THE UNIVERSITY OF CHICAGO

DYNAMIC TRNA MODIFICATION LANDSCAPE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

BY
WESLEY CONYERS CLARK

CHICAGO, ILLINOIS
AUGUST 2016

To my parents:

without your support, I would not have had the fortitude and endurance to make this anything but a hollow endeavor.

To my friends:

I dedicate this to the moments I missed with you. Forgive me.

"Von einem gewissen Punkt an gibt es keine Rckkehr mehr. Dieser Punkt ist zu erreichen."

(*"From a certain point onward there is no longer any turning back. That is the point that must be reached."*)

Franz Kafka

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to acknowledge and thank all of the Pan Lab members of the past who trained me and worked with me and members of the present who continue to shape me and my ideas. I owe special thanks to Dr. George Perdrizet for welcoming me as a friend and helping me to find my home in the Pan Lab and to Dr. Marc Parisien for training me in bioinformatics to push me forward. I also owe thanks to Dr. Guanqun Zheng and Dr. Ana Cristina Gomes for sharing in ideas and experiments that have helped me in my pursuits. I owe a great deal to my committee of Dr. Tobin Sosnick, Dr. Alex Ruthenburg, and Dr. D. Allan Drummond. Thank you to Dr. Tobin, whose enthusiasm in his course and in casual conversations always left me with more questions than I started. Next to Dr. Alex Ruthenburg, whose mentoring and advice during my preliminary exam and during my medical absence helped me to remain on the right track. And, to Dr. D. Allan Drummond, whose charisma and mentoring set me on the right path from my first year, and without whose advice I would have spent much more time with my brow furrowed. Lastly I owe thanks to Dr. Tao Pan, for were it not for him, none of this would have come to fruition.

ABSTRACT

Eukaryotic tRNAs contain on average 14 modified nucleotides per molecule. Modifications in tRNAs are responsible for their structure and their function. During stress, tRNAs have been shown to modulate and reprogram their tRNA status in order to fine-tune translational cellular responses. tRNAs deficient in even one modification have been known to cause deleterious health effects for organisms. However, full scale investigations thus far into the global biological functions of many tRNA modifications have been insufficient in scope due to the difficulty of resolution of tRNA isodecoders. Also, in order to obtain exact information regarding modifications, tools thus far used to study tRNA modifications are at best imprecise for modification fraction determination. However, in the advent of high-throughput sequencing, it is possible to determine and quantify modifications in RNA molecules. Herein, the method of tRNA-seq and application to human eukaryotic tRNA is described. Modifications impact and impair the process of reverse transcription, and due to this nuclear-encoded and mitochondrial Watson-Crick interface modifications can be identified via proxy through sequencing due to the presence of mutations and truncations in cDNA. These mutations and stops in the cDNA reads allow us to assign a metric for each position in all eukaryotic tRNAs, which we call modification index (MI). From MI, we can determine positions of modification by proxy. Using pre-existing knowledge of tRNA modifications from homology and other studies, we are able to validate existing modifications as well as discover new sites. Further bioinformatic work is done to identify modification signatures based on context, modification identity, and reverse transcriptase. This is used to identify other types of modifications more broadly in other RNA molecules.

CHAPTER 1

DYNAMIC RNA MODIFICATIONS

1.1 Introduction

All somatic cells contain the same biological information, and the content of genetic information flows from DNA to RNA and then finally to protein. However, the utilization of this information varies depending on cellular function. The cell regulates this expression of genetic information in various ways, and the most common tools are with methylations and modifications of nucleic acids in the genetic code. Reversible modifications were known to exist in DNA and protein for a long time, such as shown in Figure 1.1 [34]. However, with RNA, because the molecule itself was short-lived, all modifications that would exist were believed to be permanent. Indeed, while there were known to be more than 100 chemical modifications on RNA molecules [126], the field believed that for the studied modifications in ribosomal and transfer RNA were installed[47], and in the absence of modification, the respective molecules would be degraded. These nucleotide modifications are well-conserved, further solidifying the notion of a non-dynamic control.

Even still, while mRNA regulation of translation is tightly controlled, it was believed that abundance [124] and variation of structure [55] were the key components to tight genetic regulation [73]. However, in the advent of high throughput sequencing coupled with powerful biochemical tools, we have come to understand that the RNA itself is edited and modified. How then do we change the transcriptomic code of A, C, G, and U? Some of these modifications are permanent and non-reversible. Known pathways of irreversible adenosine-to-inosine have been elucidated [8, 80], and the obvious implications of this modification include genetic recoding at the sequence level. Thus, the A-to-I allows a permanent shift in interpretation by the eyes of the cell.

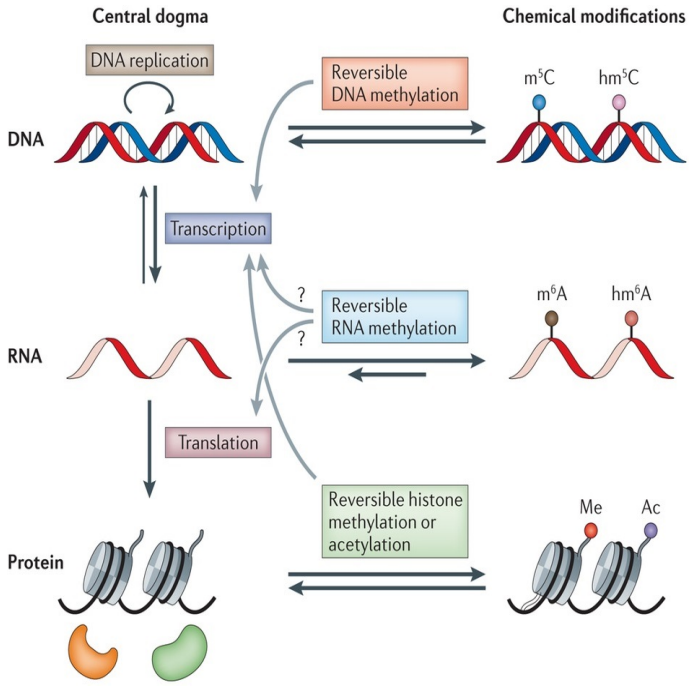

Figure 1.1: Modifications give dynamic control to biological molecules. For each molecule in the central dogma, there are known crucial modifications that modulate cellular activity to provide a range of function. Source: [34].

However, what is more interesting are the reversible modifications: the addition of chemical adducts that can be removed just as quickly as they were installed. For instance, the mRNA modification 6-methyladenosine (m6A) was demonstrated to be present across the entire transcriptome [31, 74, 75] . m6A has been shown to be tightly regulated developmentally in many organisms [9], lending a dynamic tool to quickly aid in somatic cell differentiation in embryonic cells. The pathways to both methylate and demethylate this particular alkyl addition are well-characterized and represent a fine balance of controlled cellular chaos. It is as though we are changing the A with a diacritic - accenting it to nuance its interpretation. There are also other reversible modifications that have been shown to exist ubiquitously in mRNA that are also managed by protein partners[126, 34, 59, 120, 101]. These include 5-methylcytosine (m^5C), pseudouridine (ψ), and more recently 1-methyladenosine (m^1A), each presumably with their own writers, readers, and erasers.

Invariably, cells use these epigenetic interactive pathways of reversible modifications to regulate expression and function post-transcriptionally. The temporal control afforded by reversible modifications allows for tightly-tuned dynamic regulation - a chemical signaling cascade overseeing the epitranscriptome. Furthermore, as mentioned before is the existence of many of these modifications in coding RNAs as well as on non-coding RNAs, such as ribosomal RNA [27], long non-coding RNAs such as MALAT1 [73], and transfer RNAs[93]. This raises the question, then, of what these modifications are used for outside of manipulation of expression and how these modifications are come to interact in the network of the cellular machinery.

1.2 RNA Modification Analysis via High-Throughput Sequencing Methods and Analysis

High-throughput sequencing has been the main tool in order to identify modifications genome-wide and transcriptome-wide [94]. While other methods are well-established, many of the methods that are known to the field such as primer extension rely on the ability of a reverse transcriptase to turn RNA to cDNA. Thus, it is only natural that doing this in parallel over many sequences gives a depth and breadth to the discovery of modifications that other low-resolution methods simply cannot afford. What makes high-throughput sequencing just as desirable of a tool is the ability to couple its resolution with biochemical tools such as antibody immunoprecipitation or chemical modifications[34, 65]. These couplings allow for greater specificity and enrichment or subtle changes to cDNA, respectively, in order to gain more information from the biological sample. Most major modification studies use one or the other: in the case of m6A and m^1A, antibodies have primarily been employed; and, for m^5C and ψ, chemical methods coupled with subtle shifts in sequencing output are utilized [126, 34, 59, 120, 101].

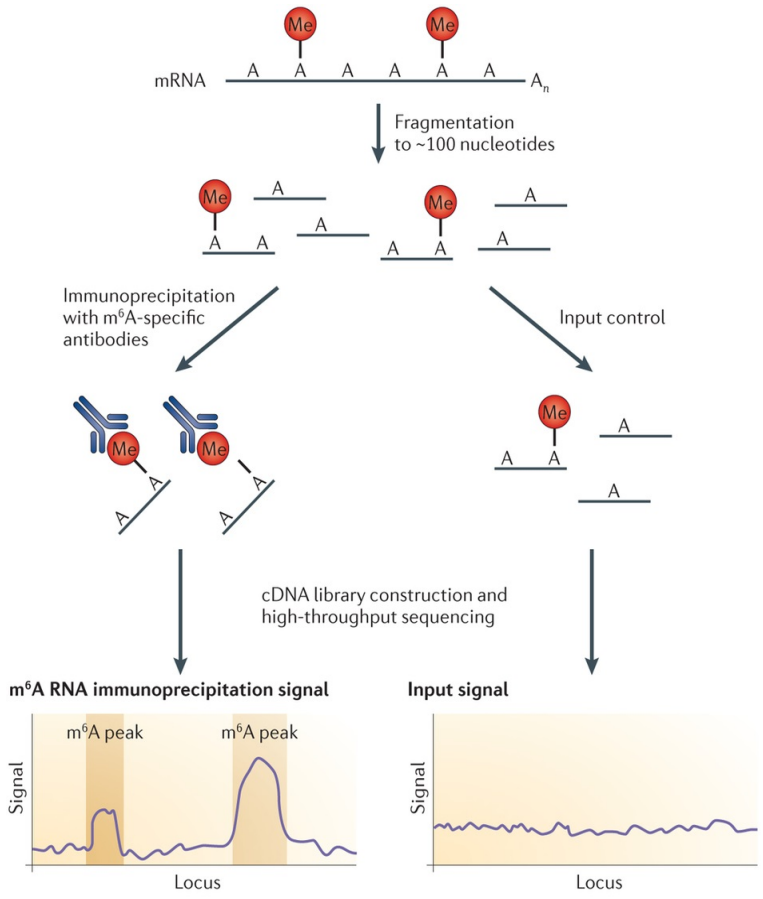

Figure 1.2: Biochemical tools assisted high-throughput sequencing for efficacy in discovering sites of modification. Specifically, the use of antibodies coupled with stringent reaction conditions allows researchers to discover and understand the roles of modifications in the transcriptome. This schematic shows the fundamentals of the Me-RIP (methyl-RNA immunoprecipitation) protocol. Source: [34]

Of course, what makes this method powerful also underlies inherent issues: for instance, with the use of antibodies, the quantitative information about modification fraction is lost due to the enrichment step. There are other artifacts that have been shown to be causally related to avidity effects or otherwise binding of the antibody to the RNA modification. Chemical treatments have their own litany of downsides, not the least of which is degradation of biological material due to stringent chemical step - such is the case of CMC-treatment for pseudouridine [16, 102], as an alkaline pH reversal step is required in order to remove the adduct from non-pseudouridine nucleosides. At this pH, RNA is much more likely to

self-cleave due to the inherent chemical properties of the molecule.

As was mentioned earlier, many methods of identification using a reverse transcriptase come from the signal provided by the experiment. What this means is how the reverse transcriptase reacts to the potential roadblock when converting RNA to cDNA[78, 53]. Thus, the true utility of the method comes from the downstream data analysis. For some modifications, the downstream effects in cDNA production allow for identification of modification by proxy [44].

What this means is that RT arrest which could lead to a truncation in cDNA is one method that researchers have used to identify sites of modification. Another is that the modified nucleobase could either be resistant to various chemical conditions. In this instance, when trying to determine sites of 5-methylcytosine in RNA, the field can use sodium bisulfite in neutral or acidic pH conditions[61]. This causes a conversion of C residues in single-stranded RNA to be chemically deaminated to U. However, 5-methylcytosine modified nucleobases are resistant to this, and naturally a ratio of C-to-U conversion can be detected in sequencing datasets.

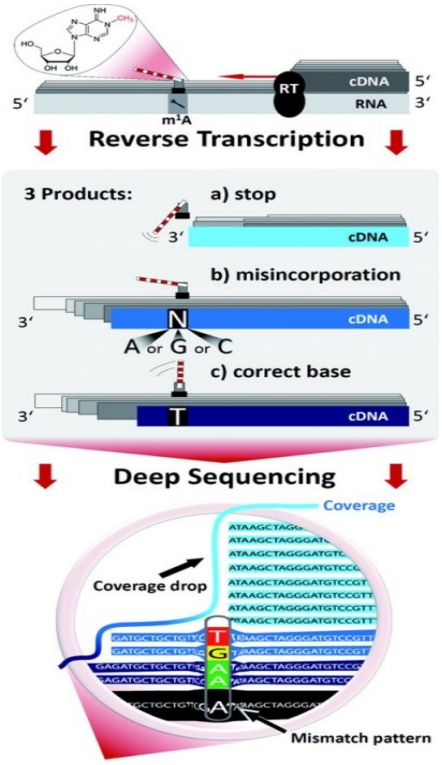

Figure 1.3: This schematic by Hauenschild et al (Source:[44]) illustrates that for each modification, we can uncover a wealth of information from the cDNA due to RT interactions with the modification.

In summary, the power of high-throughput sequencing is not only in the depth of coverage that the technology can afford to us, but also its adaptability to various biochemical-assisted methods that allows for rapid and powerful identification and even quantitation of modifica-

tions. As the hardware for this technology grows exponentially as it should, the limitations for detection exist in the experimental design and also in the data analysis.

1.3 Modification Dynamics in mRNA via regulated writers and erasers

Proteins modulate modifications. In the case of certain types of modifications, they are driven and guided by RNA, but fundamentally all modifications exist because of important enzymatic controllers. These enzymes can be subdivided into three classes - the enzymes that provide the modification; the proteins that interpret the modification; and, lastly, the proteins that remove the modifications - writers, readers, and erasers respectively[34, 73, 120]. In order to consider modification dynamics, we have to consider two things: how the modification first arrived, and the tuning of the modification by removing it. Another aspect of modifications to consider, of course, would be the different cellular components that interact and modulate function or localization or other cellular activity based on the presence or absence of a modified site. However, this purely is downstream of the constantly shifting modification landscape, so we will not consider them at this time. The introductions here mostly focus on m6A, as it was the first of many modifications to be discovered in mRNA. The other mRNA modifications that we have talked about - m^5C, ψ, and m^1A - were all discovered first in non-coding RNAs, and then subsequently found in mRNA.

1.3.1 Modification Writers

In order to annotate our transcriptomic alphabet with the diacritics - the methylations, and in the case of pseudouridine, the necessary isomerization - there must be proteins responsible[120, 126]. These are typically energetically uphill transformations in the kinetic landscape. For a methylation, the transfer of S-adenosylmethionine's activated CH3-methyl group to a donor molecule is preceded by many highly unfavourable steps involving

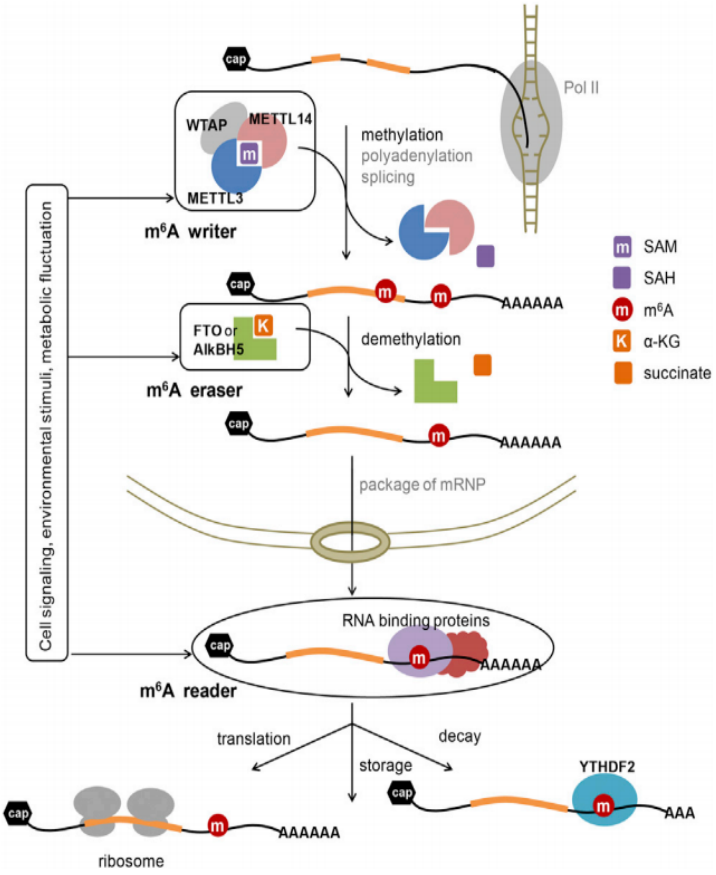

Figure 1.4: The crucial players in dynamic modifications in mRNA. For each role in modification dynamics for mRNA, there are key components that write (METTL3/METTL14 methyltransferases), erase (ABH5 and FTO), and read (YTH-domain containing proteins) which facilitate the global modulation of mRNA localization and cellular function inherent to m6A modification. Source: [120].

the cellular fuel of ATP[6]. For the isomerization of uridine to pseudouridine, this cannot occur spontaneously, and thus a protein is required. These necessary proteins are our writers.

While the discovery of modifications in mRNA was serendipitous at the time, once scientists found the methylated adenosine residue m6A [28], it was only a matter of time before they discovered the necessary enzymes. However, the characterization of enzymes from lysate in order to generate this species took 20 years[12]. A mega-dalton complex of several proteins from HeLa lysate was determined to be the enzymatic methyltransferase, and

subsequently the individual components necessary in vivo were respectively characterized: METTL3, METTL14, and WTAP. From the viewpoint of other known RNA modifications such as 5-methylcytosine and pseudouridine, their writers have also been identified and more recently characterized as well. For 5-methylcytosine, the RCMT (short for RNA cytosine methyltransferase) was first discovered in bacteria and reported to catalyze the formation of 5-methylcytosine in ribosomal RNA [77].

1.3.2 Modification Erasers

Just as important as it is to add a modification to RNA, the role of removing modifications is indispensable in the cell. The discovery of these erasers was once again from bacteria - the AlkB protein in E. coli. As it were, the protein was first discovered as a repair enzyme. AlkB catalyzes oxidative demethylation in DNA and RNA in E. coli [113]. AlkB's purpose in the cell was to remove alkyl lesions from DNA during various stresses, and the knockout of the protein caused the mutants to be highly susceptible to methylating agents[84]. A breakthrough in 2011 showed that a homologue of AlkB called FTO in humans was responsible for oxidative demethylation of m6A in polyadenylated mRNA[50]. Subsequent studies of AlkB homologues also uncovered ALKBH5 (AlkB human homologue 5), a nuclear protein also responsible for m6A demethylation[130]. Its knockout produced abnormal splicing and mRNA synthesis in humans, and the knockout of mice caused aberrant spermatogenesis. These two proteins were crucial in forging the notion that modifications were dynamic: deleterious phenotypes from the absence of these proteins implied that there must be a critical role for active demodification.

1.4 tRNA: Function, Modification, and Dynamics

Transfer RNA, or tRNA as it is more commonly referred to, is one of the most abundant RNA molecules in the cell. It is the third RNA component in translation aside from the

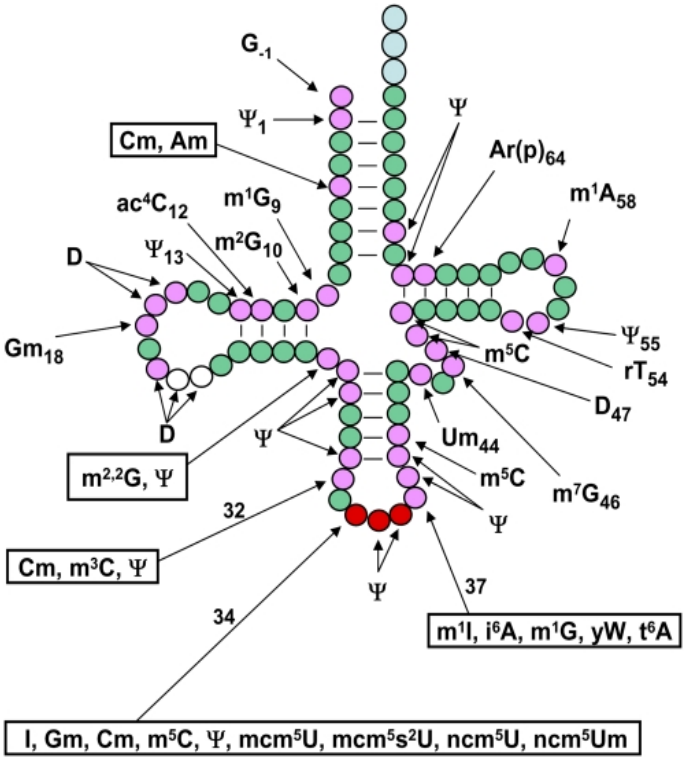

Figure 1.5: A map of eukaryotic tRNA modifications. This figure from shows a schematic of modifications found from S. cerevisiae cytoplasmic tRNA with its characteristic secondary structure. Source: [93]

messenger RNA and the ribosome and its respective RNA. tRNA gets its name by transferring its bound amino-acid to the ribosome to facilitate protein synthesis. It carries out its role in translation by base-pairing its three nucleotide anticodon to the mRNA codon and only then transferring the amino acid it carries to the ribosome. In this respect, tRNA is the mediator in the genetic information pathway from RNA to protein, serving as a physical connector in the central dogma for translation. Since each codon in the mRNA is respective for an amino acid, the tRNA is also the interpreter in translation, since each tRNA molecule has an identity formed from its anticodon sequence which determines which amino acid it transfers to the ribosome.

The cell regulates tRNAs in multitudes of ways. The first of which is controlling the abundance of tRNA themselves. During conditions of stress, it has been shown that tRNA

fragments are generated as crucial mediators as stress signalers[54]. This is accomplished by the protein angiogenin, which cleaves the tRNA near the anticodon. The resultant fragments are bound by various proteins and signal inhibitors for translation[11, 116]. Another method of tRNA regulation is to use the tRNA to misincorporate its amino acid. In mammalian cells, the methionyl-tRNA synthetase MetRS has been shown to mis-methionylate non-methionine tRNAs, allowing for incorporation of the methionine residues at non-methionine sites[79]. This modulates translation by affecting the pool of functional proteins, presumably affording a wider array of function and resistance to the available pool of proteins. This can be used quickly as a translational adjustment to stress by incorporating the oxygen-scavenging amino acid into crucial sites of proteins, affording them protecting during stress. This is a well-characterized and well-conserved regulation across all kingdoms of life. Yet one more regulation of tRNAs comes from modulating the abundance of tRNAs. For instance, during amino acid starvation, the nucleus will re-import tRNA and deplete the cytosolic pool of translational-ready tRNAs, thus allowing for cellular preservation in nutrient starvation[30].

This brings up the last point of tRNA dynamics: controlling the tRNA modifications. Despite being the most highly modified molecule in the cell, transfer RNA modifications are still not widely characterized in eukaryotes[92]. Notably, this molecule has been difficult to characterize, as its extensive secondary and tertiary structure as well as multiple modified residues make it resistant to many biochemical assays[131]. Early on, many genetic studies were carried out in yeast in order to characterize the modification landscape. Figure 1.5 shows a schematic representation of how modified a tRNA is. Recently, the field has been able to resolve identification reasonably high resolution of modification identification using powerful tools such as high-performance liquid chromatography coupled with triple-quadrupole tandem mass spectrometry. However this only is able to determine if a modification exists, and unfortunately the sequence context of such modifications is lost using this. What this tool instead affords is the opportunity to examine molar increases of particular modifications.

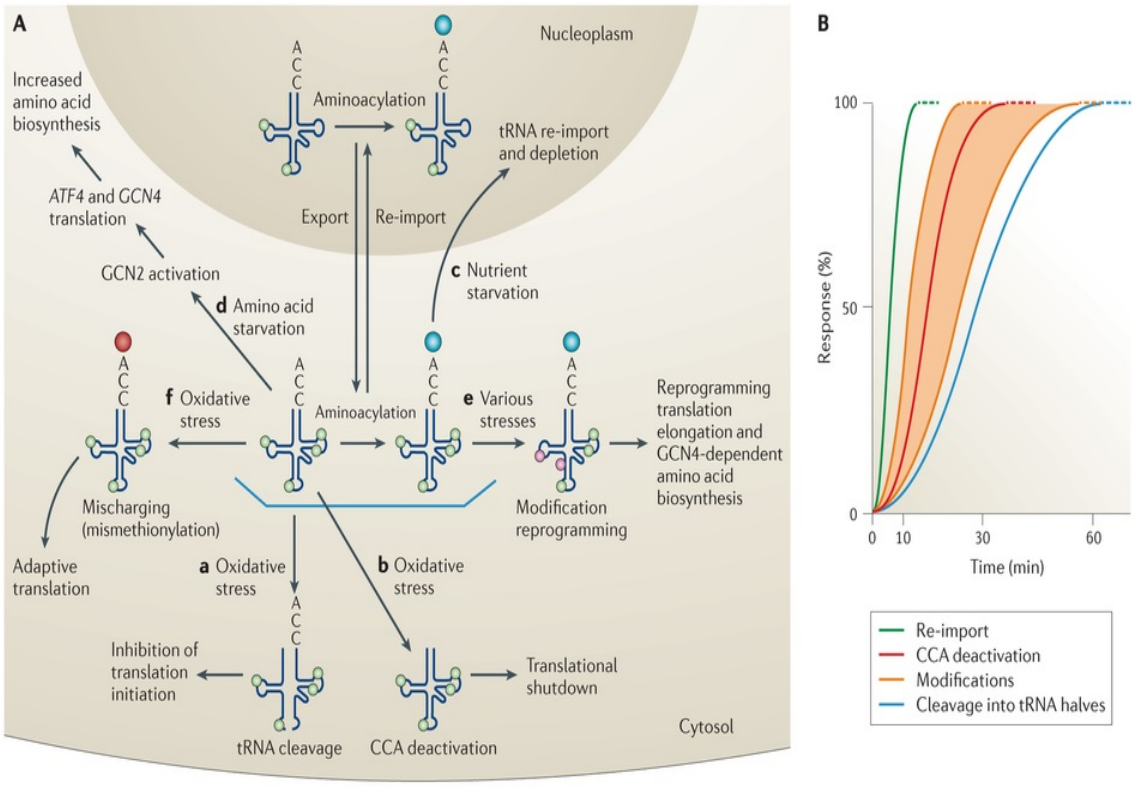

Figure 1.6: A map of eukaryotic tRNA regulatory pathways. The cell utilizes diverse pathways in order to regulate tRNA, and thereby regulate translation and extra-translational functions through tRNA. Source: [54]

From Figure 1.7 , we can see that many modifications shift under stress conditions including alkylating stress, oxidative stress, and chlorinating stress in yeast. Indeed in certain stress conditions, the Dedon group has been able to identify dynamic shifts in modification quantity in particular tRNAs. For instance, oxidative stress causes a shift in a particular wobble uridine modification, and this allows the cell to recode a particular tRNA's affinity for its cognate codon, shifting the efficiency of translation in this stress background[19]. This shift in translation creates a codon bias toward proteins whose mRNAs are enriched in this particular codon (Figure 1.7B)

Also notably, there are examples of writer proteins affording dynamic modulation of tRNA modifications. As was mentioned, tRNA fragments have been shown to be crucial

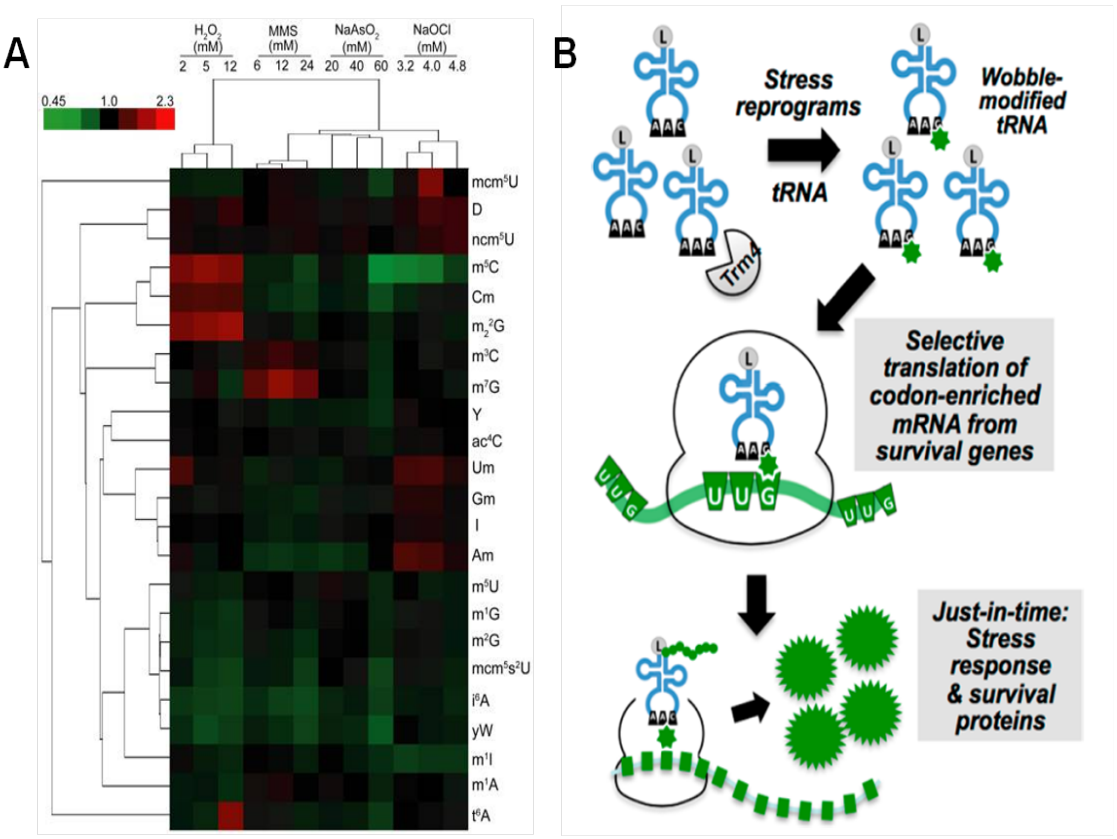

Figure 1.7: tRNA modifications are altered during various stresses. (A) Heat map of various known yeast tRNA modifications that shift during chemical stresses. (B) Schematic for stress reprogramming of translation. Source: [18]

mediators for signals of cell stress. However, the cell has a method to prevent cleavage of angiogenin substrates. This is accomplished by methylating certain tRNAs. In fact, in the presence of a 5-methylcytosine near the anticodon stem loop, these tRNAs are resistant to angiogenin cleavage, thus preserving the pool of translationally active tRNAs during stress conditions[11, 116]. The writer proteins for this methylation are Nsun2 and Dnmt2, and if a person is defective in either of these proteins, they are afflicted by an autosomal-recessive intellectual disability, further pointing to the necessity for this dynamic regulation of modification.

