mRNA DECAY PATHWAYS USE TRANSLATION FIDELITY AND COMPETING DECAPPING COMPLEXES FOR SUBSTRATE SELECTION

A Dissertation Presented

By

ALPER CELIK

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

(MAY, 15 2017)

BIOMEDICAL SCIENCE

mRNA DECAY PATHWAYS USE TRANSLATION FIDELITY AND COMPETING DECAPPING COMPLEXES FOR SUBSTRATE SELECTION

A Dissertation Presented
By
ALPER CELIK

This work was undertaken in the Graduate School of Biomedical Sciences
Interdisciplinary Graduate Program

The signature of the Thesis Advisory signifies
validation of the Dissertation content

_____
Allan Jacobson, Ph.D., Thesis Advisor

The signatures of the Dissertation Committee signify
completion and approval as to the style and content of the Dissertation

_____
Melissa Moore, Ph.D., Member of Committee

_____
Craig Peterson, Ph.D., Member of Committee

_____
Andrei Korostelev, Ph.D., Member of Committee

_____
Jeff Coller, Ph.D., External Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets
the requirements of the Dissertation Committee

_____
Reid Gilmore, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that
the student has met all graduation requirements of the School

_____
Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

MAY, 15 2017

ABSTRACT

mRNA decay is an important step in gene regulation, environmental responsiveness, and mRNA quality control. One such quality control pathway, Nonsense-mediated mRNA Decay (NMD), targets transcripts whose translation terminates prematurely. However, the scope and the defining features of NMD-targeted transcripts remain elusive. To address these issues, we re-evaluated the genome-wide expression of annotated transcripts in yeast cells harboring deletions of the *UPF1*, *UPF2*, or *UPF3* genes. The vast majority of NMD-regulated transcripts are normal-looking protein-coding mRNAs. Our bioinformatics analyses reveal that this set of NMD-regulated transcripts generally have lower translational efficiency, lower average codon optimality scores, and higher ratios of out-of-frame translation.

General mRNA decay is predominantly mediated by decapping by the Dcp1-Dcp2 complex and 5' to 3' decay by Xrn1, but the exact mechanism of decapping regulation has remained largely unknown. Several *in vitro* and *in vivo* studies have revealed the importance of the C-terminal extension of Dcp2 and the identities of many decapping regulators that interact with the decapping complex. To better understand how decapping regulation is achieved by the C-terminal extension of Dcp2 we generated RNA-Seq libraries from a Dcp2 allele that lacks this portion of Dcp2 along with libraries from strains that contain single deletions of several decapping activators. Our transcriptome-wide results indicate that the C-terminal extension of Dcp2 is crucial for efficient regulation of decapping, and different decapping activators are responsible for targeting different sets

of mRNAs. Considering the limited pool of Dcp1-Dcp2 in the cell decapping activators might be in competition for decapping complex binding.

Collectively, our results yield valuable insights into the mechanism of substrate selection for mRNA quality control and decay in yeast.

Table of Contents

Title……i

Signature Page……ii

Acknowledgements……iii

Abstract……iv

List of Tables……ix

List of Figures……x

Chapter I – Introduction: mRNA Decay in *Saccharomyces cerevisiae*……1

Overview……2

Deadenylation……3

5' to 3' Decay by Xrn1……7

3' to 5' Decay by the Exosome……9

Decapping ……11

Nonsense-mediated mRNA Decay……20

Non-stop and No-go Decay……36

Conclusions……39

Chapter II - High Resolution Profiling of NMD Targets in Yeast Reveals Translational Fidelity as a Basis for Substrate Selection……41

Summary……42

Introduction……43

Results ……47

Upf1, Upf2, and Upf3 regulate a common set of transcripts in yeast cells……47

Structural and functional classes of NMD-regulated RNAs in transcriptome 1 ……56

Validation of newly identified NMD substrates……..……57

NMD substrates are principally degraded by decapping and 5' to 3' exonucleolytic decay ……..…....60

Intron-containing pre-mRNAs targeted by NMD are engaged in translation…….63

NMD substrates similar to normal mRNAs are poorly translated regardless of the NMD status of the cell ……..…….…63

Normal-looking NMD substrates have a higher rate of out-of-frame translation…..67

Normal looking NMD substrates have lower average codon optimality and a biased distribution pattern of non-optimal codons…...……71

Discussion ……..73

A comprehensive catalog of annotated yeast NMD substrates…….73

A significant fraction of yeast intron-containing mRNAs are targeted by cytoplasmic NMD……...74

Potential targeting mechanisms of the normal-looking NMD substrates……76

Chapter III – Regulation of mRNA decapping by competing factors ……79

Summary…….……80

Introduction…….……81

Results ……85

***dcp2-N245* cells have a different gene expression profile than *dcp2E153Q-N245* and *dcp2E198Q-N245* cells……..……85**

The Dcp2 C-terminal domain is important for decapping regulation……90

Lsm1 and Pat1 regulate the same set of transcripts, which partially overlaps with the set of Dhh1-regulated transcripts……...92

Transcripts regulated by Lsm1, Pat1, and Dhh1 are different from NMD substrates …….....95

Validation of bioinformatic analyses…….…..97

Different decapping complexes are responsible for regulating genes with different biological functions ……...97

Upf proteins are in competition for Dcp2 with Dhh1, but not with Lsm1 and Pat1 ……101

Discussion.……...104

Dcp2 C-terminal extension is crucial for efficient regulation of decapping…….104

Different decapping activators regulate different subsets of transcripts……105

Competition among different decapping complexes……..…106

Chapter IV – General Discussion ……..…109

A new class of NMD substrates illustrates the role of NMD in the cell……..110

A complex network of activators regulates decapping in the cell……..…112

Potential issues in bioinformatic analyses……...…114

Future directions ……115

Appendices ……118

Appendix A – Timing and specificity of polysomal association of Upf1……119

Appendix B – Effects of Upf2, eRF1, and eRF3 on Upf1's ATPase activities ……124

Appendix C – Association of Upf2 and Upf3 with ribosomal subunits ……129

Appendix D – Materials and Methods ……133

References……..…147

List of Tables

Table 2.1. Different categories of transcripts that were analyzed, detected, and commonly up- or down-regulated in all three *UPF* deletion strains……...53

Table 2.2. Overlap of NMD-regulated transcripts identified in this study and in several previous studies……..55

List of Figures

Chapter II

Figure 2.1. Upf1, Upf2, and Upf3 regulate the same set of transcripts in yeast……49

Figure 2.2. RNA-Seq libraries generated from WT, *upf1*Δ*, upf2*Δ*, upf3*Δ, dcp1*Δ*, dcp2*Δ*, and xrn1*Δ strains show good reproducibility and correlation between biological replicates ……51**

Figure 2.3. Validation of several different classes of NMD substrates by northern blotting……..58

Figure 2.4. NMD substrates are principally degraded by decapping and 5' to 3' exonucleolytic decay……61

Figure 2.5. NMD targeted intron-containing pre-mRNAs are engaged in translation……64

Figure 2.6. NMD substrates are less efficiently translated than non-substrates independent of the NMD machinery……...65

Figure 2.7. NMD substrates have lower translation fidelity and lower codon optimality……..68

Figure 2.8. Transcripts upregulated upon *UPF* deletion have lower ribosome densities and higher out-of-frame translation regardless of iTSSs……69

Figure 2.9. Different classes of NMD substrates……..……75

Chapter III

Figure 3.1. Cells harboring catalytically inactive Dcp2 alleles *dcp2E153Q-N245* and *dcp2E198Q-N245* show similar expression profiles which are significantly different from the indiscriminate *dcp2-N245* allele……..…….86

Figure 3.2 RNA-Seq libraries generated from WT, *lsm1*Δ, *pat1*Δ**, *dhh1*Δ**, *dcp2-N245, dcp2E153Q-N245,* and *dcp2E198Q-N245* strains show good correlation between biological replicates……..……88**

Figure 3.3. Deletion of C-terminal extension of Dcp2 results in decapping deregulation not decapping or decay inactivation……...91

Figure 3.4. Lsm1 and Pat1 regulate the same set of transcripts which partially overlaps with the set regulated by Dhh1…..……..93

Figure 3.5. Transcripts regulated by Lsm1, Pat1, and Dhh1 are different from NMD substrates…….…….…..96

Figure 3.6. Validation of bioinformatic analyses by northern blotting…….…..…..98

Figure 3.7. Different decapping complexes regulate genes with different biological functions……..…..……100

Figure 3.8. Upf proteins are in competition for Dcp2 with Dhh1, but not with Lsm1 and Pat1……...103

Figure 3.9 Different decapping complexes have different substrate specificities…….108

Appendix A

Figure A.1.Stochastic binding of Upf1 to ribosomes during translation……123

Appendix B

Figure B.1. Upf2 inhibits, and eRF1 and eRF3 promote Upf1's ATPase activities……128

Appendix C

Figure C.1. Association of Upf2 and Upf3 with ribosomal subunits……..…132

Chapter I

Introduction

mRNA DECAY IN *SACCHAROMYCES CEREVISIAE*

OVERVIEW

All RNAs in the cell are subject to turnover as a part of their life cycles and differential degradation of mRNAs can play important roles in the regulation of cellular responses to external stimuli. Degradation of mRNAs in yeast occurs by two main mechanisms, both of which are initiated by shortening of the poly(A) tail, a process called deadenylation. Deadenylation in yeast is carried out by two complexes, the Pan2-Pan3 complex and the Ccr4-Not complex (Wahle and Winkler, 2013). Following deadenylation, transcripts are either degraded by decapping by the Dcp1-Dcp2 enzyme (Dunckley and Parker, 1999), followed by 5' to 3' exonucleolytic decay by Xrn1 (Long and McNally, 2003), or are subjected to 3' to 5' decay by the large exosome complex (Liu et al., 2006). Several lines of evidence suggest that, at least during exponential growth, yeast mRNAs are predominantly targeted by the 5’ to 3’ pathway. First, strains lacking components of the decapping enzyme grow very slowly, or mutations causing these defects are lethal in some strains (Beelman et al., 1996), but 3' to 5' decay mutants have significantly milder phenotypes (Anderson and Parker, 1998). Second, strains lacking proper 5' to 3' decay machinery show significant changes in both mRNA steady-state levels and decay rates (Beelman et al., 1996; He et al., 2003). Third, genome-wide mapping of mRNA decay intermediates show that only a handful of mRNAs are targeted by endonucleolytic cleavage (Harigaya and Parker, 2012).

Reports from different groups have shown that decay rates among mRNAs can vary significantly (Herrick et al., 1990; Presnyak et al., 2015; Sun et al., 2013). These

differences in decay rates can be the results of differential targeting of decapping, deadenylation, or exonucleolytic decay. Therefore, understanding how these processes are regulated, and the interplay between them, is crucial for elucidation of mRNA decay mechanisms.

In addition to normal turnover of mRNA, there are specialized mRNA decay pathways that monitor mRNA quality and translational fidelity. These pathways are No-Go Decay (NGD) (Doma and Parker, 2006); Non-Stop Decay (NSD) (Frischmeyer et al., 2002; van Hoof et al., 2002); and Nonsense-Mediated mRNA Decay (NMD) (Leeds et al., 1991; Peltz et al., 1993; Pulak and Anderson, 1993). Targets of these pathways are subject to combinations of decapping, endonucleolytic cleavage, 5' to 3' decay, or 3' to 5' decay. While these processes are markedly different from regular mRNA turnover, they utilize the same decay machinery. Understanding how these pathways recognize their substrates can enhance our understanding of not only important aspects of mRNA quality control, but also normal mRNA turnover.

In this introduction, I summarize our current understanding of different decay mechanisms and the factors that are involved in these processes. The connections between different modes of decay and the complex interaction networks of their subsequent regulators illustrate how tightly mRNA decay is modulated and underscore the importance of mRNA turnover in gene expression.

DEADENYLATION

In eukaryotes, the processing and maturation of mRNAs includes the addition of a nontemplated 5’ 7-methylguanosine cap and a 3’ poly(A) tail. These modifications have profound effects on the entire life cycle of the mRNA. They are needed for efficient

translation of the mRNA and its protection from the decay machinery (Coller et al., 1998; Gray et al., 2000). Therefore, processes that target mRNAs for degradation also target the poly(A) tail.

At steady-state, the size of the poly(A) tail of cytoplasmic mRNAs varies from <50 nt in yeast to ~100 nt in mammalian cells (Subtelny et al., 2014). mRNAs with longer tails are thought to be translated more efficiently *in vitro* and *in vivo* and display higher ribosome densities in yeast (Beilharz and Preiss, 2007; Gallie, 1991; Munroe and Jacobson, 1990). This phenomenon has been attributed to the poly(A)-binding protein (Pab1 in yeast). The interaction network of Pab1 illustrates its central role in translation. Its interactions with eIF4G are thought to establish a closed loop mRNP (Amrani et al., 2008; Gallie and Tanguay, 1994; Jacobson, 1996; Le et al., 2000), ensuring efficient translation initiation and possibly ribosome recycling for additional rounds of translation. Pab1 interactions with the eukaryotic release factor 3 (eRF3; Sup35 in yeast) (Kervestin et al., 2012; Roque et al., 2015) also appear to ensure efficient translation termination (Amrani et al., 2004; Kervestin et al., 2012) (see below)

Consistent with the importance of the poly(A) tail, regulation of deadenylation - the removal of the poly(A) tail - is an important process in gene expression. There are two main deadenylases in the yeast cytoplasm: the Ccr4-Not complex and the Pan2-Pan3 complex. These two complexes are thought to work cooperatively and regulate deadenylation in yeast.

Ccr4-Not Complex

The Ccr4-Not complex is the main deadenylase in all organisms studied (Wahle and Winkler, 2013). There are at least five canonical subunits in this complex: Ccr4,

Caf1/Pop2, Not1, Not2, and Not3/5 (Albert et al., 2000; Chen et al., 2001; Temme et al., 2004). Ccr4 is one of the catalytic subunits of the complex (the other one being Caf1, see below). Its 3' to 5' exonuclease activity is Mg^{2+}-dependent and poly(A) specific (Tucker et al., 2002; Tucker et al., 2001). Ccr4 belongs to the exonuclease-endonuclease-phosphatase (EEP) class protein family (Dlakic, 2000). Structural studies (Basquin et al., 2012; Bhaskar et al., 2013) show a heart-shaped protein with two β sheets and a central α-helix in the nuclease core of the protein. In addition to the nuclease domain, Ccr4 also contains an LLR repeat which is responsible for interactions with Caf1 (Dupressoir et al., 2001; Malvar et al., 1992). Caf1 is the second catalytic subunit of the Ccr4-Not complex and is also universally conserved in eukaryotes (Schwede et al., 2008). Although *Saccharomyces cerevisiae* Caf1 lacks the catalytic residues that are thought to be important for activity, nuclease activity has been reported for the recombinant protein (Daugeron et al., 2001; Thore et al., 2003). Its sequence and the catalytic activity is consistent with the DEDD family of nucleases (Zuo and Deutscher, 2001). The name comes from four conserved amino acid residues that are necessary for coordinating Mg^{2+} cations for substrate hydrolysis (Andersen et al., 2009; Jonstrup et al., 2007). The large Not1 protein serves as the bridge for most of the interactions (Bai et al., 1999; Maillet and Collart, 2002; Maillet et al., 2000). Previously published reports present evidence suggesting that the Ccr4-Not complex exclusively exists as a single complex (Maillet et al., 2000; Russell et al., 2002; Temme et al., 2010) and both deadenylases are required for efficient poly(A) shortening *in vitro* (Maryati et al., 2015). The Ccr4-Not complex can degrade poly(A) tails processively, but its activity is inhibited by the poly(A)-binding protein (Cosson et al., 2002; Tucker et al., 2002). This illustrates the need for a

mechanism that would assist in removing the poly(A)-binding protein or shorten the poly(A) tail and thus reduce the affinity of Pab1. The Pan2-Pan3 complex fulfills this role in the cell.

Pan2-Pan3 complex

The poly(A)-specific nuclease (PAN) was the first deadenylating enzyme discovered (Sachs and Deardorff, 1992). This complex contains two subunits Pan2 and Pan3, Pan2 being the catalytic subunit (Boeck et al., 1996; Brown et al., 1996; Uchida et al., 2004). Pan2, like Caf1, belongs to the DEDD family of exonucleases and its activity depends on divalent cations (Brown et al., 1996; Zuo and Deutscher, 2001). In sharp contrast with Ccr4-Not complex, the activity of Pan2 is enhanced by the presence of poly(A)-binding protein and Pan3 (Lowell et al., 1992; Sachs and Deardorff, 1992; Schafer et al., 2014; Uchida et al., 2004). At least *in vitro* the Pan2 can deadenylate without poly(A)-binding protein, but not without Pan3 (Jonas et al., 2014; Schafer et al., 2014). The Pan2-Pan3 complex shows 1:2 stoichiometry and a distributive mechanism of action, with kinetic enhancement of Pan2 accomplished by interactions between Pan3 and Pab1 (Simon and Seraphin, 2007; Wolf et al., 2014; Yao et al., 2007; Zhang et al., 2013). The role of the Pan complex is thought to be the trimming of the poly(A) tail (Lowell et al., 1992; Yamashita et al., 2005) before the tail is completely degraded by the Ccr4-Not complex (Brown and Sachs, 1998; Petit et al., 2012; Temme et al., 2004; Tucker et al., 2001). In *pab1*Δ cells, mRNAs have elongated tails, and kinetic analyses show a lag phase for deadenylation in the absence of Pab1 *in vitro* (Caponigro and Parker, 1995; Sachs and Davis, 1989).

Targeting specific mRNAs for deadenylation

Some cellular proteins that bind to mRNAs in a sequence specific manner are implicated in the recruitment of deadenylation complexes. The six Puf proteins for example, bind to specific sequences and can regulate almost 10% of the transcriptome (Gerber et al., 2004; Olivas and Parker, 2000; Yosefzon et al., 2011). In yeast Puf1, Puf3, Puf4, and Puf5 have all been shown to promote poly(A) shortening on different subsets of mRNAs (Goldstrohm et al., 2007; Hook et al., 2007; Olivas and Parker, 2000; Tadauchi et al., 2001; Ulbricht and Olivas, 2008). At least one factor, Puf5, accomplishes this task by direct interactions with Caf1 (Goldstrohm et al., 2006; Goldstrohm et al., 2007). In addition, these proteins have been shown to inhibit translation to further contribute to deadenylation (Chritton and Wickens, 2011). Another example of sequence-dependent deadenylation involves the Vts1 protein, which recognizes a specific stem-loop sequence (Aviv et al., 2006) and recruits the Ccr4-Not complex (Rendl et al., 2008). Further, linking decapping and deadenylation, at least in mammalian cells Pat1 has been shown to interact with Not1 (Ozgur et al., 2010). Considering the importance of deadenylation in mRNA decay there are probably additional factors that either recognize mRNA sequences or regulate deadenylation rates based on environmental cues. Poly(A) shortening, however, is only one step in exonucleolytic decay of mRNAs in the cytoplasm. 5' to 3' and 3' to 5' decay are responsible for complete degradation of mRNAs, usually after deadenylation (see below).

5’ TO 3’ DECAY BY XRN1

There are two main 5’ to 3’ exoribonucleases in yeast, nuclear Xrn2/Rat1 and cytoplasmic Xrn1. Xrn2/Rat1 is involved in ribosomal RNA maturation, mRNA transcription termination, and telomere maintenance(Wang and Pestov, 2011). Xrn1 is is

responsible for the 5’ to 3’ decay of mRNAs after decapping or endonucleolytic cleavage (Orban and Izaurralde, 2005) (see below).

The well conserved Xrn1 protein contains several domains. N-terminal CR1 and CR2 sites are shared with Xrn2, with CR1 containing the catalytic activity and the less conserved CR2 enhancing the enzymatic activity of CR1 (Xiang et al., 2009). CR1 contains seven well conserved acidic amino acids that are involved in coordinating a Mg^{2+} ion that is essential for activity. Xrn1 has additional domains that Xrn2 lacks. Its middle PAZ/Tudor, KOW, winged helix, and SH3-like domains are conserved from humans to yeast (Chang et al., 2011; Jinek et al., 2011).

Xrn1 degrades decapped mRNAs harboring a 5’ phosphate. The 5’ phosphate is recognized by a basic pocket in Xrn1 and the 5’ cap generates steric hindrance with this pocket, thus inhibiting its activity. This requirement for a 5' phosphate explains Xrn1's dependence on decapping for mRNA decay (see below). The tower domain degrades its substrates processively (Page et al., 1998). The PAZ/Tudor domain plays a structural role stabilizing the catalytic domain. The C-terminal domain structure of Xrn1 has not been elucidated, but deletion and mutagenesis analyses suggest that it is required for enzymatic activity (Page et al., 1998). The RNA channel of Xrn1 is large enough to accommodate single-stranded RNA. Xrn1 stalls when in contact with large stem-loops, although the enzyme can perform some unwinding of smaller structures (Bashkirov et al. 1997) (Bashkirov et al., 1997).

Interestingly, Xrn1 and the exosome (see below) do not co-localize in the cell despite reports showing at least partial redundancy in their activities (Bashkirov et al., 1997). Consistent with its role in mRNA decay, Xrn1 does co-localize with many factors

that are important for mRNA decapping and decay, e.g., Dcp1-Dcp2 or the Lsm1-7 complex (Braun et al., 2012; Jinek et al., 2011), and can interact with Dcp1 directly (Braun et al., 2012). There is strong evidence from the literature that at least for some mRNAs Xrn1 can also promote decay co-translationally (Hu et al., 2009; Pelechano et al., 2015).

While Xrn1 is responsible for degradation of mRNAs, these transcripts are not its only substrates. Xrn1-dependent untranslated transcripts (XUTs) (Wery et al., 2016), cryptic untranslated transcripts (CUTs) (Wery et al., 2016; Xu et al., 2009), and stable untranslated transcripts (SUTs) (Xu et al., 2009) have all been shown to be targeted by Xrn1. Some of these transcripts have been proposed to be targeted by the NMD pathway (Malabat et al., 2015; Wery et al., 2016), but the limited evidence for their translation makes this mode of decay less likely.

3’ TO 5’ DECAY BY THE EXOSOME

While Xrn1 is responsible for 5’ to 3’ mRNA decay, the large exosome complex is responsible 3’ to 5’ decay. The eukaryotic exosome is composed of ten or eleven subunits (Liu et al., 2006), which can be categorized in two different groups depending on their structural and functional attributes. The first group is composed of nine proteins comprising the exosome core, and includes small proteins largely limited to RNA binding domains (Liu et al., 2016; Liu et al., 2006). The second group is composed of Dis3 and Rrp6, which associate with the core proteins and provide catalytic activity to the exosome (Dziembowski et al., 2007; Schneider et al., 2009).

The architecture of the exosome core is well established and is principally derived from structures of the human and yeast complexes (Liu et al., 2016; Liu et al., 2006). The nine subunits of the core are arranged in a two-layer ring. The bottom layer is called the

“hexamer” which is composed of Rrp41-46 and Mtr3 (Liu et al., 2016; Liu et al., 2006) and the top layer is called the cap which contains Rrp4, Rrp40, and Csl4 (Ramos et al., 2006). The cap subunits do not contact one another. Instead, they are associated with the hexamer. The cap proteins form a three-fold symmetry on top of a trimer of dimers (Liu et al., 2016; Liu et al., 2006).

The core complex does not have any enzymatic activity. Dis3 and Rrp6 proteins that associate with the core provide the decay activity to the exosome. Dis3 is an endoribonuclease and the 3’ to 5’ exoribonuclease (Lebreton et al., 2008; Schaeffer et al., 2009; Schneider et al., 2009). Dis3's substrates are RNAs with single stranded 3’ ends but Dis3 can unwind intramolecular secondary structures (Dziembowski et al., 2007; Lorentzen et al., 2008). Rrp6 on the other hand belong the to the DEDD family nucleases (Briggs et al., 1998) but unlike Dis3, Rrp6 is a distributive RNAase (Januszyk et al., 2011; Midtgaard et al., 2006). This protein is a part of the 11-subunit nuclear exosome complex and does not contribute to cytoplasmic decay.

Even though the core complex does not have any catalytic activity, it is necessary for exosome function (Allmang et al., 1999; Dziembowski et al., 2007; Mitchell et al., 1997). Binding to the core complex markedly reduces the decay rate of RNA by impairing RNA binding activity of the exonucleases (Wasmuth and Lima, 2012), this decrease in activity is necessary to prevent indiscriminate decay of mRNAs.

In the context of mRNA decay in the cytoplasm, the activity of the exosome is regulated by the superkiller proteins (Ski) (Anderson and Parker, 1998). In yeast this complex is composed of DEVH family RNA helicase Ski2 and RNA binding proteins Ski3, Ski8 and Ski7. The Ski complex is evolutionarily conserved and has been shown to take

part in 3' to 5' decay in both yeast and metazoans (Araki et al., 2001; Synowsky and Heck, 2008; van Hoof et al., 2000; Wang et al., 2005). Ski2, Ski3 and Ski8 mediate protein-protein interactions (D'Andrea and Regan, 2003; Stirnimann et al., 2010) and Ski7 is responsible for linking the exosome to the Ski complex to modulate RNA decay. While the majority of RNA decay is done by Xrn1 in yeast cytoplasm, the substrates of NGD and NSD are degraded by the exosome (see below). Yet, there seems to be significant redundancy between the two decay mechanisms (Parker, 2012). The extent of this redundancy is not well understood and how each decay mechanism chooses its substrates is an area of active research.

DECAPPING

During transcription, eukaryotic mRNAs acquire a 5’ 7-methyl-guanosine cap which significantly promotes translation (Mitchell et al., 2010) by its interaction with eIF4E. In addition, the mRNA cap protects the transcript from non-specific attack by 5’ to 3’ exonucleases (Stevens, 1978). The removal of the 5' mRNA cap by the decapping complex is, therefore, a critical step in decay regulation and this process is closely linked with translation (see below). Unlike some other regulatory steps, e.g., translational repression, decapping is usually an irreversible process that inhibits translation initiation and commits transcripts to decay (Arribas-Layton et al., 2013). The occurrence of cytoplasmic recapping (Schoenberg and Maquat, 2009) indicates that mRNA decapping can be reversible in some instances.

Dcp1-Dcp2 decapping enzyme

The best characterized decapping protein is Dcp2. This protein was originally identified in yeast (Dunckley and Parker, 1999; He and Jacobson, 1995; Peltz et al.,

1994), but subsequent studies revealed homologs in all eukaryotic model organisms (Cohen et al., 2005; Iwasaki et al., 2007; Steiger et al., 2003; Wang et al., 2002). Purified Dcp2 has decapping activity *in vitro* for long RNA substrates, but cannot effectively hydrolyze unmethylated cap, free GTP, or short capped oligonucleotides (LaGrandeur and Parker, 1998). Dcp2 contains a Nudix domain (Dunckley and Parker, 1999; Wang et al., 2002), a structure found in enzymes that hydrolyze nucleoside diphosphates that are linked to other biological moieties (Bessman et al., 1996) through conserved glutamate residues that coordinate Mn^{2+} or Mg^{2+} ions (Piccirillo et al., 2003; Steiger et al., 2003). Sequence analysis shows that the N-terminal region of Dcp2 is highly conserved among eukaryotes while its C-terminal extension is divergent both in length and sequence (ranging from ~175 amino acids in humans to ~700 in yeast) (Wang et al., 2002). In addition to the Nudix domain, the conserved N-terminal region of Dcp2 contains a Box A motif which is responsible for Dcp1 interaction, and a Box B motif that regulates RNA binding (Lykke-Andersen, 2002; Piccirillo et al., 2003; She et al., 2006; van Dijk et al., 2002). Structural studies revealed that substrates bind to the Dcp1-Dcp2 complex via a channel that places the cap structure within the Nudix domain while the rest of the RNA wraps along the Box B domain. Dcp2 requires its substrates to be at least 12nt long for efficient decapping (Mugridge et al., 2016; She et al., 2008).

An essential *in vivo* cofactor of Dcp2 is Dcp1. This protein was originally described as the catalytic subunit of the complex and yeast strains lacking Dcp1 are highly deficient in decapping (LaGrandeur and Parker, 1998). Dcp1 directly interacts with Dcp2 and decapping activity is enhanced several orders of magnitude *in vitro* by that interaction (Steiger et al., 2003). Unlike Dcp2, Dcp1 is a small protein that contains an EVH1 domain,

which is implicated in modulating protein-protein interactions (Ball et al., 2002; She et al., 2004). In addition to this domain, Dcp1 contains a trimerization domain in metazoans (Tritschler et al., 2009b). Mutations in this domain have been shown to impair decapping activity *in vivo* (Tritschler et al., 2009b). In its interaction with Dcp2, at least in the crystal structure from *Schizosaccharomyces pombe*, Dcp1 does not utilize these conserved domains, but instead uses N-terminal sequences that are only found in yeast species (She et al., 2008). Dcp1 in yeast also lacks C-terminal extensions that are present in other eukaryotes, mainly the HLM domains. However, to make up for this, the yeast Dcp2 protein contains additional HLM sequences, some of which bind to the decapping activators Edc3 and Scd6 (Fromm et al., 2012; Harigaya et al., 2010) (see below). The presence of multiple binding domains is thought to increase affinity by avidity effects (Jonas and Izaurralde, 2013). Mutational analyses of similar sequences from *Saccharomyces cerevisiae* and two-hybrid analyses are consistent with these hypotheses (He and Jacobson, 2015a; He et al., 2014).

Structural studies revealed different conformations of the Dcp1-Dcp2 complex (She et al., 2008). The open conformation resembles that of apo-Dcp2 (She et al., 2006) and assumes a dumbbell-like shape. The closed conformation, on the other hand, places the Dcp2 N-terminus in close proximity of the Nudix domain to form a compact structure. Biochemical data indicates that this closed conformation is a close approximation of the active Dcp1-Dcp2 complex (She et al., 2008). Other structural data suggest that the enhancement of Dcp2 activity in the presence of Dcp1 is due to stabilization of the active conformation (Valkov et al., 2016). Mutations in residues that block the formation of the

closed complex result in loss of Dcp1 stimulation of Dcp2 and loss of decapping. (She et al., 2008).

Decapping and decay regulation is multifaceted

The Dcp1-Dcp2 decapping complex is stimulated by additional proteins called decapping activators. These activators are characterized by a modular architecture consisting of globular domains connected by flexible unstructured regions. For many decapping regulators these globular domains are well conserved, but the unstructured regions vary considerably both in length and in sequence among eukaryotes, yet these motifs are implicated in regulating complex assembly by binding to other folded domains (Jonas and Izaurralde, 2013). The first proteins that were identified to regulate decapping were the Edc1 and Edc2 proteins (Dunckley et al., 2001). Structural studies have shown the interaction of Edc1 and Edc2 with Dcp1 is responsible for activation of decapping *in vitro* by interacting with Dcp1's EVH1 domain (Borja et al., 2011; Wurm et al., 2016). Interestingly, deletion of Edc1 or Edc2 does not impair *in vivo* decay rates of reporter mRNAs (Dunckley et al., 2001). Yet, they were later shown to bind RNAs directly to stimulate decapping (Schwartz et al., 2003). While there are no homologs of Edc1 and Edc2 in metazoans, Edc3 is well conserved among eukaryotes (Fromm et al., 2012). This protein directly interacts with Dcp2 in yeast and stimulates the decapping reaction *in vitro* (Harigaya et al., 2010; Nissan et al., 2010). However, deletion of *EDC3* in yeast results in upregulation of only 2 transcripts (Dong et al., 2010), illustrating how tightly decapping can be regulated.

Another enhancer of decapping is Pat1. This protein was initially identified in a mutational screen for stimulators of decapping (Nissan et al., 2010; Tharun et al., 2000)

and was later found to interact directly with Dcp2 (He and Jacobson, 2015a). Pat1 and its orthologs in other organisms have been shown to associate with other decapping regulators (Haas et al., 2010; Ozgur et al., 2010; Tharun et al., 2000). One of these interacting partners is the Lsm1-7 complex. This interaction is mediated by Lsm2 and Lsm3 within the complex (Nissan et al., 2010; Sharif and Conti, 2013). In *Saccharomyces cerevisiae,* the Lsm1-7 heptameric ring associates with mRNAs that have undergone deadenylation (Parker and Song, 2004). Deletion of the *LSM1* or *PAT1* genes results in accumulation of deadenylated but capped mRNAs (Tharun et al., 2000). Recombinant Lsm1-7-Pat1 complex has an intrinsic preference for oligoadenylated mRNAs over polyadenylated and this affinity depends on the presence of Pat1 (Chowdhury et al., 2014; Chowdhury et al., 2012). Yeast strains lacking Pat1 are unable to effectively inhibit translation in response to glucose starvation and overexpression of Pat1 leads to general translation repression (Coller and Parker, 2005). In addition, Pat1 shows RNA-dependent associations with eIF4E and eIF4G, as well as Pab1 (Tharun and Parker, 2001). These observations suggest that an Lsm1-7-Pat1 complex can prime an mRNP for decapping through translational inhibition by releasing initiation factors either by direct interactions of by further stimulating deadenylation and disrupting the closed-loop complex (Kahvejian et al., 2005).

The DEAD box helicase Dhh1 is also implicated in the regulation of decapping. Again, first identified in budding yeast as a decapping activator, deletion of *DHH1* results in the accumulation of capped mRNA decay intermediates (Coller et al., 2001; Fischer and Weis, 2002). Dhh1 can inhibit translation when tethered to specific mRNAs (Sweet et al., 2012) and is thought to decrease translation efficiency by slowing the movement of

ribosomes on the mRNAs or targeting mRNAs that are already translating ineffectively due to lower codon optimality (Radhakrishnan et al., 2016). Interestingly, Dhh1 can accomplish this even when its ATPase activity is abolished by mutations (Carroll et al., 2011). Interaction studies have shown that the C-terminal domain of Dhh1 has binding sites for the Edc3 and Pat1 proteins which, therefore, can compete for Dhh1 binding (Sharif et al., 2013). These binding motifs on Dhh1 reside near the ATP-dependent RNA-binding site. Consequently, the binding of Pat1 or Edc3 to this patch of positively charged residues interferes with Dhh1 binding to RNA (Sharif et al., 2013). However, the biological implications of such regulation or the importance of the ATPase activity of Dhh1 remain unknown.

Other proteins have also been implicated in decapping activation through translational inhibition. Scd6, for example, has been shown to interact with eIF4G directly and to inhibit the joining of the 43S complex (Nissan et al., 2010; Rajyaguru et al., 2012). Scd6 can also interact with the C-terminal extension of Dcp2 and modestly increase its activity *in vitro,* linking translation inhibition to decapping (Fromm et al., 2012). However, it is not yet clear how these proteins can select their substrates for decapping, whether they associate with the decapping complex simultaneously with other decapping activators or how regulation of different transcripts is achieved. Alternatively, decapping activation could occur after a series of ordered exchanges of factors that ultimately end with the recruitment of the decapping complex.

In the cytoplasm, the mRNA cap is usually bound by the eIF4F complex. Within this structure, eIF4E directly binds to the 5' cap. In turn, this binding is stimulated by eIF4G and can decrease decapping rates *in vitro* (Ramirez et al., 2002; Schwartz and Parker,

2000; Vilela et al., 2000). The interaction between eIF4G and poly(A)-binding protein, leading to formation of the closed-loop complex, can protect the mRNA from decapping and decay (Roy and Jacobson, 2013). Mutations in eIF4E or eIF4G which slow down translation initiation also lead to increases in decay rates *in vivo* (Schwartz and Parker, 1999). It has been demonstrated that, at least in yeast, decapping is accelerated when the poly(A) tail has been shortened to about 12 nucleotides (Decker and Parker, 1993). This is approximately the minimal length required for Pab1 binding (Sachs et al., 1987). Further, *pab1* mutations result in uncoupling of deadenylation from decapping (Caponigro and Parker, 1995; Morrissey et al., 1999). These results indicate that, in the absence of Pab1, the shortening of the poly(A) tail is no longer a requirement for decapping and underscore the importance of the interactions between the poly(A) tail and the 5' cap. Based on these observations, decapping activation likely involves remodeling of the mRNP to allow access to the cap structure for hydrolysis. Therefore, it is not surprising that some decapping enhancers are also implicated in translation repression.

For general mRNA turnover, decapping occurs after mRNAs have been deadenylated. This order of events ensures that functional mRNAs are not targeted for decay prematurely. However, how decapping and deadenylation are coordinated is poorly understood. Pat1 may play important roles in this transition due to its interactions with the Ccr4-Not complex (see above) (Haas et al., 2010; Ozgur et al., 2010) and Dcp2 along with other decapping activators (Bonnerot et al., 2000; Bouveret et al., 2000; Braun et al., 2012; Fromont-Racine et al., 2000; Haas et al., 2010; Ozgur et al., 2010; Pilkington and Parker, 2008; Tharun et al., 2000; Tharun and Parker, 2001). Further, Pat1 can link decapping to 5' to 3' decay by Xrn1 (see above) (Bouveret et al., 2000; Nissan et al.,

2010), but the specifics of these interactions and the order of events are not well understood. Nonetheless, the interaction between Xrn1 and the decapping complex, either with the help of Pat1 or directly (Braun et al., 2012), ensures that decay can immediately follow decapping. Interestingly, overexpression of Xrn1 inhibits decapping in a dominant negative manner in flies (Braun et al., 2012), suggesting that excess Xrn1 can inhibit the ordered assembly of decapping complexes, possibly by sequestration of necessary components.

Substrate selection for decapping

Decapping can be activated in a transcript specific manner. In NMD (see below), direct interactions of Upf1 with the C-terminal extension of Dcp2 can recruit the decapping complex to its substrates (He and Jacobson, 2001, 2015a; Lykke-Andersen, 2002). Similarly, the ribosomal protein Rps28 and the export factor Yra1 appear to recruit Edc3 to their own mRNA/pre-mRNA to autoregulate their expression (Badis et al., 2004; Dong et al., 2010; Dong et al., 2007; He and Jacobson, 2015a; He et al., 2014; Kolesnikova et al., 2013). Further, the binding of Puf3 to the Cox17 mRNA enhances decapping of this mRNA following deadenylation (Olivas and Parker, 2000; Ulbricht and Olivas, 2008).

The recently discovered inhibitory element in the Dcp2 protein provides another regulatory mechanism for decapping (He and Jacobson, 2015a). Removal of this sequence results in a constitutively active protein that shows an overall loss of substrate specificity as evidenced by the bypassing of the requirement for Edc3 in decay of *YRA1* pre-mRNA (He and Jacobson, 2015a). This regulatory element is hypothesized to prevent non-specific decapping of mRNAs, thus ensuring a tightly regulated mode for mRNA decay and gene expression.

The potential structural switches of the Dcp1-Dcp2 complex are even more complicated when these interacting partners are considered. The C-terminal extension of budding yeast Dcp2 contains several overlapping binding sites for decapping activators (He and Jacobson, 2015a). Structural evidence indicates that these interactions can stabilize the active conformation of the Dcp1-Dcp2 complex and activate decapping (Charenton et al., 2016; Valkov et al., 2016). The complex and modular nature of the Dcp1-Dcp2 interacting factors suggests that there can be different decapping complexes and that these complexes can have different substrate specificities. By utilizing different combinations of regulators, Dcp2 can achieve a large repertoire of substrates with a limited number of regulators. Lack of conservation of the C-terminal regions of Dcp2 can also hint at the divergent needs for mRNA decay regulation from single-celled organisms to mammals. This model of several different mutually exclusive decapping complexes is also consistent with the notion that decay rates in the cytoplasm show a large range of values (Presnyak et al., 2015; Sun et al., 2013). The mode of mRNA decapping regulation for each mRNA likely depends on numerous factors that interact with the decapping proteins to help recruit them to specific mRNAs. Quality control pathways such as NMD, NGD, or NSD can also promote decapping that is independent of deadenylation and the recruitment of decay factors in these pathways shows many similarities and differences between general decapping (see below).

NONSENSE-MEDIATED mRNA DECAY (NMD)

NMD is one of several translational quality control pathways that monitor mRNA translational activities (Celik et al., 2015; Lykke-Andersen and Bennett, 2014; Shoemaker and Green, 2012). As its name suggests, NMD targets mRNAs that contain a premature

termination (nonsense) codon within an open reading frame, although recent studies revealed additional NMD targets that do not contain a PTC (see below).. Initially discovered in yeast and nematodes (Leeds et al., 1991; Peltz et al., 1993; Pulak and Anderson, 1993), NMD is a well conserved pathway present in all eukaryotes examined (Kertesz et al., 2006; Schweingruber et al., 2013). This pathway was initially thought to protect the cell from the deleterious effects of truncated polypeptides (He et al., 1993; Pulak and Anderson, 1993) arising because mutations, transcriptional errors, splicing errors, or alternative splicing events gave rise to premature termination codons (PTCs). However, genome-wide studies have revealed that NMD substrates are much more numerous and diverse than initially thought (He et al., 2003). Therefore, in addition to being a quality control pathway, NMD also serves as a key post-transcriptional regulatory mechanism. NMD can regulate approximately 5-20% of a cell's transcriptome, affecting many aspects of cell biology, from proliferation to neuronal activity (He et al., 2003; Lelivelt and Culbertson, 1999; Mendell et al., 2004; Metzstein and Krasnow, 2006; Ramani et al., 2009; Rayson et al., 2012; Rehwinkel et al., 2005; Tani et al., 2012; Weischenfeldt et al., 2012; Wittkopp et al., 2009).

Substrates of the NMD pathway

NMD targets both PTC-containing and normal-looking mRNAs in the cell. These targets can be classified into several categories. One typical category, as noted above, includes transcripts on which an elongating ribosome encounters a PTC. The PTCs can be generated by mutations (He et al., 2003), leaky scanning at the initiation codon (Welch and Jacobson, 1999), transcriptional errors, internal transcription start sites (Malabat et al., 2015), unspliced mRNAs, or alternative splicing events (Gehring et al., 2005; He et

al., 1993; Jaillon et al., 2008; Kawashima et al., 2014; Lareau et al., 2007; Ma et al., 2008; Ni et al., 2007; Weischenfeldt et al., 2012). A second category includes transcripts that do not code for proteins and are presumably being translated erroneously, e.g., long noncoding RNAs (Kurihara et al., 2009; Lykke-Andersen et al., 2014; Tani et al., 2013), transcripts of pseudogenes (He et al., 2003; McGlincy and Smith, 2008), or small RNAs derived from intergenic regions (Smith et al., 2014; Thompson and Parker, 2007). A third category encompasses transcripts that undergo programmed frameshifting, such as the mRNAs of transposable elements (He et al., 2003). In addition to these substrates, mRNAs with upstream open reading frames (uORFs) (Colombo et al., 2017; He et al., 2003; Heyer and Moore, 2016; Ingolia et al., 2009; Nyiko et al., 2009; Rehwinkel et al., 2005) or unusually long 3' UTRs (Colombo et al., 2017; Kebaara and Atkin, 2009; Kertesz et al., 2006; Muhlrad and Parker, 1999a; Singh et al., 2008; Zaborske et al., 2013) can also be targeted by NMD, although not all mRNAs with such attributes are subjected to accelerated decay. uORF-containing mRNAs can regulate their levels in a context dependent manner. For example, *CPA1* mRNA in yeast can be targeted by NMD upon certain physiological cues. The peptide encoded by the uORF of this mRNA can control ribosome stalling depending on the levels of available arginine in the cytoplasm and regulate *CPA1* mRNA degradation by NMD (Gaba et al., 2005). On a similar note, not all mRNAs with long 3' UTRs are targeted by NMD. Some of these mRNAs can contain protective sequence elements that block NMD activity to ensure full expression of these genes (Ge et al., 2016; Toma et al., 2015).

The largest category of NMD substrates is comprised of mRNAs that do not contain any of the aforementioned features, i.e., they appear to be "normal" mRNAs. How

and why these mRNAs are targeted by NMD has been an active area of research, but reports from different groups have led to conflicting results. The question of how NMD selects its substrates has, therefore, remained unanswered (see below).

