THE UNIVERSITY OF CHICAGO

DEVELOPMENT AND APPLICATIONS OF A HIGH-THROUGHPUT SMALL RNA SEQUENCING METHOD

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

GRADUATE PROGRAM IN BIOCHEMISTRY AND MOLECULAR BIOPHYSICS

BY

CHRISTOPHER PATRICK WATKINS

CHICAGO, ILLINOIS

AUGUST 2022

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am extremely grateful to the current and past members of the Pan lab, and to my advisor Dr. Tao Pan. I would also like to thank the members of my thesis committee: Dr. Jingyi Fei (Chair), Dr. A. Murat Eren, and Dr. Joseph A. Piccirilli.

Many people have contributed to this work. I would like to acknowledge Dr. Christopher D. Katanski for his development of our computation pipeline for MSR-seq; Dr. Wen Zhang for performing polysome profiling, Northern blotting, cell cultures, and much more; Adam Wylder for providing Western blots; Noah Peña for analyzing the libraries produced from cultured SARS-CoV-2 virions; Dr. Samuel E. Miller for developing and applying a computational pipeline for microbial and microbiome samples; Karen Lolans for her work in developing RNA extraction protocols for diverse samples; and Dr. Matthew Reyer for providing treated *E. coli* for analysis. I would like to thank members of the Simon lab at Mt. Sinai School of Medicine for providing cultured virions for library construction. Finally, my many thanks to the University of Chicago Sequencing Core for their service and assistance over several years during the development and application of MSR-seq.

This work was supported by the National Institutes of Health Chemical Biology Training Program grant RM1HG008935, the National Cancer Institute grant T32GM008720, the National Institutes of Health R01 grant R01GM113194, and the William M. Keck foundation. I would also like to thank the University of Chicago Biological Sciences Division for their generous support.

This thesis includes text and figures derived from articles published in *Nature Communications* by Springer Nature [1], *Frontiers in Cell and Developmental Biology* by Frontiers Media [2], and *Nucleic Acids Research* by Oxford University Press [3].

Abstract

The cellular transcriptome is predominantly composed of RNAs less than 200 nucleotides in length. These small RNAs participate in and regulate numerous biological processes, including translation, transcription, epigenetic reprogramming, apoptosis, cell-wall biosynthesis, natural product biosynthesis, and the cellular immune response. The most abundant small RNA is tRNA, comprising roughly 80% of cellular transcripts. Many classes of small RNAs, however have proven difficult to sequence owing to their high degree of secondary and tertiary structure, presence of epitranscriptomic modifications, and aminoacylation. Although sequencing methods have been developed over the past decade, they possess several drawbacks that make them impractical for studies both clinical and fundamental that require high throughputs and the ability to produce libraries from low quantities of material.

In this thesis, I detail the development and application of a high-throughput small RNA sequencing method termed MSR-seq. In Chapter 2, I describe the development of MSR-seq. In Chapter 3, I demonstrate the application of MSR-seq to studying translational regulation during stress response. In Chapter 4, I discuss a study performed to analyze the SARS-CoV-2 packageome of host and viral RNAs, the discover of a chemical treatment that renders several modifications detectable by reverse transcriptase error signatures, and new insights into the function of the host oral microbiome. In Chapter 5, I provide preliminary results on the integration of RNA structural mapping techniques to study the changes in small RNA interactions with other molecular partners. This work demonstrates the power of a new high-throughput small RNA sequencing method and its ability to inform and provide insight into previously intractable areas of biology.

Chapter 1

Introduction

Small RNAs, defined here as < 200 nucleotides in length, are critical biological regulators in every organism and constitute > 90% of cellular transcripts. Processes such as translation, transcription, cell-cell communication, apoptosis, pathogenicity, cell wall biosynthesis, viral regulation and immune suppression, and many others are regulated by small RNAs [4-11]. Multiple classes of small RNAs are observed in each kingdom of life, including transfer RNA (tRNA), tRNA fragments (tRF), the signal recognition particle RNA (SRP RNA), 5S ribosomal RNA (5S rRNA), microRNAs (miRNA), and bacterial small RNAs (sRNA). Small RNAs regulate biological processes through their abundance, post-transcriptional modification fractions, and charging fractions (for tRNA), in addition to their structure and sequence [4,5]; thus, small RNAs provide a rich source of potential insight into biology. Despite their demonstrated importance, small RNAs, particularly tRNAs and tRFs, have remained understudied transcriptome-wide due to numerous challenges in sequencing method development.

In this work, I have developed and implemented a multiplexed solid-support platform to sequence small RNAs and their modifications in a large variety of biological systems [1]. I have also developed a reaction condition that enables the detection of RNA modifications previously invisible to or inhibitory of reverse transcription [3]. This method enables simultaneous, integrative analysis of many classes of RNAs and simultaneous, quantitative detection of RNA modifications at specific sites; these data have provided insight into cellular stress response, potential roles of RNA modifications in translation and tRF biogenesis [1], viral RNA packaging [2], and microbiome characterization and activity. I have also developed a chemical treatment that

enables the detection of queuosine and 2-thio tRNA modifications, significantly expanding our capacity to report on epitranscriptomic changes [3]

In this Chapter, I will discuss some of the major small RNA classes and detail well-known aspects of their biological functions. In addition, I will discuss many of the post-transcriptional modifications that are found in small RNAs and that regulate their stability and function. Lastly, I will outline the state of the small RNA sequencing field with respect to existing methods, their benefits and limitations, and the challenges and opportunities before the field.

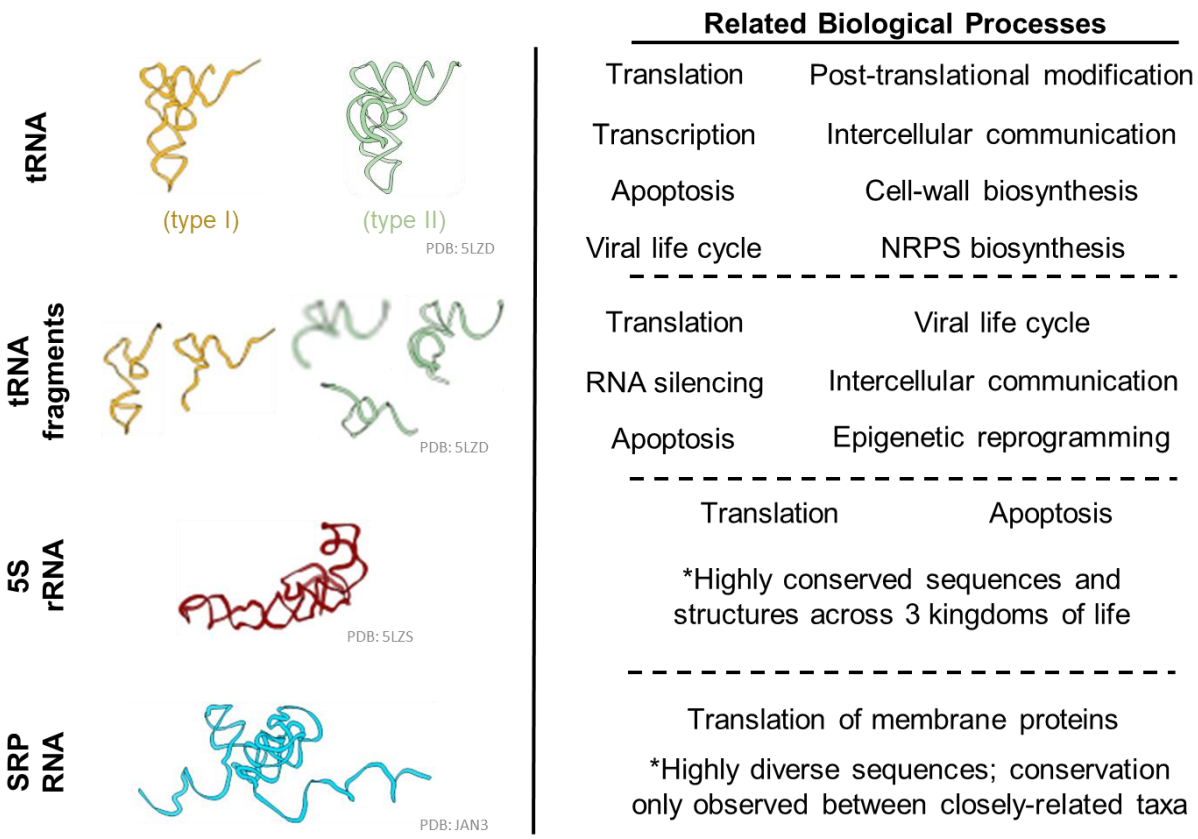

Figure 1.1: Prominent small RNAs and their biological roles. (Note: SRP RNA shown here is 7S.)

1.1 Small RNAs & Their Regulatory Roles

1.1.1. tRNA

tRNA is the most abundant cellular RNA class, comprising roughly 80% of transcripts [4]. In length, tRNAs range from 70-100 nucleotides, although exceptions can be found [12,13]. Structurally, tRNAs consist of 1) an acceptor step of with 7-9 base pairs, a discriminator base (often adenosine [A]), and a 3' CCA tail that is added post-transcriptionally; 2) a T arm with 4-5 nucleotides in the stem and a TΨC loop of 7 nucleotides (Ψ is pseudouridine, a post-transcriptional modification); 3) a variable loop of 3-5 nucleotides in type I tRNAs and 10+ nucleotides in type II tRNAs; 4) an anticodon stem-loop with a highly conserved 5 nucleotides in the stem and 7 nucleotides in the loop, including the anticodon; and 5) the D arm, consisting of a stem of 3-5 nucleotides and a loop of 4-12 nucleotides containing one or more dihydrouridines (D), another post-transcriptional modification. These strongly conserved lengths lead in part to the predictable tRNA cloverleaf secondary structure and the conserved 'bent arm' tertiary structure, where the T arm folds over to form the outer elbow region and the D arm folds under the acceptor stem. (See Ref. 12,14 for an overview.)

tRNAs are categorized first by the amino acid they are charged with (e.g., $tRNA^{Ala}$, $tRNA^{Cys}$, etc.), then into isoacceptors, which are charged with the same amino acids but possess a different anticodon, and lastly into isodecoders, which have the same anticodon but differ in sequence at other positions [5]. In general, lower eukaryotes, including *D. melanogaster*, have fewer tRNAs genes than mammals, as well as substantially fewer isodecoders; in contrast, mice and humans have high genetic diversity among their hundreds of tRNA genes [15].

Critical to the function and stability of tRNAs are their post-transcriptional modifications. Currently, there are more than 150 known RNA modifications, over half of which are found on

tRNAs [16-18]. These modifications range in chemical diversity and size from pseudouridine (Ψ), an isomer of uridine, and Watson-Crick face methylations to ring expansions and additions such as queuosine (Q) and wybutosine (yW) to thio- and phosphorous-containing functional groups. Additionally, tRNAs are the most heavily modified RNA class, with an average of 8 modifications per tRNA in prokaryotes and 13 in eukaryotes (mitochondrial tRNAs possess fewer modifications than cytosolic tRNAs) [4,19,20]. tRNA modifications can generally be classified into two groups: anticodon loop modifications and body modifications. Anticodon loop modifications occur in the 7 nucleotides of the anticodon loop, primarily at positions 32, 34 (the wobble base), 37, and 38. Modifications at positions 34 and 37 are the most chemically diverse [19,20]. Body modifications occur outside of the anticodon loop and are far less diverse in their chemical composition, being primarily Ψs, 2' OH methylations (Nm, where N is adenosine (A), cytosine (C), guanosine (G), or uracil (U)), Ds, Watson-Crick face methylations such as N^1-methyladenosine (m^1A), N^1-methylguanosine (m^1G), N^3-methylcytosine (m^3C), and $N^2{}_2$-dimethylguanosine ($m^2{}_2G$), and Hoogsteen-face modifications such as 5-methyl cytosine (m^5C) [19,20]. Indeed, tRNA modifications have been extensively documented in terms of their positions and chemical composition, though new modifications are being discovered as we improve technology, computational analyses, and study previously unexplored organisms.

Functionally, tRNA modifications have diverse roles, which depend as much on their location as the chemical nature of the modification [21-23]. Processes affected by tRNA modifications include translational efficiency and accuracy [24,25], cleavage by RNases [26], tRNA folding [22,23], and recognition by RNA binding proteins [27], among others. Modifications in the anticodon loop have been to affect the efficiency and accuracy of decoding and tRNA charging; however, this list is far from comprehensive as many modifications do not

have identified enzymes for their synthesis ('writers') or removal ('erasers') or proteins that specifically interact with the modified RNAs ('readers') [27]. Body modifications are known to participate in regulating many of the same processes as anticodon loop modifications (tRNA folding, stability, translation efficiency, and cleavage). Although modifications in the anticodon loop have much greater chemical diversity than body modifications, there is some overlap in their repertoire (e.g., m^1G, Ψ, and m^3C). It is clear from the literature that the effect of the modification is dependent upon its position; however, generalizations have been difficult to draw. Indeed, the same modification in one part of the tRNA may produce an entirely different effect as that in another region. Interestingly, while some modifications are present in the same position on many tRNAs (e.g., $m^1A57/8$, $m^1A/G9$, and $m^2{}_2G26$), others are present only in a few (m^1A14 in human $tRNA^{Phe}$). In addition to their complexity in composition and placement, many tRNA modifications are dynamic [27]. Indeed, these modifications are known to be altered in response to changing environments and cellular status and these variations create an incredibly diverse array of tRNA microspecies [5, 28, 29].

Lastly, tRNAs are charged, meaning to have an amino acid covalently linked to its 3' end. tRNA synthetases are responsible for charging tRNAs. Charged tRNAs participate in translation, as they are responsible for decoding mRNAs. tRNA charging is a dynamic process and is regulated by a variety of factors, including RNA modifications, tRNA abundance, cellular status, and amino acid availability [4]; limiting quantities of amino acids have a substantial impact on isoacceptor charging – in some cases causing the charged tRNA fraction to drop to near zero levels [30]. Indeed, selective deacylation under stress conditions is a useful mechanism for cells to control translation of essential proteins for stress response [31]. This was first demonstrated with amino acid starvation, during which *E. coli* were deprived of various amino acids, whereupon the cognate

isoacceptors displayed differences of up to 70% between charging levels of isoacceptors [30]. tRNA charging by tRNA synthetases is sine qua non for participation in the translational process; as a result, differences in tRNA optimality for aminoacylation can enable tRNAs to be utilized in processes outside of translation [7]. Thus, tRNA charging is a regulatory mechanism for potential tRNA function in cellular processes and regulate processes in addition to translation.

1.1.1.1.Translation

tRNAs are most commonly associated with translation, where they play a vital role as both a resource and a regulator of translational fidelity. The ability of tRNAs to regulate translation arises from their interactions with the mRNA, the ribosome, and neighboring tRNAs in the ribosome [32-34]. These interactions occur during initial selection and proofreading of the incoming tRNA, accommodation of the tRNA in the A site, translocation, and extension of the peptide chain. The tRNA anticodon stem-loop and elbow regions display extensive interactions with the small and large subunits, respectively [33]. During elongation, incoming tRNAs regulate translation via a tripartite interaction between the incoming tRNA, mRNA, and rRNA [35]. This proofreading step employs the kinetics of Watson-Crick pairing to prevent mistranslation, as suboptimal parings between the tRNA and mRNA are more likely to dissociate before the tRNA can be moved into the A site of the ribosome. Further, cognate anticodon-codon interactions lead to conformational changes in the rRNA of the small subunit of the ribosome so that it interacts with the anticodon-codon duplex [35]. The structure of the tRNA is also critical for the decoding step, as its stability prevents non- and near-cognate anticodon-codon pairs from distorting the anticodon into the A/T configuration, which is required for the GTPase activity necessary for the tRNA to enter the ribosome [36].

Since there fewer isoacceptors than codons in all studied organisms, tRNA modifications in the anticodon loop at positions 32, 34 (the wobble base), 37, and 38 are important for codon decoding, and thus the energetics required for the distortion of the tRNA into the A/T conformation [37-39]. As stated previously, the anticodon loop possesses the greatest chemical diversity in terms of RNA modifications, many of which interact directly with codons to efficiently discriminate between their cognate pairings and potential near- and non-cognate interactions [39]. These chemically complex modifications are frequently found to co-occur with smaller modifications such as Watson-Crick face methylations at positions 32 and 38, which have been hypothesized to act in conjunction to enable efficient and accurate codon decoding [40].

In addition to decoding, tRNAs regulate translation through interactions with the A, P, and E sites in the ribosome. The interactions between the tRNA anticodon stem-loop and the small subunit in the A site are particularly important in regulating translocation, as they work to ensure proper alignment of the A site tRNA as it moves to the P site during peptidyl transfer [33,41]. In a manner similar to the decoding process, tRNA modifications are hypothesized to stabilize anticodon-codon pairs to maintain proper alignment during translocation [42]. Indeed, tRNA hypomodification has been proposed as a means by which translation is regulated, as hypomodified $tRNA^{Pro}_{XXX}$ is associated with +1 frameshifting, which can be utilized as a mechanism for encoding multiple proteins in a single mRNA or to limit translation of proteins during stress [43]. The interactions between the tRNA elbow regions and the large ribosomal subunit have also been shown to affect gene expression. Previous work from the Pan lab found a positive correlation with tRNA m^1A levels at positions 22 and 58/59 in mouse gut Clostridia and the rate of translation of proteins enriched in amino acids corresponding to m^1A-modified tRNAs [44].

While charged tRNAs function as carriers of amino acids to continue translation elongation, uncharged tRNAs inhibit translation. Uncharged tRNA interferes with elongation through the activation of protein kinase GCN2, which a tRNA-binding domain recognizes, leading to phosphorylation of the elongation factor eIF2α and consequently inhibition of reassembly of the initiator complex [45]. Although eIF2α phosphorylation largely shuts down translation, several stress response genes require eIF2α phosphorylation, as the resulting leaky scanning mechanism enables translation of activating transcription factor 4 (ATF4) transcripts and subsequent transcription of additional stress response genes [46,47]. As a result, changes in tRNA charging fractions affect not only translation, but also transcription.

In short, tRNA sequence, structure, post-transcriptional modifications, and charging function synergistically to regulate translation at a multitude of steps. The role of post-transcriptional modifications is particularly emphasized here, as these modifications are a means by which the cell can regulate translation and shift it to favor synthesis of proteins that are preferred or needed in different conditions. That modifications play such a critical role in a central process underscores the need for sequencing methods that can detect the panoply of modifications that decorate tRNAs.

1.1.1.2.Transcription

As discussed in *1.1.1.2*, tRNAs play an important role in gene regulation at the level of transcription. While eukaryotes utilize GCN2 to sense uncharged tRNAs (described in the previous section), bacteria have been shown to employ both tRNA-protein and tRNA-RNA interactions to control global gene expression. The *E. coli* protein RelA is a ribosome-associated (p)ppGpp synthase senses the presence of uncharged tRNA in the A site of the ribosome. Uncharged tRNAs

in the A site result in ribosome stalling, which activates RelA synthesis of ppGpp [48,49]. ppGpp is an alarmone that binds the *E. coli* RNA polymerase to inhibit transcription globally while stimulating the expression of amino acid synthesis genes [48,49].

The T box riboswitches are classical examples of biological regulation by RNA-RNA interactions. T boxes are found in the leader sequences of amino acid biosynthesis genes and can bind specific tRNAs by specific interactions with the anticodon and the uncharged 3' CCA sequence [7,50,51]. T box leaders typically function as terminators until periods of amino acid starvation, when the corresponding tRNA is uncharged and can anneal with the complementary GGT sequence; charged tRNAs are sterically prevented from engaging in this interaction [50,51]. Upon binding the anticodon and the CCA end, the T box energetically favors an antiterminator secondary structure, allowing RNA polymerase to continue transcription. It should be noted that T box riboswitches can also regulate translation initiation through a similar mechanism, albeit one that operates by sequestering the Shine-Dalgarno sequence.

These examples demonstrate an additional regulatory function of tRNA, one controlled by cellular adjustments of the tRNA charging fractions. It is plausible that additional control systems and mechanisms are used to regulate transcription through the detection of tRNA abundance, post-transcriptional modifications, and/or charging. As such, developing new technologies that can report on these parameters will reveal new biological insights with potential biomedical applications.

1.1.1.3. Apoptosis

Apoptosis is a programmed cell death process that occurs via extrinsic and intrinsic pathways. Both pathways proceed via oligomerization of proteolytic caspases (caspase 8 for the extrinsic

pathway and caspase 9 for the intrinsic) [52]. Cytochrome c (Cyt c), a mitochondrial peroxidase and part of the ATP synthase complex, was discovered to play a regulatory function of tRNA in preventing apoptosis [53]. Cyt c binds tRNA, which inhibits the ability of cytochrome c to bind and trigger oligomerization of apoptotic protease activating factor-1 (Apaf-1). (Oligomerization of Apaf-1 leads to oligomerization of caspase 9, triggering apoptosis.) Cyt c binds both cytosolic and mitochondrial tRNA, thus providing a means for regulating Cyt c function once it is released into the cytosol, where it promotes apoptosis. Hydrogen/deuterium exchange mass spectrometry analyses showed that deuteration levels of Cyt c were reduced after incubation with $tRNA^{Phe}$ or total tRNA, indicating a significant conformational change toward a compact, globular structure [54]. Further analysis demonstrated that Cyt c binds tRNA with similar affinity to other tRNA-binding proteins when in its reduced form, but displays reduced affinity in its oxidized forms, which is pro-apoptotic. Interestingly, Cyt c binding affinity is not affected by tRNA charging.

Further illustrating the role of tRNA in apoptosis, transcription of tRNA and 5S rRNA are negatively regulated by intrinsic pro-apoptotic factors such as p53 and Rb, which bind directly to RNA polymerase III (Pol III), and positively regulated by the oncogenic proteins c-Myc and Ras, which activate Pol III [55]. Unsurprisingly, cancers have roughly 4-5-fold increased levels of cytosolic tRNAs and 5-10-fold increased mitochondrial tRNAs than healthy cells [56], as p53 is largely inactivated in 50% of cancers while c-Myc and Ras are upregulated in many cancers. This dysregulation of tRNA synthesis likely contributes to the reduction of intrinsic apoptosis through increased tRNA substrates for Cyt c. It remains to be seen whether tRNA modifications affect binding to Cyt c or if different isodecoders have increased binding affinity, which could provide additional levers of control over cell death.

1.1.1.4.Regulation of viruses

Viruses have been known to package host tRNAs and to utilize tRNA-like structures to regulate processes such as capsid formation and translation of viral mRNAs [57-60]. Additionally, viruses alter host tRNA pools to manipulate translation to favor synthesis of viral proteins [61]. HIV is a particularly notable example of viral hijacking of host tRNAs. As a retrotransposon, HIV must convert its RNA genome into cDNA to integrate into the host, a process which requires a primer. Packaged within the HIV virion is $tRNA^{Lus}_{UUU}$, which is remodeled by the nucleocapsid (NC) protein to expose the 3'-most 18 nucleotides [62,63]. This remodeled $tRNA^{Lys}_{UUU}$ then binds to the primer binding site (PBS) of the HIV genome, whereupon reverse transcription is initiated by the packaged HIV RT. Similar mechanisms are observed in other retroviruses, including Moloney murine leukemia virus (Mo-MLV) and Rous sarcoma virus, albeit with differences in the proteins, tRNAs, and structurally destabilized regions of the primer tRNAs [58].

The clear dependence of viruses on packaging host tRNAs to regulate their life cycles raises significant questions about how other tRNAs and small RNAs are utilized by infectious particles. My own work has delved into this topic and has found selective packaging of host small RNAs in SARS-CoV-2 virions, including $tRNA^{Lys}_{UUU}$, as well as evidence of preference for tRNAs with particular post-transcriptional modifications [2]. Further investigations will discover the functional role of these packaged tRNAs and other small RNAs in regulating the viral life cycle, potentially discovering therapeutic targets.

1.1.2. tRNA fragments

tRNA fragments (tRFs) are stable small RNAs derived from cleavage of mature tRNAs and precursor tRNAs by endoribonucleases and are roughly 100-fold less abundant than tRNA [64].

tRFs have been observed in all three kingdoms of life and they participate in a myriad of cellular processes as regulators [64,65]. As stated, cleavage of tRNAs produces tRFs; these cleavage events occur at defined positions and yield tRFs of specific lengths that likely maintain the secondary structure of the tRNA regions. Cleavage by different nucleases can produce tRFs that are from the terminal or internal regions of the tRNA; indeed, there are many known RNases implicated in the biogenesis of tRFs, the most well-studied of which are angiogenin, RNase P, RNase Z, and Dicer [64,65]. Because there are hundreds of human tRNA genes and many positions for cleavage, a large repertoire of tRFs can be generated for many regulatory purposes. (It should be noted that a variety of nomenclatures have been developed to categorize tRFs based on where they were cleaved. However, inconsistencies and disagreements in the use of terminology render these nomenclatures more confusing than clarifying. Hence, I use here the broader term tRNA fragments to refer to any small RNA produced from cleaving tRNAs.) tRNA cleavage is affect by a large number of factors, including regulation of nuclease activity, tRNA modifications, and tRNA-protein interactions, all of which can be affected by changes in cellular status [5,26,65,66].

Because of their incredible regulatory potential, tRFs have been a topic of significant interest for both fundamental science and medically relevant studies. Despite this interest, tRFs have been difficult to study, owing to their low abundance and post-transcriptional modifications. Antisense oligonucleotides have offered a means to directly perturb tRF regulation, but discovery and full characterization of tRFs in their biologically relevant states and processes remains largely elusive. One can correctly surmise that many of the same complicating factors in sequencing tRNAs apply to tRFs. Thus, developing sequencing methods for tRNAs and their post-transcriptional modifications could allow for a fuller analysis of tRFs and their biological functions, which I detail below.

1.1.2.1.Regulation of gene expression

Gene expression can be regulated by tRFs via RNA degradation, repression of translation, and epigenetic reprogramming. tRFs mediate RNA degradation with a mechanism similar to miRNA-directed target degradation [26,65,66]. Indeed, tRFs have been shown to bind to all four Argonaute proteins (Ago 1-4) and Ago 1-4-bound tRFs are 14-32 nt in length, similar to processed miRNAs [26,65-68]. A further overlap between tRFs and miRNAs is that tRFs are also processed by Dicer, though not exclusively [68,69]. Although tRFs associate with all four Ago proteins, less-structured, particularly single-stranded, tRFs bind preferentially to Ago 3 and 4. Additional work discovered that addition of oligos antisense to 3' tRFs shifted the binding affinity to Ago 2 and resulted in robust degradation of target mRNAs (> 80%) [70]; this is a point of departure from the miRNA-directed target degradation, which is repressed when complexed with an antisense oligo. Ago-bound tRFs target both long non-coding RNAs and mRNAs, the latter of which have tRF binding sites in the coding sequence and the 5' and 3' untranslated region [71,72]. tRFs also function in repression of retroviral reverse transcription, as observed with an Ago 2-bound tRNALys(UUU)-derived tRF binding to and cleaving the HIV promoter-binding site [73]. Furthermore, tRF binding sites are also found in introns, indicating a role for tRFs in determining the levels of protein isoforms expressed by the cell [74]. tRFs may also target multiple sites in the same target transcript [72]; this has been shown for Ago-independent mechanisms [75]. Recent work has shown the importance of m^1A modifications in regulating pairing of the seed region of tRFs to their target during the unfolded protein response [76]. Reduction of the m^1A fraction in the studied tRF did not significantly alter the levels of tRFs, but did reduce the unfolded protein response, indicating that epitranscriptomic modifications offer a potent means to control the cleavage of tRF-directed targets.

In addition to Ago-dependent RNA regulation, recent work has shown tRFs to regulate gene expression via Ago-independent means [77,78]. Ago-independent tRFs promote translation through binding of a complementary mRNA target [78]. Promotion has been observed to occur by unfolding of the duplexed regions that would otherwise stall the ribosome; inhibition has also been observed, which functions through displacing proteins protecting the mRNA from degradation [78,79].

Translational regulation by tRFs occurs through multiple mechanisms, including inhibition of the ribosome and displacement of initiation factors, which inhibits translation initiation, followed formation of stress granules [80,81]. These mechanisms, naturally, are controlled by different types of tRFs containing or lacking different motifs [80-82]. Unlike the tRFs associated with Ago proteins, translationally repressive tRFs do not act direct pairing to an mRNA target. It has been demonstrated that a 26-mer tRF responsible for ribosome inhibition possess a conserved 3' terminal GG dinucleotide motif [83]. Mutation of the GG motif abolished the inhibitory activity of the tRF.

Angiogenin-produced 5' tRFs (~ 30-40 nt) are more abundant during stress and inhibit translation by displacing initiation factors eIF4A, F, and G from capped and uncapped mRNA [83,84]. As a part of this process, some tRFs bind the Y-box-binding protein 1 (YB-1, also written as YBX1) to form stress granules, which further inhibits translation [83-85]. However, YB-1 is dispensable for displacement of eIF4F from capped mRNA [85]. This processes stalls translation initiation and promotes formation of stress granules. Similar to the ribosome-binding tRFs, the YB-1-binding tRFs possess a 5' terminal TOG motif, as well as a string of at least four guanosine nucleotides at the 5' end, which are predicted to form a G quadruplex structure [85,86]. Despite the role of YB-1 in translational inhibition via tRFs, knockout of YB-1 only partially reverses the

inhibitory effect of 5’ tRFs, indicating further mechanisms in mediating tRF repression of translation [85]. Recent work has also demonstrated the importance of pseudouridylation TOG-motif tRFs in their stability, showing that post-transcriptional modifications are an additional lever for regulation in the mechanisms [87].

Epigenetic reprogramming by tRFs has recently gained more attention, but precise mechanisms by which it occurs have been challenging to discern [88]. It is plausible that tRFs achieve epigenetic reprogramming by a variety of mechanisms, including pathways similar to piRNA-mediated silencing (tRFs have been shown to interact with PIWI proteins), target-directed DNA and histone methylation, and heterochromatin remodeling [88,89]. For example, a piRNA-like tRF bound PIWIL4, which then recruited histone 3 lysine 9 (H3K9) methyltransferases SETDB1 and SUV39H1 and heterochromatin protein 1β (HP1β) in a sequence-directed manner [90]. Epitranscriptomic modifications are thought to also play a significant role in regulating this form of epigenetic reprogramming, as the relevant tRFs possess dynamic modifications, which can respond to changing environmental conditions [91]. Several of the most recent studies have demonstrated the importance of tRFs on intergenerational inheritance by examining the effects of diet of mouse sperm [92,93]. Injection of tRFs from high-fat diet-fed mice were injected into zygotes, which then generated metabolic disorders in offspring and altered gene expression in the embryos. However, there are still many unknown factors, including the tRF-interacting proteins that effect epigenetic reprogramming or if some tRFs achieve epigenetic reprogramming without the need for protein partners.

1.1.2.2.Cell-cell communication

Intercellular communication by small RNAs has been investigated since the early 2000s [94,95]. These small RNAs have been found in extracellular vesicles, ribonucleoprotein complexes, and other fractions [94]. YB-1, previously described in this Chapter as a tRF-binding protein, were found to be necessary for sorting small RNAs into exosomes [96]. Interestingly, YB-1 is an intrinsically unstructured protein and predicted to bind 5-methylcytosine (m^5C) [97], consistent with a recent study that detailed miRNA sequences codes that promote or hinder their inclusion into small extracellular vesicles by the intrinsically unstructured proteins Alyref, a putative m^5C reader protein [98], HNRNPG, and Fus, among others [99].YB-1-associated tRFs were found to be enriched in T-cell-derived EVs, indicating a similar mechanistic pathway for tRF sorting into EVs [100].

Extracellular tRNAs have been reported on for several decades, but only in 2012 were tRFs first detected outside of the cell and in 2013 they were observed in biofluids [94]. Since then, tRFs have been determined to exist in ribonucleoprotein (RNP) complexes and extracellular vesicles (EVs). (Some controversial studies have shown tRFs in plasma lipoprotein fractions, but that topic is beyond the scope of this work and can be reviewed elsewhere.) Additionally, tRFs can be found in the non-EV/RNP fractions. A large body of work characterizing extracellular small RNAs has demonstrated that the extracellular transcriptome is dependent on which fraction is studied, as well as the cell type, biofluid, tissue, and organism. T cells have been shown to release specific tRFs in EVs in a signal-related manner [100]. These T-cell-derived EVs were enriched in various tRFs, with the specific tRFs being dependent upon the state of the T cell (resting or activated). Silencing of the EV-enriched tRFs promoted T cell activation, indicating a potential role for tRFs delivered by EVs in suppressing activation of neighboring T cells.

Several examples in the past decade have highlighted intercellularly trafficked tRFs as biological regulators, some of which I have described in the previous section. Companion papers in 2016 discovered that mature sperm transfer tRFs via epididymosomes to immature sperm as they pass through the epididymal tract [92,93]. Further, dietary changes in mice (high-fat versus low-fat diet) altered the expression profiles of tRFs in mature sperm, which were matched by the profiles in the epididymosomes. Injection of the tRFs from high-fat diet-fed mice into normal zygotes resulted in metabolic disorders in the offspring and altered expression of metabolic genes. As stated in the previous section, this was accomplished without changes to DNA methylation. Subsequent work demonstrated that Caput epididymosomes can deliver small RNAs to testicular spermatozoa *in vitro* and *in vivo* [101]. Together, these studies demonstrated a pathway by which paternal dietary-induced patters can be inherited by transfer of tRFs. These works represent a groundbreaking development in the field of intercellular communication by tRfs and highlight the need for future development in this niche.

1.1.3. Signal recognition particle

The sorting of nascent secretory and membrane proteins is a challenge common to all three kingdoms of life and a similar solution has been evolutionarily conserved in all of them: the signal recognition particle (SRP) [11]. Functionally, the SRP is ribonucleoprotein complex that docks the nascent peptide and ribosome to a translocon complex on the inner membrane (bacteria) or the endoplasmic reticulum (eukaryotes) [11]. Although the general SRP mechanism of translocation has been conserved in the three kingdoms, the precise details vary across the phylogenetic tree [11,102,103] and I will focus on the RNA component of the SRP complex here. Eubacteria encode a 4.5S SRP RNA that is ~ 115 nucleotides in length and homologous to helix 8 of the ~ 300

nucleotide human 7S/SRP RNA [11]. Interestingly, some prokaryotes possess a 6 or 7S SRP RNA [104]. Larger SRPs can generally be divided into two domains: 1) the S domain, which interacts with the nascent peptide and 2) the Alu domain, which arrests translation [104]; smaller SRPs, by contrast, typically lack the Alu domain and thus must co-translationally direct the bound ribosome to the translocon [104]. While the Alu domain of higher eukaryotes interacts with the Alu-binding proteins SRP9 and SRP14, no homologs of these proteins have been found in bacteria or archaea, despite the similarities in their SRP RNAs [105]. *Bacillus subtilis*, for instance, utilizes RNA-RNA interactions between the large ribosomal subunit and *Bs*SRP, which may be a more widespread mechanism in certain bacterial taxa [105]. However, all SRPs contain a site responsible for binding their respective SRP54 homolog protein, which binds the signal peptide for membrane insertion [103]. (In bacteria, this RNA region is known as domain IV.)

Mammalian SRP RNAs are known to participate in many additional, noncanonical functions, including intercellular communication, viral packaging, proliferation, and red blood cell metabolism [57,106]. These functions are mediated through a wide array of interacting partners, both protein and RNA. 4.5S RNAs have also been reported to interaction with elongation factor G (EF-G), although the functional consequence of this interaction remains unclear [107,108]. Sequence analyses of SRP RNAs have revealed only a few conserved bases in secondary structures that are difficult to predict [109]. Thus, finding new sequences that are not closely related remains a challenge. Developments in small RNA sequencing complemented with traditional biochemical assays would significantly expand opportunities in discovering new SRP sequences as well as understanding the biology of uncultivated organisms and microbial communities.

1.1.4. 5S rRNA

The 5S rRNA is an essential component of the ribosome and is found in all kingdoms of life with general conservation of its secondary and tertiary structure [110,112]. Indeed, the fitness costs on various organisms when the 5S was deleted were quite severe [111]. 5S rRNA is approximately ~ 120 nt in length and binds the large subunit near the ribosome active site [110]. It is composed of 5 helices and 5 loops [110,112]. Intriguingly, although the large and small subunit rRNAs are post-transcriptionally modified to a significant degree, the 5S possesses few modifications [113]. Despite its high degree of structural conservation, the exact function of the 5S has not been determined; however, significant advances in understanding its role in translation and ribosome-independent pathways have been made [114]. Structural and functional studies have shown that 5S is critical for tRNA binding in the A site [115]. Further, the P site may communicate with EF-G in part through interactions with the 5S [115,116]; considering the position of the 5S on the yeast ribosome, where it physically connects the functional regions, the 5S has been predicted to transduce signals between the separate regions [114,117].

For the pre-ribosome, the 5S is required for processing of the large subunit [114,118]. Although the 5S is essential for a functional ribosome, it is not clear at what stage the 5S is integrated with the rest of the ribosome [119]. Most mitochondria, interestingly, do not encode an identifiable 5S rRNA gene; in its stead, they utilized a mitochondrial Phe or Val tRNA, highlighting an additional noncanonical role of tRNAs in biological regulation [120,121].

The 5S is also known to participate in ribosome-independent pathways, the most studied of which is the regulation of p53 [121,122]. 5S ribonucleoprotein (RNP) complex associated with the 60S ribosome. However, upon sensing impaired ribosome biogenesis, 5S RNP is released from

the 60S and binds to the human and mouse Double Minute 2 (H/MDM2) protein, inhibiting its activity [121,122]; H/MDM2 repression then frees p53.

Where the SRP RNA is highly divergent, 5S rRNA is highly conserved across the kingdoms of life. As a result, these small RNAs provide an interesting opportunity to perform phylogenetic reconstructions of translational regulation, particularly in microbes that are difficult to cultivate or that exist in microbiome. Given the fact that in bacteria they both associate with EF-G, sequencing them in context with other small RNAs could provide significant insights into the function of the 5S during translation of membrane proteins in response to an ever-changing environment. Pairing small RNA sequencing of 5S, SRP, and other small RNAs with ribosome profiling may further provide insight into the conserved and divergent functional roles of 5S.

1.2. RNA Modifications

RNA modifications have been observed since the mid-20th century [27,123]. Although objects of some curiosity, limited means of studying their biological function and relevance prevented significant progress until the past decade. The study of RNA modifications, known as epitranscriptomics, has seen a veritable boom since 2015 and the roles of these modifications are seemingly innumerable, affecting nearly every aspect of biology [20]. Currently, there are approximately 150 known RNA modifications in all domains of life, with roughly 80% being found on tRNAs [18]. As stated in a previous section, tRNAs are highly co- and post-transcriptionally modified, with modifications occurring between on 10-20% of tRNA nucleotides [18,20]. An important concept to mention here is the idea of modification fraction. While some modifications are uniformly introduced to certain isodecoders or isoacceptors, some modifications

are installed only on a percentage of tRNA copies. Indeed, some modifications are dynamic, allowing their modification fraction(s) to be adjusted as needed by the cell [27].

In this section, I will discuss some of the modifications that are relevant to my work in terms of their chemical properties, biological function, and regulation.

1.2.1. Watson-Crick face methylations (m^1A, m^1G, m^3C, $m^2{}_2G$): Biology & functions

Methylations on the Watson-Crick face are highly abundant post-transcriptional modifications in biological systems, occurring in many RNA classes [124,125]. Several of these modifications are dynamic in nature, meaning that they can be added and removed in response to cellular needs. tRNAs are particularly enriched in methylation modifications, particularly in the loop regions where they participate in stabilizing tRNA structure and mRNA decoding, among other processes [124,126]. Methylations are typically installed by writer enzymes that are dependent on a metabolite known as *S*-adenosylmethionine (SAM), which supplies the methyl group [127,128]. Erasers, on the other hand, have a conserved non-heme iron dioxygenase active site that utilizes radical chemistry to oxidatively remove methylations [128]. Some methylations produce a positive charge on the nucleobase, altering their binding properties to proteins and other RNAs [27]. Many of these methyl modifications affect the ability of nucleotides to participate in Watson-Crick base pairing [27]. As to their biological function, that is highly dependent on the position of the nucleotide within the RNA [23,129-131]; for example, a methylated nucleotide in the anticodon loop may be necessary for decoding mRNA or the installation of a separate anticodon-loop modification, while the same methylated nucleotide in the T- or D-loop might be responsible for stabilizing tRNA structure or promoting translational efficiency through interactions with the ribosome [23,129-131]. I shall provide more specific details in the subsections below.

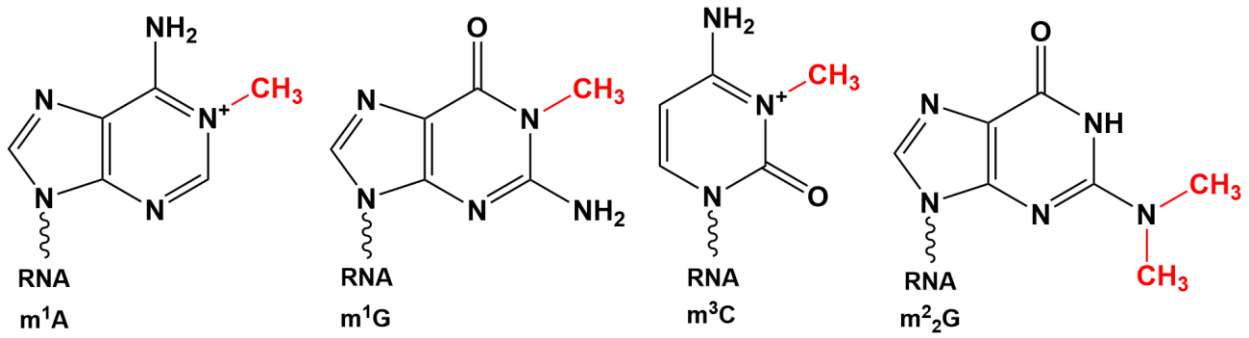

Figure 1.2: Chemical structures of the Watson-Crick face methylations m^1A, m^1G, m^3C, and $m^2{}_2G$.

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

m^1A is a ubiquitous modification in eukaryotes, found in most cytosolic and many mitochondrial tRNAs [20,27]. (m^1A has also been found in bacteria [132], but few studies have been conducted and no enzymes responsible for either writing or erasing have been reported in the literature.) m^1A, along with several other Watson-Crick face methylations, induced a positive charge on the nucleobase and significantly blocking base-pairing (Figure 1.2); both properties dramatically alter the chemical properties of the nucleotide and therefore its ability to bind proteins and RNAs. The dynamic nature of m^1A is another fascinating feature of this modification, as cells can, with relative ease, adjust the levels of m^1A in response to changes in the cellular status [24,27].

In tRNAs, m^1A can be found at positions 9, 14, 22, 57, and 58, although these are dependent on the type of organism studied [132]. Structure is critical to the synthesis of m^1A-modified substrates, as it is the key feature by which the writer enzymes identify the A nucleotide to be modified [132]. Several m^1A writers and erasers have been confirmed in mammalian cells, both in the nucleus and mitochondria, and the writers are shown to specifically methylate a single region (i.e., m^1A9 writers do not methylate at position 58) [132, 133]. Two of these writers operate as dimers, in a manner similar to the m^6A writers Mettl3 and Mettl14 [27], where only one of the

proteins is responsible for catalytic activity [133]. There are currently three known erasers for m^1A in tRNA: ALKBH1 ALKBH3, and FTO [24a,24b,134]

m^1A is reasoned to mostly participate in stabilizing the tRNA structure [23,129,135]. The most striking example of m^1A supporting tRNA stability comes in the form of m^1A9 in mt-$tRNA^{Lys}$ was necessary for the proper folding of the tRNA [129]. (Cytosolic tRNAs possess more modifications on average than mitochondrial tRNA and a single methylation likely has less impact as a result.) m^1A has been the most functionally characterized at positions 9 and 58 in tRNAs, where they are mostly observed to function as stabilizers of tRNA tertiary structure and/or promoters of proper folding [23,129,135]. m^1A58 has been found to have additional functions, such as HIV-1 reverse transcription by $tRNA^{Lys}_{UUU}$ priming [136]. Yeast $tRNA^{iMet}$ m^1A is essential in preventing transcript degradation, allowing cells a quick mean by which to control translation initiation [135]. Further, translational efficiency was found to be positively correlated with m^1A levels, which were decreased during glucose deprivation, indicating a role for m^1A in the regulation of cellular metabolism [24]. While these examples highlight phenomena of m^1A, they often lack a mechanistic connection that allows broader conclusions to be drawn. Compounding this issue is the fact that m^1A likely acts in concert with other modifications, substantially raising difficulty in dissecting the individual contributions of a single modification.

1.2.1.2.N^1*-methylguanosine (*m^1G*)*

Similar to m^1A, m^1G is a methyl modification of the endocyclic nitrogen at the 1 position of guanosine, which interferes with Watson-Crick base pairing (Figure 1.2). m^1G also occurs at position 9, where it is known to stabilize tRNA structure [131]. (m^1G9 is more common in cytosolic tRNAs, while m^1A9 is most frequently found in mitochondrial tRNAs.) In many other

respects, m^1G differs from m^1A. The anticodon loop position G37 has a high rate of methylation [137], while there has been no observation of m^1A in the anticodon loop. m^1G37 has been shown to prevent +1 frameshifting during translation, likely through weaking of tRNA-mRNA interactions [130,138]. In a fascinating contrast, m^1G37 in *S. cerevisiae* $tRNA^{Phe}$ promotes -1 programmed ribosomal frameshifting, although this may be highly context-dependent, as m^1G37 is a precursor to the wyosine modification [139]. m^1G has also been indicated to play an important role in $tRNA^{Pro}$ aminoacylation, as deletion of TrmD in *E. coli* results in a growth defect due to reduced $tRNA^{Pro}$ charging [140]; this is also observed for other tRNAs [141-143]. Similar to m^1A, writers for m^1G display structural selectivity (Trm5 methylates G9 and TrmD methylates position 37) [144]. The biological role of m^1G is thus far from exhaustively examined.

1.2.1.3.N^3*-methylcytosine (*m^3C*)*

Another dynamic methyl modification of a cyclic nitrogen, m^3C occurs at only three known positions: 1) the anticodon loop at position 32 in cytosolic and mitochondrial tRNAs; 2) position 47d (also known as position e2) in the variable loop of Ser and Leu(CAG) tRNAs; and 3) position 20 of human elongator $tRNA^{Met}$ [145]. m^3C is only known to occur in eukaryotic tRNAs; of the studied eukaryotes, m^3C is only found on two tRNA isoacceptors families in two types of yeast ($tRNA^{Ser}$ and $tRNA^{Thr}$ in *S. cerevisiae* and *S. pombe*), while human cells have eight tRNA isoacceptors families (6 cytosolic and 2 mitochondrial) with m^3C modifications [145]. In keeping with expanded repertoire of substrates, humans have at least five m^3C methyltransferases compared to the one and two for *S. cerevisiae* and *S. pombe*, respectively [145,146]. Interestingly, knockout of the individual m^3C methyltransferases in human cells did not completely eliminate the m^3C level of their tRNA substrates in all cases, indicating either a significant level of

redundancy and/or additional methyltransferases [146]. In addition, there appear to be position 37 modifications that are pre- or co-requisite for methylation by the m^3C methyltransferases [124,147,148]. It should be stated that the methyltransferase for m^3C47d and $tRNA^{Met}$ m^3C20 have yet to be identified. Currently, the only known eraser for m^3C in tRNA is ALKBH3 [134].

The effects of m^3C outside of position 32 on tRNA function are not clear, although it has been implicated in structural maintenance. *S. cerevisiae* knockouts of Trm140 do not display a detectable phenotypic difference from wild-type yeast under normal growth conditions [147]; however, m^3C may play a much more critical role during stress. m^3C32 has been found to affect translation and ribosome occupation [147,149, 150]. Furthermore, because only writers for m^3C32 in tRNAs have been identified and perturbed, we there have been no examinations to this date of the functional roles of m^3C modifications outside of the anticodon loop – apart from my work discussed in Chapter 3.

1.2.1.4.N^2_2*-dimethylguanosine (*m^2_2G*)*

m^2_2G is a dimethyl modification of the exocyclic nitrogen at position 2 of guanosine that occurs exclusively at nucleotide 26 (between the D-step and the anticodon stem loop), whether in eukaryotes or prokaryotes [151]. 80% of eukaryotic tRNAs have an m^2_2G26, although many tRNAs with a G26 are unmodified [151]; G26 sites that are known to be demethylated have high levels of methylation [152]. The C11-G24 and G10-C25 base pairs are an additional identity element necessary for the demethylation of G26 by Trm1 [153]. m^2_2G26 is another modification that contributes to stabilizing tRNA structure. In a similar fashion to m^1A9 in mt-$tRNA^{Lys}$, human $tRNA^{Asn}$ requires dimethtylation of G26 to prevent formation of a G26-C11 base pair that disrupts the tRNA structure [154,155]. (m^2_2G26 instead pairs with A44 [156].) Apart from this, however,

$m^2{}_2G$ largely remains a mystery in terms of biochemical function, although it is implicated in several translation and cellular redox homeostasis [19].

1.2.2. Anticodon loop modifications (I, Q, τm^{5}s^{2}U, mcm^{5}s^{2}U, cmnm^{5}s^{2}U, s^{2}C): Biology & functions

As stated in previous sections, anticodon loop modifications possess diverse and often elaborate chemical structures, often requiring multiple enzymatic steps for their synthesis [19,20,157]. I distinguish anticodon loop modifications from modifications such as m^1G and m^3C, which are found in the anticodon loop and other regions in the tRNA body, by the fact that anticodon loop modifications are only found in the anticodon loop. These modifications are often necessary for accurate and efficient decoding of mRNAs, and in fact have the most potent effects in decoding mRNA. Indeed, loss of many of these complex modifications are implicated in several diseases [20].

There are a significant number of tRNA U34 modifications with chemically diverse chains at carbon 5 and a thiolation of carbon 2 [19,20,157]. Structural studies have demonstrated the larger carbon 5 modifications stabilize the U34 anticodon-codon interactions with A, G, and U, while the 2-thio modifications restrict the possible pairings to purines at the third position of the codon [158,159].

Anticodon loop modifications have proven difficult to study, owing to intractable synthetic and purification methods, impractical means of genetic perturbation, *in situ* phenotypes that may only be observable during certain stress conditions, and onerous means of detection. In this section, I will discuss a few of the anticodon loop modifications that are relevant to my work.

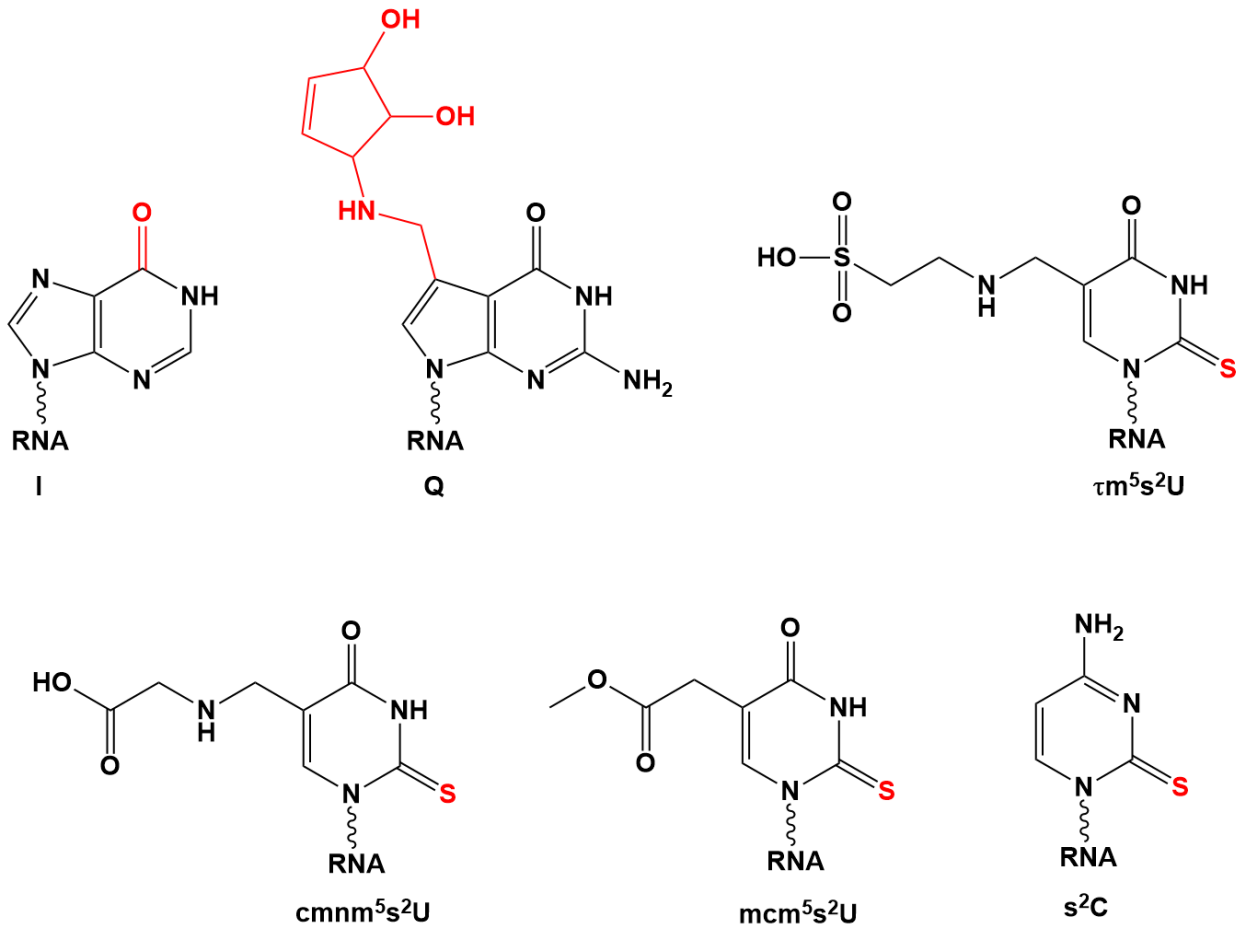

Figure 1.3: Chemical structures of the anticodon loop modifications I, Q, τm^5s^2U, $cmnm^5s^2U$, mcm^5s^2U, and s^2C.

