UNIVERSITY OF CALIFORNIA
RIVERSIDE

Adductomic Approach to Evaluate the Epitranscriptome and Disease

A Dissertation submitted in partial satisfaction of
requirements for the degree of

Doctor of Philosophy

in

Environmental Toxicology

by

Gwendolyn Michelle Gonzalez

September 2021

Dissertation Committee:
Dr. Yinsheng Wang, Chairperson
Dr. Constance I. Nugent
Dr. Joseph C. Genereux

The Dissertation of Gwendolyn Michelle Gonzalez is approved:

_____

_____

_____

Committee Chairperson

University of California-Riverside

COPYRIGHTS ACKNOWLEDGEMENTS

The text and figures in Chapter 2, in part or in full, are a reprint of the material as it appears in ***Journal of Chromatography A*** 2020, 1623, 461181 and the supporting information therein. The coauthors, Dr. Yuxiang Cui and Dr. Pengcheng Wang, contributed to the writing and synthesis of ribonucleoside standards, respectively. The coauthor, Dr. Yinsheng Wang, listed in this publication directed and supervised the research which forms the basis of this chapter.

The text and figures in Chapter 3, in part or in full, are a reprint of the material as it appears in ***Analytical Chemistry*** 2018, 90, 6380-6384 and the supporting information therein. The coauthors, Dr. Xiaoxia Dai and Dr. Tianlu Wang, contributed to the proteomic analysis, YTHDF2 EMSA, and bisulfide sequencing. The coauthor, Dr. Yinsheng Wang, listed in this publication directed and supervised the research which forms the basis of this chapter.

The text and figures in Chapter 5, in part or in full, are a reprint of the material as it appears in ***Environmental International*** 2020, 144, 106021 and the supporting information therein. The coauthors, Dr. Allison Kuspco, Brennan H. Baker, Dr. Julia M. Knox, Dr. Yinan Zheng, Dr. Sheng Wang, Dr. Dou Cheng, Dr. Joel Schwartz, and Dr. Lifang Hou contributed the RNA extraction and RWE Q-PCR analysis. The coauthors, Dr.

Yinsheng Wang and Dr. Andrea A. Baccarelli, listed in this publication directed and supervised the research which forms the basis of this chapter.

The figures in Chapter 6 are in part a reprint of the material as it appears in ***Cell Reports*** 2021, 35, 109091. The coauthor, Dr. Na Li, contributed the PCIF1 Q-PCR and Western Blot analysis. The coauthors, Dr. Yinsheng Wang and Dr. Tariq Rana, listed in this publication directed and supervised the research which forms the basis of this chapter.

ACKNOWLEGEMENTS

I would like to first thank my advisor, Professor Yinsheng Wang, for the support and guidance. Since the first day I met him, he was welcoming and kind. He showed me what it meant to be excited about science. He dedicated time and effort to build an amazing support system in his research group, through his hard work and investment in each and every student. His knowledge and patience in mass spectrometry fundamentals has allowed me to grow to become the researcher I am today. For that I thank him. I would also like to thank Shuli Zhai for taking over as lab manager to order all the supplies we needed for our experiments and providing delish meals during the holidays. She opened her heart and her home to hold holiday parties for a relatively large research group. Both Dr. Wang and Shuli have ensured that I and the other members of the Wang research group had a support system to help them succeed.

I would also like to thank Dr. Connie Nugent and Dr. Joseph Genereux for not only serving as my dissertation committee members but being a great support system. Dr. Nugent is an inspiration for women in science. She not only has a successful research group that I came to know during my undergraduate research at the University of California-Riverside (UCR), but is also the undergraduate Dean of Student Affairs. Our conversations during our yearly meetings, showed me a path toward success and what it meant to have a life-work balance. I am grateful to have Dr. Genereux's research lab so close to Dr. Wang's. He was always available when I wanted to talk about maintaining mass spectrometers and research topics. I appreciated our talks walking around the third floor of the Chemical Science building. In addition, I made long-lasting friendships with the

members of his research group, including Dr. Maureen Montoya, Melody Sykes, and Guy Quanrud.

Next, I would like to thank my peers and collaborators for being patient, kind, and supportive. I am grateful for having numerous collaborators from different institutions, including Dr. Tariq Rana's entire research group from the University of California-San Diego, Dr. Ruth Steward and Dr. Hiep Tran from Rutgers University, Dr. Andrea A. Baccarelli and Dr. Allison Kupsco from Columbia Mailman School of Public Health, and finally Dr. Xiaoxia Dai and Dr. Xiaochuan Liu from University of California-Riverside. I am thankful for Dr. Ming Huang, Dr. Yuxiang Cui, Dr. Yang Yu, Jiekai Yin, and Yen-Yu Yang for their help in maintaining and troubleshooting instruments, for insightful discussions, and most importantly, for their friendship.

Equally important, I would like to acknowledge the financial support provided by the Environmental Toxicology Graduate Program at UCR and the National Institute of Health, for without their support this work would not be possible. I am grateful for the Graduate Student Affairs officers, Dawn Loyola and Antonio Knox, whom I went to countless times to sort out the complicated financial and administrated paperwork. They worked tirelessly to ensure that my sanity remained intact.

I would like to thank the current and former lab members: Dr. Xiaoxia Dai, Dr. Kailin Yu, Dr. Nicole Williams, Dr. Preston Williams, Dr. Nicholas Amato, Dr. Nathan Price, Dr. Rong Cai, Dr. Tianlu Wang, Dr. Lin Li, Dr. David Bade, Dr. Yuxiang Sun, Xuejiao Dang, Tianyu Qi, Shiyuan Guo, Dr. Andrew Kellum, Ying Tan, Su Guo, Zi Gao,

Xin Wang, Garrit Clabaugh, Yinan Wang, Feng Tang, Zhongwen Cao, and Xingyuan Chen. They have help shape many happy memories.

I would like to give a special thank you to my fiancé, Kevin Palencia, and children, Sophia and Breanna. They are encouraging of my work and patient with me throughout the years. Graduate school is difficult with many obstacles, but knowing that I could lean on of them provided me with the confidence to persevere. I could not ask for a better support system. Last but not least, I would like to thank my parents, siblings, and friends for their love and care.

ABSTRACT OF DISSERTATION

Adductomic Approach to Evaluate the Epitranscriptome and Disease

by

Gwendolyn Michelle Gonzalez

Doctor of Philosophy, Graduate Program in Environmental Toxicology
University of California, Riverside, September 2021
Dr. Yinsheng Wang, Chairperson

Chemical modifications on RNA play critical roles in post-transcriptional gene regulation. Like DNA epigenetic modifications, e.g., 5-methyl-2'-deoxycytidine (5-mdC) and 5-hydroxymethyl-2'-deoxycytidine (5-hmdC), they provide an additional layer of regulation by suppressing or enhancing gene expression. The various types of RNA, including ribosomal RNA (rRNA), transfer RNA (tRNA) and messenger RNA (mRNA), contain over 100 chemical modifications that make up the epitranscriptome. As more modified ribonucleosides are identified, so are the demands for sensitive and accurate methods for evaluating the epitranscriptome and its relationship with disease pathology.

In chapter 2, I developed a highly sensitive method for profiling post-transcriptional modifications in RNA by using nano-liquid chromatography-multi-stage mass spectrometry (nLC-MS/MS/MS). This method enabled simultaneous profiling of 31 modified ribonucleosides. To achieve a high-throughput scheduled selected-reaction monitoring, we assigned normalized retention time (iRT) values for each modified ribonucleoside with respect to the canonical ribonucleosides. We utilized the method for the identification of 20 modified ribonucleosides with minimal RNA input.

In chapter 3, we identified a RNA 5-methylcytidine (m^5C) reader protein, YTHDF2 based on quantitative proteomics. Even though m^5C is abundant in all RNA types, the exact role of m^5C remains elusive. YTHDF2 uses the same binding pocket to bind N^6-methyladenosine (m^6A) and m^5C. In addition, global m^5C levels are altered in HEK293T cells upon CRISPR-mediated knockout of *YTHDF2* gene.

In chapter 4, I developed a method to identify and quantify a novel RNA modification, 5-hydroxymethyluridine (5-hmrU) by using LC-MS/MS. I quantified 5-hmrU in total RNA and mRNA samples isolated from mouse tissues, cultured cells, and *Drosophila melanogaster*. In addition, we identified ten-eleven translocation (Tet) enzymes, known oxidizers of 5-methylcytosine (m^5C) to 5-hydroxymethylcytosine (5-hmC) in DNA and RNA, to also mediate the oxidation of 5-methyluridine (m^5U) to 5-hmrU.

In chapters 5 and 6, we employ the epitranscriptome global profiling method to evaluate the role of the epitranscriptome on environmental toxin exposure and disease. Chapter 5 focuses on evaluating how levels of ribonucleosides are affected by air pollution and tobacco smoke. Total RNA samples isolated from peripheral blood from truck drivers and office workers were taken for LC-MS/MS/MS analysis. I found that smoking in men was associated with a decrease in global m^6A level compared to non-smokers. By contrast, black carbon exposure from air pollution resulted in an increase in m^6A level. In chapter 6, I evaluated the role of epitranscriptome modifications and its regulatory proteins upon infection with HIV-1 and SARS-CoV-2. HIV infection resulted in attenuated PCIF1 protein level resulting in reduced m^6A_m level in host MT4 cell lines. In addition, I observed

similarities in modified ribonucleoside profile in viral RNA genomes of HIV-1 and SARS-CoV-2.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Cells utilize various mechanisms to cope with homeostatic stress. Epigenetics is a form of gene expression regulation that does not alter the primary sequence but employs chemical modifications on DNA, RNA and proteins. The various epigenetic marks impart differing functions.

Some common DNA epigenetic marks, including 5-methyl-2'-deoxycytidine (m^5dC) and 5-hydroxymethyl-2'-deoxycytidine (5-hmdC), repress and enhance gene expression, respectively.[1] In the genome, 5-mdC is commonly found at CpG islands.[2] It is reported that approximately 70% of all human CpG islands found at promoters are methylated, resulting in repression of downstream genes.[2] Conversely, the presence of m^5dC's oxidization product, 5-hmdC, is an epigenetic mark in its own respect and is associated with enhancement of gene expression.[3]

DNA and histone epigenetic marks work synergistically to mediate transcription of genes.[2] m^5dC marks on CpG islands signal histone methyltransferases to hypermethylate histones. Interestingly, the different methylation states, including mono-, di-, and tri-methylations, result in varying functions. Histone 3-lysine 9 trimethylation (H3K9me3) and histone 3-lysine 27 trimethylation (H3K27me3) result in repression and are commonly found in areas where CpG islands are hypermethylated.[4] Conversely, mono-, di-, and tri-methylations of histone 3 at lysines 4, 36, and 79 are activation marks.[4] However, methylations are not the sole histone epigenetic mark. Other histone modifications include acetylations, ubiquitinations, and phosphorylations.[4] The different types of modifications as well as their positions on histone tails can result in altered gene expression.

Identifying epigenetic marks and their functions is vital in understanding their dysregulation in disease. The varying types of cancer and other diseases can result from entire reorganization of these epigenetic marks, such as hypomethylation of the genome. Therefore, expanding our knowledge in gene expression regulation can assist our understanding of disease pathology. One form of gene expression regulation that has only begun to be explored is the presence of modified ribonucleosides, which are referred to as the epitranscriptome.

1. The Epitranscriptome

An emerging concept in gene expression regulation is the presence of modified ribonucleosides in mRNA, tRNA, and rRNA that make up the epitranscriptome (**Figure 1.1**). Over 100 modifications mediate structure, function, translation, splicing, stability, and interactions with other biomolecules.[5, 6] In addition to simple base modifications, including methylations and acetylations, other modifications contain backbone methylations and bulky adducts, e.g., 5-*S*-methoxycarbonylhydroxymethyluridine (mchm^{5}U).

A few modified ribonucleosides, including 5-methylcytidine (m^{5}C) and pseudouridine (ψ), are well documented in various mammalian cell RNA, including tRNAs, rRNAs, and mRNA. [7-10] In tRNA, m^{5}C is commonly found at the 48th and 49th position near the T-ψ-C stem, suggesting an influence in tRNA stability.[8] In addition, *in-vitro* experiments demonstrated the importance of m^{5}C tRNA methylation on Mg^{2+} binding.[11] In mRNA, ψ is tightly regulated in response to environmental cues, including glucose deprivation.[12] However, in different types of RNA, the exact function of m^{5}C and

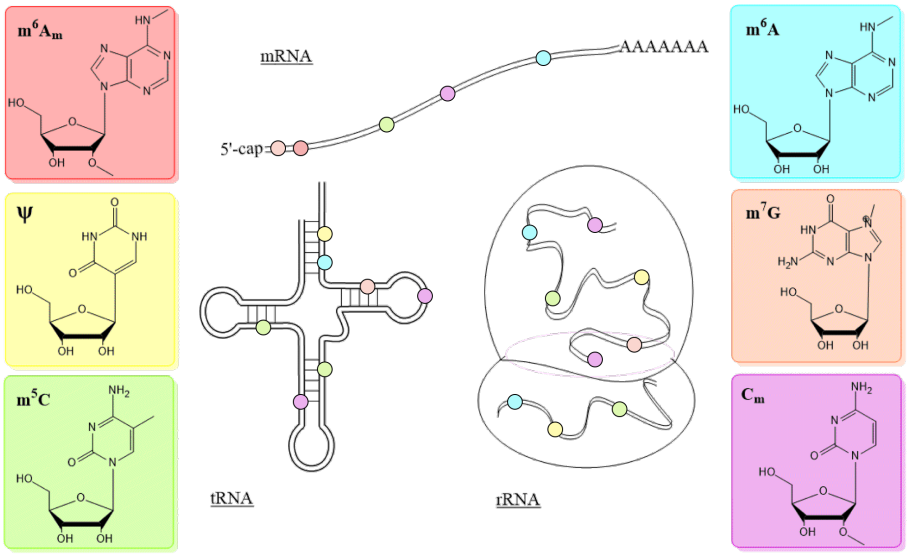


Figure 1.1 Epitranscriptome modifications in mRNA, tRNA, and rRNA.

ψ are not well understood, so are true for the majority of other modified ribonucleosides.

Over the past few decades, modified ribonucleosides on mRNA were gaining interest. Modifications, such as *N*7-methylguanosine (m^7G) and N^6,2'-*O*-dimethyladenosine (m^6A_m), are found in the 5'-cap region of mRNA, where they mediate cap-dependent translation and recruitment of small (40S) ribosomal subunit and translation initiation factors.[6, 13] In addition, the presence of m^6A_m increases mRNA stability by inhibiting the mRNA decapping enzyme, DCP2.[6] However, viral RNA and uncapped RNA are translated through a cap-independent mechanism by internal ribosomal entry site (IRES) motif and the presence of N^6-methyladenosine (m^6A) in the 5' untranslated region, respectively. [13, 14] In mammalian cells, m^6A occurs at a frequency of 0.1-0.4% per adenosine in mRNA and is one of the most abundant internal mRNA modifications.[15] Antibody-based sequencing has determined that DRACH is an m^6A consensus sequence

with D representing A/G/U, R representing A/G, and H representing A/C/U.[16, 17] In addition, m^6A is enriched at stop codons, 3'-untranslated regions (3'-UTRs), and few are found on long exons.[15] Investigations into mRNA modifications have revealed the importance of m^7G, m^6A_m, and m^6A in mRNA stability and translation efficiency. Follow up investigations are necessary to decipher the presence of other modified ribonucleosides, such as N^4-acetylcytidine (ac^4C), and their functions in mRNA metabolism. More importantly, the various aspects of mRNA metabolism governed by modified ribonucleoside regulators, like m^6A, are vital in our understanding of the origins of diseases.[15]

2. Epitranscriptomic regulators

Cells are equipped with mechanisms to rapidly respond to stimuli. Epigenetic modifications and their dynamic nature is common among all modes of epigenetic regulation. Genome and histone modifications are regulated by readers, writers, and erasers to mediate the transcriptional fate of genes. It has recently been recognized that mRNA levels correlate well with protein levels for only some protein-coding genes.[18] This discrepancy has paved the road for investigations into post-transcriptional regulation of gene expression.

Recently, several modified ribonucleosides within the epitranscriptome, including mRNA, were found to be dynamic and reversible. Tissue-dependent distribution of epitranscriptomic modifications was one of the first indications of regulatory proteins.[7, 19, 20] Readers, writers, and erasers govern the modified ribonucleoside level and function through translation and stability of the transcript of interest.[21-23] For instance, YTHDF2

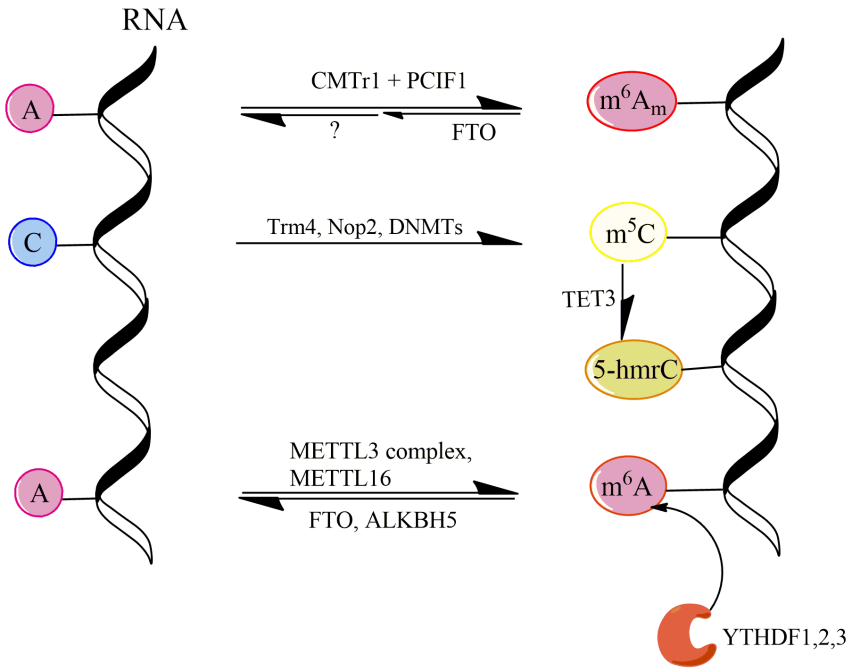


Figure 1.2 The reversible and dynamic post-transcriptional regulation of modified ribonucleosides.

binding of m^6A- containing RNA results in degradation of that particular transcript. In the following sections, I will review some of the key epitranscriptome writers (methyltransferases), erasers (demethylases), and readers (RNA binding proteins) (**Figure 1.2**).

2.1 Methyltransferases

Methylation is one of the most commonly identified and researched modifications. They occur in all types of ribonucleosides, including cytidine, uridine, adenosine and guanosine.[24, 25] The most investigated methylations exist in bacteria tRNA. However, a few of the modified ribonucleosides have also been identified in mRNA, including m^5C, m^7G, 2'-*O*-methylations (N_m), and m^6A.

Methylations are installed on RNA by an enzyme called methyltransferases. There are 3 main classes of m^5C methyltransferase: tRNA-specific methyltransferase (Trm4), nucleolar protein 2 (Nop2), and DNA methyltransferase (DNMTs). Trm4 is important in early tRNA metabolism by methylating positions 34 and 40 of $tRNA^{Leu}$ (CAA) and $tRNA^{Phe}$ (GAA).[26] Further studies identified cysteine 310 as a vital catalytic residue and the ability of the N-terminus to enhance cysteine availability for binding to m^5C.[26] Through sequence complementarity, Trm4's human homolog, NSUN2, was identified and was found to be involved in c-Myc dependent proliferation.[8] Even though NSUN2 has sequence similarity, it is unable to methylate $tRNA^{Leu}$ (GAA).[8]

Another methyltransferase is Nop2, in yeast, responsible for maturation of 60S rRNA from 26S rRNA in a methylation-dependent manner.[27] Nop2's key cysteine residue (Cys424) is essential for yeast viability.[27] The remainder of the Nop2/NSUN family are highly conserved in mammals. Deletion of critical NSUN genes, NSUN5A/B/C, results in a mental disorder called Williams-Beuren syndrome.[28]

The most well-studied nucleic acid methyltransferases are the DNA (cytosine-5)-methyltransferases (DNMTs). DNMTs are responsible for inheritance and maintenance of DNA methylation patterns.[29] Until fairly recently, DNMTs were only known to mediate DNA methylation. However, recent investigations have demonstrated its role in RNA methylation. DNMT2 was the first of this family shown to methylate RNA, specifically $tRNA^{Asp}$. Using DNMT2 modeling and *in-vitro* experiments, it was illustrated that tyrosine residue impedes DNA but not RNA binding.[30]

In addition to m^5C, there are a few modifications that have been identified in mRNA. The *N*7-methylguanosince (m^7G) cap is essential for cap-dependent mRNA translation.[31] Following the m^7G cap, mRNA typically contains a 2'-*O*-methylation on the first nucleotide, installed by CMTr1.[32] Phosphorylated serine RNA Pol II acts as a scaffold for RNMT, the *N*7-methylguanosine methyltransferase, and CMTr1 to co-transcriptionally methylate mRNA with essential modified ribonucleosides.[32] Typically, if the first nucleotide is 2'-*O*-methyladenosine (Am), a second methylation event can occur to form m^6A_m by PCIF1.[32-34] Upon further investigation, PCIF1 was found to be cap-specific and requires the 2'-*O*-methylation to be present.[32, 34] However, PCIF1 is not responsible for internal m^6A formation.[33] Human methyltransferase-like 3 (METTL3) and 14 (METTL14) form a heterodimer with Wilms' tumor-1 associating protein (WTAP) making a stable complex which mediates m^6A methylation in a sequence-dependent manner.[17, 35] The METTL3 methyltransferase complex is primarily responsible for the most common internal mRNA modification.[35] METTL3 binds *S*-adenosyl-L-methionine (SAM), while METTL14 and WTAP are important in substrate recognition and nuclear speckle localization, respectively.[17, 35-38] The METTL3-METTL14 complex methylates not only mRNA, but to a lesser degree, microRNA (miRNA), long intergenic non-coding RNA (LincRNA), and pseudogenes.[17] Recently, METTL16 was found to methylate approximately 20% of all m^6A, but differ from those methylated by METTL3-METTL14 complex, suggesting substrate and sequence specificity.[39] METTL16 is also associated with methylated U6 small nuclear RNA (snRNA) and the 3'-UTR of methionine adenosyltransferase 2A (MAT2A).[39] m^6A methyltransferases is the most heavily studied

mono-methylations in RNA. The ability for m^6A to be removed from RNA is due to regulatory proteins, including demethylases.

2.2 Iron(II) – and 2-Oxoglutarate-dependent Dioxygenases

The iron (II)-and-2-oxoglutarate-dependent dioxygenases (2OG-D), also known as the α-ketoglutarate-dependent dioxygenases, function in DNA repair and epigenetic regulation. This family consists of 80 enzymes that are known to coordinate Fe(II) at 2-histidine-1-carboxylate motif and bind 2OG within its metal ion core.[40] Using 2-OG as a cofactor with Fe(II), the 2OG-D can oxidize methyl groups into a hydroxymethyl group.[15]

Initial investigations demonstrated the 2OG-D's capability to demethylate lysine residues in proteins.[40] For example, a subset of 2OG-Ds, known as the Jumonji domain (JMJD)-containing proteins hydroxylate the 5th carbon of lysine or arginine, important for the function of RNA splicing proteins and histones.[40] The different families, like LSD and Jumonji domain protein, are not only histone lysine-dependent, but are also methylation state-dependent, e.g. mono-, di-, and tri-methylations.[41-44] The LSD family of proteins are only capable of demethylating lysine mono- and di-methylations.[41] However, the Jumonji domain proteins are capable of demethylating all methylation states.[44] That being said, the different members of the Jumonji domain family, including KDM4A, KDM4B, and KDM4C, have differing activities toward various methylated lysines on histones and their methylation states.[4]

Alkylated nucleic acids can also be oxidized by the 2OG dioxygenases. AlkB is one of the first enzymes identified within the 2OG-D family to function on nucleic acids in *E. coli*.[40] Early evidence suggested that 2OG-D functions primarily for repair of DNA

lesions induced by alkylating agents.[45] Within the past few decades, AlkB was identified as a RNA demethylase, first misunderstood as a RNA repair protein, but now recognized as a mono-methylated ribonucleoside demethylase.[46] AlkB oxidizes 3-methylcytidine (m^3C) and *N*1-methyladenosine (m^1A), resulting in the spontaneous loss of formaldehyde to form the corresponding unmethylated nucleoside.[40] The known nucleosides that AlkB can oxidize have expanded to 1-methylguanosine, 3-methylthymine, and N^4-methylcytidine.[40] In mammals, AlkB homologs are referred to as ALKBH1-8 and fat-mass and obesity associated gene (FTO). Many members within the ALKBH family have no known substrate in proteins or nucleic acids. For instance, ALKBH6 and ALKBH7 knockout in HEK293T cells has demonstrated their role in necrosis, but their binding partners and substrates remain poorly understood.[47]

FTO, alternatively called ALKBH9, is shown to mediate human obesity and energy homeostasis.[48] FTO was originally revealed to oxidatively demethylate 3-methylthymine (m^3T) and 3-methyluridine (m^3U) *in-vitro.*[49] However in 2011, N^6-methyladensine (m^6A) was identified as one of FTO's primary substrates.[48] ALKBH5 can also oxidatively demethylate m^6A *in vitro* and *in vivo.*[50] Similar to FTO, ALKBH5 mediates mRNA export, metabolism, and processing in nuclear speckles.[50] From a pathological prospective, both ALKBH5 and FTO play crucial roles in cancer, viral infection, and fertility.[48, 50, 51] However, FTO is a promiscuous eraser as it demethylates internal m^6A in mRNA and m^6A_m found in the 5' cap region of mRNA,[6] revealing the potential of dynamic regulation of multiple modified ribonucleosides by a single demethylase.

ALKBH8 is the only member in the ALKBH family to have both a methyltransferase and hydroxylase domain.[52, 53] To perform as a methyltransferase, ALKBH8 and Trm112b recognize and methylate 5-carboxymethyluridine (cm^5U) to 5-methylcarboxymethyluridine (mcm^5U) at wobble positions of tRNA.[52] ALKBH8 can subsequently oxidize mcm^5U to 5-S-methoxycarbonylhydroxymethyluridine ($mchm^5U$).[53] ALKBH8's importance in cell survival has only been monitored in *E. coli*, where its' genetic depletion results in sensitivity to DNA damaging agents.[54] However, ALKBH8's function in mammalian cells has not been investigated. Similar to many of the other members of the ALKBH family, a thorough assessment of other potential targets have yet to be identified.

Like the ALKBH family's action on DNA and RNA, ten-eleven translocation (Tet) proteins are also epigenetic and epitranscriptomic regulators.[19, 55] Tet proteins are widely understood as the 5-methyl-2'deoxycytidine (m^5dC) hydroxylase on CpG islands of DNA.[56] 5-hydroxymethyl-2'-deoxycytidine (5-hmdC) is a m^5dC oxidation product and is regarded as an epigenetic marker for active transcription on its own.[57] Tet can further oxidize 5-hmdC to 5-formyl-2'-deoxycytidine (5-fodC) and 5-carboxyl-2'-deoxycytidine (5-cadC). 5-fodC and 5-cadC in DNA play a role in the demethylation process to restore the unmethylated 2'-deoxycytidine through base-excision repair.[55] In short, thymine DNA glycosylase recognizes 5-fodC and 5-cadC to excise the modified base and signal the base excision repair (BER) pathway to restore an unmodified 2'-deoxycytidine.[56] In RNA, the 5-hydroxymethyl-, 5-formyl-, and 5-carboxyl-cytidine modifications were identified.[19, 55] However, their exact functions in RNA remain elusive.

2.3 RNA binding proteins

It is now understood that modified ribonucleosides are marks to facilitate and discriminate mRNAs for a specific fate using RNA-binding proteins. RNA-binding proteins play a critical role in RNA stability, turnover, and translation efficiency.[58-60] Investigations primarily focus on *N*7-methylguanosine (m^7G) and m^6A-mRNA binding proteins due their high abundance.

mRNA is protected from degradation by the presence of a poly-A tail and a 5'-cap.[61] The 5'-m^7G cap on mRNA is the most heavily studied. Cap-dependent degradation is mediated by RNA-binding proteins to recognize the poly-A tail and m^7G cap of a mRNA sequence of interest.[61] The process is initiated through deadenylation of the mRNA by exonucleases.[62] In yeast, Dcp1 hydrolyzes the m^7G cap to release m^7GDP.[62] In mammalian systems, hDcp2 is recognized as the human mRNA decapping enzyme that is necessary and sufficient of proper mRNA decay.[62] Further studies demonstrated the ability for other modifications to influence decapping, including m^6A_m. The presence of m^6A_m as the first nucleotide after the 5'-cap can also inhibit hDcp2 activity.

Another mechanism of mRNA degradation and stability is based on m^6A-binding proteins. Among the known modified ribonucleosides, m^6A has the best characterized readers that bind directly[21] or indirectly by secondary structures induced by the presence of m^6A.[58, 63, 64] The YTH family members are known as direct binders of m^6A, m^5C, and m^1A to influence RNA stability.[21, 65, 66] YTHDF1 and YTHDF3 enhance and promote translation, respectively, in an m^6A-dependent manner.[67, 68] YTHDF1 interacts directly with translation machinery, including translation initiation factor, Elf3, in a cap-dependent

translation manner.[15, 69] Conversely, YTHDF2 binding hinders translation by promoting mRNA degradation.[70] YTHDF2 carries the m^6A-containing mRNA to processing bodies by targeting m^6As at the stop codon, 3'-untranslated region (3'-UTR), and coding region, indicating its role in modulating mRNA stability.[70]

As described in the past few sections, the regulators of the epitranscriptome govern the fate and function of all types of ribonucleoside modifications. Some of the readers, writers and erasers have been identified for m^6A. However, the proteins mediating other modification level and function within key regulator families, including the 2OG-Ds, remain elusive.

3. Epigenetic/epitranscriptomic response to diseases and environmental exposure

Cancer is linked to aberrant epigenetic and epitranscriptomic regulation. It is now recognized that one of the hallmarks of cancer is abnormal m^6A levels.[15, 71-73] A common characteristic in various cancers, including breast cancer, is induction of hypoxia-inducible factors and known m^6A regulatory proteins, like ALKBH5, resulting in altered m^6A levels.[15, 74] In addition, high expression of FTO in acute myeloid leukemia (AML), a malignant hematopoietic cancer, is associated with increased tumor growth.[75] Moreover, FTO knockdown resulted in diminished tumor growth, indicating its ability to promote self-renewal.[75] While investigation in human lung cancer demonstrated METTL3's ability to enhance translation of oncogene by interacting with translation initiation factors, including nuclear cap-binding protein subunit 1 (CBP80) and eukaryotic translation initiation factor (EIF4E).[15]

Investigations into environmental toxins have also indicated altered gene expression. Metals and metalloids, like arsenite, found in ground water and food supply can enter cells to cause a variety of effects. Arsenite and nickel exposure can cause an array of diseases by altering gene expression and protein functions.[76-78] High exposure is common in drinking water, food and industrial waste supplies in countries where arsenite levels are not tightly regulated.[79] Investigations into chronic exposure through drinking water in mice resulted in learning and memory impairment, and anxiety-like behavior.[80] In humans, chronic exposure causes impaired cognitive abilities.[81] Some possible mechanisms are related to low levels of dopamine at the synaptic cleft.[80] At a molecular level, arsenite can displace metals, such as iron and zinc, critical for metal ion-dependent protein, like the 2OG-Ds or zinc finger proteins.[78, 82, 83] As a result, histone and DNA hypomethylation is commonly observed in exposed human subjects.[83] However, only a few studies have been conducted to evaluate the effect of arsenite exposure on ribonucleoside modifications, specifically methylations regulated by 2OG-Ds. One study revealed elevated levels of m^6A in dopaminergic neurons associated with diminished FTO expression caused by arsenite.[80]

Until recently, the impact of modified ribonucleosides was and still is not well understood especially in complex diseases resulting from environmental exposure, including air pollution. This is partly due the limited knowledge of the regulatory proteins that mediate modified RNA levels. Therefore, epidemiological studies that have been conducted are limited to m^6A in RNA.