NMD machinery

NMD is a translation dependent process that is thought to be initiated by a nonsense codon in the ribosomal A site during translation elongation (Amrani et al., 2004). NMD identifies distinguishing features of premature termination and targets PTC-containing mRNAs for rapid decay. This process requires the activity of the Upf proteins -the principal regulators of NMD- and other factors that monitor or modulate Upf activity or initiate degradation of the targeted mRNA.

Upf proteins

In all eukaryotes examined, the activity of the Upf1, Upf2, and Upf3 proteins have been shown to be essential for NMD (Cui et al., 1995; Gatfield et al., 2003; He et al., 1997; He and Jacobson, 1995; Leeds et al., 1991; Lykke-Andersen et al., 2000; Nicholson et al., 2012; Nicholson et al., 2010; Perlick et al., 1996; Salas-Marco and Bedwell, 2004). These proteins are, therefore, well conserved. Upf1 and Upf2 are cytoplasmic proteins and Upf3 is thought to shuttle between the nucleus and the cytoplasm (Lykke-Andersen et al., 2000; Serin et al., 2001) though the importance of this shuttling to NMD is unclear. Because NMD seems to recognizes termination events that appear premature, the recognition of a nonsense codon by translation of an mRNA is crucial. Consistently, the core Upf proteins interact with each other, ribosomes, and many translation and mRNA decay factors (see below).

Upf1, the only Upf protein with a defined enzymatic activity, is a superfamily I helicase. Upf1 contains two conserved domains: an N-terminal cysteine-histidine rich domain and a C-terminal ATPase-Helicase domain (Altamura et al., 1992) (Chakrabarti et al., 2011; Clerici et al., 2009; Koonin, 1992). Purified Upf1 has RNA-dependent ATPase and ATP-dependent helicase activities (Bhattacharya et al., 2000; Czaplinski et al., 1995) which are enhanced by addition of Upf2 and Upf3 (Chamieh et al., 2008; Fiorini et al., 2013). The enzymatic activities of Upf1 are essential for NMD (Chamieh et al., 2008; Fiorini et al., 2012; Fiorini et al., 2013; Weng et al., 1996), i.e., mutations that abolish these activities also result in protection of NMD substrates from rapid decay. Further, Upf proteins that lack ATP hydrolysis activities cannot associate with ribosomes (Min et al., 2013). While our understanding of the role of these properties is limited there is some indirect evidence hinting at a role in termination complex disassembly from the mRNA (Durand et al., 2016; Franks et al., 2010; Ghosh et al., 2010; Serdar et al., 2016). Considering that Upf1 is a helicase that can exert substantial mechanical force on RNA in a processsive manner. Yeast two-hybrid and other structural studies have shown that Upf1's CH domain interacts with the C-terminal region of Upf2 (He et al., 1997), the ribosomal protein Rps26 (Min et al., 2013), and the decapping enzyme Dcp2 (He and Jacobson, 1995, 2001)(see above), and self-associates (He et al., 2013). Since these interactions cannot occur simultaneously due to overlapping binding sites they hint at a sequential array of interactions that are necessary for targeting NMD substrates for decay. Metazoan Upf1 proteins have additional conserved domains that are rich in serine and glutamine (Applequist et al., 1997; Perlick et al., 1996). These residues have been shown to be important in phosphorylation/de-phosphorylation cycles, as well as

interactions with other metazoan specific regulators that are necessary for NMD (Ohnishi et al., 2003; Yamashita et al., 2001). While there are a handful of reports showing Upf1 and Upf2 phosphorylation in yeast the biological importance of these events in yeast has not been thoroughly investigated (de Pinto et al., 2004; Lasalde et al., 2014; Wang et al., 2006).

Upf2 acts as a bridge between Upf1 and Upf3. The N-terminal region of Upf2 contains three eIF4G-like (MIF4G) domains (Aravind and Koonin, 2000; Clerici et al., 2014; Kadlec et al., 2004; Ponting, 2000). The most C-terminal of these interacts with Upf3's middle RRM domain (He et al., 1997; Kadlec et al., 2004) and, in metazoans, with the C-terminal domain of Smg1 (see below). Binding of Upf2 to Upf1 switches Upf1 from its RNA binding to an RNA unwinding conformation, as evidenced by an increase in helicase and ATPase activities and reduced RNA affinity (Chakrabarti et al., 2011; Chamieh et al., 2008).

Upf3 is a smaller, highly alkaline protein. There is only one *UPF3* gene in the *Saccharomyces cerevisiae* genome, but, in humans, there are two copies, respectively encoding Upf3a and Upf3b (Lykke-Andersen et al., 2014; Mendell et al., 2004). The two isoforms have different expression patterns and while their activities are somewhat redundant their respective abilities to trigger NMD are markedly different (Kunz et al., 2006). The central RRM domain of both Upf3s interact with Upf2 (Lejeune et al., 2003; Serin et al., 2001) and its C-terminal domain interacts with the exon junction complex (EJC) in higher eukaryotes (Buchwald et al., 2010; Gehring et al., 2003). Recent experiments suggest that Upf3a also is a testes-specific NMD inhibitor (Shum et al., 2016).

Additional NMD regulators

Smg Proteins

In higher eukaryotes, there are additional factors that regulate the activities of the Upf proteins. Smg1, Smg8, and Smg9 form a protein kinase complex that is responsible for phosphorylating Upf1. Smg1 contains the enzymatic activity in the complex while Smg8 and Smg9 regulate its activity. This is considered to be the rate limiting step in NMD in metazoans (Grimson et al., 2004; Kashima et al., 2006; Yamashita et al., 2009; Yamashita et al., 2001). Upf1 contains numerous phosphorylation sites although none of these sites seem to play a significant individual role, but they may work together to regulate Upf1 activity (Durand et al., 2016). The phosphorylation of Upf1 by the Smg1 kinase requires Upf2 and Upf3, as well as other components of the exon-junction complex (EJC) (Lykke-Andersen and Bennett, 2014) (see below). Structurally related Smg5, Smg6, and Smg7 proteins are important for the de-phosphorylation of Upf1 (Page et al., 1999; Yamashita, 2013). These proteins all contain a 14-3-3 (Jonas et al., 2013) motif and bind to phosphorylated Upf1 (Chakrabarti et al., 2014; Fukuhara et al., 2005; Jonas et al., 2013; Ohnishi et al., 2003; Okada-Katsuhata et al., 2012). In addition, Smg5 and Smg6 contain PIN motifs which are related to those found in the RNase H family of proteins (Glavan et al., 2006). Consistent with this characteristic, Smg6 has endonucleolytic activity and cleaves mRNAs to mark them for decay (Boehm et al., 2014; Eberle et al., 2009; Huntzinger et al., 2008; Lykke-Andersen et al., 2014; Schmidt et al., 2015). Smg5 and Smg7 then form a heterodimer and recruit the Ccr4-Not deadenylase to promote mRNA deadenylation (Loh et al., 2013; Unterholzner and Izaurralde, 2004), and decay. To continue the phosphorylation and de-phosphorylation cycle in other rounds

of NMD, Smg5 interacts with the protein phosphatase PP2A (Anders et al., 2003; Ohnishi et al., 2003; Okada-Katsuhata et al., 2012; Page et al., 1999) to induce de-phosphorylation of Upf1.

Eukaryotic Release Factors

Recognition of an mRNA as a substrate for NMD is dependent on identification of a premature termination codon in the coding region of the mRNA. mRNAs, therefore, need to be translated and the ribosome must encounter a nonsense codon in its A site. Upf proteins have been shown to interact with the release factors, eRF1 and eRF3 (Czaplinski et al., 1995; Kashima et al., 2006; Singh et al., 2008). These interactions are thought to recruit the Upf factors to the mRNA. However, the exact role and the order of events that lead to recruitment of Upf proteins to a prematurely terminating ribosome are unclear. Upf1's interaction with eRF1 and eRF3 leads to the formation of the SURF complex (Smg1-Upf1-release-factor) in higher eukaryotes (Kashima et al., 2006). Yeast Upf2 and Upf3 both interact with eRF3 and may be in competition with eRF1 for an eRF3 interaction domain (Wang et al., 2001). Upf1 binding to GST-eRF1 or GST-eRF3 decreases its ATPase activity (Czaplinski et al., 1995), possibly leading to stabilization of an mRNA-Upf1 complex. Upon binding to Upf2 and Upf3, the stalled termination complex is resolved with re-activation of Upf1's enzymatic activities (Chamieh et al., 2008). While there are no direct observations for this order of events, there is circumstantial evidence supporting this model (see below).

Exon Junction Complex

In higher eukaryotes pre-mRNA splicing results in deposition of a protein complex near the site of splice junctions (Le Hir et al., 2000). This complex, aptly named exon

junction complex (EJC), consists of four core proteins (eIF4A3, Y14, Magoh and MLN51) (Bono and Gehring, 2011), as well as other ancillary proteins. However, the EJC is not present at all splice junctions (Sauliere et al., 2012) and could be present in non-canonical complexes (Singh et al., 2012) or show heterogeneity in its composition (Hauer et al., 2016). Nonetheless, the EJC has substantial roles determining the fates of mRNAs on which it resides. A role for the EJC in NMD is supported by several lines of evidence. First, Upf3 directly interacts with the EJC complex (Kim et al., 2001; Lykke-Andersen et al., 2000). Second, the presence of an EJC near a PTC, while not strictly required, greatly enhances the efficiency of NMD (Buhler et al., 2006; Singh et al., 2008). Third, artificially tethering EJC components near a termination codon can render those mRNAs susceptible to NMD (Gehring et al., 2005; Gehring et al., 2003; Lykke-Andersen et al., 2001). However, these lines of evidence cannot fully substantiate a role for the EJC in NMD. Since translation is needed for NMD, and the presence of an EJC can enhance translation of an mRNA (Chazal et al., 2013; Ma et al., 2008; Nott et al., 2004), an indirect effect cannot be ruled out.

NMD and translation termination

Even though both premature and normal termination events occur with a nonsense codon in the ribosome's A site, there must be substantial differences between these events. Understanding the cause and the consequences of these differences between the two termination events is the key to understanding how NMD is activated.

Normal Translation Termination

Translation termination happens when a translating ribosome reaches the end of an open reading frame and its A site is occupied by a nonsense codon. This process

includes recognition of the nonsense codon and hydrolysis of the ester bond between the tRNA and the nascent amino acid chain. The standard genetic code has three nonsense codons (UAA, UAG, and UGA) and when these codons are in the A site they are recognized by eRF1, a class 1 release factor (Ivanova et al., 2007). eRF1 is also responsible for release of the nascent peptide. The two glycines in the GGQ motif of eRF1 adopt a specific conformation that allows the placement of this motif in the ribosomal peptidyl transferase center, which in turn causes conformational changes in the rRNA and allows access to a water molecule, exposing the ester bond to attack (Jin et al., 2010; Korostelev et al., 2010; Laurberg et al., 2008; Song et al., 2000; Weixlbaumer et al., 2008).

eRF1 has three conserved domains that are present in all eukaryotes: the N-terminal (N), middle (M), and C-terminal (C) domains (Song et al., 2000). The M domain consists of a β sheet and four α helices that adopt a conformation that exposes the GGQ motif (Frolova et al., 1999; Ivanova et al., 2007). Domain N is responsible for stop codon recognition mainly via its TASNIKS and YxCxxxF motifs, where residues interact with different nucleotides of nonsense codons (Bertram et al., 2001; Brown et al., 2015; Cheng et al., 2009; des Georges et al., 2014; Frolova et al., 2002; Hatin et al., 2009; Kolosov et al., 2005; Lekomtsev et al., 2007; Pillay et al., 2016; Salas-Marco et al., 2006; Seit-Nebi et al., 2002) in conjunction with longer range structural interactions (Blanchet et al., 2014; Kryuchkova et al., 2013). Finally, the C domain of eRF1 is important for its interactions with eRF3. The solution structure of eRF1 shows an extended conformation which seems unsuitable to catalyze peptide hydrolysis and nonsense recognition simultaneously (Mantsyzov et al., 2010). This observation suggested that, at some point during

termination, eRF1 must undergo significant conformational changes to achieve its goal (Brown et al., 2015; Matheisl et al., 2015). The induction of such conformational changes is accomplished by eRF1's interaction with its partner, eRF3. Consequently, unlike its bacterial counterparts, eRF1 alone is very inefficient *in vitro* (Alkalaeva et al., 2006).

eRF3 is a class II release factor (Kisselev and Buckingham, 2000), a ribosome dependent GTPase. Via its C-terminal domain, eRF3 can form a stable complex with eRF1 (Merkulova et al., 1999), which is further enhanced by eRF1's M domain (Kononenko et al., 2008). eRF1 and 80S ribosomes are essential for stimulation of eRF3's GTPase activity, although this stimulation can occur in the absence of peptide hydrolysis (Frolova et al., 1996). Like eRF3, the activity of eRF1 is also greatly enhanced by its interacting partner (Alkalaeva et al., 2006). Upon GTP hydrolysis by eRF3 the conformational changes cause eRF1's M domain to rotate and point the GGQ motif to the ribosome peptidyl transferase center. eRF1, in accordance with its role, then assumes a tRNA like structure (Brown et al., 2015; Matheisl et al., 2015).

Following nascent peptide release, the ribosome, a tRNA in the P site, and at least eRF1 on the A site (Pisarev et al., 2010; Pisareva et al., 2007) is still associated with the mRNA. The release of these factors from the mRNA is crucial not only for re-utilization of ribosomes in other cycles of translation, but also for the remaining ribosomes on the same mRNA to finish their respective rounds of translation. The release of the 60S subunit from the mRNA following translation termination is accomplished by the essential ABC cassette ATPase ABCE1 (Rli1 in yeast) (Pisarev et al., 2010). ABCE1's Walker A and Walker B domains are important for ATPase activity, Fe-S domain mediates interactions with eRF1 C-terminal domain (Brown et al., 2015; Kiosze-Becker et al., 2016; Preis et al.,

2014). ABCE1 can only release the 60S subunit from the mRNA if it is preceded by eRF1 binding, but not by puromycin induced release, indicating the importance of sequential assembly (Pisareva et al., 2011). Similarly, ABCE1 and eRF3 bind to the same domain on eRF1, and eRF1's stimulation of ABCE1 activity is muted in the presence of eRF3 (Shoemaker and Green, 2011). The activity of ABCE1 relies on Mg^{2+} ions yet high concentrations of Mg^{2+} can also inhibit this reaction (Sims and Igarashi, 2012). Consistent with its role in ribosome release, depletion of ABCE1 in yeast results in significant nonsense codon readthrough, even at normal termination codons (Young et al., 2015).

After the release of the 60S subunit the mRNA is still associated with the 40S and a tRNA. The removal of these factors is accomplished by the concerted efforts of eIF3, eIF1, eIF1A, and the non-essential subunit eIF3j (Pisarev et al., 2007; Pisarev et al., 2010). eIF3 is a large complex that binds to the solvent exposed side of the ribosome (Siridechadilok et al., 2005), and eIF1 and eIF1A bind to the intersubunit side of the small ribosomal subunit (Kolupaeva et al., 2003; Rabl et al., 2011). Interactions of the 40S subunit with eIF1 and 1A induce conformational changes which result in opening of the mRNA channel and probably the release of the mRNA (Passmore et al., 2007). In addition, the binding on eIF3j involves interactions with the mRNA channel near the A site and this binding is thought to show negative cooperativity with the mRNA, further reducing the affinity of 40S for the mRNA (Fraser et al., 2007). While eIF1, eIF1A and eIF3j can completely dissociate *in vitro* assembled termination complexes, activities of the rest of eIF3 are crucial *in vivo* (Pisarev et al., 2007; Rodnina, 2010).

Premature termination is different from normal termination

Although both normal and premature termination events begin when a nonsense codon occupies the ribosomal A site, the two events show considerable differences. In addition to the fact that most normal termination events do not trigger NMD, two other lines of evidence support this hypothesis. First, toeprint analyses in yeast cell-free extracts fail to show any signal from ribosomes positioned at a normal termination codon unless translation termination is rendered inefficient by a temperature-sensitive mutation in the gene encoding eRF1 (Amrani et al., 2004). In contrast, premature termination events readily yield toeprints in the same assay without any release factor inactivation (Amrani et al., 2004). This observation has been corroborated by similar results in mammalian cell-free systems using the β-globin mRNA (Peixeiro et al., 2012). Second, nonsense readthrough can occur at a much higher frequency at premature termination codons than at normal termination codons (Johansson and Jacobson, 2010; Kvas et al., 2012; Maderazo et al., 2000; Roy et al., 2015; Wang et al., 2001).

Consistent with the notion that Upf proteins play important roles at premature termination events deletion of any one of these proteins results in increased readthrough at a PTC (Johansson and Jacobson, 2010; Kvas et al., 2012; Maderazo et al., 2000; Roy et al., 2015; Wang et al., 2001). However, these observations can, in part, be attributed to increased intracellular Mg^{2+} concentrations due to increased stability of the mRNAs for the uORF-containing Mg^{2+} transporters, Alr1 and Alr2 (Johansson and Jacobson, 2010). Adding another layer of complexity to the issue, the efficacy of ribosome recycling is dependent on Mg^{2+} levels, as ABCE1 activity is hindered in high Mg^{2+} concentrations (Sims and Igarashi, 2012) (see above). It is, therefore, possible that these effects are due

to decreased effects of the ABCE1 ATPase. However, there is currently no evidence linking ABCE1 to ribosome recycling at premature termination events.

Further evidence suggesting that premature termination is markedly different from normal termination is that ribosome re-utilization after premature termination is reduced *in vivo* and *in vitro* upon deletion of any of the Upf proteins (Amrani et al., 2004; Ghosh et al., 2010), suggesting that these termination events lack efficient ribosome recycling mechanisms in the absence of Upf proteins. Similarly, ATPase- or helicase-deficient Upf1 leads to accumulation of decay intermediates that are blocked in subsequent steps in mRNA decay after being targeted by NMD (Serdar et al., 2016) (see below). These observations suggest that, similar to the NSD or NGD pathways, one role of the Upf proteins is to dissociate otherwise poorly dissociable mRNP complexes (Ghosh et al., 2010) .

Different models for NMD targeting

Despite decades of research, our understanding of how NMD is activated remains incomplete. Published work from many groups has converged on three separate models of NMD activation. While there are common elements in each model they are fundamentally different in their view of how NMD selects its substrates.

Exon junction complex model

While this model does not apply to budding yeast, the inspiration for this hypothesis originates from observations that the degradation of nonsense-containing *PGK1* mRNA requires both a downstream element and a *trans*-activating factor (Hrp1) that recognizes this sequence (Gonzalez et al., 2000). Several observations made in mammalian cells led to the development of a related model. First, almost all mammalian genes contain

introns and these are marked by EJCs after pre-mRNA splicing (Le Hir et al., 2000) and the normal stop codons for most mRNAs in higher eukaryotes are located in the last exon, meaning that there is no EJC after the annotated termination event (Nagy and Maquat, 1998; Zhang et al., 1998). Second, the EJC travels with the mRNA to its location for translation and is removed during the first round of translation by the translation machinery-associated protein, Pym (Gehring et al., 2009). Finally, the EJC's association with Upf3 directly connects this complex to NMD (Kim et al., 2001; Lykke-Andersen et al., 2001). These and other observations led to the conclusion that, during the first round of translation, if a terminating ribosome is within close proximity of an EJC, that mRNA will be targeted for NMD. Following recognition of a PTC it is hypothesized that Upf1 is recruited to the mRNA and then later activated by Upf2 and Upf3, followed by mRNA decay (see below).

While intriguing, there is experimental evidence that contradicts the EJC model. For example, mRNAs can still be targeted by NMD in the absence of any introns (Rajavel and Neufeld, 2001; Singh et al., 2008). Recent reports suggested that the EJC itself is more heterogeneous than previously anticipated and not all exon junctions contain this complex (Singh et al., 2012). Further, an implication of the pioneer round of translation model is that only mRNAs that are bound to the cap binding protein, CBP, can be targeted for NMD, yet there is ample evidence suggesting that eIF4E bound mRNAs can be substrates of NMD (Durand and Lykke-Andersen, 2013; Rufener and Muhlemann, 2013). Finally, these results are in direct contrast with observations in yeast that NMD can target mRNAs at any time as evidenced by the rapid decay of hundreds of NMD substrates in a strain with GAL-regulated NMD (Johansson et al., 2007; Maderazo et al., 2003).

Upf1 3'-UTR length sensing model

Premature termination would result in a long 3'-UTR, and some WT mRNAs with unusually long UTRs are also targets for NMD. Studies investigating the consequences of a long 3'-UTR suggested that Upf1 preferentially associates with these UTRs without any regard for translation or sequence (Hogg and Goff, 2010). However, this association could be disrupted by translating ribosomes. These observations, in conjunction with CLIP-Seq experiments (Gregersen et al., 2014; Hurt et al., 2013; Kurosaki et al., 2014; Zund et al., 2013; Zund and Muhlemann, 2013), led to the model that Upf1 can non-specifically associate with mRNAs and that the extent of Upf1 occupancy of an mRNA can be a signal for NMD activation (Zund and Muhlemann, 2013). However, this model fails to account for several other observations. At least in yeast the protein levels of Upf1 are not sufficient to have several molecules of Upf1 per mRNA (Ghaemmaghami et al., 2003). Most importantly, translation independent association of Upf1 with mRNAs directly contradicts many observations indicating the essentiality of translation for NMD (Atkin et al., 1995; Atkin et al., 1997; Hu et al., 2010; Smith et al., 2014; Zhang et al., 1997).

faux-UTR model

First conceived to explain NMD in yeast (Amrani et al., 2004), the principal components of this model are supported by many observations from other model organisms (Behm-Ansmant et al., 2007a; Kertesz et al., 2006; Singh et al., 2008). This model suggests that factors present at a normal 3'-UTR, or at the poly(A) tail of an mRNA, are responsible for proper translation termination and the absence of these features or factors at a PTC is what distinguishes a premature termination event from a normal one. The main lines of evidence that support this model are: 1) The efficiency of NMD targeting

is inversely correlated to how close the premature termination is to the normal termination codon (Gatfield et al., 2003); 2) Artificial extensions of a 3'-UTR can target that mRNA for NMD (Muhlrad and Parker, 1999a); 3) Tethering of the poly(A)-binding protein or eRF3 can inhibit targeting of mRNAs by NMD (Amrani et al., 2004; Behm-Ansmant et al., 2007a; Eberle et al., 2008; Ivanov et al., 2008; Kervestin et al., 2012; Silva et al., 2008); and 4) As mentioned above, premature termination or ribosome recycling are inefficient compared to their normal counterparts (Amrani et al., 2004).

The *faux*-UTR model can explain much of the evidence gathered from studies that used reporter systems for studying NMD. All the experimental constructs in these studies contained PTCs so they could be targeted by NMD. Yet, genome-wide studies in yeast (He et al., 2003) suggested that the largest subset of endogenous substrates of NMD includes mRNAs that do not contain a PTC in their open reading frames so the mechanism for targeting these mRNAs cannot be explained by any of the NMD activation models.

Mechanism of NMD targeted mRNA decay

Regardless of the mode of Upf1 targeting, NMD substrates are degraded by multiple mechanisms. In yeast, these mRNAs are degraded by decapping by the Dcp1/Dcp2 complex, followed by 5' to 3' decay by the Xrn1 exonuclease (He and Jacobson, 2001; Muhlrad and Parker, 1994) (see above). The decapping complex is thought be recruited to the mRNP by direct interaction with Upf1 (He and Jacobson, 1995, 2015a). NMD targeted mRNAs can also be subjected to deadenylation followed by 3' to 5' decay by the exosome, although this method seems to be preferred only in the absence of an active 5' to 3' decay pathway (Mitchell and Tollervey, 2003; Takahashi et al., 2003).

In higher eukaryotes, decay events are preceded by endonucleolytic cleavage by the Smg6 protein which cleaves the mRNA near the prematurely terminating ribosome, leading to degradation of each fragment by Xrn1 or the exosome (Boehm et al., 2014; Eberle et al., 2009; Huntzinger et al., 2008).

Work over the last two and half decades has shed light into many aspects of NMD, such as the identities of the core regulators and their interacting partners, the biochemical properties of many of the proteins involved in this pathway, and the many endogenous substrates of NMD. The inherent link between translation termination and NMD is supported by many lines of evidence from all model organisms studied. However, the exact mechanism of Upf recruitment to NMD targets and why some of the endogenous NMD substrates are targeted by NMD is still unclear. Because NMD requires translation termination, the last and non-rate-limiting step of a long and intricate biological process, understanding the differences between normal and premature translation termination and the mode of NMD activation has been challenging. Future work considering translation termination dynamics and the potentially ordered assembly of translation termination and Upf factors can shed light into this important quality control pathway.

NON-STOP AND NO-GO DECAY

NMD is not the only mRNA quality control mechanism that monitors translation in the cell. Non-Stop Decay (NSD) (Frischmeyer et al., 2002; van Hoof et al., 2002) targets mRNAs that do not contain stop codons and, therefore, have ribosomes that have translated into the poly(A) tail of the mRNA. Conversely, No-Go Decay (NGD) (Doma and Parker, 2006) targets mRNAs on which ribosomes are stalled due to secondary structure and to a lesser extent long stretches of non-optimal codons. While the substrates of these

two pathways seem very different from one another they are recognized by the same factors.

Non-Stop Decay (NSD)

Initially discovered in yeast (Frischmeyer et al., 2002; van Hoof et al., 2002) NSD is a conserved mRNA quality control pathway that targets mRNAs that do not harbor a stop codon and, therefore, elongating ribosomes reach the poly(A) tail. The consequences of translating the poly(A) tail are twofold. First, this results in synthesis of a highly basic poly-lysine polypeptide which is targeted for degradation by the proteasome (Ito-Harashima et al., 2007; Wilson et al., 2007). Second, the translating ribosome is stalled at the end of the mRNA with an empty A site, which is resolved by the interacting partners, Dom34 and Hbs1 (Pisareva et al., 2011; Shoemaker et al., 2010) (see below), (Shoemaker et al., 2010).

After being selected for decay, the mRNA is initially targeted for endonucleolytic cleavage (Tsuboi et al., 2012) followed by exosome mediated decay (Frischmeyer et al., 2002; Inada and Aiba, 2005; Wilson et al., 2007). Like NMD, translation of transcripts is required for detection of NSD substrates and therefore these mRNAs are associated with polysomes (Guydosh and Green, 2014, 2017; Tsuboi et al., 2012) and inhibition of translation with cycloheximide inhibits NSD (Inada and Aiba, 2005).

Consistent with the translation of the poly(A) tail these substrates do not have Pab1 associated with their poly(A) tails (Inada and Aiba, 2005), although it is unclear whether the ribosome is responsible for the removal of the poly(A)-binding protein from the mRNA. It is possible that the absence of Pab1 can play an important role in further destabilizing

these transcripts by disrupting the closed-loop complex and repressing translation (see above).

No-Go Decay (NGD)

Another mRNA quality control pathway that utilizes the Dom34-Hbs1 pair is the NGD pathway. Again, first discovered in yeast (Doma and Parker, 2006), this pathway is also well conserved from archaea to mammals. Targets of NGD include mRNAs with substantial secondary structures (Doma and Parker, 2006) and to a lesser extent mRNAs with long stretches of non-optimal codons (Chen et al., 2010).

Like NSD, the substrates of this pathway undergo endonucleolytic cleavage (Doma and Parker, 2006), but the identity of the endonuclease that performs this reaction remains unknown. The targets of NGD are cleaved just 5' of a ribosome stall site (Chen et al., 2010). Following cleavage and ribosome dissociation by Dom34, Hbs1, and ABCE1 ATPase (Pisareva et al., 2011; Shoemaker et al., 2010) (see below), the 5' fragment is subjected to 3' to 5' decay by the exosome and the 3' fragment is degraded 5' to 3' by Xrn1 (Doma and Parker, 2006). As expected, translation of the mRNA is required for detection of NGD substrates. Therefore, introduction of stem-loops into the 5'-UTR of mRNAs does not result in NGD targeting (Doma and Parker 2006).

Dom34 and Hbs1

While the substrates of NSD and NGD seem very different, one common feature of these transcripts is that they are typified by stalled ribosomes. Considering that NSD substrates have an empty A site and NGD substrates can have any codon in the A site, a codon-independent mechanism to release these stalled complexes is needed. Consequently, the conserved proteins Dom34 (Pelota in higher eukaryotes) and Hbs1

have evolved to address this issue. Consistent with its function in releasing stalled ribosomes, the structure and activity of the Dom34-Hbs1 complex closely resembles that of eRF1-eRF3 (see above) (Becker et al., 2011; Chen et al., 2010; Graille et al., 2008) with some important distinctions. Unlike eRF1, Dom34 does not contain a GGQ motif (Chen et al., 2010); therefore, Dom34-Hbs1-GTP complex cannot perform peptide hydrolysis (Pisareva et al., 2011; Shoemaker et al., 2010). Nonetheless, GTP hydrolysis by Hbs1 induces similar conformational changes in Dom34. Followed by these changes, binding and ATP hydrolysis by the ABCE1 ATPase promotes 60S dissociation (Pisareva et al., 2011) and promotes ribosome recycling.

NGD and NSD are not the only responsibilities of Dom34. This protein is also implicated in the re-activation of translation after glucose starvation (van den Elzen et al., 2014). *dom34*Δ strains show subtle polysome defects compared to WT suggesting a general rescue mechanism for stalled ribosomes (van den Elzen et al., 2014). In line with this hypothesis, Dom34 has been implicated in releasing ribosomes that are stuck in 3'-UTRs of mRNAs (Guydosh and Green, 2014, 2017).

CONCLUSIONS

There are several avenues of research in the study of mRNA decay in yeast. While we have a thorough understanding of the factors involved in many of the decay processes, our understanding of how these pathways interact is still poorly understood. For example, the mechanism of substrate selection by the decapping complex and its interacting partners is still an open question. How the Xrn1 exonuclease is recruited to specific decay complexes or the preference for 5' to 3' as opposed to 3' to 5' decay in regular mRNA decay remains to be determined. An *in vitro* system where mRNA

translation coupled with decay could provide valuable insights into how substrates for each decay mechanism is selected. Elucidation of a more complete interaction network of decapping activators and the decapping complex will also provide a better understanding of step-wise recruitment of decay factors and potentially mutually exclusive complexes that target different mRNPs for decapping. Of course, the underlying implication is that different decay pathways regulate different aspects of the cellular machinery and that there is a direct relationship between turnover and function. Similarly, the order of assembly of Upf proteins in NMD and how these factors recruit decapping and decay factors remains to be resolved and an *in vitro* system where Upf targeting and decay can be closely monitored can shed light into the mechanism of substrate selection by NMD.

Chapter II

High Resolution Profiling of NMD Targets in Yeast Reveals Translational Fidelity as a Basis for Substrate Selection

The work presented in this chapter has been published as:

Celik A, Baker R, He F, Jacobson A.

High Resolution Profiling of NMD Targets in Yeast Reveals Translational Fidelity as a Basis for Substrate Selection. RNA (2017), 23: 735-48

Feng He prepared RNA-Seq libraries and performed Northern blotting

Richard Baker wrote scripts that generated transcriptome 1

Illumina sequencing of these libraries, was carried out at Beijing Genomics Institute (BGI).

SUMMARY

Nonsense-mediated mRNA decay (NMD) plays an important role in eukaryotic gene expression yet the scope and the defining features of NMD-targeted transcripts remain elusive. To address these issues, we re-evaluated the genome-wide expression of annotated transcripts in yeast cells harboring deletions of the *UPF1*, *UPF2*, or *UPF3* genes. Our new RNA-Seq analyses confirm previous results of studies, but also uncover hundreds of new NMD-regulated transcripts that had escaped previous detection, including many intron-containing pre-mRNAs and several non-coding RNAs. The vast majority of NMD-regulated transcripts are normal-looking protein-coding mRNAs. Our bioinformatics analyses reveal that this set of NMD-regulated transcripts generally have lower translational efficiency and higher ratios of out-of-frame translation. NMD-regulated transcripts also have lower average codon optimality scores and higher transition probability to non-optimal codons. Collectively, our results generate a comprehensive catalog of yeast NMD substrates and yield new insights into the mechanisms by which these transcripts are targeted by NMD.

INTRODUCTION

Nonsense-mediated mRNA decay (NMD) is a eukaryotic surveillance mechanism that targets mRNAs undergoing premature translation termination for rapid degradation (He and Jacobson, 2015b; Kervestin and Jacobson, 2012; Lykke-Andersen and Bennett, 2014). The pathway was initially uncovered in *Saccharomyces cerevisiae* and *Caenorhabditis elegans* (Leeds et al., 1991; Peltz et al., 1993; Pulak and Anderson, 1993) and later shown to be conserved from yeast to humans (Behm-Ansmant et al., 2007b; Schoenberg and Maquat, 2012). NMD's function was originally thought to be limited to quality control, i.e., the elimination of mRNAs derived from genes harboring nonsense mutations to prevent the accumulation of potentially deleterious truncated polypeptides (He et al., 1993; Pulak and Anderson, 1993). However, NMD also targets a significant fraction of apparently normal and physiologically functional wild-type mRNAs (Schweingruber et al., 2013), indicating that it also serves as a fundamental post-transcriptional regulatory mechanism for eukaryotic gene expression. Consistent with these important roles, NMD function is linked to diverse cellular processes, including cell growth and proliferation (Avery et al., 2011; Lou et al., 2014; Weischenfeldt et al., 2008), development and differentiation (Gong et al., 2009; Medghalchi et al., 2001; Metzstein and Krasnow, 2006; Wittkopp et al., 2009), innate immunity (Gloggnitzer et al., 2014), antiviral or stress responses (Balistreri et al., 2014; Sakaki et al., 2012), and neuronal activity or behavior (Colak et al., 2013; Giorgi et al., 2007).

In all organisms examined the activation of NMD requires a set of conserved core regulatory factors, Upf1, Upf2, and Upf3 (He and Jacobson, 2015b; Kervestin and Jacobson, 2012). These three proteins interact with each other, the ribosome, and

multiple translation and mRNA decay factors (Kervestin and Jacobson, 2012). Based on these molecular interactions, several potential functions have been proposed for the Upf factors, including remodeling terminating mRNPs (Franks et al., 2010), releasing and recycling ribosomal subunits (Ghosh et al., 2010), and recruiting mRNA decay factors (He and Jacobson, 2015a; Nicholson et al., 2014; Okada-Katsuhata et al., 2012). However, the exact roles for the Upfs, and their modes of action in NMD, remain largely unknown.

Despite the conservation of the core Upf proteins, NMD-targeted mRNAs appear to be degraded by different mechanisms in different eukaryotic cells. In yeast, NMD-targeted mRNAs are degraded predominantly through a deadenylation-independent mechanism involving decapping by the Dcp1/Dcp2 decapping enzyme and 5' to 3' exonucleolytic digestion by Xrn1 (He and Jacobson, 2001; Muhlrad and Parker, 1994). In human cells, NMD-targeted mRNAs are degraded through multiple mechanisms including endonucleolytic cleavage (Eberle et al., 2009; Huntzinger et al., 2008; Lykke-Andersen et al., 2014), deadenylation-dependent decapping (Loh et al., 2013; Unterholzner and Izaurralde, 2004; Yamashita et al., 2005), and exosome-mediated 3' to 5' decay (Lejeune et al., 2003), with endonucleolytic decay appearing to be the predominant initiating mechanism in human cells (Boehm et al., 2014). In the latter decay pathway, Smg6 cleaves its substrate mRNAs in the vicinity of PTCs and the resulting 5' and 3' fragments are degraded by the exosome and Xrn1, respectively (Boehm et al., 2014; Eberle et al., 2009; Huntzinger et al., 2008).

Depending on the organism or cell type, about 5 to 20% of the transcripts in a typical transcriptome are substrates of NMD (He et al., 2003; Lelivelt and Culbertson, 1999; Mendell et al., 2004; Ramani et al., 2009; Rehwinkel et al., 2005; Weischenfeldt et

al., 2008) and these transcripts can be classified into several general categories. One category, exemplifying typical NMD substrates, includes mRNAs with a destabilizing premature termination codon (PTC) in their coding region. These transcripts are generated from endogenous genes harboring nonsense or frameshift mutations (He et al., 2003), pseudogenes (He et al., 2003; McGlincy and Smith, 2008), non-productively rearranged genetic loci (Li and Wilkinson, 1998), or from alternative splicing events that lead to intron retention or inclusion of a PTC-containing exon (Jaillon et al., 2008; Lareau et al., 2007; Lykke-Andersen et al., 2014; Ni et al., 2007). A second category contains mRNA-like transcripts with limited or no apparent coding potential, such as long noncoding RNAs (Kurihara et al., 2009; Lykke-Andersen et al., 2014; Tani et al., 2013), small RNAs derived from intragenic regions (Smith et al., 2014; Thompson and Parker, 2007), or transcripts of inactivated transposable elements (He et al., 2003). A third category contains a subset of physiologically relevant transcripts that appear to be "normal," such as mRNAs with upstream open reading frames (uORFs) (Arribere and Gilbert, 2013; Gaba et al., 2005; He et al., 2003), or with atypically long 3'-UTRs (Kebaara and Atkin, 2009; Singh et al., 2008), or normal-looking wild-type mRNAs with no atypical features (He et al., 2003).

To generate a comprehensive and high resolution catalog of NMD-regulated transcripts, and to delineate the defining features of these transcripts in NMD targeting, we utilized RNA-Seq to re-evaluate the effects of deleting the *UPF1*, *UPF2*, or *UPF3* genes on the transcriptome-wide expression of annotated yeast genes. Our new analyses confirm previous results of microarray studies, but also uncover hundreds of new NMD-regulated transcripts that had escaped previous detection, including many intron-

containing pre-mRNAs. Our bioinformatics analyses reveal several intrinsic features of NMD-regulated transcripts, yield new insights into the mechanisms by which translation of these transcripts targets them for NMD, and provide strong support for the notion that transcripts can become NMD substrates at any time during their translational life cycle.

RESULTS

Upf1, Upf2, and Upf3 regulate a common set of transcripts in yeast cells

To obtain a high resolution catalogue of yeast NMD substrates, we analyzed expression profiles of wild-type (WT), *upf1Δ*, *upf2Δ*, and *upf3Δ* strains. RNA-Seq libraries were prepared from these strains in three biological replicates and the sequence reads from each library were aligned to: 1) a yeast transcriptome comprised of 7473 transcripts, including all annotated protein-coding sequences, functional and non-coding RNAs, and the unspliced isoforms of all intron-containing mRNAs and 2) a separate transcriptome comprising all 3569 CUT, SUT, and XUT transcripts annotated previously (Wery et al., 2016). We opted to use two separate transcriptomes because combining them resulted in a substantial loss of power to detect differential expression. Further investigation revealed that overlapping annotations were the principal cause: 1655 out of 3569 genomic coordinates of CUT, SUT, and XUT sequences had some extent of overlap, sometimes with multiple annotations (2024 total overlaps). Less than 1% of reads mapping to CUTs, SUTs, and XUTs were unique, compared to >55% for other transcripts. Based on posterior probabilities calculated by the RSEM software tool (Li and Dewey, 2011), including CUTs, SUTs, and XUTs in a combined analysis would have resulted in a substantial increase in read count quantification error for all transcripts, compromising dispersion estimates used for subsequent differential expression calculations (Anders and Huber, 2010; Robinson et al., 2010). Similarly, because of their repetitive nature, we also excluded autonomous replicating sequences and long terminal repeats of transposable elements from our analysis.

We used RSEM (Li and Dewey, 2011) for transcript quantification and the DESeq R package (Anders and Huber, 2010) for differential expression analysis. To account for replicate variability, we used a false discovery rate threshold of 0.01 instead of arbitrary fold-change as a criterion for differential expression. All libraries exhibited similar distributions of read counts with few outliers (Fig. 2.1A) and biological replicates were extremely consistent, with Pearson correlation coefficients ranging from 0.84 to 0.99 (Fig. 2.2). We refer to the set of sequences that contains all annotated transcripts except for CUTs, SUTs, and XUTs as transcriptome 1 (T_1) and the set with the sequences for CUTs, SUTs, XUTs as transcriptome 2 (T_2).