1.2.2.1.Inosine (I)

Inosine is a deaminated form of adenosine and a highly abundant and essential post-transcriptional modification in eukaryotic tRNAs (7-8 isoacceptor families, depending on the taxa), although observed in only two tRNAs at the wobble base in bacteria ($tRNA^{Arg}_{ACG}$ and $tRNA^{Leu}_{AAG}$) [19,160,161]. In tRNAs, inosine is observed at the wobble base (position 34), position 37 in $tRNA^{Ala}$ of some eukaryotic taxa, and position 57 in archaea, though the latter two are also methylated to N^1-methylinosine [162]. Adenosines at position 34 in tRNAs are nearly uniformly

deaminated to inosine, although this can be altered by controlling the levels of the deaminases since inosine is produced in precursor tRNAs [163]. (Some notable exceptions to this rule are Pro, Ser, and Thr A34 tRNAs in Firmicutes, which have no detectable levels of inosine. The presence of negative identity elements may likely explain this phenomenon [164].) Bacteria possess a single adenosine deaminase that functions as a homodimer, while eukaryotes utilize a heterodimeric deaminase complex. These proteins are known as 'adenosine deaminases acting on tRNAs,' or ADATs [162].

Inosine functions to expand the decoding capacity of the modified tRNAs, as it can pair with A-, C-, and U-ending codons [160,161]. (Inosine is structurally similar to guanosine, but lacks the exocyclic amine at position 2, providing more freedom in base pairing at the cost of the strength of the pairs [165-167].) The I-C and I-A base pairs maintain the same Watson-Crick pairing geometry, while the I-U base pair is shifted slightly to allow the O2 and N3 of uracil to pair with the N1 and O6 of inosine, respectively [160]. (N3 and O4 of uracil pairs with N1 and N6 of adenosine.) Evolutionary studies of I34 tRNAs has identified a positive correlation between the number of ADAT amino acids (Arg, Leu, Thr, Ser, Pro, Ile, Ala, and Val) and ADAT-sensitive codons and also show that genes enriched in ADAT amino acids are overrepresented in the human genome, indicating an increased dependence on I34 tRNAs as the number of inosine-modified tRNAs increase [161,164,168]. Further, a recent study of low-complexity protein domains in humans revealed an overrepresentation of ADAT-sensitive codons and that depletion of I34 tRNAs or synonymous codon substitution resulted in reduced translational efficiency [168]. Many of the genes identified in this study were transmembrane proteins involved in cell adhesion, a process associated with higher eukaryotes. This fits well with the picture of expanded coding

capacity being necessary for the evolution of higher organisms, as a subset of the translational machinery can be dedicated to genes that are necessary for multicellularity.

The discovery of inosine in uncharacterized systems is of particular interest to fields such as the human microbiome, where metabolic and translational insights are difficult to obtain. Much of this difficulty is due to the complexity of the microbiome, which can be reviewed elsewhere [169,170]. Innovations in small RNA sequencing will provide substantial opportunities for new insights here, as inosine is an easily identifiable post-transcriptional modification and small RNAs are generally quite stable and provide a wealth of biological information.

1.2.2.2.Queuosine (Q)

Queuosine is of particular interest among the large collection of epitranscriptomic modifications in mammals since it can only be derived from the diet and subsequent derivation from gut microbes [19,171]. The Q nucleobase is highly similar to unmodified G on the Watson-Crick face, with the modification being attached to a carbon at the 7 position (G has a nitrogen at position 7). Unlike most epitranscriptomic modifications, Q is not produced by modifying a preexisting G but by switching out G nucleobase with Q using a dimeric tRNA guanine transglycosylase (TGT) enzymes (e.g., human QTRT1/2) [171,172]. The anticodon motif $G_{34}U_{35}N_{36}$ is recognized by the dimeric TGT complex, which swaps the G34 nucleobase for Q. There is also an observed reliance on Q modification for DNMT2-dependent m^5C38 modification, likely to alter the anticodon loop geometry to favor translation in certain contexts [40,173].

Only four eukaryotic tRNAs have queuosine: Arg, Asn, His, and Tyr, both cytosolic and mitochondrial [19,171]. Queuosine is known to affect translational efficiency and accuracy through preventing second-position codon misreading of the glycine codon GGC by $tRNA^{Asp}$

(G/Q_{34}UN) [174]. In *S. pombe*, Q modification of $tRNA^{His}$ and $tRNA^{Asp}$ increase the speed of translation for cognate C-ending codons, while Q-modified $tRNA^{Asn}$ and $tRNA^{Tyr}$ decrease speed for cognate U-ending codons [174]. In human cells, queuine depletion reduces translational fidelity and increases the level of unfolded proteins, triggering ER stress and the unfolded protein response, as well as impairing mitochondrial function and ATP synthesis [175,176]. Queuine is also linked to protection of HeLa cells against hypoxia stress [177]. Interestingly, Q modification displayed a protective effect in tRNAs, preventing cleavage by angiogenin [178]. These findings, though insightful, underscore lack of mechanistic understanding for Q-modified tRNAs and their context-dependent effects.

Similar to the evolutionary codon usage of inosine, codon usage of Q-modified tRNAs in the *Drosophila/Sophophora* genus are found to correlate with genome-wide changes in the use of Asp, Asn, His, and Tyr, with depletion of the queuine precursor resulting in reduced translational fidelity of tRNAs that normally contain Q [179,180]; thus, eukaryotic cells possess a pathway for directly linking nutrient levels to translational accuracy. Given this context, it is unsurprising that queuine depletion and QTRT1/2 mutations are associated several diseases, including cancer and neurodegeneration [19,20,181,182]. There are many plausible pathways and mechanisms through which depletion and dysregulation of queuosine modification may cause illness and developing innovative methods to detect Q in a high-throughput manner would significantly accelerate progress in this area.

1.2.2.3.5-taurinomethyl-2-thiouridine ($\tau m^5 s^2 U$)

$\tau m^5 s^2 U$ is an exclusively mitochondrial tRNA modification that occurs in mt-$tRNA^{Glu}$, mt-$tRNA^{Gln}$, and mt-$tRNA^{Lys}$ in wobble uridines [18,183,184]. mt-$tRNA^{Gln}$ is modified with $\tau m^5 s^2 U$

in roughly two out of three transcripts [184]. As stated in *1.2.2.*, the complex modifications at C5 of uridines increase the stability of the wobble base pair – in this case, a wobble U•G pair, as mt-tRNA^{Lys} decodes AAA and AAG [183,185]. It is unclear at this time the individual effect the s^2 modification has in this particular context, although data from other modifications indicate that may restrict the ability of the wobble U to pair with only purines and enhance the rate of translocation [186]. Currently, there is only one known enzyme in the τm^5s^2U: MTU1, which is responsible for thiolation of the O2 position [187]. Lack of the τm^5s^2 modification is directly causative of the myoclonus epilepsy, ragged-red fibers (MERRF) condition, a developmental disorder that affects much of the patient's physiology [20,23,188]. Interestingly, taurine depletion results in the same mt-tRNAs (Glu, Gln, and Lys) being modified with $cmnm^5s^2U$, a bacterial modification that has a glycine in place of taurine [189].

1.2.2.4.5-methoxycarbonylmethyl-2-thiouridine (mcm^5s^2U)

The cytosolic (rough) equivalent of τm^5s^2U is mcm^5s^2U, which possesses a methyl ester rather than the sulfate group at the C5 position and is found in tRNA^{Glu}, tRNA^{Gln}, and tRNA^{Lys} [18,158]. Like τm^5s^2U, mcm^5s^2U stabilizes wobble U•G and increases translational efficiency [158,190]. More is known about the biosynthetic pathway for mcm^5s^2U. Presently, it is known that CTU2 is responsible for thiolation of O2 [191], the Trmt112-ALKBH8 dimer synthesizes methylates cm^5 to mcm^5 using the methyltransferase protein [192], and the Elongator complex protein Elp3 acetylates U to cm^5U and is thought to be the likely first step in the biosynthesis of multiple highly modified U34 nucleotides [193]. CTU2 is the medically relevant enzyme and is associated with intellectual disabilities [20].

1.2.2.5.5-carboxymethylaminomethyl-2-thiouridine (cmnm^{5}s^{2}U)

Although hypermodified uridines are frequently found in eukaryotic tRNA, they are not exclusive to that kingdom of life. The wobble U34 modification cmnm^{5}s^2 is observed in both eukaryotes and eubacteria [194]. Similar to the other 2-thio hypermodified U34 nucleotides, these nucleotides are found only in tRNA^{Glu}, tRNA^{Gln}, and tRNA^{Lys} and performs similar functions to the other xm^{5}s^{2}U modifications [20,185,195]. This modification requires a glycine, which is connected to C5 of U34 by a methylene group. (This is also observed with taurine for τm^{5}s^{2}U.)

1.2.2.6.2-thiocytosine (s^{2}C)

s^{2}C has only been detected in the tRNAs of microorganisms, where it has been shown to impair wobble base-pairing when the third position in the codon is an A for Arg codons [196]. TtcA, an iron-sulfur cluster enzyme, catalyzes the thiolation of the C2, replacing the oxygen bonded to it [197]. However, little else has been documented about this modification.

1.2.3. tRNA 'modopathies'

A simple, but effective heuristic in biological sciences is that if a process or molecule is important, then its dysregulation will result in a negative phenotype, such as disease. Indeed, this is the case for many tRNA modifications, where mutations in either the modifying enzyme or the sites of tRNA modification reduce the fitness of cultured cells or are highly associated with/causative of a disease [20]. Diseases that have been shown to be directly caused by mutations in the tRNA, particularly at sites of modification, are frequently related to defects in the mitochondria. This is unsurprising, as human mitochondria have only 22 tRNAs and possess significantly fewer

modifications on average than cytosolic tRNAs. The maladies result from a loss of stabilizing modifications, such as m^1A, and modifications that enhance translation [23].

A wider variety of pathologies, including numerous cancers, are associated with or caused by the inactivation of enzymes that generate RNA modifications. I should note that these enzymes frequently play a multitude of roles, and thus these may not be predominantly modopathies. However, in many cases, there is substantial evidence to support a role for dysregulated tRNA modifications as a contributor to the disease in question [20].

1.2.4. Non-sequencing methods for detecting RNA modifications

Prior to the development of sequencing applications for the relevant RNA modification(s), methods such as liquid chromatography-mass spectrometry (LC-MS), 2D gel analysis, and *N*-acryloyl-3-aminophenylboronic acid-supplemented gel electrophoresis and Northern blotting were applied to separate the modifications based on their chemical nature [198-202]. These methods provided quantitative or semi-quantative analysis of modification levels and even some limited capacity to determine sequence contexts in the case of tandem LC-MS (LC-MS/MS) [198]. Because of their reliability in detecting and quantifying chemical changes, they are still applied in some cases to verify newly discovered modifications or validate sequencing results. Although potent, many of these methods suffer from significant drawbacks, including high amounts of material input, exceptionally low-throughput data generation, and hazardous conditions [198-202]. An exception to this trend is a method developed by our lab termed CMC-RT and ligation assisted PCR analysis of Ψ modification (CLAP) [203], although it sacrifices quantitative prowess for these advantages.

1.2.5. Methods for sequencing RNA modifications

Sequencing-based methods for detecting and quantifying RNA modifications rely on taking advantage of chemical features of the modification. For example, inosine is read by reverse transcriptases as guanosine, allowing it to be detected easily by comparison of the tRNA cDNA to the tRNA gene. Other modifications, such as queuosine, present a more significant challenge, as they occur on the Hoogsteen face and there are no known enzymatic tools to alter the modification, or, like m^6A, are present at the Watson-Crick face, but have little impact on base pairing. However, queuosine does possess *cis* diols on the cyclopentane ring, which can be chemically oxidized [3]. Additionally, antibodies to some modifications have been generated to enable enrichment of the modified transcripts [204]; however, this approach is limited to providing a 20-30 nucleotide window where the modification could be. Some Watson-Crick face methylations, such as those discussed in previous sections, can be removed by the *E. coli* AlkB dioxygenase and thus detected by comparison to untreated samples, although this enzymatic reaction can significantly reduce the number of reads, potentially generating bias in the final libraries [205,206]. Some methods have been developed that cleave the site of modification or inhibit reverse transcription, providing a clear stop signal [207,208]. However, these methods prevent detection of anything 5' of the modifications at the cleavage site. Lastly, modifications with a 2-thio or 2-methylthio group induce abortive reverse transcription at high frequencies (~ 95% stop rate), allowing them to be detected, but with a loss of any upstream data. These modifications can potentially be chemically modified to shift the ratio of stops to mutations towards the mutation fraction, allowing for read-though by the reverse transcriptase and the reporting of upstream nucleotides or modifications.

1.3. Small RNA Sequencing Methods

Given the abundant examples of critical biological functions of small RNAs detailed above, it comes as no surprise that a litany of small RNA sequencing methods has been developed over the past two decades. Many early methods were developed specifically with miRNAs and other small interfering RNAs in mind [209], while more recent methods have been designed with highly structured and post-transcriptionally modified small RNAs in mind [1,2,206,207]. In this section, I will discuss several of the scientific and practical requirements for small RNA sequencing methods, the two larger categories that most small RNA sequencing methods fall into, the details these methods, and the existing challenges and opportunities for the field of small RNA sequencing.

1.3.1. Ligation-based methods

Ligation is a common strategy in small RNA sequencing methods, as it allows functional and informational regions of DNA or RNA to be attached efficiently. Most attractively, there is the potential to install unique molecular identifiers from the earliest steps, allowing sample multiplexing to take place near the onset. There are two forms of ligation-based sequencing strategies: 1) simultaneous and 2) staggered. Simultaneous ligation strategies attach oligos at the 5’ and 3’ ends of the RNA substrates, in theory allowing reverse transcription and PCR to be performed in quick succession. For many classes of small RNAs, such as microRNAs, these approaches are highly useful and can generate a large quantity of data in a quick fashion. However, if reverse transcription is aborted before a sufficient length of the 5’ oligo is reverse transcribed, the read is lost. Abortive reverse transcription is commonly observed in RNAs with robust secondary structure and post-transcriptional modifications; therefore, simultaneous ligation

strategies can generally not be applied to study tRNAs, tRFs, or similar classes of RNAs without the use of highly processive reverse transcriptase.

Staggered ligation strategies attach an oligo to the 3' end of the RNA substrates, with the second oligo being attached to the 3' of the cDNA after reverse transcription. These strategies are necessary for studying highly structured and post-transcriptionally modified RNAs, but frequently suffer from low ligation efficiency and/or enzymatic bias. More recent methods, including my work that will be discussed in Chapter 2, have overcome this particular obstacle and the authors have discovered ligation conditions that allow low abundance RNAs to be captured with sufficient reads to produce quality sequencing data. However, a significant drawback for these methods is that the multiple enzymatic steps require additional purification steps, including gel extraction, which can result in the loss of a considerable amount of the input. Further, these purification steps expand the time between initiation and completion of library construction and greatly hinder investigations that require more than ~ 20 samples. In the following subsection, I will detail several small RNA sequencing methods to provide a brief introduction to this topic. I will continue to elaborate on this topic in Chapter 2, where I discuss the development of my thesis project and the resulting small RNA sequencing method.

1.3.2. Template switch-based methods

The engineering of a group-II intron-derived reverse transcriptase known as TGIRT enabled the development of template switch-based sequencing methods. Where retrovirus-derived reverse transcriptases synthesize between 10-20 nucleotides before disengaging with the template and the re-engaging, group-II intron-derived RTs are highly processive and can produce cDNAs up to approximately 10 kb in a single run [210,211]. However, they are single-turnover enzymes and do

not re-engage the template after falling off [210,212]. TGIRT and related RTs require a single-nucleotide overlap between a primer and the template, whereupon TGIRT can switch from DNA/RNA hybrid duplex with single nucleotide overlap on the 3' end of the DNA strand to the RNA [212]. The single-nucleotide overlap pairs with the 3' end of the RNA, which, in the case of tRNA sequencing, is a 3' T. The ability to combine the ligation and reverse transcription steps has significantly reduced number of purification steps in a method previously developed by the Pan lab. TGIRT also has the benefit of being able to read through several modifications, including Watson-Crick face methylations.

1.3.3. Recent small RNA sequencing methods

DM-tRNA-seq [207]

DM-tRNA-seq functions via a template-switching procedure. To begin, tRNAs are separated from total RNA using gel purification. After this, half of the sample is left untreated and the other half is subjected to AlkB demethylation. For preparing the sequencing library, a DNA/RNA duplex with a 3' dT overhang is added to anneal to the 3' rA of the purified tRNA. The DNA oligo contains the requisite PCR handles for Illumina sequencing. TGIRT then switches and begins reverse transcribing the tRNA. Once the RT reaction is completed, CircLigase is employed to ligate the 5' and 3' ends of the cDNA. Afterwards the resulting circularized cDNA libraries are gel purified and amplified by PCR.

mim-tRNAseq [152]

mim-tRNAseq is a ligation-based method that utilized a pre-adenylated DNA oligo to attach the PCR handles. T4 RNA ligase 2 attaches the DNA adaptor to the 3' end of the tRNA, which is then

reverse transcribed and subsequently circularized with CircLigase. Reverse transcription can be performed here using a wide variety of RT enzymes, which may enable read-through of different tRNA modifications.

YAMAT-seq [213]

Where mim-tRNAseq utilizes a single-stranded DNA adaptor, YAMAT-seq employs a double-stranded adaptor duplex that ligates to the 5' and 3' ends of tRNA. This specifically targets tRNA as the DNA strand that ligates to the tRNA 5' end has a 3' GGTN overhang to selectively anneal to the tRNA. This method achieves high adaptor ligation efficiency. With the 5' and 3' ends of the tRNA thus ligated, reverse transcription can take place. A significant drawback of YAMAT-seq, however, is that the final cDNA libraries will only capture full-length tRNAs, as any premature RT stops will not be amplified.

LOTTE-seq [214a] (QuantM-tRNA-seq [214b])

LOTTE-seq has several similarities with YAMAT-seq in the design of the adaptor oligo – with two significant differences: 1) the adaptor is a single-stranded oligo that anneals to itself and 2) the 3' overhang is phosphorylated at the 3' end, preventing its ligation to the 5' end of the tRNA. After reverse transcription, a second adaptor is ligated to the 3' end of the cDNA. The first adaptor oligo in LOTTE-seq takes advantage of the annealing 3' overhang while still enabling truncated tRNA reads to be captured in the final libraries. QuantM-tRNA-seq utilizes a similar strategy.

AQRNA-seq [215]

The first ligation step of AQRNA-seq is unbiased ligation to the 3’ end of all RNAs, not just tRNA. Subsequently, excess linker is removed and additional steps, such as AlkB demethylation, are performed. Then, the ligated RNAs are reverse transcribed and a second ligation is used to attach a second primer binding site. The second ligation oligo possesses an NNN NNN sequence near the 3’ end of the adaptor, allowing the oligo to pair with the cDNAs produced with sufficient stability for ligation to occur. This method is a significant improvement over previous applications due to the fact that it can generate data on many RNA class and not just tRNA.

Chapter 2

Multiplex Small RNA (MSR)-seq Provides a Representative Readout of Small RNA Abundance & Modification Fractions

Acknowledgment: This chapter is derived from an article published in *Nature Communications* by Springer Nature [1]. The authors of that article were: Christopher P. Watkins, Wen Zhang, Adam C. Wylder, Christopher D. Katanski, and Tao Pan. Author contributions: Conceptualization, C.P.W., C.D.K., and T.P.; Methodology: C.P.W., C.D.K., and T.P.; Software: C.D.K.; Formal Analysis, C.D.K.; Investigation, C.P.W., C.D.K., W.Z., and A.C.W.; Writing – Original Draft, C.P.W., C.D.K., and T.P.; Writing – Review & Editing, C.P.W., W.Z., A.C.W., C.D.K., and T.P.; Supervision, T.P.

2.1 Introduction

High-throughput RNA sequencing has provided transformative insights into cellular homeostasis, dynamic response to stress and environmental change, and RNA modifications. However, significant challenges associated with the characterization of small RNAs remain. Small RNAs are less than 200 nucleotides in length and include transfer RNAs (tRNA), microRNAs, small nucleolar RNAs, tRNA fragments, and many others that play important roles in cellular pathways and physiology. Altogether, small RNAs constitute more than 90% of cellular RNAs in copy numbers; among these, tRNA is the most abundant [9]. The role of tRNA in translational regulation depends on the expression and aminoacylation (charging) levels of different tRNA species, as well as many modifications that fine-tune tRNA activity [20,21,182]. Therefore, comprehensive and high-throughput characterization of tRNA is essential for a deeper understanding of the biological function of small RNAs.

Most commonly used RNA-seq methods are incompatible with the comprehensive study of small RNAs. Many small RNA-seq techniques ligate adapter oligonucleotides to the target RNAs, followed by cDNA synthesis. Products of incomplete reverse transcription, often induced by RNA modification or structure, are not amplified and not included in downstream analysis. tRNA is the RNA family most limited by these methods, due to rigid secondary structure and extensive modification. A common drawback of other approaches is the necessity of size selection steps or sequence context requirements, which limit the investigation of RNAs to a certain size range or to specific families and uncouples the coordinated expression and response of small RNA families, e.g., tRNA with tRNA fragments. Finally, small RNA-seq procedures also lack the level of multiplexing enjoyed by mRNA sequencing [216]. Therefore, new small RNA-seq methods are still needed to better characterize tRNA properties, incorporate tRNA results with other small RNA families, and increase multiplexing capability.

Here, I describe multiplex small RNA sequencing (MSR-seq), a platform for RNA-seq library construction that provides multiplexing to greatly increase throughput. The key feature of MSR-seq is the design of a biotinylated oligonucleotide that is used for barcode adapter ligation, immobilization, on-bead reverse transcription, second adapter ligation, and PCR. This unification of multiple steps in RNA-seq library construction enables multiplexing of many samples in the same reaction which increases sample handling throughput and reduces input amount. Our method also allows for the inclusion of enzymatic and chemical treatment of RNA on-bead, thus accommodating the investigation of RNA modifications and other applications.

2.2 Results

The basic steps of small RNA-seq library construction include first adapter ligation, reverse transcription, second adapter ligation, and PCR amplification. Additions can include enzymatic or chemical treatments of the RNA to profile RNA modifications or map RNA structures. Our goal is to immobilize and barcode RNA samples at the earliest step possible, thereby enabling all subsequent reactions, including optional enzymatic/chemical reactions to proceed on a solid support, which increases multiplexing potential and minimizes sample handling (Figure 2.1(a)). First, a barcode was ligated to the 3′ of input RNA from any biological source. Since > 90% of cellular RNA in copy numbers were small RNA, they constituted most of the ligated product, thus further isolation was not necessary. Barcoded samples were pooled and immobilized on streptavidin beads. Enzymatic or chemical treatments could then be added before reverse transcription. On-bead reverse transcription was followed by second adapter ligation to the cDNA and PCR; the off-bead PCR products were readily used for sequencing.

2.2.1. Design & optimization of MSR-seq

We designed a capture hairpin oligonucleotide (CHO) that contains several features to make all steps compatible with streptavidin beads (Figure 2.1(b)). The salient features include: (i, ii) a 5′ phosphate and ribonucleotide for efficient first ligation [217]; (iii) a barcode sequence for sample pooling and multiplexing; (iv) a biotin moiety for immobilization; (v) a primer binding site embedded in the hairpin for reverse transcription and PCR; (vi) a 3′ ribonucleotide to prevent second adapter ligation to free CHO: this ribonucleotide enables periodate oxidation of the unligated CHO which converts its 2′,3′ hydroxyls to 2, 3′ aldehydes at the 3′ end. Only 2′,3′

hydroxyls allow for the ligation of the second adapter. (vii) a 3′ phosphate to prevent self-ligation of the excess CHO.

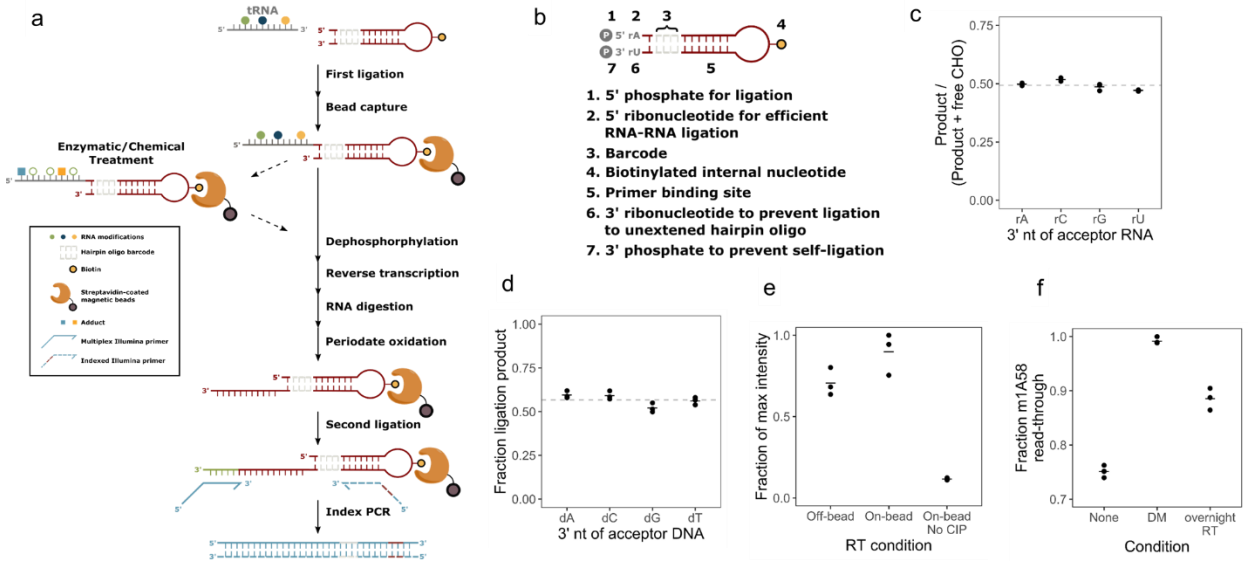

Figure 2.1: Design and optimization of MSR-seq. (a) Schematic representation of multiplex small RNA-seq. The required steps are indicated in solid arrows and optional steps in dashed arrows. Symbols are explained in the box. (b) Features of the capture hairpin oligo (CHO) with detailed descriptions for each feature. (c) Fraction of ligation products for the test RNA oligonucleotides containing rA, rC, rG, and rU at the 3′ end. n = 3 independent experiments. The mean value is shown as a bar; the mean of all replicates is shown as a dashed line. The molar ratio as measured by UV absorbance of input RNA and the CHO in the ligation reaction was ~1:2. Fraction product is calculated as (Product)/(Product + Free CHO). The expected stoichiometric amount of ligation product corresponds to ~0.5. (d) Fraction of ligation products for the test 5′ biotinylated deoxyoligonucleotides containing dA, dC, dG, and T at the 3′ end. n = 3 independent experiments. The mean value is shown as a bar; the mean of all replicates is shown as a dashed line. Fraction product is calculated as (Product)/(Product + Free 5′biotin-oligo). (e) Relative reverse transcription products with RT performed free in solution (off-bead), on streptavidin bead (on-bead), and on-bead without calf-intestine alkaline phosphatase treatment (No-CIP). Input samples were total RNA from HEK293T cells. n = 3 independent experiments. The mean value of replicates is shown as a bar. Product is defined as all bands above the terminal transferase product of the RT (TdT) band. Fraction product is calculated as (Product in each sample)/(sample with the maximum amount of product). The nonzero value from No-CIP samples was from spurious oligos associated with the beads. (f) Fraction of reverse transcriptase m^1A58 read-through product with no demethylase treatment/short RT reaction (none), demethylase treatment/short RT reaction (DM), and no demethylase treatment/long RT reaction (overnight RT). Input samples were total RNA from HEK293T cells. n = 3–4 independent experiments. The mean value of replicates is shown as a bar. Product is defined as all bands above the m^1A58 band. Product is defined as all bands above the TdT band. Fraction product is calculated as (Product)/(Product + m^1A58 stopband).

We optimized reaction conditions and demonstrated high efficiency and equity for RNA substrates [218]. To examine the bias of the 3′ end nucleotide identity of the sample RNA in the first ligation, we measured the ligation efficiency of four RNA oligonucleotides of 5′ N10X (X = A, C, G, U) as the input samples. The ligation generated stoichiometric amounts of products for all four RNA oligos under our reaction condition (Figure 2.1(c), Figure 2.2(a)). To examine the bias of the 3′ end nucleotide identity of cDNA products in the second ligation, we used four 5′ biotinylated DNA oligonucleotides ending with dN10dX (X = dA, dC, dG, T) as the input samples. The on-bead ligation generated high levels of product with little bias for the 3′ deoxynucleotide identity (Figure 2.1(d), Figure 2.2(b)). To examine the efficiency of on-bead reverse transcription, we used total HEK293T RNA as input and ran the RT reaction in parallel either on or off the streptavidin bead and measured the product amounts by RT-PCR. The on-bead RT reaction generated a higher amount of products than the off-bead RT reaction (Figure 2.1(e), Figure 2.2(c)).

We deployed two innovations to enable all steps after the first barcode ligation to be performed on-bead. First, we added a 3′ phosphate that blocked the self-ligation of CHO during the first ligation. This phosphate was removed on-bead to allow for subsequent reverse transcription from the 3′ OH. No specific product was visible without the 3′ phosphate removal by phosphatase (Figure 2.1(e), Figure 2.2(c)). Second, we designed a 3′ ribonucleotide in CHO which enabled the reduction of PCR products from the excess CHO on-bead. After cDNA synthesis, the CHO with ligated input RNA contained a terminal deoxyribonucleotide which protected it from periodate oxidation and thus made it a substrate for the second adapter ligation. In contrast, the excess, free CHO containing 3′ ribonucleotide was periodate oxidized and no longer a ligation substrate (Figure 2.2(d)).

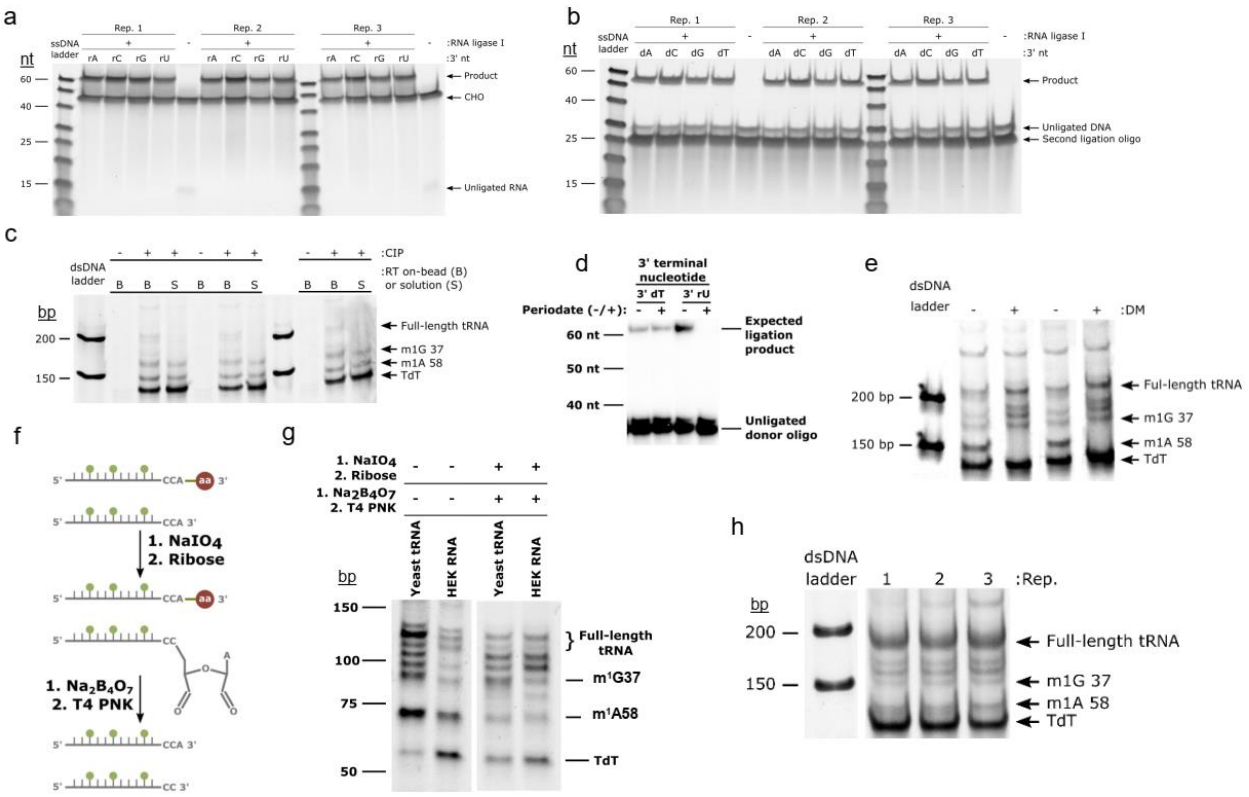

Figure 2.2: Optimization of MSR-seq reaction steps. (a) First ligation of synthetic RNA oligos to capture hairpin in solution. Samples inputs were four RNA 11mers of 5' N_{10}-A/C/G/U (labels indicate the identity of the 3' nucleotide of the substrate RNA, i.e., ribo-uricil = rU). Results are quantified in Figure 2.1(c). All data from n=3 biological replicates is shown. (b) Second ligation of synthetic DNA oligos to ligation oligo donor on streptavidin beads. Sample inputs were four 5' biotinylated DNA 26mers ending with dN_{10}-dA/dC/dG/T intended as mock cDNA products featuring different terminal nucleotides. The ligation oligo donor was the same used for ligation to the cDNA product on bead in the library construction. Results are quantified in Figure 2.1(d). All data from n=3 biological replicates is shown. For panels (c), (e), (g), (h), DNA size markers are on the left. Major RT stops caused by the base methylations (m^1A58, m^1G37) in human tRNAs are indicated on the right. TdT corresponds to the product derived from the aberrant terminal transferase activity of the RT. (c) The efficacy of reverse transcription on immobilized substrate (B) was compared to substrate in solution (S). Libraries were constructed using total RNA from HEK293T cells. Streptavadin immobilization was omitted for solution samples during the RT step. As a control, calf intestine alkaline phosphatase (CIP) was omitted from some samples. Results are quantified in Figure 2.1(e). All data from n=1 replicate shown. (d) Periodate oxidation test with 3' dT or 3'rU ends. Ligation was performed in solution with one donor oligonucleotide and two acceptor oligonucleotides containing 3' terminal dT or rU. Periodate oxidation was performed as indicated, followed by ligation. All data from n=1 replicate shown. (e) MSR-seq libraries were constructed from HEK total RNA, with (+DM) or without (-DM) demethylase treatment. Bands corresponding to RT stops at known modifications are labeled; DM treatment has a strong effect on m^1A58-induced stops. Results are quantified in Figure 2.1(f). All data from n=2 replicates shown. (f) Schematic representation of measuring tRNA charging by sequencing in one-pot periodate oxidation and β-elimination. $NaIO_4$: sodium periodate; $Na_2B_4O_7$: sodium tetraborate.

Figure 2.2, continued: (g) MSR-seq libraries were constructed with (+) and without (-) the one-pot β-elimination protocol from HEK total RNA. 'Without beta-elimination' samples were separately deacylated, cleaned through a spin column, then used for library construction. By contrast, 'with beta-elimination' samples were used directly in the first ligation reaction following chemical treatment. Final PCR products of libraries were separated on 10% TBE gel and stained with SYBR Gold. $NaIO_4$: sodium periodate, $Na_2B_4O_7$: sodium tetraborate. All data from n=1 replicate shown (h) MSR-seq libraries were constructed from HEK total RNA using overnight RT conditions that facilitate reading through certain modifications. Results are quantified in Figure 2.1(f). All data from n=3 replicates shown.

After the first ligation and streptavidin bead binding of all CHO, the sample could be optionally split in two: one mock treated, and the other subjected to chemical or enzyme treatment. To validate that enzyme treatment is compatible with bead immobilization, we compared libraries prepared with and without AlkB treatment. The on-bead AlkB demethylase reaction efficiently reduced the PCR product bands derived from tRNA m^1A/m^1G modification relative to the full-length tRNA [206] (Figure 2.1(f), Figure 2.2(e)). To facilitate tRNA charging studies [218], we modified the oxidation and β-elimination protocol to enable the sequential addition of these reagents in a single tube. The product of this β-elimination reaction is used directly in the CHO ligation so that no reaction intermediates were precipitated or purified (Figure 2.2(f,g)).

The only caveat of our approach was the known terminal deoxyribonucleotide transferase activity (TdT) of reverse transcriptases [219], which produced varying amounts of PCR products derived from the free CHO on-bead. This artificial product could be removed through gel extraction of the final PCR products.

Although demethylase treatment in our DM-tRNA-seq method improved efficiency and quantitation [206], it results in duplication of library preparation for the same sample, rendering it less practical for high volume tRNA-seq experiments. Similar to mim-tRNA-seq [152], we found an SSIV RT reaction condition (Figure 2.1(f), Figure 2.2(h)) that could more effectively read

through tRNA methylations, thus enabling the investigation of tRNA abundance, charging, and modification simultaneously with a single sequencing library.

MSR-seq also allows for chemical treatment of RNA on-bead which is useful for RNA modification studies or RNA structural mapping [220]. We used the well-established N-cyclohexyl-N′-β-(4-methylmorpholinium) ethylcarbodiimide (CMC) reaction for pseudouridine (Ψ) modification [221] to test the compatibility of MSR-seq on-bead protocol with harsh chemical treatment conditions (Figure 2.3(a)). To map the Ψ sites in human rRNA, we fragmented total RNA, ligated the fragments to the CHO, then performed the CMC reaction on-bead. We assigned each rRNA position a stop and mutation fraction and observed a good correlation between biological replicates (Figure 2.3(b)). We identified strong signals in the stop and/or mutation fractions in the CMC-treated sample at the 35 of 36 known Ψ sites [222] in the 18S rRNA (Figure 2.3(c)), validating the usefulness of our approach.

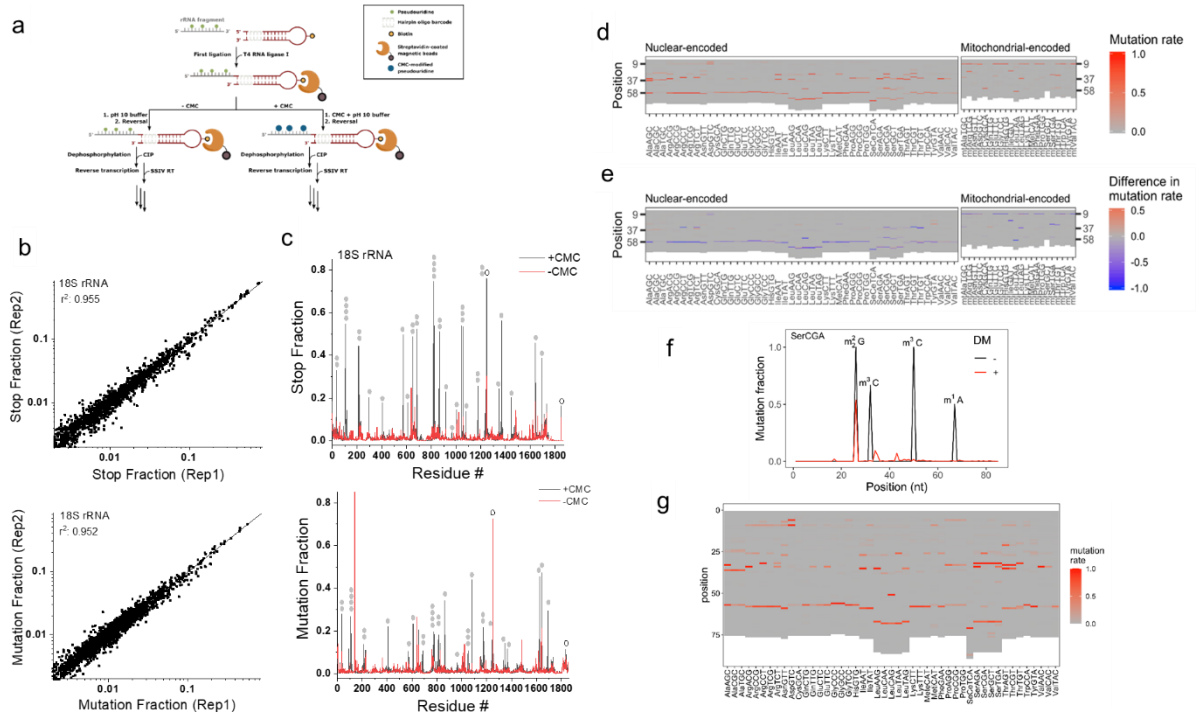

Figure 2.3: Sequencing results from chemical and enzymatic treatment on bead. (a) Schematic representation of incorporating CMC reaction in MSR-seq for pseudouridine (Ψ) mapping. Total RNA was fragmented, 3′ end-repaired, and ligated to the capture hairpin oligonucleotide. CMC reaction was done on-bead before cDNA synthesis. (b) CMC reaction mapping replicates: Top graph shows the RT stop fraction and the bottom graph shows the mutation fraction for every residue in human 18 S rRNA among the biological replicates. (c) Stop and mutation fraction along 18 S rRNA without (red) and with CMC (black) treatment: Top graph shows the RT stop fraction and the bottom graph shows the mutation fraction at each nucleotide position. Known Ψ sites are marked by filled gray ovals, known $m^1acp^3\Psi$ and $m^6{}_2A$ sites by open ovals. (d) Heat map showing the mutation signature at every nucleotide for the most abundant RNA isodecoder in each isoacceptor family, no demethylase treatment. Nuclear-encoded and mitochondrial-encoded tRNAs are shown separately. (e) Heatmap showing changes in mutation signature upon demethylase (DM) treatment at every nucleotide for the most abundant tRNA isodecoder in each isoacceptor family. Red indicates an increase and blue indicates a decrease in mutation fraction. (f) Mutation fraction across a tRNASer(CGA) isodecoder ± demethylase treatment (DM) showing the effective removal of base methylation. Known modifications are indicated. (g) Same as (d), using overnight RT condition for library construction.

To quantitatively compare the MSR-seq result with another previously published tRNA-seq method, we compared MSR-seq and DM-tRNA-seq results for tRNA isodecoder expression in HEK293T cells [206]. We observed a good correlation (Figure 2.4(a)), even though MSR-seq and DM-tRNA-seq used different RTs, steps in library construction, and input RNAs. To test the sample input limits of the MSR-seq method, we built libraries starting with 1000, 100, and 10 ng of total HEK293T RNA (Figure. 2.4(b)). tRNA abundance was well correlated between these libraries with r^2 ~0.94.

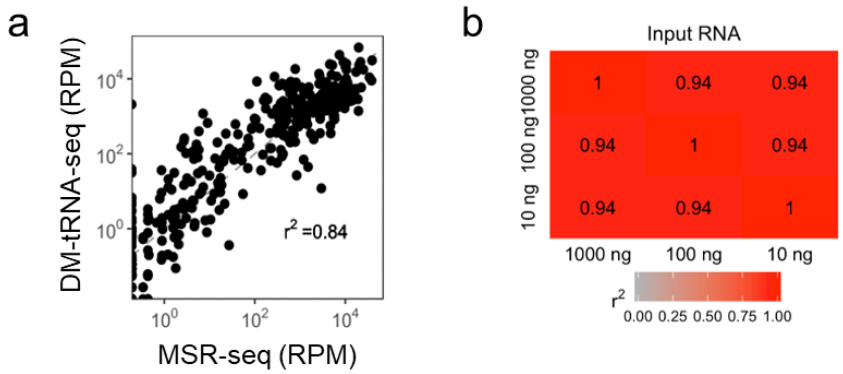

Figure 2.4: Comparison of MSR-seq to DM-tRNA-seq and evaluation of MSR-seq replicability with varying input quantiaties. (a) Correlation of tRNA isodecoder abundance from libraries constructed with MSR-seq with demethylase treatment compared to published DM-tRNA-seq [206]. RPM: reads-per-million. Sample for MSR-seq was from HEK293T total RNA; sample for DM-tRNA-seq was from gel purified HEK293T total tRNA. R^2 value corresponds to Spearmen's correlation coefficient. (b) MSR-seq libraries were constructed from HEK total RNA with varying amounts of input material. Spearman's correlation coefficients between tRNA isodecoder abundance compared between libraries is presented.

To validate that demethylase treatment worked well on bead, we found that Watson–Crick face methylations such as m^1A, m^3C, and m^1G produced substantial mutation signatures as expected from previous studies [137,223] (Figure 2.3(d)), but demethylase treatment abolished or reduced the mutation fractions associated with these methylations (Figure 2.3(e, f)). Validation of mutation signature as RNA modifications with demethylase treatment still has value when working

with less well-characterized samples. We also found that our overnight RT condition reports a very similar tRNA modification landscape as the short RT reaction (Figure 2.3(g)).

2.3 Discussion

Here I developed an RNA-seq method that enables multiplexed sequencing library preparation, on-bead enzymatic and chemical treatment, one-pot tRNA abundance, charging, and modification measurements, and simultaneous analysis of tRNA, tRNA fragment, and other small RNA families.

Advantages of carrying out sequencing library construction on solid support include the rapid exchange of buffers and reagents, thorough removal of contaminants, and elimination of size selection steps or adapter/primer removal. The solid support platform also allows for on-bead treatment of RNA with chemicals and enzymes, which has become widely used in studies of RNA modifications and other applications [224-227]. We found that the RNA libraries on streptavidin beads could withstand harsh chemical treatments such as the CMC reaction, which involves two steps at pH 8–10 and hours of incubation at 30–37 °C. Chemical and enzymatic treatments are useful in profiling RNA modifications such as Ψ, m^5C, m^1A, or m^7G [228,229] and for RNA structural mapping.

The key feature of our method is the design of the capture hairpin oligonucleotide (CHO). Our innovation is the inclusion of a 3′ ribonucleotide and 3′ phosphate which blocks the second ligation to the excess CHO upon periodate oxidation after the RT reaction so that only CHOs with cDNA product are amplified in the final PCR step. The presence of a substantial amount of the terminal transferase product (TdT) [219] in the RT reaction is the single issue that remains to be solved. We found that the amount of TdT was highly variable depending on the input sample and

reaction conditions. Currently, we remove the TdT products by size selecting the final PCR products. Possible future solutions include screening other RTs, further optimization of reaction conditions, and finding new enzymes that can selectively remove the added deoxynucleotides.

2.4 Materials & Methods

2.4.1 One-pot deacylation & β-elimination for tRNA charging

Up to 500 ng of total RNA in 7 µL was used for optional one-pot beta-elimination prior to library construction. To start, 1 µL of 90 mM sodium acetate buffer, pH 4.8 was added to 7 µL input RNA. Next, 1 µL of freshly prepared 150 mM sodium periodate solution was added for a reaction condition of 16 mM $NaIO_4$, 10 mM NaOAc pH 4.8. Periodate oxidation proceeded for 30 min at room temperature. Oxidation was quenched with the addition of 1 µL of 0.6 M ribose to 60 mM final concentration and incubated for 5 min. Next 5 µL of freshly prepared 100 mM sodium tetraborate, pH 9.5 was added for a final concentration of 33 mM. This mixture was incubated for 30 min at 45 °C. To stop β-elimination and perform 3′ end-repair, 5 µL of T4 PNK mix (200 mM Tris-HCl pH 6.8, 40 mM $MgCl_2$, 4 U/µL T4 PNK, from New England Biolabs) was added to the reaction and incubated at 37 °C for 20 min. T4 PNK was heat-inactivated by incubating at 65 °C for 10 min. This 20 µL reaction mixture can be used directly in the first barcode ligation by adding 30 µL of a ligation master mix described below.

2.4.2 Standard tRNA deacylation

Total RNA was prepared for library construction by first deacylating in a solution of 100 mM Tris-HCl, pH 9.0 at 37 °C for 30 min, then neutralized by addition of sodium acetate, pH 4.8 to a final

concentration of 180 mM. Deacylated RNA was then ethanol precipitated and resuspended in water or desalted using a Zymo Research Oligo Clean-and-Concentrator spin column.

2.4.3 General protocol for MSR-seq

2.4.3.1 First barcode ligation

Depending on the nature of the experiment described in the main text, input RNA or total nucleic acid samples were either deacylated or had undergone one-pot deacylation and β-elimination as described above. Up to 1 µg of total RNA input was used in a ligation reaction of 50 µL with the following components: 1 U/µL T4 RNA ligase I (NEB), 1x NEB T4 RNA ligase I buffer, 15% PEG 8000, 50 µM ATP, 1 mM hexamine cobalt chloride, and 5% DMSO. After adding the ligation mix to the sample, the capture hairpin oligo (CHO) was added to a final concentration of 1 µM and the samples were incubated at 16 °C overnight (12+ h).

The CHO sequences are listed below:

Abbreviations:

/5Phos/ – 5’ phosphorylation

rN – Ribonucleic acid base

/iBiodT/ – Internally biotinylated thymine

/3Phos/ – 3’ phosphorylation

Read2_BC1: /5Phos/rACT GGAA AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT TTCC AG rU/3Phos/

Read2_BC2: /5Phos/rACT CAGA AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT TCTG AG rU/3Phos/

Read2_BC3: /5Phos/rACT ACCA AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT TGGT AG rU/3Phos/

Read2_BC4: /5Phos/rACT TCAG AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT CTGA AG rU/3Phos/

Read2_BC5: /5Phos/rACT ATGG AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT CCAT AG rU/3Phos/

Read2_BC6: /5Phos/rACT GATG AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT CATC AG rU/3Phos/

Read2_BC7: /5Phos/rACT CTAC AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT GTAG AG rU/3Phos/

Read2_BC8: /5Phos/rACT TACC AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT GGTA AG rU/3Phos/

Read2_BC9: /5Phos/rACT AGTC AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT GACT AG rU/3Phos/

Read2_BC10: /5Phos/rACT TGGT AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT ACCA AG rU/3Phos/

Read2_BC11: /5Phos/rACT GTCT AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT AGAC AG rU/3Phos/

Read2_BC12: /5Phos/rACT CCTT AGA TCG GAA GAG CAC ACG at/iBiodT/ aga CGT GTG CTC TTC CGA TCT AAGG AG rU/3Phos/

2.4.3.2 Binding to dynabeads

The ligation mixture was diluted by adding an equal volume of water to reduce the viscosity of the solution. Next, streptavidin-coated MyOne C1 Dynabeads (Thermo Fisher) were added to each sample in a 1.2:1 excess over CHO (for example, a 50 µL reaction had 50 pmol capture hairpin oligo; beads were supplied at 10 mg/ml and had a binding capacity of 500 pmol biotinylated oligo per mg, so use 12 µL beads). The bead-sample mixture was incubated at room temperature for 15 min. After binding, supernatants were removed, and the beads were washed once with high salt wash buffer (1 M NaCl, 20 mM Tris-HCl, pH 7.4) and once with low salt wash buffer (100 mM NaCl, 20 mM Tris-HCl, pH 7.4). After washing, multiple individually barcoded samples can be combined for downstream steps. At this stage, enzymatic or chemical treatments can be incorporated into the library preparation protocol such as AlkB demethylase reaction or CMC treatments (see below).

2.4.3.3 Dephosphorylation

A 50 µL dephosphorylation mix containing the following was added to the multiplexed sample on bead: 0.04 U/µL calf-intestine phosphatase (Roche), 10 mM $MgCl_2$, 0.5 mM $ZnCl_2$, 20 mM HEPES, pH 7.3. The sample was incubated at 37 °C for 30 min. The sample was then washed once with high salt wash buffer and once with low salt wash buffer, then resuspended in 20 µL water.

2.4.3.4 Reverse transcription

Five microliters of SuperScript IV VILO 5x master mix (Thermo Fisher) were added to the dephosphorylated sample (final volume of 25 µL) and then incubated at 55 °C for 10 min. The sample was then washed once with high salt wash buffer and once with low salt wash buffer.

Identical results were obtained upon extending the reaction time to 60 min. For overnight reverse transcription, after the initial 10 min incubation at 55 °C, the samples were further incubated at 37 °C overnight.

2.4.3.5 RNase H digestion

Beads were then resuspended into 50 µL of RNase H master mix containing 0.4 U/uL RNase H (NEB) and 1x NEB RNase H buffer and incubated at 37 °C for 15 min. The sample was then washed once with high salt wash buffer and once with low salt wash buffer. The sample was then resuspended in 40 µL water.