3.1 Air Pollution and Tobacco Smoking

Since the industrial revolution, air quality has become the major health issue. The World Health Organization's 2016 analysis suggests that 4.2 million deaths were a result from chronic exposure to poor air quality world-wide.[84] As a whole, government organizations have created legislation for corporations and automobile emissions to lessen their carbon footprint. However, the high cost of converting to alternative resources have hindered progress. As a result, people within low-income communities or live in areas where air quality is not tightly regulated are at an increased risk of cancer and illness, including acute lower respiratory infection, lung cancer, ischemia heart disease and stroke.[84]

A few studies have been conducted to elucidate the relevant pathways mediating air pollution-driven illness. Many of the air pollution studies have also linked tobacco smoke using similar disease pathologies. For instance, tobacco smoking-related cancers is a preventable disease worldwide associated with changes in genome-wide methylation at CpG islands in whole and peripheral blood.[85] Perinatal exposure from maternal smoking during pregnancy results in DNA hypomethylation in cord and whole blood.[85] Analogously, a study conducted using peripheral blood samples from Beijing truck drivers found elevated levels of 5-hmrC in DNA in exposure to high levels of particulate matter less than 10 μm.[86] Overall the results from the described studies suggest that there is reprogramming of epigenetic regulators during disease progression. However, only a few investigations have been conducted on epitranscriptomic regulators and their relationship to environmental exposure. By expanding the technologies available to investigate the

epitranscriptome, researchers will be able to investigate how our environmental exposure confers adverse human health consequences by perturbing the epitranscriptome.

4. Methods to Assess the Epitranscriptome

Over the past 50 years, targeted analysis methods have been developed to understand the epitranscriptome topology, function, and regulation. Common techniques are based on the modifications' separation by physiochemical properties and differential chemical reactivities.[8]

Methods utilizing physicochemical separation are thin-layer chromatography (TLC) and reverse phase high-performance liquid chromatography (HPLC).[8] Two-dimensional TLC is one of the first techniques to identify the presence of modified ribonucleosides in tRNA.[87] The experimental design consists of ^{32}P labeling of the 3' end of modified ribonucleoside following enzymatic digestion, but in doing so destroying any sequence context information.[87] Similar to TLC, liquid chromatography (LC) methods also require nuclease digestion and dephosphorylation prior to separation using a reverse-phase C18 column.[24, 88] However, both techniques require substantial amounts of RNA per experiment.

Other methods utilize modified ribonucleosides' chemical reactivity for sequencing techniques. Normally, the chemical modifications on RNA do not alter base pairing properties or inhibit reverse transcription, making their identification, quantification, and sequencing difficult.[89] Therefore, methods were developed utilizing antibodies or clickable biomolecules followed by sequencing to identify modified ribonucleosides based on selective purification and mismatches compared to control, respectively. These methods

can provide sequence information, including in mRNA and tRNA, but are limited to a few number of modifications including m^6A, m^5C, pseudouridine (ψ), and 2'-O-methylations (A_m, G_m, C_m, and U_m).[5, 10, 12, 90]

For instance, antibody-based m^6A-seq analysis revealed a consensus sequence, RRACH, and m^6A topology in mRNA.[5, 89] The antibody-enriched sequence fragments are labeled with a coded primer then sequenced.[5] Currently, sequencing requires a few micrograms of mRNA and excellent resolution of approximately 18 nucleotides around the modification, but are limited by the number of antibodies commercially available for epitranscriptomic modifications and lack of specificity.[89] For instance, the frequently used m^6A antibody also recognizes m^6A_m.[5, 6] Therefore, initial m^6A-seq analysis mischaracterized the presence of m^6A at the transcription start sites (TSS).[6]

Another form of modified ribonucleosides sequencing technique is based on click chemistry. Sequencing ψ modifications through click chemistry has enhanced our ability to identify ψ at single-nucleotide resolution.[12] Treatment with *N*-cyclohexyl-*N*′-(2-morpholinoethyl)-carbodiimide metho-*p*-toluensulphonate (CMC) prior to sequencing to form N_3-[N-cylcohexyl-*N*'-β-(4-methylmorpholinium)ethylcarbodiimide]-ψ (N_3-CMC-ψ), a known reverse transcription terminator.[10, 12] However, CMC can covalently link uridine, guanosine and ψ residues to form uridine-CMC (U-CMC), guanosine-CMC (G-CMC) an N_3-CMC-ψ, respectively.[12] To remove U-CMC and G-CMC adducts, alkaline treatment results in hydrolysis of only U-CMC and G-CMC, leaving only ψ-CMC to prematurely terminate reverse transcriptase during sequencing.[10] This method has a false

discovery rate of 0.1%.[12] However, the method does not provide sequence information on the 3'-end of the ψ.

A major limitation with modified ribonucleoside sequencing methods is their inability to simultaneously evaluate multiple modifications at a transcriptome-wide scale. A single method, called RBS-Seq, can simultaneously sequence m^5C, ψ, and m^1A at single-base resolution.[91] Utilizing bisulfite treatment, m^5C and m^1A are detected using signature mismatch, while ψ is identified with 1-2 base deletions.[91] Moreover, the method has surpassed typical sequencing methods by simultaneously identifying crucial ribonucleoside modification sites within the same sequence of tRNA, rRNA, or mRNA.[91] Even with this advancement in epitranscriptome monitoring, the method lacks the ability to monitor a multitude of possible ribonucleoside modifications.

4.1 Liquid chromatography-tandem mass spectrometry methods

Tandem mass spectrometry (MS/MS) methods have provided new advancements in epitranscriptomic research. These methods combine LC separation and mass spectra information which have led to the identification of novel nucleosides in a highly sensitive manner.[8, 24, 88] In addition, LC-MS/MS methods can monitor several modifications simultaneously. Over the past decade, the number of modified ribonucleosides that could be identified in a single injection has risen from 20 to 67 nucleobases in tRNA.[24, 88] However, most methods target highly modified tRNA to identify the effect of stressors on modification level.[92]

A typical LC-MS/MS method follows a similar work flow. First, highly purified RNA is required for accurate identifications of modified ribonucleosides within a specific

RNA type, i.e. tRNA, mRNA, and microRNA. Obtaining pure samples is essential to ensure the modification and its level is not due to a contaminate of another type of RNA, such as rRNA. Total RNA consists primarily of approximately 80% rRNA and 15% tRNA.[93] Therefore, when isolating RNA types of low abundance, like tRNA or mRNA, extra purification steps and thorough assessment of purity, by bioanalyzer analysis, are required. This is important because of step 2, hydrolysis of RNA into mononucleosides. Different experimental set-ups and preferences have resulted in slightly different digestion protocols.[7, 24] In general, endo- and exo-nucleases are used to enzymatically digest RNA. The resulting mixture will consist of mononucleotides of the sample. Therefore, any contamination in a purified RNA sample of any RNA type will be indistinguishable from one another once digested into mononucleotides. In addition to nucleases, digestions can also incorporate a dephosphorylation step for nucleoside production. Finally, the sample is subject to reverse phase-liquid-chromatography mass spectrometry (LC-MS) analysis. Various C18 stationary phase materials are utilized to enhance the separation capabilities of modified ribonucleosides.[7, 88] Mono-methylated ribonucleosides, such as *N*3-methylcytidine (m^3C), N^4-methylcytidine (m^4C), and m^5C, have very close relative retention time.[94] Therefore, any chromatographic shift can result in co-elution of any two or more mono-methylated cytidines. By modifying LC gradients and optimizing stationary phase materials, methods can accurately and reproducibly separate mono-methylated nucleosides. The modified ribonucleosides are then identified and quantified using a mass spectrometer. Multi-stage tandem mass spectrometry (MS^n) has allowed for identification of different modified ribonucleosides based on characteristic fragmentation. Nucleobases,

in both DNA and RNA, have a unique fragmentation due to cleavage of the *N*-glycosidic bond.[54, 92, 95] This provides a clear way of identifying a modification of interest by its *m/z* value and fragmentation pattern.[96]

In addition to global profiling of ribonucleosides, LC-MS/MS methods have been developed for modified ribonucleoside sequencing.[97, 98] However, this LC-MS/MS method is limited to tRNA. mRNA transcripts are highly variable among cells, making it difficult to purify enough mRNA transcripts for a MS-based sequencing analysis.[99] For tRNA analysis, enzymatic digestion using RNase T1 results in cleavage at the 3' end of guanosine. The short segments are subjected to LC-MS/MS analysis to identify not just one modified ribonucleoside per tRNA segment but several,[100] thus broadening our knowledge of what modifications are present in a tRNA of interest but also provides sequence context information.

LC-MS/MS methods have provided an approach to identify the interplay of disease and modified ribonucleoside level. Cells, like yeast, are known to respond to cellular stress by altering modified RNA level.[92] Therefore, LC-MS/MS approaches provide an effective tool for understanding RNA biology.

5. Scope of this dissertation

As established by the previous sections, there is a recent appreciation of modified ribonucleosides and their impact on heath. The goal of the presented studies is to design a sensitive method to accurately identify and quantify epitranscriptome modifications. In addition, we have utilized highly sensitive LC-MS/MS methods to identify the regulators, including readers and erasers, responsible for the highly dynamic and reversible

modifications in RNA. Finally, we applied our method to evaluate the role the epitranscriptome in air pollution, tobacco smoke, and viral infection.

References

1. Tang, Y.; Zheng, S.-J.; Qi, C.-B.; Feng, Y.-Q.; Yuan, B.-F., Sensitive and Simultaneous Determination of 5-Methylcytosine and Its Oxidation Products in Genomic DNA by Chemical Derivatization Coupled with Liquid Chromatography-Tandem Mass Spectrometry Analysis. *Analytical Chemistry* **2015,** *87* (6), 3445-3452.

2. Baylin, S. B., Tying it all together: epigenetics, genetics, cell cycle, and cancer. *Science (New York, N.Y.)* **1997,** *277* (5334), 1948-1949.

3. Kroeze, L. I.; van der Reijden, B. A.; Jansen, J. H., 5-hydroxymethylcytosine: An epigenetic mark frequently deregulated in cancer. *Biochimica et biophysica acta* **2015,** *1855* (2), 144-154.

4. Zhao, Z.; Shilatifard, A., Epigenetic modifications of histones in cancer. *Genome Biology* **2019,** *20* (1), 1-16.

5. Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.; Sorek, R.; Rechavi, G., Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **2012,** *485* (7397), 201-6.

6. Mauer, J.; Luo, X.; Blanjoie, A.; Jiao, X.; Grozhik, A. V.; Patil, D. P.; Linder, B.; Pickering, B. F.; Vasseur, J.-J.; Chen, Q.; Gross, S. S.; Elemento, O.; Debart, F.; Kiledjian, M.; Jaffrey, S. R., Reversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nature* **2017,** *541* (371-375).

7. Fu, L.; Amato, N. J.; Wang, P.; McGowan, S. J.; Niedernhofer, L. J.; Wang, Y., Simultaneous Quantification of Methylated Cytidine and Adenosine in Cellular and Tissue RNA by Nano-Flow Liquid Chromatography–Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal. Chem.* **2015,** *87ha* (15), 7653-7659.

8. Motorin, Y.; Lyko, F.; Helm, M., 5-methylcytosine in RNA: Detection, enzymatic formation and biological functions. *Nucleic Acids Research* **2009,** *38* (5), 1415-1430.

9. Motorin, Y.; Helm, M., TRNA stabilization by modified nucleotides. *Biochemistry* **2010,** *49* (24), 4934-4944.

10. Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; Leon-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.; Satija, R.; Lander, E. S.; Fink, G.; Regev, A., Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **2014,** *159* (1), 148-162.

11. Chen, Y.; Sierzputowska-Gracz, H.; Guenther, R.; Everett, K.; Agris, P. F., 5-Methylcytidine is required for cooperative binding of Mg2+ and a conformational

transition at the anticodon stem-loop of yeast phenylalanine tRNA. *Biochemistry* **1993,** *32* (38), 10249-53.

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

13. Meyer, K. D.; Patil, D. P.; Zhou, J.; Zinoviev, A.; Skabkin, M. A.; Elemento, O.; Pestova, T. V.; Qian, S. B.; Jaffrey, S. R., 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* **2015,** *163* (4), 999-1010.

14. Stoneley, M.; Willis, A. E., Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene* **2004,** *23* (18), 3200-7.

15. Dai, D.; Wang, H.; Zhu, L.; Jin, H.; Wang, X., N6-methyladenosine links RNA metabolism to cancer progression. *Cell Death Dis* **2018,** *9* (2), 124.

16. Linder, B.; Grozhik, A. V.; Olarerin-George, A. O.; Meydan, C.; Mason, C. E.; Jaffrey, S. R., Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **2015,** *12* (8), 767-772.

17. Ping, X.-L.; Sun, B.-F.; Wang, L.; Xiao, W.; Yang, X.; Wang, W.-J.; Adhikari, S.; Shi, Y.; Lv, Y.; Chen, Y.-S.; Zhao, X.; Li, A.; Yang, Y.; Dahal, U.; Lou, X.-M.; Liu, X.; Huang, J.; Yuan, W.-P.; Zhu, X.-F.; Cheng, T.; Zhao, Y.-L.; Wang, X.; Rendtlew Danielsen, J. M.; Liu, F.; Yang, Y.-G., Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell research* **2014,** *24* (2), 177-89.

18. Vogel, C.; Marcotte, E. M., Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **2012,** *13* (4), 227-232.

19. Fu, L.; Guerrero, C. R.; Zhong, N.; Amato, N. J.; Liu, Y.; Liu, S.; Cai, Q.; Ji, D.; Jin, S.-G.; Niedernhofer, L. J.; Pfeifer, G. P.; Xu, G.-L.; Wang, Y., Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **2014,** *136* (33), 11582-5.

20. He, C., Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.* **2010,** *6* (12), 863-5.

21. Du, H.; Zhao, Y.; He, J.; Zhang, Y.; Xi, H.; Liu, M.; Ma, J.; Wu, L., YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. *Nature Communications* **2016,** *7*, 12626-12626.

22. Fu, Y.; Jia, G.; Pang, X.; Wang, R. N.; Wang, X.; Li, C. J.; Smemo, S.; Dai, Q.; Bailey, K. A.; Nobrega, M. A.; Han, K.-L.; Cui, Q.; He, C., FTO-mediated formation

of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nature Communications* **2013,** *4*, 1798-1798.

23. Liu, J.; Jia, G., Methylation Modifications in Eukaryotic Messenger RNA. *Journal of Genetics and Genomics* **2014,** *41* (1), 21-33.

24. Su, D.; Chan, C.; Gu, C.; Lim, K.; Chionh, Y.; McBee, M.; Russell, B.; Babu, I.; Begley, T.; Dedon, P., Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nat. Prot.* **2014,** *9* (4), 828-841.

25. Basanta-Sanchez, M.; Temple, S.; Ansari, S. A.; D'Amico, A.; Agris, P. F., Attomole quantification and global profile of RNA modifications: Epitranscriptome of human neural stem cells. *Nucleic Acids Res.* **2015**, gkv971-gkv971.

26. Motorin, Y.; Grosjean, H., Multisite-specific tRNA:m5C-methyltransferase (Trm4) in yeast Saccharomyces cerevisiae: Identification of the gene and substrate specificity of the enzyme. *RNA* **1999,** *5* (8), 1105-1118.

27. Hong, B.; Brockenbrough, J. S.; Wu, P.; Aris, J. P., Nop2p is required for pre-rRNA processing and 60S ribosome subunit synthesis in yeast. *Molecular and Cellular Biology* **1997,** *17* (1), 378-388.

28. Doll, A.; Grzeschik, K. H., Characterization of two novel genes, WBSCR20 and WBSCR22, deleted in Williams-Beuren syndrome. *Cytogenetics and Cell Genetics* **2001,** *95* (1-2), 20-27.

29. Estève, P. O.; Chin, H. G.; Benner, J.; Feehery, G. R.; Samaranayake, M.; Horwitz, G. A.; Jacobsen, S. E.; Pradhan, S., Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** *106* (13), 5076-5081.

30. Goll, M. G.; Kirpekar, F.; Maggert, K. a.; Yoder, J. a.; Hsieh, C.-L.; Zhang, X.; Golic, K. G.; Jacobsen, S. E.; Bestor, T. H., Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Sci.* **2006,** *311* (5759), 395-398.

31. Mauer, J.; Luo, X.; Blanjoie, A.; Jiao, X.; Grozhik, A. V.; Patil, D. P.; Linder, B.; Pickering, B. F.; Vasseur, J. J.; Chen, Q.; Gross, S. S.; Elemento, O.; Debart, F.; Kiledjian, M.; Jaffrey, S. R., Reversible methylation of m^6A_m in the 5' cap controls mRNA stability. *Nature* **2017,** *541* (7637), 371-375.

32. Akichika, S.; Hirano, S.; Shichino, Y.; Suzuki, T.; Nishimasu, H.; Ishitani, R.; Sugita, A.; Hirose, Y.; Iwasaki, S.; Nureki, O.; Suzuki, T., Cap-specific terminal N 6 -methylation of RNA by an RNA polymerase II–associated methyltransferase. *Science* **2019,** *363* (6423).

33. Sun, H.; Zhang, M.; Li, K.; Bai, D.; Yi, C., Cap-specific, terminal N 6-methylation by a mammalian m6Am methyltransferase. *Cell Research* **2019,** *29* (1), 80-82.

34. Boulias, K.; Tocydlowska-Socha, D.; Hawley, B. R.; Liberman-Isakov, N.; Takashima, K.; Zaccara, S.; Guez, T.; Vasseur, J.-J.; Debart, F.; Aravind, L.; Jaffrey, S. R.; Greer, E. L., Identification of the m6A methyltransferase PCIF1 reveals the location and functions of m6Am in the transcriptome. *Molecular Cell* **2019,** *75* (3), 631-643.

35. Liu, J.; Yue, Y.; Han, D.; Wang, X.; Fu, Y.; Zhang, L.; Jia, G.; Yu, M.; Lu, Z.; Deng, X.; Dai, Q.; Chen, W.; He, C., A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol* **2014,** *10* (2), 93-5.

36. Fu, Y.; Dominissini, D.; Rechavi, G.; He, C., Gene expression regulation mediated through reversible m^{6}A RNA methylation. *Nat. Rev. Genet.* **2014,** *15* (5), 293-306.

37. Wang, P.; Doxtader, K. A.; Nam, Y., Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. *Molecular Cell* **2016,** *63* (2), 306-317.

38. Wang, X.; Feng, J.; Xue, Y.; Guan, Z.; Zhang, D.; Liu, Z.; Gong, Z.; Wang, Q.; Huang, J.; Tang, C.; Zou, T.; Yin, P., Structural basis of N6-adenosine methylation by the METTL3–METTL14 complex. *Nature* **2016,** *534* (7608), 575-578.

39. Pendleton, K. E.; Chen, B.; Liu, K.; Hunter, O. V.; Xie, Y.; Tu, B. P.; Conrad, N. K., The U6 snRNA m(6)A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **2017,** *169* (5), 824-835 e14.

40. Hausinger, R. P.; Schofield, C. J. E., *2-Oxoglutarate-Dependent Oxygenases*. Royal Society of Chemistry: Cambridge, UK, 2015; p 95-122.

41. Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y., Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **2014,** *119* (7), 941-953.

42. Shi, Y.; Whetstine, J. R., Dynamic Regulation of Histone Lysine Methylation by Demethylases. *Molecular Cell* **2007,** *25* (1), 1-14.

43. Cloos, P. A. C.; Christensen, J.; Agger, K.; Helin, K., Erasing the methyl mark: Histone demethylases at the center of cellular differentiation and disease. *Genes and Development* **2008,** *22* (9), 1115-1140.

44. Cloos, P. A. C.; Christensen, J.; Agger, K.; Maiolica, A.; Rappsilber, J.; Antal, T.; Hansen, K. H.; Helin, K., The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. *Nature* **2006,** *442* (7100), 307-311.

45. Camps, M.; Eichman, B. F., Unraveling a connection between DNA demethylation repair and cancer. *Molecular Cell* **2011,** *44* (3), 343-344.

46. Lee, D. H.; Jin, S. G.; Cai, S.; Chen, Y.; Pfeifer, G. P.; O'Connor, T. R., Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. *Journal of Biological Chemistry* **2005,** *280* (47), 39448-39459.

47. Alemu, E.; He, C.; Klungland, A., ALKBHs-facilitated RNA modifications and de-modifications. *DNA Repair* **2016,** *44*, 87-91.

48. Jia, G.; Fu, Y.; Zhao, X.; Dai, Q.; Zheng, G.; Yang, Y.; Yi, C.; Lindahl, T.; Pan, T.; Yang, Y.-G.; He, C., N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chemical Biology* **2011,** *7* (12), 885-887.

49. Jia, G.; Yang, C.-g.; Yang, S.; Jian, X.; Yi, C.; Zhou, Z.; He, C., Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Letters* **2008,** *582* (23), 3313-3319.

50. Zheng, G.; Dahl, J. A.; Niu, Y.; Fedorcsak, P.; Huang, C. M.; Li, C. J.; Vagbo, C. B.; Shi, Y.; Wang, W. L.; Song, S. H.; Lu, Z.; Bosmans, R. P.; Dai, Q.; Hao, Y. J.; Yang, X.; Zhao, W. M.; Tong, W. M.; Wang, X. J.; Bogdan, F.; Furu, K.; Fu, Y.; Jia, G.; Zhao, X.; Liu, J.; Krokan, H. E.; Klungland, A.; Yang, Y. G.; He, C., ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **2013,** *49* (1), 18-29.

51. Lichinchi, G.; Zhao, Boxuan S.; Wu, Y.; Lu, Z.; Qin, Y.; He, C.; Rana, Tariq M., Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. 2016; Vol. 20, pp 666-673.

52. Songe-møller, L.; Born, E. V. D.; Leihne, V.; Vågbø, C. B.; Kristoffersen, T.; Krokan, H. E.; Kirpekar, F.; Falnes, P. Ø.; Klungland, A., Mammalian ALKBH8 Possesses tRNA Methyltransferase Activity Required for the Biogenesis of Multiple Wobble Uridine Modifications Implicated in Translational Decoding. *Molecular and Cellular Biology* **2010,** *30* (7), 1814-1827.

53. Van Den Born, E.; Vågbø, C. B.; Songe-Møller, L.; Leihne, V.; Lien, G. F.; Leszczynska, G.; Malkiewicz, A.; Krokan, H. E.; Kirpekar, F.; Klungland, A.; Falnes, P., ALKBH8-mediated formation of a novel diastereomeric pair of wobble nucleosides in mammalian tRNA. *Nature Communications* **2011,** *2*.

54. Fu, D.; Brophy, J. a. N.; Chan, C. T. Y.; Atmore, K. a.; Begley, U.; Paules, R. S.; Dedon, P. C.; Begley, T. J.; Samson, L. D., Human AlkB homolog ABH8 Is a tRNA methyltransferase required for wobble uridine modification and DNA damage survival. *Molecular and cellular biology* **2010,** *30* (10), 2449-2459.

55. Basanta-Sanchez, M.; Wang, R.; Liu, Z.; Ye, X.; Li, M.; Shi, X.; Agris, P.; Zhou, Y.; Huang, Y.; Sheng, J., TET1-mediated Oxidation of 5-formylcytosine (5fC) to 5-carboxycytosine (5caC) in RNA. *ChemBioChem* **2016**, 1-6.

56. Pastor, W. a.; Aravind, L.; Rao, A., TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature reviews. Molecular cell biology* **2013,** *14* (6), 341-56.

57. Cheng, Y.; Bernstein, A.; Chen, D.; Jin, P., 5-Hydroxymethylcytosine: A new player in brain disorders? *Experimental Neurology* **2014,** *268*, 3-9.

58. Alarcón, Claudio R.; Goodarzi, H.; Lee, H.; Liu, X.; Tavazoie, S.; Tavazoie, Sohail F., HNRNPA2B1 Is a Mediator of m6A-Dependent Nuclear RNA Processing Events. *Cell* **2015,** *162* (6), 1299-1308.

59. Aguilo, F.; Zhang, F.; Sancho, A.; Fidalgo, M.; Di Cecilia, S.; Vashisht, A.; Lee, D. F.; Chen, C. H.; Rengasamy, M.; Andino, B.; Jahouh, F.; Roman, A.; Krig, S. R.; Wang, R.; Zhang, W.; Wohlschlegel, J. A.; Wang, J.; Walsh, M. J., Coordination of m(6)A mRNA Methylation and Gene Transcription by ZFP217 Regulates Pluripotency and Reprogramming. *Cell Stem Cell* **2015,** *17* (6), 689-704.

60. Anantharaman, A.; Tripathi, V.; Khan, A.; Yoon, J.-H.; Singh, D. K.; Gholamalamdari, O.; Guang, S.; Ohlson, J.; Wahlstedt, H.; Öhman, M.; Jantsch, M. F.; Conrad, N. K.; Ma, J.; Gorospe, M.; Prasanth, S. G.; Prasanth, K. V., ADAR2 regulates RNA stability by modifying access of decay-promoting RNA-binding proteins. *Nucleic Acids Research* **2017**, gkw1304-gkw1304.

61. Houseley, J.; Tollervey, D., The Many Pathways of RNA Degradation. *Cell* **2009,** *136* (4), 763-776.

62. Wang, Z.; Jiao, X.; Carr-Schmid, A.; Kiledjian, M., The hDcp2 protein is a mammalian mRNA decapping enzyme. *PNAS* **2002,** *99* (20), 12663-12668.

63. Pavanapuresan P. Vaidyanathan, I. A. D. K. M. H. A.-H. D. H., Pseudouridine and N-6 methyladenosine modifications weaken PUF protein/RNA interactions. *RNA Journal* **2017**.

64. Liu, N.; Zhou, K. I.; Parisien, M.; Dai, Q.; Diatchenko, L.; Pan, T., N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Research* **2017,** *45* (10), 6051-6063.

65. Dai, X.; Gonzalez, G.; Li, L.; Li, J.; You, C.; Miao, W.; Hu, J.; Fu, L.; Zhao, Y.; Li, R.; Li, L.; Chen, X.; Xu, Y.; Gu, W.; Wang, Y., YTHDF2 Binds to 5-Methylcytosine in RNA and Modulates the Maturation of Ribosomal RNA. *Analytical Chemistry* **2019,** *92*, 1346-1354.

66. Dai, X.; Wang, T.; Gonzalez, G.; Wang, Y., Identification of YTH Domain-Containing Proteins as the Readers for N1-Methyladenosine in RNA. *Analytical Chemistry* **2018,** *90*, 6380-6384.

67. Jiang, Q.; Sun, B.; Liu, Q.; Cai, M.; Wu, R.; Wang, F.; Yao, Y.; Wang, Y., MTCH2 promotes adipogenesis in intramuscular preadipocytes via an m 6 A-YTHDF1 – dependent mechanism. *FASEB Journal* **2019,** *33* (2), 2971-2981.

68. Wang, X.; Zhao, B. S.; Roundtree, I. A.; Lu, Z.; Han, D.; Ma, H.; Weng, X.; Chen, K.; Shi, H.; He, C., N^6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **2015,** *161* (6), 1388-99.

69. Wang, X.; Zhao, B. S.; Roundtree, I. A.; Lu, Z.; Han, D.; Ma, H.; Weng, X.; Chen, K.; Shi, H.; He, C., N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **2015,** *161* (6), 1388-99.

70. Wang, X.; Lu, Z.; Gomez, A.; Hon, G. C.; Yue, Y.; Han, D.; Fu, Y.; Parisien, M.; Dai, Q.; Jia, G.; Ren, B.; Pan, T.; He, C., N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014,** *505* (7481), 117-20.

71. Cui, Q.; Shi, H.; Ye, P.; Li, L.; Qu, Q.; Sun, G.; Sun, G.; Lu, Z.; Huang, Y.; Yang, C. G.; Riggs, A. D.; He, C.; Shi, Y., m6A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. *Cell Reports* **2017,** *18* (11), 2622-2634.

72. Jaffrey, S. R.; Kharas, M. G., Emerging links between m6A and misregulated mRNA methylation in cancer. *Genome Med* **2017,** *9* (1), 2-2.

73. Lin, S.; Choe, J.; Du, P.; Triboulet, R.; Gregory, R. I., The m6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Molecular Cell* **2016,** *62* (3), 335-345.

74. Zhang, C.; Samanta, D.; Lu, H.; Bullen, J. W.; Zhang, H.; Chen, I.; He, X.; Semenza, G. L., Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m ⁶ A-demethylation of NANOG mRNA. *Proceedings of the National Academy of Sciences* **2016,** *113* (14), E2047-E2056.

75. Li, Z.; Weng, H.; Su, R.; Weng, X.; Zuo, Z.; Li, C.; Huang, H.; Nachtergaele, S.; Dong, L.; Hu, C.; Qin, X.; Tang, L.; Wang, Y.; Hong, G.-M.; Huang, H.; Wang, X.; Chen, P.; Gurbuxani, S.; Arnovitz, S.; Li, Y.; Li, S.; Strong, J.; Neilly, M. B.; Larson, R. A.; Jiang, X.; Zhang, P.; Jin, J.; He, C.; Chen, J., FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N6-Methyladenosine RNA Demethylase. *Cancer Cell* **2016**, 127-141.

76. Chen, H.; Costa, M., Iron- and 2-oxoglutarate-dependent Dioxygenases: an emerging group of molecular targets for nickel toxicity and carcinogenicity. *Biometals* **2009,** *22* (1), 191-196.

77. Chen, H.; Giri, N. C.; Zhang, R.; Yamane, K.; Zhang, Y.; Maroney, M.; Costa, M., Nickel ions inhibit histone demethylase JMJD1A and DNA repair enzyme ABH2 by replacing the ferrous iron in the catalytic centers. *Journal of Biological Chemistry* **2010,** *285* (10), 7374-7383.

78. Zhou, X.; Li, Q.; Arita, A.; Sun, H.; Costa, M., Effects of nickel, chromate, and arsenite on histone 3 lysine methylation. *Toxicology and Applied Pharmacology* **2009,** *236* (1), 78-84.