Deletion of *UPF1*, *UPF2*, or *UPF3* led to differential expression of a subset of transcripts in yeast cells. In each of the *UPF* deletion strains, the vast majority of differentially expressed transcripts were up-regulated and only a small number of transcripts were down-regulated (Figs. 2.1B-D). The number of up- or down-regulated transcripts in individual *UPF* deletion strains were comparable and exhibited substantial overlap (Figs. 2.1B, C). These results indicate that the three Upf factors control the expression of a common set of transcripts in yeast cells. In transcriptome 1, we identified 907 transcripts that were up-regulated and 29 that were down-regulated upon *UPF* deletion; under the same circumstances transcriptome 2 had 332 up-regulated transcripts (including 8 CUTs, 114 SUTs, and 210 XUTs; Table 2.1) and 124 down-regulated transcripts (Figs 2.1B). Pairwise comparisons of read counts for each of the 1392 differentially expressed transcripts in *upf1Δ*, *upf2Δ*, and *upf3Δ* cells manifested almost equivalent expression values in each case (Fig. 2.1E), indicating that deletion of *UPF1*, *UPF2*, or *UPF3* has almost identical quantitative effects on NMD-regulated transcripts

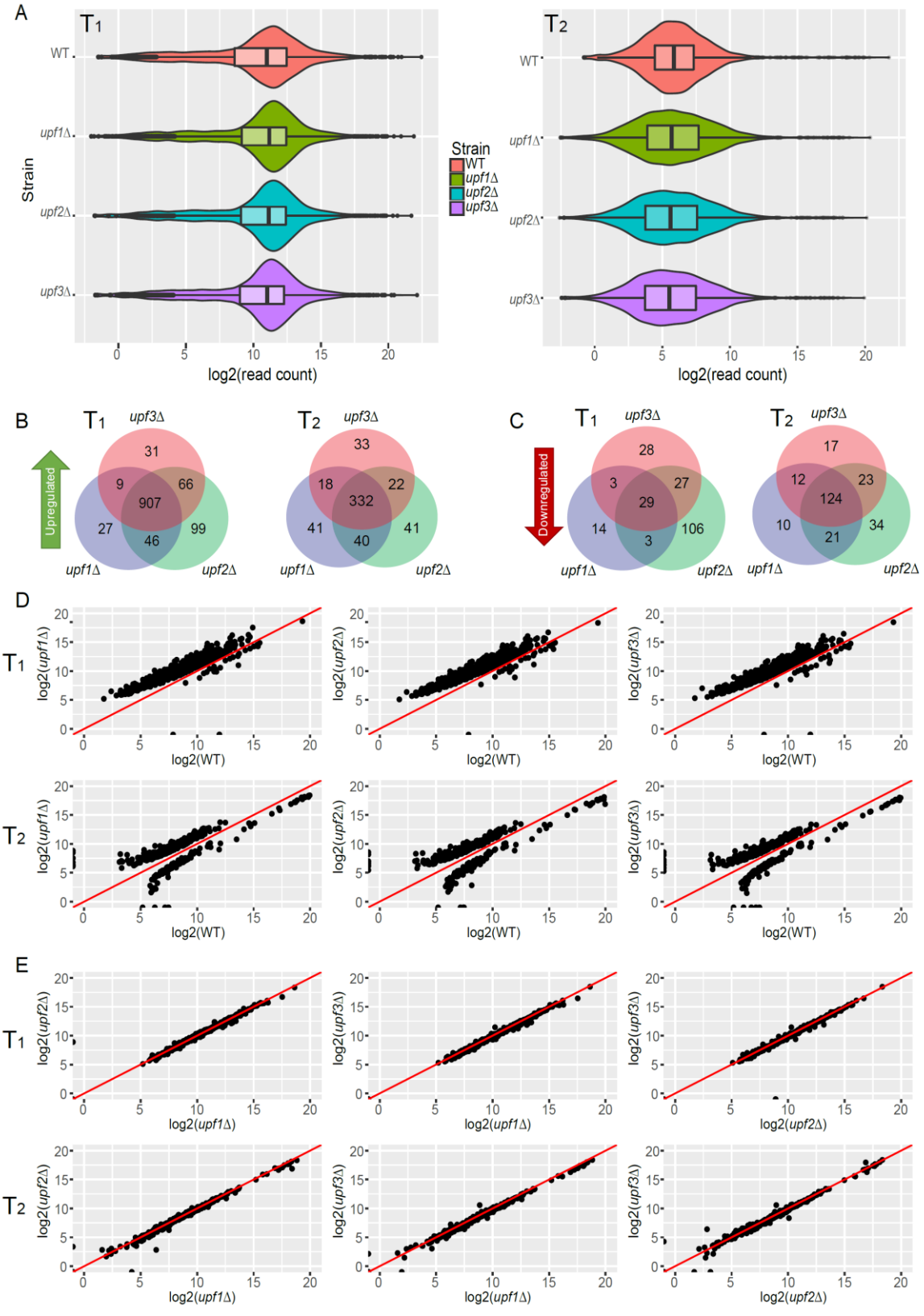
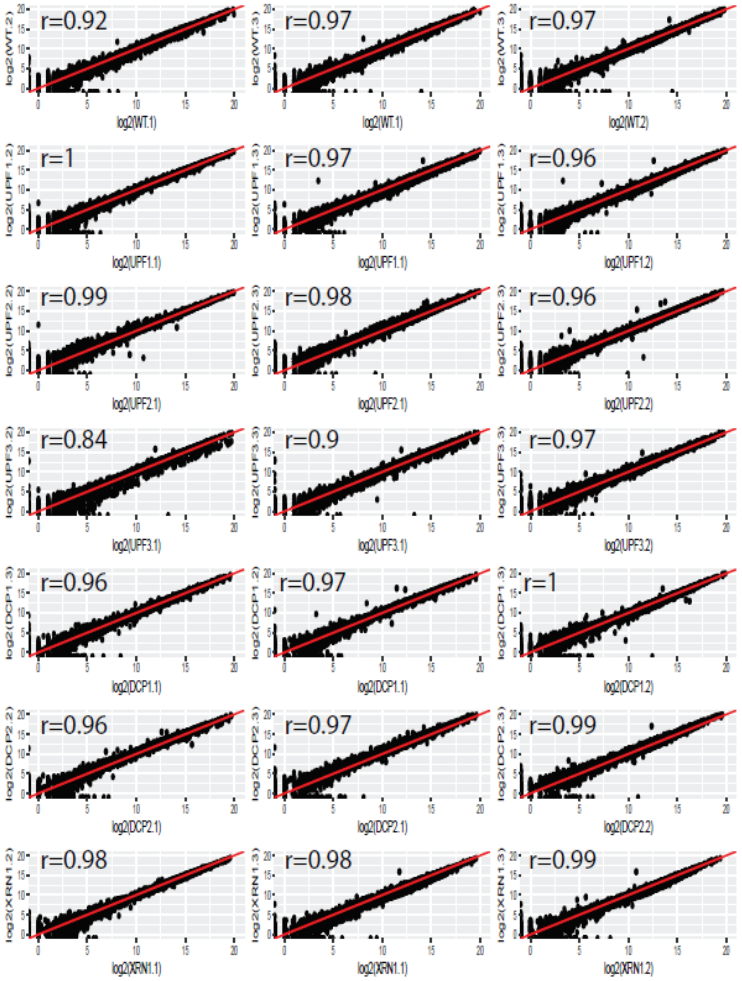

A
T1
T2
WT
upf1Δ
upf2Δ
upf3Δ
Strain
log2(read count)
B
C
Upregulated
Downregulated
D
E
log2(WT)
log2(upf1Δ)
log2(upf2Δ)
log2(upf3Δ)

Figure 2.1. Upf1, Upf2, and Upf3 regulate the same set of transcripts in yeast.
A. RNA-Seq libraries from *WT*, *upf1Δ, upf2Δ,* and *upf3Δ* strains display comparable overall read count distributions for both transcriptome 1 (T_1; left panel) and transcriptome 2 (T_2; right panel). Violin and box-plots were used to visualize the average sequence reads distribution of the transcriptomes of the indicated strains from three independent experiments. **B.** and **C.** Transcripts up- and down-regulated in *upf1Δ, upf2Δ* and *upf3Δ* strains show significant overlap. Transcripts up- or down-regulated in each of *UPF* deletion strains were identified by comparisons to the WT strain. Venn diagrams were used to display the relationships among the sets of transcripts that are up-(B) and down-(C) regulated in T_1 or T_2 of *upf1Δ, upf2Δ,* and *upf3Δ* strains. **D.** All three *UPF* deletion strains display similar genome-wide expression patterns. Scatterplots were used to compare the read count values of differentially expressed transcripts between WT and *upf1Δ, upf2Δ,* or *upf3Δ* strains. The vast majority of differentially expressed transcripts in *UPF* deletion strains showed up-regulation and a small number of transcripts showed down-regulation. The y=x line is shown in red. Top panel: pairwise comparisons of the expression levels between WT and each *UPF* deletion strain for 936 differentially expressed transcripts from transcriptome 1. Bottom panel: pairwise comparisons of the expression levels between *WT* and each *UPF* deletion strain for 456 differentially expressed transcripts from transcriptome 2. **E.** Transcripts commonly regulated by NMD each have virtually identical expression values in *upf1Δ, upf2Δ,* or *upf3Δ* strains. As in D, scatterplots were used to compare the read count values of NMD-regulated transcripts between *upf1Δ and upf2Δ*, *upf1 and upf3Δ, and upf2Δ* or *upf3Δ* strains. Top panel: differentially expressed transcripts from transcriptome 1. Bottom panel: differentially expressed transcripts from transcriptome 2.

Transcriptome 1



Transcriptome 2

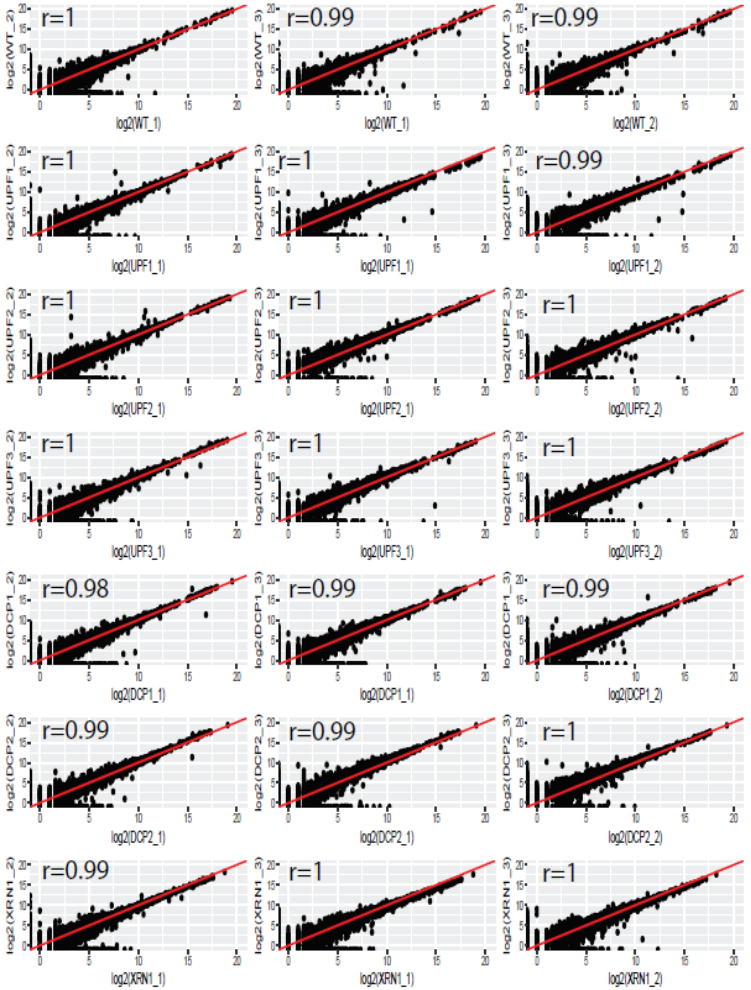


Figure 2.2. RNA-Seq libraries generated from WT, *upf1Δ, upf2Δ, upf3Δ*, *dcp1Δ, dcp2Δ, and xrn1Δ* strains show good reproducibility and correlation between biological replicates.

The figure depicts pairwise scatterplot comparisons of estimated read counts for all transcripts in either transcriptome 1(left panel) or transcriptome 2 (right panel) between different biological replicates of the same strains. The y=x line in red and the Pearson correlation coefficient for pairwise comparison are indicated in each plot.

	Total	Detected	Upregulated upon *UPF* deletion	Downregulated upon *UPF* deletion
All transcripts	11040	10438	1239	153
Protein coding	6937	6581	902	27
Intron within ORF	351	301	57	0
5'-UTR intron	24	24	5	1
uORF containing*	356	356	89	3
Transposable element	50	50	8	1
Mitochondrial	43	27	0	1
Non-Coding**	493	421	5	2
tRNA***	359	295	0	0
Intron containing tRNA	60	0	0	0
CUT	925	821	8	35
SUT	849	738	114	22
XUT	1795	1664	210	67

Table 2.1. Different categories of transcripts that were analyzed, detected, and commonly up- or down-regulated in all three *UPF* deletion strains.

Sequence reads from individual libraries of WT, *upf1Δ, upf2Δ,* and *upf3Δ* strains were separately mapped to transcriptomes 1 and 2 using the differential expression pipeline summarized in Methods. Transcripts differentially expressed in *upf1Δ, upf2Δ,* and *upf3Δ* strains were identified using the Bioconductor DESeq package. Comparisons to the WT strain were based on three independent experiments. The first nine categories of transcripts were from transcriptome 1 and the last three categories of transcripts were from transcriptome 2.

* uORF-containing transcripts based on the annotations of Ingolia et al. (2009).

** Transcripts annotated as non-coding including rRNAs, snRNAs, snoRNAs, and antisense regulatory RNAs, but not tRNAs.

*** Most tRNAs are detected at extremely low levels and their read count values are probably unreliable as our RNA-Seq library construction protocol has a strong bias against small RNAs.

These results strengthen the idea that, at least in yeast, Upf1, Upf2, and Upf3 have equivalent effects on the execution of a single NMD pathway (He et al., 1997).

Transcripts regulated by NMD have also been identified in previous analyses (He et al., 2003; Johansson et al., 2007; Malabat et al., 2015; Sayani et al., 2008; Smith et al., 2014) and we sought to compare our new transcript lists to those generated earlier. However, naming conventions for Affymetrix microarray probes are not completely consistent with annotated gene names and a significant fraction (~20%) of transcripts represented by microarray probes are not in our new reference transcriptomes. Therefore, we restricted our comparisons to microarray probes that have definitively matched transcripts in our transcriptomes. As shown in Table 2.2, there is substantial overlap (>60%) with our results and previously published datasets. These results indicate that our RNA-Seq analyses yielded comprehensive sets of NMD-regulated transcripts in yeast cells that include the majority of transcripts identified by previous analyses and numerous transcripts that were not detected in previous studies. Transcripts commonly up-regulated in all three *UPF* deletion strains did not show any significant enrichment for those encoding factors involved in signal transduction or transcriptional regulation and, although a handful of up-regulated transcripts encode transcription factors, the annotated targets for each of these transcription factors were not particularly enriched in our NMD-regulated transcript list. Although we cannot definitively rule out indirect effects that alter mRNA abundance, these observations and the known roles of the three Upf factors as positive regulators of NMD (He and Jacobson, 2001) suggest that the 1239 transcripts

	He et al. (2003)	Johansson et al. (2007)	Malabat et al. (2015)	Malabat et al. (2015) Protein coding	Smith et al. (2014)	Smith et al. (2014) Protein coding	Sayani et al. (2008)
Contains protein coding transcripts	+	+	+	+	+	+	+
Contains noncoding transcripts	–	–	+	–	+	–	–
Contains CUTs, SUTs, XUTs	–	–	+	–	+	–	–
Identified as NMD substrates	570	450	2280	872	1167	873	38
Also identified in this study	468	353	672	545	700	576	23
% Overlap	82%	78%	30%	63%	61%	66%	61%

Table 2.2. Overlap of NMD-regulated transcripts identified in this study and in several previous studies.

RNA-Seq analyses in this study identified 1239 transcripts commonly up-regulated in all three *UPF* deletion strains. This set of transcripts was compared to several other previously published datasets (He et al., 2003; Johansson et al., 2007; Malabat et al., 2015; Sayani et al., 2008; Smith et al., 2014) which had the following key features: He et al. – this list, identified by microarray analysis, includes the common set of transcripts upregulated in response to single *UPF* deletions; Johansson et al. - this list, identified by microarray analysis, includes the set of transcripts that are down regulated upon transcriptional activation of a galactose-regulated *UPF2* gene; Sayani et al. – this list, identified using tiling arrays, includes intron-containing transcripts that were >2x-upregulated in *upf1*Δ strains; Smith et al. – this dataset was generated using the bioinformatics pipeline described in this study and two independent RNA-Seq experiments from WT and *upf1*Δ strains; and Malabat et al. – the authors' analyses employed an FDR of 1% to identify the set of transcripts upregulated in *upf1*Δ cells.

commonly up-regulated in the two transcriptomes analyzed here likely constitute direct substrates of NMD. Because of the ambiguity of read alignments and lack of information about the biological role of CUT, SUT, and XUT sequences in the literature, most of our subsequent analyses were focused on the 907 up-regulated transcripts present in transcriptome 1.

Structural and functional classes of NMD-regulated RNAs in transcriptome 1

To gain insight into the mechanism targeting well annotated NMD substrates, we classified the NMD-regulated transcripts of transcriptome 1 into their respective structural and functional categories. Among the 907 up-regulated transcripts of transcriptome 1, 902 were from annotated protein-coding genes and 5 were from annotated “non-coding” RNA genes (Table 2.1). Since NMD requires ongoing translation (Hu et al., 2010; Zhang et al., 1997), our observation that some “non-coding” RNAs of transcriptome 1 are substrates of NMD suggests that these transcripts are actually translated and may encode *bona fide* polypeptides. Among the 902 NMD-regulated transcripts coming from protein coding genes, 88% appear to be “normal” mRNAs and to lack any structural features indicative of substrates of NMD (see below). The remaining 12% can be classified into five known structural classes of NMD substrates (He et al., 2003), namely: 1) mRNAs encoded by genes harboring nonsense mutations in their coding regions (e.g., *CAN1, LEU2*), 2) mRNAs utilizing frameshiftting in their translation (e.g., *YGR109W-B, YIL082W-A, and YIL009C-A*), 3) transcripts originated from pseudogenes (e.g. *YAR061W, YFL056C, YOL153C*), 4) mRNAs that contain annotated and putative upstream open reading frames (Ingolia et al., 2009) (uORFs; e.g., *CPA1*), and 5) pre-mRNAs that retain their introns and enter the cytoplasm as a consequence of inefficient

or regulated splicing (e.g., *RPL22B, RPL24B, HRB1*). The latter two classes (uORF-containing mRNAs and intron-containing pre-mRNAs) were enriched significantly in NMD substrates: 89 out of 356 putative uORF-containing yeast transcripts (Ingolia et al., 2009) and 57 out of 351 potential intron-containing pre-mRNAs were present at a higher ratio than would be expected by chance (chi-square $p=4.8 \times 10^{-9}$ and 0.0006, respectively). Our observation that NMD targets a large number of intron-containing pre-mRNAs indicates that there is a widespread entry of intron-containing transcripts into the yeast cytoplasm and that, in yeast, NMD plays a general role in the degradation of a subset of intron-containing pre-mRNAs.

Validation of newly identified NMD substrates

To validate our RNA-Seq and bioinformatics results, we assessed the levels of expression of seven newly identified NMD-regulated transcripts in wild-type, *upf1Δ*, *upf2Δ*, and *upf3Δ* strains by northern blotting. To ascertain NMD substrate specificity, we also analyzed yeast strains harboring single deletions of genes encoding the major cytoplasmic 5' to 3' exonuclease (Xrn1), the Dcp1/Dcp2 decapping enzyme, and several decapping activators (Edc3, Pat1, Lsm1, Dhh1, and Scd6). Among the seven selected NMD-regulated transcripts analyzed, four (*HRB1, RPL22B, NHP6B, MTR2*) are intron-containing pre-mRNAs, two are annotated as "non-coding" RNAs, and one utilizes frameshifting during translation (Ty4 transposon). Of the intron-containing pre-mRNAs, two (*HRB1, RPL22B*) contain an intron in their coding regions and two (*NHP6,* MTR2) contain an intron in their 5'-UTRs. As a negative control, we analyzed the intron-containing *HAC1* pre-mRNA in these strains. All seven NMD-regulated transcripts showed the expected expression patterns in these yeast strains (Fig. 2.3, A-C).

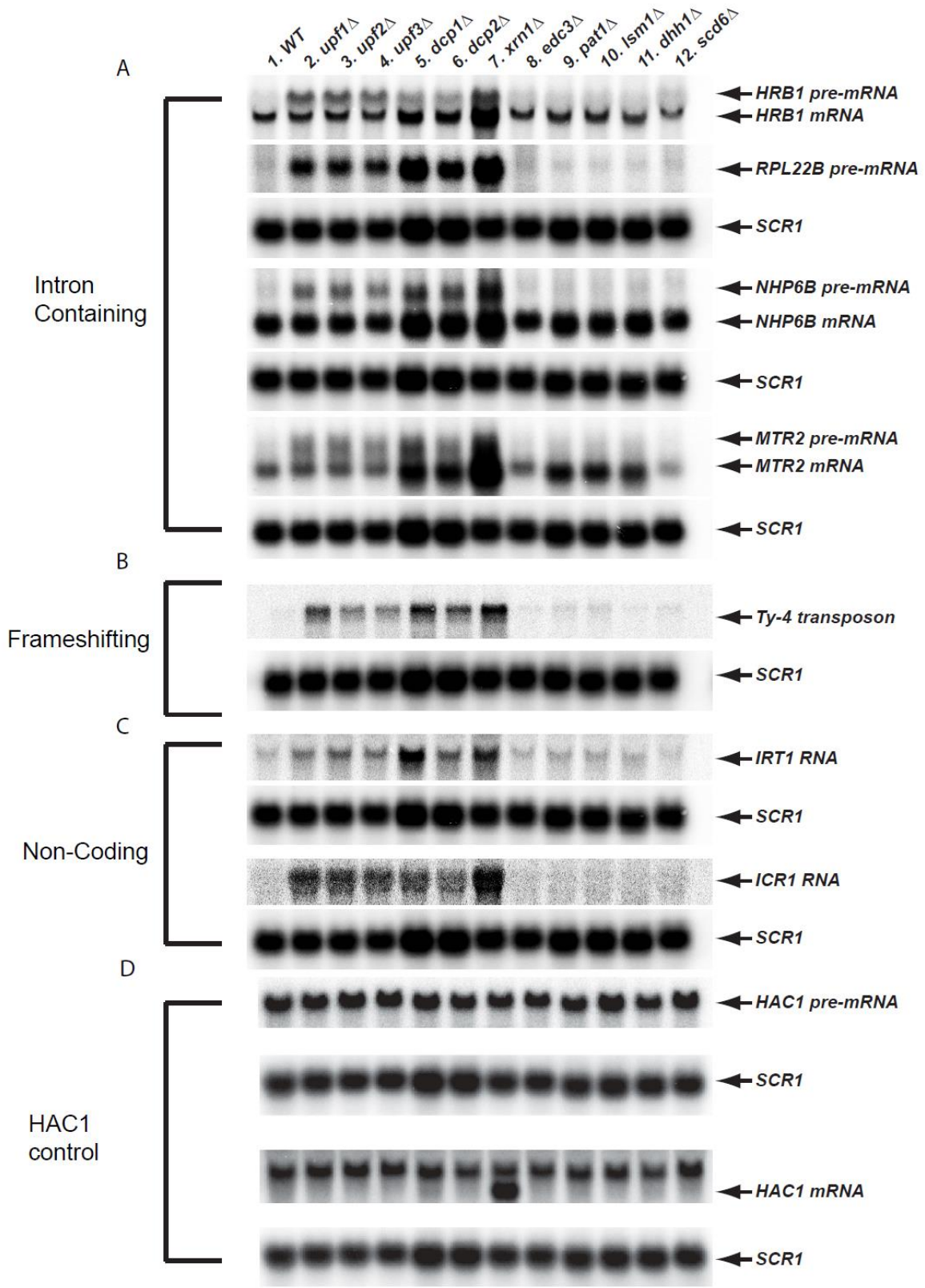

1. WT
2. upf1Δ
3. upf2Δ
4. upf3Δ
5. dcp1Δ
6. dcp2Δ
7. xrn1Δ
8. edc3Δ
9. pat1Δ
10. lsm1Δ
11. dhh1Δ
12. scd6Δ
A
Intron Containing
HRB1 pre-mRNA
HRB1 mRNA
RPL22B pre-mRNA
SCR1
NHP6B pre-mRNA
NHP6B mRNA
SCR1
MTR2 pre-mRNA
MTR2 mRNA
SCR1
B
Frameshifting
Ty-4 transposon
SCR1
C
Non-Coding
IRT1 RNA
SCR1
ICR1 RNA
SCR1
D
HAC1 control
HAC1 pre-mRNA
SCR1
HAC1 mRNA
SCR1

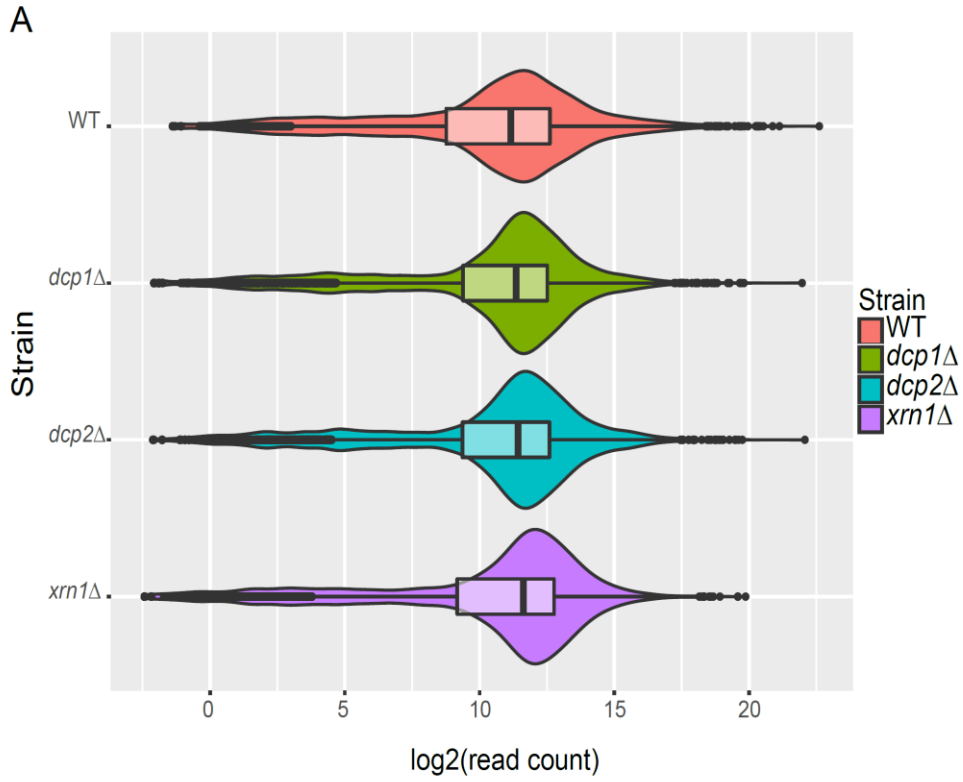
Figure 2.3. Validation of several different classes of NMD substrates by northern blotting.

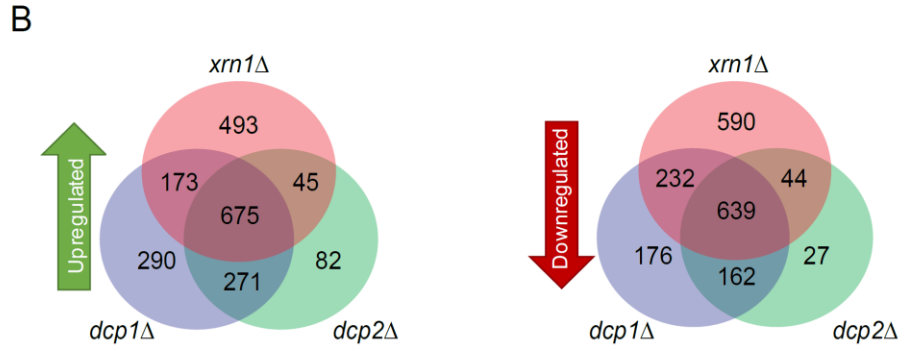
Northern blotting analyses of: **A.** Intron-containing transcripts (*HRB1, RPL22B, NHP6B* and *MTR2*), **B.** Transcripts using frameshifting during translation (Ty-4 transposons), **C.** "non-coding" RNAs (*ICR1* and *IRT1*), and **D.** negative control transcripts (*HAC1* pre-mRNA). Total RNA was isolated from the indicated strains and the steady-state levels of individual transcripts in these strains were analyzed by northern blotting. In each case, a random-primed probe was hybridized to the blot and *SCR1* served as the loading control.

Compared to their expression in wild-type cells, each of these transcripts showed increased levels in *upf1Δ*, *upf2Δ*, and *upf3Δ* strains, and also in *xrn1Δ*, *dcp1Δ*, and *dcp2Δ* strains. However, the levels of expression of each of these transcripts were unchanged in *edc3Δ*, *pat1Δ*, *lsm1Δ, dhh1Δ*, and *scd6Δ* strains. The levels of the *HAC1* pre-mRNA were essentially unchanged in all the deletion strains. Collectively, these observations confirm that all seven transcripts are NMD-regulated and degraded through decapping and 5' to 3' exonucleolytic decay. The latter results mirror our analyses of RNA-Seq data (see below).

NMD substrates are principally degraded by decapping and 5' to 3' exonucleolytic decay

Several NMD substrates have previously been shown to be degraded by decapping and 5' to 3' exonucleolytic degradation (He and Jacobson, 2001; Muhlrad and Parker, 1994), including those analyzed in Fig. 2.3. To evaluate the prevalence of this mechanism in NMD at a transcriptome-wide level, we analyzed the expression patterns of the 907 NMD regulated transcripts of transcriptome 1 in *dcp1Δ, dcp2Δ,* and *xrn1Δ* strains we prepared RNA-Seq libraries from *dcp1Δ, dcp2Δ,* and *xrn1Δ* strains and subjected them to the same analysis pipeline as described above for libraries prepared from *upf* strains. These libraries showed similar consistencies and read count distributions as single WT and *UPF* deletion libraries (Fig. 2.4 A; Fig 2.2). Consistent with current concepts of the roles of Dcp1, Dcp2, and Xrn1 in NMD and general 5' to 3' decay (Parker, 2012), deletion of *DCP1*, *DCP2*, and *XRN1* each caused up-regulation of >1,000 transcripts (Fig. 2.4 B). As expected, the up-regulated transcripts in *dcp1Δ*, *dcp2Δ*, and *xrn1Δ* strains had significant overlap, with overlapping fractions ranging from 52% to 88%


A
WT
dcp1Δ
dcp2Δ
xrn1Δ
Strain
0
5
10
15
20
log2(read count)
Strain
WT
dcp1Δ
dcp2Δ
xrn1Δ


B
xrn1Δ
493
173
45
675
290
82
271
dcp1Δ
dcp2Δ
Upregulated
xrn1Δ
590
232
44
639
176
27
162
dcp1Δ
dcp2Δ
Downregulated

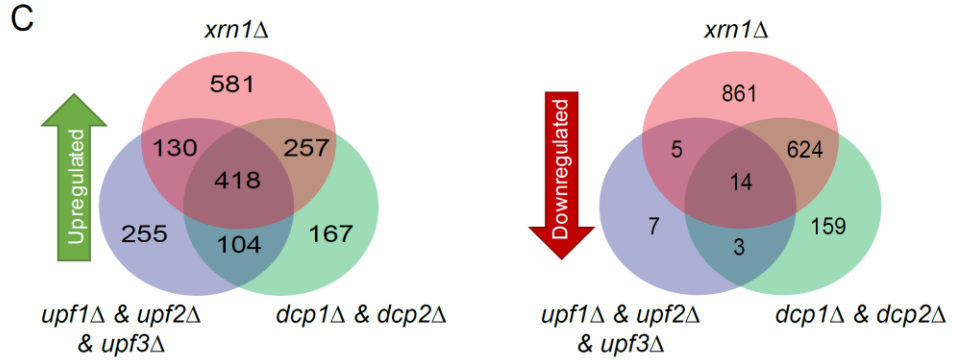

C
xrn1Δ
581
130
257
418
255
104
167
upf1Δ & upf2Δ
& upf3Δ
dcp1Δ & dcp2Δ
Upregulated
xrn1Δ
861
5
624
14
7
159
3
upf1Δ & upf2Δ
& upf3Δ
dcp1Δ & dcp2Δ
Downregulated

Figure 2.4. NMD substrates are principally degraded by decapping and 5' to 3' exonucleolytic decay.

A. RNA-Seq libraries from *WT*, *dcp1Δ, dcp2Δ,* and *xrn1Δ* strains display normal and comparable overall read count distributions. As in Fig. 2.1 A, violin and box-plots were used to visualize the average sequence reads distribution of the transcriptomes of the indicated strains from three independent experiments. **B.** Transcripts up- or down-regulated in *dcp1Δ, dcp2Δ,* and *xrn1Δ* strains show significant overlap. Transcripts up- or down-regulated in *dcp1Δ, dcp2Δ,* and *xrn1Δ* strains (upper and lower panels, respectively) were identified by comparisons to the WT strain. **C.** Transcripts commonly up- or down-regulated from transcriptome 1 in all three *UPF* deletion strains show significant overlap with transcripts up-regulated in both in *dcp1Δ* and *dcp2Δ* strains or an *xrn1Δ* strain. Venn diagrams were used to display the relationships among the up- or down-regulated transcripts from the indicated strains.

(Fig. 2.4 B). The 907 NMD-regulated transcripts had significant overlap with transcripts that were up-regulated in both *dcp1Δ* and *dcp2Δ* cells or *xrn1Δ* cells (Fig. 2.4 C). Overall, approximately 70% of NMD-regulated transcripts had increased levels of expression in *dcp1Δ, dcp2Δ,* and *xrn1Δ* strains. Consistent with our earlier microarray analyses (He et al., 2003), these results indicate that yeast NMD substrates are largely but probably not exclusively degraded by decapping and 5' to 3' exonucleolytic decay.

Intron-containing pre-mRNAs targeted by NMD are engaged in translation

To further elucidate the role of NMD in the degradation of intron-containing pre-mRNAs, we analyzed ribosome occupancy within the intronic regions of pre-mRNAs targeted by NMD (n=57) and those that are not (n=244). Using published ribosome profiling data (Young et al., 2015), we measured ribosome densities (coverage $_{\text{profiling}}$/coverage$_{\text{RNA-Seq}}$) within the intronic regions for these two groups of intron-containing transcripts. We found that introns from the pre-mRNAs targeted by NMD do in fact show a subtle, but statistically significant, higher ribosome occupancy than introns from the pre-mRNAs not targeted by NMD (two sample K.S. test p=0.038; Figs. 2.5 A, B). These results indicate that intron-containing pre-mRNAs targeted by NMD are generally engaged in translation. In support of this conclusion, several of these NMD-targeted pre-mRNAs were previously shown to be associated with polyribosomes (He et al., 1993).

NMD substrates similar to normal mRNAs are poorly translated regardless of the NMD status of the cell

The compilation of a comprehensive list of NMD substrates raises the general question of what dictates NMD specificity for these transcripts. While the NMD targeting

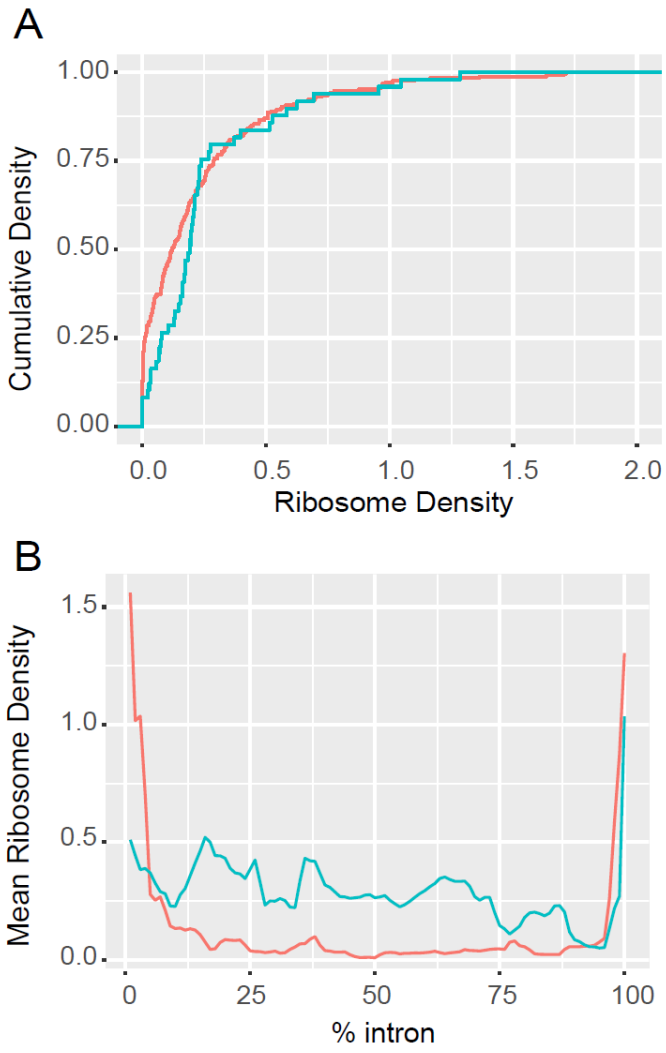


Figure 2.5. NMD targeted intron-containing pre-mRNAs are engaged in translation.

A. Cumulative density plot of ribosome density of the intronic regions for pre-mRNAs targeted (blue, n=57) or not targeted (red, n=244) by NMD. This plot illustrates the fraction (on the Y axis) of transcripts having the indicated ribosome densities (on the x axis). **B.** Distribution of mean ribosome densities over normalized intronic regions for the same two sets of intron-containing transcripts as in **A.** Plots in **A** and **B** were derived from the ribosome profiling data of WT cells by Young et al. (2015). Ribosome densities were calculated as $\text{profiling}_{\text{coverage}}$ / $\text{RNA-Seq}_{\text{coverage}}$ for each intron. Introns of NMD-targeted pre-mRNAs show higher ribosome densities than introns of the pre-mRNAs that are not targeted by NMD (two sample K.S. test p=0.038).

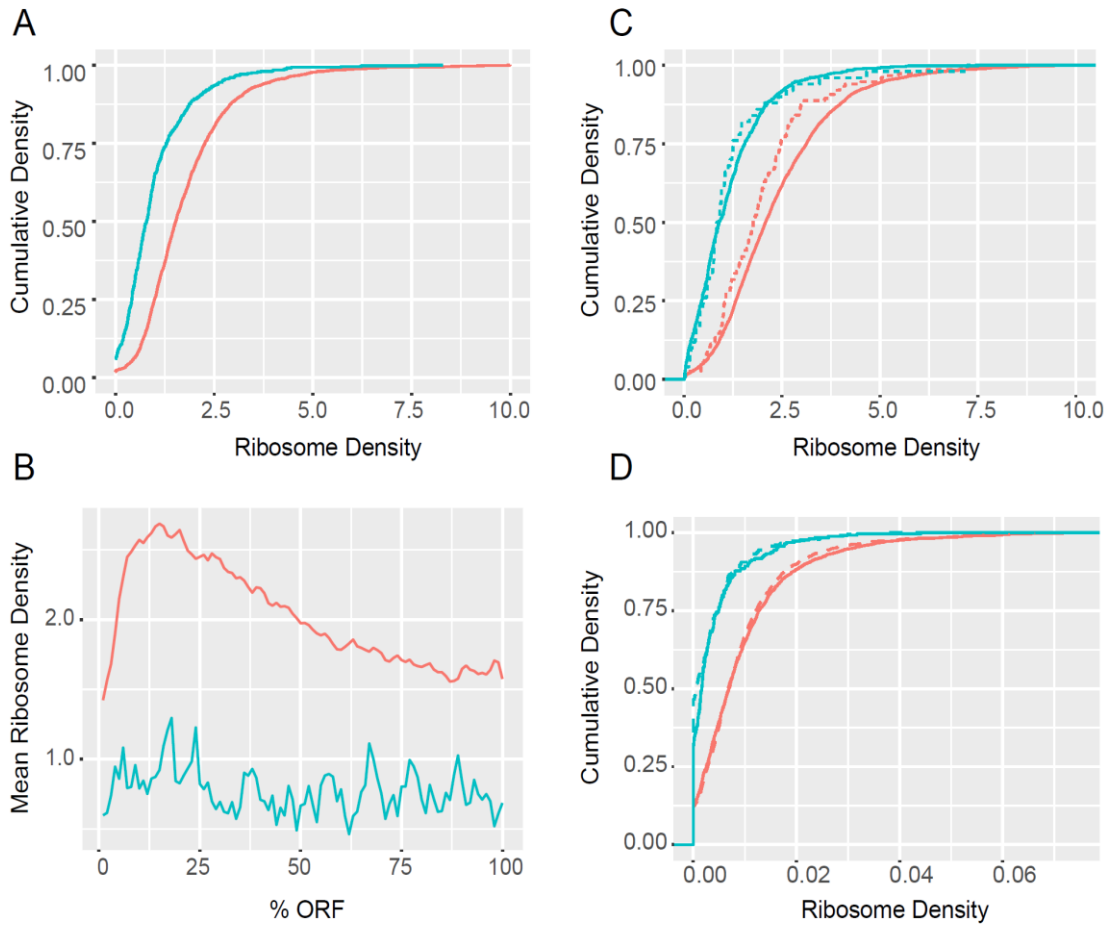


Figure 2.6. NMD substrates are less efficiently translated than non-substrates independent of the NMD machinery.

A. Cumulative density plots of ribosome densities derived from ribosome profiling data of WT cells for normal-looking NMD substrates (blue n=746) and non-NMD substrates (red=4633). **B.** Mean ribosome densities over normalized ORFs derived from the same data and for the same two sets of transcripts shown in **A. C.** Cumulative density plots of ribosome densities derived from the same data in A for uORF-containing transcripts targeted (dashed, blue n=42 NMD) or not targeted (dashed, red n=199 by NMD and for uORF-lacking transcripts targeted (solid, blue n=704) or not targeted (solid, red n=4434) by NMD. **D.** Cumulative density plots of ribosome densities derived from other ribosome profiling datasets of WT (solid) and *upf1Δ* (dashed) cells for normal-looking NMD substrates (blue) and non-NMD substrates (red) shown in **A.** Plots in **A, B**, and **C** were derived from the ribosome profiling data of Young et al (2014) and plots in **D** were derived from the ribosome profiling data of Smith et al (2015). Ribosome densities were calculated as $\text{profiling}_{\text{coverage}}$ / $\text{RNA-Seq}_{\text{coverage}}$ for each transcript. Two-sample K.S. test p values are described in the results section.

of PTC-, uORF-, and intron-containing transcripts, and “non-coding” RNAs, can all be attributed to premature translation termination, the vast majority (almost 90%) of NMD substrates in transcriptome 1 are protein-coding transcripts that look like normal, wild-type mRNAs. To identify potential features associated with this “normal-looking” group of NMD substrates, we evaluated several parameters, including 5’-UTR, ORF, and 3’-UTR lengths, and ribosome densities, for this group of NMD substrates and compared these parameters to those generated from protein-coding mRNAs not subject to NMD regulation. Using previously published annotations (Nagalakshmi et al., 2008; Pelechano et al., 2013; Xu et al., 2009), we observed conflicting results for the potential role of UTR lengths in substrate selection. Using the annotations of Nagalakshmi et al., we found no discernible difference between 5’- and 3’-UTR lengths of normal-looking mRNAs targeted vs. not targeted by NMD, while the annotations of Pelechano et al UTRs are longer for NMD substrates and the annotations of Xu et al. suggested that 5’-UTRs are shorter for NMD substrates (data not shown). These conflicting annotations precluded any conclusions about the role of UTR lengths in the determination of NMD substrate status. However, by comparing the published (Young et al., 2015) ribosome occupancies of these subsets of transcripts we observed a striking difference in normalized ribosome occupancy in wild-type cells. The normal-looking NMD substrates had significantly lower ribosomal density throughout their open reading frames than the non-NMD substrates (two sample K.S. test $p<2.2e-16$) (Figs. 2.6 A, B, blue and red lines, respectively). Based on this observation, we also analyzed the normalized ribosome occupancy of all putative uORF-containing transcripts. We separated the transcripts into two different groups: those regulated by NMD and those not regulated by NMD. Much like the normal-looking

NMD substrates, the NMD-regulated uORF-containing transcripts also exhibited lower ribosome densities than those not subject to NMD regulation (Fig. 2.6 C; two sample K.S. test p<2.2e-16 and p=3.57e-10 for transcripts without and with uORFs, respectively). Together, these results indicated that NMD substrates are probably translated less efficiently than non-NMD substrates.

In addition to triggering rapid transcript degradation, recognition of an mRNA as an NMD substrate has been suggested to lead to concomitant Upf1-dependent translational repression (Muhlrad and Parker, 1999b). To assess whether the observed lower ribosome density of normal-looking NMD substrates in wild-type cells reflected this phenomenon, we utilized published ribosome footprinting libraries (Smith et al., 2014) to compare the normalized ribosome densities of NMD substrates in wild-type and *upf1*Δ cells. These analyses demonstrated that the ribosome density profiles for both normal-looking NMD substrates and non-NMD substrates showed similar differences in ribosome density regardless of the strain (Fig. 2.6 D; two sample K.S. test p<2.2e-16 for both WT and *upf1*Δ, between NMD substrates and non-substrates), i.e., the NMD substrates were also translated less efficiently in *upf1*Δ cells. Collectively, our bioinformatics analyses indicate that low ribosome density is an intrinsic property of NMD-targeted transcripts.

