P61626	LYZ	0(0%)	0(0%)	0(0%)	3(30.4%)	0(0%)	0(0%)
Q13330	MTA1	0(0%)	0(0%)	0(0%)	3(4.5%)	0(0%)	0(0%)
P46778	RPL21	0(0%)	0(0%)	0(0%)	3(16.2%)	0(0%)	0(0%)
P61247	RPS3A	0(0%)	0(0%)	0(0%)	3(10.2%)	0(0%)	0(0%)
P61163	ACTR1A	0(0%)	0(0%)	0(0%)	3(8.8%)	0(0%)	0(0%)
Q15233	NONO	0(0%)	0(0%)	0(0%)	3(11.9%)	0(0%)	0(0%)
P51991	HNRNPA3	0(0%)	0(0%)	0(0%)	3(9.5%)	0(0%)	0(0%)
Q9NY65	TUBA8	0(0%)	0(0%)	0(0%)	3(10.5%)	0(0%)	0(0%)
P20700	LMNB1	0(0%)	0(0%)	0(0%)	3(10.1%)	0(0%)	0(0%)
Q92526	CCT6B	0(0%)	0(0%)	0(0%)	3(2.5%)	0(0%)	0(0%)
P63151	PPP2R2A	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	3(3.4%)
Q9UNM6	PSMD13	0(0%)	0(0%)	0(0%)	3(8.2%)	0(0%)	0(0%)
Q9Y265	RUVBL1	0(0%)	0(0%)	0(0%)	3(4.8%)	0(0%)	0(0%)
P35579	MYH9	0(0%)	0(0%)	0(0%)	2(1.3%)	0(0%)	0(0%)
Q14406	CSHL1	0(0%)	2(3.6%)	0(0%)	0(0%)	0(0%)	0(0%)
P28074	PSMB5	0(0%)	0(0%)	0(0%)	2(10.6%)	0(0%)	0(0%)
P28070	PSMB4	0(0%)	0(0%)	0(0%)	2(11.4%)	0(0%)	0(0%)
Q9BUA3	C11orf84	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	2(3.1%)
Q14008	CKAP5	0(0%)	0(0%)	0(0%)	2(1.8%)	0(0%)	0(0%)
Q9UBQ5	EIF3K	0(0%)	0(0%)	0(0%)	2(13.3%)	0(0%)	0(0%)
O43809	NUDT21	0(0%)	0(0%)	0(0%)	2(10.6%)	0(0%)	0(0%)
P69849	NOMO3	0(0%)	0(0%)	0(0%)	2(1.9%)	0(0%)	0(0%)
P55036	PSMD4	0(0%)	0(0%)	0(0%)	2(4.8%)	0(0%)	0(0%)
Q9H853	TUBA4B	0(0%)	0(0%)	0(0%)	2(11.6%)	0(0%)	0(0%)
P13010	XRCC5	0(0%)	0(0%)	0(0%)	2(2.9%)	0(0%)	0(0%)
P49207	RPL34	0(0%)	2(15.4%)	0(0%)	0(0%)	0(0%)	0(0%)
P07900	HSP90AA1	0(0%)	0(0%)	0(0%)	2(3.3%)	0(0%)	0(0%)

Q7Z406	MYH14	0(0%)	0(0%)	0(0%)	2(1.1%)	0(0%)	0(0%)
Q7Z6Z7	HUWE1	0(0%)	0(0%)	0(0%)	2(0.6%)	0(0%)	0(0%)
P49792	RANBP2	0(0%)	0(0%)	0(0%)	2(0.6%)	0(0%)	0(0%)
P62140	PPP1CB	0(0%)	0(0%)	0(0%)	2(9.2%)	0(0%)	0(0%)
P62244	RPS15A	0(0%)	0(0%)	0(0%)	2(12.3%)	0(0%)	0(0%)
O00571	DDX3X	0(0%)	0(0%)	0(0%)	2(4.2%)	0(0%)	0(0%)
P62314	SNRPD1	0(0%)	0(0%)	0(0%)	2(10.9%)	0(0%)	0(0%)
P60842	EIF4A1	0(0%)	0(0%)	0(0%)	2(7.6%)	0(0%)	0(0%)
P60660	MYL6	0(0%)	0(0%)	0(0%)	2(15.9%)	0(0%)	0(0%)
Q13509	TUBB3	0(0%)	0(0%)	0(0%)	2(6.4%)	0(0%)	0(0%)
P62851	RPS25	0(0%)	2(15.2%)	0(0%)	0(0%)	0(0%)	0(0%)
Q86SG5	S100A7A	0(0%)	0(0%)	0(0%)	2(23.8%)	0(0%)	0(0%)
P61313	RPL15	0(0%)	2(10.3%)	0(0%)	0(0%)	0(0%)	0(0%)
Q96L21	RPL10L	0(0%)	0(0%)	0(0%)	2(8.9%)	0(0%)	0(0%)
P01243	CSH1	0(0%)	2(3.7%)	0(0%)	0(0%)	0(0%)	0(0%)
P01241	GH1	0(0%)	2(3.7%)	0(0%)	0(0%)	0(0%)	0(0%)
P01242	GH2	0(0%)	2(3.7%)	0(0%)	0(0%)	0(0%)	0(0%)
Q13283	G3BP1	0(0%)	0(0%)	0(0%)	2(4.1%)	0(0%)	0(0%)
Q15393	SF3B3	0(0%)	2(2.2%)	0(0%)	0(0%)	0(0%)	0(0%)
Q15459	SF3A1	0(0%)	0(0%)	0(0%)	2(4.8%)	0(0%)	0(0%)
P46821	MAP1B	0(0%)	0(0%)	0(0%)	2(1%)	0(0%)	0(0%)
P61254	RPL26	0(0%)	2(11%)	0(0%)	0(0%)	0(0%)	0(0%)
P83731	RPL24	0(0%)	2(13.4%)	0(0%)	0(0%)	0(0%)	0(0%)
Q15155	NOMO1	0(0%)	0(0%)	0(0%)	2(1.9%)	0(0%)	0(0%)
O15523	DDX3Y	0(0%)	0(0%)	0(0%)	2(4.2%)	0(0%)	0(0%)
Q96DI7	SNRNP40	0(0%)	0(0%)	0(0%)	2(5.9%)	0(0%)	0(0%)
Q99417	MYCBP	0(0%)	0(0%)	0(0%)	2(19.4%)	0(0%)	0(0%)
P53618	COPB1	0(0%)	0(0%)	0(0%)	2(3.6%)	0(0%)	0(0%)
Q92616	GCN1L1	0(0%)	0(0%)	0(0%)	2(1%)	0(0%)	0(0%)
Q9UQE7	SMC3	0(0%)	0(0%)	0(0%)	2(1.6%)	0(0%)	0(0%)
Q9NYJ8	TAB2	0(0%)	0(0%)	0(0%)	2(3.5%)	0(0%)	0(0%)
P26373	RPL13	0(0%)	2(8.5%)	0(0%)	0(0%)	0(0%)	0(0%)
P18077	RPL35A	0(0%)	2(16.4%)	0(0%)	0(0%)	0(0%)	0(0%)
Q5JPE7	NOMO2	0(0%)	0(0%)	0(0%)	2(1.8%)	0(0%)	0(0%)

Q9UNX3	RPL26L1	0(0%)	2(11%)	0(0%)	0(0%)	0(0%)	0(0%)
P36542	ATP5C1	0(0%)	0(0%)	0(0%)	2(7%)	0(0%)	0(0%)
P57052	RBM11	0(0%)	0(0%)	0(0%)	2(3.9%)	0(0%)	0(0%)
Q9Y295	DRG1	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	2(4.6%)
P61978	HNRNPK	0(0%)	0(0%)	0(0%)	2(3.2%)	0(0%)	0(0%)
O75369	FLNB	0(0%)	0(0%)	0(0%)	2(0.9%)	0(0%)	0(0%)
P50990	CCT8	0(0%)	0(0%)	0(0%)	39(38%)	2(4.7%)	5(2.6%)
P39019	RPS19	2(11%)	0(0%)	0(0%)	22(34.5%)	0(0%)	0(0%)
P0CG48	UBC	2(3.6%)	4(3.6%)	0(0%)	6(3.2%)	0(0%)	0(0%)
P0CG47	UBB	2(10.9%)	4(10.9%)	0(0%)	6(9.6%)	0(0%)	0(0%)
P62979	RPS27A	2(16%)	4(16%)	0(0%)	6(14.1%)	0(0%)	0(0%)
P62987	UBA52	2(19.5%)	4(19.5%)	0(0%)	6(17.2%)	0(0%)	0(0%)
P09651	HNRNPA1	2(7.5%)	0(0%)	0(0%)	9(13.2%)	0(0%)	0(0%)
P62888	RPL30	2(14.8%)	7(32.2%)	0(0%)	2(20.9%)	0(0%)	0(0%)
Q8NHM4	PRSS3P2	2(14.6%)	2(12.1%)	0(0%)	0(0%)	0(0%)	0(0%)
P47929	LGALS7	2(18.4%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)
Q53GS9	USP39	2(3.7%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)
Q7Z7F7	MRPL55	0(0%)	0(0%)	0(0%)	0(0%)	2(11.7%)	0(0%)
Q86YZ3	HRNR	2(1.7%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)
Q92620	DHX38	2(2.1%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)
P23396	RPS3	3(14.4%)	4(17.7%)	0(0%)	15(35.8%)	0(0%)	0(0%)
P32969	RPL9	3(12%)	4(5.7%)	0(0%)	12(37.5%)	0(0%)	0(0%)
P54652	HSPA2	0(0%)	0(0%)	0(0%)	7(4.5%)	3(3.9%)	0(0%)
P35268	RPL22	3(18.8%)	4(19.5%)	0(0%)	0(0%)	0(0%)	0(0%)
P01040	CSTA	3(30.6%)	2(30.6%)	0(0%)	0(0%)	0(0%)	0(0%)
Q15208	STK38	0(0%)	0(0%)	0(0%)	0(0%)	3(11.6%)	0(0%)
P81605	DCD	4(37.3%)	2(12.7%)	0(0%)	19(531.6%)	0(0%)	2(23.6%)
P11142	HSPA8	0(0%)	4(4.2%)	0(0%)	16(13.3%)	4(5.4%)	0(0%)
P48643	CCT5	0(0%)	0(0%)	3(2.2%)	53(51.8%)	2(5.2%)	6(2.2%)
P34931	HSPA1L	2(3.9%)	7(6.1%)	0(0%)	3(5.5%)	3(4.2%)	0(0%)
P30050	RPL12	2(18.8%)	0(0%)	0(0%)	9(14.5%)	3(18.2%)	0(0%)
P02768	ALB	5(5.1%)	2(2.8%)	0(0%)	5(3.9%)	0(0%)	0(0%)
P62857	RPS28	5(33.3%)	0(0%)	0(0%)	2(23.2%)	0(0%)	0(0%)
Q16875	PFKFB3	0(0%)	0(0%)	0(0%)	0(0%)	7(7.1%)	0(0%)