2.4.3.6 Periodate oxidation

Ten microliters of 250 mM freshly prepared sodium periodate, 0.5 M sodium acetate, pH 5 were added to the RNase H digested sample and incubated at room temperature for 30 min. Afterward, ribose was added to a final concentration of 167 mM to quench excess periodate at room temperature for 5 minutes. The sample was then washed once with high salt wash buffer and once with low salt wash buffer.

2.4.3.7 Second ligation

Beads were resuspended into 50 µL of a ligation master mix with the following components: 2 U/µL T4 RNA ligase I (NEB), 1x NEB T4 RNA ligase I buffer, 2 µM second ligation oligo, 25% PEG 8000, 50 µM ATP, 7.5% DMSO, and 1 mM hexaammine cobalt chloride. After incubation at room temperature overnight (12+ hours), the reaction was diluted with 50 µL water to reduce viscosity, washed once with high salt wash buffer and once with low salt wash buffer, and then

resuspended in water. The amount of water was 6 µL per sample in the first ligation reaction, before pooling the barcoded samples. For example, if the second ligation mixture contains a pool of six samples, the amount of water used for resuspension would be 36 µL. Samples can be stored at 4 °C or frozen at - 20 °C; both can be used for the next PCR step.

The second ligation oligo sequence is as follows:

Abbreviations:

/5Phos/ – 5’ phosphorylation

/3ddC/ – 3’ dideoxycytosine

Read1_L2: /5Phos/NN NNN NGA TCG TCG GAC TGT AGA A/3ddC/

2.4.3.8 Polymerase chain reaction

A 50 µL PCR reaction was run using 5–10% of the bead slurry products from the second ligation reaction using Q5 DNA polymerase (NEB) and following the manufacturer’s instructions: 0.02 U/µL Q5 DNA polymerase, 1x Q5 reaction buffer, 0.2 mM dNTPs, 0.5 µM Illumina index primer, and 0.5 µM Illumina multiplex primer. Typical PCR cycles were 9–15 cycles at 10 s at 98 °C, 15 s at 55 °C, and 72 °C for 20 s. PCR reactions were then processed through the DNA Clean and Concentrate kit (Zymo Research).

The PCR primer sequences are listed below (index in red):

Index 1: CAA GCA GAA GAC GGC ATA CGA GAT CGT GAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 2: CAA GCA GAA GAC GGC ATA CGA GAT ACA TCG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 3: CAA GCA GAA GAC GGC ATA CGA GAT GCC TAA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 4: CAA GCA GAA GAC GGC ATA CGA GAT TGG TCA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 5: CAA GCA GAA GAC GGC ATA CGA GAT CAC TGT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 6: CAA GCA GAA GAC GGC ATA CGA GAT ATT GGC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 7: CAA GCA GAA GAC GGC ATA CGA GAT GAT CTG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 8: CAA GCA GAA GAC GGC ATA CGA GAT TCA AGT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 9: CAA GCA GAA GAC GGC ATA CGA GAT CTG ATC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 10: CAA GCA GAA GAC GGC ATA CGA GAT AAG CTA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 11: CAA GCA GAA GAC GGC ATA CGA GAT GTA GCC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 12: CAA GCA GAA GAC GGC ATA CGA GAT TAC AAG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 13: CAA GCA GAA GAC GGC ATA CGA GAT TTG ACT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 14: CAA GCA GAA GAC GGC ATA CGA GAT GGA ACT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 15: CAA GCA GAA GAC GGC ATA CGA GAT TGA CAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 16: CAA GCA GAA GAC GGC ATA CGA GAT GGA CGG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 17: CAA GCA GAA GAC GGC ATA CGA GAT CTC TAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 18: CAA GCA GAA GAC GGC ATA CGA GAT GCG GAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 19: CAA GCA GAA GAC GGC ATA CGA GAT TTT CAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 20: CAA GCA GAA GAC GGC ATA CGA GAT GGC CAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 21: CAA GCA GAA GAC GGC ATA CGA GAT CGA AAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 22: CAA GCA GAA GAC GGC ATA CGA GAT CGT ACG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 23: CAA GCA GAA GAC GGC ATA CGA GAT CCA CTC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Index 24: CAA GCA GAA GAC GGC ATA CGA GAT GCT ACC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

Multiplex primer: AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GAC GAT C

2.4.3.9 TBE-PAGE gel extraction

Following desalting, PCR products were run on 10% non-denaturing TBE gel with dsDNA size markers; lanes were cut according to the desired product size, mashed by pipette tip, and then resuspended in crush-and-soak buffer (500 mM sodium acetate, pH 5.0). The gel fragments were extracted overnight and then ethanol precipitated.

2.4.4 First ligation bias test

Four separate mixtures containing final 0.8 µM CHO and four 11-mer RNA oligonucleotides (5′ N10X, where X = rA, rC, rG, rU) were made at an approximate molar ratio of 2 CHO: 1 RNA oligo. Four microliters of each mixture were added to individual tubes (n = 3 for each RNA oligo). To each tube, 21 µL of first ligation mixture (18% PEG 8000, 1.2x NEB T4 RNA ligase I buffer, 60 µM ATP, 6% DMSO, 1.2 mM hexaammine cobalt (III) chloride, and 2 U/µL NEB T4 RNA ligase I) were added (final volume of 25 µL) and the samples were incubated at 16 °C overnight.

After overnight incubation, 2.5 µL of 18 mg/mL Roche PCR-grade Proteinase K and 15 µL of 2x RNA loading dye (9 M urea, 0.02% bromophenol blue, and 0.02% xylene cyanol) were added to each sample to degrade the ligase. The samples were then incubated at 37 °C for 25 min and subsequently boiled at 92 °C for 3 min.

The samples were loaded onto a Bio-Rad pre-cast 15% TBE-Urea PAGE gel and electrophoresed until the bromophenol blue was near the bottom of the gel. The gel was then

stained with a 10,000x dilution of SYBR Gold and imaged on a ChemiDoc imager and quantified with Quantity One software (Bio-Rad).

2.4.5 Second ligation bias test

Four separate mixtures containing the final 2 µM second ligation oligo and four 26-mer 5′ biotinylated DNA oligonucleotides (5′ biotin- CTCTTCCGATCTAGT N10X, where X = dA, dC, dG, T) were made at an approximate molar ratio of 10 s ligation oligo: 1 DNA oligo. Four microliters of each mixture were added to individual tubes (n = 3 for each biotin-DNA oligo). Three microliters of 10 mg/mL Thermo Fisher streptavidin-coated MyOne C1 Dynabeads were then added to each sample and allowed to capture the biotinylated DNA oligos for 10 min.

Then, to each tube, 21 µL of second ligation mixture (30% PEG 8000, 1.2x NEB T4 RNA ligase I buffer, 60 µM ATP, 9% DMSO, 1.2 mM hexaammine cobalt (III) chloride, and 4 U/µL NEB T4 RNA ligase I) was added (final volume of 25 µL) and the samples were incubated at room temperature (~20 °C) overnight.

After overnight incubation, 2.5 µL of 18 mg/mL Roche PCR-grade Proteinase K and 15 µL of 2x RNA loading dye (9 M urea, 0.02% bromophenol blue, and 0.02% xylene cyanol) were added to each sample to release the products from the beads and to degrade the ligase. The samples were then incubated at 37 °C for 25 min and subsequently boiled at 92 °C for 3 min.

The samples were loaded onto a Bio-Rad pre-cast 15% TBE-Urea PAGE gel and electrophoresed until the bromophenol blue was near the bottom of the gel. The gel was then stained with a 10,000x dilution of SYBR Gold and imaged on a ChemiDoc imager quantified with Quantity One software (Bio-Rad).

2.4.6 On-bead reverse transcriptase and phosphatase (CIP) treatment validation test

A first ligation mixture was made containing the following: 15% PEG 8000, 1x NEB T4 RNA ligase I buffer, 50 µM ATP, 1 mM hexaammine cobalt (III) chloride, 0.8 µM CHO, 25 ng/µL HEK total RNA, and 1.67 U/µL NEB T4 RNA ligase I (total volume of 275 µL). This was incubated at 16 °C overnight.

After overnight ligation, 75 µL (equivalent to 3 × 25 µL ligation reactions) were pipetted into a separate tube (henceforth known as "- CIP" tube) and mixed with 75 µL of 100 mM EDTA to quench the ligation reaction. The remaining ligation mixture was ethanol precipitated.

The ethanol-precipitated ligation mixture (henceforth known as "+CIP" tube) was resuspended in 92 µL of water and mixed with 100 µL of 2x dephosphorylation buffer (20 mM MgCl2, 1 mM ZnCl2, and 40 mM HEPES, pH 7.5) and 8 µL of 5 U/µL NEB Quick CIP. The "+ CIP" tube was then incubated at 37 °C for 30 min. After the 37 °C incubation, the "+CIP" tube was incubated at 65 °C for 10 min to inactivate the phosphatase. The "+ CIP" sample was then ethanol precipitated.

After ethanol precipitation, the material in the "+ CIP" tube was resuspended in 64 µL of water. Forty-eight microliters were then pipetted out and split into six separate tubes (8 µL each). To three of these tubes, 8 µL of 10 mg/mL Thermo Fisher streptavidin-coated MyOne C1 Dynabeads were added. To the "- CIP" tube, 24 µL of beads were added. The beads were incubated with the samples for 10 min to allow them to bind the biotinylated oligos. (The "−CIP" tube was then split into three separate tubes.) After incubation, the beads were then magnetized, the supernatant was removed, and the beads were washed once with 50 µL of high salt Neonate wash buffer (20 mM Tris, 1 M NaCl, and 0.1% Neonate 20, pH 7.4) and then washed once with low salt wash buffer (20 mM Tris and 0.1 M NaCl, pH 7.4).

After washing, the samples were resuspended in 8 µL of autoclaved water. To all nine tubes, 2 µL of Thermo Fisher 5x SuperScript IV VILO mix were added and the samples were then incubated at 55 °C for 10 min. After incubation, the off-bead samples were then mixed with 8 µL of beads. All samples were then washed as before.

From this point on, the MSR-seq protocol was followed, beginning at the RNase H digestion section.

2.4.7 AlkB & AlkB-D135S purification

These were adapted from the previously described protocols for DM-tRNA-seq5. Briefly, NEB T7 Expression cells were grown in LB media at 37 °C, in the presence of 50 µM kanamycin, to an A600 of 0.6–0.8. Once the cells reached the desired density, IPTG and iron sulfate were added to final concentrations of 1 mM and 5 µM, respectively. After induction, the cells were incubated overnight at 30 °C. Cells were collected, pelleted and then resuspended in lysis buffer (10 mM Tris, pH 7.4, 5% glycerol, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol) plus 300 mM NaCl. The cells were lysed by sonication and then centrifuged at 17,400 rcf for 20 min. The soluble proteins were first purified using a Ni-NTA superflow cartridge (Qiagen) with buffers A (lysis buffer plus 1 M NaCl for washing) and B (lysis buffer plus 1 M NaCl and 500 mM imidazole for elution) and then further purified by ion-exchange (Mono S GL, GE Healthcare) with buffers A (lysis buffer plus 100 mM NaCl) for column loading and B (lysis buffer plus 1.5 M NaCl) for elution.

2.4.8 AlkB treatment

Demethylase buffer conditions were modified from ref. 230. Three stock solutions are made fresh immediately before reaction: L-ascorbic acid 200 mM, 2-ketoglutarate 3 mM, and ammonium iron sulfate 5 mM. The final reaction mixture contained 2 mM L-ascorbic acid, 1 mM 2-ketoglutarate, 0.3 mM ammonium iron sulfate, 100 mM KCl, 50 mM MES pH 6, 50 ng/µL BSA, 4 µM wild-type AlkB, and 4 µM AlkB-D135S. About 50 µL of the reaction mixture was added to 5–20 µL of decanted streptavidin bead slurry after ligation, immobilization, and washing. The reaction continued for 30 min at 37 °C. Following the reaction, beads were washed once with high salt wash buffer (1 M NaCl, 20 mM Tris-HCl pH 7.4) and once with low salt wash buffer (100 mM NaCl, 20 mM Tris-HCl pH 7.4).

2.4.9 HEK cell culture & RNA extraction

HEK293T cells (ATCC, CRL-11268) were cultured with complete DMEM medium under standard conditions according to ATCC. Briefly, HEK293T cells were grown in Hyclone DMEM medium (GE Healthcare Life Sciences, SH30022.01) with 10% FBS and 1% Pen–Strep (Penicillin–Streptomycin) to 80% confluency and passaged. Cells were collected and total RNA was extracted using TRIzol (Thermo Fisher, 15596026) by following the manufacturer's protocol when cells reached 80–90% confluency.

2.4.10 MCF7 cell culture & RNA extraction

MCF7 cells (ATCC, HTB-22) were cultured in EMEM medium (ATCC, 30-2003) with 10% FBS (Thermo Fisher, 10082147), 0.01 mg/ml bovine insulin (Sigma-Aldrich, I0516), and 10 nM β-

estradiol (SigmaAldrich, E2758) to 80% confluency and passaged at ratios of 1:3. Total RNA were extracted using TRIzol.

2.4.11 CMC treatment & library construction

MCF7 total RNA sequencing libraries were constructed as follows. Small RNA (<200 nt) was first removed from 1 μg MCF7 total RNA using spin columns (Zymo Research RNA Clean & Concentrator-5, R1016) and the large RNA (>200 nt) was eluted with 18 μl sterile H2O in a microcentrifuge tube. The RNA was transferred to PCR tubes and 2 μl Magnesium RNA fragmentation buffer (NEB, E6150S) was added to each tube and the tubes were incubated at 94 °C in a thermocycler for 5 min to fragment the RNA to ~200 nt. Two microliters of RNA fragmentation stop solution were then added to each tube. The samples were diluted to 50 μl with H2O and Zymo Research spin columns were used to purify the fragmented RNA; the RNA were eluted in 16 μl sterile H2O in a microcentrifuge tube. For 3′ end-repair of the RNA fragments, 2 μl 10x T4 PNK buffer and 2 μl T4 PNK at 10 U/μl (Thermo Fisher, EK0032) were added and the mixture was incubated at 37 °C for 30 min. The fragmented, end-repaired RNA was used to build sequencing libraries using the MSR-seq protocol described above with the following modifications. The fragmented RNA was ligated to barcoded capture hairpin oligonucleotides and bound to streptavidin beads. The samples were then pooled, mixed, and split into two parts for ±CMC (N-cyclohexyl-N′-(2-morpholinoethyl) carbodiimide) treatment (+CMC:-CMC = 1.5: 1 ratio). About 12 μl sterile H2O and 24 μl TEU buffer (50 mM Tris-HCl (pH 8.3), 4 mM EDTA, 7 M urea) were first added to each tube, then 4 μl freshly prepared 1 M CMC in TEU buffer was added to +CMC samples and 4 μl sterile H2O was added -CMC samples. The samples were incubated at 30 °C for 16 h at 1400 rpm on an Eppendorf ThermoMixer. The samples were washed

twice with high salt buffer and once with low salt buffer. The samples were then resuspended with 40 µl of 50 mM sodium carbonate and 2 mM EDTA (pH 10.4) buffer and incubated at 37 °C for 6 h at 1400 rpm. The beads were washed twice with high salt buffer and once with low salt buffer and then proceeded to the MSR-seq steps such as phosphatase treatment and reverse transcription.

2.4.12 Read processing & mapping

Libraries were sequenced on Illumina Hi-Seq or NEXT-seq platforms. First, paired-end reads were split by barcode sequence using Je demultiplex with options BPOS = BOTH BM = READ_1 LEN = 4:6 FORCE = true C = false [252]. BM and LEN options were adjusted for samples with a 3 nt barcode instead of 4, and for samples where the barcode is located in read 2. Next, only the read beginning with the barcode (usually read 2) was used to map with bowtie2 (version 2.3.3.1) with the following parameters: "-q -p 10 -local -no-unal". For human sample reads were mapped to the human transcriptome, with tRNA genes shaped for a curated, nonredundant, set of high-scoring tRNA genes. This reference was a combination of HG19 ORFs, ncRNAs, and our curated tRNA list based on HG19 tRNAs curated to be non-redundant, tRNA-scan SE with score >47, and 3′ "CCA" appended. Bowtie2 output sam files were converted to bam files, then sorted using samtools. Next IGV was used to collapse reads into 1 nt window. IGV output.wig files were reformatted using custom python scripts (available on GitHub). The bowtie2 output Sam files were also used as input for a custom python script using PySam, a python wrapper for SAMTools ([253], https://github.com/pysam-developers/pysam) to sum all reads that mapped to each gene. Related custom scripts were used to divide reads based on which 10 nt window the 3′ end mapped to for each tRNA; this is for fragment analysis. Data were visualized with custom R scripts. All custom scripts are available on GitHub (https://github.com/ckatanski/CHRIS-seq).

2.4.13 Read processing from CMC reaction

Raw 100 bp paired-end sequencing reads were obtained from the Illumina Hi-Seq platform. Read1 reads were separated by barcodes with the barcodes sequence on paired read2 reads using custom python scripts. Read2 reads were separated by barcodes using fastx_barcode_splitter (fastx_toolkit, http://hannonlab.cshl.edu/fastx_toolkit/). For read1 reads, the random six nucleotides' unique molecular identifier (UMI) sequence at the start of the reads and the barcoded adapter sequence at the end of the reads were removed using Trimmomatic [254] using single-end mode with a 15 nt cutoff. For read2 reads, the 7 nt barcode sequence at the start of the reads and the UMI and adapter sequence at the end of the reads were removed by Trimmomatic using paired-end mode with a 15 nt cutoff. The reads were then mapped to human rRNA transcripts using bowtie2. The output sam files were converted to bam files and then sorted and indexed using samtools. A command-line version of "igvtools count" (IGV, http://software.broadinstitute.org/software/igv/download) were used to count nucleotide composition, insertions, and deletions at single-base resolution. "Bedtools genomecov" (bedtools, https://bedtools.readthedocs.io/en/latest/) was used to count the start and end of all reads at each position. All the output files and reference sequence were combined into a single file for each sample, and the mutation rate and the stop rate were computed by custom python scripts. The output files were analyzed to identify target pseudouridine sites.

Chapter 3

tRNA Abundance, Modifications Fractions, & Fragments Provide Insight into Stress-Related Translation Regulation

Acknowledgment: This chapter is derived from an article published in Nature Communications by Nature Portfolio [1]. The authors of that article were: Christopher P. Watkins, Wen Zhang, Adam C. Wylder, Christopher D. Katanski, and Tao Pan. Author contributions: Conceptualization, C.P.W., C.D.K., and T.P.; Methodology: C.P.W., C.D.K., and T.P.; Software: C.D.K.; Formal Analysis, C.D.K.; Investigation, C.P.W., C.D.K., W.Z., and A.C.W.; Writing – Original Draft, C.P.W., C.D.K., and T.P.; Writing – Review & Editing, C.P.W., W.Z., A.C.W., C.D.K., and T.P.; Supervision, T.P.

3.1 Introduction

The development of any new method requires evaluation by comparison of the data it generates and subsequent interpretation to previously studied systems both to determine the accuracy of the new method and the possibility of producing novel insights. For biological application of MSR-seq, we investigated the stress response of tRNA abundance, charging, and modification, as well as other small RNA families upon heat shock, exposure to hydrogen peroxide, and to arsenite which are known to strongly affect translation [231-233].

Using MSR-seq measurements of tRNA in total RNA and in the polysome, we identify specific tRNA responses in both tRNA abundance and modifications. Together with mRNA measurements, the tRNA response is consistent with stress- and tRNA-dependent translational downregulation during translational elongation. We also find native tRNA molecules lacking

several modifications as altered reservoirs for tRNA fragment biogenesis. These insights demonstrate the power of MSR-seq to identify new biological phenomena and as a tool to investigate the mechanisms and functions of small RNAs and epitranscriptomic modifications.

3.2 Results

3.2.1. Stress induces coordinated tRNA abundance & modification changes in total RNA

We applied MSR-seq to investigate the stress response of tRNAs. To broaden the scope of our study, we subjected HEK293T cells to three commonly used, but different stress types: heat shock, hydrogen peroxide, and arsenite, plus the unstressed control. Using total RNA as input, we obtained on average 15 million reads mapped to the human genome among all samples (Figure 3.1(a)). As expected, most reads were from nuclear-encoded tRNA, with the remaining from 5S and 5.8S rRNA, mitochondrial-encoded tRNA, spliceosomal RNA (snRNA), mRNA, and other RNA families.

We analyzed the nuclear-encoded tRNA results by comparing each stress condition with unstressed control. We examined tRNA abundance, charging, and modification at the isoacceptor level wherein we pooled reads mapped to tRNA genes with the same anticodon. At the isoacceptor level, tRNA abundance changes were within 1.25-fold (Figure 3.1(b)), and tRNA charging level within 1.15-fold (Figure 3.1(c)), indicating that tRNA abundance and charging in total cellular tRNA did not change widely under these stress conditions.

We searched for tRNA isodecoder modification responses under stress and identified m^3C in serine tRNAs. Using primer extension, we validated that the mutation fraction of m^3C in MSR-seq could be used to quantify the modification fraction (Figure 3.1(d)). In human tRNAs, m^3C is present in the anticodon loop of $tRNA^{Arg}$(yCU)/$tRNA^{Ser}$/$tRNA^{Thr}$ (m^3C32 in tRNA nomenclature)

and in the loop of the variable arm stem-loop of $tRNA^{Leu}$(CAG)/$tRNA^{Ser}$ (m^3C47d or m^3Ce2 in tRNA nomenclature) [16,137]. The mutation fraction of m^3C47d in $tRNA^{Ser}$, but not in $tRNA^{Leu}$(CAG) showed a marked decrease under heat and arsenite stress, and an increase under hydrogen peroxide stress (Figure 3.1(e)).

Human $tRNA^{Ser}$ isodecoder genes have either C47d or T47d. Isodecoders with T47d cannot be methylated and no mutation signature was observed. During stress, the m^3C32 for the C47d-$tRNA^{Ser}$ isodecoders generally showed an increase in the mutation fraction, whereas the T47d isodecoders showed little change (Figure 3.1(e,f)). During heat and arsenite stress only C47d-$tRNA^{Ser}$ isodecoders showed an increase in abundance, but T47d-$tRNA^{Ser}$ isodecoders showed little change (Figure 3.1(g)). By contrast, there was no significant change in the charging levels of either sets of isodecoders under stress (Figure 3.1(h)). m^3C introduces a positive charge at physiological pH. In the three-dimensional structure of $tRNA^{Ser}$ on the ribosome, m^3C32 is close to the tRNA-mRNA base pairs in the 40S, and m^3C47d is at the 40S–60S interface [234] (Figure 3.1(i)). These results suggest that m^3C47d and m^3C32 modifications can respond synergistically under stress.

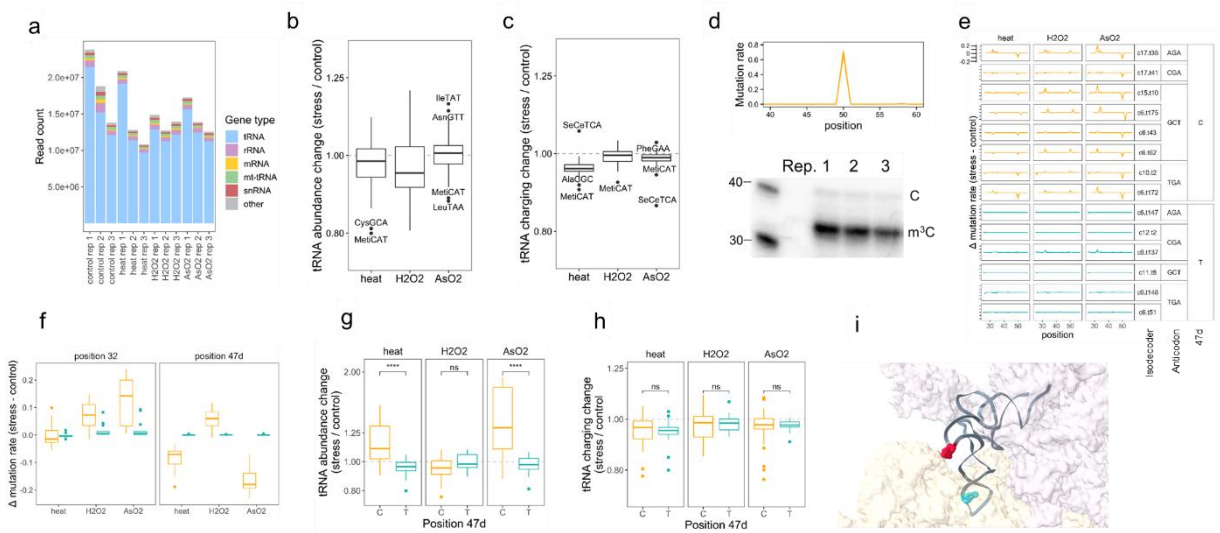

Figure 3.1: tRNA analysis in total RNA under stress. MSR-seq was performed in biological triplicates (n = 3) for each condition using HEK293T cells: unstressed, 42 °C/1 h (heat), 0.6 mM H_2O_2/2 h (H_2O_2), 0.3 mM $NaAsO_2$/2 h (AsO_2). Box and Whisker plots show median, 25th and 75th quartile, and whiskers to 1.5 times interquartile range. (a) Read coverage among RNA families for each sample. (b) Stress-induced tRNA abundance change among isoacceptor families. Abundance is calculated as summed normalized coverage of isodecoders in an isoacceptor family. The abundance change upon stress is calculated for each isoacceptor family, averaged across n = 3 independent biological replicates, displaying 46 isoacceptor families. (c) Stress-induced tRNA charging change among isoacceptor families. Charging is calculated as the ratio of the sum of charged isodecoder reads (CCA-ending) to the sum of uncharged isodecoder reads (CC-ending) for an isoacceptor family. The change in this ratio between stressed and unstressed samples is calculated for each isoacceptor family, then averaged across n = 3 independent biological replicates, displaying 46 isoacceptor families. (d) Top graph shows the mutation fraction at the 50th nucleotide position in tRNA^{Ser}(GCT) as a representative isodecoder; this position corresponds to 47d in tRNA nomenclature. The three independent biological replicates are shown. The bottom gel image shows validation of sequencing mutation fractions of m^{3}C47d in tRNA^{Ser}(GCT) by primer extension: the shorter product indicates an m^{3}C-induced RT stop, while the longer product indicates read through of hypomodified molecules. (e) The difference in mutation fraction between stressed and unstressed samples is plotted for tRNA^{Ser} isodecoders, highlighting differences in m^{3}C32 and m^{3}C47d mutation signature (n = 3 replicates shown). C47d-tRNA^{Ser} isodecoders are in orange, T47d-tRNA^{Ser} in blue. Isodecoder designations are according to the hg19 genomic tRNA database [255], e.g., c17.t35 corresponds to chromosome 17, tRNA 35. Change in mutation fraction for each replicate is calculated as mutation fraction minus the mean of control replicates. (f) Summary of data from panel (e): mutation fraction at sites corresponding to tRNA^{Ser} m^{3}C32 and m^{3}C47d change under stress. C47d-tRNA^{Ser} isodecoders are in orange and T47d-tRNA^{Ser} in blue (each point is a tRNA^{Ser} isodecoder). The mutation fraction at T47d is zero; these sites are not methylated. Data were shown for n = 3 independent biological replicates of 8 and 6 C or T tRNA^{Ser} isodecoders, respectively. (g) Abundance change for C47d-tRNA^{Ser} isodecoders in orange and T47d-tRNASer in blue. ****$p < 10^{-4}$, ns not significant; significance calculated with two-sided Wilcoxon test. Data were shown for n = 3 independent biological replicates of 8 and 6 C or

Figure 3.1, continued: T $tRNA^{Ser}$ isodecoders, respectively. P values are 1.2×10^{-5} and 1.3×10^{-5} for heat and arsenite stress. (h) Charging change for C47d-$tRNA^{Ser}$ isodecoders in orange and T47d-$tRNA^{Ser}$ in blue. ns not significant; significance calculated with two-sided Wilcoxon test. Data were shown for n = 3 independent biological replicates of 8 and 6 C or T $tRNA^{Ser}$ isodecoders, respectively. (i) Location of m^3C32 in cyan and m^3C47d in red in the 3D structure of ribosome-bound $tRNA^{Ser}$ (PDB 6Z6M from [234]). The 40 S subunit is in light yellow, 60S subunit is in light purple.

3.2.2. Stress induces coordinated tRNA abundance & modification changes in polysome

We performed an MSR-seq of RNA from polysome fractions to directly reveal tRNA involvement in translational regulation during stress. The polysome profile showed a significant decrease in global translation in all three stress conditions, with the largest decrease occurring under arsenite stress (Figure 3.2(a)). We confirmed the known increase of eIF2α phosphorylation level for arsenite stress (Figure 3.3(a)). We obtained on average 3.3 million mapped reads among all samples (Figure 3.2(b)). As expected for polysome samples, the rRNA and mRNA read portions were markedly increased, and the tRNA portion decreased in the polysome compared to total RNA mapping.

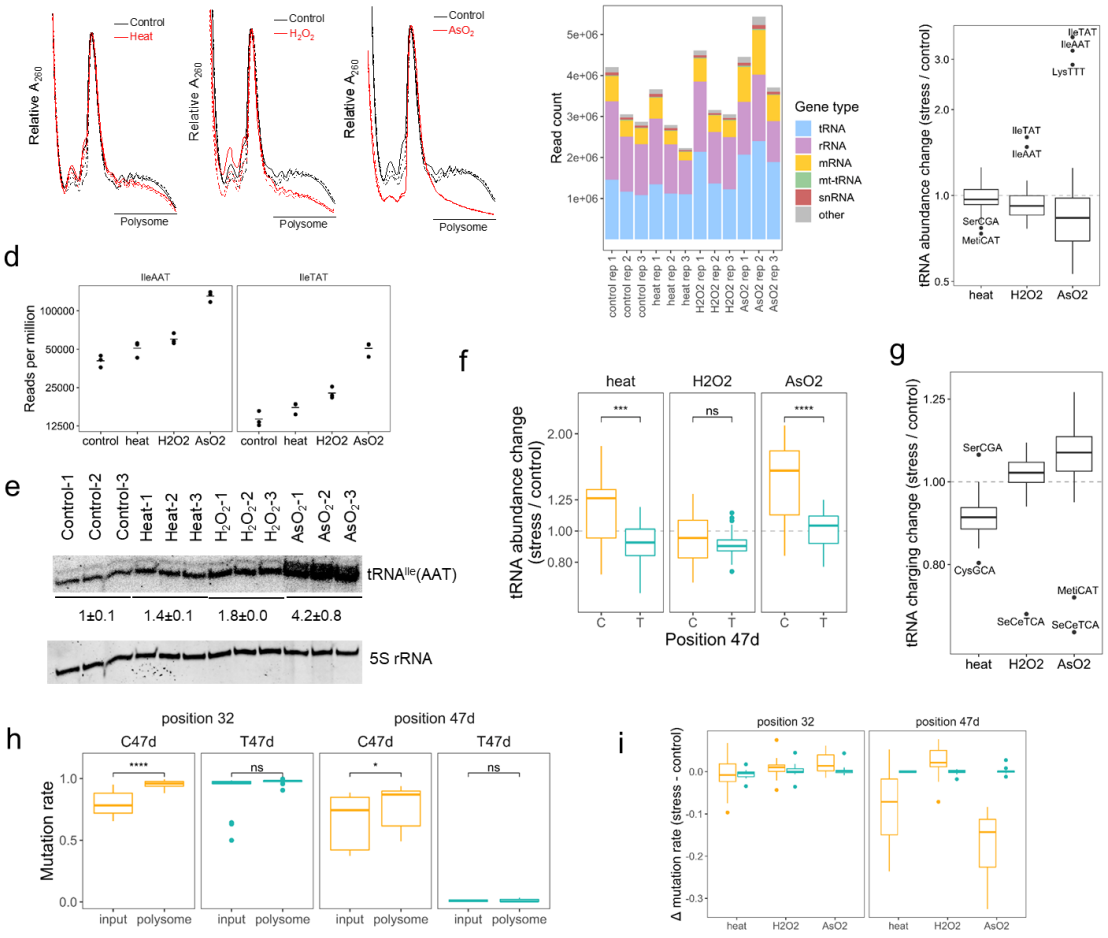

Figure 3.2: tRNA analysis of polysome under stress. MSR-seq was performed in biological triplicates (n = 3) for each condition. Box and Whisker plots show median, 25th and 75th quartile, and whiskers to 1.5 times interquartile range. (a) Polysome profile from sucrose density gradient. Unstressed control in black and stress in red. The three independent biological replicates are shown. The line "polysome" indicates the pooled fractions for MSR-seq. (b) Read coverage among RNA families for each sample. (c) Stress-induced changes in abundance of tRNA isoacceptor families on the polysome are calculated as in Figure 3.1(b), averaged across n = 3 independent biological replicates, displaying 46 isoacceptor families. (d) Normalized abundance of polysome-associated tRNA^{Ile}(AAT) and tRNA^{Ile}(TAT). (e) Validation of tRNA^{Ile}(AAT) change on polysome by Northern blot. 5 S rRNA is the loading control. All data for n = 3 independent biological replicates is shown. (f) Abundance change for C47d-tRNASer isodecoders in orange and T47d-tRNASer in blue. ****$p < 10^{-4}$, ***$p < 10^{-3}$, ns not significant; significance calculated with two-sided t-test. Data were shown for n = 3 independent biological replicates of 8 and 6 C or T tRNASer isodecoders, respectively. P values are 7.4×10^{-4} and 6.6×10^{-4} for C and T isodecoders. (g) Stress-induced change in tRNA charging on the polysome; charging is calculated as in Figure 3.1(c) averaged across n = 3 independent biological replicates, displaying 46 isoacceptor families. (h) Comparing mutation fraction of position 32 and 47d of tRNA^{Ser} isodecoders in the total RNA (input) and on polysome of the unstressed control. C47d-tRNA^{Ser} isodecoders in orange and T47d-

Figure 3.2, continued: $tRNA^{Ser}$ in blue. ****$p < 10^{-4}$, *$p < 0.05$, ns not significant; significance calculated with two-sided t-test. Data were shown for n = 3 independent biological replicates of 8 and 6 C or T tRNASer isodecoders, respectively. P values are 3.4×10^{-8} and 2.0×10^{-2} for positions 32 and 47d. (i) Stress-induced m^3C32 and m^3C47d change of tRNASer isodecoders. C47d-tRNASer isodecoders are in orange, T47d-$tRNA^{Ser}$ in blue. Data were shown for n = 3 independent biological replicates of 8 and 6 C or T tRNASer isodecoders, respectively.

The tRNA abundance change at the isoacceptor level on the polysome is the largest for the arsenite stress (Figure 3.2(c)). Specifically, $tRNA^{Ile}$(AAT), $tRNA^{Ile}$(TAT), and $tRNA^{Lys}$(TTT) showed a 3–4-fold increase (Figure 3.2(d), Figure 3.3(b)), while the bulk of other tRNAs showed a slight overall decrease. We validated the increase of $tRNA^{Ile}$(AAT) under stress by Northern blot (Figure 3.2(e)). In addition, C47d-$tRNA^{Ser}$ isodecoders showed an increase in the polysome compared to T47d-$tRNA^{Ser}$ in heat and arsenite stress (Figure 3.2(f)), similar to the $tRNA^{Ser}$ isodecoder patterns observed in the total tRNA (Figure 3.1(g)). The overall tRNA charging level on the polysome decreased slightly in heat and increased slightly in arsenite (Figure 3.2(g)). The charging levels for the bulk $tRNA^{Ser}$ isodecoders changed little (Figure 3.2(c)). The most striking result of polysome tRNA abundance and charging, therefore, was the dramatic increase of $tRNA^{Ile}$(AAT), $tRNA^{Ile}$(TAT), and $tRNA^{Lys}$(TTT) abundance in the arsenite stress. All three tRNAs read A/T-rich codons and may associate with ribosome stalling that contributes to the reduction of global translation (see below).

We examined how $tRNA^{Ser}$ m^3C modifications were associated with the polysome. In unstressed controls, polysome-associated $tRNA^{Ser}$ showed an increase in m^3C47d and m^3C32 levels compared to bulk tRNA (Figure 3.2(h), Figure 3.3(d)). In heat and arsenite stress, m^3C47d was markedly decreased on the polysome, whereas m^3C32 levels remained the same (Figure 3.2(i), Figure 3.3(e)). The $tRNA^{Ser}$ m^3C response to stress can be summarized in a model of a coordinated response of m^3C47d and m^3C32 on the polysome. In unstressed cells, m^3C47d and m^3C32 levels

are higher on the polysome than in the total tRNA, suggesting that m^3C modification enhances decoding in general. In the total RNA under arsenite and heat stress, the m^3C47d level decreases, possibly through the action of a cellular m^3C eraser [134,235]. The m^3C47d level also decreases on the polysome, possibly in response to the m^3C47d level decrease in total RNA. By contrast, the m^3C32 level increases in bulk tRNA, possibly through the action of a cellular m^3C32 writer [146]. There is no change in m3C32 levels on the polysome under stress, as m^3C32 in both C47d and T47d-$tRNA^{Ser}$ are already at nearly stoichiometric levels as indicated by their very high mutation fractions in the total RNA (Figure 3.2(h)).

3.2.3. Stress-induced change in translation efficiency is codon usage-dependent

To gain further insights into the stress response, we sequenced polyA-selected mRNA using MSR-seq for the same total RNA and polysome profiling samples (Figure 3.5(a)). As expected, many mRNA transcripts changed expression under stress (Figure 3.5(b)). For example, heat stress increased the level of the hsp1a1 transcript by ~10-fold in total RNA, and ~30-fold in polysome-associated RNA (Figure 3.4(a)); arsenite stress increased the eIF2α phosphorylation-dependent translation of ATF4 transcript by ~3.5-fold (Figure 3.4(b)). We found a decrease in translation efficiency among a set of well-detected mRNAs under each stress which was particularly pronounced in arsenite stress, both globally (Figure 3.4(c)) and for individual mRNA transcripts (Figure 3.5(c)). Gene ontology of the mRNAs with either highly increased or decreased translation efficiency showed the most affected genes belonging to metabolic processes (Figure 3.4(d)).

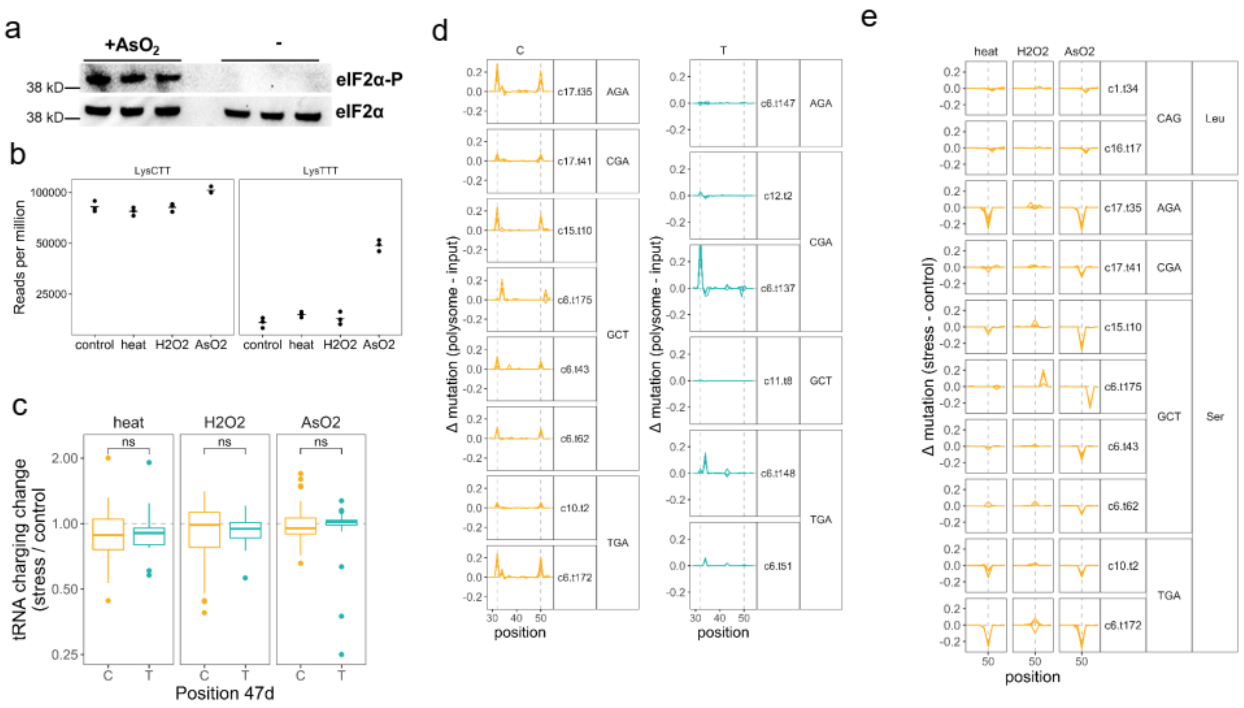

Figure 3.3: tRNA analysis without and with stress on the polysome. (a) HEK cells were stressed with Arsenite (+AsO_2) or unstressed and used for sequencing. Western blot against phosphorylated eIF2α confirms a stress response mounted in treated cells, but not control cells. Blotting against total eIF2α protein is used as a loading control. (b) Normalized abundance of $tRNA^{Lys}$ isoacceptor families in the polysome fraction is plotted for each stress condition. All data from n=3 biological replicates shown. (c) Fold-change in charging (aminoacylation level) for $tRNA^{Ser}$ isodecoders in the polysome faction is plotted. Isodecoders with C at position 47d are orange, while isodecoders with T at 47d are blue. Difference between C and T isodecoders is tested with Student's two-sided T-test; ns: not significant. Data are shown for n = 3 independent biological replicates of 8 and 6 C or T $tRNA^{Ser}$ isodecoders, respectively. Box and Whisker plots show median, 25th and 75th quartile, and whiskers to 1.5 times interquartile range. (d) Differences in mutation signature near position 32 and 47d between input and polysome fractions for control (unstressed) cells are plotted for $tRNA^{Ser}$ isodecoders. C47d-$tRNA^{Ser}$ isodecoders in orange, T47d-$tRNA^{Ser}$ in blue. Results are summarized in Figure 3.2(h). (e) Differences in mutation signature near position 47d for $tRNA^{Leu}$ and $tRNA^{Ser}$ isodecoders in polysome fractions between stressed and control cells are plotted. Traces for each of the 3 stress replicates are shown; difference is shown as subtracted mutation fraction for each stress sample from the mean mutation fraction of control replicates. Only isodecoders with C at position 47d are shown (orange). Results are partially summarized in Figure 3.2(i).

We analyzed codon use in an effort to understand how changes in polysome-associated tRNAs may affect translation, similar to those done by Begley, Dedon, and co-workers [29,236,237]. We limited our analysis to a set of mRNAs that are well detected in all stress

conditions and in both input and polysome fractions (~1700 genes). First, we observed that codon usage was related to transcript abundance, for example, abundant transcripts encode more lysine using both AAA and AAG codons (Figure 3.5(d)). Next, we compared genes with high and low translational efficiency (TE) during non-stress conditions. Genes with high TE contained significantly fewer AAA codons than genes with low TE, this pattern was absent for the synonymous codon AAG (Figure 3.4(e)). Expanding on this result, we asked how codon usage was related to stress-induced changes in TE. We observed that genes with decreased TE during arsenite stress contained significantly more AAA than genes with increased TE and found the opposite pattern for the synonymous lysine codon AAG (Figure 3.4(e, f)). We repeated this analysis for every codon in each stress to see how changes in TE were broadly related to codon usage. We found that during heat and arsenite stress, the A/T-ending codons were enriched in genes with decreasing TE, but the C/G-ending codons were enriched in the genes with increasing TE (Figure 3.4(f)).

We then examined the relationship between changes in codon preference and translational efficiency to changes in tRNA abundance. Only A/G-ending codons are analytically tractable for this analysis since T34 wobble tRNAs can have a preference for A-ending codons and C34 wobble tRNA reads only G-ending codons. Comparing the tRNA abundance change on the polysome under stress showed a preference of T34 over C34 tRNA in all cases (Figure 3.4(g)). The T34 over C34 tRNA preference was further confirmed by comparing the abundance change of every T34 over C34 tRNA of the same amino acid on the polysome, and the absence of this preference in total RNA under stress (Figure 3.4(h)).

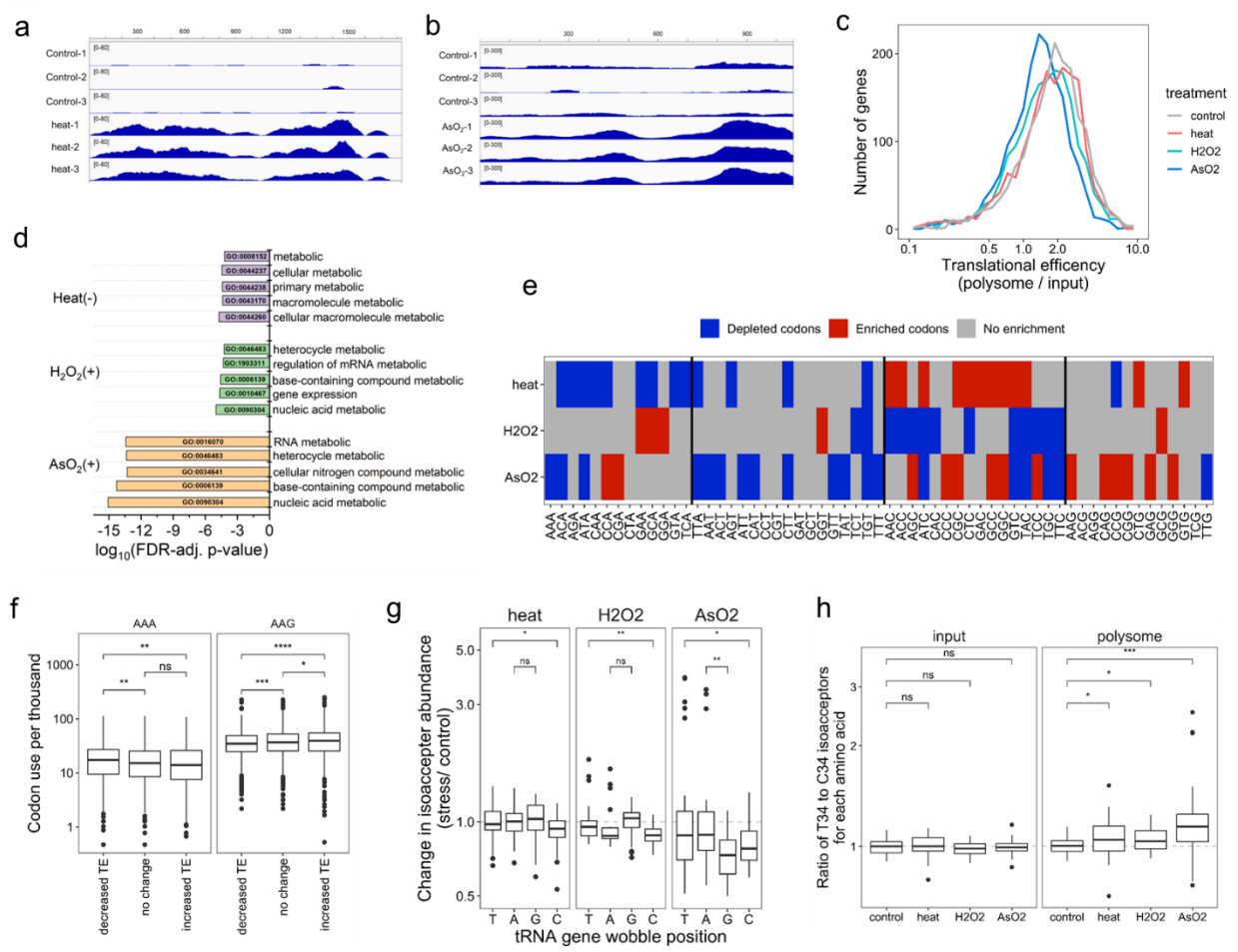

Figure 3.4: mRNA analysis of total RNA and on polysome under stress. mRNA-seq was performed in biological triplicates (n = 3) for each condition. Box and Whisker plots show median, 25th and 75th quartile, and whiskers to 1.5 times interquartile range. (a) IGV plot for the coding region of the HSP1A1 mRNA transcript, unstressed control (top), and heat stress (bottom) shows stress-increased read coverage along the length of the CDS. (b) IGV plot for the coding region of the ATF4 mRNA, unstressed control (top), and arsenite stress (bottom) showing stress-increased read coverage along the length of the CDS. (c) Histogram of translation efficiency (TE) among commonly detected mRNA genes in all conditions, in 30-gene increments. (d) Gene ontology analysis of biological processes among genes with stress-induced changes to TE. (+) Genes with highly increased translation efficiency, (−) Genes with highly decreased translation efficiency. (e) Lysine codon usage for AAA and AAG is compared between genes in the bottom, middle, and top tertile of ΔTE values during arsenite stress. ****p < 10^{-4}, ***p < 10^{-3}, **p < 0.01, *p < 0.05, ns not significant; significance calculated with two-sided Wilcoxon test. Data were shown for n = 3 independent biological replicates for n = 1747 genes, comparing tertiles. P values (top to bottom) for AAA are 2.6×10^{-3} and 9.5×10^{-3}, and for AAG 1.7×10^{-8}, 4.3×10^{-2}, and 3.5×10^{-4}. (f) Heat map showing significant differences in codon use between genes with increasing versus decreasing TE during stress. Significance is computed as Wilcox test p < 0.05 for comparing codon use between genes in the top and bottom tertile of ΔTE values for each stress. The heat map shows

Figure 3.4, continued: codons preferred by genes with increasing TE (red) or decreasing TE (blue). (g) tRNA abundance change on the polysome for isoacceptors with different wobble anticodon nucleotide in the tRNA gene. **p < 0.01, *p < 0.05, ns not significant; significance calculated with two-sided Wilcoxon test. Data were shown for n = 3 independent biological replicates for 42, 24, 24, and 48 isodecoders with wobble nucleotide T, A, G, or C, respectively. P values (top to bottom) for heat: 1.3×10^{-2} H_2O_2: 2.3×10^{-3} $NaAsO_2$: 1.5×10^{-2}, 5.9×10^{-3}. (h) tRNA abundance change of the ratios for T34 versus C34 wobble anticodon tRNA for each amino acid, total RNA input on the left, and on the polysome on the right. ***p < 10^{-3}, *p < 0.05, ns not significant; significance calculated with two-sided Wilcoxon test. Data were shown for n = 3 independent biological replicates for 11 amino acids. P values (top to bottom) for polysome are 7.4×10^{-4}, 2.4×10^{-2}, 3.7×10^{-2}.

The polysome mRNA and tRNA results can be explained by a tRNA-dependent downregulation of translational elongation during stress. Even though mRNAs on the ribosome are enriched for C/G-ending codons, tRNAs on the ribosome are enriched for those that read A/T-ending codons. This is consistent with increased ribosome stalling at A/T-ending codons corresponding to slowed elongation exacerbated by stress. Potential ribosome stalling may be particularly pronounced at the Ile-AAT, Ile-ATA, and Lys-AAA codons under arsenite stress which may explain the high level of on-polysome accumulation of $tRNA^{Ile}$ and $tRNA^{Lys}$(TTT) that read these codons.