Normal-looking NMD substrates have a higher rate of out-of-frame translation

The comparatively reduced ribosome densities of NMD substrates observed in Fig. 2.6 suggested that these mRNAs may share a common impairment. We thus tested whether normal-looking NMD substrates and non-NMD substrates may have different

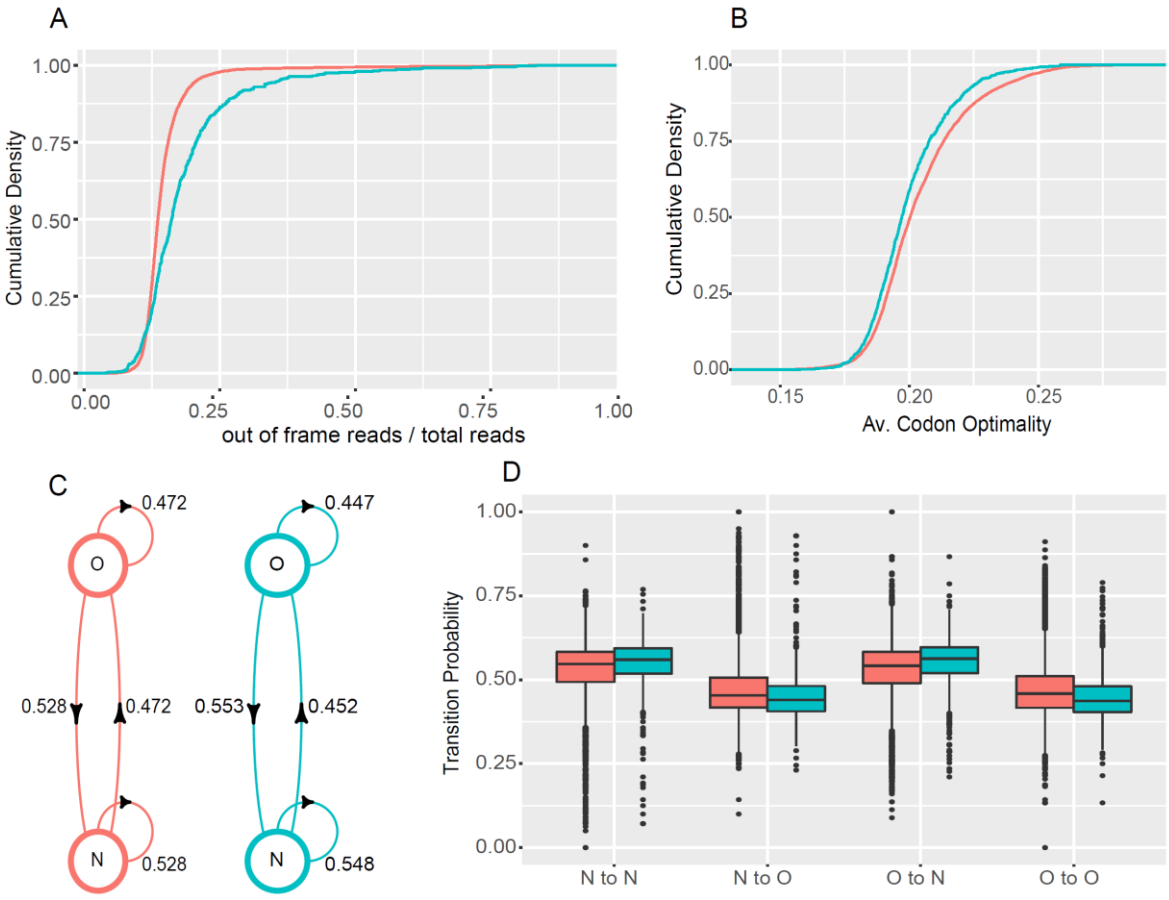


Figure 2.7. NMD substrates have lower translation fidelity and lower codon optimality

A. Cumulative density plots of in-frame read ratios over total reads derived from ribosome profiling data of WT cells for intron-lacking NMD substrates (n= 746, blue) and non-NMD substrates (n=4633, red). **B.** Cumulative density plots of mean codon optimality scores for two sets of transcripts shown in **A**. **C.** Mean transition probabilities of a two-state discrete time Markov chains between optimal (O) and non-optimal (N) codons for intron-lacking NMD substrates (blue) and non-NMD substrates (red). **D.** Distributions of Markov chain codon transition probabilities for intron-lacking NMD substrates (blue) and non-NMD substrates (red). Plots in **A** were derived from the ribosome profiling data of Young et al (2014) and plots in **B, C, and D** were based on codon optimality assignments and scores published by Pechmann and Frydman (2013). Two-sample K.S. test p values are described in the results section.

A

Filter 1

Padj unphased < 0.01
Total fold change > 1.5
Unphased fold change > 1.5
Has iTSS
No uORF

Filter:	Out of frame ratio	Ribosome occupancy
Pass:	1.2e-11	4.1e-8
Fail:	<2.2e-16	<2.2e-16

82 genes out of 5232

Filter 2

Padj unphased <0.01
Total fold change > 1.5
Unphased fold change > 2
Has iTSS

Filter:	Out of frame ratio	Ribosome occupancy
Pass:	0.02	0.004
Fail:	<2.2e-16	<2.2e-16

426 genes out of 5232

Filter 3

Unphased fold change > 1

Filter:	Out of frame ratio	Ribosome occupancy
Pass:	2.9e-14	1.7e-
Fail:	<2.2e-16	<2.2e-16

796 genes out of 5232

Filter 4

Has iTSS
No uORF

Filter:	Out of frame ratio	Ribosome occupancy
Pass:	<2.2e-16	1.6e-11
Fail:	<2.2e-16	<2.2e-16

1577 genes out of 5232

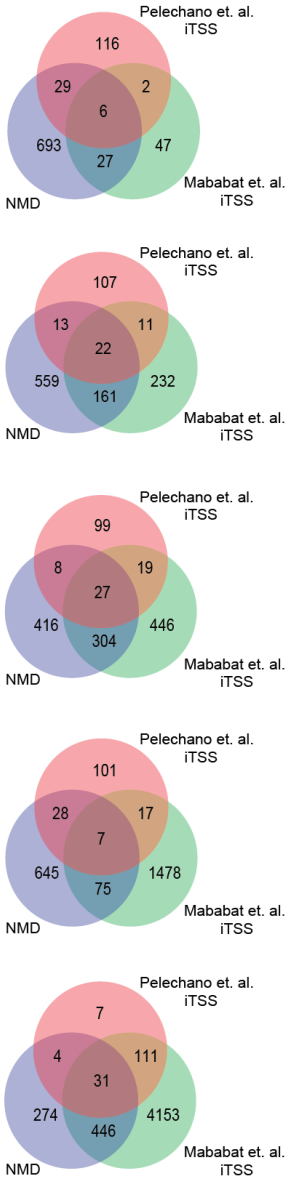
Filter 5

Has iTSS

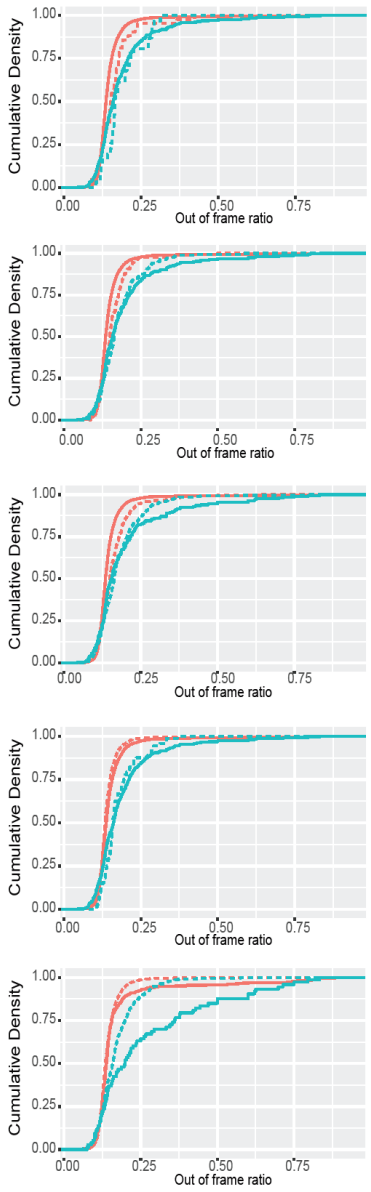
Filter:	Out of frame ratio	Ribosome occupancy
Pass:	<2.2e-16	<2.2e-16
Fail:	<2.2e-16	<2.2e-16

4741 genes out of 5232

B



C



D

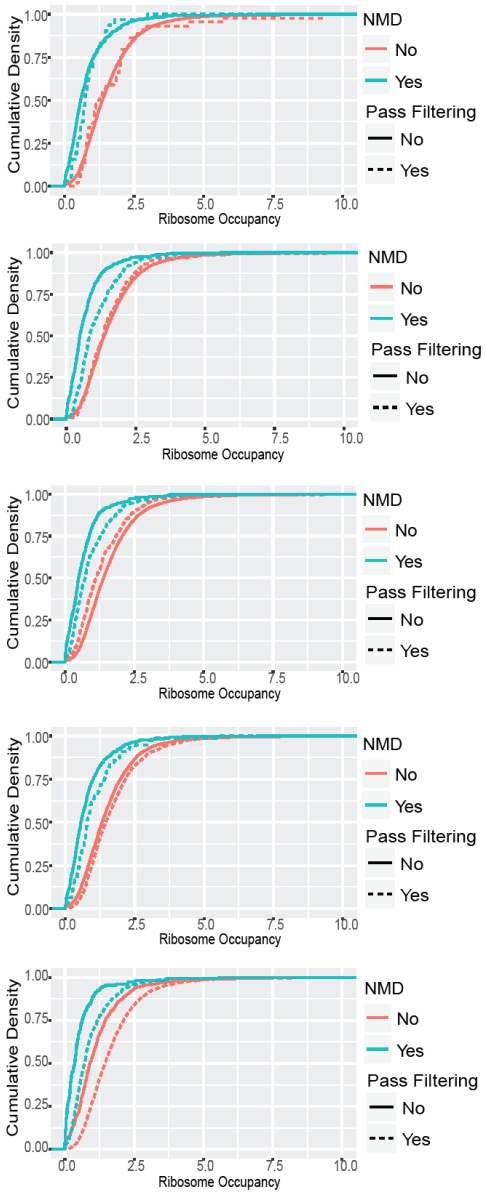


Figure 2.8. Transcripts upregulated upon *UPF* deletion have lower ribosome densities and higher out-of-frame translation regardless of iTSSs.

A. Filtering schemes employed to identify iTSS-containing transcripts in the dataset of Malabat et al. (2015) and subsequent two sample K.S. between different subsets of transcripts. **Padj, unphased**: includes genes that have an out of frame AUG as the first methionine, and that isoform is differentially expressed in *upf1*Δ strains with the indicated FDR cutoff; **Total fold change**: the fold change cutoff in the total number of reads mapping to any kind of transcription start site in WT and *upf1*Δ strains; **Unphased fold change**: fold change in the number of reads that map to a TSS that is in a different frame than the annotated start site; **Has iTSS**: genes that contain iTSSs, according to authors; **No uORF**: genes that do not contain uORFs, according to authors. Two sample K.S. tests between NMD substrates and non- substrates for out of frame read ratios and ribosome occupancies that do or do not contain iTSSs **B.** Overlap of NMD substrates and transcripts that were identified as having iTSSs by Pelechano et al. (2013) and Malabat et al. (2015). **C.** Ratio of out of frame reads in previously published ribosome footprint profiling datasets (Young et al., 2015). Blue lines indicate NMD substrates and red lines indicate non-NMD substrates. Solid lines indicate genes that did not pass the filtering cited in column A and dashed lines indicate the genes that did. **D.** Similar to C with ribosome occupancies.

amounts of out-of-frame translation in their coding regions. For this analysis, we analyzed published (Young et al., 2015) ribosome profiling data and only used read lengths that displayed a strong preference (>80% of reads) for one frame, mapping these sequence reads to NMD and non-NMD-regulated transcripts. We calculated the ratio of out-of-frame reads to total mapped reads for each of these transcripts and compared the out-of-frame ratio distributions between the normal-looking NMD and non-NMD populations. This analysis indicated that the NMD-regulated transcripts showed a significantly higher ratio of out-of-frame reads (Fig. 2.7 A; two sample K.S. test $p<2.2e-16$). These results indicate that in addition to lower ribosome density, NMD substrates also exhibit higher out-of-frame read ratios and thus a higher rate of out-of-frame translation. One potential explanation for increased out-of-frame translation could be that these genes have internal transcription start sites (iTSSs) and the subsequent isoforms are the main substrates for NMD. To test this hypothesis, we used iTSS-containing gene lists published by two independent groups (Malabat et al., 2015; Pelechano et al., 2013). We compared the overlaps between transcripts with iTSSs and our list of NMD substrates. Interestingly, we found little overlap between these three groups of transcripts. In addition, when we compared the ribosome densities and out-of-frame read ratios between transcripts that have been suggested to have iTSSs by Malabat et. al. and those that don't we observed similar differences between NMD substrates and non-substrates. That is, regardless of iTSS status, transcripts that we have identified as NMD substrates show higher rates of out-of-frame translation and lower ribosome densities (Fig. 2.8).

Normal looking NMD substrates have lower average codon optimality and a biased distribution pattern of non-optimal codons

To further understand the basis of lower ribosome density and increased out-of-frame translation of NMD substrates, we explored potential differences in codon usage for normal looking NMD and non-NMD-regulated transcripts. We used published codon optimality data (Pechmann and Frydman, 2013), to calculate average codon optimality scores for both NMD and non-NMD-regulated transcripts, and compared the score distributions of these two populations. We found that the NMD-regulated transcripts had a subtle, but statistically significant lower average codon optimality score (Fig. 2.7 B; two sample K.S. test p=3.9e-7). Based on this finding, we then recoded the codon sequences of each NMD- and non-NMD-regulated transcript as a binary series of optimal (O) or non-optimal (N) codons, treated each recoded transcript as a discrete time Markov chain, and calculated the transition probabilities from one state to another for each transcript (i.e., O to O, O to N, N to N, and N to O). We then compared the distributions of transition probabilities of the NMD and non-NMD-regulated transcripts. We found that NMD-regulated transcripts again showed a subtle but statistically significant preference towards non-optimal codons (i.e., having higher N to N and O to N, but lower N to O and O to O transition probabilities; two sample K.S. test p=2.4e-7, 1.2e-7, 5.4e-14, 5.3e-14, respectively; Figs. 2.7 C, D). Our analyses thus indicated that, as individual metrics, average codon optimality and N to N and O to O transition probabilities all seem to contribute to NMD susceptibility. Because average codon optimality is highly correlated with transition probabilities we were unable to conclude whether NMD substrates still had a higher tendency to exhibit longer stretches of non-optimal codons when controlled for overall codon optimality.

DISCUSSION

A comprehensive catalog of annotated yeast NMD substrates

Using RNA-Seq analyses, we have redefined the set of transcripts regulated by NMD in *Saccharomyces cerevisiae*. Our new comprehensive list of yeast NMD substrates originates from the well annotated genes of transcriptome 1 and includes the vast majority of NMD-regulated transcripts identified by previous analyses (He et al., 2003; Johansson et al., 2007; Malabat et al., 2015; Sayani et al., 2008; Smith et al., 2014), as well as hundreds of new transcripts that escaped prior detection. While many CUTs, SUTs, and XUTs of transcriptome 2 manifest changes in abundance in response to inactivation of each Upf factor the relevance of these transcripts to the conventional understanding of NMD remains obscure, particularly in light of the lack of precise mapping information for these transcripts. Accordingly, our attention has largely been drawn to the components of transcriptome 1. Consistent with the positive roles of Upf1, Upf2, and Upf3 in NMD activation (He and Jacobson, 2015b; Kervestin and Jacobson, 2012), almost all NMD substrates in the transcriptome 1 list are up-regulated by NMD inactivation, with only a handful of transcripts showing down-regulation under the same conditions. Further, the strict requirement for translation in NMD activation (Hu et al., 2010; Zhang et al., 1997) was reflected in the observation that 99% of the NMD substrates in transcriptome 1 were annotated as protein coding transcripts.

Each of the up-regulated transcripts shares a nearly identical quantitative response to deletion of the *UPF1*, *UPF2*, or *UPF3* genes and exhibits comparable expression levels in the three *UPF* deletion strains (Fig. 2.1 E). In addition, most transcripts in the up-regulated group also exhibit increased accumulation upon inactivation of the Dcp1/Dcp2

decapping enzyme and the 5’ to 3’ Xrn1 exonuclease (Fig. 2.4 C), two critical components that function downstream the yeast NMD pathway (Parker, 2012). The set of up-regulated transcripts also includes several known structural classes of NMD substrates including mRNAs encoded by genes harboring nonsense mutations in their coding regions, mRNAs utilizing frameshifting in their translation, pseudogene transcripts, mRNAs that contain uORFs, and pre-mRNAs that retain their introns and enter the cytoplasm as a consequence of inefficient or regulated splicing (Fig. 2.9 A). Collectively, these observations lead us to conclude that the bulk of the up-regulated transcripts identified here are likely to be *bona fide* substrates of the NMD pathway in yeast cells.

A significant fraction of yeast intron-containing mRNAs are targeted by cytoplasmic NMD

In addition to generating a comprehensive catalog of yeast NMD substrates, our RNA-Seq analyses also revealed that a significant fraction (~16%) of yeast intron-containing genes produce intron-containing pre-mRNA isoforms that are engaged with translating ribosomes and subject to NMD regulation (Fig. 2.5; Table 2.1). These observations indicate that, even under normal growth conditions, a significant fraction of intron-containing pre-mRNAs are exported from the nucleus to the cytoplasm, where they are degraded by NMD. The large number of yeast intron-containing pre-mRNAs subject to NMD regulation suggests that NMD plays a much more significant than anticipated role in intron-containing pre-mRNA degradation in yeast cells. Consistent with this conclusion, previous tiling microarray analyses also revealed a large overlapping set of intron-containing pre-mRNAs subject to NMD (Sayani et al., 2008).

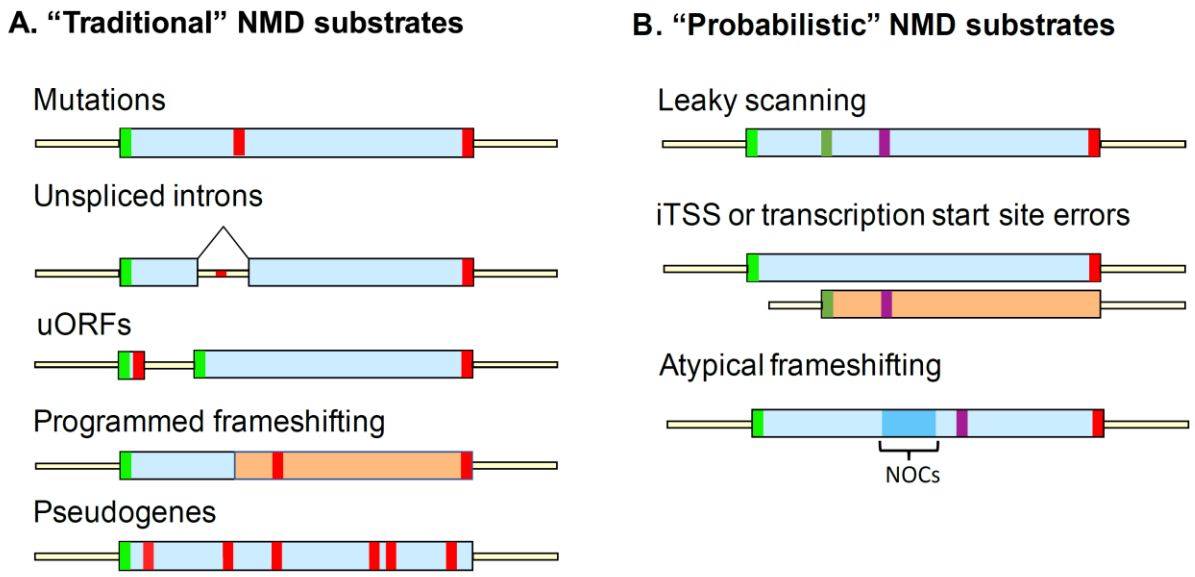


Figure 2.9. Different classes of NMD substrates.

A. "Traditional" NMD substrates. Translation of these NMD substrates commences at initiation codons located at ORF (or uORF) 5' ends, proceeds 3', and leads to an in-frame encounter with a coding region premature termination codon. Transcripts in this class include mRNAs derived from nonsense alleles, pre-mRNAs that enter the cytoplasm with unspliced introns, uORF-containing mRNAs, mRNAs in which programmed frameshifting allows a fraction of ribosomes to avoid premature termination, and mRNAs transcribed from pseudogenes. **B.** "Probabilistic" NMD substrates. These NMD substrates lack in-frame premature termination codons in their coding regions, but contain mRNA features that promote either downstream out-of-frame translational initiation or frameshifting and thus trigger premature termination in a new reading frame. Transcripts in this category include mRNAs with poor sequence context around the normal initiation codon, mRNAs whose transcription start site is internal to the principal ORF, and mRNAs with lower overall codon optimality or a long stretch of non-optimal codons (NOCs). In each of these cases, a subset of ribosomes translate the mRNA in a frame different from that of the annotated ORF. Green=initiation codon; red=stop codon; yellow=UTR; purple=stop codon encountered in the +1 or +2 reading frame; blue=cluster of non-optimal codons.

Potential targeting mechanisms of the normal-looking NMD substrates

The largest fraction of NMD substrates identified here is comprised of normal-looking mRNAs that appear to lack defining features of premature termination. Our bioinformatics analyses reveal several intrinsic properties of these normal-looking NMD substrates that suggest a potential NMD-targeting mechanism. Compared to non-NMD substrates, this group of mRNAs has lower translation efficiency, a higher rate of out-of-frame translation, lower average codon optimality, and a propensity to have stretches of non-optimal codons. Further, in contrast to an earlier proposition (Isken et al., 2008; Muhlrad and Parker, 1999b), the lower translation efficiency for this group of substrates appears to be independent of the NMD machinery.

The intrinsic properties that we uncovered for the normal-looking NMD substrates could reflect direct causes or indirect consequences of NMD targeting for these mRNAs, i.e., some of these properties may function independently or synergistically in NMD targeting. One possible mechanism of NMD targeting may be attributable to translational elongation through a stretch of non-optimal codons. The lower average codon optimality or longer stretches of non-optimal codons of the normal-looking NMD substrates might lead to less efficient translation for this group of transcripts, an increased probability that an error will be made during translation elongation, and the observed higher rate of out-of-frame translation. Clearly, the latter offers a greater likelihood for premature translation termination.

The less efficient translation and the higher rate of out-of-frame translation that we observed for normal-looking NMD substrates could also be caused by events independent of the translation of a stretch of non-optimal codons. One possibility could

be heterogeneity in the primary structure of this group of NMD substrates, i.e., the normal-looking NMD substrates may each have multiple transcript isoforms. Some of the isoforms may have very short 5'-UTRs (Arribere and Gilbert, 2013) and some of the isoforms may result from internal transcriptional initiation in protein coding regions (Malabat et al., 2015). These unusual isoforms could have lower efficiencies of translation initiation at conventional ORF start sites and higher rates of out-of-frame downstream translation initiation. However, the detection of such isoforms as NMD substrates in a transcriptome-wide study would require that they constitute a significant fraction of the mRNA isoform population for a particular gene. Significantly, the less efficient translation and the higher rate of out-of-frame translation that we observed for normal-looking NMD substrates are largely independent of iTSS status (Fig. 2.8). Similarly, alternative splicing events could also produce a subset of transcripts targeted by NMD, but previously reported alternative splicing events in yeast appear to generate only minor mRNA isoforms (Kawashima et al., 2014) and are thus unlikely to be detected as NMD substrates in our study. Therefore, the less efficient translation and higher rates of out-of-frame translation of the normal-looking NMD substrates are unlikely to be caused by transcript isoform heterogeneities, and most likely originate from the intrinsic translation properties of these mRNAs. Further, their propensity for frameshifting is most likely the cause of subsequent premature termination and NMD substrate status. In short, NMD can serve as a probabilistic quality control mechanism that allows for detection of errors during translation elongation. Collectively, atypical transcription or translation initiation, or unexpected frameshifting events, could all be targeted by NMD. In each of these cases, NMD activation is linked to premature or premature-like translation termination, as

observed for several previously characterized classes of transcripts targeted by NMD (Fig. 2.9). This mode of action for NMD is consistent with previously published results in which NMD could target transcripts even after their first round of translation (Gaba et al., 2005; Maderazo et al., 2003).

Chapter III

Regulation of mRNA decapping by competing factors

Feng He prepared RNA-Seq libraries and performed Northern blotting

Illumina sequencing of these libraries, was carried out at Beijing Genomics Institute (BGI).

SUMMARY

mRNA decay is an important step in gene regulation and mRNA decapping can be a rate-limiting step in at least one major decay pathway. Decay rates of different transcripts show large variability, but the exact mechanism by which the Dcp1-Dcp2 decapping complex is regulated is largely unknown. Several *in vitro* and *in vivo* studies have indicated that interactions among the set of decapping activator proteins and the C-terminal extension of Dcp2 may target and activate the decapping complex, but the precise regulatory role of this network of interactions is unclear. To better understand how decapping regulation is achieved by the C-terminal extension of Dcp2 in *Saccharomyces cerevisiae,* we generated RNA-Seq libraries from a *DCP2* allele, *dcp2-N245,* which lacks the sequence encoding the Dcp2 C-terminal domain, and from strains that harbor point mutations in the catalytic domain of this allele in conjunction with the C-terminal truncation. To understand the regulatory roles of the decapping activators, we also prepared and analyzed RNA-Seq libraries from strains that contain single deletions of the *LSM1*, *PAT1,* and *DHH1* genes. Our transcriptome-wide results indicate that: 1) the Dcp2 C-terminal extension is crucial for efficient regulation of decapping, 2) different decapping activators are responsible for targeting different sets of mRNAs, and 3) different decapping activators might compete for binding to the limited pool of cellular decapping enzyme. Further, our gene ontology analyses indicate that decapping activators are responsible for regulating specific biological processes, thus underscoring the biological importance of decapping regulation.

INTRODUCTION

Gene expression in eukaryotes is a tightly regulated process. From transcription to protein folding, every step of this pathway is subjected to scrutiny and modulation by numerous factors. One such step, mRNA decay, can be an effective means to regulate the extent of gene expression or to respond to rapid changes in the cellular environment (Holmes et al., 2004). Eukaryotic mRNAs are protected from nonspecific decay by two end-specific complexes: 1) the 5'-cap and its associated binding proteins and 2) the poly(A) tail and its binding partner, the poly(A)-binding protein (Mangus et al., 2003; Roy and Jacobson, 2013; Shoemaker and Green, 2012). Accordingly, transcripts targeted for decay by the predominant 5' to 3' or 3' to 5' exonucleolytic pathways are usually deadenylated by the Pan2-3 and Ccr4-Not complexes (Wahle and Winkler, 2013), and/or decapped by the Dcp1-Dcp2 complex (LaGrandeur and Parker, 1998; Steiger et al., 2003; Wang et al., 2002) prior to decay.

Dcp2 is the catalytic subunit of the decapping complex and this protein includes BoxA, Nudix, and BoxB motifs in an N-terminal domain occupying the first ~250 amino acids (Mugridge et al., 2016). The Nudix motif is responsible for catalysis of the decapping reaction (Dunckley and Parker, 1999, 2001) and BoxB is involved in RNA binding (She et al., 2006). While these Dcp2 N-terminal domains are well conserved, the C-terminal extensions beyond them show wide variability, from as short as ~170 amino acids in humans to ~700 amino acids in budding yeast (Wang et al., 2002). Dcp1 is a small protein that interacts with the BoxA motif of Dcp2 and contains a conserved EVH1 domain thought to be involved in modulating the binding of additional factors (She et al., 2004). Because the purified N-terminal fragment of Dcp2 is catalytically active and its activity is

greatly enhanced by Dcp1 (She et al., 2008; Steiger et al., 2003), most biochemical and structural studies have focused on truncated forms of Dcp2 (Arribas-Layton et al., 2013; Grudzien-Nogalska and Kiledjian, 2017). However, there is considerable evidence suggesting that decapping activators such as Pat1, Edc3, and the nonsense-mediated mRNA decay (NMD) regulator, Upf1, associate with Dcp2's C-terminal region to regulate substrate recognition and decapping *in vivo* (He and Jacobson, 2015a).

An understanding of the precise roles of the decapping activators has remained elusive. While the mechanism of decapping activation of Dcp2 by decapping regulators is unclear, it has been suggested that binding of activators stabilizes the active conformation of the complex (Valkov et al., 2016). Indeed, the Dcp1-Dcp2 complex undergoes significant conformational changes from active to inactive states (Mugridge et al., 2016; She et al., 2008). At least in yeast, where a Dcp2 negative regulatory element is thought to prevent indiscriminate decapping, this inhibitory role is suppressed by the binding of decapping activators (He and Jacobson, 2015a). Decapping activators are also thought to provide substrate specificity to the decapping complex and deletion of their binding sites results in loss of substrate specificity (He and Jacobson, 2015a).

One extensively studied decapping activator, the Lsm1-7 complex , along with its partner Pat1, which interacts with Lsm2 and Lsm3 (Sharif and Conti, 2013), is responsible for targeting mRNAs that have undergone significant poly(A) shortening (Nissan et al., 2010; Tharun and Parker, 2001). Deletion of components of this complex results in the accumulation of deadenylated but capped decay intermediates (Tharun et al., 2000). Without Pat1, the Lsm1-7 complex loses its preference for mRNAs with shorter poly(A) tails (Chowdhury et al., 2014). The Lsm1-7 complex does not associate with the

decapping complex directly. That role is reserved for Pat1, which, in yeast, has multiple binding domains on the Dcp2 C-terminal extension, some of which overlap with Upf1 binding sites (He and Jacobson, 2015a). Pat1 also associates with the 5' to 3' exonuclease Xrn1 (Nissan et al., 2010), an interaction also thought to promote mRNA decay.

Another well studied decapping regulator, Dhh1, is a conserved DEAD-box helicase (Presnyak and Coller, 2013). Recent evidence suggests that Dhh1 targets slowly translating mRNAs, possibly due to poor codon optimality (Radhakrishnan et al., 2016; Sweet et al., 2012). Tethering Dhh1 can result in even slower translation and subsequent decay in yeast (Sweet et al., 2012). Adding further complexity to the network of decapping regulators, Dhh1 can interact with Pat1 and Edc3, but their overlapping binding sites indicate that this interaction cannot happen concurrently (Tritschler et al., 2009a). While Dhh1 can associate with RNA effectively this interaction is inhibited upon Pat1 or Edc3 binding (Sharif et al., 2013). These observations led to the hypothesis that there can be multiple decapping complexes in the cytoplasm, with different activators and different substrate specificities (He and Jacobson, 2015a; Jonas and Izaurralde, 2013).

Cells expressing *dcp2-N245,* an allele encoding a large C-terminal truncation of Dcp2, have been shown to lose decapping substrate specificity, i.e., they appear to catalyze decapping indiscriminately (He and Jacobson, 2015a). To better understand the mechanism of decapping regulation, we generated and characterized RNA-Seq libraries from *dcp2-N245* cells, as well as from cells expressing two other *dcp2* alleles that encode single point mutations in the Dcp2 catalytic domain (E153Q and E198Q) in addition to the C-terminal deletion. Further, to understand the contributions of different decapping

activators, we also generated RNA-Seq libraries from *lsm1*Δ, *pat1*Δ, and *dhh1*Δ strains. Differential expression analyses of these libraries emphasize the regulatory importance of the unconserved C-terminal extension of Dcp2 and reveal distinct regulation groups determined by the activity of specific decapping activators. In light of the limited abundance of Dcp1-Dcp2, different decapping activators are likely to be in competition for Dcp1-Dcp2 binding, and tasked with degradation of different mRNAs at various stages of the mRNA life cycle. Since the sets of transcripts comprising the activator-specific regulatory groups have largely non-overlapping biological functions, this regulatory division of labor appears to address specific cellular processes.

RESULTS

dcp2-N245 cells have a different gene expression profile than *dcp2E153Q-N245* and *dcp2E198Q-N245* cells

Consistent with the notion that factors interacting with the C-terminal region of Dcp2 regulate its substrate specificity, deletion of this large region results in non-specific decapping and decay of several transcripts (He and Jacobson, 2015a). To confirm this observation and test the consequences of this loss of regulation at a transcriptome-wide level, we prepared RNA-Seq libraries from cells harboring a C-terminal truncation of *DCP2* that only contains the first 245 codons of the gene (*dcp2-N245*), and from cells expressing two other *dcp2* truncation alleles that also contain point mutations altering important glutamate residues in the active site of Dcp2. E153 of Dcp2 has been shown to function as a general base during the hydrolysis reaction and E198 is important for Mg^{2+} coordination within the Nudix domain (Aglietti et al., 2013). Consequently, mutations converting these glutamates to glutamines (*dcp2E153Q-N245* and *dcp2E198Q-N245*) severely inhibit decapping reactions *in vitro* (Aglietti et al., 2013).

Libraries prepared from all three mutants and wild-type cells showed good read count distribution (Fig. 3.1A) and notable consistency between biological replicates, with Pearson correlation coefficients ranging from 0.96 to 0.99 (Fig. 3.2 A). Utilizing previously described data analysis pipelines for transcript quantitation and assessment of differential expression (Celik et al., 2017), we identified 616 upregulated and 1025 downregulated transcripts in *dcp2-N245* cells, relative to WT cells (Fig. 3.1. B-D). A larger number of downregulated transcripts was an expected outcome since the C-terminal domain of yeast Dcp2 also contains a negative regulatory element that is responsible for hindering

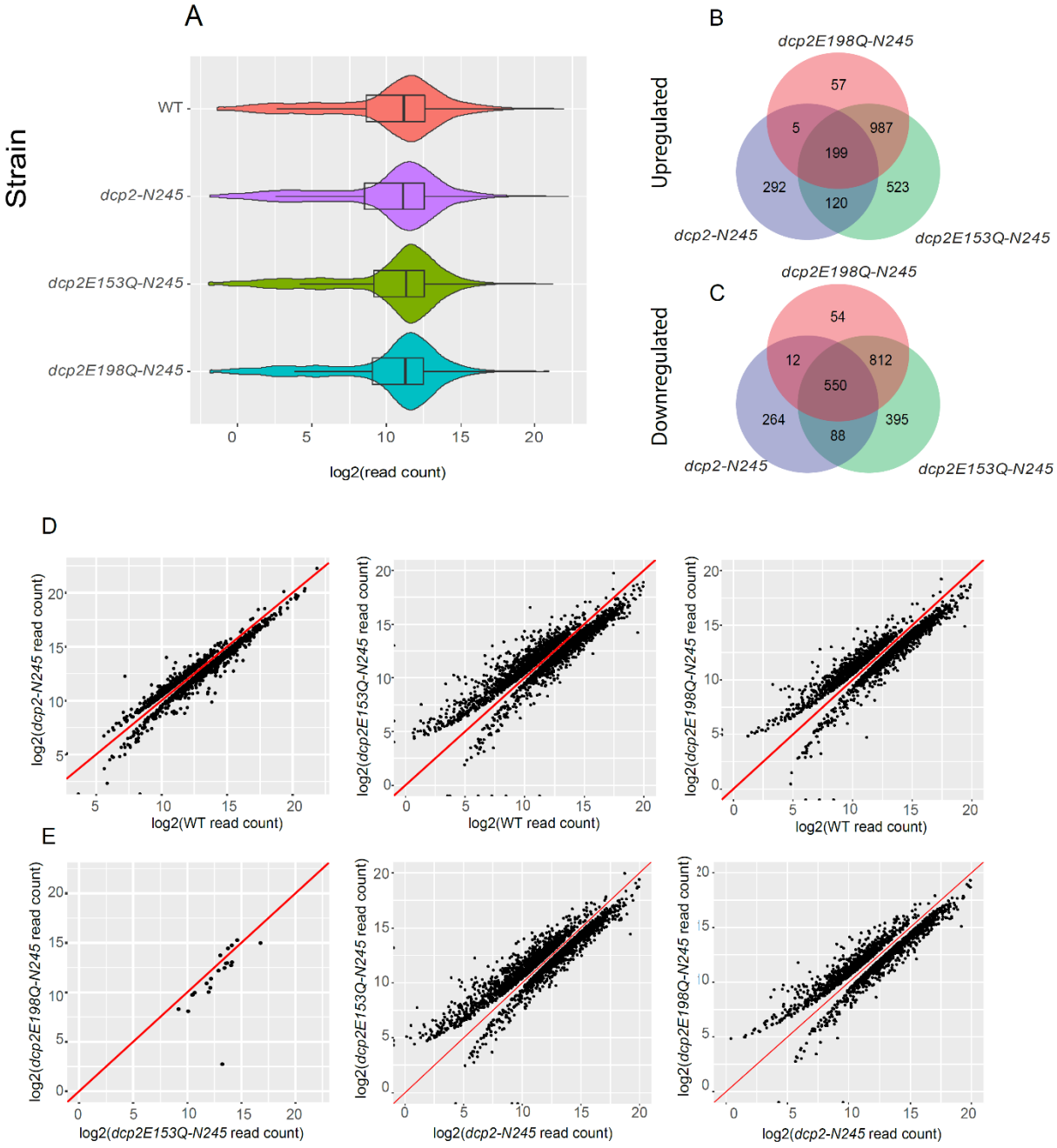

A
Strain
WT
dcp2-N245
dcp2E153Q-N245
dcp2E198Q-N245
0
5
10
15
20
log2(read count)
B
Upregulated
dcp2E198Q-N245
57
5
987
199
292
523
120
dcp2-N245
dcp2E153Q-N245
C
Downregulated
dcp2E198Q-N245
54
12
812
550
264
395
88
dcp2-N245
dcp2E153Q-N245
D
log2(dcp2-N245 read count)
log2(WT read count)
log2(dcp2E153Q-N245 read count)
log2(WT read count)
log2(dcp2E198Q-N245 read count)
log2(WT read count)
E
log2(dcp2E198Q-N245 read count)
log2(dcp2E153Q-N245 read count)
log2(dcp2E153Q-N245 read count)
log2(dcp2-N245 read count)
log2(dcp2E198Q-N245 read count)
log2(dcp2-N245 read count)

Figure 3.1. Cells harboring catalytically inactive Dcp2 alleles *dcp2E153Q-N245* and *dcp2E198Q-N245* show similar expression profiles which are significantly different from that obtained from the indiscriminate *dcp2-N245* allele.
A. RNA-Seq libraries from WT, *dcp2-N245*, *dcp2E153Q-N245,* and *dcp2E198Q-N245* strains show comparable read count distributions. Violin and box plots used to visualize the average and median read count distributions of each strain from three independent biological replicates. **B.** and **C.** Transcripts up- and downregulated in *dcp2E153Q-N245* and *dcp2E198Q-N245* strains show significant overlap which only partially overlaps with the *dcp2-N245* strain. Transcripts were identified as differentially expressed by comparing read counts to the WT strain. Venn diagrams were used to display the relationships between different datasets. Transcripts that are upregulated are presented in (B) and those downregulated are presented in (C) **D.** Differentially expressed transcripts in each single deletion strain show a wide range of expression levels. Scatterplots were used to display the log2 read count numbers of each transcript in WT, *dcp2-N245*, *dcp2E153Q-N245,* and *dcp2E198Q-N245* strains. A larger number of transcripts showed upregulation in each strain. The y=x line is shown in red. Compared to WT cells, the *dcp2-N245* strain manifested 616 up and 1025 downregulated transcripts, the *dcp2E153Q-N245* strain had 1921 up and 1845 downregulated transcripts, and the *dcp2E198Q-N245* strain had 1346 up and 1428 downregulated transcripts. **E.** *dcp2E153Q-N245* and *dcp2E198Q-N245* strains show virtually identical expression levels for almost all transcripts (leftmost panel). Differential expression analysis revealed only 21 transcripts as differentially expressed between *dcp2E153Q-N245* and *dcp2E198Q-N245* backgrounds, whereas there were 1658 upregulated transcripts and 1690 downregulated transcripts in *dcp2E153Q-N245* with respect to *dcp2-N245* strain, and 1113 upregulated transcripts and 1090 downregulated transcripts in *dcp2E198Q-N245* strain with respect to the *dcp2-N245* strain.

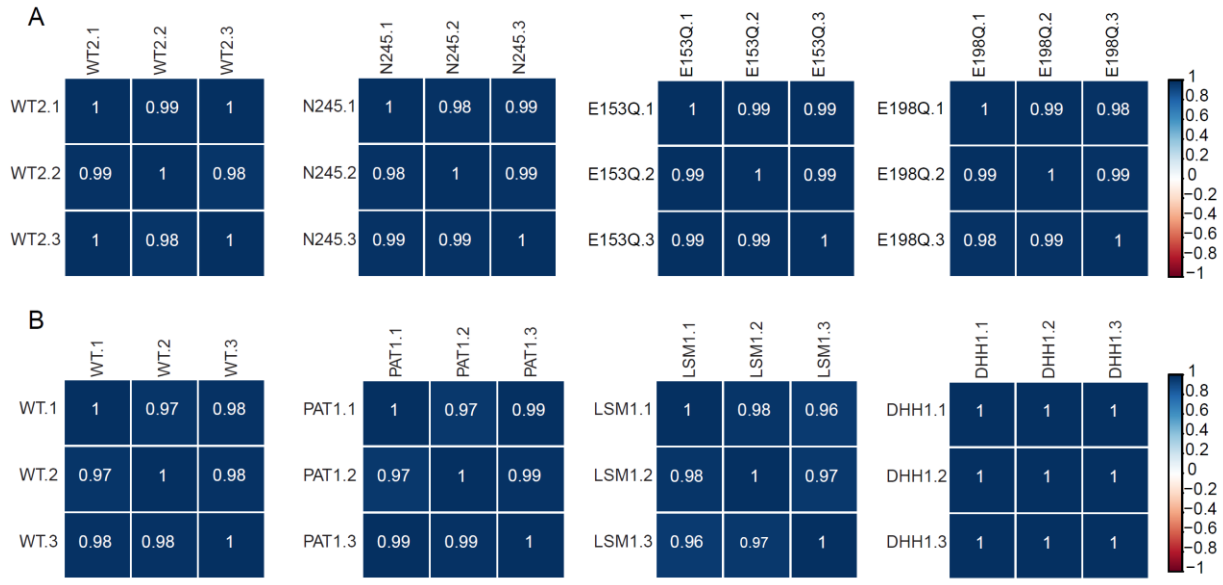


Figure 3.2 RNA-Seq libraries generated from WT, *lsm1*Δ, *pat1*Δ, *dhh1*Δ, *dcp2-N245, dcp2E153Q-N245,* and *dcp2E198Q-N245* strains show good correlation between biological replicates

RNA-Seq libraries show very high Pearson correlation coefficients among biological replicates. WT.1, WT.2, and WT.3 are WT controls for *lsm1*Δ, *pat1*Δ, and *dhh1*Δ strains and WT2.1, WT2.2, and WT2.3 are WT controls for the *dcp2-N245, dcp2E153Q-N245,* and *dcp2E198Q-N245* strains.

Dcp2 activity until it is negated by Dcp2 interaction with one of the decapping activators. This is thought to maintain the spatial and temporal regulation of decapping (He and Jacobson, 2015a). The loss of this regulation can, therefore, create an indiscriminate and hyperactive Dcp2 capable of removing 5' caps from mRNAs without regard to regulation, resulting in the downregulation of many transcripts. Additionally, we observed 1921 upregulated and 1845 downregulated transcripts in cells expressing the *dcp2E153Q-N245* allele, and 1346 upregulated and 1428 downregulated transcripts in cells expressing the *dcp2E198Q-N245* allele (Fig. 3.1. B-D). A closer examination of upregulated and downregulated transcripts revealed a significant overlap between *dcp2E153Q-N245* and *dcp2E198Q-N245* strains with 1186 and 1362 commonly upregulated and downregulated transcripts, respectively (Fig. 3.1 B,C). However, these sets only partially overlapped with transcripts that are differentially expressed in the *dcp2-N245* background, with 199 and 550 upregulated and downregulated transcripts, respectively, being common to all three backgrounds (Fig. 3.1. B, C).

Considering that both the E153 and E198 residues in Dcp2 are located in the Nudix domain, the significant overlap between these datasets was not surprising. Nonetheless, we wanted to test directly whether *dcp2E153Q-N245* cells and *dcp2E198Q-N245* cells show the same expression patterns, with the different numbers of mRNAs identified due to replicate variability. When we applied the same differential expression pipeline to the *dcp2E153Q-N245 and dcp2E198Q-N245* libraries we identified only 21 transcripts as differentially expressed, 4 of which showed higher expression values in *dcp2E198Q-N245* (Fig. 3.1. E leftmost panel). Therefore, we conclude that these two active site mutations have the same effect on the transcriptome. Interestingly, a similar comparison with the

dcp2-N245 strain showed 1658 upregulated and 1690 downregulated transcripts with respect to *dcp2E153Q-N245,* and 1113 upregulated 1090 downregulated with respect to the *dcp2E198Q-N245* strain (Fig. 3.1. E middle and right panels), indicating that *dcp2-N245* strains show considerable differences in expression profiles compared to those expressing *dcp2E153Q-N245* or *dcp2E198Q-N245*.

