2017

Killing The Messenger: Exploring Novel Triggers For Messenger Rna Decay In Eukaryotes

Lee Elliott Vandivier
University of Pennsylvania, evlee@mail.med.upenn.edu

Killing The Messenger: Exploring Novel Triggers For Messenger Rna Decay In Eukaryotes

Abstract
The lifecycle of messenger RNAs is regulated by multiple layers beyond their primary sequence. In addition to carrying the information for protein synthesis, mRNAs are decorated with RNA binding proteins, marked with covalent chemical modifications, and fold into intricate secondary structures. However, the full set of information encoded by these "epitranscriptomic" layers is only partially understood, and is often only characterized for select transcripts. Thus, it is crucial to develop and apply transcriptome-wide analytical tools to probe the location and functional relevance of epitranscriptome features. In this dissertation, I focus on applying such methods toward better understanding determinants of mRNA stability, through using 1) High Throughput Annotation of Modified Nucleotides, 2) nuclease-mediated probing of RNA secondary structure, and 3) detection of partial mRNA degradation from RNA sequencing. I observe that chemical modifications tend to mark uncapped and small RNA fragments derived from mRNAs in plants and humans, suggesting a link between modifications and mRNA stability. I then show this link is direct through showing differential stability at Arabidopsis transcripts that change modification status during long-term salt stress. By probing secondary structure, I show a link between structure, smRNA production, and co-translational RNA decay. Finally, I develop a new in silico method to detect partial RNA degradation in mouse oocytes, and identify sequence elements that appear to block complete exonucleolytic transcript cleavage during meiosis. I then identify putative RNA binding proteins that might mediate this partial decay. In summary, I apply transcriptome-wide sequencing-based methods to survey the effects of covalent modifications, secondary structure, and RNA binding proteins on mRNA stability.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Cell & Molecular Biology

First Advisor
Brian D. Gregory

Subject Categories
Bioinformatics | Biology | Genetics

KILLING THE MESSENGER: EXPLORING NOVEL TRIGGERS FOR MESSENGER RNA DECAY IN EUKARYOTES

Lee E. Vandivier

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

_____

Brian Gregory, Ph.D.
Associate Professor of Biology

Graduate Group Chairperson

_____

Dan Kessler, Ph.D.
Associate Professor of Cell and Developmental Biology

Dissertation Committee

Doris Wager, Ph.D.
Professor of Biology (Chairperson)
R. Scott Poethig, Ph.D.
John H. and Margaret B. Fassitt Professor of Biology
Zissimos Mourelatos, M.D.
Professor of Pathology and Laboratory Medicine
Blake Meyers, Ph.D.
Professor, Division of Plant Sciences, University of Missouri

KILLING THE MESSENGER: EXPLORING NOVEL TRIGGERS FOR MESSENGER RNA DECAY IN EUKARYOTES

ACKNOWLEDGMENTS

Research is both an edifying and humbling experience, and I am grateful to everyone who helped me walk on this exciting but uncertain path.

I must first thank my thesis mentor Brian Gregory, who I first encountered as an undergraduate in search of coursework in the natural sciences that I might truly enjoy. I did, decided to give up an English major, and eventually would meet Brian during my graduation ceremony – giving a pitch to my whole family as to why I should rotate in his lab. You have been a consistent supporter of my research, in all its successes and failures. You helped me push beyond my boundaries of expertise and learn genomics and computational biology. I have grown as a scientist under your guidance, and I am happy to have chosen your lab despite, in your words, “the worst rotation in history.”

I am equally thankful to the members of Brian’s lab who taught me how to code, analyze big datasets, troubleshoot everything, and set realistic expectations for research. To Ian Silverman and Matthew Willmann, for your guidance at the bench. To Fan Li, Qi Zheng, and Nathan Berkowitz for your patience in teaching a complete novice everything from how to type on a command line to how to develop and version-control custom pipelines. Your help has been invaluable.

Likewise, the Gregory lab has grown, and it has been a pleasure to know and learn from Sager Gosai, Shawn Foley, Marianne Kramer, Lucy Shan, Xiang Yu, Steve Anderson, and Zach Anderson. Zach, I am thankful for your endless patience in triple-checking the HAMR pipeline. I have also been fortunate to mentor many talented students, including Jacky Buckley, Nickole Kanyuch, and Rafael Campos. You have taught me invaluable lessons about how to listen, manage, and teach.

I must also thank my collaborators, whose expertise has enriched my research and my education. To Paul Ryvkin, Pavel Kuksa, and Li-San Wang, for developing HAMR. To Nur Selamoglu and Fevzi Daldal for all the mass spectrometry and proteomics work. To Richard Schultz and Jun Ma for having helped open my eyes to the world of mouse development, and for trusting me to contribute to your research. I am equally thankful to my committee members. To Scott Poethig and Doris Wagner, for your expertise in plant genetics and for your insights that have helped increase the relevance of my work. To Zissimos Mourelatos, for your expertise in all things RNA, and your encouragement to pursue difficult projects. To Blake Meyers, for helping to bridge the worlds of plant and RNA biology, and for continuing to support my work even after moving to the Danforth Center in St. Louis.

And I course, I would be never have made it this far without my parents John and Wendy. Your unconditional love has allowed me to pursue my interests to wherever they take me, and you have given me the passion to learn. To my father, who has taught me patience and kindness and how to cook delicious food. To my mother, who has taught me with your sharp mind and artist’s eye and activist’s clarity. To my Grandmom and Zaydee, who have taught me diligence and the joy of eating a second dinner every night. To my Grandpa Van, whose intellectual spirit has inspired me.

Most of all, I must thank my wife and true love, Minghua Hsiao. I have learned more from you than from all my years in school. You have supported me through every high and every low of graduate school. Your love is fearless, and I admire your courage as we live across nations. I am honored to live my life with you.

ABSTRACT

KILLING THE MESSENGER: EXPLORING NOVEL TRIGGERS FOR MESSENGER RNA DECAY IN EUKARYOTES

Lee E. Vandivier

Brian D. Gregory

The lifecycle of messenger RNAs is regulated by multiple layers beyond their primary sequence. In addition to carrying the information for protein synthesis, mRNAs are decorated with RNA binding proteins, marked with covalent chemical modifications, and fold into intricate secondary structures. However, the full set of information encoded by these “epitranscriptomic” layers is only partially understood, and is often only characterized for select transcripts. Thus, it is crucial to develop and apply transcriptome-wide analytical tools to probe the location and functional relevance of epitranscriptome features. In this dissertation, I focus on applying such methods toward better understanding determinants of mRNA stability, through using 1) High Throughput Annotation of Modified Nucleotides, 2) nuclease-mediated probing of RNA secondary structure, and 3) detection of partial mRNA degradation from RNA sequencing. I observe that chemical modifications tend to mark uncapped and small RNA fragments derived from mRNAs in plants and humans, suggesting a link between modifications and mRNA stability. I then show this link is direct through showing differential stability at *Arabidopsis* transcripts that change modification status during long-term salt stress. By probing secondary structure, I show a link between structure, smRNA production, and co-translational RNA decay. Finally, I develop a new *in silico* method to detect partial RNA degradation in mouse oocytes, and identify sequence elements that appear to block complete exonucleolytic transcript cleavage during meiosis. I then identify putative RNA binding proteins that might mediate this partial decay. In summary, I apply transcriptome-wide sequencing-based methods to survey the effects of covalent modifications, secondary structure, and RNA binding proteins on mRNA stability.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: THE MULTILAYERED MESSAGE – RNA MODIFICATIONS, SECONDARY STRUCTURE, AND BINDING PROTEINS

A note on use of the first-person: Throughout this dissertation, I make extensive use of the first-person to increase readability. When referring to the dissertation, I use the first-person singular. When referring to experimental methods and conclusions, I instead use the first-person plural to acknowledge the collaborative nature of my (our) work.

This section refers to work from:

Vandivier L.E. and Gregory B.D. (2017). Reading the Epitranscriptome: New Techniques and Perspectives. *The Enzymes*. 41, 269-298. PMID: 28601224

Vandivier L.E.*, Anderson, S.J.*, Foley S.W.*, and Gregory B.D. (2017). The Conservation and Function of RNA Secondary Structure in Plants. *Annual Reviews Plant Biology*. 67, 463-88. PMID: 26865341

Foley, S.W.*, Vandivier, L.E.*, Kuksa, P., Gregory, B.D. (2015). Transcriptome-wide measurement of plant RNA secondary structure. *Current Opinion in Plant Biology*. 27, 36-43. PMID: 26119389

Vandivier L.E., Li F., and Gregory B.D. (2015). High-Throughput Nuclease-Mediated Probing of RNA Secondary Structure in Plant Transcriptomes. 1284, 41-70. PMID: 25757767

*Indicates co-first author

1.1 INTRODUCTION

Messenger RNAs (mRNA) contain dense and overlapping layers of information. In addition to mediating the flow of genetic information from DNA to protein through their primary sequence, mRNAs are punctuated with RNA binding proteins (RBPs) (Glisovic et al., 2008), marked with covalent chemical modifications (Cantara et al., 2011; Dunin-Horkawicz et al., 2006; Limbach et al., 1994; Machnicka et al., 2012), and fold into intricate secondary structures (Cruz and Westhof, 2009). While not directly encoded in a transcript's gene of origin, each of these features has the ability modulate both the regulatory and coding potential of mRNAs, and thus form an "epitranscriptomic" layer of regulation (Meyer et al., 2012; Saletore et al., 2012) that is analogous to the epigenetic information encoded through DNA methylation, histone post-translational modifications, chromatin looping, and DNA binding protein occupancy.

Like epigenetic information, epitranscriptomic features of RNA create additional nodes of regulation that affect nearly every point of the complex lifecycle of mRNAs, from transcription to splicing and maturation to export, localization, translation of proteins, and ultimately decay. This enables an increase in transcript and protein diversity, contributing to the ability of complex organisms to develop, specify cell fate, and respond to environmental cues and stresses. However, our understanding of epitranscriptomic regulation is still in its infancy, and thus developing techniques to probe the breadth and function of covalent modifications, secondary structure, and protein binding is crucial to gaining a better understanding of the mRNA lifecycle and its contribution to organismal behavior.

Here, I introduce my dissertation in which I apply transcriptome-wide techniques that elucidate the relationship between the features of the epitranscriptome and mRNA decay. I begin with a brief review of the pathways of mRNA degradation. I then introduce what is known about mRNA secondary structure and covalent modifications, and give an overview of the state of the art in techniques for their detection. Finally, I give an outline of my work and contributions to the field of RNA epitranscriptomics.

1.2 MRNA DECAY

The end of an mRNA's lifespan is a highly regulated process controlled by a host of RBPs that recognize and bind to sequence elements, structural motifs, and covalently modified bases. These RBPs link mRNA decay to upstream regulatory signals, such as signal transduction pathways and stress, and also couple mRNA decay to other aspects of post-transcriptional regulation. For instance, mRNA decay and translation are tightly linked in order to prevent the production of unnecessary or aberrant proteins. Multiple RNA surveillance pathways target defects in translation, including premature stop codons (nonsense-mediated decay, NMD), ribosomal stalling (no-go decay), and ribosomal readthrough into the transcript's polyA tail (non-stop decay) (Garneau et al., 2007; Roy and Jacobson, 2013). As a result, epitranscriptomic features that interfere with translation, such as secondary structures that trigger ribosomal stalling, or covalent modifications that lead to stop codon readthrough, are likely to also trigger mRNA decay. Thus, there are numerous ways in which the epitranscriptome can modulate RNA stability, which I will discuss in more detail in **Sections 1.4 and 1.6**.

The mechanisms by which eukaryotic mRNA is decayed are diverse, though all must overcome the two most important safeguards for mRNA stability, namely the 5' 7-methylguanosine cap and the 3' polyadenosine (polyA$^+$) tail. These structures are added co-transcriptionally, and removing either is sufficient to trigger the activity of 5'-to-3' or 3'-to-5' exonucleases and direct mRNA decay. Thus, the primary mechanisms of mRNA decay involve 3' deadenylation or 5' decapping. Additionally, endonucleolytic cleavage events, which produce one uncapped fragment and another that is deadenylated, are also sufficient to trigger mRNA decay. mRNA decay can be broadly characterized based upon whether it is initiated by deadenylation, decapping, or endonucleolytic cleavage (Garneau et al., 2007). In deadenylation-dependent degradation, the polyA tail of mRNAs is first shortened by the CCR4-NOT, PAN2, or PARN deadenylases. PolyA-binding protein (PABP), which binds to and protects polyA tails, can inhibit these deadenylases though is also required for facilitating certain modes of mRNA decay (Brook and Gray, 2012). Moreover, cap-binding complexes (CBPs) inhibit PARN (Balatsos et al., 2006; Gao et al., 2000), suggesting an interplay between the stability of mRNA caps and tails. Once a transcript is deadenylated, it is degraded 3'-to-5' by the exosome complex, or decapped and degraded 5'-to-3' by XRN exonucleases. Alternatively, in deadenylation-independent mechanisms, mRNAs are directly decapped and degraded by 5'-to-3' XRN exonucleases.

These diverse triggers and mechanisms for mRNA decay enable it to integrate a broad array of regulatory inputs in *cis* and in *trans* and produce dynamic patterns of transcript stability. One of the best characterized examples of dynamic regulation involve the AU-rich elements (AREs), which can bind to factors that trigger deadenylation, but

can also be bound by competing RBPs like HuR that have the opposite effect of transcript stabilization (Schoenberg and Maquat, 2012). In mammals, these often reside in transcripts that should be repressed under “basal” (non-growth, non-stress) physiology, such as inflammatory response genes and oncogenes (Schoenberg and Maquat, 2012). In my dissertation, I also observe a preponderance of RNA secondary structure in defense-response transcripts (Vandivier et al., 2013) and covalent modifications in stress-related transcripts (Vandivier et al., 2015a), and propose that like AREs in mammals, these could be involved in the dynamic stability of plant transcripts that should normally be repressed. miRNA target sites are another key regulator of transcript stability, in which recruiting a miRNA-bound Argonaute (AGO) protein triggers either direct transcript cleavage (slicing) or translational repression. Differential expression of miRNAs can thus lead to differential transcript stability, as has been well-characterized for targets of miR156 and miR172 during vegetative phase change in plants (Wu et al., 2009).

1.3 METHODS TO STUDY MRNA DECAY

Methods to study mRNA stability can be broadly classified into those that either directly track the lifespan of mRNA using pulse-chase or transcription-free systems, or those that instead capture degradation intermediates. Techniques that probe for intermediates include Global Mapping of Uncapped Cleaved Transcripts (GMUCT) (Gregory et al., 2008; Willmann et al., 2014) and Parallel Analysis of RNA Ends (PARE) (German et al., 2008, 2009), both of which probe for uncapped transcripts. In support of

these techniques' utility for capturing actively degrading transcripts, I later show that the proportion of a transcripts in an uncapped state is a useful proxy of mRNA instability. Other techniques probe at the 3' end for shortened polyA tails. For instance, the polyA tail (PAT) assay measures transcript stability through PCR with 1) an oligo(dT) anchor primer and 2) a primer complementary to an upstream region of the transcript. This enables determining the distribution of polyA tail lengths from the size distribution of PCR products (Sallés and Strickland, 1999). A downshift in sizes implies an increase in deadenylation and subsequent decay.

Since steady state mRNA abundance is a function of both transcription and degradation rates, direct measurements of mRNA stability must control for RNA production by either removing transcription or by metabolic pulsing with a labeled ribonucleotide such as tritiated uridine (^{3}HU) (Cleary et al., 2005) or 4-thiouridine (4sU) (Dölken et al., 2008; Rabani et al., 2011) that is later washed away and chased. Alternatively, one can study a transcription-free system, such as enucleated red blood cells or actively mitotic or meiotic cells. For instance, in my analysis of partial mRNA decay in mouse oocytes (**Chapter 5**), cells are meiotic and lack condensed nuclei and transcripts. Thus, half-life can be determined simply by tracking the abundance of maternal mRNA.

Alternatively, transcription can be inhibited using drugs such as actinomycin-D, cordycepin, and α-amanitin. Again, RNA stability can be inferred simply by measuring abundance along series of timepoints. Inhibitor-based assays are simple and broadly applicable across different organisms, though they have the disadvantage of drug toxicity, which can lead to potentially confounding effects on the organism of interest.

Metabolic labelling of transcripts with RNA analogues such as 4sU largely overcomes this problem, since these are far less toxic than transcriptional inhibitors (Dölken et al., 2008; Rabani et al., 2011). 4sU is readily biotinylated and pulled out of the bulk RNA population, allowing labelled transcripts to be chased. Both transcriptional inhibition and metabolic labelling can be readily applied in a high-throughput manner to measure RNA stability across the transcriptome of interest, though careful steps must be taken to ensure equal RNA input across timepoints.

1.4 MRNA SECONDARY STRUCTURE

All RNAs have the capacity to base pair via Watson-Crick, Hoogsteen, or sugar-edge patterns of hydrogen bonds (Leontis and Westhof, 2001; Schroeder et al., 2004). Intermolecular RNA base pairing underlies the coding and replicative abilities of RNA, and enables RNA to serve as a specificity factor in guiding the activity of processes like RNA-directed DNA methylation (RdDM) and microRNA-mediated gene silencing. Intramolecular RNA base pairing is the basis of RNA secondary structure, and is a critical determinant of overall macromolecular folding. In conjunction with cofactors and RNA binding proteins (RBPs), secondary structure forms higher order tertiary structures and confers catalytic, regulatory, and scaffolding functions to RNA. In turn, disrupting the secondary structure of both coding and noncoding RNAs can cause widespread physiological perturbations. For instance, improper transfer RNA (tRNA) folding disrupts its intricate set of interactions with tRNA synthetases, cofactors, and the ribosome that are required for translation, thus impeding a process fundamental to life (Bhaskaran et

al., 2012; Demeshkina et al., 2010). Secondary structure is known to be equally necessary for the functions of ribosomal RNAs (rRNAs) (Nissen et al., 2000; Ramakrishnan, 2014; Steitz and Moore, 2003; Yusupova and Yusupov, 2014), small nuclear RNAs (snRNAs) (Fica et al., 2013; Madhani, 2013), small nucleolar RNAs (snoRNAs) (Ganot et al., 1997; Kiss, 2002; Kiss-László et al., 1996; Lestrade and Weber, 2006; Ni et al., 1997), and microRNAs (miRNAs) (Carthew and Sontheimer, 2009; Chapman and Carrington, 2007; Kurihara and Watanabe, 2004; Park et al., 2002; Reinhart et al., 2002). Additionally, recent studies are beginning to demonstrate the importance of structure in long noncoding RNAs (lncRNAs) (Novikova et al., 2012; Ponting et al., 2009; Ulitsky et al., 2011; Wang and Chang, 2011) and messenger RNAs (mRNAs) (Ding et al., 2014; Gosai et al., 2015; Li et al., 2012a; Rouskin et al., 2014). Thus, a complete understanding of the regulation and functionality of RNAs will require methods to probe and manipulate RNA secondary structure.

1.4.1 Formation of mRNA Secondary Structure

As with protein folding, the formation of RNA secondary structure is not a simple matter of maximizing the number of stable chemical bonds to minimize free energy. Instead, RNA secondary structure is constrained by transcription, steric crowding, RBPs, and interacting ions. For instance, RNA folding is co-transcriptional, leading to "sequential folding" that can vary with the speed of RNA polymerase (RNAP) elongation (Schroeder et al., 2004). Moreover, RNA folding is guided by proteins and ribozymes with RNA chaperone activity during its initial formation to avoid "kinetic folding traps"

(local free energy minima) and improper conformations (Kang et al., 2013; Lorsch, 2002; Mohr et al., 2002; Schroeder et al., 2004; Tompa and Csermely, 2004). Thus, the correct *in vivo* structure of RNA may differ substantially from structures that spontaneously form *in vitro* or the minimum free energy (MFE) structures predicted *in silico*.

In addition to chaperones, there exist a wide array of RNA binding proteins that that can constrain or actively remodel RNA secondary structure. For instance, the RNA recognition motif (RRM) (Ding et al., 1999; Oubridge et al., 1994) and K-homology (KH) domain (Backe et al., 2005; Braddock et al., 2002) specifically recognize single-stranded RNA (ssRNA), while the double-stranded RNA binding domain (dsRBD) preferentially binds double-stranded RNA (dsRNA) (Ryter, 1998). RBPs can also target specific structural patterns, as illustrated by the sterile alpha motif (SAM) protein domain that only targets stem-loops in a “shape-specific” manner (Oberstrass et al., 2006). Notably, both RNA binding elements and RBPs undergo structural rearrangements in response to binding, in a type of induced fit (Williamson, 2000). Active remodelers include ATP-dependent RNA helicases (most notably the ribosome) that actively unwind RNA, leading to the observation that RNA secondary structure is diminished *in vivo* compared to *in vitro* in a partially ATP-dependent manner (Rouskin et al., 2014). Conversely, RNA annealers such as *Hfq* and dsRBD-containing proteins speed the process of folding (Møller et al., 2002; Rajkowitsch et al., 2007). RNA secondary structure can likewise be remodeled by nonprotein ligands, such as metabolite-triggered riboswitches (Bocobza and Aharoni, 2014) and inorganic ions (Draper, 2004).

Since the formation of RNA secondary structure is dependent upon both chaperones and remodelers, methods that measure RNA secondary structure outside its

native context may in fact yield incorrect predictions. In particular, algorithms that utilize free energy minimization such as RNAFold (Hofacker, 2003) often yield very different predictions of secondary structure than empirical structure mapping techniques (Mathews et al., 2004; Vandivier et al., 2013). Thus, in my thesis work I make use of an empirical nuclease-based structure probing technique developed by the Gregory lab.

1.4.2 Functions of mRNA Secondary Structure

A growing body of evidence indicates that secondary structure regulates nearly every step of the mRNA lifecycle, including transcription (Wanrooij et al., 2010), 5' capping (Dong et al., 2007), splicing (Buratti and Baralle, 2004; Jin et al., 2011; Liu et al., 1995; Raker et al., 2009; Warf and Berglund, 2010), polyadenylation (Klasens et al., 1998; Oikawa et al., 2010), nuclear export (Grüter et al., 1998), localization (Bullock et al., 2010; Subramanian et al., 2011), translation (Kozak, 1988; Svitkin et al., 2001; Wen et al., 2008), and turnover (Goodarzi et al., 2012).The best characterized structural elements in mRNAs include internal ribosome entry sites (IRES) to recruit the ribosome (Pelletier and Sonenberg, 1988), histone stem loops to recruit stabilizing factors to non-polyadenylated histone mRNAs (Williams and Marzluff, 1995), and iron response elements (IRE) to recruit RBPs in an iron-dependent manner (Hentze et al., 1987). mRNA can likewise contain riboswitches (Miranda-Ríos, 2007), and even produce miRNAs from their introns and less often exons.

Notably, all canonical smRNAs (e.g. miRNAs, siRNAs) are processed from double-stranded precursors, suggesting that elements of high secondary structure in

mRNAs might be similarly processed. In support of this hypothesis, nuclease-based structure mapping in *Arabidopsis* has revealed a positive correlation between secondary structure and smRNA processing (Li et al., 2012a). Furthermore, highly structured transcripts are in general less abundant and transcribed from more heterochromatic regions, suggesting that smRNA derived from highly structured transcripts could initiate RdDM (Li et al., 2012a). In mammals, secondary structural elements are also known to recruit RBPs that can either stabilize or destabilize mRNAs (Goodarzi et al., 2012), so differential recruitment of RBPs might also explain the tendency of highly structured *Arabidopsis* RNAs to be less abundant. In support of this hypothesis, a recent study found that most regions of the *Arabidopsis* transcriptome that are bound by RBPs are less structured (Gosai et al., 2015).

1.5 METHODS TO STUDY MRNA SECONDARY STRUCTURE

As with the study of mRNA modifications, marrying existing biochemical techniques with high-throughput sequencing has yielded rapid advances in our understanding of mRNA secondary structure. These techniques can be broadly categorized into those that use 1) physical techniques such as X-ray crystallography and nuclear magnetic resonance (NMR), 2) chemical probes that adduce to single-stranded RNA, and 3) structure-specific nucleases to probe both double- and single-stranded RNA.

1.5.1 Physical methods

While labor intensive and targeted, physical techniques still provide the highest fidelity models of secondary structure, and when available provide a “gold standard” for transcripts. X-ray crystallography provided some of the earliest portraits of RNA secondary structure using tRNAs (Kim and Rich, 1968; Kim et al., 1974; Robertus et al., 1974), though it requires short RNAs that readily crystallize and thus has limited utility to study mRNA secondary structure. NMR, in contrast, probes transcripts in solution and can thus capture dynamic secondary structures, though not without considerable computational challenges. For instance, NMR studies have been used to define the structural rearrangements of ribozymes (Hammann et al., 2001; Hoogstraten et al., 1998). However, these techniques require extensive optimization and are laborious, and have only been applied to a select few transcripts.

1.5.2 Chemical-based methods

Chemical probing of mRNA secondary structure relies upon small molecules such as DMS (Ding et al., 2014; Rouskin et al., 2014) or NMIA (Wilkinson et al., 2006) that preferentially form adducts with solvent-accessible nucleotides (Ehresmann et al., 1987). Adduct formation, which is measured in high-throughput sequencing reads as base transitions or reverse transcriptase stalls, is thus a proxy for the lack of secondary structure in specific regions of RNA molecules (**Figure 1.2**). These techniques are powerful insofar as they can provide single nucleotide resolution and can be readily applied *in vivo* (Ding et al., 2014; Rouskin et al., 2014). Nonetheless, these methods

have the disadvantage of only measuring unpaired bases, while paired bases are merely inferred from the absence of evidence. This may also result in the selection of unstructured intermediates that are being translated by the ribosome (Qu et al., 2011; Rouskin et al., 2014), which is of less interest for determining the true functions of RNA secondary structure in the transcriptome. Furthermore, RBP binding can also block the addition of adducts (Talkish et al., 2014). Thus, unpaired regions of RNAs bound by RBPs will be inferred to be in a structured conformation, which could result in the production of incorrect models of RNA secondary structure.

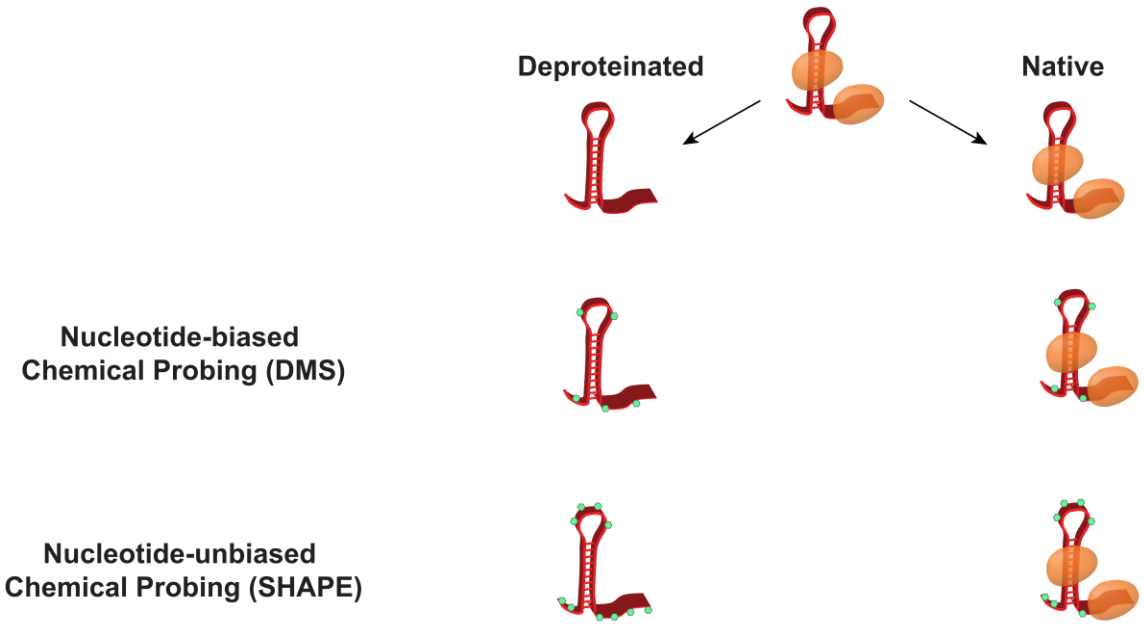

Figure 1.1: Chemical-based probing techniques for empirically determining secondary structure

Chemical probing works through reagents that preferentially form adducts with nucleotides in a single-stranded confirmation, forming covalent modifications in either a nucleotide-biased (DMS) or unbiased (SHAPE) manner.

1.5.3 Nuclease-based methods

A second class of methods relies upon structure-specific RNases (dsRNases and ssRNases) that preferentially cut the phosphodiester bonds 3' of paired or unpaired bases (Gosai et al., 2015; Kertesz et al., 2010; Li et al., 2012b, 2012a). These techniques can either be applied with single-hit stoichiometry (Kertesz et al., 2010; Underwood et al., 2010) or by digesting transcripts to completion (Li et al., 2012a; Zheng et al., 2010) (**Figure 1.2**). While the former approach is likely more accurate, it only produces a single informative nucleotide per read, and produces considerably less coverage per sequencing depth than exhaustive digestion. While nuclease-based techniques have yet to be applied *in vivo*, they have the advantage of producing complementary measurements of both paired and unpaired bases, which guards against selecting for unstructured translating RNA intermediates or incorrectly determining structure for RBP-bound sites. Moreover, measuring both paired and unpaired conformations allows detection of dynamic structures in which bases cycle between paired and unpaired conformations, and also allows for nonparametric definition of highly structured elements. However, nuclease probing is not always at single nucleotide resolution, and bulky nucleases are more likely than small chemical adducts to be occluded by RBPs and higher order structures.

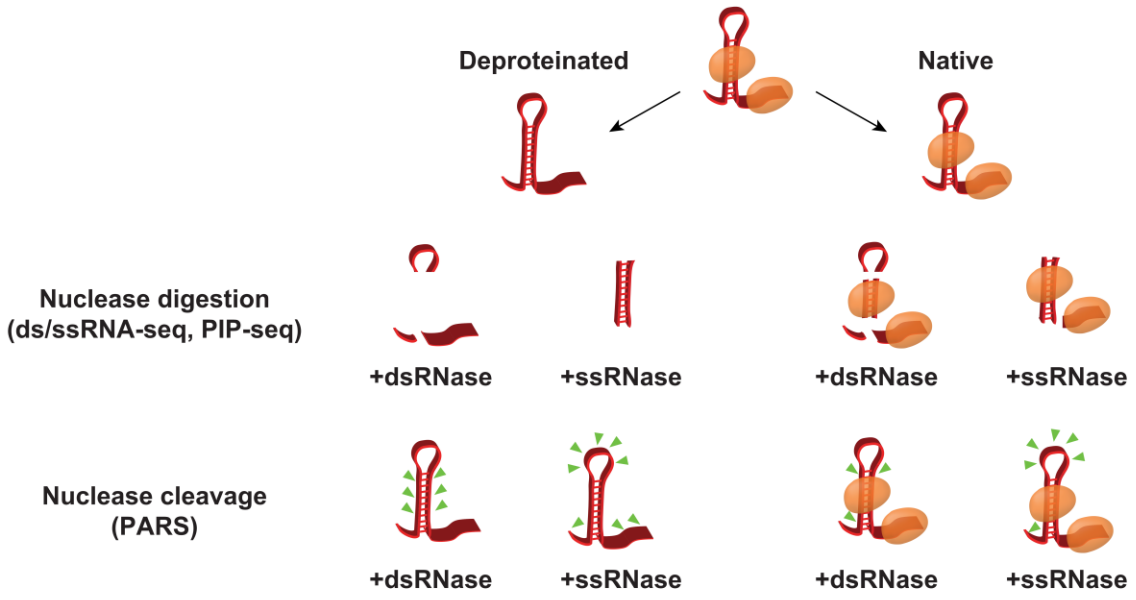

Figure 1.2: Nuclease-based probing techniques for empirically determining secondary structure

RNA can either be probed in a native state bound by RNA binding proteins (orange ovals) or deproteinated through extraction protocols or proteinase K treatment. PARS and Frag-seq assigns structure by the sites of transcript cleavage (green triangles), whereas dsRNase/ssRNase-seq and PIP-seq both work by complete digestion. It is worth noting that while multiple cleavage sites per transcript are shown, PARS and Frag-seq use single-hit stoichiometry, with one cut/modification site interrogated per sequencing read.

1.6 RNA COVALENT MODIFICATIONS

RNA chemical modifications can decorate nearly every known class of RNA, across all kingdoms of life and viruses. To date, over 100 classes of post-transcriptional modifications have been characterized (Cantara et al., 2011; Dunin-Horkawicz et al., 2006; Limbach et al., 1994; Machnicka et al., 2012), each of which can alter the chemical properties of their respective nucleotides, leading to altered base pairing and structural conformations and in turn to differential association with RNA binding proteins

(RBPs). For instance, both N^6-methyladenosine (m^6A) (Liu et al., 2015; Roost et al., 2015) and N^1-methyladenosine (m^1A) (Helm et al., 1999; Zhou et al., 2016) destabilizes double-stranded RNA, which can in turn allow for interaction with RNA binding proteins (RBPs) such as HNRNPC (Liu et al., 2015). Conversely, pseudouridine stabilizes secondary structural elements (Arnez and Steitz, 1994; Kierzek et al., 2014; Newby and Greenbaum, 2002; Sundaram et al., 2000). Modifications can likewise act as signals to direct binding of reader proteins, such as those that recognize m^6A via aromatic methyl-binding pockets (Luo and Tong, 2014; Xu et al., 2014). In turn, modifications can regulate nearly every step of the RNA lifecycle, from transcription (Patil et al., 2016) and maturation (Xiao et al., 2016; Zhao et al., 2014) to export (Fustin et al., 2013; Zheng et al., 2013), translation (Choi et al., 2016; Wang et al., 2015), and stability (Du et al., 2016; Mauer et al., 2016; Wang et al., 2014a, 2014b). Thus, establishing robust methods to survey modifications across the transcriptome is a critical component of understanding post-transcriptional regulation.

Until recently, the majority of nucleotide-resolution RNA modifications studies were limited to highly abundant and predominantly noncoding species like transfer RNAs (tRNAs) (Sprinzl and Vassilenko, 2005), ribosomal RNAs (rRNAs) (McCloskey and Rozenski, 2005), and small nuclear RNAs (snRNAs) (Massenet et al., 1998) since their methods of detection required large amounts of highly pure RNA (Gupta and Randerath, 1979; Sprinzl and Vassilenko, 2005; Tanaka et al., 1980). As a result, tRNAs are still the most thoroughly characterized of any RNA class, and their modifications remain a gold standard for measuring the true discovery rate of new techniques. However, progress in marrying biochemical techniques with high-throughput sequencing have yielded rapid

advances in the understanding of both the form and function of RNA modifications, particularly in messenger RNAs (mRNAs) and long non-coding RNAs (lncRNAs). These mRNA and lncRNA modifications, often referred to as the epitranscriptome (Meyer et al., 2012; Saletore et al., 2012), are now known to encompass N^6-methyladenosine (m^6A), N^1-methyladenosine (m^1A), 5-methylcytosine (m^5C), pseudouridine (ψ), and 2'-0-methylation of ribose (2'OMe) (**Figure 1.3**), and likely contain additional modification types that can be detected but not unambiguously defined, such as modified internal guanosines (Ryvkin et al., 2013; Vandivier et al., 2015a).

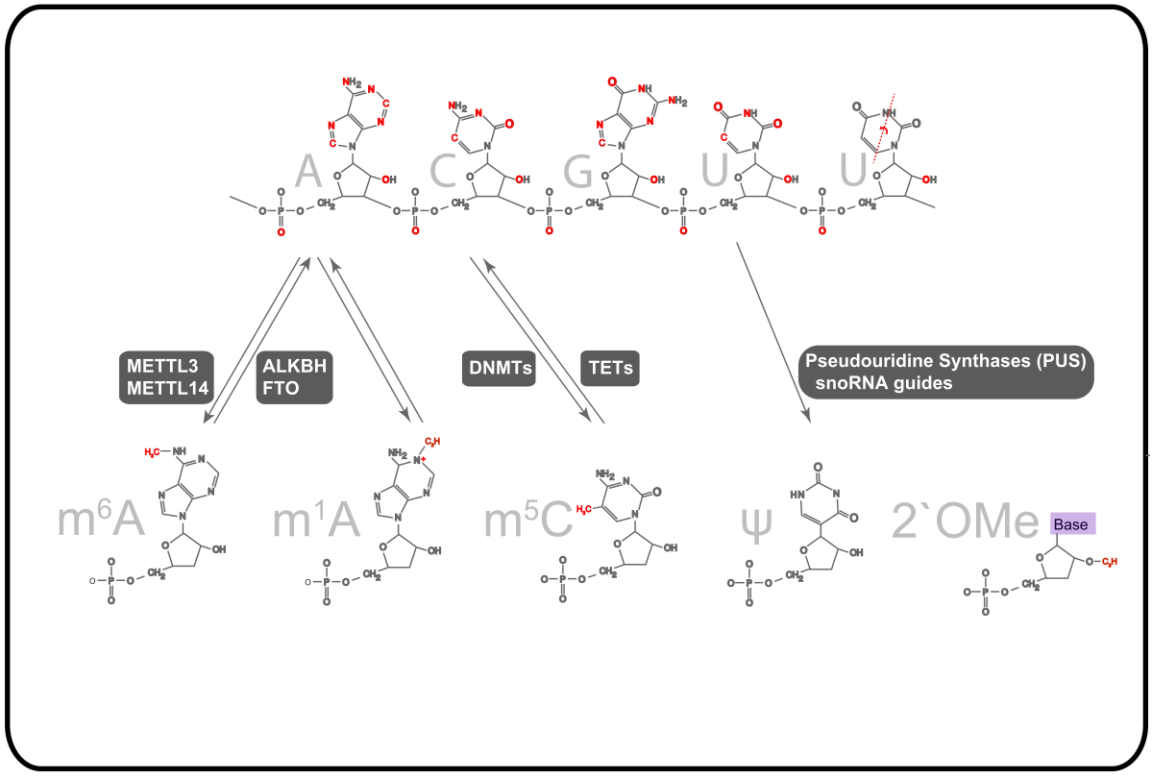

Figure 1.3: RNA harbors multiple potential modifications, though only five have been mapped to mRNAs

Unmodified ribonucleotides are shown in the panel above. All atoms corresponding to known sites of modification are labelled in red, in addition to uridine's axis of isomerization to form pseudouridine (top right). Known mRNA modifications are shown in the panel below, and include N^6-methyladenosine (m^6A), N^1-methyladenosine (m^1A), Pseudouridine (ψ), 5-methylcytosine (m^5C) and 2`-0-methylation of ribose (2`OMe). Black boxes between the panels denote writer enzymes known to catalyze the formation of these modifications, and eraser enzymes known to catalyze their removal.

1.6.1 N^6*-methyladenosine (m^{6}A)*

Of these transcriptome modifications, m^6A (**Figure 1.1**) is the most abundant and well-studied chemical mark, and has been reviewed extensively (Cantara et al., 2011; Chen et al., 2016; Fu et al., 2014b; Jia et al., 2013; Roundtree and He, 2016; Schwartz, 2016; Zhao et al., 2016). Methylation outside the 7mG cap was first detected in

mammalian mRNAs through measuring the incorporation of radiolabeled methyl groups from ^{3}H-methylmethionine (Desrosiers et al., 1974; Perry and Kelley, 1974), and was later attributed to m^{6}A specifically through various chromatographic methods such as electrophoresis, thin-layer chromatography, and high performance liquid chromatography (Dubin and Taylor, 1975; Perry et al., 1975). Subsequent studies demonstrated that m^{6}A is a widespread feature across viruses (Beemon and Keith, 1977; Canaani et al., 1979; Furuichi et al., 1975) and mRNAs from bacteria (Deng et al., 2015), actively meiotic yeast (Bodi et al., 2010), and plants like *Arabidopsis* (Zhong et al., 2008), maize (Nichols, 1979), wheat (Kennedy and Lane, 1979), and oat (Haugland and Cline, 1980). m^{6}A has likewise been detected in archaea (Edmonds et al., 1991), though archaeal mRNAs have yet to been assayed. Like DNA methylation, RNA m^{6}A tends to occur in a specific sequence context (Csepany et al., 1990; Kane and Beemon, 1985; Narayan et al., 1994). Targeted mutation studies and *in vitro* analysis of methyltransferases indicated a general motif of RRACH (R is either G or A, and H is A, C, or U) that is largely consistent across multiple organisms (Csepany et al., 1990; Deng et al., 2015; Dominissini et al., 2012; Kane and Beemon, 1985; Luo et al., 2014; Schwartz et al., 2013) , hinting at broad conservation of the machinery that deposits RNA m^{6}A. Across transcripts, m^{6}A tends to occur near stop codons, long introns, and 3' UTRs (Dominissini et al., 2013; Luo et al., 2014; Meyer et al., 2012). In plants, additional mRNA enrichment is observed at the start codon (Luo et al., 2014). In addition, like all modifications m^{6}A is not mutually exclusive with other chemical marks, and for instance has been shown to co-occur with 2'-0-methylation of ribose (2'OMe) (**Section 1.2.5**) to

form m^6A_m, a newly discovered chemical mark shown to enhance mRNA stability by inhibiting binding of decapping protein 2 (DCP2) (Mauer et al., 2016).

m^6A is also the best example of a complete epitranscriptomic regulatory system, as it possesses known writers (methyltransferase complexes), readers (RNA binding proteins), and erasers (demethylases) (Fu et al., 2014b). The first characterized m^6A writer was the mammalian methyltransferase METTL3 (Bokar et al., 1994, 1997), which was later shown to function as a heterodimer with its catalytically active paralog METTL14 (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014b), alongside cofactors such as the splicing regulator Wilms tumor 1-assocared protein (WTAP) (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014b) and KIAA1429 (Schwartz et al., 2014a). Consistent with WTAP's role in splicing, m^6A has been shown to be deposited in pre-mRNAs, and transcriptome-wide studies have shown its enrichment in long introns (Meyer et al., 2012; Schwartz et al., 2014a), suggesting that the bulk of mRNA m^6A is written in the nucleus. METTL3 and WTAP are broadly conserved across yeast (Agarwala et al., 2012), plants (Zhong et al., 2008), and non-mammalian animals (Hongay and Orr-Weaver, 2011), and within these multiple clades, disruption of m^6A writers leads to a broad range of phenotypes such as loss of stem cell differentiation, developmental defects, and impaired gametogenesis (Batista et al., 2014; Bodi et al., 2010; Geula et al., 2015; Hongay and Orr-Weaver, 2011; Zhong et al., 2008), indicating that m^6A is an ancient and physiologically relevant RNA regulatory feature.

Currently, m^6A is one of the few post-transcriptional modifications with direct evidence of *in vivo* reversibility. m^6A erasers include the alkB family proteins fat mass and obesity associated protein (FTO) (Jia et al., 2011) and alkB homolog 5 (ALKBH5)

(Zheng et al., 2013). FTO is known to catalyze oxidative demethylation, analogous to the TET DNA demethylases (Fu et al., 2014b; Jia et al., 2013), and FTO has been recently shown to more efficiently demethylate m^6A_m (Mauer et al., 2016). Intriguingly, the oxidative demethylation intermediates N^6-hydroxymethyladenosine (hm^6A) and N^6-formyladenosine (f^6A) have been observed *in vivo* in mRNAs (Fu et al., 2013), though their function remains to be elucidated. Like m^6A writers, both FTO and ALKBH5 have been shown to function primarily in the nucleus (Jia et al., 2011; Zheng et al., 2013), and FTO localizes to the nuclear speckles (Jia et al., 2011). Consequently, their genetic disruption leads to altered patterns of splicing (Zhao et al., 2014; Zheng et al., 2013) and increased mRNA export (Zheng et al., 2013). Conversely, loss of METTL3 inhibits RNA export (Fustin et al., 2013).

m^6A is recognized either through “direct readers” like the YTH domain-containing proteins (YTHDs) that contain dedicated aromatic methyladenosine-binding pockets (Luo and Tong, 2014; Xu et al., 2014), or through “indirect readers” such as HNRNPC that directly favor single stranded RNA, and are recruited via m^6A-induced relaxation of secondary structure (Liu et al., 2015; Zhao et al., 2016). m^6A has been estimated to reduce base pairing stability by 0.5-1.7 kcal/mol (Roost et al., 2015). YTHDs include YTHDC1, which has been shown to enhance exon inclusion via recruitment of SRSF3 and blocking of SRSF10 (Xiao et al., 2016), and can associate with *Xist* m^6A to facilitate X chromosome silencing (Patil et al., 2016). YTHDF1 increases translation through recruiting initiation factors (Wang et al., 2015), and YTHDF2 binding destabilizes mRNAs through localization to processing bodies (P-bodies) (Wang et al., 2014a) and recruitment of the CCR4-NOT deadenylases (Du et al., 2016). Among the indirect

readers of m^{6}A, HNRNPC and HNRNPA2B1 affect splicing (Alarcón et al., 2015; Liu et al., 2015), and HNRNPA2B1 also facilitates miRNA stem-loop processing through recruitment of the microprocessor complex (Alarcón et al., 2015).

In addition to their biochemical reversibility, m^{6}A is also notable in being dynamic across time, development, or stress. For instance, m^{6}A is a feature of clock transcripts, and reduction via METTL3 knockdown slows nuclear export, leading to a longer circadian period (Fustin et al., 2013). Additionally, m^{6}A can be rapidly upregulated in response to stress, leading to enhanced and even cap-independent translation (Wang et al., 2015; Zhou et al., 2015), which can for instance allow for translation of heat shock proteins in spite of widespread translational repression (Zhou et al., 2015). m^{6}A thus illustrates the potential of covalent chemical modifications to direct rapid and widespread post-transcriptional regulation of mRNAs. It's effects on enhancing export, increasing translation, and promoting RNA decay have led to its proposal as a "fast track" mark that speeds up the RNA lifecycle and reduces the time needed to respond to physiological stimuli (Zhao et al., 2016).

1.6.2 N^1*-methyladenosine (m^{1}A)*

Like m^{6}A, m^{1}A (**Figure 1.3**) is a widespread transcriptome mark that is known to disrupt RNA secondary structure in both coding and noncoding RNAs (Helm et al., 1999; Zhou et al., 2016). However, m^{1}A was only recently been shown to be present in mRNAs (Dominissini et al., 2016; Li et al., 2016), so most of what is known comes from studies of tRNAs and rRNAs. For instance, methyltransferases have been defined for tRNAs

(Chujo and Suzuki, 2012), though an eraser, ALKBH3, has recently been characterized to direct demethylation both *in vitro* and *in vivo* for m^1A (Li et al., 2016). In mRNAs, m^1A is has been observed to cluster around the start codon, including noncanonical starts (Dominissini et al., 2016; Li et al., 2016), as well as the most upstream splice site (Dominissini et al., 2016). Thus, it is very likely that m^1A plays a role in facilitating translation, though its precise mechanism has yet to be elucidated.

1.6.3 Pseudouridine (ψ)

Pseudouridine (**Figure 1.3**) is prevalent in rRNAs (Maden, 1990), tRNAs (Sprinzl and Vassilenko, 2005), and small nuclear RNAs (snRNAs) (Wu et al., 2011a; Yu et al., 2011), and is the most abundant modification in total cellular RNA (Ge and Yu, 2013), which enabled its early detection as the "fifth nucleotide" of RNA (Davis and Allen, 1957). Pseudouridine has been studied primarily through chromatographic methods and primer extension assays, and can be mapped with single-nucleotide resolution. Pseudouridine is formed through isomerization of uracil such that the ribose C1' binds to uracil C5 instead of N1. In turn, the more accessible N1 is free to form additional hydrogen bonds (**Figure 1.3**), leading to pseudouridine stabilizing RNA secondary structure and increasing RNA rigidity (Arnez and Steitz, 1994; Newby and Greenbaum, 2002), even though its Watson-Crick edge remains identical. The unique structural properties of pseudouridine contribute to the folding of tRNAs and rRNA, and recent studies indicate that pseudouridylation can also affect mRNA coding potential. For instance, this modification has been found to result in readthrough at stop codons

(Fernández et al., 2013; Karijolich and Yu, 2011). Given its strong conservation in tRNAs and rRNAs, is it not that surprising that pseudouridine is found across all kingdoms of life, including endosymbionts (Ofengand and Bakin, 1997).