1.5 Perspectives and Summary

Continued advances in biochemical understanding of modifications will invariably lead to a greater understanding of the ever-complex regulatory networks in the cell. Every modification causes an inevitable cascade of small changes will slowly accumulate to an avalanche of dynamic interplay in the transcriptome. Even further, as technology related to high-throughput sequencing grows, the depth of coverage for even the low abundance molecules will be sufficient to ascertain quantitative measures of modifications. We will be able to gain insight into the true dynamic nature of the system with temporal certainty gained by high-throughput methodologies. The advent of bioinformatics and machine learning will likely make this an even more feasible undertaking - no longer will we require manual curation, which inherently contains flaws due to the human nature of the user. Discovery, identification, and validation will be a constantly updating task. The beauty of this has already been applied narrowly to a small subset of modifications, but once we understand the signatures and necessary inputs, we gain innumerable outputs and far wider scope of understanding in biology.

What makes the tRNA an appealing model for modification dynamics is purely that there are so many modifications to explore. If we assume that each methylation, isomerization, thiolation, or any combination of the above has a concerted system for being written and erased, then the tRNA is at the center of what must surely be chaos. As we have already mentioned, it is the physical link in translation between mRNA and protein, so it is already a particularly prominent molecule in that regard. We can imagine that each small fine-tuned adjustment per tRNA over the course of an entire mRNA can lead to vastly drastic overall changes in translational efficiency.

In this thesis, I will set out to explore how we discovered simply that there were multiple dynamic modifications in the cell - if these were static, then presumably all tRNAs would

contain roughly the same level of modification, and this would be reproducible between biological systems. Or, as the field believed for quite some time, that all tRNA modifications were fully present. As it turns out, this is not the case. First, we actually discovered a putative tRNA demethylase that works in mitochondrial tRNA. ALKBH1, the first discovered human homologue of the AlkB enzyme, turns out to bind specific tRNAs, all of which contain a common structural modification. We use several different methods to ascertain its localization and function. Because the mitochondrial tRNA is more sparsely modified, it is much easier to exhibit dynamic control over the tRNA. Using existing models for the purpose of the targeted modification, we offer a hypothesis for why the enzyme acts on mitochondrial tRNA.

Next, using the established biochemical tool of reverse transcriptase primer extension, we are able to ascertain that two modifications common to most tRNAs - 1-methyladenosine and 1-methylguanosine - are fractional and dynamic between isoacceptors and isodecoders. We are able to fine-tune this method in order to gain a quantitative analysis for modifications that can be utilized for any Watson-Crick face modification along the tRNA.

High-throughput sequencing changed the face of modification discovery and validation. We were able to harness its power and direct it toward the landscape of tRNA. This tool allowed us insight into many modifications in parallel as well as get a better sense of tRNA abundance. Coupled with engineered AlkB mutants, we were able to validate vital Watson-Crick methylations in tRNAs, and comparative analysis driven from this demethylase-directed search allowed us to discover other modifications by proxy via cDNA truncations and mutations. We adapted bioinformatic pipelines and data analysis toward tRNA modifications. This adaptation and subsequent analysis in the replicates of data uncovered many previously unknown sites of modification in tRNA. Using biochemical methods we were able to validate some of these sites. Furthermore, we were able to use modification sig-

natures in tRNA and ribosomal RNA to begin to construct signatures of modification for 1-methyladenosine.

Lastly and briefly, bioinformatic analysis of data from tRNA and also from pseudouridine analysis using several reverse transcriptases are also reported. The results of this experiment suggest that various modifications contain unique signatures that are dependent on reverse transcriptases and on sequence context.

CHAPTER 2

ABH1 - A PUTATIVE TRNA DEMETHYLASE

2.1 Introduction

tRNA molecules undergo rapid regulation under many mechanisms. The cell can upregulate the gene expression of a tRNA, affecting the available pool of tRNAs for translation or extra-translational function[54]. The cell can misacylate tRNAs to affect the link of genetic information from mRNA to protein[79]. Most important, however, for any of these functions, the tRNA must be modified, as many modifications are crucial to function and structure[103].

In order to maintain any modification dynamics, sites must be both written and erased. While many RNA modifications are installed and remain functional into the RNA is degraded, such as the m7G cap of mRNAs or 2'-O-methyl modifications in ribosomal RNA, much emphasis has been placed on the range of dynamic modifications and modification fraction of more recently discovered RNA modifications such as m6A and pseudouridine. While there is evidence for genetic regulation of these modification writers, this still does not affect the dynamic state of the modification. And as such, we turn to the eraser proteins.

As was mentioned previously, the known RNA modification eraser proteins all come from a particular family of homologs: the AlkB family. AlkB family proteins have shown to have a variety of activity on various cellular substrates, with the most prominent members performing oxidative demethylation of m6A[50, 130]. AlkB itself primarily acted on 3-methylcytosine and 1-methyladenosine[113], the latter being a prominent structural modification in both nuclear and mitochondrial tRNAs. While the E. coli variant of the protein did not seem to have any activity toward tRNAs in vivo, the enzyme showed demethylase activity in vitro[84]. It seemed likely that with many human homologues of the AlkB pro-

tein available, one of them would likely act on tRNAs. However, early work characterizing the first discovered homologues showed minimal activity toward RNA relative to the E. coli AlkB[76].

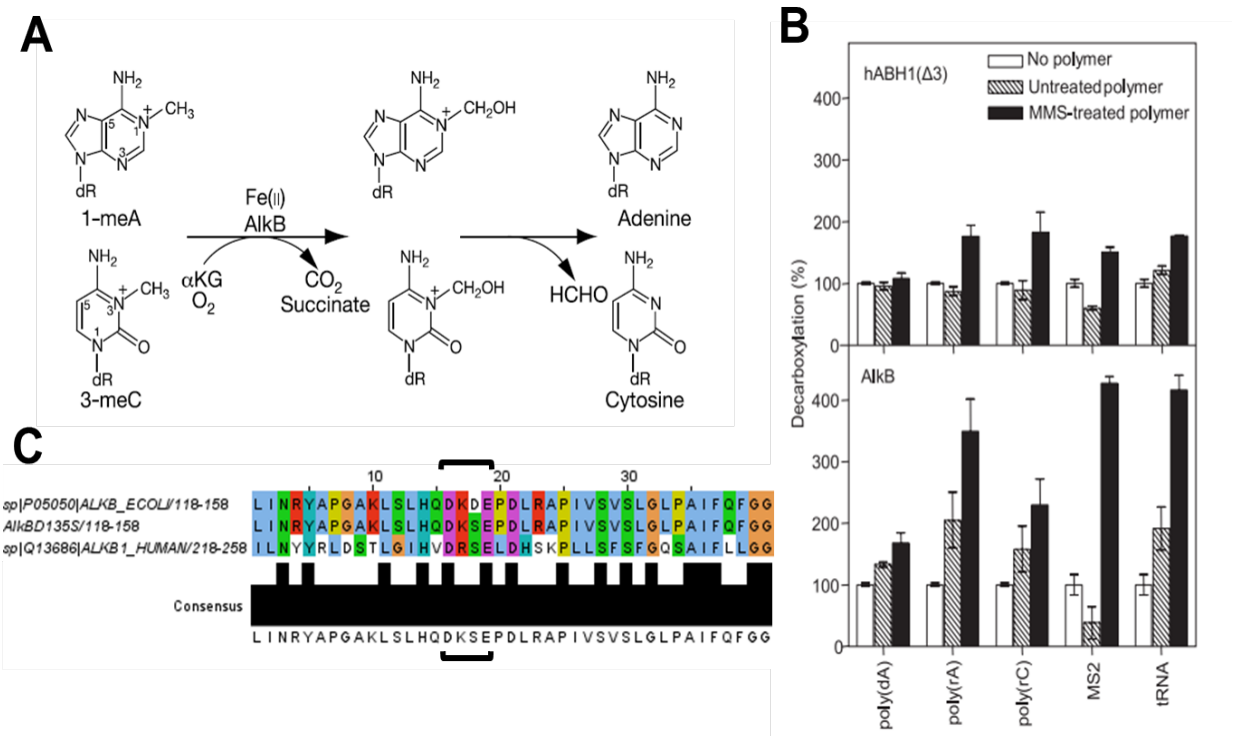

Figure 2.1: (A) Reaction of AlkB on its known methyl substrates. Using iron and oxoketoglutarate as substrates, the enzyme oxidatively demethylates 1-methyladenosine and 3-methylcytosine in ssDNA and ssRNA. (B) From source [122], human ABH1 activity compared to native E. coli AlkB enzyme on several common nucleic acid substrates. (C) Clustal analysis of AlkB, enzymatically promiscuous AlkBD135S, and ALKBH1, with the enzymatic active site highlighted by the black bracket.

However, later work characterized ABH1 by itself, and found that it had very little enzymatic activity at all to most available substrates, as shown in Figure 2.1. Moreover, one group found it to be mostly localized to the mitochondria, although ABH1 localization would become a contested matter between several different groups in different organisms and tissue types[81, 83, 115, 122]. However, what made the ABH1 protein appeal to our study was its similar active site to a particularly promiscuous mutant that we had discovered:

AlkBD135S, a change of aspartate to serine in the catalytic active site. This promiscuous mutant acted on both m^1A and m^3C, but also had markedly increased activity toward m^1G. The only molecule known to contain all of these modifications is tRNA. Thus, we returned to ABH1 to investigate its functional activity and its potential binding partners.

We find that human ABH1 cross-links to tRNA in vivo, but that it mainly crosslinks to mitochondrial tRNAs. There are 6 primary tRNA substrates of ABH1, all containing 1-methyladenosine at position 9. We find that ABH1 also seems to localize to the cytoplasm and the mitochondria, concomitant with the interaction partner result. However, despite the binding to these tRNAs, we find that disrupting ABH1 expression has minimal effect on m^1A9 modifications in its target tRNAs. Changes in ABH1 expression, however, seem to increase mitochondrial translation, and the disrupted expression levels have a protective phenotype to mitochondrial respiration under oxidative stress conditions.

2.2 Results

2.2.1 ABH1 UV-CLIP shows binding to m^1A containing mitochondrial tRNAs and nuclear-encoded tRNA AspGTC

As was mentioned before, little was known about ABH1 binding partners. It seemingly had weak activity against all common nucleic acid substrates in the decarboxylation assays used as a proxy for demethylase activity. However, perhaps there were alternative functions for ABH1, as not all members of the AlkB family primarily have demethylase activity. For instance, ABH8 is actually a modification enzyme for a particular wobble uridine modification. Thus, in order to understand its activity, we undertook an assay to find its RNA binding partners. We performed tRNA-CLIP, which is an adaptation of standard CLIP experiments using microarray technology to find tRNA binding partners.

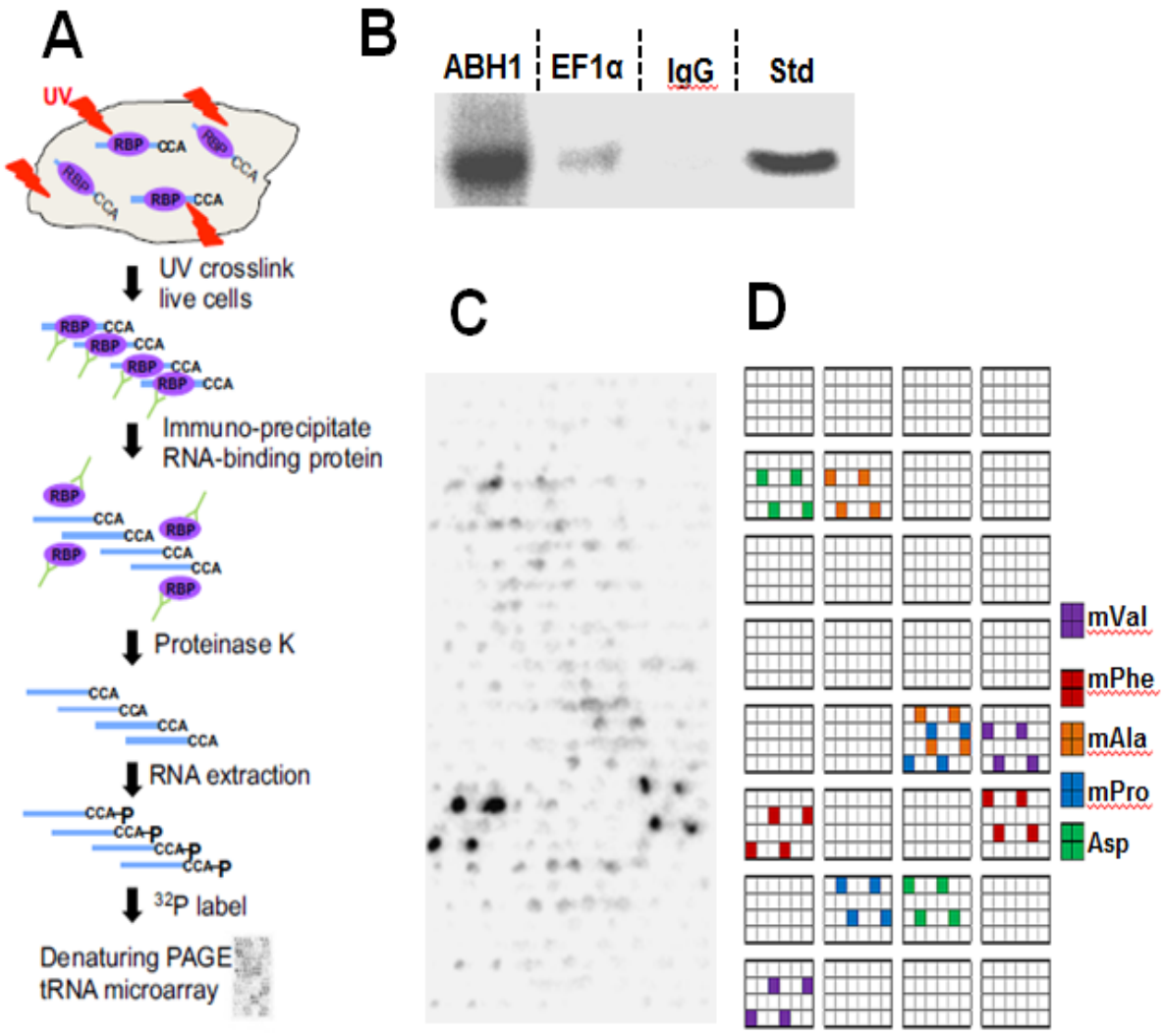

Figure 2.2: UV-tRNA-CLIP shows ABH1 binding to mitochondrial tRNA. (A) Schematic for tRNA-CLIP. Source: [88] (B) Denaturing PAGE bands for ABH1, our positive control EF-1α, our negative control of IgG, and a tRNAPhe loading control. (C and D) tRNA microarray with radioactive signal in (C) and the corresponding diagram for the high-intensity spots in (D).

Interestingly, binding signal only corresponds mostly to mitochondrial tRNA probes, as shown in the figure. The common characteristic between these 5 tRNAs is that they all contain 1-methyladenosine at position 9, a known structural modification for mitochondrial tRNAs. Also, there is signal for one nuclear tRNA, AspGTC. Relative to the positive control of elongation factor 1 alpha, the eukaryotic homologue of EF-Tu, there is enrichment for the

AspGTC and mitochondrial tRNAs.

2.2.2 ABH1 immunofluorescence and subcellular fractionation

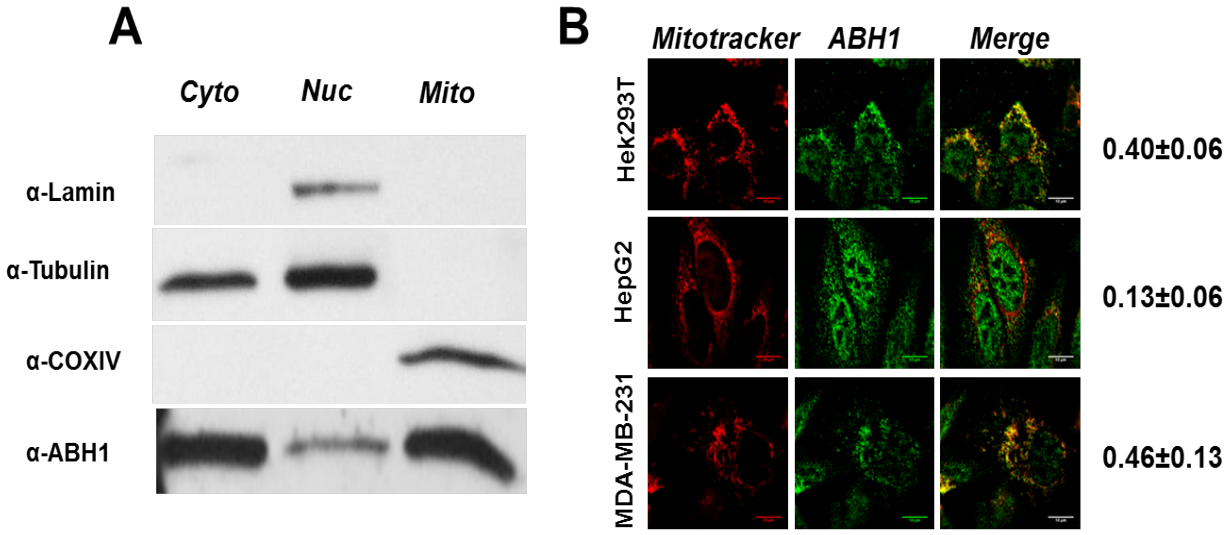

Figure 2.3: ABH1 localizes to the cytoplasm and mitochondria in HEK293T. (A) Subcellular fractionation of cell lysate followed by western blots shows that ABH1 localizes ubiquitously, but with strong signal to the cytoplasm and mitochondria. (B) Immunofluorescence staining for ABH1 and Mitotracker in 3 different tissue-specific cell lines. ABH1 localization to the mitochondria is represented as a fraction on the right side of the merge column.

The results for ABH1 tRNA substrates posed us with a question of ABH1 localization. First, we performed sub-cellular localization followed by western blotting (Figure 2.3A). Nuclear and cytoplasmic extracts were prepared separately, and then further differential centrifugation of the cytoplasmic extract, we were able to isolate and further lyse mitochondria. Western blots with compartment-specific markers show that ABH1 has a ubiquitous presence in the cytoplasm, nucleus, and mitochondria.

Interestingly, using various different predictors for localization, no consensus localization could be found. A cryptic mitochondrial localization appears to emerge after N-terminal truncation of amino acids from the primary polypeptide sequence. Thus, in order to verify,

we used immunofluorescence to identify sub-cellular localization in three different cell lines (Figure 2.3B). Interestingly, ABH1 signal could be find in both the cytoplasm as well as overlapping with mitochondrial signal depending on the cell type. This could be due to differential expression of the enzyme dependent on cell type.

2.2.3 ABH1 demethylates tRNA m^1A in vitro

Due to prior knowledge regarding engineering of AlkB, we were interesting in the demethylase activity of ABH1. As was shown in Figure 2.1, a standalone analysis using Basic Local Alignment Search Tool (BLAST) alignment of the engineered AlkB active site aligned against AlkB and ABH1. The change from the aspartate to the serine was originally what encouraged the promiscuous nature of the D135S enzyme, so naturally we were curious to see if this was preserved in ABH1. Thus in order to test ABH1 activity, we needed to express ABH1. However, E. coli expression of the protein proved not to be successful, so we instead utilized FLAG purification from human cell culture. This method, while not being so robust in expression, has the added benefit of preserving post-translational modifications of proteins. Because ABH1 is not so well-characterized, we were unsure if there were potentially activating modifications. ABH1 was spliced into a mammalian over-expression construct (graciously gifted to us from the Chuan He lab) containing a 5' terminal FLAG tag. Once transfected into HEK293T and overexpressed for 48 hours, the protein was purified from HEK293T lysate. Interestingly, residual mass shifts of approximately the weight of a tRNA can be observed in the SDS-PAGE used to confirm purification.

Next, we purified nuclear-encoded tRNA using common denaturing PAGE, followed by identifying tRNA using UV shadowing, cutting out bands of the appropriate size, and using the established crush and soak method to isolate tRNA. However, for mitochondrial tRNA, this will not work due to the variable size of mitochondrial tRNA and the much lower abundance relative to cytoplasmic tRNA. So, in order to isolate mitochondrial RNA, we

used a method of mitochondrial isolation involving paramagnetic beads conjugated to an antibody for mitochondrial outer membrane proteins. This allows us to isolate mitochondria without high centrifugation steps, which could potentially contaminate with sedimenting ribosomes and thus cytoplasmic tRNAs. Once mitochondria were purified, they were lysed, and mitochondrial RNA was isolated using acid phenol followed by ethanol precipitation.

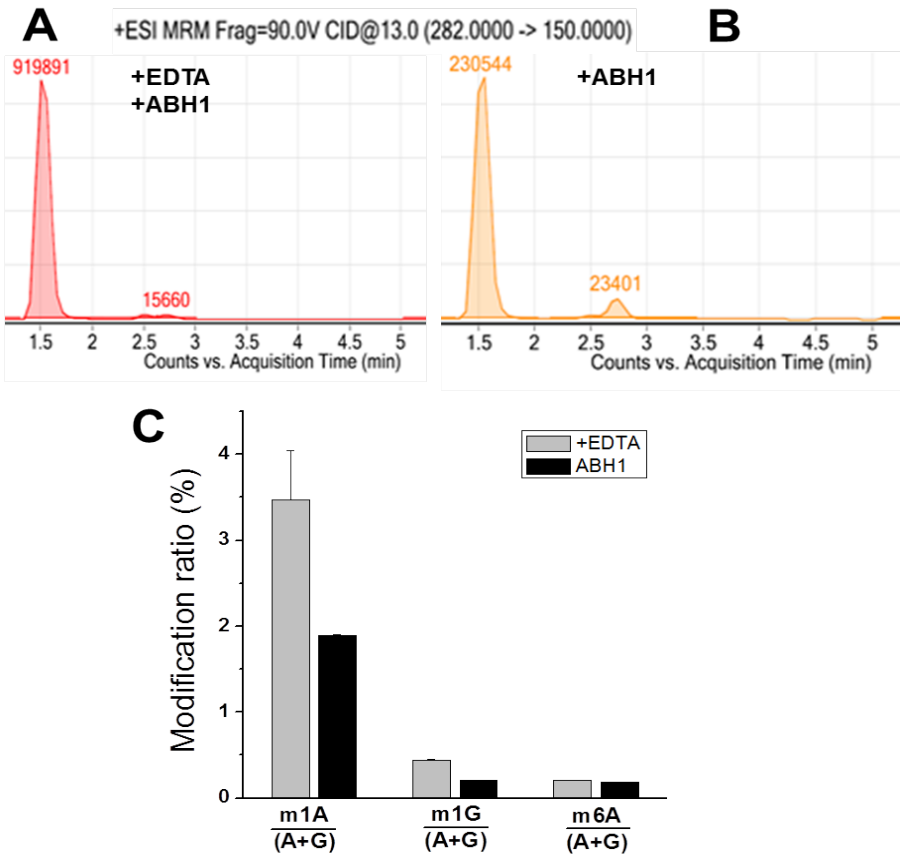

Figure 2.4: *In vitro* demethylase activity of ABH1 toward 1-methyladenosine in tRNA. (A,B) LC/MS-QQQ plots showing 1-methyladenosine levels in ABH1-treated tRNA. The fragmentation mass of 150 is the specific mass fragment generated by m^1A. EDTA chelates the iron in the reaction, thus inhibiting the ABH1 oxidative demethylation. (C) Ratio of specific RNA modifications from QQQ, showing preferential demethylase activity toward 1-methyladenosine in the ABH1 treated tRNA.

To test ABH1 demethylase activity, we incubated cytoplasmic tRNAs and mitochondrial

RNA with the purified ABH1-FLAG protein constructs in standard oxidative demethylase conditions for AlkB at a ratio of 5:1 protein to RNA. We used EDTA incubation as a negative control because AlkB oxidative demethylation is an iron-dependent reaction. After the demethylase assay, we prepared samples for LC/MS-QQQ with subsequent incubations with Nuclease P1 and Alkaline Phosphatase. Shown are the results in Figure 2.4, wherein ABH1 treatment on nuclear tRNA and mitochondrial RNA showed activity toward both m^1A and m^1G, with a strong preference of action toward m^1A.

2.2.4 ABH1 knockdown causes minimal effect on mitochondrial tRNA modifications in vivo

With ABH1 localization and function understood in vitro, we wanted to verify ABH1 activity in vivo. In order to do this, we performed transfection of an si-RNA to ABH1. Because the construct for overexpression contained an N-terminal expression tag, we figured this may impact possible mitochondrial localization, so we instead undertook the opposite direction for ABH1 expression perturbation. To this end, ABH1 knockdown should decrease demethylation on tRNA substrates, thus increasing methylation for 1-methyladenosine.

To investigate this interaction, we used the established method of reverse transcription primer extension. In theory, if the sites were 100% modified, we would see no readthrough. However, with ABH1 knockdown, with increased methylation, we would expect to see less primer extended through the modification, which we saw for mitochondrial tRNAPhe and mitochondrial tRNAAla. For other tRNAs, we saw very little readthrough, perhaps concomitant with a fully modified m^1A9.

2.2.5 ABH1 knockdown increases mitochondrial translation and increases respiration under stress

As the tRNA is the casual link in translation, and modifications are crucial for this function, we reasoned that there may be a resultant effect on translation under the effect of ABH1 expression. To test this, we pulse-labeled transfected 293T with S35-methionine to assay translation. To further explore mitochondrial translation, we treated cells with emetine, a translational inhibitor that affects cytoplasmic translation but will not bind mitochondrial ribosomes. Quantifying translation of the 13 mitochondrially encoded proteins, we find that mitochondrial translation increases nearly 1.3 fold over control cells, while there is a modest but insignificant increase in cytoplasmic translation.

With this effect on translation, we were curious of the phenomenological impact on the cells. Interestingly, many different assays showed little to no change. Knockdown of ABH1 did not affect cell proliferation, nor was there a change in mitochondrial lethality, assayed using membrane potential experiments. ROS generation remained constant in both cells, showing that there were not particular defects or subtle changes in mitochondrial health. Because the mitochondria's key function resides in cellular respiration, we thought to further perturb ABH1 knockdown cells with oxidative stressors such as arsenite and menadione. When we assayed cellular respiration under ABH1-knockdown and oxidative stress, we find that the cells function better under oxidative stress. Furthermore, they appear to produce more ATP under basal conditions and under oxidative stress.

2.3 Discussion

2.3.1 ABH1 acts on structurally similar modifications

As a member of the AlkB family, it was expected that ABH1 would have some affinity towards nucleic acids, due to the nature of the oxidative demethylation reaction. From the UV-

CLIP experiment, it shows that ABH1 preferentially binds mitochondrial tRNA, which all contain 1-methyladenosine at position 9. Critically, at the time of the experiment, the outlier binder was AspGTC, as it was known to be hypomodified for m^1A58, the common nuclear-encoded structural m^1A modification. We later understood that AspGTC also contains m^1A9, underlying a common structural motif that ABH1 recognizes and therefore binds. It must also be noted that the initial hypothesis for ABH1 was to bind 1-methylguanosine due to the similarity in active site, but seemingly it prefers this modification buried in a large structural context. This may also be why earlier experiments had such negative results, as the commonly tested substrates are all single-stranded RNA or DNA, which would not contain any noticeable structure.

2.3.2 ABH1 is a mitochondrially localized protein

Despite previous work on ABH1, endogenous ABH1 substrate and localization was contentious with reports on its action on 3-methylcytosine in mitochondria and histone demethylation in the nucleus. ABH1 knockout mice showed deleterious phenotypes in development and pachytene spermatogenesis [81], which is a stark difference from the phenotypes we observed in cell culture. However, we provide clear evidence illustrating that ABH1 is localized to the mitochondria to varying extents in different cell lines and acts on the methylations in mitochondrial tRNA. So where do the severe phenotypes in the knockout mice arise? A missing link may be the role of mitochondrial transcription and translation in spermatogenesis. Wai et al. offered a link between mitochondrial DNA copy number and mammalian fertility [119], where a decrease in mtDNA is necessary for mammalian fertility. Alcivar et al. showed that pachytene spermatocytes is accompanied with a decrease in mitochondrial RNA transcripts [3].

2.3.3 Rationalization of ABH1 model for translational effects

Why would the cell have a need for unmodified tRNA? In previous work, mitochondrial tRNALys transcripts lacking m^1A showed more accessibility to RNase cleavages and degradation. From this, we can extrapolate that perhaps the mitochondria uses ABH1's oxidative demethylation in order to tune the amount of translation that occurs. Thus, ABH1 demethylates tRNAs and they are subsequently degraded or sequestered.

Another point of mention is that tRNA sequencing shows wide variation in mitochondrial tRNA abundance. Mitochondrial tRNA genes all lie on three promoters for different polycistronic transcripts, so presumably tRNA genes lying along the promoter should all share equal abundance. This is not the case, and furthermore the first step of mitochondrial tRNA processing involves 5' end processing coupled with methylation, as the mitochondrial RNaseP is complexed with a 1-methyladenosine methyltranferase homologue. Thus, we assume that if a tRNA is processed, it is methylated. Therefore, the mitochondria may use ABH1 to fine-tune tRNA abundance immediately post-processing for degradation. However, we find that there exist hypomodified mitochondrial tRNAs, somewhat invalidating the idea that they are immediately degraded. Further experiments are necessary in order to understand the role and function of hypomodified tRNAs in the mitochondria, as well as characterizing the degradation enzymes in human mitochondria.

2.3.4 Minimal effect of ABH1 on cellular dynamics may be due to threshold effects

Lastly, we see very small effects despite large-scale perturbations of the ABH1 enzyme. In essence, if we are demethylating a position that should be modified, we are in essence causing a heteroplasmy, the presence of more than one genomic or transcriptomic allele, in the transcribed RNA. While this should have large effects, as evidenced often by the deletion

of modification enzymes in other organisms, the mitochondria appear to be more resistant to this phenomenological change. In fact, many mitochondrial diseases are undetectable in vitro and even in vivo with respect to health of the organism up until a certain percentage of heteroplasmic DNA in the host. This has been coined the mitochondrial threshold effect, indicating that after some critical point, only then do deleterious phenotypes appear.

We believe that this is likely the mechanism that prevents us from seeing large changes in phenotype for the ABH1 knockdowns. Because we are actually preventing demethylation, and thus preventing a theoretical degradation, it is especially likely that we are inducing a positive response, noting the increase in translation and the protective effect on cellular respiration. However, there may be deleterious effects unseen to us - perhaps without being able to tune translation, the long-term effects on protein folding of mitochondrial complexes is impacted. Seeing as how knockout mice are severely defective, it is likely that the mitochondria can tolerate a substantial loss in expression and still perform adequately in function. Since we are only assaying on short time scales and not robustly changing ABH1 expression, in the future, it is likely that a CRISPR knockout of ABH1 will be needed to assay the gamut of effects.