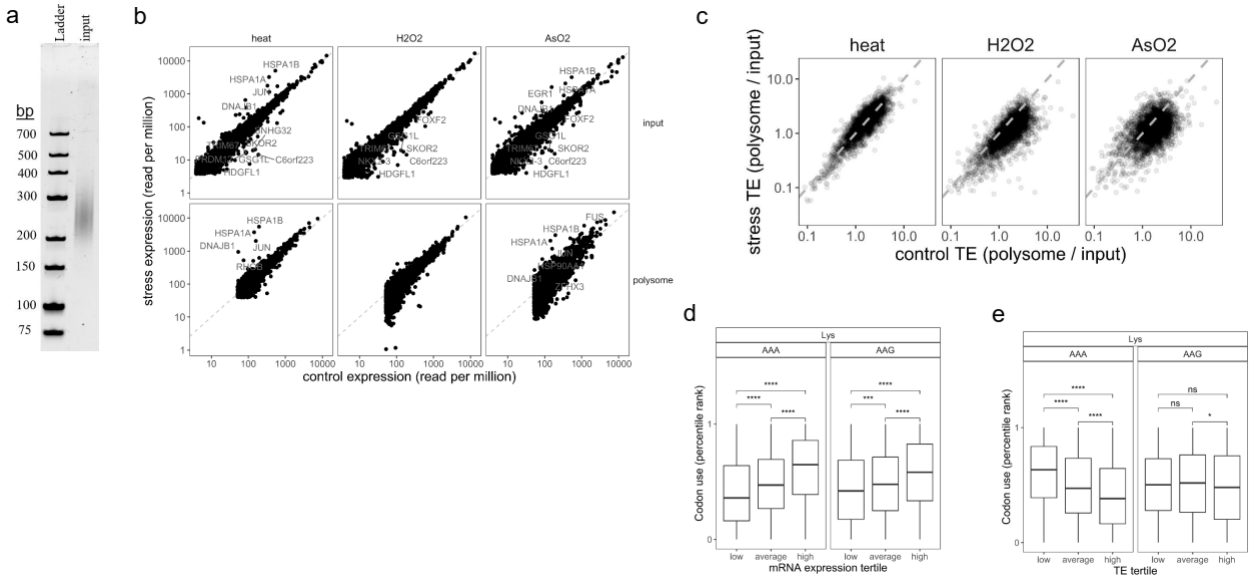

Figure 3.5: Additional analysis of mRNA libraries. (a) MSR-seq is compatible with input RNA of all kinds. As a demonstration, mRNA libraries were constructed from poly(A)-selected, then fragmented HEK RNA. Final PCR product of these libraries is shown. Representative gel from 1 of n=3 biological replicates shown. (b) Normalized abundance of mRNA is compared between stressed and unstressed cells for both total RNA and polysome fractions. Select genes with large changes in abundance are highlighted with names. Control (unstressed) abundance is plotted along the x-axis and stress abundance is plotted on the y-axis. Input samples are plotted as the top row of graphs; polysome fractions are the bottom row. (c) Translation efficiency (TE) is calculated as the ratio of normalized abundance in the polysome fraction to the input fraction. To visualize stress-induced changes in TE, the TE before stress for each gene is plotted along the x-axis and the TE after stress is plotted on the y-axis. Points below the diagonal indicate genes with reduced TE. (d) Codons are not used evenly across all genes. Well detected mRNA genes were divided into tertiles based on abundance in the input fraction from unstressed cells. The frequency of codon usage for lysine AAA and AAG for each gene is used to find a percentile rank among the set of analyzed genes. Percentile rank is represented in the tertile boxplots. Both AAA and AAG codons are used more frequently in abundance genes. Statistical comparisons are done with two-sided Wilcox's test. Data are shown for n = 3 biological replicates for n=1747 genes, comparing tertiles. Box and Whisker plots show median, 25th and 75th quartile, and whiskers to 1.5 times interquartile range. P values (top to bottom) for AAA: $<2.2x10^{-16}$, $6.3x10^{-11}$, $<2.2x10^{-16}$; AAG: $2.2x10^{-16}$, $8.1x10^{-4}$, $1.3x10^{-12}$. (e) Among unstressed cells, the TE of well detected genes is calculated. Genes are divided into tertiles based on TE (low, average, high). The codon usage for lysine AAA and AAG is calculated for each gene and used to assign a percentile rank among this set of genes. Differences in percentile rank distribution between tertiles for AAA indicates that efficiently translated genes use less AAA codon than inefficiently translated genes. This same pattern is not seen for AAG. Statistical comparisons are done with two-sided Wilcoxon test. Data are shown for n=3 independent biological replicates for n=1747 genes, comparing tertiles. Box and Whisker plots show median, 25th and 75th quartile, and whiskers to 1.5 times interquartile range for AAA: $<2.2x10^{-16}$, $<2.2x10^{-16}$, $1.1x10^{-10}$; AAG: $1.9x10^{-2}$.

3.2.4. tRNA modification affects tRNA fragment biogenesis

tRNA fragments (tRF) are a family of small RNAs that regulate many aspects of gene expression [5,238]. tRF sequencing commonly uses size-selected RNAs of 20–60 nucleotides. Although this approach obtains a high coverage of tRF and other small RNAs, potential direct connections between tRF and full-length tRNA are diminished by sequencing them separately. To simultaneously analyze tRNA and tRF in MSR-seq, we used a simplistic approach by binning reads by the 3′ ends of their mapped tRNA. The bins roughly correspond to fragments that terminate in the T stem-loop (50–60), variable loop and adjacent region (40–50), and anticodon stem-loop (30–40). Consistent with expectations, the amount of tRF mapped in this way was ~1% of the full-length tRNA (Figure 3.6(a)). The total amount of tRF did not change much under our stress conditions which were consistent with literature using specific stress conditions and cell lines. For example, arsenite stress showed a high level of tRF only at ≥ 500 μM [239,240] whereas our stress was at 300 μM. Most studies on arsenite stress were performed with HeLa cells, whereas we used HEK293T which generated lower levels of tRF at 500 μM arsenite [178]. We validated the tRF pattern of sequencing by Northern blot (Figure 3.6(b)).

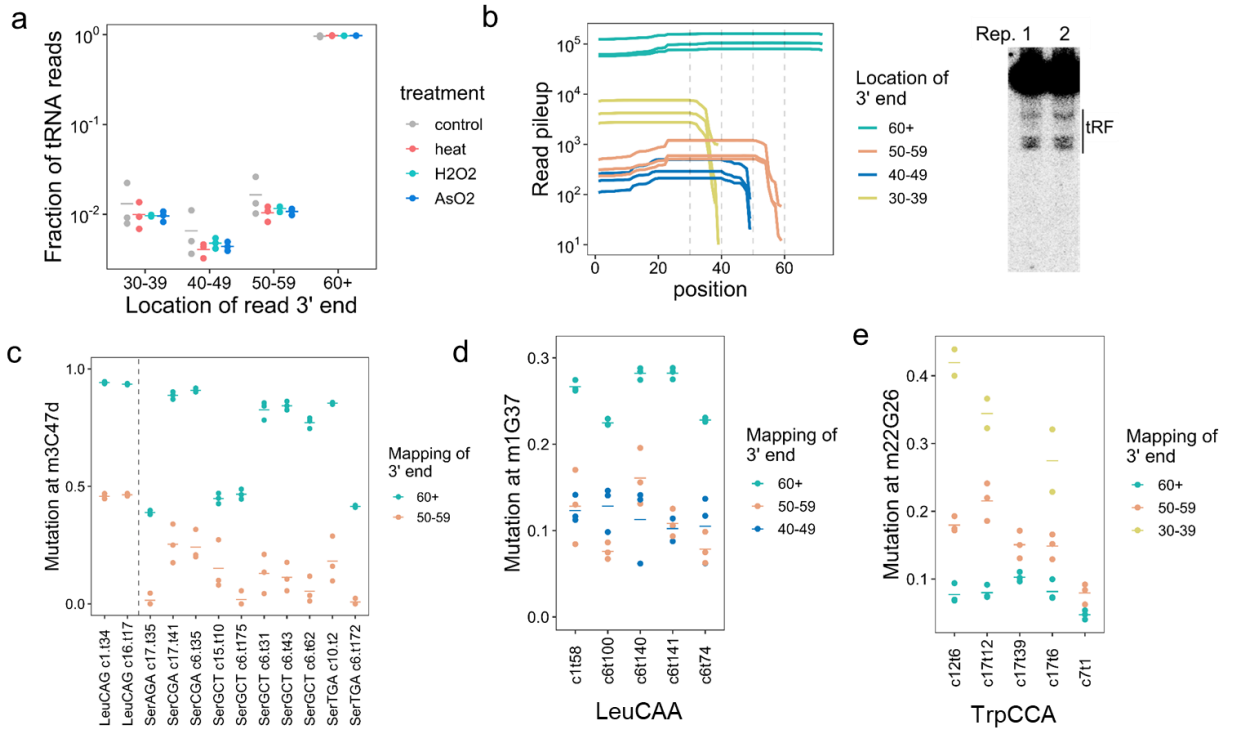

Figure 3.6: Analysis of tRNA fragment biogenesis and tRNA modification. MSR-seq was performed in biological triplicates (n = 3) for each condition. (a) Total count of tRNA fragment (tRF) relative to full-length tRNA in all four conditions. 60+ tRNA are considered full-length tRNAs; replicates are shown as points, mean of replicates is shown as a bar. (b) Read pileup of the most abundant tRF^{Gly}(CCC) isodecoder showing the amount of tRFs with 3′ ends in 30–39 (anticodon loop), 40–49 (variable loop), and 50–59 (T loop) in unstressed control samples. The gel shows Northern blot tRF^{Gly}(CCC) validation. Data for n = 2 independent biological replicates are shown. (c) Comparing mutation fraction of m3C47d in tRF versus full-length cognate tRNA for $tRNA^{Leu}$(CAG) and $tRNA^{Ser}$ isodecoders in unstressed controls. (d) Comparing mutation fraction of m^1G37 in tRF versus fulllength cognate tRNA for $tRNA^{Leu}$(CAA) isodecoders in unstressed controls. (e) Comparing mutation fraction of $m^2{}_2G26$ in tRF versus full-length cognate tRNA for $tRNA^{Trp}$ isodecoders in unstressed controls.

An important question in tRF biology is how specific tRNA modifications in the full-length tRNA affect tRF biogenesis. Reduction of modification levels by writer enzyme knock-down/knock-out has revealed that modifications such as m^5C, Q, and m^1G protect tRNA from cleavage [241]. However, it is unclear whether naturally occurring tRNA lacking a specific modification enhances or hinders tRF biogenesis. We addressed this question by simultaneously comparing the mutation fractions of the tRF and its cognate full-length tRNA of the same

sequence. The m^3C47d level was much lower in tRF compared to their cognate full-length tRNA (Figure 3.6(c)), consistent with the C47d-hypomodified tRNA being preferred substrates for tRF biogenesis. Similarly, the m^1G37 level was lower for the tRFs compared to their cognate full-length $tRNA^{Leu}$(CAA) (Figure 3.6(d)). In contrast, the $m^2{}_2G26$ level was higher for tRFs compared to their cognate full-length $tRNA^{Trp}$ (Figure 3.6(e)). These results indicate that tRNA modification can both stimulate and inhibit tRF biogenesis. Our results also show that tRNA modification affects cleavage in the distal regions of tRNA, similar to those observed previously for $m^5C48–50$ and m^1G9 modifications [242,243]. By directly comparing the modification levels of tRF and full-length tRNAs, MSR-seq seq provides another avenue to investigate the mechanisms of tRF biogenesis.

3.3 Discussion

Translational regulation in stress response has been extensively studied [244,245], and our work here reveals additional insights regarding tRNA abundance and modification. This was achieved by not only measuring the response of total tRNA, but also tRNA on the polysome under stress. Previous tRNA profiling from density gradients analyzed tRNAs in the ribo-seq 80S peak [246]. Our tRNA profiling includes the fractions of disomes and larger which measures tRNAs only in the elongating ribosomes. Most previous stress response studies deal with the regulation of global translation and translational initiation. By measuring tRNAs on elongating ribosomes, we were able to illuminate tRNA-based stress response in translational elongation.

For tRNA abundance, the most striking result is the enrichment of tRNAs that read A-ending over G-ending codons on the polysome under stress. At the same time, mRNAs from the same polysome samples show enrichment of C/G-ending over A/Tending codons. Among the 3

tRNAs that are enriched at the highest level on the polysome, $tRNA^{Ile}$(AAT) reads ATT, $tRNA^{Ile}$(TAT) reads ATA, and $tRNA^{Lys}$(TTT) reads AAA codons. These results are consistent with increased ribosome stalling at A/T-ending codons under stress; furthermore, ATT, ATA, and AAA codons may be hotspots of such stalling under arsenite stress. Stressinduced, codon-dependent tRNA response in translational elongation may also be driven in part by the post-translational modification of the eEF2 protein [247].

For tRNA modification, we focused on m^3C, which can be studied by MSR-seq at high sensitivity. m^3C32 is present in the anticodon loop of $tRNA^{Ser}$, $tRNA^{Thr}$, and $tRNA^{Arg}$(yCT) and enhances translation [146]. m^3C47d is in the variable loop of $tRNA^{Ser}$ and $tRNA^{Leu}$(CAG), which is located at the 40S–60S interface in the 80S ribosome. Our striking result is the coordinated response of m^3C32 and m^3C47d under stress. Without stress, tRNAs with higher m^3C32 and m^3C47d levels are loaded on the polysome. Heat and arsenite stresses reduce the m^3C47d level in total RNA, possibly through the action of a tRNA eraser such as ALKBH3 [134]; this results in a corresponding reduction of m3C47d level in polysome-associated $tRNA^{Ser}$. However, the same stresses also increase m^3C32 levels in $tRNA^{Ser}$ to reach nearly stoichiometric levels in total RNA and in polysome. These results are consistent with both m^3C modifications working together to fine-tune decoding under stress. At this time, however, the m3C47d installation enzyme is not known [248], so we were unable to thoroughly investigate the function of this modification on codon-dependent protein synthesis. Once the m^3C47d writer is identified, knock-down or knock-out of this enzyme separately or in combination with the m^3C32 writer could reveal how these modifications coordinate in translation regulation.

We also revealed an underappreciated relationship between tRNA modification and tRNA fragment biogenesis in native tRNA. tRFs are involved in many aspects of gene expression,

developmental biology, and human diseases. It is well known that tRNA modifications strongly influence tRF biogenesis, with a consensus that tRNA modification generally hinders tRF generation. We present two insights here through simultaneous fulllength tRNA and tRF analysis in the same data. First, certain tRFs have lower m^3C or m^1G levels compared to their cognate fulllength tRNA, indicating that naturally occurring hypomodified tRNAs can indeed be preferred reservoirs for tRF biogenesis. Second, tRFs can have higher $m^2{}_2G$ levels compared to their cognate tRNA, indicating that naturally occurring tRNA modifications can also enhance tRF biogenesis.

These results from our forays in translational regulation and tRNA fragment biogenesis demonstrate that MSR-seq has the potential to provide unprecedented insight into the biological roles of small RNAs. I anticipate that applying MSR-seq in conjunction with other omics methods will be transformative in investigations of the microbiome, virology, cancer pathology, and many other biological niches.

3.4 Methods & Materials

Note: Sequencing library construction, read processing, and mapping methods, as well as cell culture methods, are the same as in Chapter 2.

3.4.1 m^3C poisoned primer extension

About 100 ng of total RNA sample was added to a tube containing 20 pmol (2 μM final concentration) of RT primer and 2 μL of 5x annealing buffer (10 mM KCl and 150 mM Tris-HCl, pH 7.5); water was added for a total volume of 5 μL. Samples were heated to 93 °C for 2 min and then directly placed on ice. Five microliters of a Post-annealing mix, containing the following components, was added to each sample tube (10 μL final volume): 2x AMV Buffer; 2 U/μL AMV

RT [New England Biolabs]; 2x poisoned dHTP mix (2 mM dATP, 2 mM dCTP, and 4 mM ddTTP); and 2 μCi/μL α-^{32}P dGTP [Perkin Elmer]. The samples were incubated at 37 °C for 30 min; and then mixed with 10 μL of 2x urea loading dye (9 M urea, 2 mM EDTA, 0.2% xylene cyanol, 0.2 % bromophenol blue) and incubated at 93 °C for 2 min before loading on a 15% denaturing polyacrylamide gels.

The primer sequence for tRNA^{Ser}(GCT) was 5′-TGG CGA CGA GGA TGG GAT TCG AAC CCA CGC GT.

3.4.2 Stress treatments

HEK293T cells were cultured in DMEM medium (GE Healthcare Life Sciences, SH30022.01) with 10% FBS and 1% Pen–Strep (Penicillin–Streptomycin). Twenty-four 15-cm plates of HEK293T cells (5×10^6 cells each) were seeded three days before collection. On the day of polysome profiling, six plates of cells (two plates for one sample) were treated with different stress conditions: (1) unstressed control; (2) 42 °C heat shock, 1 h; (3) 0.6 mM H_2O_2, 2 h; (4) 300 μM $NaAsO_2$, 2 h. Polysome profiling was immediately performed after stress treatment.

3.4.3 Polysome profiling

Polysome profiling procedures were adapted from ref. 249. Briefly, cells were treated with 100 μg/ml cycloheximide (CHX) in DMEM for 7 min right after stress treatments. DMEM medium was removed. Cells were then collected using 10 ml ice-cold PBS with 100 μg/ml CHX and cell lifter. Cells were pelleted by centrifugation at 3000 RPM for 5 min. The cell pellet was washed twice with 5 ml ice-cold PBS with 100 μg/ml CHX. Cells from two plates were combined for one sample. Cells were resuspended in 1 ml ice-cold PBS and transferred to microcentrifuge tubes.

About 200 µl cell suspension from each sample was saved as input. Cells were pelleted by centrifugation at 3000 RPM for 5 min. For input cells samples, 500 µl TRIzol reagent was added to extract the total RNA. For polysome cells samples, the cell pellet was resuspended by 4 volumes of lysis buffer (20 mM HEPES, pH 7.6, 100 mM KCl, 5 mM MgCl2, 1% Triton X-100, 100 µg/ml CHX, freshly added 1× protease inhibitor (11873580001, Roche), 40 U/µl RNase inhibitor (AM2696, Thermo)). Cells were rotated and lysed at 4 °C cold room for 30 min. Lysed samples were centrifuged at 16,000×g for 15 min to collect the clear lysate (~600–700 µl). About 4 ul Turbo DNase was added to each lysate and the samples were incubated at room temperature for 15 min. The samples were centrifuged again at 16,000×g for 15 min to get clear lysate. Absorbance at 260 nm of each sample was measured. Samples were adjusted to the same absorbance using lysis buffer. 5–50% sucrose gradient (20 mM HEPES, pH 7.6, 100 mM KCl, 5 mM MgCl2, 100 µg/ml CHX, freshly added 1 × protease inhibitor, 40 U/µl RNase inhibitor) was prepared using a Biocomp gradient station. About 600 µl gradient buffer was removed from the top of the balanced sucrose gradient. About 600 µl lysate was loaded onto the top of the gradient slowly while gently rotating the tube. The samples were centrifuged at 28,000 RPM for 3 h at 4 °C using a Beckman SW28.1 rotor. After centrifugation, fractions were collected and measured using the Biocomp gradient station (30 fractions total). Fractions were flash-frozen and stored at –80 °C before RNA extraction and library construction. For polysome RNA extraction, fractions from disome and higher were combined and 2 volumes of TRIzol reagent were added to extract the RNA. PolyA+ RNA was extracted from the input and polysome RNA using a polyA+ RNA extraction kit from Promega (Z5310) or NEB (E7490S), respectively.

Poly(A)-selection was done with NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490S) according to the manufacturer's instructions.

One microgram of input polyA+ RNA samples and 100 ng polysome polyA+ RNA samples were used to build sequencing libraries. The RNA fragmentation and end-repair steps are the same as the CMC sequencing libraries construction. Briefly, polyA+ RNA in 18 µl were added to PCR tubes and 2 µl Magnesium RNA fragmentation buffer (NEB, E6150S) was added to each tube. The tubes were incubated at 94 °C in a thermocycler for 5 min to fragment the RNA to ~200 nt. The tubes were transferred to ice and a 2 µl RNA fragmentation stop solution was then added to each tube to stop the fragmentation. Samples were spun down and diluted to 50 µl using sterile H_2O. The fragmented RNA was purified using Zymo RNA clean and concentrator columns and eluted in 16 µl sterile H_2O. Two microliters of T4 PNK buffer and 2 µl T4 PNK were added to the tubes and the tubes were incubated at 37 °C for 30 min to repair the ends of the fragmented RNA. MSR-seq method was then used to build sequencing libraries with the fragmented RNA. Slight modifications were made to the MSR-Seq protocol. After the first ligation step, the ligation reaction was quenched by adding 50 mM EDTA. Samples were combined and ligation products over 200 nt long was purified using Zymo RNA clean and concentrator columns twice. Target cDNA products were purified using AMPure XP beads with a 1:1 ratio after PCR.

3.4.4 mRNA transcriptome mapping

Raw 100 bp paired-end sequencing reads were obtained from the Illumina Nova-Seq platform. Reads processing and trimming was the same as CMC sequencing libraries. The reads were mapped to the human transcriptome (hg38) obtained from Ensembl. The mapped "bam" files were then analyzed using bamCoverage tool of the "deeptools" [250] (https://deeptools.readthedocs.io/en/develop/) with bin size as 1 to get the bigwig sequencing depth

coverage files. The bigwig coverage files were visualized using IGV and coverage track images were obtained.

3.4.5 Northern blots

The northern blot method was adapted from ref. [201] using 32^{p} radiolabeled probes. About 500 ng of each polysome RNA sample were diluted to 9 µl in microcentrifuge tubes. About 1 µl 1 M Tris-HCl, pH 9 was added to each tube and mixed well. The samples were incubated at 37 °C for 30 min to deacylate the tRNAs. About 10 µl 2× RNA loading buffer ((9 M Urea, 100 mM EDTA, pH 8, 0.2% bromophenol blue, 0.2% Xylene cyanol)) were added. All RNA samples were loaded onto a 10% pre-run denaturing PAGE gel. The gel was stopped when the xylene cyanol band passed the middle of the gel. RNA was transferred to HybondXL Membrane (RPN303S, GE Healthcare) at 80 °C for 4 h using a gel dryer (BioRad). The membrane was soaked in deionized water with the membrane side on the top to separate the gel and the membrane. The gel was stained with SYBR gold (S11494, Thermo) and scanned using a Chemi-Doc imaging system (Bio-Rad). The membrane was UV-crosslinked twice (254 nm for 1200 mJ). The membrane was then prehybridized for 30 min twice with hybridization buffer (20 mM phosphate, pH 7, 300 mM NaCl, 1% SDS). About 40 pmol of the tRNA probes were radiolabeled by T4 PNK with γ-32P-ATP in a 10 µl reaction. The labeling mixture were diluted to 50 µl and cleaned by Illustra MicroSpin G-25 Columns (27532501, Cytiva). The membrane was incubated with 15 µl 32P radiolabeled probes for 16 h at 60 °C in the UVP Hybridizer Oven (95-0030-01, Analytik Jena). The membrane was washed twice using 50 ml washing buffer (20 mM phosphate, pH 7, 300 mM NaCl, 2 mM EDTA, and 0.1% SDS) for 30 min each. The membrane was wrapped in plastic wrap and exposed to a phosphorimager screen for 1–2 days depending on the signal strength. The screen was then

scanned using a personal molecular imager (Bio-Rad). The image was analyzed using ImageLab software.

Northern blot probes were from Integrated DNA Technologies (IDT) and gel purified. Sequences of the probes were (Y = C/T; R = A/G; W = A/T, M = A/C):

Gly: 5′-TGC ATT GGC CRG GAA TYG AAC CCG GGY CTC CCR CGT GGW AGG CGA GAA TTC TAC CAC TGM ACC ACC MAY GC-3′

Ile-AAT: 5′-TGG CCM GTA CGG GGA TCG AAC CCG CGA CCT TGG CGT TAT TAG CAC CAC GCT CTA ACC AAC TGA GCT AAC CRG CC-3′

3.4.6 Gene ontology analysis

Gene ontology analysis was performed using the default setting in ref. [251] (http://geneontology.org/).

3.4.7 Western blotting of eIF2α phosphorylation

All samples were incubated at 95 °C for 10 min, separated on a 4–12% polyacrylamide Bis-Tris protein gel (NP0322BOX, Thermo), and transferred to polyvinylidene fluoride membranes (IPVH00010, Millipore). The membranes were blocked in 10% w/v milk (1706404, Bio-Rad). The blots were probed with 1/1000 v/v EIF2S1 antibody (AHO0802, Invitrogen) or 1/500 v/v Phospho-EIF2S1 (Ser51) antibody (MA5-15133, Invitrogen), followed by 1/10000 v/v sheep anti-mouse IgG (NA931V, Cytiva) or 1/10000 v/v donkey anti-rabbit IgG conjugated to horseradish peroxidase (NA934V, Cytiva). The blots were visualized with ECL Prime Western Blotting Detection Reagents (RPN2232, Amersham) using a Bio-Rad Chemi-Doc MP.

3.4.8 Translational efficiency & mRNA codon usage analysis

For each gene, read counts mapping to all transcript variants were summed together. Read counts were normalized for total detection in each sample (i.e., reads per million) among mRNA-detection reads (only genes with "gene_biotype" as "protein_coding" were included for normalization). After normalization, genes were filtered to have more than 100 counts. Next translational efficiency (TE) was calculated as the normalized gene counts in the polysome fraction divided by the input fraction for each replicate and each stress. Next, analysis was limited to a set of well-detected genes: this set of genes was defined as genes where TE could be calculated in all four treatments (control, heat, H_2O_2, and AsO_2) for one replicate. This gave a set of ~1500 genes, with small fluctuations in detection in other replicates. TE calculations were confirmed to be roughly log-normally distributed. A Z-score was calculated for each sample based on log10 (TE) value. The mean Z-score from the control replicates was used as a reference for stress-dependent change.

For each CDS transcript in our HG19-derived reference genome, the occurrence of each codon was tallied—frequency was calculated as the number of codon instances divided by protein length. For genes with several transcripts, the median value for each codon was used. The calculation was done with a custom python script, available on GitHub. These data were combined with our TE calculations.

Next, genes were divided into three groups: low TE, average TE, and high TE. Divisions were based on the 33rd percentile and 66th percentile rank for TE in each sample. For each codon, the frequency was used to assign a percentile rank to each gene in each sample (e.g., a gene with abundant "AAA" Lys usage may be in the 99^{th} percentile for "AAA" and the 5th percentile for "CCT" Pro). Percentile ranks were used to calculate statistical differences between gene groups

for each sample and each codon; tests were calculated via the two-sided Wilcox test. Codon usage frequency is not normally distributed, so a non-parametric hypothesis test is appropriate.

Next, for each sample, genes were again divided into three groups, but based on the percentile of change in Z-score from the mean of control replicates: decreased TE, no change in TE, and increased TE. Percentile group thresholds were again 33rd and 66th. Percentile ranking in codon usage was again used for hypothesis testing between groups for each sample and each codon with a two-sided Wilcox test as above.

Chapter 4

MSR-seq Provides Insights into Diverse Biological Fields

Acknowledgment: This chapter is derived from two articles, the first published in Frontiers in Cell and Developmental Biology by Frontiers Media [2] and the second published in Nucleic Acids Research by Oxford Academic [3], and unpublished data pursued during the course developing of MSR-seq. The authors of that Frontiers article were: Noah Peña, Wen Zhang, Christopher P. Watkins, Mateusz Hałucha, Hala Alshammary, Matthew M. Hernandez, Wen-Chun Liu, Randy A. Albrecht, Adolfo Garcia-Sastre, Viviana Simon, Christopher D. Katanski, and Tao Pan. Author contributions: Conceptualization, N.P., C.P.W., V.S., C.D.K., and T.P.; Methodology: W.Z., C.P.W., M.H., H.A., M.M.H., W-C. L., R.A.A., and A. G-S.; Software: N.P. and C.D.K.; Formal Analysis, N.P.; Investigation, N.P., W.Z., C.P.W.; Writing – Original Draft, N.P., W.Z., C.P.W., C.D.K., V.S., and T.P.; Writing – Review & Editing, N.P., W.Z., C.P.W., V.S., C.D.K., and T.P.; Supervision, V.S. and T.P.

The authors of that Nucleic Acids Research were: Christopher D. Katanski, Christopher P. Watkins, Wen Zhang, Matthew Reyer, Samuel E. Miller, and Tao Pan. Author contributions: Conceptualization, C.D.K., C.P.W., and T.P.; Methodology: C.P.W., W.Z., and M.R.; Software, C.D.K.; Formal Analysis, C.D.K. and S.E.M.; Investigation, C.D.K., C.P.W., W.Z., S.E.M.; Writing – Original Draft, C.D.K., C.P.W., T.P.; Writing – Review & Editing, C.D.K., C.P.W., W.Z., M.R., S.E.M., and T.P.; Supervision, T.P.

The contributors of the unpublished data were: Christopher P. Watkins, Samuel E. Miller, Karen Lolans, Jessica N. Pan, A. Murat Eren, Christopher D. Katanski, and Tao Pan. Contributions: Conceptualization, C.P.W., A.M.E., C.D.K., and T.P.; Methodology, C.P.W. and

K.L.; Software, S.E.M., A.M.E., and C.D.K.; Formal Analysis, S.E.M., J.N.P., and C.D.K.; Investigation, C.P.W., C.D.K., and T.P.

4.1 Introduction

In Chapters 2 and 3, I described both the development and application of MSR-seq to study stress-related translational regulation by use of tRNA isodecoders and tRNA modifications, selective usage of anticodon-codon pairs, and the effects of tRNA modifications on tRF biogenesis. These experiments were performed with cultured mammalian cells, and thus represent a taste of what can be identified using readily available lab-grown samples. However, one of the key goals in the development of MSR-seq was that it should be robust enough to prepare libraries from clinical and environmental samples with similar ease. Further, we desired to expand the modifications that could be detected and quantified by MSR-seq. In this Chapter, I detail further studies that I performed to investigate small RNAs in clinically relevant samples and detail a simple, but potent chemical reaction that I applied to detect many anticodon loop modifications that are normally invisible to or inhibitory of reverse transcriptases. These results further demonstrate the potential of MSR-seq as a broadly applicable and complementary tool to analyze the function(s) of the various RNA classes that comprise the transcriptome across all three kingdoms of life, as well as the biological function and regulation of many post-transcriptional modifications.

4.2 Results

4.2.1 SARS-CoV-2 selectively package host small RNAs with specific epitranscriptomic modifications

We performed Illumina sequencing starting with total RNA extracted from VeroE6 cells and from SARS-CoV-2 viral preparations cultured on VeroE6 cells (Figure 4.1(a)). To improve efficiency and quantitative assessment of small RNA-seq, defined here as RNA of <200 nucleotides in length, we built two libraries for each sample. The first library was treated with a demethylase mixture (DM) [206] which removed many Watson-Crick face methylations in tRNA that impede reverse transcription in library construction while the second library was left untreated. As described previously, the DM-treated libraries are useful for quantitative assessments of transcript abundance, whereas the untreated samples are useful for modification analysis [137]. As expected, sequencing reads of the VeroE6 cells mostly mapped to tRNAs, followed by those from 5S to 5.8S rRNA, a small amount of SRP RNA, and others such as spliceosomal RNA (snRNA) and Y RNA (Figure 4.1(b)). In the viral preparations, a substantial proportion of reads mapped to viral genomic RNA as expected, and tRNA and SRP RNA are present at almost equally high proportions, followed by a small amount of rRNA (Figure 4.1(b)). Although one cannot exclude RNA in exosomes or extracellular RNA not associated with vesicles, our data clearly show RNAs that are differently present in our cell free preparations as compared to those in cells. Our viral preparations used for the sequencing experiments contain high levels of infectious particles outside the cell, strongly suggesting that the sequenced RNA are derived, to a large extent, from cell free virions. For example, our results show a ~150-fold enrichment of the SRP RNA over tRNA in the viral preparation samples vs. the cell samples, which suggests that we eliminated most if not all of the

cellular debris. These results indicate that SARS-CoV-2 virions also package tRNA and SRP RNA in significant proportions.

Viral isolate #	Lineage	Clinical presentation & COVID-19 outcome
Viral Isolate 1	B.1.3	90 years old Male; severe COVID-19 with fatal outcome
Viral Isolate 2	B.1	63 years old Male; severe COVID-19 with fatal outcome
Viral Isolate 3	A.2	Patient has respiratory tract infection. History of travel to Europe
Viral Isolate 4	B.1	88 years old Female; severe COVID-19 with fatal outcome
Viral Isolate 5	B.4	39 years old Female; history of travel to Iran
Viral Isolate 6	B.1.3	62 years old Male; severe COVID-19 with fatal outcome

Table 4.1: Viral isolates used in this study.

We next examined the selectivity of packaged tRNA at two levels. Mammalian genomes contain many tRNA isodecoders genes that share the anticodon but possess different body sequences; all tRNA isodecoders with the same anticodon belong to a single tRNA isoacceptor family [5,15]. Abundance of isodecoders was summed for each isoacceptor family and used to calculate the fraction of tRNA reads for each anticodon. The isoacceptor abundance fraction for each sample (three biological replicates of uninfected VeroE6 cells and six distinct viral culture supernatant preparations), was compared to the mean of Vero E6 cells. All three VeroE6 cell replicates were nearly identical, as the heat map shows close to zero values in all cases (Figure 4.1(c)). To avoid exaggerated representation of low abundant tRNAs by ratioed comparison, we subtracted the tRNA fraction in each viral preparation to its counterpart in the VeroE6 cells, so that the differences were readily identified for more abundant tRNAs (Figure 4.1(c)). We found several isoacceptor families that are significantly enriched across all six isolates. They include $tRNA^{Glu}$(TTC), $tRNA^{Lys}$(TTT), $tRNA^{Leu}$(AAG), $tRNA^{Ser}$(AGA), $tRNA^{Ser}$(GCT), and $tRNA^{Ser}$(TGA). These results indicate that SARS-CoV2 virions selectively incorporate tRNA isoacceptors.

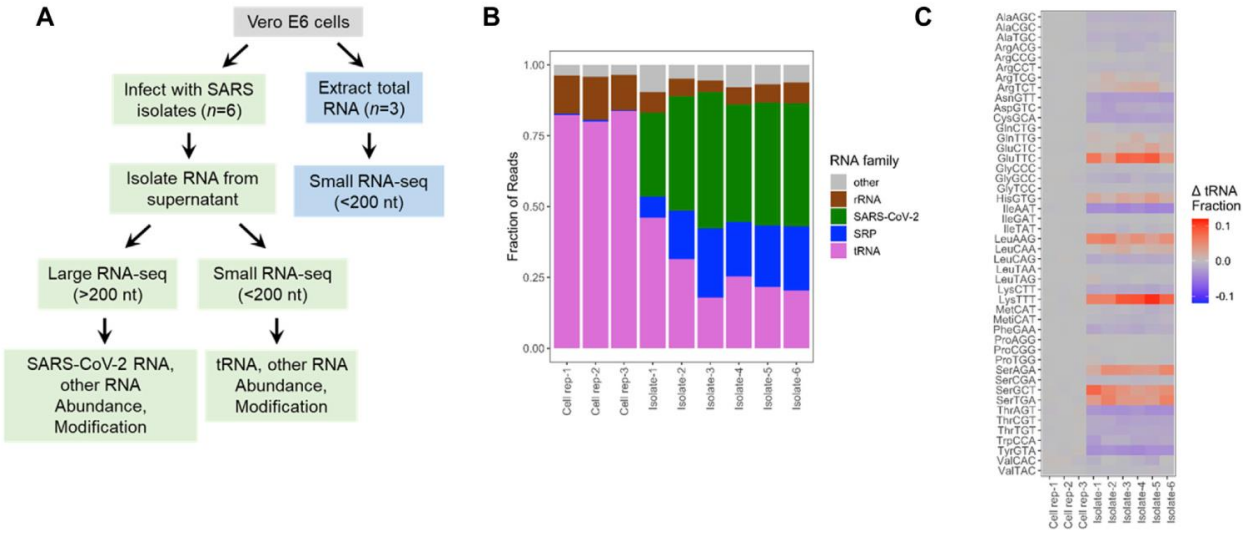

Figure 4.1: Selective enrichment of small RNAs in SARS-CoV-2 viral preparations. (a) Experimental scheme. Vero E6 cells were either infected with SARS-CoV-2 virus isolates from infected individuals (n = 6 biological isolates) or uninfected cultures (n = 3 biological replicates). Total RNA was extracted from the cells (blue boxes) or only from the cell free viral preparations (green boxes). Small RNA-seq was carried out using total RNA with and without demethylase treatment. Large RNA-seq was carried out with the RNA fraction after the removal of small RNAs of <200 nt, and chemical fragmentation. (b) Small RNA-seq results. Vero cell data are mostly tRNA and 5S/5.8S rRNA. Aside from SARS-CoV-2 RNA, virions contain significant portions of tRNA, rRNA, and signal recognition particle (SRP) RNA. (c) Enrichment and depletion of specific tRNAs in the cell free viral samples. Shown are the combined reads from all tRNA isodecoders that share the same anticodon. Heatmap shows the abundance of tRNAs for each anticodon subtracted from the mean of control cultures. Subtraction emphasizes the differences among abundant tRNAs. Enriched tRNAs are in red, depleted tRNAs in blue. Top 3 enriched tRNAs are $tRNA^{Lys}$(TTT), $tRNA^{Glu}$(TTC), and $tRNA^{Ser}$(GCT). Top 3 depleted tRNAs are $tRNA^{Ile}$(AAT), $tRNA^{Tyr}$(GTA), and $tRNA^{Asn}$(GTT).

Our downstream analysis, thus, focused on those six tRNAs enriched in the viral preparations. First, we analyzed the tRNA at the isodecoder level for all six tRNAs. Among the seven $tRNA^{Glu}$(TTC) isodecoders, four could be detected in the viral preparations. However, only two isodecoders represent almost all $tRNA^{Glu}$(TTC) in the viral preparations, even though neither is the most abundant isodecoder in VeroE6 cells (Figure 4.2(a)). In contrast, the single dominant $tRNA^{Leu}$(AAG), $tRNA^{Lys}$(TTT), $tRNA^{Ser}$(AGA) isodecoders in cells are also the ones in the viral preparations (Figures 4.2(b–d)). For $tRNA^{Ser}$(GCT) and $tRNA^{Ser}$(TGA), two isodecoders each are

present at appreciable levels, in each case, the isodecoder at the highest level is also the one in the viral preparations (Figures 4.2(e,f)).

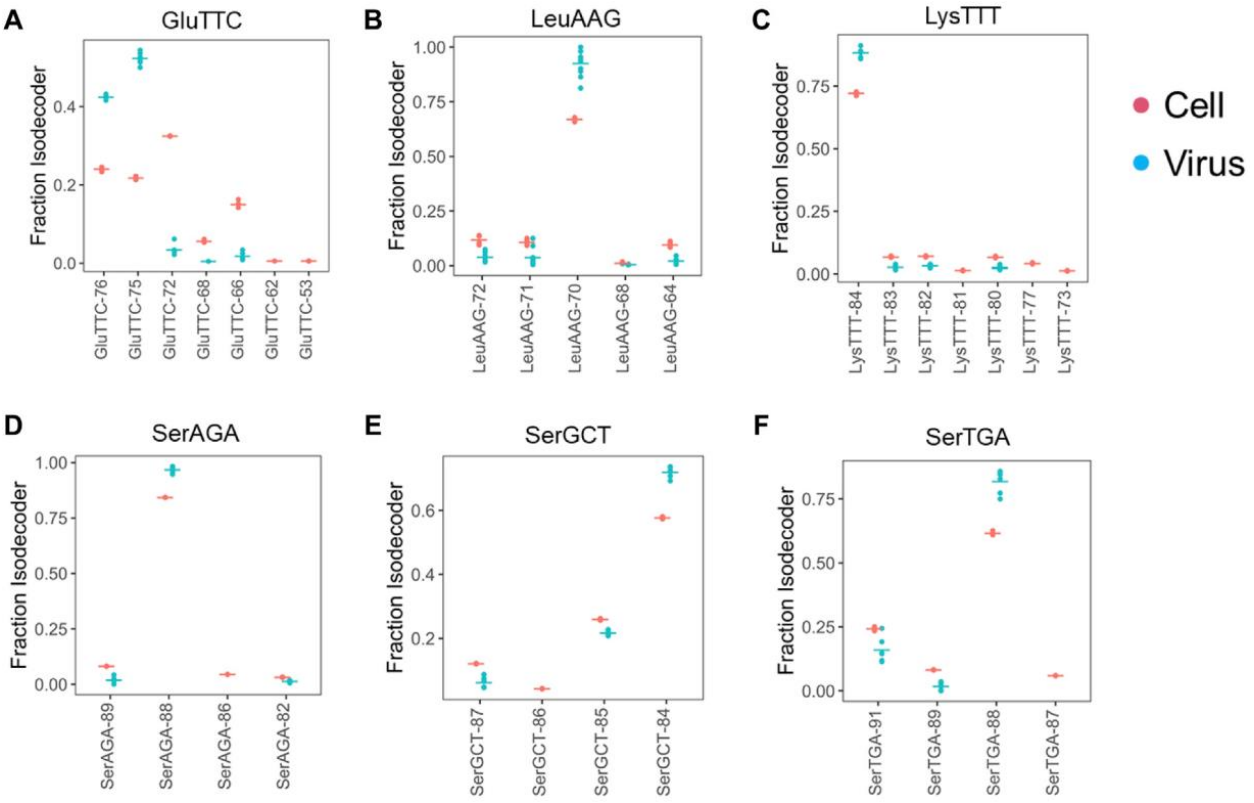

Figure 4.2: Selective enrichment of tRNA isodecoders in SARS-CoV-2 viral preparations. tRNA isodecoder fractions from uninfected Vero cell (n = 3, red) or cell free viral preparations (n = 6, blue) are shown. Mean values are shown as a horizontal bar. Isodecoder nomenclature is according to the tRNAScan score of the *Chlorocebus sabaeus* tRNA genes identified in Rfam database. (a) $tRNA^{Glu}$(TTC). (b) $tRNA^{Leu}$(AAG). (c) $tRNA^{Lys}$(TTT). (d) $tRNA^{Ser}$(AGA). (e) $tRNA^{Ser}$(GCT). (f) $tRNA^{Ser}$(TGA).

We examined the read pileup of the most abundant isodecoder in the viral preparations. By experimental design, our tRNA-seq results always start from the 3′ end of the tRNA and show a decline toward the 5′ end with sharp drops at certain tRNA modifications, an expected behavior for full-length tRNAs [206]. Three types of results are observed: first, the pileup decreases faster in the viral preparation tRNA compared to the cellular tRNA, this group includes $tRNA^{Glu}$(TTC) (Figure 4.3(a)). The pronounced drop of the viral preparation tRNA in the anticodon loop region

is consistent with GluTTT-75 in the virion being a 3′ half tRNA fragment with the 5′ end in the anticodon loop, because there is no known RT stopping modifications in this tRNA. In the second type, the read pileup decreases at about the same rate, this group includes $tRNA^{Leu}$(AAG) and $tRNA^{Lys}$(TTT) (Figures 4.3(b,c)). The similar drop off is consistent with the tRNA in the viral preparations as the full-length tRNA like those in cells, and the sharp drop offs corresponds to the N2,2-dimethyl-G at position 26 ($m^2{}_2G26$) in $tRNA^{Leu}$(AAG) which is difficult to remove by the demethylase because it is buried in the tRNA structure [256] and 2-methylthio-6-carbamoylthreonine at position 37 (ms2t6A37) intRNA^{Lys}(TTT) [16] which does not react with the demethylase. In the third type, the read pileup decreases slower in the viral preparation tRNA, this group includes all three $tRNA^{Ser}$, and the sharp drop offs correspond to the $m^2{}_2G26$ modification (Figure 4.3(d–f)). This result is consistent with the $tRNA^{Ser}$ in the viral preparations having lower modification levels in the anticodon stem-loop region which can include N^6-methyl-N^6-threonylcarbamoyladenosine at position 37 (m^6t^6A37) in $tRNA^{Ser}$ [16].

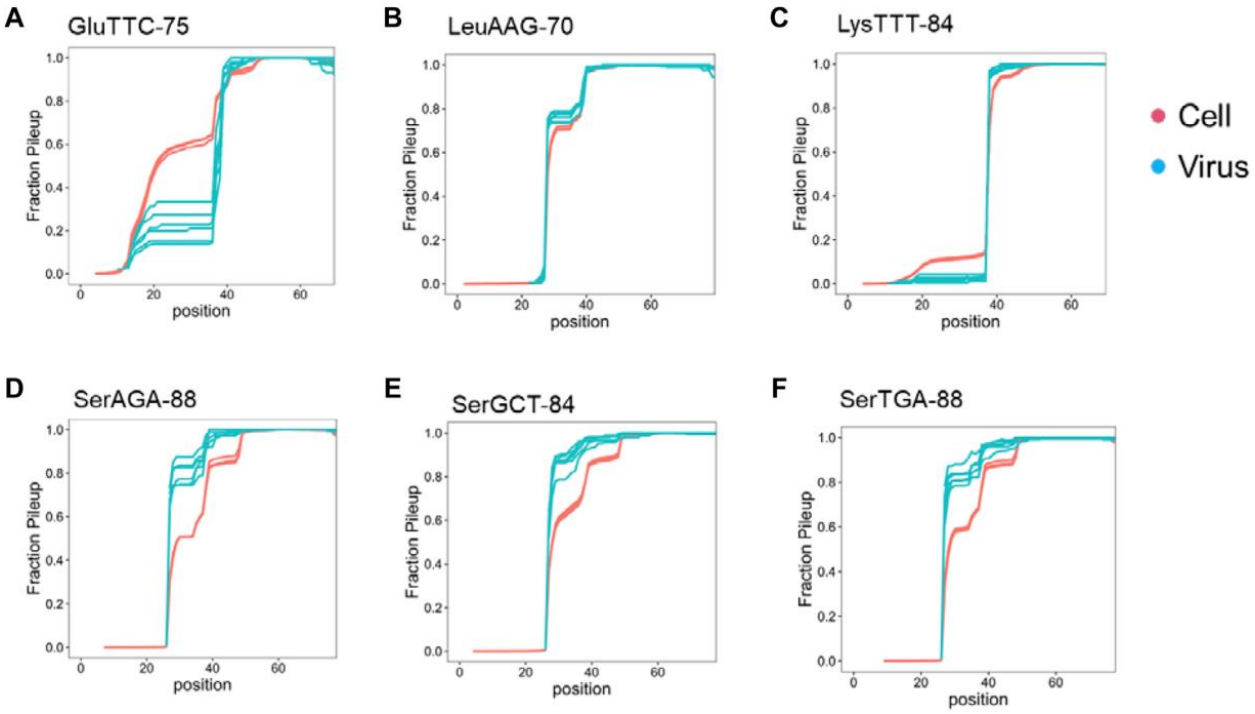

Figure 4.3: Read pileup of the enriched tRNA isodecoder in SARS-CoV-2 viral preparations. Shown are read pileups of the most abundant tRNA isodecoders in viral isolates (n = 6, blue) and their counterparts in uninfected Vero cell (n = 3, red). Isodecoder nomenclature is according to the tRNAScan score of the *Chlorocebus sabaeus* tRNA genes identified in Rfam database. (a) $tRNA^{Glu}$(TTC). This result is consistent with 3′ tRNA fragment being the dominant form in the viral preparations. (b) $tRNA^{Leu}$(AAG). This result is consistent with full-length tRNA in the viral preparations. (c) $tRNA^{Lys}$(TTT). This result is consistent with full-length or 3′ tRNA fragment with 5′ end exactly at position 39 in the viral preparations. (d) $tRNA^{Ser}$(AGA). The $tRNA^{Ser}$ results are consistent with full-length tRNA in the viral preparations. (e) $tRNA^{Ser}$(GCT). (f) $tRNA^{Ser}$(TGA).

We next looked for RNA modification differences among the tRNAs from cells and viral preparations. In tRNA-seq, certain modifications can be identified by their "mutation" signatures in the sequencing data. Vero cells are derived from *Chlorocebus sabaeus* kidney, their tRNA modifications have not been reported in the literature. We analyzed the mutation signatures of the cellular tRNAs in the libraries without demethylase treatment and validated the methylations using the results from demethylase treatment [137] to provide a comprehensive analysis of Vero cell tRNA modifications. Overall, the *C. sabaeus* tRNA modification patterns detected by sequencing are very similar to those from the human HEK293T cells [137]). A minor difference is the $m^2{}_2G26$ modification which is present in $tRNA^{Val}$ in *C. sabaeus* Vero cells but not in HEK293T cells. We detected inosine modification at the wobble anticodon position (I34) in all tRNAs that are A34 in the genome which include $tRNA^{Leu}$(AAG) and $tRNA^{Ser}$(AGA) (Figure 4.4(a), Figure 4.5(a)).

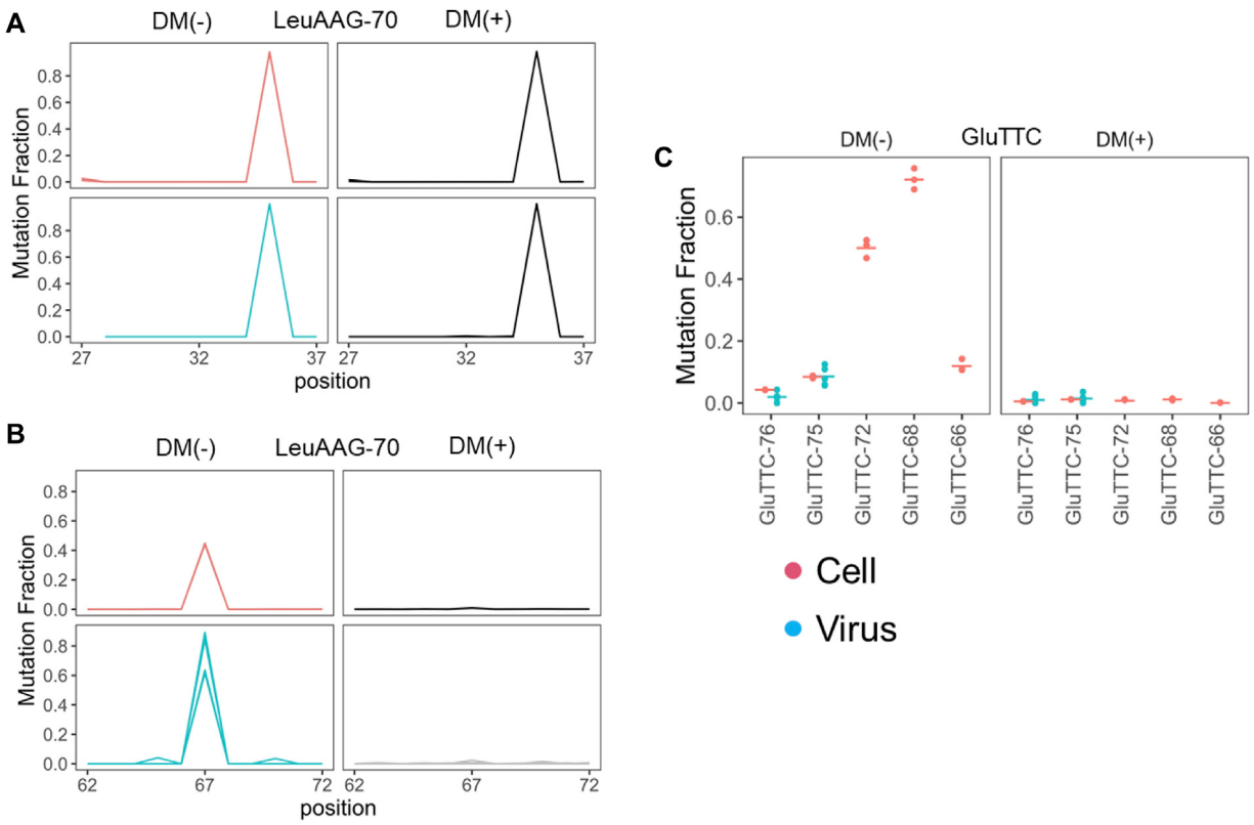

Figure 4.4: Selective enrichment of tRNA with m^1A modification profiles. Mutation fractions from uninfected Vero cell (n = 3, red) or cell free viral preparations (n = 6, blue) are shown. (a) Mutation fractions of $tRNA^{Leu}$(AAG) residues around the wobble anticodon position (35 for this tRNA) without (DM-) and with (DM+) demethylase treatment showing the I34 modification. (b) Mutation fractions of $tRNA^{Leu}$(AAG) around the residues at position 67 which corresponds to m^1A58 in the tRNA nomenclature. $tRNA^{Leu}$(AAG) shows higher mutation fraction in the viral preparations, consistent with SARS-CoV-2 selectively packaging m^1A modified $tRNA^{Leu}$(AAG). (c) Mutation fractions of the top five abundant $tRNA^{Glu}$(TTC) isodecoders at position 57 (DM-) which is validated as m^1A in the T loop upon removal by demethylase treatment (DM+). Isodecoder nomenclature is according to the tRNAScan score of each $tRNA^{Glu}$(TTC) gene. The two isodecoders enriched in the viral preparations are nearly unmodified, corresponding to their counterparts in the Vero cells.

We then compared the mutation levels between the tRNAs enriched in viral preparations and their counterparts in cells. For high confidence analysis we applied a filter of ≥50 read coverage at nucleotide positions of interest. The modification sites passing this filter among the tRNAs in viral preparations only include m^1A58 in $tRNA^{Leu}$(AAG), $tRNA^{Lys}$(TTT), and $tRNA^{Glu}$(TTC), and I34 in $tRNA^{Leu}$(AAG) and $tRNA^{Ser}$(AGA). For $tRNA^{Leu}$(AAG) and $tRNA^{Lys}$(TTT), the mutation fraction at m^1A58 is higher in the tRNA from the viral preparations than the VeroE6 cell tRNA (Figure 4.4(b), Figure 4.5(b)), suggesting preferential packaging of m^1A modified tRNA. Among the $tRNA^{Glu}$(TTC) isodecoders, m^1A58 level is variable in cells. Only two of the five abundant $tRNA^{Glu}$(TTC) isodecoders have high modification levels in VeroE6 cells, but only isodecoders with low modification fractions are present in the viral preparations (Figure 4.4(c)). Since $tRNA^{Glu}$(TTC) in the virions are likely tRNA fragments, this result is consistent with low m^1A modified $tRNA^{Glu}$(TTC) being the preferred source of $tRNA^{Glu}$(TTC) fragments in cells. tRNAs in cells and in the viral preparations are >90% modified with I34 in both $tRNA^{Leu}$(AGA) and $tRNA^{Ser}$(AGA) (Figure 4.4(a), Figure 4.5(a)).