P22626	HNRNPA2B1	8(16.4%)	9(25.8%)	0(0%)	31(30.9%)	0(0%)	8(9.1%)
P48741	HSPA7	0(0%)	0(0%)	0(0%)	10(3.3%)	8(6.8%)	0(0%)
Q9BYX7	POTEKP	9(6.7%)	8(6.7%)	0(0%)	0(0%)	0(0%)	0(0%)
Q6S8J3	POTEE	9(2.3%)	8(2.3%)	0(0%)	0(0%)	0(0%)	0(0%)
A5A3E0	POTEF	9(2.3%)	8(2.3%)	0(0%)	0(0%)	0(0%)	0(0%)
Q9Y657	SPIN1	0(0%)	0(0%)	0(0%)	0(0%)	9(6.5%)	4(6.5%)
P17066	HSPA6	2(4.5%)	7(4.5%)	0(0%)	10(1.9%)	9(6.4%)	0(0%)
P62736	ACTA2	11(9.5%)	8(6.6%)	0(0%)	0(0%)	0(0%)	0(0%)
P68133	ACTA1	11(9.5%)	8(6.6%)	0(0%)	0(0%)	0(0%)	0(0%)
P68032	ACTC1	11(9.5%)	8(6.6%)	0(0%)	0(0%)	0(0%)	0(0%)
P63267	ACTG2	11(9.6%)	8(6.6%)	0(0%)	0(0%)	0(0%)	0(0%)
P08107	HSPA1A	4(8.7%)	22(15.6%)	0(0%)	58(31.8%)	11(10.9%)	0(0%)
P60709	ACTB	15(17.1%)	9(12.3%)	0(0%)	4(12.8%)	0(0%)	0(0%)
P63261	ACTG1	15(17.1%)	9(12.3%)	0(0%)	4(12.8%)	0(0%)	0(0%)
O75688	PPM1B	0(0%)	0(0%)	0(0%)	0(0%)	17(26.7%)	0(0%)
Q9BQA1	WDR77	0(0%)	0(0%)	10(9.4%)	103(81.7%)	9(17.3%)	0(0%)
Q8WWY3	PRPF31	0(0%)	0(0%)	0(0%)	17(26.9%)	19(19.6%)	3(9.6%)
P69905	HBA1	21(23.9%)	17(28.2%)	0(0%)	0(0%)	0(0%)	0(0%)
P02042	HBD	25(25.2%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)
P68871	HBB	40(591.2%)	19(10.9%)	0(0%)	0(0%)	0(0%)	0(0%)
P98175	RBM10	40(19.9%)	2(3.2%)	0(0%)	51(21.3%)	21(10.8%)	47(11.6%)
O14744	PRMT5	7(7.2%)	5(3%)	41(11%)	246(64.1%)	37(28.9%)	13(6.3%)
P23588	EIF4B	0(0%)	0(0%)	0(0%)	0(0%)	90(40.8%)	51(25.4%)

Table S2. Primers and siRNAs used in this study

1) Primers for RT-PCR

Sense (5'-3')		antisense(5'-3')	
KO-009 RBM7 S	CAGGTACGAAAGGACTATGG	KO-010 RBM7 AS	GTCTCAAAGCACTGTTTCATCAC
KO-060 GAPDH S	ACGACCACTTTGTCAAGCTC	KO-061 GAPDH AS	TTCTCTTGTGCTCTTGCTG
KO-064 CSTF3 int pA S	CTGGAAACTGTACATTGAAGCAGAG	KO-065 CSTF3 int pA AS	TCCTCCTTGTCTAATTGGCTAT
KO-066 CSTF3 last pA S	GCCTCTGCAGAAAACCTCTG	KO-067 CSTF3 last pA AS	CCATGTGATAGAGGCACCAA
KO-068 DIDO1 int pA S	CCAAACTCTTGCCCTTGAG	KO-069 DIDO1 int pA AS	TCCTAACTCCTGCTCCCAGA
KO-070 DIDO1 last pA S	TTGGTGTGAAAGCCAAGTG	KO-071 DIDO1 last pA AS	CCTTCTCATCGTACTAGATGTTGG
KO-072 DAP int pA S	CGGCCCACTACACTAAAGGA	KO-073 DAP int pA AS	TGAGCATTGAGGCACAAGTC
KO-074 DAP last pA S	GCCTTTCTGGTGCTGTCTC	KO-075 DAP last pA AS	TTTCAAGTGTGAGGCTGTGC
KO-091 TMED4 int pA S	CAGTTGCTTGATCAGGTGGA	KO-092 TMED4 int pA AS	AGCTGATCTCCAAGCAGAA
KO-093 TMED4 last pA S	CACTTGCTGACCCTGGTTCT	KO-094 TMED4 last pA AS	GGAAAATCACTCGAGGCAAA
KO-095 CD55 int pA S	TTGCCAGAGTGCAGAGGTAA	KO-096 CD55 int pA AS	TGAAAGGTGGGTTTGCTACA
KO-097 CD55 last pA S	GGCAGTCCTGGAATCACATT	KO-098 CD55 last pA AS	TTTTCTCGTGATCCCATTC
KO-116 ZNF207 UA S	TGGGCGCTGTCTCTATCTTT	KO-126 ZNF207 UA AS	GCATGCGAATGGAAAACAG
KO-118 IGF2BP1 UA S	ACGTGGCTGGGTAGAACAAA	KO-119 IGF2BP1 UA AS	AGAAATGGGGCTGGGACTTA
KO-120 AIFM1 UA S	CCATGCTTAAGTCCAGATGCT	KO-121 AIFM1 UA AS	CCCTCACAATGGTTCCGACTT
KO-189 proRBM39 S	CATTTTTGAAGGAACGGTAG	KO-188 proRBM39 AS	GGAAATAGTGGAGAAAAGCA
KO-190 proBIRC4 S	ATCCATCATCTCACCACTT	KO-191 proBIRC4 AS	CTGGTCATACCCTGGATTTA
KO-217 proFBXO7 S	TTGGGGTGATTGTTATGCAG	KO-218 proFBXO7 AS	AGGCCTCAATGAGCTGTGAT
KO-256 CSTF3 exon2 S	GGTGAAGAAAGCGGAAAAGA	KO-065 CSTF3 int pA AS	TCCTCCTTGTCTAATTGGCTAT
KO-257 TMED4 exon3 S	TACCAGGATGGCTCTCTTCG	KO-092 TMED4 int pA AS	AGCTGATCTCCAAGCAGAA
KO-261 NEAT1 S	TTGGTTCTGAGCTGCGTCTA	KO-262 NEAT1 AS	GTGCTGTAAAGGGGAAGAAA
KO-265 RPPH1 S	TGCTACTCCACTCCCATGTC	KO-266 RPPH1 AS	CAGCCATTGAACTCACTTCG

2) Primers for RL-PAT assay

primary PCR		secondary PCR	
KO-064 CSTF3 int pA S	CTGGAAACTGTACATTGAAGCAGAG	KO-106 CSTF3 int PAT 2nd	GCCAATTAGGACAAGGAGGATT
KO-419 DIDO1 int PAT 1st	GCTCTTCTGGGAAGAATGTTTG	KO-420 DIDO1 int PAT 2nd	TGTCAGGTAAGCATTCTTCAGT
KO-418 TMED4 int PAT 1st	GTTCTCACAGTTCATCCATGTT	KO-107 TMED4 int PAT 2nd	GCATGTGTCAGAATTTCTTAAGGC
KO-132 DAP int S PAT 1st	CAGGAAGCTCTGGTGTCTTGT	KO-072 DAP int pA S	CGGCCACTACACTAAAGGA
KO-126 ZNF207 UA AS	GCATGCGAATGGAAAACAG	KO-129 ZNF207 UA PAT 2nd	AAAGATAGAGACAGGCCCA
KO-119 IGF2BP1 UA AS	AGAAATGGGGCTGGGACTTA	KO-130 IGF2BP1 UA PAT 2nd	TTTGTCTACCCAGCCACGT
KO-154 MAN1A2 UA AS	TCACCGGAGTTAATGGCTTC	KO-155 MAN1A2 UA PAT 2nd	GAGGATTCAGCAACAGGGTC

3) Primers for plasmid construction

Sense (5'-3')		antisense(5'-3')	
KO-052 Mtr4 Sall S	TTTGTGACATGGCGGACGCGCATTCGGAG AT	KO-053 Mtr4 Xbal AS	ATATCTAGACTACAAGTAGAGGCTGGCAGCA A
KO-295 NRDE2 3Flag Sall KpnI S	AAAAGTCGACGGTACCATGGCGCTGTT CCAGC	KO-296 NRDE2 3Flag Xbal AS	AATTTCTAGACTAATCCTCCAGCAGCA
KO-380 hnRNPU PstI S	TTCCCTGCAGATGAGTTCCTCGCCTGTT	KO-381 hnRNPU Sall AS	TTTTGTGCGACTCAATAATATCCTTGGTGATAA
KO-382 hnRNPM PstI S	TTTCCTGCAGATGGCGGAGGGGTCGA	KO-383 hnRNPM Sall AS	TTTTGTGCGACTTAAGCGTTTCTATCAATTC
KO-400 Prp31 Sall S	TTTTGTGCGACATGTCTCTGGCAGATGAG C	KO-401 Prp31 Xbal AS	TTTTTCTAGATCAGGTGGACATAAGGCC
KO-402 SNRNP40 Sall S	TTTTGTGCGACATGATAGAACAGCAGAAG CGT	KO-403 SNRNP40 Xbal AS	TTTTCTAGATCTTCACTGAATCTCTCCA

4) siRNAs

Name	Sense (5'-3')	antisense(5'-3')
siMtr4	CAAUUAAGGCUCUGAGUAATT	UUACUCAGAGCCUUAUUUGTT
siRBM7	GUCAUAUGGUGGAAAAUUUTT	AAUUUUCCACCAUAUGACTT
siZCCHC8	GAAAUACAACAGAAUAAAATT	UUUUUAUUCUGUUGUAUUUUCTT
siZCCHC7	CUAUUGAGCUGGUUGUAUAATT	UUAUCAACCAGCUCUAUAGTT
siZFC3H1	GAAACAAGCUGAAGAAGAATT	UUCUUCUUCAGCUUGUUUCTT



Appendix III. NRDE-2, the human homolog of fission yeast Nrl1, prevents DNA damage accumulation in human cells

Yaqiong Chen performed experiments and analyzed data in Figure 4

BRIEF COMMUNICATION



NRDE-2, the human homolog of fission yeast Nrl1, prevents DNA damage accumulation in human cells

Patricia Richard ^a, Koichi Ogami ^{a,b*}, Yaqiong Chen^a, Shuang Feng^a, James J. Moresco^c, John R. Yates III^c, and James L. Manley^a

^aDepartment of Biological Sciences, Columbia University, New York, NY, USA; ^bDepartment of Biological Chemistry, Graduate School of Pharmaceutical Sciences, Nagoya City University, Nagoya, Japan; ^cDepartment of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA

ABSTRACT

The RNA helicase Mtr4 is a versatile protein that is a crucial component of several distinct RNA surveillance complexes. Here we describe a novel complex that contains Mtr4, but has a role distinct from any of those previously described. We found that Mtr4 association with the human homolog of fission yeast Nrl1, NRDE-2, defines a novel function for Mtr4 in the DNA damage response pathway. We provide biochemical evidence that Mtr4 and NRDE-2 are part of the same complex and show that both proteins play a role in the DNA damage response by maintaining low DNA double-strand break levels. Importantly, the DNA damage response function of the Mtr4/NRDE-2 complex does not depend on the formation of R loops. We show however that NRDE-2 and Mtr4 can affect R-loop signals at a subset of distinct genes, possibly regulating their expression. Our work not only expands the wide range of Mtr4 functions, but also elucidates an important role of the less characterized human NRDE-2 protein.

ARTICLE HISTORY

Received 1 March 2018
Accepted 14 April 2018

KEYWORDS

NRDE-2; Mtr4; DNA damage;
DDR; R loop; R-loop;
double-strand breaks; DSBs;
Nrl1; lncRNAs; NRDE2

1. Introduction



Mtr4 is an RNA helicase that is the centerpiece of several distinct complexes involved in turnover of specific RNAs [1]. It is part of exosome adaptor complexes such as TRAMP (Trf4-Air-Mtr4 polyadenylation) and NEXT (nuclear exosome targeting) [2]. We and others recently identified Mtr4 as a master player of another RNA surveillance complex, Mtr4/ZFC3H1 or PAXT (poly(A) tail exosome targeting), which plays a role in turnover of polyadenylated lncRNAs, such as ptRNAs (prematurely terminated RNAs), uaRNAs (upstream antisense RNAs) and eRNAs (enhancer RNAs) [3,4].

We and others previously identified NRDE-2 (nuclear RNAi defective-2; also known as C14ORF102) as an Mtr4-interacting protein in human cells [2,3]. While the function of human NRDE-2 is unknown, the *S. pombe* NRDE-2 homolog Nrl1 (NRDE-2 like 1) has been found to interact with the Mtr4-like protein Mtl1 and with splicing factors that help the RNAi-dependent assembly of heterochromatin at loci called HOODs (heterochromatin domains) on certain specific genes [5,6]. NRDE-2 was first identified in *C. elegans* as a factor required for RNAi in the nucleus [7]. Interestingly, its association with the RNAi machinery is necessary for both trimethylation of H3K9 at genomic loci targeted by siRNAs and also inhibition of transcription elongation downstream of the siRNA-targeted sequences, most likely achieved by inducing early transcription termination. Human NRDE-2 is a ~ 130 kDa protein that contains, like the *C. elegans* and *S. pombe* orthologues, many


Half-A-Tetratricopeptide (HAT) repeats (SMART accession #: SM00386) [8], usually found in RNA-binding proteins and often involved in protein-protein interactions [9].