79. Hughes, M. F.; Beck, B. D.; Chen, Y.; Lewis, A. S.; Thomas, D. J., Arsenic exposure and toxicology: a historical perspective. *Toxicol Sci* **2011,** *123* (2), 305-32.

80. Bai, L.; Tang, Q.; Zou, Z.; Meng, P.; Tu, B.; Xia, Y.; Cheng, S.; Zhang, L.; Yang, K.; Mu, S.; Wang, X.; Qin, X.; Lv, B.; Cao, X.; Qin, Q.; Jiang, X.; Chen, C., m6A Demethylase FTO Regulates Dopaminergic Neurotransmission Deficits Caused by Arsenite. *Toxicol Sci* **2018,** *165* (2), 431-446.

81. Rosado, J. L.; Ronquillo, D.; Kordas, K.; Rojas, O.; Alatorre, J.; Lopez, P.; Garcia-Vargas, G.; Del Carmen Caamano, M.; Cebrian, M. E.; Stoltzfus, R. J., Arsenic exposure and cognitive performance in Mexican schoolchildren. *Environ Health Perspect* **2007,** *115* (9), 1371-5.

82. Zhou, X.; Sun, H.; Ellen, T. P.; Chen, H.; Costa, M., Arsenite alters global histone H3 methylation. *Carcinogenesis* **2008,** *29* (9), 1831-1836.

83. Liu, S.; Jiang, J.; Li, L.; Amato, N. J.; Wang, Z.; Wang, Y., Arsenite Targets the Zinc Finger Domains of Tet Proteins and Inhibits Tet-Mediated Oxidation of 5-Methylcytosine. *Environmental Science and Technology* **2015,** *49* (19), 11923-11931.

84. Organization, W. H., Burden of disease from ambient air pollution for 2016. *Geneva, Switzerland:* **2018,** (WHO).

85. Bauer, M.; Fink, B.; Thurmann, L.; Eszlinger, M.; Herberth, G.; Lehmann, I., Tobacco smoking differently influences cell types of the innate and adaptive immune system-indications from CpG site methylation. *Clin Epigenetics* **2015,** *7*, 83.

86. Sanchez-Guerra, M.; Zheng, Y.; Osorio-Yanez, C.; Zhong, J.; Chervona, Y.; Wang, S.; Chang, D.; McCracken, J. P.; Díaz, A.; Bertazzi, P. A.; Koutrakis, P.; Kang, C.-M.; Zhang, X.; Zhang, W.; Byun, H.-M.; Schwartz, J.; Hou, L.; Baccarelli, A. a., Effects of Particulate Matter Exposure on Blood 5-hydroxymethylation: Results from the Beijing Truck Driver Air Pollution Study. *Epigenetics* **2015,** (July), 00-00.

87. Grosjean, H.; Droogmans, L.; Roovers, M.; Keith, G., Detection of Enzymatic Activity of Transfer RNA Modification Enzymes Using Radiolabeled tRNA Substrates. *Methods in Enzymology* **2007,** *425*, 55-101.

88. Giessing, A. M. B.; Scott, L. G.; Kirpekar, F., A nano-chip-LC/MS based strategy for characterization of modified nucleosides using reduced porous graphitic carbon as a stationary phase. *J. Am. Soc. Mass Spectrom.* **2011,** *22* (7), 1242-1251.

89. Schwartz, S.; Agarwala, S. D.; Mumbach, M. R.; Jovanovic, M.; Mertins, P.; Shishkin, A.; Tabach, Y.; Mikkelsen, T. S.; Satija, R.; Ruvkun, G.; Carr, S. A.; Lander, E. S.; Fink, G. R.; Regev, A., High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **2013,** *155* (6), 1409-21.

90. Li, X.; Xiong, X.; Yi, C., Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Meth.* **2016,** *14* (1), 23-31.

91. Khoddami, V.; Yerra, A.; Mosbruger, T. L.; Fleming, A. M.; Burrows, C. J.; Cairns, B. R., Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc Natl Acad Sci U S A* **2019,** *116* (14), 6784-6789.

92. Chan, C. T. Y.; Dyavaiah, M.; DeMott, M. S.; Taghizadeh, K.; Dedon, P. C.; Begley, T. J., A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genetics* **2010,** *6* (12), 1-9.

93. Lodish H; Berk A; Zipursky SL; al, e., *Section 11.6, Processing of rRNA and tRNA*. New York: W. H. Freeman;, 2000.

94. Xu, L.; Liu, X.; Sheng, N.; Oo, K. S.; Liang, J.; Chionh, Y. H.; Xu, J.; Ye, F.; Gao, Y.-G.; Dedon, P. C.; Fu, X.-Y., Three distinct 3-methylcytidine (m3C) methyltransferases modify tRNA and mRNA in mice and humans. *Journal of Biological Chemistry* **2017**, jbc.M117.798298-jbc.M117.798298.

95. Cui, Y.; Wang, P.; Yu, Y.; Yuan, J.; Wang, Y., Normalized Retention Time for Targeted Analysis of the DNA Adductome. *Anal. Chem.* **2018,** *90*, 14111-14115.

96. Rose, R. E.; Quinn, R.; Sayre, J. L.; Fabris, D., Profiling ribonucleotide modifications at full-transcriptome level: a step toward MS-based epitranscriptomics. *RNA (New York, N.Y.)* **2015,** *21* (7), 1361-74.

97. Wetzel, C.; Li, S.; Limbach, P. A., Metabolic de-isotoping for improved LC-MS characterization of modified RNAs. *Journal of the American Society for Mass Spectrometry* **2014,** *25* (7), 1114-1123.

98. Cao, X.; Limbach, P. A., Enhanced Detection of Post-Transcriptional Modifications Using a Mass-Exclusion List Strategy for RNA Modification Mapping by LC-MS/MS. *Analytical Chemistry* **2015,** *87* (16), 8433-8440.

99. Kempe, H.; Schwabe, A.; Crémazy, F.; Verschure, P. J.; Bruggeman, F. J., The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. *Molecular Biology of the Cell* **2015,** *26* (4), 797-804.

100. Ross, R.; Cao, X.; Yu, N.; Limbach, P. A., Sequence mapping of transfer RNA chemical modifications by liquid chromatography tandem mass spectrometry. *Methods* **2016,** *107*, 73-78.

Chapter 2: Normalized Retention Time for Scheduled Liquid Chromatography-Multistage Mass Spectrometry Analysis of the Epitranscriptomic Modifications

Introduction

There are over 100 types of modified nucleosides in RNA, primarily in transfer RNA (tRNA) and ribosomal RNA (rRNA).[1-3] These modifications in tRNA and rRNA can influence their stability and maturation, respectively, thereby impacting translation efficiency and cellular response to stress.[1, 4] Recently, several types of modified nucleosides have also been identified in other types of RNA, including messenger RNA (mRNA) and microRNA (miRNA).[5-9] The biological functions of these RNA modifications, their regulatory enzymes, and their contributions to disease pathology, however, remain incompletely understood and are under intense investigation.[8]

Conventional methods for monitoring RNA modifications are often suitable for a single or a few types of modifications with relatively high modification stoichiometry, and are semi-quantitative.[1, 9-11] For instance, bisulfide sequencing is commonly employed for mapping 5-methylcytosine (m^5C) in DNA and RNA.[10] False positives can arise from imperfect deamination of cytosine to uracil and the method cannot distinguish m^5C from its oxidation products of 5-hydroxymethylcytosine (5-hmrC), 5-formylcytosine (5-forC) and 5-carboxylcytosine (5-carC).[10, 12]

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) constitutes a powerful tool for accurately identifying and quantifying modified nucleosides, and most

measurements are based on the characteristic neutral loss of a ribose or 2-deoxyribose from protonated nucleosides during collisional activation.[1, 8, 9, 13, 14] Nevertheless, previously published LC-MS/MS methods for ribonucleoside analysis require relatively large amount of RNA (100-1000 ng) per injection.[1, 7, 14] For RNAs of relatively low abundance, including mRNA and miRNA, obtaining a substantial amount of RNA is laborious and costly.

Scheduled selected-reaction monitoring (SRM) method relying on the use of normalized retention time (iRT) for the prediction of retention times of analytes under given chromatographic conditions has been employed for high-throughput analyses of peptides and modified 2′-deoxyribonucleosides.[13, 15] A scheduled SRM method is advantageous in limiting acquisition time for specific precursor ions to a few minutes, thereby allowing for shorter cycle time and providing more data points per chromatographic peak to enable robust and sensitive analyte detection.[13]

The objective of this study was to develop a nano-flow liquid chromatography-multistage MS (nLC-MS^3) method to simultaneously monitor modified ribonucleosides with low nanogram quantities of RNA. We show that the scheduled SRM method can facilitate simultaneous assessment of a large number of ribonucleosides (27 modified nucleosides and 4 unmodified canonical nucleosides, structures shown in **Figure 2.1**) in a single LC-MS^3 run without compromising detection sensitivity or accuracy.

Experimental Section

RNA sample preparation

Total RNA was isolated from HEK293T cells and digested with enzymes following previously published procedures.[9] Briefly, total RNA was isolated using E.Z.N.A. Total

RNA Kit I (Omega) according to the manufacturer's recommended procedures. The RNA samples were then enzymatically digested to mononucleosides with nuclease P1 and phosphodiesterase 2 at 37°C for 4 hrs. Alkaline phosphatase and phosphodiesterase 1 were subsequently added to the resulting solution, and the mixture was incubated at 37°C for 2 hrs. The enzymes were subsequently removed from the digestion mixture by chloroform extraction.

A mixture of ribonucleoside standards and 20 ng of enzymatically digested RNA was used for defining iRT values and for establishing iRT-RT correlation. A list of ribonucleoside standards employed in this study are shown in **Table 2.1**.

Nanoflow liquid chromatography-multistage mass spectrometry (nLC-MS^3) analysis

Unscheduled and scheduled SRM experiments were performed on an LTQ-XL linear ion trap mass spectrometer coupled with an EASY-nLC II (Thermo Fisher Scientific, San Jose, CA). The mass spectrometer was operated in the positive-ion mode with the electrospray, capillary, and tube lens voltages being 2.0 kV, 12 V, and 100 V, respectively. Ion transport tube temperature was maintained at 275°C. Precursor ions for MS^2 and MS^3 analyses of ribonucleosides are listed in Table 2.2.

To assign iRT values, a nucleoside mixture from the enzymatic digestion of total RNA and stable isotope-labeled ribonucleoside standards was loaded, at a flow rate of 2.5 μL/min, onto a 5- cm or 3.5-cm long in-house packed porous graphitic carbon (PGC, 5 μm particle size, Thermo Fisher Scientific) trapping column (150 μm i.d.) for linear and non-linear gradients, respectively. Analytes eluting from the trapping column were directed to and resolved on an 18-cm Zorbax SB-C18 (5 μm in particle size, 100 Å in pore size,

Table 2.1: A list of ribonucleoside standards employed in this study and their sources.

Modified Ribonucleoside	Abbreviation	Reference Citation or Vendor
N^3-methylcytidine	m^3C	16
N^4-methylcytidine	m^4C	17
5-methylcytidine	m^5C	9
2′-*O*-methylcytidine	C_m	
5-hydroxymethylcytidine	5-hmrC	8
5-formylmethylcytidine	5-forC	
5-carboxymethylcytidine	5-carC	
2-thiocytidine	s^2C	Carbosyth
N^4-acetylcytidine	ac^4C	Prof. Xiaochun Yu, the City of Hope
2′-*O*-methyluridine	U_m	Thermo Fisher Scientific
N^3-methyluridine	m^3U	1
5-methyluridine	m^5U	Sigma-Aldrich
5-hydroxymethyluridine	5-hmrU	18
N^1-methyladenosine	m^1A	19
N^6-methyladenosine	m^6A	3
2′-*O*-methyladenosine	A_m	
8-methyladenosine	m^8A	20
N^6-acetyladenosine	ac^6A	Prof. Xiaochun Yu, the City of Hope
$N^{6,6}$-dimethyladenosine	$m^{6,6}A$	Toronto Research Chemicals
N^6-methyl-2′-*O*-methyl-adenosine	m^6A_m	Toronto Research Chemicals
Inosine	rI	Sigma Aldrich
7-methylinosine	m^7I	19
N^1-methylguanosine	m^1G	21
N^2-methylguanosine	m^2G	
O^6-methylguanosine	m^6G	

Agilent) or Magic C18-AQ (5 μm in particle size, 100 Å in pore size, Michrom BioResources) analytical column (75 μm i.d.) at a flow rate of 300 nL/min. Mobile phases A and B were 0.1% (v/v) formic acid in water and 0.1% (v/v) formic acid in acetonitrile, respectively, and the mobile phase gradients are listed in **Table 2.3**. Briefly, linear gradients had a linear increase from 16% to 70% mobile phase B over 65 min for the fast gradient (hereafter referred to as linear, long-fast gradient). To evaluate iRT-RT correlation with complex gradients, we also employed a gradient with a step-wise increase in %B, which is referred to as non-linear gradient. For instance, non-linear, long-fast gradient consisted of 0−16% B in 5 min, 16−22% B in 23 min, 22−50% B in 17 min, 50-90% B in 5 min and finally at 90% B for 30 min.

To assign iRT values for modified ribonucleosides, the iRT values for cytidine (rC) and adenosine (rA) were arbitrarily assigned to 10 and 100, respectively, and the iRT values for the nucleoside of interest (X) was determined using the following equation: $iRT_X = 100 - [(RT_2 - RT_X)/(RT_2\text{-}RT_1)] \times 90$, where iRT_X represents the iRT value for nucleoside X, and RT_X, RT_2, and RT_1 designate the observed retention times for X, rA and rC, respectively. The predicted retention time for nucleoside X was calculated from its iRT (iRT_X) and the actual retention times observed for rC (RT_1) and rA (RT_2) in the calibration run with the following equation: $RT_X = RT_2 - (100 - iRT_X) \times (RT_2 - RT_1)/90$.

Quantifications of 5-methylcytidine and 2′-*O*-methylcytidine

For all measurements, 2.5 ng of digested RNA spiked with isotopically labeled standards was loaded onto a 4.1-cm PGC trapping column and eluted to an 18-cm Zorbax SB-C18 analytical column. The same sample was injected and analyzed using the

Table 2.2. A list of precursor ions for ribonucleosides monitored in MS^3, isolation width, and normalized collision energy (NCE). iRT scores obtained using PGC trapping column together with a Zorbax SB-C18 (long-fast, linear and non-linear gradients) or Magic C18-AQ (long-fast, linear gradient) analytical column.

	iRT			**Transitions**		**Isolation Width**		**NCE**	
	Zorbax Linear	**Zorbax Non-Linear**	**Magic Linear**	**MS^2**	**MS^3**	**MS^2**	**MS^3**	**MS^2**	**MS^3**
rC	10.0	10	10	244 →112	112 → 95	3	2	37	40
C_m	26.5	18.7	26.5	258 →112	112 → 95	3	2	37	40
m^3C	22.9	17.2	22	258 →126	126 → 109	3	2	37	40
m^4C	12.3	12.2	20.3	258 →126	126 → 108	3	2	37	40
m^5C	32.0	31.6	29.4	258 →126	126 → 108	3	2	37	40
5-hmrC	18.2	22	21.6	274 → 142	142 → 124	3	2	37	42
s^2C	62.0	57	43	260 → 128	128 → 111	3	2	43	35
ac^4C	139.9	121.6	123.1	286 → 154	154 → 112	3	2	43	37
5-forC	110.4	107.9	105	272 → 140	140 → 97	3	2	42	45
5-carC	304.9	148.9	235.8	288 → 156	156 → 138	3	2	42	45
rU	39.6	35.9	38.1	245 → 113	113 → 96	3	2	43	50
Ψ	32.8	24.1	26.9	245 → 179	179 → 151	3	2	43	50
U_m	51.4	52.4	52.1	259 → 113	113 → 96	3	2	43	50
m^3U	59.9	69.0	60.8	259 → 127	127 → 96	3	2	43	50
m^5U	67.02	79.5	65.7	259 → 127	127 →110	3	2	43	50
5-hmrU	40.3	43.2	46.3	275 → 143	143 → 125	3	2	37	42
rA	100.0	100.0	100.0	268 → 136		3	2	35	
A_m	115.0	108.8	116.1	282 → 136	136 → 94	3	2	43	40
m^1A	50.8	45.1	49.6	282 → 150	150 → 133	3	2	43	37
m^6A	153.7	130.6	157	282 → 150	150 → 133	3	2	43	37
m^8A	163.0	134.2	155.7	282 → 150	150 → 133	3	2	43	37
ac^6A	164.3	139.8	162.8	310 → 178	178 → 136	3	2	43	37
$m^{6,6}A$	260.0	196.3	259.8	296 → 164		3	2	37	
m^6A_m	170.5	138.1	173.9	296 → 150	150 → 94	3	2	37	35
rI	84.6	92.7	81.8	269 → 137		3	2	35	
m^7I	33.5	29.0	33.1	283 → 151	151 → 133	3	2	42	39
rG	173.5	138.2	164.4	284 → 152	152 → 135	3	2	42	39
G_m	186.4	143.6	168.5	298 → 152	152 → 135	3	2	42	39
m^1G	153.4	134.0	148.7	298 → 166	166→ 148	3	2	42	39
m^2G	154.3	134.3	148.7	298 → 166	166→ 148	3	2	42	39
m^6G	241.1	230.9	233.4	298 → 166	166→ 148	3	2	42	39

Table 2.3: A list of A) linear and B) non-linear mobile phase gradients employed in the present study.

A. Linear Gradients

Short and Fast		Long		Long and fast	
Time	**% B**	**Time**	**% B**	**Time**	**% B**
0	0	0	0	0	0
5	16	5	16	5	16
35	75	70	50	70	75
37	90	72	90	72	90
47	90	80	90	80	90

B. Non-linear Gradients

Short		Short and Fast		Long		Long and fast	
Time	**% B**	**Time**	**% B**	**Time**	**% B**	**Time**	**% B**
0	0	0	0	0	0	0	0
5	16	5	16	5	16	5	16
14	20	14	22	28	20	28	22
23	30	23	50	45	30	45	50
25	90	25	90	50	90	50	90
40	90	40	90	80	90	80	90

scheduled LC-MS/MS/MS method, with precursor ions for the stable isotope-labeled standards being incorporated into the transition list.

Results

To assign iRT values and to assess iRT-RT correlation, we injected a solution containing nucleoside mixture from the digestion of 5 ng of total RNA isolated from HEK293T cells and synthetic standard ribonucleosides. For accurate prediction of

retention times, we employed the four canonical ribonucleosides (i.e. cytidine, uridine, guanosine and adenosine) as standards, where we assigned the iRT values of cytidine and adenosine as 10 and 100, respectively. The long-fast gradient was employed for calculating iRT scores for uridine, guanosine and modified ribonucleosides using retention time and linear regression analysis (**Figures 2.1 and 2.2, Table 2.2**).

Traditionally, linear gradients were employed for establishing RT-iRT correlation.[13, 15] To assess the reliability of iRT-RT correlation under different chromatography settings, we considered different gradients speeds (long, long fast, and short fast), and examined both linear and non-linear gradients. Our results showed that all different gradients yielded linear iRT-RT relationships with good correlation coefficients: linear gradients (long, R^2 = 0.992; long-fast, R^2= 0.997; short-fast, R^2 = 0.987) and non-linear gradients (long, R^2 = 0.969; long-fast, R^2 = 0.995; short, R^2 = 0.975; short-fast, R^2 = 0.972) (**Figure 2.3**). Therefore, normalized retention time values can also be employed for more complex gradients. Typically, non-linear gradients allow for more efficient separation of closely eluting species.

We also observed that short gradients exhibited poor separation efficiencies, and did not allow for the elution of a few modified hydrophobic nucleosides. Longer gradients, especially those with higher content of mobile phase B, including long and long-fast gradients, facilitate the elution of all or nearly all target nucleosides. Consistent with the previous study,[13] the number of analytes eluted in a single run depends on the length of the gradient and the maximum percentage of mobile phase B.

Next, we examined if a similar iRT-RT correlation could be attained using an

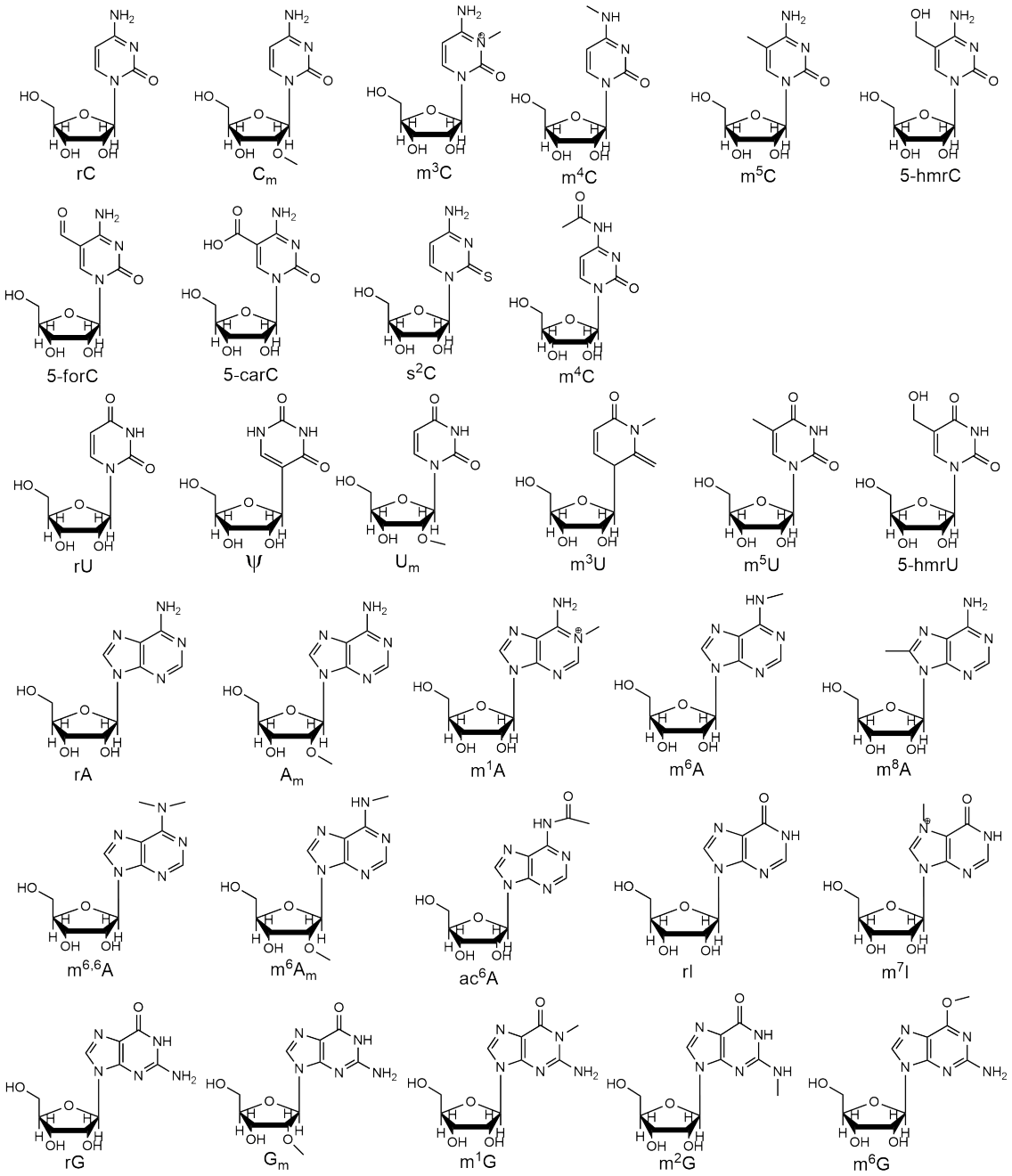


Figure 2.1: The chemical structures of modified ribonucleosides used in the present study.

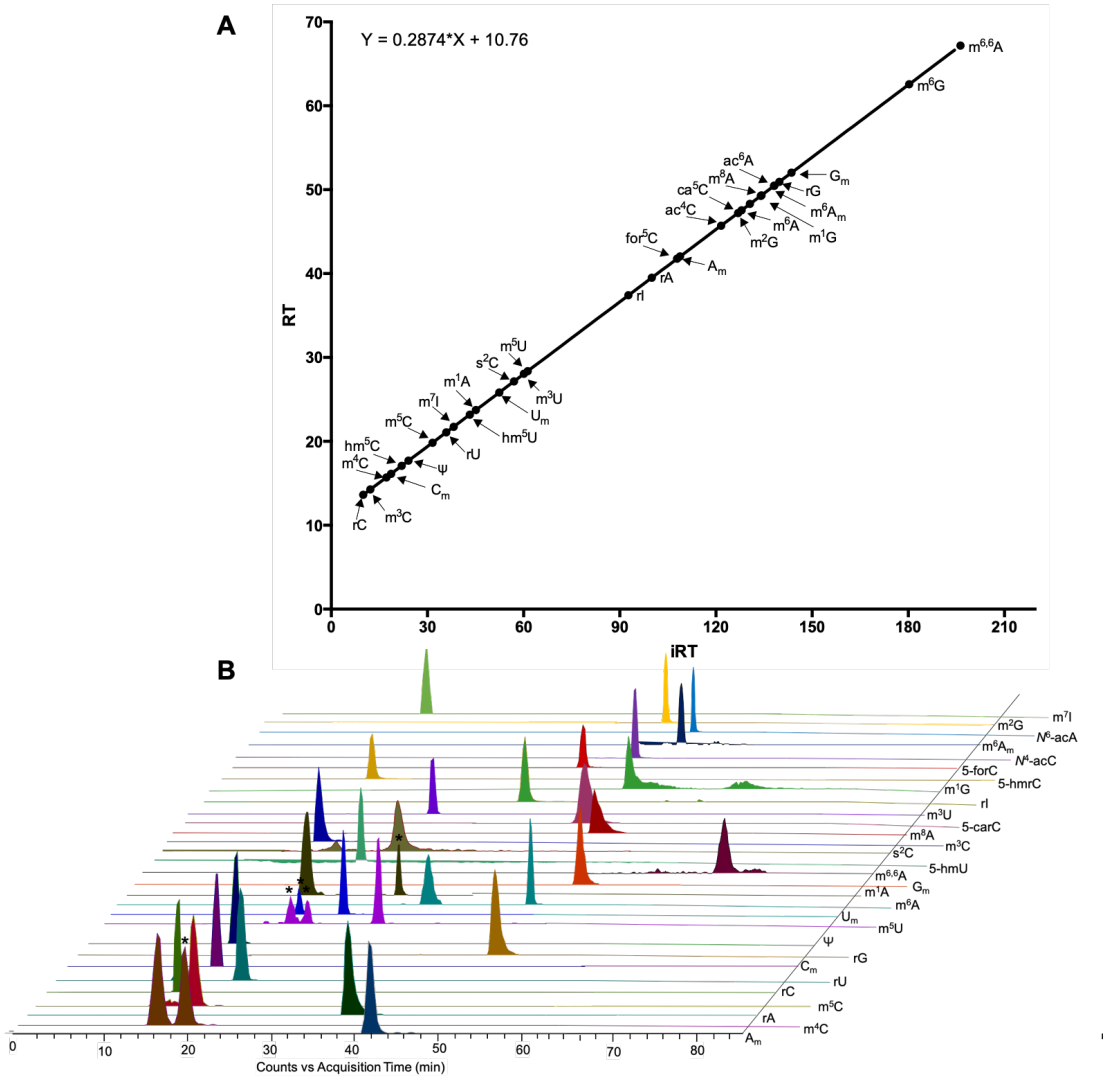


Figure 2.2: Scheduled SRM for the analysis of modified ribonucleosides. A) The iRT-RT correlation for ribonucleosides on a Zorbax SB-C18 column with the use of long-fast gradient. Ribonucleosides are labeled on the line. Complete names and chemical structures are shown in Figure 1 and Table 2.1. B) Representative selected-ion chromatograms (SICs) for rA, rG, and their mono-methylated derivatives using a scheduled SRM LC-MS^3 method with a non-linear, long-fast gradient. Interference peaks are denoted with *.

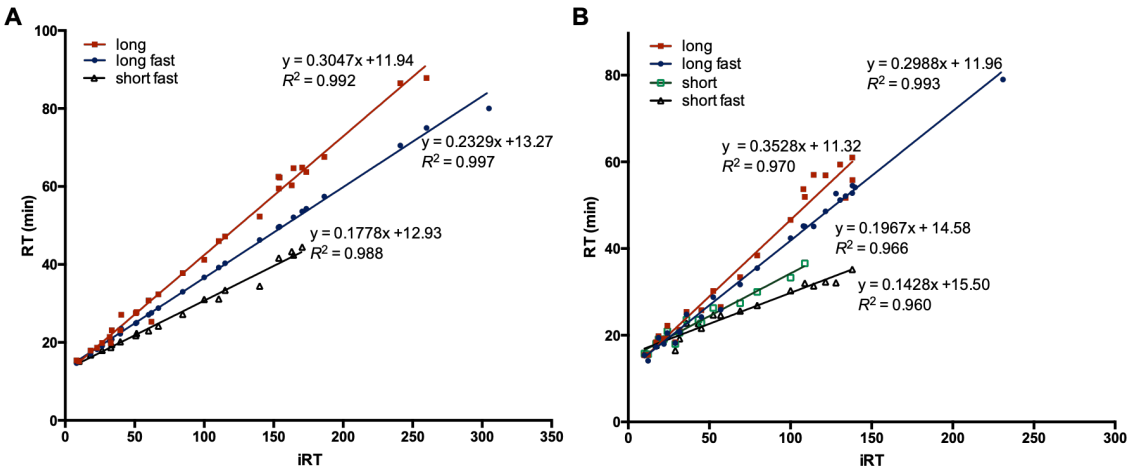


Figure 2.3: RT-iRT correlations acquired from the use of a porous graphitic carbon (PGC) trapping column and a Zorbax SB-C18 analytical column with different mobile phase gradients. A) linear gradients with modified gradient speeds (long, long-fast, and short-fast); and B) non-linear gradients with different gradient speeds (long, long-fast, short-fast, short).

analytical column packed with a different stationary phase material, Magic C18-AQ. We found that, with the use of different linear gradients, the iRT values exhibit excellent linear correlations with the observed RTs on this analytical column (**Figures 2.4**). Moreover, iRT values did not differ substantially with the use of Magic C18-AQ or Zorbax SB-C18 analytical column (**Figure 2.4, Table 2.2**), suggesting that iRT values and RT predictions are transferable to other reversed-phase C18 stationary phase materials.