Pseudouridine writers are termed the pseudouridine synthases (PUSs), and are known to function via two different mechanisms. First, the RNA-dependent pathway involves the formation of a ribonucleoprotein (RNP) complex containing a PUS, cofactors, and box H/ACA or C/D small nucleolar RNAs (snoRNAs). The snoRNAs act as guides that recognize targets with sequence complementarity, thus directing pseudouridylation in a site-specific manner (Ganot et al., 1997; Ni et al., 1997). Alternatively, the RNA-independent pathway relies upon direct recognition of targets by PUS complexes (Ma et al., 2003; Sibert and Patton, 2012), often at conserved structural or sequence motifs. For instance, RNA-independent pseudouridylation of noncoding RNAs tends to occur within paired structures, and has been shown to be base pairing-dependent (Ganot et al., 1997; Ma et al., 2003; Urban et al., 2009). In addition, certain PUS enzymes have been shown to target specific sequence motifs (Behm-Ansmant et al., 2003; Decatur and Schnare, 2008), and recent transcriptome-wide analyses of coding RNAs have confirmed these motifs (Carlile et al., 2014; Li et al., 2015; Schwartz et al., 2014b). Thus, there is evidence that coding and noncoding RNAs share the same pseudouridine writers.

While there is no known mechanism by which pseudouridine is reversed, this chemical mark is still known to be dynamic across development and stress. For instance, inducible pseudouridylation has been observed upon rapamycin treatment (Courtes et al., 2014), heat stress (Li et al., 2015; Schwartz et al., 2014b; Wu et al.,

2011b), nutrient deprivation (Carlile et al., 2014; Li et al., 2015; Wu et al., 2011b), and oxidative stress (Li et al., 2015). Differential pseudouridylation has also been observed within telomerase RNA in Dyskeratosis Congenita cells (Schwartz et al., 2014b). The precise regulatory outcomes of these changes in mRNA pseudouridylation has yet to be clearly defined, though it has been speculated that pseudouridylation stabilizes secondary structures to alter translation efficiency, RNA localization, and RNA stability (Carlile et al., 2014; Schwartz et al., 2014b).

1.6.4 5-methylcytosine (m5C) and 5-hydroxymethylcytosine (hm5C)

Early studies using ^{3}H-methylmethionine radiolabeling coupled with various chromatographic methods (analogous to those used to define m^{6}A) demonstrated that m^{5}C (**Figure 1.3**) can mark mRNAs (Dubin and Taylor, 1975), tRNAs (Motorin and Grosjean, 1999), and viral RNAs (Dubin et al., 1977). Nonetheless, this chemical mark has primarily been studied as the characteristic mark of DNA methylation, and has not been extensively characterized in mRNAs until recently (Hussain et al., 2013a; Khoddami and Cairns, 2013; Squires et al., 2012). As in DNA, RNA m^{5}C can be readily detected at single nucleotide resolution through bisulphite conversion (Hussain et al., 2013a; Squires et al., 2012). Additional techniques rely upon antibody pulldown, or upon cytidine analogues that remain bound to their methyltransferases (Hussain et al., 2013a; Khoddami and Cairns, 2013), and will be covered in **Section 3.1**.

Known m^{5}C writers were first characterized through their methylation of tRNAs, and include the yeast tRNA:m^{5}C methyltransferase (Trm4) (Motorin and Grosjean,

1999), their animal homologue NOP2/SUN RNA methyltransferase family member 2 (Nsun2) (Brzezicha et al., 2006), and the tRNA aspartic acid methyltransferase Dnmt2, which is conserved across plants and animals (Goll et al., 2006). Loss of Dnmt2 leads to reduced stress tolerance, in part through leading to an increase in stress-induced tRNA cleavage (Schaefer et al., 2010). Loss of Nsun2 has also been linked to developmental disability (Abbasi-Moheb et al., 2012) and to impaired male germ cell differentiation (Hussain et al., 2013b). At the transcript level, loss of Nsun2 leads to an increase in aberrant vault RNA cleavage (Hussain et al., 2013c), suggesting that m^5C may have a general role in protecting RNAs from cleavage.

m^5C is also known to be reversible, and the oxidative demethylation intermediates 5-hydroxymethylcytosine (hm^5C) and 5-formylcytosine (f^5C) have been observed *in vivo* (Delatte et al., 2016; Fu et al., 2014a; Zhang et al., 2016). The ten-eleven (TET) family demethylases were previously known to direct DNA demethylation, and have been shown to be necessary (Delatte et al., 2016; Fu et al., 2014a) and sufficient (Fu et al., 2014a) to direct the formation hm^5C in RNA, and thus comprise the first set of known RNA m^5C erasers. hm^5C (**Figure 1.3**) has not been studied extensively, but is known to be enriched among polysome-associated RNA, suggesting a role in facilitating or demarcating active translation (Delatte et al., 2016). Additionally, loss of Drosophila Tet (dTet) has been shown to both reduce hm^5C and disrupt brain development (Delatte et al., 2016).

To date, the precise function of RNA m^5C in coding RNAs is still unclear, though transcriptome-wide experiments have shown it to be enriched in the UTRs (Squires et al., 2012). Notably, some m^5C marks in introns have been shown to reside in regions

with homology to tRNAs, though it is not clear whether these regulate pre-mRNAs directly (Hussain et al., 2013a).

1.6.5 2'-0-methylation of ribose (2'OMe)

The first studies to characterize mRNA methylation also detected 2'-0-methylation of ribose (**Figure 1.3**), which can modify any ribonucleotide (A_m, C_m, G_m, U_m) (Desrosiers et al., 1974). 2'OMe is best characterized as a plant-specific marker that stabilizes smRNAs and is added by the HEN1 methyltransferase (Li et al., 2005; Park et al., 2002). Very little is known about the function of 2'OMe in mRNAs, other than that it has the potential to inhibit adenosine deamination (Yi-Brunozzi et al., 1999) and mRNA stability by inhibiting decapping when marking cap-proximal m^6A (m^6A_m) (Mauer et al., 2016). Given the availability of targeted (Dong et al., 2012) and high-throughput (Birkedal et al., 2015; Marchand et al., 2016) methods for mapping 2'OMe, the function of 2'OMe in the epitranscriptome remains an open but approachable question.

1.7 METHODS TO STUDY RNA COVALENT MODIFICATIONS

Recent advances in merging existing biochemical techniques with high-throughput sequencing have enabled more rapid progress in the study of covalent RNA modifications. Broadly, these methods can be classified based upon their reliance on 1) antibody pulldowns, 2) chemical conversion and adducts, and 3) *in silico* detection from high-throughput RNA sequencing data. Future detection methods may also involve high-

throughput single molecule sequencing with technologies such as Oxford Nanopore, though these techniques are still in their infancy and are beyond the scope of this dissertation. Given that nearly all high-throughput techniques for detecting modified ribonucleotides are based upon existing biochemical approaches, I begin by introducing these targeted techniques for detecting modifications.

1.7.1 Targeted Biochemical Techniques

The earliest methods for detecting RNA modifications relied upon various one- and two-dimensional chromatographic methods such as high performance liquid chromatography, electrophoresis, or thin layer chromatography to resolve modified ribonucleotides based upon changes in their migration properties (Davis and Allen, 1957; Desrosiers et al., 1974). (**Figure 1.4A**). When paired with direct sequencing, these approaches could also map modifications to single-nucleotide resolutions within abundant, readily purified transcripts. In direct sequencing, fragments are resolved at base resolution on a gel, followed by separation in an additional dimension to determine modification status. (Gupta and Randerath, 1979; Sprinzl and Vassilenko, 2005; Tanaka et al., 1980) (**Figure 1.4B**). While these techniques cannot be applied on a transcriptome-wide scale, they have found new use in combination with techniques to purify less abundant species of RNA, such as splint-ligation (Liu et al., 2013).

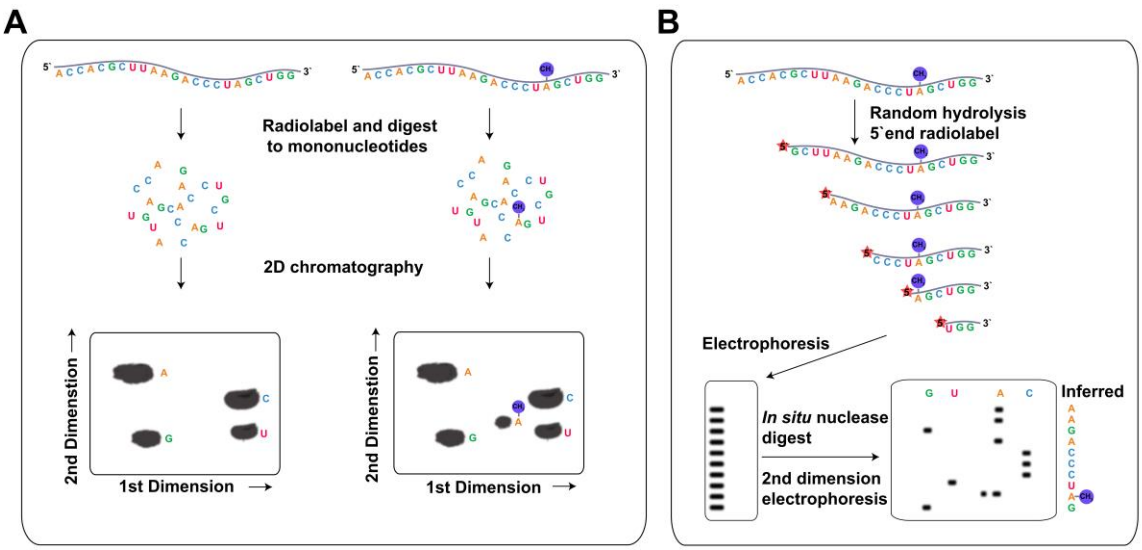

Figure 1.4: Early methods for mapping RNA modifications

(A) Bulk mapping of RNA base modification through complete nuclease digestion followed by 2-dimensional chromatography. Methylated adenosine migrates differently than other nucleotides, and thus appears as a fifth dot. **(B)** Direct sequencing of RNA modifications through random hydrolysis to form a “ladder” of fragments differing in size by one nucleotide. Several representative fragments are shown. Fragments are then radiolabeled, and separated by gel electrophoresis. Fragments are then digested to single nucleotides, and separated again by electrophoresis. Radiolabeled 5’ mononucleotides can be visualized by their different migration patterns.

The study of RNA modification was further aided with the advent of mass spectrometry, and new interest has emerged regarding its application to the direct characterization of RNA modifications (Castleberry et al., 2001; Gaston and Limbach, 2014; Meng and Limbach, 2006; Wetzel and A. Limbach, 2016). Unlike chromatographic methods, mass spectrometry can in principle detect any modification that causes a change in mass, and can even be applied to mass-neutral modifications like pseudouridine through treatment with the pseudouridine-specific adduct CMC (Mengel-Jørgensen and Kirpekar, 2002). Mass spectrometry can also find new modifications in

an unbiased manner by defining mass shifts without any *a priori* knowledge of modification structure. As such, mass spectrometry was critical in the initial characterization of most known modifications (Gaston and Limbach, 2014), and has now begun to uncover novel modifications such as tRNA geranylation (Dumelin et al., 2012), cyclic N^6-threonylcarbamoyladenosine (ct^6A) (Miyauchi et al., 2013), and even novel combinations of known and novel modifications (Dumelin et al., 2012).

Another set of methods relies upon the tendency of RNA modifications to disrupt RNA base pairing and interact with RNA binding proteins. In turn, RNA polymerases such as reverse transcriptase (RT) behave differentially upon encountering a modified base, leading either to base misincorporation or termination of transcription. This has enabled the development of primer extension, a technique that has allowed for base-resolution mapping of modifications by priming with an oligo of homology to an *a priori* defined region, followed by RT extension (**Figure 1.5**). Thus, one of the major advantages of primer extension is that it can target transcripts in a heterogeneous pool of RNA, in contrast to approaches like direct sequencing that require large volumes of high-purity RNA (Motorin et al., 2007). Thus, primer extension is an ideal approach for studying less abundant RNAs, and helped enable the mapping of mRNA modifications. The first primer extension assays were often coupled to gel-based dideoxynucleoside sequencing, in which aberrant RT termination events could be inferred to result from modified ribonucleotides (Brownlee and Cartwright, 1977; Lane et al., 1985). However, distinguishing modification-induced stalling (signal) from normal variation in RT movement (noise) is difficult (Motorin et al., 2007).

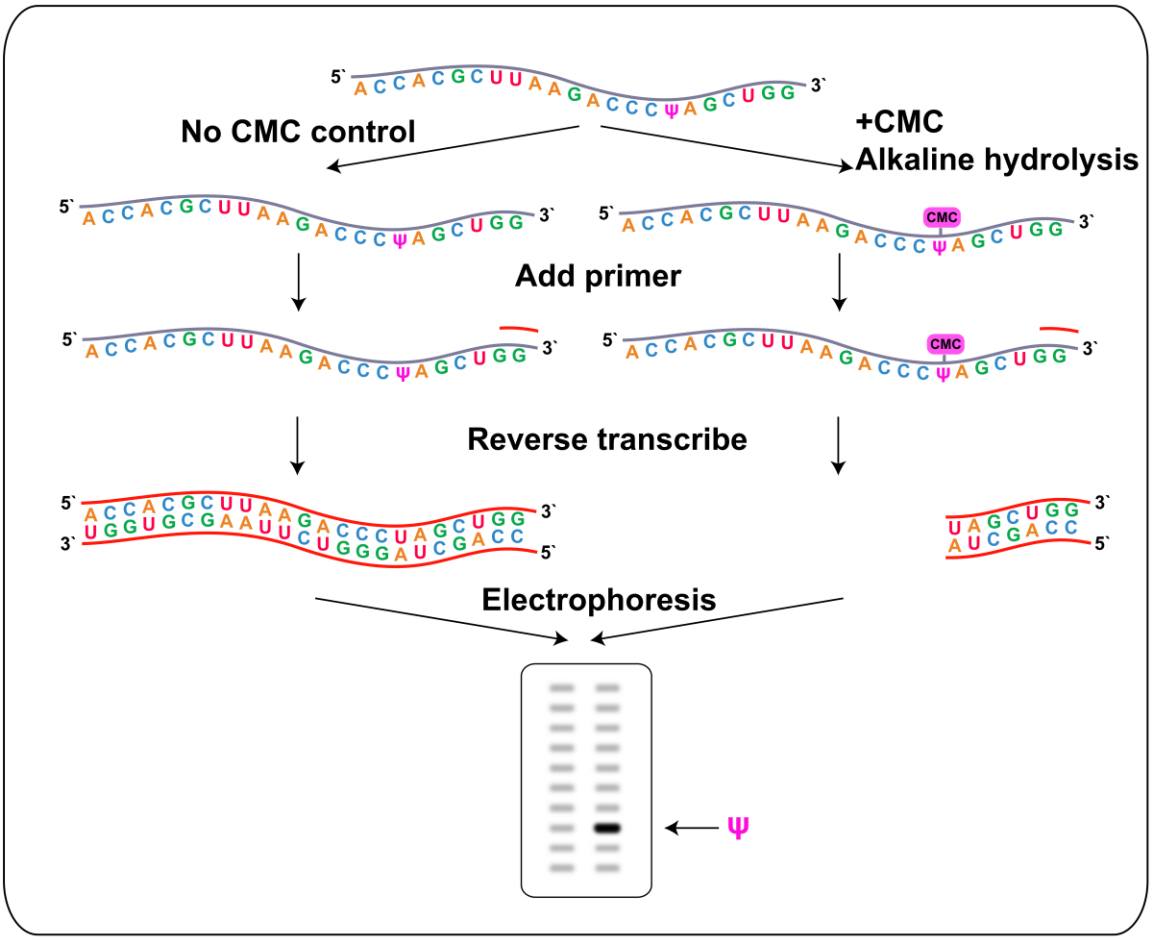

Figure 1.5: Early reverse-transcriptase-based methods for mapping RNA modification

A reverse-transcriptase (RT)-based method involving labelling of pseudouridines with CMC, followed by reverse transcription and observation of stalling sites. Stalls enriched in +CMC over the –CMC control are inferred to be pseudouridylated (darker band).

Later iterations overcame these difficulties through incorporating reagents that specifically target modified ribonucleotides and lead either to large adducts or RNA cleavage. Modifications are inferred from RT stalls that are enriched upon addition of the reagent (**Figure 1.5**). For instance, pseudouridine is known to preferentially react with

the adduct N-cyclohexyl-N'-(2-morpholinoethyl)carbodiimide metho-p-toluenesulphonate (CMC), which forms adducts that induce RT stalls (Bakin and Ofengand, 1993). CMC initially labels all uridines, while alkaline hydrolysis only removes CMC from unmodified uridines. The development of CMC treatment enabled the rapid survey of pseudouridines across rRNAs from all kingdoms of life (Ofengand and Bakin, 1997), and is now the basis for a variety of high-throughput pseudouridine sequencing methods (Carlile et al., 2014; Li et al., 2015; Lovejoy et al., 2014; Schwartz et al., 2014b). Analogously, recent techniques have also relied upon comparing RT stalling in the presence or absence of modification eraser proteins to define sites of m^1A (Li et al., 2016).

1.7.2 Antibody-based global methods

Reliable antibodies have been raised against modified ribonucleotide epitopes, including m^6A, m^1A, pseudouridine, m^5C, and hm^5C. In turn, they have enabled the development of RNA immunoprecipitation (RIP)-based sequencing methods that allow unbiased surveys of these modifications across a transcriptome of interest (**Figure 1.6**). Antibody-based methods were first used to map m^6A via methyl RIP-seq (MeRIP-seq) (Meyer et al., 2012) and m^6A-seq (Dominissini et al., 2012), m^1A via m^1A-seq (Dominissini et al., 2016) and m^1A-ID-seq (Li et al., 2016), and m^5C (Hussain et al., 2013b) and hm^5C via methyl and hydroxymethyl RIP-seq (MeRIP-seq and hMeRIP-seq), respectively (Delatte et al., 2016). Some of these methods involve simple pulldown and sequencing, and are directly analogous to chromatin IP (ChIP) and RNA binding protein

crosslinking and IP (CLIP), drawing upon similar experimental and computational protocols. Others layer on additional chemical treatment and RT-based detection methods, using antibodies primarily to purify out an informative RNA subpopulation.

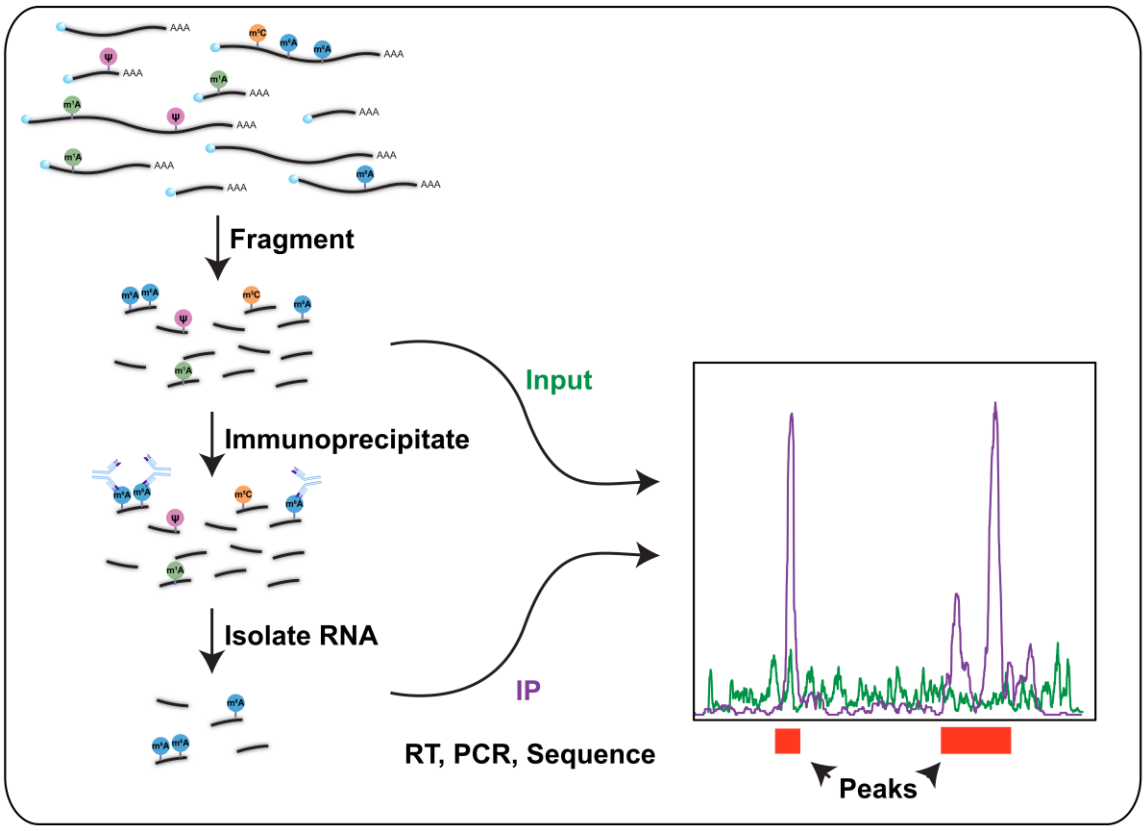

Figure 1.6: Antibody-based methods for mapping RNA modifications

Antibody-based approaches, which rely upon antibodies recognizing modified ribonucleotide epitopes (shown in figure) or epitopes from modification writer proteins. Immunoprecipitated (IP) fragments are sequenced and compared to an input (shown in figure) or isotype control library, from which peaks of significant IP enrichment are calculated.

Simple pulldown methods have bene used extensively and successfully to map m^6A (Batista et al., 2014; Dominissini et al., 2012; Geula et al., 2015; Luo et al., 2014; Meyer et al., 2012; Schwartz et al., 2014a; Zhou et al., 2015) and more recently hm^5C

(Delatte et al., 2016). For this approach, RNA is first fragmented to a suitable size range, before purification with bead-linked antibodies (**Figure 1.6**). A related method involves the use of “suicide inhibitor” nucleotide analogues such as 5-azacytodine, which irreversibly bind its methyltransferase (Khoddami and Cairns, 2013). Modified regions can thus be pulled down through IP of a writer protein. In both methods, RNA fragments are then sequenced and compared to reads from a control library composed of input RNA or RNA immunoprecipitated with an antibody isotype control. Sites of RNA modification are inferred from sequencing read peaks in the modification-specific antibody pulldown as compared to the background control, and thus a major drawback of most antibody-based methods is that modification sites cannot be defined with single nucleotide resolution (**Figure 1.6**).

Nonetheless, several experimental and computational approaches have been taken to improve resolution. A simple approach is to infer modification sites from the presence of consensus motifs within the identified sequence read peaks, although this assumption is vulnerable to false negatives of modification at nonconsensus sites, and false positives when multiple consensus sequences exist in the same peak. More elaborate inferences might also take into account secondary structure, given for instance that m^6A disrupts structure and tends to occur in single-stranded RNA regions (Schwartz et al., 2013). Another approach is to reduce fragment size in order to reduce peak width and improve density (Schwartz et al., 2013), akin to the strategy of treating ChIP samples with exonucleases in the ChIP-exo approach (Rhee and Pugh, 2001). Alternatively, a more direct approach incorporates crosslinking into the RIP protocol in order to define modification sites based upon crosslinking-induced mutations (CIMS)

(Linder et al., 2015). This technique was first applied to m^6A mapping and adapts the method called cross-linking and immunoprecipitation sequencing (CLIP-seq) that was first developed to map RBP-RNA interactions (Kishore et al., 2011), and is thus called m^6A individual nucleotide CLIP (miCLIP) (Linder et al., 2015).

Another approach that bypasses this resolution limit is to couple antibody pulldown to chemical modification and assays of RT stalling. For instance, both m^1A-seq (Dominissini et al., 2016) and m^1A-ID-seq (Li et al., 2016) first utilize anti-m^1A antibodies to pull down methylated RNA fragments. Unlike m^6A, m^1A affects the Watson-Crick base pairing edge and causes reverse transcriptase stalling, which can be used to infer the location of these modifications. To unambiguously define m^1A-induced stalling events, both methods involve comparison to an m^1A-depleted control library, prepared either through *in vitro* addition of demethylases (m^1A-ID-seq) (Li et al., 2016) or through inducing Dimroth rearrangements in which m^1A isomerizes to m^6A and thus no longer blocks RT (Dominissini et al., 2016).

1.7.3 Chemical-based global methods

Combining existing compounds that specifically target or exclude modified ribonucleotides with high-throughput sequencing has yielded powerful, single nucleotide resolution techniques for determining the location of modification sites (**Figure 1.7**). For instance, bisulphite conversion has been used extensively in mapping DNA epigenetic m^5C, and has recently been applied to mapping the same modification in RNA (Hussain et al., 2013a; Squires et al., 2012). In bisulphite sequencing, unmodified cytosines are

converted to inosine, while m^5C is unchanged. Thus, every read gives information regarding the modification status of its cytidines, allowing global and quantitative detection of m^5C. However, resulting reads are also less complex, and global mappability is reduced, leading to potential false negatives. CMC treatment has likewise been used to develop at least four different protocols for global detection of pseudouridine sites, including Pseudo-Seq (Carlile et al., 2014), ψ-seq (Schwartz et al., 2014b), PSI-seq (Lovejoy et al., 2014), and CeU-seq (Li et al., 2015). RT stalling and overall read coverage is then compared in the presence or absence of CMC (**Figure 1.7**).

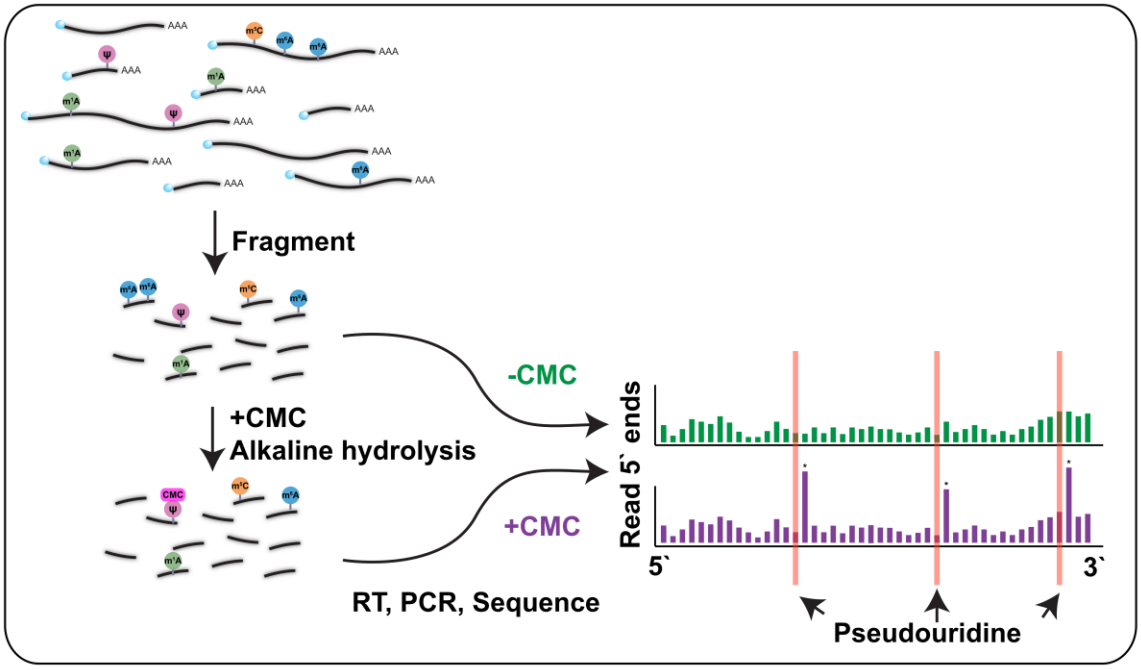

Figure 1.7: Chemical-based methods for mapping RNA modifications

An example chemical adduct-based approach that involves CMC treatment to induce RT stalling at pseudouridines. CMC-treated and -untreated libraries are sequenced, and significant enrichment of RT stalls indicate the presence of a pseudouridine one base upstream (red lines).

1.7.4 In silico *methods*

Even in the absence of additional chemical adducts, chemical modifications that lie at the Watson-Crick base pairing edge will interfere with base pairing and alter the behavior of RNA-dependent polymerases such as RT. Given that most high-throughput RNA sequencing methods rely upon RT for sequencing library preparation, it follows that the presence of modified ribonucleotides will lead to apparent mismatches from the expected sequence. In fact, this was observed when comparing mismatches in small RNA sequencing data to known tRNA modified bases (Ebhardt et al., 2009). This logic underlies High-throughput Annotation of Modified Ribonucleotides (HAMR), a novel *in silico* method for retrospective detection of RNA modifications from any RNA sequencing dataset (Ryvkin et al., 2013) (**Figure 1.8**). This technique recapitulated existing tRNA modifications (Ryvkin et al., 2013; Vandivier et al., 2015a), and found characteristic mismatch profiles for different modification types, allowing nearest neighbor-based prediction of novel modification identity (Ryvkin et al., 2013). Throughout my thesis work, I made extensive use of HAMR and contributed to its development.

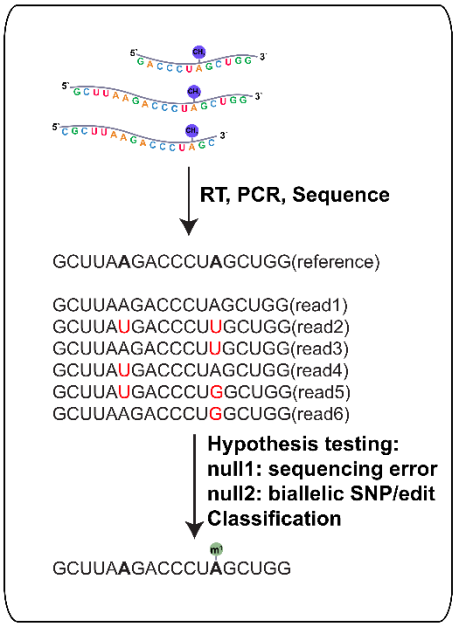

Figure 1.8: *In silico* methods for mapping RNA modifications

In silico determination of modifications using the High-Throughput Annotation of Modified Ribonucleotides (HAMR) pipeline (Ryvkin et al., 2013). Observed mismatches in sequencing data are tabulated, and sites are tested against null hypotheses that 1) sequencing error explains the pattern of mismatches, and 2) biallelic genotypes explain the pattern of mismatches. Sites inferred to be modified are then classified using machine learning trained on known tRNA modifications.

Like patterns of reverse transcriptase stalling, patterns of mismatches can be quite messy and lead to artifacts if not properly controlled. HAMR requires multiple steps to ensure that a set of observed mismatches is not due to sequencing error, alignment algorithm error, or single nucleotide polymorphisms. To account for sequencing error, only reads with high quality score (less than 1/1000 probability of sequencing error) are considered, and bases are only retained if they have significantly more mismatches than expected by sequencing error alone (binomial test). Remaining bases are then tested to ensure that no biallelic genotype can explain the observed pattern of mismatches (ensemble of binomial tests), ruling out RNA editing or polymorphism (Ryvkin et al., 2013). As a result, HAMR is limited to diploid and haploid organisms. Moreover, this relatively high bar for modification calling results in a method with low false positives but high false negative rates, and HAMR is consistently far from saturation (full census of the genome) even at very high read coverage, a problem it shares with other methods like m6A IP. To address this, we define HAMR mods as a proportion of total “accessible bases” with sufficient read coverage for analysis (Vandivier et al., 2015a).

Nonetheless, HAMR has the advantage of being able to probe modifications retrospectively, and can be readily applied to existing data and in meta-analyses. Moreover, it can be applied to specialized library types (such as global mapping of uncapped transcripts (Gregory et al., 2008; Willmann et al., 2014) to survey uncapped, degrading RNAs) that would not normally be amenable to IP or chemical treatment. This has enabled novel observations such as the strong enrichment of modifications in actively degrading transcripts (Vandivier et al., 2015a). Moreover, HAMR can survey multiple modification subtypes simultaneously, so long as they affect the Watson-Crick

base pairing edge, and is currently the only high-throughput technique that can detect modified guanosines in the body of mRNAs (Ryvkin et al., 2013; Vandivier et al., 2015a). HAMR is thus a powerful technique that has multiple applications toward the study of RNA modifications.

1.8 OUTLINE OF DISSERTATION

In this dissertation, I aim to demonstrate new links between the multiple layers of the epitranscriptome and mRNA stability. In **Chapter 2**, I further develop normalization methods for the High Throughput Annotation of Modified Ribonucleotides (HAMR) pipeline to enable direct comparison of predicted modifications from different library types. We then apply this technique to libraries of small RNAs (smRNA-seq), capped and polyadenylated RNAs (RNA-seq), and uncapped degrading RNAs (GMUCT) across both human cells and *Arabidopsis*. In both species, we observe a strong enrichment of HAMR-predicted modifications in uncapped degrading mRNAs and to a lesser extent small RNAs, suggesting a relationship between mRNA decay and covalent modification. Moreover, the number of modifications per transcript correlated with a monotonic increase in the proportion in the uncapped state (proportion decapping), suggesting modifications in uncapped RNAs are a hallmark of unstable transcripts. Finally, we show that transcripts with such modifications tend to be involved in stress response in *Arabidopsis* and cell death in humans, suggesting that these modifications could dampen expression of these transcripts under basal conditions.

In **Chapter 3**, I further develop the link between modifications in uncapped transcripts and mRNA decay by applying the HAMR pipeline to the same three populations of RNA in salt-stressed *Arabidopsis*. With this approach, we are able to identify differentially modified bases that either gain or lose modifications upon salt stress. We then use this as a model system in which to test the effect of HAMR-predicted modifications on mRNA stability, and observe that gain of modifications correlates with higher proportion decapping and less ribosome occupancy, suggesting most modifications destabilize transcripts, possibly by interfering with translation. We then directly test transcript stability with a transcriptional inhibitor-based assay, and show that certain modified bases appear stabilizing, while others appear destabilizing. Notably, some of these differentially modified, differentially stable transcripts are involved in response to salt stress, indicating that differential modification could a part of stress response. Finally, we indicate a possible mechanism for differential modification-induced transcript decay by showing that modifications are strongly enriched for ribosome pausing sites, and that gains of modifications are associated with an increase in co-translational decay.

In **Chapter 4**, I attempt to show a link between mRNA secondary structure and DICER-LIKE-mediated decay in *Arabidopsis*. We develop a method of structure mapping that removes contamination from duplex (intermolecular) RNAs, and then demonstrate that transcripts with high degrees of secondary structure tend to possess regions that are cleaved into small RNAs in a DCL1-dependent manner. We also show that structured region length correlates with DCL1 targeting, suggesting that long regions of structure are more readily processed, consistent with the long miRNA stem loop

structures normally processed by DCL1. We then present preliminary evidence that highly structured transcripts that are cleaved by DCL1 to smRNAs are stabilized upon loss of DCL1, suggesting that mRNA secondary structure can be a direct target for endonucleolytic cleavage.

In **Chapter 5**, I develop a novel approach for detecting partial mRNA degradation using RNA-seq data, and apply this technique to maternal mRNAs in the developing mouse oocyte. We define putative boundary elements that prevent full mRNA decay, and assay for RBPs that could bind to these elements and prevent full exonucleolytic cleavage.

in **Chapter 6** I then discuss the impact of these findings on the field of post-transcriptional regulation and epitranscriptomics, and discuss future directions and open questions.

CHAPTER 2: CHEMICAL MODIFICATIONS MARK UNCAPPED MESSENGER RNAS IN ARABIDOPSIS AND HUMANS

This section refers to work from:

Vandivier L.E., Campos R., Kuksa P.P., Silverman I.M., Wang L.S., and Gregory BD (2015). Chemical Modifications Mark Alternatively Spliced and Uncapped Messenger RNAs in Arabidopsis. *Plant Cell*. 27, 3024-37. PMID: 26561561

2.1 INTRODUCTION

RNA chemical modification is both widespread and physiologically relevant across prokaryotes and eukaryotes. While modifications are best characterized in noncoding transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), mRNAs are also modified with N^6-methyladenosine (m^6A) (Dominissini et al., 2012; Horowitz et al., 1984; Meyer et al., 2012), N^1-methyladenosine (m^1A) (Dominissini et al., 2016; Li et al., 2016), 5-methylcytosine (m^5C) (Squires et al., 2012), pseudouridine (Y) (Carlile et al., 2014; Schwartz et al., 2014b), and 2'-0-methylation of ribose (2'OMe) (Mauer et al., 2016). Additionally, there is a growing body of evidence to support the functional significance of RNA modifications within mRNAs, which is discussed in detail in **Section 1.6**. For instance, spliceosome assembly disruption and changes in mRNA localization were observed upon knockdown of the oxidative demethylase ALKBH5, which removes methyl groups from RNA (Zheng et al., 2013). Furthermore, the presence of certain

methylated bases in human cell lines anti-correlates with mRNA stability (Schwartz et al., 2014a). However, coding and noncoding RNAs likely share the same modifying enzymes (Lee et al., 2014) and specifically testing the function of mRNA modification through genetic ablation of these proteins is difficult. Thus, the functional consequences of most mRNA modifications are still unclear.

The best characterized mRNA modification to date is m^6A, which has known readers, writers, and erasers and thus represents the most complete example of an epitranscriptomic mark. m^6A is enriched around the stop codon, suggesting interplay with the translation and degradation machinery (Meyer et al., 2012). This mark is also enriched at alternatively spliced introns and over long exons (Dominissini et al., 2012), suggesting a role in modulating splicing. m^6A (Liu et al., 2015; Roost et al., 2015) and N^1-methyladenosine (m^1A) (Helm et al., 1999; Zhou et al., 2016) can also disrupt RNA secondary structure, while pseudouridine modifications stabilize secondary structures (Kierzek et al., 2014; Schwartz et al., 2014a; Sundaram et al., 2000), and may do the same in mRNAs in which they are incorporated (Carlile et al., 2014; Schwartz et al., 2014b). Similarly, as tRNA modifications are known to direct cleavage of internally transcribed spacers, mRNA modifications can likewise direct transcript cleavage and subsequent turnover (Du et al., 2016; Hughes and Ares, 1991; Kiss-László et al., 1996; Wang et al., 2014b). Thus, chemical modifications likely have widespread and varied effects across the eukaryotic transcriptome. However, our knowledge of the mRNA modification sites and their functional consequences is currently limited.

Here, we comprehensively identify mRNA modifications using High-throughput Annotation of Modified Ribonucleotides (HAMR) (Ryvkin et al., 2013). HAMR exploits the

tendency of certain covalent RNA modifications, including those known to be common in tRNAs, to interfere with Watson-Crick base pairing and cause reverse transcriptase (RT) to stall and/or misincorporate nucleotides during reverse transcription. This in turn produces a characteristic pattern of RT errors, which present in deep sequencing as mismatches from the reference genome. Working on this premise, HAMR tabulates high confidence (quality score > 30, error probability < 1/1000) mismatches and tests for significance by 1) ruling out that the changes are merely sequencing error and 2) excluding single nucleotide polymorphisms (SNPs) or editing sites (**Figure 2.1**). To this end, we focus on modification-induced errors that have a tri-nucleotide substitution pattern and do not have a clear bias toward any single base misincorporation in order to avoid SNPs and sites of RNA editing (Ryvkin et al., 2013). These stringent filtering steps require high read coverage, and as a result HAMR is designed to minimize false positives at the expense of likely missing a portion of the modified transcriptome. Moreover, modifications such as m^6A, which do not significantly affect the Watson-Crick base pairing edge, will not be detected by HAMR. Nonetheless, this algorithm provides a high-throughput, robust, and generalized *in silico* method to detect RNA modifications that affect Watson-Crick base pairing in eukaryotic transcriptomes. Such HAMR-predicted modifications include but are not limited to 3-methyl cytosine (m^3C); 1-methyl guanosine (m^1G); and 1-methyl adenosine (m^1A) (Ryvkin et al., 2013). This algorithm also incorporates a validated (Ryvkin et al., 2013) machine learning step into the analysis that allows prediction of modification identity (e.g. m^3C) based on the specific tri-nucleotide substitution pattern that we observe at every HAMR-predicted modification site. This analytical approach is based on our previous observation that each type of

covalent RNA modification directs a distinct tri-nucleotide reverse transcriptase (RT) incorporation pattern based on their differential base-pairing properties (Ryvkin et al., 2013).

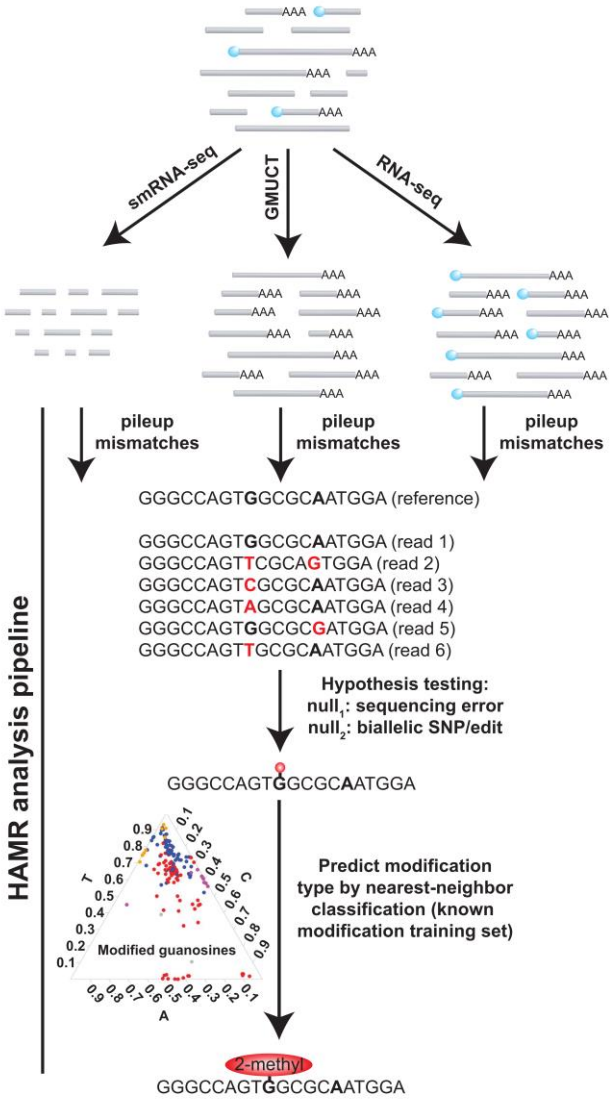

Figure 2.1: Study design to comprehensively identify covalent, HAMR-predicted modifications in the Arabidopsis transcriptome

smRNA, polyA$^+$-selected RNA, and polyA$^+$-selected GMUCT (Gregory et al., 2008; Willmann et al., 2014) libraries were constructed in parallel. GMUCT specifically captures transcripts without a 7-methylguanosine cap (light blue circles). The HAMR analysis pipeline was then run on the resulting datasets. Specifically, reads are mapped to their reference genome, and mismatches (red bases) for each base (bolded bases) are tabulated. After two rounds of hypothesis testing, predicted modifications are then classified, based on a training set of known tRNA modifications from *Saccharomyces cerevisiae*.

Here, we apply the HAMR analysis pipeline to RNA sequencing data for the polyA$^+$ and small portions of the transcriptome (RNA-seq and smRNA-seq, respectively), as well as uncapped and degrading RNAs via global mapping of uncapped and cleaved transcripts (GMUCT) (Gregory et al., 2008; Willmann et al., 2014). We identify, classify, and functionally characterize RNA modifications in *Arabidopsis*, and then test whether the results generalize to human RNAs (**Figure 2.1**). In total, our results provide a global view of HAMR-predicted modifications across eukaryotic transcriptomes, allowing us to begin teasing apart their functional significance in post-transcriptional regulation.

2.2 RESULTS AND DISCUSSION

2.2.1 Using HAMR to predict RNA modification sites that affect the Watson-Crick base pairing edge throughout the Arabidopsis transcriptome

In general, uncapped fragments derived from mRNAs in eukaryotic transcriptomes are generated by decapping or endonucleolytic cleavage, and these RNA fragments are then rapidly recognized and degraded by 5' to 3' (e.g. XRN4) (Gazzani et al., 2004a) and 3' to 5' (e.g. exosome) (Chekanova et al., 2007) exonucleases. Thus, they represent the degrading fraction of the transcriptome (**Section 1.3**). Through Global Mapping of Uncapped and Cleaved Transcripts (GMUCT) (Gregory et al., 2008; Willmann et al., 2014), we surveyed the polyadenylated, uncapped, degrading transcriptome of *Arabidopsis thaliana* (hereafter *Arabidopsis*) unopened flower buds. We then paired this data with small RNA sequencing (smRNA-seq) and polyA$^+$-selected

RNA sequencing (RNA-seq) from this same tissue to identify HAMR-predicted modifications at multiple levels of the plant transcriptome (**Figure 2.1**).

To do this, we ran the HAMR pipeline on the set of uniquely mapping reads from these three RNA-seq approaches (see Materials and Methods). From this analysis, we observed differing numbers of HAMR-predicted modifications for different classes of RNA at the three different levels of the transcriptome. For instance, we found that long noncoding RNAs (lncRNAs) and small nucleolar RNAs (snoRNAs) contained the most HAMR-predicted modifications within the GMUCT dataset, while a few and none were identified when analyzing the smRNA- and RNA-seq datasets, respectively (**Figure 2.2A**). These results suggest that there may be a link between HAMR-predicted modifications and degradation for lncRNAs and snoRNAs. In contrast, HAMR-predicted modifications in miRNAs were most abundant within smRNA-seq compared to GMUCT and RNA-seq datasets (**Figure 2A**). Among mRNAs, we observed an average of 5,368 HAMR-predicted modifications in two replicates of GMUCT data. In contrast, an average of only 58 modifications were observed in two replicates of smRNA-seq, and 27 in four replicates of RNA-seq data (**Figure 2B**). Thus, we observed a strong enrichment of HAMR-predicted modifications within degrading mRNAs, as compared to stable, polyA$^+$ mRNAs (hereafter stable mRNAs) and mRNA-derived smRNAs (**Figure 2B**).