2.4 Materials and Methods

2.4.1 FLAG purification of ABH1

AlkBH1-FLAG plasmid was transfected using Lipofectamine 3000 using the manufacturers protocol. Cells were plated at 20% seeding density and harvested after 48 hours. AlkBH1-FLAG was purified from HEK293T cells overexpressing the AlkBH1-FLAG construct using Sigma FLAG M2 Affinity Gel. Purified protein was run on 4-20% Tris-Glycine gels and stained to examine for purity. Protein was quantified by taking A280 absorbance on Nanodrop.

2.4.2 Biochemical subcellular fractionation

For cytoplasmic and nuclear fractionation, 293T cells were harvested from a 15cm plate in 1mL PBS and pelleted at 500xg for 5 minutes. Fractionation was performed according to the protocol for NE-PER Cytoplasmic and Nuclear Extraction Reagents. Mitochondria were isolated using Miltenyi Biotec Mitochondria Isolation Kit. Briefly, 293T cells were harvested on ice in 10mL PBS and spun at 300xg for 10 minutes at 4C. The supernatant was aspirated, and cells were lysed using the lysis buffer provided. Cells were homogenized using a 26G needle, and the Separation Buffer and anti-TOM22 paramagnetic Micro Beads were incubated at 4C for 1 hour to label the mitochondria. Mitochondria were then separated on a MACS separator. Mitochondria were then lysed using Cellytic M lysis buffer.

2.4.3 Isolation of Nuclear tRNA and Mitochondrial RNA

Total RNA was isolated from HEK293T cells using the Trizol method. RNA was separated via 10% denaturing PAGE and bands corresponding to tRNA were excised from the gel. tRNA was eluted from the gel via the crush and soak method as previously described, resuspended in 50ul H2O, and stored at -20C. Mitochondrial RNA isolation was performed by by lysing mitochondria with Cellytic M followed by acid phenol/chloroform extraction.

2.4.4 In vitro demethylation

200 pmol of purified AlkBH1 protein was incubated with 40 pmol of either purified tRNA or mitochondrial RNA and the following demethylation buffer: 25 mM MES, pH 6, 100 mM KCl, 20mM MgCl2, 2mM Ascorbic Acid, 0.3 mM 2-ketoglutarate, 150 M of Fe2+, 1 U RNasin. 2 mM EDTA was added to samples as an inactive control. This solution was incubated at room temperature for 2 hours.

2.4.5 LC/MS-QQQ of ABH1-treated RNAs

100 ng of RNA were treated digested with Nuclease P1 in 100 mM ammonium acetate at 37C for 2 hours. 3 l of freshly prepared 1M ammonium carbonate was added with 1 U of Alkaline Phosphatase (Roche) at 37C for 2 hours. Mixture was centrifuged through a 0.22 m PVDF syringe filter (Millex-GV) to remove decontaminants. Remaining sample was then analyzed using an Agilent 6460 Triple Quad MS-MS with attached 1290 UHPLC.

2.4.6 UV-CLIP with tRNA array

The protocol for identification of protein-tRNA interactions was followed as described previously (ref.). Briefly, live HEK 293T cells were UV crosslinked. Pellets of crosslinked cells were lysed in 0.5 ml RIPA lysis buffer and freshly prepared protease inhibitor cocktail. Cell lysates were pre-cleared with Dynabeads protein A conjugated beads and then transferred to a tube containing conjugated Dynabead-antibody complexes for each antibody (EF1, IgG, AlkBH1) overnight at 4C and the supernatant was discarded. The beads were washed three times each in a high salt buffer and then a normal wash buffer. The sample was then treated with proteinase K followed by phenol/chloroform extraction with ethanol precipitation to isolate RNA. The RNA was then 3 32P-labeled using radioactive [32P]pCp and T4 RNA ligase. 2X loading dye was added and the mixture separated on 10% denaturing PAGE. Bands corresponding to tRNA were excised from the gel and eluted via crush and soak. tRNA microarray preparation and hybridization were carried out as described [88].

2.4.7 ABH1 Immunofluorescence

To determine the subcellular localization of the AlkBH1 protein, cells were seeded onto 1.5mm coverslips and stained with 0.5 M Mito-Tracker Deep Red for 2h. After fixation with PBS 3.7% paraformaldehyde cells were permeabilized using 0.25% Triton X. Incubation with 1:100 rabbit monoclonal anti-AlkBH1 (abcam) was done at 4C overnight and followed by

extensive washing with PBS. The Alexa Fluor 488 coupled secondary antibody was then added in a 1:1000 dilution for 1h at room temperature. Coverslips were mounted on microscope slides using SlowFade Gold antifade reagent and images captured using an Olympus DSU spinning disk confocal microscope with an Evolve EM-CCD camera. Co-localization was determined using the JACoP plugin for ImageJ.

2.4.8 S35 in vivo pulse labeling

This protocol was followed as described previously [100]. Briefly, cells were cultured as mentioned above in 6 well plates. 30 minutes prior to labeling, culture media was aspirated and DMEM without methionine and cysteine is added to cells. Chloramphenicol or emetine was added to a final concentration of 100g/ml to inhibit either whole cell translation or cytoplasmic translation. 200Ci of EXPRE35S35S Protein Labeling Mix, [35S]-, EasyTag (Perkin-Elmer) was used to label for 60 minutes. Cells were then chased with regular DMEM for 10 minutes, and washed 3 times in PBS before cells were harvested. Cells were lysed 100l in Cellytic M, and 250 U of Benzonase (Millipore) was added. Protein concentrations were determined via Bradford assay. 50 g of protein was suspended in loading dye and loaded on an SDS-PAGE gel. Gel was exposed for 48 hours and subsequently analyzed on a PMI Molecular Imager.

2.4.9 si-RNA Knockdown of ABH1

Cells were reverse transfected with 10nM of scrambled or AlkBH1 siRNA from Origene (GGUAUAAAGAAGCGACUAAACGGAG) using Lipofectamine RNAimax via the manufacturers protocol. When necessary, cells were treated with Sodium Arsenite and Menadione sodium bisulfite 36 hours after transfection and collected after 12h incubation.

2.4.10 SEAHORSE assay

Cells were seeded at 1.5x104 cells per well in a XF 96-well cell culture microplate (Seahorse Bioscience), transfected and grown for 48h. Before the assay the growth media was replaced with bicarbonate- free, serum-free XF Base medium supplemented with 10mM glucose, 4mM glutamine and 1mM pyruvate. The cell culture microplate was then was incubated for 1h at 37C in a non-CO2 incubator. Oxygen consumption rate (OCR) was measured in a XF96 Extracellular Flux Analyzer (Seahorse Bioscience) in accordance to manufacturers instructions. After measurement of baseline OCR cells were treated sequentially with 2M oligomycin, 1M FCCP and 0.5M Rotenone/antimycin, using the XF Cell Mito stress kit (Seahorse Bioscience). Cell numbers for data normalization were then determined using the CyQUANT Direct from Life Technologies.

CHAPTER 3

MANY TRNA METHYLATIONS ARE HYPOMODIFIED

3.1 Introduction

Processing of tRNA molecules is a many step process, beginning with transcription by a particular RNA polymerase Pol III. After transcription, the ends of the tRNA are processed by RNase P and RNaseZ at the 5' and 3' ends of the tRNA molecule respectively[93, 109]. Next, splicing occurs in a handful of tRNAs, typically between positions 37 and 38 of the mature tRNA. Despite only a handful of tRNAs containing introns, defects in the tRNA splicing endonuclease have been associated with the disease pontocerebellar hypoplasia[14]. Lastly, of course, the tRNA is modified at a myriad of positions, contributing to the final mature tRNA product. This pathway of tRNA processing is conserved from bacteria to eukaryotes, and each step is crucial to production of the mature tRNA. Abnormalities at any stage cause highly deleterious phenotypes.

Of course, all of the modifications are critical in their own right, and in yeast and higher eukaryotes, defects in the modification enzymes lead to growth defects. For instance, in the absence of A-to-I editing at position 34, yeast are inviable, and mice and humans have been associated with a wide range of diseases including cancer and autoimmune disorders[36, 93]. It is known that strains lacking modifications at the 37 position, the position +1 of the anticodon, are extremely sick due to likely frameshifting[43]. Similarly, missing most modifications cause growth phenotypes or other issues, whether they occur near the anticodon or within the body of the tRNA. Absence of the anticodon loop modifications would imply that there are issues with translation, so logically these phenotypes make sense. Lacking body tRNA modification likely imposes a flaw in the folding of the tRNA molecule or its stability, and as such contribute to the sickly phenotypes[92]. What's more is that the cell has a nuclear surveillance complex designed to degrade specific tRNAs that lack an m^1A

modification at position 58, which is conserved and present in most eukaryotic tRNAs[52]. tRNAs deficient in other modifications have also been shown to be degraded in various other degradation pathways dubbed the rapid tRNA decay pathway[56]. Given all of this evidence, one would assume that tRNAs are likely fully modified and the cell attempts to make hypomodified or unmodified tRNAs as scarce as possible.

That being said, there is evidence for hypomodifications in tRNA. Despite the nuclear surveillance pathways that exist for the m^1A at position 58, there are other tRNAs that do not appear to undergo the same quality control. A genomic tRNA microarray approach showed that certain species of tRNAs were hypomodified at position 58 for m^1A in various human cell culture lines[97]. Naturally knockdown of the methyltransferase showed a marked decrease in growth. This indicates that while the cell can tolerate some hypomodifications, the modification is still critical for translational function.

However, we were curious as to the fraction of hypomodification. The previous research could not give quantitative information in this regard, so we set out to design a quantitative primer extension protocol. Also, despite the cell's imperative to keep the position +1 of the anticodon modified, we were curious to see if all tRNAs were fully modified for the frameshift prevention modification. To this end, we developed a quantitative assay for determining fraction of modification for Watson-Crick modifications. We find that many 1-methyladenosine and 1-methylguanosines in humans are indeed hypomodified, and also that hypomodification is conserved also in bacteria as it is in humans.

3.2 Results

3.2.1 Development of method: quantitation of fractional modifications

While the previous microarray technology could detect readthroughs and therefore hypomodifications, it lacked the ability to quantify the fraction on the scale from 0 to1 . In order to do this, we adjusted the primer extension protocol to account for total quantitation of sample as well as fractional quantitation of the modification. By using di-deoxyNTPs (ddNTPs) in the total quantitation, we allow priming of all samples, extend one position by incorporating a radioactive nucleotide, and quench with a ddCTP or ddTTP at the m^1G or m^1A respectively. Then for fractional modification, we use a radioactive nucleotide complimentary to the m^1G or m^1A, in this case α-32P-CTP or α-32P-TTP allowing readthrough using Avian Myeloblastosis Virus (AMV) reverse transcriptase if the site is hypomodified. AMV has the added advantage of being more thermostable and well reported for use in primer extension studies showing arrest at these Watson-Crick face methylations.

Quantification of modification fraction is based on the signal from the gel at presumed bands. This is likely due to RT stops only appearing at sites of modification, and thus we can roughly estimate the number of radioactive nucleotides incorporated at each visible band on the gel.

3.2.2 1-methylguanosine hypomodifications

We investigated all known tRNAs with the +1 nucleotide from the anticodon loop containing a G. For tRNA isodecoders, we used the sequence from highest scoring tRNA from the isodecoder family for the hg19 dataset. We posited that if the site contained a G, it would be modified to a 1-methylguanosine in order to preserve fidelity of the tRNA.

A number of tRNAs show readthrough for m^1G37, as shown in Figure 3.1. Interestingly,

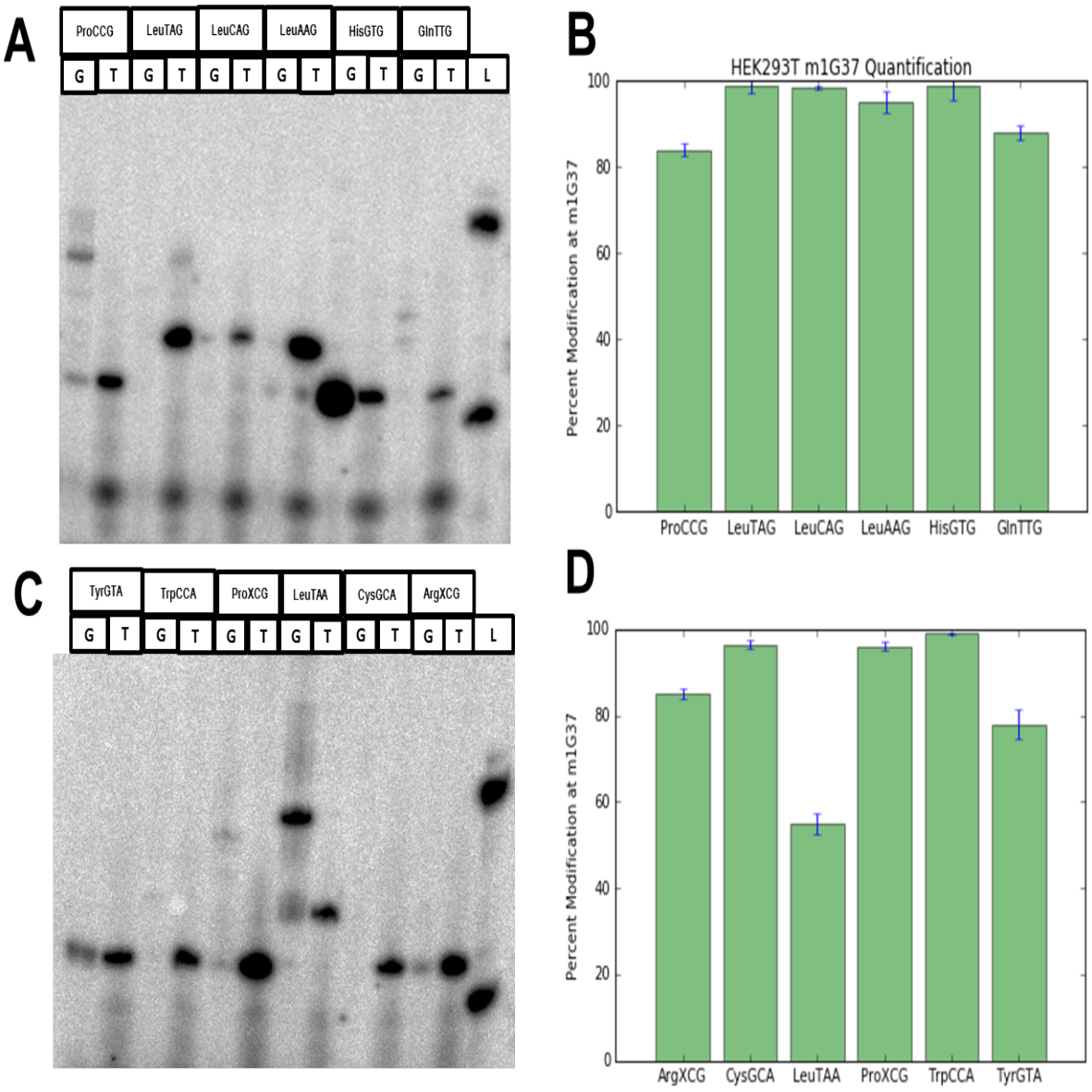

Figure 3.1: Quantitative reverse primer extension assay shows appreciable hypomodification for various isodecoder m^1G37. In (A) and (C), G stands for lanes where the reverse transcriptase incorporates a radioactive nucleotide complimentary to the hypomodified m^1G37 site, and T stands for the +1 incorporation followed by RT 'quenching' with dideoxynucleotide for the total isodecoder signal. (A,C) Various tRNA isodecoders containing 1-methylguanosine at position 37 show modest readthrough in the various G lanes. (B,D) Quantification of the readthrough as a percentage of modified m^1G37.

according to our assay, all but two appear to be modified greater than 75% of the time. The two which stand out for largely hypomodified are LeuTAA, a less common isoacceptor for

Leucine, and TyrGTA. Also, there appears to be quite a bit of signal from HisGTG in the readthrough lane. However, the fraction quantified clearly appears larger than the input sample. We deduce this may be a steric hindrance due to the likely queuosine modification that occurs at G34 and the incorporation of three radioactive nucleotides in succession.

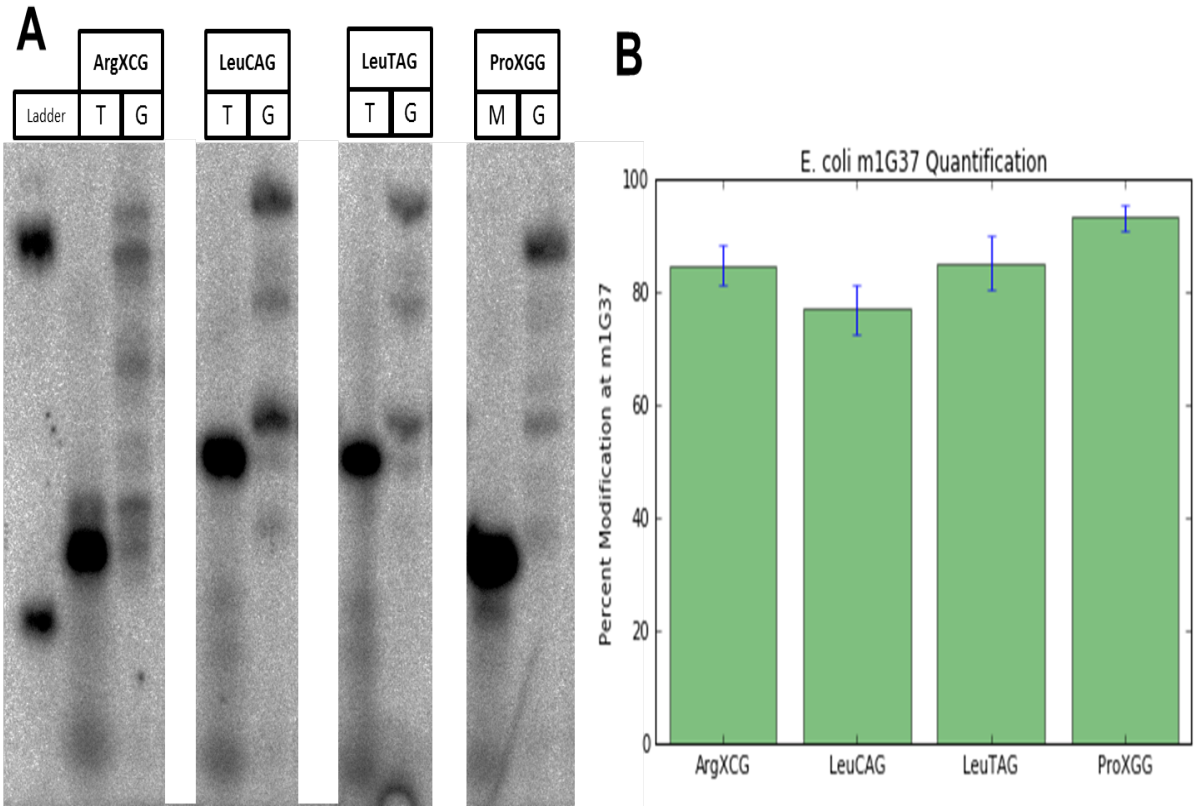

Figure 3.2: Quantitative reverse primer extension for *E. coli* tRNAs with a G +1 to the anticodon loop. Similar to the previous figure, T stands for total signal, and G stands for readthrough reaction. (A) The tRNAs show significant readthrough, but because this is the incorporation of multiple Gs, relative to the total signal, these sites also contain modest hypomodification. (B) Quantification of percent modification of m^1G37.

We also explored hypomodification in E. coli arginine, leucine, and proline isoaccepting families in Figure 3.2. There are very few +1 G containing tRNAs, and these three families contain the only known +1 G nucleotides. Also due to the lack of isodecoders, there was no need to select based on alignment score from the tRNA database. E. coli tRNAs appear to be much more hypomodified than their human counterparts. All tRNAs are at least twenty

percent hypomodified, with the LeuCAG isodecoder being the most hypomodified at roughly 55% modification.

3.2.3 1-methyladenosine hypomodifications

The number of tRNAs containing m^1A58 is much larger than the perceived number of m^1G37 containing tRNAs. Our strategy was to investigate tRNAs either with known quantifications, either through empirical knowledge or experimental validation. Thus, we chose initiator methionine tRNA, GluTTG, and GlyGCC. We report modification fractions for these tRNAs in Figure 3.3. Initiator methionine has known decay pathways for incompletely modified tRNAs, and as such it was our control for seeing how frequently AMV RT would read-through m^1A modified nucleotides. Glutamate tRNA has been shown previously to be hypomodified, and we report roughly 70% modification at the m^1A58. Lastly, GlyGCC is an interesting tRNA in that the tRNA is a known Dicer substrate at the 3' end. We were curious to know if there may be variable modification at the site, as that could modulate Dicer binding to the tRNA. We report that the Gly is also roughly 65% modified, indicating another hypomodified species.

3.3 Discussion

3.3.1 Hypomodification exists in bacterial and eukaryotic tRNA

There is not a lot understood about why the cell would hypomodify a tRNA or any molecule. However, the interesting facet to the result is that it seems to be conserved at least in bacteria and humans for m^1G37. Counterintuitively, a possible reason for this hypomodification at position 37 is to aid in frameshifting. Translational frameshifting is important for a small handful of proteins, and while it seems odd that a frameshift prevention modification would be hypomodified to ensure small latent amounts of frameshifting, there is small circumstantial evidence for this. The first known instance of +1 frameshifting occurs

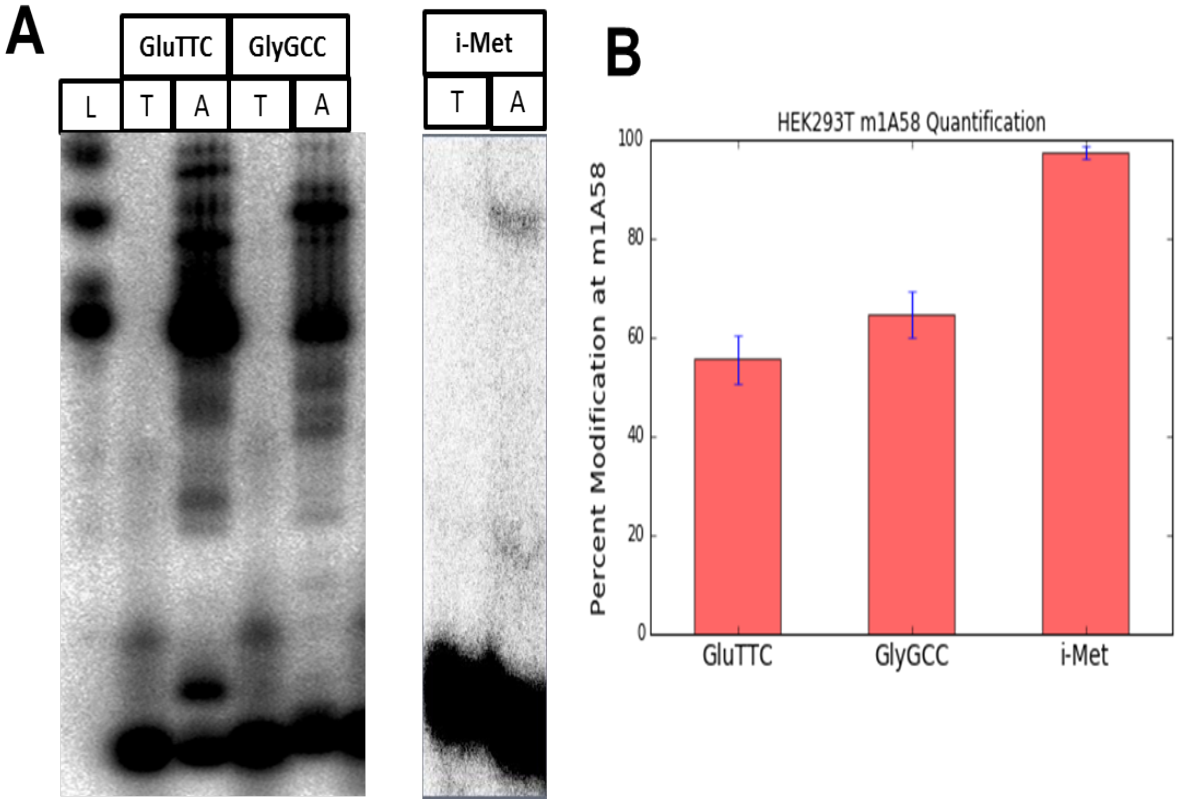

Figure 3.3: UV-tRNA-CLIP shows ABH1 binding to mitochondrial tRNA. As before, T stands for total signal, whereas A stands for readthrough due to a hypomodified A at m^1A58. (A) Quantitative RT primer extension used to validate m^1A58 positions with previously known hypomodifications, as well as initiator methionine for the positive control. (B) Quantification of percent modification at position 58 for m^1A.

in the prfB gene of E. coli, which produces the peptide chain release factor 2 (RF2). In the event of frameshifting, the slippage codon is actually decoded by a tRNA-Leu in order to allow for the +1 frameshifted gene product to be made.

Hypomodification at position 58 is even more of a mystery. The nuanced effect of m^1A58 in cells is only noticed when the methyltransferase is deleted. The cell perishes due to the degradation of hypomodified tRNA-iMet, but overexpression of the unmodified transcript restores viability. However, other tRNAs that contain the modification are not at all affected by the deletion. The hypomodified tRNAs are still associated with polysomes and are

seemingly used in translation, indicating that it may purely be a housekeeping modification for surveillance. As was mentioned before, there is the curious case of the GlyGCC tRNA that is a Dicer substrate. Appearance of the unmodified tRNA arm as well as the modified tRNA arm appear in certain sequencing data sets for B Cells, where the microRNA can be found. Thus, the fine-tuning of individual tRNA's hypomodification may serve crucial functions that are yet to be explored.

3.3.2 Isodecoder convolution in human tRNA

One confounding factor with the reported hypomodification quantitations is the issue of isodecoders. For most eukaryotic tRNAs, every codon has 1-3 tRNAs on average that can decode it. However, human tRNAs have a reported 625 tRNA genes, which including very low scoring genes and pseudogenes. Even removing duplicate scoring tRNA genes that occur on different chromosomes, this still only lowers the number to 462. That number means that every codon has on average 7.5 tRNAs that can decode it. For most of these species, there are significant enough differences in sequence to tell them apart.

However, for the sake of the primer extension experiments, there exist tRNA genes that contain only one nucleotide difference between them. Other tRNAs contain differences only in the 5' half of the tRNA, meaning that our primer sequences do not encounter the differences in sequence. Shown is an alignment for elongator methionine in Figure 3.4, illustrating the subtle differences in many isodecoder genes.

While our experiments provide evidence for hypomodification at these modification sites, without the aid of sequencing, it would be impossible to deconvolute the signal we see on the radioactive gels. Furthermore, this could skew quantitation: perhaps two isodecoders are primed by our oligonucleotide. One isodecoder maintains 100% modification, while the other isodecoder only maintains 50%. Our gel would show an ensemble average of 75%,

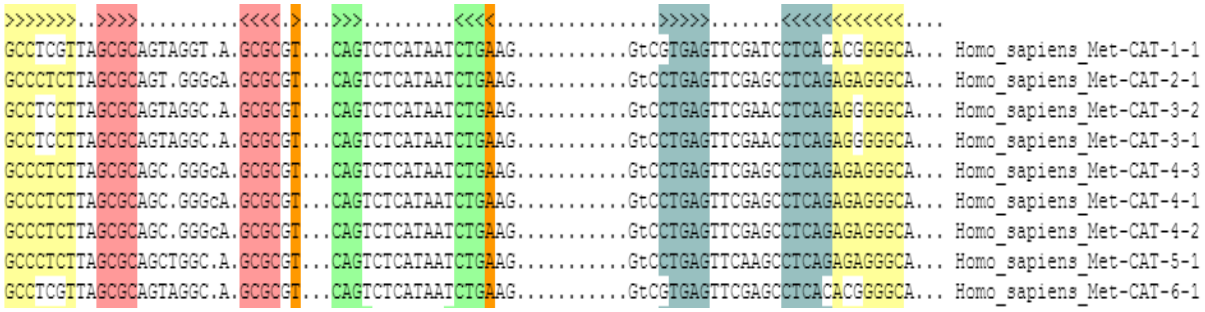

Figure 3.4: Alignment for elongator methionine tRNA genes, displaying the subtle diversity in isodecoder genes. These small differences are impossible to discern via the quantitative RT primer extension. Source: [21]

and in this hypothetical experiment we would only be able to discern that average. We can gain information about isoacceptor hypomodification, but that is the limit of the method. It still provides significant improvement to the microarray method for quantitation, and even identification, as many isodecoder families are clustered together with a doped sequence for the hybridized oligonucleotides.

3.4 Materials and Methods

3.4.1 DNA oligonucleotides and chemicals

All DNA oligonucleotides were purchased from IDT and based on tRNA sequences from tRNADB. For m^1G37 oligos, the oligonucleotide was complimentary for positions 39-76, and for m^1A58 oligos, the oligonucleotide was complimentary from positions 60-76 of the tRNA. RNaseH was purchased from Epicenter Biotechnologies. Proteinase K was purchased from Epicenter Biotechnologies. Ultra-pure dNTPs and ddNTPs were purchased from Affymetrix.

3.4.2 RT Primer Extension Strategy for total and fractional Quantifications

The protocol for primer extension was followed as described (ref.). Briefly, 100ng of gel purified tRNA were incubated with 1l of 5M of primer for respective tRNA (for mitochondrial

tRNA, primer is complementary to nucleotides 11-76) and annealing buffer and heated to 95C for 3 minutes and then cooled to RT. A 2.5X reverse transcription master mix containing 32P-dCTP and ddATP/ddGTP/ddTTP mixture (m^1A+A lane) dNTP mixture (A lane) was added. The mixture was incubated at 42C for 30 minutes. 5U RNase H was added for 5 minutes. 2X PAGE loading buffer was added and samples were separated on a 10% denaturing PAGE. Subsequent bands were analyzed on a PMI.

CHAPTER 4

TRNA-SEQUENCING AS A METHOD TO INVESTIGATE TRNA DYNAMICS

4.1 Introduction

High-throughput RNA sequencing revolutionized our understanding of gene expression and RNA biology. Because of RNA's role as the mediator between the genome and the proteome, RNA-sequencing (RNA-seq) became an invaluable tool for discovering transcripts and quantifying gene expression. Beyond the standard toolkit of RNA-seq, the technique can be adapted to a myriad of techniques, including immunoprecipitation in order to understand RNA-protein interactions, or chemical treatment with particular structural endonucleases in order to ascertain transcriptome-wide RNA structure [126, 34, 59, 120, 101, 104]. Naturally, the reconstruction of the transcriptome using bioinformatic methods of analysis is just as crucial as the design of the experiment, but the entire RNA-seq pipeline affords molecular biologists the ability to investigate many aspects of RNA biology in parallel and with great depth and breadth [58, 112].

However, native RNA-seq is not without its limitations, and the biology of the tRNA is almost contrived to be its foil. Because of the extensive secondary and tertiary structure of tRNA, most reverse transcriptases are unable to melt and traverse tRNA. In order to unfold the tRNA, extreme heat and depletion of cations are often required, but both inhibit the natural processivity of commercially available reverse transcriptases [85].

The other obvious limitation to sequencing RNA is the landscape of modifications which cover its surface. Most of these modifications are in the Watson-Crick face, and as such often cause arrest of reverse transcriptases [78, 93]. Even if tRNAs are hypomodified at some stochastic rate, it would be nearly impossible to push through the many roadblocks of

the tRNA sequence. There had been one instance of a group explicitly trying to sequence tRNAs by exploiting the lack of modifications at the 3' end, however their overall coverage is low[87]. This failure means that they are unable to capture the full depth of quantitation, nor are they able to understand the full scope of the modification landscape.