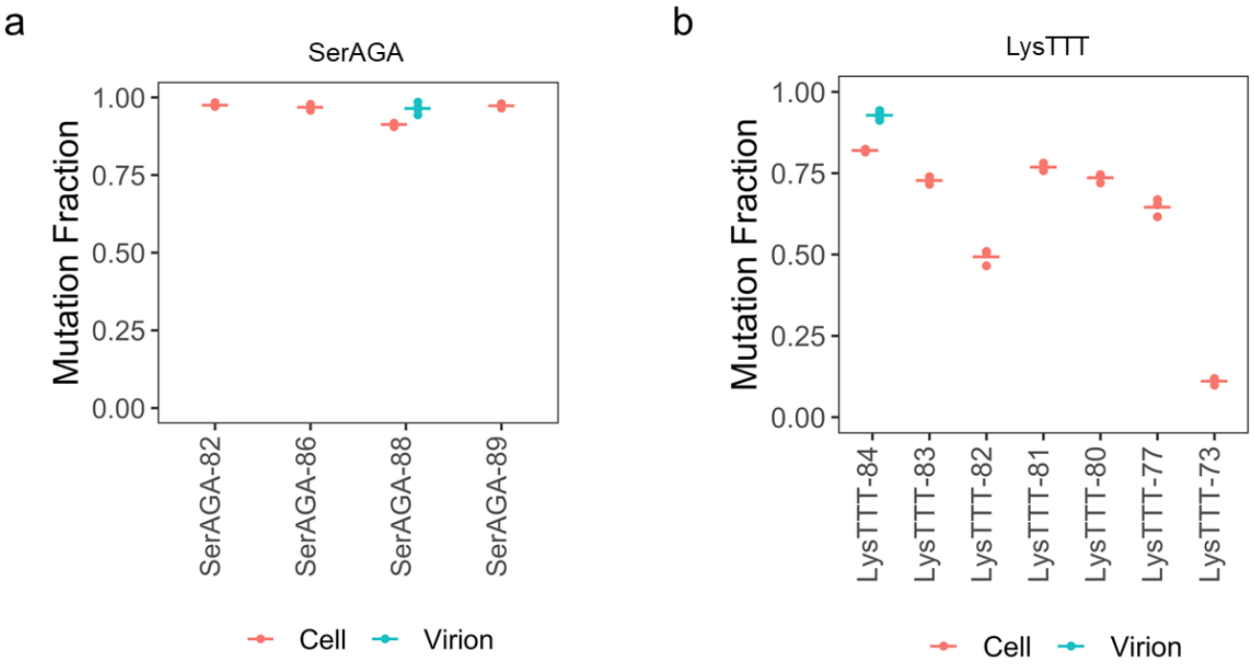

Figure 4.5: Selective enrichment of other tRNA with modification profiles. (a) Mutation fractions of $tRNA^{Ser}$(AGA) isodecoders from uninfected Vero cell (n=3, red) or viral preparations (n=6, blue) of the wobble anticodon position (I34). (b) Mutation fractions of $tRNA^{Lys}$(TTT) isodecoders of position m^1A58.

4.2.2 SARS-CoV-2 viral RNAs contain a variety of modifications

We next carried out large RNA-seq (>200 nt) of the viral preparations to characterize the viral RNA and its candidate modifications. We first removed small RNAs by size-selection, followed by chemical fragmentation and library construction. As expected, most of the reads mapped to the SARS-CoV-2 genome (Wuhan reference). We used a mutation threshold of >90% to identify single nucleotide polymorphisms (SNPs) in these samples (Figure 4.6(a)). These SNPs did not change upon our enzymatic or chemical treatment described below (data not shown). Our samples are derived from distinct viral isolates obtained from patients infected during the first pandemic wave in the spring of 2020, at the time when most of the circulating SARS-CoV-2 viruses still had only a few sequence changes [257]. We measured the read counts of the SARS-CoV-2 genomic RNA and the 18S + 28S rRNA in our viral preparations (Figure 4.6(b)). We found an average ratio of SARS-CoV-2/rRNA of ~9.5, or a molar ratio of SARSCoV-2/rRNA of ~2. Given that the cellular ribosomes and viral particles produced upon infection [258] is higher than 100:1, our results show a >200-fold enrichment of SARSCoV-2 viral RNA over rRNA in our culture supernatant preparations.

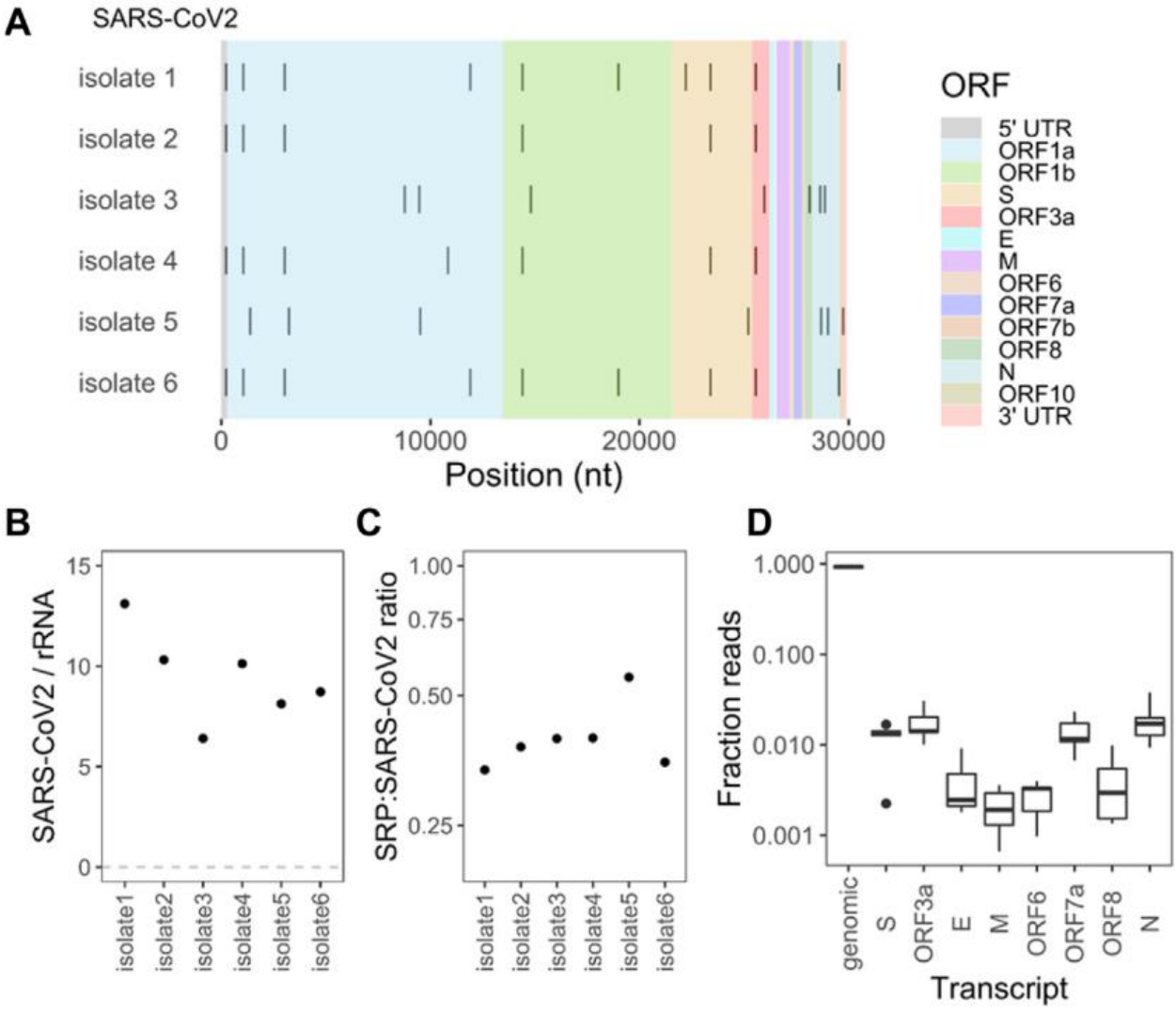

Figure 4.6: Large RNA sequencing identifies viral sequence variants, subgenomic viral RNAs, and signal recognition particle RNA. (a) Single nucleotide polymorphisms (SNPs) for each viral isolate identified by >90% mutation fraction from the Wuhan SARS-CoV-2 reference genome. (b) Mapped read count ratio of SARS-CoV2 genomic RNA to large ribosomal RNA (18S and 28S) in the viral preparations. (c) Normalized ratio of SRP RNA reads to SARS-CoV-2 genomic RNA reads in the viral preparations using the transcript size of 300 nucleotides for SRP, and 29,903 nucleotides for SARS-CoV-2. (d) Relative fraction of reads that bridge the junction between the 5′ leader region and the genomic RNA (set at 1) and between individual subgenomic RNA. Box and Whisker plot for n = 6 individual isolates.

We also measured the ratio of reads from the SRP RNA, which was the most abundant RNA in the viral preparations in the small RNA-seq data. We counted the read coverage for the six SARS-CoV-2 genomes and those mapping to the *C. sabaeus* SRP from the Rfam database [259]. We then used a normalization factor that accounts for the length difference of SARS-CoV-2 genome (29,903 nt) and SRP RNA (300 nt) to get a ratio of SRP to SARS-CoV-2 RNA (Figure 4.6(c)). We obtained a ratio of .3–.55 for the six viral preparations. Given the intermediate size of the SRP RNA that may incur losses during the size-selection step, we estimate that a virion contains up to one SRP RNA transcript.

We also asked whether our cell free viral preparations contain SARS-CoV-2 subgenomic RNA. Subgenomic RNAs are generated in the infected cells by joining a 5′ leader sequence to each of the SARS-CoV-2 protein coding genes (Figure 4.7). We found 2,000–3,000 reads that cover the junction region of the 5′ leader with Orf1a, which corresponds to the full-length SARS-CoV-2 RNA. Normalizing the reads covering other genes to this junction, we found up to 2% of subgenomic RNAs containing S, Orf3a, Orf7a, and E genes in all six isolates (Figure 4.6(d)). Altogether, up to 10% of the viral RNAs in the cell free viral preparations are subgenomic RNAs. This low level is consistent with our viral preparations comprising primarily of SARS-CoV-2 virions, and most virions in the culture containing the full-length genomic RNA. The subgenomic RNAs may represent in part virions devoid of the full-length genomic RNA (i.e., defective viral particles).

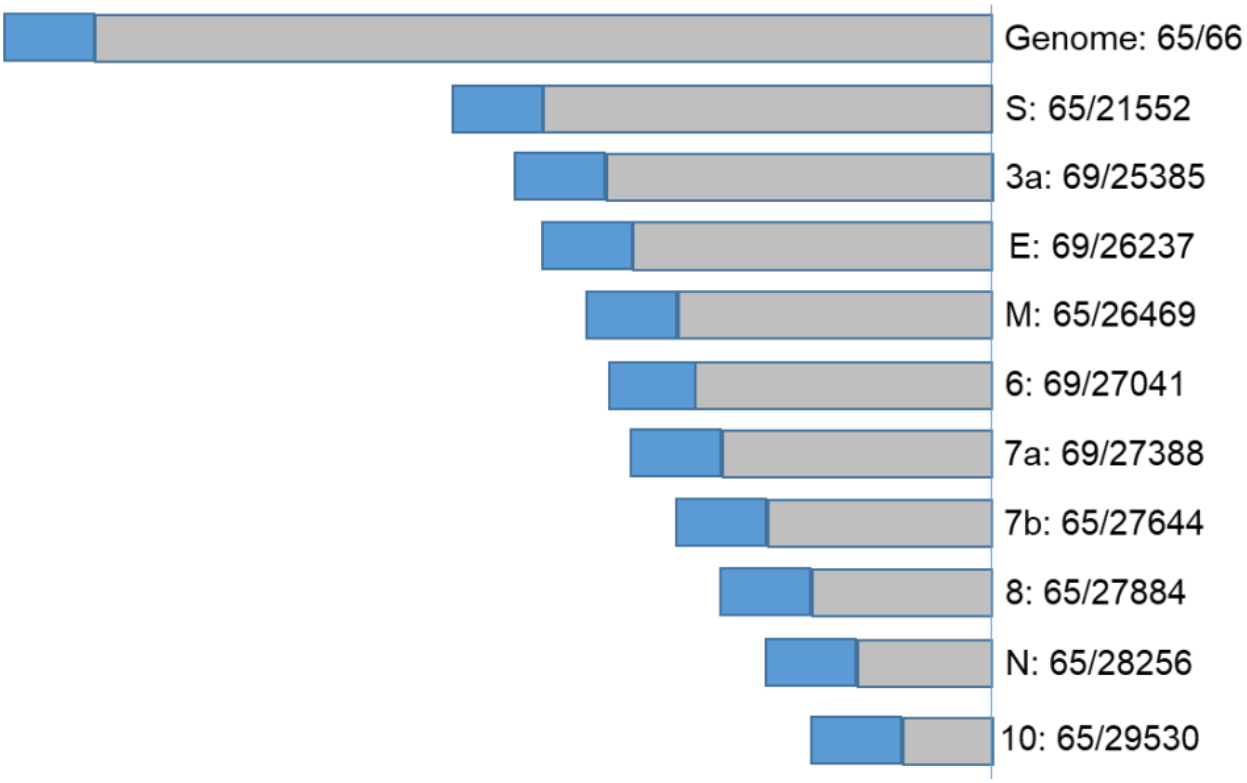

Figure 4.7: Subgenomic junction sites used in our large RNA analysis. Residue numbers are according to the Wuhan SARS-CoV-2 reference genome. Residues linking the 5' leader and the subgenomic regions are indicated.

Finally, we carried out experiments to identify novel modifications of SARS-CoV-2 vRNA. To facilitate this identification, we added two consecutive steps in the large RNA library construction, one with demethylase treatment (DM), and the other with 1-cyclohexyl-(2-morpholinoethyl) carbodiimide (CMC) treatment (Figure 4.8(a)). Demethylase treatment generally detects Watson-Crick methylations such as m1A in our tRNA studies, whereas CMC treatment is a standard procedure for detecting pseudouridine (Ψ) modification in mRNA [221]. We first carried out a threshold analysis in four pairwise comparisons of ±DM, ±CMC, ±DM (plus CMC treatment), ±CMC (plus DM treatment) using the filters for mutation difference of 5%, stop difference of 15%, deletion difference of 2%, and insertion difference of 2%. We filtered next the candidate sites from the threshold analysis to only include those that show the same signature

change in at least four of the six viral preparations. Only the mutation signature of five sites passed these two filters (Figure 4.8(b)). These sites fall into four groups: 1) U8323 and U20331 are DM-sensitive without CMC, CMC-sensitive without DM, not detected in DM with CMC, nor in CMC with DM. These two sites may represent N^3-methyl-U (m^3U) derivatives. 2) A29517 is DM-sensitive without CMC, CMC-sensitive without DM, the mutation signals are reversed in DM with CMC, or in CMC with DM. This site may represent N^1-methyl-A (m^1A) derivatives. 3) U3877 is DM-insensitive without CMC, CMC-sensitive without DM, DM-sensitive with CMC, not detected in CMC with DM. This site may represent N^1-methyl or N^3-methyl-pseudouridine ($m^1\Psi$, $m^3\Psi$) derivatives. 4) A29780 is DM-insensitive without CMC, CMC-sensitive without DM, DM-sensitive with CMC, not detected in CMC with DM. We do not recognize an A modification at this time that would generate such signatures. No stop signature was observed for any of these five sites (Figure 4.9).

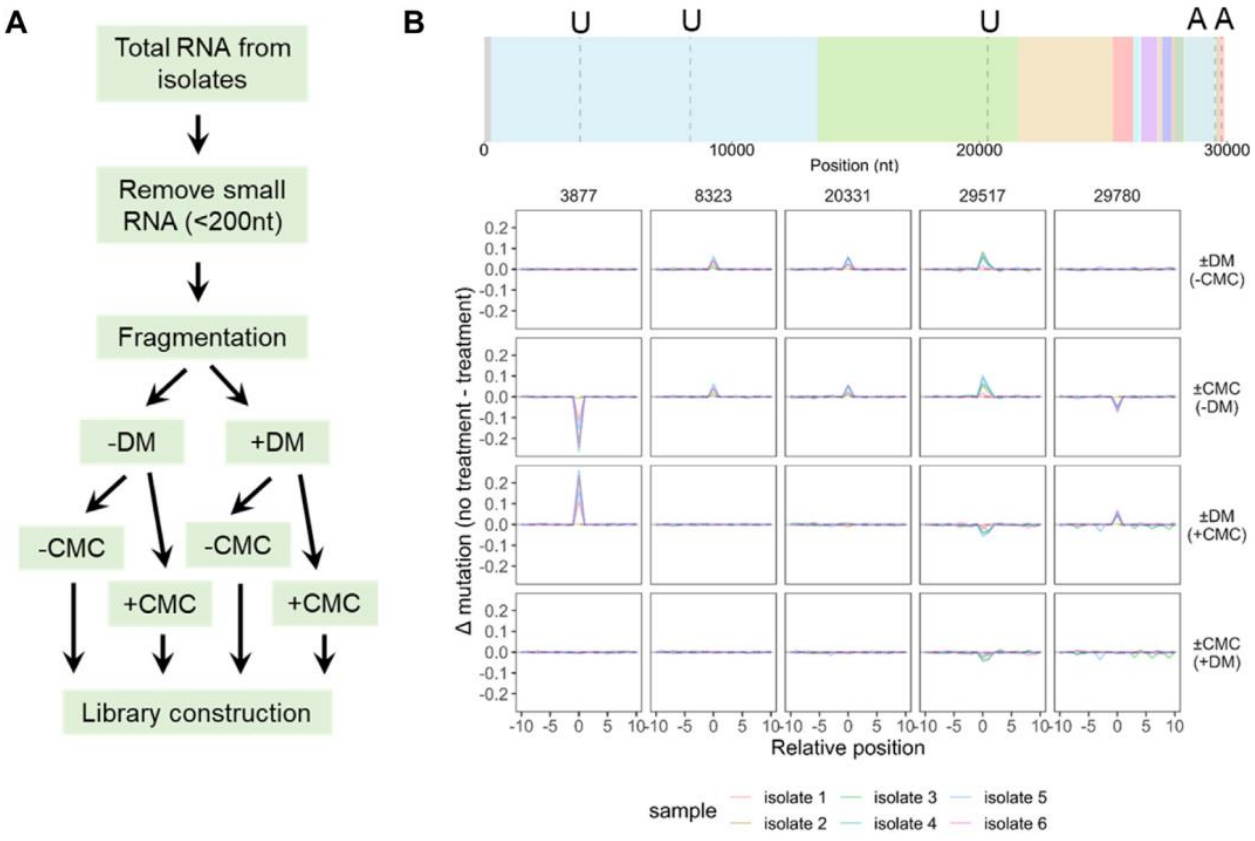

Figure 4.8: Large RNA sequencing identifies candidate SARS-CoV-2 modifications. (a) Scheme of modification detection. An enzyme treatment (DM) and a chemical treatment (CMC) are added before library construction, producing four combinations for each sample. (b) Candidate modifications from comparing four treatment conditions. Site locations are indicated in a dashed line, nucleotide identity indicated on top of the SARS-CoV-2 gene map. Data from top to bottom: with and without demethylase only (±DM, -CMC); with and without CMC only (±CMC, -DM), with and without demethylase, also CMC (±DM, +CMC); with and without CMC, also DM (±CMC, +DM). Positions with mutations >5% in at least 4 of the 6 isolates but excluding the SNP positions in Figure 4.6(a) are shown.

It is intriguing to note that these five candidate modification sites in SARS-CoV-2 do not follow the expected pattern of changes for well-characterized RNA modifications. In all five cases, the mutation levels are at most 20%, which may reflect either low levels of modification at these sites or under-counting the modification levels because of the unknown nature of these modifications. Future studies are needed to pinpoint the exact nature of these modifications.

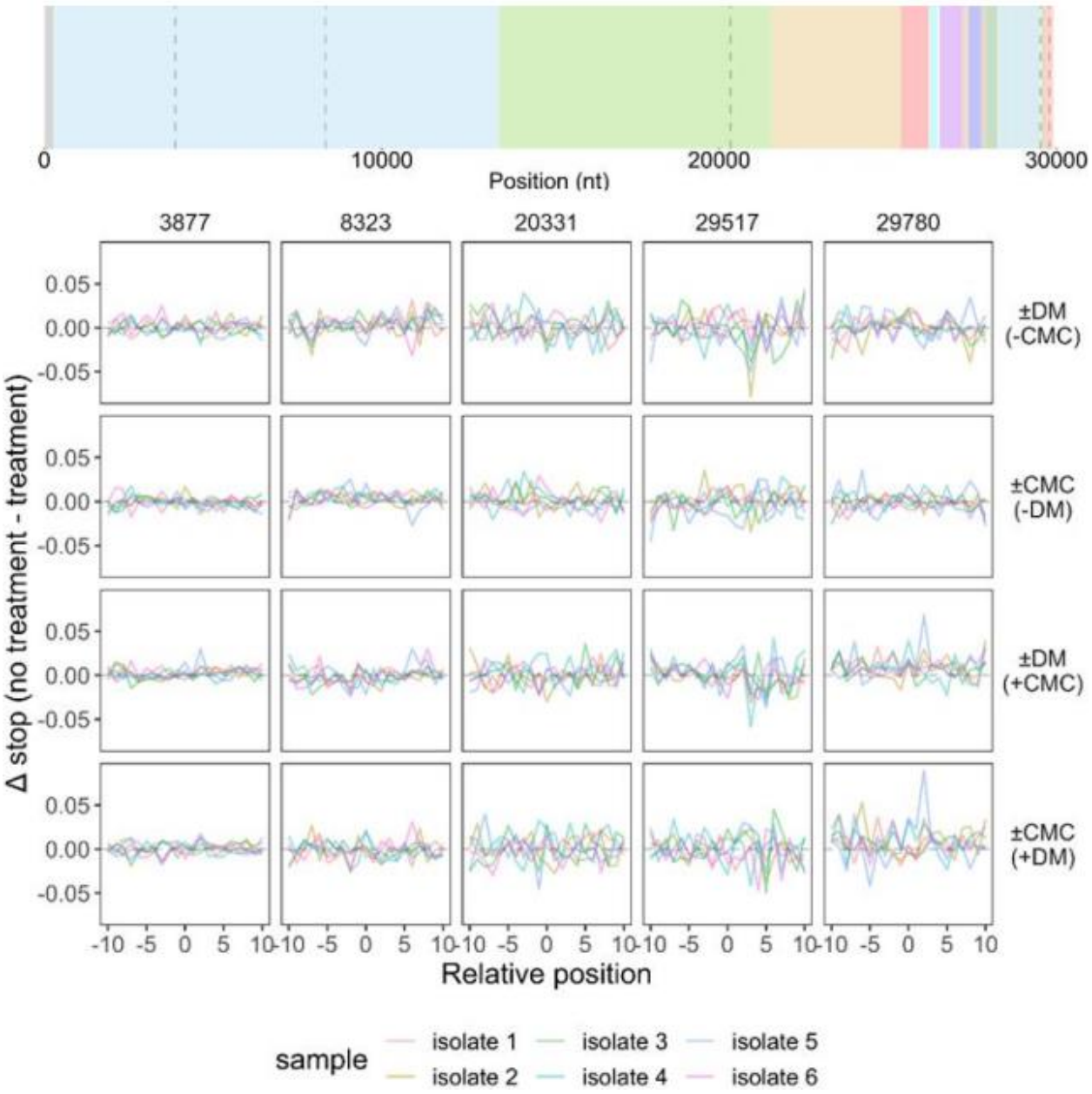

Figure 4.9: Stop signatures of the five sites in Figure 4.8(b).

4.2.3 Periodate oxidation renders the anticodon loop modifications Q, τm5s2U, mcm5s2U, cmnm5s2U, and s2C by deletion & mutation signatures

4.2.3.1.Periodate treatment produces deletion signatures for Q-modification in sequencing

Periodate is known to oxidize cis-diol groups into aldehydes. Periodate oxidation can be used to study tRNA aminoacylation levels by chromatography, microarrays, or sequencing [30,218,260]. The Q-base has a cis-diol group that is a known substrate for periodate oxidation, a common reaction used to confirm the presence of Q modification in APB gel electrophoresis [180,200]. In our tRNA-seq procedure to measure tRNA charging levels, periodate treatment is a pre-requisite step in the sequencing library construction before reverse transcription (30,31). Using total RNA from HEK293T cells (Figure 4.10(b)) that are either completely devoid of tRNA Q-modification (0Q, (8)) or fully modified with Q (100Q), we found unexpectedly that $tRNA^{His}$ from 100Q cells produced deletion in the high throughput sequencing data at the Q34 nucleotide at ~5% level only in the periodate treated, but not in the untreated sample (Figure 4.10(c)). The signature is absent in 0Q cells. The periodate-dependent deletion signature of the Q nucleotide in sequencing is the most pronounced among the other signatures analyzed such as misincorporation of bases (mutation), addition of extra bases (insertion), or the location of the 5’ end of the sequencing products (stop; Figure 4.10(d)).

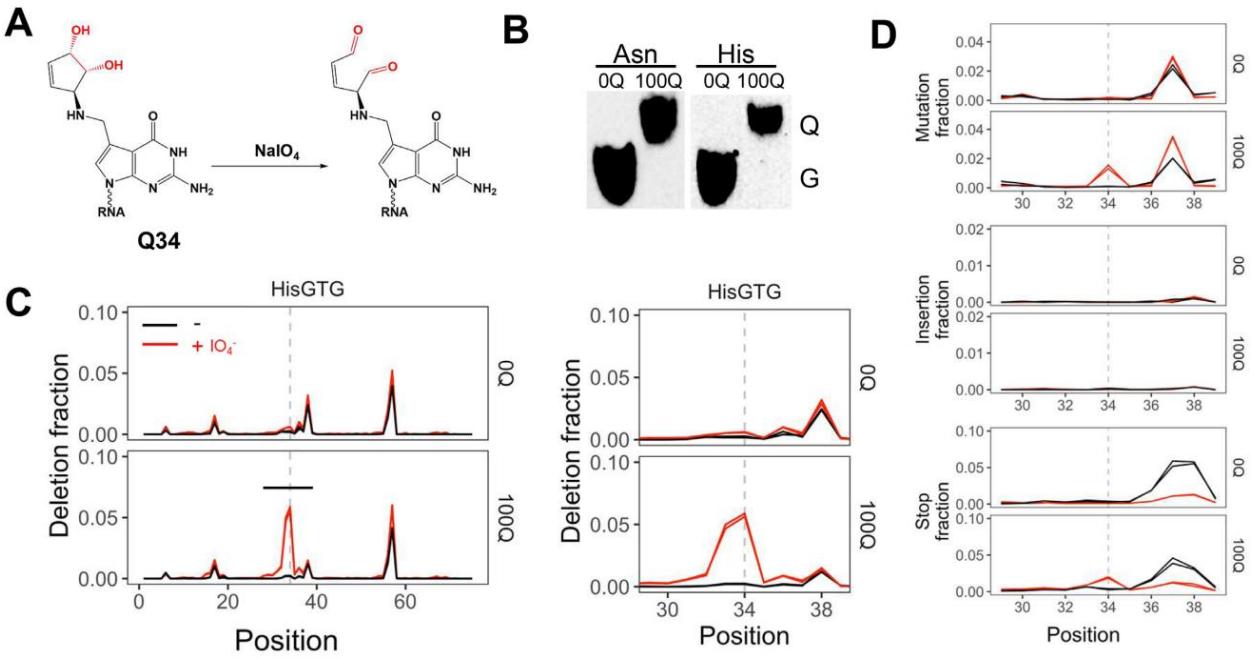

Figure 4.10: Q-modification in tRNA generates deletion signature after periodate treatment. (A) Chemical structure of the Q-base and its proposed periodate oxidized form. (B) Northern blot of APB gel showing the $tRNA^{Asn}$ and $tRNA^{His}$ Q-levels in 0Q and 100Q samples used in sequencing. (C) Deletion fraction of $tRNA^{His}$, in 0Q and 100Q samples, ±periodate. Expanded view in the region of ±5 nt to Q34 residue (dashed line). The biological replicates are overlaid in each graph. Only the most abundant $tRNA^{His}$ isodecoder is shown. A known modification that also produces deletion signature is m^1G at position 37. (D) Same samples and legend as panel C showing mutation, insertion, and stop fractions in the region of ±5 nt to Q34 residue.

To further evaluate how a fully Q-modified nucleotide only results in a ~5% deletion signature, we explored how different reverse transcriptases (RT) interact with periodate-oxidized queuosine. Among six different RT enzymes tested, the Q-associated, periodate-dependent signature from mutation, deletion, or stops are highly idiosyncratic (Figure 4.11(a)). Each enzyme leaves a distinct misincorporation signature composed of different fractions of mutations, deletions, insertions, and stops (Figure 4.11(b)). The deletion fraction generated by the RT used in this work, SuperScript IV (SSIV), produced the most robust signal among the RTs tested.

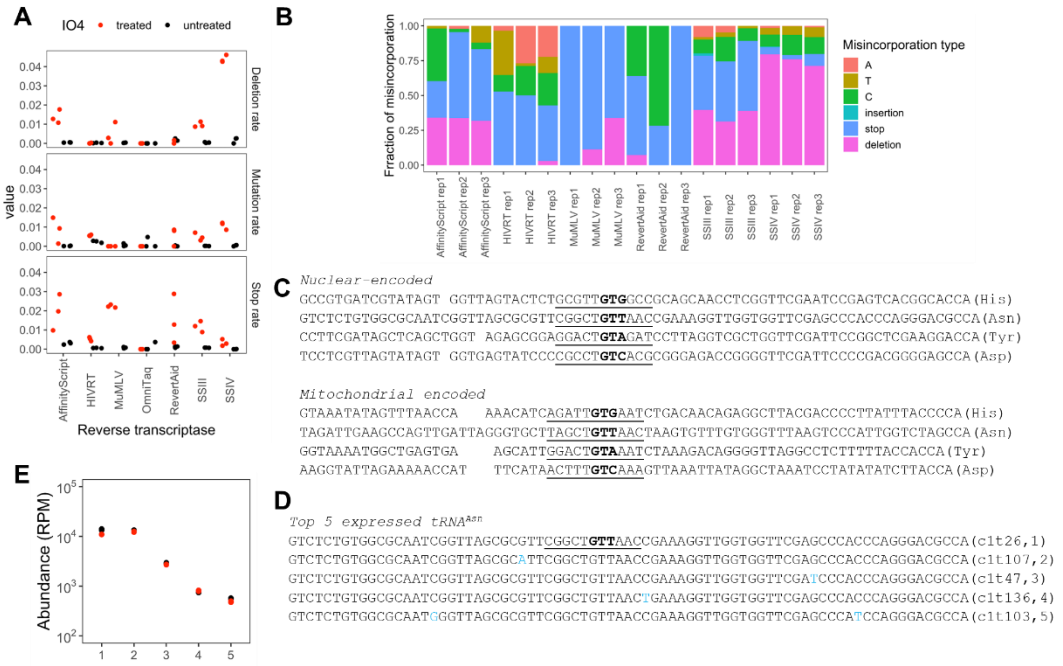

Figure 4.11: Additional results and information on Q-modified tRNAs. (a) Libraries for 100Q samples constructed using different reverse transcriptases (RT); the periodate-dependent sequencing signature at Q34 of tRNAHis is shown. n=3 biological replicates. X-axis shows the names of the commercial RT used. (b) For each sample shown in (A), the relative fractions for each misincorporation type are shown: mutation from G to A, T, C, insertion, stop, and deletion. (c) Sequence of Q-modified tRNAs shown in Figures 4.10 and 4.12. The anticodon nucleotides are in bold, and the ±5 nt region is underlined. (d) Top 5 expressed tRNAAsn isodecoder in HEK293T cells. The anticodon nucleotides are in bold, and the ±5 nt region is underlined in the highest expressed isodecoder. Sequence differences among the isodecoders are shown in blue. (E) Abundance of the top 5 tRNA^{Asn} isodecoders. The numbers 1-5 correspond to the tRNA sequences in panel (d).

A total of eight tRNAs in human cells can be modified with Q. The nucleus-encoded $tRNA^{His}$ and $tRNA^{Asn}$ are modified with Q, whereas $tRNA^{Tyr}$ and $tRNA^{Asp}$ are further modified to galactosyl-Q and mannosyl-Q, respectively [261,262]. In addition, the mitochondria-encoded tRNAs for these same 4 amino acids are also modified with Q [184]. We first examined the deletion signatures for other Q-modified tRNAs. The cytosolic $tRNA^{Asn}$ displayed a relatively high deletion fraction of ~13% at the Q34 position that is Q-modification and periodate-dependent (Figure 4.12(a)). All four mitochondrial tRNAs show deletion signatures in the same manner at the Q-modified nucleotide as well, ranging from ~4% in mt-$tRNA^{Asp}$ to ~20% to mt-$tRNA^{Asn}$ (Figure 4.12(b–e)). These results confirm that Q-modified nucleotide can indeed be detected using periodate-treated RNA-seq libraries.

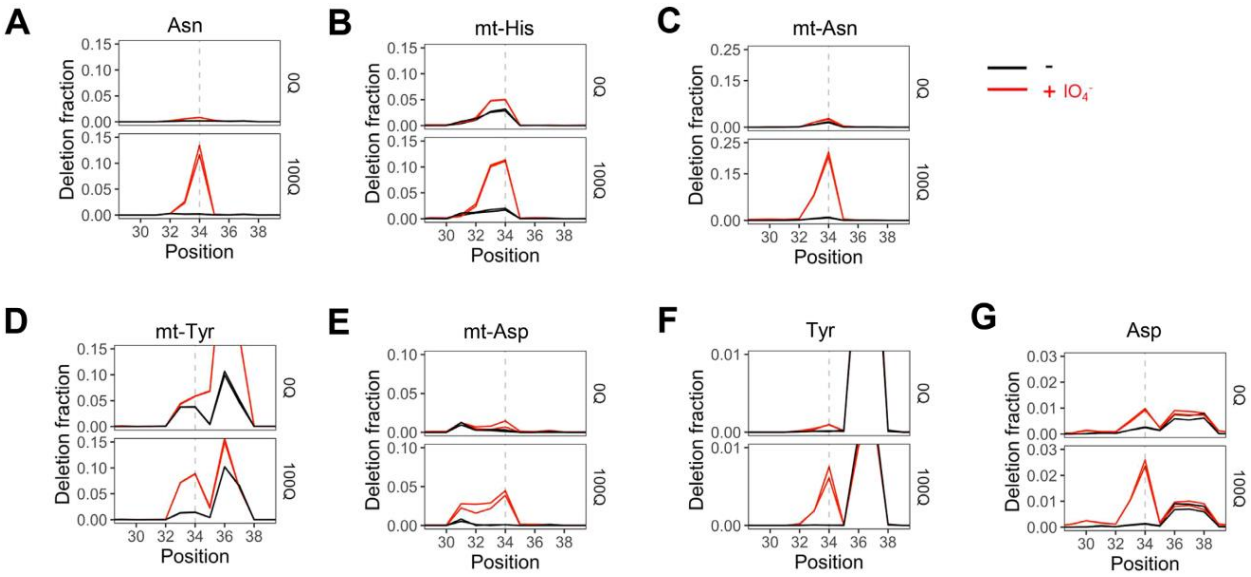

Figure 4.12: Periodate-treatment dependent deletion signatures in other tRNAs containing Q or glycoQ modifications. Shown are regions ±5 nt to the Q34 residue (dashed line) in each tRNA. Biological replicates are overlaid in each graph. For nucleus-encoded tRNAs, only the most abundant isodecoder for Asn/Tyr/Asp is shown. All residue numbers are according to the standard tRNA nomenclature, i.e. the wobble anticodon nucleotide is at position 34. (a) $tRNA^{Asn}$, (b) Mitochondria-encoded $tRNA^{His}$. (c) Mitochondria-encoded $tRNA^{Asn}$. (d) Mitochondria-encoded $tRNA^{Tyr}$. A known modification that also produces deletion signature is ms^2i^6A at position 37. (e) Mitochondria-encoded $tRNA^{Asp}$. (f) $tRNA^{Tyr}$. A known modification that also produces deletion signature is m^1G at position 37. (g) $tRNA^{Asp}$. The periodate-independent deletion signature at positive 37 is unknown.

Why does oxidized Q-base induce a deletion signature in the reverse transcriptase (RT) reaction in the sequencing library construction? Q-base affects anticodon-codon pairing through altering anticodon loop geometry and increasing its rigidity [40,262,263]. In vitro studies of codon-anticodon complexes show a 3-fold increase in stabilization of Q–U pairings over G-U, while pairings with C were destabilized [263]. The 5-member ring of the Q-base is located in the major groove of the RNA-DNA hybrid in the active site of reverse transcriptase. Periodate oxidation opens the ring which may lead to increased flexibility and steric occupancy of the oxidized moiety in the major groove, thereby inducing the RT to skip the oxidized Q nucleotide. However, the deletion fraction has a strong dependence on the context of the Q-modified nucleotide (Figure 4.11(c–e)). One empirical factor is the nucleotide sequences immediately upstream of the Q34 residue. $tRNA^{Asn}$ has an upstream C32 (5’ GGCUQUU) and a high deletion fraction, whereas $tRNA^{His}$ has an upstream U32 (5’ CGUUQUG) and a low deletion fraction. Mitochondrial mt-$tRNA^{Asn}$ (5’ AGCUGUU) has C32 and mt-$tRNA^{His}$ (5’ GAUUGUG) has U32 which again is consistent with the observed high and low deletion fraction for these two tRNAs. Mitochondrial mt-$tRNA^{Asp}$ (5’ CUUUGUC) has U32 and U31 which may correlate with the observed deletion signature spanning four nucleotides (Figure 4.12(e)). Furthermore, other modifications near the Q34 nucleotide may also generate additional context dependence of our results. For example, mt-$tRNA^{Tyr}$ (5’ GACUGUA) could have a high deletion fraction, but this may be obscured by the ms^2i^6A37 modification that has a much larger deletion signature with and without periodate treatment (Figure 4.12(d)).

The glyco-Q modified tRNAs do not significantly react with the boronic acid derivative to cause a gel shift like the Q-modified tRNAs. However, both galactose and mannose can form a small proportion of furanose tautomer that contain *cis*-diol in equilibrium with their major

pyranose tautomer [265]. Our 100Q sample is known to have nearly stoichiometric amount of glycosylated Q-modification as measured by a combination of APB and acid denaturing gel electrophoresis [201]. We found that both nucleus-encoded $tRNA^{Tyr}$ and $tRNA^{Asp}$ also display deletion signatures using periodate treatment in our sequencing (Figure 4.12(f,g)) albeit the deletion fraction was only ~0.5% or 2% for $tRNA^{Tyr}$ and $tRNA^{Asp}$, which are substantially lower than Q-modified tRNAs. Since the deletion background in our sequencing is <0.1%, these low deletion fractions are still useful in detecting glyco-Q modifications, especially in tRNA sequencing where the coverage at the glyco-Q nucleotides can easily reach >1000. The differences in the deletion fraction among $tRNA^{Tyr}$ and $tRNA^{Asp}$ may be related to the periodate reacted product of galQ and manQ, both tRNAs have C32 in their upstream sequences ($tRNA^{Tyr}$ 5' GACUGUA, $tRNA^{Asp}$ 5' GCCUGUC).

4.2.3.2.Deletion fraction can be used to quantify Q-modification levels

An important application to use sequencing to study Q-modification is the potential ability to quantify Q modification fraction in any biological sample, which enables simultaneous assessment of transcriptome-wide tRNA properties associated with Q-modification. To assess whether the deletion signature can be used to quantify Q levels, we systematically mixed two biological samples of 0Q and 100Q HEK293T cells at varying ratios between 0 and 100% and performed sequencing reactions after periodate treatment. As expected, the deletion fraction at the Q34 position for $tRNA^{His}$ and $tRNA^{Asn}$ steadily increases between the mixture with increasing proportion of 100Q RNA (Figure 4.14(a,b)). Changes in the deletion fraction has a better fit with an exponential dependence of Q-modification fraction (r^2 = 0.964, 0.985, Figure 4.14(c)) than linear fits (r^2 = 0.901, 0.923, Figure 4.13). A non-linear fit can be explained by Q34 tRNA being

reverse transcribed less efficiently than unmodified (G34) tRNA upon periodate treatment. Because only up to 20% of Q34 tRNA produces deletion, the quantitative production of the deletion-containing cDNA is skewed when the Q-level is low. Similar results have been observed for other modifications that reduce the RT efficiency such as N^1-methyladenosine (m^1A) compared to unmodified RNA [266,267].

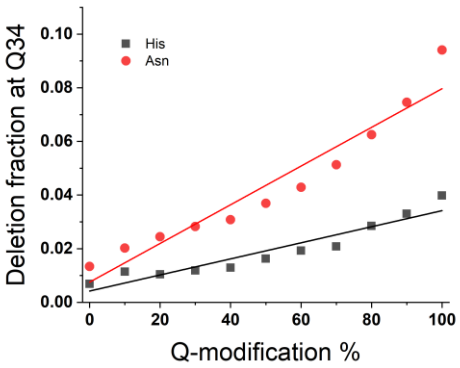

Figure 4.13: Linear fit plots of $tRNA^{His}$ and $tRNA^{Asn}$ Q deletion fractions. The r^2-value of the fit is 0.901 and 0.923 for His and Asn, respectively. For exponential fit (Fig. 3C), the r^2-value of the fit is 0.964 and 0.985 for His and Asn, respectively.

Interestingly, both $tRNA^{His}$ and $tRNA^{Asn}$ show a very similar slope in the Q-level calibration curve (Figure 4.11(c)). This result is consistent with the absolute deletion fraction being dependent on the sequence context of Q34 in individual tRNAs, but the changing Q-levels respond the same way to the RT reaction.

Five $tRNA^{Asn}$ isodecoders comprise >95% of total $tRNA^{Asn}$ (Figure 4.14(d)), whereas a single $tRNA^{His}$ isodecoder comprises >99% of total $tRNA^{His}$ in our HEK293T RNA samples. Examination of the deletion fraction of the five $tRNA^{Asn}$ isodecoders in 100Q samples show nearly identical deletion rates, indicating all are modified at the same level, which is consistent with the identical sequence in the 11-nucleotide window around the Q34 residue (Figure 4.11(b)).

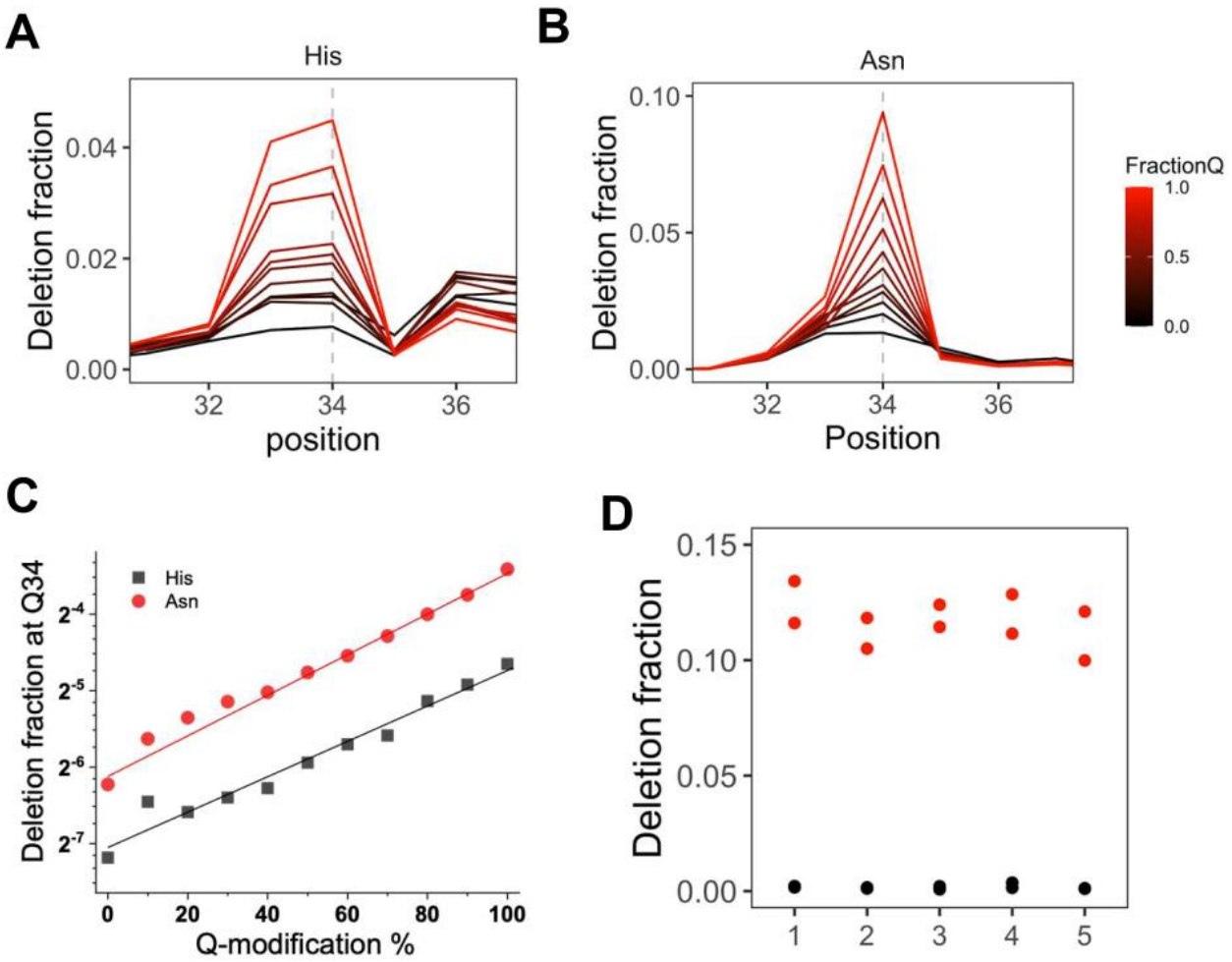

Figure 4.14: Quantitative assessment of Q-modification levels in $tRNA^{His}$ and $tRNA^{Asn}$. Shown are regions ±3 nt to the Q34 residue (dashed line) in each tRNA. (a) Overlay of deletion fraction for $tRNA^{His}$ of the 11 samples that are pre-mixed with proportions of 0Q and 100Q RNAs. (b) Overlay of deletion fraction for $tRNA^{Asn}$ of the 11 samples that are pre-mixed with proportions of 0Q and 100Q RNAs. (c) The deletion fraction at Q34 can be fit to the equation $\log_2 y = a + bx$, where a is the intercept and b is the slope. The $tRNA^{His}$ curve has a fit of a = –7.0, b = 0.023 and $r^2 = 0.964$. The $tRNA^{Asn}$ curve has a fit of a = –6.0, b = 0.025, and $r^2 = 0.985$. (d) Isodecoder Q levels of the top 5 expressed $tRNA^{Asn}$.

4.2.3.3.Periodate treatment also produces sequencing signatures in 2-thio-modifications

Periodate is also known to oxidize sulfides and thiol groups [268-270] which could alter the sequencing signatures of thio modifications in tRNA. Thio modifications in human tRNAs are primarily 2-thio (s^2), which substitutes the oxygen atom with a sulfur at the 2-position of uridine (Figure 4.15(a)). The 2-thio modification is in the minor groove of the DNA–RNA hybrid in the active site of the reverse transcriptase. 2-thio oxidation may alter the proof-reading mechanism of the RT, resulting in detectable signatures in RNA sequencing.

In mitochondria-encoded tRNAs, the s^2U-modification is present in the wobble anticodon position of $tRNA^{Gln}$, $tRNA^{Glu}$ and $tRNA^{Lys}$ in the context of 5-taurinomethyl-2-thiouridine (τm^5s^2U, [184]). We found a strong mutation signature for mt-$tRNA^{Gln}$ and mt-$tRNA^{Glu}$ at the modified nucleotide that are also accompanied with a strong double deletion signature 1-2 nucleotides upstream from the modified nucleotide that are periodate-dependent but not Q-dependent, as expected (Figure 4.15(b)). Mitochondrial $tRNA^{Lys}$ shows periodate-dependent deletion consistent with a 2-thio modification but no mutation (Figure 4.15(b)). This result may be derived from the unique sequence and/or other modifications around the 5-methyltaurine modified nucleotide. Among the τm^5s^2U34 modified tRNAs, only mt-$tRNA^{Lys}$ has a N^6-threonylcarbamoyladenosine (t^6A) modification at position 37 [184], which may influence obtaining a mutation signature in the RT reaction. Another possibility is that the mt-$tRNA^{Lys}$ in our specific sample (total RNA from HEK293T cells) may not contain a 2-thio modification at the U34 position.

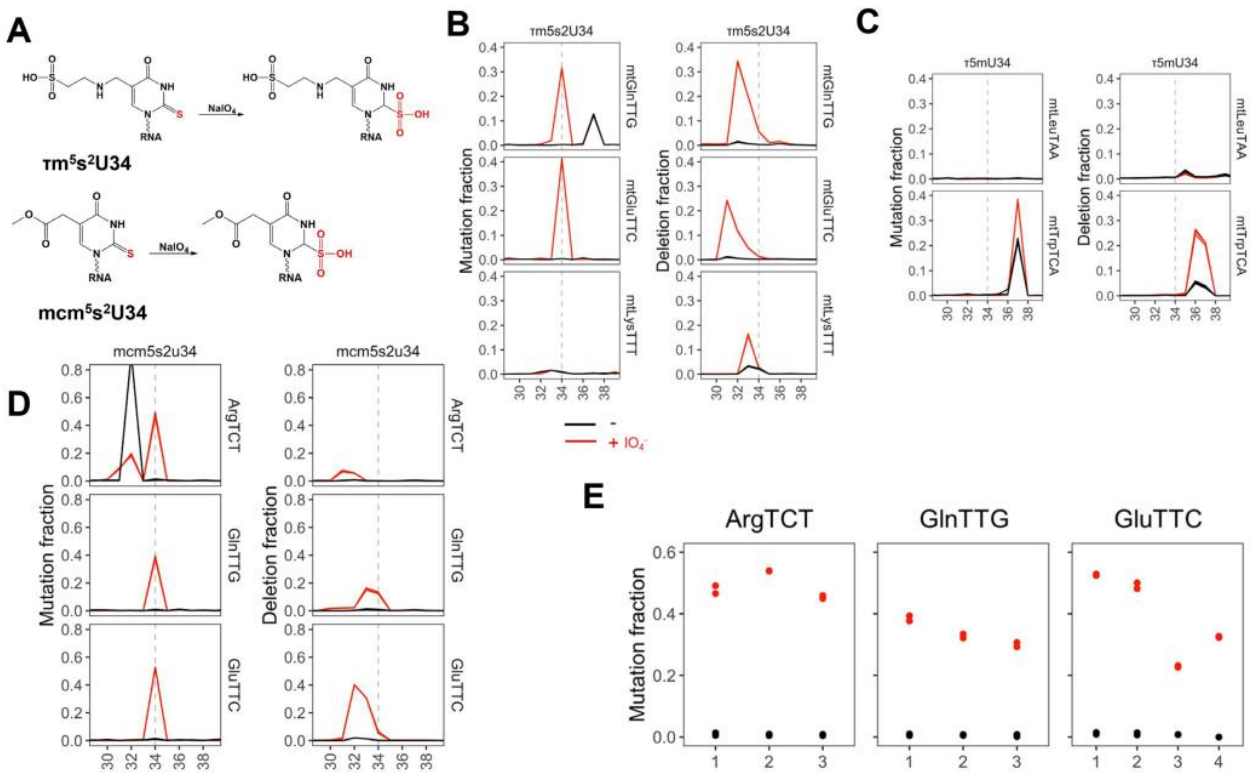

Figure 4.15: Periodate-treatment dependent analysis of 2-thio tRNA modifications. All residue numbers are according to the standard tRNA nomenclature, i.e. the wobble anticodon nucleotide is at position 34. (a) Chemical structure of the 2-thio-modifications and their proposed periodate oxidized forms. Shown in (b–d) are mutation and deletion signatures in regions ±5 nt to the relevant residue (dashed line) in each human tRNA, 0Q samples with (red) and without (black) periodate treatment. Biological replicates are overlaid in each graph. (b) Mitochondrial $tRNA^{Gln}$, $tRNA^{Glu}$, and $tRNA^{Lys}$ known to contain 5-taurinomethy-2-thio-U (τm^5s^2U) at wobble anticodon position. (c) Mitochondrial tRNALeu(TAA), and tRNATrp known to contain 5-taurinomethy-U (τm^5U) at wobble anticodon position. (d) Nucleus-encoded tRNAs known to contain 5-methoxycarbonylmethyl-2-thio-U (mcm^5s^2U)34. A known modification that also produces signatures is m^3C at position 32 for $tRNA^{Arg}$(TCT). (e) mcm^5s^2U34 mutation rates and abundance for isodecoders of $tRNA^{Arg}$(TTC), $tRNA^{Gln}$(TTG) and $tRNA^{Glu}$(TTC) with and without periodate treatment.

Importantly, the two mitochondrial tRNAs that have 5-methyltaurine, but no 2-thio modification, mt-tRNA^{Leu}(TAA) and mt-tRNA^{Trp} [184], do not show periodate-dependent mutation or deletion signature at the τm^{5}U position (Figure 4.15(c)), which lends support that the periodate-dependent mutation and deletion in mt-tRNA^{Gln} and mt-tRNA^{Glu} are indeed derived from the 2-thio modification. Unexpectedly, strong, periodate-dependent mutation and deletion signatures are present in mt-tRNA^{Trp} that corresponds to the known ms^{2}i^{6}A37 in this tRNA. The sulfur atom in ms^{2}i^{6}A may also be subject to the thio-modification, which likely contributes to the periodate-dependent signatures in sequencing.