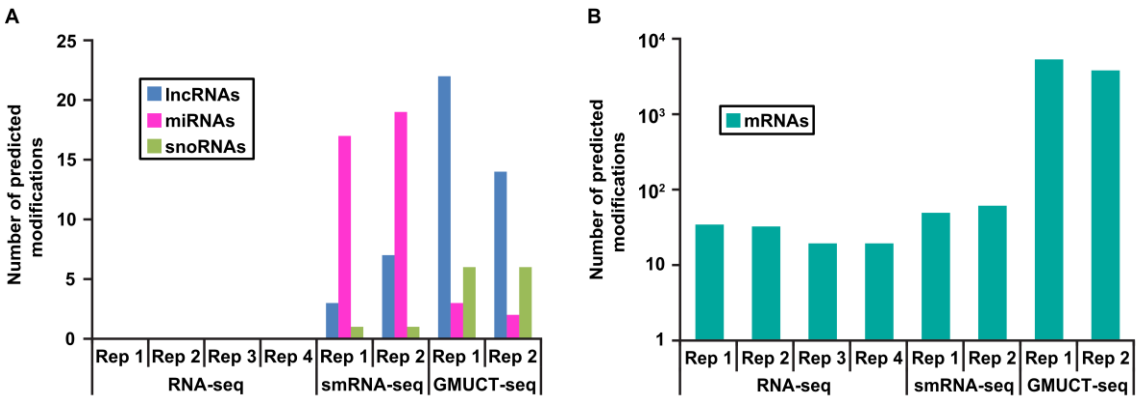

Figure 2.2: HAMR-predicted modifications in Arabidopsis thaliana tend to mark uncapped transcripts

Total number of modifications predicted in (A) noncoding RNAs and (B) coding mRNAs are plotted for each dataset

Interestingly, this strong enrichment of modifications within uncapped, degrading mRNAs as compared to stable mRNAs or mRNA-derived smRNAs was also seen using the same three RNA sequencing data types from two human cell lines (ENCODE Project Consortium, 2012; Huelga et al., 2012; Willmann et al., 2014) (**Figures 2.3A and 2.3B**), suggesting that our observations generalize to other eukaryotic organisms.

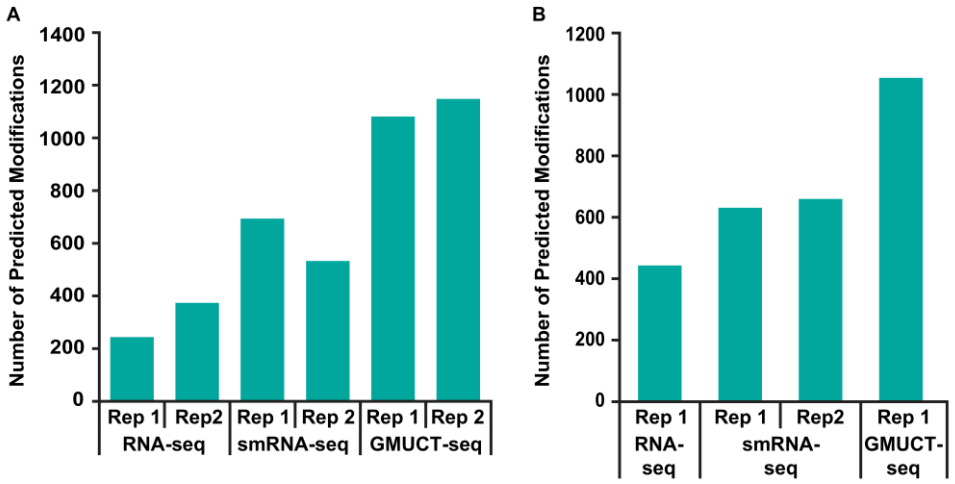

Figure 2.3: HAMR-predicted modifications in human cell lines mark uncapped and alternative spliced transcripts

Total number of HAMR-predicted modification sites from analyzing the three RNA-seq datasets (RNA-seq, smRNA-seq, and GMUCT) for HeLa (A) and HEK293T (B) cells.

Since the statistical power of HAMR depends upon sequencing depth (Ryvkin et al., 2013), we took several approaches to ensure that our observed differences in HAMR-predicted modifications were not artifacts of varying sequencing coverage of transcriptome nucleotides, spurious read mapping, or differential processing of sequencing reads that are a consequence of the different library preparations necessary for each sequencing technique. To first test that potential differences in sequencing coverage of transcriptome nucleotides between libraries was not leading to the differential identification of HAMR-predicted modifications, we downsampled all libraries to equal numbers of uniquely mapping reads. We then looked at the total sequencing read coverage of each nucleotide of the *Arabidopsis* transcriptome. From this analysis, we found that different libraries displayed varying distributions of read coverage, notably

with GMUCT and RNA-seq skewed toward higher read coverage, with GMUCT having a few nucleotides that had extremely high read depth, while smRNA-seq showed lower overall coverage (**Figure 2.4A**). This suggests that GMUCT could have more RNA bases with sufficient read coverage for HAMR to call a modification site ("HAMR accessible bases") than smRNA- and to a lesser extent RNA-seq. From this analysis, we also found that for all three sequencing approaches the minimum coverage at a HAMR-predicted modification site was 50 reads covering that base (**Figure 2.4A, black dashed line**), so we defined "HAMR accessible bases" as those with at least this level of depth. We then normalized total modification number to total "HAMR accessible bases" for the datasets from all three sequencing approaches, and found that mRNAs still have an average of 1207 HAMR-predicted modifications per million accessible bases in GMUCT, compared to 602 in smRNA-seq and 15 in RNA-seq (**Figure 2.4B**). This jump in the number of smRNA-seq predicted modifications suggests that mRNA-derived smRNAs may have more modifications that are simply not called by the HAMR pipeline due to the generally low levels of small RNA processing from mRNAs (**Figure 2.4A**). Since this normalization might not fully control for the proportion of nucleotides that have very high read depth in GMUCT experiments as compared to both RNA- and smRNA-seq (**Figure 2.4A, right hand side of the graph**), we also defined a set of different coverage thresholds (1000, 500, 250, and 100 reads) above which modifications were ignored (**Figure 2.4C**). Again, the major trends in numbers of modifications were not altered, even when setting the upper thresholds to relatively low numbers of sequencing reads (e.g. 100 reads) (**Figures 2.4C**). This discrepancy in HAMR-predicted modifications between the different sequencing approaches was still observed even after combining

this upper limit thresholding with normalization to “HAMR accessible bases” (**Figure 2.4D**). In total, these results indicate that the overall differences in HAMR-predicted modifications between the three RNA-seq approaches are not a consequence of differential sequencing depth at RNA nucleotides.

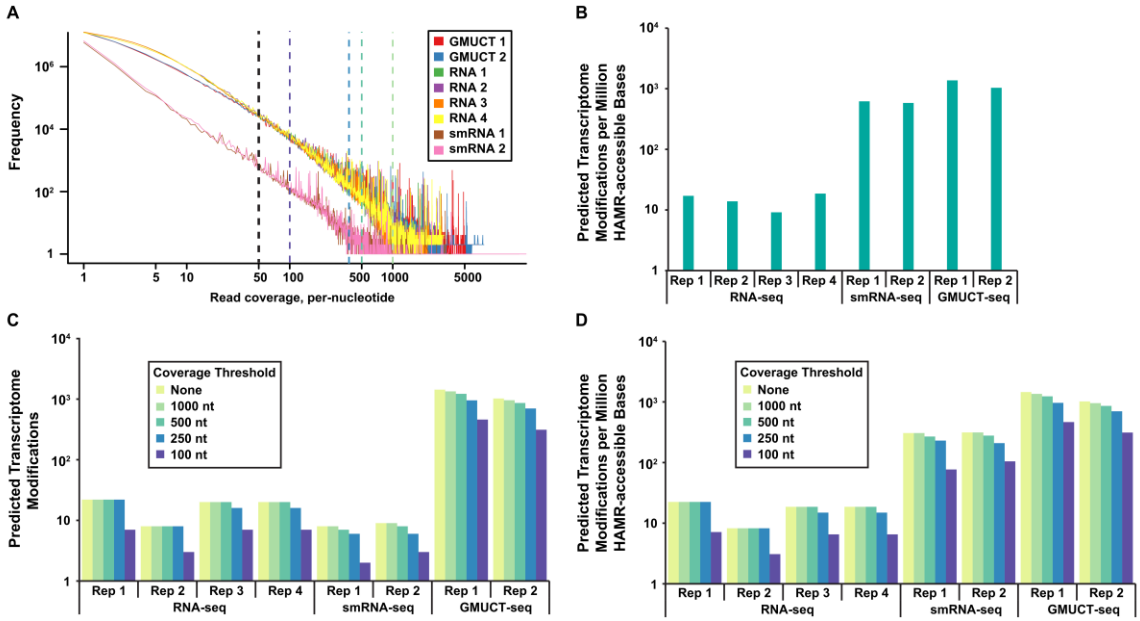

Figure 2.4: Differences in the number of HAMR-predicted modifications are not artifacts of differences in overall size or transcriptome coverage

(A) All Arabidopsis libraries were randomly down-sampled to the number of reads from the smallest library (~3 million), and a histogram of coverage at all TAIR10 mRNA transcriptome bases is plotted in log-log scale. The black dashed line indicates the 50x minimum coverage observed at a HAMR-predicted modification site (“HAMR accessible bases”), and colored dashed lines indicate various maximum coverage thresholds used in Figures 2.4C and 2.4D. (B) Total number of HAMR modifications identified for each RNA-seq dataset were normalized to the number of “HAMR accessible bases” available from those experiments. (C) HAMR was rerun on down-sampled data, and modifications with greater than 100x, 250x, 500x, or 1000x coverage were excluded from the analysis. (D) Total number of HAMR modifications identified for each RNA-seq dataset after down-sampling were normalized to the number of “HAMR accessible bases” available from those experiments, and modifications with greater than 100x, 250x, 500x, or 1000x coverage were excluded from the analysis.

We had previously demonstrated that HAMR results were consistent across an array of high-throughput sequence read mapping software programs even when analyzing the highly repetitive human transcriptome (Ryvkin et al., 2013). However, certain high-throughput sequence read mapping software may produce spurious "uniquely mapping" reads without exhaustively searching for matches across the whole transcriptome. Therefore, although *Arabidopsis* mRNAs do not generally contain large amounts of repetitive sequence, we still controlled for this possibility by repeating our analysis on repeat-masked (Smit, AFA, Hubley, R & Green, P. (2013); RepeatMasker Open-4.0, http://www.repeatmasker.org) data, and observed no change in the number of HAMR-predicted modifications for GMUCT or RNA-seq, and only a slight reduction in the number of modifications on smRNAs (**Figures 2.5, Repeat-masked data**). Finally, the different types of RNA-seq libraries were subjected to different adaptor trimming strategies based on the relation between sequencing read size (50 nucleotide reads) and expected fragment size (see Materials and Methods). To address this, we ran the uniform strategy of concatenating all reads (reads with and without adapter trimming) for all three library types. Once again, treating all libraries the same and analyzing all reads together did not alter the observed trends in differential modification calls between the three different sequencing libraries (**Figures 2.5, All concatenated data**). In total, these control analyses verify that uncapped, degrading mRNAs are strongly enriched for RNA modifications that affect the Watson-Crick base-pairing edge, as compared to stable mRNAs or mRNA-derived smRNAs.

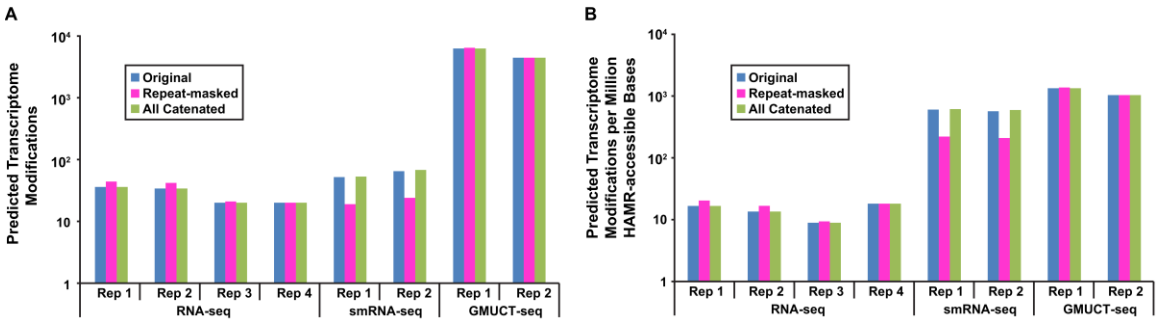

Figure 2.5: Differences in the number of HAMR-predicted modifications are not artifacts of differences in library preparation or spurious designation of unique mappers

(A) To exclude artifacts from mapping and read handling, HAMR was rerun on data from the three RNA-seq approaches that had been mapped to a repeat-masked (Smit, AFA, Hubley, R & Green, P. (2013) RepeatMasker Open-4.0. http://www.repeatmasker.org) TAIR10 transcriptome, and on RNA-seq and smRNA-seq data for which adapter-trimmed and untrimmed reads were concatenated in the same way that was done for GMUCT data (see methods). (B) The same analysis as in A in which the total number of HAMR modifications identified for each RNA-seq dataset were normalized to the number of "HAMR accessible bases" available from those experiments.

2.2.2 Validation of HAMR-predicted modification sites in the Arabidopsis transcriptome

Many of the covalent modifications within yeast tRNAs have been identified and characterized through years of extensive research (G R Bjork et al., 1987; Grosjean et al., 1997; Hopper and Phizicky, 2003; Machnicka et al., 2012; Yacoubi et al., 2012). For this reason, the machine learning algorithm that HAMR uses to classify the type of modification occurring at each predicted site uses the substitution patterns from a yeast smRNA-seq dataset at known tRNA modification sites as its training set (Ryvkin et al., 2013). Furthermore, through homology comparisons of yeast tRNAs to those from other organisms, the orthologous modification sites can be identified (Ryvkin et al., 2013).

Therefore, as a positive control that HAMR was detecting bona fide modification sites in the *Arabidopsis* transcriptome, we derived "known" *Arabidopsis* tRNA modification sites as those with extensive homology to known modified sites in *Saccharomyces cerevisiae*.

Specifically, the yeast data were compiled from the Modomics database (Dunin-Horkawicz et al., 2006), and aligned to *Arabidopsis* tRNAs. Modifications within regions of homology were mapped from yeast to *Arabidopsis* using a custom pipeline incorporating tRNAscan (Lowe and Eddy, 1997) and LocARNA (Will et al., 2007) (see **Appendix A.3.4**). As tRNA loci are highly duplicated, we then filtered our two smRNA-seq datasets to allow multi-mapping reads that align exclusively to tRNAs. Additionally, we cannot unambiguously determine modifications at specific tRNA loci, so we perform all analyses at the level of tRNA family consensus sequences. After running HAMR on two replicates of smRNA-seq, we observed that 23 of 48 (48%) and 24 of 52 (46%) of predicted modification sites correspond to these well-defined modification sites. This level of overlap between HAMR-predicted and known modification sites is significantly (p-value < $1x10^{-7}$, Fisher's exact test) higher than random sampling alone (~11% success rate) (**Figure 2.6A**). To ensure these results are not specific to our library preparation, we also analyzed a species- and tissue-matched smRNA dataset generated by another group (Li et al., 2014), and observed comparable levels of known modification sites identified in tRNAs (p-value < $1x10^{-7}$, Fisher's exact test) (**Figure 2.6B**). Finally, we tested the true positive rate versus the false positive rate at various threshold settings (receiver operating characteristic) for HAMR identification of these known tRNA modification sites (see Materials and Methods), which confirmed the ability of HAMR to identify known modification sites in *Arabidopsis* tRNAs (AUC = 69.87)

(**Figures 2.6C and 2.6D**). Thus, HAMR identifies a significant number of tRNA modification sites in the *Arabidopsis* transcriptome with known homology to yeast, demonstrating its predictive power for studying these covalent additions in plant RNA.

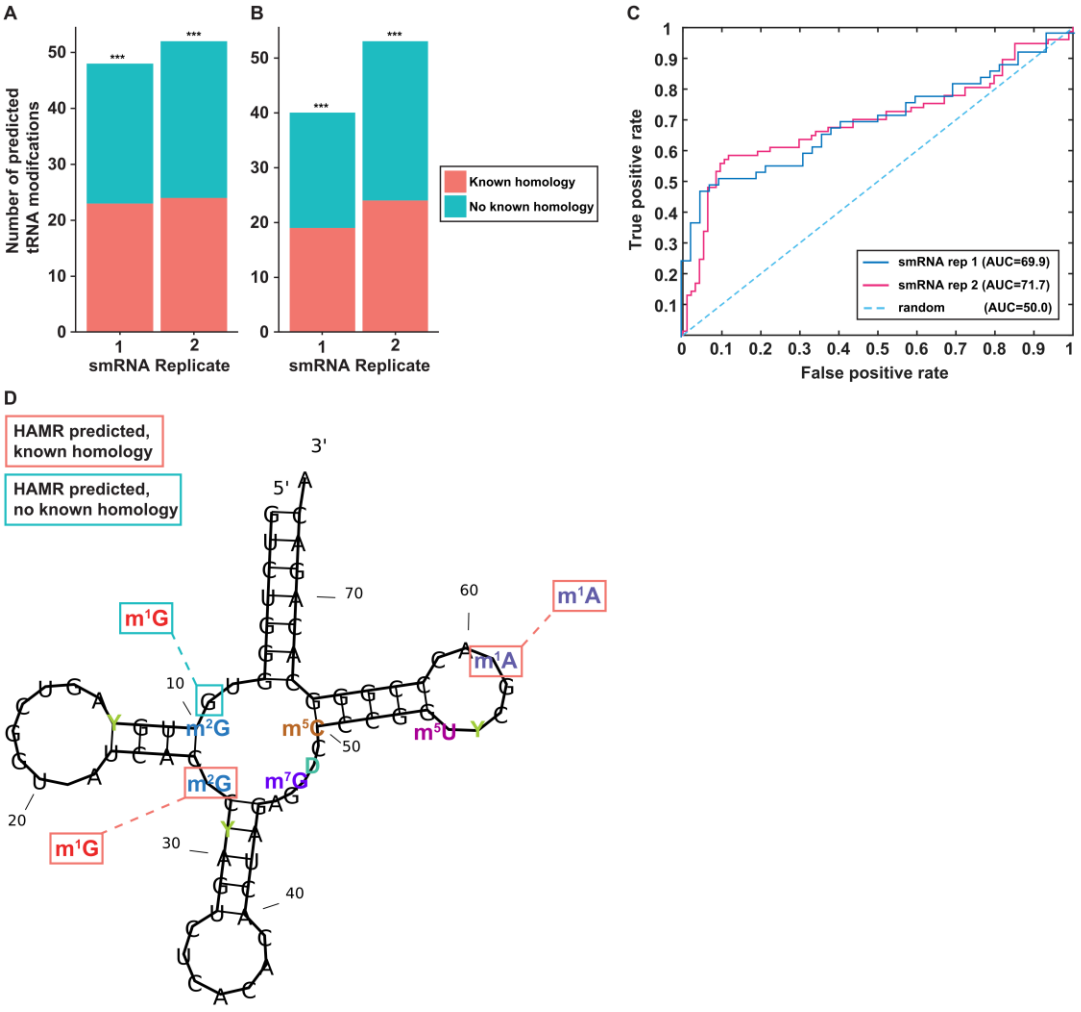

Figure 2.6: HAMR captures a large proportion of known tRNA modification sites in the Arabidopsis transcriptome

HAMR modifications from (A) our smRNA sequencing data and (B) a previously published, tissue matched smRNA sequencing dataset (Li et al., 2014) are overlapped with known tRNA modifications, as determined by homology to yeast tRNAs. The total number of HAMR-predicted modifications are plotted on the y-axis. P-values were calculated by Fisher's exact test, over a background of all tRNA consensus bases (see methods). *** denotes p-value < $1x10^{-7}$. (C) Receiver operating characteristic curves for datasets from both replicates of our smRNA-seq experiments. AUC = area under curve. (D) An example tRNA, *tRNA-Val* (anticodon:CAC), with known modifications labeled as bold, colored letters across the structure backbone (black line). HAMR-predicted modification sites are labeled as known (red boxes) or novel (light blue boxes) with boxes across the structure backbone, while HAMR predicted modification types at those predicted nucleotide positions are shown as outlying boxes connected with dashed lines.

HAMR takes advantage of the propensity of RT to misincorporate nucleotides at modification sites that affect the Watson-Crick base pairing edge. However, another consequence of RT encountering such a modification is to stall, terminate elongation, and fall off the template (Foley et al., 2015). For this reason, such blocks to RT extension have been used for previous identification of covalent modifications to tRNA molecules (Talkish et al., 2014; Woodson et al., 1993). Therefore, to further validate HAMR-predicted modification sites in *Arabidopsis* mRNAs, we tested whether these specific nucleotide positions coincide with reverse transcriptase (RT) stalls that were recently identified in the control samples for dimethyl sulphate (DMS) sequencing (Structure-seq) experiments (Ding et al., 2014). Unlike our RNA-seq data, these Structure-seq libraries are not fragmented, and unambiguously define RT stalls at the 5' terminal nucleotide of their sequencing reads (Ding et al., 2014). Importantly, these Structure-seq control datasets measure RT extension inhibition in the absence of DMS treatment, which indicates they are unrelated to the addition of exogenous DMS adducts and are specifically measuring blocks to normal RT extension by the presence of an RNA modification that affects the Watson-Crick base pairing edge. Using this approach, we found that HAMR-predicted modification sites in the degrading fraction of mRNAs identified by GMUCT significantly coincide with RT extension inhibition sites (all p-values $< 1x10^{-20}$, Fisher's exact test) (**Figure 2.7A**), and relatedly overlap with a greater number of RT stalls per site (all p-values $< 1x10^{-39}$, Wilcoxon Rank Sum test) (**Figure 2.7B**) as measured in the DMS control experiments compared to a background of all mRNA bases. In total, these findings provide strong evidence that HAMR detects bona fide

modification sites in *Arabidopsis* mRNAs, and that this class of covalent additions are enriched in the degrading fraction of these molecules.

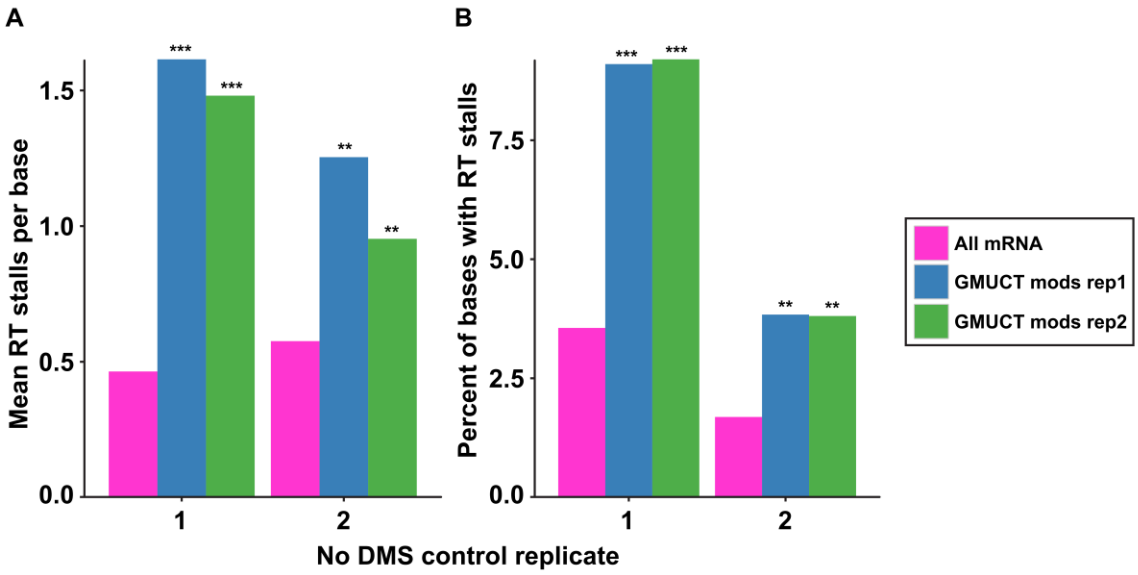

Figure 2.7: Sites of HAMR-predicted modifications are enriched in reverse transcriptase (RT) stalls

RT stalls from no DMS control experiment datasets for Structure-seq (Ding et al., 2014) are tabulated across all mRNA bases (red bars), and across mRNAs predicted to contain modifications based upon GMUCT sequencing (blue and green bars). (A) The mean RT stalls per base and (B) the percent of bases with any number of RT stalls are plotted. Significance was determined for A with a Wilcoxon Rank Sum test (mean RT stalls per base) and for B with a Fisher's exact test (percent of bases with RT stalls) over a background of all mRNA bases. ** denotes p-value < $1x10^{-20}$ and *** denotes p-value < $1x10^{-50}$.

2.2.3 Characterization of HAMR-predicted modifications in the Arabidopsis transcriptome

To better understand the potential functions of HAMR-predicted RNA modifications, we determined whether they were enriched in any particular regions of

Arabidopsis mRNA molecules. From this analysis, we found that modifications called using HAMR on *Arabidopsis* GMUCT data tended to localize within the coding sequence (CDS) and 3' untranslated region (3' UTR), whereas HAMR-predicted modifications from the RNA-seq datasets were almost exclusively localized to introns (**Figure 2.8A**). In regards to the human transcriptome, we found that these results for the GMUCT and RNA-seq datasets are entirely recapitulated in both HEK293T and HeLa cell lines (**Figure 2.8B**). Furthermore, modifications in mRNAs called by HAMR using the HEK293T and HeLa smRNA-seq dataset are mostly found in mRNA introns, where the majority of human miRNA stem-loop precursors are known to reside (**Figure 2.8B**). In contrast, modification sites in *Arabidopsis* mRNAs identified by HAMR using smRNA-seq data display no real bias toward any specific mRNA region (**Figure 2.8A**), consistent with the relative paucity of miRNA precursors residing in *Arabidopsis* introns or other mRNA sequences.

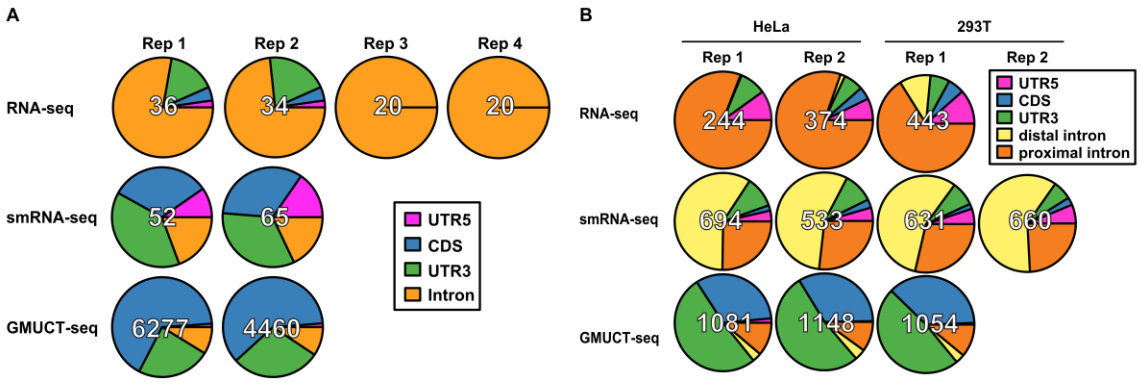

Figure 2.8: HAMR-predicted modifications from different RNA populations mark different transcriptomic regions

Relative transcript location of predicted modifications in mRNAs in (A) *Arabidopsis* and (B) human cell lines. Modifications that lie outside of mRNAs are excluded from this analysis. Intronic modification sites are proximal if within 500 nucleotides (nt) of a known constitutive or alternative splice donor/acceptor site, and distal if further than 500 nt from these sites. *Arabidopsis* introns are short and thus proximal/distal intron classification is omitted.

Intriguingly, a closer inspection of all of HAMR-predicted modification sites in stable mRNAs identified using the RNA-seq datasets from both *Arabidopsis* and humans revealed that these covalent additions are significantly enriched (all p-values < $1x10^{-12}$, Fisher's exact test) in or near introns annotated as being alternatively spliced (**Figure 2.9**). Analysis of an expanded *Arabidopsis* transcriptome annotation (atRTD) (Zhang et al., 2015) yields comparable results (**Figure 2.9A**).

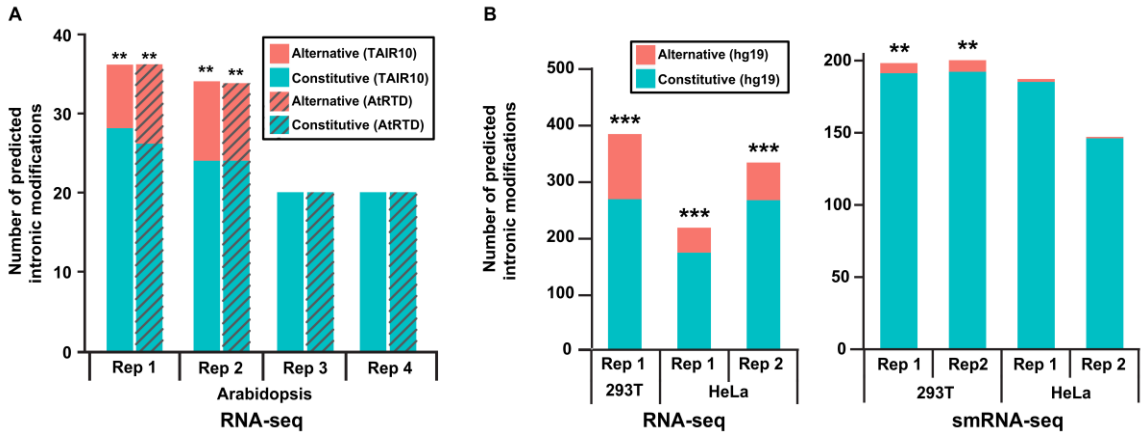

Figure 2.9: HAMR-predicted modifications mark alternatively spliced introns

Localization of modifications to alternative versus constitutive introns in A) *Arabidopsis* and B) humans. Enrichment was calculated with a Fisher's exact test. ** denotes p-value < $1x10^{-10}$ and *** denotes p-value < $1x10^{-50}$. Analysis was performed using transcriptome annotations from TAIR10 (solid bars) or AtRTD (hatched bars) (Zhang et al., 2015) in *Arabidopsis* and UCSC hg19 in human cells.

Furthermore, seven modification sites identified with both RNA-seq replicates 1 and 2 lie within the splice donor site (first six nucleotides) of introns within *AT1G3710*, *AT4G19110*, *AT4G25080*, and *AT4G38510* (**Figure 2.10A)**. It is worth noting that even those that are currently annotated as constitutively spliced introns are most likely novel retained intron events given that they can be captured by a polyA$^+$-selected RNA-seq approach. In support of this idea, over 50% of the HAMR-predicted modification sites lie within the *Arabidopsis* ribosomal protein L3 gene (*AT1G43170*), which has 9 annotated isoforms and a known retained intron event within the 3' UTR, as well as a novel retained intron in the 5' UTR identified by our analysis here (**Figure 2.10A**). Similar examples exist for other transcripts with modifications predicted by HAMR using the

RNA-seq data (**Figure 2.10A**), but are less common for transcripts with modifications predicted by analyzing data from the GMUCT approach (**Figure 2.10B**).

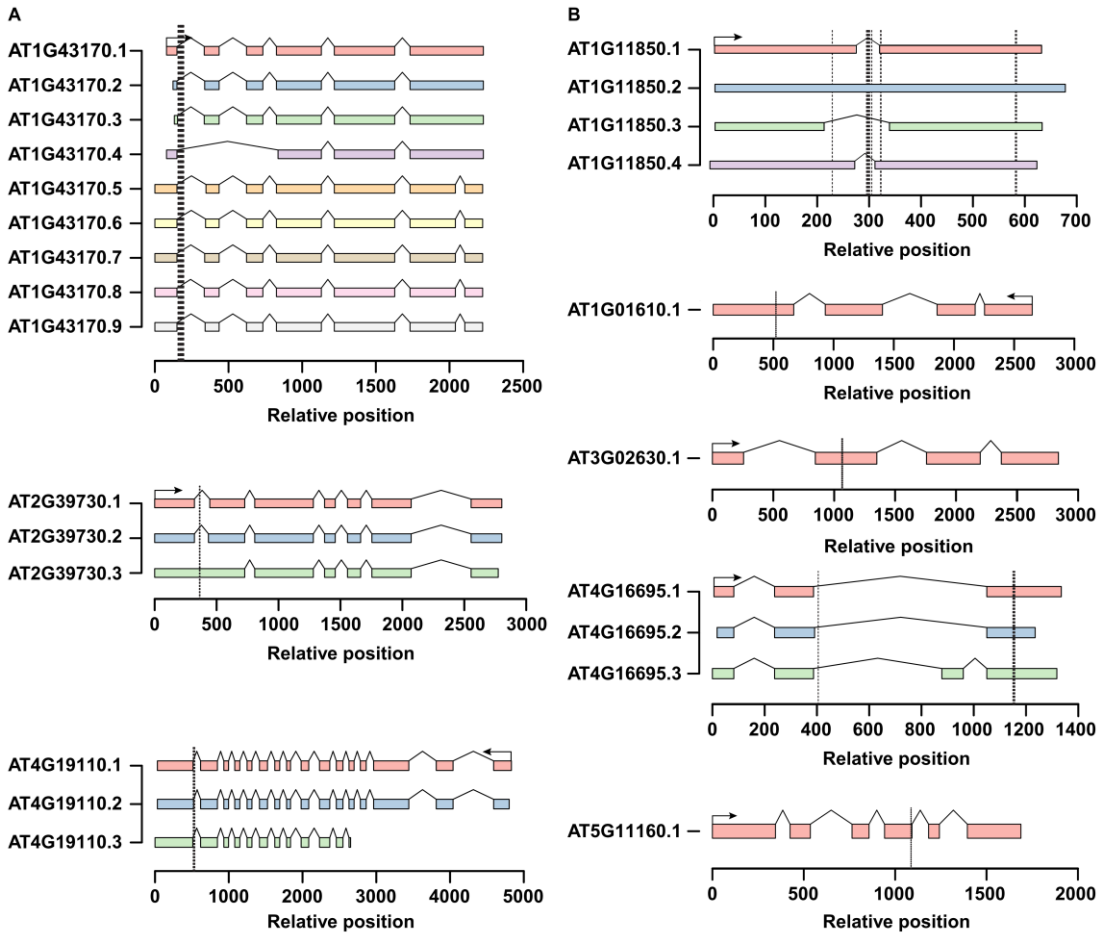

Figure 2.10: HAMR-predicted modifications mark various transcriptome features

HAMR modifications predicted in A) three specific *Arabidopsis* transcripts with HAMR-predicted modifications identified by analyzing GMUCT datasets (uncapped RNAs). B) Five specific *Arabidopsis* transcripts with HAMR-predicted modifications identified by analyzing the RNA-seq datasets (stable mRNAs). For both A and B, the vertical dashed, black lines indicate the relative position of each modification. In plus strand transcripts, relative position 0 indicates the very 5′ end. In minus strand transcripts, relative position 0 indicates the 3′ end. All known splice variants of these seven transcripts are shown in these figures.

We also observed a significant enrichment (p-value → 0, Fisher's exact test) of HAMR-predicted modifications identified in human stable mRNAs using the human RNA-seq data within introns that were annotated to be alternatively spliced (ENCODE Project Consortium, 2012; Huelga et al., 2012). However, this bias was either much less common or was not observed for HAMR-predicted modifications identified using the smRNA-seq data from the two different cell lines for this analysis (**Figure 2.8B**). In total, our findings for HAMR-predicted modifications identified in both *Arabidopsis* and human stable mRNAs using RNA-seq data suggests a role for this class of modifications in regulation of alternative splicing. This hypothesis is further supported by the fact that most of these modification sites are proximal to the splice donor/acceptor sites of these alternatively spliced introns (**Figures 2.11**), with some lying directly within donor site sequences. In total, these results reveal that modifications in uncapped, degrading mRNAs are prevalent in the CDS and 3' UTR, while those in stable transcripts are associated with specific alternative splicing events in both plants and humans. It is noteworthy that another RNA chemical modification, m^6A, has also been found to cluster near specific alternatively spliced exons and introns (Dominissini et al., 2012). Taken together, this combination of findings suggests that in general RNA modifications in stable mRNAs may play a significant role in regulating the processes of alternative splicing in eukaryotic transcriptomes. This hypothesis will require further testing.

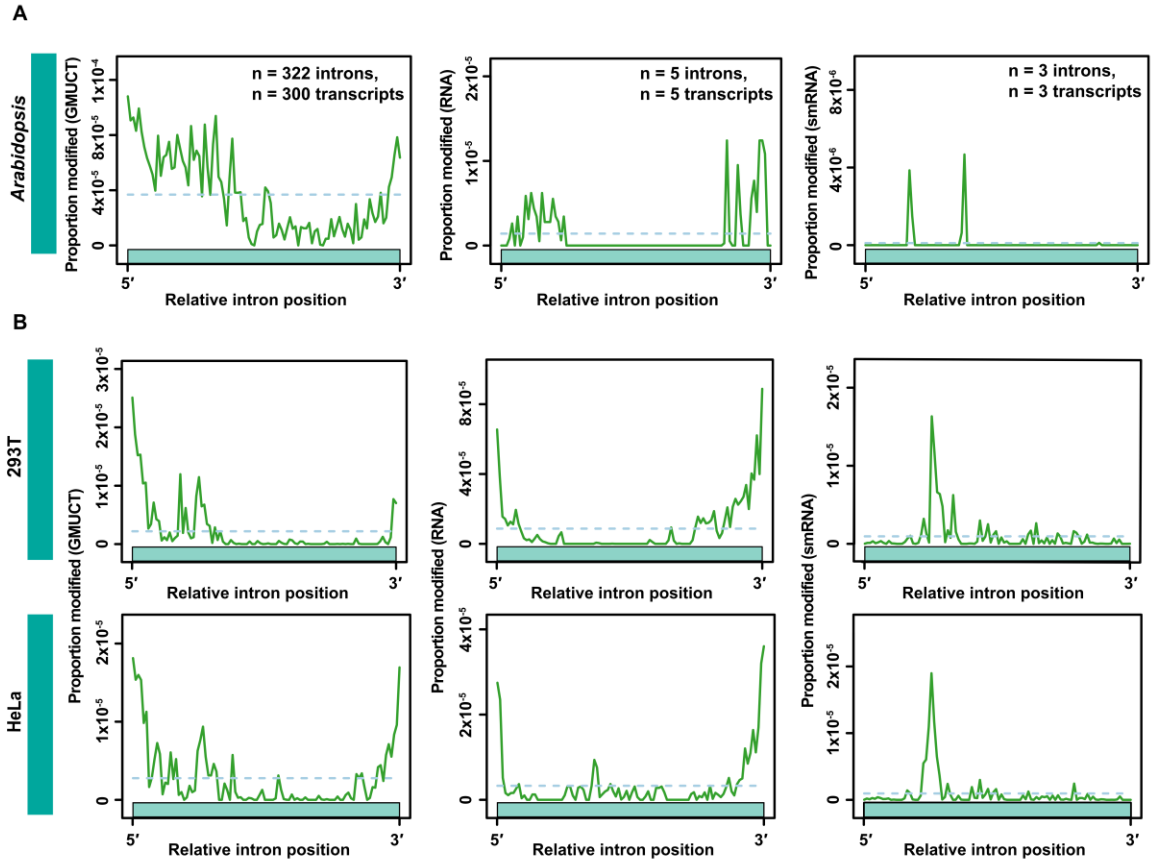

Figure 2.11: HAMR-predicted modifications mark intron termini

From left to right, the relative position of intron-localized HAMR-predicted modification sites using the data from GMUCT RNA-seq, and smRNA-seq are plotted across the length-normalized average of all introns in A) *Arabidopsis* and B) human cell lines.

2.2.4 Uncapped and stable mRNAs contain different proportions of specific RNA modifications

As described above, the HAMR analysis pipeline includes a step to determine the actual modification at each predicted site based on a machine learning approach where known modification sites in yeast tRNAs are used as the training set (Ryvkin et

al., 2013). As a first test that this approach could identify the actual modification at predicted sites in *Arabidopsis*, we tested if the classifier would call the correct identity at "known" modification sites as determined by homology with yeast tRNAs (**Figures 2.6A and 2.6B**). From this analysis, we found that the HAMR modification classifier correctly predicted the exact modification type at ~50% of these known modification sites in *Arabidopsis* tRNAs (**Figure 2.6D**). Therefore, we were comfortable using this approach to determine the identity of the specific modifications predicted using the three different RNA-seq approaches.

Using this machine learning-based classifier (**Figure 2.1**), we identified a wide range of modification types in both noncoding (**Figure 2.12A**) and coding RNAs (**Figure 2.12B**). Interestingly, the modification types between different classes of RNAs (lncRNAs, miRNAs, snoRNAs, and mRNAs) were quite distinct in their total quantities, but in general mostly consisted of the same few types of modifications. The most common types of modifications that HAMR could distinguish were m^3C, Y, m^1A, m^1G, dihydrouridylation (D), N^6- isopentenyladenosylation (i^6A), and threonylcarbamoyladenosylation (t^6A). In lncRNAs, D and Y sites were only identified for HAMR-predicted modification sites found with GMUCT data (**Figure 2.12A**), while m^1G, i^6A/t^6A, m^3C, and m^1A sites were found using both GMUCT and smRNA-seq data. In miRNAs, we revealed that Y, m^1A, i^6A/ t^6A, and m^2G are only observed in smRNA-seq data, but the modification sites identified with the GMUCT data were classified mostly as m^1G or D (**Figure 2.12A**). For snoRNAs, we uncovered only a single predicted m^3C site in both replicates. Conversely, HAMR-predicted modification sites for the GMUCT datasets were a mix of m^1A, i^6A/t^6A, D, Y, and m^3C (Figure 3A). In total, these results

reveal that different collections of modifications that affect Watson-Crick base pairing are found in non-coding RNAs, including lncRNAs, that have been processed into smRNAs, as compared to those that are uncapped.

In coding mRNAs, we found that the identified modifications included previously characterized adenosine methylation (m^1A) and Y sites (Carlile et al., 2014; Schwartz et al., 2014b; Squires et al., 2012), as well as novel cytosine (m^3C) and guanosine methylation (m^1G), dihydrouridylation (D), N^6- isopentenyladenosylation (i^6A), and threonylcarbamoyladenosylation (t^6A) (**Figures 2.12B and 2.12C**). As in noncoding RNAs, the distribution of these modification types is distinct between stable RNA, smRNA, and uncapped, degrading transcripts. For instance, m^3C and m^1G modifications tend to be much more common in stable RNAs and mRNA-derived smRNAs, respectively, as compared to the overall distribution of these covalent additions in uncapped, degrading transcripts identified by GMUCT in both *Arabidopsis* and human data (Figures 3B and S6). Conversely, uncapped, degrading mRNAs as identified by HAMR analysis of GMUCT data demonstrate much higher levels of D and i^6A/t^6A as compared to stable mRNAs and mRNA-derived smRNAs in both plants and humans (**Figures 2.12B and 2.12C**), suggesting that these modifications may be the cause or consequence of protein-coding transcript turnover in eukaryotic transcriptomes. In total, these results reveal that the different collections of transcripts in eukaryotic transcriptomes are marked by distinct distributions of covalent modifications that affect the Watson-Crick base pairing edge.

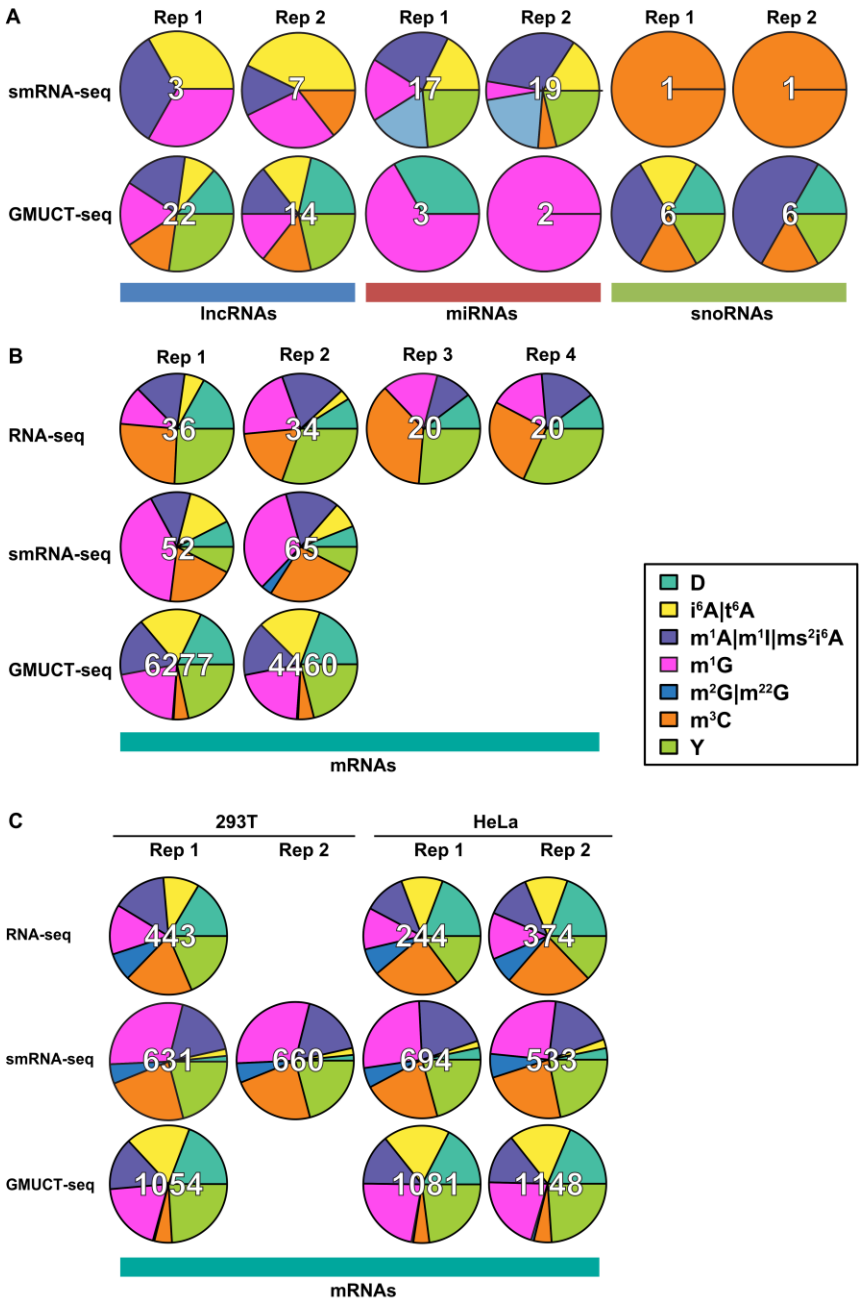

Figure 2.12: HAMR predicts a variety of known and novel modification types

Distribution of the predicted identity of HAMR modifications in (A) *Arabidopsis* noncoding RNAs, (B) *Arabidopsis* coding mRNAs, and C) human coding mRNAs, as determined by nearest-neighbor classification using a training set of known tRNA modifications from *Saccharomyces cerevisiae*.

To experimentally validate both HAMR and the machine learning-based prediction of modification identity, we performed m^3C RNA immunoprecipitations (IP) on RNAs predicted to contain this modification alongside negative controls with no predicted m^3C. Using reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) on fractions of RNAs immunoprecipitated with either an antibody specific for m^3C or an IgG control, we measured the abundance of two mRNAs predicted to contain m^3C using the RNA-seq data, five mRNAs predicted using the GMUCT data, and six mRNAs which were not predicted to contain such modification sites in any of the HAMR analyses (**Figure 2.13**). We normalized qPCR measurements in the two IP fractions to *tRNA-ala* (anticodon:AGC), which is known to be devoid of m^3C in all other eukaryotic organisms, and which HAMR does not predict to contain m^3C in *Arabidopsis*. Thus, this RNA serves as the most confident negative control locus for our analyses. We found that six of the seven transcripts tested (86%) were significantly (all p-values < 0.01, Student's t-test) enriched in the m^3C fractions, compared to the nonspecific antibody control (**Figure 2.13**). Notably, one of these transcripts (*AT4G25080*) contained a predicted m^3C site within the splice donor sequences (**Figure 2.10**). For the one mRNA (*AT2G15580*) that was predicted to contain an m^3C site but that was not validated by this approach, this result could be a consequence of an incorrect modification site call (part of the 5% false discovery rate) or misclassification by the machine learning approach of the HAMR pipeline. Regardless, 86% of the predicted m^3C sites could be experimentally validated, providing evidence for the robustness of the identification and classification of modification sites by the HAMR approach (**Figure 2.13**). For the putative negative

control loci (those predicted not to contain an m^3C site), we found that all of these RNAs had similar or significantly (all p-values < 0.01, Student's t-test) lower levels in the m^3C IP fractions as compared to the IgG control (**Figure 2.13**). These results supported the HAMR prediction that these loci truly lack an m^3C modification site. In total, these results indicated that in general HAMR identified and classified bona-fide covalent modification sites that affect the Watson-Crick base pairing edge within the *Arabidopsis* and human (Ryvkin et al., 2013) transcriptomes, and that these modifications are enriched within degrading mRNAs.