To this end, we developed tRNA-sequencing using approaches to both of these inherent obstacles to RNA-sequencing. This strategy is outlined schematically in Figure 4.1 [131]. First, we utilized the AlkB enzyme, the known aforementioned modification eraser. We were able to exploit the geometry of the active site to develop a more promiscuous mutant (AlkBD135S) that tackled not only the known substrates of m^1A and m^3C, but also was able to demethylate m^1G. Furthermore, we utilized the thermostable group II intron reverse transcriptase (TGIRT). This allowed for the production of cDNA at high enough temperatures to melt the tRNA structure.

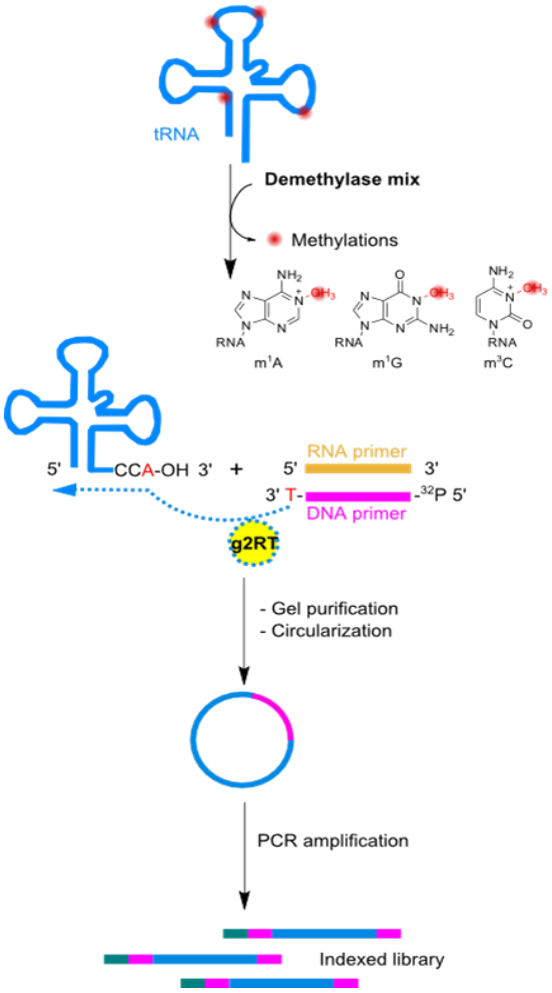

Figure 4.1: Source: [131]. Schematic for generation of libraries for tRNA-sequencing, which uses a demethylase mixture coupled with the thermophilic RT TGIRT in order to break down the structure and modified landscape of tRNA.

We find that using these two accessories to RNA-sequencing in tandem we are able to generate far more full-length tRNA reads compared to technical replicates without AlkB treatment. We find that many sites of modification are demethylated entirely by

the demethylase cocktail, although there are still positions that remain modified. We find that relative to previous results using microarray technology to quantify tRNA abundance, which some families of tRNA isoacceptors correlate well.

4.2 Results

4.2.1 Demethylase treatment leads to an increase in full-length tRNA reads

We first wanted to assess the status of full-length tRNA reads. We reasoned that without full-length tRNA reads, isodecoder deconvolution would be an issue, as there are several tRNA species with subtle changes in the 5' half of the tRNA. We were also curious as to how m^1A58 would impact the assignment of reads during alignment, because the 15-nucleotide stretch prior to the modification may not be sufficiently different in many species to differentiate which tRNA the cDNA read comes from.

We performed Illumina sequencing on HEK293T tRNAs with and without demethylase treatment using both gel-purified tRNAs and total RNA. In order to properly align the reads, we first needed to cull the existing database of tRNA genes. We removed same-scoring tRNA genes as well as pseudogenes. The former would muddle the algorithm for multiple alignment, and we presume the latter should not have tRNAs mapped to them because we are specifically gel-purifying tRNA from other cellular RNAs. After mapping, we see that there is an interesting many fold change across tRNA read distribution, and even more curiously we find that there exists a bi-modal distribution of reads.

In order to illustrate the change in full-length reads, we examined the distribution of cDNA truncations along tRNAs. Therefore, for each position, we can see what fraction of reads stops. Stochastically, we should expect roughly 2-2.5% across the tRNA, since we cannot have stops shorter than 18 nucleotides long due to raw read processing. We find that demethylase treatment significantly increases full-length reads in the LeuAAG and

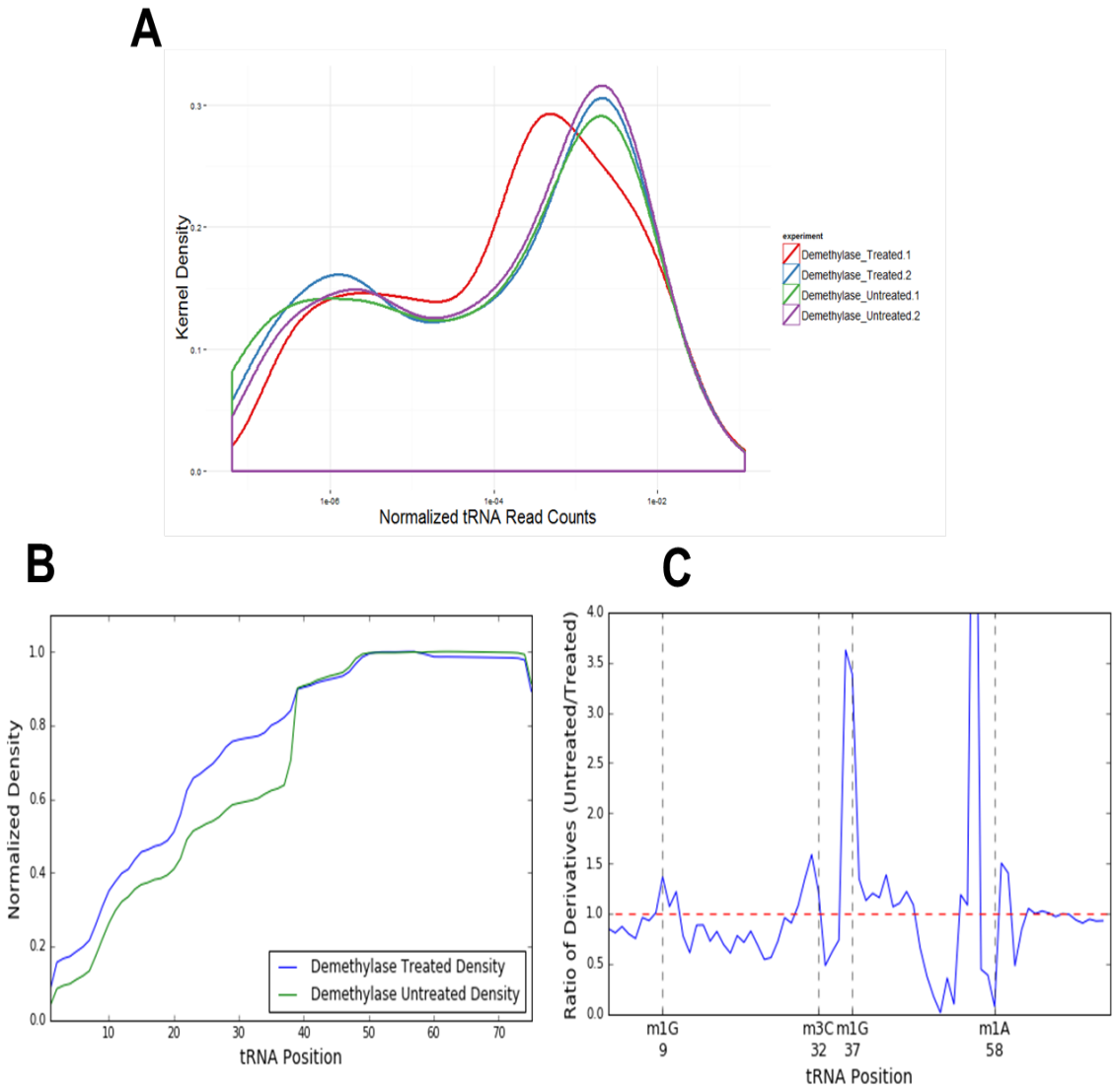

Figure 4.2: Interesting features of the tRNA sequencing data.. (A) Kernel density plot for the replicate data, showing a bi-modal distribution for tRNA normalized isodecoder counts. This means likely that there not only can we find sequencing reads for low-abundant tRNAs, but that the cell keeps these tRNAs at a low-abundance likely for some extra-translational function. (B) Cumulative normalized density for type I tRNAs, showing the number of reads covering all of the position in tRNA for each set of data. (C) Ratio of derivatives for the cumulative density plot. Slopes greater than 1 show where the slope of the untreated sample is larger, meaning more of a decrease in density. These positions where slope ¿ 1 corresponds with known methylation sites in tRNA.

GlnCTG shown in Figure 4.3. The LeuAAG also demonstrates that TGIRT is still somewhat processive despite encountering roadblocks, for in both the treated and untreated samples,

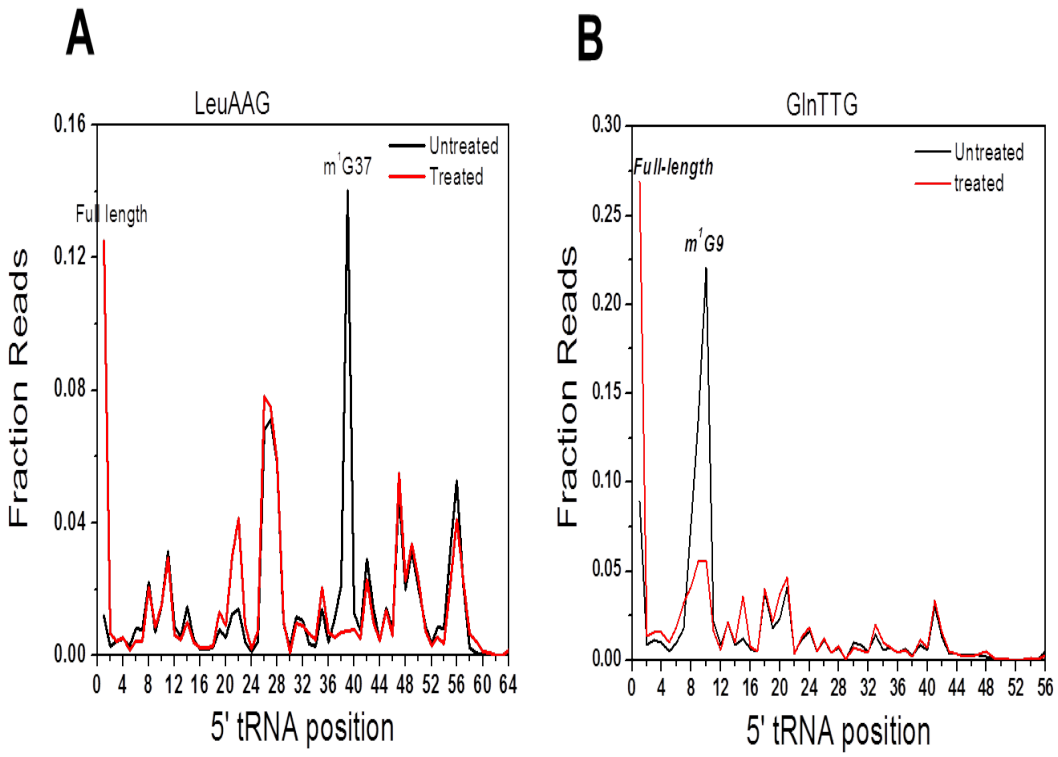

Figure 4.3: Representative tRNA isodecoders showcasing the efficacy of the demethylase treatment for tRNA-sequencing. (A,B) Plot of the fraction of reads whose 5' end aligns with the tRNA position. For instance, if the cDNA read is a full-length read, its 5' tRNA position is position 1. (A) LeuAAG contains a m^1G37 signal, which disappears in the treated sample, leading to a significant increase in full-length reads. There are a number of other modification roadblocks which we assume to be a result of di-methylguanosine and other unidentified sites. (B) GlnTTG isodecoder contains only m^1G9, the signal of which decreases significantly in the treated sample, leading to nearly 30% of the reads for this isodecoder to be full-length.

it traverses through the known $m^2{}_2G$ modification at position 26. The stark difference in full length reads for the LeuAAG is observed due to removing the m^1G37 which had caused considerable RT arrest and created cDNA truncations at position 38. Similarly, the same result appears for the GlnTTG tRNA. Once the m^1G9 is demethylated, there exist even more full-length reads. The GlnTTG is a rather sparsely modified tRNA relative to the average, and for tRNAs containing only one or two modifications, we can see a large increase in full-length coverage across the tRNA gene.

4.2.2 Many fold variation in tRNA abundances tracks well with previous experimental results

As expected, there is a large variation in tRNA abundance between isodecoders and even between tRNA isoaccepting families. In Figure 4.4, we compared isoacceptor tRNA count fractions in both treated and untreated samples relative to the isoacceptor gene copy number. There is a modest correlation, which is consistent with tissue-specific tRNA gene expression in humans. There are still a few fair outliers which exist in both the treated and untreated samples, especially with the Methionine and Valine fractions which lie far above the expected regression line.

We were curious if within isoaccepting families we were capturing the experimental distribution of tRNAs within a family. To do this, we used existing fluorescence signal data for tRNA abundance, a previously established method for examining and estimating tRNA abundance. We chose to examine the Arginine isoaccepting family, since it is a six box tRNA (meaning that it can decode six different codons) and has noticeable distinction in fluorescence array signal for the family members. For the more highly abundant species in the fluorescence, the demethylase fraction tracks well.

4.2.3 Demethylation of tRNA samples allows for identification of various modifications

There are many annotations that exist for tRNA modifications, due in part either to organismal homology, such as mouse and bovine samples, or to experimental validations using large-scale column purifications followed by fragmentation and HPLC. Most tRNAs contain m^1A58, and tRNAs that contain a G +1 from the anticodon are likely modified to m^1G. In vitro the enzymatic cocktail was able to remove a large fraction of the methylation from the respective nucleoside. The use of examining demethylase untreated data compared

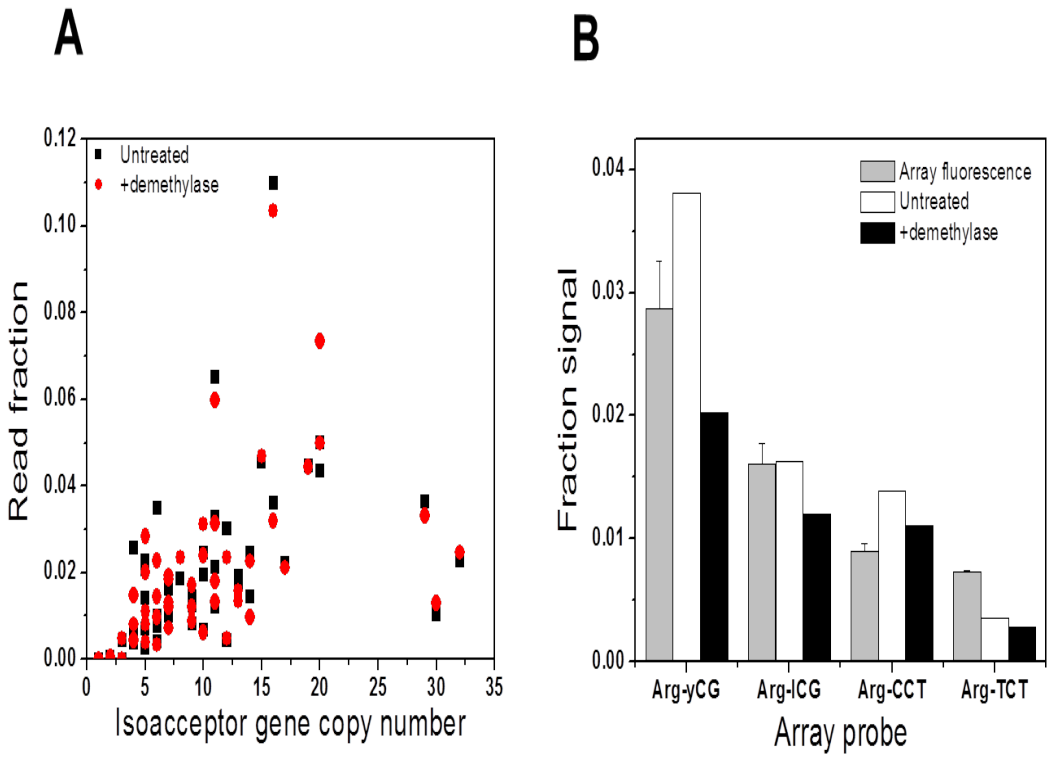

Figure 4.4: tRNA sequencing data shows that isoacceptor features that differ from conventional tRNA knowledge (A) Plot of isoacceptor gene copy number from various tRNA gene containing locus versus the read fraction for each isoacceptor. There is a modest correlation, but there exist heavy outliers such as methionine and valine. (B) Fraction of sequencing data signal versus established fluorescent array tRNA abundance results, showing that tRNA-seq is able to capture more depth with respect to the fluorescent arrays. However, the signal still tracks well with the sequencing results.

side-by-side with the demethylase treated data is apparent when we examine known sites of modification. As such, we would expect from the sequencing data to see a conversion of the cDNA truncations and nucleotide misincorporations to disappear, so we would be left with proper nucleotide incorporation with respect to the gene template.

We can detect these by examining the absence of mismatch in our demethylase treated samples. In Figure 4.5, we notice that for m^1A58 and m^3C in various tRNAs, all mismatches and stops disappear in the demethylase treated samples. Signal for m^1G in both

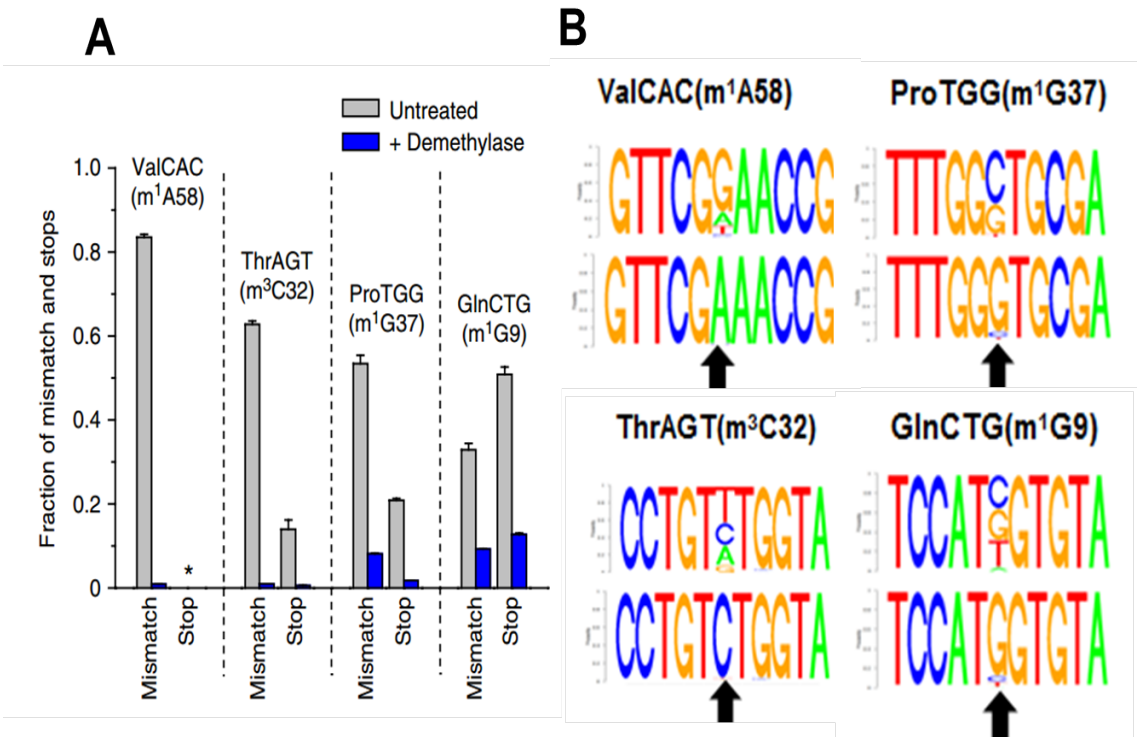

Figure 4.5: tRNA sequencing gives us site-resolved information about modification fractions. (A) Plot showing specific isodecoder modifications for untreated and demethylase treated samples. Concomitant with known AlkB substrates, the mismatch disappears for m^1A and m^3C, and the mutant is able to ablate a significant amount of m^1G signal. (B) Sequence logos for the respective tRNA isodecoders.

the anticodon stem loop and m^1G9 near the acceptor stem loop are both only partially demethylated. Furthermore, looking purely at reconstructing the sequence in our sequence logo, we spot the presence of heavy misincorporations relative to the native sequence. The reduction in mismatch signal also indicates that we are not dealing with polymorphisms or misalignment.

4.3 Discussion

4.3.1 Comparative analysis strategy

Strictly speaking, there is often a lack of a control when dealing with bioinformatics and stand-alone sequencing data. There are often many confounding factors in the singular

data set, and a point of reference needs to be established. This is why ribosome profiling relies heavily on an accompanying mRNA-seq; why immunoprecipitations require an IgG antibody control; or, why one uses an input RNA relative to some enrichment step. The hope for these is to separate the signal from the noise, although there are not necessarily clear-cut standards for bioinformatics at the current stage. Needless to say, there are many confounding factors researchers would need to control for, but they are often quite beyond the scope of the experiment and the experimenter at any rate.

Our strategy for understanding sequencing is necessarily no different. We change two 'variables' from normal sequencing: first, we use a thermostable reverse transcriptase; and, second, we use the demethylase cocktail. Trying to control around the first variable may potentially be difficult, as there has already been considerable trouble sequencing around the highly structured tRNA. Instead, the use of the demethylase cocktail enables the reverse transcriptase to process farther along the tRNA, despite its own latent ability to read through modifications to some degree, as we notice with modifications like $m^2{}_2G$. We can determine differences in modification dynamics by looking at these sites of incomplete conversion at single-base resolution.

There is an interesting phenomenon that we do spot from the comparisons between samples. Certain isodecoders are much more abundant in the demethylase treated samples, which is likely due to the multiple hurdles beset on the tRNA by many Watson-Crick modifications. Thus, in order to ascertain abundance, demethylase treatment is necessary, and this is reflected in the relation of fluorescence data to sequencing data for the Arginine iso-family.

By comparing tRNAs from demethylase untreated versus demethylase treated, we are able to piece together the full landscape for the tRNA. We are able to view differences in modification for m^1A, m^1G, and m^3C with confidence at single-base resolution while also

determining what appears to be modification dynamics. What also aids in our search is that modifications are rarely 100% from the sequencing - possibly a limitation in that TGIRT is powerful enough to properly incorporate regardless of being fully modified, or a symptom of hypomodification.

4.3.2 tRNA database scoring does not reflect species' abundance

Our expectations using the annotated database of human tRNAs was that the highest scoring tRNA genes would be the isodecoders we would expect to find the most from codon family to codon family. Surprisingly, this was not always reflected. The scoring for tRNAscan-SE involves comparison against covariation from E. coli tRNA genes as well as structural comparison against previously found tRNAs and analysis of tRNA transcriptional regions in the genome[68]. The end result is a list of tRNA genes that fold properly similarly to tRNA genes as well as demonstrating covariation from known tRNAs. The scoring is then broken down into regimes, wherein a score greater than 60 involves a tRNA likely used in translation, and scores less than 60 as having extra-translational function but not necessarily upkept in the translational machinery.

We find that often the highest scoring tRNA from tRNAscan-SE is not the most abundant tRNA by our sequencing data, and it is reflected in the technical replicates and the biological replicates. We do not find that the lower-scoring extra-translational tRNAs are ever the most abundant, but often in certain species they contain a significant number of read counts. It is possible that we have stumbled upon tRNA genes that are much more prevalent than previously expected considering gene copy and previous bioinformatic methods.

One further point regarding tRNAscan scoring. The comparative analysis shows that the most abundant tRNAs in the demethylase untreated sample as well do not always contain fully modified sites, perhaps elucidating that even the most used tRNAs in cellular machinery

still undergo the dynamics of modification that threads together the research thus far.

4.4 Methods

4.4.1 Sequencing Alignment

All reads were sequenced on an Illumina HiSeq 2000 with paired-end mapping with read lengths of 100 base pairs. Standard quality control via FastQC was performed after sequencing and after subsequent trimming mentioned next. Sequencing reads were aligned using Bowtie to a modified tRNA genome file containing nuclear-encoded tRNAs, mitochondrially-encoded tRNAs and human rRNAs (found at the GEO accession for this paper). Splicing of tRNAs was accounted for in the modified genome file and only mature tRNAs were aligned against. Briefly, the tRNA library was adapted from the tRNAScan-SE library ((http://gtrnadb.ucsc.edu/Hsapi19/,[17]) by appending 3CCA to tRNAs from the genomic tRNA database. Isodecoders with identical SEscan scores from the genomic tRNA database were consolidated for ease of identity assignment, so the total number of annotated tRNA genes and pseudogenes were reduced from 625 to 462. Prior to mapping, reads were processed using Trimmomatic v0.32 as well as further trimming using custom Python scripts to remove any further artifacts from demultiplexing and removal of primers, adapters, or any other low quality sequences. Sequences greater than 15 base pairs were then aligned to the library using Bowtie 1.0 with sensitive options using the highest allowed mismatch settings for Bowtie 1.0. Mapping to all references occurred simultaneously. Only one alignment (best with Bowtie1 or k=1 with Bowtie2, k refers to the number of allowed distinct alignments per raw read) declared as valid by the respective mapping software was reported for each read. Analysis was also performed using k = 3 to allow for up to 3 alignments per read and also v = 0,1,2,3 (how many mismatches allowed in the seed sequence per raw read) in Bowtie 1 to allow for fewer mismatches.

CHAPTER 5

USING TRNA-SEQUENCING METHODS TO IDENTIFY TRNA AND RRNA MODIFICATION FRACTIONS

5.1 Introduction

The technical advance afforded by tRNA-sequencing allows us sequence human tRNAs with several million reads, with confidence in quantification of tRNA abundance. As well, the power of comparative analysis between demethylase treated and demethylase untreated samples gives us site-resolved information. In the demethylase untreated samples, there exist specific nucleotide mutations in the cDNA as well as truncations upstream of either known or putative mutation sites. However, in the demethylase treated samples, these sequencing "aberrations" return to the native primary sequence of the tRNA. Yet, there still persist sites with the misincorporations or truncations even in the demethylase treated samples, and we use these remaining peculiarities to identify modifications further. There is recent precedence for using mutations and truncations in sequencing analysis (despite for a long time regarded as noise or otherwise junk sequencing reads): residual antibody artifacts in the cDNA reads afforded Linder et al site specific information as to the location of m6A in immunoprecipitated RNA[65, 114].

We are able to assign a metric to every single position in the tRNA genome of human eukaryotic tRNAs based on the proportion of mutated or truncated reads per position from the demethylase untreated samples. This metric we refer to this as modification index, or MI. We find that in demethylase treated samples, cumulative modification index decreases at certain positions and remains at others, indicating other modification sites unaffected by the demethylase cocktail.

We are able to identify many different tRNA modifications patterns in sequencing through

homology or existing tRNA knowledge. We also validate this for mitochondrial tRNA modifications and ribosomal RNA modifications for the sake of a control.

We assign a semi-quantitative measure to modification index, indicating that the higher the modification index, the more likely a site is modified, and using primer extension we quantify these sites of modification with good correlation between sequencing and experimental results. We use this modification index to identify new sites of modification that have been previously unannotated in current modification databases. Furthermore, we validate the existence of two newly discovered tRNA modifications using a suite of biochemical tools such as MALDI-TOF and LC/MS-QQQ by using an ingenious method of isolating fragments containing the modification of interest from specific tRNAs. Lastly, we find interesting variations in methylation patterns between two common cell culture lines.

5.2 Results

5.2.1 Modification Index

Previous tRNA-sequencing bioinformatics relied on analyzing known sites of modification. This was due to the knowledge that tRNAs have many common sites of modification that occur at the same position across most tRNAs, such as m^1G at position 37 and m^1A at position 58. Thus, given sequence context, inferring positions of modification relied on homology or experimentally verified sites. However, to identify all sites of modification from the sequencing data, we could not just rely on these factors, as isodecoders often contain sequence difference, and homology typically only allowed for identification of the high-scoring tRNAs from the tRNA database.

In order to search for modifications across the tRNA-ome, we needed to derive a scoring function based on what we perceived to be mismatch due to modification. For our purposes,

mismatch is the combination of mutations at a site relative to the aligned tRNA sequence and cDNA reads stopped directly 3' of the position. In this way, we can score how modified a site is at single-nucleotide resolution. This metric, which we label as modification index (MI), is calculated via the following:

$$MI = \frac{m + s}{c} \tag{5.1}$$

where m designates mutated nucleotides counted at the position, s designates reads stopped +1 of the position, and c designates the sum of all correct reads, incorrect incorporations, and read stopped prior to that position. This is demonstrated schematically in Figure 5.1 for a hypothetically methylated position.

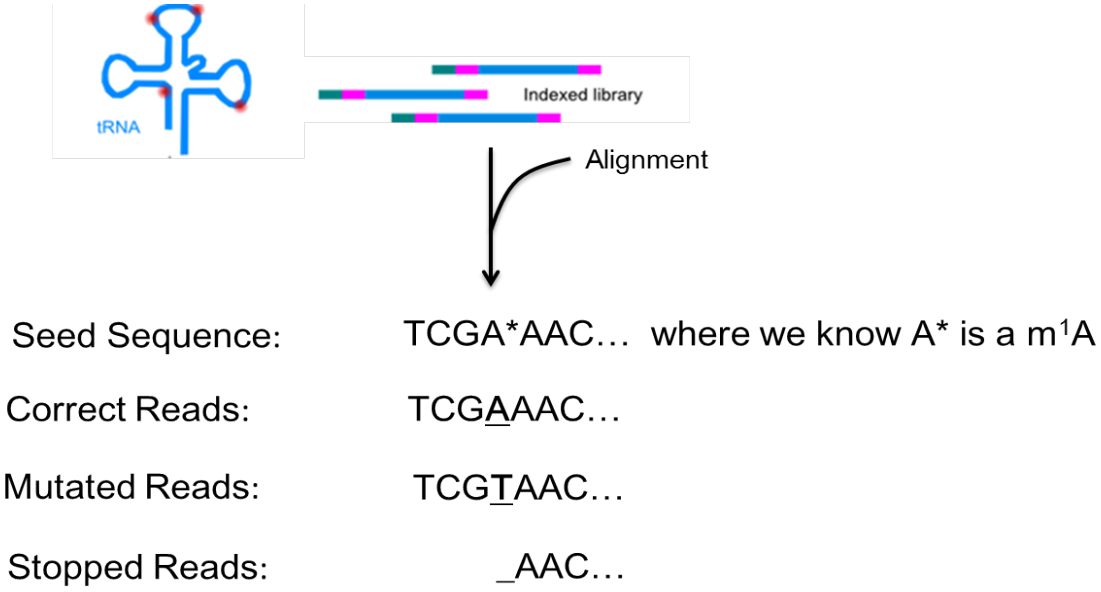

Figure 5.1: For a hypothetical known modified site, we can determine what constitutes mutations and stops based on how the cDNA reads are aligned to the seed tRNA gene sequence.