The Dcp2 C-terminal domain is important for decapping regulation

To further assess the concept that the *dcp2-N245* allele lacks regulation, but is not defective in decapping, we compared expression levels of all transcripts in this strain to those we previously published for *dcp1*Δ*, dcp2*Δ*,* and *xrn1*Δ strains (Celik et al., 2017) in addition to those assessed in *dcp2E153Q-N245* and *dcp2E198Q-N245* cells. Because the *dcp1*Δ*, dcp2*Δ*, xrn1*Δ*,* and *dcp2* C-terminal truncation libraries were prepared at different times with different WT controls we chose to improve consistency by comparing fold changes of each transcript with respect to the appropriate WT control. As expected, results from *dcp1*Δ and *dcp2*Δ strains, or from *dcp2E153Q-N245* and *dcp2E198Q-N245* strains showed excellent correlation (Pearson correlation coefficient = 0.868 and 0.932, respectively) (Fig 3.3). Consistently, *dcp1*Δ and *dcp2*Δ strains were also well correlated with those from *xrn1*Δ, *dcp2E153Q-N245,* and *dcp2E198Q-N245* cells (Pearson correlation coefficients = 0.748, 0.812, 0.805 for *dcp1*Δ and 0.734, 0.772, 0.789 for *dcp2*Δ strains, respectively) (Fig. 3.3). In line with the notion that the *dcp2-N245* allele lacks regulation and targets transcripts at random the changes in transcript expression levels showed poor correlations of the *dcp2-N245* library with those from *dcp1*Δ, *dcp2*Δ, *xrn1*Δ, *dcp2E153Q-N245,* or *dcp2E198Q-N245* strains (Pearson correlation coefficients = 0.371,

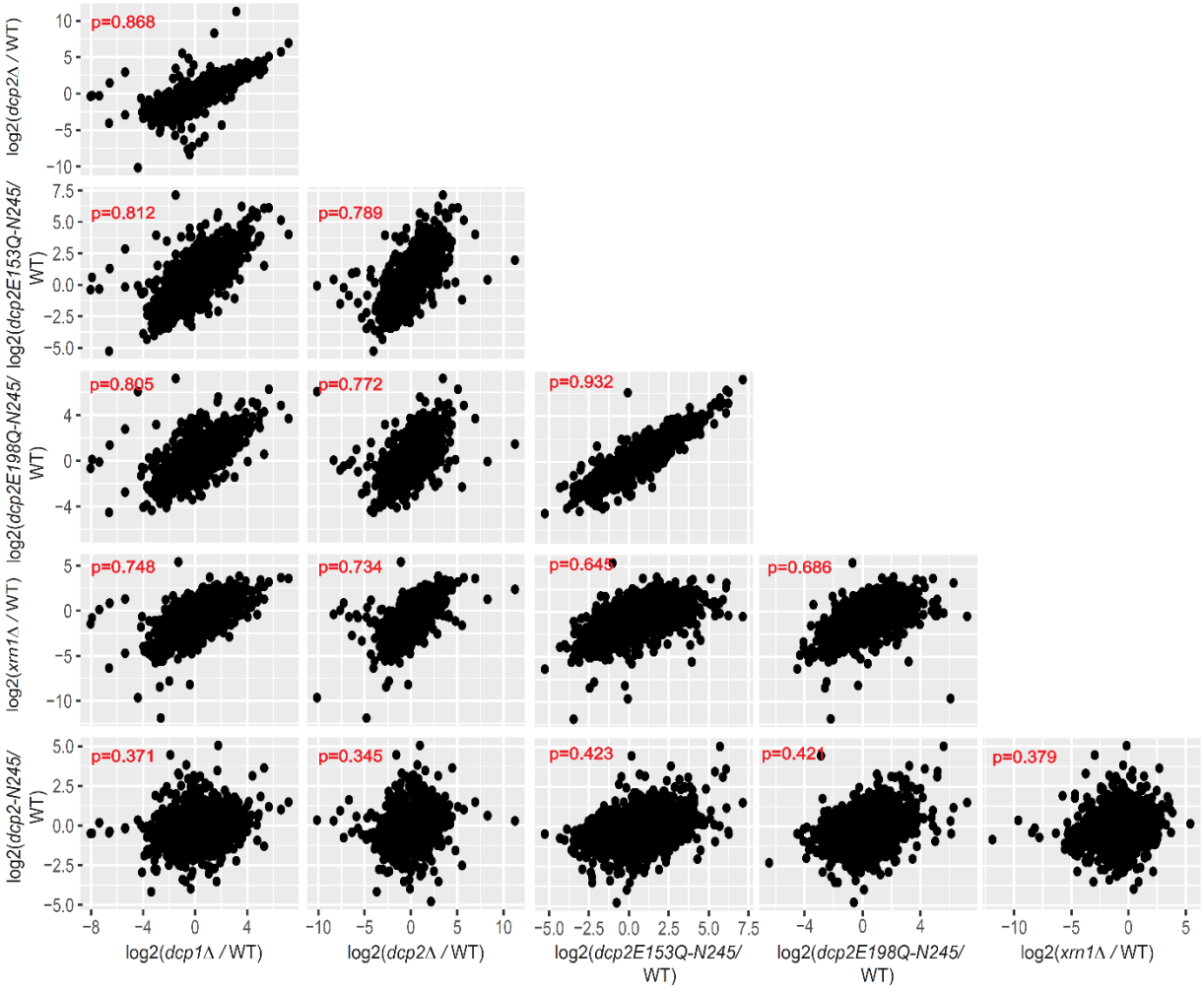


Figure 3.3. Deletion of the C-terminal extension of Dcp2 results in decapping deregulation, but not decapping or decay inactivation

Scatterplot matrix of log2 fold changes between previously described *dcp1*Δ, *dcp2*Δ, and *xrn1*Δ libraries (Celik et al., 2017) and *dcp2-N245, dcp2E153Q-N245, dcp2E198Q-N245* libraries compared to their respective WT controls. Pearson correlation coefficients of log2 fold changes for each scatterplot comparison are shown in red. Log2 fold changes compared to WT in *dcp1*Δ and *dcp2*Δ strains, and *dcp2E153Q-N245* and *dcp2E198Q-N245* strains show excellent correlation. They are also well correlated with each other and with log2 fold changes in an *xrn1*Δ strain. The *dcp2-N245* strain shows poor correlation with all five decapping and decay mutants.

0.345, 0.379, 0.423 and 0.421 with *dcp1*Δ, *dcp2*Δ, *xrn1*Δ *dcp2E153Q-N245* and *dcp2E198Q-N245* libraries, respectively) (Fig. 3.3).

Lsm1 and Pat1 regulate the same set of transcripts, which partially overlaps with the set of regulated by Dhh1

Having established the importance of the C-terminal extension of Dcp2 for decapping regulation we wanted to focus on individual factors that regulate decapping activity and better understand the mechanism of substrate selection by some of the factors that associate with the decapping complex. Therefore, we generated RNA-Seq libraries, in three biological replicates, from yeast strains harboring single deletions of the *LSM1, PAT1,* or *DHH1* genes. Our libraries again showed good read count distribution (Fig. 3.4 A) and notable consistency between biological replicates (Fig. 3.2 B). Using the same analysis pipeline employed in the experiments of Fig 3.1, we identified 955 upregulated and 681 downregulated transcripts in *lsm1*Δ cells, 940 upregulated and 685 downregulated transcripts in *pat1*Δ cells, and 1098 upregulated and 788 downregulated transcripts in *dhh1*Δ cells (Fig. 3.4 B-D). A comparison of the lists of transcripts differentially expressed in each strain indicated a striking overlap between the results obtained in cells with *lsm1*Δ or *pat1*Δ backgrounds (864 common upregulated transcripts and 583 and common downregulated transcripts). However, these sets of transcripts only overlapped partially with the transcripts differentially expressed in the *dhh1*Δ strain (482 common upregulated transcripts and 290 common downregulated transcripts in all three strains) (Fig. 3.4 B, C).

Given the substantial overlap of differentially expressed transcripts between *lsm1*Δ and *pat1*Δ strains, and the interactions between Pat1 and the Lsm1-7 complex (Wu et al.,

2014), we tested whether Lsm1 and Pat1 did indeed control the expression of the same set of transcripts. Again, utilizing the same differential expression analysis pipeline, comparisons between *lsm1*Δ and *pat1*Δ cells manifested overall expression levels of all transcripts that were remarkably consistent (Fig. 3.4 E, leftmost panel), with only four transcripts differentially expressed between the two strains, two of which were accounted for by the respective deletions (Fig. 3.1 E, leftmost panel, red dots). We therefore concluded that Lsm1 and Pat1 target the same set of transcripts for decapping.

When we compared the expression levels of transcripts in *dhh1*Δ and *lsm1*Δ or *pat1*Δ strains we identified 1385 upregulated transcripts and 1037 downregulated transcripts with respect to the *lsm1*Δ strain, and 1332 upregulated transcripts and 874 downregulated transcripts with respect to the *pat1*Δ strain (Fig.1E, last two panels).

Transcripts regulated by Lsm1, Pat1, and Dhh1 are different from NMD substrates

In a recent study (Celik et al., 2017), we identified the substrates of the yeast nonsense-mediated mRNA decay (NMD) pathway. Upf1, the main regulator of NMD, has been shown to interact with Dcp2 to target NMD substrates for decapping, and one of its interaction domains with Dcp2 overlaps with a Pat1-interacting domain (He and Jacobson, 2015a). Our identification of discrete Dcp2 domains that interact with decapping activators, along with other observations (Jonas and Izaurralde, 2013), led to the hypothesis that there can be distinct decapping complexes in the cell with different substrate specificities (He and Jacobson, 2015a). To test aspects of this hypothesis, we compared the list of transcripts that were differentially expressed in all single *upf* deletions (Celik et al., 2017) to the sets of differentially expressed transcripts described in Fig. 3.1.

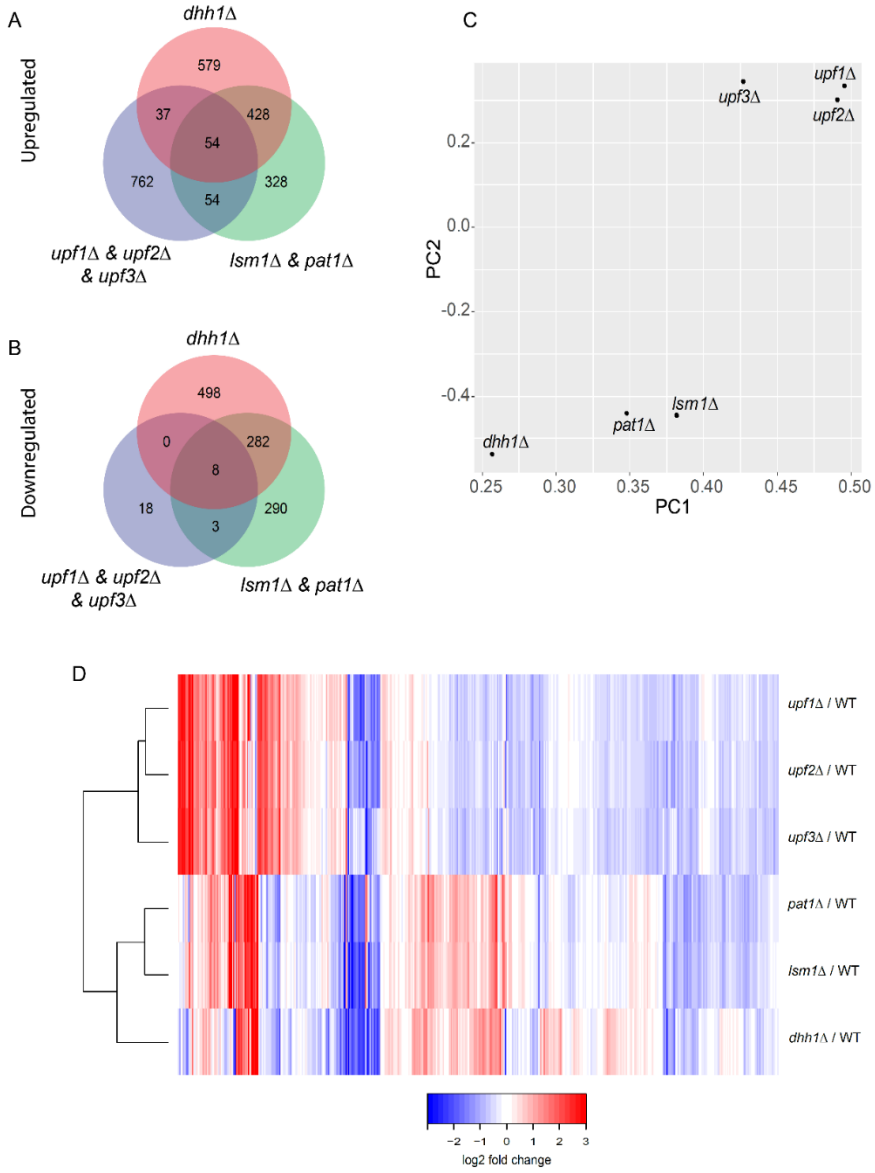


Figure 3.5. Transcripts regulated by Lsm1, Pat1, and Dhh1 are different from NMD substrates

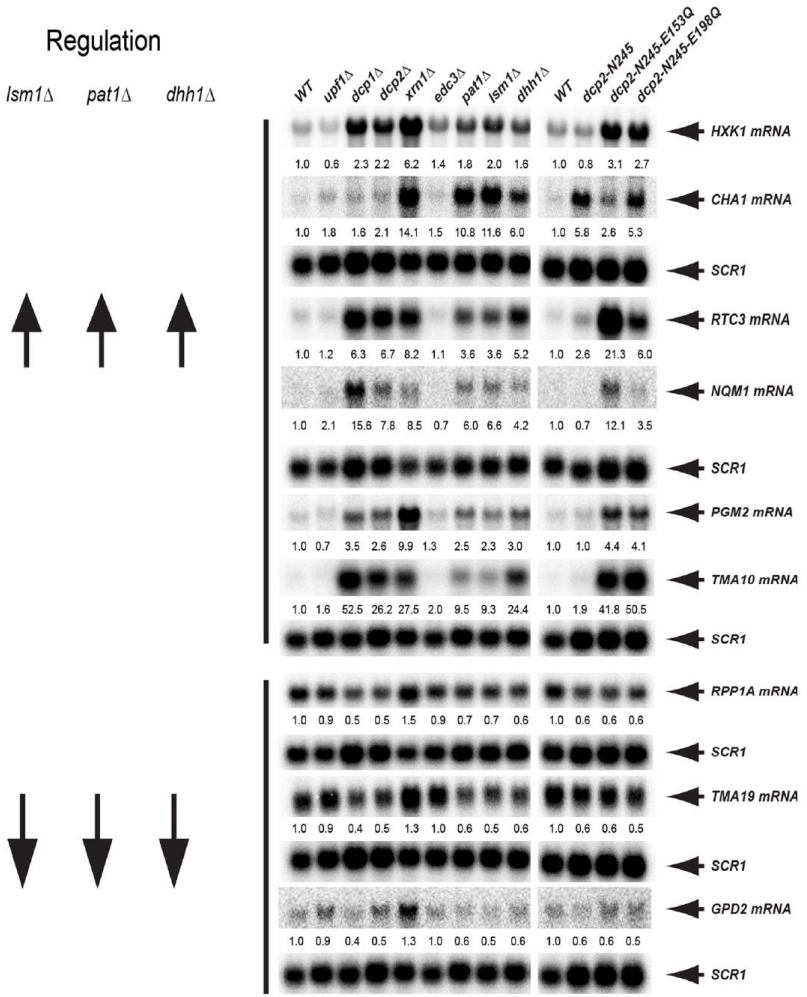
A. and **B.** Overlaps of upregulated (A) and downregulated (B) transcripts in *dhh1*Δ, *lsm1*Δ and *pat1*Δ strains and previously identified NMD substrates (Celik et al., 2017) using the same analysis pipeline. NMD substrates show minimal overlap with differentially expressed transcripts in *dhh1*Δ, *lsm1*Δ, and *pat1*Δ backgrounds. **C.** First two principal components of principal component using log2 fold changes for transcripts compared to WT reveal three distinct groups of expression levels: single upf deletions, *lsm1*Δ, *pat1*Δ, and *dhh1*Δ. **D.** Heatmap of log2 fold changes in *upf1*Δ, *upf2*Δ, *upf3*Δ, *pat1*Δ, *lsm1*Δ, and *dhh1*Δ strains. Red indicates upregulation and blue indicates downregulation.

We observed minimal overlap between NMD substrates and transcripts upregulated or downregulated in *lsm1*Δ*, pat1*Δ*,* or *dhh1*Δ strains (Fig. 3.5 A, B). When we compared the overall expression levels of all transcripts by principal component analysis or hierarchical clustering of log2 fold changes of all transcripts compared to their WT control we observed three distinct groups of transcript expression levels (Fig. 3.5 C, D): 1) *dhh1*Δ alone, 2) *lsm1*Δ and *pat1*Δ*,* and 3) single *upf* deletions. These observations support our hypothesis that different proteins that interact with the decapping complex regulate the expression of different mRNAs (He and Jacobson, 2015a).

Validation of bioinformatic analyses

We validated our bioinformatic analyses by using northern blotting to assess the levels of distinct mRNAs that were representative of transcripts identified in the RNA-Seq experiments. Among the selected transcripts, six (*HXK1, CHA1, RTC3, NQM1, PGM2*, and *TMA2)* were upregulated and three (*RPP1, TMA19, GPD2*) were downregulated in RNA-Seq libraries of all three strains, three (*CIT2, SDS23,* and *HOS2*) were upregulated only in *dhh1*Δ cells, and three (*DIF1, AGA1,* and *BUR7*) were upregulated and one (*GTT2*) was downregulated only in *lsm1*Δ and *pat1*Δ backgrounds. We also compared expression levels in *dcp2E153Q-N245* and *dcp2E198Q-N245* strains for all these genes. In all cases, we observed similar expression level changes in both the northern and bioinformatics analyses (Fig. 3.6). As additional controls, we included *upf1*Δ and *edc3*Δ in our northern blot analyses.

Different decapping complexes are responsible for regulating genes with different biological functions


Regulation
lsm1Δ pat1Δ dhh1Δ
WT upf1Δ dcp1Δ dcp2Δ xrn1Δ edc3Δ pat1Δ lsm1Δ dhh1Δ WT dcp2-N245 dcp2-N245-E153Q dcp2-N245-E198Q
HXK1 mRNA
1.0 0.6 2.3 2.2 6.2 1.4 1.8 2.0 1.6 1.0 0.8 3.1 2.7
CHA1 mRNA
1.0 1.8 1.6 2.1 14.1 1.5 10.8 11.6 6.0 1.0 5.8 2.6 5.3
SCR1
RTC3 mRNA
1.0 1.2 6.3 6.7 8.2 1.1 3.6 3.6 5.2 1.0 2.6 21.3 6.0
NQM1 mRNA
1.0 2.1 15.6 7.8 8.5 0.7 6.0 6.6 4.2 1.0 0.7 12.1 3.5
SCR1
PGM2 mRNA
1.0 0.7 3.5 2.6 9.9 1.3 2.5 2.3 3.0 1.0 1.0 4.4 4.1
TMA10 mRNA
1.0 1.6 52.5 26.2 27.5 2.0 9.5 9.3 24.4 1.0 1.9 41.8 50.5
SCR1
RPP1A mRNA
1.0 0.9 0.5 0.5 1.5 0.9 0.7 0.7 0.6 1.0 0.6 0.6 0.6
SCR1
TMA19 mRNA
1.0 0.9 0.4 0.5 1.3 1.0 0.6 0.5 0.6 1.0 0.6 0.6 0.5
SCR1
GPD2 mRNA
1.0 0.9 0.4 0.5 1.3 1.0 0.6 0.5 0.6 1.0 0.6 0.6 0.5
SCR1

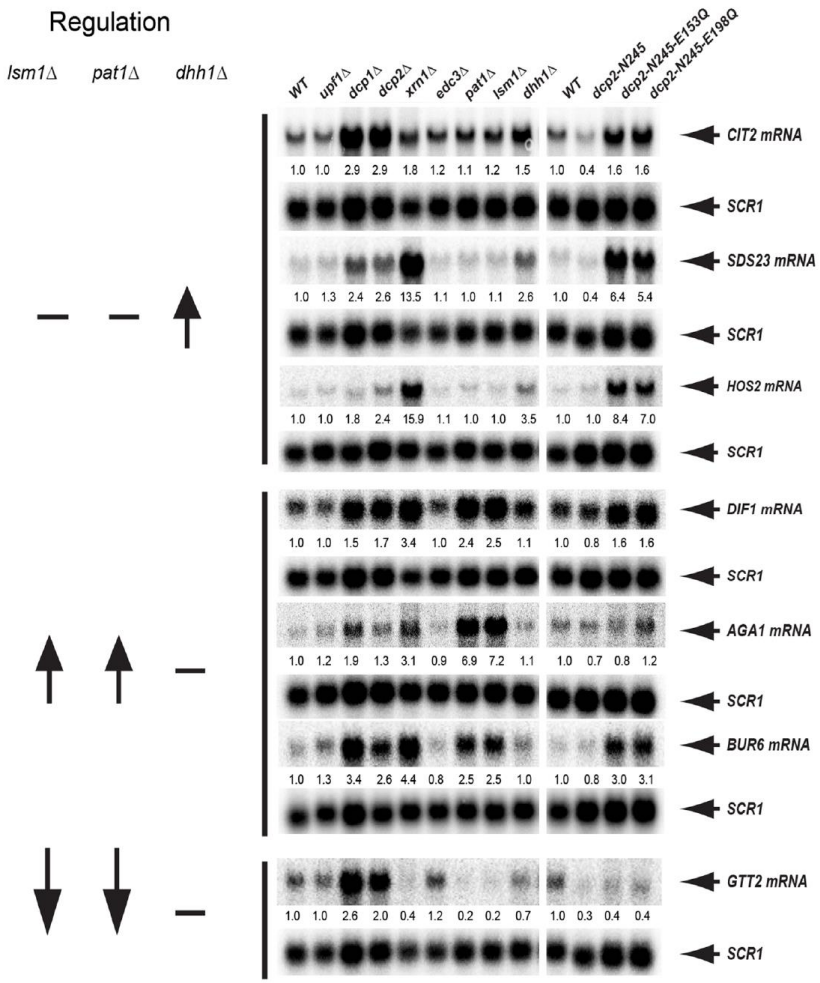

Regulation
lsm1Δ pat1Δ dhh1Δ
WT upf1Δ dcp1Δ dcp2Δ xrn1Δ edc3Δ pat1Δ lsm1Δ dhh1Δ WT dcp2-N245 dcp2-N245-E153Q dcp2-N245-E198Q
CIT2 mRNA
1.0 1.0 2.9 2.9 1.8 1.2 1.1 1.2 1.5 1.0 0.4 1.6 1.6
SCR1
SDS23 mRNA
1.0 1.3 2.4 2.6 13.5 1.1 1.0 1.1 2.6 1.0 0.4 6.4 5.4
SCR1
HOS2 mRNA
1.0 1.0 1.8 2.4 15.9 1.1 1.0 1.0 3.5 1.0 1.0 8.4 7.0
SCR1
DIF1 mRNA
1.0 1.0 1.5 1.7 3.4 1.0 2.4 2.5 1.1 1.0 0.8 1.6 1.6
SCR1
AGA1 mRNA
1.0 1.2 1.9 1.3 3.1 0.9 6.9 7.2 1.1 1.0 0.7 0.8 1.2
SCR1
BUR6 mRNA
1.0 1.3 3.4 2.6 4.4 0.8 2.5 2.5 1.0 1.0 0.8 3.0 3.1
SCR1
GTT2 mRNA
1.0 1.0 2.6 2.0 0.4 1.2 0.2 0.2 0.7 1.0 0.3 0.4 0.4
SCR1

Figure 3.6. Validation of bioinformatic analyses by northern blotting
Northern blotting analyses of **A.** Transcripts upregulated in all *lsm1*Δ, *pat1*Δ and *dhh1*Δ strains (*HXK1, CHA1, RTC3, NQM1, PGM2, TMA10*), **B.** Transcripts downregulated in all *lsm1*Δ, *pat1*Δ and *dhh1*Δ strains (*RPP1, TMA19, GPD2*), **C.** Transcripts upregulated only in *dhh1*Δ background (*CIT2*, *SDS23, HOS2*, Transcripts **D.** upregulated (*DIF1, AGA1*) and **E.** downregulated (*GTT2*) only in *lsm1*Δ and *pat1*Δ strains. Total RNA was isolated from the indicated strains and the steady-state levels of individual transcripts in these strains were analyzed by northern blotting. In each case, a random-primed probe was hybridized to the blot and *SCR1* served as the loading control. Normalized fold changes after *SCR1* loading control correction compared to WT are presented under each blot.

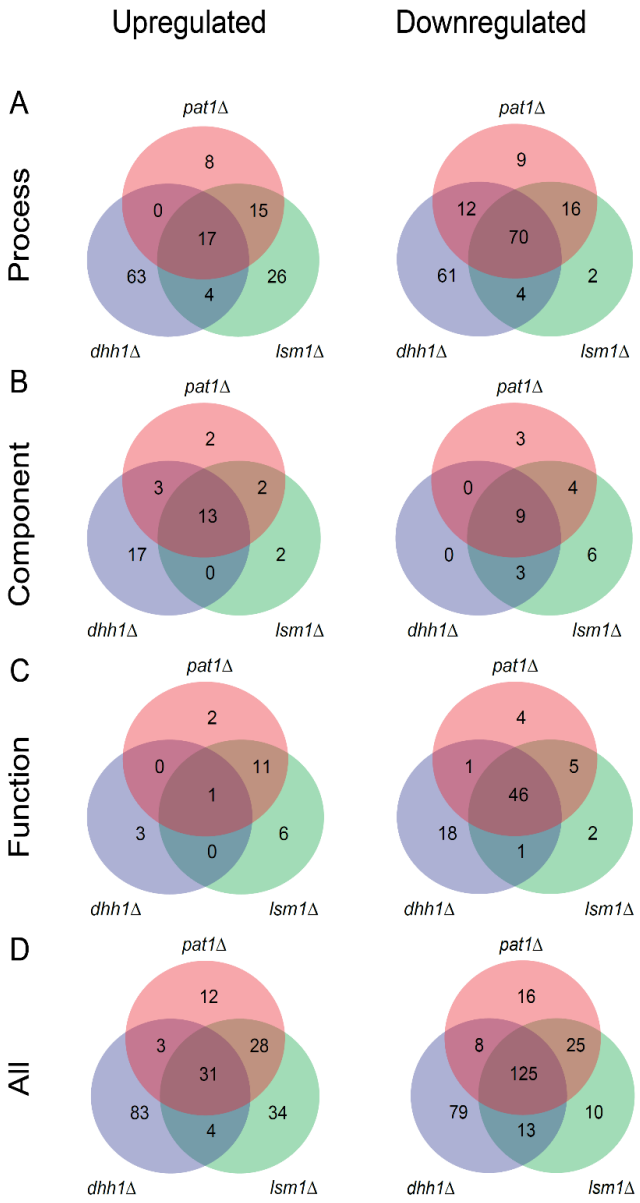


Figure 3.7. Different decapping complexes regulate genes with different biological functions

Gene ontology (GO) term enrichment analysis of biological process, **A.** cellular component, **B.** function, **C.** and all three combined **D.** enriched GO terms manifested a similar pattern to transcript expression. Upregulated and downregulated transcripts in *lsm1*Δ and *pat1*Δ strains showed substantial overlap that only partially overlaps with GO terms enriched in *dhh1*Δ strain.

To determine whether the mRNAs regulated by Lsm1, Pat1, or Dhh1 have overlapping or non-overlapping biological functions we used the GOrilla web interface (Eden et al., 2009) to calculate gene ontology (GO) term enrichment for all the regulation groups. Upregulated and downregulated transcripts in *pat1*Δ and *lsm1*Δ showed significant overlap (Fig. 3.7 A-D), with only partial overlap with transcripts differentially expressed in *dhh1*Δ, a pattern closely resembling the differential expression analysis of Fig 3.4. These results suggest that transcripts targeted by different decapping complexes also have different biological functions in the cell.

Upf proteins in competition for Dcp2 with Dhh1, but not with Lsm1 and Pat1

We determined whether transcripts upregulated upon single *upf* deletions were down regulated in *dhh1*Δ*, lsm1*Δ, or *pat1*Δ strains. Because both Dhh1 and Upf1 have been implicated in co-translational targeting of their substrates (Hu et al., 2010; Hu et al., 2009; Radhakrishnan et al., 2016) we first asked whether transcripts downregulated exclusively in *dhh1*Δ cells might contain NMD substrates. We observed a significant enrichment for NMD substrates (Fisher's exact test, $p=7.3 \times 10^{-15}$) among transcripts downregulated only in *dhh1*Δ cells (Fig. 3.8 A). Surprisingly, transcripts downregulated only in *lsm1*Δ and *pat1*Δ cells were depleted of NMD substrates (Fisher's exact test, $p=6.2 \times 10^{-6}$) (Fig. 3.8 C) and transcripts downregulated in all three decapping activator-deficient strains showed no enrichment for NMD substrates (Fisher's exact test p=0.12) (Fig. 3.8 B). A similar comparison of transcripts exclusively upregulated in *dhh1*Δ strains with transcripts downregulated only in *lsm1*Δ and *pat1*Δ cells showed no enrichment (Fisher's exact test p=0.3079) (Fig.3.8 D). Yet, the inverse comparison, i.e. transcripts upregulated only in *lsm1*Δ and *pat1*Δ strains versus transcripts downregulated only in *dhh1*Δ cells

showed a subtle depletion compared to what would be expected by chance (Fisher's exact test p=0.00117) (Fig. 3.8 E). These results suggest the that different decapping complexes can target mRNAs for decay at different stages in the mRNA life cycle and decapping complexes that target during similar processes can compete for the Dcp1-Dcp2 complex.

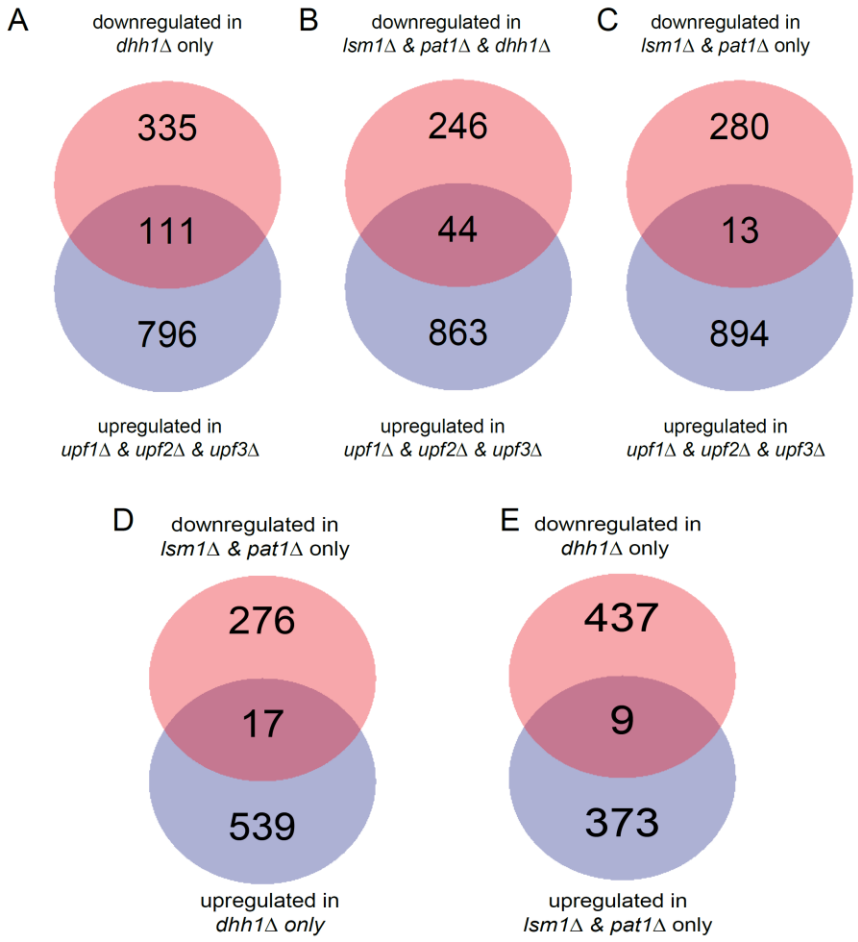


Figure 3.8. Upf proteins are in competition for Dcp2 with Dhh1, but not with Lsm1 and Pat1

A. NMD substrates were represented at a higher ratio than would be expected than by chance among transcripts that are downregulated in *dhh1*Δ **B.** There was no enrichment for NMD substrates among transcripts that are regulated by all three decapping regulators **C.** NMD substrates were depleted among transcripts regulated by Lsm1 and Pat1 only. **D.** Transcripts exclusively upregulated in *dhh1*Δ strains showed no enrichment among downregulated only in *lsm1*Δ and *pat1*Δ background **E.** The inverse comparison from (D) revealed a slight depletion. See text for Fisher's exact test p values.

DISCUSSION

Dcp2 C-terminal extension is crucial for efficient regulation of decapping

While the N-terminal domain of Dcp2 is well conserved, the C-terminal extension of this protein shows variability in length and sequence among different organisms (Wang et al., 2002). This segment of the protein is not essential for catalysis of decapping *in vitro* (LaGrandeur and Parker, 1998; Mugridge et al., 2016; Tharun and Parker, 1999; Valkov et al., 2016) and a significant number of biochemical and structural studies have, therefore, only focused on the conserved N-terminal domains of Dcp2. Our recent results indicate that the Dcp2 C-terminal extension is important for decapping *in vivo* (He and Jacobson, 2015a) and raise questions about experimental approaches that omit this domain. We have found that this portion of Dcp2 in yeast is the site of several binding domains for decapping activators and contains an inhibitory element that is responsible for preventing non-specific mRNA decapping. Because of these sites, deletion of this large section results in an overactive and indiscriminate decapping protein that can target mRNAs at random (He and Jacobson, 2015a). Using RNA-Seq analyses, we identified transcripts that are differentially expressed in cells harboring the *dcp2-N245* allele that yields a large deletion of the *DCP2* C-terminal regulatory domain. As controls, we also compared expression levels of *dcp2E153Q-N245* and *dcp2E198Q-N245* strains (Fig 3.1 B-D) which, in addition to expressing the same truncation also encode point mutations in the active site of Dcp2 that renders these alleles virtually inactive (Aglietti et al., 2013). Differential comparisons of the transcripts from these strains revealed that the expression profile of *dcp2-N245* is drastically different from those of *dcp2E153Q-N245* and *dcp2E198Q-N245* strains. On the other hand, log2 fold changes between the transcript

sets of *dcp1*Δ, *dcp2*Δ, *xrn1*Δ, *dcp2E153Q-N245,* or *dcp2E198Q-N245* cells manifested high correlation, supporting the idea that these alleles are defective in decapping and/or decay (Fig 3.3). The lack of correlation of expression profiles from cells harboring the *dcp2-N245* allele with those with *dcp1*Δ, *dcp2*Δ, *xrn1*Δ, *dcp2E153Q-N245,* or *dcp2E198Q-N245* mutations provides transcriptome-wide confirmation that the decapping enzyme derived from the *dcp2-N245* allele lacks substrate specificity, but not decapping activity (Fig 3.3).

Different decapping activators regulate different subsets of transcripts

We identified transcripts that are regulated by the decapping activators Lsm1, Pat1, and Dhh1. Consistent with biochemical and structural data, a substantial overlap among the transcripts upregulated or downregulated in *lsm1*Δ or *pat1*Δ strains (Fig. 3.4 B, C) provides transcriptome-wide evidence for the functional interdependency of these two proteins. In contrast, transcripts regulated by Dhh1 showed only partial overlap with Lsm1 and Pat1 decapping substrates (Fig 3.4 B, C), suggesting that this decapping regulator has functions that are distinct from those of Lsm1 or Pat1. Previously published results suggesting interactions between Dhh1 and Pat1 (Sharif et al., 2013) raise the possibility that an Lsm1-7-Pat1-Dhh1 complex (Fig. 3.9) can target a different set of mRNAs than those targeted by Lsm1-7-Pat1 or Dhh1 alone. This set of regulatory patterns among decapping regulators seems to have a biological significance as well, i.e., the enrichment of GO terms mirrors the sets of differentially expressed transcripts and does not appear to associate cellular functions and decapping substrates arbitrarily (Fig, 3.7 A-D).

Based on results from our lab and others (He and Jacobson, 2015a; He et al., 2014; Jonas and Izaurralde, 2013), we hypothesized that there can be different decapping complexes in the cell with different substrate specificities. The targets of these decapping complexes are thought to be dependent on the specific set of activators associated with Dcp2, allowing regulation of a diverse set of transcripts with a small number of regulatory proteins. Recent results from our lab also implied that some of the interactions of decapping activators with Dcp2 can be mutually exclusive (He and Jacobson, 2015a). The patterns of overlap of transcripts regulated by Dhh1, Lsm1 and Pat1, and Upf1 are in line with this model (Fig 3.5).

Competition among different decapping complexes

Our results reinforce the notion that multiple mutually exclusive decapping complexes target different transcripts in the cell. If this hypothesis is correct, then elimination of one type or related group of decapping activators might result in an increase in available Dcp1-Dcp2 for interaction with other decapping regulators. In support of this possibility, a subset of NMD substrates was enriched among transcripts that were downregulated in *dhh1*Δ cells (Fig 3.8 A). However, this was not the case for transcripts downregulated only in *lsm1*Δ and *pat1*Δ strains or in cells harboring all three deletions (Fig. 3.8 B, C). This raises the possibility of a more focused hypothesis, i.e., that mRNAs targeted for decay co-translationally are in competition for the Dcp1-2 complex. Considering the abundances of Upf1, Dhh1, Dcp2, and Dcp1 (6090, 42900, 8530, 2880 molecules per cell, respectively) (Ghaemmaghami et al., 2003; Huh et al., 2003) competition for the Dcp1-Dcp2 complex seems plausible. An alternative explanation is that the rate limiting step of the decay mechanism utilized by the Lsm1-7-Pat1 complex

is not binding to Dcp2, or substrate selection, but another downstream process, and the competition of this complex for Dcp2 will not be reflected by steady-state mRNA levels.

A general drawback of measuring steady-state mRNA levels is our inability to distinguish direct vs. indirect targets of Lsm1, Pat1, and Dhh1. Because these proteins are decay factors it is likely that a majority of upregulated transcripts represent direct targets. For the downregulated transcripts, we propose that these represent transcripts that are targeted by other decapping regulators due to increased availability of Dcp1-Dcp2 complexes. Dhh1 and the Lsm1-7-Pat1 complex are not the only decapping regulators in the cell; the Upfs, Edc1-3, Scd6, and potentially other proteins, alone or in combination, can also target transcripts for decapping (Arribas-Layton et al., 2013). Enrichment of NMD substrates among transcripts that are downregulated in *dhh1*Δ strains provides strong support for this hypothesis. The large repertoire of mutually exclusive interactions (Jonas and Izaurralde, 2013; Sharif et al., 2013; Tritschler et al., 2009a) is also in line with this mode of decapping regulation (Fig 3.9).

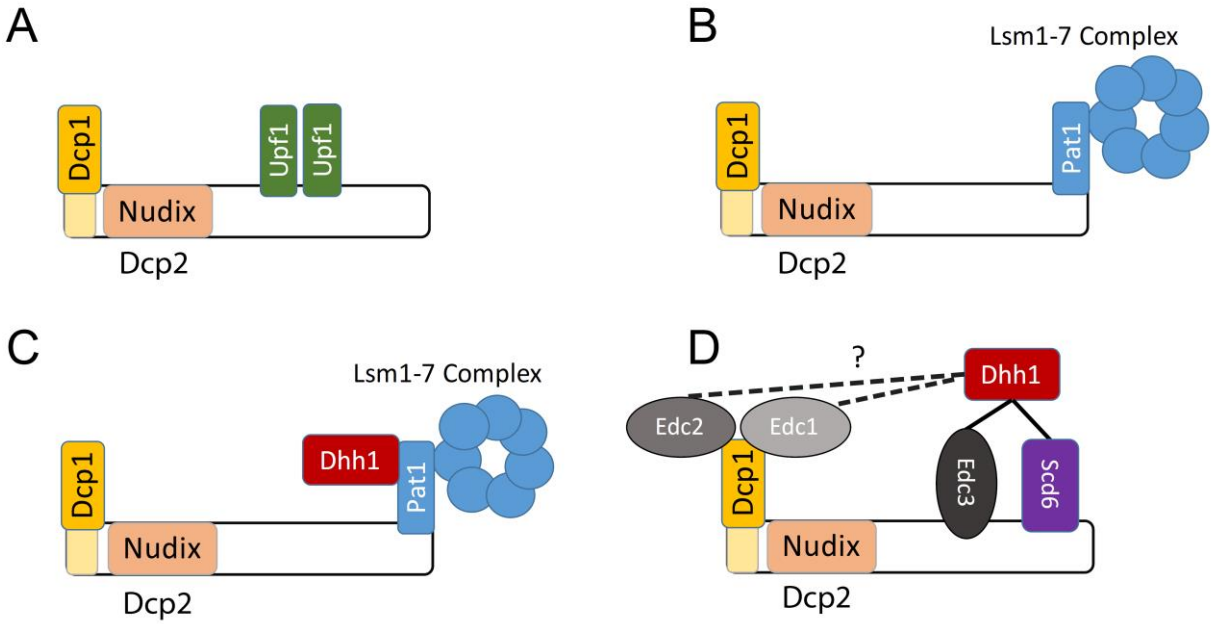


Figure 3.9 Different decapping complexes have different substrate specificities **A.** NMD substrates are targeted for decay by interactions between Upf1 and Dcp2 (He et al., 2013; He and Jacobson, 2015a). **B.** and **C.** The substrate specificity of the Lsm1-7-Pat1 complex is different from that of the Lsm1-7-Pat1-Dhh1 complex. **D.** Other decapping factors can interact with Dhh1 to target additional transcripts for decay. Dhh1 can associate with Edc3 or Scd6 and can, in turn, interact with Dcp2 (Fromm et al., 2012; He and Jacobson, 2015a; He et al., 2014; Tritschler et al., 2009a). Yeast specific decapping regulators Edc1 and Edc3 can bind to Dcp1's EVH1 domain (Borja et al., 2011). Unconfirmed interactions are shown as dashed lines, whereas confirmed interactions are shown as solid lines.

Chapter IV

General Discussion

General Discussion

Studies done by many groups in recent decades greatly increased our understanding of mRNA decay. The factors involved and their network of interactions have been revealed for the mRNA quality control pathways NMD, NGD, and NSD, as well as those required for general mRNA decay. Structural data have also been reported for many complexes and components, and have elucidated the significant conformational changes that they undergo. However, a mechanistic understanding of how mRNAs are targeted for decay has been limited. My thesis work focused on two different branches of mRNA decay (quality control and general decay) and different aspects of substrate selection for these pathways. My bioinformatic analyses provided insight into the reason why many "normal-looking" transcripts are targeted by NMD (Chapter II) and how a diverse pool of transcripts can be targeted for decapping and decay using a limited set of regulatory proteins (Chapter III).

A new class of NMD substrates illustrates the role of NMD in the cell

Work presented in Chapter II attempted to accomplish two goals: 1) to determine and confirm a high resolution, high confidence list of endogenous NMD substrates in yeast and 2) to determine the reason(s) why these substrates are targeted by NMD. NMD has widely been regarded as an mRNA quality control pathway. Decades of study dedicated to understanding the underpinnings of this pathway relied heavily on reporter genes that have been engineered to contain premature termination codons (PTCs). Yet these nonsense mutations, however deleterious (Peltz et al., 2013), are rare and do not provide a suitable explanation for the universal evolutionary conservation of NMD from yeast to plants to mammals. In addition, previous genome wide studies (He et al., 2003;

Malabat et al., 2015) identified many endogenous substrates that did not contain nonsense codons. In light of these results, it was proposed that the role of NMD was two-fold; 1) to detect PTC-containing mRNAs and 2) to regulate the expression of subsets of genes specific to each model organism and cell type examined (Celik et al., 2015; He and Jacobson, 2015b). While factors modulating NMD are well conserved, the subsets of genes that are targeted by NMD did not seem to have a functional conservation.

Several recent papers made unique observations about endogenous NMD substrates that previously escaped detection. Studies conducted by Kawashima and colleagues (Kawashima et al., 2014) indicated that pre-mRNA splicing in yeast is less precise than previously thought and that there are many “differentially” spliced mRNAs that are targeted by NMD. On a similar note, Malabat et al. (Malabat et al., 2015) observed that most genes can be expressed in many different minor isoforms. Some of these isoforms have transcription start sites that are inside the annotated open reading frame and the first ATG the ribosome would encounter is out of frame with respect to the annotated start. Clearly, translation of such mRNAs would likely result in premature termination and subsequent targeting by NMD.