Importantly, Nrl1 has been found to protect the *S. pombe* genome from instability by resolving R loops and promoting DNA repair through homologous recombination (HR) [8]. R loops are conserved co-transcriptional structures that arise from hybridization of a nascent RNA with the DNA template, and are thought to cover ~5% of mammalian genomes [10]. While R loops function in several important cellular processes, such as immunoglobulin class switch recombination in activated B cells [11], mitochondria replication [12,13], protection against epigenetic silencing at promoters [10] and transcription termination [10,14,15], their persistence or formation at inappropriate locations can lead to mutations, DNA double-strand breaks (DSBs) and chromosome rearrangements causing genome instability [16,17]. Indeed, R-loop accumulation has been linked to many diseases, from cancer to neurological disorders [18,19].

In this study, we have investigated the role of NRDE-2 and an Mtr4/NRDE-2 complex in human cells. We first conducted co-immunoprecipitation (co-IP) experiments of FLAG-tagged derivatives of both NRDE-2 and Mtr4 stably expressed in HEK293 cells. We identified common partners that confirm the existence of a specific Mtr4/NRDE-2 complex that interacts with splicing factors, similar to its yeast counterpart. We also identified new interacting proteins, including several involved in chromatin remodeling/DNA damage response (DDR), associated with the proteasome, or are cytoskeletal proteins. Importantly, we provide

CONTACT Patricia Richard  pr2232@columbia.edu  Department of Biological Sciences, Columbia University, New York, NY 10027, USA

*These authors contributed equally to this work.

 Supplemental data for this article can be accessed [here](#).

© 2018 Taylor & Francis Group, LLC

evidence that NRDE-2 plays a role in the DDR. However, in contrast with Nrl1, DSBs that accumulate after NRDE-2 or Mtr4 depletion are not dependent on R-loop accumulation. Finally, we found that NRDE-2 and Mtr4 siRNA-mediated knockdowns (KDs) can modestly but significantly affect R-loop profiles at specific and distinct loci, suggesting an independent connection of NRDE-2 and Mtr4 with transcriptional regulation through R-loop formation/resolution.

2. Results

2.1. Identification of human Mtr4/NRDE-2-associated proteins

With the goal of understanding NRDE-2 function, we first set out to identify NRDE-2-interacting proteins. NRDE-2 was previously identified as an Mtr4-interacting protein [2,3]; however, proteins associated with Mtr4/NRDE-2 remain undetermined. To identify NRDE-2-interacting proteins, we prepared extracts from HEK293 cells stably producing NRDE-2-3FLAG with expression equivalent to endogenous NRDE-2 (Fig. S1A). Similar to its *S. pombe* counterpart [6], NRDE-2-3FLAG localizes to the nucleus (Fig. S1B). Considering the possible association of NRDE-2 with insoluble chromatin-

binding proteins, and to avoid obtaining possible indirect interactions mediated by DNA and/or RNA, we treated the extracts with Benzonase and RNase A (Fig. S1C). By treating with Benzonase, strong chromatin-binding proteins, such as histones, can be efficiently extracted even under physiological salt concentrations [20]. Nuclease-treated lysates were used for co-IP with FLAG antibody, and co-IPed proteins were eluted with 3×FLAG peptide (silver stain; Figure 1(a)) and subjected to mass spectrometry (MS) (Table S1). As expected, Mtr4 was detected with the highest peptide counts [2,3,6,8]. This interaction was further validated with another NRDE-2-3FLAG co-IP experiment performed under the same conditions as the one subjected to MS (Fig. S1D). Interestingly, despite the high abundance of Mtr4 in the MS analysis, no exosome subunits were identified. Additionally, no NEXT (RBM7 and ZCCHC8), TRAMP (PAPD5 and ZCCHC7) or Mtr4/ZFC3H1 (or PAXT) subunits were detected, confirming that the Mtr4/NRDE-2 complex is distinct from these complexes. Unlike in *C. elegans* [7], but similar to *S. pombe* [6,8,21], no RNAi factors were obtained.

Next, we wished to determine which NRDE-2-associated proteins also co-purify with Mtr4. Since Mtr4 exists in multiple distinct protein complexes, we performed gel filtration before co-IP/MS analysis. Protein complexes in

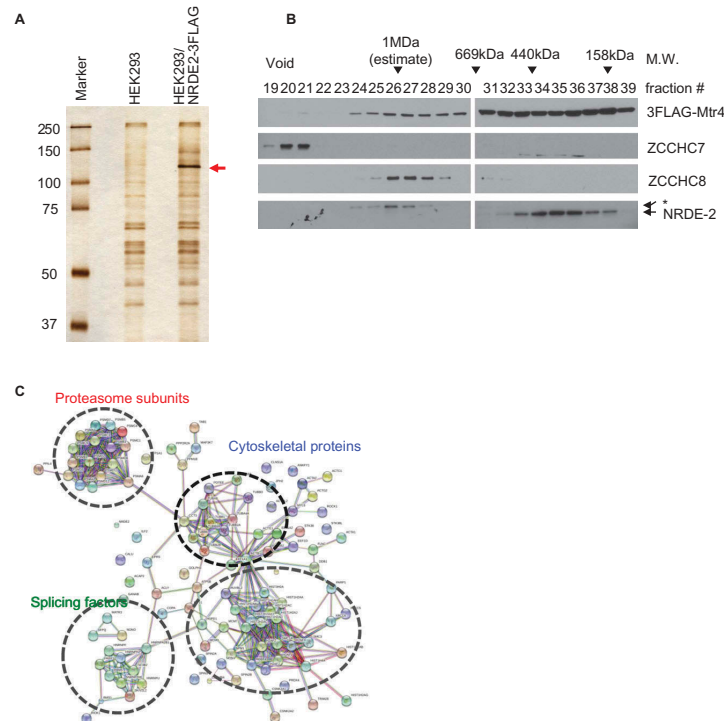


Figure 1. Identification of NRDE-2-interacting proteins by co-IP and mass spectrometry.

(a) Silver staining of FLAG co-IPs from control and NRDE-2-3FLAG stable HEK293 cell lysates used for mass spectrometry. (b) Fractions from Superose 6 gel filtration of 3FLAG-Mtr4-expressing HEK293 cell lysates treated with Benzonase/RNase A were analyzed by WB, using antibodies against proteins shown on the right. Approximate molecular sizes are indicated at the top and fractions pooled at the bottom. The asterisk indicates a non-specific band. (c) The protein-protein interaction network among the proteins shared in NRDE-2-3FLAG and 3FLAG-Mtr4 co-IP that are absent in HEK293 control co-IP was constructed using STRING v10.5 (<http://string-db.org>) with the high confidence setting. Disconnected nodes are not shown in the network.

Benzonase/RNase A-treated extracts prepared from the stable HEK293 cell line expressing 3FLAG-Mtr4 [3] were fractionated by size using a Superose 6 column (Figure 1(b)), and the NRDE-2-containing fractions (158–669 kDa) were used for anti-FLAG co-IP followed by MS. We previously reported that 3FLAG-Mtr4 was broadly distributed between 158 kDa and void fractions in RNase A-treated extracts [3]. However, 3FLAG-Mtr4 in >1 MDa ~ void fractions was absent following Benzonase treatment (compare Figure 1(b) with Figure 2(b) in [3]). There were no significant changes in the distribution patterns of ZCCHC7, ZCCHC8 and NRDE-2 due to Benzonase (Figure 1(b), see Figure 2(b) in [3]). MS analysis successfully detected known Mtr4 partners, including NRDE-2 and exosome subunits (Table S1). Among the proteins that were not detected in a HEK293 control, 98 were detected in both NRDE-2 and Mtr4 interactomes. We note that 103 out of 252 (40.8%) of the proteins from the Benzonase/RNase A-treated 3FLAG-Mtr4 co-IP/MS overlap with the RNase A only-treated 3FLAG-Mtr4 co-IP/MS described in our previous work [3], despite the absence of most of the DNA-associated proteins (presence of benzonase) in our current analysis.

To gain better insight into the Mtr4/NRDE-2-interacting proteins, we uploaded the shared proteins to STRING v10.5 database [22] and created high confidence interaction networks (Figure 1(c)). The proteins were clustered largely into four groups: splicing factors, histone/chromatin/DDR-related proteins, cytoskeletal proteins and proteasome subunits. The interaction of Mtr4/NRDE-2 with splicing factors is reminiscent of the fission yeast Mtl1/Nrl1 complex, although an additional conserved protein, Ctr1, stably interacts with Mtl1/Nrl1 and splicing factors [6,8,21], but its human homolog CCDC174 was absent in both NRDE-2-

3FLAG and 3FLAG-Mtr4 co-IP/MS analyses. Altogether these co-IP experiments highlight the existence of a human Mtr4/NRDE-2 complex distinct from any previously identified Mtr4 complexes, and which appears to share several features with the fission yeast complex.

2.2. NRDE-2 prevents DNA damage accumulation

For some time, splicing factors have been linked to genome stability maintenance through the prevention of R-loop accumulation [23]. Importantly, deletion of *Nrl1* in fission yeast leads to an accumulation of R loops and DSBs [8]. Since we found NRDE-2 interacting with splicing factors, and to begin to test whether NRDE-2 is also involved in maintaining genome stability in human cells, we monitored, by western blot (WB) and immunofluorescence (IF), levels of γ H2AX, a marker of DSBs, following siRNA-mediated KD of NRDE-2 in HeLa cells. Strikingly, the IFs, and quantification of γ H2AX signal, showed a drastic increase (~ two-fold relative to control) of DNA damage in KDed cells compared to non-transfected cells or cells transfected with a control siRNA (NC) (Figure 2 + Fig. S2). Since NRDE-2 interacts with Mtr4, and Mtr4 KO in mouse B cells shows an increase in R loops at certain loci [24], we asked whether Mtr4 KD also leads to DSBs. Indeed, Mtr4 KD also resulted in a significant increase in DSBs as shown by the quantification of the γ H2AX signal (Fig. S2). These data suggest that the Mtr4/NRDE-2 complex is involved in DNA damage prevention or repair.

2.3. DNA damage in NRDE-2 and Mtr4 depleted cells is R-loop independent

Since Nrl1 and mouse Mtr4 have both been linked to R-loop metabolism [8,24], we next asked whether the DSBs detected after

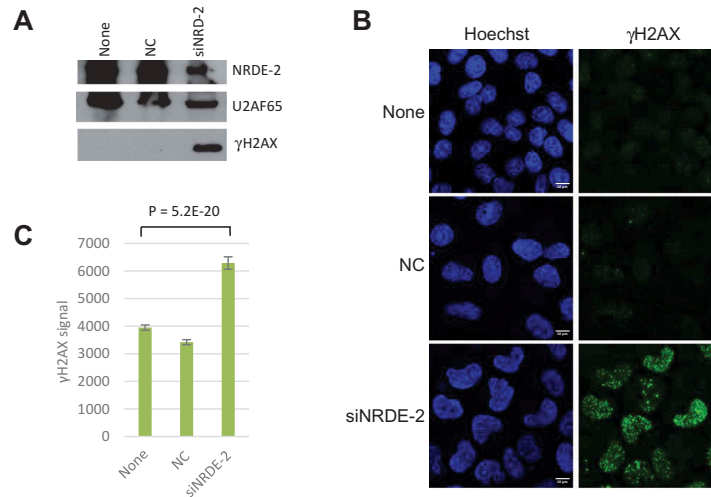


Figure 2. NRDE-2 KD leads to increased DSBs.

(a) γ H2AX protein levels in non-transfected HeLa cells (None), cells transfected with an siRNA control (NC) or siNRDE-2 analyzed by WB. U2AF65 is used as a loading control. (b) γ H2AX signal visualized by IFs in HeLa cells as in A. (c) Quantification of images as in B. n = 267 cells, SE is shown.