On average, the RTs predicted from iRTs were within 2 and 2.5 min of the actual RTs with the use of the linear and non-linear gradients, respectively (**Figure 2.5 and Table 2.4**). The iRT values were assigned based on the long-fast gradient; thus, the elution times of analytes with the use of long-fast gradient were more accurately predicted and overall display smaller variations from the predicted RTs than those with other modified gradients (**Figure 2.5**). In addition, the analytes exhibiting the largest uncertainty in RT prediction

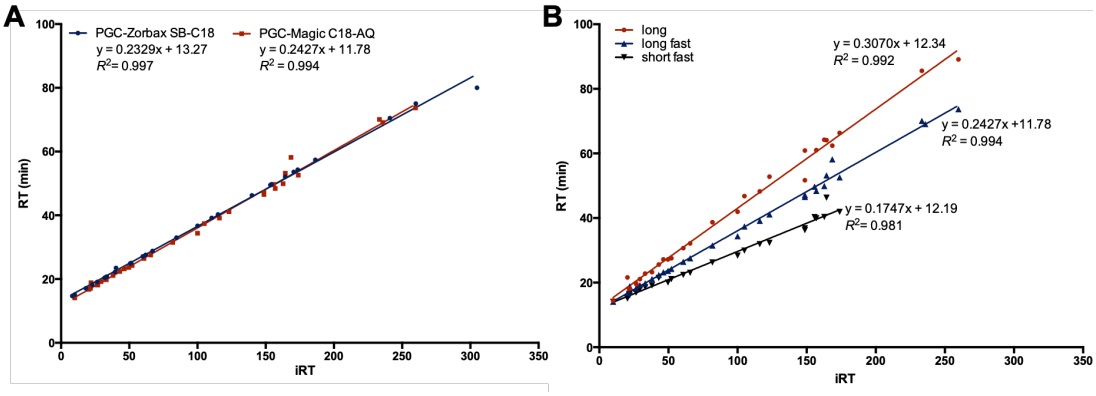


Figure 2.4: A) A comparison of RT-iRT correlations obtained from the use of Zorbax SB-C18 and Magic C18-AQ analytical columns. Data in B) RT and iRT correlations acquired from LC-MS^3 analysis using PGC trapping column and Magic C18-AQ analytical column with different linear gradients (long, long-fast, and short-fast).

eluted toward the end of the gradients, e.g. rG and its mono-methylated derivatives (**Figure 2.2 and 2.5**). Expanding the acquisition time window for later-eluting analytes will compensate for drift in RT without compromising measurement efficiency because a relatively small number of modified nucleosides elute from the column toward the end of the gradient.

The major motivation for developing this method was to achieve simultaneous quantifications of modified ribonucleosides with a scheduled SRM method. By using an LTQ-XL linear ion trap mass spectrometer, we divided data acquisition time into 10-min retention time windows and monitored the transitions for ribonucleosides in individual windows based on their predicted retention times. In doing so, we were able to monitor a small number of analytes per acquisition time segment rather than monitoring all analyte transitions throughout the entire gradient. However, some scan events are redundant between segments because several analytes have similar transitions. For example, N^1-methyladneosine (m^1A) and N^6-methyladenosine (m^6A) share the same MS/MS transition,

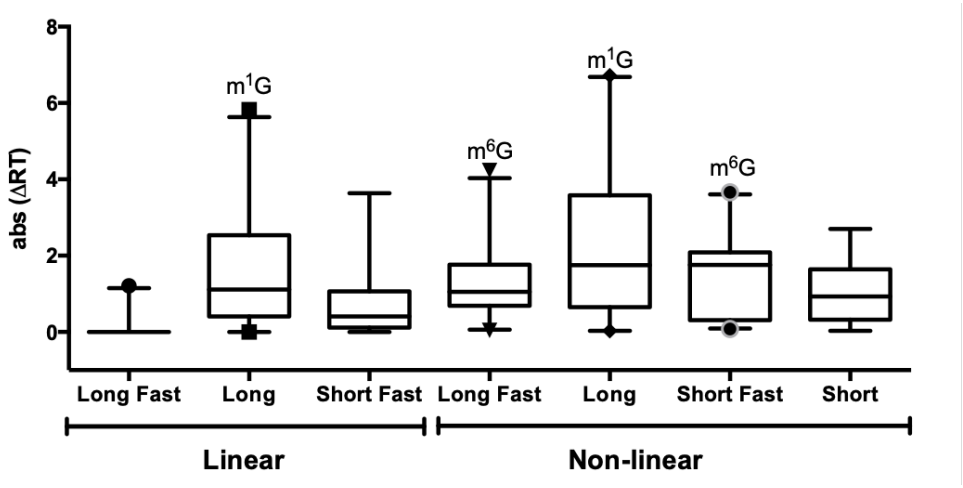


Figure 2.5: Precision and accuracy of the high-throughput scheduled SRM method determined by: A) absolute difference (in minutes) between the predicted and observed RT using linear and non-linear, modified gradients; and B) a comparison of quantification results from scheduled SRM LC-MS^3 using long-fast, linear and non-linear gradients with those obtained with low-throughput method. The whiskers plotted in A) correspond to the 5th and 95th percentiles and representative outliers are shown based on data acquired from an average of four LC-MS^3 runs. On average, divergences of the predicted RTs from the observed RTs were within 2 and 2.5 min for linear and non-linear gradients, respectively. Detailed average difference from predicted RT for all gradients are found in Table 2.3.

Gradient	Average Deviation from Predicted RT (min)
Linear Long-fast	0.1
Linear Long	1.6
Linear Short-fast	0.7
Nonlinear Long-fast	1.3
Non-linear Long	2.3
Non-Linear Short-fast	1.5
Non-linear Short	1.0

Table 2.4: Average absolute difference between predicted and observed RT for different linear and non-linear gradients.

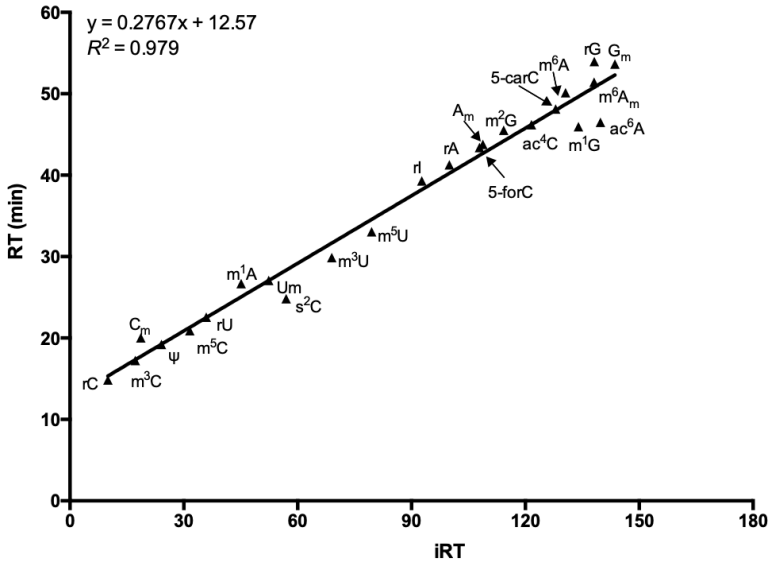


Figure 2.6. iRT-RT correlation analysis of ribonucleosides from the digestion of total RNA isolated from HEK-293T cells with the use of a non-linear gradient. Ribonucleosides are labeled on line.

but differ in MS/MS/MS (data not shown) and retention time (**Figure 2.2**).Thus, common precursor ions may be monitored in more than one retention time segment for isomeric ribonucleosides; nevertheless, they are easily distinguishable from one another based on retention time and MS^3.[3, 9]

Next, we utilized the method to identify all possible modified ribonucleosides in the nucleoside mixture of 2.5 ng of total RNA isolated from HEK293T cells. Conventional LC- MS/MS-based global screening methods require considerable amount of RNA digestion mixture per injection.[1, 2, 14] Another drawback of these methods resides in the throughput, where a single or few analytes of interest are monitored per run to enable short cycle time and robust quantification.[6, 9] Our scheduled LC-MS^3 method incorporated more analytes per run while ensuring cycle time in each segment is short enough for robust

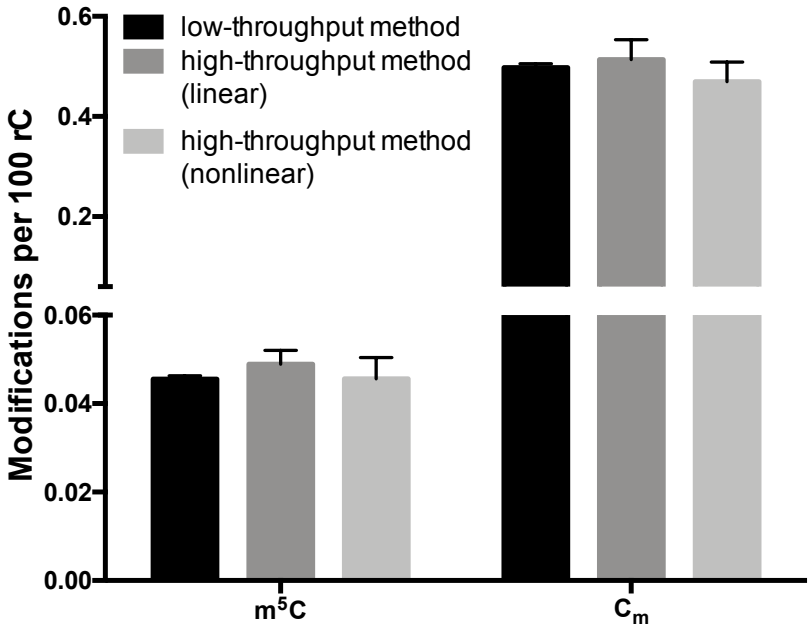


Figure 2.7: Data represents the mean and standard deviation (n = 3) between low-throughput method monitoring 6 precursor ions throughout the gradient and high-throughput scheduled SRM method monitoring 33 unlabeled and 3 labeled precursor ions. *p*-values were calculated using two-tailed Student's *t*-test are found in Table 2.5.

Method	m^5C/rC (%)			C_m/rC (%)		
	Mean	***p*-value**	**LOQ (amol)**	**Mean**	***p*-value**	**LOQ (amol)**
Low Through-put Method	0.0455	NA	361 ± 45	0.4975	NA	502 ± 22
High Through-put Method (Linear Gradient)	0.0489	0.1451	511 ± 75	0.5135	0.5342	1026 ± 288
High Through-put Method (Non-linear Gradient)	0.0456	0.9915	520 ± 108	0.4695	0.2937	1054 ± 251

Table 2.5. Comparable quantification results obtained from the current scheduled LC-MS/MS/MS method and those published previously. The data represent the mean and *p*-values calculated using two-tailed Student's *t*-test (n = 3).

quantification. Furthermore, the use of isotopically labeled standards provides unambiguous identification and reliable quantification, as isotopically labeled standards exhibit identical chromatographic behaviors and fragmentation pathways to the unlabeled analytes.[9] Our results revealed the presence of 20 modified ribonucleosides and the 4 canonical ribonucleosides in the nucleoside mixture of 2.5 ng of total RNA (**Figure 2.6**).

We also compared the quantification results for 5-methylcytidine (m^5C) and 2′-*O*-methylcytidine (C_m) obtained from the scheduled LC-MS/MS/MS method with a previously published stable-isotope dilution method based on unscheduled SRM analysis of two modified ribonucleosides.[9] The scheduled method allowed for monitoring significantly more types of modifications, including isotopically labeled standards. Meanwhile, the quantification results for m^5C and C_m are as reliable as those obtained from the previously published low-throughput method, as manifested by the lack of statistically significant differences in the quantification results between the two methods (**Figure 2.7 and Table 2.5**). Furthermore, we examined the limits of quantification (LOQ), which is defined as the amount analyte that gives rise to a signal to noise level of 10. The results indicate that LOQ could be obtained in the mid to high attomole range for rC, m^5C, and C_m using the scheduled LC-MS/MS/MS method (**Table 2.5**). Therefore, the scheduled SRM method offers high-throughput analysis without diminishing quantification efficiency.

Discussion

LC-MS/MS is a widely used tool for global RNA modification screening. Recently published LC-MS/MS-based methods have overcome some long-standing analytical

challenges, including assessing the levels of modified nucleosides of low abundance. Nano-flow LC with a PGC column afforded improved sensitivity and global screening capabilities for the analyses of modified ribonucleosides.[14, 22] Another need in the analysis of modified ribonucleosides resides in the differentiation of regioisomers. MS/MS-based methods, through monitoring the neutral loss of a ribose, do not allow for the differentiation of regioisomeric mono-methylated nucleobase modifications, e.g. m^1A and m^6A. Nevertheless, MS/MS/MS on a linear ion-trap mass spectrometer can provide diagnostic fragment ions for distinguishing some of these regioisomers.[9] In addition, retention time provides another dimension of information to corroborate identification of modified ribonucleosides based on MS/MS and/or MS/MS/MS.

Future studies would benefit from expanding the number of modified ribonucleosides to be monitored. The method presented here targeted modified ribonucleosides with elution time between rC and $m^{6,6}A$. Other more hydrophobic modified ribonucleosides are known to exist, including N^2,N^2-dimethylguanosine and N^6-isopentenyladenosine.[1, 23] The analyses of these nucleosides may entail the use of alternative stationary phase materials for efficient elution, as well as using other more hydrophobic nucleosides as standards for iRT determination.

In conclusion, we developed a scheduled SRM method for targeted analyses of modified ribonucleosides. We established iRT values for 27 modified and 4 canonical ribonucleosides. We also observed consistent iRT scores for ribonucleosides with the use of two types of stationary phase materials, Zorbax SB-C18 and Magic C18-AQ, and with different mobile phase gradients. We reason that the iRT values should be easily

transferable on different chromatographic settings and in different laboratories for reliable retention time prediction. We also showed that the precision and accuracy of the scheduled LC-MS/MS/MS method were comparable to those obtained previously with the low-throughput, unscheduled SRM method. Therefore, the scheduled SRM-based LC-MS/MS/MS method allows for robust and high-throughput analysis of modified ribonucleosides, which holds great potential for investigations into the dynamic regulation of the epitranscriptome.

References

1. Su, D.; Chan, C.; Gu, C.; Lim, K.; Chionh, Y.; McBee, M.; Russell, B.; Babu, I.; Begley, T.; Dedon, P., Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nat. Prot.* **2014,** *9* (4), 828-841.

2. Chen, Q.; Shi, J.; Peng, H.; Zhang, X.; Zhang, Y.; Qian, J.; Duan, E.; Zhai, Q.; Zhou, Q., Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Sci.* **2016,** *351* (6271).

3. Ross, R.; Cao, X.; Yu, N.; Limbach, P. A., Sequence mapping of transfer RNA chemical modifications by liquid chromatography tandem mass spectrometry. *Methods* **2016,** *107*, 73-78.

4. Calo, E.; Flynn, R. A.; Martin, L.; Spitale, R. C.; Chang, H. Y.; Wysocka, J., RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* **2015,** *518* (7538), 249-53.

5. Gaston, K. W.; Limbach, P. A., The identification and characterization of non-coding and coding RNAs and their modified nucleosides by mass spectrometry. *RNA Bio.* **2014,** *11* (12), 1568-1585.

6. Yamauchi, Y.; Nobe, Y.; Izumikawa, K.; Higo, D.; Yamagishi, Y.; Takahashi, N.; Nakayama, H.; Isobe, T.; Taoka, M., A mass spectrometry-based method for direct determination of pseudouridine in RNA. *Nucleic Acids Res.* **2015,** *44* (6).

7. Basanta-Sanchez, M.; Temple, S.; Ansari, S. A.; D'Amico, A.; Agris, P. F., Attomole quantification and global profile of RNA modifications: Epitranscriptome of human neural stem cells. *Nucleic Acids Res.* **2015**, gkv971-gkv971.

8. Fu, L.; Guerrero, C. R.; Zhong, N.; Amato, N. J.; Liu, Y.; Liu, S.; Cai, Q.; Ji, D.; Jin, S.-G.; Niedernhofer, L. J.; Pfeifer, G. P.; Xu, G.-L.; Wang, Y., Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **2014,** *136* (33), 11582-5.

9. Fu, L.; Amato, N. J.; Wang, P.; McGowan, S. J.; Niedernhofer, L. J.; Wang, Y., Simultaneous Quantification of Methylated Cytidine and Adenosine in Cellular and Tissue RNA by Nano-Flow Liquid Chromatography–Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal. Chem.* **2015,** *87ha* (15), 7653-7659.

10. Helm, M.; Motorin, Y., Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* **2017,** *18* (5), 275-291.

11. Li, X.; Xiong, X.; Yi, C., Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Meth.* **2016,** *14* (1), 23-31.

12. Booth, M.; Ost, T.; Beraldi, D.; Bell, N.; Branco, M.; Reik, W.; Balasubramanian, S., Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat. Prot.* **2013,** *8* (10), 1841-1851.

13. Cui, Y.; Wang, P.; Yu, Y.; Yuan, J.; Wang, Y., Normalized Retention Time for Targeted Analysis of the DNA Adductome. *Anal. Chem.* **2018,** *90*, 14111-14115.

14. Giessing, A. M. B.; Scott, L. G.; Kirpekar, F., A nano-chip-LC/MS based strategy for characterization of modified nucleosides using reduced porous graphitic carbon as a stationary phase. *J. Am. Soc. Mass Spectrom.* **2011,** *22* (7), 1242-1251.

15. Escher, C.; Reiter, L.; MacLean, B.; Ossola, R.; Herzog, F.; Chilton, J.; MacCoss, M.; Rinner, O., Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **2012,** *12* (8), 1111-1121.

16. Yamuchi, K.; Hosokawa, T.; Kinoshita, M., POSSIBLE EFFECT OF HYDROGEN-BONDING ON METHYLATION OF PYRIMIDINE AND PYRIDONE NUCLEOSIDES. *J. Chem. Soc. Perkin Trans. 1* **1989,** (1), 13-15.

17. Mahto, S.; Chow, C., Synthesis and solution conformation studies of the modified nucleoside N-4,2 '-O-dimethylcytidine (m(4)Cm) and its analogues. *Bioorg. Med. Chem.* **2008,** *16* (19), 8795-8800.

18. LaFrancois, C.; Fujimoto, J.; Sowers, L., Synthesis and characterization of isotopically enriched pyrimidine deoxynucleoside oxidation damage products. *Chem. Res. Tox.* **1998,** *11* (1), 75-83.

19. Jones, J.; Robins, R., PURINE NUCLEOSIDES .3. METHYLATION STUDIES OF CERTAIN NATURALLY OCCURRING PURINE NUCLEOSIDES. *J. Am. Chem. Soc.* **1963,** *85* (2), 193-&.

20. Vanaerschot, A.; Mamos, P.; Weyns, N.; Ikeda, S.; Declercq, E.; Herdewijn, P., ANTIVIRAL ACTIVITY OF C-ALKYLATED PURINE NUCLEOSIDES OBTAINED BY CROSS-COUPLING WITH TETRAALKYLTIN REAGENTS. *J. Med. Chem.* **1993,** *36* (20), 2938-2942.

21. You, C.; Dai, X.; Wang, Y., Position-dependent effects of regioisomeric methylated adenine and guanine ribonucleosides on translation. *Nucleic Acids Res.* **2017,** *45* (15), 9059-9067.

22. Sarin, L.; Kienast, S.; Leufken, J.; Ross, R.; Dziergowska, A.; Debiec, K.; Sochacka, E.; Limbach, P.; Fufezan, C.; Drexler, H.; Leidel, S., Nano LC-MS using capillary columns enables accurate quantification of modified ribonucleosides at low femtomol levels. *RNA* **2018,** *24* (10), 1403-1417.

23. Iwanami, Y.; Brown, G., METHYLATED BASES OF RIBOSOMAL RIBONUCLEIC ACID FROM HELA CELLS. *Arch. Biochem. Biophys.* **1968,** *126* (1), 8-15.

Chapter 3: YTHDF2 Binds to 5-Methylcytosine in rRNA

Introduction

RNA harbors more than 100 distinct types of modifications, which modulate its structure and functions.[1] Recent transcriptome-wide mapping studies revealed the widespread occurrence of 5-methylcytidine (m^5C),[2, 3] N^6-methyladenosine (m^6A),[4, 5] N^6,2′-*O*-dimethyladenosine (m^6A_m),[6] and pseudouridine (Ψ)[7-9] in mRNA. In addition, proteins involved in the installation (writers),[10-12] removal (erasers)[13, 14] and recognition (readers)[4, 15-17] of m^6A have been discovered and found to play important roles in modulating the localization, stability, and translational efficiencies of mRNA. These recent exciting findings suggest that post-transcriptional modifications of RNA, similar as methylation of cytosine in DNA and post-translational modifications of histones, may play an epigenetic role in gene expression.[18]

Recent studies also offered some insights into the functions of m^5C in RNA.[3, 19, 20] In this respect, m^5C in tRNA and rRNA were shown to stabilize the secondary structure of tRNA and regulate translational fidelity, respectively.[19, 21] In addition, m^5C is known to be present in mRNA, where a previous transcriptome-wide mapping study revealed the enrichment of m^5C in the untranslated regions of mRNA in HeLa cells,[2] though a recent study showed that m^5C in mRNA may not be as widespread as initially thought.[3] Hence, the functions of m^5C in mRNA remain unclear. NSUN2 and TRDMT2 are two known methyltransferases for the formation of m^5C in eukaryotes (writers),[22, 23] and m^5C in RNA can be converted to 5-hydroxymethylcytidine by ten-eleven translocation (Tet) enzymes

(erasers).[24, 25] A recent study showed that m^5C can interact with mRNA export adaptor ALYREF (reader).[26] However, it is unknown whether other cellular proteins are also involved in the recognition of m^5C in RNA.

In this study, we discovered, by employing an unbiased quantitative proteomics method, a number of candidate protein readers of m^5C, including YTH domain-containing proteins. We also showed that YTH domain-containing family protein 2 (YTHDF2) binds to m^5C-carrying RNA *in vitro* and in cells. In addition, genetic ablation of *YTHDF2* elicited substantial elevations in the levels of m^5C at multiple loci in rRNA. Moreover, YTHDF2 could modulate rRNA maturation in human cells. Together, our study expanded the functions of YTHDF2 and provided a foundation for understanding better the biological functions of m^5C in RNA.

Experimental Section

Cell culture

HeLa and HEK293T cells (ATCC) were cultured at 37°C in Dulbecco's Modified Eagle Medium (DMEM) containing 10% fetal bovine serum (Invitrogen) and 100 units ml^{-1} penicillin and 100 μg ml^{-1} streptomycin (Life Technologies) in an incubator containing 5% CO_2.

For SILAC experiments, DMEM medium without lysine or arginine was obtained from Fisher Scientific. The complete light and heavy DMEM media were prepared by the addition of light or heavy lysine and arginine ([$^{13}C_6$,$^{15}N_2$]-L-lysine and [$^{13}C_6$]-L-arginine, Sigma), along with dialyzed fetal bovine serum, to the above medium. The cells were

cultured in a 37°C incubator for at least 10 days (more than 5 cell doublings) to ensure complete stable isotope incorporation.

Quantitative discovery of m^5C-binding proteins

Biotin-labeled oligoribonucleotides with the sequence of 5′-biotin-ACUGGCUCCUUCCACGUCUCACXAGGCAGACAGU-3′ (X=C or m^5C) were obtained from Integrated DNA Technologies (IDT). HeLa cells cultured in SILAC medium were harvested at 70-80% confluence, washed with PBS and lysed in CelLytic M cell lysis buffer (Sigma). The lysates were centrifuged at 13,000 rpm and at 4°C for 10 min. The supernatant was pre-cleared at 4°C for 1 h by incubation with streptavidin-conjugated agarose beads (Thermo Scientific). Biotinylated RNA baits (3 μg) were incubated, at 4°C for 2 h, with pre-cleared cell lysates in a binding buffer containing 10 mM Tris-HCl (pH 7.5), 150 mM KCl, 1.5 mM $MgCl_2$, 0.05% (v/v) IGEPAL CA-630, 0.5 mM DTT, and 0.4 units μl^{-1} RNase inhibitor (New England Biolabs). Streptavidin-conjugated agarose beads were then added to the mixture, which was kept in a shaker at 4°C for 2 h. The beads were extensively washed, and the heavy and light lysates were then combined. In forward SILAC experiments, the m^5C and the control probes were incubated with the heavy and light isotope-labeled lysates, respectively. The opposite incubations were conducted in reverse SILAC experiments. The samples were separated on a 10% (w/v) SDS-PAGE gel for a short distance (1 cm) and stained with Coomassie blue. The gel was then destained. The proteins were subsequently reduced and alkylated with dithiothreitol and iodoacetamide, respectively, and then digested in gel with trypsin (Roche) at 37°C for 16

h. The resulting tryptic peptides were subsequently extracted from the gel with 5% acetic acid, desalted and analyzed using LC-MS/MS.

LC-MS/MS experiments were performed as previously described.[27] Briefly, the peptides were separated on an EASY-nLC II and analyzed on an LTQ Orbitrap Velos mass spectrometer equipped with a nanoelectrospray ionization source (Thermo). The trapping column (150 μm × 50 mm) and separation column (75 μm × 120 mm) were both packed with ReproSil-Pur C18-AQ resin (3 μm in particle size, Dr. Maisch HPLC GmbH, Germany). The peptide samples were firstly loaded onto the trapping column in CH_3CN/H_2O (2:98, v/v) at a flow rate of 4.0 μl/min, and resolved on the separation column with a 120-min linear gradient of 2-40% acetonitrile in 0.1% formic acid and at a flow rate of 300 nl/min. The LTQ-Orbitrap Velos mass spectrometer was operated in the positive-ion mode, and the spray voltage was 1.8 kV. The full-scan mass spectra (*m*/*z* 300-2000) were acquired with a resolution of 60,000 at *m*/*z* 400 after accumulation to a target value of 500,000 in the linear ion trap. MS/MS data were obtained in a data-dependent scan mode where one full MS scan was followed with 20 MS/MS scans.

Protein identification and quantification were performed using Maxquant,[28] Version 1.2.2.5 against International Protein Index (IPI) database, version 3.68. The maximum number of miscleavages for trypsin was two per peptide. Cysteine carbamidomethylation and methionine oxidation were set as fixed and variable modifications, respectively. The search was performed with the tolerances in mass accuracy of 10 ppm and 0.6 Da for MS and MS/MS, respectively. The required false-positive discovery rate was set at 1% at both the peptide and protein levels, with the

minimal required peptide length being set at 6 amino acids. For obtaining reliable results, the quantification of the protein expression ratio was based on six independent SILAC labeling experiments, which included three forward and three reverse labelings.

Vector construction and protein expression

The human *YTHDF2* gene was amplified from mRNA isolated from HEK293T cells by reverse transcription-PCR to introduce a 5' XbaI site and a 3' BamHI site, and subcloned into pRK7-3×FLAG vector. The pGEX-4T-1-YTHDF2 vector was a gift from Prof. Chuan He.[16] The vector for the YTHDF2-W432A mutant was constructed by site-directed mutagenesis using primers containing the indicated mutations. The primers are listed in **Table 3.1**.

Recombinant YTHDF2 and YTHDF2-W432A proteins were obtained by inducing transformed Rosetta (DE3) pLysS *Escherichia coli* cells with 1 mM isopropyl 1-thio-β-D-galactopyranoside when OD_{600} of the culture reached approximately 0.6, and culturing at room temperature overnight. Subsequently, the recombinant proteins were extracted from the lysate with glutathione agarose (Pierce) following the manufacturer's recommended procedures. The proteins were concentrated and purified using Microcon YM-30 ultracentrifugal filters (Millipore).

Cellular RNA sample preparation and LC-MS/MS/MS measurement

Total RNA was extracted from HEK293T cells using TRI reagent (Sigma), and mRNA was isolated and purified by using PolyATtract mRNA Isolation System IV (Promega) and RiboMinus Transcriptome Isolation Kit (Invitrogen) according to the manufacturers' instructions.

Table 3.1. The primers and probes used in the present study.

Primer Name	Primer Sequence
pRK7-YTHDF2-F	5'- AAATCTAGAATGTCGGCCAGCAGCCTCTTG -3'
pRK7-YTHDF2-R	5'- AAAGGATCCTTTCCCACGACCTTGACGTTCC-3'
YTHDF2W432A-F	5'-GTTCCATTAAGTATAATATTGCGTGCAGCACAGAGC-3'
YTHDF2W432A-R	5'-GCTCTGTGCTGCACGCAATATTATACTTAATGGAAC-3'
ITS1	5'-CCTCGCCCTCCGGGCTCCGTTAATGATC-3'
ITS2	5'-CGCACCCCGAGGAGCCCGGAGGCACCCCCGG-3'

The *in vitro* pull-down experiment was performed using a previously reported method with minor changes.[16] Briefly, recombinant YTHDF2 protein purified from *E. coli* was pretreated with RNase to remove any residual RNA from bacteria cells, washed thoroughly, and incubated with mRNA from HEK293T cells in IPP buffer (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.1% IGEPAL CA-630, 0.5 mM DTT, 40 units ml^{-1} RNase inhibitor) at 4°C for 2 h. GST-affinity beads (Pierce) were then added to the mixture, and the mixture was incubated at 4°C with shaking for another 2 h. Unbound mRNA was recovered from the aqueous phase as the flow-through fraction.

The beads were washed for four times and the YTHDF2-bound mRNA was extracted from the beads using TRI reagent. The procedures for *in vitro* RNA cross-linking and immunoprecipitation (CLIP) were similar as the *in vitro* RNA pull-down experiment, with the following modifications: Before immunoprecipitation with GST-affinity beads,

the RNA-protein mixture was cross-linked by irradiating on ice for three times with 0.15 J/cm^2 of 254 nm UV light each time; after UV cross-linking, the RNA-protein mixture was subjected to RNase T1 digestion (1 unit μl^{-1} RNase T1 for 8 min at 22°C); after immunoprecipitation and washing, the RNA was detached from GST-affinity beads by treating with proteinase K (1 mg ml^{-1}) at 50°C for 30 min, and the RNA was further recovered by using Zymo RNA Clean and Concentrator.

For cellular pull-down experiment, HEK293T cells were transfected with a plasmid encoding FLAG-tagged YTHDF2 or the corresponding W432A mutant. After a 48-h incubation, the cells were washed with PBS and lysed in a lysis buffer, which contained 10 mM HEPES (pH 7.5), 150 mM KCl, 2 mM EDTA, 0.5% IGEPAL CA-630, 0.5 mM DTT, protease inhibitor (Sigma), and 40 units ml^{-1} RNase inhibitor. The supernatant was incubated with anti-FLAG M2 beads (Sigma) at 4°C overnight. The beads were washed for three times with a washing buffer containing 50 mM HEPES (pH 7.5), 200 mM NaCl, 2 mM EDTA, 0.05% IGEPAL CA-630, and 0.5 mM DTT. The beads were then incubated with proteinase K (1.2 mg/ml) at 55°C for 1 h. The sample was centrifuged at 5000 rpm for 1 min and the RNA was recovered from the supernatant.

The LC-MS/MS/MS measurement of RNA samples was performed using previously reported methods with some modifications.[29] Briefly, 100 ng of mRNA was digested with 1 unit of nuclease P1 in a 25-μl buffer containing 25 mM NaCl and 2.5 mM $ZnCl_2$. The mixture was then incubated at 37°C for 2 h, and to the mixture were added 0.5 unit of alkaline phosphatase and 3 μl of 1.0 M NH_4HCO_3. After incubating at 37°C for an additional 2 h, the digestion mixture was dried and reconstituted in 100 μl ddH_2O.