In my thesis work, we observed that the majority of the mRNAs that were upregulated upon single deletion of the *UPF* genes did not contain any distinguishing features (e.g., upstream open reading frames, unspliced introns, programmed frameshifting, etc.) that could lead to a premature-like termination event. Upon closer inspection using ribosome footprint profiling data, we realized that these transcripts showed poorer translation fidelity which could only be partially explained by the transcription isoforms mentioned above. Therefore, these mRNAs are more prone to

mistranslation, possibly due to their lower codon optimalities which have been shown to stall ribosomes or slow translation elongation (Brule and Grayhack, 2017; Gamble et al., 2016; Letzring et al., 2010).

These observations lead to the conclusion that the primary role of NMD is not addressing the consequences of nonsense mutations, but rather to ensure that proper protein production takes place. This hypothesis can also explain the seemingly arbitrary selection of biological processes that are endogenously regulated by NMD in each organism because the role of NMD is not to regulate gene expression for functional purposes but is to monitor protein production. These "probabilistic" NMD substrates not only explain why genome wide studies in many different organisms identified normal looking mRNAs as substrates, but also provide indirect evidence against the pioneer-round of NMD activation model (Hosoda et al., 2005; Ishigaki et al., 2001; Lejeune et al., 2002; Maquat et al., 2010; Sato et al., 2008). If NMD substrates can only be targeted during the first round of translation, then the cell would accumulate significant amounts of truncated non-functional peptides translated at a completely different frame than the annotated gene. This is in direct contrast with the proposed function of NMD, i.e. to protect the cell from the effects of potentially dominant-negative effects of truncated or aberrant peptides. Being able to detect mistranslation events at any point in time renders NMD a much more effective quality control pathway.

A complex network of activators regulates decapping in the cell

Independent of quality control mechanisms such as NMD mRNAs in the cell need to be targeted for decay for effective regulation of gene expression. Decapping by the Dcp1-Dcp2 complex can occur *in vitro* (Dunckley and Parker, 1999; Dunckley et al.,

2001), but decay rates of mRNAs can vary greatly in *vivo,* implying that decapping and decay are subject to regulation. Work done in Chapter III of my thesis suggests a possible mechanism for how precise regulation of decapping is achieved in the cell.

Almost all biochemical and structural work concerning the decapping complex used a C-terminal truncation of the Dcp2 protein (Charenton et al., 2016; Mugridge et al., 2016; Valkov et al., 2016). While this truncated form of Dcp2 is catalytically active *in vitro* (Dunckley and Parker, 1999), we have recently shown that it lacks important aspects of regulation and substrate selection (He and Jacobson, 2015a). Deletion of different sections within the Dcp2 C-terminal domain resulted in loss of regulation for at least one mRNA (He and Jacobson, 2015a), suggesting that different transcripts are targeted by different decapping complexes that associate with this segment of Dcp2. My work in Chapter III provides transcriptome-wide evidence for this hypothesis and underscores the universal importance of the C-terminal extension of Dcp2. Differential expression analysis revealed that Lsm1 and Pat1, Dhh1, and Upf proteins regulate different subsets of mRNAs. In addition to targeting different transcripts for decay, the mechanisms with which different decapping complexes target their substrates also show divergence. While Dhh1 and the Upf factors have been shown to target transcripts co-translationally, and the mechanism for targeting by the Lsm1-7-Pat1 complex is likely to be different from that of the Dhh1 and Upf proteins. Comparisons of upregulated and downregulated transcripts also suggested the possibility of competition for Dcp2 between Dhh1 and Upf1, but not by the Lsm1-7-Pat1 complex. This observation revealed another aspect of substrate selection by different decapping complexes: different factors can confer substrate specificity by targeting transcripts at different stages in their life cycle.

Unlike its N-terminal domain, the C-terminal extension of Dcp2 shows poor evolutionary conservation (Wang et al., 2002). This observation may reflect the idea that factors that interact with this domain are responsible for providing substrate specificity. Hypothetically, different organisms and different tissues require different spatial and temporal regulation of gene expression, so the ability to maintain effective regulation with a limited number of decapping activators may have led to selective pressure on the Dcp2 C-terminal domain to evolve different interaction networks that are organism specific. Some of the unstructured domains in the yeast Dcp2 C-terminal extension show similarities for unconserved amino acid sequences in Dcp1 in metazoans. In combination with multiple Dcp2 proteins in higher eukaryotes (Grudzien-Nogalska and Kiledjian, 2017) regulation of decapping by conserved factors could be achieved by using divergent binding sites (Jonas and Izaurralde, 2013).

Potential issues in bioinformatic analyses

Our bioinformatic analyses provided valuable insight into substrate selection mechanisms of NMD and general decapping. Due to the design of our experiments, however, unambiguous interpretation of some results is difficult. Because we prepared our RNA-Seq libraries from strains with single deletions of several genes, at steady-state, we cannot definitively differentiate between direct and indirect effects of these genetic manipulations. One potential way to resolve this issue is to compare decay rates of the transcripts that differed between WT and single deletion strains. Significant changes in decay rates or decay mechanisms could certainly reveal direct substrates of each decapping factor. Addition of spike-ins for RNA-Seq libraries can also provide additional controls for model fitting and differential expression. This additional control would

especially improve data analysis in *dcp1*Δ*, dcp2*Δ*, dcp2-N245,* and *xrn1*Δ strains, where the general mRNA population might be different from WT and therefore have unintended consequences during model fitting.

Another issue that needs attention is the ribosome profiling analysis pipeline (Ingolia et al., 2013). Alignment of ribosome profiling libraries is often performed with substantially higher stringencies than in RNA-Seq experiments. For organisms that have undergone genome duplication like *Saccharomyces cerevisiae,* this can result in significant underestimation of ribosome occupancies of some mRNAs. Bayesian methods for transcript quantitation and Markov-Chain Monte-Carlo methods for confidence interval estimation have been successfully applied to RNA-Seq data previously (Li and Dewey, 2011; Li et al., 2010). These methods can greatly increase read count accuracies for transcripts with orthologs in the genome. Application of similar methods or other probabilistic approaches to read assignment for ribosome profiling data can alleviate some of the inherent problems with existing pipelines. For example, adaptation of Bayesian methods proposed by Li et al. (Li et al., 2010) for ribosome profiling by taking three nucleotide periodicities or abundances of different orthologs into account can result in more accurate estimations of ribosome occupancy.

Future directions

A) Mechanism of NMD activation

My work presented in Chapter II has improved our understanding of the role of NMD in the cell and provided an explanation for "normal-looking" endogenous NMD substrates. However, many aspects of NMD remain unanswered. To gain a better

understanding of how Upf proteins are recruited to mRNPs and how these mRNAs are targeted for decay requires that some key questions be addressed:

1) What is the timing and order of Upf protein association during translation?

While translation is required for activation of NMD the specific timing and order of Upf factor association has remained a controversial issue. To answer these questions at a transcriptome-wide level ribosome footprint profiling libraries from single deletions of the *UPF* genes as well as numerous mutants of these genes can be used to provide additional insight. In combination with mutations in translation termination and ribosome recycling factors we can obtain a dynamic picture of translation termination and NMD *in vivo*.

2) What are the immediate consequences of Upf association?

Previous work from our lab has provided indirect evidence suggesting that Upf proteins might be responsible for resolving stable premature termination complexes, a role reserved for Rli1 during normal termination (Amrani et al., 2004; Ghosh et al., 2010). A direct examination of this question using a reconstituted translation system from purified components can yield substantial information about the exact role of Upf proteins in the cell, such as enzymatic activities, or the roles of each Upf protein or their subsequent domains.

B) Different decapping complexes and their substrates

Results presented in Chapter III further confirmed the existence of different decapping complexes tasked with targeting different subsets of mRNAs. However, transcripts identified as differentially expressed in Chapter III constituted a small fraction of all transcripts in the cell, suggesting that there can be additional decapping complexes

regulating the expression of many other transcripts, potentially with some redundancy. There are two key questions that can greatly improve our understanding of decapping and decay regulation:

1) What are the components of the different decapping complexes in the cell?

Serial pulldown experiments followed by mass spectrometry can reveal the contents of different decapping complexes. These experiments can later be complemented by systematic two-hybrid experiments to determine individual binding sites.

2) Which transcripts are directly regulated by these decapping complexes?

Determination of transcriptome-wide decay rates from strains that contain deletions or disruptions of crucial components of different decapping complexes can reveal their direct substrates. Supervised and unsupervised clustering methods and thorough statistical analyses can also shed light into decay kinetics of these transcripts. In combination with deletions of the *XRN1* and *SKI2* or *SKI7* genes to target 5' to 3' decay and 3' to 5' decay pathways we might understand the decay preferences of different decapping complexes.

APPENDICES

APPENDIX A

TIMING AND SPECIFICITY OF THE POLYSOMAL ASSOCIATION OF Upf1

Contributions to this appendix

Robin Ganesan performed the affinity purifications of Upf1-bound 80S ribosomes and prepared next generation sequencing libraries.

Illumina sequencing of these libraries, was carried out at Beijing Genomics Institute (BGI).

Mass spectrometry sample preparation and data analyses were done by Robin Ganesan and John Leszyk (Proteomics and Mass Spectrometry Facility, UMass Medical School).

INTRODUCTION

Nonsense-mediated mRNA decay (NMD) is a cytoplasmic surveillance mechanism that degrades mRNA transcripts that contain premature termination codons (PTCs) (Kervestin and Jacobson, 2012). The mechanism by which nonsense-containing mRNAs are recognized and selectively targeted by NMD remains to be determined. Overexpression of Upf1 can compensate for mutations in Upf2 and Upf3, but not *vice versa* (Maderazo et al., 2000) and the maximal *in vitro* activation of the Upf1 ATPase and helicase activities requires both Upf2 and Upf3 (Chamieh et al., 2008). These observations imply that Upf1 is the key effector of NMD, whereas Upf2 and Upf3 are likely to be regulators of Upf1 function. Some current models of NMD activation by premature termination postulate that an inappropriate mRNP structure neighboring the PTC allows the recruitment of Upf1 to its substrates (Kervestin and Jacobson, 2012). Upf1 co-localizes with polyribosomes, 80S ribosomes, and 40S ribosomal subunits (Atkin et al., 1995; Ghosh et al., 2010; Mangus and Jacobson, 1999; Min et al., 2013), and interacts with the release factors eRF1 and eRF3 (Czaplinski et al., 1998; Kashima et al., 2006; Singh et al., 2008) and a specific ribosomal protein, Rps26 (Min et al., 2013). To map the transcriptome positions of ribosomes when they associate with Upf1, we employed the selective ribosome profiling approach (Becker et al., 2013).

RESULTS AND DISCUSSION

To assess the timing and specificity of Upf1 association with ribosomes, we established methods for selecting Upf1-associated ribosomes. In these experiments, 80S ribosomes with or without Upf1 selection were isolated, and libraries were generated from the respective ribosome-protected mRNA fragments in two biological replicates. Because

pulldown experiments using strains expressing Upf1 at physiological levels manifested very low recovery, we opted to utilize Upf1 overexpression strains to increase yield (Fig A.1 A). We then determined Upf1 enrichment using mass spectrometry analysis of the FLAG-Upf1-associated ribosomes. We observed a ~0.7:1 ratio of ribosomal proteins:Upf1 by Mascot Distiller quantitation, at least a 7-10-fold enrichment in Upf1-overexpressing cells (Fig A.1 B). Upf2 and Upf3 were largely absent from the Upf1-selected 80S preparations (data not shown), consistent with previous experiments indicating that Upf2 and Upf3 are not required for Upf1:80S association (Min et al., 2013).

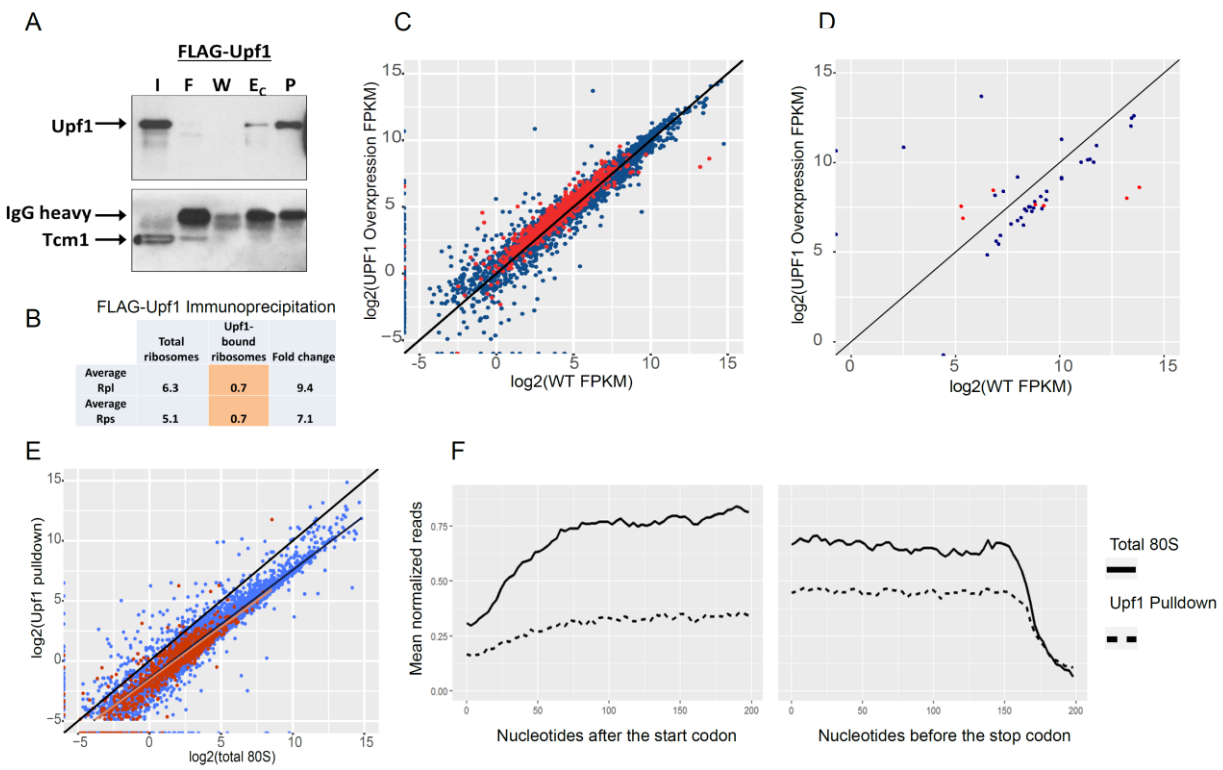
We first wanted to establish the consequences of Upf1 overexpression on the yeast transcriptome. We utilized the same differential expression pipeline described in Chapter II. Mean FPKM values showed good correlation between WT and overexpression RNA-Seq libraries (Fig A.1. C) and we only identified 46 transcripts to be differentially expressed, 11 of which were upregulated in Upf1 overexpression strains (Fig A.1. D). Within these 46 transcripts 7 were among the NMD substrates described in Chapter II. Subsequent chi-square analysis did not reveal any enrichment or depletion for NMD substrates ($p=0.681$). Therefore, we concluded that mRNA expression levels in Upf1 overexpression strains were mostly comparable to WT cells.

We then compared ribosome footprint profiling libraries between WT and Upf1 overexpression strains. Interestingly, we observed that a significant portion (~50%) of our sequencing reads came from ribosomes translating Upf1 (data not shown). This amount of enrichment was much larger than what would be expected due to overexpression of Upf1. We therefore hypothesized that in addition to obtaining Upf1 bound ribosomes our

samples also contain Upf1 nascent chains. We subsequently removed all reads that map to the *UPF1* ORF before further analysis.

Consistent with our previous unpublished results (Min, 2015), we did not observe any significant enrichment for NMD substrates in Upf1 pulldown libraries. FPKM distributions in pulldown libraries were highly correlated with total ribosomes, but showed slightly higher ribosome occupancies in total 80S libraries even when accounted for library sizes. This slight shift in ribosome occupancies, however, was comparable between NMD substrates and non-substrates (linear regression slope=0.901±0.013 and 0.914±0.003 for NMD substrates and non-substrates respectively) (Fig A.1. E). Finally, we evaluated whether Upf1 associated uniformly within ORFs or whether these ribosomes were enriched at the 3' or 5' end of transcripts. Unlike our previous observations (Min, 2015), we did not observe an enrichment at the 3' ends of ORFs (Fig. A.1. F).

Results presented here indicate that Upf1 can interact with ribosomes stochastically. Upf bound ribosomes do not show any significant enrichment for NMD substrates or any specific location within the ORF. There is one major caveat that needs to be addressed in this study. Upf1 ribosome association might appear non-specific in our studies because of excess Upf1 compared to Upf2 or Upf3. Therefore, establishing protocols that would allow us to isolate Upf1-bound ribosomes from cells expressing this protein at physiological levels in future experiments is crucial for determining the dynamics of Upf1 association with ribosomes during translation.



Stochastic binding of Upf1 to ribosomes during translation

A. Western blot of input (I), flowthrough (F), final wash (W), concentrated eluate (E_c) and immunoprecipitated (FLAG-Upf1) ribosomes after pelleting through a 10% sucrose cushion (P). **B.** Average of Upf1-normalized Mascot Distiller quantitation values of all detected ribosomal proteins normalized against Upf1 from total ribosomes (input) and Upf1-associated ribosomes (pellet) from two (FLAG) pulldown experiments. Fold change indicates the change between ribosomal protein:Upf1 ratios before and after selective pulldown. **C-E.** Scatterplots of log2(FPKM) values from libraries prepared from FLAG-Upf1-containing strains. Reads mapping to Upf1 ORF were omitted from Upf1-associated ribosome profiling libraries to correct for pulldown of the nascent FLAG-Upf1 peptide. NMD substrates identified in Chapter II are in red. **(C)** mean log2(FPKM) values between RNA-Seq libraries prepared from two single-copy (endogenous) and two episomal (overexpressed) Upf1-containing strains show good correlation. **(D)** Transcripts that were differentially expressed between Upf1 overexpression and WT strains. **(E)** log2(FPKM) values of all transcripts from different ribosome footprint profiling libraries. Orange line and dark blue line represent linear regression fits to NMD substrates and non-substrates respectively. **F.** Distribution of mean normalized reads over the first 200 nucleotides after the start codon and the last 200 nucleotides before the stop codon. Reads from ribosome profiling libraries were normalized against their respective RNA-Seq libraries per transcript.

APPENDIX B

EFFECTS OF Upf2, eRF1, AND eRF3 ON Upf1 ATPase ACTIVITIES

INTRODUCTION

Nonsense-mediated mRNA decay (NMD) is a translation quality control pathway that recognizes mRNAs whose translation is terminating prematurely (Celik et al., 2015). NMD is regulated by three well conserved proteins, Upf1, Upf2, and Upf3 (Cui et al., 1995; Gatfield et al., 2003; He et al., 1997; He and Jacobson, 1995; Leeds et al., 1991; Lykke-Andersen et al., 2000; Nicholson et al., 2012; Nicholson et al., 2010; Perlick et al., 1996; Salas-Marco and Bedwell, 2004). Upf1 contains RNA-dependent ATPase and ATP-dependent RNA helicase activities (Bhattacharya et al., 2000; Czaplinski et al., 1995) that have been implicated in mRNP remodeling upon PTC recognition (Ghosh et al., 2010). The conserved N-terminal cysteine-histidine rich domain of Upf1 interacts with the C-terminal domain of Upf2 and, in turn, Upf2 interacts with Upf3 (He et al., 1997). In addition, Upf1 can associate with the 40S ribosomal subunit (Min et al., 2013) and eukaryotic release factors (eRF1 and eRF3) (Czaplinski et al., 1995; Kashima et al., 2006; Singh et al., 2008).

Previously published results using recombinant Upf fragments indicated that Upf2 and Upf3 can promote (Chamieh et al., 2008) while eRF1 and eRF3 can inhibit (Czaplinski et al., 1995) the ATPase and helicase activities of Upf1. However, the enhancement of ATP hydrolysis of Upf1 was only observed using a small fragment of Upf2 (Chamieh et al., 2008). Similarly, experiments illustrating the effects of release factors on Upf1 were performed using recombinant proteins that contained bulky Glutathione S-transferase (GST) tags on their N-termini (Czaplinski et al., 1995). We therefore re-assessed the influence of Upf2, eRF1, and eRF3 on Upf1. Contrary to previously published results, we

observed that Upf2 inhibits, and release factors promote activation of Upf1's ATPase activities.

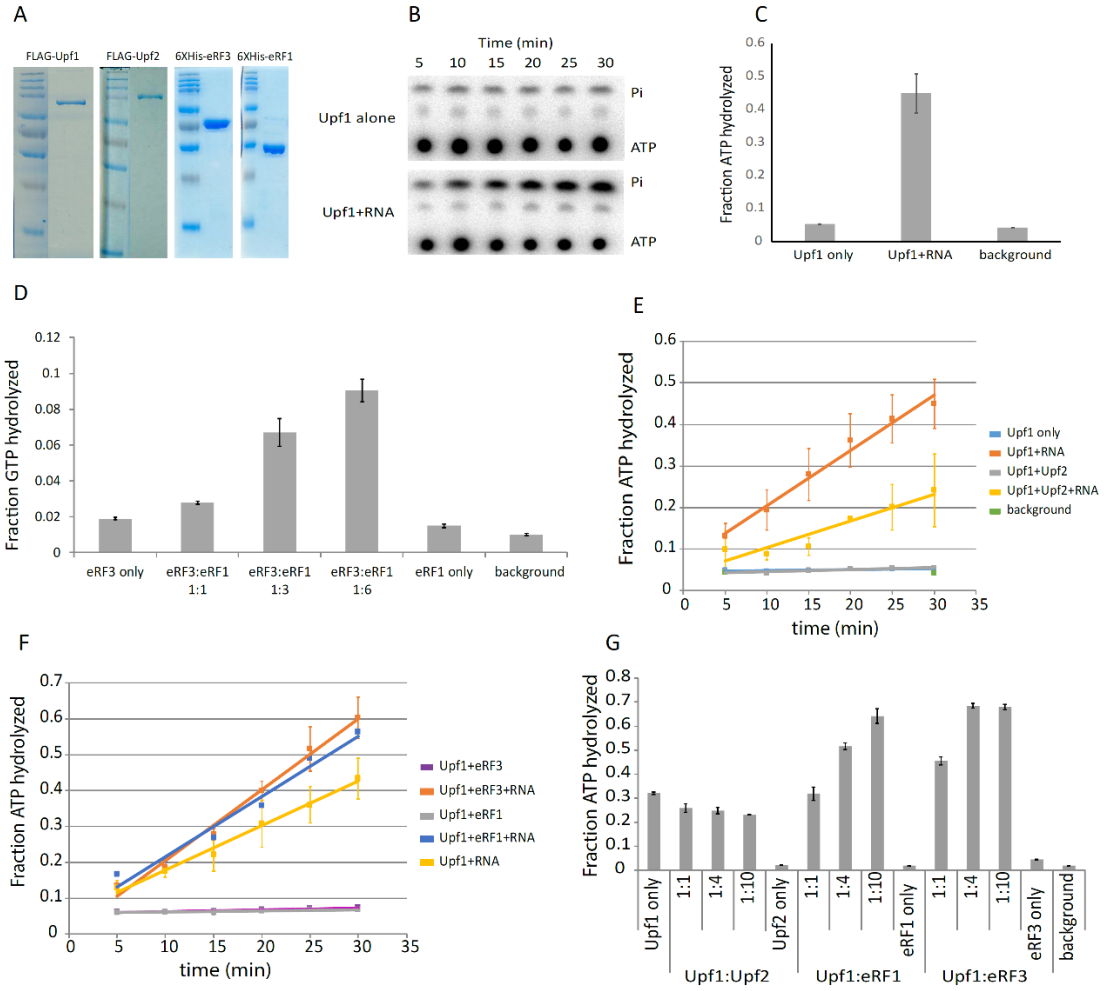
RESULTS AND DISCUSSION

To determine the consequences of Upf2 or eRF binding on Upf1's enzymatic activities, we purified full-length FLAG-Upf1 and FLAG-Upf2 directly from *Saccharomyces cerevisiae* and full-length recombinant yeast 6xHis-eRF1 and an N terminally truncated (codons 166-685) 6xHis-eRF3 from *Escherichia coli*. We opted to use a truncated version of eRF3 for two reasons: 1) the N-terminal region of eRF3 in yeast contains a prion domain that can aggregate at high concentrations, potentially interfering with derivation of meaningful results. In addition, this prion domain is responsible for the [PSI+] phenotype in yeast, which results in reduced translation termination fidelity (Bradley et al., 2003; Cosson et al., 2002; Serio et al., 2001; Serio et al., 1999; Ter-Avanesyan et al., 1994); 2) Previous results have shown that this N-terminal region is dispensable for NMD, translation termination, or growth *in vivo* (Roque et al., 2015), and would likely have minimal effects in our *in vitro* analyses.

Our purification protocols yielded high purity proteins from yeast and bacteria (Fig B.1 A). In line with previous observations, purified Upf1 showed RNA-dependent ATPase activity (Fig B.1. B, C). Similarly, the GTPase activity of eRF3 was stimulated by addition of eRF1 in a dose dependent manner, and eRF1 alone did not have any discernible GTPase activity (Fig B.1. D). We then tested the effect of Upf2 on Upf1's ATPase activity. Surprisingly, we observed a reduction of ATP hydrolysis when Upf2 was present alongside Upf1 (Fig B.1. E). eRF1 or eRF3, on the other hand, manifested the opposite result to Upf2, leading to stimulation of ATP hydrolysis by Upf1 (Fig B.1 F). This inhibition

or enhancement of ATP hydrolysis was dependent on Upf2 or eRF concentration, namely, higher concentrations of Upf2 or eRFs resulted in further inhibition or stimulation of Upf1 respectively (Fig B.1. G). Our observations are in direct conflict with previously published results and underscore the importance of utilizing full-length proteins with small epitope tags for *in vitro* studies whenever possible.

Previously it was hypothesized that Upf1 association with eRF1 and eRF3 on a prematurely terminating ribosome could result in stabilization of RNA-Upf1 interactions. Upon binding to Upf2 and 3 this interaction is resolved with re-activation of Upf1's enzymatic activities (Chamieh et al., 2008). Yet the results presented here suggest the opposite mechanism, where a stable Upf complex can bind to a prematurely terminating ribosome and dissociation of Upf2 and Upf3 from Upf1 and binding to release factors can improve Upf1's enzymatic activities and result in efficient remodeling of the mRNP. Our results are also consistent with the observation that yeast Upf2 and Upf3 both interact with eRF3 and may compete with eRF1 for eRF3 binding (Wang et al., 2001). These interactions can play roles in early steps in recruiting Upf1 to a prematurely terminating ribosome.



Upf2 inhibits, eRF1 and eRF3 promote Upf1's ATPase activities

A. Coomassie Brilliant Blue-stained SDS-PAGE gels showing purity of yeast FLAG-Upf1, FLAG-Upf2, 6xHIS-eRF1, and 6xHIS-eRF3. **B-C.** Thin layer chromatography showing ATPase activities of Upf1 in the presence and absence of poly(U_{15}) RNA. Aliquots were taken at each time point. Reactions were quenched with 400mM EDTA and reaction products were separated by thin layer chromatography and quantitated using autoradiography. Error bars in (C) indicate standard deviations. **D.** Similar to (C), but with γ-labeled GTP and without poly(U_{15}) RNA. eRF3's GTPase activity was stimulated in the presence of eRF1 in a concentration dependent manner. eRF1 alone shows minimal GTPase activity. **E.** Time course experiments indicate that Upf2 diminishes Upf1 enzymatic activity in the presence of RNA; error bars show standard deviations. Lines represent linear regression fits for each experimental condition. **F.** Time course ATP hydrolysis assays show that eRF1 and eRF3 stimulate Upf1's ATPase activity in an RNA dependent manner. Error bars indicate standard deviations of each time point. Lines represent linear regression fits for each experimental condition. **G.** Concentration dependent inhibition by Upf2 and stimulation by eRF1 or eRF3. Upf2, eRF3, or eRF1 alone show negligible ATPase activities.

APPENDIX C

ASSOCIATION OF Upf2 AND Upf3 WITH RIBOSOMAL SUBUNITS

Contributions to this appendix

Feng He prepared HA-Upf2 and HA-Upf3 plasmids.

Nadia Amrani generated single, double, and triple *upf* deletion strains.

INTRODUCTION

Nonsense-mediated mRNA decay (NMD) is a cytoplasmic surveillance mechanism responsible for degrading transcripts that contain premature termination codons (Celik et al., 2015). Factors that regulate NMD in the yeast *Saccharomyces cerevisiae* include the highly conserved but non-essential proteins, Upf1, Upf2, and Upf3. All three Upf proteins associate with polysomes and numerous translation factors (Celik et al., 2015). Previous results have indicated that Upf1 can interact with the small ribosomal subunit and this interaction dependent on retention of Upf1's ATP hydrolysis activities (Min et al., 2013), but not Upf2 or Upf3.

To better understand the mechanics of Upf assembly on ribosomes, we determined which ribosomal subunits Upf2 and Upf3 interact with and whether this interaction is mediated by other Upf proteins. Our results indicate that both Upf2 and Upf3 can bind to both ribosomal subunits and, while Upf2 binding did not rely on other Upf factors, Upf3's association with 40s relies on the presence of either Upf1 or Upf2. These results suggest that a complex and dynamic restructuring of prematurely terminating mRNPs occurs upon PTC recognition and provide evidence for all three Upf proteins' involvement with ribosomes.

RESULTS AND DISCUSSION

Upf2 associates with both ribosomal subunits independent of Upf1 or Upf3

Previously published results place Upf1 on the small ribosomal subunit (Min et al., 2013). We therefore wanted to investigate whether this also was the case for Upf2. Because there are no available antibodies for yeast Upf2 we transformed *upf2*Δ cells with a plasmid that contained a full length, N-terminal HA tagged Upf2, purified individual ribosomal subunits as previously described (Min et al., 2013), and investigated the

presence of Upf2 by western blotting. Interestingly, Upf2 associated equally well with both ribosomal subunits (Fig C.1 A). When we repeated this experiment in *upf1*Δ*, upf3*Δ*,* and *upf1*Δ and *upf3*Δ cells we observed the same phenotype, suggesting that stable association of Upf2 did not depend on the presence of either Upf1 or Upf3 (Fig C.1 B-D).

Upf3 associates with both ribosomal subunits, but its 40S association requires the presence of either Upf1 or Upf2

We then wanted to determine whether Upf3 showed a similar association pattern with ribosomes. Using a similar N-terminal HA-tagged Upf3 protein in *upf3*Δ strains, in the presence of Upf1 and Upf2, Upf3 also associated with both ribosomal subunits (Fig C.1 E). Single deletions of *UPF1* or *UPF2* did not alter this association (Fig C.1 F, G respectively). However, deletion of both *UPF1* and *UPF2* resulted in inhibition of Upf3 association with the 40S (Fig C.1 H).

Upf proteins have been implicated in resolving poorly dissociable premature termination complexes (Ghosh et al., 2010). Association of Upf2 and Upf3 with both ribosomal subunits is consistent with this hypothesis. At PTCs, the Upf complex can therefore assume the roles played by ribosome recycling complexes at normal termination events (Jackson et al., 2012). Upf3's requirement for either Upf1 or Upf2 also suggests a dynamic assembly of these factors upon PTC recognition. However, the stable interactions determined in these experiments cannot distinguish whether both ribosomal subunits are occupied with Upf2 and/or Upf3 simultaneously. Further experimentation is needed to determine the exact location and the order of Upf2 and Upf3 binding to the ribosome.

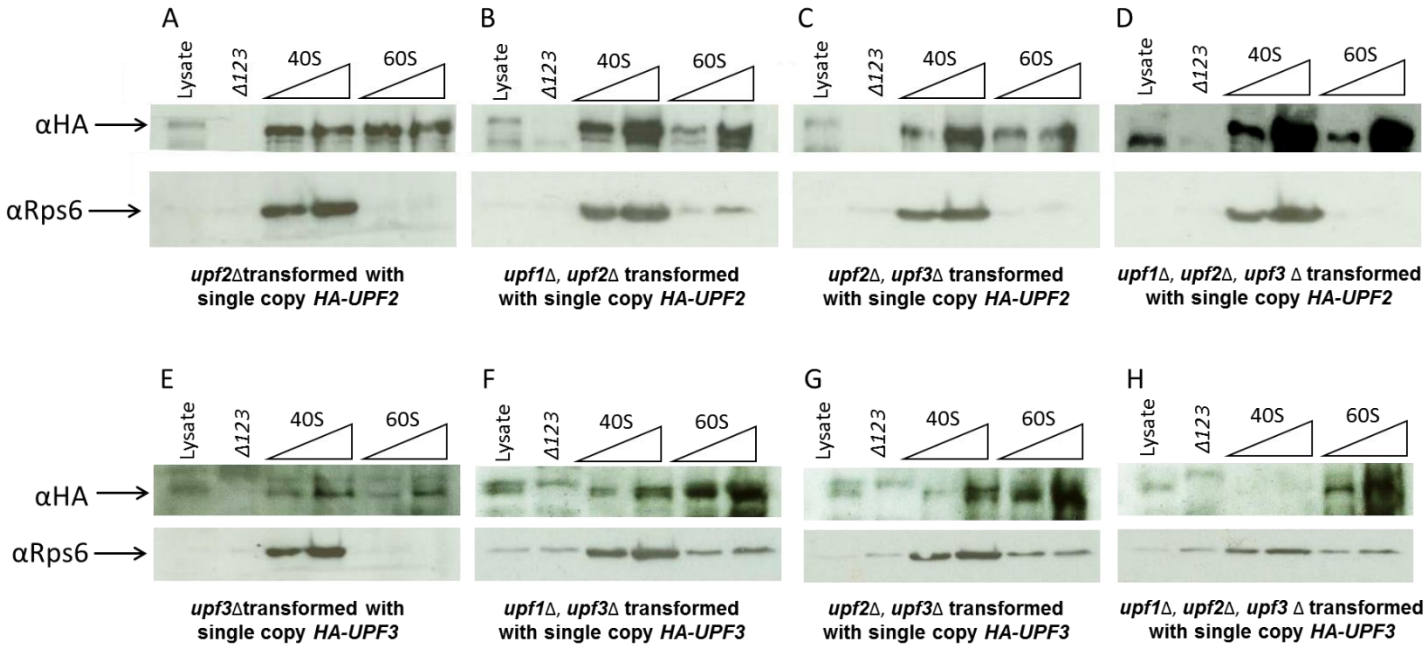


Figure C.1. Association of Upf2 and Upf3 with ribosomal subunits.

Upf2 associates with both ribosomal subunits in WT conditions (A), in the absence of Upf1 (B), in the absence of Upf3, or in the absence of both Upf1 and Upf3 (D). Upf3 associates with both subunits under WT conditions (E), in the absence of Upf2 (F), in the absence of Upf1 (G). However, in *upf1Δ/upf2Δ* strains Upf3 only associates with the 60S subunit (H). HA-tagged Upf1 or Upf3 were detected with α-HA antibody (upper blots). Western blots were also probed with α-Rps6 antibody to determine 40S contamination in 60S samples (lower panels). Triple *upf* deletion strains were used as control (second lane in each panel).

APPENDIX D

MATERIALS AND METHODS

CHAPTER II

Materials and Methods

Yeast strains

All strains used in this study are in the W303 background. The wild-type strain (HFY114) and its isogenic derivatives harboring deletions of *UPF1* (HFY871), *UPF2* (HFY116), *UPF3* (HFY861), *DCP1* (HFY1067), or *XRN1* (HFY1080) were described previously (He et al. 2013), as were isogenic strains harboring deletions of *DCP2* (CFY1016), *EDC3* (CFY25), *PAT1* (SYY2674), *LSM1* (SYY2680), or *DHH1* (SYY2686) (He and Jacobson, 2015a). A strain harboring a deletion of *SCD6* (SSY2352) was constructed by gene replacement (Guthrie and Fink, 1991) using a DNA fragment harboring the *scd6::KanMX6* null allele.

Cell growth and RNA isolation

Cells were all grown in YEPD media at 30°C. In each case, cells (15 ml) were grown to an OD_{600} of 0.7 and harvested by centrifugation. Cell pellets were frozen on dry ice and then stored at -80°C until RNA isolation. The procedures for RNA isolation were as previously described (He and Jacobson, 1995).

RNA-Seq library preparation

Total RNA was treated with Baseline-Zero DNase (Epicenter) to remove any genomic DNA contamination. Five micrograms of DNase-treated total RNA was then depleted of rRNA using the Illumina yeast RiboZero Removal Kit and the resulting RNA was used for RNA-Seq library preparation. Multiplex strand-specific cDNA libraries were constructed

using the Illumina TruSeq Stranded mRNA LT Sample Prep Kit. Three independent cDNA libraries were prepared for each yeast strain analyzed.

RNA sequencing

Total RNA cDNA libraries were sequenced on the Illumina HiSeq4000 platform at Beijing Genomics Institute. Four independent libraries were pooled into a single lane and single-end 50-cycle sequencing was carried out for all cDNA libraries.

Northern analysis

Procedures for northern blotting were as previously described (He and Jacobson 1995). In each case, the blot was hybridized to a random primed probe for a specific transcript, with *SCR1* serving as a loading control. Transcript-specific signals on northern blots were determined with a FUJI BAS-2500 analyzer. Specific PCR fragments from the following genes were used as probes for northern blotting analyses presented in Fig. 4: *RPL22B*, entire 321-nt intron; *RPS13*, entire 539-nt intron; *HAC*, entire 252-nt intron, *HRB1*, exon2 nt 784-1278; *NHP6B*, CDS nt 1-300; *MTR2,* CDS nt 1-555; *ICR1*, nt 2461-3040; *IRT1*, nt 811-1340; and *TY4*, CDS nt 4801-5410.

Bioinformatic methods

General computational methods

All statistical analyses were carried out using the R statistical programming environment, versions 3.2.4 and 3.2.5. R packages ggplot2, gplots, plyr, reshape2, and gridExtra were used for data pre-processing and visualization, and foreach, doSNOW, and doParallel were used for parallel processing.

Differential expression analysis

The R64-2-1 S288C reference genome assembly (sacCer3) (*Saccharomyces* Genome Database Project) was used for sequencing reads mapping and transcriptome construction. We generated a yeast transcriptome comprised of 7473 transcripts that includes all annotated protein-coding sequences, functional, and non-coding RNAs, and the unspliced isoforms of all intron-containing genes. Because of their repetitive nature, autonomous replicating sequences and long terminal repeats of transposable elements were excluded from the transcriptome. The RSEM program (Li and Dewey 2011) was used to map the sequence reads to the transcriptome and to quantify individual transcript levels with settings --bowtie-m 30 --no-bam-output --forward-prob 0. The expected read counts for individual transcripts from RSEM were considered as the number of reads mapped to each transcript and were then imported into the Bioconductor DESeq package (Anders and Huber, 2010) for differential expression analysis.

The Benjamini-Hochberg procedure was used for multiple testing corrections. To account for replicate variability, we used a false discovery threshold of 0.01 (1%) instead of an arbitrary fold change cutoff as the criterion for differential expression. We repeated the same pipeline for a separate transcriptome that contained CUT, SUT and XUT sequences. These sequences were extracted from the yeast genome based on previous annotations (Wery et al., 2016).

Ribosome footprint profiling analysis

We generated a second transcriptome for ribosome profiling analysis that only included mRNAs and unspliced isoforms from verified protein-coding genes. Because there are no

formal 5'- and 3'-UTR annotations for most yeast transcripts, we used sequences 300 nt upstream of start codons or downstream of stop codons as the 5'- and 3'-UTR sequences. For transcripts that have annotated 5'-UTR introns, 300nt immediately upstream of the annotated introns were considered as their 5'-UTRs. We used raw data from previously published ribosomal profiling experiments (Smith et al., 2014; Young et al., 2015). Raw fastq files and sequence reads were trimmed for adapter sequences with cutadapt with settings *-a CTGTAGGCA -q 10 --trim-n -m 10*. After adapter trimming, sequence reads were mapped to the transcriptome with bowtie (Langmead et al., 2009) with settings *-m 4 -n 2 -l 15 --suppress 1,6,7,8 --best –strata*. Because our transcriptome contains both spliced and unspliced isoforms from hundreds of the intron-containing genes, we allowed as many as 4 multiple mappings. After *bowtie* alignment, the riboSeqR (Chung et al., 2015) was used for initial visualizations and frame calling. All other analyses were carried out by R scripts written in-house.

Ribosome density calculations

The ratio of $\text{profiling}_{\text{coverage}}$ / $\text{RNA-Seq}_{\text{coverage}}$ for each transcript along the entirety of either intronic or coding regions (Fig. 3A, Fig. 5 A, C), or over 100 bins (percentages) of the entire intron or coding regions were calculated using in-house scripts. For this analysis, we only used ribosome footprint read lengths that showed a strong preference (>80%) to a specific reading frame.

Calculation of in- and out-of-frame read ratios

We used only ribosome footprint read lengths that showed a strong preference (>80%) to a specific reading frame. Accounting for A-site occupancy, we mapped the reads from

each length to our transcriptome and calculated the number of reads mapped to each transcript and the number of out-of-frame reads for each transcript for each read length. We then pooled the total and out-of-frame reads together and calculated the ratio of out-of-frame over total reads for each transcript.

Codon optimality calculations

We used previously published codon optimality assignments and scores (Pechmann and Frydman, 2013) in our analyses. The average codon optimality score for each transcript in our ribosome profiling transcriptome was calculated using the Biostrings R package and in-house scripts. We took the sum of optimality scores for all codons in a transcript and then divided the sum by the total number of codons in the corresponding transcript. For discrete time Markov chain analysis, we labelled each codon as optimal (O) or non-optimal (N) and then calculated the transition probabilities using maximum likelihood estimates as an unbiased measure for each transcript using the markovchain R package.

Statistical tests

We used chi-square tests with Yates continuity correction to assess different subsets of transcripts for either enrichment or depletion of a particular group of transcripts. Because the data for ribosome densities, transcript in-frame reads ratios, average codon optimality, and codon transition probabilities did not show normal (Gaussian) distributions, we used non-parametric two-sample Kolmogrov-Smirnov (K.S.) tests to assess the significance between different groups of transcripts. As K.S. tests compare the empirical distributions of two population samples, for consistency we used cumulative density plots in Figures 4-6.

The data discussed in this chapter have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE86428.

CHAPTER III

Yeast strains

All strains used in this study are in the W303 background. The wild-type strain (HFY114) and its isogenic derivatives harboring deletions of *UPF1*, *UPF2*, *UPF3*, *DCP1*, or *XRN1* were described previously (He et al., 2013), as were isogenic strains harboring deletions of *DCP2*, *EDC3*, *PAT1*, *LSM1*, or *DHH1* (He and Jacobson, 2015a). Construction of the *DCP2* alleles *dcp2-N245, dcp2E153Q-N245,* and *dcp2E198Q-N245* was also described previously (He and Jacobson, 2015a).

Cell growth and RNA isolation

Cells were all grown in YEPD media at 30^{o}C. In each case, cells (15 ml) were grown to an OD600 of 0.7 and harvested by centrifugation. Cell pellets were frozen on dry ice and then stored at -80^{o}C until RNA isolation. The procedures for RNA isolation were as previously described (He and Jacobson 1995).

RNA-Seq library preparation

Total RNA was treated with Baseline-Zero DNase (Epicenter) to remove any genomic DNA contamination. Five micrograms of DNase-treated total RNA was then depleted of rRNA using the Illumina yeast RiboZero Removal Kit and the resulting RNA was used for RNA-Seq library preparation. Multiplex strand-specific cDNA libraries were constructed

using the Illumina TruSeq Stranded mRNA LT Sample Prep Kit. Three independent cDNA libraries were prepared for each yeast strain analyzed.