In nucleus-encoded tRNAs, the s^2-modification is present in the wobble anticodon position of tRNA^{Arg}(TCT), tRNA^{Gln}(TTG) and tRNA^{Glu}(TTC) in the context of 5-(carboxy)methylaminomethyl-2-thiouridine (mnm^{5}s^{2}U34). Indeed, we found a strong mutation signature right at the modified nucleotide for all three tRNAs in both 0Q and 100Q cells that is periodate-dependent, but not Q-dependent (Figure 4.15(d)). A double deletion signature is also present on or upstream of the modified nucleotide depending on the tRNA species. These results indicate that periodate treatment is capable of detecting mnm^{5}s^{2}U34 modifications, although the absolute mutation and deletion signatures likely depend on the context of the neighboring sequences and other modifications.

We compared the mutation signatures for the abundant tRNA^{Arg}(TCT), tRNA^{Gln}(TTG) and tRNA^{Glu}(TTC) isodecoders in our samples (Figure 4.15(e)). Among the isodecoders of tRNA^{Arg}(TCT) and tRNA^{Gln}(TTG), the mutation rates are comparable to each other despite the variations of their abundances. This result may be expected as the sequence differences of these isodecoders are all outside of the ±5 nucleotide window of the modification. A ~2.5-fold difference was observed among the tRNA^{Glu}(TTC) isodecoders. This difference, however, may be

attributed to the substantial sequence or other modification difference between these isodecoders within the ±5 nucleotide window of the thio-modification, rather than a biological difference of the mnm^5s^2U34 modification fraction.

4.2.3.4.Thio-modifications in E. coli & in stress response

2-Thio modifications are also present in E. coli tRNAs (Figure 4.16(a)). We performed sequencing of *E. coli* tRNA with and without periodate treatment, and found strong, periodate-dependent mutation and deletion signatures for the known 5-carboxymethylaminomethyl-2-thiouridine ($cmnm^5s^2U$) modification at the wobble anticodon position in $tRNA^{Gln}$(TTG) and $tRNA^{Glu}$(TTC) (Figure 4.16(b)). Again, the mutation is at the modified nucleotide, whereas the deletion is at or immediately upstream of the modification nucleotide.

E. coli tRNA contains 2-thio-C (s^2C) and 4-thio-U (s^4U) modifications that are absent in human tRNA. We found strong, periodate-dependent mutation for the known s^2C32 modification in all 5 tRNAs right at the modified nucleotide (Figure 4.16(c)). In each case, a low level of deletion was also observed at 1–3 nucleotides upstream of the s^2C modified nucleotide. On the other hand, s4U modification at position 8 only shows an expected mutation signature [44] that is independent of periodate treatment and no deletion signature (Figure 4.16(d)). The 2-thio modification is located in the minor groove, whereas the 4-thio modification is in the major groove of the DNA–RNA hybrid in the active site of RT. These results are consistent with thio-oxidation in the minor groove interfering with the proof-reading activity of reverse transcription.

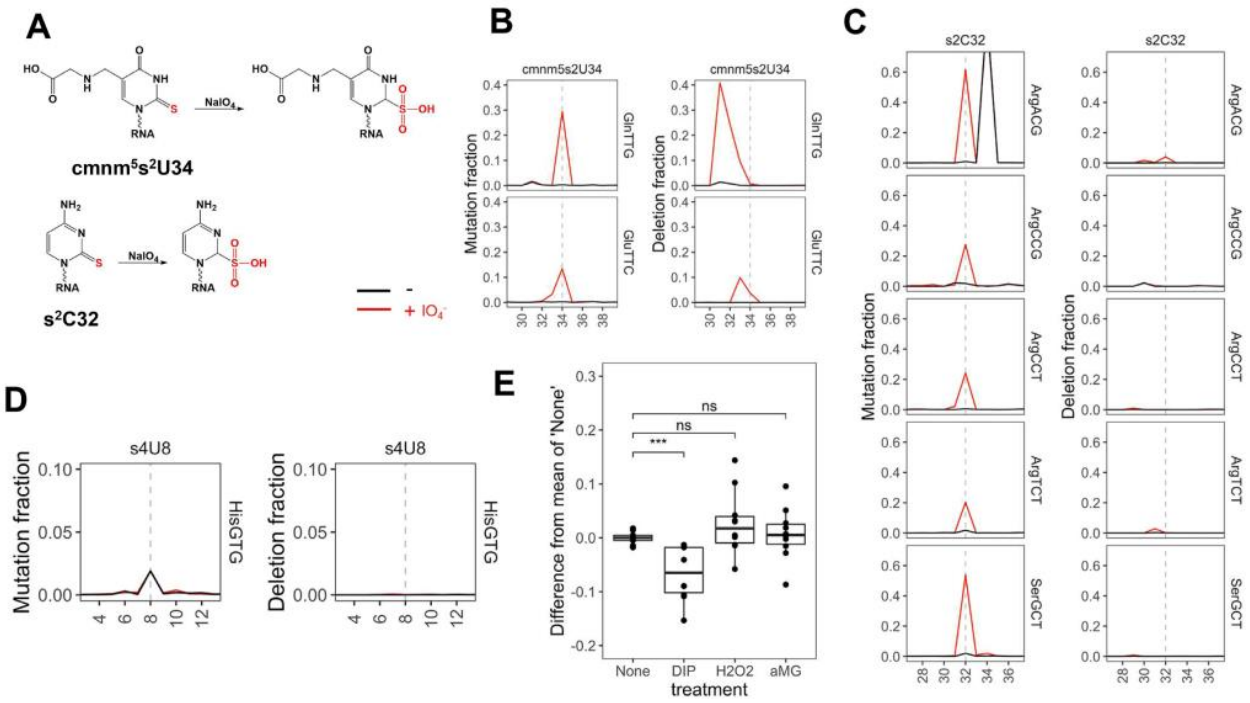

Figure 4.16: 2-thio tRNA modifications in E. coli tRNA and response to stress. Shown are mutation and deletion signatures in regions ±5 nt to the relevant residue (dashed line) in each tRNA. (A) Chemical structure of the 2-thio-modifications and their proposed periodate oxidized forms. (B) $tRNA^{Gln}$(TTG) and $tRNA^{Glu}$(TTC) known to contain 5-carboxymethylaminomethyl-2-thio-U ($cmnm^5s^2U$) 34. (C) $tRNA^{Arg}$(ACG), $tRNA^{Arg}$(CCG), $tRNA^{Arg}$(CCT), $tRNA^{Arg}$(TCT), and $tRNA^{Ser}$(GCT) known to contain 2-thio-C (s^2C) at position 32. A known modification that also produces signatures is I34 in $tRNA^{Arg}$(ACG). (D) $tRNA^{His}$ known to contain 4-thio-U (s^4U) at position 8. (E) s^2C32 response to 2,2'-dipyridyl (DIP), hydrogen peroxide (H_2O_2), and methyl α-D-glucopyranoside (αMG), biological replicates are shown in each graph. ***$P < 10^{-3}$, ns: not significant.

To examine the biological response of 2-thiomodifications, we subjected *E. coli* to the 2,2'-dipyridyl (DIP) stress, which chelates Fe^{2+}, oxidative stress with H_2O_2, or glucose starvation with αMG and performed tRNA-seq of the control and stressed samples. The mutation rates for the cmnm^{5}s^{2}U34 modified tRNA^{Gln}(TTG) and tRNA^{Glu}(TTC) did not change under these stress conditions. In contrast, the s^{2}C32 levels in all tRNAs were reduced in the DIP stress but not in H_2O_2 or αMG stress (Figure 4.16(e)). 2-Thio-C modification is installed by the enzyme TtcA which contains an iron-sulfur cluster in the active site [197], the s^{2}C32 reduction under iron chelation is consistent with a reduction of the TtcA activity. It remains to be determined whether the reduction of s^{2}C32 level affects decoding of specific codons (CGN and AGN) read by these modified tRNAs.

4.2.4 Microbiome characterization by small RNAs indicates global coordination of translation regulation

Efficient small RNA-seq has an immediate impact on revealing new insights into translational regulation of microbiomes. Applying MSR-seq to tongue scrapes from 8 human individuals, we obtained on average 6.1 million mapped tRNA reads among all samples. We simultaneously examined the expression of tRNA and two other small RNA families, the signal recognition particle (SRP) RNA and 5S rRNA (Figure 4.17(a), Figure 4.18(a)). SRP RNA is required for translating a subset of cellular mRNA of membrane proteins and secreted proteins [271], while 5S rRNA is an integral component of the ribosome (Figure 4.17(b)). The abundance among each pair of tRNA, SRP RNA, and 5S RNA was proportional (Figure 4.17(c-e)), suggesting globally coordinated expression of these essential components of the translation machinery. The coordinated expression explained the similarities of tRNA, 5S rRNA or SRP RNA based

taxonomies, while the RNA transcript-based taxonomy was qualitatively identical, but quantitatively distinct from the 16S rRNA gene amplicons (Figure 4.18(b)). A major feature of MSR-seq is the identification of tRNA modifications within the microbiome. We previously analyzed N^1-methyladenosine (m^1A) tRNA modification in the gut microbiome [44]; the same analysis here also identified tRNA m^1A in several bacterial taxa on the tongue (Figure 4.18(c)).

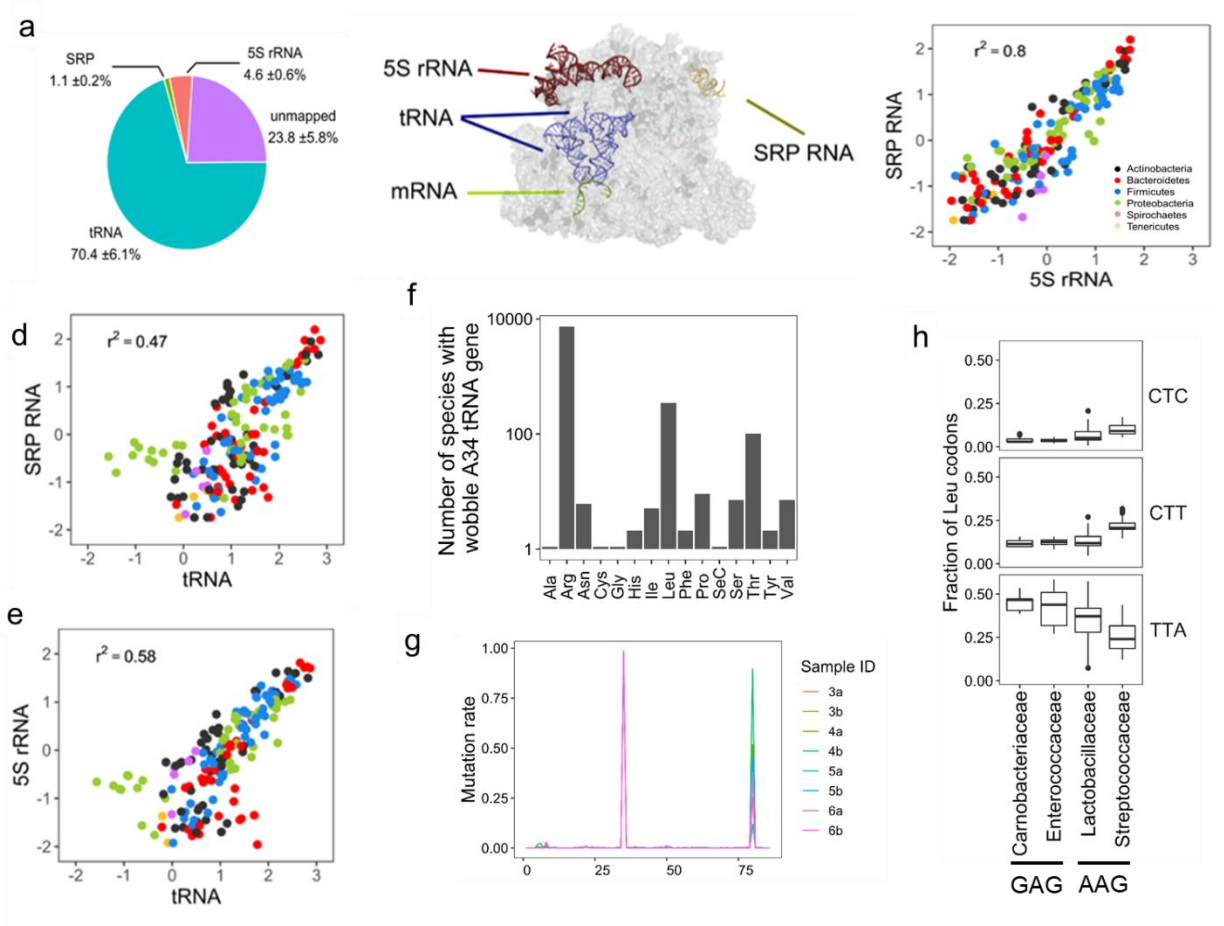

Figure 4.17: Microbiome analysis of multiple RNA families and inosine modification. MSR-seq of tongue scrapes was performed with n = 8 individuals. (a) Averaged read content among RNA families. (b) 3D structure of the *E. coli* 70S ribosome (PDB: 5GAF [290]) showing the location of tRNA, 5S rRNA, and signal recognition particle (SRP) RNA [235]. (c-e) Correlation of expression between 5S-SRP RNA (c), tRNA-SRP RNA (d), and tRNA-5S rRNA (e) at the taxonomic order level. Colors indicate orders from different phyla indicated in (c). (f) Number of bacterial species with annotated tRNA genes containing A34 wobble anticodon from genomic tRNA database [255]. (g) Mutation rate of $tRNA^{Leu}$(AAG) in streptococcus of all 8 samples. The high mutation rate at ~35 corresponds to wobble A-to-I modification. Variable mutation rates at position ~80

Figure 4.17, continued: correspond to the varying amounts of different streptococcus species among these individuals. (h) Leucine codon shifts for the bacterial families belonging to the same order containing either Leu(AAG) or Leu(GAG) tRNA genes.

4.2.5. Inosine is present in Streptococcus tRNA^{Leu}(AAG) & Leu(IAG) is correlated with evolutionary divergence in codon usage

To reveal new aspects of translational regulation, we investigated the wobble anticodon A34-to-inosine (I) modification in the microbiome. Eukaryotic tRNAs contain many genomic A34 tRNA genes that are modified to inosine in tRNA transcript to expand decoding [16,272]. In prokaryotes, the only characterized A34-to-I case is the Arg(ACG) tRNA gene. Our survey of the annotated tRNA genes in the genomic tRNA database [255] showed that among the 4037 bacterial genomes, Arg(ACG) was present in 3698 (92%) genomes. We also found A34 in many other tRNA genes, the most pronounced of which was Leu(AAG) in 349 (8.6%) bacterial genomes (Figure 4.17(f)). Intriguingly, the Leu(AAG) tRNA genes were restricted to several closely related taxonomic families whose tRNAs were also expressed at high levels on human tongue, and there was mutual exclusivity between Leu(AAG) and the more commonly present Leu(GAG) tRNA genes in bacterial genomes (Figure 4.18(d,e)). The A34 nucleotide in Streptococcus tRNALeu(AAG) transcript was mutated to G in all individuals, consistent with A34-to-I modification in the tongue microbial tRNA (Figure 4.17(g)). As a control, the same tRNALeu(AAG) transcript data showed a wide variant distribution at another nucleotide location which was derived from different Streptococcus species among the individuals.

We analyzed genomic leucine codon usage to associate a potential role of Leu(AAG) tRNA genes in translational regulation. We restricted our analysis to the genomes of 2 families each with Leu(AAG) or Leu(GAG) tRNA gene within the same taxonomic order, all are A/T-rich and have very similar G/C content (Figure 4.18(f)). We observed a substantial increase of the CTT and CTC

codons in the Leu(AAG)-containing families at the expense of the TAA codon, while CTA/CTG/TTG codon usages remained roughly the same (Figure 4.17(h), Figure 4.18(g)). Since CTT/CTC codons are read by both $tRNA^{Leu}$(IAG) and $tRNA^{Leu}$(GAG), this result suggests that $tRNA^{Leu}$(IAG) may efficiently decode CTT/CTC codons than $tRNA^{Leu}$(GAG).

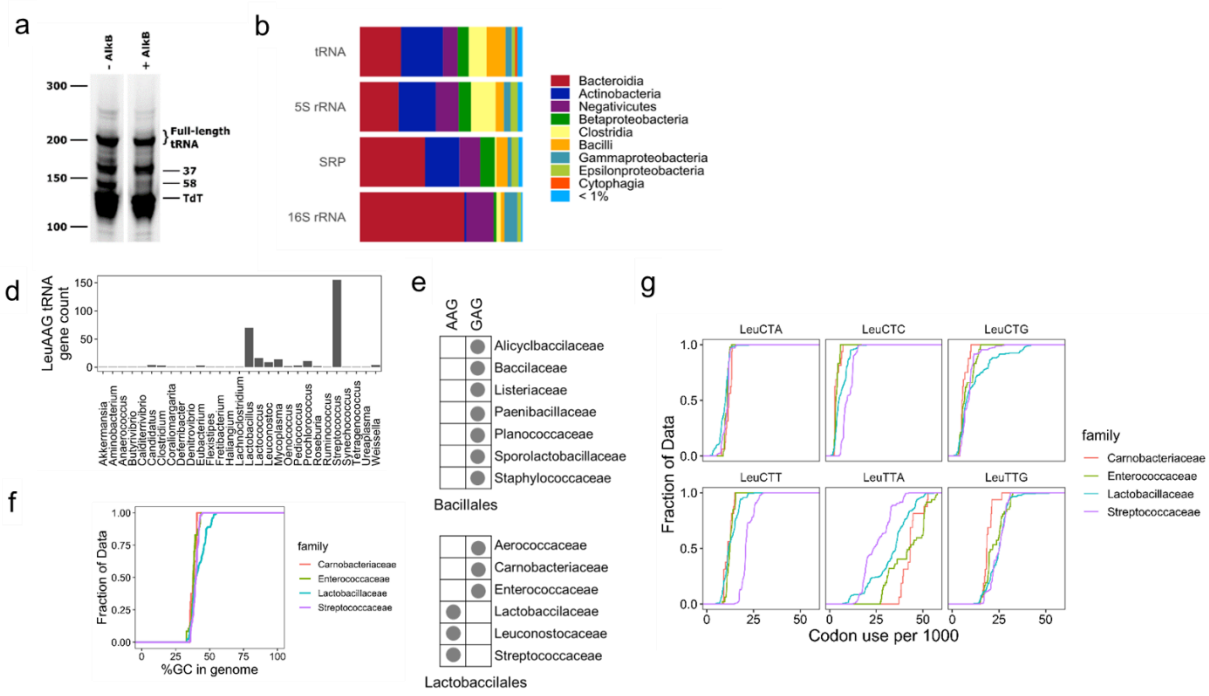

Figure 4.18: More microbiome MSR-seq analysis. (a) Final PCR product of MSR-seq libraries of the tongue scrapes, with and without demethylase treatment. Each lane contains 8 sequencing libraries. (b) Comparing tRNA, 5S rRNA, and SRP RNA based taxonomy to those of 16S rRNA gene amplicons at the bacterial class level. (c) m^1A identification at position 22 in the tongue scrapes. Heatmap shows different mutation rates in samples with and without demethylase treatment. (d) Bacterial Leu(AAG) gene counts from the genomic tRNA database. For the two genus with highest gene counts, Streptococcus is under Streptococcaceae, and Lactobacillus under Lactobaccilaceae. (e) The presence of annotated Leu(AAG) and Leu(GAG) tRNA genes in the orders of class Bacilli. No genome in this class contains both Leu(AAG) and Leu(GAG) tRNA genes. (f) G/C-content of the genomes in 4 families of Lactobaccilales, 2 contain Leu(AAG) tRNA gene and 2 Leu(GAG) tRNA gene. (g) Leucine codon usage in the genomes of the same 4 families as (f). Shown are cumulative, individual Leu codons among the species in each bacterial family.

4.3 Discussion

In this Chapter, I present four projects that demonstrate the applicability of MSR-seq to diverse biological questions ranging from the identification of host small RNAs packaged into virions to transcriptome analysis of the microbiome, a field that has largely been intractable to omics methods other than metagenomics.

4.3.1 Summary of SARS-CoV-2 packaging of host small RNAs (4.2.1.) & viral sequencing and modifications (4.2.2.)

In this work, we identified selective enrichment of host tRNAs and SRP RNA in cell free SARS-CoV-2 viral preparations, differences in tRNA modification between the tRNA in viral preparations and in cells, and candidate modification sites in the packaged SARS-CoV-2 genomic RNA. We estimate that a SARSCoV-2 virion contains up to one molecule of SRP RNA. Given the roughly equivalent read counts of SRP and tRNA in the viral preparation (Figure 4.1(b)), and full-length tRNAs are approximately one fourth in size to SRP RNA, we estimate that a SARS-CoV-2 virion also contains up to four tRNA molecules.

How and why specific tRNAs and the SRP RNA are enriched in virions remains unclear. Packaging of the HIV primer $tRNA^{Lys}$(TTT) is facilitated by the host lysyl-tRNA synthetase (LysRS) and gag protein interaction [58]. As a consequence, $tRNA^{Lys}$(CTT) is also packaged at similar levels. In our case, however, $tRNA^{Lys}$(CTT) was not enriched in the virions.

On the other hand, the enrichment of multiple $tRNA^{Ser}$ isoacceptors may go through this mechanism of facilitating tRNA packaging through host seryl-tRNA synthetase (SerRS) and SARS-CoV-2 protein interactions. Retroviruses also package a large amount of SRP RNA into virions [57,273]. Our results add SARS-CoV-2 to this list. Cells release tRNA fragments into

extracellular medium in many ways [274]. tRNA modifications influence tRNA fragment biogenesis, and the secreted tRNA fragments often match the abundance profiles of those fragments in cells. For the $tRNA^{Glu}$(TTC) fragment, its levels and m^1A58 modification profile match in the viral preparations and in cells. However, the full-length $tRNA^{Leu}$(AGA) and $tRNA^{Lys}$(TTT) in the viral preparations have higher m^1A58 levels than their counterparts in cells. m^1A58-modified tRNA can interact differently with cellular components such as eEF1A compared to the hypo-modified tRNA [24]. The higher tRNA m^1A58 level in the viral preparations may be related to their enhanced interaction with viral proteins. We also identified candidate modifications in the SARS-CoV-2 genomic RNA. It is surprising that none of the five modification sites described here could be readily assigned to well characterized modifications, just like those sites reported by nanopore sequencing [275]. Viral RNA modification studies have only taken off recently and future studies will be needed to reveal the chemical nature and the functional consequences of these modifications.

This proof-of-concept study was performed using the viral isolates cultured early in the pandemic (March/April 2020). In future work, we will produce larger amounts of SARS-CoV-2isolates representative of the early circulating strains as well as the viral variants of concern that have dominated most of the pandemic in 2021. We will culture these isolates on ACE2-expressing human cells rather than on African green monkey VeroE6 cells. These studies will firmly establish the dependency of host RNA packaging on SARS-CoV-2 variants and on host cell source.

4.3.2 Summary of queuosine & 2-thio modification detection and analysis

In summary, we found that periodate treatment of total RNA samples before reverse transcription enables the detection and semi-quantitation of Q-modified tRNAs through periodate-dependent

deletion signature in high throughput sequencing. Even though the deletion fractions are not high, the sequencing coverage of tRNA can easily reach hundreds to thousands per nucleotide position so that even glyco-Q34 and mitochondrial Q-modifications can be assessed. Although Q-modification can be precisely quantified using boronate affinity electrophoresis (APB gel) or LC/MS, these methods cannot simultaneously measure the effects of Q-modification on other tRNAs in the cell to study tRNA fragment biogenesis and linkage to the tRNA abundance and modification at the transcriptome level. These previously hidden, Q-modification dependent properties can now be assessed using the PAQS-seq approach.

We also found that periodate treatment of RNA samples enables detection of 2-thio-modifications in tRNA by high throughput sequencing. Like Q-modification, the ability to analyze 2-thio-modifications together with their association with other tRNA properties in the cell should reveal new biological insights on the tRNA transcriptome. At this time, we do not have a calibration curve of 2-thio-modifications for more precise quantitation, although these can be readily obtained upon chemical synthesis of oligonucleotides containing these modifications. Nevertheless, the ability to study these modifications by sequencing when comparing biological samples from different conditions can already examine the biological response involving these modifications. The 2-thio-modification generated mutation fraction in sequencing is sufficiently high to enable its application to identify new modification sites in low abundant mRNAs.

Chemical and enzymatic treatment of RNA samples before RT reaction has proven to be an extremely versatile tool for RNA modification studies [220,276]. We now add periodate treatment to this list for the studies of Q and 2-thiomodifications. As periodate treatment is part of the procedure to quantify tRNA charging by sequencing [152,218], Q and 2-thio-modification detection and semi-quantitation are already present in these sequencing data. Both deletion and

mutation have low background in the standard sequencing results so that these modifications can be studied at high sensitivity.

4.3.3 Summary of microbiome characterization & Streptococcus tRNA^{Leu}(AAG) deamination & evolutionary divergence of tRNA genes and codon usage

We analyzed multiple RNA families required for translation in microbiomes and found a strong correlation of their expression, thus enabling independent, multiple taxonomic assignments using tRNA, 5S rRNA and SRP RNA. Searching for a new aspect of translational regulation in the microbiome, we found phylogenetic clustering of tRNA^{Leu}(AAG) genes whose transcripts were modified to inosine in the microbiome. Importantly, the tRNA^{Leu}(IAG) modification is strongly associated with the unique leucine codon usage of these bacterial taxa present at high levels in humans, suggesting a fine-tuned translational regulation through wobble inosine modification in the human microbiome that has only been investigated so far in eukaryotes.

4.4 Materials & Methods

Note: The one-pot deacylation and β-elimination, standard tRNA deacylation, and MSR-seq protocol were performed as described in the Materials and Methods section of Chapter 2.

4.4.1 Isolation of SARS-CoV-2 clones

Residual nasopharyngeal swab specimens were collected after completion of the diagnostic process as part of the Mount Sinai Pathogen Surveillance Program. To culture SARS-CoV-2 isolates, .1 ml of viral transport media was inoculated into one well of a six-well plate seeded with a confluent monolayer of VeroE6 cells. Culture supernatants were harvested when CPE

(cytopathic effect) became visible, aliquoted, and stored at −80°C. All work related to SARS-CoV-2 culture was performed in a BSL3 biocontainment facility by trained personnel and in accordance with the research registration approved by the Institutional Biosafety Committee (IBC).

We cultured six distinct SARS-CoV-2 isolates representing the early lineages of the pandemic (Table 1). After isolation of the clinical isolates on VeroE6 cells, we determined the infectious viral titers for each of the viral culture supernatants by plaque assay (Table 1). All six isolates displayed medium plaque phenotypes. We shared aliquots of the viral stocks analyzed in this study with the NIH BEI repository in the early summer of 2020 (see [2] for specifics).

4.4.2 RNA isolation

The viral culture supernatants were spun at 3,000 rpm for 10 min to remove particulates. vRNA from each viral preparation and total RNA from VeroE6 cells was performed using QIAamp Viral RNA Kits (Qiagen) following the manufacturer's instructions. We did not filter the supernatants as that could result in some viruses being absorbed to the membrane and/or causing a loss of viral infectivity.

4.4.3 MSR-seq of virion-packaged small RNAs

MSR-seq was applied as described in the Materials and Methods second of Chapter 2. In brief, 10 µL of of total RNA from total RNA viral culture supernatants or uninfected cells (containing up to 1 µg of RNA, as determined by NanoDrop) was deacylated. The deacylated material was then processed into cDNA using the MSR-seq protocol.

4.4.4 MSR-seq of viral RNAs

1 µg of total RNA from infected Vero cell culture preparations was diluted to 50 µl in microcentrifuge tubes. Zymo RNA Clean and Concentrator-5 columns (R1016, Zymo) were used to remove small RNAs ≤200 nt by following the manufacturer's protocol. The large RNA (>200 nt) was eluted in 18 µl RNase free water. PCR machine was preheated at 95°C for RNA fragmentation experiment. Eluted RNA was then transferred to PCR tubes, 2 µl RNA fragmentation buffer (E6150S, NEB) was added to each tube and mixed well. Samples were incubated at 95°C on PCR machine for 6 min (target fragmentation is 200–500 nt) followed by putting on ice immediately. 2 µl RNA fragmentation stop buffer (E6150S, NEB) was added to each tube and mixed well. Samples were transferred to microcentrifuge tubes and diluted to 50 µl. Zymo RNA Clean and Concentrator-5 columns were used to cleanup RNA (>200 nt). RNA was eluted in 8 µl RNase free water. 1 µl T4 PNK buffer and 1 µl 10 U/µl T4 PNK were added to each tube and mixed well. The samples were incubated at 37°C for 30 min to repair RNA fragment 3′ ends. The samples were spun down and incubated at 75°C for 10 min followed by immediately putting on ice to inactivate T4 PNK. All samples (~10 µl reaction mixture each) were then processed using the MSR-seq method. After pooling, the beads were then split into two equal parts for ± demethylase treatment. After demethylase treatment, the beads were split again to two parts for −CMC and +CMC treatment (1:1.5 ratio) [203]. The ± demethylase and ±CMC treated beads were then used to continue the library construction.

4.4.5 Sequencing analysis of virion-associated RNAs

Reference RNA sequences from *Chlorocebus sabaeus* that included non-coding RNA and tRNA were downloaded from Rfam database (https://rfam.xfam.org/, [259]). *C. sabeaus* tRNA

sequences from Rfam were processed through tRNAScan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/, Lowe and Chan, 2016), only sequences with high confidence (i.e., tRNAScan score ≥50) were used as reference. Following this, tRNA sequences were appended by adding CCA at their 3′ end as well as removing intron sequences. These processed *C. sabeaus* cytosolic tRNA, mitochondrial tRNA, and noncoding RNA sequences such as SRP from Rfam were combined with the Wuhan SARS-CoV-2 genome sequence (MN908947.3) to generate a custom reference database.

Raw reads following sequencing were designated reads 1 and reads 2 and were merged together using bbmerge.sh present within the bbmap package (https://github.com/BioInfoTools/BBMap), which results in merged fastq files. These merged fastq were converted to fasta file format using reformat.sh present within the bbmap package. These fasta files were aligned to our custom reference genome using bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml, [277]) with the following parameters: -f -p 10 --local —no-unal. The aligned reads were then used to determine RNA sequence abundance using custom python script. RNA modifications were detected based on aligned reads using samtools sort (http://www.htslib.org/, [253]) feature sort the reads in a bam file format. Then IGVtools count (https://software.broadinstitute.org/software/igv/igvtools) count feature was utilized to output wig files using the following parameters: -z 5 -w 1 -e 250 —bases. The resulting wig files were processed using a custom python script to identify nucleotide mutations as well as coverage of aligned reads.

4.4.6. HEK293T culturing conditions (± queuosine)

HEK293T cells were cultured with complete DMEM medium under normal conditions. 0Q HEK293T cells were obtained by culturing the cells with dialyzed FBS for certain passages, and 100Q HEK293T cells were obtained by treating 0Q cells with 1 M queuine for 24 hours (8). Briefly, HEK293T cells were grown in complete DMEM medium (Cytiva Hyclone SH30022.01) with 10% dialyzed FBS (Thermo Fisher Scientific 26400044) and 1% PenicillinStreptomycin (Thermo Fisher Scientific 15070063) to 80% confluency and passaged. TRIzol reagent (Thermo Fisher Scientific 15596026) was used to extract total RNAs at each passage by following the manufacturer's manual. Q levels in $tRNA^{His/Asn}$ were constantly examined at each passage by APB gel based Northern blot. Q modification fractions of $tRNA^{His/Asn}$ dropped to below detection after ~10 passages, these cells are designated as 0Q. 100Q cells were obtained by culturing 0Q cells to 60–80% confluency followed by incubation with 1 M queuine for 24 hours.

4.4.7. APB Northern blots

Northern blots were performed as previously described (12). Three g of total RNA was added to each microcentrifuge tube and diluted to 9 l with H_2O. 1 µL of 1M Tris–HCl (pH 9.0) was added to the tube with mixing followed by incubation at 37◦C for 30 min to deacylate tRNAs. 10 µL 2x RNA loading dye (8 M urea, 0.1 M EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) were added to each tube. All samples were loaded to a pre-run hand-cast 10% denaturing PAGE gel containing 0.5% (g/ml) acrylamidophenylboronic acid (APB). The gel was run at 18W for ~2–3 hours until the xylene cyanol band was ~1–2 cm to the bottom in the 4 °C cold room using 1× TAE buffer. The gel area containing the target tRNAs was saved and the slightly larger sized Hybond-XL membrane (GE Healthcare, RPN303S) was put on top of the gel to take the gel out of the plate with caution. Dry RNA transfer was then performed using a gel dryer (Bio-Rad, 1651745)

for 4 hours at 80 °C. The gel and membrane were separated by soaking in distilled water. The RNA was crosslinked to the membrane by UV for two times (254 nm, 1200 mJ). The membrane was then blocked for 30 minutes twice with hybridization buffer (20 mM phosphate, pH 7, 300 mM NaCl, 1% SDS) at room temperature. The membrane was incubated with 50 mL 3 pmol/ml biotinylated tRNA probes for 16 hours at 60 °C in the UVP Hybridizer Oven (Analytik Jena 95-0030-01) followed by washing with 50 mL washing buffer (20 mM phosphate, pH 7, 300 mM NaCl, 2 mM EDTA, and 0.1% SDS) for 30 minutes twice in the UVP Hybridizer Oven. The membrane was then incubated with streptavidin-HRP conjugate (Genscript M00091) in 30 mL hybridization buffer (1:5000–1:10 000 dilution) for 30 minutes at room temperature followed by three washes for 5 minutes each in 25 mL washing buffer. The membrane was then transferred to plastic wrap with the RNA-side facing up. Peroxidase detection reagent 1 and 2 (Bio-Rad 1705061) were mixed (0.1 ml per 1 cm2 membrane) and applied to the top of the membrane by pipetting. The membrane was incubated with the reagent mixture for 5 mintues in darkness. The membrane was then transferred to a new plastic wrap to remove extra detection reagent. The membrane was scanned using the ChemiDoc imaging system (Bio-Rad) and the data was analyzed using ImageLab.

The oligonucleotide probe sequences were:

$tRNA^{His}$: 5’-biotin-TGC CGT GAC TCG GAT TCG AAC CGA GGT TGC TGC GGC CAC AAC GCA GAG TAC TAA CCA CTA TAC GAT CAC GGC

$tRNA^{Asn}$: 5’-biotin-CGT CCC TGG GTG GGC TCG AAC CAC CAA CCT TTC GGT TAA CAG CCG AAC GCG CTA ACC GAT TGC GCC ACA GAG AC

4.4.8. E. coli *culture and RNA extraction*

E. coli MG1655 cells were grown in LB to a A600 of 0.4 before subjecting to the stress conditions. Cells were harvested by centrifuging 25 mL culture for 1 minutes at 12,000 rcf and decanting media. Mock treated cells, 25 mL, were left to grow for 10 minutes. Iron depletion stress was done by adding to 25 mL cells 2,2'-dipyridl (DIP) to 250 M final concentration for 10 min. Hydrogen peroxide stress was done by adding H_2O_2 to 25 mL cells to a final concentration of 0.5% for 10 min. Glucose phosphate stress was done by adding to 25 ml cells α-methyl glucoside-6-phosphate (αMG) to a final concentration of 1 mM for 10 minutes. Cells were resuspended in 0.5 mL ice cold lysis buffer (150 mM KCl, 2 mM EDTA, 20 mM HEPES pH 7.5) then flash frozen in liquid nitrogen. RNA was extracted by a hot acid-phenol protocol. Briefly, 0.5 ml of acid-buffer phenol (pH 4.5 citrate) was added to frozen samples. Samples were incubated in a heat block with shaking at 50 °C for 30 minutes. The aqueous phase was extracted for another round of phenol extraction and two rounds of chloroform extraction before ultimately precipitating with glycoblue, 300 mM sodium acetate, and 3 volumes of ethanol. Samples were incubated for 1 hour at -80 °C, then centrifuged at maximum speed (20,000 rcf) for 45 minutes to pellet RNA. Pellets were washed twice with 70% ethanol, then resuspended in water.

4.4.9. On-bead periodate oxidation

During library preparation, RNA samples that underwent standard tRNA deacylation were treated with periodate when immobilized on streptavidin beads. Beads were resuspended in final concentration of 50 mM $NaIO_4$, 0.1 M NaOAc/HOAc, pH 5, then incubated at room temperature for 30 minutes. Reaction was quenched with 10 µL of 1 M ribose for 5 minutes. Beads were washed twice with high salt wash buffer.

4.4.10. Preparation of queuosine calibration samples

The queuosine calibration samples were mixed using a combination of 0% queuosine (queuosine-depleted) HEK total RNA and 100% queuosine (queuosine-abundant) HEK total RNA to a final volume of 10 l. (Queuosine modification levels were quantified by Northern blot.) The calibration samples ranged from 0% queuosine to 100% queuosine in 10% intervals.

Samples on beads from the barcode ligation and multiplexing above were resuspended in 40 µL of deionized, autoclaved water and 10 µL of a 0.25 M $NaIO_4$, 0.5 M NaOAc/HOAc, pH 5 (final concentration: 50 mM $NaIO_4$, 0.1 M NaOAc/HOAc, pH 5) were added. The reaction proceeded at room temperature for 30 minutes and was quenched by addition of 10 µL of 1 M ribose for 5 min. After quenching, the samples were washed as stated for all other steps in the MSR-seq protcol.

4.4.11. Reverse transcriptase screen

Libraries were prepared as above with standard deacylation conditions for three replicates of 100Q samples, however each sample was used in two different ligation reactions with different barcodes. After ligation, samples were pooled into two groups, differing only by the barcode used for each replicate. One group was further treated with periodate on bead as described above, while the other was not treated. Next all samples were combined – use of 6 barcodes allows differentiation of the three samples with periodate treatment and the three without. This combined mixture was used for further library construction. Before reverse transcription, the sample was divided into six aliquots, each used for reverse transcription with a different RT enzyme. The enzymes used were AffinityScript (Agilent 600105), HIV-RT (Worthington Biochemical Corporation, LS05003), MuMLV (NEB M0253L), RevertAid (ThermoFisher EP0441), SuperScript III (ThermoFisher

18080093), and SuperScript IV (ThermoFisher 11756050). Each RT product was kept separate for the remainder of library construction and PCR amplified with a different index for sequencing.

4.4.12. Data analysis

Libraries were sequenced on Illumina Hi-Seq or NEXT-seq platform. First, paired end reads were split by barcode sequence using Je demultiplex with options BPOS = BOTH BM = READ 1 LEN = 4:6 FORCE = true C = false 6. BM and LEN options were adjusted for samples with a 3 nt barcode instead of 4, and for samples where the barcode is located in read 2. Barcode sequences are available on Github at https://github.com/ckatanski/Q paper. Next read 2 files were used to map with bowtie2 (26) with the following parameters: -q -p 10 –local —no-unal. Reads were mapped to curated list of non-redundant tRNA genes with tRNAScane score >40 for respective organisms (human and E. coli). Bowtie2 output sam files were converted to bam files, then sorted using samtools. Next IGV was used to collapse reads into 1 nt window. IGV output.wig files were reformatted using custom python scripts (available on GitHub at https://github.com/ckatanski/Q paper). The bowtie2 output Sam files were also used as input for a custom python script using PySam, a python wrapper for SAMTools [278](27) to sum all reads that mapped to each gene. Data was visualized with custom R scripts (available on GitHub at https://github.com/ckatanski/Q paper). ‘Reads per million’ normalization was calculated by dividing the number or reads mapped to a specific gene by the total number of tRNA-mapped reads in that sample, and scaling by a factor of 1 000 000. Unless otherwise stated, analysis was limited to genes and positions with read coverage >100 reads. For presentation, the position value of each tRNA gene was adjusted to match canonical tRNA numbering (anticodon in positions 34, 35, 36). For calibration curve, Origin was used to fit linear or semiology line of best fit using least squares regression and calculated r^2

statistics. Comparing change in this modification during stress, an unpaired two-sided Wilcox test (Mann–Whitney): ns indicates a P-values >0.05, * <0.05, **<0.01, *** <0.001.

4.4.13. Stool & oral (tongue scrape) microbiome sample collection

Oral Cavity: Tongue dorsum scrapings were collected from 1 female and 3 male volunteers (two samples per volunteer) on two consecutive days [A & B sample]. Sample collection used BreathRx Gentle Tongue Scraper (Philips Sonicare) and was performed in the morning hours prior to eating, drinking or performing oral hygiene. Starting as far back as possible on the tongue, the scraper was passed forward over the entire surface three sequential times. The scrapings were combined with 500-µl RNAlater Stabilization solution (Invitrogen, AM7020) and stored at −80 °C until extraction.

Gastrointestinal tract: Stool specimens were self-collected by 1 female and 1 male volunteer. Volunteers were provided with a commercial "toilet hat" stool specimen collection kit (Fisherbrand Commode Specimen Collection System; ThermoFisher Scientific, 02-544-208). Specimens were immediately transported to the laboratory (<1-hr) and thoroughly homogenized. 100-mg stool was transferred into a cryovial using a sterile spatula and 700-µl RNAlater Stabilization solution was added. Specimens were stored at −80°C until extraction.

Total RNA Extraction: RNAlater was first removed from tongue dorsum and stool samples by centrifugation at 17,200 rcf for 10 minutes at 4 °C. Pelleted material was lysed in 400 µL of 0.3M NaOAc/HOAc,10mM EDTA, pH 4.8 with an equal volume of acetate-saturated phenol:chloroform pH 4.5 (Invitrogen, AM9722). After addition of 1.0 mm glass lysing beads (Bio-Spec Products, 11079110) in a 1:1 ratio (bead:sample weight), samples were placed in a reciprocating bead beater (Mini-Beadbeater-16, Bio-Spec Products) for two 1-minute intervals on

maximum intensity. Samples were centrifuged at 17,200 rcf for 15 minutes at 4 °C before re-extraction and isopropanol precipitation of total RNA. Pellets were washed with 75% ethanol before resuspension in an acid-buffered elution buffer (10mM NaOAc, 1mM EDTA, pH 4.8).

4.4.14. 16S rRNA gene amplicon sequencing

16S ribosomal RNA (rRNA) gene amplicon sequencing was performed using the same acid-phenol extracted stool and oral (tongue scrape) samples used in MSR-Seq. 5 µl of input material was subjected to RNaseA (NEB, T3018L) treatment, per manufacturer's instructions, followed by sample clean-up using an Oligo Clean-and-Concentrator spin column (Zymo Research, D4060). Nucleic acids were eluted into 8 µl TE buffer, pH 8.0 and quantified using a Qubit fluorometer. The V4 hypervariable region of the bacterial 16S rRNA gene was amplified using the custom fusion primers [290]. Briefly, the fusion primers contained a 16S rRNA gene locus-specific sequence [515F-Y (forward, GTGYCAGCMGCCGCGGTAA) or 806RB (reverse, GGACTACNVGGGTWTCTAAT)], appropriate Illumina adapter (P5 or P7), and an 8-nt multiplexing index barcode. Dual-indexed PCR amplicons were generated in 25 µL reaction volumes with an amplification cocktail containing 0.4 U Phusion High-Fidelity DNA Polymerase (ThermoFisher, F-530L), 1x High-Fidelity enzyme buffer (includes MgCl2), 200 µM of each dNTP (NEB, N04475), and 0.2 µM of each primer. 1 ng template DNA was added to each PCR and a no-template control was run for each primer pair used. PCR amplification conditions were as follows: initial denaturation step at 98 °C, 30s; 35 cycles of 98 °C for 10s, 56 °C for 30s, and 72 °C for 60s; with a final extension at 72 °C for 5 min. For stool library preparation, each sample was amplified in triplicate to generate three uniquely barcoded amplicon libraries per sample. In contrast, only a single amplicon library per sample was generated for oral samples. Each DNA

amplicon was first analyzed on a 1.5% (w/v) agarose gel, along with its corresponding No Template Control (NTC), to confirm amplification. The barcoded DNA amplicon libraries were pooled by combining 10 µL of each sample. After running on a 1.5% agarose gel, the pool was excised and purified for sequencing using the GeneJet Gel Extraction Kit (ThermoScientific™, K0701) following the manufacturer's protocol, eluting DNA into 50-µl of buffer EB.

To prepare our normalized amplicon libraries for sequencing, the MiniSeq Denature and Dilute Libraries Guide (Protocol A, Illumina 2019) was followed. Stool and oral libraries were sequenced in separate runs where the final sequencing library combined 250 µl of the denatured and diluted 16S library (1.8 pM) with 250 µl of denatured and diluted Illumina generated PhiX (1.5 pM). Libraries were sequenced by a 150-cycle paired-end run using an Illumina MiniSeq platform, a Mid-Output reagent kit and spiking in the custom sequencing primers designed by [290].

4.4.15. Microbiome tRNA analysis

The methods of the previously published pipeline 10 were used with some modifications. Raw paired-end sequence reads of 75 or 100 nucleotides were processed by Illumina-utils (available at https://github.com/merenlab/illumina-utils) 11. Inserts contained a 7-nucleotide sample barcode and a random 6 nucleotide unique molecular identifier (UMI). Given that tRNA molecules range in length from 74-96 nucleotides, forward and reverse 100 nucleotide reads fully covered some tRNA sequences and partially overlapped for others. We upgraded the Illumina-utils 'iu-merge-pairs' command to merge both fully and partially overlapping reads, while trimming overhanging adapter sequences in the case of more than full overlap (the command line flag, '--marker-gene-stringent', enables consideration of full as well as partial overlap). We minimized erroneous base

calls, important for our analysis of modification-induced mutations, by retaining reads that matched with zero mismatches in the overlapping region (option '--max-num-mismatches 0') 11.

The procedure for de novo profiling of full-length and fragmentary tRNA reads introduced in tRNA-seq-tools 10 was modified as follows. tRNA structural features were identified from the 3' end of the read. The minimum criteria for tRNA identification were the correct distance to the 7 canonically conserved nucleotides in the T arm, of which 5 must be found, and base pairing in of the T stem. A full-length read must contain a base-paired acceptor stem and all features in between.

tRNA sequences were taxonomically annotated by using the Global Alignment for Sequence Taxonomy (GAST) tool 12 to search a set of reference tRNA sequences that tRNAscan-SE (v1.3.1) 13 identified from 4,235 gold-standard bacterial genomes (non-endosymbiont genomes with an assembly level of "chromosome") stored in the Ensembl Genomes 2016 database 14.

Specific nucleotide positions were selected from tRNA sequences for modification analysis. Positions were identified relative to profiled structural features. For example, canonical position 22, a site of m^1A modification in many tRNA species, is identified as being 5 nucleotides from the 5' nucleotide of the anticodon stem, canonical position 27. The workflow analyzed the distribution of nucleotides at positions of interest in each taxon, grouping tRNA species by anticodon. tRNA species represented by at least 50 reads in both demethylated and untreated sample splits were selected. Mutations likely to be caused by modifications were separated from other sources of nucleotide variants, such as related tRNA sequences with a single nucleotide polymorphism, by only considering tRNA species with 2 other nucleotides beside the most abundant nucleotide in at least 5% of reads from the untreated split. A significantly reduced

mutation signature in the demethylated split confirmed the putative modification (χ^2 p-value < 0.001, from a χ^2 test comparing the observed numbers of the 4 nucleotides in the demethylated experiment to the expected numbers of the 4 nucleotides given the distribution from the untreated experiment).

4.4.16. Microbiome 16S rRNA analysis

To characterize microbial community composition in our samples through 16S rRNA gene amplicons, the program ‘iu-merge-pairs’ in illumina-utils 11 was used first to merge partially overlapping short reads. Minimum Entropy Decomposition 15 was then used, which is available in the oligotyping package v3 16, with default parameters to identify amplicon sequence variants (ASVs) across samples and used GAST 12 to assign taxonomic affiliation to each ASV.

4.4.17. Microbiome 5S rRNA analysis

Reference sequences for 5S rRNA were downloaded from the 5S rRNA database (http://combio.pl/rrna/) 17. Sequences were combined from Bacteria (n=7291), Archaea (n=319), Eukaryota (n=2861), mitochondria (n=110), and plastids (n=838). Full lineages were assigned to each reference using the ETE3 NCBITaxa toolkit in python. Sequencing reads were processed the same as E. coli reads but mapped to the combined 5S reference set using bowtie2 with the same mapping parameters as for E. coli. Mapping data were further processed as described for E. coli for base-wise mapping information as well as by-gene counting. These data were combined with SRP mapping data (below) by species. Mapping data for species within the same genus was summed. Data was then grouped by order or class, for Figures 4.17 and 4.18, respectively, then summed.

4.4.18. Microbiome SRP RNA analysis

Reference sequences for SRP RNA were obtained from Rfam [220]. Two Rfam families of sequence were combined: bacterial small SRP (RF00169) (n=5207) and bacterial long SRP (RF01854) (n=1143). Lineages for each reference sequence were assigned via entrez taxonomy lookup. Reads were mapped and processed as for 5S rRNA, described above using the SRP reference sequences.

Chapter 5

Small RNA Sequencing Combined with Structural Mapping Enables the Detection of Changes in Nucleotide Pairing

Acknowledgment: This Chapter is derived from an ongoing project in the Pan lab. The contributors of the data discussed here are: Christopher P. Watkins, Noah Peña, and Tao Pan. Scientific contributions: Conceptualization, C.P.W., N.P., and T.P.; Methodology: C.P.W.; Software: C.D.K.; Formal Analysis, N.P.; Investigation, C.P.W. and N.P.; Supervision, T.P.

5.1 Introduction

Based on the robustness of MSR-seq in detecting epitranscriptomic modifications and the vast biological applications our lab has employed it towards, we proposed to integrate existing RNA structural mapping techniques with MSR-seq. Our motivation was based on our previous findings detailed in Chapters 2-4. In brief, small RNAs are regulated and regulate based on the conditions and needs of the cell. To make the necessary changes, there are processes that will alter structure and the interactions of the small RNAs. We see this clearly in the polysome profiles of arsenite-treated cells, where the polysome fraction is almost entirely lost. Further, we observed an enrichment of specific tRNAs on the remaining polysomes and a reduction of many others. This almost certainly extends to other classes of small RNAs, including tRFs. Thus, cells alter not only their coding transcripts in response to changing conditions, but the usage and composition of their entire transcriptomes and the interactions of transcripts with other biomolecules – all of which is a rich source of information that many groups wish to explore. These changes in biomolecular interactions will result in changes in the exposure of nucleotides to their compartmental media, which can be monitored by treatment with highly reactive chemicals.

Structural mapping techniques rely on the use of chemical agents that modify unpaired nucleotides or crosslink nucleotides [279-281]. Typically, these methods extrapolate secondary structure of mRNAs from the pattern of modified (unpaired) and unmodified (base-paired) nucleotides [280-283]. However, these chemical agents will also be unable to modify nucleotides that are paired with residues of proteins. Since tRNA structures are largely invariant, we hypothesized that we could detect shifts in small RNA-protein interactions and, combined with cellular perturbations and computational prowess, predict the specific interactions between individual small RNAs and proteins. Because of the diverse regulatory functions of small RNAs, this strategy could provide detailed information on the biological shifts that accompany changes in cellular status and potentially predict new biologically relevant interactions between proteins and small RNAs.

In this Chapter, I describe the adaptation of *in vivo* dimethyl sulfite (DMS)-seq mapping method to MSR-seq [284,285]. I performed this using unstressed and arsenite-stressed HEK293T, as we have already characterized these cells via MSR-seq profiling of the total cellular RNA and polysome-associated RNA using the same conditions. I also outline our unique ability to differentiate DMS-induced modifications from endogenous epitranscriptomic modifications using MSR-seq in a biologically relevant context. Lastly, I present the nascent results of our investigation and detail the future steps to be taken with this project.

5.2 Results

Using the arsenite condition detailed in Chapter 3, I stressed HEK293T cells followed by a 5-minute incubation with 5% v/v DMS, with a no-DMS control for comparison. This concentration was selected because previous DMS structural mapping methods observed that 5% DMS in yeast

and human cells resulted in the labeling of many sites, allowing the exposure of many nucleotides on the same transcript to be determined by RT mutation signatures [284,285]. DMS is known to produce m^1A and m^3C methylations on unpaired A and C nucleotides [283]. To control for the effects of arsenite stress, I used unstressed HEK293T cells that were also incubated with 5% v/v DMS, also with a no-DMS control. As a further control, I extracted HEK293T total cellular RNA and then subjected 4 µg of the extracted RNA to 5% v/v DMS treatment *in vitro* with the same no-DMS control. Unique to our approach to this question was the inclusion of ± AlkB treatment on-bead to confirm mutation signals as arising from DMS-induced modifications, as DMS preferentially produces m^1A and m^3C methylations.