Using this metric, we were able to identify over 500 sites of modification index greater than .15, which was our threshold for signal relative to noise. However, upon inspection of

many of these sites, they appeared near the 5' end of the tRNA gene, and it was often due to TGIRT aborting transcription. Thus, we needed to further refine our selection process for sites of modification.

5.2.2 Creating a framework for identifying modifications - Optimization of tRNA-Seq Analysis

In order to identify modifications properly, we first needed to re-examine the data. Upon first inspection, we noticed that many cDNA reads often contained a 5' nucleotide that was added randomly to the extending cDNA by the TGIRT enzyme. This caused many issues with Bowtie2 analysis, because it would often lead to dissimilar comparisons between isodecoders even among full-length tRNAs. Globally inspecting the data, we noticed this led to an inflation of our modification index. As such during our trimming process, we incorporated a one nucleotide trimming of the 5' end, which cleaned up our background significantly with respect to un-modified sites.

5.2.3 Modifications are reflected via truncations and mis-incorporations but not indels

In order to further qualify what was counted as a mismatch, we were interested to see if modifications could cause insertions of extra nucleotides near the modification, or deletions at the site altogether as though the reverse transcriptase were skipping the modification altogether. We plotted several high abundance tRNA mismatches at well-known sites of modification using various parameters in common aligners. Plotted here are the most common Bowtie1 and Bowtie2 options for mismatches, as well as cumulative mismatch over the tRNA gene. We note that in comparing modification index at positions, the various parameters allowing for maximum mismatches are nearly identical when assigning modification index. However, the difference occurs when we use Bowtie1 versus Bowtie2 for the sake of

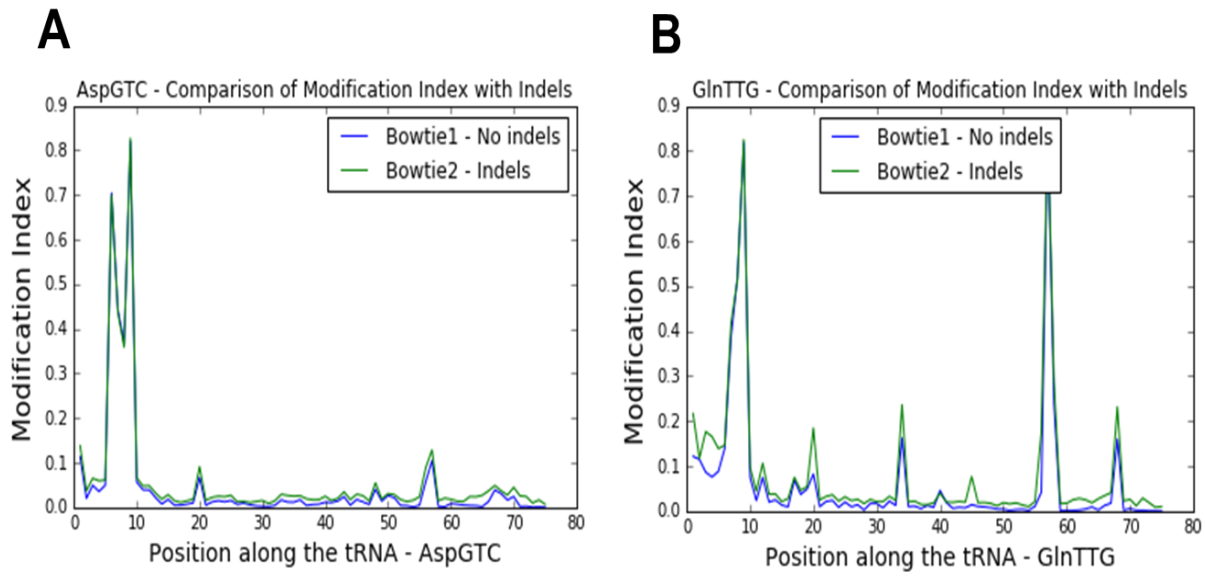

Figure 5.2: Modifications do not cause insertions or deletions in the cDNA relative to the seed tRNA sequence. (A,B) Demethylase untreated AspGTC and GlnTTG plots for modification index along the respective tRNA gene position. If there were insertions and deletions, there would be noticeable translations in the modification index peaks and/or a high baseline modification index across the entire gene.

tRNA-wide modification index, noted in Figure 5.2. Bowtie2 alignment seemed to cause bins of high noise over certain regions on both the 5' end and 3' end of tRNAs, despite sites of modification being equal in modification index. Thus, we used the Bowtie1 aligner with highest allowed mismatches to facilitate modification identification.

5.2.4 Cumulative tRNA Analysis

We calculated modification index for every tRNA from our sequencing data, and they are plotted cumulatively in Figure 5.3 for type I and type II tRNAs. Within each plot, we notice large decreases in modification index for positions where known AlkB substrates exist, such as m^1A58. We further identified other sites of global modification index peaks using homology to classify dihydrouridine peaks within the D loop region for tRNAs, as well as other assumed peaks such as $m^2{}_2G$ and inosine.

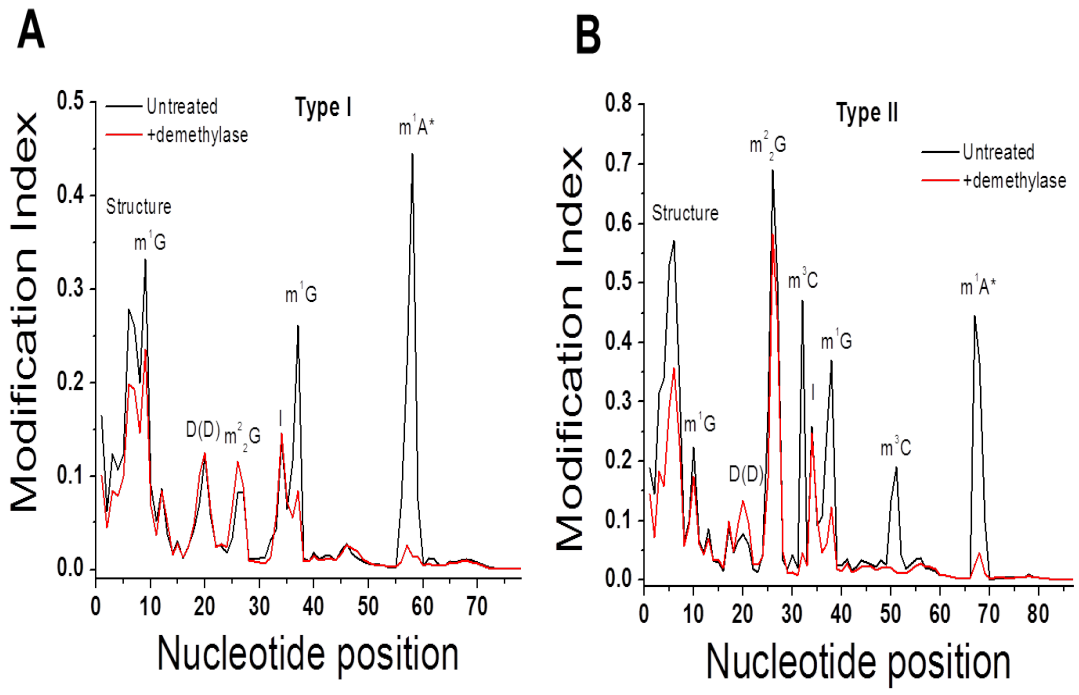

Figure 5.3: Modification index plotted over the cumulative tRNA landscape. (A,B) Modification index plotted over type I (A) and type II (B) tRNAs. Significant decrease in modification index occurs at known AlkB substrate positions, leaving us with modifications that impact reverse transcriptase function, causing misincorporations and cDNA truncation, which we can infer as positions of modifications. Known modifications that correlate with sites of modification index are labelled.

5.2.5 Deconvoluting Modification Index: Identifying Modifications

However, globally, we understand from the original tRNA sequencing analysis that certain modifications appear to misincorporate more than they induce RT arrests. Thus, to further get at identifying modifications, we deconvoluted the cumulative modification index into its respective components of induced mutations and induced stop, shown in Figure 5.4. Comparing the demethylase treated and untreated samples, we developed a decision tree shown in Figure 5.5 for identifying modifications based on each modification's respective modification index components or features related to the comparative analysis. We can use the respective signatures to classify and distinguish many modifications from each other. These signatures include a large change in modification index between demethylase treated

and untreated, or respective combinations of modification index components.

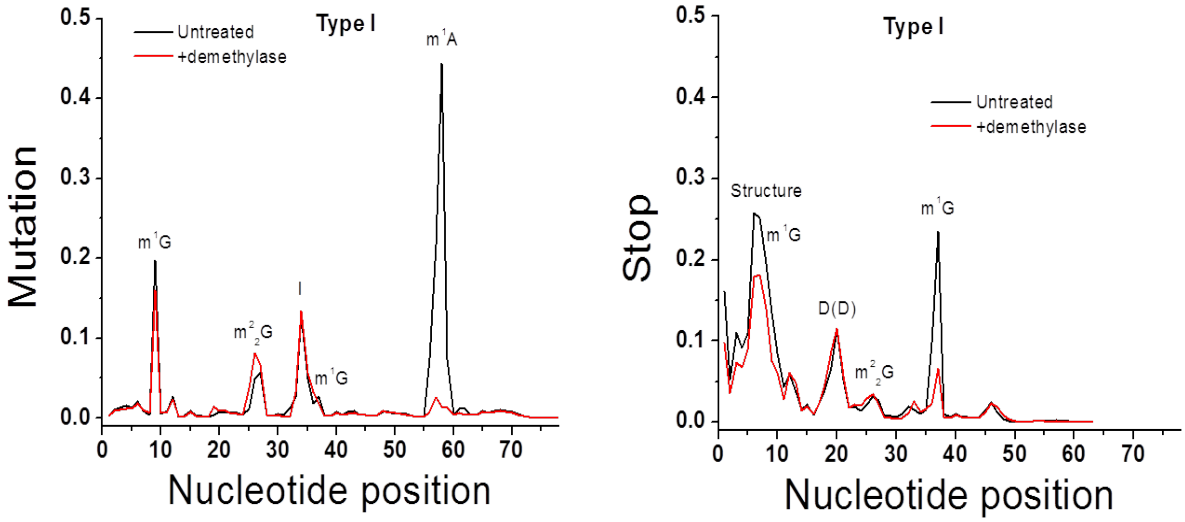

Figure 5.4: Mutation and stop signals for every nucleotide position are plotted for type I tRNAs, showing that modifications display different mutational or stop signatures that we can use to determine the type of modification at the position.

5.2.5.1 AlkB substrate modifications: m^1A, m^1G, and m^3C (Figure 5.6)

The first classification involves AlkB substrate methylations. These sites of modification are most apparent from the data, because of the generally large change in modification index. For example, we will take a tRNA isodecoder which appears to contain all three common AlkB substrate methylations: m^1G37, m^1A58, and m^3C in the variable loop of this type II tRNA. For this tRNA, we see a large drop-off in modification index intensity at the three assumed sites of modification. Another known modification, $m^2{}_2G$, does not see a similar decrease, because the modification is not an AlkB substrate, and thus its MI signal persists.

5.2.5.2 Inosine

In the introduction, inosine was referred to as the static modification in the RNA landscape. Indeed, this deamination of an adenosine to inosine (chemical structures shown in

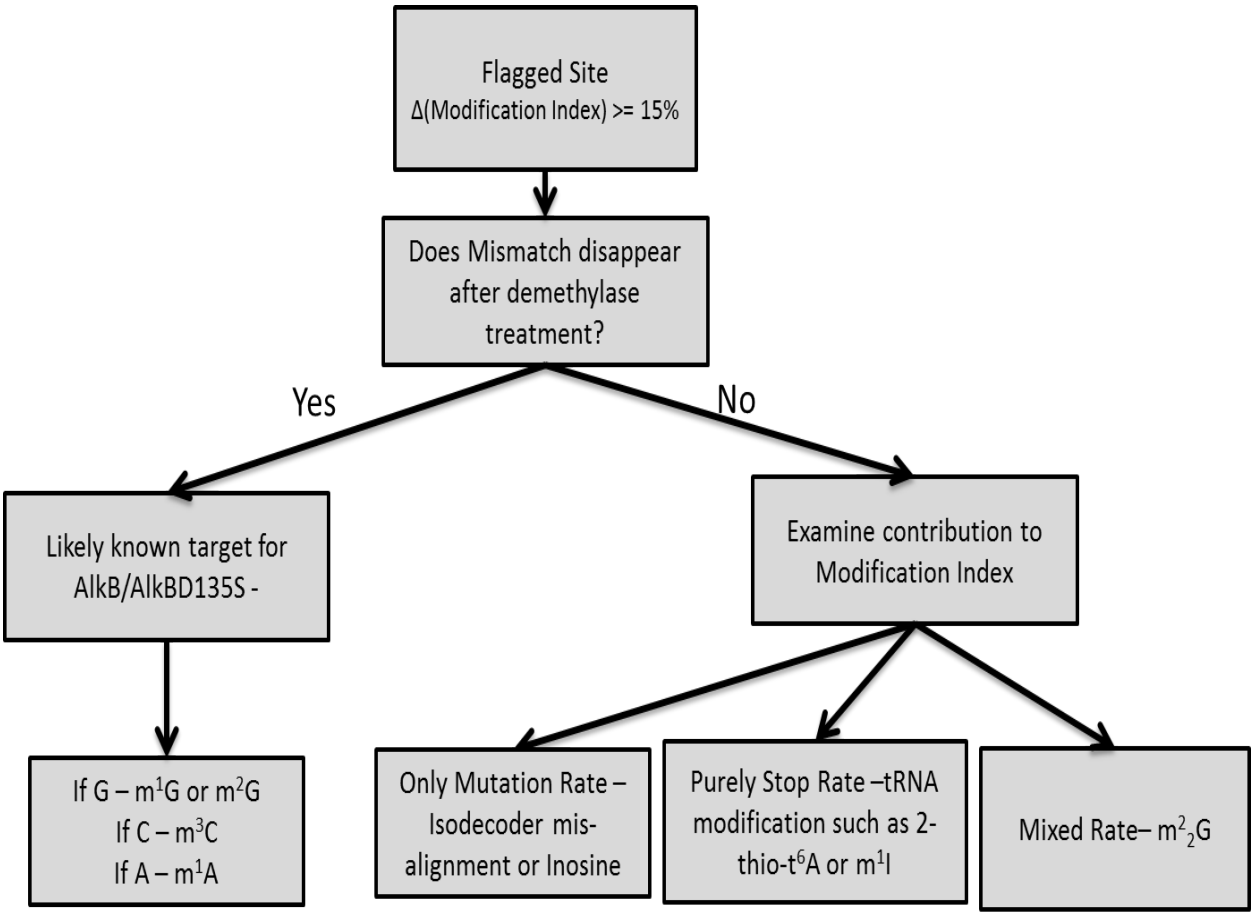

Figure 5.5: Detailed breakdown of how modifications are annotated and assigned based on modification index.

Figure 5.7 is not known to be reversible. This change in Watson-Crick face from a hydrogen bond donor to a hydrogen bond acceptor in the change from an amine to a ketone. Identifying these sites is somewhat easy, as A-to-I editing is common in tRNA anticodons with an A34. This A-to-I allows for the wobble base pairing in order to facilitate a wider range of codon decoding. This is reflected in the sequencing data, where these positions are entirely designated by a modification index greater than .95 with contributions only from the mutational component. Figure 5.7 shows a side by side comparison of the modification index and its mutational component for ValAAC, which contains A34, A35, and C36 in the anticodon, with the A34 being converted to an inosine.

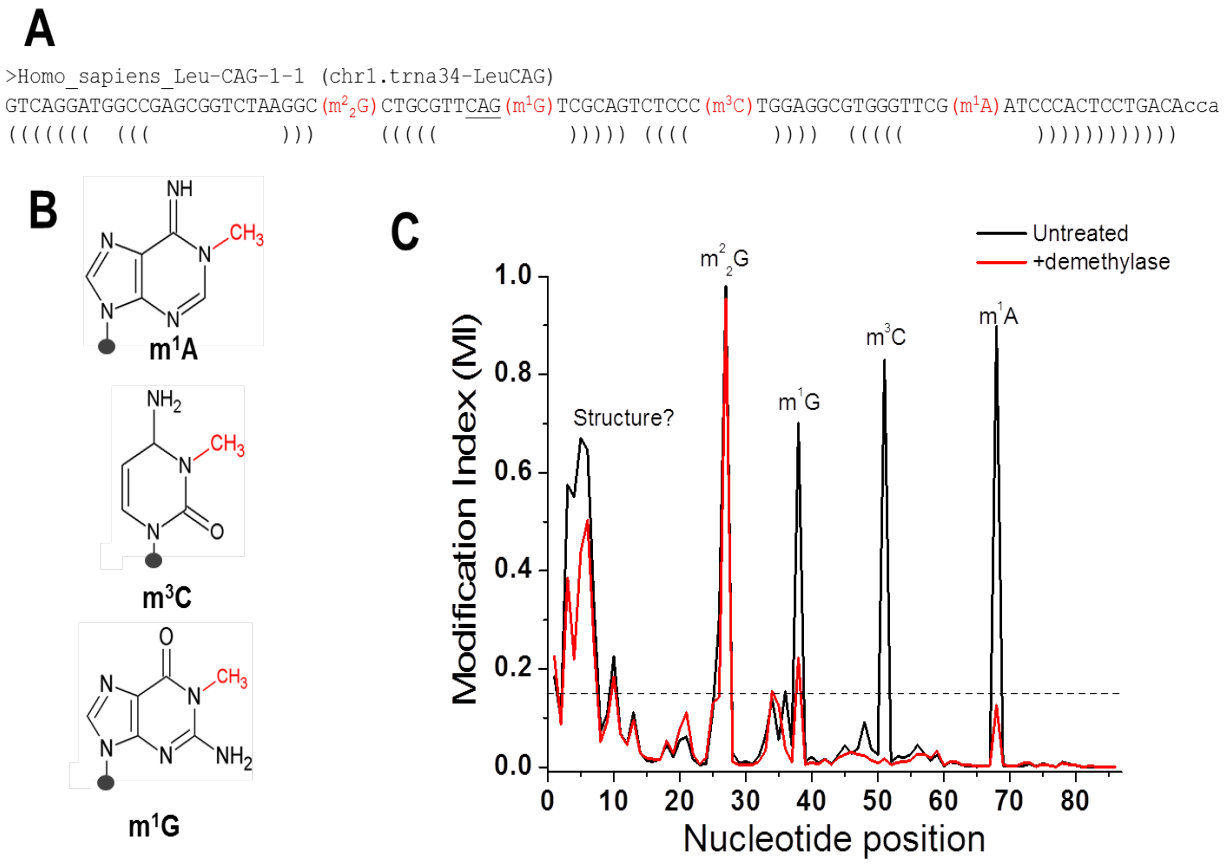

Figure 5.6: AlkB substrates show characteristic loss in modification index. (A) Primary sequence including annotated modifications from our study. (B) Chemical structure of AlkB substrate modifications. (C) Modification index plotted for LeuCAG nucleotide positions, with the known AlkB substrates showing loss of signal in the demethylase treated sample (red).

5.2.5.3 Dihydrouridine , 1-methylinosine , and 2-methylthio modifications

Conversely, there are modifications designated by contributions from only the stop or truncation component of the modification index. Interestingly, dihydrouridine has been shown to disrupt local structure in tRNAs, but the nature of the base would not necessarily reflect a change in the Watson-Crick face of the modification, so it was surprising that we have signal for this modified nucleobase. However, the local disruption is what gives dihydrouridine its signature, as examining modification index at known sites of dihydrouridine shows a striking multi-nucleobase signature. The disruption in local structure causes the RT

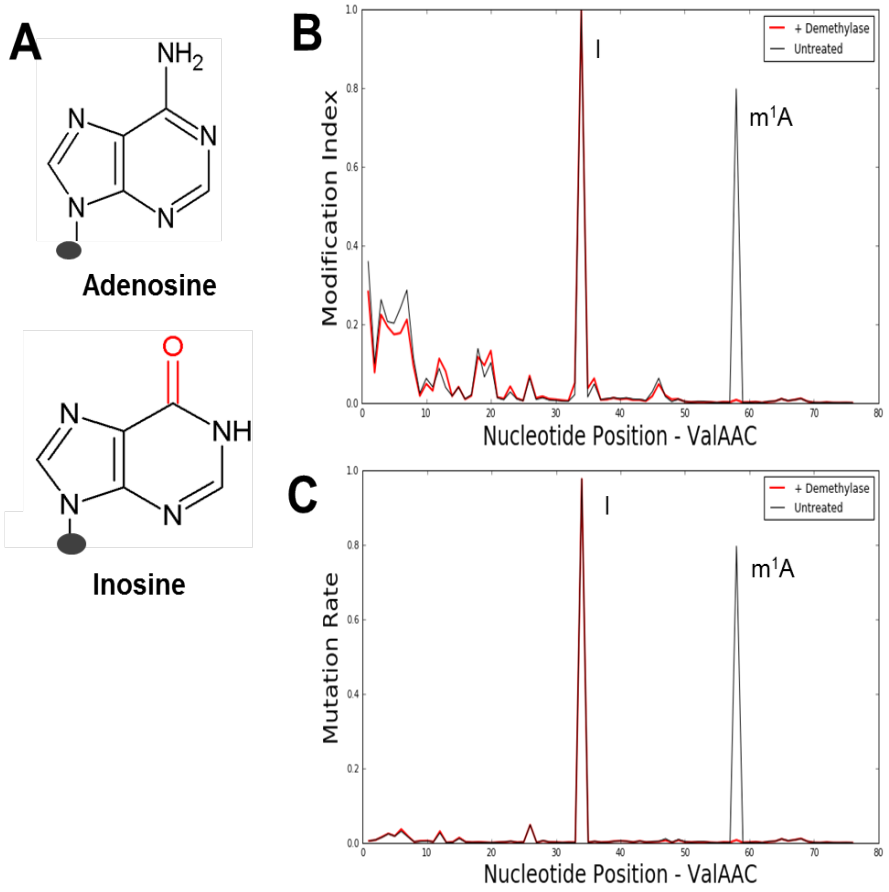

Figure 5.7: Inosine shows persistent mutational signal through demethylase treatment, concomitant with static RNA editing of the nucleotide (A) Chemical structure for adenosine (A) and inosine (I). (B,C) Modification index and mutation rate respectively for ValAAC, in which there is a persistence of MI signal throughout demethylase treatment.

to stop +1 of the dihydrouridine but also in some cases +2 as well.

Also striking is the how the methylation of inosine converts it from a purely mutational component to a purely stop component, but this can be recognized, as 1-methylinosine appears to be demethylated by AlkB, and thus the conversion at a position from highly truncated MI to a highly mutated MI with no overall delta in MI is its latent signature, shown in Figure 5.8.

Lastly, the 2-methylthio modifications are exceptionally sterically bulky, so it is natural

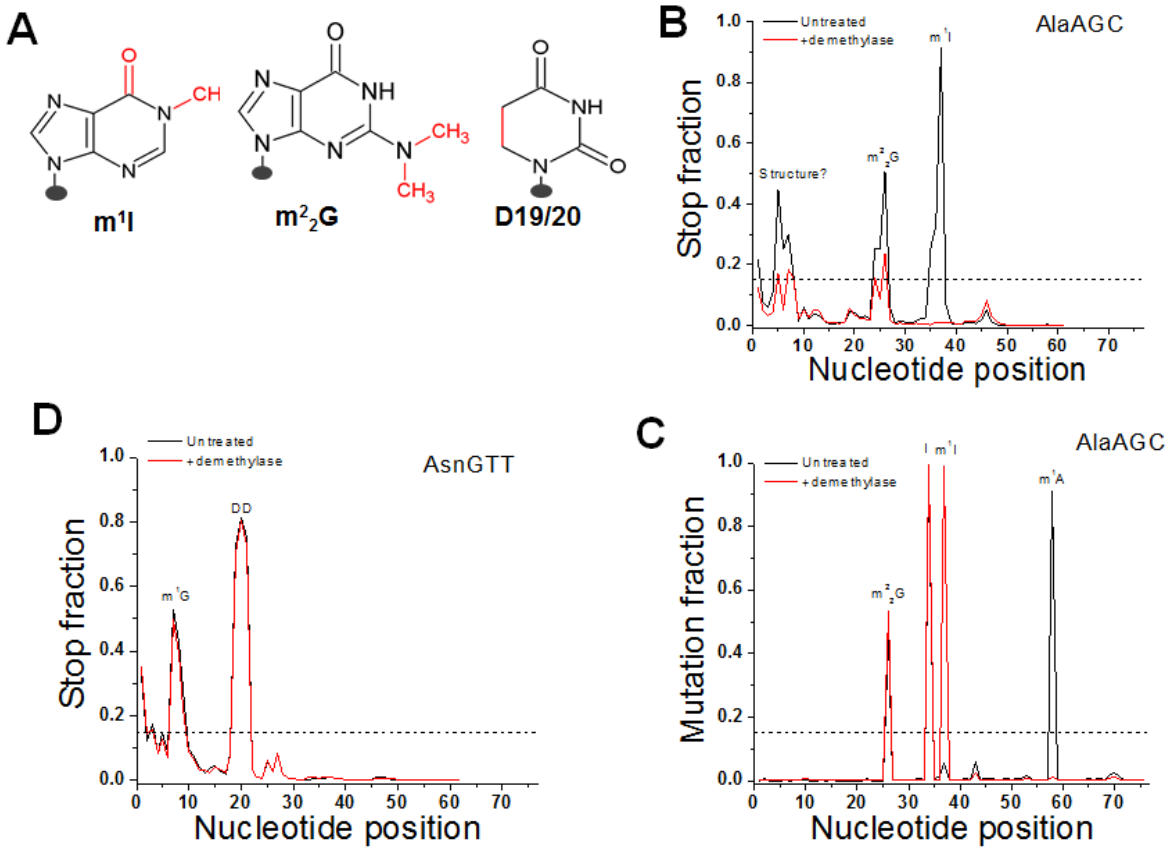

Figure 5.8: Modifications that contain purely signal from cDNA truncations. (A) Chemical structure for 1-methylinosine (m1I), di-methylguanosine ($m^2{}_2G$), and dihydrouridine. (B) Stop component plot for AlaAGC, which contains signal at position 38 for 1-methylinosine. (C) Mutational component plot for AlaAGC, which shows mutational signature for that position after AlkB treatment. (D) Stop component plot for AsnGTT, which has a well-known consecutive dihydrouridine modification in the D loop from Modomics.

that TGIRT is unable to process through the modification. Characteristic of the size of the modification is the size of the stop component - on average, 95% of the reads for a tRNA are stopped at the modification, with very little readthrough.

5.2.5.4 Dimethylguanosine

Dimethylguanosine, or $m^2{}_2G$, is the most interesting of the modifications that we encounter. It is mildly responsive to AlkB treatment, as can be seen in the decrease in stop

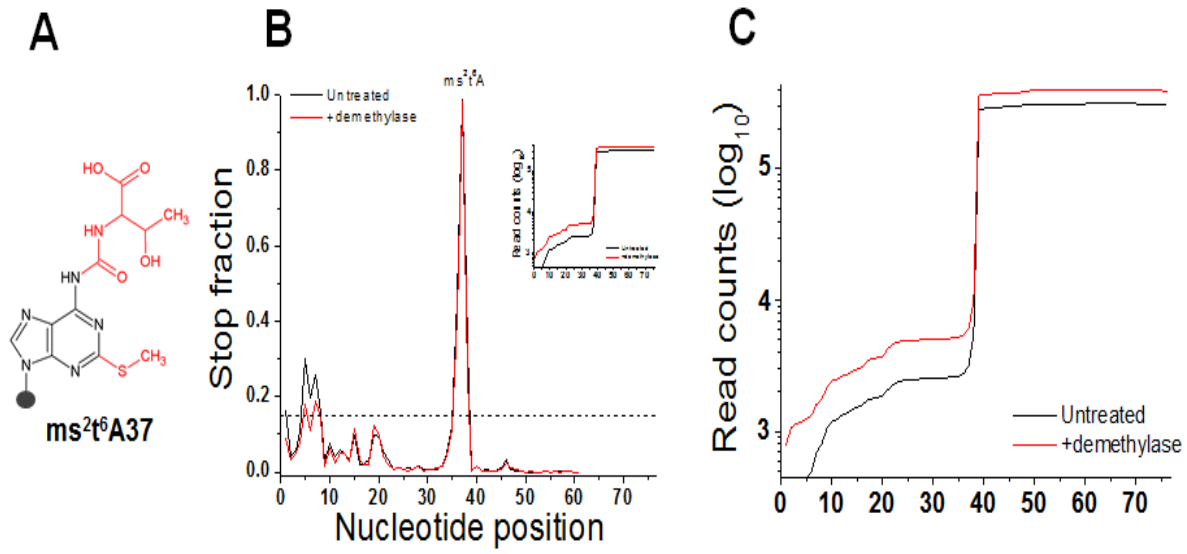

Figure 5.9: 2-methylthio modifications cause a large steric block, completely impeding TGIRT. (A) Chemical structure for the cytoplasmic 2-methylthio modification ms2t6A37. (B) Stop fraction for LysTTT, showing the demethylase resistance of the site. (C) Log read counts for LysTTT, showing a tremendous drop-off in read density over LysTTT reads.

fraction modification index in Figure 5.8 for AlaAGC, although no noticeable change appears for the mutational component. It is most notably the only modification we encounter with a sizeable stop component and mutation component after AlkB treatment, and that signature is enough for its identification in many tRNA species.

5.2.5.5 rRNA Modifications: 3-methyluridine and $m^1Acp3\psi$

Lastly, in ribosomal RNA, many sites of modification are carefully curated and annotated. Most modifications are known to be crucial for biogenesis and function, and as such these sites allowed us to examine certain modifications out of the tRNA context. From ribosomal RNA, we were also able to find two well-characterized Watson-Crick modifications, each with their own unique signatures, as shown in Figure 5.10. These modifications are particularly rare and not known to be found in any other RNA species, so their identification allows us just to compile a list of modifications we can identify and flag from sequencing data.