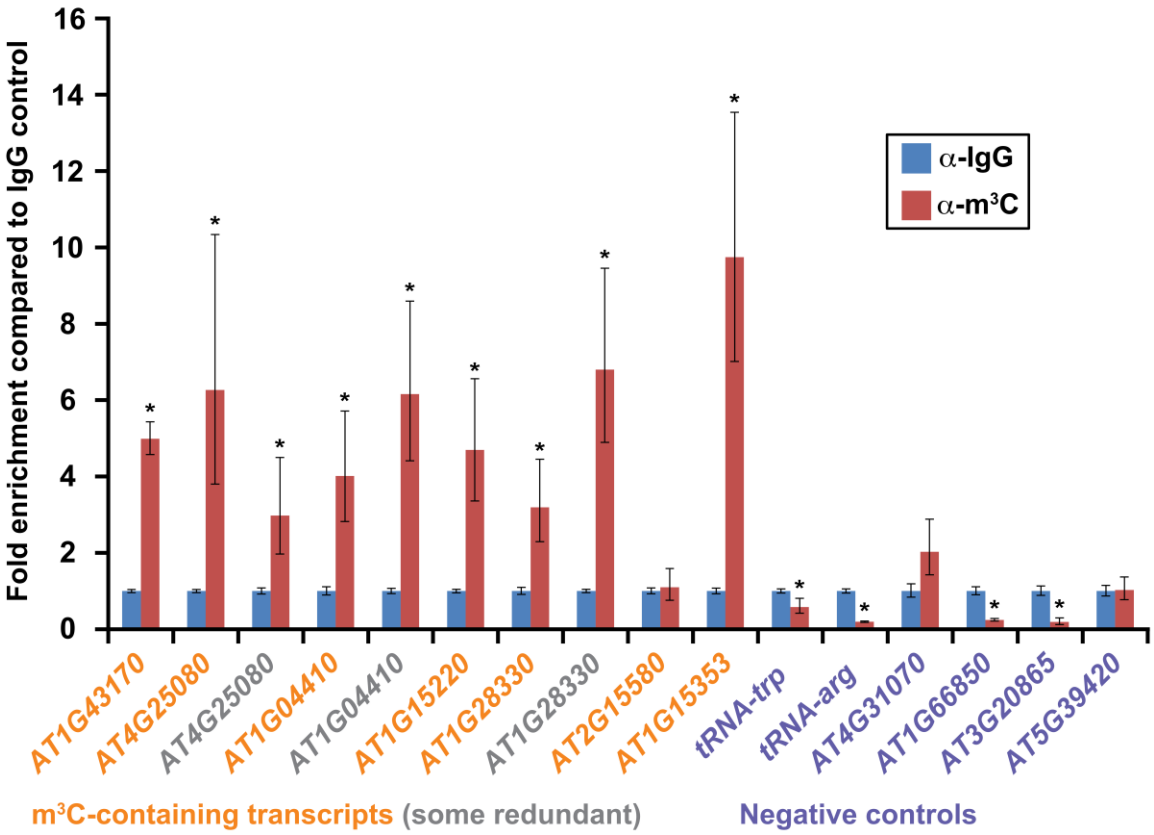

Figure 2.13: Validation of HAMR predicted 3-methylcytosines

Immunoprecipitations of transcripts predicted to contain m^3C modifications. qPCR analysis of two transcripts (*AT1G43170*, *AT4G25080*) predicted to contain m^3C based upon RNA-seq data, five transcripts (*AT1G04410*, *AT1G15220*, *AT1G28330*, *AT2G15580*, *AT3G15353*) predicted to contain m^3C based upon GMUCT, and six transcripts/tRNA families (*tRNA-Arg* (anticodon: AGT), *tRNA-Trp* (anticodon: CCA), *AT1G66850*, *AT3G20865*, *AT4G31070*, and *AT5G39420*) not predicted to contain m^3C. The qPCR data for all transcripts was normalized to *tRNA-ala* (anticodon:AGC), which is well known to not contain m^3C in any other organism, making it the most reliable negative control. Fold enrichment over an IgG nonspecific antibody control (y-axis) is plotted for each transcript. qPCRs were performed in at least duplicate. P-values were calculated with a Student's t-test, as previously described (Ryvkin et al., 2013). * denotes p-value < 0.05.

2.2.5 The proportion of uncapped transcripts and number of HAMR-predicted modifications positively correlate for Arabidopsis mRNAs

We found that uncapped, degrading transcripts as interrogated by GMUCT were the most enriched class of transcripts for HAMR-predicted covalent modifications within our analyses (**Figures 2.2 and 2.3**). Therefore, we wanted to test whether these Watson-Crick base pairing edge affecting modifications correlate with the proportion of steady state transcripts in an uncapped state (proportion uncapped) (**Figures 2.14**), as measured by GMUCT reads (steady state uncapped population) normalized to RNA-seq reads (steady state total transcript population). We have previously used this measure as an approximation of the overall percentage of transcripts that are undergoing turnover (Li et al., 2012a), and in **Chapter 3** demonstrate that it is a valid proxy for mRNA stability. Using this approach, we observed a monotonic increase in the total levels of transcripts that are found in the uncapped and likely degrading fraction of transcripts as the number of predicted modification sites in mRNAs increases (**Figure 2.14A**). Interestingly, the majority of these stepwise increases were significant (all p-values < 0.01, Wilcoxon Rank Sum test), and comparison of all transcripts containing HAMR-predicted modifications to all transcripts that are not identified as containing these modifications also yields highly significant differences ($p \rightarrow 0$, Wilcoxon Rank Sum test). Furthermore, we observed the same trends across two independent replicates of GMUCT and RNA-seq (**Figure 2.14A**). Similar trends were also observed in human (HEK293T and HeLa) cells, though not all stepwise comparisons reached detectable significance in our analyses (**Figure 2.14C**).

Interestingly, modified lncRNAs and snoRNAs, but not miRNAs, likewise showed a similar trend where transcripts with HAMR-predicted modifications had a higher proportion of their populations in the uncapped, degrading portion of the transcriptome as compared to those without these covalent additions, although not at detectable significance. However, this lack of significance is most likely a consequence of the low numbers of detected modification sites in these classes of RNAs (**Figures 2.2A and 2.14B**). In summary, these findings reveal that higher levels of HAMR-predicted covalent modifications in mRNAs in both plants and humans correlate with increased proportions of those transcripts in the uncapped, degrading fraction of transcripts as measured by GMUCT. In total, these findings suggest that covalent RNA modifications that affect the Watson-Crick base pairing edge are a cause or consequence of mRNA turnover in eukaryotic transcriptomes.

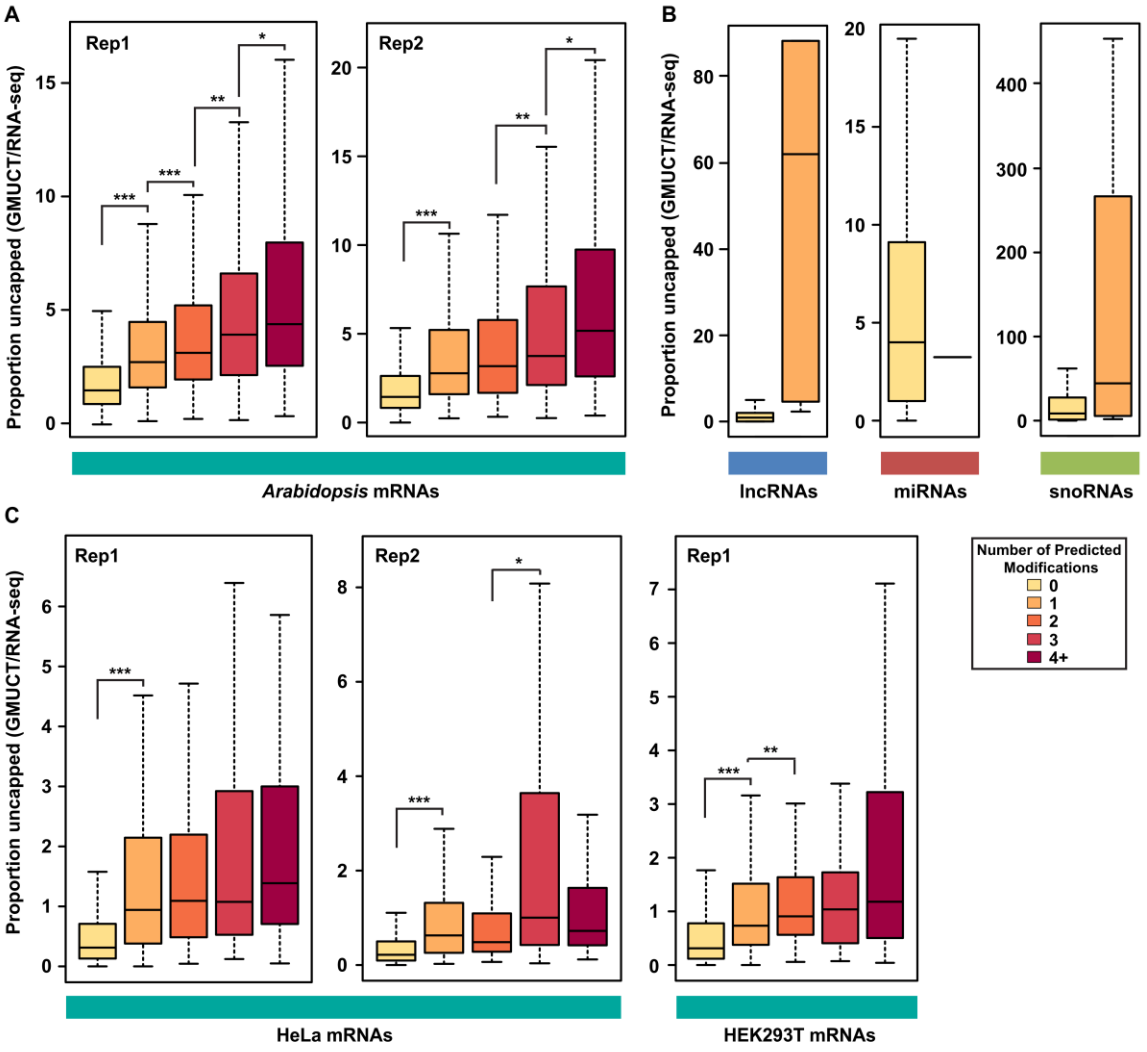

Figure 2.14: mRNAs with HAMR-predicted modifications have higher levels of uncapped transcripts

Distribution of proportion uncapped (total GMUCT reads per transcript normalized to total RNA-seq reads) per transcript for (A) *Arabidopsis* coding mRNAs, (B) a representative replicate for *Arabidopsis* noncoding RNAs, and C) human coding mRNAs. P-values were calculated with a Wilcoxon Rank Sum test; * denotes p-value < 0.01, ** denotes p-value < 0.001, *** denotes p-value < $1x10^{-5}$. Only a single miRNA was predicted to contain a modification using GMUCT data, so it is represented as a single line.

Since GMUCT maps the precise position of RNA cleavage events in detected transcripts, we then sought to determine whether the predicted modified positions within mRNAs were in close proximity to specific cleavage events. We tested this because such a finding would suggest that these modifications could be the signal for an RNA cleaving enzyme to initiate the degradation process. To test this idea, we examined the 50 nucleotides up- and downstream of HAMR-predicted modification sites (**Figure 2.15A**). This analysis revealed no specific peak or pattern in GMUCT cleavage signal in this 100-nucleotide window surrounding HAMR-predicted modification sites (**Figure 2.15A**). These results suggest modification-associated uncapping and RNA turnover does not require a specific cleavage event related to the site of covalent addition, but is either a consequence of the degradation process and/or induces the turnover of these transcripts by normal 5' to 3' and 3' to 5' exonucleolytic mechanisms. Intriguingly, seven transcripts containing HAMR-predicted modifications in the GMUCT datasets overlapped with the set of 33 transcripts recently found to undergo nonsense-mediated decay (NMD) in an alternative-splicing dependent manner (Kalyna et al., 2012), suggesting NMD might be one such turnover mechanism. In contrast, HAMR-predicted modification sites in the human (HEK293T and HeLa) cells showed a small peak in average GMUCT cleavage signal directly upstream (**Figure 2.15B**) of HAMR-predicted modification sites, suggesting that a mechanism of modification-induced cleavage may be active in humans. Thus, HAMR-predicted modifications may function differently in plants and humans. However, this hypothesis will require future testing.

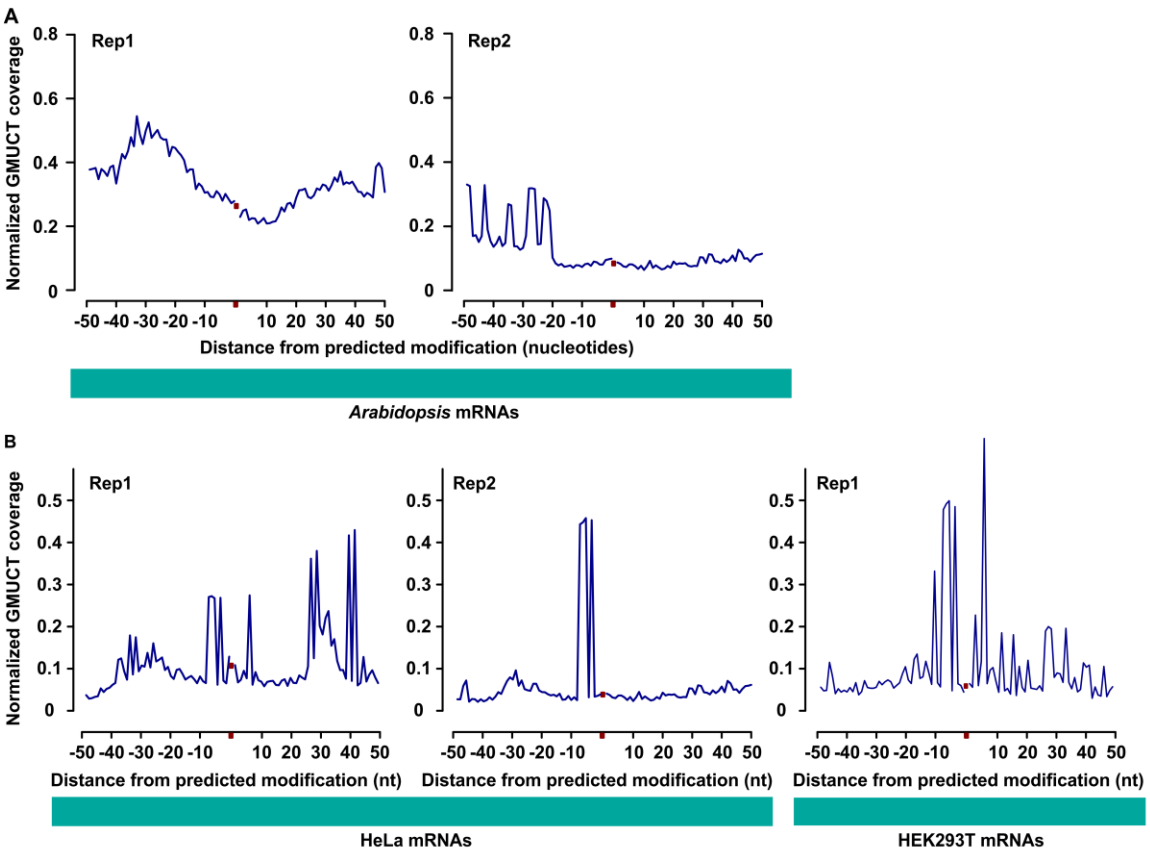

Figure 2.15: HAMR-predicted modifications do not coincide with precise cleavage peaks

Averaged GMUCT coverage profiles 50 nt up- and downstream of all predicted mRNA modification sites, normalized to RNA-seq read abundance, for A) *Arabidopsis* and B) human cell lines. Red dots indicate the position of the predicted modification, and are plotted within 50 nt up- and downstream flanking regions. Modifications within 50 nt of the mRNA 5' or 3' ends were given correspondingly shorter flanking regions.

2.2.6 Stress responsive mRNAs are enriched for RNA modifications that affect the Watson-Crick base pairing edge

Our finding that HAMR-predicted covalent modifications were enriched in degrading mRNAs as identified by GMUCT (**Figures 2.2, 2.3, and 2.14**) suggested the

intriguing possibility that this could be a mechanism for regulating the levels of mRNAs encoding proteins with common cellular functions. To test this hypothesis, we searched for overrepresented Gene Ontology (GO) terms among the collection of modified mRNAs identified using the GMUCT data. To reduce any bias in reporting GO terms for this collection of mRNAs, we identified all GO terms within three branches of the "biological process" and "molecular function" roots, as determined by a depth first search (Vandivier et al., 2013). From this analysis, we observed a significant (FDR < 0.05) enrichment for transcripts encoding ribosomal proteins for both *Arabidopsis* and human uncapped transcripts identified by GMUCT (**Figures 2.16 and 2.17**). Additionally, for *Arabidopsis* uncapped, degrading transcripts containing HAMR-predicted modifications, we also observed a significant (FDR < 0.05) enrichment of transcripts encoding proteins involved in photosynthesis, as well as a variety of biotic and abiotic stress response terms, including "defense response", "response to water", "response to cold", "response to heat", "response to radiation", and "response to oxidative stress" (**Figure 2.16A**). Relatedly, for human uncapped, degrading transcripts containing HAMR-predicted modifications identified by GMUCT, we found significant (FDR < 0.05) enrichment of transcripts encoding proteins involved in "cell death" and "cell cycle" (**Figure 2.17A**). Conversely, we did not observe any measurable enrichment for the transcripts with HAMR-predicted modifications in our smRNA-seq and RNA-seq datasets, which is likely a consequence of the low levels of these covalent additions identified by HAMR analysis of data from these approaches. In total, the overrepresentation of certain biological functions such as stress responses and cell cycle among uncapped transcripts with HAMR-predicted modifications but not in stable mRNAs or mRNA-derived smRNAs

suggests that addition of modifications that affect the Watson-Crick base pairing edge targets specific sets of transcripts for degradation to maintain their proper levels in the cell. This hypothesis will require further testing.

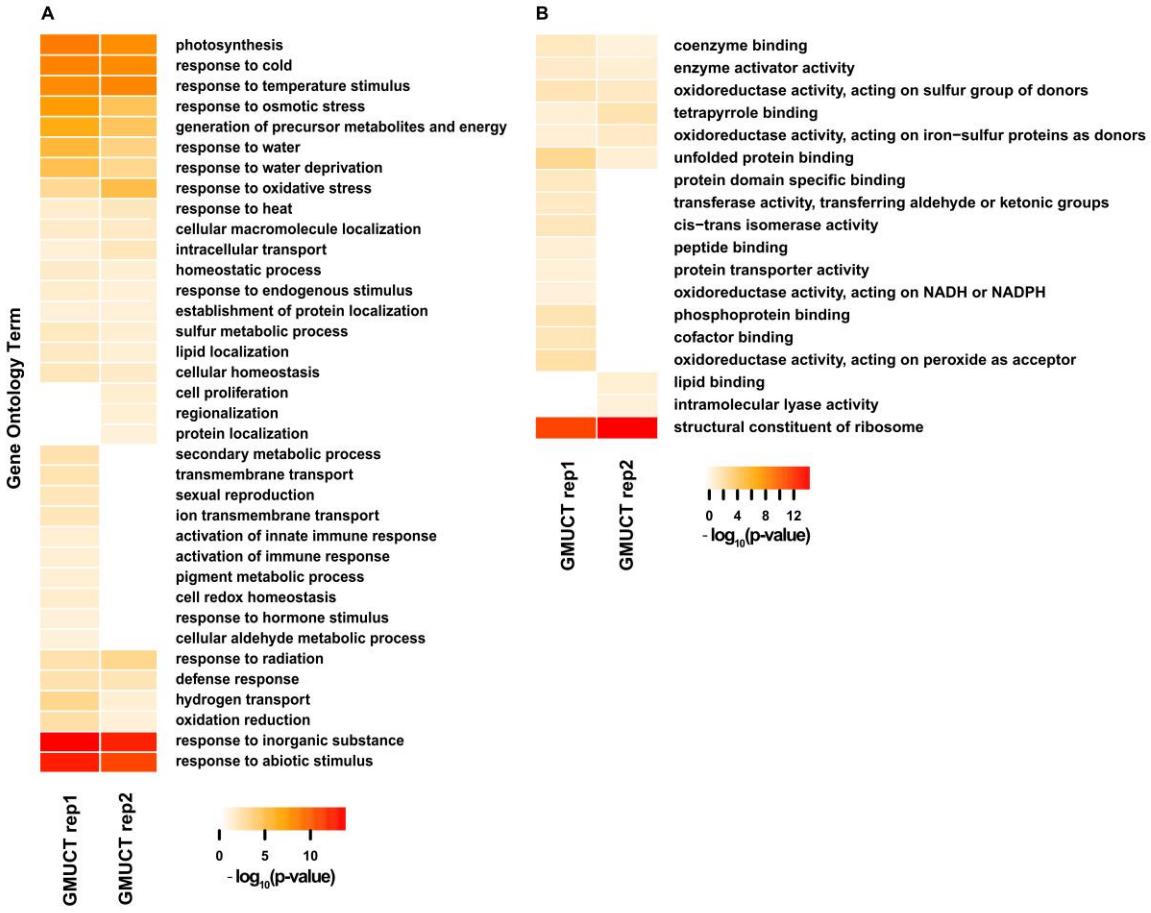

Figure 2.16: Arabidopsis transcripts with HAMR-predicted modifications encode proteins with coherent functions

(A) Biological process and (B) molecular function Gene Ontology (GO) terms are reported if they are significantly enriched (FDR < 0.05), over a background of all “HAMR accessible transcripts” with at least 100 uniquely mapping reads. Analyses were performed using the DAVID package (Huang et al., 2009). Furthermore, terms are only reported if they are separated from their ancestor term by no more than two parents, as determined by a depth first search as previously described (Vandivier et al., 2013). Lack of color denotes lack of significance.

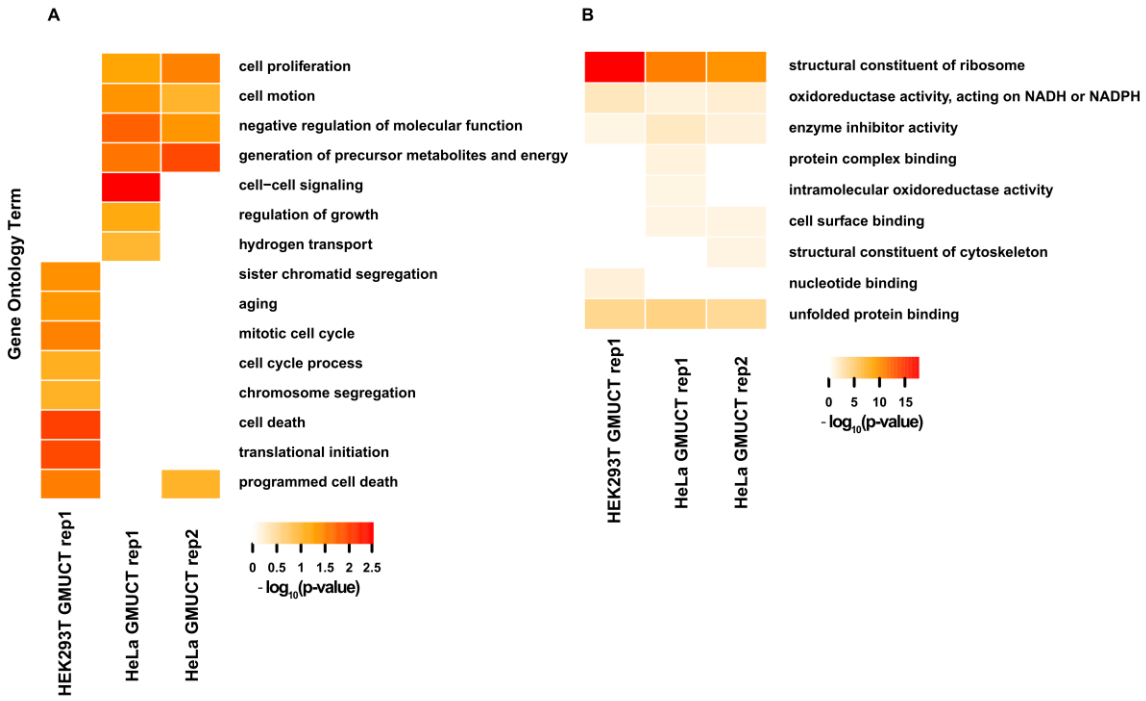

Figure 2.17: Human transcripts with HAMR-predicted modifications encode proteins with coherent functions

(A) Biological process and (B) molecular function Gene Ontology (GO) terms are reported if they are significantly enriched (FDR < 0.05), over a background of all “HAMR accessible transcripts” with at least 10 uniquely mapping reads. Analyses were performed using the DAVID package (Huang et al., 2009). Furthermore, terms are only reported if they are separated from their ancestor term by no more than two parents, as determined by a depth first search as previously described (Vandivier et al., 2013). Lack of color denotes lack of significance.

2.3 CONCLUSIONS

Here, we present evidence that covalent modifications of mRNA bases that affect the Watson-Crick base pairing edge are strongly enriched in uncapped, degrading mRNAs in both *Arabidopsis* and two human cell lines, and are usually found within exonic portions of these transcripts. In contrast, the identified modifications in stable

mRNAs tend to occur in alternatively spliced introns of protein-coding transcripts, and often accumulate in or near the splice donor and acceptor sites. Together, these results suggest a potential role for HAMR-predicted modifications in modulating specific alternative splicing events. Moreover, we found that specific HAMR-predicted modifications tend to occur in stable mRNAs (e.g. m^3C), whereas others tend to label uncapped, degrading transcripts (e.g. i^6A). These results suggest that certain classes of chemical modifications mark transcripts that are being degraded in eukaryotic transcriptomes. However, whether this is a cause or consequence of the RNA degradation process requires further investigation. Finally, we found that mRNA modifications mark transcripts that encode proteins with specific functions, many of which are involved in stress responses in both *Arabidopsis* and humans. These results suggest that modifications mark these classes of mRNA molecules for degradation to maintain them as mostly unstable during normal development, as was profiled in our experiments here. However, this hypothesis will require future testing during specific stress responses in both *Arabidopsis* and humans, which we describe in **Chapter 3**. In total, our study provides a resource for studying mRNA chemical modifications that affect the Watson-Crick base pairing edge, and identifies a potentially novel mechanism for initiating and/or maintaining mRNA degradation in eukaryotic transcriptomes.

CHAPTER 3: DIFFERENTIAL MESSENGER RNA MODIFICATION ALTERS TRANSCRIPT STABILITY UPON LONG TERM SALT STRESS

This section refers to work from:

Vandivier L.E., Anderson, Z.D., and Gregory BD (2017). Differential messenger RNA modification alters transcript stability upon long term salt stress. In preparation.

3.1 INTRODUCTION

Covalent chemical modifications are a widespread feature and physiologically relevant regulator of the messenger RNA lifecycle. In **Chapter 2**, we showed that these modifications mark uncapped, degrading mRNAs involved in stress response. Here, we apply the High Throughput Annotation of Modified Ribonucleotides (HAMR) pipeline to investigate the dynamics of modifications in response to long-term salt stress, which mimics the effects of irrigation-induced hypersalinity in agriculture.

Uncapped transcripts stem from decapping and/or endonucleolytic cleavage events and are readily degraded by both 5' and 3' exonucleases (Chekanova et al., 2007; Gazzani et al., 2004b), and thus represent actively degrading mRNAs. It follows that modifications are either a cause or a consequence of mRNA destabilization, or alternatively modifications may stabilize uncapped, degrading mRNAs. This is consistent with the known ability of m^6A to destabilize mRNAs (Du et al., 2016; Wang et al., 2014b), and led us to hypothesize that mRNA modifications identified in the uncapped,

degrading portion of the transcriptome stem from stability-altering modifications in capped, intact mRNAs.

To test this, we sought to exploit naturally occurring dynamic modifications that change in response to stress or stimuli. Notably, modified uncapped transcripts are enriched for stress-related functional annotations, such as “programmed cell death” in human transcripts and response to various abiotic stresses (e.g. water, salt, heat) in *Arabidopsis*. Thus, we chose one such stressor (salt stress) in *Arabidopsis* as a natural model system in which to test the effects of changing modifications on mRNA stability. Salt stress is also a notable agricultural problem stemming from an increasing reliance upon irrigation. Even when irrigation water is not saline, it can indirectly lead to a buildup of soil osmolytes through raising water tables and dissolving normally inaccessible soil minerals (Jorenush and Sepaskhah, 2003). Thus, we chose to implement a low-amplitude (100mM), long-duration (3 weeks) salt stress treatment that more closely mimics what would be observed in the field. Here, we define a portion of the salt-responsive epitranscriptome through applying the HAMR pipeline to the transcriptomes of salt-stressed and control unstressed *Arabidopsis*. We then demonstrate their functional relevance through measuring changes in RNA stability, ribosome occupancy, and co-translational RNA decay. In summary, we demonstrate that the epitranscriptome changes during long-term salt stress, leading to changes in RNA stability that could result from ribosome pausing and co-translational decay. We also provide some of the first evidence for functionally relevant internal modified guanosines in mRNA.

3.2 RESULTS AND DISCUSSION

3.2.1 Long-term salt stress has little effect on the total number of mRNA modifications

To survey the dynamics of RNA modifications in response to low-amplitude, long-duration salt stress, we treated *Arabidopsis* one-week-old seedlings with a single 50mM NaCl watering, followed by an additional three 100mM NaCl waterings, all in 0.25x Hoagland's solution (**Figure 3.1A**). Control plants were treated at identical intervals with 0.25x Hoagland's media. We then harvested RNA from pre-bolting rosettes (approximately 25 days old) and prepared libraries for total polyadenylated RNAs (RNA-seq), as well as two RNA populations that often capture degradation products: small RNAs (smRNA-seq), and uncapped, degrading polyadenylated RNAs via the GMUCT method (Gregory et al., 2008; Willmann et al., 2014) (**Figure 3.1A**). To verify the efficacy of salt treatment, we observed both physical and transcriptomic phenotypes. Salt-treated plants are both smaller and darker, consistent with downregulated growth and expression of stress-related anthocyanins (**Figure 3.1A**). Significantly upregulated genes (FDR < 0.05) are also significantly enriched (FDR < 0.05) for anthocyanin biosynthesis, as well as "response to salt stress", "response to abscisic acid" (ABA), and "response to osmotic stress" Gene Ontology terms (**Figure 3.1B**). In contrast, significantly downregulated genes are significantly enriched for response to biotic stresses and response to auxin, a key mediator of plant growth (**Figure 3.1B**). Thus, both physical and transcriptomic phenotypes are consistent with salt stress.

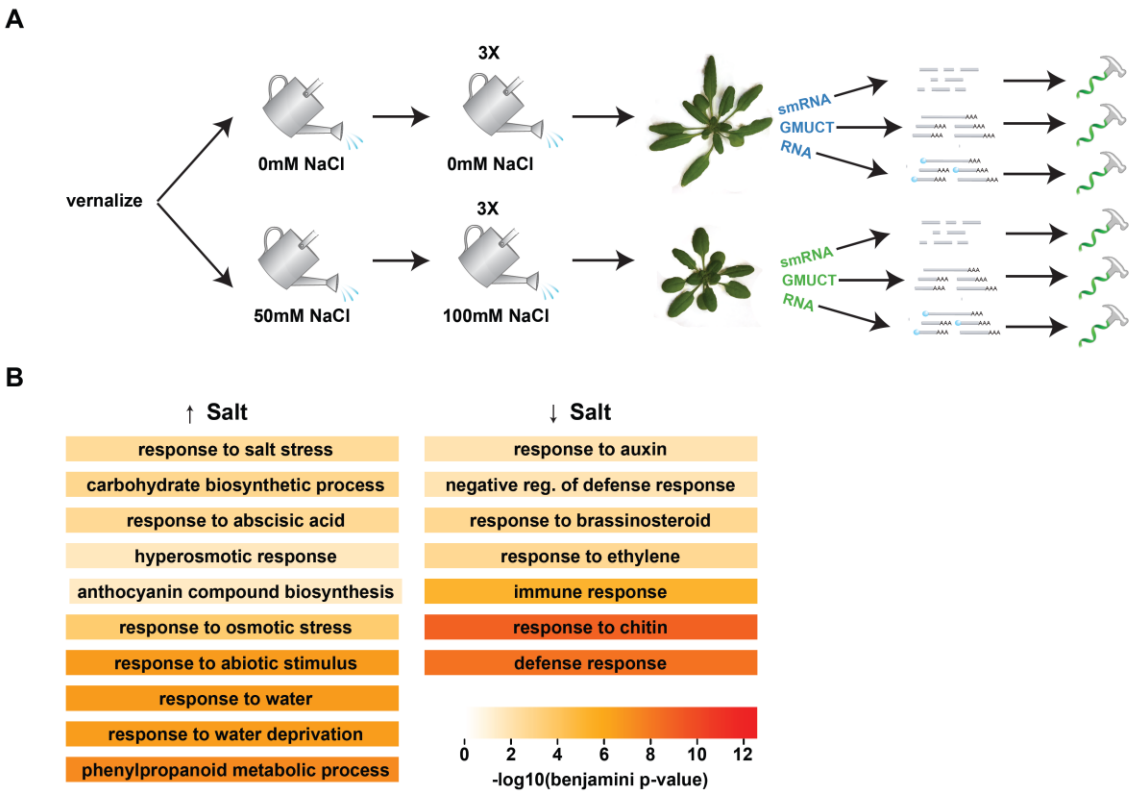

Figure 3.1: Experimental overview and validation of salt stress

(A) Experimental overview, in which plants are either subjected to long-term, low-amplitude salt stress or treated with control media, with two biological replicates in each treatment group. In salt stress, plants are first treated with 50mM NaCl in 0.25x Hoagland's solution after one week of growth, followed by 3 additional 100mM NaCl treatments. Representative individual plants are shown at the time of harvest. RNA extraction from salt-stress and control plants are then used to construct smRNA, capped polyadenylated RNA, and uncapped degrading RNA libraries (GMUCT). HAMR is applied across trimmed, uniquely mapping reads from all libraries. Blue spheres denote 7mG caps. (B) Significantly enriched gene ontology terms among transcripts that are significantly up- or down-regulated (black arrows) in salt stress. Heatmap colors denote Benjamini-corrected p-values.

We then performed our HAMR analysis pipeline on salt-treated and control transcriptomes. Since the distribution of mismatches showed bias toward the read termini, we masked any mismatches from read ends or from the random-hexamer region

used in construction of GMUCT libraries. Since the statistical power of HAMR is dependent upon sufficient read depth, we then normalized the total number of HAMR-predicted modifications to the number of “accessible bases” with at least 50x read coverage (Vandivier et al., 2015a). In control plants, HAMR-predicted modifications over two biological replicates are most abundant in uncapped RNAs, with a total of 799 and 830 modifications per million accessible bases (MPM) (**Figure 3.2A**). smRNAs possess 200 and 198 MPM, while modifications in total mRNAs are least abundant (21 and 14 MPM) (**Figure 3.2A**), consistent with previous observations (Vandivier et al., 2015a). Under salt stress, total numbers of MPM do not change appreciably, with 820 and 840 in uncapped RNAs, 168 and 165 in smRNAs, and 14 and 13 in total mRNAs (**Figure 3.2A**). Similarly, modification type and subtranscript localization are relatively constant across stress and control treatments and across replicates (**Figures 3.2B and 3.2C**). Nonetheless, there are differences between library types, as observed previously (Vandivier et al., 2015a). Modified cytosines are relatively rare in uncapped transcripts, but are more common in smRNAs and RNAs. Uncapped transcripts and RNAs are enriched for modified adenosines and uracils, while modified uracils are nearly absent from smRNAs (**Figure 3.2B**). Modifications in uncapped transcripts also tend to be in the coding sequence (CDS) and 3’ untranslated region (UTR), while those in polyA$^+$ mRNAs are enriched for both the 5’ and 3’ UTRs (**Figure 3.2C**). Both RNA and GMUCT are polyA-selected, so the additional 3’ bias in GMUCT is unlikely to be a simple artifact of polyA selection. Moreover, modified transcripts are significantly enriched (FDR < 0.05) for stress-response and photosynthesis annotations (**Figure 3.2D**), as observed previously.

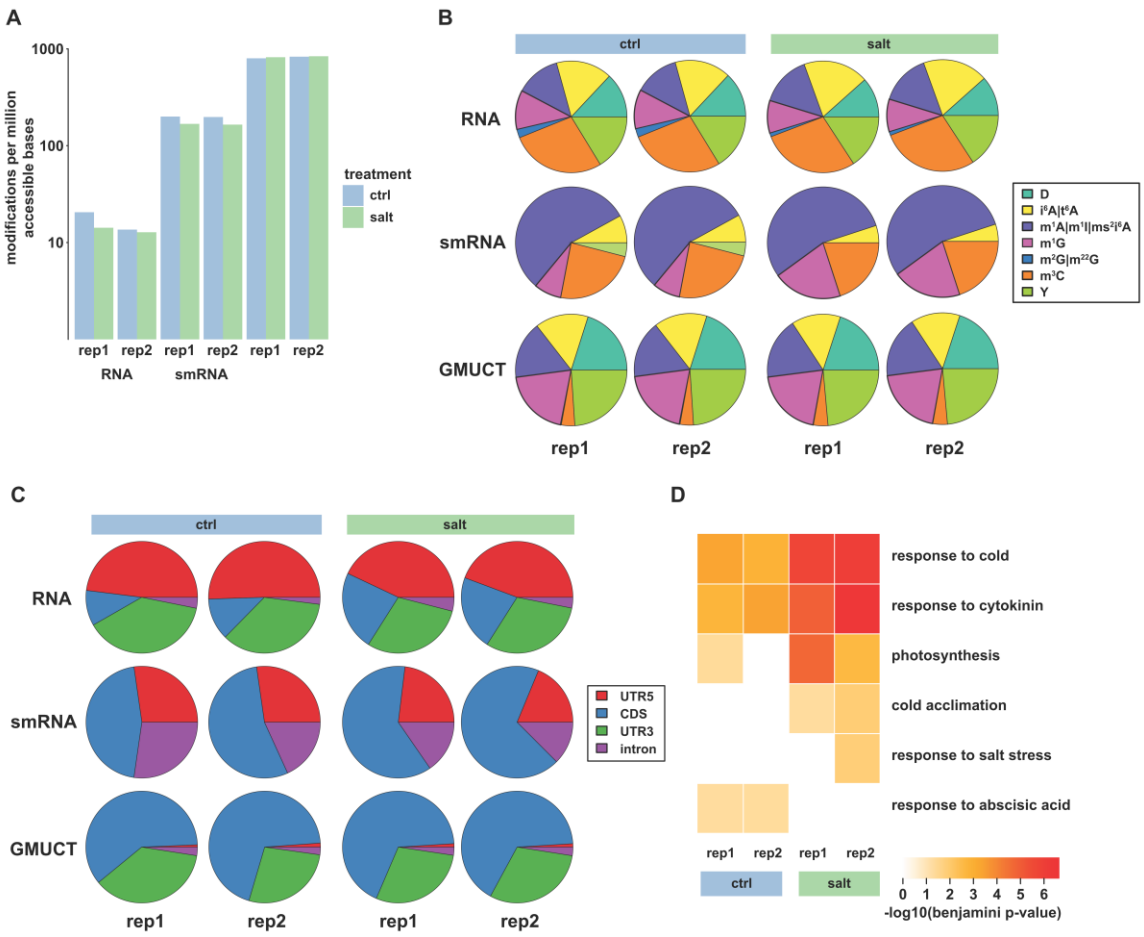

Figure 3.2: Long-term salt stress has little effect on the total number of mRNA modifications

(A) Total number of modifications normalized to millions of HAMR-accessible bases. (B) Identity of modified bases predicted with HAMR, including 3-methylcytosine (m^3C), N^1-methyladenosine (m^1A), 1-methylguanosine (m^1G), 2-methylguanosine (m^2G), 2,2-dimethyl guanosine ($m^{22}G$), pseudouridine (Y), dihydrouridine (D), N^6-isopentenyladenosine (i^6A), and threonylcarbamoyladenosine (t^6A). One replicate is shown. (C) Subtranscriptomic localization of modifications to 5' and 3' untranslated regions (UTRs), coding sequence (CDS), and introns. (D) Gene Ontology terms significantly (FDR < 0.05) enriched among uncapped, degrading transcripts with modifications.

Rarefaction analyses in which GMUCT libraries are randomly downsampled

indicate that the detection of modifications and number of HAMR accessible bases is far

from saturation (**Figures 3.3A and 3.3B**), although the total number of MPM remains stable as read coverage increases (**Figure 3.3C**), indicating that the differences in MPM across different library types are unlikely to be an artifact of different read coverages. Additionally, estimating the total number of modifications indicates that HAMR only captures ~5% of total modifications in uncapped bases, consistent with its low false-positive, high false-negative design and indicating that more modifications could be captured with more sequencing.

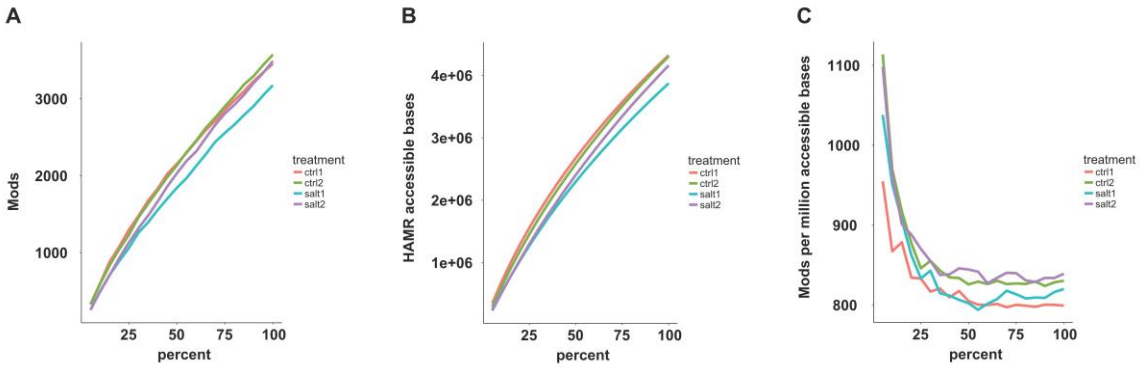

Figure 3.3: Rarefaction curves for HAMR analyses

Libraries are randomly sampled at 5% intervals. HAMR is run to analyze A) total modifications, B) HAMR-accessible bases with at least 50x coverage, and C) modifications per million accessible bases (MPM), as plotted again the percent of reads sampled.

3.2.2 Long-term salt stress leads to changes in the epitranscriptome

Despite the lack of changes in overall modification abundance, we were still able to define salt-responsive modifications that are either gained or lost upon salt treatment. To ensure that differences in modification status are not simply due to differential HAMR accessibility, with first constrained our analysis to bases that are HAMR accessible in

both treatments. Since the degree of overlap between replicates (**Figure 3.4A**) is only slightly higher than the degree of overlap between salt and control modifications (**Figure 3.4B**), we also required that salt-specific modifications be present in both salt replicates and absent in both control replicates, while control-specific modifications must be present in both control replicates and absent in both salt replicates. Using this approach, we defined 230 modifications in uncapped mRNAs that are lost upon salt stress, and 189 that are gained (**Figure 3.4C**). Transcripts with these differential modifications are enriched in stress response annotations ("response to metal ions" and "response to bacterium", FDR < 0.05). Fewer than 10 such differential modifications were observed in either total mRNAs or smRNAs, and thus we focused our analysis on uncapped mRNAs.

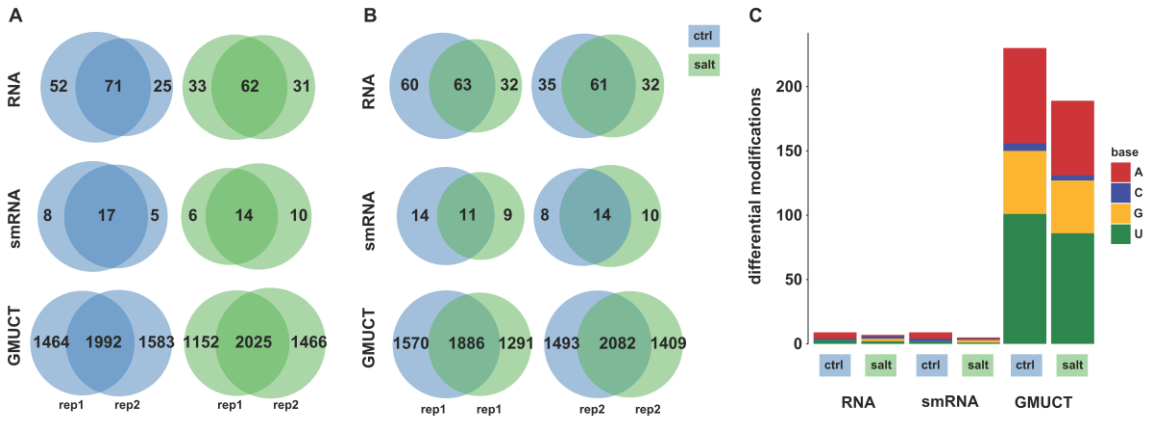

Figure 3.4: Long-term salt stress leads to changes in the epitranscriptome

The degree of overlap between modifications predicted across (A) biological replicates and (B) salt and control treatments. (C) Total numbers of differential modifications, defined as modifications present in both replicates of one treatment of absent in both replicates of the other. Color denotes the modified base.

To validate these differential modifications, we performed RNA immunoprecipitation using antibodies raised against 3-methylcytosine (m^3C), as

described previously (Ryvkin et al., 2013; Vandivier et al., 2015a). We only tested transcripts with “coherent” patterns in differential modification, either exclusively gaining or exclusively losing modifications upon salt stress. We first normalized to samples pulled down with a control anti-IgG antibody, and then renormalized to the average of four transcripts that are 1) equally abundant in salt and control-treated samples and 2) unlikely to be modified. To define which transcripts are unlikely to be modified, we tested out several potential correlates with number of predicted modifications, including number of mismatched reads as a proportion of total read coverage (**Figure 3.5A**) and raw numbers of mismatched reads (**Figure 3.5B**). Since raw numbers of mismatched reads correlate best with number of predicted modifications (**Figure 3.5**), we defined unmodified genes by 1) minimizing number of mismatches to less than 10, 2) maximizing total read coverage among these candidates, and 3) giving preference to transcripts with stable steady state abundance between salt and control treatments.