Uniformly ^{15}N-labeled cytidine and [$^{13}C_5$]-m^5C were employed as internal standards for the quantifications of cytidine and m^5C, respectively, and $^{13}C_5$-labeled adenosine and [D_3]-m^6A were used as internal standards for the quantifications of adenosine and m^6A, respectively. The enzymes in the digestion mixture were removed by extraction using chloroform:isoamyl alcohol (24:1). The aqueous layer was dried, reconstituted in 10 μL of ddH_2O, and injected for LC-MS/MS/MS analysis on an LTQ-XL linear ion trap mass spectrometer equipped with nanoelectrospray ionization source and an EASY-nLC II (Thermo). The instrument conditions and scan events were previously described.[29] For cytidine and m^5C quantification, the pre-column and analytical column were packed with porous graphitic carbon (PGC) and Zorbax-SB C18, respectively, where a gradient of 0-15% B in 10 min, 15-35% B in 40 min, 35-90% B in 1 min, and 90% B in 15 min was used. For adenosine and m^6A quantification, both the pre-column and analytical column were packed with Magic C18 AQ, where a gradient of 0-15% B in 40 min, 55-90% B in 1 min, and 90% B in 10 min was employed.

Results

A quantitative proteomics method led to the identification of multiple putative m^5C-interacting proteins

Systematic identifications of m^5C-binding proteins constitute an important step toward understanding the biological functions of m^5C in RNA. Hence, we employed a quantitative proteomics method to screen for proteins in HeLa and HEK293T cells that can bind to m^5C-bearing RNA, where we employed metabolic labeling with SILAC

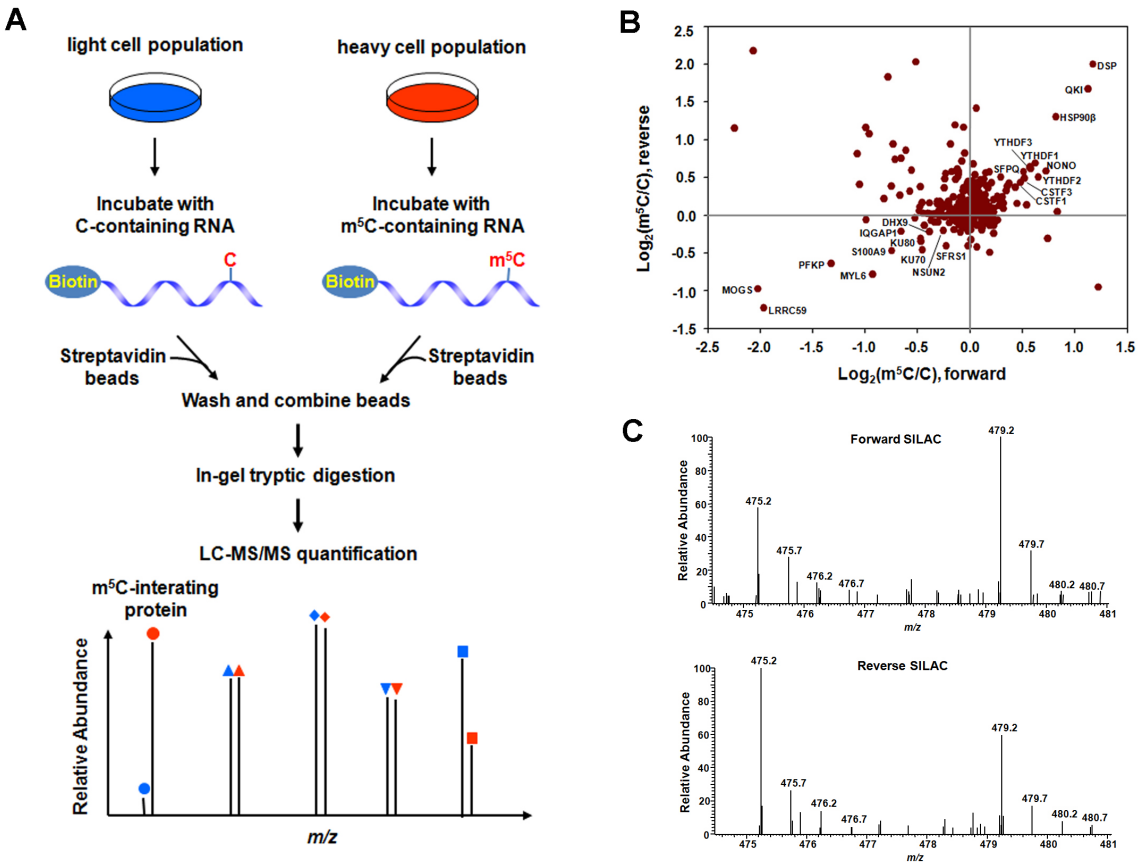


Figure 3.1. The identification of m^{5}C-interacting proteins. (**A**) A schematic overview of the SILAC-based quantitative proteomics method for discovering m^{5}C reader proteins. Shown is the workflow for a forward SILAC labeling experiment. (**B**) A scatter plot showing the proteins identified in RNA pull-down assay in HeLa cells. Displayed are results based on three forward and three reverse SILAC labeling experiments. (**C**) Representative ESI-MS for the $[M+2H]^{2+}$ ions of YTHDF2 peptide SINNYNPK revealing the preferential binding of YTHDF2 toward the m^{5}C probe in both forward (up) and reverse (bottom) SILAC experiments.

(stable isotope labeling by amino acid in cell culture) (**Figure 3.1A**).[30] In this respect, we chose an RNA sequence derived from the mRNA of the human *CINP* gene, which was recently shown to carry an m^{5}C at position 748 (with approximately 46% methylation in HeLa cells),[2] as the probe bait and the corresponding unmethylated sequence as the control bait.

Our results led to the identification of multiple proteins exhibiting preferential binding toward the m^5C-bearing RNA over control (**Figure 3.1**). These proteins include mRNA cleavage stimulation factors CSTF1-3, YTH domain-containing family proteins 1-3 (YTHDF1-3), pre-mRNA splicing factors SFPQ/NONO, and others (ratio of $m^5C/C>1.4$, **Tables 3.2 and 3.3**). **Figure 3.1C** depicts the representative electrospray ionization-mass spectrometry (ESI-MS) results for a tryptic peptide derived from YTHDF2 (MS/MS for the peptide are shown in **Figure 3.2**), which supports the preferential binding of YTHDF2 toward the m^5C probe.

YTHDF2 is a reader for m^5C-containing RNA

Because YTHDF2 was previously found to be a reader protein for m^6A and N^1-methyladenosine,[4, 31] we decided to choose this protein for further investigation. In this context, we examined whether YTHDF2 can bind directly to m^5C in RNA by performing electrophoretic mobility shift assay (EMSA) with recombinant YTHDF2 protein purified from *E. coli*. Our results showed that YTHDF2 binds more strongly to an m^5C-carrying RNA substrate than its unmethylated counterpart, though the binding affinity is much weaker than that toward m^6A-containing RNA (**Figure 3.3**).

The X-ray crystal structure of YTHDF2 revealed three aromatic amino acid residues in the hydrophobic pocket of YTHDF2 that are crucial for its recognition of m^6A.[32, 33] To explore whether this hydrophobic pocket also assumes an important role in binding toward m^5C, we conducted EMSA experiment with a mutant form of YTHDF2 protein where the conserved Trp^{432} was mutated to an alanine (W432A). The result indeed showed that the mutation led to a reduction in binding affinity toward the m^5C-containing

Table 3.2. A list of proteins with relative binding ratios toward m^5C- over C-containing RNA identified from SILAC-based affinity screening experiments with the use of lysate of HeLa cells.

Protein name	**Average (m^5C/C)**	**S.D.**
CSTF3 Cleavage stimulation factor 77 kDa subunit	1.452	0.058
CSTF2 Beta CstF-64 variant 2	1.449	0.052
CSTF1 Cleavage stimulation factor subunit 1	1.421	0.117
YTHDF1 YTH domain family protein 1	1.576	0.050
YTHDF3 YTH domain family protein 3	1.516	0.071
YTHDF2 Isoform 1 of YTH domain family protein 2	1.496	0.140
YTHDC2 Probable ATP-dependent RNA helicase YTHDC2	1.106	0.105
SFPQ Isoform Long of Splicing factor, proline- and glutamine-rich	1.514	0.202
NONO Non-POU domain-containing octamer-binding protein	1.577	0.173

Table 3.3. A list of proteins with relative binding ratios toward m^5C- over C-containing RNA identified from SILAC-based affinity screening experiments with the use of lysate of HEK293T cells.

Protein name	**Average (m^5C/C)**	**S.D.**
CSTF2 Isoform 1 of Cleavage stimulation factor 64 kDa subunit	2.822	1.889
CSTF1 Cleavage stimulation factor 50 kDa subunit	1.657	0.191
CSTF3 Cleavage stimulation factor 77 kDa subunit	1.550	0.602
YTHDF2 Isoform 1 of YTH domain family protein 2	2.340	1.633
SFPQ Isoform Long of Splicing factor, proline- and glutamine-rich	4.763	2.770
NONO Non-POU domain-containing octamer-binding protein	4.108	1.956

Figure 3.2. Representative MS/MS data of a tryptic peptide from YTHDF2 in SILAC experiments. Shown are the MS/MS for the $[M+2H]^{2+}$ ions of YTHDF2 peptide SINNYNPK (**A**) and SINNYNPK* (**B**, 'K*' designates the heavy lysine), respectively.

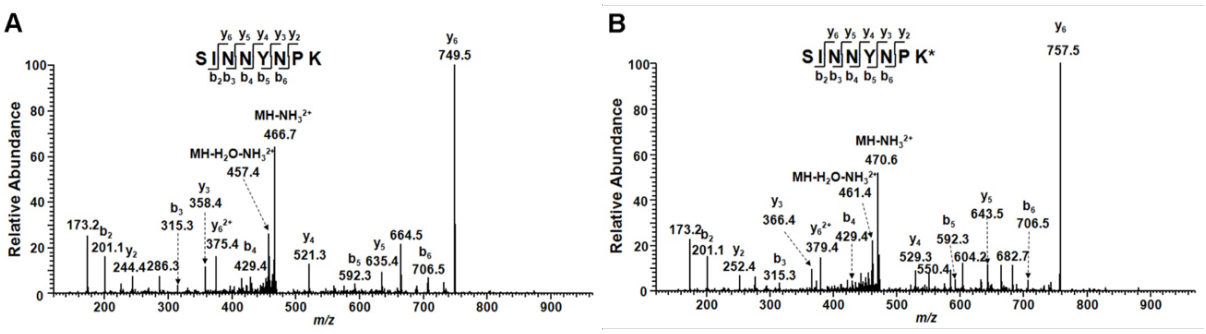
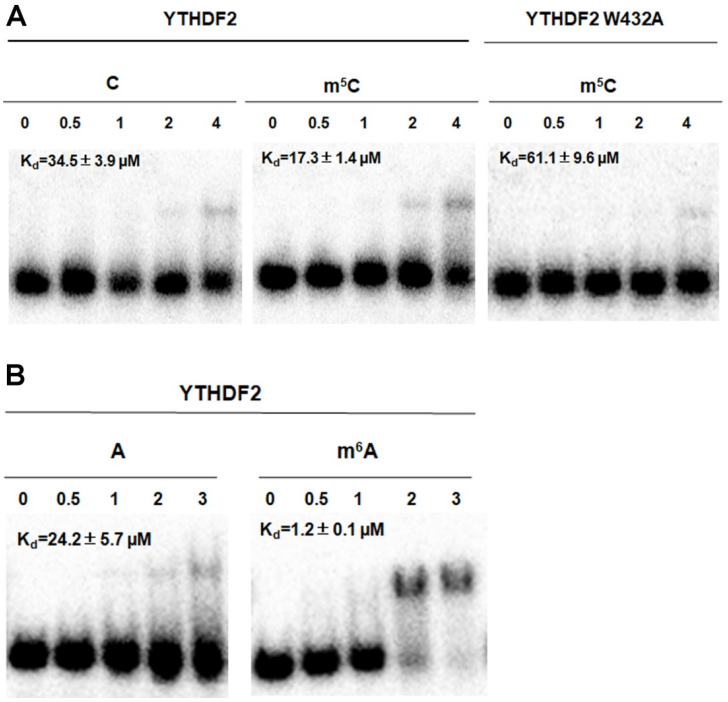


Figure 3.3. Electrophoretic mobility shift assay for measuring the binding affinities of YTHDF2 and W432A mutant proteins with methylated and unmethylated RNA probes. (**A**) The binding affinity of YTHDF2 and W432A mutant proteins with m^5C- and C-containing RNA probes. (**B**) The binding affinity of YTHDF2 with m^6A- and A-containing RNA probes. Protein concentrations ranged from 0.5 to 4 μM. The dissociation constants (K_d) are listed in individual figure panels, and the data represent the mean ± S. D. from three separate EMSA experiments.



probe (**Figure 3.3**), suggesting that m^5C may bind to the same hydrophobic pocket in YTHDF2 that is required for m^6A recognition.

To further substantiate the above findings, we performed *in vitro* pull-down assay to determine if recombinant YTHDF2 can allow for the enrichment of m^5C-carrying mRNA. LC-MS/MS/MS analysis of the mononucleoside mixture arising from the enzymatic digestion of the poly(A)-tailed mRNA samples revealed that the level of m^5C was significantly higher in the YTHDF2-bound fraction than the input or flow-through fraction (**Figure 3.4A**). When the YTHDF2 protein was cross-linked with its associated RNA using UV light and the cross-liked RNA was partially digested using RNase T1, the enrichment of m^5C in the YTHDF2-bound fraction was increased compared to that without RNase T1 digestion (**Figure 3.4A**). Similar observations were made for m^6A (**Figure 3.4B**), which is in keeping with the previous finding of YTHDF2 binding m^6A.[16] We further expressed FLAG-tagged wild-type YTHDF2 and the W432A mutant in HEK293T cells, immunoprecipitated the proteins using anti-FLAG beads, and quantified the levels of m^5C in the total RNA samples isolated from the immunoprecipitated proteins. Our results showed that the levels of m^5C were significantly higher in the pull-down samples of wild-type YTHDF2 than those of the W432A mutant (**Figure 3.4C**). Together, the above results support that YTHDF2 can bind directly to m^5C, and this binding entails the intact hydrophobic pocket of YTHDF2 that is also involved with m^6A binding.

We realized that the exact mechanism of recognition of m^5C by YTHDF2 necessitates structural studies. We have attempted, but failed to obtain the crystal structure

of complex with m^5C-bearing RNA. Therefore, we performed structural modeling for YTHDF2 in YTHDF2-m^5C-RNA using the crystal structures of YTH domain in complex with m^6A-carrying RNA as a reference model.[33, 34] The results showed that m^6A is docked in the aromatic cage of the YTH domain comprised of W432, W486 and W491 (**Figure 3.4D**). Aside from the hydrophobic interaction, the side chain carbonyl oxygen of D422 forms a hydrogen bond with the *N*1 of m^6A at an average distance of 3.0 Å. The mode of binding is nearly identical to that observed in the crystal structure,[33] which validates our docking method. The results from the corresponding docking of m^5C-carrying RNA showed that m^5C is sandwiched between W432 and W491 at the hydrophobic cage (**Figure 3.4E**). In addition, the side-chain carbonyl oxygen of D422 forms a hydrogen bond with the exocyclic N^4 nitrogen of m^5C at an average distance of 3.0 Å. Hence, the comparison of the docking structures showed a similar mode of recognition of m^5C and m^6A by the YTH domain of YTHDF2.

YTHDF2 regulates the m^5C profile in RNA of human cells

Others revealed that YTHDF2 interacts physically with components of m^6A writers[35] and it protects the m^6A in the 5′ UTR of stress-induced transcripts by limiting the FTO-mediated demethylation of m^6A.[36] Thus, we next investigated how the absence of YTHDF2 in HEK293T cells alters the distribution of m^5C in RNA by employing a recently reported method of bisulfite conversion and next-generation sequencing analysis,[2, 3] where we knocked out the *YTHDF2* gene by using the CRISPR-Cas9 system and conducted the experiments in three replicates.

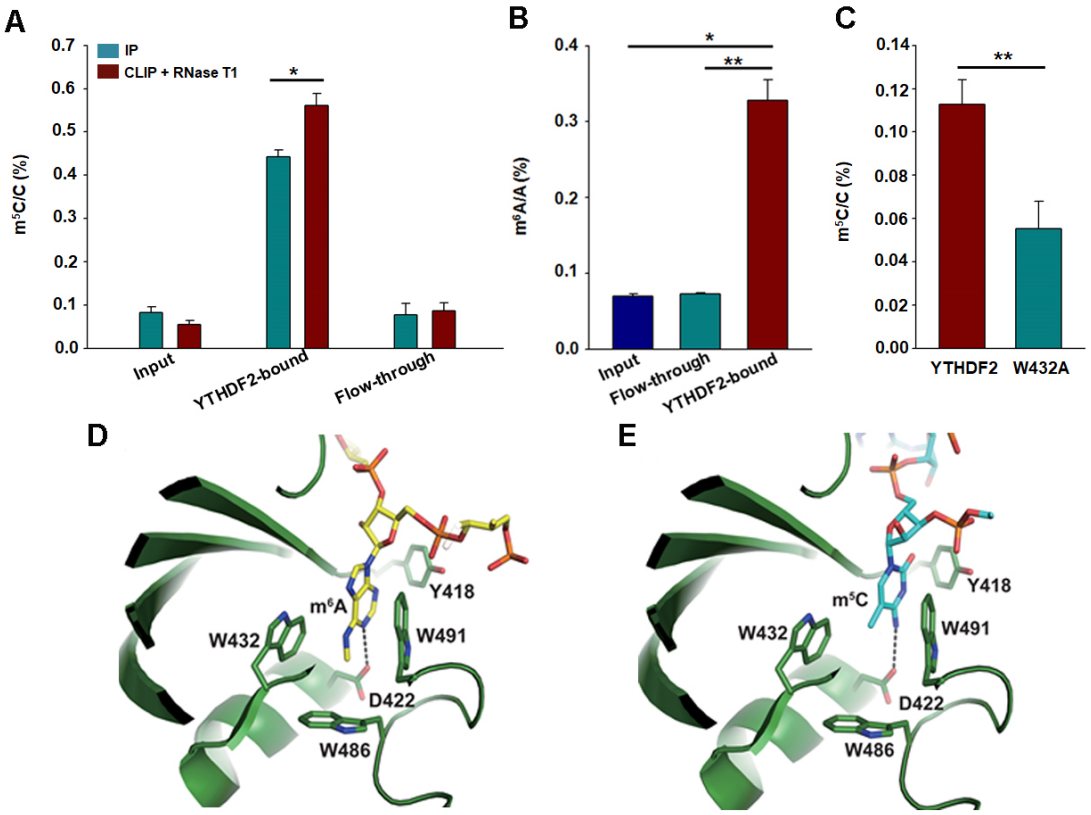


Figure 3.4. YTHDF2 is an m^5C-binding protein. (**A**) LC-MS/MS/MS quantification results showed that recombinant YTHDF2 protein can enrich m^5C-containing RNA from poly(A)-tailed mRNAs of HEK293T cells. When YTHDF2 was cross-linked with its associated RNA by UV light and partially digested using RNase T1, the enrichment of m^5C in the YTHDF2-bound fraction was increased compared to that without RNase T1 treatment. (**B**) LC-MS/MS/MS results showed that m^6A was enriched in YTHDF2-bound mRNA than in the input or flow-through samples. (**C**) Relative enrichment of m^5C in total RNA products immunoprecipitated with wild-type YTHDF2 over W432A mutant protein from HEK293T cells. The data in (**A**-**C**) represent the mean and SEM of results from three (**A**, **B**) or six (**C**) technical replicates, i.e. three parallel pull-down experiments each for data in (**A**) and (**B**), and six independent transfection and pull-down experiments for data in (**C**). '*', $P< 0.05$; '**', $P< 0.01$. The *P* values were calculated using unpaired (**A**, **C**) or paired (**B**) two-tailed Student's *t*-test. (**D**-**E**) Closed-up view of the docking models of the YTH domain of YTHDF2 with the 3-mer RNA housing an m^6A (**D**) or m^5C (**E**) in similar orientation. The comparison shows similar mode of recognition of m^6A and m^5C. RNAs and the residues involved in the recognition are depicted in stick representation. Hydrogen bonds are shown in dashed lines. The nitrogen and oxygen atoms are colored in blue and red, respectively. YTH domain is colored in green, m^6A- and m^5C-containing RNA, i.e. G(m^6A)C and G(m^5C)C, are displayed in yellow and cyan, respectively.

To process the bisulfite sequencing data, we employed a recently reported statistical method to calculate the adjusted *p* values,[3] which was used to define the final list of the 'authentic' m^5C sites, and the mean m^5C rate, which was used for comparing the levels of m^5C between HEK293T cells and the isogenic YTHDF2 knockout cells. To examine whether there is a global change in m^5C level, we performed a linear regression analysis using the m^5C rates between HEK293T and YTHDF2 knockout cells in three groups of loci derived from rRNA, mitochondrial RNA (mtRNA), and others (primarily mRNAs). The results revealed no appreciable change in m^5C rates in mtRNAs upon deletion of *YTHDF2* gene [$f(x) = 1.0046x + 0.0009$; $R^2 = 0.9928$, where 'x' and '$f(x)$' represent m^5C rates in HEK293T and YTHDF2 knockout cells, respectively] (**Figure 3.5**). This result underscores that the sample processing (bisulfite treatment, cloning and sequencing) and data analysis workflow does not elicit a global bias favoring the YTHDF2 knockout cells or the parental HEK293T cells.

The m^5C levels in the rRNAs, however, increased globally in the YTHDF2 knockout cells over the parental cells, where there is an approximately 1.6-fold increase in m^5C levels upon genetic ablation of *YTHDF2* [$f(x) = 1.6327x + 0.0028$ and $R^2 = 0.9307$. **Figure 3.6B**]. Interestingly,the sites with high m^5C levels are clustered in 5 regions of mature rRNAs, including one on the 18S rRNA and 4 on the 28S rRNA (**Figure 3.6A**). In this regard, LC-MS/MS/MS analysis of 18S rRNA also revealed that the level of m^5C, but not m^6A, in 18S rRNA was significantly higher in the YTHDF2 knockout cells (**Figure 3.6C**).

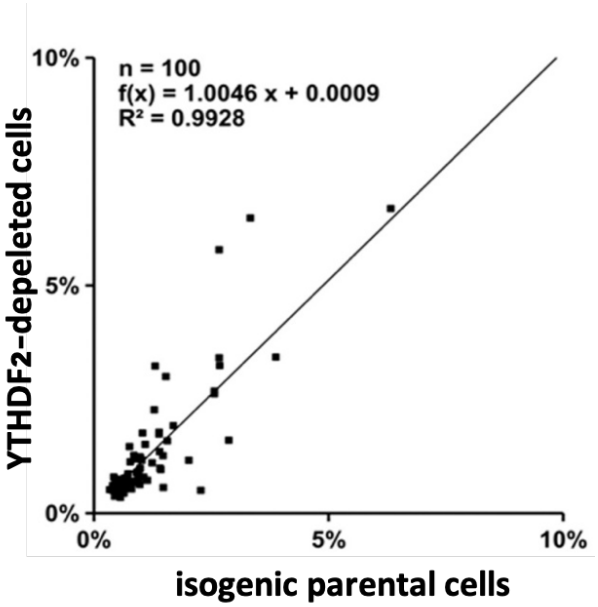


Figure 3.5 Linear regression analysis to show the overall levels of m^5C in mtRNA of the YTHDF2-depeleted cells (Y axis) and the isogenic parental cells (X axis). '*x*' and '*f(x)*' represent m^5C rates in HEK293T and YTHDF2 knockout cells, respectively.

Discussion

m^5C is one of the most prevalent post-transcriptional modifications of RNA in human cells [3, 19, 20]. In this study, we employed a state-of-the-art bioanalytical method, which involves SILAC-based quantitative proteomics, to screen for interacting proteins of m^5C, and this experiment led to the identification of multiple putative readers of m^5C, including three YTH domain-containing proteins (YTHDF1-3, **Figure 3.1**). We further showed that the purified recombinant YTHDF proteins can bind directly to m^5C-containing RNA *in vitro* (**Figure 3.3**). Moreover, robust quantification by using LC-MS/MS/MS with the stable isotope-dilution method revealed that affinity pull-down using recombinant YTHDF2 and RNA-protein photocross-linking followed by pull-down of ectopically expressed YTHDF2 both lead to enrichment of m^5C (**Figure 3.4**). Similar LC-MS/MS/MS

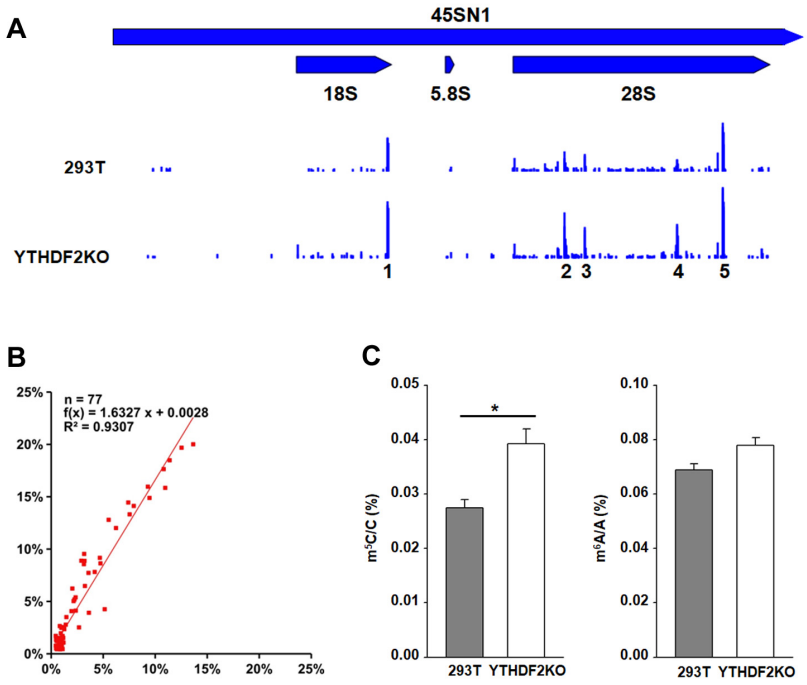


Figure 3.6 YTHDF2 regulates the m^{5}C profile in rRNA. (**A**) The locations and levels of m^{5}C in mature rRNA of HEK293T and the isogenic YTHDF2 knockout cells. The sites with high m^{5}C levels are clustered in 5 regions, including one in the 18S rRNA and four in the 28S rRNA. (**B**) Linear regression analysis in the rRNAs showing that an approximately 1.6- fold increase in m^{5}C levels upon genetic ablation of *YTHDF2*. '*x*' (X axis) and '*f(x)*' (Y axis) represent m^{5}C rates in HEK293T and YTHDF2 knockout cells, respectively. (**C**) LC-MS/MS/MS quantification results showing the total levels of m^{5}C (left) and m^{6}A (right) in 18S rRNA of HEK293T and the isogenic YTHDF2 knockout cells. Data represent the mean ± SEM (n = 3). '*', $p < 0.05$; The p values were calculated using unpaired two-tailed Student's *t*-test.

measurement showed that genetic ablation of YTHDF2 led to augmented levels of m^{5}C in 18S rRNA (**Figure 3.6**).

Different members of the YTH domain-carrying proteins are known to exert distinct functions, where bindings of m^{6}A-bearing mRNA with YTHDF2 and YTHDF1 result in mRNA decay[16] and promotion of translation,[17] respectively. The characterizations

of reader proteins of m^5C is an important step toward understanding the biological functions of m^5C in RNA.

Our findings, along with recent structural and functional studies of YTH-domain family proteins, support that YTHDF2 is a versatile reader that can recognize m^6A, m^1A and m^5C,[4, 31] though the binding affinity toward m^5C is much weaker than that toward m^6A. In this vein, several published structural studies of YTH domain-containing proteins uncovered a hydrophobic pocket comprised of two or three aromatic residues that are instrumental for the recognition of methyl group in m^6A [32-34, 37, 38]. It is conceivable that the same hydrophobic pocket may be able to accommodate the 5-methyl group of m^5C. This is indeed supported by our findings that mutation of one of the conserved aromatic residues, i.e. W432, at the hydrophobic pocket to an alanine reduces markedly the binding affinity of the protein toward m^5C-carrying RNA. In addition, our docking studies revealed a very similar mode of recognition of m^5C and m^6A by the YTH domain of YTHDF2. Furthermore, all YTH domain-housing proteins, except YTHDC1 which preferentially binds m^6A in GG(m^6A)C sequence, bind m^6A regardless of sequence context [37], despite the fact that m^6A is known to exist in RRAC ('R' is a G or A) and RGAC consensus sequence motifs in mammalian and yeast species, respectively [4, 5]. The lack of recognition of the consensus flanking nucleobase(s) of m^6A is consistent with the notion that YTH domain-containing proteins can recognize, in addition to m^6A, other post-transcriptionally modified ribonucleosides (i.e. m^1A and m^5C).

We also found that YTHDF2 modulates the distribution of m^5C in both coding and non-coding RNA, where its depletion led to pronounced increases in the levels of m^5C at

multiple sites in rRNA and decreases in m^5C levels at many sites in mRNA. These results suggest that YTHDF2 may play versatile roles in regulating m^5C levels. In this vein, previous reports showed that binding of YTHDF2 with m^6A-bearing mRNA decreases m^6A levels by P-body-mediated mRNA degradation,[16] but increases the m^6A levels of some stress-induced transcripts by protecting them from demethylation.[39] Moreover, m^5C in rRNA is known to modulate translational fidelity;[21] hence, the findings made from the present study suggest that the function of m^5C in this process may involve, at least partly, its recognition by YTHDF2.

In summary, we identified, by using an unbiased quantitative proteomic method, YTHDF2 protein as a reader of m^5C in RNA, which expands the repertoire of RNA modifications that can be regulated by this protein. In this vein, polymorphism in *YTHDF2* gene was previously found to be associated with longevity in humans [40]; thus, the results from the present study suggest that the function of YTHDF2 in this process could be attributed, in part, to its recognition of m^5C in RNA. Furthermore, the data from our SILAC-based quantitative proteomic screening showed that m^5C may also be recognized by other cellular proteins. Further characterizations of these proteins in m^5C recognition and RNA metabolism will promote our understanding about the biological functions of m^5C in RNA.

References

1. Cantara, W. A.; Crain, P. F.; Rozenski, J.; McCloskey, J. A.; Harris, K. A.; Zhang, X.; Vendeix, F. A.; Fabris, D.; Agris, P. F., The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.* **2011,** *39* (Database issue), D195-201.

2. Squires, J. E.; Patel, H. R.; Nousch, M.; Sibbritt, T.; Humphreys, D. T.; Parker, B. J.; Suter, C. M.; Preiss, T., Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* **2012,** *40* (11), 5023-33.

3. Legrand, C.; Tuorto, F.; Hartmann, M.; Liebers, R.; Jacob, D.; Helm, M.; Lyko, F., Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res.* **2017,** *27* (9), 1589-1596.

4. Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.; Sorek, R.; Rechavi, G., Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **2012,** *485* (7397), 201-6.