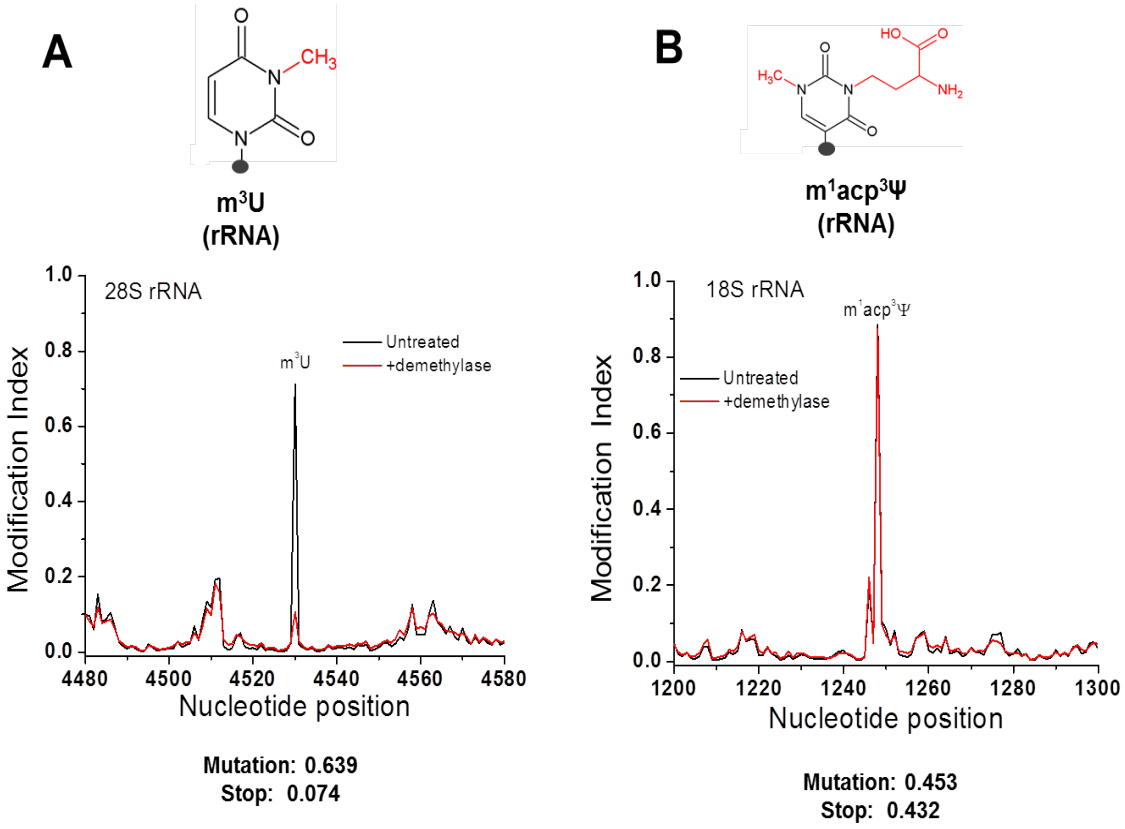

Figure 5.10: Expanding modification index into ribosomal modification (A) Chemical structure and modification index plot for the known 3-methyluridine site in 28S rRNA. (B) Chemical structure and modification index plot for the methylated pseudouridine in 18S rRNA.

5.2.5.6 Summary

From all of these identified modifications, we can compile a list and compare it to known databases of modifications, namely Modomics and the Sprinzl database[51, 69]. For each isodecoder, we compiled a list of sites of modification, which can be seen in Table 5.1 . We also compiled the list of modifications for mitochondrial tRNAs as well, shown in Table 5.2, and upon comparison to the recently published comprehensive mitochondrial modification list from Suzuki et al in bovine liver [108], we identify two new modifications as well. All in all, we are able to identify 11 distinct nucleobase modifications from sequencing data consisting of tRNA and rRNA data.

tRNA[a]	$m^1A58^{b,c}$	m^1G37^b	$m^2{}_2G26^b$	m^1G9^b	m^3C^b
AlaAGC	+++		+++		
AlaCGC	++[d]		+++		
AlaTGC	++[d]		+++		
CysGCA	+++[d]	+++			
AspGTC				+++ (A)	
GluCTC	+			+	
GluTTC	++			++	
GlyTCC	+++				
HisGTG	+++	++			
IleTAT	+++[d]		+++	+++	
LeuCAG	+++	++	+++		+++ (47d)(CAG)
LeuTAG	+++	+	+++		
Met-e	+++		+++		+++ (20)
ArgCCG	+++	+++	+++	+++	
ArgTCG	+++[d]	+++	+++	+++	
SecTCA	+++				
SerCGA	+++[d]		+++		++ (32) + (47d)
SerTGA	+++		+++		+++ (32) + (47d)
ThrCGT	+++		+++	+++	++ (32)
ThrTGT	+++		+++	+++	++ (32)
ValTAC	+++				

Table 5.1: Table of newly annotated modifications from tRNA sequencing data.(a) A black colour denotes previously known in human or mammalian tRNA, whereas red is previously unknown. (b) +: MI = .15-.30. ++: MI = .30-.50. +++: MI .50-1.0. (c) MI values include modification information only. (d) Described in source

tRNA[a]	$m^1A58^{b,c}$	m^1G37^b	$m^2{}_2G26^b$	$m^1A/G9^b$	Other[b]
mtAlaTGC		+++		++	
mtArgTCG				+++	+ (m^1A16)
mtAsnGTT			+	+++	
mtAspGTC				+++	
mtCysGCA				++	
mtGlnTTG		+++		+++	
mtGluTTC				+++	
mtGlyTCC				+++	
mtHisGTG				+++	
mtIleGAT			+++	+++	
mtLeuTAA	+++			+++	
mtLeuTAG		++		+++	
mtLysTTT	+			+++	
mtMetCAT					
mtPheGAA				++	
mtProTGG		+++		+++	
mtSerGCT					
mtSerTGA	+				+ (m^3C32)
mtThrTGT				+++	++ (m^3C32)
mtTrpTCA				+++	
mtTyrGTA				+++	
mtValTAC				+++	

Table 5.2: a. Black: known in human mitochondrial tRNA (10). Red: previously unknown in human and other mammalian mitochondrial tRNA, but can be inferred from the modification maps of bovine mitochondrial tRNAs (25). Blue: New modifications in human mitochondrial tRNA. b. +: MI = 0.15-0.30; ++: MI = 0.30-1.00; c. MI values for m^1A58 include mutations only

5.2.6 *Variation in isodecoder modifications*

Given the large change in abundance between different tRNA isodecoders, we decided to investigate how different isodecoders were modified by examining the top quartile of isodecoders for each amino acid family. Shown in Figure 5.11is a heat map for Leucine isodecoder, generated by comparing modification index at each position. We observed variation not only in methylation patterns but also with the existence of certain methylations, namely the m^3C species that had been identified.

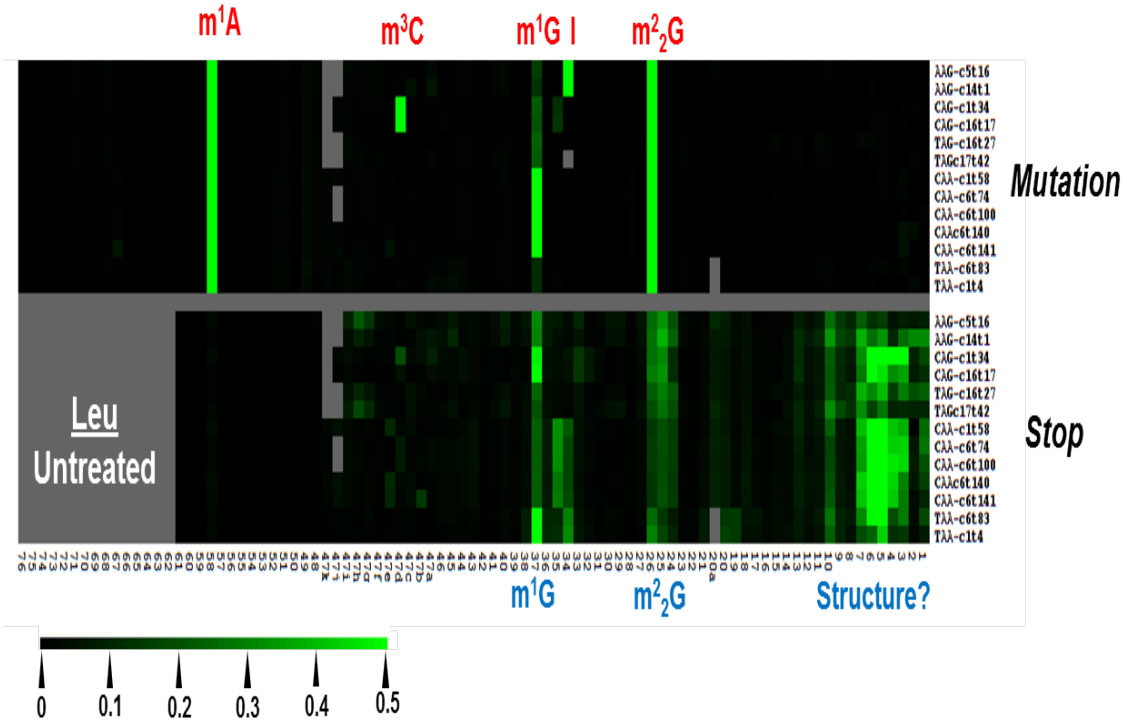

Figure 5.11: Heat map for demethylase untreated for the entire Leucine tRNA family. The scale for the heatmap is displayed below.

5.2.7 *Validation of newly discovered tRNA modifications*

From compiling the table of modifications, we realized quickly that many modifications had been previously uncurated for human tRNAs, and we set out to identify two rather novel sites of modification. The first of which is m^1A9 in AspGTC. 1-methyladenosine itself

is not an uncommon modification for nuclear-encoded tRNAs, but m^1A at position 9 was formerly believed to only in mitochondrial tRNAs. The second site is m^3C51 in LeuCAG. The variable loop of type II tRNAs has been shown to be modified in some Ser and Leu tRNAs, and even rat and human tRNA Serine has been shown to contain a 3-methylcytosine, but this had not been shown in human leucyl tRNA.

In order to isolate these particular tRNAs, we decided to innovate a strategy for isolation of an isodecoder. The principle of isolating a particular fragment of interest from a full-length tRNA was designed to make MALDI analysis of the fragment more pertinent. This is compared to traditional isolation of a specific tRNA using a biotinylated oligo, however with subsequent digestion of the tRNA with RNase T1, we would have had confounding fragments from the full-length tRNA versus our fragment. We were able to isolate the tRNAs we wanted as verified from the MALDI spectra (Figure 5.12. Furthermore, we could identify specific mass shifts related to the addition of a methyl (-CH3) in the fragments of interest.

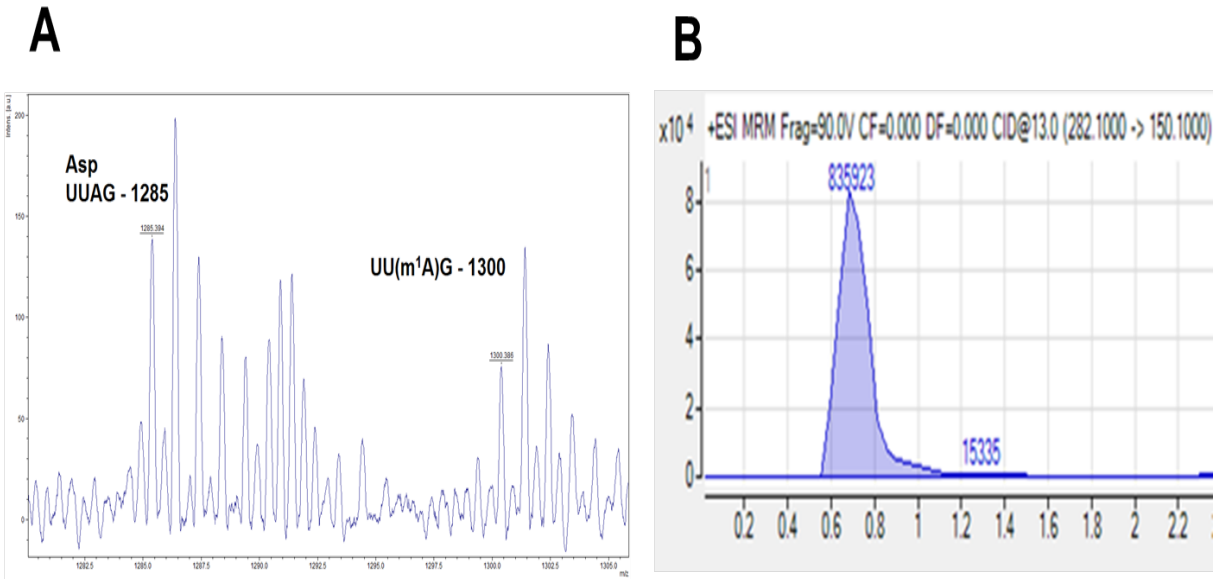

Figure 5.12: Biochemical validation of putative m^1A9 in AspGTC. (A) MALDI-ToF mass/charge plot for digested AspGTC fragment showing the isolated T1 digested fragment and the 15 mass shift due to methylation. (B) LC/MS-QQQ peak for 1-methyladenosine.

To further validate that these methyl shifts were not any other particular modification, we further digested the RNase T1 fragments with Nuclease P1, and we subjected the resultant sample to LC/MS-QQQ to validate what modifications were there. We found that AspGTC with the 5' fragment contained 1-methyladenosine, and the LeuCAG 3' fragment also contained 3-methylcytosine, which validated our sequencing identification results.

5.2.8 Quantifying tRNA modifications

One question still loomed over the sequencing results. That is the notion of how quantitative our modification index correlates with the how modified a site is. To do this, we performed the quantitative primer extension as described previously for four separate tRNA m^1A modifications. These modifications were specifically chosen to have a variation in modification fraction. The AspGTC m^1A at position 9 was our newly discovered position, but due to its proximity to the 5' end of the tRNA end of the molecule and TGIRT's issue with processivity, we wanted a better certainty of its quantitation. AspGTC is known also to be hypomodified at position 58 as well. Similarly, we had previously analyzed GlyGCC in employing this method. ArgTCG was chosen because it maintained a very high modification index in many separate tRNA-sequencing samples. The results of this comparison are shown in Figure 5.13. All sites of modification seemed to track closely in the experimental results as compared to their sequencing MI counterpart, allowing us some certainty in our quantitative analysis.

5.2.9 Differences in Cell Culture Line Modifications - m^1G and m^3C

Different tissue types have different tRNA expression. This has been well-studied and characterized. Similarly, we were wondering if tRNA-sequencing was sensitive enough to notice these changes in abundance. As well, we wondered if there were trends in modification difference. To do this, we averaged the modification indices from replicates of HEK293T data

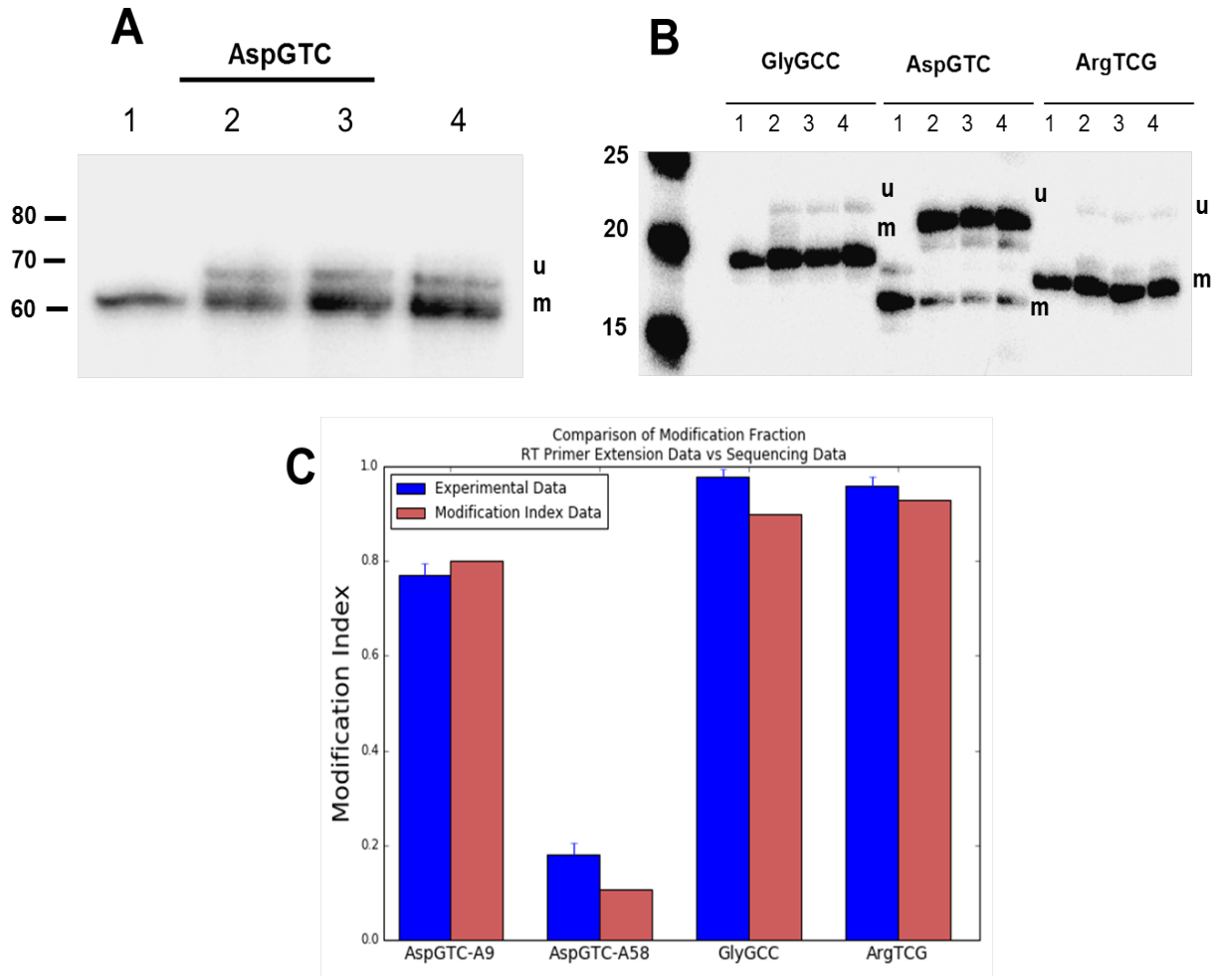

Figure 5.13: Quantitative RT Primer Extension for various 1-methyladenosine modifications. In (A,B) 'm' stands for the methylated stop band and 'u' stands for the readthrough quantification band. (A) AspGTC m^1A9. Lane 1 contains only a single nucleotide extension followed by ddNTP incorporation, and lanes 2,3,4 are replicates of the extended product. (B) GlyGCC, AspGTC, and ArgTCG m^1A58 quantitation. Lane setup is similar to (A). (C) Comparison of the experimental quantitation versus modification index values.

and HeLa data for demethylase untreated data. We applied certain stringent statistics to the data to ensure outliers would not impact the average modification index, and then we subtracted modification index for modified positions across all isodecoders. We then only flagged absolute differences in modification index of .15 or greater, a similar threshold that we use to distinguish signal from noise. These differences are shown in Figure 5.14. We find

that there is a trend of similar changes in 1-methylguanosine and 3-methylcytosine globally across the cell types, with HEK having larger 1-methylguanosine MI than HeLa and lower 3-methylcytosine MI.

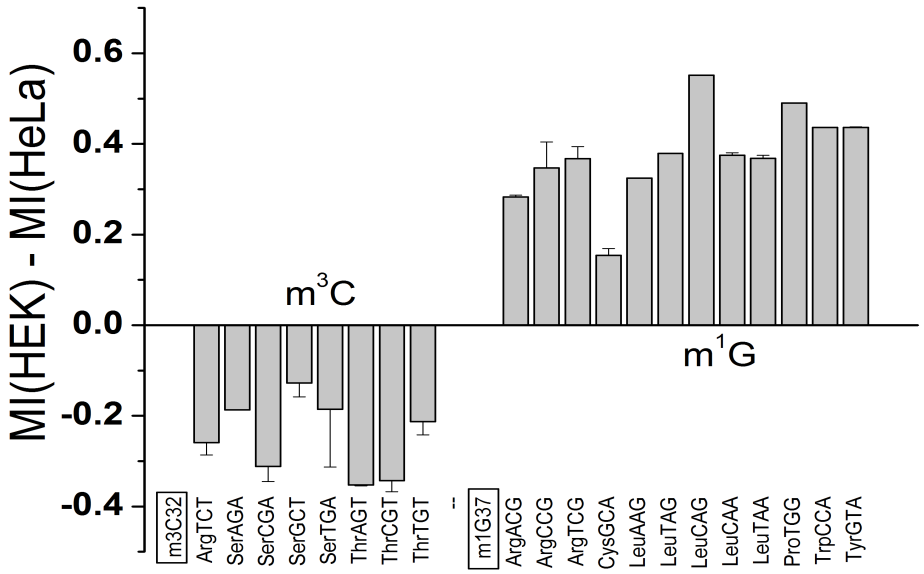

Figure 5.14: Modification trends between two common cell culture lines. The y-axis is the δMI between the two cell lines, with the various tRNA isodecoders on the x-axis.

5.3 Discussion

5.3.1 Novel human tRNA methylations

This work demonstrates that RNA methylations at the Watson-Crick face can be precisely identified and importantly, their modification fractions assessed by the newly developed tRNA-sequence methods. We show that the high extent of readthrough of the modifications by the thermophilic, group II intron RT (TGIRT) generates adequate amount of reads to enable transcriptome-wide analysis of 5 tRNA and 2 rRNA methylations. The tRNA methylations we detect and assess quantitatively cover approximately one third of all human tRNA

modification sites. We were able to identify many previously known sites in the widely used RNA modification databases as well as new sites for which no prior experimental data exist. Examples of completely novel sites include the m^3C47d site in LeuCAG, m^1A9 in AspGTC and m^1G37 in mitochondrial tRNAAla (Tables 1,2).

5.3.2 Variation in modification signature

Comparing the parallel sequenced, demethylase treated and untreated data lends high confidence regarding the m^1A, m^3C, m^1G and $m^2{}_2G$ methylations that are widespread in the tRNA transcriptome. Interestingly, the particular mutation and stop patterns also allow for discrimination of the modification types for the same base such as m^1G and $m^2{}_2G$: m^1G is manifested by more significant stop than mutation fractions, whereas $m^2{}_2G$ is manifested by more mutation than stop signatures.

A confounding factor not investigated here is the effect of sequence context of the modification site. A recent work by Hauenschild et al [23] shows differences in mutation and stop fractions for m^1A with a different +1 nucleotide sequence, indicating that sequence context can also be an important parameter in determining the ratio of mutations and stop at each site. Despite these potential caveats, MI values can be interpreted as providing a lower bound on the modification fraction for each methylation sites.

5.3.3 Causative explanation of HeLa/HEK modification differences

Even more surprising is the result that compared to HEK293T cells, m^3C32 modification fraction uniformly increases, but the m^1G37 modification fraction uniformly decreases in HeLa cells. Differences in specific tRNA modifications have been well documented for the 5-methoxycarbonylmethyl (mcm5) and 5-methoxy-carbonyl-methyl-2-thio (mcm5s2) U34 modifications in tRNAArg(UCU) and tRNAGlu(UUC) and 5-methyl (m^5C)-C34 in tRNALeu(CAA) to enhance stress response [42-45]. In yeast, global levels of many modi-

fication types such as m^5C, $m^2{}_2G$, and 2 O-methyl-C (Cm) can change significantly when cells are exposed to distinct types of chemical stressors [24, 46]. Our HeLa/HEK comparison is derived from the standard culturing conditions where no stress was applied.

5.4 Methods

5.4.1 Modification Index

Mapping pipeline proceeded by conversion from the SAM output from Bowtie using custom C and Python scripts to separate isodecoders based on previous alignment. From here, further cleanup based on redundancy (any read that could inherently map due to misalignment) was also discarded. For each position in the reference file, the following was calculated: at each position = n, how many counts existed (total counts) = c, how many misincorporations or mutations = m, and how many aligned reads stopped at the position = s. This leads to a full modification index (MI) calculated at each position by (mutations at position + stops at position directly 3 to position)/(total counts at position), or (m+s)/c, and individual mutation and stop components to the metric can be further reduced by calculating m/c and s/c, respectively.

For detection purposes, sites corresponding to a MI value of 15% or below were discarded as noise due to sequencing error or basal level of misincorporation from the RT reaction. For simplicity and ease of discussion, sites with MI 15% were categorized in the range of 15-50% as low, 50%-80% as medium and 80-100% as highly or completely modified. Positional shifts of modifications due to tRNA length (either due to variations in length of type I tRNAs or type I versus type II tRNAs) were recognized and adjusted for in calculations for heat maps. Modification index in regions of variable loop are accounted for accordingly. Modification index 3 to $m^1A57/58/59/68$ (m^1A58 in standard tRNA nomenclature) is only derived from mutations due to the lack of stop information at these positions because of the short reads

needed for stops at this m^1A position.

We first look for the effect of demethylase treatment at each flagged site. m^1A, m^3C, m^1G, $m^2{}_2G$ and m^3U sites show decreased MI values upon demethylase treatment (left branch). Sites that are unaffected by demethylase treatment are further assessed based on whether the MI is composed of only mutations, only stops or a mixture of mutations and stops (right branch). Among the right branch, mutations only correspond to A-to-Inosine modification or isodecoder misalignment, stops only correspond to the presence of bulky modifications in the Watson-Crick face and dihydrouridine (D), and a mixture of mutations and stops corresponds to other, uncommon modifications in the Watson-Crick face. This type of analysis enables us to identify the specifics of the modification using the MI values.

5.4.2 Isolating specific tRNA fragments of interest

AspGTC: First, total tRNA was gel purified from HEK293T total RNA. Then the total tRNA was 5' radiolabeled, and incubated with a DNA oligonucleotide complimentary to positions 26-76 of the AspGTC isodecoder of interest. The mixture was treated with RNaseH, and run on denaturing PAGE. The radiolabeled lower molecular weight fragments were then excised from the gel, shown in Figure 5.15.

LeuCAG: Similar to the AspGTC preparation, total tRNA was gel purified from HEK293T total RNA. Then, total tRNA was 3' radiolabeled, and incubated witha DNA oligonucleotide complimentary to positions 1-42 of the LeuCAG isodecoder of interest. The mixture was similarly treated with RNaseH and run on denaturing PAGE. The fragments separate from the total radiolabeled tRNA were excised from the gel.

5.4.3 MALDI-ToF analysis

1 pmol of the solution was mixed with an equal amount of MALDI matrix, which was composed by 9:1 (v:v) ratio of 20 ,40 ,60 -trihydroxyacetophenone (THAP, 10 mg/ml 1

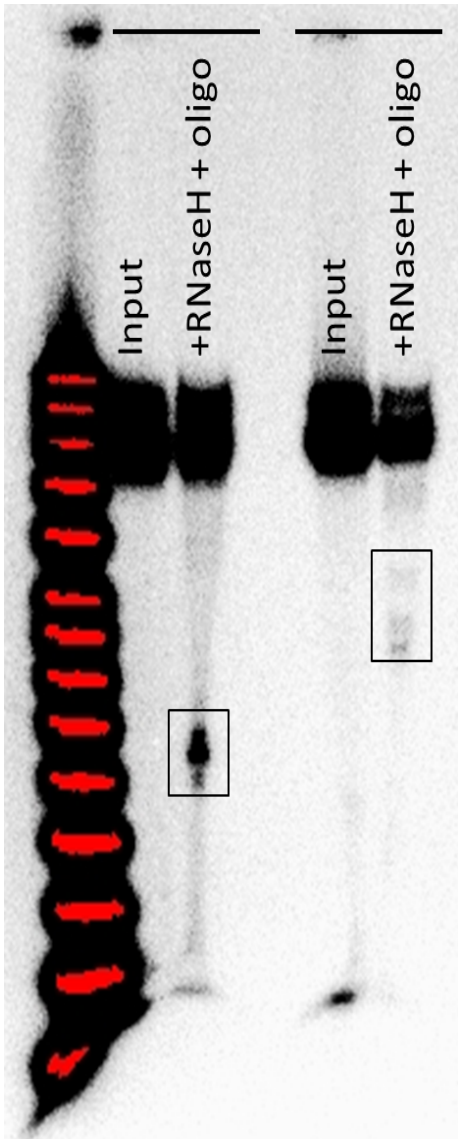

Figure 5.15: Isolating tRNA fragments of interest. With specific oligonucleotides complimentary to tRNAs of interest, we can degrade away regions of non-interest (regions where the modification that we are investigating is not present) in order to further validate the modification. Multiple bands in the RNaseH + oligo lane correspond to incomplete RNaseH degradation.

in 50% CH3CN/H2O): diammonium citrate (50 mg ml 1 in H2O). The mixture was then spotted on a MALDI sample plate, dried under vacuum and analysed by a Bruker Ultraflextreme MALDI-TOF-TOF Mass Spectrometers in reflector, positive mode.

CHAPTER 6

EXPANDING RNA MODIFICATION BIOINFORMATICS

6.1 Introduction

Given the wealth of sequencing data that currently exists, the task of combing through sequencing reads is becoming insurmountable. Public genome repositories estimate that we generate roughly 15 petabytes of data per year, as a typical RNA-seq study from an Illumina HiSeq2500 produces hundreds of gigabytes of raw data per run [107]. Bioinformatics as a field has turned from algorithmic problem solving largely into data analysis as the volume of information approaches that of 'big data.' Thus, the field must start to incorporate these data science techniques like machine learning to begin to scale the mountains of data that are piling up. To this end, one technique of machine learning known as supervised learning has already been utilized in identifying modifications[114]. Supervised learning allows scientists to mine data with known outputs or categories in mind: in this case, Hauenschild et al were able to classify sites as either containing 1-methyladenosine or other adenosine modifications with some certainty, given inputs of contextual sequencing signature based on arrest rate using a particular reverse transcriptase[44].

We previously were able to identify many different modifications from a small subset of sequencing data. We posit that for each modification, there is likely some combination of truncation and mutation that is inherent to the function of reverse transcription acting on modification, or: $f_{rt}(mod) = signature$. We find that for tRNA modifications, the identity of the modification coupled with the nucleotide in the +1 position cause a variety of signatures. We also find that this persists for non-biological adducts such as the addition of CMC for identifying pseudouridine.

6.2 Results

6.2.1 tRNA modifications signatures depend on +1 nucleotide

Sites of modification were flagged similarly from before, with a difference of .15 for modification index between samples used to determine AlkB substrate modifications. However analysis was independent of position for flagged modifications, allowing for the training set of modifications to contain off-targets. Modifications were sorted by nucleotide identity at the site of high MI, and then sorted again based on +1 and -1 nucleotide identity. The modification index was then deconvoluted into stop component and mutation component to allow for determination of signature.

m^3C has the fewest sites in tRNA with 39 identified sites. Because of the location of m^3C in the various locations in the tRNAs, all within loop structures, there is no double-strandedness to affect the RT. The modification signature for m^3C is shown in Figure 6.1. m^3C shows large variation in how the +1 position tunes how the reverse transcriptase interacts with the modification, showing an almost complete misincorporation with the +1 T which contrasts with the complete RT arrest from a +1 G.

m^1G resides mostly in two positions in tRNA, at positions 9 and 37. There are stark structural differences between these two sites, with the anticodon stem loop position being mostly exposed and unpaired, whereas the m^1G9 is contained in a highly structural region of the tRNA. However, there are no correlations between position and +1 identity. Much like m^3C, m^1G also has different signatures depending on +1 nucleotide, seen in Figure 6.2. Diverging from the m^3C, modification, the +1 G nucleotide causes the most misincorporations, whereas the +1 T is all cDNA truncation due to RT arrest. However, +1 A is similar between the two modifications.