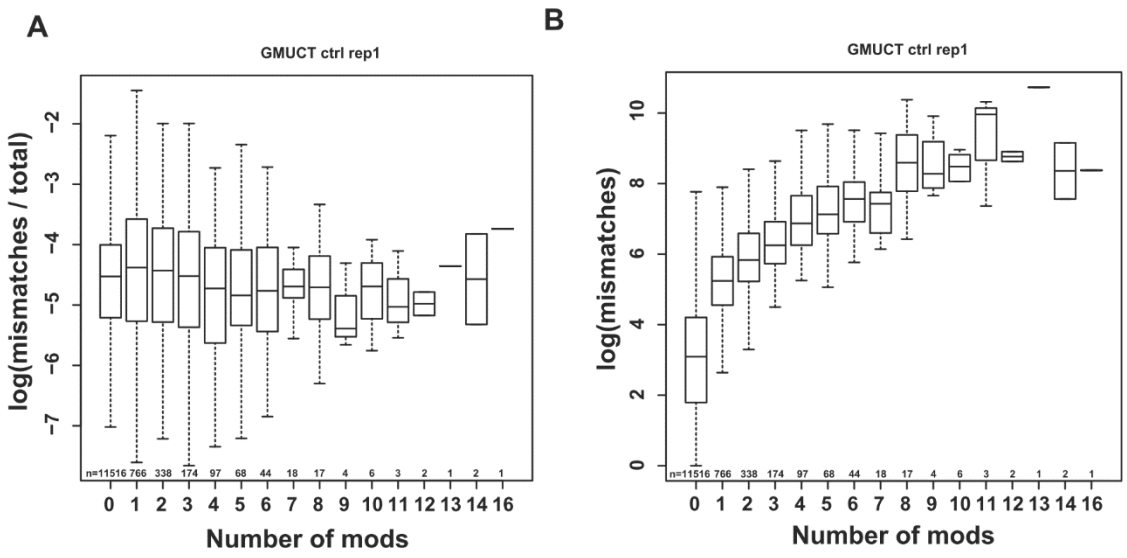

Figure 3.5: Statistics for determining unmodified transcripts

A) Mismatched reads as a proportion of total reads and B) mismatched reads are plotted again number of predicted modifications for each gene. Y-axes are logarithmic for ease of visualization.

Using these transcripts to renormalize, 4 out of 4 tested transcripts that lose modifications upon salt stress show both enrichment in anti-m^3C pulldowns over IgG pulldowns (m^3C/IgG) in control plants, and reduced m^3C/IgG enrichment upon salt stress (3 significantly, $p < 0.05$, Student's t-test). 1 out of 2 tested transcripts that gain modifications show significant m^3C/IgG enrichment in salt-treated plants, and significantly increased m^3C/IgG enrichment (**Figure 3.6**). Thus, the majority of tested differential modifications validate with an independent method of measurement, demonstrating the predictive power of HAMR to call modifications that change over stress conditions.

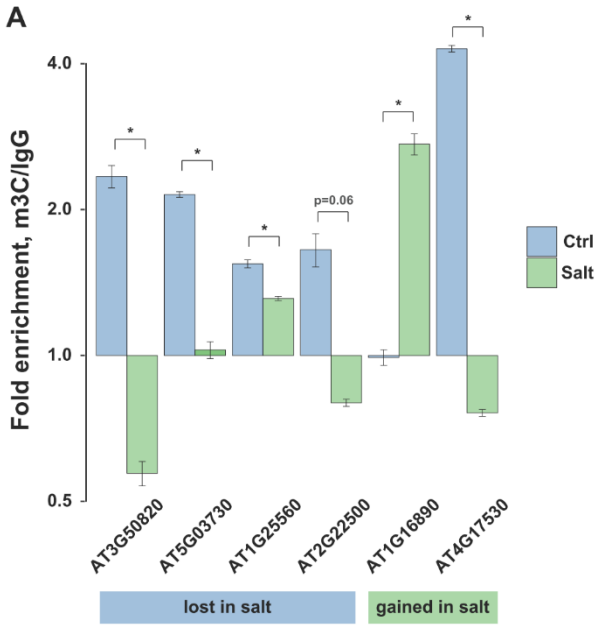

Figure 3.6: Validation of differential modifications through m^3C immunoprecipitation

Each bar denotes enrichment in anti-m^3C pulldowns over IgG control antibody pulldowns (m^3C/IgG), and is further normalized to an array of unmodified genes. Error bars are +/- standard error of the mean, and * denotes $p < 0.05$ as calculated with a Student's t-test.

Intriguingly, most modified bases, including those that are responsive to salt stress, share a common UGGAA motif directly downstream of the site of modification (**Figure 3.7**). This motif resembles known consensus binding motifs for RNA Recognition Motif (RRM)-containing RBPs across multiple species, such as Dmel\mod in *Drosophila* and Pp_0237 in the moss *Physcomitrella Patens* (Bailey et al., 2009; Ray et al., 2013). Thus, it is possible that modifications are either deposited or stabilized by RRM-containing RBPs. Unexpectedly, this motif is common to all modified bases, which could stem from the tendency of modified bases to form clusters. In addition, several modified adenosines and uracils are also preceded by a string of cytosines (**Figure 3.7A**). In

summary, there appears to be a degree of sequence-specificity in the location of modifications within uncapped mRNAs, though the exact mechanism by which these modifications are deposited has yet to be elucidated.

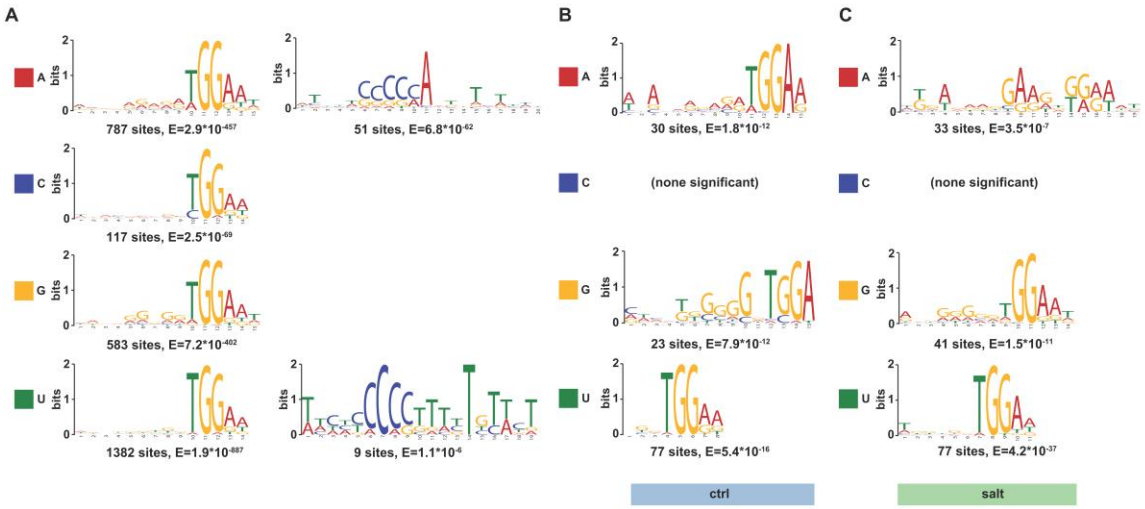

Figure 3.7: Sequence context of modified bases

Motifs identified with the MEME Suite (Bailey et al., 2009) for 20bp windows centered around each predicted modification, for A) all modifications, B) modifications lost upon salt stress, and C) modifications gained upon salt stress.

3.2.3 Differential modification alters transcript stability

We then looked to use salt-responsive modifications to probe the functional consequences of mRNA modifications on transcript stability. We first tested whether differential modifications correlate with changes in a transcript's proportion decapping, which is defined as the ratio of GMUCT to RNA-seq read coverage and which we have previously proposed to be a proxy for transcript instability (Li et al., 2012a; Vandivier et al., 2015a). Overall, loss of modifications upon salt stress leads to significantly lower

proportion decapping ($p = 9.1 \times 10^{-13}$, Wilcoxon test), while gain of modifications leads to significant increases ($p = 0.007$, Wilcoxon test), suggesting that most modifications destabilize mRNAs (**Figures 3.8A and 3.8B**). On average, these effects do not depend on the identity of the modified base (**Figure 3.8A**), although modifications in the 3' UTR appear to have a greater effect on stability than those in the CDS (**Figure 3.8B**). This could relate to the known enrichment of RNA stability elements within the 3'UTR across multiple species. These include AU-rich elements (Chen and Shyu, 1995; Narsai et al., 2007), mRNA secondary structural elements (Goodarzi et al., 2012), miRNA target sites (Carthew and Sontheimer, 2009), and various sequence elements enriched in highly stable or unstable *Arabidopsis* transcripts (Narsai et al., 2007). Notably, modified bases are approximately 2-fold more likely than unmodified bases to coincide with miRNA target sites predicted with psRNATarget (Dai and Zhao, 2011) (**Figure 3.8C**). While overlaps between miRNA target sites and differential modifications are rare, one such modified guanosine in transcripts from the AT5G20450 gene is lost upon salt stress and coincides with a decrease in proportion decapping.

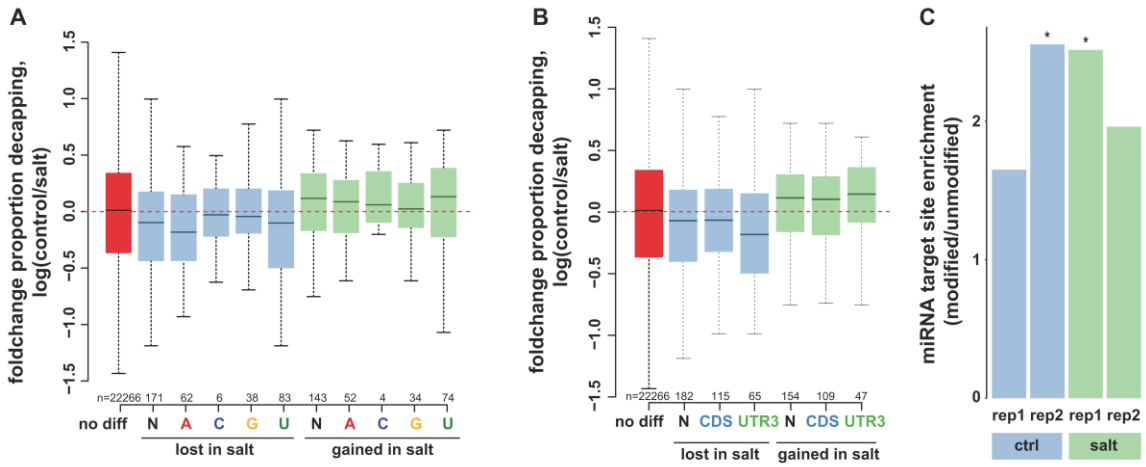

Figure 3.8: Differential modification associates with altered proportion decapping

A) The foldchange in GMUCT read coverage normalized to RNA read coverage (proportion decapping), from salt to control, is plotted for mRNAs with or without differential modifications in their population of uncapped degrading transcripts. Differential modifications are split into those that are gained or lost upon salt stress. Transcripts with differential modifications are further stratified by the type of differentially modified base. (B) Transcripts are again stratified by the location of each differentially modified base. Transcripts with such modifications in the 5` UTR are rare and are not shown. (C) Overlap between differential modifications and miRNA target sites. * denotes $p < 0.05$ as calculated with a Fisher's exact test.

To more directly assay for changes in mRNA stability, we then tracked the decay of differentially modified transcripts in protoplasts treated with the transcriptional inhibitors actinomycin and cordycepin (**Figure 3.9**). Of the 17 transcripts tested, 11 show significant changes in RNA half-life ($p < 0.05$, Student's t-test) upon salt stress (**Figure 3.10**), including two transcripts from genes (AT3G48360 and AT2G03440) with annotated salt-stress functions. Of the negative control transcripts used for normalizing immunoprecipitation experiments (**Figure 3.6**), all show a slight decrease in half-life upon salt stress (2 of 4 significant, $p < 0.05$, Student's t-test) (**Figure 3.10**), consistent with the small global increase in proportion decapping among transcripts with no

differential modifications (**Figure 3.8A**). Stratifying differential modifications based upon their modified base (A, C, G, or U) reveals base-dependent trends. For instance, modified adenosines appear to be stabilizing (4 of 4 transcripts tested). These modified adenosines are distinct from m^6A, which is methylated outside the Watson-Crick base pairing edge and cannot be detected by HAMR. Thus, these observations do not necessarily contradict the known tendency of m^6A to trigger transcript destabilization (Du et al., 2016; Wang et al., 2014b) Modified cytosines likewise appear stabilizing (3 of 4 transcripts tested), consistent with the known stabilizing effects of m^5C in noncoding RNAs (Hussain et al., 2013d; Schaefer et al., 2010). In contrast, modified guanosines appear destabilizing (3 of 5 transcripts tested), as are modified uracils (3 of 4 transcripts tested) (**Figure 3.10**).


A
rep1
rep2
Control
Salt
0hr
4hr
19hr
AT3G48360
AT1G80920
AT5G20700
AT1G68790
lost in salt
gained in salt
B
C
AT3G50820
AT5G03730
AT1G25560
AT4G17530
C
G
AT4G27450
AT3G61140
AT3G22790
AT5G66400
AT4G26400
D
U
ATCG00190
AT5G02380
AT3G21055
AT2G03440
E
Ia Diff
AT4G25150
AT4G28300
AT3G08550
AT3G63510

Figure 3.9: Decay curves after treatment with actinomycin and cordycepin

Decay curves, shown as the proportion of a transcript remaining (y-axis) as a function of time following treatment with transcriptional inhibitors (x-axis), for transcripts with differentially modified A) adenosines, B) cytosines, C) guanosines, or D) uracils. E) Decay curves for unmodified transcripts (no diff).

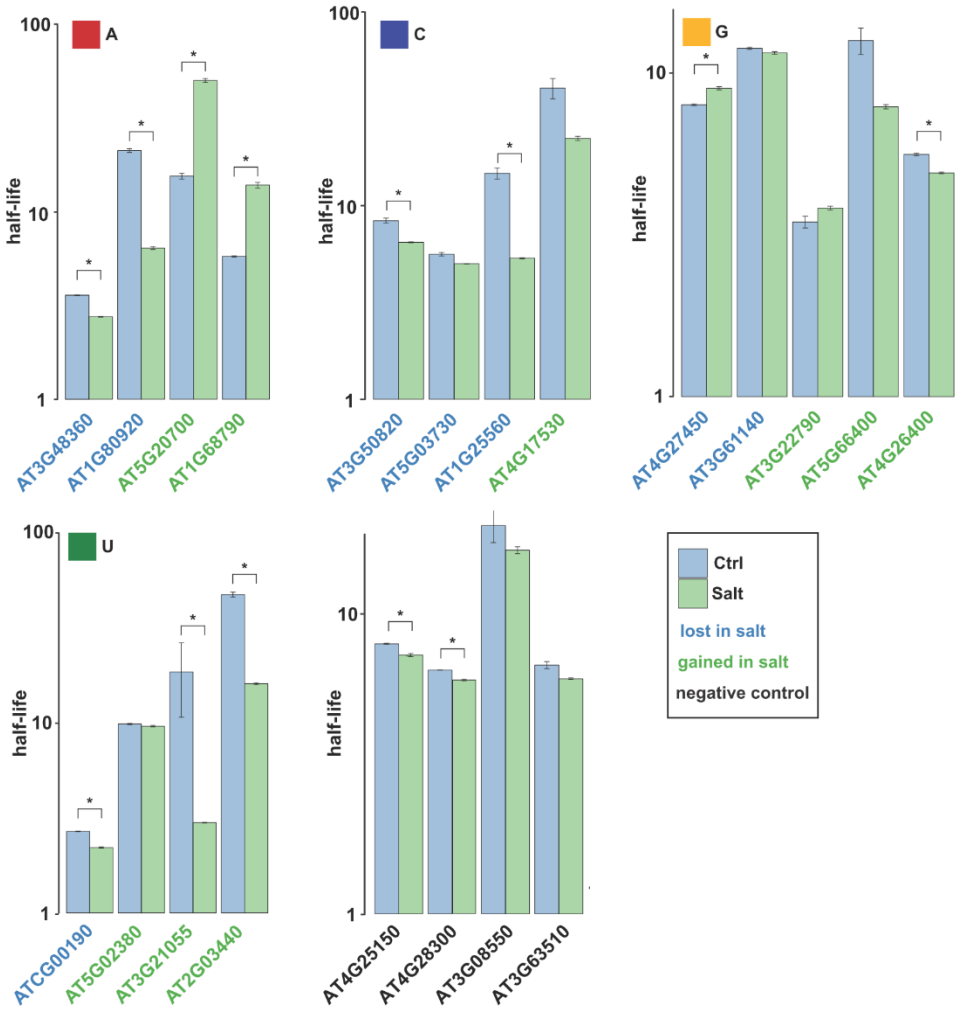

Figure 3.10: Differential modification alters transcript stability

Comparison of half-lives calculated from decay curves for a representative biological replicate. Error bars are +/- standard error of the mean, and * denotes $p < 0.05$ as calculated with a Student's t-test.

Notably, for 15 of 17 transcripts the change in half-life is consistent with the change in proportion decapping, that is to say an increase in half-life corresponds to a decrease in proportion decapping, and vice-versa. This suggests that proportion decapping is in fact a valid proxy for transcript instability. Consistently, modified adenosines and uracils show both the greatest change in decapping (**Figure 3.8A**) and the largest magnitude changes in half-life. While the effect of differential modified guanosines is small in magnitude, to our knowledge this is the first report suggesting a functional significance of modified internal (non-cap) guanosines in mRNAs.

3.2.4 Differential modification associates with altered ribosome dynamics

Finally, we sought to investigate potential mechanisms through which differential modifications could lead to differential stability. We hypothesized a role for changes in translation since 1) translation and RNA decay are known to be linked (Bazzini et al., 2012; Pelechano et al., 2015; Roy and Jacobson, 2013), 2) modified bases in uncapped RNAs across all treatments and replicates are over 15 times as likely to coincide with a ribosomal pause site than are unmodified bases ($p < 10^{-48}$, Fisher Exact test) (**Figure 3.11A**), and 3) multiple known mRNA modifications such as m^6A, m^1A, and pseudouridine are known to modulate rates of translation (Choi et al., 2016; Wang et al., 2015), correlate with the start and stop positions of open reading frames (ORFs) (Dominissini et al., 2012; Li et al., 2016), and change a base's coding potential (Fernández et al., 2013; Karijolich and Yu, 2011). Thus, we measured both ribosome

occupancy (ribo-seq normalized to RNA-seq) and co-translational decay at differentially modified transcripts.

Overall, ribosome occupancy anticorrelates with proportion decapping at tested differentially modified transcripts, as transcripts that lose modifications upon salt stress tend to show an increase in ribosome occupancy, while transcripts that gain modifications tend to show a decrease (**Figures 3.11B and 3.11C**). These trends are not changed when stratifying by modified base identity (**Figure 3.11B**) or by location to the CDS versus 3' UTR (**Figure 3.11C**), though these trends are not apparent when considering all detectable transcripts (**Figure 3.11D**).

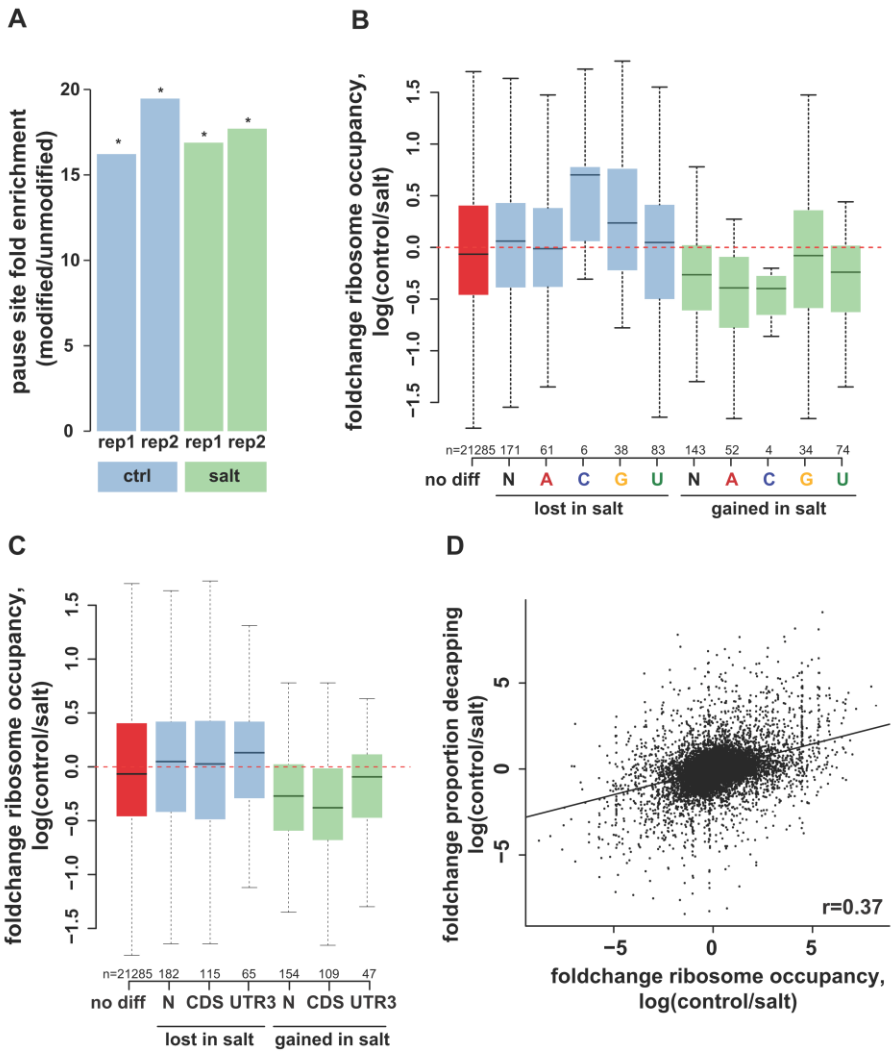

Figure 3.11: Differential modification alters ribosome dynamics

A) Fold enrichment of ribosome pause sites in modified bases, as compared to unmodified bases. Pause sites are determined as runs of nucleotides with ribosome footprint coverage at least 25-fold over the median coverage for each transcript. B) The foldchange in ribo-seq read coverage normalized to RNA read coverage (ribosome occupancy), from salt to control, is plotted for mRNAs with or without differential modifications in their population of uncapped degrading transcripts. Differential modifications are split into those that are gained or lost upon salt stress. Transcripts with differential modifications are further stratified by the type of differentially modified base. (C) Transcripts are again stratified by the location of each differentially modified base. Transcripts with differential modifications in the 5' UTR are rare and are not shown. (D) Foldchange in ribosome occupancy (x-axis) is plotted against foldchange proportion decapping (y-axis) for all detectable genes.

We then used our GMUCT data to probe for co-translational decay, as first described by Steinmetz and colleagues (Pelechano et al., 2015). Specifically, we focused on the buildup of cleavage sites at approximately 17 nucleotides upstream of the stop codon, which corresponds to the boundary of a stop codon-stalled ribosome (Pelechano et al., 2015). In transcripts that lose modifications upon salt stress, this cleavage site peak is smaller in salt-treated than in control plants, indicating less co-translational decay in salt stress (**Figure 3.12**). A subtler, though opposite trend is apparent in transcripts that gain modifications in salt stress, suggesting increased co-translational decay in salt stress (**Figure 3.12**). Notably, the magnitude of co-translational decay is greater in transcripts with differential modifications in the 3'UTR, consistent with greater changes in proportion decapping (**Figure 3.8B**). Thus, one mechanism by which modifications might affect transcript stability is by triggering ribosome pausing and co-translational decay, consistent with the known effects of m^6A and m^1A.

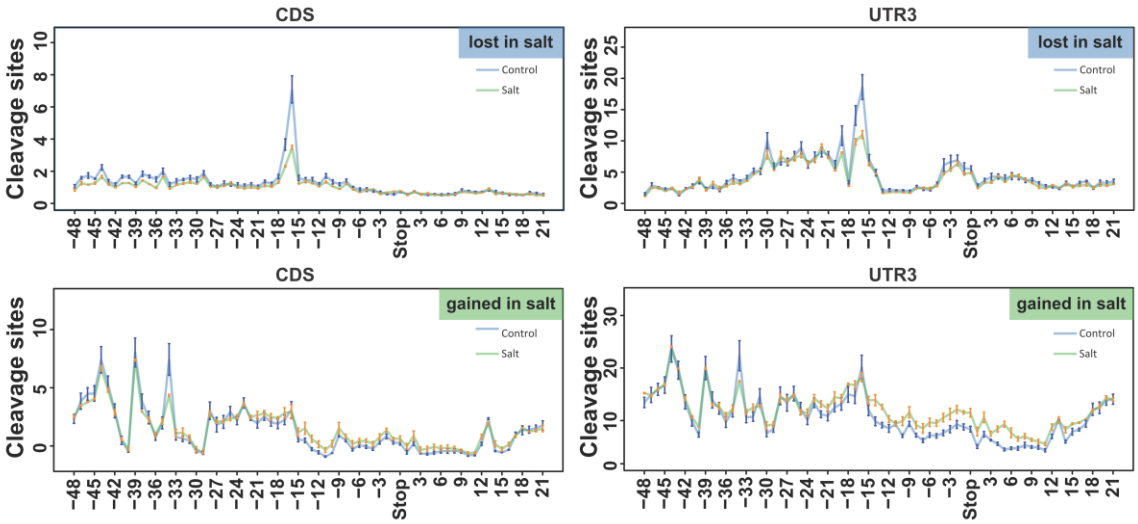

Figure 3.12: Differential modification alters co-translational decay

Co-translational decay, as determined by cleavage site accumulation ~17nt upstream of the stop codon. Cleavage sites are defined as the 5' termini of GMUCT reads. Plots are shown as averaged metaprofiles for all differentially modified transcripts that gain (green lines, orange error bars) or lose (blue lines, dark blue error bars) modifications upon salt stress. Plots are further stratified by location of differential modifications to the 3' UTR or CDS.

Future studies will address this hypothesis through assaying the epitranscriptome in mutants deficient in co-translational decay, such as the components of the no-go decay pathway, which triggers degradation of transcripts containing paused ribosomes (Doma and Parker, 2006). Just as modifications can cause processive enzymes like reverse transcriptase to stall, they could likewise act as direct steric inhibitors of ribosome procession. m^6A, for instance is known to disrupt elongation by interfering with tRNA selection (Choi et al., 2016). A buildup of modifications in capped, polyadenylated RNAs within decay pathway mutants would support this hypothesis.

Intriguingly, total numbers of HAMR predicted modifications also vary in mutants for key mediators of mRNA stability such as the nuclear mRNA cap-binding complex ABH1 (Hugouvieux et al., 2001) and Rail1, which promotes activity of the *Arabidopsis* XRN2 5’ to 3’ exonuclease. Loss of ABH1 (Yu et al., 2016), which should globally destabilize mRNAs, associates with an increase in modifications among uncapped, degrading mRNAs (**Figure 3.13**). Conversely, loss of Rail1, which should globally stabilize mRNAs, associates with decreases in both smRNAs and uncapped, degrading mRNAs (**Figure 3.13**). This raises the possibility that in addition to their known ability to destabilize mRNAs, some modifications could in fact be deposited after mRNA decapping, and are thus downstream of the process of mRNA decay.

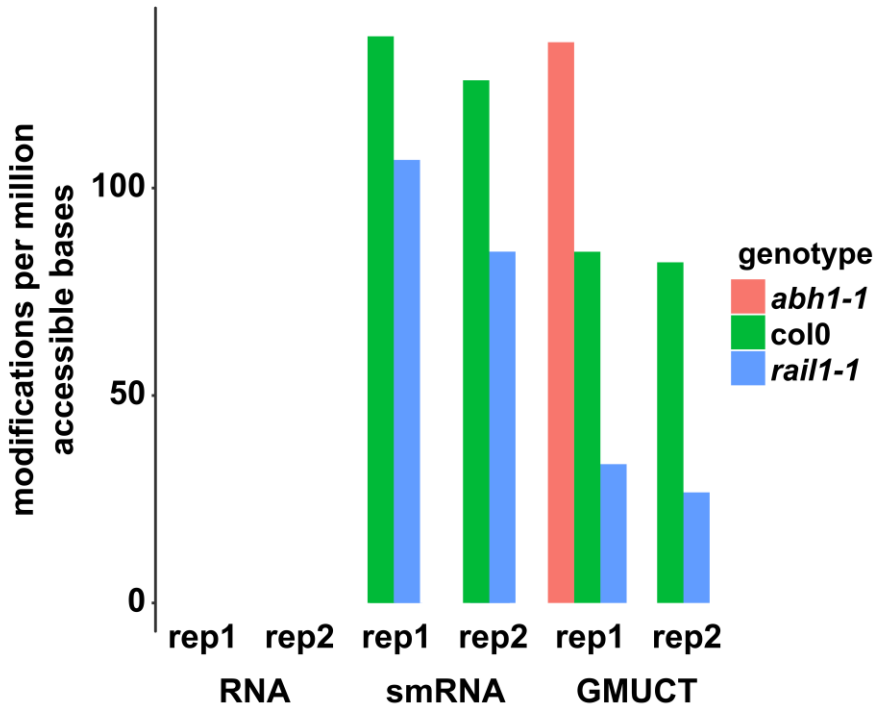

Figure 3.13: Disrupting cap stability or exonuclease activity changes modification abundance in uncapped mRNAs

Modifications per million accessible bases predicted from RNA-seq, smRNA-seq, and GMUCT libraries for two replicates of WT (col0) and a single replicate of *abh1-1* mutants. Modifications per million accessible bases from RNA-seq and GMUCT are also plotted for two replicates of *rail1-1*. No bar indicates no modifications, except where data is missing (no replicate 2 of *abh1-1*, and no smRNA-seq for *rail1-1*).

3.3 CONCLUSIONS

Here, we characterize the response of the *Arabidopsis* epitranscriptome to long-term salt stress, and uncover numerous constitutive and differential modifications in uncapped, degrading RNAs. Gain of modifications in the uncapped RNA population tends to correlate with a decrease in RNA stability, decrease in ribosome occupancy,

and increase in co-translational decay. More nuanced trends emerge when stratifying by base or location. For instance, modifications in the 3'UTR tend to have a greater effect on stability, consistent with the known enrichment in this region of sequence and structure elements that regulate stability (Carthew and Sontheimer, 2009; Chen and Shyu, 1995; Goodarzi et al., 2012). Additionally, many modifications appear to stabilize transcripts, in particular those that mark adenosines and cytosines. Overall, we present evidence that salt-stress-responsive modifications of all four RNA bases present in uncapped, degrading mRNAs correlate with the stability of capped, polyadenylated transcripts, and may do so based upon altering ribosome dynamics.

CHAPTER 4: A LINK BETWEEN MRNA SECONDARY STRUCTURE AND DICER-LIKE-MEDIATED DECAY IN *ARABIDOPSIS*

This section refers to work from:

Vandivier L.E.*, Anderson, S.J.*, Foley S.W.*, and Gregory B.D. (2017). The Conservation and Function of RNA Secondary Structure in Plants. *Annual Reviews Plant Biology*. **67:**463-88. PMID: 26865341

Foley, S.W.*, Vandivier, L.E.*, Kuksa, P., Gregory, B.D. (2015). Transcriptome-wide measurement of plant RNA secondary structure. *Current Opinion in Plant Biology*. **27**:36-43. PMID: 26119389

Vandivier L.E., Li F., and Gregory B.D. (2015). High-Throughput Nuclease-Mediated Probing of RNA Secondary Structure in Plant Transcriptomes. **1284**:41-70. PMID: 25757767

Vandivier L**.**, Li F, Zheng Q, Willmann M, Chen Y, Gregory B. (2013). Arabidopsis mRNA secondary structure correlates with protein function and domains. *Plant Signaling and Behavior*. **8**:e24301. PMID: 23603972

Li F., Zheng Q., Vandivier L.E., Willmann M.R., Chen Y., Gregory B.D. (2012) Regulatory impact of RNA secondary structure across the Arabidopsis transcriptome. *Plant Cell*. **24**:4346-59. PMID: 23150631

*Indicates co-first author

4.1 INTRODUCTION

Both coding and noncoding RNAs fold into intricate secondary structures via intramolecular base-pairing. These secondary structures, often in conjunction with RNA-binding proteins (RBPs), form the basis for higher-order tertiary structures that can direct catalysis, form scaffolds, and regulate RNA posttranscriptionally (Cruz and Westhof, 2009). In turn, RNA secondary structure regulates multiple steps of the RNA lifecycle, including transcription (Wanrooij et al., 2010), addition of the 5' cap (Dong et al., 2007), splicing (Buratti and Baralle, 2004; Jin et al., 2011; Liu et al., 1995; Raker et al., 2009; Warf and Berglund, 2010) polyadenylation (Klasens et al., 1998; Oikawa et al., 2010), nuclear export (Grüter et al., 1998), subcellular localization (Bullock et al., 2010; Subramanian et al., 2011), translation (Kozak, 1988; Svitkin et al., 2001; Wen et al., 2008), and turnover (Goodarzi et al., 2012). Additionally, specific classes of RNAs, such as microRNAs (miRNAs) and transfer RNAs (tRNAs) require secondary structure for correct processing and subsequent functionality (Bhaskaran et al., 2012; Carthew and Sontheimer, 2009; Francklyn and Minajigi, 2010). Structure likewise enables many long noncoding RNAs (lncRNAs) (Rinn and Chang, 2012; Tsai et al., 2010), ribosomal RNAs (rRNAs) (Korostelev and Noller, 2007), and tRNAs to function as structural scaffolds. Thus, determining the patterns of RNA folding across the transcriptome is crucial to fully understanding RNA function and regulation.

In previous work, we observed a link between mRNA secondary structure and production of mRNA-derived smRNAs. Transcripts with higher levels of paired bases demonstrate higher levels of smRNAs and lower overall transcript abundance (Li et al., 2012a), suggesting that secondary structure could lead to targeting transcripts for

smRNA production, resulting in endonucleolytic cleavage and consequent decay (see **Section 1.2**). Notably, all known targets of the Dicer endonucleases are in a double-stranded conformation, perhaps due to the evolutionary origins of RNAi in targeting viral duplex RNAs (Grimson et al., 2008). For instance, miRNAs adopt intramolecular fold back structures (Rajagopalan et al., 2006; Reinhart et al., 2002; Ruby et al., 2007), natural antisense siRNAs are processed from pairs of overlapping transcripts (Borsani et al., 2005; Katiyar-Agarwal et al., 2006), and endogenous siRNAs and trans-acting siRNAs (Gasciolli et al., 2005; Howell et al., 2007; Xie et al., 2005; Yoshikawa et al., 2005) rely upon an RNA-dependent RNA polymerase (RDR) to form double-stranded RNA. Additionally, noncanonical Dicer substrates like tRNAs, snoRNAs, rRNAs, and hairpin RNAs also possess a high degree of secondary structure. It has also been observed that certain mRNAs that are cleaved into smRNAs possess structural elements resembling known miRNA precursors (Burroughs et al., 2011). Thus, we hypothesized one mechanism by which highly structured mRNAs are degraded is by direct targeting by Dicer.

The canonical mechanisms by which Dicers function involves targeting of either intramolecular RNA secondary structure or intermolecular RNA duplexes to generate small RNAs that are then loaded onto Argonaute (AGO) proteins (Bartel, 2004; Meister and Tuschl, 2004; Jones-Rhoades et al., 2006; Carthew and Sontheimer, 2009). These AGO-bound smRNAs then direct either translational repression or cleavage of their target transcripts. Importantly, the mechanism we hypothesize is distinct from canonical miRNA-mediated cleavage, since we propose that mRNAs are smRNA precursors rather

than targets, and the resulting smRNAs do not necessarily need to be loaded onto AGO or to be functional *in trans*.

To test this hypothesis, we developed a structure probing technique that enriches for intramolecular RNA secondary structure, thus helping to disentangle it from intermolecular RNA duplexes. We then measured the levels of smRNAs and mRNAs from a panel of *Arabidopsis* DICER-LIKE (DCL) mutants, in order to define regions of DCL-dependent production of mRNA-derived smRNAs. We also paneled similar libraries for mutants in the best-characterized RDRs (RDR1, RDR2, and RDR6), which are the primary enzymes responsible for production of duplex RNA. Here, we observe a high degree of secondary structure in regions that produce smRNAs in a DICER-LIKE1 (DCL1) dependent but RDR-independent manner, consistent with the known role of DCL1 in targeting imperfectly paired miRNA precursor secondary structures. We then show a link between these regions and transcript steady state abundance, and give preliminary evidence that this could be due to DCL1-dependent transcript destabilization.

4.2 RESULTS AND DISCUSSION

4.2.1 PolyA$^+$ selection reduces the duplex RNA signal in structure mapping

To map RNA secondary structure across the whole transcriptome, we used a nuclease-based approach that probes both single- and double-stranded RNAs (ssRNA and dsRNA) (**Figure 4.1**) (Li et al., 2012b, 2012a). Briefly, RNA was cut to completion with single-stranded-specific nucleases to generate ssRNA, and with double-stranded-specific nucleases to generate dsRNA (**Figure 4.1A**). From these two pools of RNA, we

constructed and sequenced libraries, and computed coverage across the genome (**Figure 4.1B**). From these data, we could then compute statistics describing the likelihood of base pairing. For instance, we computed a numerical 'structure score' defined as a generalized log ratio of dsRNA to ssRNA (**Figure 4.1C**). From these scores, we could then define bases with a high or low probability of being base-paired, and use these either/or statistics to constrain RNA folding algorithms (**Figure 4.1D**). Finally, we defined peaks of either high or low structure using ChIP-seq-like peakcalling software (**Figure 4.1E**).

We first sought to demonstrate the necessity of empirical structure mapping over *in silico* free energy minimization. Thus, we looked to see if computational structure prediction via RNAFold (Zuker and Stiegler, 1981) could recapitulate our previously observed correlations between secondary structure, RNA abundance, and smRNA production. To do so, we compared computationally predicted mRNA secondary structure with mRNA abundance, ribosome association, and smRNA processing from these transcripts. We found a weak positive correlation (Pearson correlation r = 0.109) between computationally predicted structure scores and overall mRNA abundance, which is both lower in magnitude than what we observe with empirical data (r = -0.45) and also contradicts previous qPCR-based validation in which five highly structured transcripts were significantly less abundant than seven lowly structured transcripts (Li et al., 2012a). We also observed that computationally predicted structure scores have a weaker and opposite-in-sign correlation with smRNA production (r = -0.29), when compared to empirical structure versus smRNA production (r = 0.62). We found that both measured correlations using computationally predicted structure are significantly weaker

and opposite-in-sign when compared to correlations with experimental data. In total, our results indicate that experimentally-based structure mapping data are necessary to uncover the regulatory functions of RNA folding in eukaryotic transcriptomes.

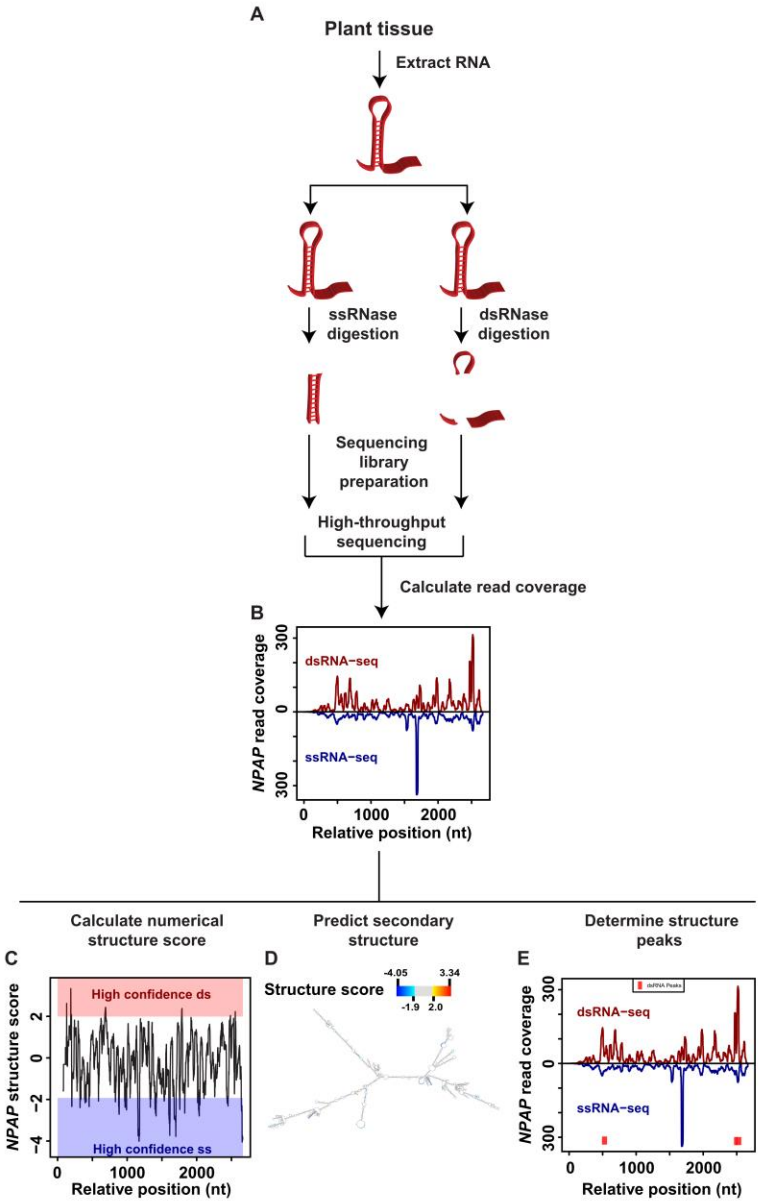

Figure 4.1: An overview of the nuclease-based structure probing used in this study

A) We began by extracting RNA from plant tissue, before splitting into an ssRNA and dsRNA treatment groups (dsRNase and ssRNase, respectively). We then constructed and sequenced libraries, and B) defined transcriptome-wide coverage. From coverage, we could C) compute a numerical structure score from the generalized log-ratio of coverages, D) constrain folding algorithms, and E) define peaks of high or low structure.

We then further refined this technique in order to disentangle bona-fide intramolecular secondary structure from intermolecular RNA duplexes, which will both register as double-stranded upon nuclease treatment. DCL processing of mRNAs that have been processed by RDRs is a well-established phenomenon distinct from the mechanism we propose. We defined RDR target transcripts as those that show significant loss of smRNAs upon genetic ablation of any RDR with well-characterized functions (RDR1, RDR2, or RDR6). We hypothesized that polyA$^+$-selection should be effective at reducing duplex RNA signal, so long as we assume that the lag time between RDR activity and duplex RNA cleavage is short. Since RDRs process along their template 3' to 5', any form of 3' selection should reduce the signal from RDR targets that have already been cleaved. Accordingly, we find that polyA$^+$-selection is effective at reducing the apparent structure scores of targets of RDR2, which is responsible for the bulk of mRNA-derived duplex RNA (**Figure 4.2**).

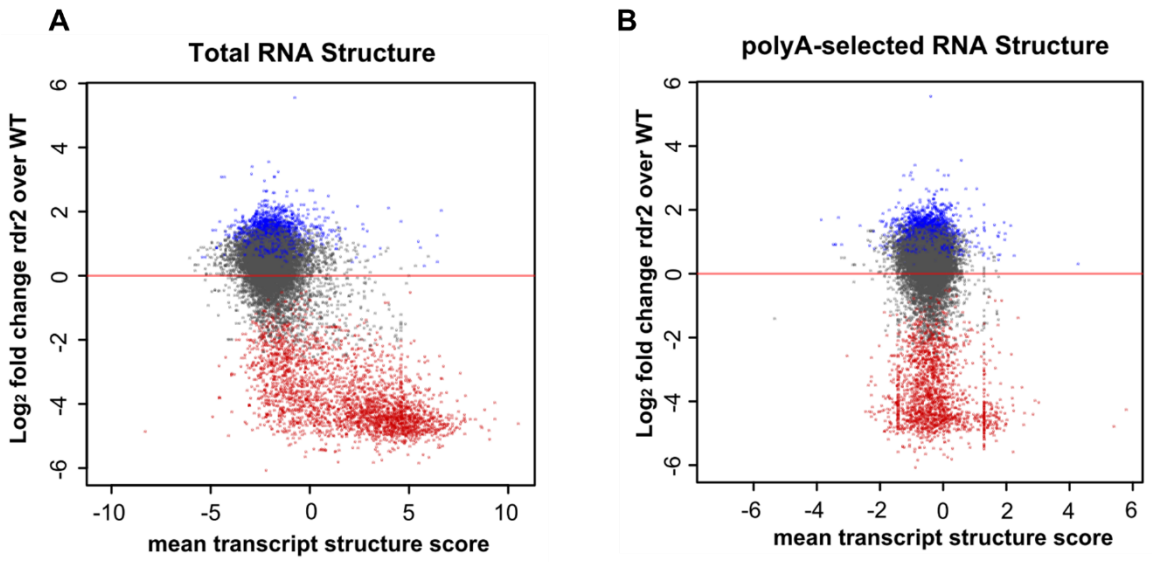

Figure 4.2: PolyA-selection reduces duplex RNA contamination

Mean transcript structure score (x-axis) is plotted again the log-foldchange of smRNA production in *rdr2-1* over WT plants (y-axis). Plots are shown for A) total RNA structure and B) structure of polyA-selected RNA. Red dots denote significant loss of smRNAs, and blue dots denote significant gain. Red dots thus signify RDR2 targets.

4.2.2 Fine-scale transcriptome binning enables identification of DCL-dependent, RDR-independent foci of smRNA production

We then sought to determine the functional outcome of mRNA secondary structure with respect to DCL-dependent smRNA production. Analogous to our methods to remove duplex RNA contamination, we sought to distinguish, at high-resolution, regions within mRNAs that produce smRNAs in a DCL-dependent but RDR-independent manner. We looked for DCL and RDR-dependent smRNA production, as defined by a significant decrease in smRNAs in each respective mutant over WT, at the resolution of either whole transcripts or 50-nucleotide bins (**Figure 4.3A**). Consistent with their known

partnership with RDRs in RNA-directed DNA methylation (RdDM) (Zilberman et al., 2004) and trans-acting siRNA (tasiRNA) production (Peragine et al., 2004; Vazquez et al., 2004; Xie et al., 2005), respectively, transcripts with DCL3- and DCL4-dependent smRNAs show the greatest degree of overlap with those that are dependent upon RDR1, RDR2, and/or RDR6 (**Figure 4.3A**). Very few transcripts demonstrate DCL-dependence and RDR-independence (**Figure 4.3A**), indicating that whole transcripts are the wrong scale at which to find these foci. At finer-scale bins, however, the separation between DCL and RDR dependence increases (**Figure 4.3A**), and is greatest for DCL1 and DCL2. Notably, smaller bins also better recapitulate the DCL1-dependence and RDR-independence that is expected for known miRNA precursors (**Figure 4.3B**). Thus, we performed all subsequent analyses with 50 nucleotide bins.

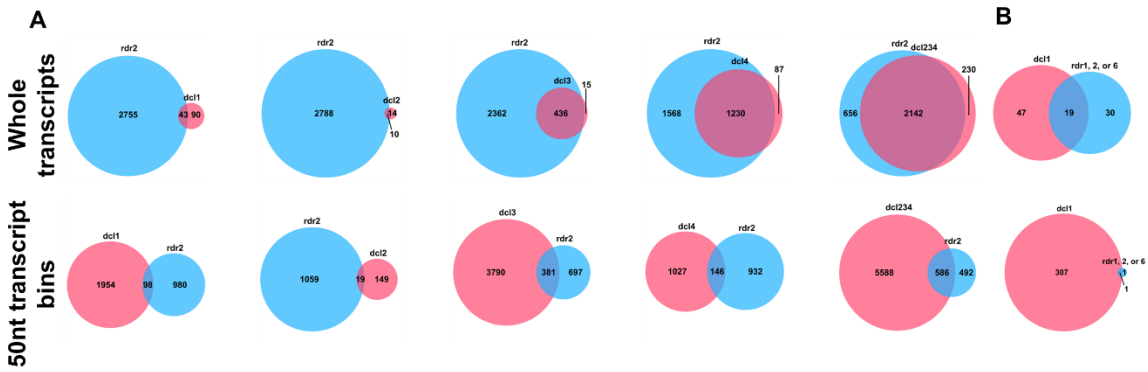

Figure 4.3: Fine-scale transcriptome binning enables detection of DCL-dependent, RDR-independent smRNA production

(A) smRNA abundance was counted at all detectable TAIR10 mRNAs and at 50-nucleotide transcriptome bins. Transcripts and bins with significantly fewer smRNAs in *dcl* mutants were overlapped with those producing significantly fewer smRNAs in *rdr* mutants. (B) As a control, overlap was calculated at TAIR10 miRNAs, which should be DCL1-depedent and RDR-independent.