NRDE-2 and Mtr4 KDs were caused by an increase in R-loop levels. To address this, we overexpressed GFP-RNase H1 in HeLa cells KDed for Mtr4 or NRDE-2 and quantified the γ H2AX levels by WB analysis (Fig. S3). Confirming our IF data, Mtr4 and NRDE-2 KDs led to an accumulation of DSBs as shown by an increased γ H2AX protein level compared to control cells. Interestingly, Mtr4 KD led to ~ 4 times more γ H2AX than did NRDE-2 KD, a result that was not apparent by IF (Fig. S2), and which is discussed below. Importantly, RNase H1 overexpression did not significantly decrease γ H2AX levels in NRDE-2 or Mtr4 KDed cells. To confirm these data, we performed IF to monitor γ H2AX and GFP-RNase H1 signals simultaneously (Figure 3). The IF images show that γ H2AX and high levels of RNase H1 expression can co-exist (see cells with white arrows) and that consequently there was no statistical difference in γ H2AX staining after NRDE-2 KD between cells expressing RNase-H1 or not, as calculated by the quantification of the γ H2AX signal in the RNase H1-expressing cells. However, RNase H1 overexpression had a significant and unexpected effect on DSB accumulation in Mtr4 KDed cells, as the γ H2AX signal in these cells actually increased by 30% after RNase H1 overexpression. While we also detect a 10% increase of the γ H2AX signal after RNase H1 overexpression in NRDE-2 KDed cells, this difference doesn't appear significant (Figure 3, left histogram). These data confirm that DNA damage induced by NRDE-2 and/or Mtr4 KDs is not due to accumulation of R loops.

We also examined a possible connection between R-loop formation and NRDE-2 and/or Mtr4 more directly. Specifically, we performed DRIP assays in HeLa cells after NRDE-2 and Mtr4 KDs, using the S9.6 antibody to IP RNA: DNA hybrids [25,26]. We examined R-loop formation at known positive loci, the *RPL13A* 3' end and the *BACT* 5' pause site, as well as an R-loop negative locus, *EGR1* [10,14] (Figure 4(a-c)). While the DRIP assay detected the expected percentage of input at *RPL13A* ($\sim 10\%$), *BACT* ($\sim 2\%$) and *EGR1* ($< 0.1\%$) loci in untransfected and siRNA control (NC) transfected cells [10], R-loop enrichment increased by 20% compared to control at *RPL13A* and *BACT* after NRDE-2 KD. In contrast, R-loop enrichment at the 5' pause site of *BACT* did not change after Mtr4 KD, while $\sim 20\%$ less R-loop signal was detected at *RPL13A*. Although we only examined a very limited number of loci, the DRIP data suggests that while DSB accumulation after NRDE-2 or Mtr4 KDs is R-loop independent, it is very likely that NRDE-2 and Mtr4 play minor but significant and independent roles in R-loop resolution or formation at specific genes, possibly regulating their transcription.

Since Mtr4 has recently been shown to regulate lncRNAs turnover as part of the Mtr4/ZFC3H1 (or PAXT) complex, we performed NRDE-2 ChIP assays to ask whether NRDE-2 could also play a role at genomic loci including those associated with lncRNAs expression. We indeed found that NRDE-2 binds various loci including eRNA (*eRNA17*), ptRNAs intronic PAS (*DAP*-PAS and *TMED4*-PAS), PROMPT (*pro-RBM39*) as well as the promoter of *CSTF3* (Figure 4(d); note that NRDE-2 and Mtr4 KDs did not significantly affect Mtr4 or NRDE-2 protein levels, respectively (Figure 4(c))). Importantly, NRDE-2 KD led to a weak but general increase in *eRNA17*, ptTMED4, ptDAP, proRBM39

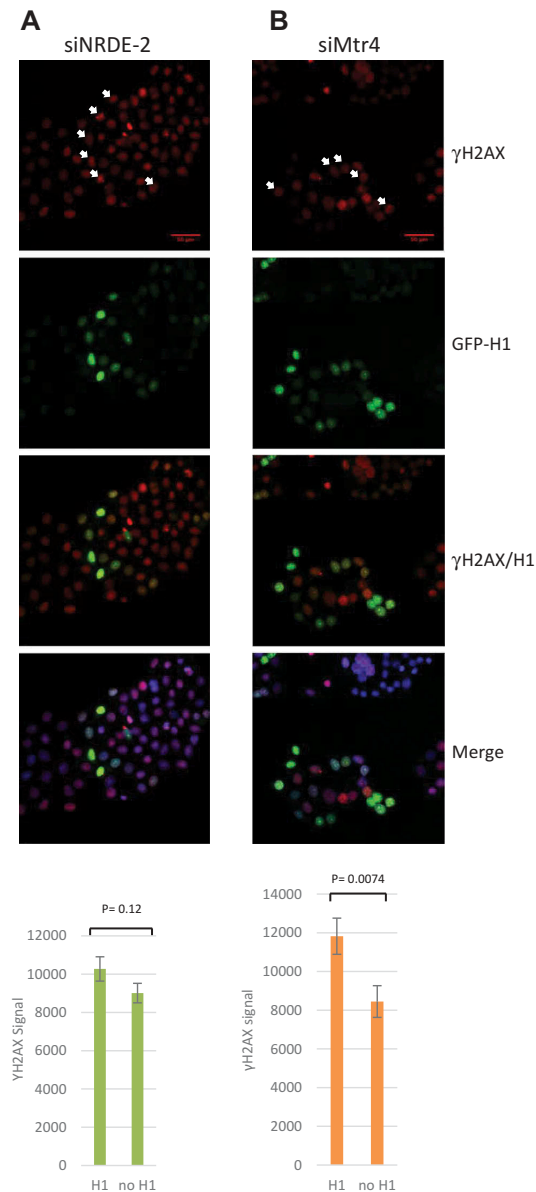


Figure 3. DSBs induced by NRDE-2 or Mtr4 KDs are R-loop independent. (a) IF of γ H2AX signal in NRDE-2 KDed HeLa cells for 72h and transfected (H1) or not (no H1) with GFP-RNase H1 for 48h. Signal quantifications are shown at the bottom. $n = 189$ cells, SE shown. (b) Quantification of γ H2AX signal in Mtr4 KDed cells expressing GFP-RNase H1 (H1) or not (no H1). $n = 78$ cells, SE shown. White arrows show cells expressing GFP-RNase H1 and high level of γ H2AX signal.

lncRNAs (Fig. S4). *CSTF3* mRNA levels also increased significantly after NRDE-2 KD (three-fold relative to control), however *CSTF3* ptRNA decreased (Fig. S4). While these data show some disparity, they indicate that NRDE-2 might function as a negative regulator of gene expression. While we could not detect R loops at *eRNA17* loci, we were able to detect DRIP signals at ptRNAs intronic PAS (*DAP* and

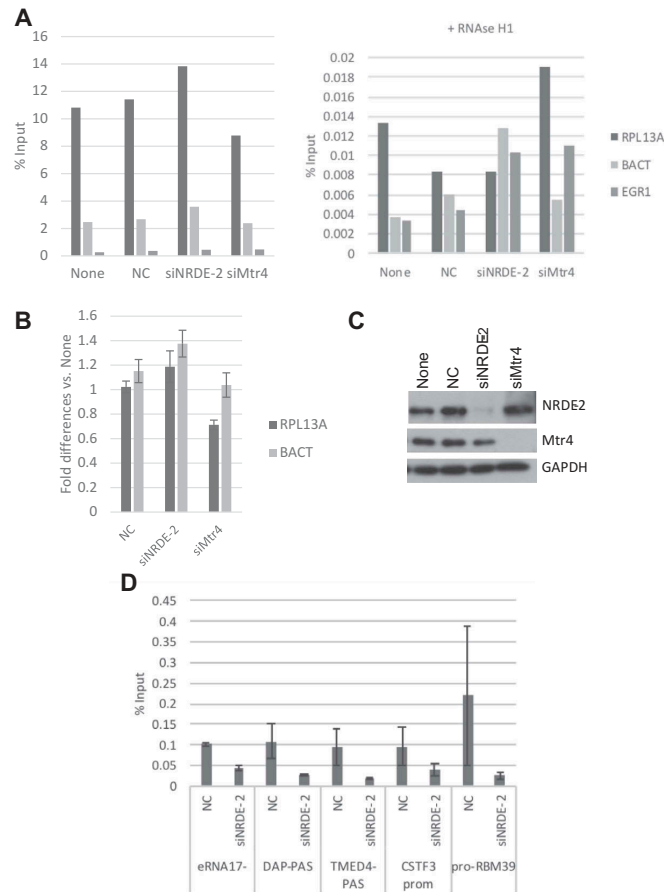


Figure 4. R-loop levels at *RPL13A*, *BACT* and *EGRI* are not affected by NRDE-2 or Mtr4 KD.

(a) DRIP assays were performed in HeLa cell after 72 h of siRNAs transfection. DRIP signal was also measured after RNase H1 treatment (right). (b) DRIP signal was normalized to the signal in non-transfected cell. Error bars represent the average of two different experiments. SD shown. (c) NRDE-2 and Mtr4 KDs confirmation by WB. None = no transfection, NC: siRNA control transfection. (d) NRDE-2 ChIP at various loci. ChIP assays were performed in HeLa cell after 72 h of siRNAs transfection. n = 2

TMED4). These did not, however, show significant differences after NRDE-2 or Mtr4 KD compared to controls (data not shown), indicating that the potential NRDE-2 negative regulation of eRNA17⁻, ptDAP and ptTMED4 expression is R-loop independent. These data indicate that it is very likely that NRDE-2 plays a role in transcription regulation that in some cases might involve R-loop resolution or formation (Figure 5).

3. Discussion

In this study, we investigated the functions and properties of the conserved human NRDE-2 protein. We identified several NRDE-2 associated proteins as well as proteins associated with one of its previously identified partners, the RNA helicase Mtr4. As strong evidence that NRDE-2 and Mtr4 are core components of a novel NRDE-2/Mtr4 complex, we found that the two proteins share many

interacting factors that can be classified into four main categories. We identified proteins from the chromatin remodeling/DDR pathways, splicing factors, cytoskeletal proteins and proteasome subunits. The association of NRDE-2 with splicing factors has previously been observed in fission yeast, where *nrl1* deletion is indeed associated with splicing defects of a subset of genes [6,8]. While it is therefore likely that human NRDE-2 also plays a role in pre-mRNA splicing we have not investigated this possibility.

Similar to the *nrl1Δ* strain, we found that NRDE-2 depleted cells accumulate DSBs. Because genome instability in *nrl1Δ* yeast has been attributed to an increase in R loops [8], we were in fact very surprised to find that the DSB increase after NRDE-2 or Mtr4 depletion was R-loop independent. Indeed, overexpression of RNase H1, which should eliminate R loops, did not reduce DNA damage after NRDE-2 or Mtr4 KD. On the contrary, RNase H1 overexpression appears to exacerbate DSBs, at least

in the Mtr4 KDed cells. This data was however only observed by IF and its significance is not yet clear. While WBs would seem likely to be more quantitative than IF, we believe that the stress level (passage number, harvesting time, etc.) and/or cell cycle might have a significant impact on DSB accumulation. While we noticed that γ H2AX levels can vary from one experiment to another, levels were always significantly higher after either NRDE-2 or Mtr4 KD. Nevertheless, our IF data showed that under certain circumstances, RNase H1 overexpression can trigger an increase in γ H2AX signal in an Mtr4 KD background.

While we know that RNase H1 eliminates R loops, we also know that its overexpression or inhibition can have other effects [27], which include changes in gene expression and inhibition of HR-mediated DSB repair, at least in fission yeast [28,29]. It is thus possible that RNase H1 overexpression affects expression of genes involved in the DDR or interferes with repair in the Mtr4 KD background. The latter scenario implies that R loops might be required for repair of some of the DSBs triggered by Mtr4 KD. Another explanation would involve a role for Mtr4 in RNA-mediated HR repair [30]. While we cannot eliminate the last two possibilities, we find them unlikely because, as discussed below, Mtr4- and NRDE-2-interacting proteins favor a function for the complex in NHEJ rather than HR. It has however been shown in budding yeast that RNA can be used as a homologous template to repair DSBs [31]. Considering this scenario, RNase H1 overexpression could interfere with the repair process, which could explain the observed γ H2AX signal increase in the Mtr4 KD background. However, despite the evidence that RNA oligos can function in repair of DSBs in humans [32], this kind of HR repair mechanism needs to be investigated more thoroughly.