5. Meyer, K. D.; Patil, D. P.; Zhou, J.; Zinoviev, A.; Skabkin, M. A.; Elemento, O.; Pestova, T. V.; Qian, S. B.; Jaffrey, S. R., 5' UTR m^6A promotes cap-independent translation. *Cell* **2015,** *163* (4), 999-1010.

6. Mauer, J.; Luo, X.; Blanjoie, A.; Jiao, X.; Grozhik, A. V.; Patil, D. P.; Linder, B.; Pickering, B. F.; Vasseur, J. J.; Chen, Q.; Gross, S. S.; Elemento, O.; Debart, F.; Kiledjian, M.; Jaffrey, S. R., Reversible methylation of m^6A_m in the 5' cap controls mRNA stability. *Nature* **2017,** *541* (7637), 371-375.

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

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

9. Khoddami, V.; Yerra, A.; Mosbruger, T. L.; Fleming, A. M.; Burrows, C. J.; Cairns, B. R., Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc. Natl. Acad. Sci. USA* **2019,** *116* (14), 6784-6789.

10. Bokar, J. A.; Shambaugh, M. E.; Polayes, D.; Matera, A. G.; Rottman, F. M., Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA* **1997,** *3* (11), 1233-47.

11. Liu, J.; Yue, Y.; Han, D.; Wang, X.; Fu, Y.; Zhang, L.; Jia, G.; Yu, M.; Lu, Z.; Deng, X.; Dai, Q.; Chen, W.; He, C., A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **2014,** *10* (2), 93-5.

12. Ping, X. L.; Sun, B. F.; Wang, L.; Xiao, W.; Yang, X.; Wang, W. J.; Adhikari, S.; Shi, Y.; Lv, Y.; Chen, Y. S.; Zhao, X.; Li, A.; Yang, Y.; Dahal, U.; Lou, X. M.; Liu, X.; Huang, J.; Yuan, W. P.; Zhu, X. F.; Cheng, T.; Zhao, Y. L.; Wang, X.; Rendtlew Danielsen, J. M.; Liu, F.; Yang, Y. G., Mammalian WTAP is a regulatory subunit of the RNA N^6-methyladenosine methyltransferase. *Cell Res.* **2014,** *24* (2), 177-89.

13. Jia, G.; Fu, Y.; Zhao, X.; Dai, Q.; Zheng, G.; Yang, Y.; Yi, C.; Lindahl, T.; Pan, T.; Yang, Y. G.; He, C., N^6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **2011,** *7* (12), 885-7.

14. Zheng, G.; Dahl, J. A.; Niu, Y.; Fedorcsak, P.; Huang, C. M.; Li, C. J.; Vagbo, C. B.; Shi, Y.; Wang, W. L.; Song, S. H.; Lu, Z.; Bosmans, R. P.; Dai, Q.; Hao, Y. J.; Yang, X.; Zhao, W. M.; Tong, W. M.; Wang, X. J.; Bogdan, F.; Furu, K.; Fu, Y.; Jia, G.; Zhao, X.; Liu, J.; Krokan, H. E.; Klungland, A.; Yang, Y. G.; He, C., ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **2013,** *49* (1), 18-29.

15. Schwartz, S.; Agarwala, S. D.; Mumbach, M. R.; Jovanovic, M.; Mertins, P.; Shishkin, A.; Tabach, Y.; Mikkelsen, T. S.; Satija, R.; Ruvkun, G.; Carr, S. A.; Lander, E. S.; Fink, G. R.; Regev, A., High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **2013,** *155* (6), 1409-21.

16. Wang, X.; Lu, Z.; Gomez, A.; Hon, G. C.; Yue, Y.; Han, D.; Fu, Y.; Parisien, M.; Dai, Q.; Jia, G.; Ren, B.; Pan, T.; He, C., N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014,** *505* (7481), 117-20.

17. Wang, X.; Zhao, B. S.; Roundtree, I. A.; Lu, Z.; Han, D.; Ma, H.; Weng, X.; Chen, K.; Shi, H.; He, C., N^6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **2015,** *161* (6), 1388-99.

18. He, C., Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.* **2010,** *6* (12), 863-5.

19. Motorin, Y.; Lyko, F.; Helm, M., 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Res.* **2010,** *38* (5), 1415-30.

20. Hussain, S.; Aleksic, J.; Blanco, S.; Dietmann, S.; Frye, M., Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol.* **2013,** *14* (11), 215.

21. Chow, C. S.; Lamichhane, T. N.; Mahto, S. K., Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. *ACS Chem. Biol.* **2007,** *2* (9), 610-9.

22. Brzezicha, B.; Schmidt, M.; Makalowska, I.; Jarmolowski, A.; Pienkowska, J.; Szweykowska-Kulinska, Z., Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the pre-tRNA Leu (CAA). *Nucleic Acids Res.* **2006,** *34* (20), 6034-43.

23. Goll, M. G.; Kirpekar, F.; Maggert, K. A.; Yoder, J. A.; Hsieh, C. L.; Zhang, X.; Golic, K. G.; Jacobsen, S. E.; Bestor, T. H., Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* **2006,** *311* (5759), 395-8.

24. Fu, L.; Guerrero, C. R.; Zhong, N.; Amato, N. J.; Liu, Y.; Liu, S.; Cai, Q.; Ji, D.; Jin, S. G.; Niedernhofer, L. J.; Pfeifer, G. P.; Xu, G. L.; Wang, Y., Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **2014,** *136* (33), 11582-5.

25. Huber, S. M.; van Delft, P.; Mendil, L.; Bachman, M.; Smollett, K.; Werner, F.; Miska, E. A.; Balasubramanian, S., Formation and abundance of 5-hydroxymethylcytosine in RNA. *ChemBioChem* **2015,** *16* (5), 752-5.

26. Yang, X.; Yang, Y.; Sun, B. F.; Chen, Y. S.; Xu, J. W.; Lai, W. Y.; Li, A.; Wang, X.; Bhattarai, D. P.; Xiao, W.; Sun, H. Y.; Zhu, Q.; Ma, H. L.; Adhikari, S.; Sun, M.; Hao, Y. J.; Zhang, B.; Huang, C. M.; Huang, N.; Jiang, G. B.; Zhao, Y. L.; Wang, H. L.; Sun, Y. P.; Yang, Y. G., 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res.* **2017,** *27* (5), 606-625.

27. Dai, X.; Otake, K.; You, C.; Cai, Q.; Wang, Z.; Masumoto, H.; Wang, Y., Identification of novel alpha-*N*-methylation of CENP-B that regulates its binding to the centromeric DNA. *J. Proteome Res.* **2013,** *12* (9), 4167-75.

28. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008,** *26* (12), 1367-72.

29. Fu, L.; Amato, N. J.; Wang, P.; McGowan, S. J.; Niedernhofer, L. J.; Wang, Y., Simultaneous quantification of methylated cytidine and adenosine in cellular and tissue

RNA by nano-flow liquid chromatography-tandem mass spectrometry coupled with the stable isotope-dilution method. *Anal. Chem.* **2015,** *87* (15), 7653-9.

30. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002,** *1* (5), 376-86.

31. Dai, X.; Wang, T.; Gonzalez, G.; Wang, Y., Identification of YTH domain-containing proteins as the readers for *N*1-methyladenosine in RNA. *Anal. Chem.* **2018,** *90*, 6380-6384.

32. Zhu, T.; Roundtree, I. A.; Wang, P.; Wang, X.; Wang, L.; Sun, C.; Tian, Y.; Li, J.; He, C.; Xu, Y., Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N^6-methyladenosine. *Cell Res.* **2014,** *24* (12), 1493-6.

33. Li, F.; Zhao, D.; Wu, J.; Shi, Y., Structure of the YTH domain of human YTHDF2 in complex with an m^6A mononucleotide reveals an aromatic cage for m^6A recognition. *Cell Res.* **2014,** *24* (12), 1490-2.

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

35. Schwartz, S.; Mumbach, M. R.; Jovanovic, M.; Wang, T.; Maciag, K.; Bushkin, G. G.; Mertins, P.; Ter-Ovanesyan, D.; Habib, N.; Cacchiarelli, D.; Sanjana, N. E.; Freinkman, E.; Pacold, M. E.; Satija, R.; Mikkelsen, T. S.; Hacohen, N.; Zhang, F.; Carr, S. A.; Lander, E. S.; Regev, A., Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* **2014,** *8* (1), 284-96.

36. Zhou, J.; Wan, J.; Gao, X.; Zhang, X.; Jaffrey, S. R.; Qian, S. B., Dynamic m^6A mRNA methylation directs translational control of heat shock response. *Nature* **2015,** *526* (7574), 591-4.

37. Xu, C.; Liu, K.; Ahmed, H.; Loppnau, P.; Schapira, M.; Min, J., Structural basis for the discriminative recognition of N^6-methyladenosine RNA by the human YT521-B homology domain family of proteins. *J. Biol. Chem.* **2015,** *290* (41), 24902-13.

38. Luo, S.; Tong, L., Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proc. Natl. Acad. Sci. USA* **2014,** *111* (38), 13834-9.

39. Zhou, J.; Wan, J.; Gao, X.; Zhang, X.; Jaffrey, S. R.; Qian, S. B., Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **2015,** *526* (7574), 591-4.

40. Cardelli, M.; Marchegiani, F.; Cavallone, L.; Olivieri, F.; Giovagnetti, S.; Mugianesi, E.; Moresi, R.; Lisa, R.; Franceschi, C., A polymorphism of the YTHDF2 gene (1p35) located in an Alu-rich genomic domain is associated with human longevity. *J. Gerontol. A Biol. Sci. Med. Sci.* **2006,** *61* (6), 547-56.

Chapter 4: Tet-mediated oxidation of 5-methyluridine RNA

Introduction

In cells, the four canonical ribonucleosides are heavily modified, resulting in over 100 diverse chemical structures.[1-4] N^6-methyladenosine (m^6A) and 5-methylcytidine (m^5C) are examples of modified ribonucleosides with known regulatory proteins, i.e., readers, writers, and erasers.[5-8] The dynamic regulation of modified nucleobases differ among the various types of RNA. For example, m^5C has been identified in transfer RNA (tRNA), messenger RNA (mRNA), and ribosomal RNA (rRNA).[7, 9] However, the regulatory proteins are RNA type-specific, as demonstrated by DNA methyltransferase 2 (DNMT2) and NSUN2, which are responsible for m^5C methylation in tRNA and rRNA, respectively.[10-12] Perturbation of epitranscriptomic regulators has been linked to modified ribonucleoside levels which in turn lead to disease, including cancers, viral infection, and diabetes.[2, 13-15] Therefore, it is imperative to identify modified ribonucleosides and their regulatory proteins.

One potentially important, but insufficiently investigated modified ribonucleoside is 5-methyluridine (m^5U). For over 50 years, we have known that m^5U is present in tRNA.[16, 17] Global modified ribonucleoside profiling has consistently identified m^5U as a prevalent modification in total RNA, rRNA and small nucleolar RNA (snRNA).[1, 2, 4] Even though, tRNA and rRNA methyltransferases, including TrmA, TrmU54, and RlmC, have been identified in *E. coli,* none have been identified in mammalian or other higher-order organisms.[18-20] In addition, investigations have yet to identify potential eraser or reader

proteins for m^5U; however, ten-eleven translocation (Tet) enzymes have been shown to act on similar substrates.[21]

Tet proteins are a family of enzymes in the Fe^{2+} and 2-oxoglutarate-dependent dioxygenases that oxidize *N*-alkylated nucleobases in both DNA and RNA.[22] Tet proteins can sequentially oxidatively demethylate 5-methylcytosine to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-foC), and 5-carboxycytosine (5-caC) in DNA and RNA.[23-25] Tet proteins were also shown to oxidize thymidine to 5-hydroxymethyl-2'-deoxyuridine in DNA.[21] Due to the structural similarity between 5-methylcytosine and 5-methyluridine and Tet's ability to oxidize thymidine, we were prompted to investigate the possibility of Tet to produce 5-hydroxymethyluridine from m^5U in RNA.

Experimental Section

Materials

All chemicals and enzymes unless otherwise specified were from Sigma-Aldrich (St. Louis, MO). We obtained *erythro*-9-(2-hydroxy-3-nonyl)adenine (EHNA) hydrochloride, [1,3-$^{15}N_2$]-cytidine, and 5-methylcytidine were obtained from Tocris Bioscience (Ellisville, MO), Cambridge Isotope Laboratories (Andover, MA), and Berry & Associates (Dexter, MI), respectively. HEK293T human embryonic kidney cells, HeLa human cervical cancer cells, and WM-266-4 human melanoma cells, and cell culture reagents were purchased from ATCC (Manassas, VA, USA). Expression vectors for the catalytic domains of human Tet1[26], human Tet2 (amino acids 1129-2002)[27], mouse Tet3

contained amino acids 697-1669, and those of the full-length Tet1-Tet3, were prepared as previously described.[24]

Total RNA samples from mouse embryonic stem (ES) cells and mouse ES cells with the depletion of all three Tet genes were isolated previously.[28] In addition, total RNA of tissues from 19-21 week old mice were also employed for m^5U and 5-hmrU analyses.

Cell Culture and Transfection

HEK293T cell were cultured using ATCC's Dulbecco's Modified Eagle Medium, while HeLa and WM-266-4 cells were cultured in Eagle's Minimum Essential Medium (ATCC), supplemented with 10% fetal bovine serum and penicillin (100 IU/mL). The cells were maintained at 37°C in a 5% CO_2 atmosphere.

Transfections were carried out using a previously published method.[24] Briefly, at 50% cell confluence level, cells were transfected with 1.5 μg of pGEM-T Easy plasmid (Pormega), catalytic domain or full-length Tet1-3 plasmids with Mirus Bio TransIT transfection reagent (Thermo Fisher) using the manufacturer's recommended protocol. However, full-length Tet2 plasmid was transfected with 3.0 μg of plasmid. All cells were harvested for total RNA extraction at 48 hours post transfection.

RNA extraction, digestion, and off-line HPLC enrichment

Total RNA was isolated from HEK293T, WM266-4, HeLa cultured cells and mouse pancreas, brain, and spleen tissues by Tri-Reagent (Sigma-Aldrich) following the manufacturer's recommended procedures. RNA digestion to individual ribonucleosides was performed following previously published procedures.[9] Briefly, the total RNA samples were enzymatically digested to ribonucleosides with nuclease P1 and phosphodiesterase 2

at 37 °C for 4 hrs. A second enzyme mixture, comprising of alkaline phosphatase and phosphodiesterase 1, was added to the resulting RNA solution and incubated at 37 °C for 2 hrs. The digestion mixture (50 ng) was aliquoted for 5-methyluridine (m^5U) analysis. To the remaining digestion mixture (9,950 ng) of total RNA, 200 fmol of [1,3-$^{15}N_2$]5-hmrC and 200 fmol of [1,3-$^{15}N_2$]5-hmrU were added. The enzymes were subsequently removed from the digestion mixture by chloroform extraction followed by drying of the aqueous layer. The resulting residues were reconstituted in doubly distilled water and subjected to off-line HPLC separation for enrichment of 5-hmrC and 5-hmrU.

PolyA mRNA was obtained using PolyATtract mRNA Isolation Systems (Promega) followed by removal of rRNA contaminations using the RiboMinus Transcriptome Isolation Kit (Invitrogen). Digestion was scaled down for a 1 μg digestion following the above procedures. From the resulting digestion mixture 30 ng was aliquoted for m^5U analysis. The remaining digestion mixture was supplemented with 20 fmol of [1,3-$^{15}N_2$]5-hmrC and 20 fmol of [1,3-$^{15}N_2$]5-hmrU standards. Chloroform extraction, drying, and off-line HPLC separation were also performed for 5-hmrC and 5-hmrU enrichment.

Enrichment of 5-hmrC and 5-hmrU was performed using a Beckman HPLC system with pump module 125 and a UV detector (module 126) connected with a 150 x 4.6 mm Kinetex C18 column (5 μm in particle size, 100 Å pore size, Phenomenex) for ribonucleoside separation. Mobile phases A and B were 10 mM ammonium formate (pH 8.0) and methanol, respectively. Ribonucleosides were separated using the following gradient: 42 min 0-0.5% B and 27 min 0.5-8% B, and the flow rate was 0.35 mL/min. A representative HPLC trace is shown in **Figure 4.2B**. 5-hmrC was collected in the time

window of 12-14 min, while 5-hmrU was collected from 16-21 min. All fractions were dried by Speed-vac and reconstituted in doubly distilled water for LC-MS/MS analysis.

Nano-flow liquid chromatography tandem mass spectrometry (nLC-MS/MS and MS/MS/MS)

5-hmrU and 5-hmrC were analyzed using a Q-Exactive plus quadrupole-orbitrap mass spectrometer coupled with an Ultimate 3000 (Thermo Fisher Scientific, San Jose, CA). The mass spectrometer was operated in the positive-ion mode with voltages of 1.8 kV and 50 V for electrospray and tube lens, respectively. The ion transport tube temperature was maintained at 275°C. PRM transitions were: 5-hmrC (*m/z* 274 to 142.061), 5-[1,3-$^{15}N_2$]5-hmrC (*m/z* 276 to 144.056), hmrU (*m/z* 275 to 125.035), and bl (*m/z* 277 to 127.029). Normalized collision energy (NCE) was set at 42%. 5-hmrC and 5-hmrU HPLC fractions were loaded on a 3.5-cm long in-housed packed porous graphitic carbon (PGC, 5 μm particle size, Thermo Fisher Scientific) column and resolved on a 20 cm Zorbax SB-C18 (5 μm in particle size, 100 Å in pore size, Agilent) column at a flow rate of 2 μL/min for loading and 300 nL/min for elution. Mobile phases A and B were 0.1% (v/v) formic acid in water and 0.1% (v/v) formic acid in 80% acetonitrile in water, respectively. Analytes were eluted using a gradient of 0-60% B in 20 min. The concentrations of 5-hmrC and 5-hmrU were determined using calibration curves. The calibration curve for 5-hmrU was previously published.[24] The calibration curve for 5-hmrU was constructed by analyzing mixtures containing 200 fmol of [1,3-$^{15}N_2$]5-hmrU labeled standard and 0.254 – 63.1 fmol of unlabeled 5-hmrU standard (**Figure 4.3**).

5-methyluridine analyses were performed on an LTQ-XL linear ion trap mass spectrometer coupled with an EASY-nLC II (Thermo Fisher Scientific, San Jose, CA). The mass spectrometer electrospray, capillary, and tube lens voltages being 2.0 kV, 12 V, and 100 V, respectively. The ion transport tube temperature was maintained at 275°C. The instrument was operated in the positive-ion mode, SRM transitions were: m^5U (*m/z* 259 to 127), 5-methyl-$^{13}C_5$-uridine (*m/z* 264 to 127), uridine (*m/z* 245 to 113), and ^{15}N-uridine (*m/z* 247 to 116). The columns described above were also used for this analysis. Nucleoside samples (2.5 ng each) were loaded at a flow rate of 2.5 µL/min and eluted with a gradient of 0−16% B in 5 min, 16−22% B in 25 min, and 22-41% B in 10 min at a flow rate of 300 nL/min. The uridine (rU) calibration curve was constructed by analyzing mixtures containing using 2800 fmol of labeled ^{15}N-uridine standard and 185 - 2200 fmol of unlabeled rU standard (**Figure 4.3**). The m^5U calibration curve was constructed by analyzing mixtures containing 5.6 fmol of 5-methyl-$^{13}C_5$-uridine standard and 0.305 – 61.6 fmol of unlabeled m^5U standard.

A summary of the LC-MS transitions and parameters can be found in **Table 4.1**

Results

Quantifications of m^5U and 5-hmrU

Global screening analysis of modified ribonucleoside using tandem mass spectrometry has previously identified 5-methyluridine (m^5U) as a prevalent modification in transfer RNA (tRNA) and ribosomal RNA (rRNA) of micro-organisms.[1, 16, 17, 19, 20] However, these investigations focused on reporting sequence context but have not reported

	Transitions		Isolation Width		NCE	
	MS^2	MS^3	MS^2	MS^3	MS^2	MS^3
rU	245 → 113	113 → 96	3	2	43	50
^{15}N-rU	247 → 115	115 → 97	3	2	43	50
m^5U	259 → 127	127 →110	3	2	43	50
$^{13}C_5$-m^5U	264 →127	127 → 110	3	2	43	50
5-hmrU	275 → 125.035		2		42	
[1,3-$^{15}N_2$] 5-hmrU	275 → 127.029		2		42	
5-hmrC	274 →142.061		2		42	
[1,3-15N2]5-hmrC	276 → 144.056		2		42	

Table 4.1: A list of precursor ions for ribonucleosides monitored in MS^2 or MS^3, isolation width, and normalized collision energy (NCE). rU and m5U are monitored on the LTQ-XL using MS^3, while 5-hmrU is monitored exclusively on the Q-Exactive plus using MS^2.

a global quantification level. We developed a method to reliably quantify m^5U levels in RNA. Using MS^3 on a linear ion-trap mass spectrometer, we measured m^5U level using a stable isotope-dilution method. The labeled standard, 5-methyl-$^{13}C_5$-uridine, had similar retention time and MS^3 spectra as its unlabeled m^5U standard (**Figure 4.1**). In the MS^3 spectra we observed a characteristic loss of H_2O and NH_3, respectively, in labeled and unlabeled standards (**Figure 4.1**). Calibration curves were generated for an accurate quantification (**Figure 4.3**).

Second, we assessed the conditions necessary to identify and quantify 5-hydroxymethyluridine (5-hmrU) in RNA. To the best of our knowledge there is no current evidence to suggest 5-hmrU is present in RNA of any type, potentially due to its low abundance. In addition, global modified ribonucleoside screening by LC-MS/MS analysis may have overlooked or been unable to detect 5-hmrU due to its inherent poor ionization efficiency. As with m^5U, we employed the use of stable isotope-dilution method to

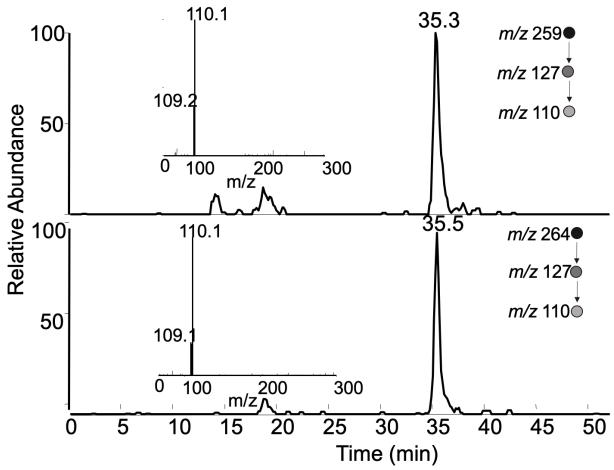


Figure 4.1: MS^3 spectra and SIC of 5-methyuridine (top) and $^{13}C_5$-5-methyluridine (bottom) on LTQ-XL.

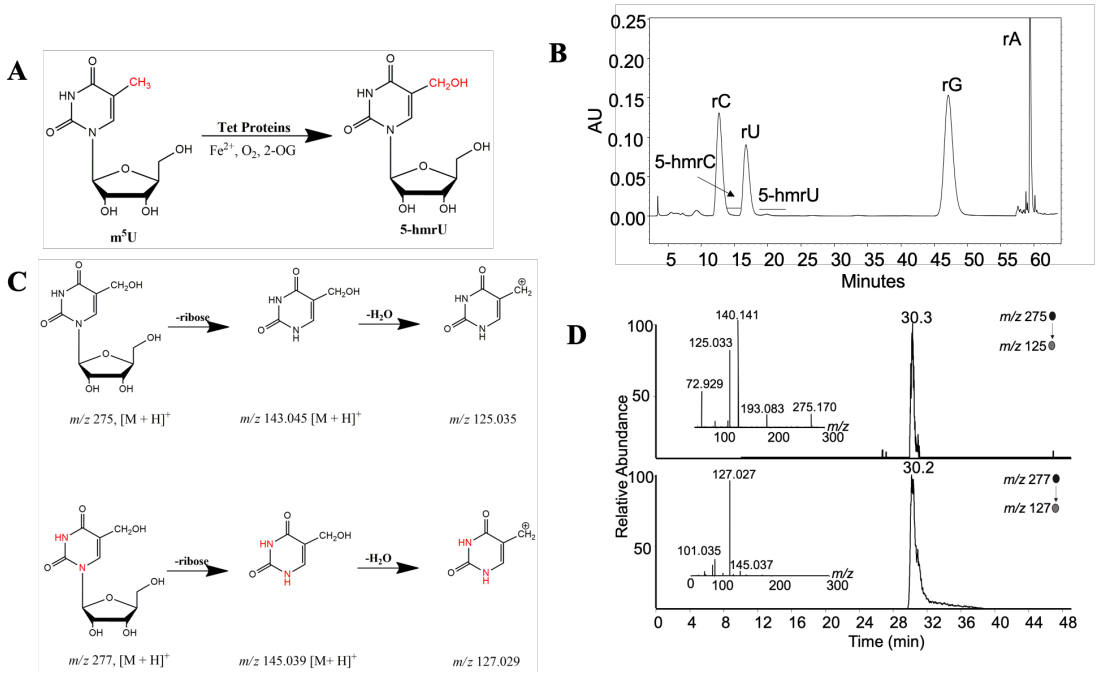


Figure 4.2 LC-MS/MS for the detection and quantification of 5-hmrU. A) Depiction of Tet-mediated formation of 5-hmrU from m^{5}U. B) Representative HPLC trace and elution time of 5-hmrC and 5-hmrU. C) schematic diagrams showing the proposed fragmentation pathways of the $[M + H]^+$ ions of 5-hmrU and [1,3-$^{15}N_2$]-5-hmrU. ^{15}N labeling is denotes in red. D) Representative selected ion chromatogram and MS/MS for 5-hmrU and labeled [1,3-$^{15}N_2$]-5-hmrU.

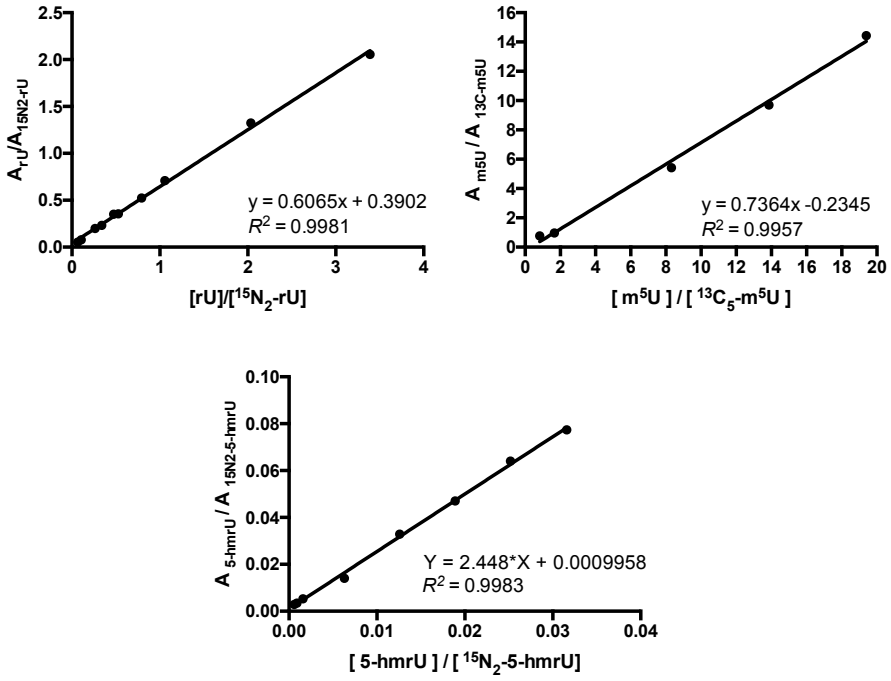


Figure 4.3: Calibration curves for uridine (rU), 5-methyluridine (m^5U), and 5-hmrU. Uridine (rU) calibration curve was constructed using 2800 fmol of labeled ^{15}N-uridine standard and 185 - 2200 fmol of unlabeled rU standard. 5.6 fmol of $^{13}C_5$-5-methyluridine (m^5U) isotope labeled standard with unlabeled m^5U standard ranged from 0.305 – 61.6 fmol was used to construct m^5U calibration curve. For 5-hmrU analysis, 200 fmol of [1,3-$^{15}N_2$]5-hmrU internal standard and unlabeled 5-hmrU standard ranged from 0.254 – 63.1 fmol was used.

accurately identify and quantify 5-hmrU, while ensuring that the correct time frame was employed during off-line HPLC enrichment of this modified ribonucleoside. Based on our off-line HPLC trace, both 5-hmrC and 5-hmrU elute similar to their unmodified nucleoside counterparts (**Figure 4.2B**). The retention time of 5-hmrC aligns with the previously published HPLC trace.[24] In addition, the HPLC trace obtained for 5-hydroxymethyl-2'-deoxyuridine (5-hmdU) has a similar retention time to its ribonucleoside counterpart.[29]

The protonated ions of the unlabeled and labeled 5-hmrU undergo losses of both the ribose and water upon collisional activation to give fragment ions of *m/z* 125.0346 and 127.029, respectively (**Figure 4.2C and 4.2D**). The isotopically labeled 5-hmrU standard has identical elution time and MS^2 to the unlabeled analyte, allowing for unambiguous identification and accurate quantification when combined with a 5-hmrU calibration curve (**Figure 4.2D and Figure 4.3**).

Identification of 5-hmrU in RNA samples isolated from mammalian tissues and cultured human cells

We employed the above-described method to measure the levels of 5-hmrU in the total RNA samples isolated from several types of mouse tissues. Our results showed that the levels of 5- hmrU were 2.92, 1.37, and 1.92 per 10^8 ribonucleosides in mouse pancreas, spleen and brain, respectively (**Figure 4.4A**). The 5-hmrU level in pancreas is significantly higher than spleen, suggesting a tissue-type dependence. This parallels previous observations made for 5-hmrC, which also demonstrated that the levels of the modified ribonucleoside varied with tissue types, though the levels of 5-hmrC in heart and brain were higher than the other types of tissues analyzed.[24] Furthermore, the 5-hmrC level in pancreas was the lowest compared to the other tissues[24]; however our analysis revealed the pancreas exhibits higher level of 5-hmrU compared to pancreas and spleen.