4.2.3 DCL1-dependent, RDR-independent foci of smRNA production are highly structured

We then sought to link DCL-dependent, RDR-independent smRNA production with mRNA secondary structure. To this end, we compared the secondary structure, as determined by polyA-selected dsRNA/ssRNA-seq, for our identified smRNA-producing foci. Consistent with their paucity of overlap to duplex RDR-dependent smRNA bins, DCL1 and DCL2-dependent smRNA bins that are also RDR-independent show the highest degree of (intramolecular) secondary structure (**Figure 4.4A**). In contrast, DCL3, DCL4, and a combination of DCL2, 3, and 4 (*dcl2-1/3-1/4-1* mutant) show little difference when compared to all bins. Thus, DCL1 and DCL2 are the most promising candidates for cleaving mRNA secondary structure. We chose to focus on DCL1, since its bins are more numerous, and DCL1 is a well-characterized component of plant miRNA production. Within DCL1-dependent bins, those that are RDR-independent also have comparable structure to known miRNAs ($p = 0.18$, t-test), while those that overlap with RDR-dependent bins do not ($p < 0.05$, t-test) (**Figure 4.4B**). Both mRNA and miRNA-derived DCL1-dependent smRNA producing bins are significantly more structured than all mRNA bins and miRNA bins, respectively ($p < 0.05$, t-test) (**Figure 4.4B**). This implies that our strategy can also distinguish processed double-stranded from unprocessed single-stranded regions of primary miRNAs.

As a complementary analysis, we also determined the overlap between DCL-dependent smRNA producing bins and structure peaks (**Figure 4.4C**). All DCL-dependent bins show significant enrichment of structure peaks ($p < 0.05$, chi-square test) when compared to the random peaks constructed by permuting dsRNA and ssRNA

reads (**Figure 4.4C**). In summary, DCL-dependent peaks coincide with regions of high mRNA secondary structure, consistent with the known requirement of DCL for double-stranded precursors.

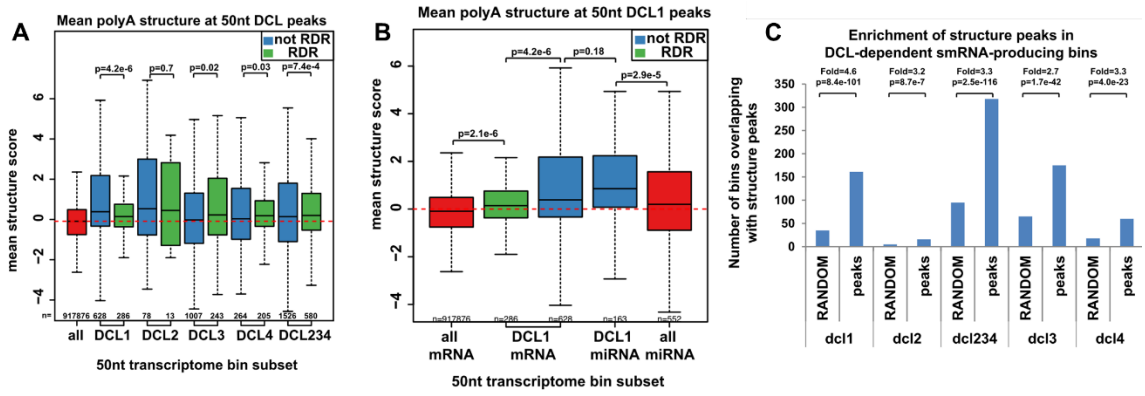

Figure 4.4: Mean structure scores and enrichment of high-structure peaks within DCL-dependent smRNA-producing bins

(A) Mean structure scores for all DCL-dependent smRNA producing bins within mRNAs that either do or do not overlap with RDR-dependent bins. (B) Comparison of mean structure for DCL1-dependent bins within mRNAs and miRNAs. P-values are measured with a t-test. (C) Overlap of DCL-dependent bins within mRNAs with high-confidence structure peaks (FDR < 0.05), as well as an equally-sized random control. Enrichment is measured as the fold-change of structure peaks over random peaks. P-values are measured by a chi-square test.

4.2.4 DCL1-dependent smRNA-producing structure peaks are longer and possess predicted stem-loop structures

Given that 1) we identified DCL1, a known miRNA processor, and 2) DCL1-dependent smRNA-producing bins in mRNAs have comparable structure to miRNAs, we reasoned that DCL1 bins in mRNAs may have similar stem-loop secondary structures to miRNAs. Since miRNA stem-loops tend to be quite long (larger than the 50-nucleotide

bins used in our analysis), we first tested whether DCL1-dependent smRNA-producing bins of high structure tend to be long as well. We observe that those DCL1 bins overlapping with structure peaks are on average longer than DCL1 bins overlapping with low structure “valleys” or random regions (**Figure 4.5A**). We then generated constrained structure models for known miRNAs and mRNAs with DCL1-dependent, smRNA producing, high structure bins. Known miRNAs display characteristic, long stem loops (**Figure 4.5B**). Interestingly, DCL1-dependent smRNA-producing bins (black arrows) in mRNAs also coincide with shorter stem-loop structures (**Figure 4.5B**), suggesting that miRNA-like stem-loops within mRNAs are recognized as miRNA precursors. Some of these stem-loops could also be bona-fide novel miRNAs, should they act in *trans* to direct transcript silencing.

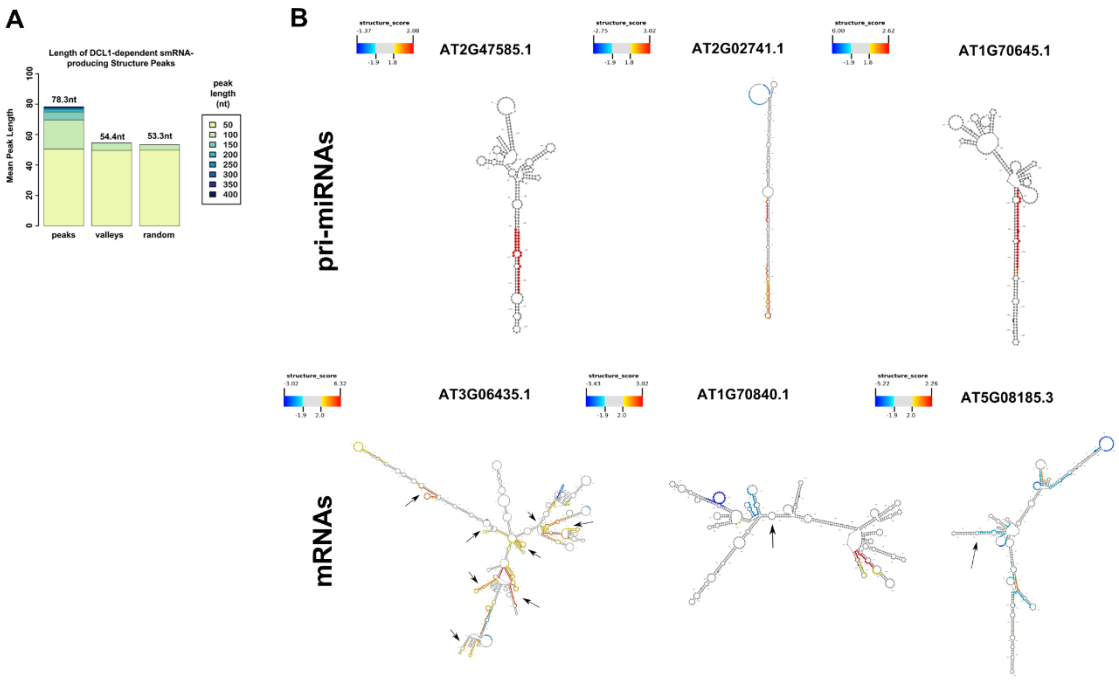

Figure 4.5: DCL1-dependent smRNA-producing structure peaks are longer and possess predicted stem-loop structures

(A) Length distribution of DCL1-dependent smRNA peaks overlapping structure peaks, valleys, or random regions. (B) Structure models of primary miRNA stem-loops and mRNAs containing DCL1-dependent smRNA-producing structure peaks but without any predicted miRNA target sites. Structure was predicted by constraining RNAFold (Zuker and Stiegler, 1981) with high-confidence paired and unpaired bases (**Appendix A.3.12**). Arrows point to DCL1-dependent smRNA-producing structure peaks within mRNAs.

4.2.5 DCL1-dependent smRNA-producing structure peaks are repressed by DCL1

Finally, we aimed to demonstrated that DCL1-dependent smRNA-producing high structure bins are functionally active in directing DCL1 targeting and transcript decay. For instance, upon loss of DCL1 we observe a significant ($p < 0.05$, t-test) increase in mRNA steady state abundance among transcripts with DCL1-dependent, RDR-independent high structure bins, but no significant increase among comparable peaks

that are also RDR-dependent (**Figure 4.6**). The latter are likely RDR-processed duplex RNAs targeted by other DCL enzymes.

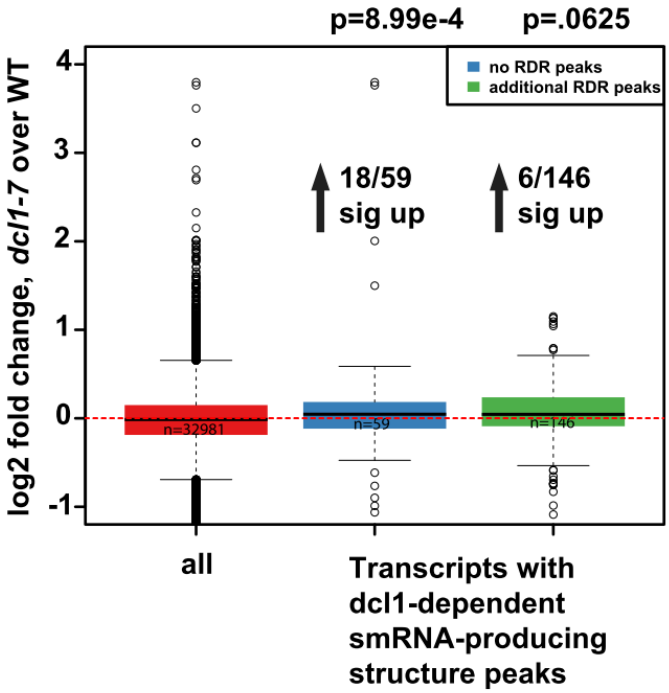

Figure 4.6: Presence of a DCL1-dependent smRNA-producing structure peak correlates with a DCL1-dependent decrease in steady-state RNA abundance

Log_2 fold change of RNA abundance in *dcl1-7* over WT plants for transcripts containing DCL1-dependent smRNA producing peaks. P-values are calculated with a t-test.

To show that this increase in steady state abundance is the result of transcript stabilization, we used actinomycin and cordycepin to inhibit transcription in protoplasts derived from *dcl1-7* and WT plants (**see Section 3.2.3**). We then tracked the rate of decay for mRNAs with DCL1-dependent smRNA-producing high structure bins (**Figure 4.7A**). Overall, we observe that for 8 of 12 such transcripts, a greater proportion of

transcript remains in *dcl1-7* mutants compared to WT, suggesting increased stability. However, these trends are subtle and assigning significance has been difficult thus far. Thus, our data is preliminary and needs further validation.

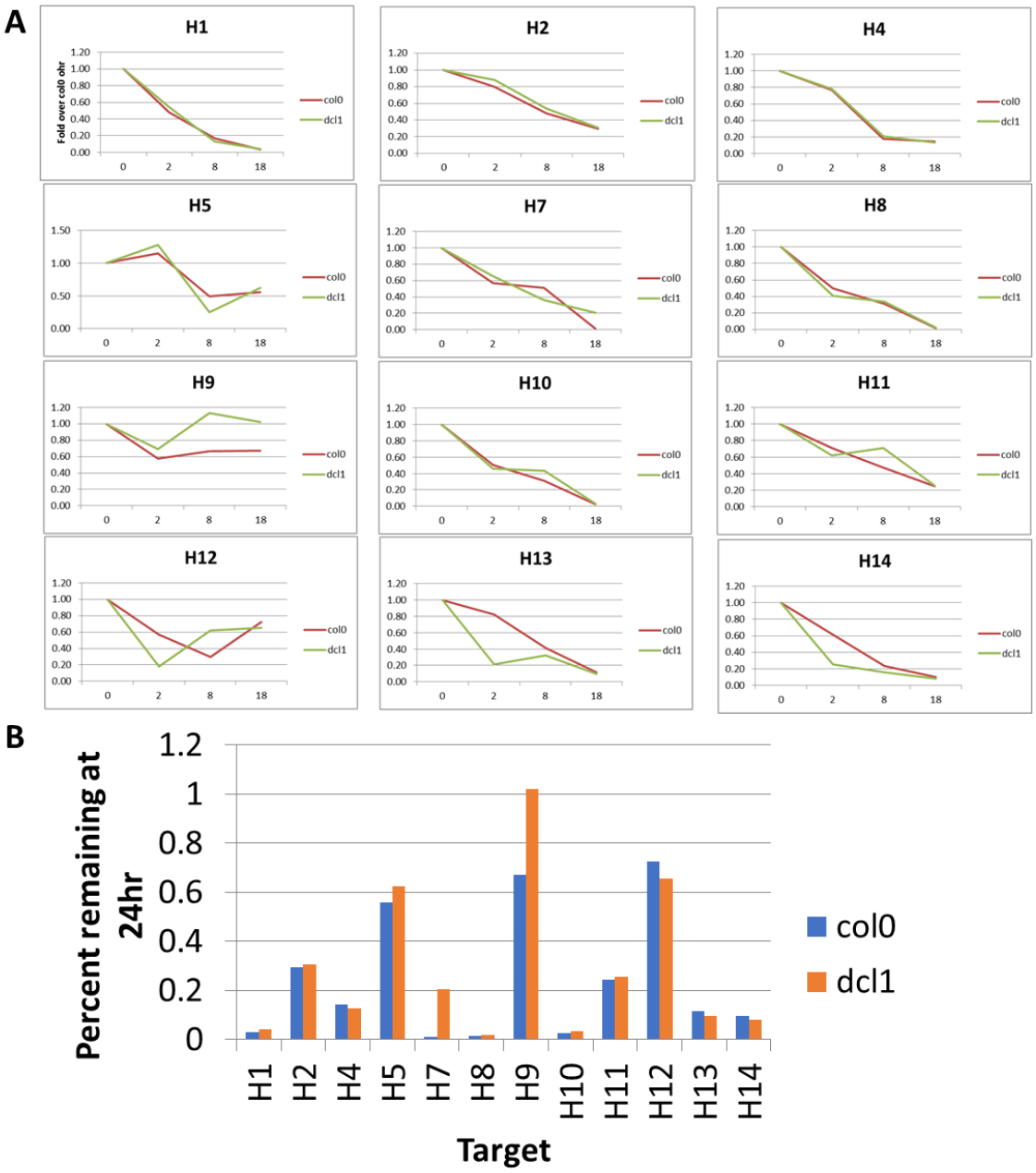

Figure 4.7: Preliminary evidence that the presence of a DCL1-dependent smRNA-producing structure peak triggers DCL1-dependent transcript destabilization

A) Decay curves, shown as proportion transcript remaining (y-axis) as a function of time following treatment with transcriptional inhibitors (x-axis). B) Proportion remaining at 24hr. “H” denotes structure “hotspots” (peaks).

4.3 CONCLUSIONS

Here, we present preliminary evidence that DCL1 targets miRNA-like elements within mRNAs. Since organisms possessing RDR enzymes possess a mixture of intermolecular mRNA duplexes and intramolecular mRNA secondary structure, we first developed a method to enrich for bona-fide secondary structure by simply polyA$^+$-selecting dsRNA/ssRNA-seq libraries. We then identified small transcriptome bins that produce smRNAs in a DCL-dependent but RDR-independent manner. For DCL1 and DCL2, these bins are on average of higher structure, and in some cases lie in predicted stem-loop structures. We then present preliminary evidence that DCL1-dependent smRNA-producing high structure bins correlate with a DCL1-dependent decrease in steady state abundance, possibly due to transcript destabilization. Future experiments are required to show this definitively.

CHAPTER 5: PARTIAL MESSENGER RNA DECAY IN THE DEVELOPING MOUSE OOCYTE

This section refers to unpublished, collaborative work done with Drs. Richard Schultz, Jun Ma, Nur Selamoglu, and Fevzi Daldal.

5.1 INTRODUCTION

The proper regulation of mRNA stability is critical for regulating the series of developmental steps that give rise to a preimplantation embryo, in particular through the maternal-to-zygotic transition (MZT) in gene expression. Prior to ovulation, oocytes enter into a long growth phase in which they accumulate maternal mRNAs and macromolecules, synthesize proteins, and greatly expand in volume in order sustain their development prior to implantation in the womb (Schultz and Wassarman, 1977). The maternal mRNAs within oocytes are highly stable (Brower et al., 1981; Jahn et al., 1976), enabling them to persist for weeks prior to ovulation. This stability is likely conferred by the binding of MSY2 RBPs, which upon phosphorylation by cyclin-dependent kinase 1 (CDK1) or genetic ablation triggers widespread RNA destabilization (Medvedev et al., 2008, 2011). Notably, a constitutive phosphomimic of MSY2 triggers premature RNA degradation (Medvedev et al., 2008), and loss of MSY2 leads to impaired maturation and female sterility (Medvedev et al., 2011), suggesting that MSY2-mediated maternal mRNA stabilization is critical for oocyte development.

As oocytes mature to become eggs, they begin to clear away maternal mRNAs, in part through MSY2 phosphorylation. While maternal mRNA decay during the maternal-to-zygotic transition is widespread and often rapid, it is also a selective process that targets distinct and functionally related sets of transcripts during different stages of oocyte development (Alizadeh et al., 2005; Clift and Schuh, 2013; Su et al., 2007; Zeng et al., 2004). For instance, as oocytes mature they degrade transcripts in involved in oxidative phosphorylation, protein synthesis, and RNA metabolism, consistent with an exit from the growth stage (Su et al., 2007). In contrast, protein kinase transcripts, which are involved in maintaining pre-fertilization metaphase II arrest, are stable during the early stages of oocyte maturation (Su et al., 2007). Likewise, certain transcripts involved in oogenesis are specifically degraded after fertilization (Alizadeh et al., 2005). To complete the maternal-to-zygotic transition, zygotes begin to transcribe their own genes. In mice, this begins as early as the 1-cell stage upon condensation of the maternal and paternal pronucleii (Latham et al., 1992; Ram and Schultz, 1993), but ramps up considerably at the 2-cell-stage (Hamatani et al., 2004; Zeng and Schultz, 2005). Maternal mRNA is also mostly cleared by the 2-cell stage (Clift and Schuh, 2013; Piko and Clegg, 1982).

While the timing and functional implications of maternal mRNA clearance is well understood, the precise mechanisms by which many of these transcripts are targeted for decay remain elusive. Even the mechanism by which MSY2 phosphorylation triggers mRNA stabilization is still unclear. Thus, we sought to address this gap in our understanding by performing high-depth, high-throughput RNA-sequencing (RNA-seq) at dense timepoints during oocyte maturation, during which maternal mRNA decay is active

but transcription is inactive. We defined time 0hr as the germinal vesicle (GV) stage, during which prophase I is arrested. After inducing maturation, we measured RNA abundance at 4-, 8-, 12- and 16-hour intervals. Notably, previous studies of maternal mRNA decay kinetics all relied upon microarrays, so we reasoned that RNA-seq should give a detailed picture of RNA decay at sub-transcript, single-nucleotide resolution, enabling a much finer scale survey of how maternal mRNAs decay.

In doing so, we uncover a class of maternal mRNAs that is only partially degraded from the 3'-end through a process that we refer to as nibbling. These transcripts are enriched for certain annotations, including involvement in RNA and protein metabolism, consistent with the groups of transcripts known to be decayed during oocyte maturation (Su et al., 2007). We then define consensus sequence elements that mark the boundary between the stable and unstable portions of each transcript, and thus could serve as roadblocks against 3'-to-5' exonucleolytic decay. Through RNA affinity purification, we then identify putative RNA binding proteins (RBPs) that mediate transcript nibbling and generate mechanistic hypotheses. For instance, we identify a GU-repeat element that is likely of high secondary structure and appears to trap RNA helicase.

5.2 RESULTS AND DISCUSSION

5.2.1 Distinct kinetic classes of degrading maternal mRNAs

To define the temporal dynamics of maternal mRNA decay at single-nucleotide resolution, we performed RNA-seq at 5 separate timepoints during preimplantation

development across three biological replicates, with a total of 25 oocytes per sample. Luciferase mRNA was spiked-in to control for the expected widespread decreases in total RNA abundance. Time 0hr was defined as oocytes in the germinal vesicle (GV) stage, which are arrested at meiotic prophase I and reside in the ovaries for long periods of an animal's lifetime. After collecting GV-stage oocytes, we induced maturation *in vitro* by incubating in milrinone-free media and collected maturing oocytes at 4hr intervals. Since these mature oocytes have yet to complete meiosis or be fertilized, they are free of transcription, and changes in RNA abundance should directly reflect RNA stability. Accordingly, the overwhelming majority of genes that are significantly (FDR < 0.05) differentially expressed across development are downregulated with the exception of time 12hr **(Figure 5.1A**), and genes that are downregulated at later timepoints have a high degree of overlap with those downregulated at earlier timepoints, suggesting this downregulation is not transient (**Figures 5.1B and 5.1C**).

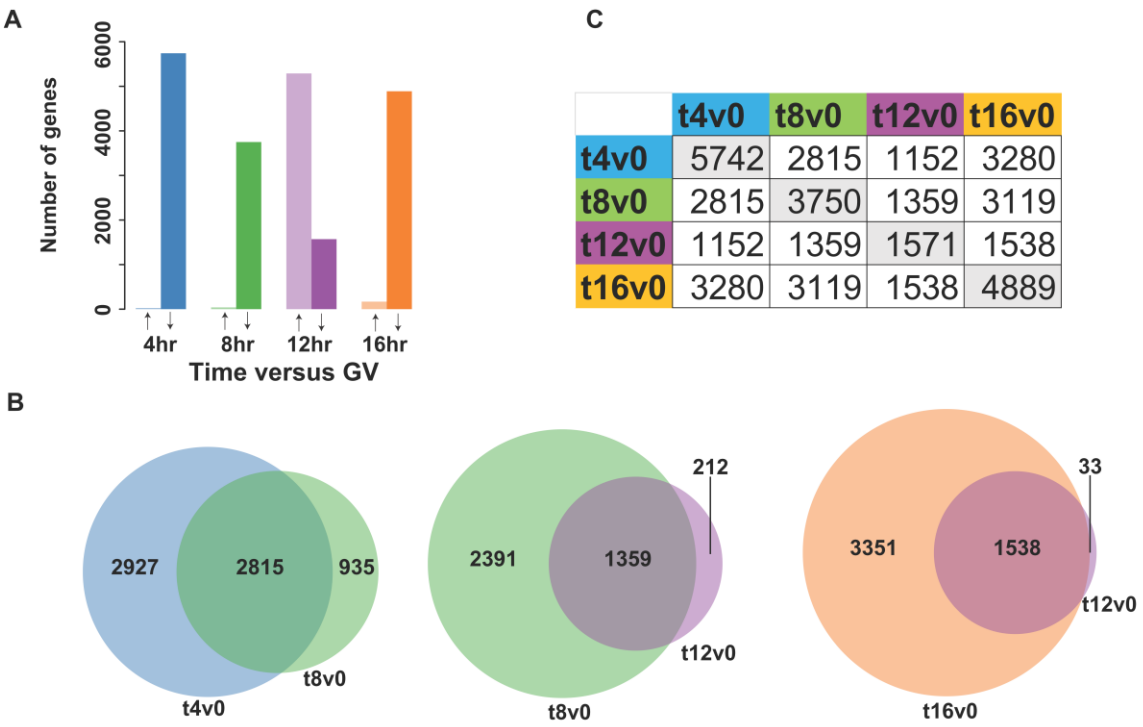

Figure 5.1: Widespread maternal mRNA decay over oocyte development

A) mRNA decay was measured by counting read abundance across all transcript isoforms per gene, and then performing differential expression analysis with the EdgeR (Robinson et al., 2010) pipeline. Up- versus down-regulation is indicated with black arrows below bars. All comparisons are between a later timepoint and GV (time 0hr) We then plot overlap between downregulated transcripts at B) successive timepoints, and C) all possible pairs of timepoints.

We then defined distinct kinetic classes of genes based upon when their downregulation was first detected over this dense series of timepoints. Our stringent definition required that for a gene to be called as significantly downregulated at a certain timepoint, it had to also be significantly downregulated at all later timepoints. With this approach, we identify 1,075 genes downregulated after 4hrs, 272 at 8hr, 121 at 12hr, and 928 at 16hr (Figure 5.2A), suggesting that there are two major waves of mRNA

decay across early preimplantation development. Conversely, we also defined highly stable mRNAs using an ANOVA-like generalized linear model that tests for significant changes between any pair of timepoints (Robinson et al., 2010). Given that maternal mRNA decay is known to target specific functional classes at different times in development (Alizadeh et al., 2005; Clift and Schuh, 2013; Su et al., 2007; Zeng et al., 2004), we then determined Gene Ontology functional enrichment among each kinetic class. Stable mRNAs are significantly (FDR < 0.05) enriched for signal transduction, development, cell adhesion, and transmembrane transport function, while genes down by 4hr (early-down) were significantly enriched for ribosomal biogenesis and redox functions (Figure 5.2B).

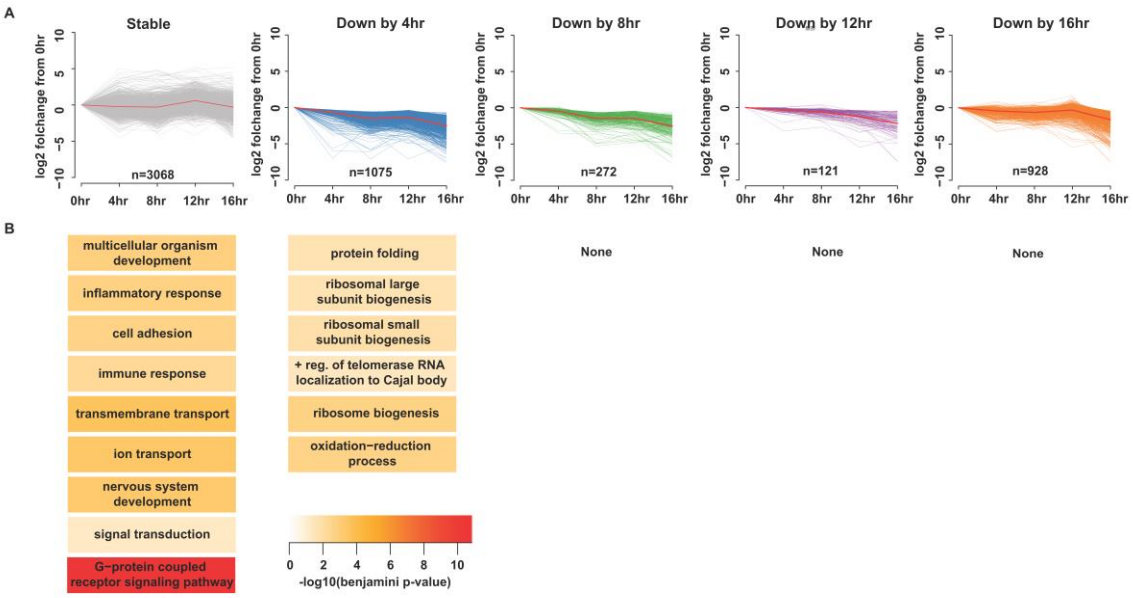

Figure 5.2: Distinct kinetic classes of degrading maternal mRNAs

A) Abundance of transcripts over time (x-axis), as a proportion of their initial abundance at 0hr (GV) (y-axis). Stable transcripts were defined as those that do not change significantly over time, usign an ANOVA-like generalized linear model in the EdgeR suite (Robinson et al., 2010). Unstable transcripts were defined as those that significantly (FDR < 0.05) decrease at a given timepoint and all subsequent timepoints. Analysis is per-gene. B) Gene Ontology term enrichment among each kinetic class. Significance is plotted as a heatmap.

5.2.2 Single nucleotide resolution RNA-seq reveals partial transcript decay

When visually inspecting profiles of RNA abundance across various transcript bodies, we noticed what appeared to be partial mRNA decay, in which one portion of a transcript was stable over time and another was disproportionately degraded (**Figure 5.3**). This suggested incomplete exonucleolytic cleavage, which we refer to as nibbling. To survey these events in an unbiased manner across the transcriptome, we define a

measure of disproportionate transcript terminus downregulation by tabulating read coverage ratios (later timepoint over 0hr) at sets of windows close to the 5' and 3' ends (**Figure 5.3, orange boxes**), and then comparing these ratios to sets of windows at the center of a transcript (**Figure 5.3, dark grey box**). Nibbled transcripts were defined as those with termini decaying at least 10x faster than the center region. We also filtered out transcripts with excessively skewed read distributions (**Appendix A.3.16**). We then sought to determine the boundary of nibbling demarcating the stable from unstable regions of a given transcript. To this end, we used a sliding-window approach, and determined at which position the change in coverage of a terminal set of windows converged to within 1.5x of the central set of windows (**Figure 5.3**). This ratio was chosen among a panel of potential ratios, based upon the subjective assertion that it most closely represented what was apparent to the human eye (**Appendix A.3.16**).

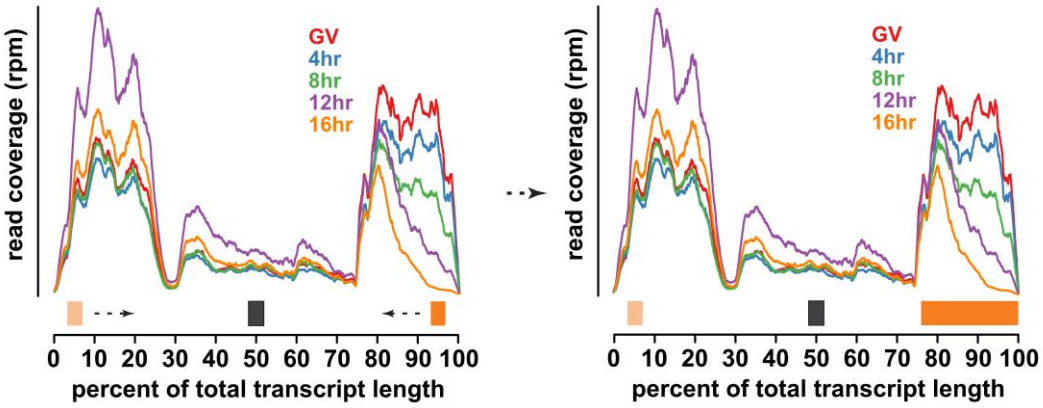

Figure 5.3: Detecting transcript nibbling

Overview of the methods for determining nibbling for an example nibbled transcript. Colors indicate each developmental timepoint, and boxes indicate window sets used to determine ratio of coverage between time 0hr and a later timepoint. Window sets are centered at the 5th, 50th, and 95th percentile of a transcript's length. Once nibbling is detected, the precise region of nibbling is determined by sliding window sets toward the transcript center (dotted arrows next to windows) until their ratio of coverage converges to within 1.5x of the ratio of the central window set (**Appendix A.3.16**).

Using this approach, we were able to separate “unnibbled” transcripts with uniform rates of transcript decay (**Figure 5.4A**) from nibbled transcripts with disproportionate terminus decay (**Figure 5.4B**). After 16hr, we identified 883 putative transcripts nibbled 3’-to-5’, but only 8 that were nibbled 5’-to-3’ (**Figure 5.4C**), indicating that the bulk of this nibbling is mediated by 3’-to-5’ RNA decay mechanisms. Most of this nibbling is only apparent by 12hr or 16hr (**Figure 5.4C**). Intriguingly, among nibbled transcripts, the size of the nibbled regions often increases monotonically from timepoint to timepoint, suggesting that nibbling could be a gradual process (**Figure 5.4B**). Moreover, nibbled transcripts do not appear stable, but are instead a subset of downregulated transcripts since the majority of nibbled transcripts are downregulated (**Figure 5.4C**) even though most downregulated transcripts are not nibbled (**Figure 5.4D**). We suspect that gene-wise downregulation is detected from nibbled regions, even though transcript decay is incomplete. We then looked to see if nibbling targets specific functional classes of transcripts, and observe significant (FDR < 0.05) enrichment of Gene Ontology annotations for mRNA processing, ribosome biogenesis, and vesicle-mediated transport (**Figure 5.4E**).

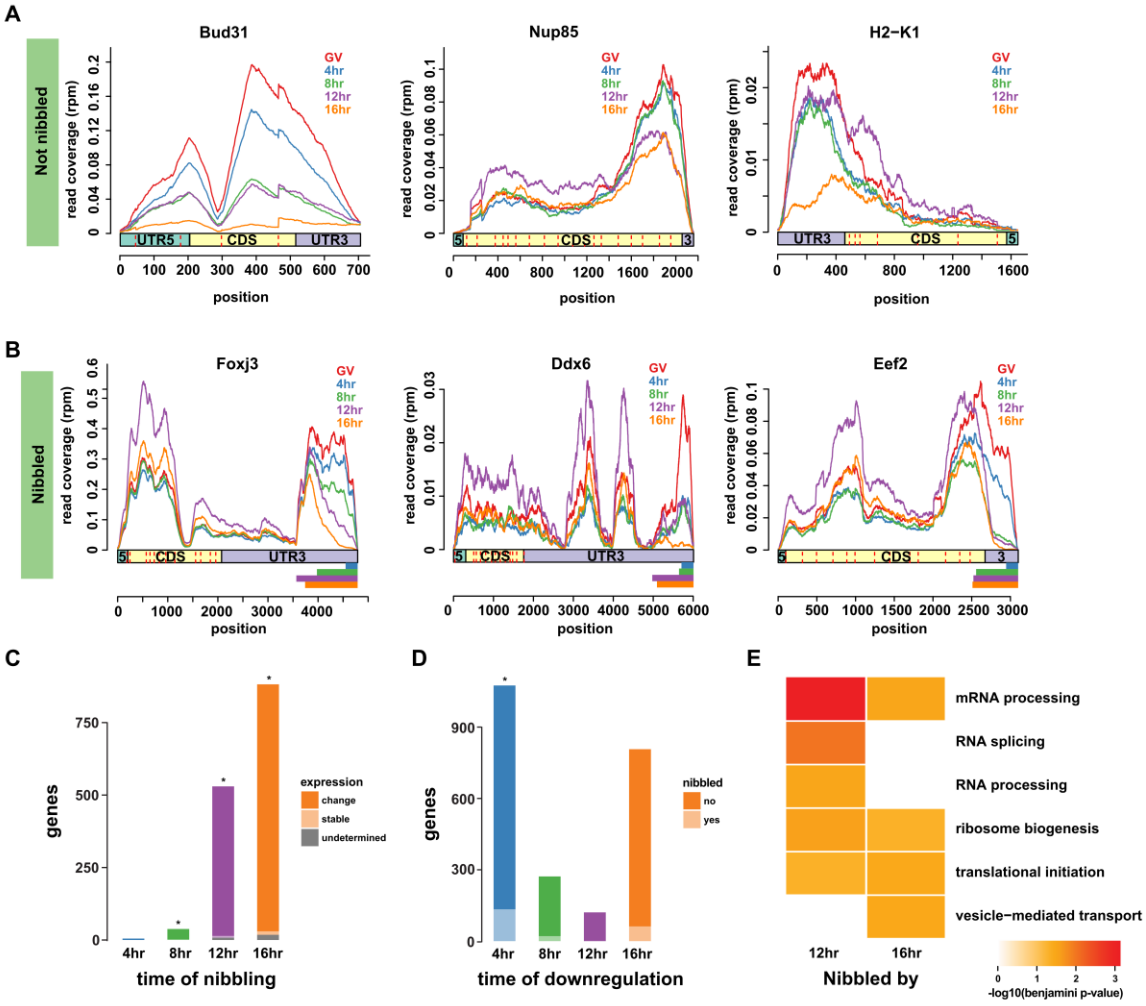

Figure 5.4: Nibbled transcripts are a subset of downregulated transcripts

A) Examples of unnibbled transcripts that decay at a more uniform rate across the transcript body. B) Examples of nibbled transcripts with disproportionate 3' end decay. Colors indicate each developmental timepoint, and boxes below indicate regions of nibbling. C) Overlap between nibbled genes and those that are stable or unstable over time. D) Overlap between significantly downregulated genes and nibbled genes. * denotes $p < 0.05$, Fisher exact test. E) Gene ontology terms significantly (FDR < 0.05) enriched among genes producing nibbled transcripts.

5.2.3 Sequence elements demarcate regions of transcript nibbling, potentially through RBP recruitment

We then sought to investigate potential mechanisms underlying the phenomenon of mRNA nibbling. Given that RNA binding proteins (RBPs) have the potential to modulate mRNA decay, and often bind to mRNAs in a sequence-specific manner, we first looked to see if there were repeated sequence elements demarcating the boundaries of nibbling. We analyzed windows from 30 nucleotides upstream to 30 nucleotides downstream of each boundary, and searched for motifs with the MEME suite (Bailey et al., 2009). We observe C/U-rich, $(GU)_{10}$, and polyA motifs (distinct from the polyA tail) (**Figure 5.5**) marking approximately 10% of nibbled genes. PolyA sequences are known to bind stabilizing RBPs such as polyA binding protein (PABP), or could alternatively be refractory to certain exoribonucleases. As well, both polyA and $(GU)_{10}$ repeats closely resemble the known binding elements of the piwiRNA (piRNA) binding protein Aubergine (Ma et al., 2017).

To identify putative RBPs in an unbiased manner, we performed RNA affinity purification using the identified sequence motifs, controlled with a scrambled (random nucleotide) 20-mer sequence. We performed this experiment in murine embryonic stem cells (mESCs) since 1) oocyte material is extremely limiting and direct proteomics are unfeasible and 2) additional validation is required regardless of the tissue type used. Top hits for polyA include polyA binding protein 1 (PABP1), the U3 SnRNP, and peroxiredoxin-1 (**Figure 5.5A**). For C/U-rich motifs, PABP1 is also pulled down, as is the

polyU splicing factor PUF60. Other hits include nucleolysin TIAR, the elongation factor eIF1D, and the DNA replication factor MCM7 (**Figure 5.5C**). For $(GU)_{10}$ motifs, hits include peroxiredoxin 1, FUSE-BP3, and the RNA helicase DHX36 (**Figure 5.5B**). Notably, G and U can form stable Hoogstein base pairs in RNA, and thus the $(GU)_{10}$ is likely of high secondary structure, which could explain why it recruits (and potentially stalls) RNA helicase. Future experiments will use siRNA injection of oocytes to knock down candidate RBPs involved in nibbling. If an RBP's potential involvement is true, then nibbled transcripts should be completely degraded upon RBP knockdown, leading to reduced overall expression. A second, though more laborious approach, is to repeat RNA-seq experiments in oocytes derived from genetic knockouts of RBPs.

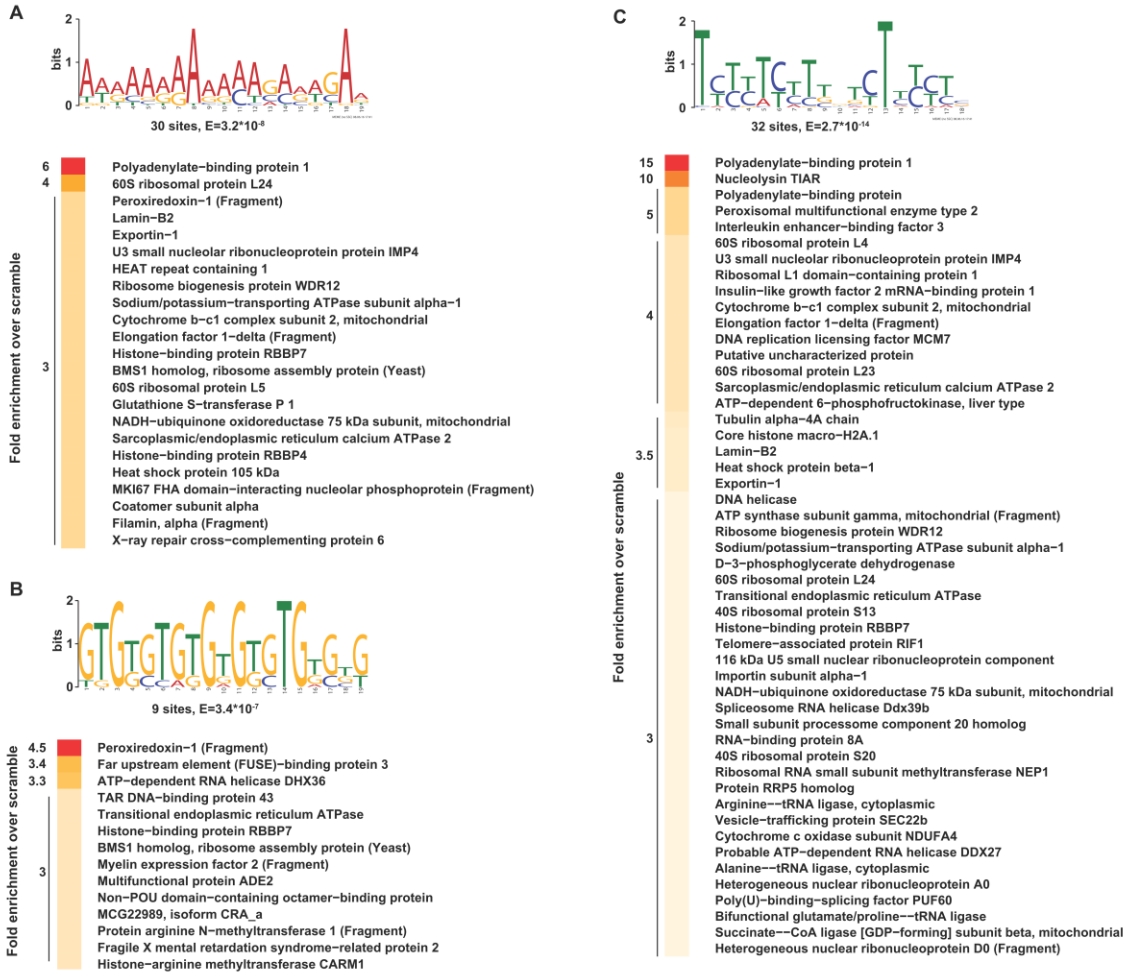

Figure 5.5: Candidate RNA binding proteins recognizing nibbling boundary elements

Using MEME, motifs were elicited from sequences between 30 nucleotides upstream and 30 nucleotides downstream of nibbled boundaries. Identified motifs were then used to construct baits for RNA affinity chromatography in mESCs. Putative RBP partners are plotted for the A) polyA motif, B) $(GU)_{10}$ motif, and C) C/U-rich motif. Fold enrichment over a scrambled RNA control bait is plotted both numerically (y-axis) and as a heatmap.

5.3 CONCLUSIONS

In summary, we provide evidence for a potentially novel form of partial mRNA decay in the developing, preimplantation mouse oocyte. The single-nucleotide resolution of RNA-seq allows us to look for regions at the termini of a transcript that are disproportionately degraded relative to the transcript body. We speculate that this partial mRNA decay could be a mechanism to retain a readily available source of nucleic acids in concentrated regions such as P-bodies. Unlike free NTPs, partially degraded mRNAs could be selectively localized and become readily available during the burst of zygotic transcription required to complete the maternal-to-zygotic transition. At this point, however, our work is still preliminary, and additional validation is required to show the requirement of candidate RBPs to mediate transcript nibbling.

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, I present a series of transcriptome-wide studies that contribute to our understanding of how features of the epitranscriptome influence mRNA stability. Moreover, I contribute to the development of high-throughput methods that are broadly applicable across biological systems and can often be applied in retrospect to existing RNA sequencing (RNA-seq) datasets. In **Chapter 2**, I uncover a link between mRNA decapping and covalent modifications in both plants and humans, suggesting that modifications influence mRNA stability. In **Chapter 3**, I survey dynamic mRNA modifications during plant salt stress to directly show that modifications in uncapped, degrading mRNAs associate with changes in mRNA stability. In **Chapter 4**, I identify structural elements in mRNAs that resemble miRNA stem-loops and are processed to smRNAs in a DCL1-dependent manner. In **Chapter 5**, I exploit the single nucleotide resolution of RNA-seq to identify partial 3'-to-5' exonucleolytic decay in the maturing mouse oocyte, and identify potential sequence elements and RNA binding proteins (RBPs) that mediate this phenomenon. Here, I further discuss the broader impacts of these studies, frame their biological relevance, and provide future directions to build upon my dissertation.

6.1 RNA COVALENT MODIFICATIONS: NOVEL INSIGHTS AND FUTURE DIRECTIONS

While there are a host of well-established techniques for studying individual mRNA modifications across the whole transcriptome, our High Throughput Annotation of Modified Ribonucleotides (HAMR) pipeline used in **Chapters 2 and 3** is unique in its ability to probe a host of different mRNA modifications, so long as they alter Watson-Crick base pairing (Ryvkin et al., 2013; Vandivier et al., 2015a). Thus, while there are existing studies demonstrating that modifications such as m^6A can alter mRNA stability (Du et al., 2016; Wang et al., 2014b), we provide some of the first evidence of this effect for modifications across all four bases (**Figures 2.12 and 3.2**). These include modified guanosines, which have not been previously characterized in-depth in mRNAs.

HAMR is also unique insofar as it can be applied retrospectively to any RNA-seq dataset that does not contain experimentally-induced mutations (for instance, cross-linking induced mutations). Thus, it can be readily deployed across any organism of interest, and also over a variety of specialized libraries types that would be difficult to combine with reverse-transcriptase (RT) stalling or antibody pulldown based methods (**Section 1.7**). One such library is Global Mapping of Uncapped Transcripts (GMUCT), which probes for uncapped degrading mRNAs (Gregory et al., 2008; Willmann et al., 2014), in which we were able to observe a strong enrichment of modified ribonucleotides, suggesting that they are a cause or consequence of mRNA turnover. Likewise, we have applied HAMR to small RNAs (smRNAs) and ribosomal footprints, and have observed unique modification profiles across each mRNA subpopulation. While numerous existing studies profile mRNA modifications across time and stress, our

work described in **Chapters 2 and 3** is among the first to profile changes across diverse populations and fragments of mRNAs.

In doing so, our work also provides insights into potential mechanisms by which mRNA modifications might lead to transcript decay. For instance, bases with modifications observed in uncapped, degrading mRNAs are approximately twice as likely as unmodified bases to coincide with miRNA target sites (**Figure 3.8**), and over 15 times more likely to occur within ribosome pause sites (**Figure 3.11**). Thus, mRNA modifications could modulate stability by altering the ability of miRNA-bound RISC complexes to target transcript cleavage, or by triggering ribosome stalling and subsequent no-go decay. To address these potential mechanisms, future studies will apply HAMR to mutants in the no-go decay pathway, as well as in various components of RNA decay. If modifications trigger no-go decay, then their knockout should lead to a decrease in modifications in uncapped RNAs and a buildup of modifications in stable polyA$^+$ RNAs. Similar patterns should be observed in mutants lacking functional RNA exonucleases, though preliminary analysis of mutants impaired for exonucleolytic cleavage (*rail1-1*) show a decrease in uncapped mRNA modifications, suggesting that some modifications could be a consequence rather than a cause of mRNA decay (**Figure 3.13**).