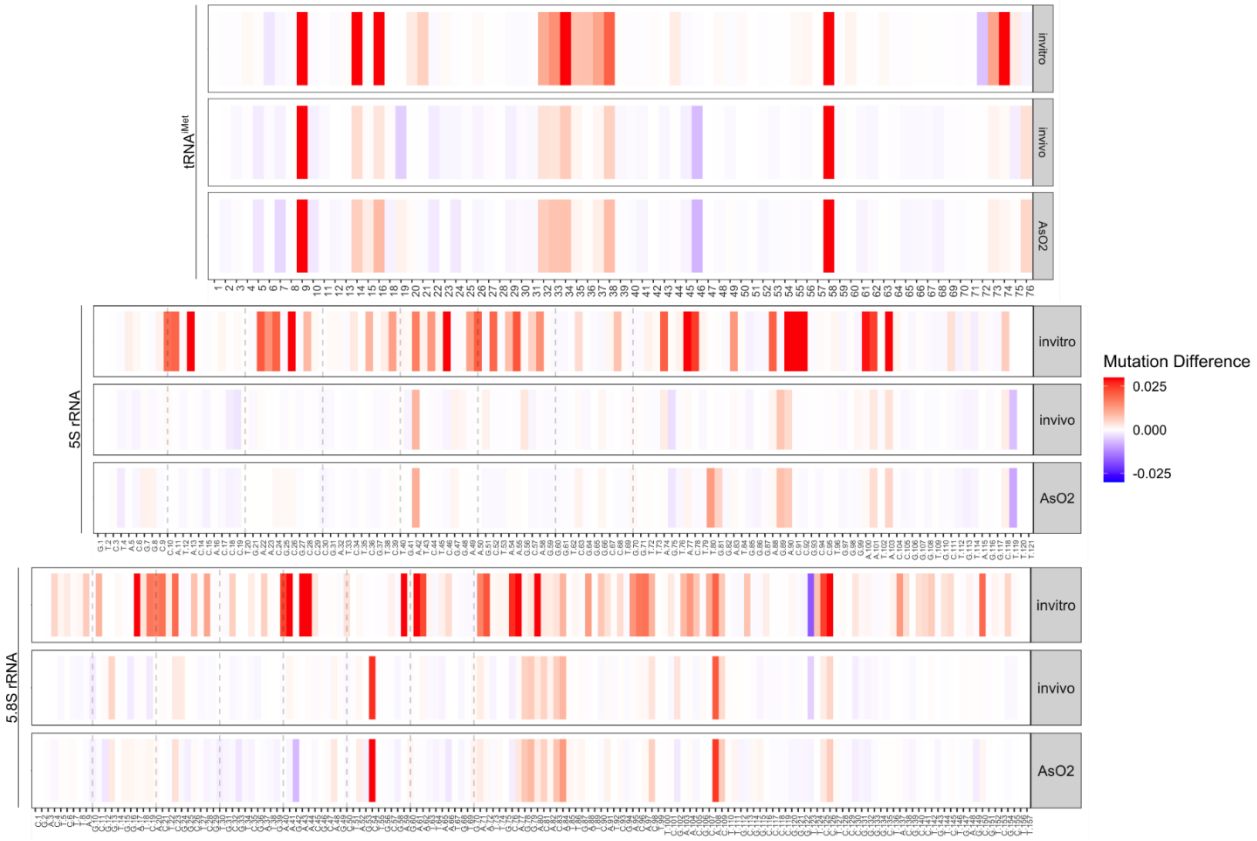

Figure 5.1: Heatmaps of $tRNA^{iMet}$, the 5S rRNA, and 5.8S rRNA. Positive Mutation Difference (red) indicates increased exposure of the nucleotide to DMS labeling.

Following MSR-seq library construction, we began investigating the differences between the *in vitro* (DMS labeling of deproteinized RNA), *in vivo* (DMS labeling of RNA in unstressed HEK293T cells), and AsO_2 (DMS labeling of RNA in arsenite-treated HEK293T cells) samples. We selected $tRNA^{iMet}$, 5S rRNA, 5.8S rRNA, and the U1, U2, U4, and U5A snRNAs as a starting point for our investigation, as these RNAs are well characterized in terms of their structure, interactions with proteins, and, in the case of tRNA and 5S, known sites of tRNA modifications. We developed an initial metric called the mutation difference, with the following equation:

$$MD_{Treatment} = \left(\frac{(Rep.1 + Rep.2 + Rep.3)_{-DM}}{3}\right) - \left(\frac{(Rep.1 + Rep.2 + Rep.3)_{+DM}}{3}\right)$$

where $MD_{Treatment}$ is the Mutation Difference of the treated samples (*in vitro*, *in vivo*, or AsO_2), Rep. # refers to the biological sample replicate, -DM means the sample was not treated with AlkB demethylase during library construction, and +DM means the sample was treated with AlkB demethylase during library construction. We restricted our subsequent evaluation to A and C nucleotides based on the reports known types of methylations detected using DMS labeling. Although there is evidence that G and U nucleotides are methylated by DMS, the interpretation of the results based on the mutation signatures from RT is more fraught. Using our established sequencing pipeline [1,3], we observed several significant changes between these select small RNAs between the *in vivo* and AsO_2 samples, which correspond to DMS-induced methylations. Based on these observations, our immediate future steps will be to incorporate structural data to highlight sites of DMS methylation in context with the small RNAs interactions with proteins and RNAs, particularly complexes such as the ribosome and spliceosome.

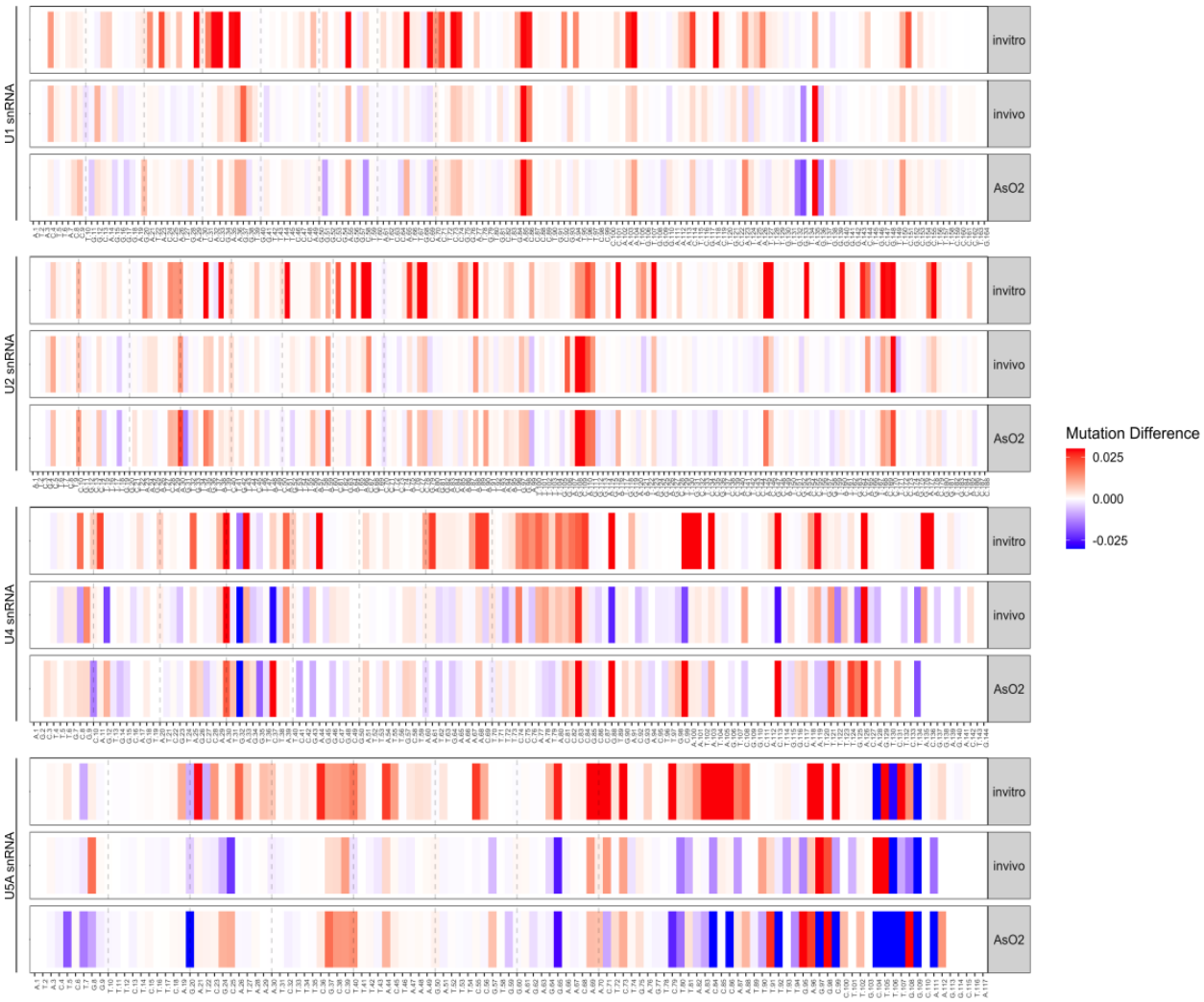

Figure 5.2: Heatmaps of the U1, U2, U4, and U5A snRNAs.

snRNA	Nucleotide (N,#)	$MD_{In\ vivo}$	$MD_{Arsenite}$
U4	C37	Decreased	Increased
	A39	Increased	Decreased
	C74	Increased	Decreased
	A77	Increased	Decreased
	A78	Increased	Decreased
	C99	Decreased	Increased
	A103	Decreased	Increased
	C108	Increased	Decreased
	C113	Decreased	Increased
	A119	Increased	Decreased
	A128	Decreased	Increased
	C131	Decreased	Increased
	A135	Increased	Decreased
U5A	C25	Decreased	Increased
	A88	Decreased	Increased
	A96	Decreased	Increased
	C104	Increased	Decreased
	A112	Decreased	Increased

Table 5.1: Mutation Difference at sites in U4 and U5A snRNA.

Despite the preliminary nature of our data, we observe significant differences between the *in vivo* and AsO_2 snRNAs U1, U2, U4, and U5A. These snRNAs are known to be regulators of mRNA splicing via the spliceosome. Interestingly, there are significant changes in the MD of U4 and U5A that differ between the *in vivo* control and arsenite-stressed cells, indicative of increased exposure to DMS by these nucleotides. In particular, the 3' region of the U4 snRNA is significantly more exposed to DMS labeling under conditions of oxidative stress (Table 5.1). Although it is too early to determine the role this change might have for the function, regulation, and use of these snRNAs, these data to point to altered interactions that may have significant biological significance in stress response, perhaps through regulating the splicing isoforms produced during arsenite stress.

5.3 Discussion

Although these results are preliminary, they demonstrate the modularity of MSR-seq as well as the potential to identify changes in RNA exposure that may correlate with global changes in small RNA usage by the cell. I anticipate that future results will greatly expand on our preliminary work. A long-term goal of this project is to be able to not only map protein-RNA interactions, but to predict interactions that have yet to be confirmed by methods such as crosslinking. This application may be especially amenable to studying changes in low abundance small RNAs, which may produce too few reads to by sequencing analysis downstream from crosslinking methods.

5.4 Materials & Methods

Note: The conditions for HEK293T culture and stress treatments, MSR-seq library construction, and read mapping were performed as described in Chapter 2 and 3.

5.4.1. Dimethyl sulfite labeling of unpaired nucleotides

To a 10 cm dish of HEK293T cells (media volume ~ 15 mL), 750 µL of DMS were added. After addition of DMS, cells were incubated at 37 °C for 5 minutes. The media was then aspirated off and the cells were washed with 10 mL of warm (~ 37 °C) PBS, which was then aspirated off. After the PBS wash, 5 mL of PBS were added to the cells, the cells were scraped off using a cell scraper, and added to a 15 mL Falcon tube. Cells were then pelleted at 3,000 x g for 3 minutes.

5.4.2. Total RNA extraction

After pelleting, the cells were resuspended in 1 mL of TRIzol and transferred to a 1.5 mL Eppendorf tube where the cells were homogenized. 200 µL of chloroform were then added to the

homogenized cells in TRIzol and allowed to separate for 3-5 minutes. The samples were then centrifuged at 12,000 x g for 15 minutes at 4 °C. The aqueous layer was then transferred to a new 1.5 mL Eppendorf tube and an additional 200 µL of chloroform and the aqueous was transferred to a new Eppendorf tube. 1 µL of GlycoBlue was added to the sample and a volume of isopropanol equivalent to the aqueous layer was also added.

The samples were then precipitated in a -80 °C freezer for at least one hour. Afterwards, the samples were centrifuged at 17,000 x g for 1 hour at 4 °C. After centrifugation, the supernatant was removed, and the remaining isopropanol was allowed to evaporate off for 10 minutes. Then, the RNA pellets were dissolved in doubly deionized, autoclaved water and stored in a -80 °C freezer for long-term preservation.

Chapter 6

Conclusions & Future Directions

6.1. Integrative Analysis of Small RNA Classes in Biological Studies & Disease

The vast diversity of small noncoding RNA classes seen throughout the three kingdoms of life are key regulators of nearly every biological process. The ability to study these RNA classes has advanced our understanding of biological regulation and even provided opportunities for developing therapeutic strategies. However, several RNA classes, including tRNA (the most abundant RNA class), pose substantial challenges to sequencing-based strategies, which have the potential to provide the most thorough data in a high-throughput manner. In Chapter 2 of this thesis, I detailed the development of a new small RNA sequencing method (MSR-seq) that can simultaneously report the abundance, modification fraction, and charging fraction (between samples) [1]. MSR-seq relies upon the Capture Hairpin Oligo I designed, which allows for ligation to the 3' ends of any RNA without any sequence bias. tRNA, having the most abundant number of transcripts, is the most likely to be ligated and sequenced; however, MSR-seq can also capture data from low abundance RNA classes such as tRNA fragments and, when enriched, others such as mRNA. Notably, MSR-seq is not significantly inhibited by many post-transcriptional modifications, allowing the simultaneous quantification of multiple microspecies of the same parent transcript. Furthermore, even when epitranscriptomic modifications do produce a stop, the read is not lost due to the staggered ligation approach.

In Chapter 3 of this thesis, I demonstrate the ability of MSR-seq to identify translational regulation by means of $tRNA^{Ser}$ C47d isodecoder enrichment on the polysome during arsenite

stress, enrichment of $tRNA^{Ser}$ C47d microspecies with m^3C32, but not m^3C47d, and stress-related codon usage patterns that are likely to slow the rate of translation in response to arsenite. MSR-seq also proves robust in identifying tRNA modifications that promote or inhibit tRF biogenesis. While these findings, combined with the robustness of the method, show the utility and promise of MSR-seq, there remains far more to explore even in data collected from cultured cells. In particular, sequencing methods that can report multiple RNA classes have yet to be applied to investigate the integration of the changes observed with these RNA classes to place the data in a much broader and comprehensive context. Studying the collective change in the transcriptome will offer a substantially greater level of insight into cellular adaptation to the environment and may well advance our ability to diagnose and treat human illnesses, as well as to respond to looming threats such as future pandemics and developing drought and heat resistant crops for regions heavily impacted by climate change.

To achieve this goal of a fully integrated transcriptomic analysis, multiple complementary methods will need to be employed, including small RNA sequencing. Further, this will likely require substantial improvements in spatially-resolved transcriptomics and single-cell small RNA sequencing in order to provide the resolution and biological contexts that these transcriptomic changes take place in.

6.2. Additive Effects of RNA Modifications on Small RNA Biology

tRNAs have, on average, between ~ 10-20% of their nucleotides post-transcriptionally modified. Although several of these modifications have been found to function as by stabilizing or altering tRNA secondary and tertiary structure, it has been difficult to evaluate the synergistic effects of these modification on the whole of the tRNA and the many potential biological roles they may

impact. Several factors have complicated this goal, including the lethality of mutating writers, the presence of redundant writers, incomplete knowledge of enzymes involved in the modification biosynthesis pathway, and the fact that writers and erasers modify a wide variety of tRNA substrates. Despite these challenges, significant progress has been made, especially for the additive effects of anticodon loop modifications [40,148]. Structural studies have shown the effects of anticodon loop modifications on both the structure of the anticodon loop and the ability of the anticodon to pair with its cognate codon [148]. Paired with biological studies and RNA sequencing, we have further been able to deduce that several anticodon loop modifications are pre- or co-requisite for other modifications in the anticodon loop [40,148]. Investigations into pathogenic mutations of mitochondrial tRNAs have shown that the lack of m^1A58 on mt-$tRNA^{Lys}$ prevents the installation of the τm^5s^2U modification of the wobble base, further contributing to the translational defect that is causative of MERRF [23].

Developments in small RNA sequencing have greatly improved our ability to address this problem. The most critical improvement for this goal has been the engineering of RTs and discovery of conditions that can read through and detect modifications without inducing a stop. These developments have given the field access to data on small RNA microspecies, i.e., small RNAs that have the same nucleotide composition, but differ in their modifications; this, in concert with applications such as polysome profiling, protein-RNA crosslinking, and structural mapping, allows the functional outputs of different small RNA microspecies to be determined based on their biological interactions. I have demonstrated in Chapter 3 that $tRNA^{Ser}$ C47d isodecoder microspecies with m^3C32 but lacking m^3C47d are significantly enriched on the polysome during arsenite stress. Cryo-EM shows that C47d is contacting the small ribosomal subunit, while C32 is near the anticodon. Because tRNAs interact with both the ribosome as well as the mRNA and since

these modifications are differentially regulated, it is likely that they are performing separate functions during translation. However, what is not known is how or by what means these modifications are regulated. Could there be pre- or co-requisite modifications for methylation of either C32 or C47d? Since there are many modifications that are invisible to MSR-seq, this scenario is plausible. Further, there are additional modifications that support the structure and function of $tRNA^{Ser}$ and likely impact its biological function. Because many of the body modifications are methylations, many of which can be dynamically altered, there are high numbers of microspecies that can be generated, each of which may be enriched in specific cellular processes.

Understanding the individual contributions of these modifications to the biological function(s) of small RNAs is important, but biological systems sense both the individual and collective properties of biomolecules. Thus, a future goal for investigations into small RNAs and the role played by their modifications should be a greater understanding of how the individual epitranscriptomic modifications contribute to the overall structure and function of the individual small RNA. Currently, the Pan lab is evaluating the crosstalk between tRNA modifications based on their rate of co-occurrence. Pairing this with the other previously mentioned methods promises to expand our understanding of this field.

6.3. Development of Treatments to Detect Additional Modifications

In Chapters 3 and 4 of this thesis, I demonstrated the ability of MSR-seq to representatively quantify modification fractions of the Watson-Crick face methylations m^1A, m^1G, m^3C, and $m^2{}_2G$, Ψ, Q, and several complex 2-thio modifications. The ease with which these modification detection steps can be integrated with the MSR-seq protocol will accelerate discoveries in this field

substantially. However, roughly ~ 140 post-transcriptional modifications do not have conditions that can readily detect them, whether enzymatic or chemical. An outstanding goal of the epitranscriptomics field is to detect, quantify, and understand the biological significance of each modification. To that end, an ongoing effort in the Pan lab is to identify chemical or enzymatic treatments to greatly expand the repertoire of RNA modifications that can be detected using MSR-seq.

6.4. Disease Diagnosis & Prognosis by Small RNA Abundance & Modification Fractions

Small RNAs have been observed in numerous biofluids and host small RNAs have been found in viruses, as studies from several decades have shown [57-60,286,287]. Furthermore, the composition of extracellular RNAs and intracellular RNAs are known to change during stress [94,100], defined here as any potentially harmful stimuli that requires the cell to shift from homeostasis to respond to the stimulus. Such changes include infection, environmental toxins, and especially the stresses often observed with cancer – proteotoxic stress, reactive oxygen species, and acidification of the tumor microenvironment. Changes in the abundance of these small RNAs has been observed in various disease states and during T cell activation [100], indicating a potential role for these small RNAs as diagnostics and prognostics. Preliminary work by myself, others in the Pan lab, and our collaborators is currently focused on evaluating the roles of small RNAs in fixed biopsy slides from patients with prostate cancer and colorectal cancer biopsy samples along with their stool microbiomes. Because MSR-seq is able to capture virtually any small RNA class from samples, we anticipate that we will be able to develop medical tools that enhance the detection and treatment of patients.

6.5. Analysis of Microbiome Activity & Its Impact on Health

In Chapter 4 of this thesis, I reported on the application of MSR-seq to the taxonomic characterization and translational regulation of the human oral microbiome on the tongue. Microbiomes are highly diverse and complex communities of microorganisms and have proven themselves to be intractable to many transcriptomics approaches [288,289]. This is due to a variety of reasons, including the diverse chemical composition of microbial membranes, the presence of both gram-positive and -negative bacteria, and general degradation of the mRNA during sample extraction. However, small RNAs are generally more stable than mRNA, and provide substantial information on translational regulation, as well as other cellular processes.

The most significant obstacle facing the analysis of microbiome small RNAs is no longer the generation of data, but the development of algorithms and computational pipelines and data analysis. Currently, we are working on a collaborative project to develop a pipeline for *de novo* tRNA discover, which will substantially aid our effects. Part of the difficulty is that tRNA phylogeny is itself difficult to evaluate given the ability of microbes to horizontally transfer genes. However, preliminary results from our collaborator show that some tRNAs are in fact specific down to the species level of taxonomy. Although this level of taxonomic specific is not necessary for our anticipated goals, it is a demonstration that tRNAs can assigned to specific taxa. In addition to the computational developments, innovations in LC-MS/MS methods will improve our ability to validate post-transcriptional modifications, although epitranscriptomic analysis of microbiome tRNAs, let alone other small RNAs, is still a long way off. Despite the challenges that remain for characterizing this enigmatic but critical aspect of biology, innovations such as MSR-seq and developments in computational analysis make this a possibility rather than a theoretical exercise. Given the abundant evidence that human microbiomes impact many diverse parts of our

physiology, the potential to make substantial contributions to this field and to understand it regulation and interactions with host cells is an incredible opportunity.

6.6. Conclusion

The work I have presented in this thesis spans over many biological niches but is united by the development and application of the high-throughput small RNA sequencing method MSR-seq. In my study of the regulation of small RNA responses during cellular stress, I demonstrated that MSR-seq reports on the changes in tRNA RNA abundance and modification fractions, providing potential insight into the role of epitranscriptomic modifications at varying positions. I also showed proof that the pairing of suboptimal tRNAs and codons is employed more broadly during translation under conditions of oxidative stress by arsenite. Next, I observed that different modifications impact the rate of tRF biogenesis, with some positively and some negatively correlating.

Subsequent investigations identified SARS-CoV-2 virion packaged host small RNAs, as well as viral RNA modifications. In keeping with previous studies of virion packaged host RNAs, we found several tRNAs, a possible tRNA fragment, and the human SRP RNA contained within SARS-CoV-2 virions cultured in Vero E6 cells. The overlap in the identified small RNAs between HIV-1 (a retrovirus) and SARS-CoV-2 (an RNA virus) may point to a function of the packaged RNAs in evading the innate immune system or providing other advantages for the infectious particle. We further discovered epitranscriptomic modifications of the SARS-CoV-2 viral RNA and identified parts of the modification. Our investigation into the microbiome also demonstrated the prowess of MSR-seq in characterizing a complex community of organisms, discovering potential coordination of expression of small RNAs that regulate translation, and identifying a new

context for inosine modification in prokaryotes and its relationship to the evolutionary divergence of codon usage by microorganisms.

Additionally, I developed a condition that significantly expands the number of modifications that can be detected by RT signature, which is critical for Illumina sequencing applications to study epitranscriptomic modifications. In particular, the Pan lab is interested in the detection of queuosine by high-throughput small RNA sequencing given the role it is shown to have in the prevention of numerous maladies such as cancer and neurodegeneration.

MSR-seq has the potential to transform the field of small RNA biology and thus the many biological niches that it touches upon. Its high-throughput nature and its adaptability with many other biochemical methods means that it can be quickly adopted by other groups and integrated in new and fascinating ways to produce new and unexpected insights into the regulation of translation, epigenetic reprogramming, virology, stress response, the microbiome, and innumerable other fields of study.

References

1. Christopher P. Watkins, Wen Zhang, Adam Wylder, Christopher D. Katanski, and Tao Pan. A multiplex platform for small RNA sequencing elucidates multifaceted tRNA stress response and translational regulation. *Nature Communications*, 13(2491): May 2022.

2. Noah Peña, Wen Zhang, Christopher P. Watkins, Mateusz Hałucha, Hala Alshammary, Matthew M. Hernandez, Wen-Chun Liu, Randy A. Albrecht, Adolfo Garcia-Sastre, Viviana Simon, Christopher Katanski and Tao Pan. Profiling Selective Packaging of Host RNA and Viral RNA Modification in SARS-CoV-2 Viral Preparations. *Frontiers in Cell and Developmental Biology*, 10: February 2022.

3. Christopher D. Katanski, Christopher P. Watkins, Wen Zhang, Matthew Reyer, Samuel Miller and Tao Pan. Analysis of queuosine and 2-thio tRNA modifications .by high throughput sequencing. *Nucleic Acids Research*: June 2022.

4. Tao Pan. Modifications and functional genomics of human transfer RNA. *Cell Research*, 28: 1-10, February 2018.

5. Paul Schimmel. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nature Reviews Molecular Cell Biology*, 19: 45-58, January 2018.

6. Irem Avcilar-Kucukgoze and Anna Kashina. Hijacking tRNAs From Translation: Regulatory Functions of tRNAs in Mammalian Cell Physiology. *Frontiers in Molecular Biosciences*, 7: December 2020.

7. Medha Raina and Michael Ibba. tRNAs as regulators of biological processes. *Frontiers in Genetics*, 5: June 2014.

8. Kiley Dare and Michael Ibba. Roles of tRNA in cell wall biosynthesis. *WIREs RNA*, 3: 247-264, January 2012.

9. Eric M. Phizicky and Anita K. Hopper. tRNA biology charges to the front. *Genes & Development*, 24: 1832-1860, September 2010.

10. G. M. Gongadze. 5S rRNA and Ribosome. *Biochemistry (Moscow)*, 76 (13): 1450-1464, July 2011.

11. Martin R. Pool. Signal recognition particles in chloroplasts, bacteria, yeast and mammals. *Molecular Membrane Biology*, 22(1): 3-15, April 2005.

12. Daewoo Paka, Robert Root-Bernsteinb, and Zachary F. Burton. tRNA structure and evolution and standardization to the three nucleotide genetic code. *Transcription*, 8(4): 205-219, June 2017.

13. Natalie Krahn, Jonathan T. Fischer, and Dieter Söll. Naturally Occurring tRNAs With Non-canonical Structures. *Frontiers in Microbiology*: October 2020.

14. Mathias Sprinzl and Konstantin S. Vassilenko. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research*, 33: D139-D140, January 2005.

15. Jeffrey M Goodenbour and Tao Pan. Diversity of tRNA genes in eukaryotes. *Nucleic Acids Research*, 34(21): 6137-6146, November 2006.

16. Magdalena A. Machnicka, Kaja Milanowska, Okan Osman Oglou, Elzbieta Purta, Malgorzata Kurkowska, Anna Olchowik, Witold Januszewski, Sebastian Kalinowski, Stanislaw Dunin-Horkawicz, Kristian M. Rother, Mark Helm, Janusz M. Bujnicki, and Henri Grosjean. MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Research*, 41: D262-D267, January 2013.

17. Pietro Boccaletto, Magdalena A. Machnicka, Elzbieta Purta, Pawel Piatkowski, Blazej Baginski, Tomasz K. Wirecki, Valérie de Crécy-Lagard, Robert Ross, Patrick A. Limbach, Annika Kotter, Mark Helm, and Janusz M. Bujnicki. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Research*, 46(D1): D303-D307, January 2018.

18. Pietro Boccaletto, Filip Stefaniak, Angana Ray, Andrea Cappannini, Sunandan Mukherjee, Elżbieta Purta, Małgorzata Kurkowska, Niloofar Shirvanizadeh, Eliana Destefanis, Paula Groza, Gülben Avşar, Antonia Romitelli, Pınar Pir, Erik Dassi, Silvestro G Conticello, Francesca Aguilo, and Janusz M Bujnicki. MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Research*, 50(D1): D231-D235, December 2021.

19. Wen Zhang, Marcus Foo, A. Murat Eren, and Tao Pan. tRNA modification dynamics from individual organisms to metaepitranscriptomics of microbiomes. *Molecular Cell*, 82: 891-906, March 2022.

20. Tsutomo Suzuki. The expanding world of tRNA modifications and their disease relevance. *Nature Reviews Molecular Cell Biology*, 22: 375-392, June 2021.

21. Basma El Yacoubi, Marc Bailly, and Valerie de Crécy-Lagard. Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. *Annual Review of Genetics*, 46: 69-95, August 2012.

22. Mark Helm, Richard Giegé, and Catherine Florentz. A Watson−Crick Base-Pair-Disrupting Methyl Group (m1A9) Is Sufficient for Cloverleaf Folding of Human Mitochondrial $tRNA^{Lys}$. *Biochemistry*, 38(40): 13338–13346, September 1999.

23. Uwe Richter, Molly E. Evans, Wesley C. Clark, Paula Marttinen, Eric A. Shoubridge, Anu Suomalainen, Anna Wredenberg, Anna Wedell, Tao Pan & Brendan J. Battersby. RNA modification landscape of the human mitochondrial $tRNA^{Lys}$ regulates protein synthesis. *Nature Communications*, 9: September 2018.

24. (a) Fange Liu, Wesley Clark, Guanzheng Luo, Xiaoyun Wang, Ye Fu, Jiangbo Wei, Xiao Wang, Ziyang Hao, Qing Dai, Guanqun Zheng, Honghui Ma, Dali Han, Molly Evans, Arne Klungland, Tao Pan, Chuan He. ALKBH1-Mediated tRNA Demethylation Regulates Translation. *Cell*, 167(3): 816-828, October 2016. (b) Jiangbo Wei, Fange Liu, Zhike Lu, Qili Fei, Yuxi Ai, P. Cody He, Hailing Shi, Xiaolong Cui, Rui Su, Arne Klungland, Guifang Jia, Jianjun Chen, and Chuan He. Differential m^6A, m^6A_m, and m^1A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Molecular Cell*, 71(6): 973-985, September 2018.

25. Danny D. Nedialkova and Sebastian A. Leidel. Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity. *Cell*, 161: 1606-1618, June 2015.

26. Shawn M. Lyons, Marta M. Fay, and Pavel Ivanov. The role of RNA modifications in the regulation of tRNA cleavage. *FEBS Letters*, 592: 2828-2844, August 2018.

27. Ian A. Roundtree, Molly E. Evans, Tao Pan, Chuan He. Dynamic RNA Modifications in Gene Expression Regulation. *Cell*, 169(7): 1187-1200, June 2017.

28. Ying Zhou, Jeffrey M. Goodenbour, Lucy A. Godley, Amittha Wickrema, Tao Pan. High levels of tRNA abundance and alteration of tRNA charging by bortezomib in multiple myeloma. *Biochemistry and Biophysics Research Community*, 385(2): 160-164, July 2009.

29. Clement T.Y. Chan, Yan Ling Joy Pang, Wenjun Deng, I. Ramesh Babu, Madhu Dyavaiah, Thomas J. Begley & Peter C. Dedon. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nature Communications*, 3: July 2012.

30. Kimberly A. Dittmar, Michael A. Sørensen, Johan Elf, Måns Ehrenberg, and Tao Pan. Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Reports*, 6(2): 151-157, January 2005.

31. Marc Torrent, Guilhem Chalancon, Natalia S. de Groot, Arthur Wuster, and M. Madan Babu. Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Science Signaling*, 11(456): September 2018.

32. Scott C. Blanchard, Harold D. Kim, Ruben L. Gonzalez, Jr., Joseph D. Puglisi, and Steven Chu. tRNA dynamics on the ribosome during translation. *Proceedings of the National Academy of Sciences*, 101(35): 12893-12898, August 2004.

33. Dmitri Graifer and Galina Karpova. Interaction of tRNA with Eukaryotic Ribosome. *International Journal of Molecular Sciences*, 16: 7173-7194, March 2015.

34. Drew Smith and Michael Yarus. tRNA-tRNA interactions within cellular ribosomes. *Proceedings of the National Academy of Sciences*, 86: 4397-4401, June 1989.

35. Marina V. Rodnina and Wolfgang Wintermeyer. Fidelity of Aminoacyl-tRNA Selection on ther Ribosome: Kinetic and Structural Mechanisms. *Annual Review of Biochemistry*, 70: 415-435, 2001.

36. Mikel Valle, Jayati Sengupta, Neil K. Swami, Robert A. Grassucci, Nils Burkhardt, Knud H. Nierhaus, Rajendra K. Agrawal, Joachim Frank. Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. *The EMBO Journal*, 21(13): 3557-3567, July 2002.

37. Francesca Tuorto and Frank Lyko. Genome recoding by tRNA modifications. *Open Biology*, 6(12): December 2016.

38. Albert Weixlbaumer, Frank V Murphy, IV, Agnieszka Dziergowska, Andrzej Malkiewicz, Franck A. P. Vendeix, Paul F. Agris, V. Ramakrishnan. Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nature Structural and Molecular Biology*, 14(6): 498-502, June 2007.

39. Namit Ranjan and Marina V. Rodnina. Thio-Modification of tRNA at the Wobble Position as Regulator of the Kinetics of Decoding and Translocation on the Ribosome. *Journal of the American Chemical Society*, 139(16): 5857-5864, April 2017.

40. Ann E. Ehrenhofer-Murray. Cross-Talk between Dnmt2-Dependent tRNA Methylation and Queuosine Modification. *Biomolecules*, 7(14): February 2017.

41. Chen-Yu Liu, Mohd Tanvir Qureshi, and Tae-Hee Lee. Interaction Strengths between the Ribosome and tRNA at Various Steps of Translocation. *Biophysical Journal*, 100: 2201-2208, May 2011.

42. Markus Duechler, Grażyna Leszczyńska, Elzbieta Sochacka, and Barbara Nawrot. Nucleoside modifications in the regulation of gene expression: focus on tRNA. *Cellular and Molecular Life Sciences*, 73: 3075-3095, April 2016.

43. Ha An Nguyen, Eric D. Hoffer, and Christine M. Dunham. Importance of a tRNA anticodon loop modification and a conserved, noncanonical anticodon stem pairing in $tRNA^{Pro}_{CGG}$ for decoding. *The Journal of Biological Chemistry*, 294(14): 5281-5291, April 2019.

44. Michael H. Schwartz, Haipeng Wang, Jessica N. Pan, Wesley C. Clark, Steven Cui, Matthew J. Eckwahl, David W. Pan, Marc Parisien, Sarah M. Owens, Brian L. Cheng, Kristina Martinez, Jinbo Xu, Eugene B. Chang, Tao Pan, and A. Murat Eren. Microbiome characterization by high-throughput transfer RNA sequencing and modification analysis. *Nature Communications*, 9: December 2018.

45. Sheree A. Wek, Shuhao Zhu, and Ronald C. Wek. The Histidyl-tRNA Synthetase-Related Sequence in the eIF-2a Protein Kinase GCN2 Interacts with tRNA and Is Required for Activation in Response to Starvation for Different Amino Acids. *Molecular and Cellular Biology*, 15(8): 4497-4506, August 1995.

46. Thomas D. Baird and Ronald C. Wek. Eukaryotic initiation factor 2 phosphorylation and translational control in metabolism. *Advances in Nutrition*, 3(3): 307-321, May 2012.

47. Karolina Pakos-Zebrucka, Izabela Koryga, Katarzyna Mnich, Mila Ljujic, Afshin Samali, and Adrienne M. Gorman. The integrated stress response. *EMBO Reports*, 17: 1375-1395, September 2016.

48. Kristoffer Skovbo Winther, Mohammad Roghanian, and Kenn Gerdes. Activation of the Stringent Response by Loading of RelA-tRNA Complexes at the Ribosomal A-Site. *Molecular Cell*, 70: 95-105, April 2018.

49. Matthew F. Traxler, Sean M. Summers, Huyen-Tran Nguyen, Vineetha M. Zacharia, G. Aaron Hightower, Joel T. Smith, and Tyrrell Conway. The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Molecular Microbiology*, 68(5): 1128-1148, June 2008.

50. Kiel D. Kreuzer and Tina M. Henkin. The T-Box Riboswitch: tRNA as an Effector to Modulate Gene Regulation. *Microbiology Spectrum*: July 2018.

51. Tina M. Henkin. The T box riboswitch: A novel regulatory RNA that utilizes tRNA as its ligand. *Biochimica et Biophysica Acta*, 1839: 959-963 (2014).

52. Yide Mei, Jeongsik Yong, Aaron Stonestrom, and Xiaolu Yang. tRNA and cytochrome c in cell death and beyond. *Cell Cycle*, 9(15): 2936-2939, August 2010.

53. Yide Mei, Jeongsik Yong, Hongtu Liu, Yigong Shi, Judy Meinkoth, Gideon Dreyfuss, and Xiaolu Yang. tRNA Binds to Cytochrome c and Inhibits Caspase Activation. *Molecular Cell*, 37(5): 668-678, March 2010.

54. Yi-Ting Lo, Hung-Wei Huang, Yi-Chuan Huang, Jui-Fen Chan, and Yuan-Hao Howard Hsu. Elucidation of tRNA–cytochrome c interactions through hydrogen/deuterium exchange mass spectrometry. *Biochimica et Biophysica Acta*, 1865: 539-546, February 2017.

55. Davide Ruggero and Pier Paolo Pandolfi. Does the ribosome translate cancer? *Nature Reviews Cancer*, 3: 179-192, March 2003.

56. Mariana Pavon-Eternod, Suzanna Gomes, Renaud Geslain, Qing Dai, Marsha Rich Rosner, and Tao Pan. tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Research*, 37(21): 7268-7280, November 2009.

57. Matthew J. Eckwahl, Helene Arnion, Siarhei Kharytonchyk, Trinity Zang, Paul D. Bieniasz, Alice Telesnitsky, and Sandra L. Wolin. Analysis of the human immunodeficiency virus-1 RNA packageome. *RNA*, 22: 1228-1238, August 2016.

58. Shan Cen, Hassan Javanbakht, Sunghoon Kim, Kiyotaka Shiba, Rebecca Craven, Alan Rein, Karla Ewalt,7 Paul Schimmel,7 Karin Musier-Forsyth, and Lawrence Kleiman. Retrovirus-Specific Packaging of Aminoacyl-tRNA Synthetases with Cognate Primer tRNAs. *Journal of Virology*, 76(24): 13111-13115, December 2002.

59. Min Jiang, Johnson Mak, Azim Ladha, Eric Cohen, Michel Klein, Benjamin Rovinski, and Lawrence Kleiman. Identification of tRNAs Incorporated into Wild-Type and Mutant Human Immunodeficiency Virus Type 1. *Journal of Virology*, 67(6): 3246-3253, June 1993.

60. Cheryl L. Isaac and Jack D. Keene. Transfer RNAs Associated with Vesicular Stomatitis Virus. *Journal of General Virology*, 56: 141-151, February 1981.

61. Anna van Weringh, Manon Ragonnet-Cronin, Erinija Pranckeviciene, Mariana Pavon-Eternod, Lawrence Kleiman, and Xuhua Xia. HIV-1 Modulates the tRNA Pool to Improve Translation Efficiency. *Molecular Biology and Evolution*, 28(6): 1827-1834, June 2011.

62. Dona Sleiman, Pierre Barraud, Franck Brachet, and Carine Tisne. The Interaction between $tRNA^{Lys}_3$ and the Primer Activation Signal Deciphered by NMR Spectroscopy. *PLoS One*, 8(6): June 2013.

63. Michele R. S. Hargittai, Anil T. Mangla, Robert J. Gorelick, and Karin Musier-Forsyth. HIV-1 Nucleocapsid Protein Zinc Finger Structures Induce $tRNA^{Lys,3}$ Structural Changes but are not Critical for Primer/Template Annealing. *Journal of Molecular Biology*, 312: 985-997, August 2001.

64. Vera Oberbauer and Matthias R. Schaefer. tRNA-Derived Small RNAs: Biogenesis, Modification, Function and Potential Impact on Human Disease Development. *Genes*, 9(607): December 2018.

65. Yijing Shen, Xiuchong Yu, Linwen Zhu, Tianwen Li, Zhilong Yan, and Junming Guo. Transfer RNA-derived fragments and tRNA halves: biogenesis, biological functions and their roles in diseases. *Journal of Molecular Medicine*, 96: 1167-1176, September 2018.

66. Xiuchong Yu, Yaoyao Xie, Shuangshuang Zhang, Xuemei Song, Bingxiu Xiao1, and Zhilong Yan. tRNA-derived fragments: Mechanisms underlying their regulation of gene expression and potential applications as therapeutic targets in cancers and virus infections. *Theranostics*, 11(1): 461-469, January 2021.

67. Pankaj Kumar, Jordan Anaya, Suresh B Mudunuri, and Anindya Dutta. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biology*, 12(78): October 2014.

68. Arianna Di Fazio, Margarita Schlackow, Sheng Kai Pong, Adele Alagia, and Monika Gullerova. Dicer dependent tRNA derived small RNAs promote nascent RNA silencing. *Nucleic Acids Research*, 50(3): 1734-1752, February 2022.

69. Canan Kuscu, Pankaj Kumar, Manjari Kiran, Zhangli Su, Asrar Malik, and Anindya Dutta. tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA*, 24(8): 1093–1105, August 2018.

70. Dirk Haussecker, Yong Huang, Ashley Lau, Poornima Parameswaran, Andrew Z. Fire, and Mark A. Kay. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA*, 16(4): 673-695, April 2010.

71. Yiran Zhou, Haoran Peng, Qinghua Cuiac, and Yuan Zhou. tRFTar: Prediction of tRF-target gene interactions via systemic re-analysis of Argonaute CLIP-seq datasets. *Methods*, 187: 57-67, March 2021.

72. Ningshan Li, Nayang Shan, Lingeng Lu, and Zuoheng Wang. tRFtarget: a database for transfer RNA-derived fragment targets. *Nucleic Acids Research*, 49(D1): D254-260, October 2020.

73. Man Lung Yeung, Yamina Bennasser, Koichi Watashi, Shu-Yun Le, Laurent Houzet, and Kuan-Teh Jeang. Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. *Nucleic Acids Research*, 37(19): 6575–6586, October 2009.

74. Lingyu Guan, Spyros Karaiskos, and Andrey Grigoriev. Inferring targeting modes of Argonaute-loaded tRNA fragments. *RNA Biology*, 17(8): 1070-1080, October 2019.

75. Hak Kyun Kim, Jianpeng Xu, Kirk Chu, Hyesuk Park, Hagoon Jang, Pan Li, Paul N.Valdmanis, Qiangfeng Cliff Zhang, and Mark A.Kay. A tRNA-Derived Small RNA Regulates Ribosomal Protein S28 Protein Levels after Translation Initiation in Humans and Mice. *Cell Reports*, 12(17): 3816-3824, December 2019.

76. Zhangli Su, Ida Monshaugen, Briana Wilson, Fengbin Wang, Arne Klungland, Rune Ougland, and Anindya Dutta. TRMT6/61A-dependent base methylation of tRNA-derived fragments regulates gene-silencing activity and the unfolded protein response in bladder cancer. *Nature Communications*, 13: April 2022.

77. M Nashimoto. Specific cleavage of target RNAs from HIV-1 with 5' half tRNA by mammalian tRNA 3' processing endoribonuclease. *RNA*, 2(6): 523-524, 1996.

78. Hak Kyun Kim, Gabriele Fuchs, Shengchun Wang, Wei Wei, Yue Zhang, Hyesuk Park, Biswajoy Roy-Chaudhuri, Pan Li, Jianpeng Xu, Kirk Chu, Feijie Zhang, Mei-Sze Chua, Samuel So, Qiangfeng Cliff Zhang, Peter Sarnow, and Mark A. Kay. A transfer-RNA-derived small RNA regulates ribosome biogenesis. *Nature*, 552: 57-62, November 2017.

79. Hani Goodarzi, Xuhang Liu, Hoang C. B. Nguyen, Steven Zhang, Lisa Fish, and Sohail F. Tavazoie. Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. *Cell*, 161(4): 790-802, May 2015.

80. Junchao Shi, Yunfang Zhang, Tong Zhou, and Qi Chen. tsRNAs: The Swiss Army Knife for Translational Regulation. *Trends in Biochemical Sciences*, 44(3): 185-189, March 2019.

81. Shengqian Dou,Yirong Wang, and Jian Lu. Metazoan tsRNAs: Biogenesis, Evolution and Regulatory Functions. *Non-Coding RNA*, 5(1): February 2019.

82. Rogan Magee and Isidore Rigoutsos. On the expanding roles of tRNA fragments in modulating cell behavior. *Nucleic Acids Research*, 48(17): 9433–9448, September 2020.

83. Pavel Ivanov, Mohamed M. Emara, Judit Villen, Steven P. Gygi, and Paul Anderson. Angiogenin-induced tRNA fragments inhibit translation initiation. *Molecular Cell*, 43(4): 613-23, August 2011.

84. Maxim P. Nekrasov, Maria P. Ivshina, Konstantin G. Chernov, Adri A.M. Thomas, John W.B. Hershey, and Lev P. Ovchinnikov. The mRNA-binding Protein YB-1 (p50) Prevents Association of the Eukaryotic Initiation Factor eIF4G with mRNA and Inhibits Protein Synthesis at the Initiation Stage. *Journal of Biological Chemistry*, 278(16): 13936-13943, April 2003.

85. Shawn M. Lyons, Chris Achorn, Nancy L. Kedersha, Paul J. Anderson, and Pavel Ivanov. YB-1 regulates tiRNA-induced Stress Granule formation but not translational repression. *Nucleic Acids Research*, 44(14): 6949–6960, August 2016.

86. Pavel Ivanov, Elizabeth O'Day, Mohamed M. Emara, Gerhard Wagner, Judy Lieberman, and Paul Anderson. G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. *Proceedings of the National Academy of Sciences*, 111 (51): 18201-18206, November 2014.

87. Nicola Guzzi, Maciej Cieśla, Phuong Cao Thi Ngoc, Stefan Lang, Sonali Arora, Marios Dimitriou, Kristyna Pimková, Mikael N.E. Sommarin, Roberto Munita, Michal Lubas, Yiting Lim, Kazuki Okuyama, Shamit Soneji, Göran Karlsson, Jenny Hansson, Göran Jönsson, Anders H. Lund, Mikael Sigvardsson, Eva Hellström-Lindberg, Andrew C.Hsieh, and Cristian Bellodi. Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. *Cell*, 173(5): 1204-1216, May 2018.

88. German Martinez. tRNA-derived small RNAs: New players in genome protection against retrotransposons. *RNA Biology*, 15(2): 170–175, December 2017.

89. Ksenia Skvortsova, Nicola Iovino, and Ozren Bogdanović. Functions and mechanisms of epigenetic inheritance in animals. *Nature Reviews Molecular Cell Biology*, 19: 774–790, November 2018.

90. Xue Zhang, Xin He, Chao Liu, Jun Liu, Qifei Hu, Ting Pan, Xiaobing Duan, Bingfeng Liu, Yiwen Zhang, Jingliang Chen, Xingru Ma, Xu Zhang, Haihua Luo, and Hui Zhang. IL-4 Inhibits the Biogenesis of an Epigenetically Suppressive PIWI-Interacting RNA To

Upregulate CD1a Molecules on Monocytes/Dendritic Cells. *Journal of Immunology*, 196(4): 1591-603, February 2016.

91. Yunfang Zhang, Xudong Zhang, Junchao Shi, Francesca Tuorto, Xin Li, Yusheng Liu, Reinhard Liebers, Liwen Zhang, Yongcun Qu, Jingjing Qian, Maya Pahima, Ying Liu, Menghong Yan, Zhonghong Cao, Xiaohua Lei, Yujing Cao, Hongying Peng, Shichao Liu, Yue Wang, Huili Zheng, Rebekah Woolsey, David Quilici, Qiwei Zhai, Lei Li, Tong Zhou, Wei Yan, Frank Lyko, Ying Zhang, Qi Zhou, Enkui Duan, and Qi Chen. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nature Cell Biology*, 20(5): 535-540, May 2018.

92. Qi Chen, Menghong Yan, Zhonghong Cao, Xin Li, Yunfang Zhang, Junchao Shi, Gui-hai Feng, Hongying Peng, Xudong Zhang, Ying Zhang, Jingjing Qian, Enkui Duan, Qiwei Zhai, and Qi Zhou. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science*, 351(6271): 397-400, January 2016.

93. Upasna Sharma, Colin C. Conine, Jeremy M Shea, Ana Boskovic, Alan G. Derr, Xin Y. Bing, Clemence Belleannee, Alper Kucukural, Ryan W Serra, Fengyun Sun, Lina Song, Benjamin R. Carone, Emiliano P Ricci, Xin Z. Li, Lucas Fauquier, Melissa J. Moore, Robert Sullivan, Craig C. Mello, Manuel Garber, and Oliver J. Rando. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science*, 351(6271): 391-396, January 2016.

94. Juan Pablo Tosar and Alfonso Cayota. Extracellular tRNAs and tRNA-derived fragments. *RNA Biology*, 17(8): 1149–1167, February 2020.

95. Xuemei Chen and Oded Rechavi. Plant and animal small RNA communications between cells and organisms. *Nature Reviews Molecular Cell Biology*, 23: 185–203, October 2021.

96. Matthew J. Shurtleff, Jun Yao, Yidan Qin, Ryan M. Nottingham, Morayma M. Temoche-Diaz, Randy Schekman, and Alan M. Lambowitz. Broad role for YBX1 in defining the small noncoding RNA composition of exosomes. *Proceedings of the National Academy of Sciences*, 114(43): E8987-E8995, October 2017.

97. Xin Chen, Ang Li, Bao-Fa Sun, Ying Yang, Ya-Nan Han, Xun Yuan, Ri-Xin Chen, Wen-Su Wei, Yanchao Liu, Chun-Chun Gao, Yu-Sheng Chen, Mengmeng Zhang, Xiao-Dan Ma, Zhuo-Wei Liu, Jun-Hang Luo, Cong Lyu, Hai-Lin Wang, Jinbiao Ma, Yong-Liang Zhao, Fang-Jian Zhou, Ying Huang, Dan Xie, and Yun-Gui Yang. 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nature Cell Biology*, 21: 978–990, July 2019.

98. Xin Yang, Ying Yang, Bao-Fa Sun, Yu-Sheng Chen, Jia-Wei Xu, Wei-Yi Lai, Ang Li, Xing Wang, Devi Prasad Bhattarai, Wen Xiao, Hui-Ying Sun, Qin Zhu, Hai-Li Ma, Samir Adhikari, Min Sun, Ya-Juan Hao, Bing Zhang, Chun-Min Huang, Niu Huang, Gui-Bin Jiang, Yong-Liang Zhao, Hai-Lin Wang, Ying-Pu Sun, and Yun-Gui Yang. 5-

methylcytosine promotes mRNA export — NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell Research*, 27: 606–625, April 2017.

99. Ruben Garcia-Martin, Guoxiao Wang, Bruna B. Brandão, Tamires M. Zanotto, Samah Shah, Sandip Kumar Patel, Birgit Schilling, and C. Ronald Kahn. MicroRNA sequence codes for small extracellular vesicle release and cellular retention. *Nature*, 601(7893): 446-451, January 2022.

100. Ni-Ting Chiou, Robin Kageyama, and K. Mark Ansel. Selective Export into Extracellular Vesicles and Function of tRNA Fragments during T Cell Activation. *Cell Reports*, 25(12): 3356-3370, December 2018.

101. Upasna Sharma, Fengyun Sun, Colin C. Conine, Brian Reichholf, Shweta Kukreja, Veronika A. Herzog, Stefan L. Ameres, and Oliver J. Rando. Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm. *Developmental Cell*, 46(4): 481-494, August 2018.

102. Pascal F. Egea, Robert M. Stroud, and Peter Walter. Targeting proteins to membranes: structure of the signal recognition particle. *Current Opinion in Structural Biology*, 15: 213–220, March 2005.

103. Robert J. Keenan, Douglas M. Freymann, Robert M. Stroud, and Peter Walter. The Signal Recognition Particle. *Annual Review of Biochemistry*, 70: 755–775, 2001.

104. Kiyoshi Nagai, Chris Oubridge, Andreas Kuglstatter, Elena Menichelli, Catherine Isel, and Luca Jovine. Structure, function and evolution of the signal recognition particle. *The EMBO Journal*, 22: 3479-3485, July 2003.

105. Bertrand Beckert, Alexej Kedrov, Daniel Sohmen, Georg Kempf, Klemens Wild, Irmgard Sinning, Henning Stahlberg, Daniel N Wilson, and Roland Beckmann. Translational arrest by a prokaryotic signal recognition particle is mediated by RNA interactions. *Nature Structural and Molecular Biology*, 22(10): 767-773, October 2015.

106. Camilla Faoro and Sandro F. Ataide. Noncanonical Functions and Cellular Dynamics of the Mammalian Signal Recognition Particle Components. *Frontiers in Molecular Biosciences*, 8: May 2021.

107. Madi Bidya Sagar, Louise Lucast, and Jennifer A. Doudna. Conserved but nonessential interaction of SRP RNA with translation factor EF-G. *RNA*, 10: 772–778, 2004.

108. Kouji Nakamura, Hideki Miyamoto, Satoru Suzuma, Taiichi Sakamoto, Gota Kawai, and Kunio Yamane. Minimal Functional Structure of Escherichia coli 4.5 S RNA Required for Binding to Elongation Factor G. *The Journal of Biological Chemistry*, 276(25): 22844–22849, April 2001.

109. Magnus Alm Rosenblad, Niels Larsen, Tore Samuelsson, and Christian Zwieb. Kinship in the SRP RNA family. *RNA Biology*, 6(5): 508-516, November 2009.

110. Feng-Jie Sun and Gustavo Caetano-Anollés. The Evolutionary History of the Structure of 5S Ribosomal RNA. *Journal of Molecular Evolution*, 69: 430–443, July 2009.

111. David Ammons, Joanne Rampersad, and George E. Fox. 5S rRNA gene deletions cause an unexpectedly high fitness loss in *Escherichia coli*. *Nucleic Acids Research*, 27(2): 637-642, January 1999.