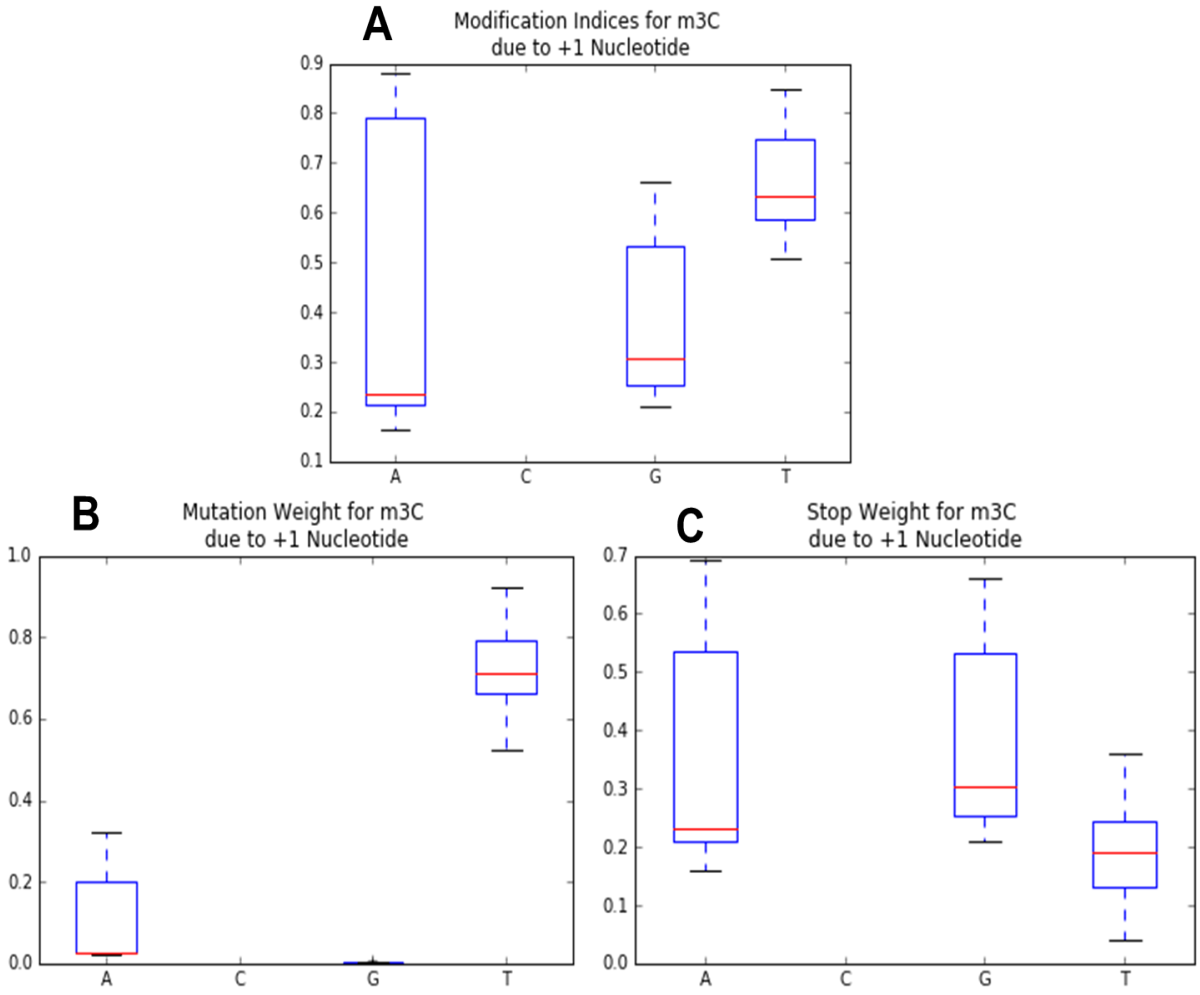

Figure 6.1: Modification Signature for m^3C in tRNA: The x-axis for plots A, B, and C all show the +1 nucleotide from the site of modification. (A) Modification index plot for all 3-methylcytosine modifications in tRNA isodecoders. (B,C) Mutation fraction and Stop fraction respectively for the corresponding +1 nucleotide from 3-methylcytosine. The sum of the box and whisker plots from (B) and (C) add to 1.

Because most m^1A data from the tRNA sequencing data sets is obtained from the 58 position, where little to no stop information is presented, all modification index component for m^1A are mutationally restricted. As shown in Figure 6.3, this leads to strong signal for showing how the +1 positions for m^1A cause large shifts in what the reverse transcriptase will incorporate into the growing cDNA.

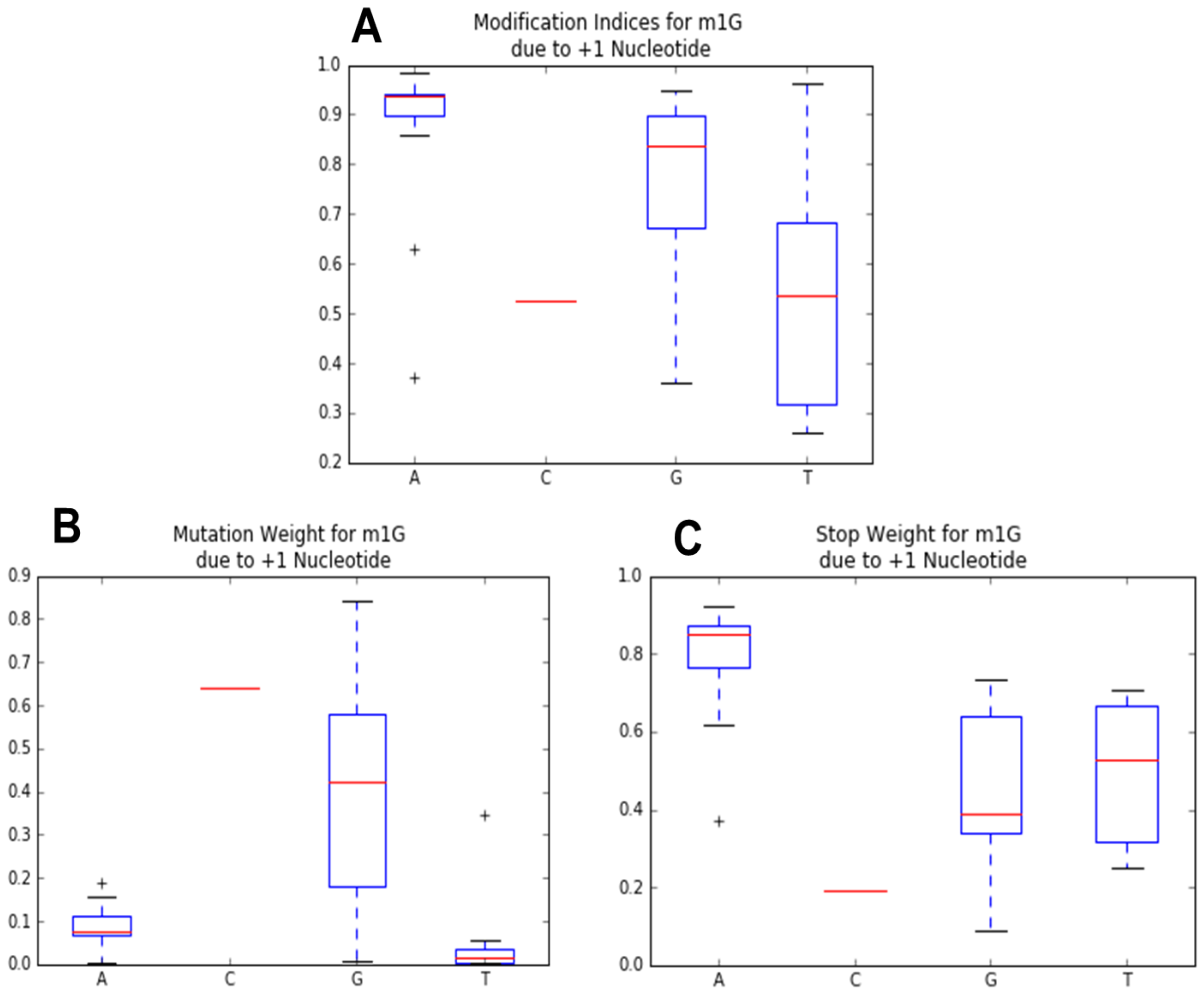

Figure 6.2: Modification Signature for m^1G in tRNA: The x-axis for plots A, B, and C all show the +1 nucleotide from the site of modification. (A) Modification index plot for all 1-methylguanosine modifications in tRNA isodecoders. (B,C) Mutation fraction and Stop fraction respectively for the corresponding +1 nucleotide from the m^1G. The sum of the box and whisker plots from (B) and (C) add to 1.

6.2.2 Pseudouridine Signature Analysis from rRNA

Unpublished results on pseudouridine (ψ) show similar variation in mutation and RT arrest. While pseudouridine by itself does not cause a reverse transcriptase to misincorporate a nucleotide nor does it cause RT arrest, the common method of identifying pseudouridines is with treatment of N-cyclohexyl-N'-(2-morpholinoethyl)carbodiimide metho-p-toluenesulphonate (CMC), which selectively modifies the ψ modification [16, 102]. This

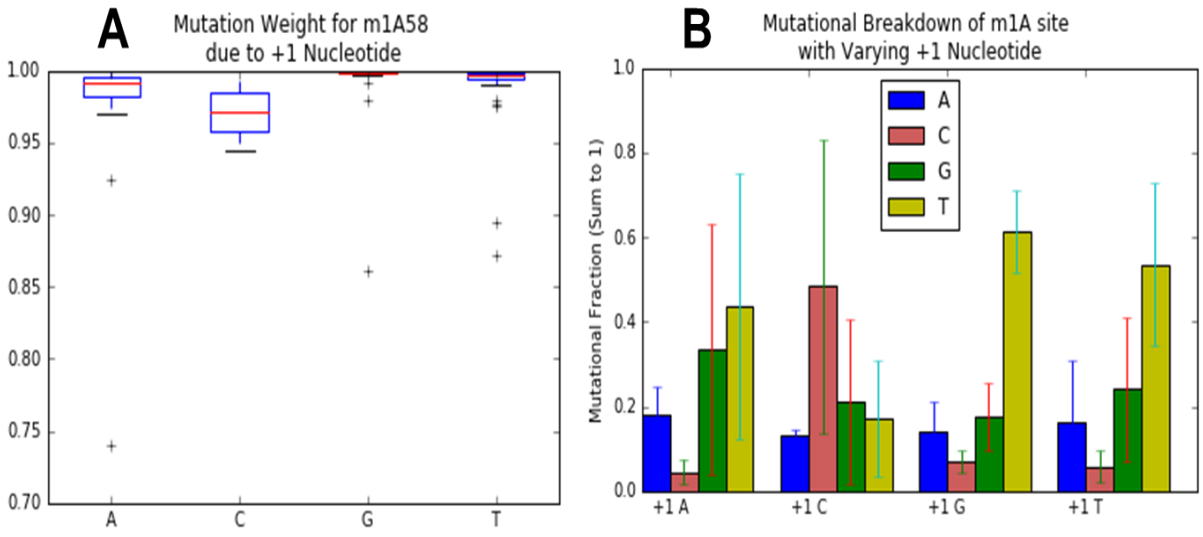

Figure 6.3: m^1A's sequencing profile allows for identification of RT misincorporation preference. (A) Mutation fraction for modification indices at position 58 for m^1A in tRNA. (B) For each +1 nucleotide, the corresponding misincorporated nucleotide fraction is shown.

large adduct is what invariably interacts with reverse transcriptases. However, the three main pseudouridine studies all focused on the RT arrest which causes eventual cDNA truncations, which were identified with trough calling in the analysis. They were only able to identify confidently a small handful of sites, however likely due to low modification fraction and poor resolution with only RT arrest information.

We were able to identify a reaction condition which causes the CMC-ψ to misincorporate and cause arrest instead of only the latter. Using this, we were able identify modification signatures from known ribosomal RNA pseudouridine sites. We used ribosomal RNA allowing us to focus purely on validation of the method. However, we were able to identify a small handful of new sites of pseudouridylation in the 18S and 28S ribosomal RNA. Shown here are plots for the modification index, mutational component, and stop component. Due to inefficiency of CMC chemical treatment, the modification index has an artifical ceiling around .4. However, this was still sufficient to identify large differences in fractional mutational and stop comonents, as shown in Figure 6.4. +1 A and +1 T seem to cause the most

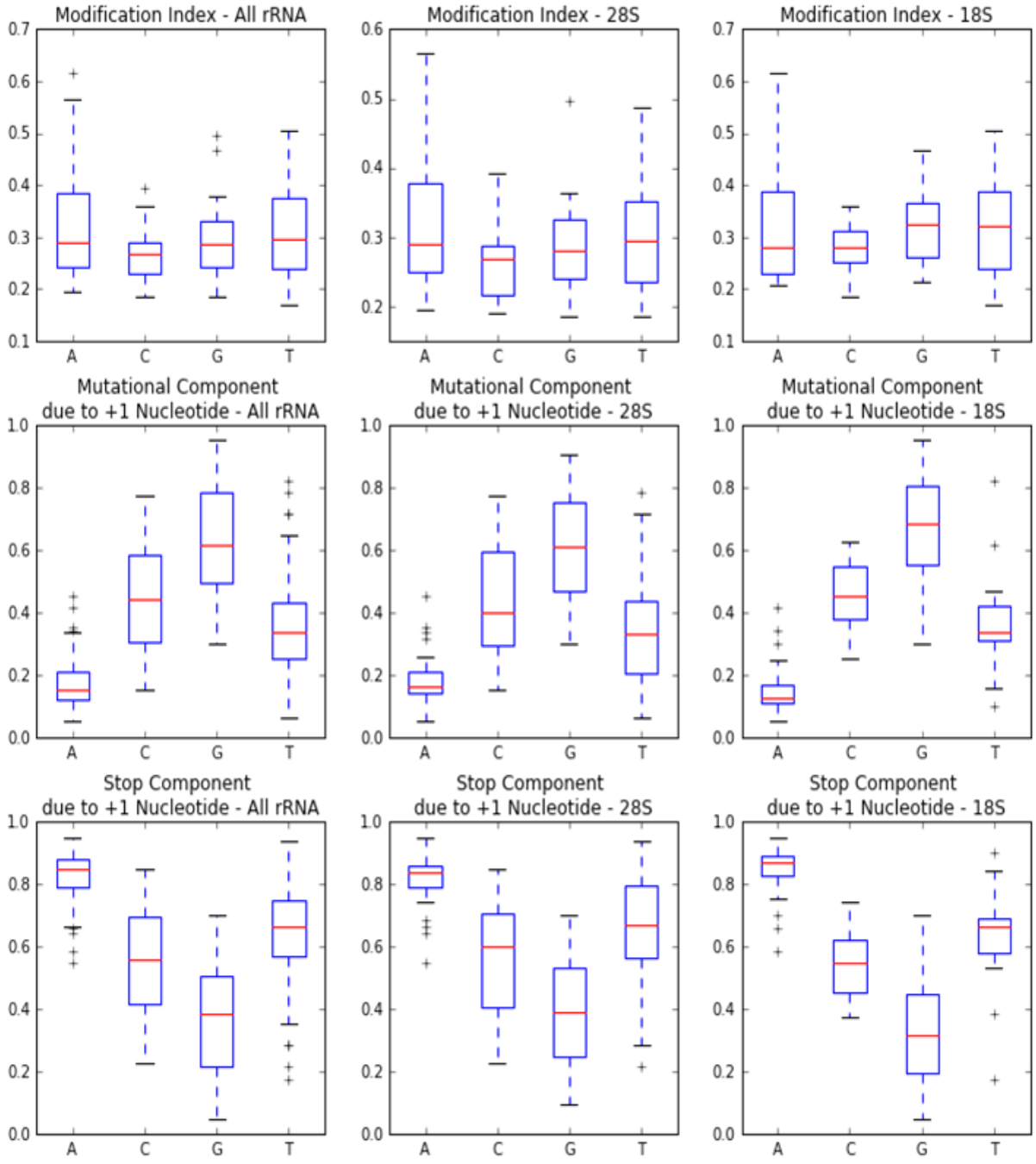

Figure 6.4: Modification index, mutational fraction, and stop fraction for combined ribosomal RNA, 28S, and 18S. The x-axis shows the nucleotide +1 to the CMC-ψ. Signatures for each position are similar, showing that native ribosomal structure may not be a determining factor for modification signature.

arrest in our conditions, whereas +1 C and +1G misincorporate more frequently.

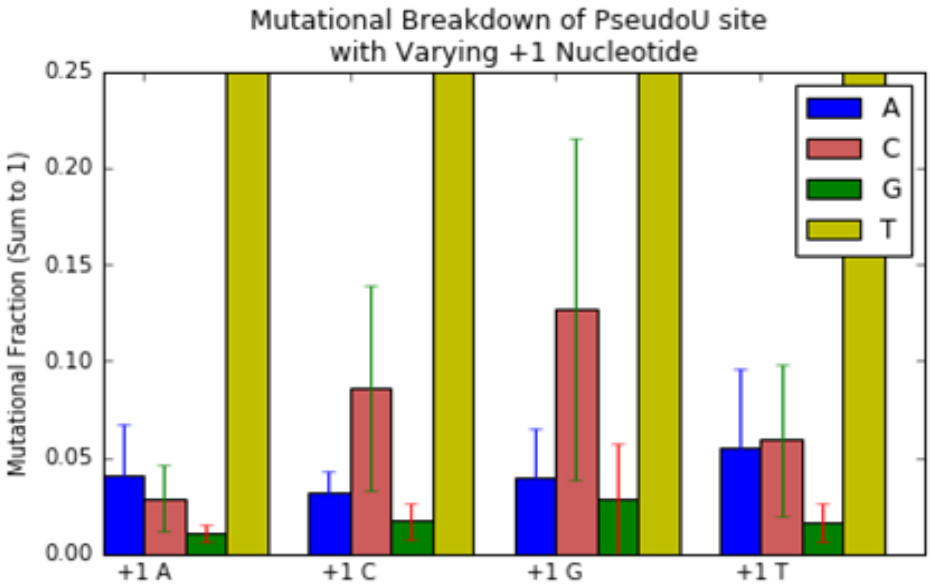

Figure 6.5: Breakdown of nucleotide incorporation at known sites of pseudouridine modification in ribosomal RNA from our experimental conditions. A strong tendency toward incorporation of dGTP across from the modified residue can be seen.

A drawback of the artificial modification index ceiling for pseudouridine is identifying the prominent misincorporated nucleotide. Shown here in Figure 6.5 is a zoomed in breakdown of the incorporated nucleotides at reported pseuouridine sites. The RT seems to favor misincorporation with a deoxy-guaonsine reading the position as a ribo-cytidine as opposed to the uridine.

6.2.3 Predicting Sites of Modification

Lastly, we can use modification signature for the ribosomal RNA for non-annotated sites. For instance, we find that a site in the 28S, U2843 has a modification index in the +CMC sample of .25, and it contains a ΔMI of .22, which initially flags the site. We can see this difference in MI in Figure 6.6. Furthermore, the mutation component and stop component at this site are .055 and .196 respectively, which give a mutation fraction and stop fraction of .20 and .80. Comparing this to Figure 6.4, a +1A tends to give the signature of a low mutational component and a high stop component, so we confidently identify this site as a

potential pseudouridine. Furthermore, this site was actually flagged in Carlile et al's work as a putative novel pseudouridine site, which gives us confidence in identifying this position from the signature we describe [16].

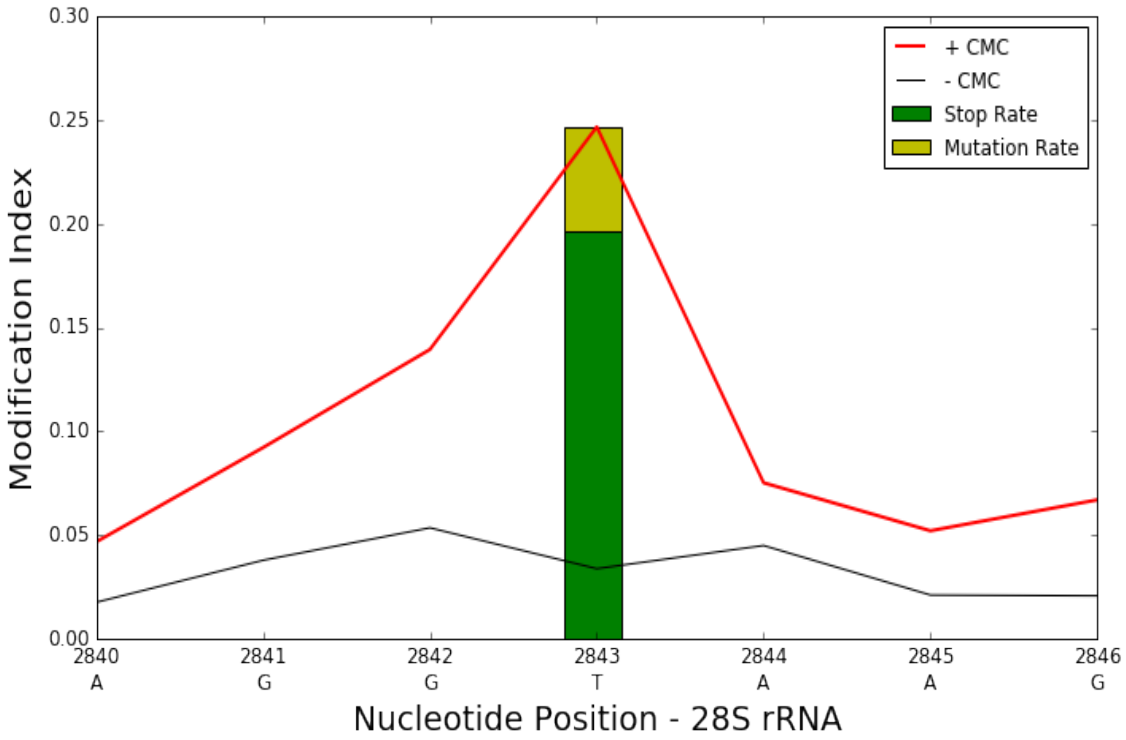

Figure 6.6: Using modification signature to predict sites of pseudouridylation. Analyzing the CMC treated and untreated data in parallel, we are able to discern new sites of modification using signature as a confident predictor of locations of modification. The red line shows modification index from CMC-treated data, peaking at .25 at the U2843. Similarly, in the untreated data, this site maintains a .03 modification index. The stacked bar graph shows the individual components of the modification index.

6.3 Discussion

Our work with tRNA sequencing demonstrates the power of coupling biochemical methods with bioinformatics for identifying sites of modification. Not only can be identify modifications given our methods, but furthermore we can assign modifications found in other contexts and other RNA molecules now without requiring prior knowledge and homology

as when we first began with tRNA. Modification signatures give us at a glance a strong statistical test by which we can test if a modification is what we believe it to be based on the reverse transcriptase used.

Furthermore, we can do this not with just native methylations in RNA molecules, but even using chemical adducts, we find that reverse transcriptases have particular propensities to either misincorporate or arrest at a site. We can expect that many reverse transcriptases will act in different ways, but there may be a possible underlying feature within the sequence context itself that lends to the signature decision point. Take, for instance, the m^3C and m^1G modifications in tRNA - given a +1 nucleotide identity of A, both showed strong stop signatures for the reverse transcriptase. This was even reflected in the pseudouridine result for a +1 A using an entirely different reverse transcriptase. As such, we modify our original assumption that the signature is a function of RT alone to something like this:

$$f_{rt}(mod, sequence) = signature \tag{6.1}$$

Why may this be the case? Well, perhaps simply there are just different energetics for the DNA-RNA hybrids in the reverse transcriptase active site that cause certain penchants for mutations or stops. However, there is a potential counterpoint to this in our data. The first is what has been mentioned is that m^3C and m^1G actually swap their tendencies for mutations with a +1 T, which we may imagine to be one of the weakest energetic reactions - 2 hydrogens bonds from the ribonucleic T to the deoxynucleic A.

Furthermore, there may be a biological imperative to exclude certain sequence contexts. With tRNA m^1G and m^3C, both excluded a +1 C over 100 possible sites in combination. This may be just due to a reasonably small sample size of modified nucleosides, or there may be a biological reason with respect to the methyltransferase that is beyond the scope of our

examination.

This work with modification signatures lends some credence to the idea that modifications themselves may be able to be identifying in a statistical learning fashion. A current downside to annotation and analysis is that it requires human intervention, but the closer we can get to identifying sites using multiple criteria, the closer we can come to bridging bioinformatic searches with established methods of data mining to push the current limitations on modification analysis.

CHAPTER 7

CONCLUSIONS, FUTURE DIRECTIONS, AND CHALLENGES

7.1 Conclusions

Regulation of RNA modifications plays a crucial role in dynamic tuning of cellular function. For mRNA events, heat shock has been shown to induce pseudouridylation in up-regulated mRNAs during heat shock, one could posit that the pseudouridine either induces stability or causes localization changes in mRNAs[102]. Similarly, global reprogramming of m6A has been shown to affect splicing events by causing local structure switches in mRNA[66], which in turn recruits protein factors from hnRNP complexes. There are several well-documented examples of stresses in various organisms causing a shift in modification status of specific positions in tRNA. This fluctuation in tRNA alone can cause events such as tRNA fragmentation which functions as a global signal, or in another instance the modification change in the anticodon allows for dynamic recoding of translational events. Because of the myriad of modified positions in tRNA, each with a potential host of protein partners, this same dynamic regulation is likely to occur for each.

For dynamic events in RNA modifications, this requires both a writer protein to establish and install a modified nucleotide, and an eraser protein to negate the work of the writer protein. For tRNA, we establish evidence of an AlkB homologue binding to mitochondrial tRNA targets in vivo and demethylating tRNA modifications in vitro. This protein-RNA interaction is what sets the foundation for tRNA modification dynamics at first. We now have a writer and eraser pair for a modification, which has only been observed for the m6A modification in mRNA and m^5C in DNA.

My work with establishing methods for tRNA modification dynamics highlights fundamental modifications that may be at the center of dynamic regulation pathways. From both

the reverse transcriptase primer extension work and the bioinformatics with tRNA sequencing, we find evidence for hypomodified sites of modification, evidence for trends in distinct modification patterns in different cell lines, and also evidence for novel methylated sites in tRNA. All of these results show that there is a rather uncharted landscape for tRNA modifications in terms of both the dynamics of stoichiometry and also for modifications that have yet to be curated and annotated.

Lastly, my bioinformatic work regarding modification signatures poses an interesting proof of concept for identifying sites of modification. While modification signature had been investigated for a particular modification, our work shows that different reverse transcriptases act on modifications in different sequence contexts to generate a wealth of distinct information for each modified nucleoside. Furthermore, even chemical adducts such as CMC in pseudouridine can provide signature. This facet of the modification and RT interaction can be utilized in the discovery of modifications in the transcriptome.

7.2 Challenges

For many modifications such as m6A, the coupling of biochemical tools such as antibodies with high throughput sequencing leads to sufficient identification across the transcriptome. This strategy has successfully identified myriad sites, For other modifications, chemically modifying the respective sites leads to similar hallmarks, especially if it is possible to further enrich for a modification. Conversely, chemically modifying sites changes one nucleobase to another, in which distinctly identifying sites requires a comparative analysis similar to what we use for tRNA-sequencing. Inherently, though, we are currently limited in discovery by the depth of power of sequencing.

Naturally, we are reliant on technology not only for the analysis of our biological data but now for the discovery therein. This poses problems with high-throughput sequencing

data when the limitations of the hardware, in this case the sequencer, can allow for only so much depth of coverage. What this means is that confidence in the discovery of sites through sequencing data is low due to a lack of coverage. Lack of coverage in this event means that truncation events may occur randomly due to the nature of the experiment as opposed to defined RT arrests. Lack of coverage also means that stochastic mutations that are latent within the sequencer can cause error in identification, leading to many more false positives. For instance, in analysis of our own data, without a difference of modification index of .15, the number of sites identified increased significantly, which would have caused a spurious misrepresentation of the number of modified sites in tRNA. And to avoid using these stringent criteria for identification is tantamount to fraudulent behavior, as anything less would mean needless false positives and misrepresentations of the data. There are those who believe that events that occur 1 in 20 (or even lower, 1 in 50) - meaning one mutation or one truncation in 20 reads - can constitute an event of modification. However, this absurd representation of the data means that for very low coverage mRNAs, the stochasticity of error inherent to the machine can lead to calling of sites that should be well below the threshold of identification and the statistical rigor necessary.

7.3 Future Directions

The contemporary world received a bit of shock when the program AlphaGo was able to beat one of the world's strongest players in a five game match, a feat people on the side of AI development and those well-versed in the ancient game thought near impossible[105]. However, beneath the facade of Google's strongest computing was an intricate algorithm for learning and understanding the game's layers of complexity. The program was able to overcome the natural limitations of the game's enormity of search space with a combination of ingenuity from the programmers. Using supervised learning algorithms from thousands of expert games coupled with the ability to simulate games of self-play, it was able to enforce certain value networks to ascertain with little doubt the correct move.

We may ask why this is important. Surprisingly, the field of bioinformatics contains much of the similar groundwork that was present for AlphaGo's triumph. Instead of thousands of games of expert play, we can utilize supervised learning to sift through the already petabytes of human genome data that exist from multiple cancer lines, myriad of tissue samples, and various reaction conditions utilizing common reverse transcriptases. In fact, the development of a neural network for discovery of modifications would really entail simple scoring algorithms based on inputs readily available from analytic and data mining techniques used in bioinformatics (Figure 7.1). These inputs are derived experimentally by generating sequencing data for each possible Watson-Crick modification (m^1A, m^3C, m^1G, etc.) using oligonucleotides containing singular modification sites with varying sequence contexts, i.e., NN(m^1G)NN. One can devise a set of training data for such a statistical learning method. By combining various percentages of modified and unmodified oligonucleotides under standard conditions, we can create a calibration curve to flesh out the various likelihoods of cDNA truncations and nucleotide misincorporations for every modification to fit our validated data.

Moving forward, the modification field has at its fingertips many tools with which it can push the boundary of discovery. The tantalizing prospects of discovering every modification is only limited by the biochemical tools at the field's disposal. But, by incorporating the potent power of machine learning with the vast reservoirs of sequencing data and biochemical knowledge, there should be no obstacles that remain for discovering the modes of dynamic modification regulation in the cell.

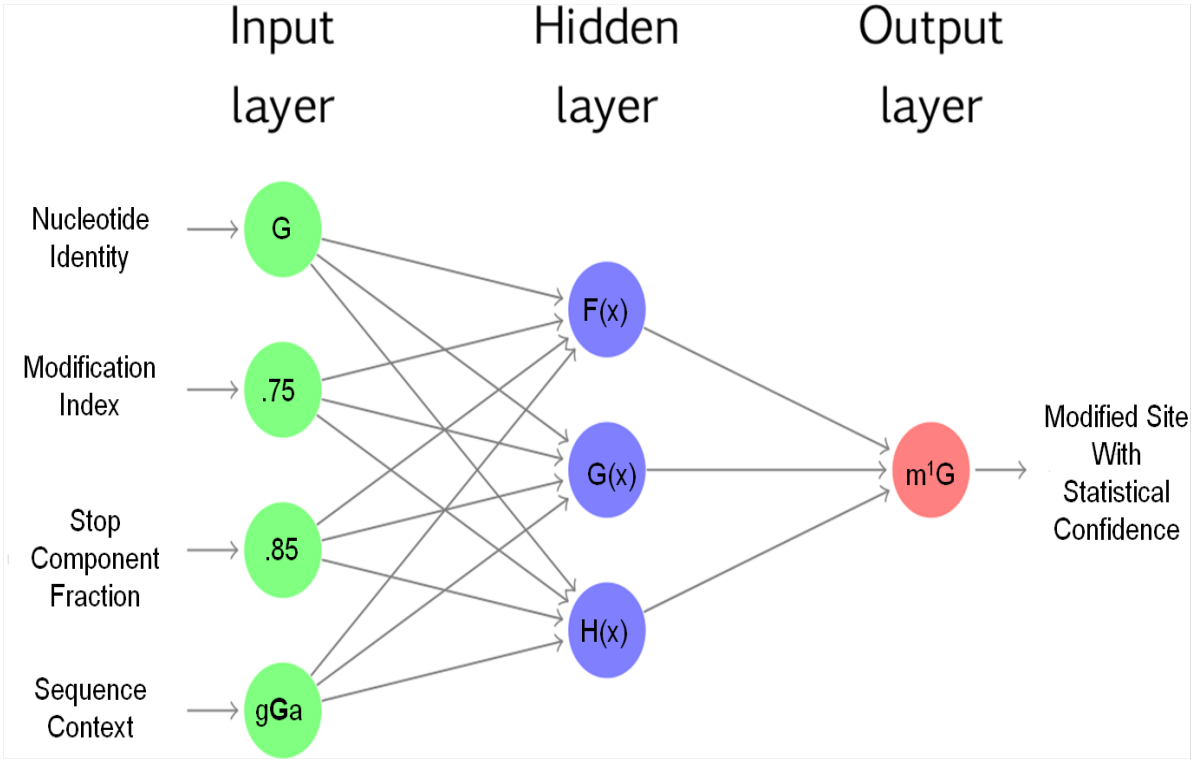

Figure 7.1: A theoretical example of how sequencing data analysis could be incorporated into a neural network. The input layer is composed of what prior knowledge from the analysis, based on standard bioinformatic results. This would include what the nucleotide identity for a putative modified site would be, the modification index at a site, how much of the modification index is stopped cDNA reads relative to RT misincorporations, and what the sequence context of the putative position is. The hidden layer is comprised of weighted functions that score each of these inputs. The combined results of the weighted functions result in an output discerned based on some supervised knowledge. In this case, using modification index information from Figure 6.2 with our example, we would expect the result to show that our site is a 1-methylguanosine.