RNA sequencing

Total RNA cDNA libraries were sequenced on the Illumina HiSeq4000 platform at Beijing Genomics Institute. Four independent libraries were pooled into a single lane and single-end 50-cycle sequencing was carried out for all cDNA libraries.

Northern analysis

Procedures for northern blotting were as previously described (He and Jacobson 1995). In each case, the blot was hybridized to a random primed probe for a specific transcript, with SCR1 serving as a loading control. Transcript-specific signals on northern blots were determined with a FUJI BAS-2500 analyzer. Specific PCR fragments from the ORFs of following genes were used as probes for the northern blotting analyses presented: *HXK1* nt 1-470, *CHA1* nt 1-500, *RTC3* nt 1-336, *NQM1* nt 481-2001, *RPGM2* nt 1201-1710, *TMA10*- nt 1-260, *RPP1A* nt 1-321, *TMA19* nt 1-500, *GPP2* nt 1-500, *CIT2* nt 1-500, *SDS23* nt 1-500, *HOS2* nt 1-500, *DIF1* nt 1-400, *AGA1* nt 1-500, *BUR6* nt 1-429, and *GTT2* nt 1-500.

Bioinformatic methods

General computational methods

All statistical analyses were carried out using the R statistical programming environment, versions 3.3.0 and 3.3.1. R packages ggplot2, gplots, plyr, reshape2, and gridExtra were used for data pre-processing and visualization. Biostrings, BiocParallel, snow, and

parallel were used for parallel processing. Statistical tests were performed using built-in functions in base R distributions.

Differential expression analysis

The R64-2-1 S288C reference genome assembly (sacCer3) (*Saccharomyces* Genome Database Project) was used for sequencing reads mapping and transcriptome construction. We used Transcriptome 1 described in Chapter 2 for differential expression using previously published pipelines (Celik et al. 2017).

Statistical tests

Hierarchical clustering was performed using Euclidian distances between libraries and transcripts with complete linkage. Non-finite (division by 0), undefined (0 divided by 0) values were removed prior to clustering. The heights of the clustering tree branches indicate distance between two libraries. We used Fisher's exact test to assess different subsets of transcripts for either enrichment or depletion of a particular group of transcripts.

Gene ontology term enrichment analysis

We used the GOrilla web interface (Eden et al. 2009) for gene ontology (GO) term enrichment analysis for all regulation subgroups. We utilized hypergeometric tests between a set of differentially expressed transcripts versus all transcripts. To minimize false discovery, we employed a stringent (0.001) p value cutoff for significance.

APPENDIX A

Cells and treatments

The pG1-FLAG-Upf1 plasmid (Czaplinski et al., 1995) was transformed into HFY871 (He et al., 2013) using previously described methods (Schiestl and Gietz, 1989). Yeast strains

were grown on -trp minimal medium until the OD_{600} reached 0.6-0.8. Cells were harvested by vacuum filtration on 80 um filters (Millipore) and resuspended in lysis buffer (10 mM Tris, pH 7.4, 100 mM NaCl, 30 mM $MgCl_2$, 1 mM DTT, 1 mM PMSF, 200 ug/ml Heparin), and dripped into liquid N_2. Frozen cell pellets were cryogenically ground at 10Hz for 15min in a mixer mill (Retsch). Frozen cell powder was thawed on ice and centrifuged for 10 min at 5000 rpm. The supernatant was centrifuged at 13,000 RPM for 10 min at 4°C. Absorbance at 260 nm was measured and 1 OD_{260} unit was treated with 0.15 ul of RNase I (100 Units/ul of Ambion) for 1 hr at room temperature. 5 ul of Superase-In (20 U/ul; Ambion AM2694) was added to the RNase I-treated samples and subjected to sucrose gradient centrifugation (see below).

Ribosome isolation

Lysates were loaded on 34 ml 10-50% sucrose gradients, prepared in low-salt polysome buffer (0.5 M Tris acetate pH 7.4, 0.5 M NH_4Cl, 0.12 M $MgCl_2$, 10 mM DTT), and centrifuged at 28,000 rpm in an SW-28 Ti rotor for 3.15 hr at 4°C. Sucrose gradient fractions containing 80S ribosomes were collected and concentrated with 100 kDa Amicon filter units (Millipore) to a volume of about 500-1000 ul and saved in aliquots at -80°C.

Affinity purification of Upf1-bound ribosomes

Ribosomes (0.05 pmol) were incubated with 50 µl of magnetic FLAG-M2 beads (Sigma) pre-equilibrated with 50mM Tris 7.4, 1M NaCl, 1% NP40, 0.1% SDS, 0.5% NaDOC, and 50x protease inhibitor cocktail (Roche) for 10 min at 4°C. The magnetic beads were washed five times with the same buffer. FLAG-Upf1-associated mRNP complexes were eluted with Buffer A supplemented with 10 mM FLAG-Peptide (pH 7.0) by incubating for

10 min at 4°C. The eluted FLAG-Upf1-80S samples were combined and concentrated by centrifugation through 100 µl sucrose cushions (50 mM Tris, pH 7.0, 500 mM KOA, 25 mM $Mg[OAc]_2$, 5 mM β-mercaptoethanol, 1 M sucrose, protease inhibitor, RNase inhibitor) at 50,000 rpm for 180 min in a TLS 55 rotor at 4°C. The resulting pellet was dissolved in 30 ul of Buffer B (20 mM HEPES, pH 7.0, 100 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, 0.5 mM spermidine, 0.05% Nikkol, 0.5% protease inhibitor, and 0.1 U/ml RNAsin) for 1 hr on a rotating shaker at 4°C. Aliquots of concentrated Upf1-associated ribosomes were flash frozen in liquid nitrogen and stored at −80°C.

Mass spectrometry

After affinity purification, Upf1-eniriched ribosomes were subjected to 8% SDS-PAGE analysis. Gels were silver stained using ProteoSilver Silver Stain Kit (Sigma, PROTSIL1-1KT) and processed for mass spectrometry (MS) analysis. Destained gel bands were denatured, reduced with dithiothreitol, and alkylated with iodoacetamide, and subjected to in-gel trypsin digestion. The resulting peptides were subjected to Liquid Chromatography/Tandem MS Analysis (LC-MS) on a Thermo LTQ. Protein identification was performed with the Mascot Server (version 2.4; Matrix Sciences, Ltd.) using the UniProt index of *Saccharomyces cerevisiae.* Quantitation was performed using Mascot Distiller (version 2.4; Matrix Science, Inc.).

Bioinformatic methods

Ribosome footprint profiling libraries were analyzed as described in Chapter II methods with some modifications. Because Upf1-enriched libraries contained significant amounts of reads that mapped to the Upf1 ORF, these were removed before further processing

and meta-gene analysis. In addition, due to a lack of discernible three nucleotide periodicities we used all sequencing reads between 15-36 nucleotides for our analyses.

APPENDIX B

Plasmid construction

For Upf1, a previously described plasmid was used for expression and purification (Czaplinski et al., 1995). FLAG-Upf2 was amplified from genomic DNA with primers that contain the FLAG tag in frame with the ORF. This DNA fragment was cloned into the YEplac112 plasmid that contains the *ADH1* promoter and terminator (Roy et al., 2015). Full length eRF1, and the eRF3 ORF containing codons 166-685 amplified from genomic DNA, were cloned into the PET15b (Novagen) plasmid.

Protein purification

Upf1 and Upf2 protein purification was performed as previously described (Bhattacharya et al., 2000) with the exception that cells were cryogenically ground as described in Appendix A methods. eRF1 and eRF3 plasmids were transformed into BL21 bacterial strains and grown under ampicillin selection. Overnight cultures were diluted to OD600 0.1 and grown until they reached OD600 0.6 in 2L LB-Amp media. Cells were then induced with 0.5 mM IPTG and transferred to a 16^{o}C water bath and grown overnight. Cells were collected by centrifugation and were suspended in 100mM Hepes pH 7.4, 100mM NaCl, 10% v/v glycerol. Cell lysis was performed by sonication with 10x 10-second pulses at 30% power, with 30 second on-ice cooling between pulses. Cell lysates were then cleared by centrifugation at 15,000 g for 15 minutes and protein purification was performed using 1.5 ml TALON resins (Clonetech) per the manufacturer's

recommendations. Elution fractions containing the highest protein concentrations determined by Bradford assays were then pooled and dialyzed overnight in 100mM Hepes pH 7.4, 100mM NaCl, 10% v/v glycerol. Final protein concentration was measured using Bradford assays with BSA as a protein standard. Proteins were aliquoted and stored at -80^{o}C. Protein purity was assessed by SDS-PAGE and Coomassie Brilliant Blue staining.

ATP-hydrolysis assays

A final concentration of 100nM Upf1 protein was combined with 1mM non-radioactive ATP, 0.1ul of gamma-labeled ATP (Perkin Elmer), and additional factors or BSA in 1X reaction buffer that contained 20mM HEPES, 30mM NaCl, 1mM ZnCl, and 2mM $MgCl_2$, and 1uM U_{15} RNA oligonucleotide to a final volume of 20ul per reaction. The reactions were incubated at 30^{o}C and 1ul aliquots were taken at each time point and quenched with 1ul 400mM EDTA. 1ul of this mixture was than spotted on thin layer chromatography plates and air dried for 10 minutes. Plates were developed using 0.15M formic acid and 0.15M $LiCl_2$, dried using heat lamps, and then exposed to phosphoimager cassettes and scanned using a FUJI BAS-2500 analyzer. Signal quantification was performed using Multi-gauge software. All experiments were done in triplicate.

APPENDIX C

Yeast strains used

All strains are derivatives of *S. cerevisiae* strain MBS (*MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100* [rho +] L-o, M-o) (Iizuka and Sarnow, 1997). NA19 (*upf2::HIS)* (Amrani et al., 2004) and NA24 (*upf3::*HIS) (Ghosh et al., 2010) were

previously described. NA35 (*upf3::HIS3, upf1::URA2*), NA38 (*upf1::HIS3, upf2::URA3*), NA39 (*upf3::HIS3, upf2::URA3*), and NA44 (*upf3::HIS3, upf2::URA3, upf1::LEU2*) were constructed as previously described (Guthrie and Fink, 1991). The *HA-UPF2* plasmid was constructed by ligating a 1.2 kb PCR fragment amplified from the region upstream of the *UPF2* gene to an oligonucleotide containing the 3xHA sequence and a 4kb fragment of the *UPF2* ORF and downstream sequences to the pRS315 plasmid. Similarly, a 400nt PCR fragment amplified from the region upstream of the *UPF3* gene was ligated to an oligonucleotide containing the 3xHA sequence and a 2kb fragment of the *UPF3* ORF and downstream sequences to the pRS315 plasmid. Corresponding *HA-UPF2* and *HA-UPF3* sequences were then cloned into YCplac112.

Plasmids and ribosomal subunit purification

Cells were transformed with *HA-UPF2* or *HA-UPF3* as previously described (Schiestl and Gietz, 1989). Ribosomal subunit purification was performed using the methods described by Min *et al.* (Min et al., 2013). All subunit purifications were done in duplicate from independent isolates. Aliquots from subunit purifications were analyzed by western blotting.

References

Aglietti, R.A., Floor, S.N., McClendon, C.L., Jacobson, M.P., and Gross, J.D. (2013). Active site conformational dynamics are coupled to catalysis in the mRNA decapping enzyme Dcp2. Structure *21*, 1571-1580.

Albert, T.K., Lemaire, M., van Berkum, N.L., Gentz, R., Collart, M.A., and Timmers, H.T. (2000). Isolation and characterization of human orthologs of yeast CCR4-NOT complex subunits. Nucleic Acids Res *28*, 809-817.

Alkalaeva, E.Z., Pisarev, A.V., Frolova, L.Y., Kisselev, L.L., and Pestova, T.V. (2006). In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. Cell *125*, 1125-1136.

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999). The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases. Genes Dev *13*, 2148-2158.

Altamura, N., Groudinsky, O., Dujardin, G., and Slonimski, P.P. (1992). *NAM7* nuclear gene encodes a novel member of a family of helicases with a Zn-ligand motif and is involved in mitochondrial functions in *Saccharomyces cerevisiae*. J Mol Biol *224*, 575-587.

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A *faux* 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature *432*, 112-118.

Amrani, N., Ghosh, S., Mangus, D.A., and Jacobson, A. (2008). Translation factors promote the formation of two states of the closed-loop mRNP. Nature *453*, 1276-1280.

Anders, K.R., Grimson, A., and Anderson, P. (2003). SMG-5, required for C.elegans nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. EMBO J *22*, 641-650.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Andersen, K.R., Jonstrup, A.T., Van, L.B., and Brodersen, D.E. (2009). The activity and selectivity of fission yeast Pop2p are affected by a high affinity for Zn2+ and Mn2+ in the active site. RNA *15*, 850-861.

Anderson, J.S., and Parker, R.P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the *SKI2* DEVH box protein and 3' to 5' exonucleases of the exosome complex. EMBO J *17*, 1497-1506.

Applequist, S.E., Selg, M., Raman, C., and Jack, H.M. (1997). Cloning and characterization of HUPF1, a human homolog of the *Saccharomyces cerevisiae* nonsense mRNA-reducing UPF1 protein. Nucleic Acids Res *25*, 814-821.

Araki, Y., Takahashi, S., Kobayashi, T., Kajiho, H., Hoshino, S., and Katada, T. (2001). Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast. EMBO J *20*, 4684-4693.

Aravind, L., and Koonin, E.V. (2000). Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. Genome Res *10*, 1172-1184.

Arribas-Layton, M., Wu, D., Lykke-Andersen, J., and Song, H. (2013). Structural and functional control of the eukaryotic mRNA decapping machinery. Biochim Biophys Acta *1829*, 580-589.

Arribere, J.A., and Gilbert, W.V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Res *23*, 977-987.

Atkin, A.L., Altamura, N., Leeds, P., and Culbertson, M.R. (1995). The majority of yeast UPF1 co-localizes with polyribosomes in the cytoplasm. Mol Biol Cell *6*, 611-625.

Atkin, A.L., Schenkman, L.R., Eastham, M., Dahlseid, J.N., Lelivelt, M.J., and Culbertson, M.R. (1997). Relationship between yeast polyribosomes and Upf proteins required for nonsense mRNA decay. J Biol Chem *272*, 22163-22172.

Avery, P., Vicente-Crespo, M., Francis, D., Nashchekina, O., Alonso, C.R., and Palacios, I.M. (2011). Drosophila Upf1 and Upf2 loss of function inhibits cell growth and causes animal death in a Upf3-independent manner. RNA *17*, 624-638.

Aviv, T., Lin, Z., Ben-Ari, G., Smibert, C.A., and Sicheri, F. (2006). Sequence-specific recognition of RNA hairpins by the SAM domain of Vts1p. Nat Struct Mol Biol *13*, 168-176.

Badis, G., Saveanu, C., Fromont-Racine, M., and Jacquier, A. (2004). Targeted mRNA degradation by deadenylation-independent decapping. Mol Cell *15*, 5-15.

Bai, Y., Salvadore, C., Chiang, Y.C., Collart, M.A., Liu, H.Y., and Denis, C.L. (1999). The CCR4 and CAF1 proteins of the CCR4-NOT complex are physically and functionally separated from NOT2, NOT4, and NOT5. Mol Cell Biol *19*, 6642-6651.

Balistreri, G., Horvath, P., Schweingruber, C., Zund, D., McInerney, G., Merits, A., Muhlemann, O., Azzalin, C., and Helenius, A. (2014). The host nonsense-mediated mRNA decay pathway restricts Mammalian RNA virus replication. Cell host & microbe *16*, 403-411.

Ball, L.J., Jarchau, T., Oschkinat, H., and Walter, U. (2002). EVH1 domains: structure, function and interactions. FEBS Lett *513*, 45-52.

Bashkirov, V.I., Scherthan, H., Solinger, J.A., Buerstedde, J.M., and Heyer, W.D. (1997). A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. J Cell Biol *136*, 761-773.

Basquin, J., Roudko, V.V., Rode, M., Basquin, C., Seraphin, B., and Conti, E. (2012). Architecture of the nuclease module of the yeast Ccr4-not complex: the Not1-Caf1-Ccr4 interaction. Mol Cell *48*, 207-218.

Becker, A.H., Oh, E., Weissman, J.S., Kramer, G., and Bukau, B. (2013). Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. Nature protocols *8*, 2212-2239.

Becker, T., Armache, J.P., Jarasch, A., Anger, A.M., Villa, E., Sieber, H., Motaal, B.A., Mielke, T., Berninghausen, O., and Beckmann, R. (2011). Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. Nat Struct Mol Biol *18*, 715-720.

Beelman, C.A., Stevens, A., Caponigro, G., LaGrandeur, T.E., Hatfield, L., Fortner, D.M., and Parker, R. (1996). An essential component of the decapping enzyme required for normal rates of mRNA turnover. Nature *382*, 642-646.

Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., and Izaurralde, E. (2007a). A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. EMBO J *26*, 1591-1601.

Behm-Ansmant, I., Kashima, I., Rehwinkel, J., Sauliere, J., Wittkopp, N., and Izaurralde, E. (2007b). mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. FEBS Lett *581*, 2845-2853.

Beilharz, T.H., and Preiss, T. (2007). Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. RNA *13*, 982-997.

Bertram, G., Innes, S., Minella, O., Richardson, J., and Stansfield, I. (2001). Endless possibilities: translation termination and stop codon recognition. Microbiology *147*, 255-269.

Bessman, M.J., Frick, D.N., and O'Handley, S.F. (1996). The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. J Biol Chem *271*, 25059-25062.

Bhaskar, V., Roudko, V., Basquin, J., Sharma, K., Urlaub, H., Seraphin, B., and Conti, E. (2013). Structure and RNA-binding properties of the Not1-Not2-Not5 module of the yeast Ccr4-Not complex. Nat Struct Mol Biol *20*, 1281-1288.

Bhattacharya, A., Czaplinski, K., Trifillis, P., He, F., Jacobson, A., and Peltz, S.W. (2000). Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. RNA *6*, 1226-1235.

Blanchet, S., Cornu, D., Argentini, M., and Namy, O. (2014). New insights into the incorporation of natural suppressor tRNAs at stop codons in Saccharomyces cerevisiae. Nucleic Acids Res *42*, 10061-10072.

Boeck, R., Tarun, S., Jr., Rieger, M., Deardorff, J.A., Muller-Auer, S., and Sachs, A.B. (1996). The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. J Biol Chem *271*, 432-438.

Boehm, V., Haberman, N., Ottens, F., Ule, J., and Gehring, N.H. (2014). 3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. Cell reports *9*, 555-568.

Bonnerot, C., Boeck, R., and Lapeyre, B. (2000). The two proteins Pat1p (Mrt1p) and Spb8p interact in vivo, are required for mRNA decay, and are functionally linked to Pab1p. Mol Cell Biol *20*, 5939-5946.

Bono, F., and Gehring, N.H. (2011). Assembly, disassembly and recycling: the dynamics of exon junction complexes. RNA Biol *8*, 24-29.

Borja, M.S., Piotukh, K., Freund, C., and Gross, J.D. (2011). Dcp1 links coactivators of mRNA decapping to Dcp2 by proline recognition. RNA *17*, 278-290.

Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Seraphin, B. (2000). A Sm-like protein complex that participates in mRNA degradation. EMBO J *19*, 1661-1671.

Bradley, M.E., Bagriantsev, S., Vishveshwara, N., and Liebman, S.W. (2003). Guanidine reduces stop codon read-through caused by missense mutations in SUP35 or SUP45. Yeast *20*, 625-632.

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

Briggs, M.W., Burkard, K.T., and Butler, J.S. (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.

Brown, A., Shao, S., Murray, J., Hegde, R.S., and Ramakrishnan, V. (2015). Structural basis for stop codon recognition in eukaryotes. Nature *524*, 493-496.

Brown, C.E., and Sachs, A.B. (1998). Poly(A) tail length control in *Saccharomyces cerevisiae* occurs by message-specific deadenylation. Mol Cell Biol *18*, 6548-6559.

Brown, C.E., Tarun, S.Z., Jr., Boeck, R., and Sachs, A.B. (1996). *PAN3* encodes a subunit of the Pab1p-dependent poly(A) nuclease in *Saccharomyces cerevisiae*. Mol Cell Biol *16*, 5744-5753.

Brule, C.E., and Grayhack, E.J. (2017). Synonymous Codons: Choose Wisely for Expression. Trends Genet *33*, 283-297.

Buchwald, G., Ebert, J., Basquin, C., Sauliere, J., Jayachandran, U., Bono, F., Le Hir, H., and Conti, E. (2010). Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. Proc Natl Acad Sci U S A *107*, 10050-10055.

Buhler, M., Steiner, S., Mohn, F., Paillusson, A., and Muhlemann, O. (2006). EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length. Nat Struct Mol Biol *13*, 462-464.

Caponigro, G., and Parker, R. (1995). Multiple functions for the poly(A)-binding protein in mRNA decapping and deadenylation in yeast. Genes Dev *9*, 2421-2432.

Carroll, J.S., Munchel, S.E., and Weis, K. (2011). The DExD/H box ATPase Dhh1 functions in translational repression, mRNA decay, and processing body dynamics. J Cell Biol *194*, 527-537.

Celik, A., Baker, R., He, F., and Jacobson, A. (2017). High-resolution profiling of NMD targets in yeast reveals translational fidelity as a basis for substrate selection. RNA *23*, 735-748.

Celik, A., Kervestin, S., and Jacobson, A. (2015). NMD: At the crossroads between translation termination and ribosome recycling. Biochimie *114*, 2-9.

Chakrabarti, S., Bonneau, F., Schussler, S., Eppinger, E., and Conti, E. (2014). Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and SMG6. Nucleic Acids Res *42*, 9447-9460.

Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., and Conti, E. (2011). Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. Mol Cell *41*, 693-703.

Chamieh, H., Ballut, L., Bonneau, F., and Le Hir, H. (2008). NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat Struct Mol Biol *15*, 85-93.

Chang, J.H., Xiang, S., Xiang, K., Manley, J.L., and Tong, L. (2011). Structural and biochemical studies of the 5'-->3' exoribonuclease Xrn1. Nat Struct Mol Biol *18*, 270-276.

Charenton, C., Taverniti, V., Gaudon-Plesse, C., Back, R., Seraphin, B., and Graille, M. (2016). Structure of the active form of Dcp1-Dcp2 decapping enzyme bound to m7GDP and its Edc3 activator. Nat Struct Mol Biol *23*, 982-986.

Chazal, P.E., Daguenet, E., Wendling, C., Ulryck, N., Tomasetto, C., Sargueil, B., and Le Hir, H. (2013). EJC core component MLN51 interacts with eIF3 and activates translation. Proc Natl Acad Sci U S A *110*, 5903-5908.

Chen, J., Rappsilber, J., Chiang, Y.C., Russell, P., Mann, M., and Denis, C.L. (2001). Purification and characterization of the 1.0 MDa CCR4-NOT complex identifies two novel components of the complex. J Mol Biol *314*, 683-694.

Chen, L., Muhlrad, D., Hauryliuk, V., Cheng, Z., Lim, M.K., Shyp, V., Parker, R., and Song, H. (2010). Structure of the Dom34-Hbs1 complex and implications for no-go decay. Nat Struct Mol Biol *17*, 1233-1240.

Cheng, Z., Saito, K., Pisarev, A.V., Wada, M., Pisareva, V.P., Pestova, T.V., Gajda, M., Round, A., Kong, C., Lim, M.*, et al.* (2009). Structural insights into eRF3 and stop codon recognition by eRF1. Genes Dev *23*, 1106-1118.

Chowdhury, A., Kalurupalle, S., and Tharun, S. (2014). Pat1 contributes to the RNA binding activity of the Lsm1-7-Pat1 complex. RNA *20*, 1465-1475.

Chowdhury, A., Raju, K.K., Kalurupalle, S., and Tharun, S. (2012). Both Sm-domain and C-terminal extension of Lsm1 are important for the RNA-binding activity of the Lsm1-7-Pat1 complex. RNA *18*, 936-944.

Chritton, J.J., and Wickens, M. (2011). A role for the poly(A)-binding protein Pab1p in PUF protein-mediated repression. J Biol Chem *286*, 33268-33278.

Chung, B.Y., Hardcastle, T.J., Jones, J.D., Irigoyen, N., Firth, A.E., Baulcombe, D.C., and Brierley, I. (2015). The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. RNA *21*, 1731-1745.

Clerici, M., Deniaud, A., Boehm, V., Gehring, N.H., Schaffitzel, C., and Cusack, S. (2014). Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2. Nucleic Acids Res *42*, 2673-2686.

Clerici, M., Mourao, A., Gutsche, I., Gehring, N.H., Hentze, M.W., Kulozik, A., Kadlec, J., Sattler, M., and Cusack, S. (2009). Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. EMBO J *28*, 2293-2306.

Cohen, L.S., Mikhli, C., Jiao, X., Kiledjian, M., Kunkel, G., and Davis, R.E. (2005). Dcp2 Decaps m2,2,7GpppN-capped RNAs, and its activity is sequence and context dependent. Mol Cell Biol *25*, 8779-8791.

Colak, D., Ji, S.J., Porse, B.T., and Jaffrey, S.R. (2013). Regulation of axon guidance by compartmentalized nonsense-mediated mRNA decay. Cell *153*, 1252-1265.

Coller, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. Cell *122*, 875-886.

Coller, J.M., Gray, N.K., and Wickens, M.P. (1998). mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. Genes Dev *12*, 3226-3235.

Coller, J.M., Tucker, M., Sheth, U., Valencia-Sanchez, M.A., and Parker, R. (2001). The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. RNA *7*, 1717-1727.

Colombo, M., Karousis, E.D., Bourquin, J., Bruggmann, R., and Muhlemann, O. (2017). Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. RNA *23*, 189-201.

Cosson, B., Couturier, A., Chabelskaya, S., Kiktev, D., Inge-Vechtomov, S., Philippe, M., and Zhouravleva, G. (2002). Poly(A)-binding protein acts in translation termination via eukaryotic release factor 3 interaction and does not influence [PSI(+)] propagation. Mol Cell Biol *22*, 3301-3315.

Cui, Y., Hagan, K.W., Zhang, S., and Peltz, S.W. (1995). Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. Genes Dev *9*, 423-436.

Czaplinski, K., Ruiz-Echevarria, M.J., Paushkin, S.V., Han, X., Weng, Y., Perlick, H.A., Dietz, H.C., Ter-Avanesyan, M.D., and Peltz, S.W. (1998). The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. Genes Dev *12*, 1665-1677.

Czaplinski, K., Weng, Y., Hagan, K.W., and Peltz, S.W. (1995). Purification and characterization of the Upf1 protein: a factor involved in translation and mRNA degradation. RNA *1*, 610-623.

D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. Trends Biochem Sci *28*, 655-662.

Daugeron, M.C., Mauxion, F., and Seraphin, B. (2001). The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. Nucleic Acids Res *29*, 2448-2455.

de Pinto, B., Lippolis, R., Castaldo, R., and Altamura, N. (2004). Overexpression of Upf1p compensates for mitochondrial splicing deficiency independently of its role in mRNA surveillance. Mol Microbiol *51*, 1129-1142.

Decker, C.J., and Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. Genes Dev *7*, 1632-1643.

des Georges, A., Hashem, Y., Unbehaun, A., Grassucci, R.A., Taylor, D., Hellen, C.U., Pestova, T.V., and Frank, J. (2014). Structure of the mammalian ribosomal pre-termination complex associated with eRF1.eRF3.GDPNP. Nucleic Acids Res *42*, 3409-3418.

Dlakic, M. (2000). Functionally unrelated signalling proteins contain a fold similar to Mg2+-dependent endonucleases. Trends Biochem Sci *25*, 272-273.

Doma, M.K., and Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature *440*, 561-564.

Dong, S., Jacobson, A., and He, F. (2010). Degradation of YRA1 Pre-mRNA in the cytoplasm requires translational repression, multiple modular intronic elements, Edc3p, and Mex67p. PLoS Biol *8*, e1000360.

Dong, S., Li, C., Zenklusen, D., Singer, R.H., Jacobson, A., and He, F. (2007). YRA1 autoregulation requires nuclear export and cytoplasmic Edc3p-mediated degradation of its pre-mRNA. Mol Cell *25*, 559-573.

Dunckley, T., and Parker, R. (1999). The DCP2 protein is required for mRNA decapping in *Saccharomyces cerevisiae* and contains a functional MutT motif. EMBO J *18*, 5411-5422.

Dunckley, T., and Parker, R. (2001). Yeast mRNA decapping enzyme. Methods Enzymol *342*, 226-233.

Dunckley, T., Tucker, M., and Parker, R. (2001). Two related proteins, Edc1p and Edc2p, stimulate mRNA decapping in *Saccharomyces cerevisiae*. Genetics *157*, 27-37.

Dupressoir, A., Morel, A.P., Barbot, W., Loireau, M.P., Corbo, L., and Heidmann, T. (2001). Identification of four families of yCCR4- and Mg2+-dependent endonuclease-related proteins in higher eukaryotes, and characterization of orthologs of yCCR4 with a conserved leucine-rich repeat essential for hCAF1/hPOP2 binding. BMC Genomics *2*, 9.

Durand, S., Franks, T.M., and Lykke-Andersen, J. (2016). Hyperphosphorylation amplifies UPF1 activity to resolve stalls in nonsense-mediated mRNA decay. Nat Commun *7*, 12434.

Durand, S., and Lykke-Andersen, J. (2013). Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. Nat Struct Mol Biol *20*, 702-709.

Dziembowski, A., Lorentzen, E., Conti, E., and Seraphin, B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. Nat Struct Mol Biol *14*, 15-22.

Eberle, A.B., Lykke-Andersen, S., Muhlemann, O., and Jensen, T.H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat Struct Mol Biol *16*, 49-55.

Eberle, A.B., Stalder, L., Mathys, H., Orozco, R.Z., and Muhlemann, O. (2008). Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. PLoS Biol *6*, e92.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res *30*, 207-210.

Fiorini, F., Bonneau, F., and Le Hir, H. (2012). Biochemical characterization of the RNA helicase UPF1 involved in nonsense-mediated mRNA decay. Methods Enzymol *511*, 255-274.

Fiorini, F., Boudvillain, M., and Le Hir, H. (2013). Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. Nucleic Acids Res *41*, 2404-2415.

Fischer, N., and Weis, K. (2002). The DEAD box protein Dhh1 stimulates the decapping enzyme Dcp1. EMBO J *21*, 2788-2797.

Franks, T.M., Singh, G., and Lykke-Andersen, J. (2010). Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense- mediated mRNA decay. Cell *143*, 938-950.

Fraser, C.S., Berry, K.E., Hershey, J.W., and Doudna, J.A. (2007). eIF3j is located in the decoding center of the human 40S ribosomal subunit. Mol Cell *26*, 811-819.

Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrerio, A.L., Parker, R., and Dietz, H.C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science *295*, 2258-2261.

Frolova, L., Le Goff, X., Zhouravleva, G., Davydova, E., Philippe, M., and Kisselev, L. (1996). Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase. RNA *2*, 334-341.

Frolova, L., Seit-Nebi, A., and Kisselev, L. (2002). Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. RNA *8*, 129-136.

Frolova, L.Y., Tsivkovskii, R.Y., Sivolobova, G.F., Oparina, N.Y., Serpinsky, O.I., Blinov, V.M., Tatkov, S.I., and Kisselev, L.L. (1999). Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. RNA *5*, 1014-1020.

Fromm, S.A., Truffault, V., Kamenz, J., Braun, J.E., Hoffmann, N.A., Izaurralde, E., and Sprangers, R. (2012). The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. EMBO J *31*, 279-290.

Fromont-Racine, M., Mayes, A.E., Brunet-Simon, A., Rain, J.C., Colley, A., Dix, I., Decourty, L., Joly, N., Ricard, F., Beggs, J.D.*, et al.* (2000). Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. Yeast *17*, 95-110.

Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., and Conti, E. (2005). SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. Mol Cell *17*, 537-547.

Gaba, A., Jacobson, A., and Sachs, M.S. (2005). Ribosome occupancy of the yeast *CPA1* upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. Mol Cell *20*, 449-460.

Gallie, D.R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev *5*, 2108-2116.

Gallie, D.R., and Tanguay, R. (1994). Poly(A) binds to initiation factors and increases cap-dependent translation *in vitro*. J Biol Chem *269*, 17166-17173.

Gamble, C.E., Brule, C.E., Dean, K.M., Fields, S., and Grayhack, E.J. (2016). Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. Cell *166*, 679-690.

Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. EMBO J *22*, 3960-3970.

Ge, Z., Quek, B.L., Beemon, K.L., and Hogg, J.R. (2016). Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. eLife *5*.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Mol Cell *20*, 65-75.

Gehring, N.H., Lamprinaki, S., Kulozik, A.E., and Hentze, M.W. (2009). Disassembly of exon junction complexes by PYM. Cell *137*, 536-548.

Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W., and Kulozik, A.E. (2003). Y14 and hUpf3b form an NMD-activating complex. Mol Cell *11*, 939-949.

Gerber, A.P., Herschlag, D., and Brown, P.O. (2004). Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol *2*, E79.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Ghosh, S., Ganesan, R., Amrani, N., and Jacobson, A. (2010). Translational competence of ribosomes released from a premature termination codon is modulated by NMD factors. RNA *16*, 1832-1847.

Giorgi, C., Yeo, G.W., Stone, M.E., Katz, D.B., Burge, C., Turrigiano, G., and Moore, M.J. (2007). The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. Cell *130*, 179-191.

Glavan, F., Behm-Ansmant, I., Izaurralde, E., and Conti, E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. EMBO J *25*, 5117-5125.

Gloggnitzer, J., Akimcheva, S., Srinivasan, A., Kusenda, B., Riehs, N., Stampfl, H., Bautor, J., Dekrout, B., Jonak, C., Jimenez-Gomez, J.M.*, et al.* (2014). Nonsense-mediated mRNA decay modulates immune receptor levels to regulate plant antibacterial defense. Cell host & microbe *16*, 376-390.

Goldstrohm, A.C., Hook, B.A., Seay, D.J., and Wickens, M. (2006). PUF proteins bind Pop2p to regulate messenger RNAs. Nat Struct Mol Biol *13*, 533-539.

Goldstrohm, A.C., Seay, D.J., Hook, B.A., and Wickens, M. (2007). PUF protein-mediated deadenylation is catalyzed by Ccr4p. J Biol Chem *282*, 109-114.

Gong, C., Kim, Y.K., Woeller, C.F., Tang, Y., and Maquat, L.E. (2009). SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNAs. Genes Dev *23*, 54-66.

Gonzalez, C.I., Ruiz-Echevarria, M.J., Vasudevan, S., Henry, M.F., and Peltz, S.W. (2000). The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. Mol Cell *5*, 489-499.

Graille, M., Chaillet, M., and van Tilbeurgh, H. (2008). Structure of yeast Dom34: a protein related to translation termination factor Erf1 and involved in No-Go decay. J Biol Chem *283*, 7145-7154.

Gray, N.K., Coller, J.M., Dickson, K.S., and Wickens, M. (2000). Multiple portions of poly(A)-binding protein stimulate translation *in vivo*. EMBO J *19*, 4723-4733.

Gregersen, L.H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., and Landthaler, M. (2014). MOV10 Is a 5' to 3' RNA Helicase Contributing to UPF1 mRNA Target Degradation by Translocation along 3' UTRs. Molecular Cell *54*, 573-585.

Grimson, A., O'Connor, S., Newman, C.L., and Anderson, P. (2004). SMG-1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in Caenorhabditis elegans. Mol Cell Biol *24*, 7483-7490.

Grudzien-Nogalska, E., and Kiledjian, M. (2017). New insights into decapping enzymes and selective mRNA decay. Wiley Interdiscip Rev RNA *8*.

Guthrie, C., and Fink, G.R. (1991). Methods in Enzymology: Molecular Biology of *Saccharomyces cerevisiae*., Vol 194 (NY: Academic Press).

Guydosh, N.R., and Green, R. (2014). Dom34 rescues ribosomes in 3' untranslated regions. Cell *156*, 950-962.

Guydosh, N.R., and Green, R. (2017). Translation of poly(A) tails leads to precise mRNA cleavage. RNA *23*, 749-761.

Haas, G., Braun, J.E., Igreja, C., Tritschler, F., Nishihara, T., and Izaurralde, E. (2010). HPat provides a link between deadenylation and decapping in metazoa. J Cell Biol *189*, 289-302.

Harigaya, Y., Jones, B.N., Muhlrad, D., Gross, J.D., and Parker, R. (2010). Identification and analysis of the interaction between Edc3 and Dcp2 in Saccharomyces cerevisiae. Mol Cell Biol *30*, 1446-1456.

Harigaya, Y., and Parker, R. (2012). Global analysis of mRNA decay intermediates in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *109*, 11764-11769.

Hatin, I., Fabret, C., Rousset, J.P., and Namy, O. (2009). Molecular dissection of translation termination mechanism identifies two new critical regions in eRF1. Nucleic Acids Res *37*, 1789-1798.

Hauer, C., Sieber, J., Schwarzl, T., Hollerer, I., Curk, T., Alleaume, A.M., Hentze, M.W., and Kulozik, A.E. (2016). Exon Junction Complexes Show a Distributional Bias toward Alternatively Spliced mRNAs and against mRNAs Coding for Ribosomal Proteins. Cell reports *16*, 1588-1603.

He, F., Brown, A.H., and Jacobson, A. (1997). Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. Mol Cell Biol *17*, 1580-1594.

He, F., Ganesan, R., and Jacobson, A. (2013). Intra- and intermolecular regulatory interactions in Upf1, the RNA helicase central to nonsense-mediated mRNA decay in yeast. Mol Cell Biol *33*, 4672-4684.

He, F., and Jacobson, A. (1995). Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. Genes Dev *9*, 437-454.

He, F., and Jacobson, A. (2001). Upf1p, Nmd2p, and Upf3p regulate the decapping and exonucleolytic degradation of both nonsense-containing mRNAs and wild-type mRNAs. Mol Cell Biol *21*, 1515-1530.

He, F., and Jacobson, A. (2015a). Control of mRNA decapping by positive and negative regulatory elements in the Dcp2 C-terminal domain. RNA *21*, 1633-1647.

He, F., and Jacobson, A. (2015b). Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. Annu Rev Genet *49*, 339-366.

He, F., Li, C., Roy, B., and Jacobson, A. (2014). Yeast Edc3 targets RPS28B mRNA for decapping by binding to a 3' untranslated region decay-inducing regulatory element. Mol Cell Biol *34*, 1438-1451.

He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., and Jacobson, A. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. Mol Cell *12*, 1439-1452.

He, F., Peltz, S.W., Donahue, J.L., Rosbash, M., and Jacobson, A. (1993). Stabilization and ribosome association of unspliced pre-mRNAs in a yeast *upf1*- mutant. Proc Natl Acad Sci U S A *90*, 7034-7038.

Herrick, D., Parker, R., and Jacobson, A. (1990). Identification and comparison of stable and unstable mRNAs in Saccharomyces cerevisiae. Mol Cell Biol *10*, 2269-2284.

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

Hogg, J.R., and Goff, S.P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. Cell *143*, 379-389.

Holmes, L.E., Campbell, S.G., De Long, S.K., Sachs, A.B., and Ashe, M.P. (2004). Loss of translational control in yeast compromised for the major mRNA decay pathway. Mol Cell Biol *24*, 2998-3010.

Hook, B.A., Goldstrohm, A.C., Seay, D.J., and Wickens, M. (2007). Two yeast PUF proteins negatively regulate a single mRNA. J Biol Chem *282*, 15430-15438.

Hosoda, N., Kim, Y.K., Lejeune, F., and Maquat, L.E. (2005). CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. Nat Struct Mol Biol *12*, 893-901.

Hu, W., Petzold, C., Coller, J., and Baker, K.E. (2010). Nonsense-mediated mRNA decapping occurs on polyribosomes in Saccharomyces cerevisiae. Nat Struct Mol Biol *17*, 244-247.

Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E., and Coller, J. (2009). Co-translational mRNA decay in Saccharomyces cerevisiae. Nature *461*, 225-229.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686-691.

Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. RNA *14*, 2609-2617.

Hurt, J.A., Robertson, A.D., and Burge, C.B. (2013). Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. Genome Res *23*, 1636-1650.

Iizuka, N., and Sarnow, P. (1997). Translation-competent extracts from *Saccharomyces cerevisiae*: effects of L-A RNA, 5' cap, and 3' poly(A) tail on translational efficiency of mRNAs. Methods *11*, 353-360.

Inada, T., and Aiba, H. (2005). Translation of aberrant mRNAs lacking a termination codon or with a shortened 3'-UTR is repressed after initiation in yeast. EMBO J *24*, 1584-1595.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2013). Genome-wide annotation and quantitation of translation by ribosome profiling. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 4*, Unit 4 18.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L.E. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell *106*, 607-617.

Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W., and Maquat, L.E. (2008). Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. Cell *133*, 314-327.

Ito-Harashima, S., Kuroha, K., Tatematsu, T., and Inada, T. (2007). Translation of the poly(A) tail plays crucial roles in nonstop mRNA surveillance via translation repression and protein destabilization by proteasome in yeast. Genes Dev *21*, 519-524.

Ivanov, P.V., Gehring, N.H., Kunz, J.B., Hentze, M.W., and Kulozik, A.E. (2008). Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. EMBO J *27*, 736-747.

Ivanova, E.V., Kolosov, P.M., Birdsall, B., Kelly, G., Pastore, A., Kisselev, L.L., and Polshakov, V.I. (2007). Eukaryotic class 1 translation termination factor eRF1--the NMR structure and dynamics of the middle domain involved in triggering ribosome-dependent peptidyl-tRNA hydrolysis. Febs J *274*, 4223-4237.

Iwasaki, S., Takeda, A., Motose, H., and Watanabe, Y. (2007). Characterization of Arabidopsis decapping proteins AtDCP1 and AtDCP2, which are essential for post-embryonic development. FEBS Lett *581*, 2455-2459.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2012). Termination and post-termination events in eukaryotic translation. Adv Protein Chem Struct Biol *86*, 45-93.

Jacobson, A. (1996). Poly(A) metabolism and translation: the closed loop model. In Translational Control, J.W. Hershey, M.B. Mathews, and N. Sonenberg, eds. (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press), pp. 451-480.

Jaillon, O., Bouhouche, K., Gout, J.F., Aury, J.M., Noel, B., Saudemont, B., Nowacki, M., Serrano, V., Porcel, B.M., Segurens, B.*, et al.* (2008). Translational control of intron splicing in eukaryotes. Nature *451*, 359-362.

Januszyk, K., Liu, Q., and Lima, C.D. (2011). Activities of human RRP6 and structure of the human RRP6 catalytic domain. RNA *17*, 1566-1577.

Jin, H., Kelley, A.C., Loakes, D., and Ramakrishnan, V. (2010). Structure of the 70S ribosome bound to release factor 2 and a substrate analog provides insights into catalysis of peptide release. Proc Natl Acad Sci U S A *107*, 8593-8598.

Jinek, M., Coyle, S.M., and Doudna, J.A. (2011). Coupled 5' nucleotide recognition and processivity in Xrn1-mediated mRNA decay. Mol Cell *41*, 600-608.

Johansson, M.J., He, F., Spatrick, P., Li, C., and Jacobson, A. (2007). Association of yeast Upf1p with direct substrates of the NMD pathway. Proc Natl Acad Sci U S A *104*, 20872-20877.

Johansson, M.J., and Jacobson, A. (2010). Nonsense-mediated mRNA decay maintains translational fidelity by limiting magnesium uptake. Genes Dev *24*, 1491-1495.

Jonas, S., Christie, M., Peter, D., Bhandari, D., Loh, B., Huntzinger, E., Weichenrieder, O., and Izaurralde, E. (2014). An asymmetric PAN3 dimer recruits a single PAN2 exonuclease to mediate mRNA deadenylation and decay. Nat Struct Mol Biol *21*, 599-608.

Jonas, S., and Izaurralde, E. (2013). The role of disordered protein regions in the assembly of decapping complexes and RNP granules. Genes Dev *27*, 2628-2641.

Jonas, S., Weichenrieder, O., and Izaurralde, E. (2013). An unusual arrangement of two 14-3-3-like domains in the SMG5-SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. Genes Dev *27*, 211-225.