How might Mtr4/NRDE-2 function in DNA repair? In *S. pombe*, it was proposed that the R loops formed in *nrl1 Δ* strains sequester HR factors that are then no longer available to repair damage, leading to an increase in DSBs [8]. While Nrl1 has not been found to interact with HR factors, we found that NRDE-2 and Mtr4 both associated with several DDR factors, including PARP1, Ku70 and Ku80 as well as many histones and chromatin remodeling factors. The heterodimer Ku is part of the NHEJ DSB repair machinery, functioning by recognizing the breaks [33]. PARP1 is involved in numerous aspects of the DDR and functions in both the NHEJ and HR pathways [34]. Thus, it appears that while NRDE-2 and Nrl1 both function in the DDR, they likely do so by distinct mechanisms. We suggest that DSBs occurring after NRDE-2 and Mtr4 KDs result from a defect in the NHEJ repair process due to a lack of break recognition by the heterodimer Ku70/80 and/or PARP1 we found interacting with both proteins (Figure 5). The interaction with histones and chromatin remodeling factors make perfect sense in that scenario since their presence and particular organization at DNA damage sites provides the environment necessary for efficient and proper repair [35].

In *C. elegans*, NRDE-2 is essential for the nuclear RNAi pathway, which involves siRNA-directed H3K9me. NRDE-2 is directed to nascent transcripts through association with an siRNA-incorporated Argonaut (AGO) protein, NRDE-3, to introduce H3K9 methylation and block RNAP II elongation [7]. Although no RNAi factors were identified in the NRDE-2 (Nrl1)-proteomics analysis performed by us (human) and others (yeast), RNAi factor-directed introduction of H3K9me is

observed in both organisms [15,36]. Notably, there are interconnections between R loops and the modification status of chromatin (reviewed by [37,38]). In human cells, R loops formed over RNAP II pause-site termination regions induce antisense transcription and consequent double-stranded RNA generation, which recruits RNAi factors such as DICER, AGOs, and the histone methyltransferase G9a to direct H3K9me at transcriptional termination sites [15]. In *S. pombe*, depletion of Nrl1 selectively abrogates RNAi-dependent heterochromatin assembly by decreasing H3K9me at HOODs [6]. The authors further suggested that RNAi-mediated heterochromatin assembly is defined by cryptic introns within HOODs and the spliceosome, since deletion of the cryptic introns or splicing factors resulted in reduced H3K9me3 levels [6]. Considering that splicing factor deficiency causes R-loop stabilization [23], it is possible that R-loop structures are enriched at specific loci such as HOODs in Nrl1-depleted cells and that H3K9 methylation is coupled with their resolution. It is then interesting to speculate that NRDE-2 is involved in R-loop resolution and H3K9 methylation at transcriptional termination sites, since NRDE-2 KD leads to an increase of R loops at the 3' end of *RPL13A* and *BACT*.

Several studies have suggested that changes in NRDE-2 expression could be involved in human health and disease. Array-based genome-wide copy number aberration analyses have suggested that *NRDE-2* (*14q32.11*) haploinsufficiency can be associated with schizophrenia [39] and can be observed in circulating tumor cells that have been detected in the blood of patients with metastatic melanoma [40]. Moreover, an intronic single-nucleotide polymorphism (SNP) in *NRDE-2* (refSNP Cluster Report ID: rs4904670) is associated with reduced lifespan, physiological aging changes, and major diseases like cancer and coronary heart disease [41]. Future analyses may uncover the possible link between the function of NRDE-2 and/or Mtr4/NRDE-2 and disease.

We identified here yet another complex containing Mtr4 that surprisingly has a function distinct from RNA surveillance. While our data reveal an R-loop independent role of the Mtr4/NRDE-2 complex in the DDR, each protein seems to individually (or part of two different complexes) affect R-loop levels at specific and distinct genes. It will be of great interest in the near future not only to dissect the molecular mechanisms that allow the Mtr4/NRDE-2 complex to protect the cell from DNA damage, but also to understand the independent role each protein plays at specific genes where R loops form. Our work shows that Mtr4 has an even broader role than expected from previous studies. In fact, it is very likely, according to our MS data, that the Mtr4/NRDE-2 complex participates in many other aspects of cell physiology through its interaction with subunits of the proteasome and cytoskeleton proteins (Figure 5).

4. Materials and methods

4.1. Gel filtration

Superose 6 gel filtration was performed as previously described [3] except that cell lysates were treated with >250 U/mL Benzonase (Sigma: #E1014) and 10 μ g/mL RNase A (Sigma: #R5250) and that eluates were collected every 5 min (1 mL, flow rate = 0.2 mL/min).

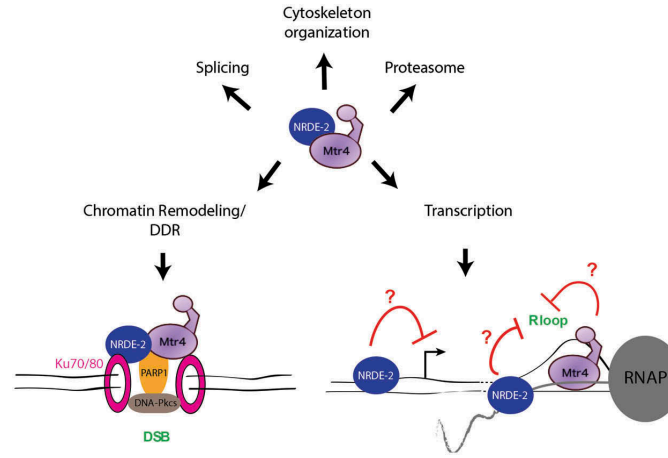


Figure 5. Possible functions of the Mtr4/NRDE-2 complex in various cellular processes.

The role of the Mtr4/NRDE-2 complex in the DDR might involve interaction with the Ku70/80 heterodimer, which recognizes DSBs. Its interaction with PARP1 may stimulate the activation of the DNA-dependent protein kinase catalytic subunit (DNA-Pkcs), a NHEJ factor that also interacts with the Ku complex. NRDE-2 and Mtr4 appear to have independent roles in transcription regulation. While NRDE-2 can have a negative effect on transcription, it seems to resolve R loops at the 3' end of certain genes. According to [24], Mtr4 might also promote R-loop resolution, at least in certain circumstances.

4.2. Co-immunoprecipitation (co-IP) and mass spectrometry (MS)

co-IP using FLAG antibody and MS analysis were performed as previously described [3] except that HEK293 and HEK293/NRDE-2-3FLAG cells were lysed in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40, 1x protease inhibitor cocktail (Biotools), 10µg/mL RNase A (Sigma: #R5250) and >250U/mL Benzonsae (Sigma: #E1014).

4.3. Immunofluorescence (IF)

γH2AX IF were performed with HeLa cells after siRNAs transfection at 20 nM with RNAiMAX (Invitrogen: #13778) for 72h, methanol fixation for 10 mins and permeabilization with acetone for 1 min. FLAG IF in HEK293 cells were performed after fixation with 4% paraformaldehyde for 10 mins followed by fixation with 0.5% Triton X 100 for 10 mins. γH2AX antibody (Cell signaling Technology: #2577S) was used at 1:200 and anti-FLAG (Sigma: #F1804) at 1:250 for 2h. Secondary anti-rabbit Alexa 488 (Invitrogen: #A11008) and anti-mouse Alexa 568 (#A11031) at 1:500 for 1h at RT. Images were acquired using Zeiss LSM 700 confocal microscope and 20x/1.4 and 40x/1.4 oil objectives were used. Nuclear immunofluorescence signals were analyzed with ImageJ. 200 ng of GFP-RNase H1 and empty plasmid was transfected with Lipofectamine 2000 (Invitrogen: #11668) 24h after siRNA transfection for 48h.

4.4. siRNAs transfection and Western Blots

HeLa cells were not transfected (Ctrl) or transfected with an siRNA control (NC: TTCTCCGAACGTGTCACGT, GenePharma), siNRDE-2 (GCAAGCAGGUUGAACGCUA), siNRDE-2#2 (ID: S30064, Cat#: 4,427,037, Thermofisher),

siMtr4 (CAAUUAAGGCUCUGAGUAA, GenePharma) for 72h at 20 nM with RNAiMAX (Invitrogen: #13778) prior extracts. The following antibodies were used: γH2AX (Cell signaling Technology: #2577S), U2AF65 (Sigma: #U4758), NRDE-2 (proteintech: #24968-I-AP), Mtr4 (Bethyl: #A300-614A), GFP (abm: #G095), GAPDH (Sigma: #G9545), FLAG (Sigma: #F1804). Western Blot quantifications were performed using ImageJ.

4.5. DRIP assay and qPCR

DRIP was performed as described in Ginno et al. [26]. Briefly, HeLa cells transfected with 20 nM siRNAs for 72h were lysed O/N in SDS and proteinase K. 50 µl of extracted gDNA was then digested O/N with a cocktail of restriction enzymes at 40 U each (HindIII/EcoRI/BsrGI/XbaI/SspI). After phenol/chloroform extraction and EtOH precipitation, 4.4 µg of digested DNA was treated O/N with 3 µl of Ribonuclease H (NEB, #M0297) as negative control. Treated and un-treated DNA were IPed O/N at 4°C with 10 µg of S9.6 antibody in binding buffer (10 mM NaPO4 pH 7.0, 140 mM NaCl, 0.5% Triton X-100). Next day, agarose A/G beads (Pierce #20421) were added for 2 hours. After washes, IPs were eluted at 55°C for 45 min in 250 µl elution buffer (50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS) supplemented with 7 µl of Proteinase K at 20 mg/ml. DNA was analyzed by qPCR using the following primers: RPL13A (3' of the gene): F: AGGTGC CTTGCTCACAGAGT, R: GGTTCATTGCCCTCATTAC, βactin (5' pause site): F: TTACCCAGAGTGCAGGTGTG, R: CCCCAATAAGCAGGAACAGA, EGR1 (intergenic region downstream of EGR1): F: GAACGTTACAGCCTCGTTCTC, R: GGAAGGTGGAAGGAAACACA.

The primers used to detect ptRNAs, full-length mRNA and proRBM39 after NRDE-2 KD are described in [3]. Primers for eRNA17- are the same as the ones used for ChIP.

4.6. ChIP assay

A confluent 10 cm dish of HeLa cells were cross-linked in 1% formaldehyde for 10 minutes at room temperature (RT). Cross-links were quenched in 125 mM glycine for 5 mins, cells were rinsed in PBS and then collected by centrifugation (500xg for 5 mins). Cells were collected in 400 μ l RIPA buffer (150 mM NaCl, 1% NP40, 0.5% DOC, 0.1% SDS, 50 mM Tris.Cl pH8, 5 mM EDTA pH 8) supplemented with PMSF and a protease inhibitor cocktail (0.2 mM Pepstatin A, 72 μ M Leupeptin, 26 μ M Aprotinin) and sonicated (30s x 10 times using a bioruptor sonicator). Extracts were clarified by centrifugation at 13000xg for 15 mins. NRDE-2 ChIP was performed O/N at 4°C with 2 μ g of antibody (proteintech: #24968-I-AP) and 20 μ l of protein A/G sepharose (Invitrogen). No antibody control was performed in parallel. Beads were washed 3x in RIPA buffer without SDS. Immune complexes were eluted in 0.1M NaHCO₃/1%SDS (15 mins rotation at RT). Cross-links were reversed for 5 hours at 65°C (250mM NaCl and 1 μ g RNase A). DNA was purified using QIAquick PCR purification kit (#28106) from Qiagen. Generally, 1/50th was used for each q-PCR reaction. The primers used for DAP int PAS, TMED4 int PAS and proRBM39 detection are described in [3]. eRNA17- and CSTF3 primers are (F: GAGCCATGGATGGGTGATAA, R: AACCCATCTTGTGAGGCAGA) and (F: ACTGATTTGGGGGTGGTTTT, R: GGCCTCAGCTGATTACAACG), respectively.