REFERENCES

[1] J. A. Abbott, C. S. Francklyn, and S. M. Robey-Bond. Transfer rna and human disease. *Front Genet*, 5:158, 2014.

[2] P. F. Agris, F. A. Vendeix, and W. D. Graham. trna's wobble decoding of the genome: 40 years of modification. *J Mol Biol*, 366(1):1–13, 2007.

[3] A. A. Alcivar, L. E. Hake, C. F. Millette, J. M. Trasler, and N. B. Hecht. Mitochondrial gene expression in male germ cells of the mouse. *Dev Biol*, 135(2):263–71, 1989.

[4] A. Alexandrov, I. Chernyakov, W. Gu, S. L. Hiley, T. R. Hughes, E. J. Grayhack, and E. M. Phizicky. Rapid trna decay can result from lack of nonessential modifications. *Mol Cell*, 21(1):87–96, 2006.

[5] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–10, 1990.

[6] Q. M. Anstee and C. P. Day. S-adenosylmethionine (same) therapy in liver disease: a review of current evidence and clinical utility. *J Hepatol*, 57(5):1097–109, 2012.

[7] J. F. Atkins and G. R. Bjork. A gripping tale of ribosomal frameshifting: extragenic suppressors of frameshift mutations spotlight p-site realignment. *Microbiol Mol Biol Rev*, 73(1):178–210, 2009.

[8] B. L. Bass. Rna editing by adenosine deaminases that act on rna. *Annu Rev Biochem*, 71:817–46, 2002.

[9] P. J. Batista, B. Molinie, J. Wang, K. Qu, J. Zhang, L. Li, D. M. Bouley, E. Lujan, B. Haddad, K. Daneshvar, A. C. Carter, R. A. Flynn, C. Zhou, K. S. Lim, P. Dedon, M. Wernig, A. C. Mullen, Y. Xing, C. C. Giallourakis, and H. Y. Chang. m6a rna modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell*, 15(6):707–19, 2014.

[10] G. R. Bjork, P. M. Wikstrom, and A. S. Bystrom. Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science*, 244(4907):986–9, 1989.

[11] S. Blanco, S. Dietmann, J. V. Flores, S. Hussain, C. Kutter, P. Humphreys, M. Lukk, P. Lombard, L. Treps, M. Popis, S. Kellner, S. M. Holter, L. Garrett, W. Wurst, L. Becker, T. Klopstock, H. Fuchs, V. Gailus-Durner, M. Hrabe de Angelis, R. T. Karadottir, M. Helm, J. Ule, J. G. Gleeson, D. T. Odom, and M. Frye. Aberrant methylation of trnas links cellular stress to neuro-developmental disorders. *EMBO J*, 33(18):2020–39, 2014.

[12] J. A. Bokar, M. E. Rath-Shambaugh, R. Ludwiczak, P. Narayan, and F. Rottman. Characterization and partial purification of mrna n6-adenosine methyltransferase from hela cell nuclei. internal mrna methylation requires a multisubunit complex. *J Biol Chem*, 269(26):17697–704, 1994.

[13] A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–20, 2014.

[14] B. S. Budde, Y. Namavar, P. G. Barth, B. T. Poll-The, G. Nurnberg, C. Becker, F. van Ruissen, M. A. Weterman, K. Fluiter, E. T. te Beek, E. Aronica, M. S. van der Knaap, W. Hohne, M. R. Toliat, Y. J. Crow, M. Steinling, T. Voit, F. Roelenso, W. Brussel, K. Brockmann, M. Kyllerman, E. Boltshauser, G. Hammersen, M. Willemsen, L. Basel-Vanagaite, I. Krageloh-Mann, L. S. de Vries, L. Sztriha, F. Muntoni, C. D. Ferrie, R. Battini, R. C. Hennekam, E. Grillo, F. A. Beemer, L. M. Stoets, B. Wollnik, P. Nurnberg, and F. Baas. trna splicing endonuclease mutations cause pontocerebellar hypoplasia. *Nat Genet*, 40(9):1113–8, 2008.

[15] W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. Vendeix, D. Fabris, and P. F. Agris. The rna modification database, rnamdb: 2011 update. *Nucleic Acids Res*, 39(Database issue):D195–201, 2011.

[16] T. M. Carlile, M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoli, and W. V. Gilbert. Pseudouridine profiling reveals regulated mrna pseudouridylation in yeast and human cells. *Nature*, 515(7525):143–6, 2014.

[17] C. T. Chan, W. Deng, F. Li, M. S. DeMott, I. R. Babu, T. J. Begley, and P. C. Dedon. Highly predictive reprogramming of trna modifications is linked to selective expression of codon-biased genes. *Chem Res Toxicol*, 28(5):978–88, 2015.

[18] C. T. Chan, M. Dyavaiah, M. S. DeMott, K. Taghizadeh, P. C. Dedon, and T. J. Begley. A quantitative systems approach reveals dynamic control of trna modifications during cellular stress. *PLoS Genet*, 6(12):e1001247, 2010.

[19] C. T. Chan, Y. L. Pang, W. Deng, I. R. Babu, M. Dyavaiah, T. J. Begley, and P. C. Dedon. Reprogramming of trna modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat Commun*, 3:937, 2012.

[20] J. C. Chan, J. A. Yang, M. J. Dunn, P. F. Agris, and T. W. Wong. The nucleotide sequence of a glutamate trna from rat liver. *Nucleic Acids Res*, 10(15):4605–8, 1982.

[21] P. P. Chan and T. M. Lowe. Gtrnadb: a database of transfer rna genes detected in genomic sequence. *Nucleic Acids Res*, 37(Database issue):D93–7, 2009.

[22] P. Chen, G. Jager, and B. Zheng. Transfer rna modifications and genes for modifying enzymes in arabidopsis thaliana. *BMC Plant Biol*, 10:201, 2010.

[23] D. A. Clayton and G. S. Shadel. Isolation of mitochondria from cells and tissues. *Cold Spring Harb Protoc*, 2014(10):pdb top074542, 2014.

[24] A. E. Cozen, E. Quartley, A. D. Holmes, E. Hrabeta-Robinson, E. M. Phizicky, and T. M. Lowe. Arm-seq: Alkb-facilitated rna methylation sequencing reveals a complex landscape of modified trna fragments. *Nat Methods*, 12(9):879–84, 2015.

[25] F. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–3, 1970.

[26] J. J. Dalluge, T. Hashizume, A. E. Sopchik, J. A. McCloskey, and D. R. Davis. Conformational flexibility in rna: the role of dihydrouridine. *Nucleic Acids Res*, 24(6):1073–9, 1996.

[27] W. A. Decatur and M. J. Fournier. rrna modifications and ribosome function. *Trends Biochem Sci*, 27(7):344–51, 2002.

[28] R. Desrosiers, K. Friderici, and F. Rottman. Identification of methylated nucleosides in messenger rna from novikoff hepatoma cells. *Proc Natl Acad Sci U S A*, 71(10):3971–5, 1974.

[29] J. M. Dewe, J. M. Whipple, I. Chernyakov, L. N. Jaramillo, and E. M. Phizicky. The yeast rapid trna decay pathway competes with elongation factor 1a for substrate trnas and acts on trnas lacking one or more of several modifications. *RNA*, 18(10):1886–96, 2012.

[30] K. A. Dittmar, M. A. Sorensen, J. Elf, M. Ehrenberg, and T. Pan. Selective charging of trna isoacceptors induced by amino-acid starvation. *EMBO Rep*, 6(2):151–7, 2005.

[31] D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, and G. Rechavi. Topology of the human and mouse m6a rna methylomes revealed by m6a-seq. *Nature*, 485(7397):201–6, 2012.

[32] D. Dominissini, S. Nachtergaele, S. Moshitch-Moshkovitz, E. Peer, N. Kol, M. S. Ben-Haim, Q. Dai, A. Di Segni, M. Salmon-Divon, W. C. Clark, G. Zheng, T. Pan, O. Solomon, E. Eyal, V. Hershkovitz, D. Han, L. C. Dore, N. Amariglio, G. Rechavi, and C. He. The dynamic n(1)-methyladenosine methylome in eukaryotic messenger rna. *Nature*, 530(7591):441–6, 2016.

[33] L. Endres, P. C. Dedon, and T. J. Begley. Codon-biased translation can be regulated by wobble-base trna modification systems during cellular stress responses. *RNA Biol*, 12(6):603–14, 2015.

[34] Y. Fu, D. Dominissini, G. Rechavi, and C. He. Gene expression regulation mediated through reversible m(6)a rna methylation. *Nat Rev Genet*, 15(5):293–306, 2014.

[35] H. B. Gamper, I. Masuda, M. Frenkel-Morgenstern, and Y. M. Hou. Maintenance of protein synthesis reading frame by ef-p and m(1)g37-trna. *Nat Commun*, 6:7226, 2015.

[36] A. P. Gerber and W. Keller. An adenosine deaminase that generates inosine at the wobble position of trnas. *Science*, 286(5442):1146–9, 1999.

[37] M. G. Goll and T. H. Bestor. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem*, 74:481–514, 2005.

[38] J. M. Goodenbour and T. Pan. Diversity of trna genes in eukaryotes. *Nucleic Acids Res*, 34(21):6137–46, 2006.

[39] T. E. Gorochowski, Z. Ignatova, R. A. Bovenberg, and J. A. Roubos. Trade-offs between trna abundance and mrna secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res*, 43(6):3022–32, 2015.

[40] H. Grosjean, S. Auxilien, F. Constantinesco, C. Simon, Y. Corda, H. F. Becker, D. Foiret, A. Morin, Y. X. Jin, M. Fournier, and J. L. Fourrey. Enzymatic conversion of adenosine to inosine and to n1-methylinosine in transfer rnas: a review. *Biochimie*, 78(6):488–501, 1996.

[41] Henri Grosjean. *Modification and editing of RNA: historical overview and important facts to remember*, pages 1–22. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[42] C. Gu, T. J. Begley, and P. C. Dedon. trna modifications regulate translation during cellular stress. *FEBS Lett*, 588(23):4287–96, 2014.

[43] T. G. Hagervall, T. M. Tuohy, J. F. Atkins, and G. R. Bjork. Deficiency of 1-methylguanosine in trna from salmonella typhimurium induces frameshifting by quadruplet translocation. *J Mol Biol*, 232(3):756–65, 1993.

[44] R. Hauenschild, L. Tserovski, K. Schmid, K. Thuring, M. L. Winz, S. Sharma, K. D. Entian, L. Wacheul, D. L. Lafontaine, J. Anderson, J. Alfonzo, A. Hildebrandt, A. Jaschke, Y. Motorin, and M. Helm. The reverse transcription signature of n-1-methyladenosine in rna-seq is sequence dependent. *Nucleic Acids Res*, 43(20):9950–64, 2015.

[45] M. Helm, C. Florentz, A. Chomyn, and G. Attardi. Search for differences in post-transcriptional modification patterns of mitochondrial dna-encoded wild-type and mutant human trnalys and trnaleu(uur). *Nucleic Acids Res*, 27(3):756–63, 1999.

[46] M. Helm, R. Giege, and C. Florentz. A watson-crick base-pair-disrupting methyl group (m1a9) is sufficient for cloverleaf folding of human mitochondrial trnalys. *Biochemistry*, 38(40):13338–46, 1999.

[47] T. Uechi T. Higa-Nakamine, S.Suzuki, A. Chakraborty, Y. Nakajima, M. Nakamura, N. Hirano, and N. Kenmochi. Loss of ribosomal rna modification causes developmental defects in zebrafish. *Nucleic Acids Res*, 40(1):391–8, 2012.

[48] S. L. Hiley, J. Jackman, T. Babak, M. Trochesset, Q. D. Morris, E. Phizicky, and T. R. Hughes. Detection and discovery of rna modifications using microarrays. *Nucleic Acids Res*, 33(1):e2, 2005.

[49] H. T. Hornig-Do, G. Gunther, M. Bust, P. Lehnartz, A. Bosio, and R. J. Wiesner. Isolation of functional pure mitochondria by superparamagnetic microbeads. *Anal Biochem*, 389(1):1–5, 2009.

[50] G. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, C. Yi, T. Lindahl, T. Pan, Y. G. Yang, and C. He. N6-methyladenosine in nuclear rna is a major substrate of the obesity-associated fto. *Nat Chem Biol*, 7(12):885–7, 2011.

[51] F. Juhling, M. Morl, R. K. Hartmann, M. Sprinzl, P. F. Stadler, and J. Putz. trnadb 2009: compilation of trna sequences and trna genes. *Nucleic Acids Res*, 37(Database issue):D159–62, 2009.

[52] S. Kadaba, A. Krueger, T. Trice, A. M. Krecic, A. G. Hinnebusch, and J. Anderson. Nuclear surveillance and degradation of hypomodified initiator trnamet in s. cerevisiae. *Genes Dev*, 18(11):1227–40, 2004.

[53] S. Kellner, J. Burhenne, and M. Helm. Detection of rna modifications. *RNA Biol*, 7(2):237–47, 2010.

[54] S. Kirchner and Z. Ignatova. Emerging roles of trna in adaptive translation, signalling dynamics and disease. *Nat Rev Genet*, 16(2):98–112, 2015.

[55] M. Kozak. Regulation of translation via mrna structure in prokaryotes and eukaryotes. *Gene*, 2005.

[56] E. B. Kramer and A. K. Hopper. Retrograde transfer rna nuclear import provides a new level of trna quality control in saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, 110(52):21042–7, 2013.

[57] P. Kumar, S. B. Mudunuri, J. Anaya, and A. Dutta. trfdb: a database for transfer rna fragments. *Nucleic Acids Res*, 43(Database issue):D141–5, 2015.

[58] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol*, 10(3):R25, 2009.

[59] S. Li and C. E. Mason. The pivotal regulatory landscape of rna modifications. *Annu Rev Genomics Hum Genet*, 15:127–50, 2014.

[60] X. Li, X. Xiong, K. Wang, L. Wang, X. Shu, S. Ma, and C. Yi. Transcriptome-wide mapping reveals reversible and dynamic n(1)-methyladenosine methylome. *Nat Chem Biol*, 12(5):311–6, 2016.

[61] Y. Li and T. O. Tollefsbol. Dna methylation detection: bisulfite genomic sequencing analysis. *Methods Mol Biol*, 791:11–21, 2011.

[62] X. H. Liang, Q. Liu, and M. J. Fournier. Loss of rrna modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rrna processing. *RNA*, 15(9):1716–28, 2009.

[63] M. W. Libbrecht and W. S. Noble. Machine learning applications in genetics and genomics. *Nat Rev Genet*, 16(6):321–32, 2015.

[64] P. A. Limbach and M. J. Paulines. Going global: the new era of mapping modifications in rna. *Wiley Interdiscip Rev RNA*, 2016.

[65] B. Linder, A. V. Grozhik, A. O. Olarerin-George, C. Meydan, C. E. Mason, and S. R. Jaffrey. Single-nucleotide-resolution mapping of m6a and m6am throughout the transcriptome. *Nat Methods*, 12(8):767–72, 2015.

[66] N. Liu, Q. Dai, G. Zheng, C. He, M. Parisien, and T. Pan. N(6)-methyladenosine-dependent rna structural switches regulate rna-protein interactions. *Nature*, 518(7540):560–4, 2015.

[67] N. Liu, M. Parisien, Q. Dai, G. Zheng, C. He, and T. Pan. Probing n6-methyladenosine rna modification status at single nucleotide resolution in mrna and long noncoding rna. *RNA*, 19(12):1848–56, 2013.

[68] T. M. Lowe and P. P. Chan. trnascan-se on-line: integrating search and context for analysis of transfer rna genes. *Nucleic Acids Res*, 2016.

[69] M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, and H. Grosjean. Modomics: a database of rna modification pathways–2013 update. *Nucleic Acids Res*, 41(Database issue):D262–7, 2013.

[70] N. Manickam, K. Joshi, M. J. Bhatt, and P. J. Farabaugh. Effects of trna modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Res*, 44(4):1871–81, 2016.

[71] R. L. Maute, C. Schneider, P. Sumazin, A. Holmes, A. Califano, K. Basso, and R. Dalla-Favera. trna-derived microrna modulates proliferation and the dna damage response and is down-regulated in b cell lymphoma. *Proc Natl Acad Sci U S A*, 110(4):1404–9, 2013.

[72] K. M. McKenney and J. D. Alfonzo. From prebiotics to probiotics: The evolution and functions of trna modifications. *Life (Basel)*, 6(1), 2016.

[73] K. D. Meyer and S. R. Jaffrey. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat Rev Mol Cell Biol*, 15(5):313–26, 2014.

[74] K. D. Meyer, D. P. Patil, J. Zhou, A. Zinoviev, M. A. Skabkin, O. Elemento, T. V. Pestova, S. B. Qian, and S. R. Jaffrey. 5' utr m(6)a promotes cap-independent translation. *Cell*, 163(4):999–1010, 2015.

[75] K. D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason, and S. R. Jaffrey. Comprehensive analysis of mrna methylation reveals enrichment in 3' utrs and near stop codons. *Cell*, 149(7):1635–46, 2012.

[76] Y. Mishina, C. H. Lee, and C. He. Interaction of human and bacterial alkb proteins with dna as probed through chemical cross-linking studies. *Nucleic Acids Res*, 32(4):1548–54, 2004.

[77] Y. Motorin, F. Lyko, and M. Helm. 5-methylcytosine in rna: detection, enzymatic formation and biological functions. *Nucleic Acids Res*, 38(5):1415–30, 2010.

[78] Y. Motorin, S. Muller, I. Behm-Ansmant, and C. Branlant. Identification of modified residues in rnas by reverse transcription-based methods. *Methods Enzymol*, 425:21–53, 2007.

[79] N. Netzer, J. M. Goodenbour, A. David, K. A. Dittmar, R. B. Jones, J. R. Schneider, D. Boone, E. M. Eves, M. R. Rosner, J. S. Gibbs, A. Embry, B. Dolan, S. Das, H. D. Hickman, P. Berglund, J. R. Bennink, J. W. Yewdell, and T. Pan. Innate immune and chemically triggered oxidative stress modifies translational fidelity. *Nature*, 462(7272):522–6, 2009.

[80] K. Nishikura. Functions and regulation of rna editing by adar deaminases. *Annu Rev Biochem*, 2010.

[81] L. M. Nordstrand, J. Svard, E. Larsen, A. Nilsen, R. Ougland, K. Furu, G. F. Lien, T. Rognes, S. H. Namekawa, J. T. Lee, and A. Klungland. Mice lacking alkbh1 display sex-ratio distortion and unilateral eye defects. *PLoS One*, 5(11):e13827, 2010.

[82] T. Ohira and T. Suzuki. Retrograde nuclear import of trna precursors is required for modified base biogenesis in yeast. *Proc Natl Acad Sci U S A*, 108(26):10502–7, 2011.

[83] R. Ougland, D. Lando, I. Jonson, J. A. Dahl, M. N. Moen, L. M. Nordstrand, T. Rognes, J. T. Lee, A. Klungland, T. Kouzarides, and E. Larsen. Alkbh1 is a histone h2a dioxygenase involved in neural differentiation. *Stem Cells*, 30(12):2672–82, 2012.

[84] R. Ougland, C. M. Zhang, A. Liiv, R. F. Johansen, E. Seeberg, Y. M. Hou, J. Remme, and P. O. Falnes. Alkb restores the biological function of mrna and trna inactivated by chemical methylation. *Mol Cell*, 16(1):107–16, 2004.

[85] F. Ozsolak and P. M. Milos. Rna sequencing: advances, challenges and opportunities. *Nat Rev Genet*, 12(2):87–98, 2011.

[86] P. S. Pallan, C. Kreutz, S. Bosio, R. Micura, and M. Egli. Effects of n2,n2-dimethylguanosine on rna structure and stability: crystal structure of an rna duplex with tandem m2 2g:a pairs. *RNA*, 14(10):2125–35, 2008.

[87] Y. L. Pang, R. Abo, S. S. Levine, and P. C. Dedon. Diverse cell stresses induce unique patterns of trna up- and down-regulation: trna-seq for quantifying changes in trna copy number. *Nucleic Acids Res*, 42(22):e170, 2014.

[88] M. Parisien, X. Wang, 2nd Perdrizet, G., C. Lamphear, C. A. Fierke, K. C. Maheshwari, M. J. Wilde, T. R. Sosnick, and T. Pan. Discovering rna-protein interactome by using chemical context profiling of the rna-protein interface. *Cell Rep*, 3(5):1703–13, 2013.

[89] M. Pavon-Eternod, S. Gomes, R. Geslain, Q. Dai, M. R. Rosner, and T. Pan. trna over-expression in breast cancer and functional consequences. *Nucleic Acids Res*, 37(21):7268–80, 2009.

[90] D. A. Peattie. Direct chemical method for sequencing rna. *Proc Natl Acad Sci U S A*, 76(4):1760–4, 1979.

[91] H. Peng, J. Shi, Y. Zhang, H. Zhang, S. Liao, W. Li, L. Lei, C. Han, L. Ning, Y. Cao, Q. Zhou, Q. Chen, and E. Duan. A novel class of trna-derived small rnas extremely enriched in mature mouse sperm. *Cell Res*, 22(11):1609–12, 2012.

[92] E. M. Phizicky and J. D. Alfonzo. Do all modifications benefit all trnas? *FEBS Lett*, 584(2):265–71, 2010.

[93] E. M. Phizicky and A. K. Hopper. trna biology charges to the front. *Genes Dev*, 24(17):1832–60, 2010.

[94] J. A. Reuter, D. V. Spacek, and M. P. Snyder. High-throughput sequencing technologies. *Mol Cell*, 58(4):586–97, 2015.

[95] M. Roovers, J. Wouters, J. M. Bujnicki, C. Tricot, V. Stalon, H. Grosjean, and L. Droogmans. A primordial rna modification enzyme: the case of trna (m1a) methyltransferase. *Nucleic Acids Res*, 32(2):465–76, 2004.

[96] P. Ryvkin, Y. Y. Leung, I. M. Silverman, M. Childress, O. Valladares, I. Dragomir, B. D. Gregory, and L. S. Wang. Hamr: high-throughput annotation of modified ribonucleotides. *RNA*, 19(12):1684–92, 2013.

[97] M. Saikia, Y. Fu, M. Pavon-Eternod, C. He, and T. Pan. Genome-wide analysis of n1-methyl-adenosine modification in human trnas. *RNA*, 16(7):1317–27, 2010.

[98] Y. Saletore, K. Meyer, J. Korlach, I. D. Vilfan, S. Jaffrey, and C. E. Mason. The birth of the epitranscriptome: deciphering the function of rna modifications. *Genome Biol*, 13(10):175, 2012.

[99] T. Samuelsson, T. Boren, T. I. Johansen, and F. Lustig. Properties of a transfer rna lacking modified nucleosides. *J Biol Chem*, 263(27):13692–9, 1988.

[100] F. Sasarman and E. A. Shoubridge. Radioactive labeling of mitochondrial translation products in cultured cells. *Methods Mol Biol*, 837:207–17, 2012.

[101] S. Schwartz. Cracking the epitranscriptome. *RNA*, 2016.

[102] S. Schwartz, D. A. Bernstein, M. R. Mumbach, M. Jovanovic, R. H. Herbst, B. X. Leon-Ricardo, J. M. Engreitz, M. Guttman, R. Satija, E. S. Lander, G. Fink, and A. Regev. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncrna and mrna. *Cell*, 159(1):148–62, 2014.

[103] Raghuvaran Shanmugam, Jacob Fierer, Steffen Kaiser, Mark Helm, Tomasz P. Jurkowski, and Albert Jeltsch. Cytosine methylation of trna-asp by dnmt2 has a role in translation of proteins containing poly-asp sequences. *Cell Discovery*, 1:15010, 2015.

[104] N. A. Siegfried, S. Busan, G. M. Rice, J. A. Nelson, and K. M. Weeks. Rna motif discovery by shape and mutational profiling (shape-map). *Nat Methods*, 11(9):959–65, 2014.

[105] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–9, 2016.

[106] C. X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C. H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, and C. He. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol*, 29(1):68–72, 2011.

[107] L. D. Stein, B. M. Knoppers, P. Campbell, G. Getz, and J. O. Korbel. Data analysis: Create a cloud commons. *Nature*, 523(7559):149–51, 2015.

[108] T. Suzuki. A complete landscape of post-transcriptional modifications in mammalian mitochondrial trnas. *Nucleic Acids Res*, 42(11):7346–57, 2014.

[109] T. Suzuki and A. Nagao. Human mitochondrial trnas: biogenesis, function, structural aspects, and diseases. *Annu Rev Genet*, 45:299–329, 2011.

[110] J. Talkish, G. May, Y. Lin, Jr. Woolford, J. L., and C. J. McManus. Mod-seq: high-throughput sequencing for chemical probing of rna structure. *RNA*, 20(5):713–20, 2014.

[111] A. G. Torres, E. Batlle, and L. Ribas de Pouplana. Role of trna modifications in human diseases. *Trends Mol Med*, 20(6):306–14, 2014.

[112] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter. Differential gene and transcript expression analysis of rna-seq experiments with tophat and cufflinks. *Nat Protoc*, 7(3):562–78, 2012.

[113] S. C. Trewick, T. F. Henshaw, R. P. Hausinger, T. Lindahl, and B. Sedgwick. Oxidative demethylation by escherichia coli alkb directly reverts dna base damage. *Nature*, 419(6903):174–8, 2002.

[114] L. Tserovski, V. Marchand, R. Hauenschild, F. Blanloeil-Oillo, M. Helm, and Y. Motorin. High-throughput sequencing for 1-methyladenosine (ma) mapping in rna. *Methods*, 2016.

[115] K. Tsujikawa, K. Koike, K. Kitae, A. Shinkawa, H. Arima, T. Suzuki, M. Tsuchiya, Y. Makino, T. Furukawa, N. Konishi, and H. Yamamoto. Expression and sub-cellular localization of human abh family molecules. *J Cell Mol Med*, 11(5):1105–16, 2007.

[116] F. Tuorto, R. Liebers, T. Musch, M. Schaefer, S. Hofmann, S. Kellner, M. Frye, M. Helm, G. Stoecklin, and F. Lyko. Rna cytosine methylation by dnmt2 and nsun2 promotes trna stability and protein synthesis. *Nat Struct Mol Biol*, 19(9):900–5, 2012.

[117] J. Urbonavicius, Q. Qian, J. M. Durand, T. G. Hagervall, and G. R. Bjork. Improvement of reading frame maintenance is a common function for several trna modifications. *EMBO J*, 20(17):4863–73, 2001.

[118] I. D. Vilfan, Y. C. Tsai, T. A. Clark, J. Wegener, Q. Dai, C. Yi, T. Pan, S. W. Turner, and J. Korlach. Analysis of rna base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J Nanobiotechnology*, 11:8, 2013.

[119] T. Wai, A. Ao, X. Zhang, D. Cyr, D. Dufort, and E. A. Shoubridge. The role of mitochondrial dna copy number in mammalian fertility. *Biol Reprod*, 83(1):52–62, 2010.

[120] X. Wang and C. He. Dynamic rna modifications in posttranscriptional regulation. *Mol Cell*, 56(1):5–12, 2014.

[121] F. Y. Wei, T. Suzuki, S. Watanabe, S. Kimura, T. Kaitsuka, A. Fujimura, H. Matsui, M. Atta, H. Michiue, M. Fontecave, K. Yamagata, and K. Tomizawa. Deficit of trna(lys) modification by cdkal1 causes the development of type 2 diabetes in mice. *J Clin Invest*, 121(9):3598–608, 2011.

[122] M. P. Westbye, E. Feyzi, P. A. Aas, C. B. Vagbo, V. A. Talstad, B. Kavli, L. Hagen, O. Sundheim, M. Akbari, N. B. Liabakk, G. Slupphaug, M. Otterlei, and H. E. Krokan. Human alkb homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in dna and rna. *J Biol Chem*, 283(36):25046–56, 2008.

[123] J. M. Whipple, E. A. Lane, I. Chernyakov, S. D'Silva, and E. M. Phizicky. The yeast rapid trna decay pathway primarily monitors the structural integrity of the acceptor and t-stems of mature trna. *Genes Dev*, 25(11):1173–84, 2011.

[124] X. Wu and G. Brewer. The regulation of mrna stability in mammalian cells: 2.0. *Gene*, 2012.

[125] C. Yarian, H. Townsend, W. Czestkowski, E. Sochacka, A. J. Malkiewicz, R. Guenther, A. Miskiewicz, and P. F. Agris. Accurate translation of the genetic code depends on trna modified nucleosides. *J Biol Chem*, 277(19):16391–5, 2002.

[126] C. Yi and T. Pan. Cellular dynamics of rna modification. *Acc Chem Res*, 44(12):1380–8, 2011.

[127] M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J. H. Min, P. Jin, B. Ren, and C. He. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, 149(6):1368–80, 2012.

[128] Y. Yue, J. Liu, and C. He. Rna n6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev*, 29(13):1343–55, 2015.

[129] J. M. Zaborske, V. L. DuMont, E. W. Wallace, T. Pan, C. F. Aquadro, and D. A. Drummond. A nutrient-driven trna modification alters translational fidelity and genome-wide protein coding across an animal genus. *PLoS Biol*, 12(12):e1002015, 2014.

[130] G. Zheng, J. A. Dahl, Y. Niu, P. Fedorcsak, C. M. Huang, C. J. Li, C. B. Vagbo, Y. Shi, W. L. Wang, S. H. Song, Z. Lu, R. P. Bosmans, Q. Dai, Y. J. Hao, X. Yang, W. M. Zhao, W. M. Tong, X. J. Wang, F. Bogdan, K. Furu, Y. Fu, G. Jia, X. Zhao, J. Liu, H. E. Krokan, A. Klungland, Y. G. Yang, and C. He. Alkbh5 is a mammalian rna demethylase that impacts rna metabolism and mouse fertility. *Mol Cell*, 49(1):18–29, 2013.

[131] G. Zheng, Y. Qin, W. C. Clark, Q. Dai, C. Yi, C. He, A. M. Lambowitz, and T. Pan. Efficient and quantitative high-throughput trna sequencing. *Nat Methods*, 12(9):835–7, 2015.