Corresponding analyses were performed for m^5U of the mouse tissues. The m^5U level was the highest in mouse brain compared to the other tissue types, where 0.98, 0.86, and 0.75 modifications per 100 uridines were detected in total RNA samples of mouse brain, spleen, and pancreas, respectively (**Figure 4.5B**). This result differs from 5-hmrC

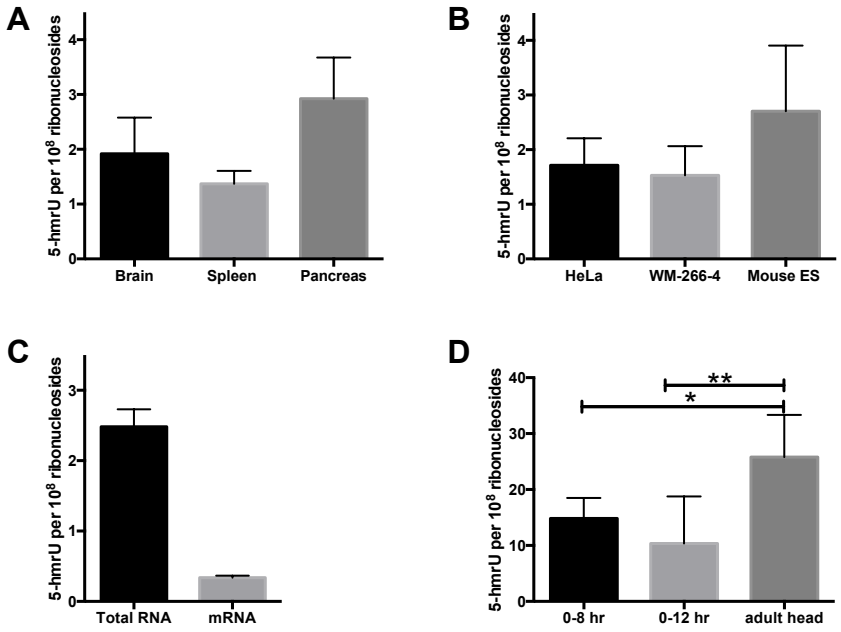


Figure 4.4: 5-hmrU is present in RNA of mammalian cells in a cell- and RNA type-dependent manner. A) 5-hmrU levels in mouse tissues. (n = 3) B) 5-hmrU level of cultured cancer cell lines, HeLa and WM-266-4 (n = 3). C) 5-hmrU level in total RNA and mRNA of HEK293T cells (n = 3). D) The levels of 5-hmrU in total RNA of Drosophila larval at different developmental stages (6 technical replicates). p-values were obtained using unpaired student's t test.

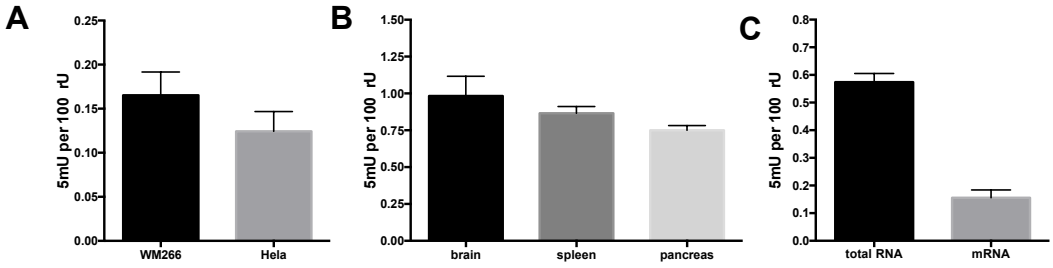


Figure 4.5 m^5U level in mammalian cells m^5U level in A) human cultured cell lines, WM266 and HeLa (n = 3) and B) mouse tissues (n = 3). RNA type specificity in total RNA and mRNA of HEK293T cells (n = 3).

and m^5C where tissues displaying higher level of m^5C correlated with higher level of 5-hmrC.[24] Interestingly, the level of m^5U was the lowest in the mouse pancreas; however, 5-hmrU was highest in pancreas. Overall, our results suggest 5-hmrU level is tissue type-dependent. (**Figure 4.4A and 4.5B**).

Next we evaluated if 5-hmrU was detectable in cultured mouse embryonic stem cells and human cancer cell lines, including HeLa cervical cancer and WM-266-4 melanoma cells. We observed 5-hmrU at levels of 2.7, 1.7, and 1.5 per 10^8 ribonucleosides, respectively (**Figure 4.4B**). Our results suggest that 5-hmrU is elevated in mouse ES cells compared to the human cultured cancer cell lines (**Figure 4.4A and 4.4B**). We then identified the corresponding m^5U levels. Mouse ES cells had the highest amount of m^5U compared to the cultured human cell lines. This correlates well with the 5-hmrU data, where mouse ES cells also have higher level of m^5U and 5-hmrU (**Figure 4.4B and Figure 4.5A**).

We also sought to determine if 5-hmrU may be RNA-type specific. We isolated poly(A) mRNA from HEK293T cells and measured the basal total RNA 5-hmrU level. The results indicated a lower level of 5-hmrU in mRNA than total RNA of HEK293T cells (**Figure 4.4C**). Moreover, mRNA samples also contained lower levels of m^5U compared to the total RNA from HEK293T cells (**Figure 4.5C**). Taken together, we conclude that 5-hmrU is present in total RNA and mRNA of mammalian cells.

Identification of 5-hmrU in *Drosophila melanogaster* total RNA

Our results have demonstrated that 5-hmrU is present in mammalian systems. We wanted to evaluate if 5-hmrU is present in *Drosophila melanogaster*. We analyzed the level

of 5-hmrU in the developing drosophila larva at 0-8 hours, 8-12 hours, and in adult head. The results revealed that 5-hmrU is present at the various developmental stages at levels of 14.8, 10.4, and 25.8 modifications per 10^8 ribonucleosides, respectively (**Figure 4.4D**). The overall level of 5-hmrU is approximately 10 times higher in Drosophila than in mammalian systems, indicating not only is 5-hmrU a stable modification in Drosophila RNA but it potentially has a gene regulatory function (**Figure 4**).

Ten-eleven translocation (Tet) proteins demethylate m^5U to 5-hmrU

Having demonstrated the presence of 5-hmrU in mammalian RNA, we wanted to determine if Tet proteins are responsible for oxidizing m^5U to 5-hmrU. Employing a sensitive LC-MS/MS method, we evaluated Tet1-3's ability to oxidize m^5U in total RNA. Upon overexpression of the catalytic domains of Tet1-Tet3, we observed a statistically significant decreases in global m^5U levels using in HEK293T cells (**Figure 4.6A**). The results indicate that Tet proteins may play a pivotal role in m^5U regulation. Interestingly, previous reports demonstrated that Tet overexpression did not influence global m^5C levels.[24] Therefore, Tet protein may play a more influential role on m^5U regulation than m^5C.

To corroborate the m^5U results, we assessed 5-hmrU in cells overexpressing the catalytic domains of Tet1-3. We identified a statistically significant increase in 5-hmrU level in cells overexpressing Tet1 and Tet3, but not Tet2 (**Figure 4.6B**). In addition, the levels of 5-hmrC are elevated upon overexpression of the catalytic domains of Tet1-3 (**Figure 4.7**).[24] Our results indicate that Tet1 and Tet3 proteins' catalytic domain has the capability to oxidatively demethylate m^5U to 5-hmrU (**Figure 4.6A and 4.6B**). The

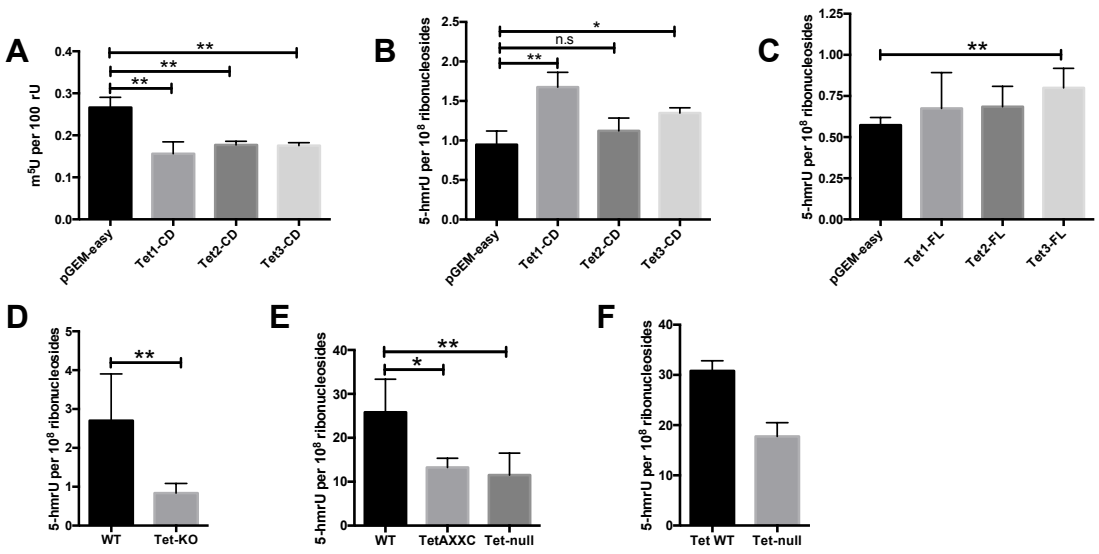


Figure 4.6: Tet protein mediated m^5U demethylation to 5-hmrU. A) HEK293T cells overexpression catalytical domains of Tet1-3. pGEM-easy refers to transfection using control pGEM-T easy plasmid. Panels A) and B) depict the levels of m^5U and 5-hmrU upon overexpression of catalytic domains of Tet1-3 in HEK293T cells. The data represent the mean and standard deviation of results obtained from three independent transfections. C) The levels of 5-hmrU in HEK293T cells upon ectopic expression of full length Tet1-3 (n = 3). D-E) The levels of 5-hmrU in wild-type mouse ES cells and Tet knock-out mouse ES cells. Levels of 5-hmrU in E) total RNA and F) mRNA of *Drosophila Melanogaster* with Tet protein deletion. The data in E) represents total RNA of 4 technical replicates of each treatments, while the preliminary data in F) represents mRNA of WT (n = 2) and Tet-null (n = 3). All p-values were obtained using unpaired student's t test.

possibility arose that the other regions of Tet protein may alter its binding affinity and activity. We next overexpressed the full-length protein in HEK293T cells. The results indicate that only full-length Tet3 protein can produce 5-hmrU (**Figure 4.6C**). Tet3 protein is primarily found in the cytosol[30] and since the analysis was performed on total RNA which primarily consists of ribosomal RNA (rRNA), the results are in line with Tet3's ability to mediate the formation of 5-hmrU levels in rRNA in the cytosol.

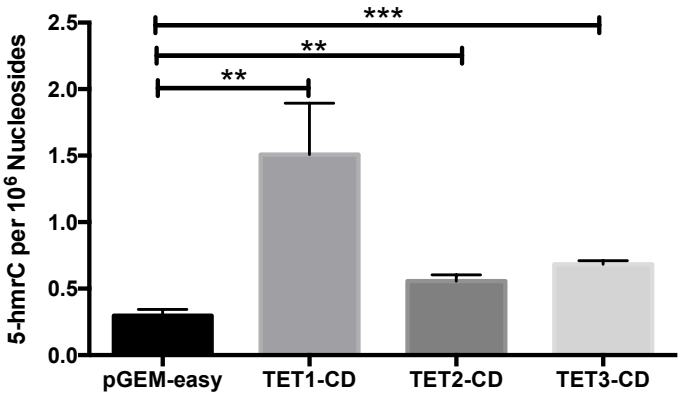


Figure 4.7 5-hmrC level in HEK293T cells overexpressing catalytic domain of Tet1-3. The data represent the mean and standard deviation of three independent transfections. All p-values were obtained using unpaired student's t test.

We also examined how the loss of function of Tet proteins in *Drosophila melanogaster* affects the level of 5-hmrU. Drosophila is a model organism to evaluate Tet activity in RNA. In mammalian cells, Tet proteins is most commonly associated with the ability to demethylate 5-methyl-2'-deoxycytidine (5-mdC).[25, 31] However, there is nearly no detectable 5-mdC in Drosophila [32], suggesting that 5-mdC is not Tet's primary substrate in Drosophila.[33] Therefore, we sought to examine if the 5-hmrU measured in Drosophila RNA is Tet-mediated. Since the Tet protein is highly expressed in drosophila's central nervous system[34], we quantified the level of 5-hmrU in Drosophila head with genetic ablation of Tet protein. We observed a decrease in 5-hmrU levels in the Tet-null total RNA (**Figure 4.6E**). In addition, mutations on Tet's DNA binding domain resulted in decreased levels of 5-hmrU **(Figure 4.6E**). Moreover, preliminary data from Tet-null Drosophila exhibited a decreased level of 5-hmrU in brain mRNA (**Figure 4.6F**). These results indicate that m^5U is an important substrate of Tet proteins in Drosophila. As a whole, these results suggest that Tet mediates demethylation of m^5U to 5-hmrU in total RNA and mRNA.

Conclusions

In this study, we developed a method to identify a previously unrecognized modified ribonucleoside, 5-hmrU, in mammalian cells, as well as in mouse and Drosophila tissues. We were not only able to identify 5-hmrU in total RNA but also in the mRNA of HEK23T cells. Its presence in RNA suggests that 5-hmrU may play a role in epitranscriptomic regulation or potentially as a transition intermediate toward demethylation. However, the analysis also suggests the existence of other modifications beyond N^6-methyladenosine (m^6A), N^6,2'-*O*-dimethyladenosine (m^6A_m), 5-methylcytosine (m^5C), m^5U and 2'-*O* nucleosides (N_m) in mRNA. Therefore, more thorough investigations are necessary to evaluate the potential for other modifications in mRNA and their respective readers, writers, and erasers.

Our results corroborate those from Pfaffeneder, et.al, where Tet enzyme displays the ability to oxidize substrates besides cytidine and its respective oxidation products.[21] Their findings demonstrated that Tet proteins can also oxidize thymine in DNA to yield 5-hmdU.[21] Our findings suggest Tet has an additional substrate of interest, m^5U, which is used to catalyze the formation of 5-hmrU in high order organisms *in vivo*. However, additional experiments are necessary to determine if Tet proteins can catalyze the conversion of 5-hmrU to 5-formyluridine and 5-carboxyuridine. Overall, Tet's ability to oxidize DNA and RNA opens the possibility of other members of 2-oxoglutarate dependent dioxygenases to function beyond their known targets.

Reference

1. Su, D.; Chan, C.; Gu, C.; Lim, K.; Chionh, Y.; McBee, M.; Russell, B.; Babu, I.; Begley, T.; Dedon, P., Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nat. Prot.* **2014,** *9* (4), 828-841.

2. Yan, M.; Wang, Y.; Hu, Y.; Feng, Y.; Dai, C.; Wu, J.; Wu, D.; Zhang, F.; Zhai, Q., Modi fi cations in Mouse Liver and Their Correlation with Diabetes. **2013**.

3. Roundtree, I. A.; Evans, M. E.; Pan, T.; Chuan, H., Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **2017,** *169*, 1187-1187.

4. Basanta-Sanchez, M.; Temple, S.; Ansari, S. A.; D'Amico, A.; Agris, P. F., Attomole quantification and global profile of RNA modifications: Epitranscriptome of human neural stem cells. *Nucleic Acids Res.* **2015**, gkv971-gkv971.

5. Bai, L.; Tang, Q.; Zou, Z.; Meng, P.; Tu, B.; Xia, Y.; Cheng, S.; Zhang, L.; Yang, K.; Mu, S.; Wang, X.; Qin, X.; Lv, B.; Cao, X.; Qin, Q.; Jiang, X.; Chen, C., m6A Demethylase FTO Regulates Dopaminergic Neurotransmission Deficits Caused by Arsenite. *Toxicol Sci* **2018,** *165* (2), 431-446.

6. Chen, W.; Zhang, L.; Zheng, G.; Fu, Y.; Ji, Q.; Liu, F.; Chen, H.; He, C., Crystal structure of the RNA demethylase ALKBH5 from zebrafish. *FEBS Letters* **2014,** *588* (6), 892-898.

7. Dai, X.; Gonzalez, G.; Li, L.; Li, J.; You, C.; Miao, W.; Hu, J.; Fu, L.; Zhao, Y.; Li, R.; Li, L.; Chen, X.; Xu, Y.; Gu, W.; Wang, Y., YTHDF2 Binds to 5-Methylcytosine in RNA and Modulates the Maturation of Ribosomal RNA. *Analytical Chemistry* **2019,** *92*, 1346-1354.

8. Fu, Y.; Jia, G.; Pang, X.; Wang, R. N.; Wang, X.; Li, C. J.; Smemo, S.; Dai, Q.; Bailey, K. A.; Nobrega, M. A.; Han, K.-L.; Cui, Q.; He, C., FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nature Communications* **2013,** *4*, 1798-1798.

9. Fu, L.; Amato, N. J.; Wang, P.; McGowan, S. J.; Niedernhofer, L. J.; Wang, Y., Simultaneous Quantification of Methylated Cytidine and Adenosine in Cellular and Tissue RNA by Nano-Flow Liquid Chromatography–Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal. Chem.* **2015,** *87ha* (15), 7653-7659.

10. Li, Q.; Li, X.; Tang, H.; Jiang, B.; Dou, Y.; Gorospe, M.; Wang, W., NSUN2-Mediated m5C Methylation and METTL3/ METTL14-Mediated m6A Methylation Cooperatively Enhance p21 Translation. **2017,** *2598* (January), 2587-2598.

11. Yang, X.; Yang, Y.; Sun, B.-F.; Chen, Y.-S.; Xu, J.-W.; Lai, W.-Y.; Li, A.; Wang, X.; Bhattarai, D. P.; Xiao, W.; Sun, H.-Y.; Zhu, Q.; Ma, H.-L.; Adhikari, S.; Sun, M.; Hao, Y.-J.; Zhang, B.; Huang, C.-M.; Huang, N.; Jiang, G.-B.; Zhao, Y.-L.; Wang, H.-L.; Sun, Y.-P.; Yang, Y.-G., 5-methylcytosine promotes mRNA export — NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell Research* **2017,** *27* (5), 606-625.

12. Goll, M. G.; Kirpekar, F.; Maggert, K. a.; Yoder, J. a.; Hsieh, C.-L.; Zhang, X.; Golic, K. G.; Jacobsen, S. E.; Bestor, T. H., Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Sci.* **2006,** *311* (5759), 395-398.

13. Dai, D.; Wang, H.; Zhu, L.; Jin, H.; Wang, X., N6-methyladenosine links RNA metabolism to cancer progression. *Cell Death Dis* **2018,** *9* (2), 124.

14. Li, N.; Hui, H.; Bray, B.; Gonzalez, G. M.; Zeller, M.; Anderson, K. G.; Knight, R.; Smith, D.; Wang, Y.; Carlin, A. F.; Rana, T. M., METTL3 regulates viral m6A RNA modification and host cell innate immune responses during SARS- CoV-2 infection. *Cell Reports* **2021,** (35), 109091.

15. Lichinchi, G.; Zhao, Boxuan S.; Wu, Y.; Lu, Z.; Qin, Y.; He, C.; Rana, Tariq M., Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. 2016; Vol. 20, pp 666-673.

16. Bjã, G. R.; Neidhardt, F. C., Analysis of 5-Methyluridine Function in the Transfer RNA of Escherichia coli1. *Cancer Research* **1971,** *31* (May), 706-709.

17. Rk, G. R. B. J.; Neidhardt, F. C., Physiological and Biochemical Studies on the Function of 5-Methyluridine in the Transfer Ribonucleic Acid of Escherichia coli. *Journal of Bacteriology* **1975,** *124* (1), 99-111.

18. Urbonavičius, J.; Auxilien, S.; Walbott, H.; Trachana, K.; Golinelli-Pimpaneau, B.; Brochier-Armanet, C.; Grosjean, H., Acquisition of a bacterial RumA-type tRNA(uracil-54, C5)-methyltransferase by Archaea through an ancient horizontal gene transfer. *Molecular Microbiology* **2008,** *67* (2), 323-335.

19. Alian, A.; Lee, T. T.; Griner, S. L.; Stroud, R. M.; Finer-Moore, J., Structure of a TrmA-RNA complex: A consensus RNA fold contributes to substrate selectivity and catalysis in m5U methyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105* (19), 6876-6881.

20. Madsen, C. T.; Mengel-Jørgensen, J.; Kirpekar, F.; Douthwaite, S., Identifying the methyltransferases for m5U747 and m5U1939 in 23S rRNA using MALDI mass spectrometry. *Identifying the methyltransferases for m5U747 and m5U1939 in 23S rRNA using MALDI mass spectrometry* **2003,** *31* (16), 4738-4746.

21. Pfaffeneder, T.; Spada, F.; Wagner, M.; Brandmayr, C.; Laube, S. K.; Eisen, D.; Truss, M.; Steinbacher, J.; Hackner, B.; Kotljarova, O.; Schuermann, D.; Michalakis, S.; Kosmatchev, O.; Schiesser, S.; Steigenberger, B.; Raddaoui, N.; Kashiwazaki, G.; Müller, U.; Spruijt, C. G.; Vermeulen, M.; Leonhardt, H.; Schär, P.; Müller, M.; Carell, T., Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nature Chemical Biology* **2014,** *10* (7), 574-581.

22. Hausinger, R. P., *2-Oxoglutarate-Dependent Oxygenases*. 2015; p 1-58.

23. Basanta-Sanchez, M.; Wang, R.; Liu, Z.; Ye, X.; Li, M.; Shi, X.; Agris, P.; Zhou, Y.; Huang, Y.; Sheng, J., TET1-mediated Oxidation of 5-formylcytosine (5fC) to 5-carboxycytosine (5caC) in RNA. *ChemBioChem* **2016**, 1-6.

24. Fu, L.; Guerrero, C. R.; Zhong, N.; Amato, N. J.; Liu, Y.; Liu, S.; Cai, Q.; Ji, D.; Jin, S.-G.; Niedernhofer, L. J.; Pfeifer, G. P.; Xu, G.-L.; Wang, Y., Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **2014,** *136* (33), 11582-5.

25. Ichiyama, K.; Chen, T.; Wang, X.; Yan, X.; Kim, B.-S.; Tanaka, S.; Ndiaye-Lobry, D.; Deng, Y.; Zou, Y.; Zheng, P.; Tian, Q.; Aifantis, I.; Wei, L.; Dong, C., The methylcytosine dioxygenase Tet2 promotes DNA demethylation and activation of cytokine gene expression in T cells. *Immunity* **2015,** *42* (4), 613-26.

26. Guo, J. U.; Su, Y.; Zhong, C.; Ming, G. L.; Song, H., Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **2011,** *145* (3), 423-434
.

27. Ko, M.; Huang, Y.; Jankowska, A. M.; Pape, U. J.; Tahiliani, M.; Bandukwala, H. S.; An, J.; Lamperti, E. D.; Koh, K. P.; Ganetzky, R.; Liu, X. S.; Aravind, L.; Agarwal, S.; MacIejewski, J. P.; Rao, A., Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **2010,** *468* (7325), 839-843.

28. Hu, X.; Zhang, L.; Mao, S. Q.; Li, Z.; Chen, J.; Zhang, R. R.; Wu, H. P.; Gao, J.; Guo, F.; Liu, W.; Xu, G. F.; Dai, H. Q.; Shi, Y. G.; Li, X.; Hu, B.; Tang, F.; Pei, D.; Xu, G. L., Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* **2014,** *14* (4), 512-522.

29. Liu, S.; Dunwell, T. L.; Pfeifer, G. P.; Dunwell, J. M.; Ullah, I.; Wang, Y., Detection of oxidation products of 5-methyl-2′-deoxycytidine in arabidopsis DNA. *PloS one* **2013,** *8* (12), 8-13.

30. Arioka, Y.; Watanabe, A.; Saito, K.; Yamada, Y., Activation-Induced Cytidine Deaminase Alters the Subcellular Localization of Tet Family Proteins. *PLoS ONE* **2012,** *7* (9), e45031.

31. Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. a.; He, C.; Zhang, Y., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Sci.* **2011,** *333* (6047), 1300-1303.

32. Frank Lyko, B. H. R. R. J., DNA methylation in Drosophila melanogaster. *Nat.* **2000,** *408*, 538-540.

33. Delatte, B.; Wang, F.; Ngoc, L. V.; Collignon, E.; Bonvin, E.; Deplus, R.; Calonne, E.; Hassabi, B.; Putmans, P.; Awe, S.; Wetzel, C.; Kreher, J.; Soin, R.; Creppe, C.; Limbach, P. A.; Gueydan, C.; Kruys, V.; Brehm, A.; Minakhina, S.; Defrance, M.; Steward, R.; Fuks, F., Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Sci.* **2016,** *351* (6270), 282-285.

34. Wang, F.; Minakhina, S.; Tran, H.; Changela, N.; Kramer, J.; Steward, R., Tet protein function during Drosophila development. *PLoS ONE* **2018,** *13* (1), 1-20.

Chapter 5: Associations of smoking and air pollution with peripheral blood RNA N^6-methyladenosine

Introduction

Similar to epigenetic modifications of DNA that impact gene expression, chemical modifications of RNA play a critical role in post- transcriptional regulation. The epitranscriptome is composed of over 100 different types of modifications on messenger RNAs (mRNAs), ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) and small RNAs. The most abundant internal mRNA modification is N^6-methyladenosine (m^6A), with approximately one to three modifications per transcript.[1] However, m^6A modifications are also abundant on rRNA, microRNA and long non-coding RNAs.[2] Modified ribonucleosides play important roles in mRNA translation, stability, export, and splicing, suggesting their wide-reaching implications for cellular functions.[1] m^6A is regulated by proteins that interpret, add, and remove nucleic acid modifications, termed readers, writers, and erasers (RWEs). The N^6 position of adenosine in mRNA are methylated co-transcriptionally in the nucleus by a methyltransferase complex, which includes methyl-transferase-like protein 3 (METTL3), METTL14, and Wilms' tumor 1-associating protein (WTAP) (m^6A writers).[1] m^6A undergoes clear active demethylation catalyzed by Fat mass and obesity-associated protein (FTO) and AlkB homolog 5 (ALKBH5) (m^6A erasers). Furthermore, m^6A reader proteins provide an additional layer of transcriptome regulation, such as YTH domain-containing family protein 2 (YTHDF2), which promotes the decay of m^6A-containing mRNAs.[3] Through its role in RNA metabolism, m^6A is a key player in

cell fate, with research demonstrating the importance of m^6A in cell development and differentiation.[4-7] Additionally, growing evidence implicates m^6A and its RWEs in many facets of human health and disease. Altered m^6A has been linked to cancer progression,[8] immunity and viral infection,[9, 10] obesity and adipogenesis,[11-13] diabetes,[14, 15] and male infertility.[16, 17]

Most importantly for the field of environmental health, early research suggests that m^6A is responsive to cellular stress conditions and may play a functional role in stress response. Increased m^6A content in mRNA molecules from oxidative stress leads to accumulation of mRNA triage in stress granules,[18] while YTHDF2 protects m^6A from demethylation in stress-induced mRNA transcripts, promoting selective translation.[19] Recently, research demonstrated that *in vitro* exposure of A549 human lung cancer cells to several environmental toxicants induced dose-dependent decreases in global m^6A levels.[20] Furthermore, parental diet during pregnancy was associated with m^6A modifications in the hypothalamus of offspring[11] and arsenite exposure was associated with elevated m^6A and decreased FTO in mouse neurons.[21] These studies suggest that m^6A may be responsive to stress in a tissue-specific manner. However, to date, no study has examined the response of m^6A to environmental exposures in a human population. We hypothesized that global m^6A and its RWEs in human blood may serve as a mechanism and biomarker of exposure to environmental stressors, such as cigarette smoke and air pollution. Particulate air pollutants and cigarette smoking have profound health impacts, leading to mortality, cardiovascular diseases, cancers, and declines in lung function.[22-24] Furthermore, these toxicants regulate epigenetics via induction of oxidative stress.[25, 26] Thus, our objective was

to conduct the first human population study to determine the associations between total RNA m^6A and its RWEs with smoking and air pollution exposures in a population of highly exposed adults from Beijing, China. We further aimed to investigate associations between gene expression of RWEs and total m^6A levels in blood in a healthy population, which has been previously unexplored. To accomplish these goals, we measured, using a sensitive mass-spectrometry approach, global m^6A levels in total RNA, and examined gene expression of six RWEs in peripheral blood from 106 truck drivers and office workers at the end of an eight-hour workday. We then examined the correlations between m^6A and RWEs, and their associations with indicators of smoking status, personal PM2.5 and black carbon (BC), and ambient PM10 measured throughout the workday.

Experimental Section

Study population

We leveraged existing samples and data from the Beijing Truck Driver Air Pollution Study, which recruited 60 truck drivers and 60 indoor office workers from June 15th to July 27th, 2008, with the goal of examining the health effects of personal air pollutant exposures.[27] All study participants had been on their current jobs for great than two years, had no significant medical illness interfering with their ability to work, and were not pregnant. Participants were matched by age (within 5 years), sex, smoking status, and education, and assessed on two independent workdays at a 1–2-week interval. In the present study, we only included data and samples from the first study visit with blood samples collected at the end of the study day. In-person interviews were conducted to

determine demographic and lifestyle characteristics. Self-administered questionnaires were collected at the end of the study day to determine any relevant exposures on that day (alcohol or tea consumption, cigarette smoking, traffic, etc.). All participants provided individual written informed consent before the start of the study and institutional review board approval was obtained from all participating institutions prior to recruitment.

Ambient and personal air pollutant exposures

Ambient PM_{10} exposure data during the study period were computed from 27 monitoring stations from the Beijing Municipal Environmental Bureau to estimate daily average PM_{10} levels in Beijing.[28] To assess personal $PM_{2.5}$ and BC exposures, study participants wore gravimetric samplers during the eight hour work day as previously described.[28] The sampler was carried in a belt pack with the inlet clipped near the breathing zone. Each sampler setup included an Apex pump (Casella Inc., Bed- ford, UK), a Triplex Sharp-Cut Cyclone separator (BGI Inc., Waltham, MA), and a 37-mm Teflon filter placed on top of a drain disc and inside a metal filter holder. Filters were stored in atmosphere-controlled conditions before and after sampling and were weighed with a microbalance (Mettler-Toledo Inc., Columbus, OH). We calculated a time-weighted average of $PM_{2.5}$ concentration by dividing the change in filter weight before and after sampling by the volume of air sampled. Technical validation tests of personal $PM_{2.5}$ measurement found high reproducibility ($r = 0.944$) of results between samplers.[28] The blackness of the filters used to measure PM2.5 was assessed using an EEL Model M43D Smoke Stain Reflectometer (Diffusion Systems Ltd., London, UK), by applying standard black-smoke index calculations of the absorption coefficients based on reflectance.[29] We assumed a

factor of 1.0 for converting the absorption coefficient to BC mass,[30, 31] which was then divided by the sampled air volume to calculate average BC exposure concentration.[29] We obtained daily outdoor average temperature and dew point data for Beijing from the National Oceanic and Atmospheric Administration (NOAA) online database.

Smoking status exposures

Information on smoking status was collected from in-person interviews. Information on the number of cigarettes smoked on the test day and the exposure to environmental tobacco smoke (ETS) was collected by self-administered questionnaires at the end of the test day. Smokers were categorized into ever/never smokers and current/former smokers. Pack years was divided into three categories: none, <3.8 pack years, and >3.8 pack years (3.8 pack years was the median). As approximately half of the current smokers smoked on the day of testing, we created a binary variable called "smoked while wearing the monitor" to indicate acute tobacco smoke exposure and serve as a covariate for the air pollutant models. As there was only one self-identified female smoker, analyses on smoking status, pack years, and smoking while wearing the monitor, were conducted only in men.

Sample collection and RNA isolation

Peripheral blood samples from the first study visit were collected into PAXgene Blood RNA Tubes (Qiagen, Germantown, MD) at each post-work exam (between 4:00 pm and 6:00 pm hours). Total RNA was extracted with the PAXgene Blood-RNA Kit Qiagen-763134 (Qiagen, Germantown, MD) and RNA quality and quantity were confirmed with an Implen spectrophotometer (Implen, Westlake Village, CA). There were 108 samples

from this visit with enough RNA yield for global m^6A analysis and 101 of the original 108 RNA samples with enough yield for gene expression analysis. Samples were stored at −80 °C until use.