Acknowledgments

We thank Robert Crouch for the GFP-RNase H1 plasmid, Frédéric Chédin and Lionel Sanz for the S9.6 antibody and for their insightful advices and comments to perform DRIP assay. We also thank Juan Izarrry-Cole for his help with the IF quantifications.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Muscular Dystrophy Association [MDA 377780]; National Institutes of Health [P41 GM 103533]; National Institutes of Health [R35 GM 118136];

ORCID

Patricia Richard  <http://orcid.org/0000-0002-9515-4509>

Koichi Ogami  <http://orcid.org/0000-0001-9380-9666>

References

- Ogami K, Chen Y, Manley JL. RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Non-Coding RNA*. 2018 Mar 11;4(1):8. PubMed Central PMCID: PMC5886371.
- Lubas M, Christensen MS, Kristiansen MS, et al. Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell*. 2011 Aug 19;43(4):624–637. PubMed PMID: 21855801.
- Ogami K, Richard P, Chen Y, et al. An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev*. 2017 Jul 21. PubMed PMID: 28733371; PubMed Central PMCID: PMC5558927. DOI:10.1101/gad.302604.117
- Meola N, Domanski M, Karadoulama E, et al. Identification of a nuclear exosome decay pathway for processed transcripts. *Mol Cell*. 2016 Nov 3;64(3):520–533. PubMed PMID: 27871484.
- Yamanaka S, Mehta S, Reyes-Turcu FE, et al. RNAi triggered by specialized machinery silences developmental genes and retro-transposons. *Nature*. 2013 Jan 24;493(7433):557–560. PubMed PMID: 23151475; PubMed Central PMCID: PMC3554839.
- Lee NN, Chalamcharla VR, Reyes-Turcu F, et al. Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell*. 2013 Nov 21;155(5):1061–1074. PubMed PMID: 24210919; PubMed Central PMCID: PMC3974623.
- Guang S, Bochner AF, Burkhart KB, et al. Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature*. 2010 Jun 24;465(7301):1097–1101. PubMed PMID: 20543824; PubMed Central PMCID: PMC2892551.
- Aronica L, Kasperek T, Ruchman D, et al. The spliceosome-associated protein Nrl1 suppresses homologous recombination-dependent R-loop formation in fission yeast. *Nucleic Acids Res*. 2016 Feb 29;44(4):1703–1717. PubMed PMID: 26682798; PubMed Central PMCID: PMC4770224.
- Preker PJ, Keller W. The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem Sci*. 1998 Jan;23(1):15–16. PubMed PMID: 9478129.
- Sanz LA, Hartono SR, Lim YW, et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol Cell*. 2016 Jun 29. PubMed PMID: 27373332. DOI:10.1016/j.molcel.2016.05.032
- Yu K, Chedin F, Hsieh CL, et al. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol*. 2003 May;4(5):442–451. PubMed PMID: 12679812.
- Xu B, Clayton DA. RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. *The EMBO Journal*. 1996 Jun 17;15(12):3135–3143. PubMed PMID: 8670814; PubMed Central PMCID: PMC450256.
- Pohjoismaki JL, Holmes JB, Wood SR, et al. Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *J Mol Biol*. 2010 Apr 16;397(5):1144–1155. PubMed PMID: 20184890; PubMed Central PMCID: PMC2857715.
- Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell*. 2011 Jun 24;42(6):794–805. PubMed PMID: 21700224; PubMed Central PMCID: PMC3145960. eng.
- Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*. 2014 Oct 5. PubMed PMID: 25296254. DOI:10.1038/nature13787
- Aguilera A, Garcia-Muse T. R loops: from transcription byproducts to threats to genome stability. *Mol Cell*. 2012 Apr 27;46(2):115–124. PubMed PMID: 22541554; eng.
- Chan YA, Hieter P, Stirling PC. Mechanisms of genome instability induced by RNA-processing defects. *Trends Genet*. 2014 Jun;30(6):245–253. PubMed PMID: 24794811; PubMed Central PMCID: PMC4039741.
- Richard P, Manley JL. R loops and links to human disease. *J Mol Biol*. 2016 Sep 4. PubMed PMID: 27600412; PubMed Central PMCID: PMC5478472. DOI:10.1016/j.jmb.2016.08.031
- Groh M, Gromak N. Out of balance: R-loops in human disease. *PLoS Genet*. 2014 Sep;10(9):e1004630. PubMed PMID: 25233079; PubMed Central PMCID: PMC4169248.
- Anindya R, Aygun O, Svejstrup JQ. Damage-induced ubiquitylation of human RNA polymerase II by the ubiquitin ligase Nedd4, but not Cockayne syndrome proteins or BRCA1. *Mol Cell*. 2007 Nov 9;28(3):386–397. PubMed PMID: 17996703.
- Zhou Y, Zhu J, Schermann G, et al. The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat Commun*. 2015 May 20;6:7050. PubMed PMID: 25989903; PubMed Central PMCID: PMC4455066.

22. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015 Jan;43(Database issue):D447–D452. PubMed PMID: 25352553; PubMed Central PMCID: PMC4383874.
23. Li X, Manley JL. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell.* 2005 Aug 12;122(3):365–378. PubMed PMID: 16096057; eng.
24. Lim J, Giri PK, Kazadi D, et al. Nuclear proximity of Mtr4 to RNA exosome restricts DNA mutational asymmetry. *Cell.* 2017 Apr 20;169(3):523–537 e15. PubMed PMID: 28431250; PubMed Central PMCID: PMC5515252.
25. Boguslawski SJ, Smith DE, Michalak MA, et al. Characterization of monoclonal antibody to DNA:RNA and its application to immunodetection of hybrids. *J Immunol Methods.* 1986 May 1;89(1):123–130. PubMed PMID: 2422282.
26. Ginno PA, Lott PL, Christensen HC, et al. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell.* 2012 Mar 30;45(6):814–825. PubMed PMID: 22387027; PubMed Central PMCID: PMC3319272. eng.
27. Vanoosthuysse V. Strengths and weaknesses of the current strategies to map and characterize R-loops. *Non-Coding RNA.* 2018;4(2):9.
28. Hartono SR, Malapert A, Legros P, et al. The affinity of the S9.6 antibody for double-stranded RNAs impacts the accurate mapping of R-loops in fission yeast. *J Mol Biol.* 2017 Dec 28. PubMed PMID: 29289567. DOI:10.1016/j.jmb.2017.12.016
29. Ohle C, Tesorero R, Schermann G, et al. Transient RNA-DNA hybrids are required for efficient double-strand break repair. *Cell.* 2016 Nov 3;167(4):1001–1013 e7. PubMed PMID: 27881299.
30. Meers C, Keskin H, Storici F. DNA repair by RNA: templated, or not templated, that is the question. *DNA Repair.* 2016 Aug;44:17–21. PubMed PMID: 27237587; PubMed Central PMCID: PMC4958532.
31. Keskin H, Shen Y, Huang F, et al. Transcript-RNA-templated DNA recombination and repair. *Nature.* 2014 Nov 20;515(7527):436–439. PubMed PMID: 25186730; PubMed Central PMCID: PMC4899968.
32. Shen Y, Nandi P, Taylor MB, et al. RNA-driven genetic changes in bacteria and in human cells. *Mutat Res.* 2011 Dec 1;717(1–2):91–98. PubMed PMID: 21515292.
33. Mahaney BL, Meek K, Lees-Miller SP. Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J.* 2009 Feb 1;417:639–650.
34. Ray Chaudhuri A, Nussenzweig A. The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat Rev Mol Cell Biol.* 2017 Oct;18(10):610–621. PubMed PMID: 28676700.
35. Escargueil AE, Soares DG, Salvador M, et al. What histone code for DNA repair? *Mutat Res.* 2008 Mar–Apr;658(3):259–270. PubMed PMID: 18296106.
36. Holloch D, Moazed D. RNAi in fission yeast finds new targets and new ways of targeting at the nuclear periphery. *Genes Dev.* 2012 Apr 15;26(8):741–745. PubMed PMID: 22508721; PubMed Central PMCID: PMC3337448.
37. Santos-Pereira JM, Aguilera A. R loops: new modulators of genome dynamics and function. *Nat Rev Genet.* 2015 Oct;16(10):583–597. PubMed PMID: 26370899.
38. Chedin F. Nascent connections: R-loops and chromatin patterning. *Trends Genet.* 2016 Dec;32(12):828–838. PubMed PMID: 27793359; PubMed Central PMCID: PMC5123964.
39. Maiti S, Kumar KH, Castellani CA, et al. Ontogenetic de novo copy number variations (CNVs) as a source of genetic individuality: studies on two families with MZD twins for schizophrenia. *PLoS One.* 2011 Mar 2;6(3):e17125. PubMed PMID: 21399695; PubMed Central PMCID: PMC3047561.
40. Chiu CG, Nakamura Y, Chong KK, et al. Genome-wide characterization of circulating tumor cells identifies novel prognostic genomic alterations in systemic melanoma metastasis. *Clin Chem.* 2014 Jun;60(6):873–885. PubMed PMID: 24718909.
41. Yashin AI, Wu D, Arbeeve LS, et al. Genetics of aging, health, and survival: dynamic regulation of human longevity related traits. *Front Genet.* 2015;6:122. PubMed PMID: 25918517; PubMed Central PMCID: PMC4394697.