Jonstrup, A.T., Andersen, K.R., Van, L.B., and Brodersen, D.E. (2007). The 1.4-A crystal structure of the S. pombe Pop2p deadenylase subunit unveils the configuration of an active enzyme. Nucleic Acids Res *35*, 3153-3164.

Kadlec, J., Izaurralde, E., and Cusack, S. (2004). The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. Nat Struct Mol Biol *11*, 330-337.

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

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev *20*, 355-367.

Kawashima, T., Douglass, S., Gabunilas, J., Pellegrini, M., and Chanfreau, G.F. (2014). Widespread use of non-productive alternative splice sites in Saccharomyces cerevisiae. PLoS Genet *10*, e1004249.

Kebaara, B.W., and Atkin, A.L. (2009). Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. Nucleic Acids Res *37*, 2771-2778.

Kertesz, S., Kerenyi, Z., Merai, Z., Bartos, I., Palfy, T., Barta, E., and Silhavy, D. (2006). Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. Nucleic Acids Res *34*, 6147-6157.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nat Rev Mol Cell Biol *13*, 700-712.

Kervestin, S., Li, C., Buckingham, R., and Jacobson, A. (2012). Testing the faux-UTR model for NMD: analysis of Upf1p and Pab1p competition for binding to eRF3/Sup35p. Biochimie *94*, 1560-1571.

Kim, V.N., Kataoka, N., and Dreyfuss, G. (2001). Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. Science *293*, 1832-1836.

Kiosze-Becker, K., Ori, A., Gerovac, M., Heuer, A., Nurenberg-Goloub, E., Rashid, U.J., Becker, T., Beckmann, R., Beck, M., and Tampe, R. (2016). Structure of the ribosome post-recycling complex probed by chemical cross-linking and mass spectrometry. Nat Commun *7*, 13248.

Kisselev, L.L., and Buckingham, R.H. (2000). Translational termination comes of age. Trends Biochem Sci *25*, 561-566.

Kolesnikova, O., Back, R., Graille, M., and Seraphin, B. (2013). Identification of the Rps28 binding motif from yeast Edc3 involved in the autoregulatory feedback loop controlling RPS28B mRNA decay. Nucleic Acids Res *41*, 9514-9523.

Kolosov, P., Frolova, L., Seit-Nebi, A., Dubovaya, V., Kononenko, A., Oparina, N., Justesen, J., Efimov, A., and Kisselev, L. (2005). Invariant amino acids essential for decoding function of polypeptide release factor eRF1. Nucleic Acids Res *33*, 6418-6425.

Kolupaeva, V.G., Lomakin, I.B., Pestova, T.V., and Hellen, C.U. (2003). Eukaryotic initiation factors 4G and 4A mediate conformational changes downstream of the initiation codon of the encephalomyocarditis virus internal ribosomal entry site. Mol Cell Biol *23*, 687-698.

Kononenko, A.V., Mitkevich, V.A., Dubovaya, V.I., Kolosov, P.M., Makarov, A.A., and Kisselev, L.L. (2008). Role of the individual domains of translation termination factor eRF1 in GTP binding to eRF3. Proteins *70*, 388-393.

Koonin, E.V. (1992). A new group of putative RNA helicases. Trends Biochem Sci *17*, 495-497.

Korostelev, A., Zhu, J., Asahara, H., and Noller, H.F. (2010). Recognition of the amber UAG stop codon by release factor RF1. EMBO J *29*, 2577-2585.

Kryuchkova, P., Grishin, A., Eliseev, B., Karyagina, A., Frolova, L., and Alkalaeva, E. (2013). Two-step model of stop codon recognition by eukaryotic release factor eRF1. Nucleic Acids Res *41*, 4573-4586.

Kunz, J.B., Neu-Yilik, G., Hentze, M.W., Kulozik, A.E., and Gehring, N.H. (2006). Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. RNA *12*, 1015-1022.

Kurihara, Y., Matsui, A., Hanada, K., Kawashima, M., Ishida, J., Morosawa, T., Tanaka, M., Kaminuma, E., Mochizuki, Y., Matsushima, A.*, et al.* (2009). Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. Proc Natl Acad Sci U S A *106*, 2453-2458.

Kurosaki, T., Li, W., Hoque, M., Popp, M.W., Ermolenko, D.N., Tian, B., and Maquat, L.E. (2014). A post-translational regulatory switch on UPF1 controls targeted mRNA degradation. Genes Dev *28*, 1900-1916.

Kvas, S., Gloor, G.B., and Brandl, C.J. (2012). Loss of nonsense mediated decay suppresses mutations in Saccharomyces cerevisiae TRA1. BMC Genet *13*, 19.

LaGrandeur, T.E., and Parker, R. (1998). Isolation and characterization of Dcp1p, the yeast mRNA decapping enzyme. EMBO J *17*, 1487-1496.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature *446*, 926-929.

Lasalde, C., Rivera, A.V., Leon, A.J., Gonzalez-Feliciano, J.A., Estrella, L.A., Rodriguez-Cruz, E.N., Correa, M.E., Cajigas, I.J., Bracho, D.P., Vega, I.E.*, et al.* (2014). Identification and functional analysis of novel phosphorylation sites in the RNA surveillance protein Upf1. Nucleic Acids Res *42*, 1916-1929.

Laurberg, M., Asahara, H., Korostelev, A., Zhu, J., Trakhanov, S., and Noller, H.F. (2008). Structural basis for translation termination on the 70S ribosome. Nature *454*, 852-857.

Le, H., Browning, K.S., and Gallie, D.R. (2000). The phosphorylation state of poly(A)-binding protein specifies its binding to poly(A) RNA and its interaction with eukaryotic initiation factor (eIF) 4F, eIFiso4F, and eIF4B. J Biol Chem *275*, 17452-17462.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J *19*, 6860-6869.

Lebreton, A., Tomecki, R., Dziembowski, A., and Seraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature *456*, 993-996.

Leeds, P., Peltz, S.W., Jacobson, A., and Culbertson, M.R. (1991). The product of the yeast *UPF1* gene is required for rapid turnover of mRNAs containing a premature translational termination codon. Genes Dev *5*, 2303-2314.

Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L.E. (2002). The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. EMBO J *21*, 3536-3545.

Lejeune, F., Li, X., and Maquat, L.E. (2003). Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. Mol Cell *12*, 675-687.

Lekomtsev, S., Kolosov, P., Bidou, L., Frolova, L., Rousset, J.P., and Kisselev, L. (2007). Different modes of stop codon restriction by the Stylonychia and Paramecium eRF1 translation termination factors. Proc Natl Acad Sci U S A *104*, 10824-10829.

Lelivelt, M.J., and Culbertson, M.R. (1999). Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. Mol Cell Biol *19*, 6710-6719.

Letzring, D.P., Dean, K.M., and Grayhack, E.J. (2010). Control of translation efficiency in yeast by codon-anticodon interactions. RNA *16*, 2516-2528.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A., and Dewey, C.N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics *26*, 493-500.

Li, S., and Wilkinson, M.F. (1998). Nonsense surveillance in lymphocytes? Immunity *8*, 135-141.

Liu, J.J., Niu, C.Y., Wu, Y., Tan, D., Wang, Y., Ye, M.D., Liu, Y., Zhao, W., Zhou, K., Liu, Q.S.*, et al.* (2016). CryoEM structure of yeast cytoplasmic exosome complex. Cell Res *26*, 822-837.

Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell *127*, 1223-1237.

Loh, B., Jonas, S., and Izaurralde, E. (2013). The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. Genes Dev *27*, 2125-2138.

Long, R.M., and McNally, M.T. (2003). mRNA decay: x (XRN1) marks the spot. Mol Cell *11*, 1126-1128.

Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. (2008). Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. Mol Cell *29*, 717-728.

Lou, C.H., Shao, A., Shum, E.Y., Espinoza, J.L., Huang, L., Karam, R., and Wilkinson, M.F. (2014). Posttranscriptional control of the stem cell and neurogenic programs by the nonsense-mediated RNA decay pathway. Cell reports *6*, 748-764.

Lowell, J.E., Rudner, D.Z., and Sachs, A.B. (1992). 3'-UTR-dependent deadenylation by the yeast poly(A) nuclease. Genes Dev *6*, 2088-2099.

Lykke-Andersen, J. (2002). Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. Mol Cell Biol *22*, 8114-8121.

Lykke-Andersen, J., and Bennett, E.J. (2014). Protecting the proteome: Eukaryotic cotranslational quality control pathways. J Cell Biol *204*, 467-476.

Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. (2000). Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. Cell *103*, 1121-1131.

Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. Science *293*, 1836-1839.

Lykke-Andersen, S., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A., and Jensen, T.H. (2014). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. Genes Dev *28*, 2498-2517.

Ma, X.M., Yoon, S.O., Richardson, C.J., Julich, K., and Blenis, J. (2008). SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. Cell *133*, 303-313.

Maderazo, A.B., Belk, J.P., He, F., and Jacobson, A. (2003). Nonsense-containing mRNAs that accumulate in the absence of a functional nonsense-mediated mRNA decay pathway are destabilized rapidly upon its restitution. Mol Cell Biol *23*, 842-851.

Maderazo, A.B., He, F., Mangus, D.A., and Jacobson, A. (2000). Upf1p control of nonsense mRNA translation is regulated by Nmd2p and Upf3p. Mol Cell Biol *20*, 4591-4603.

Maillet, L., and Collart, M.A. (2002). Interaction between Not1p, a component of the Ccr4-not complex, a global regulator of transcription, and Dhh1p, a putative RNA helicase. J Biol Chem *277*, 2835-2842.

Maillet, L., Tu, C., Hong, Y.K., Shuster, E.O., and Collart, M.A. (2000). The essential function of Not1 lies within the Ccr4-Not complex. J Mol Biol *303*, 131-143.

Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mRNA decay. eLife *4*.

Malvar, T., Biron, R.W., Kaback, D.B., and Denis, C.L. (1992). The CCR4 protein from Saccharomyces cerevisiae contains a leucine-rich repeat region which is required for its control of ADH2 gene expression. Genetics *132*, 951-962.

Mangus, D.A., Evans, M.C., and Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. Genome Biol *4*, 223.

Mangus, D.A., and Jacobson, A. (1999). Linking mRNA turnover and translation: assessing the polyribosomal association of mRNA decay factors and degradative intermediates. Methods *17*, 28-37.

Mantsyzov, A.B., Ivanova, E.V., Birdsall, B., Alkalaeva, E.Z., Kryuchkova, P.N., Kelly, G., Frolova, L.Y., and Polshakov, V.I. (2010). NMR solution structure and function of the C-terminal domain of eukaryotic class 1 polypeptide chain release factor. Febs J *277*, 2611-2627.

Maquat, L.E., Tarn, W.Y., and Isken, O. (2010). The pioneer round of translation: features and functions. Cell *142*, 368-374.

Maryati, M., Airhihen, B., and Winkler, G.S. (2015). The enzyme activities of Caf1 and Ccr4 are both required for deadenylation by the human Ccr4-Not nuclease module. Biochem J *469*, 169-176.

Matheisl, S., Berninghausen, O., Becker, T., and Beckmann, R. (2015). Structure of a human translation termination complex. Nucleic Acids Res *43*, 8615-8626.

McGlincy, N.J., and Smith, C.W. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? Trends Biochem Sci *33*, 385-393.

Medghalchi, S.M., Frischmeyer, P.A., Mendell, J.T., Kelly, A.G., Lawler, A.M., and Dietz, H.C. (2001). Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. Hum Mol Genet *10*, 99-105.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet *36*, 1073-1078.

Merkulova, T.I., Frolova, L.Y., Lazar, M., Camonis, J., and Kisselev, L.L. (1999). C-terminal domains of human translation termination factors eRF1 and eRF3 mediate their in vivo interaction. FEBS Lett *443*, 41-47.

Metzstein, M.M., and Krasnow, M.A. (2006). Functions of the nonsense-mediated mRNA decay pathway in Drosophila development. PLoS Genet *2*, e180.

Midtgaard, S.F., Assenholt, J., Jonstrup, A.T., Van, L.B., Jensen, T.H., and Brodersen, D.E. (2006). Structure of the nuclear exosome component Rrp6p reveals an interplay between the active site and the HRDC domain. Proc Natl Acad Sci U S A *103*, 11898-11903.

Min, E.E. (2015). PhD Thesis, Graduate School of Biomedical Sciences, Univ of Massachusetts, Worcester.

Min, E.E., Roy, B., Amrani, N., He, F., and Jacobson, A. (2013). Yeast Upf1 CH domain interacts with Rps26 of the 40S ribosomal subunit. RNA *19*, 1105-1115.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. Cell *91*, 457-466.

Mitchell, P., and Tollervey, D. (2003). An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. Mol Cell *11*, 1405-1413.

Mitchell, S.F., Walker, S.E., Algire, M.A., Park, E.H., Hinnebusch, A.G., and Lorsch, J.R. (2010). The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. Mol Cell *39*, 950-962.

Morrissey, J.P., Deardorff, J.A., Hebron, C., and Sachs, A.B. (1999). Decapping of stabilized, polyadenylated mRNA in yeast *pab1* mutants. Yeast *15*, 687-702.

Mugridge, J.S., Ziemniak, M., Jemielity, J., and Gross, J.D. (2016). Structural basis of mRNA-cap recognition by Dcp1-Dcp2. Nat Struct Mol Biol *23*, 987-994.

Muhlrad, D., and Parker, R. (1994). Premature translational termination triggers mRNA decapping. Nature *370*, 578-581.

Muhlrad, D., and Parker, R. (1999a). Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. RNA *5*, 1299-1307.

Muhlrad, D., and Parker, R. (1999b). Recognition of yeast mRNAs as "nonsense containing" leads to both inhibition of mRNA translation and mRNA degradation: implications for the control of mRNA decapping. Mol Biol Cell *10*, 3971-3978.

Munroe, D., and Jacobson, A. (1990). mRNA poly(A) tail, a 3' enhancer of translational initiation. Mol Cell Biol *10*, 3441-3455.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344-1349.

Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci *23*, 198-199.

Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E., and Ares, M., Jr. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev *21*, 708-718.

Nicholson, P., Joncourt, R., and Muhlemann, O. (2012). Analysis of nonsense-mediated mRNA decay in mammalian cells. Current protocols in cell biology / editorial board, Juan S Bonifacino [et al] *Chapter 27*, Unit27 24.

Nicholson, P., Josi, C., Kurosawa, H., Yamashita, A., and Muhlemann, O. (2014). A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. Nucleic Acids Res *42*, 9217-9235.

Nicholson, P., Yepiskoposyan, H., Metze, S., Zamudio Orozco, R., Kleinschmidt, N., and Muhlemann, O. (2010). Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. Cell Mol Life Sci *67*, 677-700.

Nissan, T., Rajyaguru, P., She, M., Song, H., and Parker, R. (2010). Decapping activators in Saccharomyces cerevisiae act by multiple mechanisms. Mol Cell *39*, 773-783.

Nott, A., Le Hir, H., and Moore, M.J. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. Genes Dev *18*, 210-222.

Nyiko, T., Sonkoly, B., Merai, Z., Benkovics, A.H., and Silhavy, D. (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. Plant Mol Biol *71*, 367-378.

Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K.R., Grimson, A., Hachiya, T., Hentze, M.W., Anderson, P., and Ohno, S. (2003). Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. Mol Cell *12*, 1187-1200.

Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., and Ohno, S. (2012). N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. Nucleic Acids Res *40*, 1251-1266.

Olivas, W., and Parker, R. (2000). The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. EMBO J *19*, 6602-6611.

Orban, T.I., and Izaurralde, E. (2005). Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. RNA *11*, 459-469.

Ozgur, S., Chekulaeva, M., and Stoecklin, G. (2010). Human Pat1b connects deadenylation with mRNA decapping and controls the assembly of processing bodies. Mol Cell Biol *30*, 4308-4323.

Page, A.M., Davis, K., Molineux, C., Kolodner, R.D., and Johnson, A.W. (1998). Mutational analysis of exoribonuclease I from *Saccharomyces cerevisiae*. Nucleic Acids Res *26*, 3707-3716.

Page, M.F., Carr, B., Anders, K.R., Grimson, A., and Anderson, P. (1999). *SMG-2* is a phosphorylated protein required for mRNA surveillance in *Caenorhabditis elegans* and related to Upf1p of yeast. Mol Cell Biol *19*, 5943-5951.

Parker, R. (2012). RNA degradation in Saccharomyces cerevisae. Genetics *191*, 671-702.

Parker, R., and Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. Nat Struct Mol Biol *11*, 121-127.

Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R., and Ramakrishnan, V. (2007). The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. Mol Cell *26*, 41-50.

Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol *20*, 237-243.

Peixeiro, I., Inacio, A., Barbosa, C., Silva, A.L., Liebhaber, S.A., and Romao, L. (2012). Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations. Nucleic Acids Res *40*, 1160-1173.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127-131.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. Cell *161*, 1400-1412.

Peltz, S.W., Brown, A.H., and Jacobson, A. (1993). mRNA destabilization triggered by premature translational termination depends on at least three *cis*-acting sequence elements and one *trans*-acting factor. Genes Dev *7*, 1737-1754.

Peltz, S.W., He, F., Welch, E., and Jacobson, A. (1994). Nonsense-mediated mRNA decay in yeast. Prog Nucleic Acid Res Mol Biol *47*, 271-298.

Peltz, S.W., Morsy, M., Welch, E.M., and Jacobson, A. (2013). Ataluren as an agent for therapeutic nonsense suppression. Annu Rev Med *64*, 407-425.

Perlick, H.A., Medghalchi, S.M., Spencer, F.A., Kendzior, R.J., Jr., and Dietz, H.C. (1996). Mammalian orthologues of a yeast regulator of nonsense transcript stability. Proc Natl Acad Sci U S A *93*, 10928-10932.

Petit, A.P., Wohlbold, L., Bawankar, P., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2012). The structural basis for the interaction between the CAF1 nuclease and the NOT1 scaffold of the human CCR4-NOT deadenylase complex. Nucleic Acids Res *40*, 11058-11072.

Piccirillo, C., Khanna, R., and Kiledjian, M. (2003). Functional characterization of the mammalian mRNA decapping enzyme hDcp2. RNA *9*, 1138-1147.

Pilkington, G.R., and Parker, R. (2008). Pat1 contains distinct functional domains that promote P-body assembly and activation of decapping. Mol Cell Biol *28*, 1298-1312.

Pillay, S., Li, Y., Wong, L.E., and Pervushin, K. (2016). Structural characterization of eRF1 mutants indicate a complex mechanism of stop codon recognition. Sci Rep *6*, 18644.

Pisarev, A.V., Hellen, C.U., and Pestova, T.V. (2007). Recycling of eukaryotic posttermination ribosomal complexes. Cell *131*, 286-299.

Pisarev, A.V., Skabkin, M.A., Pisareva, V.P., Skabkina, O.V., Rakotondrafara, A.M., Hentze, M.W., Hellen, C.U., and Pestova, T.V. (2010). The role of ABCE1 in eukaryotic posttermination ribosomal recycling. Mol Cell *37*, 196-210.

Pisareva, V.P., Hellen, C.U., and Pestova, T.V. (2007). Kinetic analysis of the interaction of guanine nucleotides with eukaryotic translation initiation factor eIF5B. Biochemistry *46*, 2622-2629.

Pisareva, V.P., Skabkin, M.A., Hellen, C.U., Pestova, T.V., and Pisarev, A.V. (2011). Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. EMBO J *30*, 1804-1817.

Ponting, C.P. (2000). Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. Trends Biochem Sci *25*, 423-426.

Preis, A., Heuer, A., Barrio-Garcia, C., Hauser, A., Eyler, D.E., Berninghausen, O., Green, R., Becker, T., and Beckmann, R. (2014). Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. Cell reports *8*, 59-65.

Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R.*, et al.* (2015). Codon optimality is a major determinant of mRNA stability. Cell *160*, 1111-1124.

Presnyak, V., and Coller, J. (2013). The DHH1/RCKp54 family of helicases: an ancient family of proteins that promote translational silencing. Biochim Biophys Acta *1829*, 817-823.

Pulak, R., and Anderson, P. (1993). mRNA surveillance by the *Caenorhabditis elegans smg* genes. Genes Dev *7*, 1885-1897.

Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A., and Ban, N. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. Science *331*, 730-736.

Radhakrishnan, A., Chen, Y.H., Martin, S., Alhusaini, N., Green, R., and Coller, J. (2016). The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. Cell *167*, 122-132 e129.

Rajavel, K.S., and Neufeld, E.F. (2001). Nonsense-mediated decay of human HEXA mRNA. Mol Cell Biol *21*, 5512-5519.

Rajyaguru, P., She, M., and Parker, R. (2012). Scd6 targets eIF4G to repress translation: RGG motif proteins as a class of eIF4G-binding proteins. Mol Cell *45*, 244-254.

Ramani, A.K., Nelson, A.C., Kapranov, P., Bell, I., Gingeras, T.R., and Fraser, A.G. (2009). High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective Caenorhabditis elegans. Genome Biol *10*, R101.

Ramirez, C.V., Vilela, C., Berthelot, K., and McCarthy, J.E. (2002). Modulation of eukaryotic mRNA stability via the cap-binding translation complex eIF4F. J Mol Biol *318*, 951-962.

Ramos, C.R., Oliveira, C.L., Torriani, I.L., and Oliveira, C.C. (2006). The Pyrococcus exosome complex: structural and functional characterization. J Biol Chem *281*, 6751-6759.

Rayson, S., Arciga-Reyes, L., Wootton, L., De Torres Zabala, M., Truman, W., Graham, N., Grant, M., and Davies, B. (2012). A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in Arabidopsis thaliana NMD mutants. PLoS One *7*, e31917.

Rehwinkel, J., Letunic, I., Raes, J., Bork, P., and Izaurralde, E. (2005). Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. RNA *11*, 1530-1544.

Rendl, L.M., Bieman, M.A., and Smibert, C.A. (2008). S. cerevisiae Vts1p induces deadenylation-dependent transcript degradation and interacts with the Ccr4p-Pop2p-Not deadenylase complex. RNA *14*, 1328-1336.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Rodnina, M.V. (2010). Protein synthesis meets ABC ATPases: new roles for Rli1/ABCE1. EMBO Rep *11*, 143-144.

Roque, S., Cerciat, M., Gaugue, I., Mora, L., Floch, A.G., de Zamaroczy, M., Heurgue-Hamard, V., and Kervestin, S. (2015). Interaction between the poly(A)-binding protein Pab1 and the eukaryotic release factor eRF3 regulates translation termination but not mRNA decay in Saccharomyces cerevisiae. RNA *21*, 124-134.

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

Roy, B., Leszyk, J.D., Mangus, D.A., and Jacobson, A. (2015). Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. Proc Natl Acad Sci U S A *112*, 3038-3043.

Rufener, S.C., and Muhlemann, O. (2013). eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. Nat Struct Mol Biol *20*, 710-717.

Russell, P., Benson, J.D., and Denis, C.L. (2002). Characterization of mutations in NOT2 indicates that it plays an important role in maintaining the integrity of the CCR4-NOT complex. J Mol Biol *322*, 27-39.

Sachs, A.B., and Davis, R.W. (1989). The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. Cell *58*, 857-867.

Sachs, A.B., Davis, R.W., and Kornberg, R.D. (1987). A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. Mol Cell Biol *7*, 3268-3276.

Sachs, A.B., and Deardorff, J.A. (1992). Translation initiation requires the *PAB*-dependent poly(A) ribonuclease in yeast. Cell *70*, 961-973.

Sakaki, K., Yoshina, S., Shen, X., Han, J., DeSantis, M.R., Xiong, M., Mitani, S., and Kaufman, R.J. (2012). RNA surveillance is required for endoplasmic reticulum homeostasis. Proc Natl Acad Sci U S A *109*, 8079-8084.

Salas-Marco, J., and Bedwell, D.M. (2004). GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. Mol Cell Biol *24*, 7769-7778.

Salas-Marco, J., Fan-Minogue, H., Kallmeyer, A.K., Klobutcher, L.A., Farabaugh, P.J., and Bedwell, D.M. (2006). Distinct paths to stop codon reassignment by the variant-code organisms Tetrahymena and Euplotes. Mol Cell Biol *26*, 438-447.

Sato, H., Hosoda, N., and Maquat, L.E. (2008). Efficiency of the pioneer round of translation affects the cellular site of nonsense-mediated mRNA decay. Mol Cell *29*, 255-262.

Sauliere, J., Murigneux, V., Wang, Z., Marquenet, E., Barbosa, I., Le Tonqueze, O., Audic, Y., Paillard, L., Roest Crollius, H., and Le Hir, H. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. Nat Struct Mol Biol *19*, 1124-1131.

Sayani, S., Janis, M., Lee, C.Y., Toesca, I., and Chanfreau, G.F. (2008). Widespread impact of nonsense-mediated mRNA decay on the yeast intronome. Mol Cell *31*, 360-370.

Schaeffer, D., Tsanova, B., Barbas, A., Reis, F.P., Dastidar, E.G., Sanchez-Rotunno, M., Arraiano, C.M., and van Hoof, A. (2009). The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. Nat Struct Mol Biol *16*, 56-62.

Schafer, I.B., Rode, M., Bonneau, F., Schussler, S., and Conti, E. (2014). The structure of the Pan2-Pan3 core complex reveals cross-talk between deadenylase and pseudokinase. Nat Struct Mol Biol *21*, 591-598.

Schiestl, R.H., and Gietz, R.D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Curr Genet *16*, 339-346.

Schmidt, S.A., Foley, P.L., Jeong, D.H., Rymarquis, L.A., Doyle, F., Tenenbaum, S.A., Belasco, J.G., and Green, P.J. (2015). Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of endogenous NMD targets in human cells. Nucleic Acids Res *43*, 309-323.

Schneider, C., Leung, E., Brown, J., and Tollervey, D. (2009). The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. Nucleic Acids Res *37*, 1127-1140.

Schoenberg, D.R., and Maquat, L.E. (2009). Re-capping the message. Trends Biochem Sci *34*, 435-442.

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

Schwartz, D., Decker, C.J., and Parker, R. (2003). The enhancer of decapping proteins, Edc1p and Edc2p, bind RNA and stimulate the activity of the decapping enzyme. RNA *9*, 239-251.

Schwartz, D.C., and Parker, R. (1999). Mutations in translation initiation factors lead to increased rates of deadenylation and decapping of mRNAs in *Saccharomyces cerevisiae*. Mol Cell Biol *19*, 5247-5256.

Schwartz, D.C., and Parker, R. (2000). mRNA decapping in yeast requires dissociation of the cap binding protein, eukaryotic translation initiation factor 4E. Mol Cell Biol *20*, 7933-7942.

Schwede, A., Ellis, L., Luther, J., Carrington, M., Stoecklin, G., and Clayton, C. (2008). A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. Nucleic Acids Res *36*, 3374-3388.

Schweingruber, C., Rufener, S.C., Zund, D., Yamashita, A., and Muhlemann, O. (2013). Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. Biochim Biophys Acta *1829*, 612-623.

Seit-Nebi, A., Frolova, L., and Kisselev, L. (2002). Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. EMBO Rep *3*, 881-886.

Serdar, L.D., Whiteside, D.L., and Baker, K.E. (2016). ATP hydrolysis by UPF1 is required for efficient translation termination at premature stop codons. Nat Commun *7*, 14021.

Serin, G., Gersappe, A., Black, J.D., Aronoff, R., and Maquat, L.E. (2001). Identification and characterization of human orthologues to *Saccharomyces cerevisiae* Upf2 protein and Upf3 protein (*Caenorhabditis elegans SMG-4*). Mol Cell Biol *21*, 209-223.

Serio, T.R., Cashikar, A.G., Kowal, A.S., Sawicki, G.J., and Lindquist, S.L. (2001). Self-perpetuating changes in Sup35 protein conformation as a mechanism of heredity in yeast. Biochem Soc Symp, 35-43.

Serio, T.R., Cashikar, A.G., Moslehi, J.J., Kowal, A.S., and Lindquist, S.L. (1999). Yeast prion [psi +] and its determinant, Sup35p. Methods Enzymol *309*, 649-673.

Sharif, H., and Conti, E. (2013). Architecture of the Lsm1-7-Pat1 complex: a conserved assembly in eukaryotic mRNA turnover. Cell reports *5*, 283-291.

Sharif, H., Ozgur, S., Sharma, K., Basquin, C., Urlaub, H., and Conti, E. (2013). Structural analysis of the yeast Dhh1-Pat1 complex reveals how Dhh1 engages Pat1, Edc3 and RNA in mutually exclusive interactions. Nucleic Acids Res *41*, 8377-8390.

She, M., Decker, C.J., Chen, N., Tumati, S., Parker, R., and Song, H. (2006). Crystal structure and functional analysis of Dcp2p from Schizosaccharomyces pombe. Nat Struct Mol Biol *13*, 63-70.

She, M., Decker, C.J., Sundramurthy, K., Liu, Y., Chen, N., Parker, R., and Song, H. (2004). Crystal structure of Dcp1p and its functional implications in mRNA decapping. Nat Struct Mol Biol *11*, 249-256.

She, M., Decker, C.J., Svergun, D.I., Round, A., Chen, N., Muhlrad, D., Parker, R., and Song, H. (2008). Structural basis of dcp2 recognition and activation by dcp1. Mol Cell *29*, 337-349.

Shoemaker, C.J., Eyler, D.E., and Green, R. (2010). Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. Science *330*, 369-372.

Shoemaker, C.J., and Green, R. (2011). Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. Proc Natl Acad Sci U S A *108*, E1392-1398.

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. Nat Struct Mol Biol *19*, 594-601.

Shum, E.Y., Jones, S.H., Shao, A., Dumdie, J., Krause, M.D., Chan, W.K., Lou, C.H., Espinoza, J.L., Song, H.W., Phan, M.H.*, et al.* (2016). The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. Cell *165*, 382-395.

Silva, A.L., Ribeiro, P., Inacio, A., Liebhaber, S.A., and Romao, L. (2008). Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. RNA *14*, 563-576.

Simon, E., and Seraphin, B. (2007). A specific role for the C-terminal region of the Poly(A)-binding protein in mRNA decay. Nucleic Acids Res *35*, 6017-6028.

Sims, L.M., and Igarashi, R.Y. (2012). Regulation of the ATPase activity of ABCE1 from Pyrococcus abyssi by Fe-S cluster status and Mg(2)(+): implication for ribosomal function. Arch Biochem Biophys *524*, 114-122.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell *151*, 750-764.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS Biol *6*, e111.

Siridechadilok, B., Fraser, C.S., Hall, R.J., Doudna, J.A., and Nogales, E. (2005). Structural roles for human translation factor eIF3 in initiation of protein synthesis. Science *310*, 1513-1515.

Smith, J.E., Alvarez-Dominguez, J.R., Kline, N., Huynh, N.J., Geisler, S., Hu, W., Coller, J., and Baker, K.E. (2014). Translation of small open reading frames within unannotated RNA transcripts in Saccharomyces cerevisiae. Cell reports *7*, 1858-1866.

Song, H., Mugnier, P., Das, A.K., Webb, H.M., Evans, D.R., Tuite, M.F., Hemmings, B.A., and Barford, D. (2000). The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. Cell *100*, 311-321.

Steiger, M., Carr-Schmid, A., Schwartz, D.C., Kiledjian, M., and Parker, R. (2003). Analysis of recombinant yeast decapping enzyme. RNA *9*, 231-238.

Stevens, A. (1978). An exoribonuclease from Saccharomyces cerevisiae: effect of modifications of 5' end groups on the hydrolysis of substrates to 5' mononucleotides. Biochem Biophys Res Commun *81*, 656-661.

Stirnimann, C.U., Petsalaki, E., Russell, R.B., and Muller, C.W. (2010). WD40 proteins propel cellular networks. Trends Biochem Sci *35*, 565-574.

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

Sun, M., Schwalb, B., Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol Cell *52*, 52-62.

Sweet, T., Kovalak, C., and Coller, J. (2012). The DEAD-box protein Dhh1 promotes decapping by slowing ribosome movement. PLoS Biol *10*, e1001342.

Synowsky, S.A., and Heck, A.J. (2008). The yeast Ski complex is a hetero-tetramer. Protein Sci *17*, 119-125.

Tadauchi, T., Matsumoto, K., Herskowitz, I., and Irie, K. (2001). Post-transcriptional regulation through the HO 3'-UTR by Mpt5, a yeast homolog of Pumilio and FBF. EMBO J *20*, 552-561.

Takahashi, S., Araki, Y., Sakuno, T., and Katada, T. (2003). Interaction between Ski7p and Upf1p is required for nonsense-mediated 3'-to-5' mRNA decay in yeast. EMBO J *22*, 3951-3959.

Tani, H., Imamachi, N., Salam, K.A., Mizutani, R., Ijiri, K., Irie, T., Yada, T., Suzuki, Y., and Akimitsu, N. (2012). Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA Biol *9*, 1370-1379.

Tani, H., Torimura, M., and Akimitsu, N. (2013). The RNA degradation pathway regulates the function of GAS5 a non-coding RNA in mammalian cells. PLoS One *8*, e55684.

Temme, C., Zaessinger, S., Meyer, S., Simonelig, M., and Wahle, E. (2004). A complex containing the CCR4 and CAF1 proteins is involved in mRNA deadenylation in Drosophila. EMBO J *23*, 2862-2871.

Temme, C., Zhang, L., Kremmer, E., Ihling, C., Chartier, A., Sinz, A., Simonelig, M., and Wahle, E. (2010). Subunits of the Drosophila CCR4-NOT complex and their roles in mRNA deadenylation. RNA *16*, 1356-1370.

Ter-Avanesyan, M.D., Dagkesamanskaya, A.R., Kushnirov, V.V., and Smirnov, V.N. (1994). The SUP35 omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [psi+] in the yeast Saccharomyces cerevisiae. Genetics *137*, 671-676.

Tharun, S., He, W., Mayes, A.E., Lennertz, P., Beggs, J.D., and Parker, R. (2000). Yeast Sm-like proteins function in mRNA decapping and decay. Nature *404*, 515-518.

Tharun, S., and Parker, R. (1999). Analysis of mutations in the yeast mRNA decapping enzyme. Genetics *151*, 1273-1285.

Tharun, S., and Parker, R. (2001). Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. Mol Cell *8*, 1075-1083.

Thompson, D.M., and Parker, R. (2007). Cytoplasmic decay of intergenic transcripts in Saccharomyces cerevisiae. Mol Cell Biol *27*, 92-101.

Thore, S., Mauxion, F., Seraphin, B., and Suck, D. (2003). X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. EMBO Rep *4*, 1150-1155.

Toma, K.G., Rebbapragada, I., Durand, S., and Lykke-Andersen, J. (2015). Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay. RNA *21*, 887-897.

Tritschler, F., Braun, J.E., Eulalio, A., Truffault, V., Izaurralde, E., and Weichenrieder, O. (2009a). Structural basis for the mutually exclusive anchoring of P body components EDC3 and Tral to the DEAD box protein DDX6/Me31B. Mol Cell *33*, 661-668.

Tritschler, F., Braun, J.E., Motz, C., Igreja, C., Haas, G., Truffault, V., Izaurralde, E., and Weichenrieder, O. (2009b). DCP1 forms asymmetric trimers to assemble into active mRNA decapping complexes in metazoa. Proc Natl Acad Sci U S A *106*, 21591-21596.

Tsuboi, T., Kuroha, K., Kudo, K., Makino, S., Inoue, E., Kashima, I., and Inada, T. (2012). Dom34:hbs1 plays a general role in quality-control systems by dissociation of a stalled ribosome at the 3' end of aberrant mRNA. Mol Cell *46*, 518-529.

Tucker, M., Staples, R.R., Valencia-Sanchez, M.A., Muhlrad, D., and Parker, R. (2002). Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in *Saccharomyces cerevisiae*. EMBO J *21*, 1427-1436.

Tucker, M., Valencia-Sanchez, M.A., Staples, R.R., Chen, J., Denis, C.L., and Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in S*accharomyces cerevisiae*. Cell *104*, 377-386.

Uchida, N., Hoshino, S., and Katada, T. (2004). Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. J Biol Chem *279*, 1383-1391.

Ulbricht, R.J., and Olivas, W.M. (2008). Puf1p acts in combination with other yeast Puf proteins to control mRNA stability. RNA *14*, 246-262.

Unterholzner, L., and Izaurralde, E. (2004). SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. Mol Cell *16*, 587-596.

Valkov, E., Muthukumar, S., Chang, C.T., Jonas, S., Weichenrieder, O., and Izaurralde, E. (2016). Structure of the Dcp2-Dcp1 mRNA-decapping complex in the activated conformation. Nat Struct Mol Biol *23*, 574-579.

van den Elzen, A.M., Schuller, A., Green, R., and Seraphin, B. (2014). Dom34-Hbs1 mediated dissociation of inactive 80S ribosomes promotes restart of translation after stress. EMBO J *33*, 265-276.

van Dijk, E., Cougot, N., Meyer, S., Babajko, S., Wahle, E., and Seraphin, B. (2002). Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. EMBO J *21*, 6915-6924.

van Hoof, A., Frischmeyer, P.A., Dietz, H.C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science *295*, 2262-2264.

van Hoof, A., Staples, R.R., Baker, R.E., and Parker, R. (2000). Function of the ski4p (Csl4p) and Ski7p proteins in 3'-to-5' degradation of mRNA. Mol Cell Biol *20*, 8230-8243.

Vilela, C., Velasco, C., Ptushkina, M., and McCarthy, J.E. (2000). The eukaryotic mRNA decapping protein Dcp1 interacts physically and functionally with the eIF4F translation initiation complex. EMBO J *19*, 4372-4382.

Wahle, E., and Winkler, G.S. (2013). RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. Biochim Biophys Acta *1829*, 561-570.

Wang, L., Lewis, M.S., and Johnson, A.W. (2005). Domain interactions within the Ski2/3/8 complex and between the Ski complex and Ski7p. RNA *11*, 1291-1302.

Wang, M., and Pestov, D.G. (2011). 5'-end surveillance by Xrn2 acts as a shared mechanism for mammalian pre-rRNA maturation and decay. Nucleic Acids Res *39*, 1811-1822.

Wang, W., Cajigas, I.J., Peltz, S.W., Wilkinson, M.F., and Gonzalez, C.I. (2006). Role for Upf2p phosphorylation in Saccharomyces cerevisiae nonsense-mediated mRNA decay. Mol Cell Biol *26*, 3390-3400.

Wang, W., Czaplinski, K., Rao, Y., and Peltz, S.W. (2001). The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. EMBO J *20*, 880-890.

Wang, Z., Jiao, X., Carr-Schmid, A., and Kiledjian, M. (2002). The hDcp2 protein is a mammalian mRNA decapping enzyme. Proc Natl Acad Sci U S A *99*, 12663-12668.

Wasmuth, E.V., and Lima, C.D. (2012). Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. Mol Cell *48*, 133-144.

Weischenfeldt, J., Damgaard, I., Bryder, D., Theilgaard-Monch, K., Thoren, L.A., Nielsen, F.C., Jacobsen, S.E., Nerlov, C., and Porse, B.T. (2008). NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. Genes Dev *22*, 1381-1396.

Weischenfeldt, J., Waage, J., Tian, G., Zhao, J., Damgaard, I., Jakobsen, J.S., Kristiansen, K., Krogh, A., Wang, J., and Porse, B.T. (2012). Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. Genome Biol *13*, R35.

Weixlbaumer, A., Jin, H., Neubauer, C., Voorhees, R.M., Petry, S., Kelley, A.C., and Ramakrishnan, V. (2008). Insights into translational termination from the structure of RF2 bound to the ribosome. Science *322*, 953-956.

Welch, E.M., and Jacobson, A. (1999). An internal open reading frame triggers nonsense-mediated decay of the yeast SPT10 mRNA. EMBO J *18*, 6134-6145.

Weng, Y., Czaplinski, K., and Peltz, S.W. (1996). Identification and characterization of mutations in the *UPF1* gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. Mol Cell Biol *16*, 5491-5506.

Wery, M., Descrimes, M., Vogt, N., Dallongeville, A.S., Gautheret, D., and Morillon, A. (2016). Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. Mol Cell *61*, 379-392.

Wilson, M.A., Meaux, S., and van Hoof, A. (2007). A genomic screen in yeast reveals novel aspects of nonstop mRNA metabolism. Genetics *177*, 773-784.

Wittkopp, N., Huntzinger, E., Weiler, C., Sauliere, J., Schmidt, S., Sonawane, M., and Izaurralde, E. (2009). Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. Mol Cell Biol *29*, 3517-3528.

Wolf, J., Valkov, E., Allen, M.D., Meineke, B., Gordiyenko, Y., McLaughlin, S.H., Olsen, T.M., Robinson, C.V., Bycroft, M., Stewart, M.*, et al.* (2014). Structural basis for Pan3 binding to Pan2 and its function in mRNA recruitment and deadenylation. EMBO J *33*, 1514-1526.

Wu, D., Muhlrad, D., Bowler, M.W., Jiang, S., Liu, Z., Parker, R., and Song, H. (2014). Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. Cell Res *24*, 233-246.

Wurm, J.P., Overbeck, J., and Sprangers, R. (2016). The S. pombe mRNA decapping complex recruits cofactors and an Edc1-like activator through a single dynamic surface. RNA *22*, 1360-1372.

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., and Tong, L. (2009). Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. Nature *458*, 784-788.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033-1037.

Yamashita, A. (2013). Role of SMG-1-mediated Upf1 phosphorylation in mammalian nonsense-mediated mRNA decay. Genes Cells *18*, 161-175.

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

Yamashita, A., Izumi, N., Kashima, I., Ohnishi, T., Saari, B., Katsuhata, Y., Muramatsu, R., Morita, T., Iwamatsu, A., Hachiya, T.*, et al.* (2009). SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. Genes Dev *23*, 1091-1105.

Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., and Ohno, S. (2001). Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. Genes Dev *15*, 2215-2228.

Yao, G., Chiang, Y.C., Zhang, C., Lee, D.J., Laue, T.M., and Denis, C.L. (2007). PAB1 self-association precludes its binding to poly(A), thereby accelerating CCR4 deadenylation in vivo. Mol Cell Biol *27*, 6243-6253.

Yosefzon, Y., Koh, Y.Y., Chritton, J.J., Lande, A., Leibovich, L., Barziv, L., Petzold, C., Yakhini, Z., Mandel-Gutfreund, Y., Wickens, M.*, et al.* (2011). Divergent RNA binding specificity of yeast Puf2p. RNA *17*, 1479-1488.

Young, D.J., Guydosh, N.R., Zhang, F., Hinnebusch, A.G., and Green, R. (2015). Rli1/ABCE1 Recycles Terminating Ribosomes and Controls Translation Reinitiation in 3'UTRs In Vivo. Cell *162*, 872-884.

Zaborske, J.M., Zeitler, B., and Culbertson, M.R. (2013). Multiple transcripts from a 3'-UTR reporter vary in sensitivity to nonsense-mediated mRNA decay in Saccharomyces cerevisiae. PLoS One *8*, e80981.

Zhang, C., Lee, D.J., Chiang, Y.C., Richardson, R., Park, S., Wang, X., Laue, T.M., and Denis, C.L. (2013). The RRM1 domain of the poly(A)-binding protein from Saccharomyces cerevisiae is critical to control of mRNA deadenylation. Mol Genet Genomics *288*, 401-412.

Zhang, J., Sun, X., Qian, Y., and Maquat, L.E. (1998). Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. RNA *4*, 801-815.

Zhang, S., Welch, E.M., Hogan, K., Brown, A.H., Peltz, S.W., and Jacobson, A. (1997). Polysome-associated mRNAs are substrates for the nonsense-mediated mRNA decay pathway in *Saccharomyces cerevisiae*. RNA *3*, 234-244.

Zund, D., Gruber, A.R., Zavolan, M., and Muhlemann, O. (2013). Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs. Nat Struct Mol Biol *20*, 936-943.

Zund, D., and Muhlemann, O. (2013). Recent transcriptome-wide mapping of UPF1 binding sites reveals evidence for its recruitment to mRNA before translation. Translation (Austin) *1*, e26977.

Zuo, Y., and Deutscher, M.P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. Nucleic Acids Res *29*, 1017-1026.