It will also be critical to establish whether mRNA modifications are necessary and sufficient to direct changes in stability. To establish necessity, future studies will assay for HAMR-predicted modifications in mutants lacking mRNA modification writers and erasers such as methyltransferases or pseudouridine synthases. Global changes in mRNA abundance and stability will suggest a role for these enzymes in mediating

transcript stability, though it will be hard to rule out pleiotropic effects given these enzymes are often responsible for modifying many other classes of RNAs, such as tRNAs. More targeted approaches will involve reporter genes that either do or do not contain modified ribonucleotides. We will assay for both reporter abundance and stability with either metabolic labelling or transcriptional inhibitor-based assays (**Section 1.3**) in the presence or absence of modifications to establish the sufficiency of modifications to direct transcript decay.

Through **Chapters 2 and 3**, we also observed a link between mRNA modifications and a variety of gene functions not associated with basal physiology, such as stress response in plants and cell death in animals (**Figures 2.16 and 2.17**). We then showed that in plants, numerous modifications in uncapped degrading mRNAs are gained or lost during salt stress, correlating with changes in stability. While we were able to observe some of these modifications in salt stress-related transcripts, we have yet to show that mRNA modifications are in fact mediators of salt stress response. Thus, we also hope to test whether mRNA modifications are required for salt stress, for instance by point mutation of differentially modified bases in stress response transcripts followed by measurement of sensitivity to salt.

In summary, we demonstrate the utility of the HAMR pipeline in surveying mRNA modifications across diverse organisms and RNA subpopulations. Since HAMR is broadly applicable to most RNA-seq datasets and can be applied to many RNA subpopulations of interest, it has the potential to become an important resource for the broader field of RNA epitranscriptomics and post-transcription regulation.

6.2 MRNA SECONDARY STRUCTURE AND DICER-LIKE-MEDIATED DECAY

In **Chapter 4**, we apply nuclease mediated structure probing (**see Section 1.5.3**) with the aim of identifying miRNA precursor-like elements in *Arabidopsis* coding mRNAs. In doing so, we identify a link between high secondary structure and the production of smRNAs in a DICER-LIKE1 (DCL1)-dependent manner, and show that mRNAs containing these elements tend to be repressed by DCL1. This suggests that the miRNA processing machinery can directly recognize mRNAs as miRNA-generating precursors, and raises the question of how this machinery generally distinguishes between "bona-fide" miRNA precursors and the numerous other mRNAs with high degrees of base pairing. It has been argued previously that certain "licensing factors" bind to regions of primary miRNAs to specify their identity as miRNA precursors (Auyeung et al., 2013). Our work also suggests that the size of structural elements could be important in specifying processing by the miRNA machinery, as DCL1-processed structural elements tend to be longer (**Figure 4.5**). It is thus possible that smRNA processing is avoided by keeping mRNA structural elements within a shorter length window.

Our work also raises the intriguing possibility that mRNA-derived smRNAs could act in *trans* to trigger downstream effects on gene silencing. Future experiments will test this possibility by looking at whether mRNA-derived smRNAs produced in a DCL-dependent manner are incorporated into ARGONAUTE (AGO) proteins, which are key components of smRNA-induced RNA silencing. We will also determine if AGO-incorporated mRNA-derived smRNAs can target in *trans* by 1) *in silico* prediction of

target sites, and 2) determining if knocking out an mRNA “precursor” or mutating a putative target sequence alters the abundance and stability of the target mRNA. Alternatively, mRNA-derived smRNAs could simply be a byproduct of a novel pathway of mRNA degradation. In either case, we should be able to describe a novel mode of mRNA degradation.

Currently, our work seeks to remove potentially confounding duplex RNA secondary structures with a combination of polyA$^+$ selection before structure mapping and with single RDR mutants. Future work will approach this problem more directly by constructing combined *rdr1/2/6* and *dcl* mutants, in which DCL-dependent smRNA production in the absence of the most common RDRs is directly quantified. This will provide stronger evidence of DCL-dependent, RDR-independent production of smRNAs from mRNA structural elements.

We will then aim to show a causal role for mRNA secondary structure in directing DCL cleavage. Though challenging, one approach would be to introduce point mutants in putative structural elements within reporter genes with the goal of inducing their unwinding. If our hypothesis is true, we expect to see a decrease in smRNA production and an increase in both steady state abundance and transcript stability upon disruption of secondary structure.

Regardless of their mechanism of action, any structural element that is sufficient to direct transcript destabilization could have important biotechnological applications in tuning down expression of transgenes. Thus, we will also determine if adding structural elements to reporters is sufficient to direct transcript destabilization. Ideally, we will

observe varying degrees of destabilization, would could create a useful “toolbox” of elements that may be added to transgenes.

6.3 DEFINING THE MECHANISM AND RELEVANCE OF PARTIAL MRNA DECAY

Beyond quantitating transcript abundance, RNA-seq is a rich and often untapped source of potentially novel biological data. In **Chapter 5**, I describe a method of detecting partial mRNA decay from RNA-seq data, which like HAMR can help to expand the amount of data gathered from a given RNA sequencing experiment. This method exploits disproportionate changes in read abundance across a transcript body to define transcripts in which the 3’ or 5’ ends degrade faster than expected, and thus suggest a new form of partial exonucleolytic mRNA decay that we call nibbling. While limited mRNA nibbling (on the order of 10 nucleotides) has been described in yeast (He and Parker, 2001), to our knowledge this is one of the first studies to suggest longer nibbling (on the order of 100 nucleotides) in higher eukaryotes.

We have also identified consensus sequence elements that mark the boundaries of nibbling within approximately 10% of nibbled transcripts (**Figure 5.5**). We then used RNA affinity purification to identify candidate RBPs that could mediate nibbling (**Figure 5.5**). For instance, RNA helicase is pulled down by a $(GU)_{10}$ repeat that of likely of high secondary structure, and could serve as a roadblock toward unwinding RNA during the course of exonucleolytic decay. However, these candidate RBPs will need to be validated, for instance by siRNA-mediated knockdown of these RBPs *in vivo* and

measurement of whether nibbled transcripts containing an RBP's target sequence elements are totally decayed instead of nibbled.

Since developing oocytes must prepare for a burst of zygotic transcription after fertilization, we suspect that these partially degrading mRNAs could serve as a store of nucleotides that are both readily metabolized (similar to free nucleotides) but also able to specifically localize to subcellular compartments (similar to intact mRNAs). To garner more evidence for this hypothesis, future studies will focus on whether nibbled mRNAs tend to display punctate localization patterns via RNA fluorescence *in situ* hybridization (FISH). If so, we will then test whether these foci overlap with regions of active zygotic transcription.

6.4 CONCLUDING REMARKS

mRNAs live a complex lifecycle and encode information with both their primary sequence and through epitranscriptomic features such as covalent ribonucleotide modifications, secondary structure, and RBPs. In this dissertation, I explore novel links between these features and mRNA stability across both plant and mammalian transcriptomes. In **Chapter 2**, I uncover a link between covalent modifications and uncapped, degrading mRNAs in both plants and mammals. In **Chapter 3**, I further develop this work and show that changes in the modification status of uncapped degrading mRNAs in response to long term salt stress associates with changes in mRNA stability. In **Chapter 4**, I uncover miRNA-like structural elements within mRNAs that appear to be targeted for smRNA production by DICERs. In **Chapter 5**, I uncover

sequence elements and their associated RBPs that appear to demarcate regions of partial mRNA decay in the developing mouse oocyte. Throughout my dissertation, I illuminate how the epitranscriptome can modulate known and novel forms of mRNA decay, and identify multiple new paths for future study. Moreover, I contribute to the development and application of transcriptome-wide techniques for probing the breadth and functional relevance of the epitranscriptome. These techniques are broadly applicable across different biological contexts, and should provide invaluable resources for the field of RNA post-transcriptional regulation.

APPENDIX A: MATERIALS AND METHODS

This section refers to work from:

Vandivier L.E., Anderson, Z.D., and Gregory BD (2017). Differential messenger RNA modification alters transcript stability upon long term salt stress. In preparation.

Vandivier L.E., Campos R., Kuksa P.P., Silverman I.M., Wang L.S., and Gregory BD (2015). Chemical Modifications Mark Alternatively Spliced and Uncapped Messenger RNAs in Arabidopsis. *Plant Cell*. 27, 3024-37. PMID: 26561561

Vandivier L.E., Li F., and Gregory B.D. (2015). High-Throughput Nuclease-Mediated Probing of RNA Secondary Structure in Plant Transcriptomes. 1284, 41-70. PMID: 25757767

Vandivier L., Li F, Zheng Q, Willmann M, Chen Y, Gregory B. (2013). Arabidopsis mRNA secondary structure correlates with protein function and domains. *Plant Signaling and Behavior*. 8, e24301. PMID: 23603972

Here, I outline materials and methods for all experiments described in this dissertation, over four subsections. In **Section A.1**, I present biological materials and model organisms. In **Section A.2**, I give an overview of experimental techniques used to generate raw data. In **Section A.3**, I describe computational, statistical, and analytical methods used to process this data. In **Section A.4**, I designate accession numbers for

data generated in these studies and for the previously published data analyzed as part of these studies. I also point to relevant software repositories.

A.1 BIOLOGICAL MATERIALS AND MODEL ORGANISMS

A.1.1 Arabidopsis tissue

For all experiments, plant material was derived from the Columbia (Col-0) ecotype of *Arabidopsis* grown under 16-hours light/8-hours dark. In salt stress experiments, we used UBQ:NTF/ACT2p:BirA (Col-0 ecotype) plants transformed with a nuclear label for the Isolation of Nuclei Tagged in Specific Cell Types (INTACT) method (Wang and Deal, 2015) so that results could be direct compared with other ongoing salt stress studies in the Gregory lab that focus on purified nuclei. Seeds were vernalized for 3 days on thoroughly soaked soil at 4°C before starting growth (day 0 of age). For the experiments described in **Chapters 2 and 4**, plants were fertilized once with 1.25 cc of 20-20-20 fertilizer per flat, and bottom watered with tap water approximately twice per week. We harvested immature flower bud clusters (inflorescences) for plants of approximately 5 weeks of age. For the salt stress experiments described in **Chapter 3**, plants were bottom watered with 1L Hoagland's solution per flat at 4-day intervals. For salt stress, 50mM NaCl (pH5.5) was added to the first watering at 1 week, and 100mM NaCl (pH5.5) was added to subsequent waterings. We then harvested pre-bolting rosettes of approximately 25 days of age.

A.1.2 Arabidopsis protoplasts

For RNA stability experiments in **Chapters 3 and 4**, mesophyll protoplasts were isolated from pre-bolting rosettes using a solution of 1.5% cellulase R10 (Yakult), 0.4% macroenzyme R10 (Yakult), 0.4 M mannitol (Sigma), 20 mM KCl, 20 mM MES at pH 5.7 (Sigma-Aldrich), 10 mM CaCl2, and 0.1% BSA, based on a protocol from the laboratory of Dr. Jen Sheen (Yoo et al., 2007). Approximately 50 leaves were used for each treatment/replicate combination. Protoplasts were filtered and washed in W5 media (Yoo et al., 2007), and intact cells were enriched by collecting only those that settled in wash media after 30 minutes. RNA yield was determined to be approximately 1 µg per $1.17*10^5$ protoplasts

A.1.3 Arabidopsis genotypes used in this study

All plant lines used in this study are outlined in **Table A.1**. ABRC refers to the Arabidopsis Biological Resource Center (https://abrc.osu.edu/).

Line	Source	Used in Chapter
Col-0	ABRC #CS70000	2,3,4
dcl1-7	Dr. Scott Poethig	4
dcl2-1	ARBC #CS66078	4
dcl3-1	ABRC #CS16390	4
dcl4-1	Gasciolli *et al.* 2005	4
dcl2-1/dcl3-1/dcl4-1	Dr. Xuemei Chen	4
***UBQ*:NTF/*ACT2p*:BirA**	Deal and Wang, 2015	3
rail1-1	ABRC	3
abh1-1	ABRC #CS66124	3
rdr1-1	ABRC #CS66077	4
rdr2-1	ABRC #CS66076	4
rdr6-15	ABRC #CS879578	4

Table A.1: *Arabidopsis* genotypes used in this study

A.1.4 Human cell lines

HeLa and HEK293T cells were seeded in 15-centimeter standard Corning tissue culture dishes (Sigma, St Louis, MO), grown to 90% confluence (approximately 18 million cells) in DMEM media (Life Technologies, San Diego, CA) supplemented with L-glutamine, 4.5 g/L D-glucose, 10% fetal bovine serum (FBS) (Atlanta Biologics, Atlanta, GA), and Pen/Strep (Fisher Scientific, Waltham, MA).

A.1.5 Mouse cell lines

V6.5 murine embryonic stem cells (mESCs) were grown, without feeder cells, to 100% confluence on 10cm gelatinized plates with ESCM (DMEM with 15% FBS, Pen/Strep, L-Glutamine, 0.1 mM NEAA, 0.1 mM beta-mercaptoethanol, and LIF). Cells were then harvested with Trypsin-EDTA from a total of six plates and were lysed in 25mM Tris-HCL (pH7.5), 150mM KCl, 5mM EDTA, 0.5% NP-40, and cOmplete™ Protease Inhibitor Cocktail (Sigma Aldrich 1 tablet per 50ml).

A.1.6 Mouse oocytes

Germinal vesicle (GV)-stage oocytes were collected as previously described (Ma et al., 2001; Schultz et al., 1983). GV oocytes were cultured in Chatot Ziomek Brinster (CZB) media with 2.5 µM milrinone (Sigma), which inhibits GV breakdown (GVBD) through inhibition of cyclic nucleic acid phosphodiesterases. To induce *in vitro* maturation, oocytes were transferred to CZB media without milrinone.

A.2 EXPERIMENTAL METHODS

A.2.1 RNA extraction – Arabidopsis

Arabidopsis bud tissue and rosettes were ground with a mortar and pestle under liquid nitrogen, and suspended in Qiazol (Qiagen). Tissue was then homogenized with a QiaShredder column (Qiagen), and RNA was extracted via 5:1 phenol:chloroform

extraction. RNA was further purified with the miRNeasy Mini Kit (Qiagen) per manufacturer's protocol.

A.2.2 RNA extraction – Human cells

Cells were scraped, pelleted, and homogenized before suspension in Qiazol (Qiagen). Tissue was then homogenized with a QiaShredder column (Qiagen), and RNA was extracted via 5:1 phenol:chloroform extraction. RNA was further purified with the miRNeasy Mini Kit (Qiagen) per manufacturer's protocol.

A.2.3 RNA extraction – Mouse oocytes

RNA was extracted from 25 oocytes per sample using PicoPure RNA isolation Kit (Thermo Fisher Scientific) with on-column genomic DNA digestion per manufacturer's protocol. 0.2pg of *in vitro* synthesized Renilla Luciferase mRNA was spiked in before extraction to allow accurate normalization despite widespread mRNA decay.

A.2.4 RNA stability assays

Arabidopsis mesophyll protoplasts were isolated as described in **Appendix A.1.2**. To ensure at least 1 µg of RNA yield, equal numbers of protoplasts (in excess of $1.17*10^5$ per sample) were then added to W1 incubation media (Yoo et al., 2007) spiked with 33 µg/ml actinomycin-D (Research Products International) and 100 µg/ml cordycepin (Sigma-Aldrich), as described in a published *Arabidopsis* protoplast-based RNA stability assay (Leonhardt et al., 2004). Cells were incubated for 4 or 19 hours at

room temperature under ambient light. Incubated controls cells were designated as time 0hr. RNA was harvested as described above, though grinding and tissue homogenization were omitted. Instead, protoplasts were lysed by vortexing in Qiazol for 5 seconds and incubating on ice for 10 minutes before extracting RNA. After qPCR as described below, all cycle thresholds (CTs) were normalized to ACTIN7 (AT5G09810), which was the most stable gene identified in a whole-transcriptome stability assay in *Arabidopsis* (Narsai et al., 2007). These delta(CTs) (dCTs) were renormalized to 0hr samples.

A.2.5 RNA Immunoprecipitation

Total RNA was immunoprecipitated with an immunoglobulin G (IgG) nonspecific control antibody (Cell Signaling) or an anti-3-methylcytosine (m^3C) antibody (Active Motif). 40 μl of Dynabeads Protein A (Thermo Fisher Scientific) were washed with 1x Dulbecco's phosphate-buffered saline (DPBS, Thermo Fisher Scientific), and coupled to 10 μg of antibody in DPBS by rocking at room temperature for 1 hour. Beads were washed again twice with DPBS. 5 μg of RNA was denatured at 70°C for 5 minutes, placed on ice for 3 minutes, and then incubated with the bead-linked antibodies in immunoprecipitation (IP) buffer (140mM NaCl, 0.05% v/v Triton X-100, 10mM Tris, all from ultrapure, RNase-free stocks dissolved in DEPC-treated water and filter sterilized at 0.22 μM). Bead/RNA mix was rocked at 4°C for 2 hours. Bound RNA was washed three times in IP buffer and then eluted in Trizol (Thermo Fisher Scientific), precipitated, and washed.

A.2.6 Quantitative PCR

For RNA stability experiments, RNA was reverse transcribed using oligo-dT to enrich for intact, polyadenylated transcripts. For RNA IP, RNA was reverse transcribed using random hexamers since antibody-bound fragments will not necessarily be polyadenylated. cDNA was then preamplified with the SsoAdvanced™ PreAmp Supermix (Bio-Rad Laboratories) for 12 cycles, per the manufacturer's protocol and using a pool of all primers (500nM each) to be used in downstream analyses. Quantitative PCR was performed using SYBR Green 2x master mix (Thermo Fisher for **Chapters 2 and 4**, BioTool for **Chapter 3**) in a QuantStudio 3 machine (Thermo Fisher Scientific).

A.2.7 RNA-seq library preparation – Arabidopsis and human cells

RNA-seq were constructed as previously described (Li et al., 2012a). 5µg of total RNA was fragmented, subjected to two rounds of polyA$^+$ selection using oligo-dT Dynabeads (Thermo Fisher Scientific), ligated to TruSeq smRNA adaptors (Illumina), amplified and indexed, and then sequenced on an Illumina HiSeq 2000 (Illumina) using 50 bp single-end geometry. All sequencing was carried out by the High Throughput Genomics Shared Resource at the Huntsman Cancer Institute, University of Utah (**Chapters 2 and 3**) or the University of Pennsylvania Next Generation Sequencing Core (**Chapter 4**).

A.2.9 RNA-seq library preparation – mouse oocytes

RNA-seq libraries for mouse oocytes were generated using the Ovation RNA-seq system V2 (NuGEN) coupled to the Ovation Ultralow Library system / DR Multiplex System (NuGEN). RNA was extracted from a total of 25 oocytes per sample (**Section A.2.3**). Libraries were sequenced on an Illumina HiSeq 2000 (Illumina Inc.) using 125 bp paired-end geometry. All sequencing was carried out by the High Throughput Genomics Shared Resource at the Huntsman Cancer Institute, University of Utah.

A.2.10 smRNA-seq library preparation – Arabidopsis and human cells

smRNA-seq were constructed as previously described (Li et al., 2012a). 25 µg of total RNA was size-selected to fragments between 15 to 50 nucleotides. Fragments were then ligated to TruSeq smRNA adaptors (Illumina), amplified and indexed, and sequenced on an Illumina HiSeq 2000 (Illumina) using 50 bp single-end geometry. All sequencing was carried out by the High Throughput Genomics Shared Resource at the Huntsman Cancer Institute, University of Utah (**Chapters 2 and 3**) or the University of Pennsylvania Next Generation Sequencing Core (**Chapter 4**).

A.2.11 GMUCT library preparation – Arabidopsis and human cells

GMUCT libraries were constructed as previously described (Gregory et al., 2008; Willmann et al., 2014). 25 µg of total RNA was subjected to two rounds of polyA$^+$ selection using oligo-dT Dynabeads (Thermo Fisher Scientific), and then directly ligated to 5’ TruSeq smRNA adaptors (Illumina) such that only fragments with free 5’ phosphate

groups are captured. 5’ adaptor-linked fragments were polyA^{+}-selected again to remove free adaptors, and then reverse transcribed with primers containing both a downstream random hexamer and an upstream 3’ adapter. cDNA was amplified and indexed, and then sequenced on an Illumina HiSeq 2000 (Illumina) using 50 bp single-end geometry. All sequencing was carried out by the High Throughput Genomics Shared Resource at the Huntsman Cancer Institute, University of Utah.

A.2.12 Ribo-seq library preparation – Arabidopsis cells

Ribosome footprinting libraries were constructed by adapting a protocol from Mustroph and colleagues (Mustroph et al., 2009), in which we isolate polysomes by ultracentrifugation of tissue lysates through a sucrose cushion. Approximately 3ml of ground rosette tissue was used per sample. To isolate ribosome-bound footprints, polysomes were treated with 18.75 µl of E. coli RNase I (Ambion). Both of these steps were done with added cycloheximide and chloramphenicol to freeze ribosomes in place by preventing elongation These footprints were then ligated to TruSeq smRNA adaptors (Illumina), amplified and indexed, and then sequenced on an Illumina HiSeq 2000 (Illumina) using 50 bp single-end geometry. All sequencing was carried out by the High Throughput Genomics Shared Resource at the Huntsman Cancer Institute, University of Utah.

A.2.13 Structure mapping with dsRNA/ssRNA-seq

Structure mapping with structure-specific nucleases was performed as previously described (Li et al., 2012a; Vandivier et al., 2015b). For each sample, 100 µg of total RNA was split into dsRNA and ssRNA treatment groups. dsRNA groups were incubated with 1 µl of the ssRNase RNase ONE (Promega) in RNase ONE 1x Reaction Buffer (Promega), and ssRNA groups were incubated with 5 µl of the dsRNase RNase V1 (Life Technologies, discontinued by manufacturer) in 1x Structure Buffer (Life Technologies, discontinued by manufacturer). Both incubations were at 37°C for 1 hour, and were designed to cut to completion. dsRNA and ssRNA fragments were then size-selected, ligated to TruSeq smRNA adaptors (Illumina), amplified and indexed, and sequenced on an Illumina HiSeq 2000 (Illumina) using 50 bp single-end geometry. All sequencing was carried out by the University of Pennsylvania Next Generation Sequencing Core.

A.2.14 RNA affinity pulldowns

RNA affinity pulldowns were performed as previously described (Foley et al., 2017). To link RNA probes to breads, 500pmol of each probe was 1) diluted 1:10 in 5 mM Sodium m-Periodate, 2) ethanol precipitated, and 3) mixed overnight with 300 µl adipic acid dihydrazide agarose bead 50% solution (Sigma). To pull down proteins, at least 50 µg of protein lysate was added to each bead/probe sample in binding buffer (3.2 mM $MgCl_2$, 20 mM creatine phosphate, 1 mM ATP, 1.3% polyvinyl alcohol, 25 ng of yeast tRNA, 70 mM KCl, 10 mM Tris, pH 7.5, 0.1 mM EDTA) and mixed for 90 minutes. After four rounds of stringent washing with GFB-200 (20 mM TE, 200 mM KCl)

plus 6 mM $MgCl_2$ and one round of washing with 20 mM Tris-HCl (pH 7.4), probe-bound proteins were digested *in situ* with 6 ng/μl trypsin (Promega) in 100 mM NH_4HCO_3, overnight at 37C. After removing the beads, free digested peptides were lyophilized, extracted with 1%HCOOH/2%CH_3CN, and then extracted several times with 50% CH_3CN. Peptide extracts were lyophilized, desalted with a ZipTip procedure, and resuspended in ~5-10 μl LC buffer A (0.1% HCOOH (v/v) in 5:95 CH_3CN:H_2O). Samples were then analyzed by LC/MS.

A.2.15 Quantitative PCR primers (Chapter 2)

Primers were designed using PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). tRNA primers were designed against tRNA family consensus sequences. Primer sequences are as follows:

AT1G43170 forward: TGGGCACAGCATTTGAGTGA
AT1G43170 reverse: ACTGCTTAGCGTACCCAGTG
AT4G25080 forward: CCCAGGGCCATCAAAAGCTA
AT4G25080 reverse: TCCAGCCGACTTTACCCAAC
AT4G25080 forward (additional primer set): TCGTGGAAGACATGCAGATTC
AT4G25080 reverse (additional primer set): GTTTGTACAGACCGTCCTCCT
AT1G04410 forward: GCTGCAATCATCAAGGCGAG
AT1G04410 reverse: TGGAAACGAACGTACCCCTC
AT1G04410 forward (additional primer set): ACAACAGGGCTTTGGGACAG
AT1G04410 reverse (additional primer set): GACAGGCTTCTCTCCAGACG
AT1G15220 forward: CAACACGAGCCCGAAGAGT
AT1G15220 reverse: AGAAAGTGAACGACTGAGGCT
AT1G28330 forward: GCGGAAGATCAGGTCACCAT
AT1G28330 reverse: TGGGGTGTTTGCAGGTTGTA

AT1G28330 forward (additional primer set): TAAAGACGCTCCTCCACACG
AT1G28330 reverse (additional primer set): GAGCAGCAGTAAGGTGGTGA
AT2G15580 forward: GAGAAACTTGACGGAGCAGC
AT2G15580 reverse: TGTACGTGGTGGGATTCTCAG
AT3G15353 forward: CTGTGCTGACAAGACCCAGT
AT3G15353 reverse: CTCCTGAGTCTCGACGATGT
AT4G08620 forward: CCCGGAATCTTGATCATCC
AT4G08620 reverse: CGGCATGCCATATTCCTTAG
AT3G21170 forward: TGAGGCAGGGTCGTCTTATC
AT3G21170 reverse: CACGCCACTGGTGATATTTG
AT1G66850 forward: GCCATCAAAGCCGAAGACAC
AT1G66850 reverse: ACGCAGGGTTCTTAGCGAAA
AT3G20865 forward: GGAGTCTCCAGCACCATCAC
AT3G20865 reverse: GAAGAGCCAAGAAGGCGGAG
AT5G39420 forward: CAAGGAGATTGGGCGGTTCT
AT5G39420 reverse: CCAACTTCTGGAACGCCTCT
AT4G31070 forward: CTGAAGGGTTTGGTGTCGGA
AT4G31070 reverse: CTGTGAAGCCATTGGTCCCT
tRNA-Arg (anticodon: AGT) forward: CCGCGTGGCCTAATGGATAA
tRNA-Arg (anticodon: AGT) reverse: GATCACGGTGGGACTCGAAC
tRNA-Trp (anticodon: CCA) forward: GATCCGTGGCGCAATGGTAG
tRNA-Trp (anticodon: CCA) reverse: TGAACCCGACGTGAATCGAA
tRNA-ala (anticodon:AGC) forward: GGGGATGTAGCTCAGATGGT
tRNA-ala (anticodon:AGC) reverse: TGGAGATGCGGGGTATCG

A.2.16 Quantitative PCR primers (Chapter 3)

Primers were designed using PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primer sequences are as follows:

AT1G16890 forward: GTGCAGGACTCCACTTGTCT
AT1G16890 forward: TCACTTCATTCGATCCTTCTCCT

AT1G25560 forward: TGGATTCAGAGAACGGCGTC
AT1G25560 reverse: CTCCCCATCTTCCGTTAGGC
AT1G68790 forward: CACAAGCATCAAAGGGTGCC
AT1G68790 reverse: CCTCCTACTTCCAACTGCGG
AT1G80920 forward: CGACGAGGGATTGAACGGAA
AT1G80920 reverse: TAATCACGGGTTCCTGCTCC
AT2G03440 forward: TCGTCACGAGTCAGACAAGC
AT2G03440 reverse: CCTCCAACATTACCGTGGCT
AT2G22500 forward: GCAAGCCGATGGTCGTTTAC
AT2G22500 reverse: CACATGAGTCCCAAGCCCAT
AT3G08550 forward: TGCCTCTGCTTCTGTTCGTT
AT3G08550 reverse: AGTCGGAGGAGGATTGGTGA
AT3G21055 forward: CATTTGCACGAAAATCATATTTGGA
AT3G21055 reverse: GGAGGCTCTGACTACGGAGA
AT3G22790 forward: CACACAGAAGCAAAGACCGC
AT3G22790 reverse: GGATTCTGGTTCAGCAGCCT
AT3G48360 forward: TTGCAAGCGGATGCTTCAAC
AT3G48360 reverse: AAATTGCCTGCAGAGAGGGA
AT3G50820 forward: GATCAAACCAAACCGTGGGC
AT3G50820 reverse: CGGAGCATTTTCCAGCGAAG
AT3G61140 forward: GAAGCGAGTGGACCGATGAT
AT3G61140 reverse: GGTTTCCTCCGCAGTGGTTA
AT3G63510 forward: CCAATGCTCAAACCCACTGC
AT3G63510 reverse: TGTCCAATCCATCATCGGGG
AT4G17530 forward: TCGATCCAGCTCCGAGATCA
AT4G17530 reverse: GCACGACTTTCCAACACCAG
AT4G25150 forward: TCCTCTGGCGTTCTCCAATG
AT4G25150 reverse: GCGAGATTGTTTGTCTCCGC
AT4G26400 forward: CAATCAGGCACGGGAGTCTT
AT4G26400 reverse: AGTTCATATCGGCACACGGG
AT4G27450 forward: ACGTTTTTCCTTCCCACAGGA
AT4G27450 reverse: CAGATGCCGGACTGTTGAGT
AT4G28300 forward: ACGAAGAACCACGCTTTTGC

AT4G28300 reverse: CCGAAGACCTTGCCATCCTT
AT5G02380 forward: CTGTGGTTGTGGATCTGCCT
AT5G02380 reverse: GAGCAACACCGAGGACAAGA
AT5G03730 forward: CCCATGTGGAAGGAGTGCAT
AT5G03730 reverse: CGAAGCGGCATCGTCTCTAT
AT5G20700 forward: ATGACGAGCCCTAAAAGCCC
AT5G20700 reverse: GACCAACAGAGCCACCAAGA
AT5G66400 forward: CACCACGCCGACATTTTCTG
AT5G66400 reverse: TTTCCGTACTCGTCAGTGGC
ATCG00190 forward: GATTGGTGATTGGGGGTCGT
ATCG00190 reverse: TTCTTGGCGGTATCGAGCTG

A.2.17 Quantitative PCR primers (Chapter 4)

Primers were designed using PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primer sequences are as follows:

AT1G05570 (H1) forward: ACTGACAGGCCTGGAACATGAACACCAGTTACATTA
AT1G05570 (H1) reverse: CAGAGACGAGATCCATGGGTTTTGCAAGAACAGCCAATAAG
AT1G35460 (H2) forward: ACTGACAGGCCTTCGAGTAGGGTTTCAATCCAAG
AT1G35460 (H2) reverse: CAGAGACGAGATCCATGGAGGAGGAGGAGAGGTGAGTC
AT1G63020 (H4) forward: ACTGACAGGCCTACCTTTTGCCTTCCACCTAAAG
AT1G63020 (H4) reverse: CAGAGACGAGATCCATGGGAAGCCTCCTTTGCCAGCTT
AT2G07741 (H5) forward: ACTGACAGGCCTTATAGTGGGACAGGCGGCG
AT2G07741 (H5) reverse: CAGAGACGAGATCCATGGGCATAGGAATCAATGGGACAATCT
AT3G44070 (H7) forward: ACTGACAGGCCTGTTGGTGTAAGGGTCAAAGACG
AT3G44070 (H7) reverse: CAGAGACGAGATCCATGGTACCTCGTGTAGTTTGTTTGGT
AT5G27630 (H8) forward: ACTGACAGGCCTTGGTAACATGTCAGGGAATGTC
AT5G27630 (H8) reverse: CAGAGACGAGATCCATGGGGTCTGCGATCTCGATCTGTC
AT4G19690 (H9) forward: ACTGACAGGCCTAACTGCATTTTGATGCTACCTTGA
AT4G19690 (H9) reverse: CAGAGACGAGATCCATGGCATGGCACTCGTGGATCTTC
AT5G38420 (H10) forward: ACTGACAGGCCTCCGGATACTATGATGGACGATACT

AT5G38420 (H10) reverse: CAGAGACGAGATCCATGGTCAACACTTGAGCGGAGTCG
AT5G60120 (H11) forward: ACTGACAGGCCTACGAAGGGCTGAGAAGGATGA
AT5G60120 (H11) reverse: CAGAGACGAGATCCATGGACGAAGGGCTGAGAAGGATGA
ATCG00150 (H12) forward: ACTGACAGGCCTACCATATTCTTCTCCAATCTGGGT
ATCG00150 (H12) reverse: CAGAGACGAGATCCATGGTGACGGCCAAAATTTCTTTGAAT
ATCG00490 (H13) forward: ACTGACAGGCCTACAAAGGACGATGCTACCACA
ATCG00490 (H13) reverse: CAGAGACGAGATCCATGGACGCAATAAATTGAGTTTCTTCTCC
ATCG00740 (H14) forward: ACTGACAGGCCTGGTATGTTCTCAGATTTTGCACG
ATCG00740 (H14) reverse: CAGAGACGAGATCCATGGTAGGCATTGCGATGCGAAGA

A.3 COMPUTATIONAL, STATISTICAL, AND ANALYTICAL METHODS

A.3.1 Statistical analyses

All statistical analyses were performed using the R software package (http://www.r-project.org/), including p-values for all hypothesis testing. See results sections and figure legends from each chapter for specific statistical tests used to assess significance.

A.3.2 Genome annotations

All analyses in *Arabidopsis* were performed using the TAIR10 genome assembly. All analyses in humans were performed using the UCSC hg19 RefSeq assembly. All analyses in mice were performed using the UCSC mm10 RefSeq assembly. Alternative and constitutive introns in *Arabidopsis* were identified using the TAIR10 transcriptome annotation, as well as the AtRTD alternate transcriptome annotation

(https://ics.hutton.ac.uk/atRTD/) (Zhang et al., 2015). Repeat-subtracted genomes (repeat-masked) for TAIR10 were produced with the RepeatMasker package (Smit, AFA, Hubley, R & Green, P. (2013); RepeatMasker Open-4.0, http://www.repeatmasker.org).

A.3.3 mRNA read processing and alignment

Read processing and alignment were performed as previously described (Li et al., 2012a) with slight modifications. Briefly, sequencing reads were first trimmed to remove 3' sequencing adapters with Cutadapt (version 1.2.1 with parameters -e 0.06 -O 6 -m 14). 50 base pair single-end reads were aligned to their respective genome using Tophat (version 2.0.10 with parameters --library-type fr-secondstrand --read-mismatches 2 --read-edit-dist 2 --max-multihits 10 --b2-very-sensitive --transcriptome-max-hits 10 --no-coveragesearch --no-novel-juncs). Longer 125 base pair paired-end reads were trimmed in a similar manner, but allowing for more mismatches (--read-mismatches 8 --read-edit-dist 8).

A.3.4 tRNA read processing and alignment (Arabidopsis smRNAs)

tRNA amino acid-anticodon families were annotated with tRNAscan (Lowe and Eddy, 1997). For each amino acid-anticodon family of tRNAs, a consensus sequence was constructed through multiple alignment of all loci with LocARNA (Will et al., 2007) and selection of the most abundant nucleotide at each aligned position. Any consensus nucleotides with biallelic SNPs were retained since HAMR will filter these in hypothesis

testing, while a few rare triallelic SNPs were excluded since these could potentially lead to HAMR artifacts. smRNA reads were first aligned to the *Arabidopsis thaliana* genome version TAIR10, allowing multimappers. Reads that mapped exclusively to tRNAs were retained. This subset of reads was then remapped to the tRNA consensus sequence set. Downstream analyses were performed using consensus coordinates, as described previously (Ryvkin et al., 2013).

A.3.5 High Throughput Annotation of Modified Ribonucleotides (HAMR)

HAMR was performed as previously described (Ryvkin et al., 2013). For each set of mapped reads, deviations from the reference sequence (mismatches) with a quality score greater than 30 (error rate < 0.001) are tabulated for each base in either the *Arabidopsis thaliana* genome version TAIR10, human genome version hg19, or TAIR10 tRNA consensus sequence set. Each base with mismatches was tested for significant enrichment of mismatches using a binomial distribution, with the conservative assumption that the sequencing error rate is 0.01. Bases that pass this filter are then tested against the null hypothesis that the genotype is biallelic. Each possible biallelic genotype is tested, again using a binomial distribution. Significant deviation from all possible biallelic genotypes is used as evidence of modification, as modification-induced errors should be semi-random and not have a clear bias toward any single base substitution, as would be true with SNPs or RNA editing (Ryvkin et al., 2013). Each predicted modified base was then classified using nearest-neighbor machine learning, as described previously (Ryvkin et al., 2013). Known tRNA modifications in

Saccharomyces cerevisiae (from the MODOMICS database) (Dunin-Horkawicz et al., 2006) were used previously (Ryvkin et al., 2013) to construct the training set.

A.3.6 Definition of HAMR-accessible bases and transcripts

In *Arabidopsis,* the minimum base coverage at an observed modification in either GMUCT, smRNA-seq, or RNA-seq was always 50 reads per base (50x). Thus, any base with at least 50x coverage was designated as HAMR-accessible. For comparison, the minimum coverage for humans, though not included in any analyses, was 10x. The minimum number of uniquely mapping reads to call a transcript as modified was 100 for *Arabidopsis* and 10 for humans. Thus, transcripts with at least 100 or 10 uniquely mapping reads were designated as HAMR-accessible in *Arabidopsis* and humans, respectively.

A.3.7 Predicting unmodified genes (Chapter 3)

Unmodified genes were defined as those with minimal mismatches (less than 10), despite high overall read coverage. We also required that unmodified genes have stable steady-state RNA abundance across control and salt stress-treated plants. Unmodified genes used in **Chapter 3** consist of AT4G25150, AT4G28300, AT3G08550, and AT3G63510.

A.3.8 Ribosome pause sites and occupancy

Ribo-seq reads were only retained for analysis when their size was within 26 to 34 nucleotides, close to the expected size of monoribosomal footprints (~28nt). Ribosome occupancy was calculated, per gene, by normalizing the abundance of ribo-seq reads to capped polyadenylated RNA abundance. Ribosome pause sites were determined from runs of nucleotides with ribo-seq coverage at least 25-fold higher than the transcript median. Pause sites within 50 nucleotide proximity were merged to form single peaks.

A.3.9 Gene Ontology enrichment

We tested for enriched of Gene Ontology terms using the DAVID online tool (Huang et al., 2009) and plotted results in heatmap form as previously described (Vandivier et al., 2013, 2015b). The background set was defined as all detectable transcripts (those with at least one mapped read) for differential expression analyses and all HAMR-accessible transcripts for HAMR analyses.

A.3.10 Structure scores

We defined numeric secondary structure scores (S_i) by calculating a log-odds ratio of base-pairing probability. For each position in the genome with at least one ssRNA-seq or dsRNA-seq read, we applied a generalized (zero-tolerant pseudocount) log ratio of normalized dsRNA-seq coverage (ds_i) over normalized ssRNA-seq coverage

(ss_i). Raw coverage (rds_i and rss_i) is normalized to a ratio of the number of mapped reads in each library (N_{ds} and N_{ss}).

$$S_i = glog(ds_i) - glog(ss_i) = log_2\left(ds_i + \sqrt{1 + ds_i^2}\right) - log_2\left(ss_i + \sqrt{1 + ss_i^2}\right)$$

$$ds_i = rds_i \cdot \frac{N_{ds}}{N_{ss}} \; ; \; ss_i = rss_i \cdot \frac{N_{ss}}{N_{ds}}$$

A.3.11 Structure peaks

We called secondary structure “peaks” and “valleys” using the ChIP-seq Analysis in R (CSAR) program (**Figure 1E**). We begin by using shuffled coverage to call peaks. We then use the Poisson distribution to determine peak scores. We define an empirical 5% false discovery rate (FDR) threshold based upon the peak score threshold above which lies the top 5% of shuffled peaks. To call structure hotspots, we use dsRNA-seq coverage as a signal and ssRNA-seq coverage as a control (analogous to an antibody control in ChIP-seq). To call structure valleys, we use ssRNA-seq coverage as the signal, and dsRNA-seq coverage as a control. We only retain hotspots and valleys with scores greater than the 5% FDR threshold.

A.3.12 Constrained prediction of RNA folding

To define high-confidence paired and unpaired bases, we generated a background distribution of coverage and structure scores by randomly permuting reads between dsRNA-seq and ssRNA-seq libraries. We computed structure scores as

described in **Appendix A.3.10**, and then determined the 97.5th and 2.5th percentiles. Bases with scores from real data above the 97.5th shuffled percentile were called as high-confidence paired nucleotides. Bases with scores from real data below the 2.5th shuffled percentile were called as high-confidence unpaired nucleotides. We then used these high-confidence bases to constrain the RNA folding algorithm RNAFold (Zuker and Stiegler, 1981) such that high-confidence paired nucleotides must be double-stranded and high-confidence unpaired bases must be single-stranded.

A.3.13 Differential expression analysis

Reads within exons of any isoform of a given gene were tabulated with HTseq-count (mode=union) (http://www-huber.embl.de/HTSeq/). Differential expression was performed with the EdgeR package (https://bioconductor.org/packages/release/bioc/html/edgeR.html), using internal library size normalization methods except for when spike-in control data was available, in which case normalization factors were set based upon number of reads mapping to spike-ins.

A.3.14 Motif analysis

All motifs were generated via the MEME suite (Bailey et al., 2009), with parameters as follows: -nmotifs 100 -maxw 20 -evt 0.01 -maxsize 10000000.

A.3.15 Mass spectral data analyses and protein identification

Experimentally collected MS/MS tandem data were searched against the latest version of the mouse NCBI proteome database from NCBI using Thermo Proteome Discoverer 1.4 software. We required full trypsin digestion with at most 3 missed cleavages, and allowed for potential modifications to methionine (oxidation) and cysteine (carbamidomethylation). All other parameters were standard for LCQ Deca XP+ instrumentation. Peptide filters were Xcorr = (1.5, 2.0, 2.25, 2.5) for charges (+1, +2, +3, +4), respectively.

A.3.16 Nibbled transcript identification

To identify nibbled transcripts, we define three sets of windows. Set 5 was centered at the 5' terminus (5^{th} percentile of transcript length), set M at the middle (50^{th} percentile of transcript length), and set 3 at the 3' terminus (95^{th} percentile of transcript length). Each set was composed of three smaller, 20nt windows from -50 to -30, -10 to +10, and +30 to +50 nucleotides away from the center. For each timepoint, we computed average RNA-seq coverage at each these windows, and took the mean for each set. If windows lay outside the mRNA of interest (for instance, in transcripts less than 1000nt in length), they were ignored. We then calculated the ratio of mean coverage, for each set, between a given later timepoint and time 0hr. We then calculated a ratio of ratios between set 3 and set M, and set 5 and set M to define 3' and 5' nibbling, respectively. When this ratio was met or exceeded 10, we called a transcript as nibbled. We also filtered for excessively skewed distributions of read coverage by

ignoring transcripts for which set 3 mean coverage was over 5x that of the mean coverage for the 5'-most three length deciles (5' terminus to 30th percentile), or for which set 5 mean coverage was over 5x that of the mean coverage for the 3'-most three length deciles (70th percentile to 3' terminus).

After identifying nibbled transcripts, we defined the exact boundary between nibbled and unnibbled transcripts by progressively sliding set 3 or set 5 windows toward the center until their ratio converged to within 1.5x of the ratio of set M.

A.4 ACCESSIONS AND REPOSITORIES

A.4.1 Chapter 2 previously published Datasets

Human RNA-seq data for HeLa cells were downloaded from the ENCODE Caltech RNA-seq compendium (GEO accession GSM958739) (ENCODE Project Consortium, 2012). Human RNA-seq data for HEK293T cells were downloaded from GEO accession GSE34995 (Huelga et al., 2012). Human GMUCT data were downloaded from GEO accession GSE47121 (Willmann et al., 2014). Additional plant smRNA-seq data were downloaded from GEO accession GSE57215 (Li et al., 2014). Reverse transcriptase stalling data (Structure-seq) were downloaded from SRA accession SRP027216 (Ding et al., 2014).

A.4.2 Chapter 2 accession numbers

All smRNA-seq, RNA-seq, and GMUCT data generated for this study were deposited in the Gene Expression Omnibus (GEO) under accession number GSE66224. Additionally, HAMR-predicted modifications are available under the same GEO accession or at http://gregorylab.bio.upenn.edu/HAMR_degradome/.

A.4.3 Chapter 3 previously published datasets

Abh1-1 GMUCT libraries were downloaded from GEO accession GSE71913 (Yu et al., 2016).

A.4.4 HAMR Software

The latest version of the HAMR pipeline is available from https://github.com/GregoryLab/HAMR.

REFERENCES

Abbasi-Moheb, L., Mertel, S., Gonsior, M., Nouri-Vahid, L., Kahrizi, K., Cirak, S., Wieczorek, D., Motazacker, M.M., Esmaeeli-Nieh, S., Cremer, K., et al. (2012). Mutations in NSUN2 Cause Autosomal- Recessive Intellectual Disability. Am. J. Hum. Genet. *90*, 847–855.

Agarwala, S.D., Blitzblau, H.G., Hochwagen, A., and Fink, G.R. (2012). RNA Methylation by the MIS Complex Regulates a Cell Fate Decision in Yeast. PLOS Genet. *8*, e1002732.

Alarcón, C.R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S.F. (2015). N6-methyladenosine marks primary microRNAs for processing. Nature *519*, 482–485.

Alizadeh, Z., Kageyama, S.-I., and Aoki, F. (2005). Degradation of maternal mRNA in mouse embryos: Selective degradation of specific mRNAs after fertilization. Mol. Reprod. Dev. *72*, 281–290.

Arnez, J.G., and Steitz, T.A. (1994). Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. Biochemistry (Mosc.) *33*, 7560–7567.

Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013). Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing. Cell *152*, 844–858.

Backe, P.H., Messias, A.C., Ravelli, R.B.G., Sattler, M., and Cusack, S. (2005). X-Ray Crystallographic and NMR Studies of the Third KH Domain of hnRNP K in Complex with Single-Stranded Nucleic Acids. Structure *13*, 1055–1067.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208.

Bakin, A., and Ofengand, J. (1993). Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: Analysis by the application of a new sequencing technique. Biochemistry (Mosc.) *32*, 9754–9762.

Balatsos, N.A.A., Nilsson, P., Mazza, C., Cusack, S., and Virtanen, A. (2006). Inhibition of mRNA Deadenylation by the Nuclear Cap Binding Complex (CBC). J. Biol. Chem. *281*, 4517–4522.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., et al. (2014). m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells. Cell Stem Cell *15*, 707–719.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. Science *336*, 233–237.

Beemon, K., and Keith, J. (1977). Localization of N6-methyladenosine in the Rous sarcoma virus genome. J. Mol. Biol. *113*, 165–179.

Behm-Ansmant, I., Urban, A., Ma, X., Yu, Y.-T., Motorin, Y., and Branlant, C. (2003). The Saccharomyces cerevisiae U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. RNA *9*, 1371–1382.

Bhaskaran, H., Rodriguez-Hernandez, A., and Perona, J.J. (2012). Kinetics of tRNA folding monitored by aminoacylation. RNA *18*, 569–580.

Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J., and Nielsen, H. (2015). Profiling of Ribose Methylations in RNA by High-Throughput Sequencing. Angew. Chem. *127*, 461–465.

Bocobza, S.E., and Aharoni, A. (2014). Small molecules that interact with RNA: riboswitch-based gene control and its involvement in metabolic regulation in plants and algae. Plant J. *79*, 693–703.