Figure S1

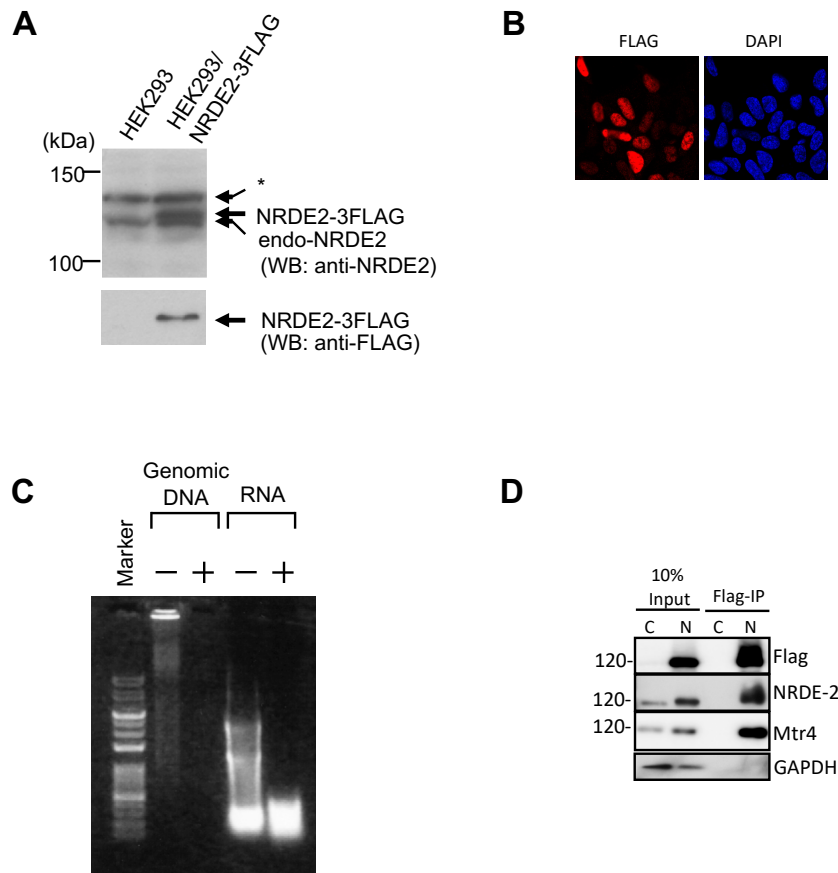


Figure S2

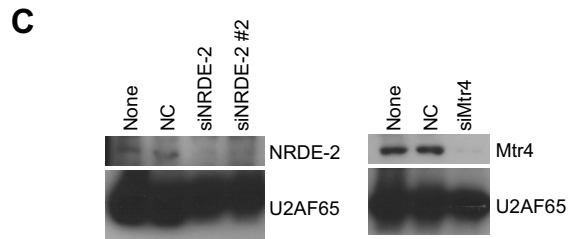
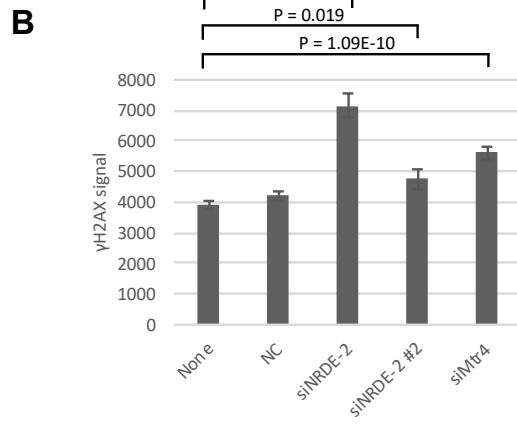
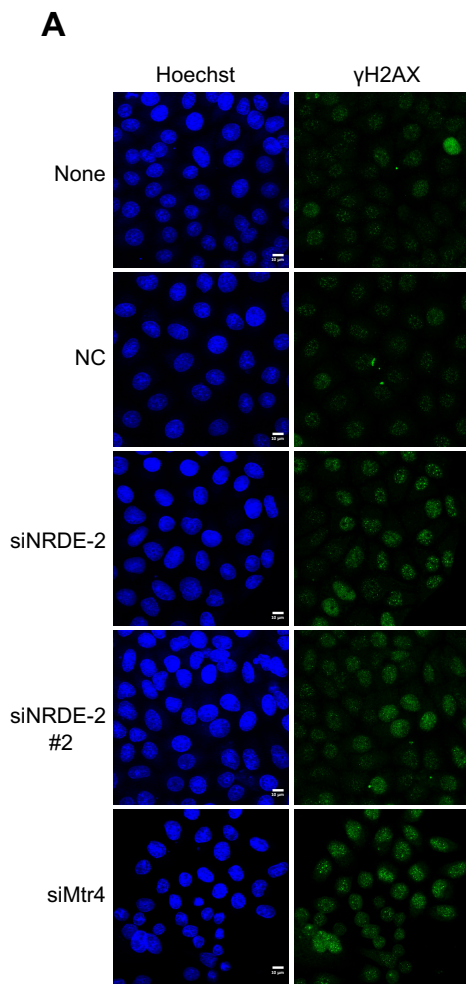
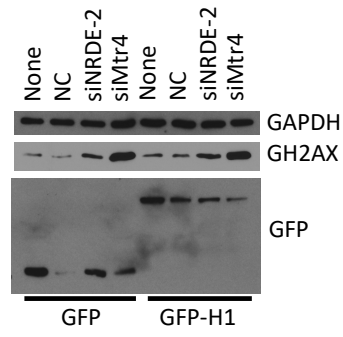


Figure S3

A



B

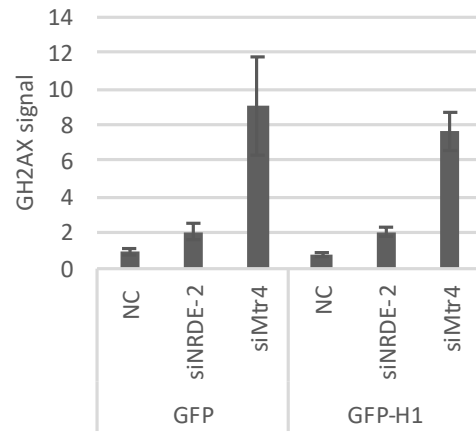
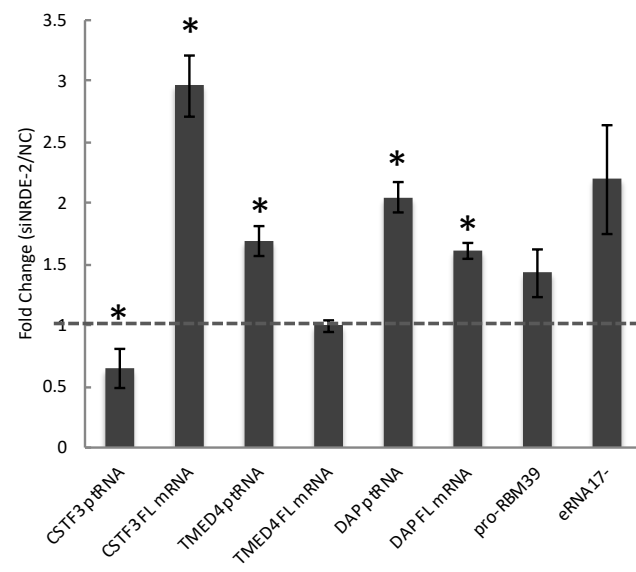


Figure S4

A



B

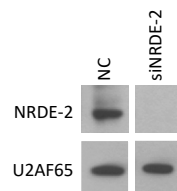


Table S1. Proteins identified in co-IP/MS analysis

Protein Name	HEK293	NRDE2-3FLAG	3FLAG-Mtr4
Mtr4	0 (0%)	228 (41.3%)	454 (49.5%)
NRDE2	0 (0%)	226 (45.7%)	57 (23.1%)
STK38	0 (0%)	34 (36.8%)	4 (14.6%)
SPIN1	0 (0%)	25 (39.3%)	50 (67.6%)
PRPF31	0 (0%)	24 (21.2%)	34 (39.5%)
CCT3	0 (0%)	23 (42.2%)	32 (40.6%)
C11orf84	0 (0%)	20 (30.7%)	80 (51.4%)
TUBB2B	0 (0%)	18 (25.4%)	38 (20.2%)
TUBB2A	0 (0%)	18 (25.4%)	38 (20.2%)
TUBB3	0 (0%)	18 (23.3%)	24 (20.9%)
EEF1A1	0 (0%)	17 (27.7%)	15 (23.4%)
SPIN2A	0 (0%)	17 (8.1%)	6 (20.2%)
SPIN2B	0 (0%)	17 (8.1%)	6 (20.2%)
EEF1A1P5	0 (0%)	16 (22.7%)	15 (23.4%)
ANKFY1	0 (0%)	14 (9.8%)	55 (37.3%)
ATP5B	0 (0%)	12 (25.5%)	3 (9.1%)
TUBB4A	0 (0%)	11 (27.7%)	45 (34.5%)
XRCC5	0 (0%)	11 (15.7%)	38 (28.7%)
EEF1A2	0 (0%)	11 (19.0%)	4 (4.1%)
SPIN3	0 (0%)	10 (12.0%)	19 (29.8%)
PARP1	0 (0%)	10 (6.4%)	9 (11.7%)
HNRNPA2B1	0 (0%)	9 (12.2%)	18 (27.5%)
TUBA4A	0 (0%)	9 (27.2%)	17 (27.2%)
HIST1H4A	0 (0%)	8 (42.7%)	10 (35.9%)
STK38L	0 (0%)	8 (13.4%)	2 (8.6%)
ACTB	0 (0%)	7 (23.7%)	19 (5.9%)
ACTG1	0 (0%)	7 (23.7%)	19 (5.9%)
PRDX4	0 (0%)	7 (24.4%)	3 (14.4%)
PSMC1	0 (0%)	7 (16.4%)	2 (8.2%)
HNRNPU	0 (0%)	6 (9.9%)	31 (21.8%)
TUBB6	0 (0%)	6 (11.7%)	28 (8.7%)
TRIM28	0 (0%)	6 (11.7%)	19 (23.2%)
NPM1	0 (0%)	6 (16.7%)	3 (18.4%)
TAB1	0 (0%)	5 (16.1%)	16 (22.6%)
SFPQ	0 (0%)	5 (5.2%)	15 (12.7%)

TUBA8	0 (0%)	5 (15.4%)	7 (21.6%)
PSMB5	0 (0%)	5 (16.7%)	5 (23.2%)
TUBB8	0 (0%)	4 (9.9%)	27 (9.0%)
CLNS1A	0 (0%)	4 (23.2%)	25 (34.6%)
XRCC6	0 (0%)	4 (8.5%)	10 (19.4%)
MCM7	0 (0%)	4 (4.7%)	8 (9.9%)
EEF1D	0 (0%)	4 (18.9%)	6 (26.0%)
EEF1B2	0 (0%)	4 (37.3%)	5 (16.0%)
PSMA2	0 (0%)	4 (20.9%)	4 (20.9%)
Putative tubulin beta chain-like protein	0 (0%)	3 (7.8%)	26 (6.7%)
Tubulin beta-8 chain-like protein	0 (0%)	3 (6.5%)	26 (5.6%)
ACTA2	0 (0%)	3 (7.2%)	19 (5.8%)
NONO	0 (0%)	3 (4.7%)	19 (25.9%)
ACTA1	0 (0%)	3 (7.2%)	19 (5.8%)
ACTC1	0 (0%)	3 (7.2%)	19 (5.8%)
ACTG2	0 (0%)	3 (7.2%)	19 (5.9%)
POTEE	0 (0%)	3 (3.6%)	16 (1.0%)
MAP3K7	0 (0%)	3 (4.1%)	9 (11.1%)
CALU	0 (0%)	3 (13.7%)	7 (34.6%)
RBM5	0 (0%)	3 (0.9%)	7 (1.2%)
RIOK1	0 (0%)	3 (8.3%)	6 (12.1%)
DDB1	0 (0%)	3 (3.0%)	6 (5.3%)
RBBP4	0 (0%)	3 (6.4%)	5 (12.7%)
HNRNPF	0 (0%)	3 (8.7%)	5 (8.2%)
H2AFJ	0 (0%)	3 (7.0%)	4 (20.9%)
HIST1H2AJ	0 (0%)	3 (7.0%)	4 (21.1%)
HIST2H2AC	0 (0%)	3 (7.0%)	4 (20.9%)
PSMA6	0 (0%)	3 (12.2%)	4 (17.9%)
HIST1H2AH	0 (0%)	3 (7.0%)	4 (21.1%)
HIST1H2AG	0 (0%)	3 (6.9%)	4 (20.8%)
HIST2H2AA3	0 (0%)	3 (6.9%)	4 (20.8%)
HIST1H2AD	0 (0%)	3 (6.9%)	4 (20.8%)
PSMB1	0 (0%)	3 (19.9%)	4 (29.0%)
HIST1H2AC	0 (0%)	3 (6.9%)	3 (12.3%)
H2AFX	0 (0%)	3 (6.3%)	3 (11.2%)
HIST1H2AB	0 (0%)	3 (6.9%)	3 (12.3%)
HIST3H2A	0 (0%)	3 (6.9%)	3 (12.3%)

HIST1H2AA	0 (0%)	3 (6.9%)	3 (12.2%)
H2AFZ	0 (0%)	3 (7.0%)	3 (12.5%)
PSMD2	0 (0%)	3 (4.0%)	3 (2.9%)
H2AFV	0 (0%)	3 (7.0%)	3 (12.5%)
PSMD7	0 (0%)	3 (7.4%)	3 (12.0%)
PSMC4	0 (0%)	3 (9.1%)	2 (6.9%)
PSMC3	0 (0%)	3 (4.1%)	2 (5.7%)
HSPD1	0 (0%)	3 (5.9%)	2 (6.3%)
ILF2	0 (0%)	3 (10.0%)	2 (5.1%)
ACAP2	0 (0%)	2 (4.6%)	16 (12.7%)
MATR3	0 (0%)	2 (3.8%)	12 (13.8%)
MCM3	0 (0%)	2 (2.6%)	11 (17.3%)
ACLY	0 (0%)	2 (3.9%)	8 (10.1%)
GANAB	0 (0%)	2 (2.9%)	7 (5.7%)
PPIL4	0 (0%)	2 (7.1%)	5 (12.4%)
PSMB3	0 (0%)	2 (16.6%)	5 (32.2%)
RUVBL2	0 (0%)	2 (7.1%)	5 (16.4%)
PSMD12	0 (0%)	2 (5.3%)	4 (12.9%)
PPM1B	0 (0%)	2 (6.5%)	4 (11.1%)
PSMD11	0 (0%)	2 (2.6%)	3 (8.8%)
ROCK1	0 (0%)	2 (2.7%)	3 (4.7%)
PPP2R2A	0 (0%)	2 (3.4%)	3 (13.9%)
SNRNP200	0 (0%)	2 (1.8%)	3 (1.9%)
HNRNPH2	0 (0%)	2 (3.8%)	2 (7.6%)
TUBA4B	0 (0%)	2 (8.7%)	2 (11.6%)
COPA	0 (0%)	2 (2.4%)	2 (1.8%)
JPH2	0 (0%)	6 (3.2%)	0 (0%)
PSMD3	0 (0%)	4 (10.3%)	0 (0%)
KIF11	0 (0%)	4 (5.5%)	0 (0%)
MYL6	0 (0%)	4 (16.6%)	0 (0%)
FLNC	0 (0%)	3 (1.7%)	0 (0%)
DPH7	0 (0%)	3 (1.5%)	0 (0%)
SMC1A	0 (0%)	3 (3.2%)	0 (0%)
TUBB1	0 (0%)	3 (6.4%)	0 (0%)
ACTBL2	0 (0%)	3 (7.2%)	0 (0%)
PSMC5	0 (0%)	3 (4.7%)	0 (0%)
SF3B2	0 (0%)	3 (3.6%)	0 (0%)