112. Maciej Szymanski, Miroslawa Z. Barciszewska, Volker A. Erdmann, and Jan Barciszewski. 5S Ribosomal RNA Database. *Nucleic Acids Research*, 30(1): 176–178, January 2002.

113. Menghong Yan, Yuangao Wang, Yanan Hu, Yan Feng, Changgui Dai, Jingxia Wu, Dongmei Wu, Fang Zhang, and Qiwei Zhai. A High-Throughput Quantitative Approach Reveals More Small RNA Modifications in Mouse Liver and Their Correlation with Diabetes. *Analytical Chemistry*, 85(24): 12173–12181, November 2013.

114. G. M. Gongadze. 5S rRNA and Ribosome. *Biochemistry (Moscow)*, 76(13): 1450-1464, December 2011.

115. Petr Sergiev, Svetlana Dokudovskaya, Elena Romanova, Andrey Topin, Alexey Bogdanov, Richard Brimacombe, and Olga Dontsova. The environment of 5S rRNA in the ribosome: cross-links to the GTPase-associated area of 23S rRNA. *Nucleic Acids Research*, 26(11): 2519-2525, June 1998.

116. Stanislav Chladek. Possible relationship of peptidyl transferase binding sites, 5S RNA and peptidyl-tRNA. *Biochemical and Biophysical Research Communications*, 45(3): 695-700, November 1971.

117. Shijie Huang, Nikolay A. Aleksashin, Anna B. Loveland, Dorota Klepacki, Kaspar Reier, Amira Kefi, Teresa Szal, Jaanus Remme, Luc Jaeger, Nora Vázquez-Laslop, Andrei A. Korostelev and Alexander S. Mankin. Ribosome engineering reveals the importance of 5S rRNA autonomy for ribosome assembly. *Nature Communications*, 11: June 2020.

118. A. M. Dechampesme, O Koroleva, I. Leger-Silvestre, N. Gas, and S. Camier. Assembly of 5S ribosomal RNA is required at a specific step of the pre-rRNA processing pathway. *Journal of Cell Biology*, 145(7): 1369-1380, June 1999.

119. Martin Ciganda and Noreen Williams. Eukaryotic 5S rRNA biogenesis. *Wiley Interdisciplinary Reviews RNA*, 2(4): 523-533, February 2012.

120. Joanna Rorbach, Fei Gao, Christopher A. Powell, Aaron D'Souza, Robert N. Lightowlers, Michal Minczuk, and Zofia M. Chrzanowska-Lightowlers. Human mitochondrial ribosomes can switch their structural RNA composition. *Proceedings of the National Academy of Sciences*, 113(43): 12198-12201, October 2016.

121. Carmine Onofrillo, Alice Galbiati, Lorenzo Montanaro, and Massimo Derenzini. The pre-existing population of 5S rRNA effects p53 stabilization during ribosome biogenesis inhibition. *Oncotarget*, 8(3): 4257-4267, January 2017.

122. Giulio Donati, Suresh Peddigari, Carol A. Mercer, and George Thomas. 5S Ribosomal RNA Is an Essential Component of a Nascent Ribosomal Precursor Complex that Regulates the Hdm2-p53 Checkpoint. *Cell Reports*, 4: 87-98, July 2013.

123. Henri Grosjean. RNA modification: the Golden Period 1995–2015. *RNA*, 21(4): 625–626, April 2015.

124. Pierre Barraud and Carine Tisné. To Be or Not to Be Modified: Miscellaneous Aspects Influencing Nucleotide Modifications in tRNAs. *IUBMB Life*, 71(8): 1126-1140, August 2019.

125. Cole J.T. Lewis, Tao Pan and Auinash Kalsotra. RNA modifications and structures cooperate to guide RNA–protein interactions. *Nature Reviews Molecular Cell Biology*, 18: 202-210, March 2017.

126. Chen Gu, Thomas J. Begley, and Peter C. Dedon. tRNA modifications regulate translation during cellular stress. *FEBS Letters*, 588: 4287–4296, October 2014.

127. William E. Swinehart and Jane E. Jackman. Diversity in mechanism and function of tRNA methyltransferases. *RNA Biology*, 12(4): 398–411, April 2015.

128. Hailing Shi, Jiangbo Wei, and Chuan He. Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Molecular Cell*, 74(4): 640-650, May 2019.

129. Mark Helm, Richard Giegé, and Catherine Florentz. A Watson−Crick Base-Pair-Disrupting Methyl Group (m^1A9) Is Sufficient for Cloverleaf Folding of Human Mitochondrial $tRNA^{Lys}$. *Biochemistry*, 38(40): 13338–13346, September 1999.

130. Glenn R. Björk, P. Mikael Wikström, and Anders S. Byström. Prevention of Translational Frameshifting by the Modified Nucleoside 1-Methylguanosine. *Science*, 244(4907): 986-989, May 1989.

131. Elisa Vilardo, Fabian Amman, Ursula Toth, Annika Kotter, Mark Helm, and Walter Rossmanith. Functional characterization of the human tRNA methyltransferases TRMT10A and TRMT10B. *Nucleic Acids Research*, 48(11): 6157-6169, June 2020.

132. Stephanie Oerum, Clément Dégut, Pierre Barraud, and Carine Tisné. m^1A Post-Transcriptional Modification in tRNAs. *Biomolecules*, 7(1): March 2017.

133. Chi Zhang and Guifang Jia. Reversible RNA Modification N1-methyladenosine (m^1A) in mRNA and tRNA. *Genomics, Proteomics & Bioinformatics*, 16(3): 155-161, June 2018.

134. Zhuojia Chen, Meijie Qi, Bin Shen, Guanzheng Luo, Yingmin Wu, Jiexin Li, Zhike Lu, Zhong Zheng, Qing Dai, and Hongsheng Wang. Transfer RNA demethylase ALKBH3 promotes cancer progression via induction of tRNA-derived small RNAs. *Nucleic Acids Research*, 47(5): 2533-2545, March 2019.

135. R. Basavappa and P.B. Sigler. The 3 A crystal structure of yeast initiator tRNA: functional implications in initiator/elongator discrimination. *The EMBO Journal*, 10: 3105-3111, October 1991.

136. Hiroyuki Fukuda, Takeshi Chujo, Fan-Yan Wei, Sheng-Lan Shi, Mayumi Hirayama, Taku Kaitsuka, Takahiro Yamamoto, Hiroyuki Oshiumi, and Kazuhito Tomizawa. Cooperative methylation of human $tRNA^{Lys}_3$ at positions A58 and U54 drives the early and late steps of HIV-1 replication. *Nucleic Acids Research,* 49(20): 11855–11867, November 2021.

137. Wesley C. Clark, Molly E. Evans, Dan Dominissini, Guanqun Zheng, and Tao Pan. tRNA base methylation identification and quantification via high-throughput sequencing. *RNA*, 22: 1771-1784, September 2016.

138. Ya-Ming Hou, Isao Masuda, and Howard Gamper. Codon-Specific Translation by m^1G37 Methylation of tRNA. *Frontiers in Genetics,* 9: January 2019.

139. William F. Waas, Zhanna Druzina, Melanie Hanan, and Paul Schimmel. Role of a tRNA Base Modification and Its Precursors in Frameshifting in Eukaryotes. *The Journal of Biological Chemistry*, 282(36): 26026–26034, September 2007.

140. Ben E Clifton, Muhammad A Fariz, Gen-Ichiro Uechi, and Paola Laurino. Evolutionary repair reveals an unexpected role of the tRNA modification m^1G37 in aminoacylation. *Nucleic Acids Research*, 49(21): 12467–12485, December 2021.

141. Feilong Meng, Mi Zhou, Yun Xiao, Xiaoting Mao, Jing Zheng, Jiaxi Lin, Tianxiang Lin, Zhenzhen Ye, Xiaohui Cang, Yong Fu, Meng Wang, and Min-Xin Guan. A deafness-associated tRNA mutation caused pleiotropic effects on the m^1G37 modification, processing, stability and aminoacylation of tRNAIle and mitochondrial translation. *Nucleic Acids Research*, 49(2): 1075–1093, January 2021.

142. Scott I. Hauenstein and John J. Perona. Redundant synthesis of cysteinyl-tRNACys in *Methanosarcina mazei*. *The Journal of Biological Chemistry*, 283(32): 22007-22017, August 2008.

143. Chun-Mei Zhang, Cuiping Liu, Simon Slater, and Ya-Ming Hou. Aminoacylation of tRNA with phosphoserine for synthesis of cysteinyl-tRNA(Cys). *Nature Structural and Molecular Biology*, 15(5): 507-514, May 2008.

144. W.M.Holmes, C.Andraos-Selim, and M.Redlak. tRNA-m^1G methyltransferase interactions: Touching bases with structure. *Biochemie*, 77(1-2): 62-65, 1995.

145. Katherine E. Bohnsack, Nicole Kleiber, Nicolas Lemus-Diaz, and Markus T. Bohnsack. Roles and dynamics of 3-methylcytidine in cellular RNAs. *Trends in Biochemical Sciences*, 47(7): 596-608, July 2022.

146. Luang Xu, Xinyu Liu, Na Sheng, Kyaw Soe Oo, Junxin Liang, Yok Hian Chionh, Juan Xu, Fuzhou Ye, Yong-Gui Gao, Peter C. Dedon, and Xin-Yuan Fu. Three distinct 3-methylcytidine (m^3C) methyltransferases modify tRNA and mRNA in mice and humans. *The Journal of Biological Chemistry*, 292(35): 14695–14703, June 2017.

147. Sonia D'Silva, Steffen J. Haider, and Eric M. Phizicky. A domain of the actin binding protein Abp140 is the yeast methyltransferase responsible for 3-methylcytidine modification in the tRNA anti-codon loop. *RNA*, 17(6): 1100-1110, June 2011.

148. Lu Han and Eric M. Phizicky. A rationale for tRNA modification circuits in the anticodon loop. *RNA*, 24(10): 1277–1284, July 2018.

149. Nicole Kleiber, Nicolas Lemus-Diaz, Carina Stiller, Marleen Heinrichs, Mandy Mong-Quyen Mai, Philipp Hackert, Ricarda Richter-Dennerlein, Claudia Höbartner, Katherine E. Bohnsack, and Markus T. Bohnsack. The RNA methyltransferase METTL8 installs m^3C_{32} in mitochondrial $tRNAs^{Thr/Ser(UCN)}$ to optimise tRNA structure and mitochondrial translation. *Nature Communications*, 13: January 2022.

150. Valentina V. Ignatova, Steffen Kaiser, Jessica Sook Yuin Ho, Xinyang Bing, Paul Stolz, Xim Tan, Chee Leng Lee, Florence Pik Hoon Gay, Palma Rico Lastres, Raffaele Gerlini, Birgit Rathkolb, Antonio Aguilar-Pimentel, Adrián Sanz-Moreno, Tanja Klein-Rodewald, Julia Calzada-Wack, Emil Ibragimov, Magdalena Valenta, Saulius Lukauskas, Andrea Pavesi, Susan Marschall, Stefanie Leuchtenberger, Helmut Fuchs, Valerie Gailus-Durner, Martin Hrabe de Angelis, Sebastian Bultmann, Oliver J. Rando, Ernesto Guccione, Stefanie M. Kellner, and Robert Schneider. METTL6 is a tRNA m^3C methyltransferase that regulates pluripotency and tumor cell growth. *Science Advances*, 6(35): August 2020.

151. Ville Y. P. Väre, Emily R. Eruysal, Amithi Narendran, Kathryn L. Sarachan and Paul F. Agris. Chemical and Conformational Diversity of Modified Nucleosides Affects tRNA Structure and Function. *Biomolecules*, 6(29): 2016.

152. Andrew Behrens, Geraldine Rodschinka, and Danny D. Nedialkova. High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Molecular Cell*, 81(8): 1802-1815, April 2021.

153. Jaunius Urbonavičiusa, Jean Armengaudb, and Henri Grosjean. Identity Elements Required for Enzymatic Formation of N^2,N^2-dimethylguanosine from N^2-monomethylated Derivative and its Possible Role in Avoiding Alternative Conformations in Archaeal tRNA. *Journal of Molecular Biology*, 357(2): 387-399, March 2006.

154. Sergey Steinberg and Robert Cedergren. A correlation between N2-dimethylguanosine presence and alternate tRNA conformers. *RNA*, 1(9): 886-91, November 1995.

155. Pradeep S. Pallan, Christoph Kreutz, Silvia Bosio, Ronald Micura, and Martin Egli. Effects of N^2,N^2-dimethylguanosine on RNA structure and stability: Crystal structure of an RNA duplex with tandem $m^2{}_2$G:A pairs. *RNA*, 14(10): 2125–2135, October 2008.

156. Romina Oliva, Luigi Cavallo, and Anna Tramontano. Accurate energies of hydrogen bonded nucleic acid base pairs and triplets in tRNA tertiary interactions. *Nucleic Acids Research*, 34(3): 865–879, February 2006.

157. Francesca Tuorto and Frank Lyko. Genome recoding by tRNA modifications. *Open Biology*, 6(12): December 2016.

158. Marcus J.O. Johansson, Anders Esberg, Bo Huang, Glenn R. Björk, and Anders S. Byström. Eukaryotic Wobble Uridine Modifications Promote a Functionally Redundant Decoding System. *Molecular and Cellular Biology*, 28(10): 3301–3312, May 2008.

159. Paul F. Agris, Amithi Narendran, Kathryn Sarachan, Ville Y.P. Väre, and Emily Eruysal. The Role of RNA Modifications in Translational Fidelity. *The Enzymes*, 41: 1-50, 2017.

160. Sundaramoorthy Srinivasan, Adrian Gabriel Torres, and Lluís Ribas de Pouplana. Inosine in Biology and Disease. *Genes*, 12(4): April 2021.

161. Àlbert Rafels-Ybern, Adrian Gabriel Torres, Xavier Grau-Bove, Iñaki Ruiz-Trillo, and Lluís Ribas de Pouplana. Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two Bacterial phyla. *RNA Biology*, 15(4-5): 500–507, 2018.

162. Adrian Gabriel Torres,David Piñeyro,Liudmila Filonava,Travis H. Stracker, Eduard Batlle, and Lluis Ribas de Pouplana. A-to-I editing on tRNAs: Biochemical, biological and evolutionary implications. *FEBS Letters*, 588(23): 4279-4286, September 2014.

163. Adrian Gabriel Torres, David Piñeyro, Marta Rodríguez-Escribà, Noelia Camacho, Oscar Reina, Adélaïde Saint-Léger, Liudmila Filonava, Eduard Batlle, and Lluís Ribas de Pouplana. Inosine modifications in human tRNAs are incorporated at the precursor tRNA level. *Nucleic Acids Research*, 43(10): 5145–5157, May 2015.

164. Àlbert Rafels-Ybern, Adrian Gabriel Torres, Noelia Camacho, Andrea Herencia-Ropero, Helena Roura Frigolé, Thomas F Wulff, Marina Raboteg, Albert Bordons, Xavier Grau-Bove, Iñaki Ruiz-Trillo, and Lluís Ribas de Pouplana. The Expansion of Inosine at the Wobble Position of tRNAs, and Its Role in the Evolution of Proteomes. *Molecular Biology and Evolution*, 36(4): 650-662, April 2019.

165. Naďa Špačková and Kamila Réblová. Role of Inosine–Uracil Base Pairs in the Canonical RNA Duplexes. *Genes*, 9(7): July 2018.

166. Daniel J. Wright, Christopher R. Force, and Brent M. Znosko. Stability of RNA duplexes containing inosine·cytosine pairs. *Nucleic Acids Research*, 46(22): 12099–12108, October 2018.

167. Keri Marie Nelson, Braden Meason, Haewon Min, Lara Senekjian, David S. Booth, and Neena Grover. Thermodynamic Stabilities of Inosine Base Pairs. *The FASEB Journal*, 21(6): A1027-A1027, April 2007.

168. Adrian Gabriel Torres, Marta Rodríguez-Escribà, Marina Marcet-Houben, Helaine Graziele Santos Vieira, Noelia Camacho, Helena Catena, Marina Murillo Recio, Àlbert Rafels-Ybern, Oscar Reina, Francisco Miguel Torres, Ana Pardo-Saganta, Toni Gabaldón, Eva Maria Novoa, and Lluís Ribas de Pouplana. Human tRNAs with inosine 34 are essential to efficiently translate eukarya-specific low-complexity proteins. *Nucleic Acids Research*, 49(12): 7011-7034, July 2021.

169. Thomas S.B. Schmidt, Jeroen Raes, and Peer Bork. The Human Gut Microbiome: From Association to Modulation. *Cell*, 172(6): 1198-1215, March 2018.

170. Verónica Lloréns-Rico, Joshua A. Simcock, Geert R.B. Huys, and Jeroen Raes. Single-cell approaches in human microbiome research. *Cell*, 185(15): 2725-2738, July 2022.

171. Claire Fergus, Dominic Barnes, Mashael A. Alqasem and Vincent P. Kelly. The Queuine Micronutrient: Charting a Course from Microbe to Man. *Nutrients*, 7: 2897-2929, 2015.

172. Sven Johannsson, Piotr Neumann, and Ralf Ficner. Crystal Structure of the Human tRNA Guanine Transglycosylase Catalytic Subunit QTRT1. *Biomolecules*, 8(81): August 2018.

173. Albert Jeltsch, Ann Ehrenhofer-Murray, Tomasz P. Jurkowski, Frank Lyko, Gunter Reuter, Serge Ankri, Wolfgang Nellen, Matthias Schaefer, and Mark Helm. Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. *RNA Biology*, 14(9): 1108-1123, September 2017.

174. Martin Müller, Carine Legrand, Francesca Tuorto, Vincent P. Kelly, Yaser Atlasi, Frank Lyko, and Ann E. Ehrenhofer-Murray. Queuine links translational control in eukaryotes to a micronutrient from bacteria. *Nucleic Acids Research*, 47(7): 3711–3727, April 2019.

175. Francesca Tuorto, Carine Legrand, Cansu Cirzi, Giuseppina Federico, Reinhard Liebers, Martin Müller, Ann E. Ehrenhofer-Murray, Gunnar Dittmar, Hermann-Josef Gröne, and Frank Lyko. Queuosine-modified tRNAs confer nutritional control of protein translation. *The EMBO Journal*, 37(18): September 2018.

176. Patti Hayes, Claire Fergus, Magda Ghanim, Cansu Cirzi, Lyubomyr Burtnyak, Callum J. McGrenaghan, Francesca Tuorto, Derek P. Nolan, and Vincent P. Kelly. Queuine

Micronutrient Deficiency Promotes Warburg Metabolism and Reversal of the Mitochondrial ATP Synthase in Hela Cells. *Nutrients*, 12(3): March 2020.

177. Thomas Reisser, Werner Langgut, and Helga Kersten. The nutrient factor queuine protects HeLa cells from hypoxic stress and improves metabolic adaptation to oxygen availability. *European Journal of Biochemistry*, 221(3): 979-86, May 1994.

178. Xiaoyun Wang, Zaneta Matuszek, Yong Huang, Marc Parisien, Qing Dai, Wesley Clark, Michael H. Schwartz, and Tao Pan. Queuosine modification protects cognate tRNAs against ribonuclease cleavage. *RNA*, 24(10): 1305–1313, October 2018.

179. Ylenia Chiari, Kirstin Dion, James Colborn, Aristeidis Parmakelis, and Jeffrey R Powell. On the possible role of tRNA base modifications in the evolution of codon usage: queuosine and Drosophila. *Journal of Molecular Evolution*, 70(4): 339-345, April 2010.

180. John M. Zaborske, Vanessa L. Bauer DuMont, Edward W. J. Wallace, Tao Pan, Charles F. Aquadro, and D. Allan Drummond. A Nutrient-Driven tRNA Modification Alters Translational Fidelity and Genome-wide Protein Coding across an Animal Genus. *PLoS Biology*, 12(12): December 2014.

181. Jilei Zhang, Rong Lu, Yongguo Zhang, Żaneta Matuszek, Wen Zhang, Yinglin Xia, Tao Pan, and Jun Sun. tRNA Queuosine Modification Enzyme Modulates the Growth and Microbiome Recruitment to Breast Tumors. *Cancers*, 12(3): March 2020.

182. Andrea Bednářová, Marley Hanna, Isabella Durham, Tara Van Cleave, Alexis England, Anathbandhu Chaudhuri, and Natraj Krishnan. Lost in Translation: Defects in Transfer RNA Modifications and Neurological Disorders. *Frontiers in Molecular Neuroscience*, 10: May 2017.

183. Takeo Suzuki, Tsutomu Suzuki, Takeshi Wada, Kazuhiko Saigo, and Kimitsuna Watanabe. Taurine as a constituent of mitochondrial tRNAs: new insights into the functions of taurine and human mitochondrial diseases. *The EMBO Journal*, 21(23): 6581-6589, December 2002.

184. Takeo Suzuki, Yuka Yashiro, Ittoku Kikuchi, Yuma Ishigami, Hironori Saito, Ikuya Matsuzawa, Shunpei Okada, Mari Mito, Shintaro Iwasaki, Ding Ma, Xuewei Zhao, Kana Asano, Huan Lin, Yohei Kirino, Yuriko Sakaguchi, and Tsutomu Suzuki. Complete chemical structures of human mitochondrial tRNAs. *Nature Communications*, 11: August 2020.

185. Shinya Kurata, Albert Weixlbaumer, Takashi Ohtsuki, Tomomi Shimazaki, Takeshi Wada, Yohei Kirino, Kazuyuki Takai, Kimitsuna Watanabe, V. Ramakrishnan, and Tsutomu Suzuki. Modified Uridines with C5-methylene Substituents at the First Position of the tRNA Anticodon Stabilize U•G Wobble Pairing during Decoding. *The Journal of Biological Chemistry*, 283(27): 18801–18811, July 2008.

186. Elzbieta Sochacka, Roman H. Szczepanowski, Marek Cypryk, Milena Sobczak, Magdalena Janicka, Karina Kraszewska, Paulina Bartos, Anna Chwialkowska, and Barbara Nawrot. 2-Thiouracil deprived of thiocarbonyl function preferentially base pairs with guanine rather than adenine in RNA and DNA duplexes. *Nucleic Acids Research*, 43(5): 2499–2512, March 2015.

187. Noriko Umeda, Takeo Suzuki, Masashi Yukawa, Yoshikazu Ohya, Heisaburo Shindo, Kimitsuna Watanabe, and Tsutomu Suzuki. Mitochondria-specific RNA-modifying enzymes responsible for the biosynthesis of the wobble base in mitochondrial tRNAs. Implications for the molecular pathogenesis of human mitochondrial diseases. *The Journal of Biological Chemistry*, 280(2): 1613-1624, January 2005.

188. Takehiro Yasukawa, Yohei Kirino, Norie Ishii, Ian J. Holt, Howard T. Jacobs, Takao Makifuchi, Nobuyoshi Fukuhara, Shigeo Ohta, Tsutomu Suzuki, and Kimitsuna Watanabe. Wobble modification deficiency in mutant tRNAs in patients with mitochondrial diseases. *FEBS Letters*, 579(13): 2948-2952, May 2005.

189. Kana Asano, Takeo Suzuki, Ayaka Saito, Fan-Yan Wei, Yoshiho Ikeuchi, Tomoyuki Numata, Ryou Tanaka, Yoshihisa Yamane, Takeshi Yamamoto, Takanobu Goto, Yoshihito Kishita, Kei Murayama, Akira Ohtake, Yasushi Okazaki, Kazuhito Tomizawa, Yuriko Sakaguchi, and Tsutomu Suzuki. Metabolic and chemical regulation of tRNA modification associated with taurine deficiency and human disease. *Nucleic Acids Research*, 46(4): 1565–1583, February 2018.

190. Vanessa Anissa Nathalie Rezgui, Kshitiz Tyagi, Namit Ranjan, Andrey L. Konevega, Joerg Mittelstaet, Marina V. Rodnina, Matthias Peter, and Patrick G.A. Pedrioli. tRNA tK^{UUU}, tQ^{UUG}, and tE^{UUC} wobble position modifications fine-tune protein translation by promoting ribosome A-site binding. *Proceedings of the National Academy of Sciences*, 110(30): 12289–12294, July 2013.

191. Monique Dewez, Fanélie Bauer, Marc Dieu, Martine Raes, Jean Vandenhaute, and Damien Hermand. The conserved Wobble uridine tRNA thiolase Ctu1–Ctu2 is required to maintain genome integrity. *Proceedings of the National Academy of Sciences*, 105(14): 5459-5464, April 2008.

192. Lene Songe-Møller, Erwin van den Born, Vibeke Leihne, Cathrine B. Vågbø, Terese Kristoffersen, Hans E. Krokan, Finn Kirpekar, Pål Ø. Falnes, and Arne Klungland. Mammalian ALKBH8 Possesses tRNA Methyltransferase Activity Required for the Biogenesis of Multiple Wobble Uridine Modifications Implicated in Translational Decoding. *Molecular and Cellular Biology*, 30(7): 1814–1827, April 2010.

193. Ting-Yu Lin, Nour El Hana Abbassi, Karol Zakrzewski, Andrzej Chramiec-Głąbik, Małgorzata Jemioła-Rzemińska, Jan Różycki, and Sebastian Glatt. The Elongator subunit Elp3 is a non-canonical tRNA acetyltransferase. *Nature Communications*, 10: February 2019.

194. Naoki Shigi. Biosynthesis and functions of sulfur modifications in tRNA. *Frontiers in Genetics*, 5(67): April 2014.

195. Annia Rodriguez-Hernandez, Jessica L. Spears, Kirk W. Gaston, Patrick A. Limbach, Howard Gamper, Ya-Ming Hou, Rob Kaiser, Paul F. Agris, and John J. Perona. Structural and Mechanistic Basis for Enhanced Translational Efficiency by 2-Thiouridine at the tRNA Anticodon Wobble Position. *Journal of Molecular Biology*, 425(20): 3888–3906, October 2013.

196. Sweta Vangaveti, William A. Cantara, Jessica L. Spears, Hasan DeMirc, Frank V. Murphy IV, Sri V. Ranganathan, Kathryn L. Sarachan, and Paul F. Agris. A Structural Basis for Restricted Codon Recognition Mediated by 2-thiocytidine in tRNA Containing a Wobble Position Inosine. *Journal of Molecular Biology*, 432(4): 913-929, February 2020.

197. Denis Bouvier, Natty Labessan, Martin Clémancey, Jean-Marc Latour, Jean-Luc Ravanat, Marc Fontecave, and Mohamed Atta. TtcA a new tRNA-thioltransferase with an Fe-S cluster. *Nucleic Acids Research*, 42(12): 7960-7970, July 2014.

198. Richard Lauman and Benjamin A. Garcia. Unraveling the RNA modification code with mass spectrometry. *Molecular Omics*, 16: 305-315, August 2020.

199. Laura Antoine and Philippe Wolff. Mapping of Posttranscriptional tRNA Modifications by Two-Dimensional Gel Electrophoresis Mass Spectrometry. *Methods in Molecular Biology*, 2113: 101-110, 2020.

200. Gabor L. Igloi and Hans Kössel. Affinity electrophoresis for monitoring terminal phosphorylation and the presence of queuosine in RNA. Application of polyacrylamide containing a covalently bound boronic acid. *Nucleic Acids Research*, 13(19): 6881-6898, 1985.

201. Wen Zhang, Ruyi Xu, Żaneta Matuszek, Zhen Cai, and Tao Pan. Detection and quantification of glycosylated queuosine modified tRNAs by acid denaturing and APB gels. *RNA*, 26(9): 1291-1298, September 2020.

202. Nian Liu, Marc Parisien, Qing Dai, Guanqun Zheng, Chuan He, and Tao Pan. Probing N^6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA*, 19(12): 1848–1856, December 2013.

203. Wen Zhang, Matthew Eckwahl, Katherine Zhou and Tao Pan. Sensitive and quantitative probing of pseudouridine modification in mRNA and long non-coding RNA. *RNA*, 25(9): 1218-1225, September 2019.

204. Dan Dominissini, Sharon Moshitch-Moshkovitz, Schraga Schwartz, Mali Salmon-Divon, Lior Ungar, Sivan Osenberg, Karen Cesarkas, Jasmine Jacob-Hirsch, Ninette Amariglio, Martin Kupiec, Rotem Sorek, and Gideon Rechavi. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, 485(7397): 201-6, April 2012.

205. Per Arne Aas, Marit Otterlei, Pål Ø. Falnes, Cathrine B. Vågbø, Frank Skorpen, Mansour Akbari, Ottar Sundheim, Magnar Bjørås, Geir Slupphaug, Erling Seeberg, and Hans E. Krokan. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature*, 421: 859–863, February 2003.

206. Guanqun Zheng, Yidan Qin, Wesley C. Clark, Qing Dai, Chengqi Yi, Chuan He, Alan M. Lambowitz, and Tao Pan. Efficient and quantitative high-throughput tRNA sequencing. *Nature Methods*, 12: 835–837, July 2015.

207. Jia Cui, Qi Liu, Erdem Sendinc, Yang Shi, and Richard I. Gregory. Nucleotide resolution profiling of m3C RNA modification by HAC-seq. *Nucleic Acids Research*, 49(5): March 2021.

208. Schraga Schwartz, Douglas A. Bernstein, Maxwell R. Mumbach, Marko Jovanovic, Rebecca H. Herbst, Brian X. León-Ricardo, Jesse M. Engreitz, Mitchell Guttman, Rahul Satija, Eric S. Lander, Gerald Fink, and Aviv Regev. Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell*, 159: 148–162, September 2014.

209. H. Alexander Ebhardt, Herbert H. Tsang, Denny C. Dai, Yifeng Liu, Babak Bostan, and Richard P. Fahlman. Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. *Nucleic Acids Research*, 37(8): 2461–2470, May 2009.

210. Chen Zhao, Fei Liu, and Anna Marie Pyle. An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA*, 24(2): 183–195, February 2018.

211. Li-Tao Guo, Rebecca L. Adams, Han Wan, Nicholas C. Huston, Olga Potapova, Sara Olson, Christian M. Gallardo, Brenton R. Graveley, Bruce E. Torbett, and Anna Marie Pyle. Sequencing and Structure Probing of Long RNAs Using MarathonRT: A Next-Generation Reverse Transcriptase. *Journal of Molecular Biology*, 432(10): 3338-3352, May 2020.

212. Alfred M. Lentzsch, Jun Yao, Rick Russell, and Alan M. Lambowitz. Template-switching mechanism of a group II intron-encoded reverse transcriptase and its implications for biological function and RNA-Seq. *The Journal of Biological Chemistry*, 294(51): 19764-19784, December 2019.

213. Megumi Shigematsu, Shozo Honda, Phillipe Loher, Aristeidis G. Telonis, Isidore Rigoutsos, and Yohei Kirino. YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Research*, 45(9): May 2017.

214. (a) Lieselotte Erber, Anne Hoffmann, Jörg Fallmann, Heike Betat, Peter F. Stadler, and Mario Mörl. LOTTE-seq (Long hairpin oligonucleotide based tRNA high-throughput sequencing): specific selection of tRNAs with 3'-CCA end for high-throughput sequencing. *RNA Biology*, 17(1): 23–32, September 2019. (b) Otis Pinkard, Sean McFarland, Thomas

Sweet, and Jeff Coller. Quantitative tRNA-sequencing uncovers metazoan tissue-specific tRNA regulation. *Nature Communications*, 11: August 2020.

215. Jennifer F. Hu, Daniel Yim, Duanduan Ma, Sabrina M. Huber, Nick Davis, Jo Marie Bacusmo, Sidney Vermeulen, Jieliang Zhou, Thomas J. Begley, Michael S. DeMott, Stuart S. Levine, Valérie de Crécy-Lagard, Peter C. Dedon, and Bo Cao. Quantitative mapping of the cellular small RNA landscape with AQRNA-seq. *Nature Biotechnology*, 39: 978–988, April 2021.

216. Feiyang Ma, Brie K. Fuqua, Yehudit Hasin, Clara Yukhtman, Chris D. Vulpe, Aldons J. Lusis, and Matteo Pellegrini. A comparison between whole transcript and 3' RNA sequencing methods using Kapa and Lexogen library preparation methods. *BMC Genomics*, 20: January 2019.

217. Thomas E. England and Olke C. Uhlenbeck. Enzymic oligoribonucleotide synthesis with T4 RNA ligase. *Biochemistry*, 17(11): 2069–2076, May 1978.

218. Molly E. Evans, Wesley C. Clark, Guanqun Zheng, and Tao Pan. Determination of tRNA aminoacylation levels by high-throughput sequencing. *Nucleic Acids Research*, 45(14): August 2017.

219. Dayue Chen and John T. Patton. Reverse Transcriptase Adds Nontemplated Nucleotides to cDNAs During 5′-RACE and Primer Extension. *BioTechniques*, 30(3): 574-580, March 2001.

220. Mark Helm and Yuri Motorin. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nature Reviews Genetics*, 18: 275–291, February 2017.

221. Wendy V. Gilbert, Tristan A. Belland, and Cassandra Schaening. Messenger RNA modifications: Form, distribution, and function. *Science*, 352(6292): 1408-1412, June 2016.

222. Dorota Piekna-Przybylska, Wayne A. Decatur, and Maurille J. Fournier. The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Research*, 36: D178-D183, January 2008.

223. Ralf Hauenschild, Lyudmil Tserovski, Katharina Schmid, Kathrin Thüring, Marie-Luise Winz, Sunny Sharma, Karl-Dieter Entian, Ludivine Wacheul, Denis L. J. Lafontaine, James Anderson, Juan Alfonzo, Andreas Hildebrandt, Andres Jäschke, Yuri Motorin, and Mark Helm. The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Research*, 43(20): 9950-9964, November 2015.

224. Eva Hrabeta-Robinson, Erin Marcus Aaron E. Cozen, Eric M. Phizicky, and Todd M. Lowe. High-Throughput Small RNA Sequencing Enhanced by AlkB-Facilitated RNA de-Methylation (ARM-Seq). *Methods in Molecular Biology*, 1562: 231-243, 2017.

225. Jinghui Song, Yuan Zhuang, Chenxu Zhu, Haowei Meng, Bo Lu, Bingteng Xie, Jinying Peng, Mo Li, and Chengqi Yi. Differential roles of human PUS10 in miRNA processing and tRNA pseudouridylation. *Nature Chemical Biology*, 16: 160–169, December 2019.

226. Zhang Zhang, Li-Qian Chen, Yu-Li Zhao, Cai-Guang Yang, Ian A. Roundtree, Zijie Zhang, Jian Ren, Wei Xie, Chuan He, and Guan-Zheng Luo. Single-base mapping of m(6)A by an antibody-independent method. *Science Advances*, 5(7): July 2019.

227. Miguel Angel Garcia-Campos, Sarit Edelheit, Ursula Toth, Modi Safra, Ran Shachar, Sergey Viukov, Roni Winkler, Ronit Nir, Lior Lasman, Alexander Brandis, Jacob H. Hanna, Walter Rossmanith, and Schraga Schwartz. Deciphering the "m^6A Code" via Antibody-Independent Quantitative Profiling. *Cell*, 178(3): 731-747, July 2019.

228. Dan Dominissini, Sigrid Nachtergaele, Sharon Moshitch-Moshkovitz, Eyal Peer, Nitzan Kol, Moshe Shay Ben-Haim, Qing Dai, Ayelet Di Segni, Mali Salmon-Divon, Wesley C. Clark, Guanqun Zheng, Tao Pan, Oz Solomon, Eran Eyal, Vera Hershkovitz, Dali Han, Louis C. Doré, Ninette Amariglio, Gideon Rechavi, and Chuan He. The dynamic N^1-methyladenosine methylome in eukaryotic messenger RNA. *Nature*, 530: 441–446, February 2016.

229. Shuibin Lin, Qi Liu, Victor S. Lelyveld, Junho Choe, Jack W. Szostak, and Richard I. Gregory. Mettl1/Wdr4-Mediated m^7G tRNA Methylome Is Required for Normal mRNA Translation and Embryonic Stem Cell Self-Renewal and Differentiation. *Molecular Cell*, 71(2): 244-255, July 2018.

230. Manqing Li, Elaine Kao, Dane Malone, Xia Gao, Jean Y. J. Wang, and Michael David. DNA damage-induced cell death relies on SLFN11-dependent cleavage of distinct type II tRNAs. *Nature Structural and Molecular Biology*, 25: 1047–1058, October 2019.

231. David Balchin, Manajit Hayer-Hartl, F. Ulrich Hartl. In vivo aspects of protein folding and quality control. *Science*, 353(6294): July 2016.

232. Helmut Sies and Dean P. Jones. Reactive oxygen species (ROS) as pleiotropic physiological signalling agents. *Nature Reviews Molecular Cell Biology*, 21: 363–383, March 2020.

233. Nandita Medda, Subrata Kumar De, and Smarajit Maiti. Different mechanisms of arsenic related signaling in cellular proliferation, apoptosis and neo-plastic transformation. *Ecotoxicology and Environmental Safety*, 208: January 2021.

234. Jennifer N. Wells, Robert Buschauer, Timur Mackens-Kiani, Katharina Best, Hanna Kratzat, Otto Berninghausen, Thomas Becker, Wendy Gilbert, Jingdong Cheng, and Roland Beckmann. Structure and function of yeast Lso2 and human CCDC124 bound to hibernating ribosomes. *PLoS Biology*, 18(7): July 2020.

235. Yuko Ueda, Ikumi Ooshio, Yasuyuki Fusamae, Kaori Kitae, Megumi Kawaguchi, Kentaro Jingushi, Hiroaki Hase, Kazuo Harada, Kazumasa Hirata, and Kazutake Tsujikawa. AlkB

homolog 3-mediated tRNA demethylation promotes protein synthesis in cancer cells. *Science Reports*, 7: February 2017.

236. Wenjun Deng, I. Ramesh Babu, Dan Su, Shanye Yin, Thomas J. Begley, and Peter C. Dedon. Trm9-catalyzed tRNA modifications regulate global protein expression by codon-biased translation. *PLoS Genetics*, 11(12): December 2015.

237. Ulrike Begley, Madhu Dyavaiah, Ashish Patil, John P. Rooney, Dan DiRenzo, Christine M. Young, Douglas S. Conklin, Richard S. Zitomer, and Thomas J. Begley. Trm9-catalyzed tRNA modifications link translation to the DNA damage response. *Molecular Cell*, 28(5): 860-870, December 2007.

238. Paul Anderson and Pavel Ivanov. tRNA fragments in human health and disease. *FEBS Letters*, 588(23): 4297-4304, November 2014.

239. Satoshi Yamasaki, Pavel Ivanov, Guo-Fu Hu, and Paul Anderson. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *Journal of Cell Biology*, 185(1): 35-42, April 2009.

240. Sherif Rashad, Xiaobo Han, Kanako Sato, Eikan Mishima, Takaaki Abe, Teiji Tominaga, and Kuniyasu Niizuma. The stress specific impact of ALKBH1 on tRNA cleavage and tiRNA generation. *RNA Biology*, 17(8): 1092-1103, August 2020.

241. Sylvain Delaunay and Michaela Frye. RNA modifications regulating cell fate in cancer. *Nature Cell Biology*, 21(5): 552-559, May 2019.

242. Sandra Blanco, Sabine Dietmann, Joana V. Flores, Shobbir Hussain, Claudia Kutter, Peter Humphreys, Margus Lukk, Patrick Lombard, Lucas Treps, Martyna Popis, Stefanie Kellner, Sabine M. Hölter, Lillian Garrett, Wolfgang Wurst, Lore Becker, Thomas Klopstock, Helmut Fuchs, Valerie Gailus-Durner, Martin Hrabě de Angelis, Ragnhildur T. Káradóttir, Mark Helm, Jernej Ule, Joseph G. Gleeson, Duncan T. Odom, and Michaela Frye. Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. *The EMBO Journal*, 33(18): 2020-2039, September 2014.

243. Cristina Cosentino, Sanna Toivonen, Esteban Diaz Villamil, Mohamed Atta, Jean-Luc Ravanat, Stéphane Demine, Andrea Alex Schiavo, Nathalie Pachera, Jean-Philippe Deglasse, Jean-Christophe Jonas, Diego Balboa, Timo Otonkoski, Ewan R. Pearson, Piero Marchetti, Décio L. Eizirik, Miriam Cnop, and Mariana Igoillo-Esteve. Pancreatic β-cell tRNA hypomethylation and fragmentation link TRMT10A deficiency with diabetes. 46(19): 10302-10318, November 2018.

244. Lauren Endres, Peter C. Dedon, and Thomas J. Begley. Codon-biased translation can be regulated by wobble-base tRNA modification systems during cellular stress responses. *RNA Biology*, 12(6): 603-614, 2015.

245. Botao Liu and Shu-Bing Qian. Translational reprogramming in cellular stress response. *Wiley Interdisciplinary Reviews RNA*, 5(3):301-315, June 2014.

246. Chien-Wen Chen and Motomasa Tanaka. Genome-wide Translation Profiling by Ribosome-Bound tRNA Capture. *Cell Reports*, 23(2): 608-621, April 2018.

247. Colin Chih-Chien Wu, Boris Zinshteyn, Karen A. Wehner, and Rachel Green. High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. *Molecular Cell*, 73(5): 959-970, March 2019.

248. Valérie de Crécy-Lagard, Pietro Boccaletto, Carl G. Mangleburg, Puneet Sharma, Todd M. Lowe, Sebastian A. Leidel, and Janusz M. Bujnicki. Matching tRNA modifications in humans to their known and predicted enzymes. *Nucleic Acids Research*, 47(5): 2143-2159, March 2019.

249. Xiao Wang, Zhike Lu, Adrian Gomez, Gary C. Hon, Yanan Yue, Dali Han, Ye Fu, Marc Parisien, Qing Dai, Guifang Jia, Bing Ren, Tao Pan, and Chuan He. N^6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, 505: 117-120, January 2014.

250. Fidel Ramírez, Devon P. Ryan, Björn Grüning, Vivek Bhardwaj, Fabian Kilpert, Andreas S. Richter, Steffen Heyne, Friederike Dündar, and Thomas Manke. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1): W160-165, July 2016.

251. Huaiyu Mi, Anushya Muruganujan, Dustin Ebert, Xiaosong Huang, and Paul D. Thomas. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. 47(D1): D419-D426, January 2019.

252. Charles Girardot, Jelle Scholtalbers, Sajoscha Sauer, Shu-Yi Su, and Eileen E.M. Furlong. Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinformatics*, 17: October 2016.

253. Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16): 2078-2079, August 2009.

254. Anthony M. Bolger, Marc Lohse, Bjoern Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114-2120, August 2014.

255. Patricia P. Chan and Todd M. Lowe. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research*, 44(D1): D184-189, January 2016.

256. Qing Dai, Guanqun Zheng, Michael H. Schwartz, Wesley C. Clark, and Tao Pan. Selective Enzymatic Demethylation of N^2,N^2-Dimethylguanosine in RNA and Its Application in High-Throughput tRNA Sequencing. *Angewandte Chemie International Edition*, 56(18): 5017-5020, April 2017.

257. Ana S. Gonzalez-Reiche, Matthew M. Hernandez, Mitchell J. Sullivan, Brianne Ciferri, Hala Alshammary, Ajay Obla, Shelcie Fabre, Giulio Kleiner, Jose Polanco, Zenab Khan, Bremy Alburquerque, Adriana van de Guchte, Jayeeta Dutta, Nancy Francoeur, Betsaida Salom Melo, Irina Oussenko, Gintaras Deikus, Juan Soto, Shwetha Hara Sridhar, Ying-Chih Wang, Kathryn Twyman, Andrew Kasarskis, Deena R. Altman, Melissa Smith, Robert Sebra, Judith Aberg, Florian Krammer, Adolfo García-Sastre, Marta Luksza, Gopi Patel, Alberto Paniz-Mondolfi, Melissa Gitman, Emilia Mia Sordillo, Viviana Simon, and Harm van Bakel. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science*, 369(6501): 297-301, July 2020.

258. Ron Sender, Yinon M. Bar-On, Shmuel Gleizer, Biana Bernshtein, Avi Flamholz, Rob Phillips, and Ron Milo. The total number and mass of SARS-CoV-2 virions. *Proceedings of the National Academy of Sciences*, 118(25): June 2021.

259. Ioanna Kalvari, Eric P Nawrocki, Nancy Ontiveros-Palacios, Joanna Argasinska, Kevin Lamkiewicz, Manja Marz, Sam Griffiths-Jones, Claire Toffano-Nioche, Daniel Gautheret, Zasha Weinberg, Elena Rivas, Sean R. Eddy, Robert D. Finn, Alex Bateman, and Anton I. Petrov. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research*, 49(D1): D192-D200, January 2021.

260. James A. Lewis and Bruce N. Ames. Histidine regulation in Salmonella typhimurium. XI. The percentage of transfer RNA His charged in vivo and its relation to the repression of the histidine operon. *Journal of Molecular Biology*, 66(1): 131-142, April 1972.

261. Hiroshi Kasai, Yoshiyuki Kuchino, Kayoko Nihei, and Susumu Nishimura. Distribution of the modified nucleoside Q and its derivatives in animal and plant transfer RNA's. *Nucleic Acids Research*, 2(10): 1931–1940, October 1975.

262. H. Kasai, Z. Ohashi, F. Harada, S. Nishimura, N. J. Oppenheimer, P. F. Crain, J. G. Liehr, D. L. Von Minden, and J. A. McCloskey. Structure of the modified nucleoside Q isolated from Escherichia coli transfer ribonucleic acid. 7-(4,5-cis-Dihydroxy-1-cyclopenten-3-ylaminomethyl)-7-deazaguanosine. *Biochemistry*, 14(19): 4198–4208, September 1975.

263. H. J. Grosjean, S. de Henau, and D. M. Crothers. On the physical basis for ambiguity in genetic coding interactions. *Proceedings of the National Academy of Sciences*, 75(2): 610-614, February 1978.

264. Rana C. Morris, Kenneth, G. Brown, and Mark S. Elliott. The effect of queuosine on tRNA structure and function. *Journal of Biomolecular Structure and Dynamcis*, 16(4): 757-774, June 1998.

265. Xin Wu, Zhao Li, Xuan-Xuan Chen, John S. Fossey, Tony D. James, and Yun-Bao Jiang. Selective sensing of saccharides using simple boronic acids and their aggregates. *Chemical Society Reviews*, 42(20): 8032-8048, October 2013.

266. Xiaoyu Li, Xushen Xiong, Kun Wang, Lixia Wang, Xiaoting Shu, Shiqing Ma, and Chengqi Yi. Transcriptome-wide mapping reveals reversible and dynamic N^1-methyladenosine methylome. *Nature Chemical Biology*, 12(5): 311-316, May 2016.

267. Huiqing Zhou, Simone Rauch, Qing Dai, Xiaolong Cui, Zijie Zhang, Sigrid Nachtergaele, Caraline Sepich, Chuan He, and Bryan C. Dickinson. Evolution of a reverse transcriptase to map N^1-methyladenosine in human messenger RNA. *Nature Methods*, 16: 1281–1288, September 2019.

268. Elzbieta Sochacka, Karina Kraszewska, Marek Sochacki, Milena Sobczak, Magdalena Janicka, and Barbara Nawrot. The 2-thiouridine unit in the RNA strand is desulfured predominantly to 4-pyrimidinone nucleoside under in vitro oxidative stress conditions. *Chemical Communications*, 47(17): 4914-4916, May 2011.

269. Arumugam Sudalai, Alexander Khenkin, and Ronny Neumann. Sodium periodate mediated oxidative transformations in organic synthesis. *Organic and Biomolecular Chemistry*, 13(15): 4374-4394, April 2015.

270. Jingyi Wang, Jiachen Shang, Zichen Qin, Aijun Tong, and Yu Xiang. Selective and sensitive fluorescence "turn-on" detection of 4-thiouridine in nucleic acids via oxidative amination. *Chemical Communications*, 55: 13096-13099, October 2019.

271. Xin Zhang and Shu-ou Shan. Fidelity of cotranslational protein targeting by the signal recognition particle. *Annual Review of Biophysics*, 43: 381-408, 2014.

272. Eva Maria Novoa, Mariana Pavon-Eternod, Tao Pan, and Lluís Ribas de Pouplana. A role for tRNA modifications in genome structure and codon usage. *Cell*, 149(1): 202-213, March 2012.

273. Matthew J. Eckwahl, Soyeong Sim, Derek Smith, Alice Telesnitsky, Sandra L. Wolin. A retrovirus packages nascent host noncoding RNAs from a novel surveillance pathway. *Genes and Development*, 29(6): 646-657, March 2015.

274. Adrian Gabriel Torres and Eulàlia Martí. Toward an Understanding of Extracellular tRNA Biology. *Frontiers in Molecular Biosciences*, 8: April 2021.

275. Dongwan Kim, Joo-Yeon Lee, Jeong-Sun Yang, Jun Won Kim, V. Narry Kim, and Hyeshik Chang. The Architecture of SARS-CoV-2 Transcriptome. *Cell*, 181(4): 914-921, May 2020.

276. Stephan Werner, Lukas Schmidt, Virginie Marchand, Thomas Kemmer, Christoph Falschlunger, Maksim V. Sednev, Guillaume Bec, Eric Ennifar, Claudia Höbartner, Ronald Micura, Yuri Motorin, Andreas Hildebrandt, and Mark Helm. Machine learning of reverse

transcription signatures of variegated polymerases allows mapping and discrimination of methylated purines in limited transcriptomes. *Nucleics Acids Research*, 48(7): 3734-3746, April 2020.

277. Ben Langmead, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biology*, 10(3): 2009.

278. Petr Danecek, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, Thomas Keane, Shane A. McCarthy, Robert M Davies, Heng and Li. Twelve years of SAMtools and BCFtools. *Gigascience*, 10(2): February 2021.

279. Pablo Cordero, Wipapat Kladwang, Christopher C. VanLang, and Rhiju Das. Quantitative Dimethyl Sulfate Mapping for Automated RNA Secondary Structure Inference. *Biochemistry*, 51(36): 7037–7039, August 2012.

280. Robert C. Spitale, Ryan A. Flynn, Qiangfeng Cliff Zhang, Pete Crisalli, Byron Lee, Jong-Wha Jung, Hannes Y. Kuchelmeister, Pedro J. Batista, Eduardo A. Torre, Eric T. Kool, and Howard Y. Chang. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, 519: 486-490, March 2015.

281. Xiaocheng Weng, Jing Gong, Yi Chen, Tong Wu, Fang Wang, Shixi Yang, Yushu Yuan, Guanzheng Luo, Kai Chen, Lulu Hu, Honghui Ma, Pingluan Wang, Qiangfeng Cliff Zhang, Xiang Zhou, and Chuan He. Keth-seq for transcriptome-wide RNA structure mapping. *Nature Chemical Biology*, 16(5): 489-492, February 2020.

282. Pilar Tijerina, Sabine Mohr, and Rick Russell. DMS Footprinting of Structured RNAs and RNA-Protein Complexes. *Nature Protocols*, 2(10): 2608–2623, October 2007.

283. Xi-Wen Wang, Chu-Xiao Liu, Ling-Ling Chen, and Qiangfeng Cliff Zhang. RNA structure probing uncovers RNA structure-dependent biological functions. *Nature Chemical Biology*, 17: 755-766, June 2021.

284. Meghan Zubradt, Paromita Gupta, Sitara Persad, Alan M. Lambowitz, Jonathan S. Weissman, and Silvi Rouskin. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nature Methods*, 14: 75-82, November 2016.

285. Silvi Rouskin, Meghan Zubradt, Stefan Washietl, Manolis Kellis, and Jonathan S. Weissman. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, 505: 701-705, January 2014.

286. Paula M. Godoy, Nirav R. Bhakta, Andrea J. Barczak, Hakan Cakmak, Susan Fisher, Tippi C. MacKenzie, Tushar Patel, Richard W. Price, James F. Smith, Prescott G. Woodruff, and David J. Erle. Large Differences in Small RNA Composition Between Human Biofluids. *Cell Reports*, 25(5): 1346-1358, October 2018.

287. Srimeenakshi Srinivasan, Ashish Yeri, Pike See Cheah, Allen Chung, Kirsty Danielson, Peter De Hoff, Justyna Filant, Clara D. Laurent, Lucie D. Laurent, Rogan Magee, Courtney Moeller, Venkatesh L. Murthy, Parham Nejad, Anu Paul, Isidore Rigoutsos, Rodosthenis Rodosthenous, Ravi V. Shah, Bridget Simonson, Cuong To, David Wong, Irene K. Yan, Xuan Zhang, Leonora Balaj, Xandra O. Breakefield, George Daaboul, Roopali Gandhi, Jodi Lapidus, Eric Londin, Tushar Patel, Robert L. Raffai, Anil K. Sood, Roger P. Alexander, Saumya Das, and Louise C. Laurent. Small RNA Sequencing across Diverse Biofluids Identifies Optimal Methods for exRNA Isolation. *Cell*, 177(2): 446-462, April 2019.

288. Thomas S.B. Schmidt, Jeroen Raes, and Peer Bork. The Human Gut Microbiome: From Association to Modulation. *Cell*, 172(6): 1198-1215, March 2018.

289. Verónica Lloréns-Rico, Joshua A. Simcock, Geert R.B. Huys, and Jeroen Raes. Single-cell approaches in human microbiome research. *Cell*, 185(15): 2725-2738, July 2022.

290. Ahmad Jomaa, Daniel Boehringer, Marc Leibundgut, and Nenad Ban. Structures of the *E. coli* translating ribosome with SRP and its receptor and with the translocon. *Nature Communications*, 7: January 2017.