Bodi, Z., Button, J.D., Grierson, D., and Fray, R.G. (2010). Yeast targets for mRNA methylation. Nucleic Acids Res. *38*, 5327-5335

Bokar, J.A., Rath-Shambaugh, M.E., Ludwiczak, R., Narayan, P., and Rottman, F. (1994). Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. J. Biol. Chem. *269*, 17697–17704.

Bokar, J.A., Shambaugh, M.E., Polayes, D., Matera, A.G., and Rottman, F.M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA *3*, 1233–1247.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.-K. (2005). Endogenous siRNAs Derived from a Pair of Natural cis-Antisense Transcripts Regulate Salt Tolerance in Arabidopsis. Cell *123*, 1279–1291.

Braddock, D.T., Louis, J.M., Baber, J.L., Levens, D., and Clore, G.M. (2002). Structure and dynamics of KH domains from FBP bound to single-stranded DNA. Nature *415*, 1051–1056.

Brook, M., and Gray, N.K. (2012). The role of mammalian poly(A)-binding proteins in co-ordinating mRNA turnover. Biochem. Soc. Trans. *40*, 856–864.

Brower, P.T., Gizang, E., Boreen, S.M., and Schultz, R.M. (1981). Biochemical studies of mammalian oogenesis: Synthesis and stability of various classes of RNA during growth of the mouse oocyte in vitro. Dev. Biol. *86*, 373–383.

Brownlee, G.G., and Cartwright, E.M. (1977). Rapid gel sequencing of RNA by primed synthesis with reverse transcriptase. J. Mol. Biol. *114*, 93–117.

Brzezicha, B., Schmidt, M., Makałowska, I., Jarmołowski, A., Pieńkowska, J., and Szweykowska-Kulińska, Z. (2006). Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the. Nucleic Acids Res. *34*, 6034–6043.

Bullock, S.L., Ringel, I., Ish-Horowicz, D., and Lukavsky, P.J. (2010). A'-form RNA helices are required for cytoplasmic mRNA transport in Drosophila. Nat. Struct. Mol. Biol. *17*, 703–709.

Buratti, E., and Baralle, F.E. (2004). Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. Mol. Cell. Biol. *24*, 10505–10514.

Burroughs, A.M., Ando, Y., de Hoon, M.L., Tomaru, Y., Suzuki, H., Hayashizaki, Y., and Daub, C.O. (2011). Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. RNA Biol. *8*, 158–177.

Canaani, D., Kahana, C., Lavi, S., and Groner, Y. (1979). Identification and mapping of N6-methyladenosine containing sequences in Simian Virus 40 RNA. Nucleic Acids Res. *6*, 2879–2899.

Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A.P., Fabris, D., and Agris, P.F. (2011). The RNA modification database, RNAMDB: 2011 update. Nucleic Acids Res. *39*, D195–D201.

Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M., and Gilbert, W.V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature *515*, 143–146.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642–655.

Castleberry, C.M., Chou, C.-W., and Limbach, P.A. (2001). Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry of Oligonucleotides. In Current Protocols in Nucleic Acid Chemistry (John Wiley & Sons, Inc.). 33:10.1.1-10.1.21.

Chapman, E.J., and Carrington, J.C. (2007). Specialization and evolution of endogenous small RNA pathways. Nat. Rev. Genet. *8*, 884–896.

Chekanova, J.A., Gregory, B.D., Reverdatto, S.V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., et al. (2007). Genome-Wide High-Resolution Mapping of Exosome Substrates Reveals Hidden Features in the Arabidopsis Transcriptome. Cell *131*, 1340–1353.

Chen, C.-Y.A., and Shyu, A.-B. (1995). AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem. Sci. *20*, 465–470.

Chen, K., Zhao, B.S., and He, C. (2016). Nucleic Acid Modifications in Regulation of Gene Expression. Cell Chem. Biol. *23*, 74–85.

Choi, J., Ieong, K.-W., Demirci, H., Chen, J., Petrov, A., Prabhakar, A., O'Leary, S.E., Dominissini, D., Rechavi, G., Soltis, S.M., et al. (2016). N6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. Nat. Struct. Mol. Biol. *23*, 110–115.

Chujo, T., and Suzuki, T. (2012). Trmt61B is a methyltransferase responsible for 1-methyladenosine at position 58 of human mitochondrial tRNAs. RNA *18*, 2269–2276.

Cleary, M.D., Meiering, C.D., Jan, E., Guymon, R., and Boothroyd, J.C. (2005). Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nat. Biotechnol. *23*, 232–237.

Clift, D., and Schuh, M. (2013). Restarting life: fertilization and the transition from meiosis to mitosis. Nat. Rev. Mol. Cell Biol. *14*, 549–562.

Courtes, F.C., Gu, C., Wong, N.S.C., Dedon, P.C., Yap, M.G.S., and Lee, D.-Y. (2014). 28S rRNA is inducibly pseudouridylated by the mTOR pathway translational control in CHO cell cultures. J. Biotechnol. *174*, 16–21.

Cruz, J.A., and Westhof, E. (2009). The Dynamic Landscapes of RNA Architecture. Cell *136*, 604–609.

Csepany, T., Lin, A., Baldick, C.J., and Beemon, K. (1990). Sequence specificity of mRNA N6-adenosine methyltransferase. J. Biol. Chem. *265*, 20117–20122.

Dai, X., and Zhao, P.X. (2011). psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. *39*, W155-159.

Davis, F.F., and Allen, F.W. (1957). Ribonucleic acids from yeast which contain a fifth nucleotide. J. Biol. Chem. *227*, 907–915.

Decatur, W.A., and Schnare, M.N. (2008). Different mechanisms for pseudouridine formation in yeast 5S and 5.8S rRNAs. Mol. Cell. Biol. *28*, 3089–3100.

Delatte, B., Wang, F., Ngoc, L.V., Collignon, E., Bonvin, E., Deplus, R., Calonne, E., Hassabi, B., Putmans, P., Awe, S., et al. (2016). Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science *351*, 282–285.

Demeshkina, N., Jenner, L., Yusupova, G., and Yusupov, M. (2010). Interactions of the ribosome with mRNA and tRNA. Curr. Opin. Struct. Biol. *20*, 325–332.

Deng, X., Chen, K., Luo, G.-Z., Weng, X., Ji, Q., Zhou, T., and He, C. (2015). Widespread occurrence of N6-methyladenosine in bacterial mRNA. Nucleic Acids Res. *43*, 6557–6567.

Desrosiers, R., Friderici, K., and Rottman, F. (1974). Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells. Proc. Natl. Acad. Sci. *71*, 3971–3975.

Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R., and Xu, R.-M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev. *13*, 1102–1115.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696–700.

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

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

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N., and Rechavi, G. (2013). Transcriptome-wide mapping of N6-methyladenosine by m6A-seq based on immunocapturing and massively parallel sequencing. Nat. Protoc. *8*, 176–189.

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M.S., Dai, Q., Di Segni, A., Salmon-Divon, M., Clark, W.C., et al. (2016). The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. Nature *530*, 441–446.

Dong, H., Ray, D., Ren, S., Zhang, B., Puig-Basagoiti, F., Takagi, Y., Ho, C.K., Li, H., and Shi, P.-Y. (2007). Distinct RNA Elements Confer Specificity to Flavivirus RNA Cap Methylation Events. J. Virol. *81*, 4412–4421.

Dong, Z.-W., Shao, P., Diao, L.-T., Zhou, H., Yu, C.-H., and Qu, L.-H. (2012). RTL-P: a sensitive approach for detecting sites of 2′-O-methylation in RNA molecules. Nucleic Acids Res. *40*, e157–e157.

Draper, D.E. (2004). A guide to ions and RNA structure. RNA *10*, 335–343.

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., and Wu, L. (2016). YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. Nat. Commun. *7*, 12626.

Dubin, D.T., and Taylor, R.H. (1975). The methylation state of poly A-containing-messenger RNA from cultured hamster cells. Nucleic Acids Res. *2*, 1653–1668.

Dubin, D.T., Stollar, V., Hsuchen, C.-C., Timko, K., and Guild, G.M. (1977). Sindbis virus messenger RNA: the 5′-termini and methylated residues of 26 and 42 S RNA. Virology *77*, 457–470.

Dumelin, C.E., Chen, Y., Leconte, A.M., Chen, Y.G., and Liu, D.R. (2012). Discovery and biological characterization of geranylated RNA in bacteria. Nat. Chem. Biol. *8*, 913–919.

Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M.J., Feder, M., Grosjean, H., and Bujnicki, J.M. (2006). MODOMICS: a database of RNA modification pathways. Nucleic Acids Res. *34*, D145–D149.

Ebhardt, H.A., Tsang, H.H., Dai, D.C., Liu, Y., Bostan, B., and Fahlman, R.P. (2009). Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. Nucleic Acids Res. *37*, 2461–2470.

Edmonds, C.G., Crain, P.F., Gupta, R., Hashizume, T., Hocart, C.H., Kowalak, J.A., Pomerantz, S.C., Stetter, K.O., and McCloskey, J.A. (1991). Posttranscriptional modification of tRNA in thermophilic archaea (Archaebacteria). J. Bacteriol. *173*, 3138–3148.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.-P., and Ehresmann, B. (1987). Probing the structure of RNAs in solution. Nucleic Acids Res. *15*, 9109–9128.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Fernández, I.S., Ng, C.L., Kelley, A.C., Wu, G., Yu, Y.-T., and Ramakrishnan, V. (2013). Unusual base pairing during the decoding of a stop codon by the ribosome. Nature *500*, 107–110.

Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J.A. (2013). RNA catalyses nuclear pre-mRNA splicing. Nature *503*, 229–234.

Foley, S.W., Vandivier, L.E., Kuksa, P.P., and Gregory, B.D. (2015). Transcriptome-wide measurement of plant RNA secondary structure. Curr. Opin. Plant Biol. *27*, 36–43.

Foley, S.W., Gosai, S.J., Wang, D., Selamoglu, N., Sollitti, A.C., Köster, T., Steffen, A., Lyons, E., Daldal, F., Garcia, B.A., et al. (2017). A Global View of RNA-Protein Interactions Identifies Post-transcriptional Regulators of Root Hair Cell Fate. Dev. Cell *41*, 204–220.e5.

Francklyn, C.S., and Minajigi, A. (2010). tRNA as an active chemical scaffold for diverse chemical transformations. FEBS Lett. *584*, 366–375.

Fu, L., Guerrero, C.R., Zhong, N., Amato, N.J., Liu, Y., Liu, S., Cai, Q., Ji, D., Jin, S.-G., Niedernhofer, L.J., et al. (2014a). Tet-Mediated Formation of 5-Hydroxymethylcytosine in RNA. J. Am. Chem. Soc. *136*, 11582–11585.

Fu, Y., Jia, G., Pang, X., Wang, R.N., Wang, X., Li, C.J., Smemo, S., Dai, Q., Bailey, K.A., Nobrega, M.A., et al. (2013). FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. Nat. Commun. *4*, 1798.

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014b). Gene expression regulation mediated through reversible m6A RNA methylation. Nat. Rev. Genet. *15*, 293–306.

Furuichi, Y., Shatkin, A.J., Stavnezer, E., and Bishop, J.M. (1975). Blocked, methylated 5′-terminal sequence in avian sarcoma virus RNA. Nature *257*, 618–620.

Fustin, J.-M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., Isagawa, T., Morioka, M.S., Kakeya, H., Manabe, I., et al. (2013). RNA-Methylation-Dependent RNA Processing Controls the Speed of the Circadian Clock. Cell *155*, 793–806.

G R Bjork, J U Ericson, C E D Gustafsson, T G Hagervall, Y H Jonsson, and Wikstrom, P.M. (1987). Transfer RNA Modification. Annu. Rev. Biochem. *56*, 263–285.

Ganot, P., Bortolin, M.-L., and Kiss, T. (1997). Site-Specific Pseudouridine Formation in Preribosomal RNA Is Guided by Small Nucleolar RNAs. Cell *89*, 799–809.

Gao, M., Fritz, D.T., Ford, L.P., and Wilusz, J. (2000). Interaction between a Poly(A)-Specific Ribonuclease and the 5′ Cap Influences mRNA Deadenylation Rates In Vitro. Mol. Cell *5*, 479–488.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nat. Rev. Mol. Cell Biol. *8*, 113–126.

Gasciolli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. (2005). Partially Redundant Functions of Arabidopsis DICER-like Enzymes and a Role for DCL4 in Producing trans-Acting siRNAs. Curr. Biol. *15*, 1494–1500.

Gaston, K.W., and Limbach, P.A. (2014). The identification and characterization of non-coding and coding RNAs and their modified nucleosides by mass spectrometry. RNA Biol. *11*, 1568–1585.

Gazzani, S., Lawrenson, T., Woodward, C., Headon, D., and Sablowski, R. (2004a). A Link Between mRNA Turnover and RNA Interference in Arabidopsis. Science *306*, 1046–1048.

Gazzani, S., Lawrenson, T., Woodward, C., Headon, D., and Sablowski, R. (2004b). A link between mRNA turnover and RNA interference in Arabidopsis. Science *306*, 1046–1048.

Ge, J., and Yu, Y.-T. (2013). RNA pseudouridylation: new insights into an old modification. Trends Biochem. Sci. *38*, 210–218.

German, M.A., Pillay, M., Jeong, D.-H., Hetawal, A., Luo, S., Janardhanan, P., Kannan, V., Rymarquis, L.A., Nobuta, K., German, R., et al. (2008). Global identification of microRNA–target RNA pairs by parallel analysis of RNA ends. Nat. Biotechnol. *26*, 941–946.

German, M.A., Luo, S., Schroth, G., Meyers, B.C., and Green, P.J. (2009). Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nat. Protoc. *4*, 356–362.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S., et al. (2015). m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science *347*, 1002–1006.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. *582*, 1977–1986.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA Methyltransferase Homolog Dnmt2. Science *311*, 395–398.

Goodarzi, H., Najafabadi, H.S., Oikonomou, P., Greco, T.M., Fish, L., Salavati, R., Cristea, I.M., and Tavazoie, S. (2012). Systematic discovery of structural elements governing stability of mammalian messenger RNAs. Nature *485*, 264–268.

Gosai, S.J., Foley, S.W., Wang, D., Silverman, I.M., Selamoglu, N., Nelson, A.D.L., Beilstein, M.A., Daldal, F., Deal, R.B., and Gregory, B.D. (2015). Global Analysis of the RNA-Protein Interaction and RNA Secondary Structure Landscapes of the Arabidopsis Nucleus. Mol Cell *57*, 376–388.

Gregory, B.D., O'Malley, R.C., Lister, R., Urich, M.A., Tonti-Filippini, J., Chen, H., Millar, A.H., and Ecker, J.R. (2008). A Link between RNA Metabolism and Silencing Affecting Arabidopsis Development. Dev. Cell *14*, 854–866.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193–1197.

Grosjean, H., Szweykowska-Kulinska, Z., Motorin, Y., Fasiolo, F., and Simos, G. (1997). Intron-dependent enzymatic formation of modified nucleosides in eukaryotic tRNAs: A review. Biochimie *79*, 293–302.

Grüter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K., and Izaurralde, E. (1998). TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. Mol. Cell *1*, 649–659.

Gupta, R.C., and Randerath, K. (1979). Rapid print-readout technique for sequencing of RNA's containing modified nucleotides. Nucleic Acids Res. *6*, 3443–3458.

Hamatani, T., Carter, M.G., Sharov, A.A., and Ko, M.S.H. (2004). Dynamics of Global Gene Expression Changes during Mouse Preimplantation Development. Dev. Cell *6*, 117–131.

Hammann, C., Norman, D.G., and Lilley, D.M.J. (2001). Dissection of the ion-induced folding of the hammerhead ribozyme using 19F NMR. Proc. Natl. Acad. Sci. *98*, 5503–5508.

Haugland, R.A., and Cline, M.G. (1980). Post-transcriptional Modifications of Oat Coleoptile Ribonucleic Acids. Eur. J. Biochem. *104*, 271–277.

He, W., and Parker, R. (2001). The Yeast Cytoplasmic LsmI/Pat1p Complex Protects mRNA 3′ Termini From Partial Degradation. Genetics *158*, 1445–1455.

Helm, M., Giegé, R., and Florentz, C. (1999). A Watson−Crick Base-Pair-Disrupting Methyl Group (m1A9) Is Sufficient for Cloverleaf Folding of Human Mitochondrial tRNALys. Biochemistry (Mosc.) *38*, 13338–13346.

Hentze, M.W., Caughman, S.W., Rouault, T.A., Barriocanal, J.G., Dancis, A., Harford, J.B., and Klausner, R.D. (1987). Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. Science *238*, 1570–1573.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res. *31*, 3429–3431.

Hongay, C.F., and Orr-Weaver, T.L. (2011). Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. Proc. Natl. Acad. Sci. *108*, 14855–14860.

Hoogstraten, C.G., Legault, P., and Pardi, A. (1998). NMR solution structure of the lead-dependent ribozyme: evidence for dynamics in RNA catalysis11Edited by I. Tinoco. J. Mol. Biol. *284*, 337–350.

Hopper, A.K., and Phizicky, E.M. (2003). tRNA transfers to the limelight. Genes Dev. *17*, 162–180.

Horowitz, S., Horowitz, A., Nilsen, T.W., Munns, T.W., and Rottman, F.M. (1984). Mapping of N6-methyladenosine residues in bovine prolactin mRNA. Proc. Natl. Acad. Sci. *81*, 5667–5671.

Howell, M.D., Fahlgren, N., Chapman, E.J., Cumbie, J.S., Sullivan, C.M., Givan, S.A., Kasschau, K.D., and Carrington, J.C. (2007). Genome-Wide Analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 Pathway in Arabidopsis Reveals Dependency on miRNA- and tasiRNA-Directed Targeting. Plant Cell Online *19*, 926–942.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., et al. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep. *1*, 167–178.

Hughes, J.M., and Ares, M. (1991). Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. EMBO J. *10*, 4231–4239.

Hugouvieux, V., Kwak, J.M., and Schroeder, J.I. (2001). An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in Arabidopsis. Cell *106*, 477–487.

Hussain, S., Aleksic, J., Blanco, S., Dietmann, S., and Frye, M. (2013a). Characterizing 5-methylcytosine in the mammalian epitranscriptome. Genome Biol. *14*, 215.

Hussain, S., Tuorto, F., Menon, S., Blanco, S., Cox, C., Flores, J.V., Watt, S., Kudo, N.R., Lyko, F., and Frye, M. (2013b). The Mouse Cytosine-5 RNA Methyltransferase NSun2 Is a Component of the Chromatoid Body and Required for Testis Differentiation. Mol. Cell. Biol. *33*, 1561–1570.

Hussain, S., Sajini, A.A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., Paramor, M., Gleeson, J.G., Odom, D.T., Ule, J., et al. (2013c). NSun2-Mediated Cytosine-5 Methylation of Vault Noncoding RNA Determines Its Processing into Regulatory Small RNAs. Cell Rep. *4*, 255–261.

Hussain, S., Aleksic, J., Blanco, S., Dietmann, S., and Frye, M. (2013d). Characterizing 5-methylcytosine in the mammalian epitranscriptome. Genome Biol. *14*, 215.

Jahn, C.L., Baran, M.M., and Bachvarova, R. (1976). Stability of RNA synthesized by the mouse oocyte during its major growth phase. J. Exp. Zool. *197*, 161–171.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.-G., et al. (2011). N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol. *7*, 885–887.

Jia, G., Fu, Y., and He, C. (2013). Reversible RNA adenosine methylation in biological regulation. Trends Genet. *29*, 108–115.

Jin, Y., Yang, Y., and Zhang, P. (2011). New insights into RNA secondary structure in the alternative splicing of pre-mRNAs. RNA Biol. *8*, 450–457.

Jorenush, M.H., and Sepaskhah, A.R. (2003). Modelling capillary rise and soil salinity for shallow saline water table under irrigated and non-irrigated conditions. Agric. Water Manag. *61*, 125–141.

Kalyna, M., Simpson, C.G., Syed, N.H., Lewandowska, D., Marquez, Y., Kusenda, B., Marshall, J., Fuller, J., Cardle, L., McNicol, J., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res. *40*, 2454–2469.

Kane, S.E., and Beemon, K. (1985). Precise localization of m6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. Mol. Cell. Biol. *5*, 2298–2306.

Kang, H., Park, S.J., and Kwak, K.J. (2013). Plant RNA chaperones in stress response. Trends Plant Sci. *18*, 100–106.

Karijolich, J., and Yu, Y.-T. (2011). Converting nonsense codons into sense codons by targeted pseudouridylation. Nature *474*, 395–398.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Zhu, J.-K., Staskawicz, B.J., and Jin, H. (2006). A pathogen-inducible endogenous siRNA in plant immunity. Proc. Natl. Acad. Sci. *103*, 18002–18007.

Kennedy, T.D., and Lane, B.G. (1979). Wheat embryo ribonucleates. XIII. Methyl-substituted nucleoside constituents and 5'-terminal dinucleotide sequences in bulk poly(AR)-rich RNA from imbibing wheat embryos. Can. J. Biochem. *57*, 927–931.

Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y., and Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. Nature *467*, 103–107.

Khoddami, V., and Cairns, B.R. (2013). Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nat. Biotechnol. *31*, 458–464.

Kierzek, E., Malgowska, M., Lisowiec, J., Turner, D.H., Gdaniec, Z., and Kierzek, R. (2014). The contribution of pseudouridine to stabilities and structure of RNAs. Nucleic Acids Res. *42*, 3492–3501.

Kim, S.-H., and Rich, A. (1968). Single Crystals of Transfer RNA: An X-Ray Diffraction Study. Science *162*, 1381–1384.

Kim, S.H., Suddath, F.L., Quigley, G.J., McPherson, A., Sussman, J.L., Wang, A.H., Seeman, N.C., and Rich, A. (1974). Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science *185*, 435–440.

Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat. Methods *8*, 559–564.

Kiss, T. (2002). Small Nucleolar RNAs: An Abundant Group of Noncoding RNAs with Diverse Cellular Functions. Cell *109*, 145–148.

Kiss-László, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M., and Kiss, T. (1996). Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs. Cell *85*, 1077–1088.

Klasens, B.I., Das, A.T., and Berkhout, B. (1998). Inhibition of polyadenylation by stable RNA secondary structure. Nucleic Acids Res. *26*, 1870–1876.

Korostelev, A., and Noller, H.F. (2007). The ribosome in focus: new structures bring new insights. Trends Biochem. Sci. *32*, 434–441.

Kozak, M. (1988). Leader length and secondary structure modulate mRNA function under conditions of stress. Mol. Cell. Biol. *8*, 2737–2744.

Kurihara, Y., and Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc. Natl. Acad. Sci. *101*, 12753–12758.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. *82*, 6955–6959.

Latham, K.E., Solter, D., and Schultz, R.M. (1992). Acquisition of a transcriptionally permissive state during the 1-cell stage of mouse embryogenesis. Dev. Biol. *149*, 457–462.

Lee, M., Kim, B., and Kim, V.N. (2014). Emerging Roles of RNA Modification: m6A and U-Tail. Cell *158*, 980–987.

Leonhardt, N., Kwak, J.M., Robert, N., Waner, D., Leonhardt, G., and Schroeder, J.I. (2004). Microarray Expression Analyses of Arabidopsis Guard Cells and Isolation of a Recessive Abscisic Acid Hypersensitive Protein Phosphatase 2C Mutant. Plant Cell *16*, 596–615.

Leontis, N.B., and Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. RNA *7*, 499–512.

Lestrade, L., and Weber, M.J. (2006). snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic Acids Res. *34*, D158–D162.

Li, F., Zheng, Q., Vandivier, L.E., Willmann, M.R., Chen, Y., and Gregory, B.D. (2012a). Regulatory Impact of RNA Secondary Structure across the Arabidopsis Transcriptome. Plant Cell Online *24*, 4346–4359.

Li, F., Zheng, Q., Ryvkin, P., Dragomir, I., Desai, Y., Aiyer, S., Valladares, O., Yang, J., Bambina, S., Sabin, L.R., et al. (2012b). Global Analysis of RNA Secondary Structure in Two Metazoans. Cell Rep. *1*, 69–82.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation Protects miRNAs and siRNAs from a 3′-End Uridylation Activity in Arabidopsis. Curr. Biol. *15*, 1501–1507.

Li, S., Vandivier, L.E., Tu, B., Gao, L., Won, S.Y., Li, S., Zheng, B., Gregory, B.D., and Chen, X. (2014). Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis. Genome Res. *25*, 235-245.

Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., and Yi, C. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat. Chem. Biol. *11*, 592–597.

Li, X., Xiong, X., Wang, K., Wang, L., Shu, X., Ma, S., and Yi, C. (2016). Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. Nat. Chem. Biol. *12*, 311–316.

Limbach, P.A., Crain, P.F., and McCloskey, J.A. (1994). Summary: the modified nucleosides of RNA. Nucleic Acids Res. *22*, 2183–2196.

Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E., and Jaffrey, S.R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat. Methods *12*, 767–772.

Liu, H.X., Goodall, G.J., Kole, R., and Filipowicz, W. (1995). Effects of secondary structure on pre-mRNA splicing: hairpins sequestering the 5’ but not the 3’ splice site inhibit intron processing in Nicotiana plumbaginifolia. EMBO J. *14*, 377–388.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95.

Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. (2013). Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. RNA *19*, 1848–1856.

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature *518*, 560–564.

Lorsch, J.R. (2002). RNA Chaperones Exist and DEAD Box Proteins Get a Life. Cell *109*, 797–800.

Lovejoy, A.F., Riordan, D.P., and Brown, P.O. (2014). Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine Synthases Modify Specific mRNAs in S. cerevisiae. PLOS ONE *9*, e110799.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res. *25*, 0955-964.

Luo, S., and Tong, L. (2014). Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. Proc. Natl. Acad. Sci. *111*, 13834–13839.

Luo, G.-Z., MacQueen, A., Zheng, G., Duan, H., Dore, L.C., Lu, Z., Liu, J., Chen, K., Jia, G., Bergelson, J., et al. (2014). Unique features of the m6A methylome in Arabidopsis thaliana. Nat. Commun. *5*, 5630.

Ma, J., Svoboda, P., Schultz, R.M., and Stein, P. (2001). Regulation of Zygotic Gene Activation in the Preimplantation Mouse Embryo: Global Activation and Repression of Gene Expression. Biol. Reprod. *64*, 1713–1721.

Ma, X., Zhao, X., and Yu, Y.-T. (2003). Pseudouridylation (Ψ) of U2 snRNA in S.cerevisiae is catalyzed by an RNA-independent mechanism. EMBO J. *22*, 1889–1897.

Ma, X., Zhu, X., Han, Y., Story, B., Do, T., Song, X., Wang, S., Zhang, Y., Blanchette, M., Gogol, M., et al. (2017). Aubergine Controls Germline Stem Cell Self-Renewal and Progeny Differentiation via Distinct Mechanisms. Dev. Cell *41*, 157–169.e5.

Machnicka, M.A., Milanowska, K., Oglou, O.O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M., et al. (2012). MODOMICS: a database of RNA modification pathways—2012 update. Nucleic Acids Res. *41*, D262-D267.

Maden, B.E.H. (1990). The Numerous Modified Nucleotides in Eukaryotic Ribosomal RNA. In Progress in Nucleic Acid Research and Molecular Biology, W.E.C. and K. Moldave, ed. (Academic Press), pp. 241–303.

Madhani, H.D. (2013). snRNA Catalysts in the Spliceosome's Ancient Core. Cell *155*, 1213–1215.

Marchand, V., Blanloeil-Oillo, F., Helm, M., and Motorin, Y. (2016). Illumina-based RiboMethSeq approach for mapping of 2′-O-Me residues in RNA. Nucleic Acids Res. *44*, e135–e135.

Massenet, S., Mougin, A., and Branlant, C. (1998). Posttranscriptional Modifications in the U Small Nuclear RNAs. 201–227.

Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. *101*, 7287–7292.

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A.V., Patil, D.P., Linder, B., Pickering, B.F., Vasseur, J.-J., Chen, Q., et al. (2016). Reversible methylation of m6Am in the 5′ cap controls mRNA stability. Nature *541*, 371–375.

McCloskey, J.A., and Rozenski, J. (2005). The Small Subunit rRNA Modification Database. Nucleic Acids Res. *33*, D135–D138.

Medvedev, S., Yang, J., Hecht, N.B., and Schultz, R.M. (2008). CDC2A (CDK1)-mediated phosphorylation of MSY2 triggers maternal mRNA degradation during mouse oocyte maturation. Dev. Biol. *321*, 205–215.

Medvedev, S., Pan, H., and Schultz, R.M. (2011). Absence of MSY2 in Mouse Oocytes Perturbs Oocyte Growth and Maturation, RNA Stability, and the Transcriptome. Biol. Reprod. *85*, 575–583.

Meng, Z., and Limbach, P.A. (2006). Mass spectrometry of RNA: linking the genome to the proteome. Brief. Funct. Genomic. Proteomic. *5*, 87–95.

Mengel-Jørgensen, J., and Kirpekar, F. (2002). Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry. Nucleic Acids Res. *30*, e135–e135.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. Cell *149*, 1635–1646.

Miranda-Ríos, J. (2007). The THI-box Riboswitch, or How RNA Binds Thiamin Pyrophosphate. Structure *15*, 259–265.

Miyauchi, K., Kimura, S., and Suzuki, T. (2013). A cyclic form of N6-threonylcarbamoyladenosine as a widely distributed tRNA hypermodification. Nat. Chem. Biol. *9*, 105–111.

Mohr, S., Stryker, J.M., and Lambowitz, A.M. (2002). A DEAD-Box Protein Functions as an ATP-Dependent RNA Chaperone in Group I Intron Splicing. Cell *109*, 769–779.

Møller, T., Franch, T., Højrup, P., Keene, D.R., Bächinger, H.P., Brennan, R.G., and Valentin-Hansen, P. (2002). Hfq: A Bacterial Sm-like Protein that Mediates RNA-RNA Interaction. Mol. Cell *9*, 23–30.

Motorin, Y., and Grosjean, H. (1999). Multisite-specific tRNA:m5C-methyltransferase (Trm4) in yeast Saccharomyces cerevisiae: identification of the gene and substrate specificity of the enzyme. RNA *5*, 1105–1118.

Motorin, Y., Muller, S., Behm-Ansmant, I., and Branlant, C. (2007). Identification of Modified Residues in RNAs by Reverse Transcription-Based Methods. B.-M. in Enzymology, ed. (Academic Press), pp. 21–53.

Mustroph, A., Zanetti, M.E., Jang, C.J.H., Holtan, H.E., Repetti, P.P., Galbraith, D.W., Girke, T., and Bailey-Serres, J. (2009). Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis. Proc. Natl. Acad. Sci. *106*, 18843–18848.

Narayan, P., Ludwiczak, R.L., Goodwin, E.C., and Rottman, F.M. (1994). Context effects on N6-adenosine methylation sites in prolactin mRNA. Nucleic Acids Res. *22*, 419–426.

Narsai, R., Howell, K.A., Millar, A.H., O'Toole, N., Small, I., and Whelan, J. (2007). Genome-Wide Analysis of mRNA Decay Rates and Their Determinants in Arabidopsis thaliana. Plant Cell *19*, 3418–3436.

Newby, M.I., and Greenbaum, N.L. (2002). Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. Proc. Natl. Acad. Sci. *99*, 12697–12702.

Ni, J., Tien, A.L., and Fournier, M.J. (1997). Small Nucleolar RNAs Direct Site-Specific Synthesis of Pseudouridine in Ribosomal RNA. Cell *89*, 565–573.

Nichols, J.L. (1979). 'Cap' structures in maize poly(A)-containing RNA. Biochim. Biophys. Acta BBA - Nucleic Acids Protein Synth. *563*, 490–495.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. Science *289*, 920–930.

Novikova, I.V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2012). Sizing up long non-coding RNAs: Do lncRNAs have secondary and tertiary structure? BioArchitecture *2*, 189–199.

Oberstrass, F.C., Lee, A., Stefl, R., Janis, M., Chanfreau, G., and Allain, F.H.-T. (2006). Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. Nat. Struct. Mol. Biol. *13*, 160–167.

Ofengand, J., and Bakin, A. (1997). Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. J. Mol. Biol. *266*, 246–268.

Oikawa, D., Tokuda, M., Hosoda, A., and Iwawaki, T. (2010). Identification of a consensus element recognized and cleaved by IRE1 alpha. Nucleic Acids Res. *38*, 6265–6273.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.-H., and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature *372*, 432–438.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in Arabidopsis thaliana. Curr. Biol. *12*, 1484–1495.

Patil, D.P., Chen, C.-K., Pickering, B.F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S.R. (2016). m6A RNA methylation promotes XIST-mediated transcriptional repression. Nature *537*, 369–373.

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

Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature *334*, 320–325.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes Dev. *18*, 2368–2379.

Perry, R.P., and Kelley, D.E. (1974). Existence of methylated messenger RNA in mouse L cells. Cell *1*, 37–42.

Perry, R.P., Kelley, D.E., Friderici, K., and Rottman, F. (1975). The methylated constituents of L cell messenger RNA: Evidence for an unusual cluster at the 5′ terminus. Cell *4*, 387–394.

Piko, L., and Clegg, K.B. (1982). Quantitative changes in total RNA, total poly(A), and ribosomes in early mouse embryos. Dev. Biol. *89*, 362–378.

Ping, X.-L., Sun, B.-F., Wang, L., Xiao, W., Yang, X., Wang, W.-J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.-S., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell Res. *24*, 177–189.

Ponting, C.P., Oliver, P.L., and Reik, W. (2009). Evolution and Functions of Long Noncoding RNAs. Cell *136*, 629–641.

Qu, X., Wen, J.-D., Lancaster, L., Noller, H.F., Bustamante, C., and Tinoco, I. (2011). The ribosome uses two active mechanisms to unwind messenger RNA during translation. Nature *475*, 118–121.

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

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. (2006). A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev. *20*, 3407–3425.

Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M.F., Konrat, R., Bläsi, U., and Schroeder, R. (2007). RNA Chaperones, RNA Annealers and RNA Helicases. RNA Biol. *4*, 118–130.

Raker, V.A., Mironov, A.A., Gelfand, M.S., and Pervouchine, D.D. (2009). Modulation of alternative splicing by long-range RNA structures in Drosophila. Nucleic Acids Res. *37*, 4533–4544.

Ram, P.T., and Schultz, R.M. (1993). Reporter gene expression in G2 of the 1-cell mouse embryo. Dev. Biol. *156*, 552–556.

Ramakrishnan, V. (2014). The Ribosome Emerges from a Black Box. Cell *159*, 979–984.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev. *16*, 1616–1626.

Rhee, H.S., and Pugh, B.F. (2001). ChIP-exo Method for Identifying Genomic Location of DNA-Binding Proteins with Near-Single-Nucleotide Accuracy. In Current Protocols in Molecular Biology, (John Wiley & Sons, Inc.). 100:21.24.1-21.24.14.

Rinn, J.L., and Chang, H.Y. (2012). Genome Regulation by Long Noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F., and Klug, A. (1974). Structure of yeast phenylalanine tRNA at 3 A resolution. Nature *250*, 546–551.

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.

Roost, C., Lynch, S.R., Batista, P.J., Qu, K., Chang, H.Y., and Kool, E.T. (2015). Structure and Thermodynamics of N6-Methyladenosine in RNA: A Spring-Loaded Base Modification. J. Am. Chem. Soc. *137*, 2107–2115.

Roundtree, I.A., and He, C. (2016). RNA epigenetics — chemical messages for posttranscriptional gene regulation. Curr. Opin. Chem. Biol. *30*, 46–51.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature *505*, 701–705.

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

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83–86.

Ryter, J.M. (1998). Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. EMBO J. *17*, 7505–7513.

Ryvkin, P., Leung, Y.Y., Silverman, I.M., Childress, M., Valladares, O., Dragomir, I., Gregory, B.D., and Wang, L.-S. (2013). HAMR: high-throughput annotation of modified ribonucleotides. RNA *19*, 1684–1692.

Saletore, Y., Meyer, K., Korlach, J., Vilfan, I.D., Jaffrey, S., and Mason, C.E. (2012). The birth of the Epitranscriptome: deciphering the function of RNA modifications. Genome Biol. *13*, 175.

Sallés, F., and Strickland, S. (1999). Analysis of Poly(A) Tail Lengths by PCR: The PAT Assay. In RNA-Protein Interaction Protocols, S. Haynes, ed. (Humana Press), pp. 441–448.

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., and Lyko, F. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes Dev. *24*, 1590–1595.

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

Schroeder, R., Barta, A., and Semrad, K. (2004). Strategies for RNA folding and assembly. Nat. Rev. Mol. Cell Biol. *5*, 908–919.

Schultz, R.M., and Wassarman, P.M. (1977). Biochemical studies of mammalian oogenesis: Protein synthesis during oocyte growth and meiotic maturation in the mouse. J. Cell Sci. *24*, 167–194.

Schultz, R.M., Montgomery, R.R., and Belanoff, J.R. (1983). Regulation of mouse oocyte meiotic maturation: Implication of a decrease in oocyte cAMP and protein dephosphorylation in commitment to resume meiosis. Dev. Biol. *97*, 264–273.

Schwartz, S. (2016). Cracking the epitranscriptome. RNA *22*, 169–174.

Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G., et al. (2013). High-Resolution Mapping Reveals a Conserved, Widespread, Dynamic mRNA Methylation Program in Yeast Meiosis. Cell *155*, 1409–1421.

Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., et al. (2014a). Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5′ Sites. Cell Rep. *8*, 284–296.

Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., León-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S., et al. (2014b). Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. Cell *159*, 148–162.

Sibert, B.S., and Patton, J.R. (2012). Pseudouridine synthase 1: a site-specific synthase without strict sequence recognition requirements. Nucleic Acids Res. *40*, 2107–2118.

Sprinzl, M., and Vassilenko, K.S. (2005). Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. *33*, D139–D140.

Squires, J.E., Patel, H.R., Nousch, M., Sibbritt, T., Humphreys, D.T., Parker, B.J., Suter, C.M., and Preiss, T. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res. *40*, 5023–5033.

Steitz, T.A., and Moore, P.B. (2003). RNA, the first macromolecular catalyst: the ribosome is a ribozyme. Trends Biochem. Sci. *28*, 411–418.

Su, Y.-Q., Sugiura, K., Woo, Y., Wigglesworth, K., Kamdar, S., Affourtit, J., and Eppig, J.J. (2007). Selective degradation of transcripts during meiotic maturation of mouse oocytes. Dev. Biol. *302*, 104–117.

Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J.-L., and Moine, H. (2011). G-quadruplex RNA structure as a signal for neurite mRNA targeting. EMBO Rep. *12*, 697–704.

Sundaram, M., Durant, P.C., and Davis, D.R. (2000). Hypermodified Nucleosides in the Anticodon of tRNALys Stabilize a Canonical U-Turn Structure. Biochemistry (Mosc.) *39*, 12575–12584.

Svitkin, Y.V., Pause, A., Haghighat, A., Pyronnet, S., Witherell, G., Belsham, G.J., and Sonenberg, N. (2001). The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. RNA N. Y. N *7*, 382–394.

Talkish, J., May, G., Lin, Y., Woolford, J.L., and McManus, C.J. (2014). Mod-seq: high-throughput sequencing for chemical probing of RNA structure. RNA *20*, 713–720.

Tanaka, Y., Dyer, T.A., and Brownlee, G.G. (1980). An improved direct RNA sequence method; its application to Vida faba 5.8S ribosomal RNA. Nucleic Acids Res. *8*, 1259–1272.

Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J. *18*, 1169–1175.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. Science *329*, 689–693.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537–1550.

Underwood, J.G., Uzilov, A.V., Katzman, S., Onodera, C.S., Mainzer, J.E., Mathews, D.H., Lowe, T.M., Salama, S.R., and Haussler, D. (2010). FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. Nat. Methods *7*, 995–1001.

Urban, A., Behm-Ansmant, I., Branlant, C., and Motorin, Y. (2009). RNA sequence and two-dimensional structure features required for efficient substrate modification by the Saccharomyces cerevisiae RNA:{Psi}-synthase Pus7p. J. Biol. Chem. *284*, 5845–5858.

Vandivier, L., Li, F., Zheng, Q., Willmann, M., Chen, Y., and Gregory, B. (2013). Arabidopsis mRNA secondary structure correlates with protein function and domains. Plant Signal. Behav. *8*, e24301.

Vandivier, L.E., Campos, R., Kuksa, P.P., Silverman, I.M., Wang, L.-S., and Gregory, B.D. (2015a). Chemical Modifications Mark Alternatively Spliced and Uncapped Messenger RNAs in Arabidopsis. Plant Cell *27*, 3024–3037.

Vandivier, L.E., Li, F., and Gregory, B.D. (2015b). High-throughput nuclease-mediated probing of RNA secondary structure in plant transcriptomes. Methods Mol. Biol. *1284*, 41–70.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.-L., Bartel, D.P., and Crété, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol. Cell *16*, 69–79.

Wang, D., and Deal, R. (2015). Epigenome Profiling of Specific Plant Cell Types Using a Streamlined INTACT Protocol and ChIP-seq. In Plant Functional Genomics, J.M. Alonso, and A.N. Stepanova, eds. (Springer New York), pp. 3–25.

Wang, K.C., and Chang, H.Y. (2011). Molecular Mechanisms of Long Noncoding RNAs. Mol. Cell *43*, 904–914.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014a). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N6-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388–1399.

Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014b). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat. Cell Biol. *16*, 191–198.

Wanrooij, P.H., Uhler, J.P., Simonsson, T., Falkenberg, M., and Gustafsson, C.M. (2010). G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. Proc. Natl. Acad. Sci. *107*, 16072–16077.

Warf, M.B., and Berglund, J.A. (2010). Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem. Sci. *35*, 169–178.

Wen, J.-D., Lancaster, L., Hodges, C., Zeri, A.-C., Yoshimura, S.H., Noller, H.F., Bustamante, C., and Tinoco, I. (2008). Following translation by single ribosomes one codon at a time. Nature *452*, 598–603.

Wetzel, C., and A. Limbach, P. (2016). Mass spectrometry of modified RNAs: recent developments. Analyst *141*, 16–23.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat. Protoc. *1*, 1610–1616.

Will, S., Reiche, K., Hofacker, I.L., Stadler, P.F., and Backofen, R. (2007). Inferring Noncoding RNA Families and Classes by Means of Genome-Scale Structure-Based Clustering. PLoS Comput Biol *3*, e65.

Williams, A.S., and Marzluff, W.F. (1995). The sequence of the stem and flanking sequences at the 3′ end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. Nucleic Acids Res. *23*, 654–662.

Williamson, J.R. (2000). Induced fit in RNA–protein recognition. Nat. Struct. Mol. Biol. *7*, 834–837.

Willmann, M.R., Berkowitz, N.D., and Gregory, B.D. (2014). Improved genome-wide mapping of uncapped and cleaved transcripts in eukaryotes—GMUCT 2.0. Methods *67*, 64–73.

Woodson, S.A., Muller, J.G., Burrows, C.J., and Rokita, S.E. (1993). A primer extension assay for modification of guanine by Ni(ll) complexes. Nucleic Acids Res. *21*, 5524–5525.

Wu, G., Park, M.Y., Conway, S.R., Wang, J.-W., Weigel, D., and Poethig, R.S. (2009). The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis. Cell *138*, 750–759.

Wu, G., Yu, A.T., Kantartzis, A., and Yu, Y.-T. (2011a). Functions and mechanisms of spliceosomal small nuclear RNA pseudouridylation. Wiley Interdiscip. Rev. RNA *2*, 571–581.

Wu, G., Xiao, M., Yang, C., and Yu, Y.-T. (2011b). U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. EMBO J. *30*, 79–89.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., et al. (2016). Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. Mol. Cell *61*, 507–519.

Xie, Z., Allen, E., Wilken, A., and Carrington, J.C. (2005). DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. Proc. Natl. Acad. Sci. *102*, 12984–12989.

Xu, C., Wang, X., Liu, K., Roundtree, I.A., Tempel, W., Li, Y., Lu, Z., He, C., and Min, J. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. Nat. Chem. Biol. *10*, 927–929.

Yacoubi, B.E., Bailly, M., and Crécy-Lagard, V. de (2012). Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. Annu. Rev. Genet. *46*, 69–95.

Yi-Brunozzi, H.Y., Easterwood, L.M., Kamilar, G.M., and Beal, P.A. (1999). Synthetic substrate analogs for the RNA-editing adenosine deaminase ADAR-2. Nucleic Acids Res. *27*, 2912–2917.

Yoo, S.-D., Cho, Y.-H., and Sheen, J. (2007). Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. *2*, 1565–1572.

Yoshikawa, M., Peragine, A., Park, M.Y., and Poethig, R.S. (2005). A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev. *19*, 2164–2175.

Yu, A.T., Ge, J., and Yu, Y.-T. (2011). Pseudouridines in spliceosomal snRNAs. Protein Cell *2*, 712–725.

Yu, X., Willmann, M.R., Anderson, S.J., and Gregory, B.D. (2016). Genome-Wide Mapping of Uncapped and Cleaved Transcripts Reveals a Role for the Nuclear mRNA Cap-Binding Complex in Cotranslational RNA Decay in Arabidopsis. Plant Cell *28*, 2385–2397.

Yusupova, G., and Yusupov, M. (2014). High-Resolution Structure of the Eukaryotic 80S Ribosome. Annu. Rev. Biochem. *83*, 467–486.

Zeng, F., and Schultz, R.M. (2005). RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo. Dev. Biol. *283*, 40–57.

Zeng, F., Baldwin, D.A., and Schultz, R.M. (2004). Transcript profiling during preimplantation mouse development. Dev. Biol. *272*, 483–496.

Zhang, H.-Y., Xiong, J., Qi, B.-L., Feng, Y.-Q., and Yuan, B.-F. (2016). The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. Chem. Commun. *52*, 737–740.

Zhang, R., Calixto, C.P.G., Tzioutziou, N.A., James, A.B., Simpson, C.G., Guo, W., Marquez, Y., Kalyna, M., Patro, R., Eyras, E., et al. (2015). AtRTD – a comprehensive reference transcript dataset resource for accurate quantification of transcript-specific expression in Arabidopsis thaliana. New Phytol. *208*, 96–101.

Zhao, B.S., Roundtree, I.A., and He, C. (2016). Post-transcriptional gene regulation by mRNA modifications. Nat. Rev. Mol. Cell Biol. *18*, 31–42.

Zhao, X., Yang, Y., Sun, B.-F., Shi, Y., Yang, X., Xiao, W., Hao, Y.-J., Ping, X.-L., Chen, Y.-S., Wang, W.-J., et al. (2014). FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. Cell Res. *24*, 1403–1419.

Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.-M., Li, C.J., Vågbø, C.B., Shi, Y., Wang, W.-L., Song, S.-H., et al. (2013). ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. Mol. Cell *49*, 18–29.

Zheng, Q., Ryvkin, P., Li, F., Dragomir, I., Valladares, O., Yang, J., Cao, K., Wang, L.-S., and Gregory, B.D. (2010). Genome-Wide Double-Stranded RNA Sequencing Reveals the Functional Significance of Base-Paired RNAs in Arabidopsis. PLOS Genet. *6*, e1001141.

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., and Fray, R.G. (2008). MTA Is an Arabidopsis Messenger RNA Adenosine Methylase and Interacts with a Homolog of a Sex-Specific Splicing Factor. Plant Cell *20*, 1278–1288.

Zhou, H., Kimsey, I.J., Nikolova, E.N., Sathyamoorthy, B., Grazioli, G., McSally, J., Bai, T., Wunderlich, C.H., Kreutz, C., Andricioaei, I., et al. (2016). m1A and m1G disrupt A-RNA structure through the intrinsic instability of Hoogsteen base pairs. Nat. Struct. Mol. Biol. *23*, 803–810.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.-B. (2015). Dynamic m6A mRNA methylation directs translational control of heat shock response. Nature *526*, 591–594.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr. Biol. CB *14*, 1214–1220.

Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. *9*, 133–148.