HIST2H2AB	0 (0%)	3 (6.9%)	0 (0%)
CSNK2A1	0 (0%)	3 (11.0%)	0 (0%)
GOLPH3	0 (0%)	2 (3.7%)	0 (0%)
ATP1A1	0 (0%)	2 (3.8%)	0 (0%)
EPRS	0 (0%)	2 (1.5%)	0 (0%)
CSNK2A2	0 (0%)	2 (10.6%)	0 (0%)
PSMA3	0 (0%)	2 (2.4%)	0 (0%)
PSMD6	0 (0%)	2 (6.2%)	0 (0%)
SMC3	0 (0%)	2 (2.1%)	0 (0%)
BMS1	0 (0%)	2 (1.9%)	0 (0%)
SHC1	0 (0%)	0 (0%)	112 (1.4%)
EXOSC10	0 (0%)	0 (0%)	55 (30.4%)
PABPC1	0 (0%)	0 (0%)	23 (20.3%)
EXOSC2	0 (0%)	0 (0%)	22 (29.4%)
PABPC3	0 (0%)	0 (0%)	19 (12.7%)
ZFC3H1	0 (0%)	0 (0%)	19 (12.4%)
C1QBP	0 (0%)	0 (0%)	19 (39.7%)
POTEKP	0 (0%)	0 (0%)	16 (2.9%)
POTEI	0 (0%)	0 (0%)	16 (1.0%)
EXOSC6	0 (0%)	0 (0%)	16 (26.8%)
POTEF	0 (0%)	0 (0%)	16 (1.0%)
HSP90AA5P	0 (0%)	0 (0%)	15 (4.5%)
CERS3	0 (0%)	0 (0%)	15 (5.0%)
EXOSC1	0 (0%)	0 (0%)	14 (29.2%)
NT5C2	0 (0%)	0 (0%)	14 (15.2%)
PFKFB3	0 (0%)	0 (0%)	14 (23.8%)
EXOSC3	0 (0%)	0 (0%)	13 (26.5%)
ZCCHC8	0 (0%)	0 (0%)	13 (18.1%)
RCN2	0 (0%)	0 (0%)	12 (37.9%)
CSHL1	0 (0%)	0 (0%)	11 (3.6%)
CSH1	0 (0%)	0 (0%)	11 (3.7%)
GH1	0 (0%)	0 (0%)	11 (3.7%)
GH2	0 (0%)	0 (0%)	11 (3.7%)
EXOSC9	0 (0%)	0 (0%)	11 (21.9%)
IPO8	0 (0%)	0 (0%)	11 (11.8%)
GTF2I	0 (0%)	0 (0%)	11 (8.2%)
MTHFD1L	0 (0%)	0 (0%)	10 (14.1%)

DDX3X	0 (0%)	0 (0%)	10 (14.0%)
SRRT	0 (0%)	0 (0%)	9 (13.0%)
USP15	0 (0%)	0 (0%)	8 (12.5%)
SNRPD3	0 (0%)	0 (0%)	8 (15.1%)
ALYREF	0 (0%)	0 (0%)	8 (24.5%)
IMPDH2	0 (0%)	0 (0%)	8 (12.3%)
MCM6	0 (0%)	0 (0%)	7 (8.3%)
JAK1	0 (0%)	0 (0%)	7 (4.9%)
USP7	0 (0%)	0 (0%)	7 (8.2%)
EXOSC5	0 (0%)	0 (0%)	7 (22.1%)
DHX15	0 (0%)	0 (0%)	7 (9.3%)
EXOSC7	0 (0%)	0 (0%)	7 (32.6%)
DDX3Y	0 (0%)	0 (0%)	7 (6.8%)
EXOSC8	0 (0%)	0 (0%)	7 (21.0%)
CCT6B	0 (0%)	0 (0%)	7 (5.7%)
PFKFB2	0 (0%)	0 (0%)	6 (6.5%)
PRMT1	0 (0%)	0 (0%)	6 (19.7%)
SNRPD2	0 (0%)	0 (0%)	6 (29.7%)
EXOSC4	0 (0%)	0 (0%)	5 (11.0%)
NAP1L1	0 (0%)	0 (0%)	5 (19.9%)
MCM5	0 (0%)	0 (0%)	5 (8.6%)
PSMA4	0 (0%)	0 (0%)	5 (33.7%)
PPP2R1A	0 (0%)	0 (0%)	5 (9.5%)
DDX5	0 (0%)	0 (0%)	5 (7.7%)
DHX38	0 (0%)	0 (0%)	5 (3.7%)
YWHAZ	0 (0%)	0 (0%)	5 (20.0%)
KPNB1	0 (0%)	0 (0%)	5 (6.3%)
PUF60	0 (0%)	0 (0%)	5 (13.8%)
YWHAG	0 (0%)	0 (0%)	5 (23.1%)
EIF3E	0 (0%)	0 (0%)	4 (10.1%)
NUDT21	0 (0%)	0 (0%)	4 (14.1%)
SRSF11	0 (0%)	0 (0%)	4 (8.1%)
PSMA8	0 (0%)	0 (0%)	4 (15.2%)
KPNA2	0 (0%)	0 (0%)	4 (14.6%)
LRPPRC	0 (0%)	0 (0%)	4 (5.0%)
CANX	0 (0%)	0 (0%)	4 (8.8%)
HNRNPA1	0 (0%)	0 (0%)	4 (11.0%)

NCL	0 (0%)	0 (0%)	4 (7.3%)
SF3B3	0 (0%)	0 (0%)	4 (4.4%)
HNRNPA1L2	0 (0%)	0 (0%)	4 (12.8%)
HNRNPA3	0 (0%)	0 (0%)	4 (12.4%)
TAB2	0 (0%)	0 (0%)	4 (10.0%)
RBM6	0 (0%)	0 (0%)	4 (0.6%)
EIF3B	0 (0%)	0 (0%)	4 (8.7%)
U2AF2	0 (0%)	0 (0%)	4 (9.7%)
SNRPN	0 (0%)	0 (0%)	4 (15.4%)
HSPH1	0 (0%)	0 (0%)	4 (5.4%)
PAICS	0 (0%)	0 (0%)	4 (13.2%)
YWHAB	0 (0%)	0 (0%)	4 (15.0%)
SNRPB	0 (0%)	0 (0%)	4 (15.4%)
PSMA5	0 (0%)	0 (0%)	3 (17.0%)
PSMB4	0 (0%)	0 (0%)	3 (25.0%)
HSP90AA2	0 (0%)	0 (0%)	3 (9.3%)
PPP2CA	0 (0%)	0 (0%)	3 (12.9%)
SPIN4	0 (0%)	0 (0%)	3 (2.8%)
HNRNPD	0 (0%)	0 (0%)	3 (7.6%)
HNRNPR	0 (0%)	0 (0%)	3 (6.2%)
COPRS	0 (0%)	0 (0%)	3 (15.8%)
SRP9	0 (0%)	0 (0%)	3 (25.6%)
TAF4	0 (0%)	0 (0%)	3 (4.1%)
PABPC1L	0 (0%)	0 (0%)	3 (5.4%)
RBBP7	0 (0%)	0 (0%)	3 (12.7%)
YWHAQ	0 (0%)	0 (0%)	3 (7.3%)
PPP2CB	0 (0%)	0 (0%)	3 (12.9%)
SMC2	0 (0%)	0 (0%)	3 (3.9%)
HNRNPDL	0 (0%)	0 (0%)	3 (6.4%)
PABPC4	0 (0%)	0 (0%)	3 (6.5%)
YWHAH	0 (0%)	0 (0%)	3 (7.3%)
2-Sep	0 (0%)	0 (0%)	3 (12.5%)
PSMD1	0 (0%)	0 (0%)	3 (5.2%)
PRDX2	0 (0%)	0 (0%)	3 (14.1%)
MPHOSPH6	0 (0%)	0 (0%)	3 (11.2%)
SLC25A6	0 (0%)	0 (0%)	3 (11.4%)
DDX17	0 (0%)	0 (0%)	3 (3.2%)

SNRPB2	0 (0%)	0 (0%)	3 (12.0%)
OAT	0 (0%)	0 (0%)	3 (17.3%)
SFN	0 (0%)	0 (0%)	3 (7.3%)
DHX9	0 (0%)	0 (0%)	3 (3.5%)
HNRNPL	0 (0%)	0 (0%)	3 (4.6%)
COPB2	0 (0%)	0 (0%)	2 (4.0%)
KLC1	0 (0%)	0 (0%)	2 (4.7%)
ACTN4	0 (0%)	0 (0%)	2 (2.5%)
CKAP5	0 (0%)	0 (0%)	2 (1.8%)
SF3B14	0 (0%)	0 (0%)	2 (20.8%)
EIF4A2	0 (0%)	0 (0%)	2 (5.7%)
RBM7	0 (0%)	0 (0%)	2 (6.8%)
TLL12	0 (0%)	0 (0%)	2 (4.0%)
EIF3A	0 (0%)	0 (0%)	2 (1.9%)
SNRNPE	0 (0%)	0 (0%)	2 (17.1%)
SRP14	0 (0%)	0 (0%)	2 (17.6%)
CPVL	0 (0%)	0 (0%)	2 (3.2%)
GAPDH	0 (0%)	0 (0%)	2 (12.8%)
TBC1D4	0 (0%)	0 (0%)	2 (2.0%)
SPAG9	0 (0%)	0 (0%)	2 (2.4%)
SNRPA	0 (0%)	0 (0%)	2 (2.8%)
PHGDH	0 (0%)	0 (0%)	2 (1.5%)
SLC25A5	0 (0%)	0 (0%)	2 (7.0%)
HNRNPM	0 (0%)	0 (0%)	2 (4.7%)
AP3B1	0 (0%)	0 (0%)	2 (2.7%)
KLC2	0 (0%)	0 (0%)	2 (6.9%)
PSMD14	0 (0%)	0 (0%)	2 (12.6%)
RPS7	0 (0%)	0 (0%)	2 (5.7%)
SRPK1	0 (0%)	0 (0%)	2 (6.1%)
PPIB	0 (0%)	0 (0%)	2 (11.6%)
SNRPG	0 (0%)	0 (0%)	2 (17.1%)
SNRPE	0 (0%)	0 (0%)	2 (13.0%)
EEF2	0 (0%)	0 (0%)	2 (4.4%)
EIF4A1	0 (0%)	0 (0%)	2 (5.7%)
CPSF6	0 (0%)	0 (0%)	2 (5.1%)
XPO1	0 (0%)	0 (0%)	2 (2.8%)
SHMT2	0 (0%)	0 (0%)	2 (6.9%)

ROCK2	0 (0%)	0 (0%)	2 (1.9%)
HSPA4L	0 (0%)	0 (0%)	2 (3.6%)
G3BP1	0 (0%)	0 (0%)	2 (5.6%)
HNRNPA0	0 (0%)	0 (0%)	2 (6.6%)
HBB	0 (0%)	0 (0%)	2 (19.7%)
ACTN1	0 (0%)	0 (0%)	2 (2.6%)
ARHGEF10	0 (0%)	0 (0%)	2 (2.2%)
PRSS3P2	0 (0%)	0 (0%)	2 (10.1%)
COPB1	0 (0%)	0 (0%)	2 (3.0%)
HDAC2	0 (0%)	0 (0%)	2 (6.4%)
KHSRP	0 (0%)	0 (0%)	2 (3.1%)
ARCN1	0 (0%)	0 (0%)	2 (4.5%)
DDX6	0 (0%)	0 (0%)	2 (5.4%)
SNRNP70	0 (0%)	0 (0%)	2 (5.0%)
STIP1	0 (0%)	0 (0%)	2 (5.5%)
GART	0 (0%)	0 (0%)	2 (2.4%)
ATP5C1	0 (0%)	0 (0%)	2 (8.4%)
EIF3L	0 (0%)	0 (0%)	2 (4.8%)