Quantitation of global m^6A from total RNA

We digested 100 ng of total RNA with 1 unit of nuclease P1 (Sigma-Aldrich, St. Louis, MO) in a buffer containing 25 mM NaCl and 2.5 mM $ZnCl_2$. The mixture was then incubated at 37 °C for 2 hrs., followed by the addition of 2.5 units Antarctic phosphatase and Antarctic phosphatase buffer (New England Biolabs, Ipswich, MA). After incubating at 37 °C for an additional 2 hrs., the digestion mixture was dried using a Speed-vac and reconstituted in 100 μL ddH_2O. To quantify the levels of global m^6A in total RNA, we measured the levels of global m^6A and adenosine (rA) with a previously reported stable isotope-dilution method.[32] In this vein, we added 42.5 fmol of D^3-N^6- methyladenosine and 8 pmol of $^{15}N_3$-labeled adenosine as internal standards to 10 ng of digested RNA. All enzymes from the digestion mixture were removed by chloroform extraction. The aqueous layer was dried and reconstituted in ddH_2O. Approximately 25% of the nucleoside mixture was injected for nLC-MS^3 analysis. The nLC-MS^3 measurements were conducted on an LTQ-XL linear ion trap mass spectrometer equipped with a nanoelectrospray ionization source and an EASY-nLC II (Thermo Fisher Scientific, San Jose, CA). The in-house-prepared trapping column (150 μm × 50 mm) and analytical column (75 μm × 25 mm) were packed with stationary phase materials of porous graphitic carbon (5 μm in particle size, Thermo Fisher Scientific) and Zorbax SB-C18 (5 μm beads, 100 Å in pore size, Agilent), respectively. The LC separation was conducted by using mobile phases A (0.1%

formic acid in ddH_2O) and B (0.1% formic acid in acetonitrile) to elute the sample. The samples were initially loaded onto the trapping column with mobile phase A at a flow rate of 2.5 μL/min. The nucleoside mixture was resolved on the analytical column using an 80-minute gradient of 0–16% B in 5 min, 16–22% B in 23 min, 22–50% B in 17 min, 50–90% B in 5 min, and finally, 90% B in 30 min at a flow rate of 300 nL/min. The mass spectrometer was operated in the positive-ion mode. The electrospray, capillary and tube lens voltages were maintained at 2.0 kV, 12 V, and 100 V, respectively, and the ion transport tube temperature was 275 °C. The average relative standard deviation of results obtained from replicate measurements of all samples was 12%, and was less than 22% for each sample. Results are presented as percent m^6A per total adenosine (%m^6A/A).

Gene expression analysis

We performed real-time-quantitative polymerase chain reaction (RT-qPCR) to assess the mRNA expression of six m^6A RWE genes, including one reader (YTHDF2), three writers (METTL3, METTL14, WTAP) and two erasers (FTO, ALKBH5), according to MIQE (minimum information for publication of quantitative real-time PCR experiments) guidelines.[33] First, 500 ng of cDNA was synthesized with the Promega Reverse Transcription System with random primers according to the manufacturer's instructions (Promega, Madison, WI). Internal reference control genes were chosen based on previous literature that determined the best reference genes for immune cells and whole blood: TBP (TATA box-binding protein), SDHA (succinate dehydrogenase complex flavoprotein subunit A), and GAPDH (Glyceraldehyde 3-phosphate dehydrogenase).[34-36] Primers were selected from published literature: YTHDF2, WTAP, METTL14, METTL3,

and ALKBH5 were based on Yang et. al 2016,[17] FTO and GAPDH were based on Shen et. al 2015,[14] SDHA was from,[37] and TBP was from Rydbirk et al., 2016 (Integrated DNA Technologies, Coralville, IA, USA).[38] Each qPCR reaction contained 7.5 ng of cDNA in 4 µL of water, 5 µL iTaq Universal SYBR Green Supermix (BioRad, Hercules, CA), 0.5 µL of 5 µM forward primer, and 0.5 µL of 5 µM of reverse primer in a 10 µL final volume in a 384-well plate. Each qPCR reaction was carried out in triplicate and three no template controls were added to each plate. Cycling conditions were 95 °C for 5 mins, followed by 40 cycles of 95 °C for 15 sec and 57 °C (YTHDF2, WTAP, METTL14, METTL3, FTO, TBP) or 53 °C (ALKBH5, GAPDH, SDHA) for 30 sec, followed by a melting curve analysis. Each gene was analyzed on a separate plate, with all samples randomized on one plate per gene. The cutoff C_q value for not expressed was 34 and there was no detectable expression of no template controls. Melting curve analysis suggested a single product, which was confirmed with 1% agarose gel electrophoresis (data not shown). The mean (standard deviation) of the percent coefficient of variation between replicate reactions for each gene was ALKBH5: 0.33 (0.18)%, FTO: 0.36 (0.22)%, GAPDH: 0.34 (0.19)%, METLL3: 0.35 (0.24)%, METTL14: 0.39 (0.27)%, SDHA: 0.37 (0.23)%, TBP: 0.54 (0.33)%, WTAP: 0.30 (0.16)%, and YTHDF2: 0.33 (0.22)%. All samples were standardized to a 5-point internal standard curve ranging from 1.875 ng to 30 ng of cDNA input (standard curve R2: 0.98–0.99), and no samples fell outside the curve. However, house- keeping genes were associated with cigarette smoking on the day of data collection and thus associations of gene expression with cigarettes smoked while wearing the monitor were not included in the analysis. Housekeeping genes were not associated with any other

outcomes or covariates. Samples were then normalized to the geometric mean of three reference genes (SDHA, GAPDH, and TBP),[39] then natural log transformed for analysis (N = 101).

Statistical analysis

From the global m^6A analysis, one extreme outlier was removed (greater than six standard deviations from the mean) and a second sample failed lab QC (Final Nm^6A = 106 and NRWEs = 101, Differences due to RNA availability). Global m^6A and gene expression data were natural log transformed before analysis to improve model fit. Correlations between gene expression of RWEs and global m^6A levels were calculated using Pearson's correlation coefficients. We investigated potential nonlinear associations between exposures and global m^6A with generalized additive models (data not shown), and when associations were determined to be linear, we fit crude and adjusted linear models between each exposure variable and global m^6Aor RWE to determine the potential confounding effects. Since the global m^6A and RWEs were log transformed prior to analysis, effect estimates were expressed as the relative percent change of the outcome per change in the exposure. This was calculated from the antilog of the regression coefficient, with the 95% confidence interval (CI). Covariates were determined a priori and crude and adjusted models were adjusted for technical batch variables (analytical chemistry: RNA digest batch; qPCR: RNA concentration and quality). Occupation and BMI were considered as covariates, but as they had no statistical effect on the outcome ($P \leq 0.05$), they were not included in final models. P-values from models on RWEs were adjusted for multiple

comparisons using false discovery rate (FDR) and $P \leq 0.05$ was considered statistically significant.

Analysis of m^6A with smoking variables

As only three individuals reported as former smokers, the analysis was conducted on the categories of ever and never smokers. Models for smoking status, pack years, and cigarette smoking on the day of the exposure were performed only in men, as there was only one woman who reported being a smoker. To remove confounding effects from first-hand smoking, models for environmental tobacco smoke (ETS) exposure were conducted only in nonsmokers. Additionally, models examining cigarettes smoked on the day of exposure were conducted only in smokers. All smoking models were adjusted for age and RNA digest batch.

Analysis of m^6A with air pollutants

Models on air pollutants were adjusted for age, sex, smoking status, smoking while wearing the monitor, average air temperature, average dew point, and day of the week. The study was conducted within a single month in Beijing, China, and so no adjustment for season was necessary.

Sensitivity analyses

To ensure the robustness of our findings, we performed multiple sensitivity analyses for this study. Firstly, all smoking models were repeated in the full population and adjusted for sex, with models for ETS and smoking on the day of the exposure additionally co-adjusted for smoking status. Second, all air pollutant models were repeated in the subpopulation of individuals that did not smoke on the day of the exposure, to ensure that

all residual confounding from cigarette smoke was removed. A third sensitivity analysis stratified the population by occupation (indoor office workers and truck drivers). Since the original study cohort was designed to recruit high and low air pollution exposure by profession, we wanted to ensure that our results were consistent across occupations. Finally, as we only had male smokers enrolled in this study, we repeated ETS and air pollutant models stratified by sex. As the sample size of each group was low, we only performed sensitivity analyses with crude models with the exception of air pollutant models in males, which were co-adjusted for smoking and cigarettes smoked on the day of exposure. All statistical analyses were performed in the software package R.[40]

Results

Population and exposure characteristics

Our study population was 63.2% male with an average age of 32.6 years and BMI of 23.4 kg/m^2 (**Table 5.1**). The majority were non-smokers (62.3%), with three individuals self-reporting as former smokers. Among smokers, the median pack years were 3.8. Approximately half of the smokers smoked on the testing day, and half of the population was exposed to environmental tobacco smoke on the testing day. One female reported smoking. Average personal $PM_{2.5}$ was 118 $\mu g/m^3$ and average PM_{10} was 120 $\mu g/m^3$.

Global m^6A and gene expression characterization

RWE gene expression levels were positively correlated across individuals (R = 0.24 to 0.61), with the exception of WTAP, which was not correlated with other RWEs (**Fig. 5.1A**). Additionally, ALKBH5 and FTO expression levels were not correlated, suggesting

Table 5.1 Population Characteristics, Air Pollution Exposures, and Smoking Habits
PM is particulate matter

Characteristics	Full Population (N = 106) Mean ± S.D or N (%)	Nonsmokers (N = 63) Mean ± S.D or N (%)	Smokers (N = 43) Mean ± S.D or N (%)
Peripheral blood m^6A (% m^6A/A)	0.0503 ± 0.0088	0.0511 ± 0.0087	0.0491 ± 0.0089
Age (Years)	32.6 ± 6.88	33.8 ± 6.30	30.8 ± 7.39
BMI (kg/m^2)	23.4 ± 3.16	23.3 ± 2.99	23.6 ± 3.41
Personal $PM_{2.5}$ ($\mu g/m^3$)	118 ± 77.1	93.2 ± 55.6	155 ± 89.1
Ambient PM_{10} ($\mu g/m^3$)	120 ± 60.5	107 ± 56.8	141 ± 60.4
Personal Black Carbon ($\mu g/m^3$)	16.1 ± 6.66	14.1 ± 5.83	19.0 ± 6.83
Average Daily Temperature (℃)	25.2 ± 2.11	25.7 ± 1.98	24.6 ± 2.16
Average Dew Point	19.8 ± 2.01	20.3 ± 1.97	19.2 ± 1.88
Sex			
Female	39 (36.8)	38 (60.3)	1 (2.3)
Male	67 (63.2)	25 (39.7)	42 (97.7)
Occupation			
Indoor Worker	50 (47.2)	30 (47.6)	20 (46.5)
Truck Driver	56 (52.8)	33 (52.4)	23 (53.5)
Smoke While Wearing Monitor			
No	85 (80.2)	63 (100)	22 (51.2)
Yes	21 (19.8)		21 (48.8)
Pack Years			
0 Years	64 (60.4)	63 (100)	1 (2.3)
≤ 3.8 Years	21 (19.8)		21 (48.8)
> 3.8 Years	21 (19.8)		21 (48.8)
Environmental Tobacco Smoke			
No	52 (49.1)	37 (58.7)	15 (34.9)
Yes	54 (50.9)	26 (41.3)	28 (65.1)

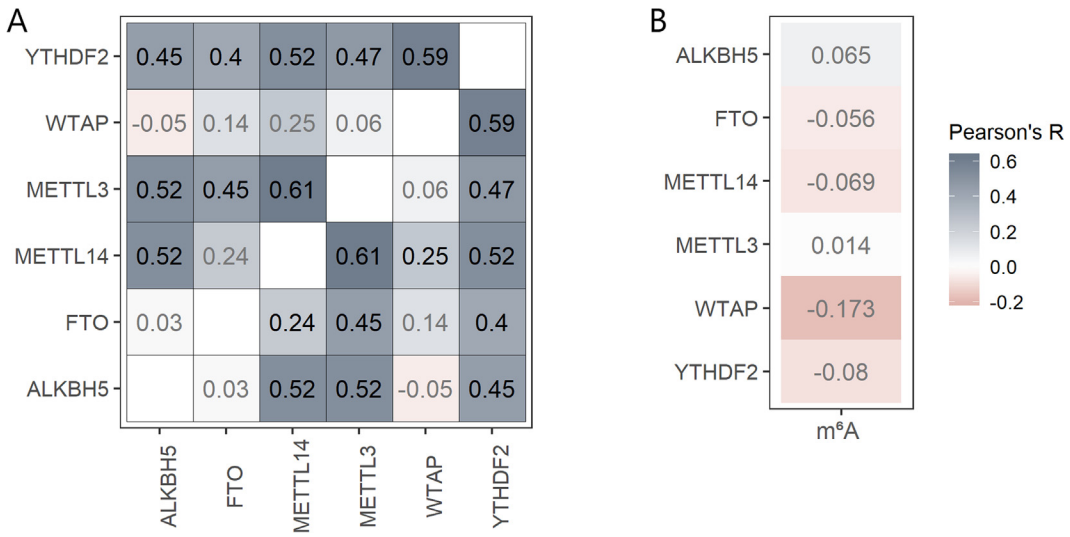


Figure 5.1 A) Pearson's correlation plot for mRNA expression levels of m^{6}A readers, writers and erasers in whole blood; and B) Pearson's correlation coefficients for each gene with m^{6}A. Number and gradient represent the correlation coefficient, and text color represents the significance of the correlation at $p \leq 0.05$ (black: significant, or gray: insignificant)

differential regulation or function. Gene expression levels were further not correlated with global m^{6}A levels (**Fig. 5.1B**).

Association of smoking status and air pollutants with m^{6}A

Ever smoking was associated with a relative 10.7% decrease in global m^{6}A (95% CI: −18.7–−1.75; p = 0.02) in men in comparison to nonsmokers (**Table 5.2**). Male smokers had on average 0.048% (95% CI: 0.046–0.051) m^{6}A/A in their blood, while male never smokers had on average 0.054% (95% CI: 0.050–0.058) m^{6}A/A (**Fig. 5.2A**). Pack years were also negatively associated with lower global m6A in men (**Fig. 5.2D**). Compared to non-smoking, smoking less than 3.8 years was associated with a 9.44% decrease in global m^{6}A (95% CI: −19.1–1.35; p = 0.08), while smoking greater than 3.8 years was associated with a 14.9% decrease in global m^{6}A (CI: –23.4–−5.35; p = 0.004)

Table 5.2 Effect estimates from linear models examining the effects of air pollutants and smoking status on global m^6A levels in whole blood RNA

CI: Confidence Interval; ETS, Environmental tobacco smoke.

[a] All models adjusted for batch variables.

[b] Smoking models adjusted for: age, sex, digestion batch; Air pollutant models adjusted for: age, sex, smoking status, smoking on day of exposure, average daily temperature, average dew point, day of week and digestion batch.

[c] Beta values are presented as the relative percent change in m^6A per 10 $\mu g/m^3$ increase in air pollutant or change from reference to smoking status. Values are on the multiplicative scale and were obtained by exponentiating the regression coefficient from the log-transformed mode

Type	Exposure	Model N	Crude Model[a]		Adjusted Model[b]	
			β (95% CI)[c]	P-Value	β (95% CI)[c]	P-Value
Smoking Status						
	Ever Smoker (Men)	67	-9.56 (-17.4- -0.95)	0.03	-10.7 (-18.7--1.75)	0.02
	Smoked While Wearing (Men)	67	-4.99 (-16.2-7.67)	0.41	-4.58 (-16.3-8.73)	0.47
	≤ 3.8 pack years (Men)	67	-8.45 (-17.5-1.52)	0.09	-9.44 (-19.1-1.35)	0.08
	>3.8 pack years (Men)	67	-14.8 (-23.3- -5.36)	< 0.01	-14.9 (-23.4--5.35)	< 0.01
	ETS (Nonsmokers)	43	-3.63 (-11.5-4.98)	0.39	-4.34 (-12.2-4.16)	0.3
Air Pollutant						
	Black Carbon	100	5.27 (-0.46-11.3)	0.07	5.95 (-0.96-13.3)	0.09
	PM_{10}	101	0.39 (-0.18-0.96)	0.18	0.09 (-0.79-0.99)	0.84
	$PM_{2.5}$	101	0.004 (0.45-0.45)	0.98	-0.03 (-0.66-0.60)	0.92

(**Table 5.2**). Men that smoked less than 3.8 pack years had on average 0.049% (95% CI: 0.045–0.054) m^6A/A, while those smoking more than 3.8 pack years had on average 0.046% (95% CI: 0.043–0.050) m^6A/A (**Fig. 5.2D**). ETS and smoking while wearing the

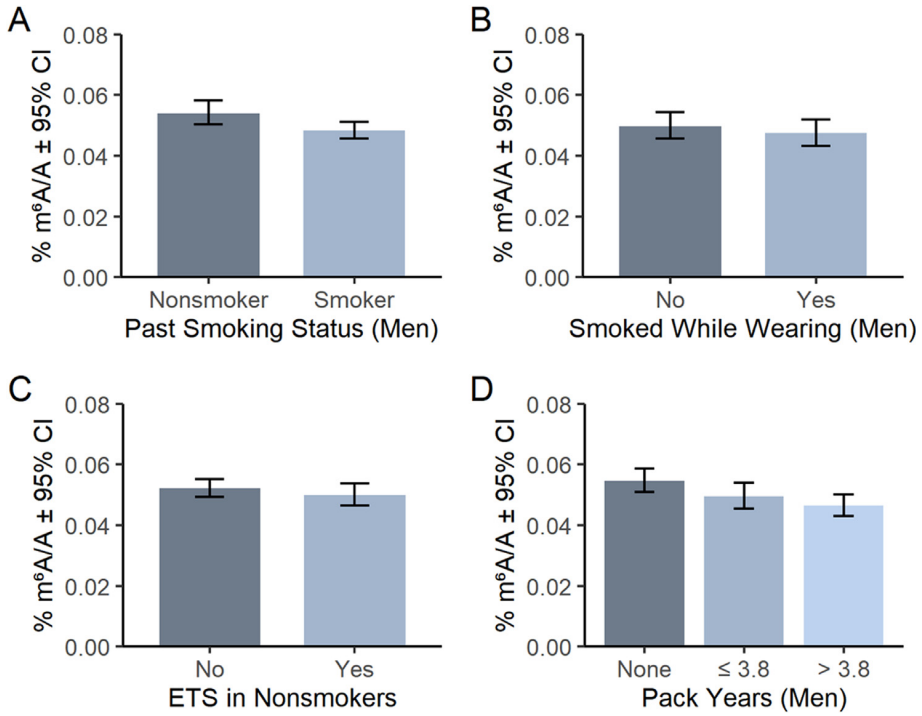


Figure 5.2 Adjusted mean percent global m^6A/A with 95% confidence intervals by smoking ca- tegory for: A) Ever/never smokers; B) Smoking while wearing the monitor; C) Environmental tobacco smoke (ETS); and D) Pack Years. Models adjusted for: age, sex, and digestion batch. ETS models also adjusted for ever/never smoking status.

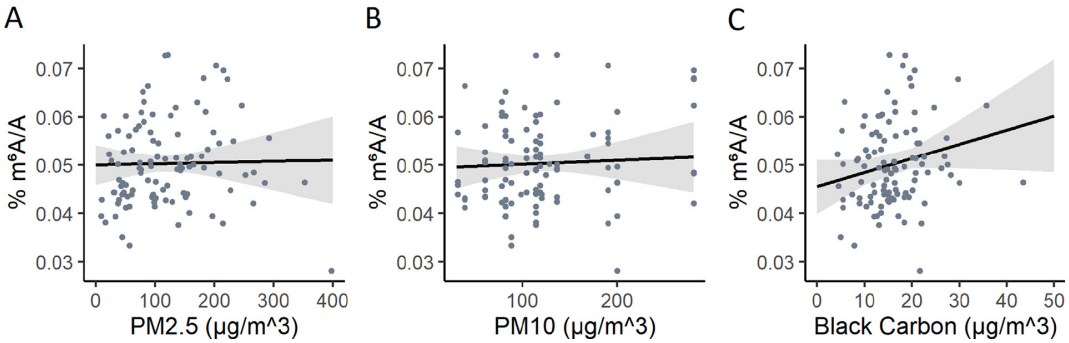


Figure 5.3 Predicted values for linear models of percent m^6A/A with A) PM2.5 ($\mu g/m^3$); B) PM_{10} ($\mu g/m^3$); and C) Black carbon ($\mu g/m^3$) with 95% confidence intervals and adjusted for: age, sex, smoking status, smoking on day of exposure, average daily temperature, average dew point, day of week and digestion batch. Points represent actual values.

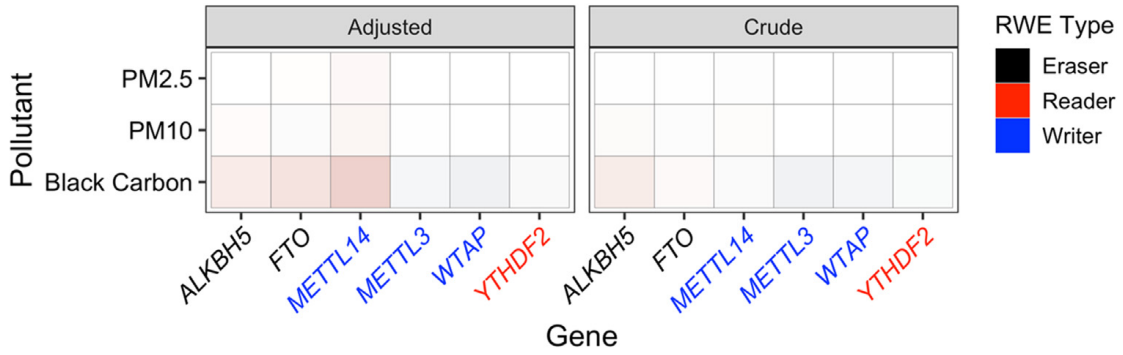


Figure 5.4 Heat map of effect estimates from adjusted and crude models as relative percent change in gene expression according to A) smoking status; or B) following a 6-hour exposure to air pollutants with a 10-unit change in air quality measure. Values are on the multiplicative scale and were obtained by exponentiating the regression coefficient from the log-transformed model. FDR adjusted P values ≤ 0.2 are indicated. ETS: Environmental tobacco smoke.

monitor were not associated with global m^6A (**Table 5.2, Fig. 5.2B & 5.2C**). Average BC exposure was borderline positively associated with peripheral blood m^6A in crude and adjusted models (5.95% per 10 $\mu g/m^3$ increase in BC; 95% CI: −0.96–13.3; p = 0.09) (**Table 5.2; Fig. 5.3C**). Global m^6A was not associated with acute exposure to $PM_{2.5}$ or PM_{10} (**Table 5.2; Fig. 5.3A & 5.3B**).

Associations of smoking status and air pollutants with RWE gene expression

After FDR correction, METTL3 expression was also positively associated with men smoking less than 3.8 pack years (20.6% increase in expression; 95% CI: 3.73–40.1; adjusted p = 0.03). Air pollutants were not associated with RWE expression levels (**Fig. 5.4B**). Full results with confidence intervals are available in **Tables 5.3 and 5.4**.

Sensitivity analyses

We tested the associations of smoking variables with m^6A and RWEs in the full population, which did not change our results (**Table 5.5; Figure 5.5**). We also examined

associations between air pollutants and m^6A RWEs in the population of individuals that did not smoke on the day of the exposure, to remove any confounding by smoking. Limiting the population slightly strengthened the positive associations of BC with m^6A (**Table 5.5**) but did not alter associations with RWEs (**Figure 5.6**). Next, we stratified the population by profession (office worker or truck driver) and repeated our crude analyses in the full population to retain a large enough sample size for analysis. Results between global m^6A and air pollutants were consistent with full, unstratified models (**Table 5.6**). Associations of m^6A with smoking variables was slightly greater among truck drivers than office workers (**Table 5.6**). Ultimately the direction of effect was conserved for all models.

Finally, to assess any sex-specific associations of air pollutants with m^6A and RWE gene expression, we repeated our ETS and air pollutant models stratified by sex. The association between ETS and m^6A levels appeared greater in male nonsmokers (-10.8%; 95% CI: −24.2–5.02; p = 0.16) in comparison to female nonsmokers (-0.85%; 95% CI: −11.5–11.1; p = 0.87) (**Table 5.7**). Associations between air pollutants and m^6A were largely similar between males and females (**Table 5.7**). Males and females appeared to have different associations between ETS and air pollutants with RWE gene expression (**Table 5.8**). ETS exposure was negatively associated with YTHDF2 expression in male nonsmokers (-19.4%; 95%CI: −29, −8.4) but no association was observed in female nonsmokers (0.05%; 95% CI: −10.4–11.8). Furthermore, both METTL14 and METTL3 trended towards negative associations with ETS exposure in males, but not in females. Trends also suggested sex-specific associations of ALKBH5, FTO, METTL14 and

Table 5.3. Effect estimates from linear models examining the effects of smoking status on reader, writer and eraser mRNA expression in whole blood RNA.
CI: Confidence Interval; ETS, environmental tobacco smoke; FDR: False discovery rate
[a]All models adjusted for RNA concentration and quality
[b]Smoking models adjusted for: age, sex, RNA concentration, and quality

		Crude Model[a]		Adjusted Model[b]	
Exposure	Gene	β (95% CI)[c]	FDR P-Value	β (95% CI)[c]	FDR P-Value
Ever Smoker (Men)	*ALKBH5*	11.4 (-2.44 – 27.2)	0.15	7.94 (-5.19 – 22.9)	0.24
	FTO	5.86 (-3.2 – 15.8)	0.63	4.83 (-4.3 – 14.8)	0.68
	METTL14	12.04 (-1.25 – 27.1)	0.10	13.6 (-0.15 – 29.2)	0.09
	METTL3	1.35 (-7.4 – 10.9)	0.99	1.83 (-7.17 – 11.7)	0.87
	WTAP	0.8 (-9.32 – 12.1)	0.88	2.45 (-7.91 – 14)	0.65
	YTHDF2	1.1 (-6.38 – 9.18)	0.78	1.46 (-6.24 – 9.8)	0.71
≤ 3.8 pack years (Men)	*ALKBH5*	15.59 (-0.93 – 34.8)	0.11	10.5 (-5.32 – 28.8)	0.24
	FTO	5.13 (-5.37 – 16.8)	0.65	3.33 (-7.29 – 15.2)	0.66
	METTL14	17.54 (1.55 – 36.1)	0.05	20.6 (3.73 – 40.2)	0.03
	METTL3	3.28 (-7.02 – 14.7)	0.81	4.43 (-6.37 – 16.5)	0.65
	WTAP	-2.08 (-13.2 – 10.4)	0.73	-0.31 (-11.9 – 12.8)	0.96
	YTHDF2	1.7 (-7.04 – 11.3)	0.71	2.2 (-6.94 – 12.3)	0.68
> 3.8 pack years (Men)	*ALKBH5*	3.44 (-11.2 – 20.5)	0.66	4.27 (-10.04 – 20.7)	0.57
	FTO	4.99 (-5.39 – 16.5)	0.65	5.31 (-5.08 – 16.9)	0.66
	METTL14	8.57 (-6.05 – 25.5)	0.32	8.09 (-6.4 – 24.8)	0.34
	METTL3	-1.83 (-11.5 – 8.94)	0.81	-2.02 (-11.8 – 8.78)	0.84
	WTAP	10.21 (-2.12 – 24.1)	0.13	9.86 (-2.42 – 23.7)	0.21
	YTHDF2	2.94 (-5.82 – 12.5)	0.65	2.85 (-5.97 – 12.5)	0.68
ETS (Nonsmokers)	*ALKBH5*	-9.24 (-18.8 – 1.37)	0.11	-9.33 (-19 – 1.48)	0.17
	FTO	-2.95 (-12.2 – 7.22)	0.85	-2.49 (-11.79 – 7.78)	0.84
	METTL14	-1.16 (-15.5 – 15.6)	0.88	-0.94 (-15.5 – 16.1)	0.90
	METTL3	-2.2 (-10.2 – 6.54)	0.80	-2.39 (-10.2 – 6.09)	0.77
	WTAP	-6.92 (-16.3 – 3.56)	0.18	-6.73 (-16.3 – 3.91)	0.48
	YTHDF2	-7.2 (-14.6 – 0.82)	0.29	-7.16 (-14.6 – 0.88)	0.24

Table 5.4 Effect estimates from linear models examining the effects of air pollutants on reader, writer and eraser mRNA expression in whole blood RNA.

Note: CI: Confidence Interval; ETS, environmental tobacco smoke; FDR: False discovery rate; PM, particulate matter

[a]All models adjusted for RNA concentration and quality

Air pollutant models adjusted for: age, sex, smoking status, smoking on day of exposure, average daily temperature, average dew point, day of week, and RNA concentration and quality

		Crude Model[a]		Adjusted Model[b]	
Exposure	Gene	β (95% CI)[c]	FDR P-Value	β (95% CI)[c]	FDR P-Value
Black Carbon	*ALKBH5*	-2.09 (-8.88 – 5.19)	0.82	-2.18 (-9.51 – 5.73)	0.74
	FTO	-0.59 (-5.56 – 4.63)	0.82	-3 (-9.24 – 3.65)	0.74
	METTL14	0.88 (-6.74 – 9.13)	0.82	-5.02 (-13.63 – 4.44)	0.74
	METTL3	2.25 (-2.75 – 7.51)	0.82	1.57 (-4.62 – 8.17)	0.74
	WTAP	1.77 (-4.08 – 8)	0.82	2.45 (-5.08 – 10.59)	0.74
	YTHDF2	0.97 (-3.44 – 5.6)	0.82	1.01 (-4.83 – 7.22)	0.74
PM_{10}	*ALKBH5*	-0.5 (-1.26 – 0.26)	0.59	-0.43 (-1.53 – 0.67)	0.86
	FTO	0.48 (-0.06 – 1.04)	0.51	0.72 (-0.23 – 1.69)	0.41
	METTL14	-0.35 (-1.19 – 0.48)	0.81	-1.08 (-2.41 – 0.25)	0.41
	METTL3	-0.02 (-0.57 – 0.51)	0.92	0.23 (-0.65 – 1.14)	0.90
	WTAP	-0.08 (-0.72 – 0.55)	0.92	0.05 (-1.03 – 1.14)	0.93
	YTHDF2	-0.08 (-0.57 – 0.39)	0.92	-0.1 (-0.96 – 0.75)	0.93
$PM_{2.5}$	*ALKBH5*	-0.08 (-0.67 – 0.5)	0.92	0.01 (-0.74 – 0.76)	0.98
	FTO	0.17 (-0.25 – 0.59)	0.92	-0.14 (-0.8 – 0.51)	0.98
	METTL14	-0.17 (-0.8 – 0.47)	0.92	-0.75 (-1.65 – 0.16)	0.63
	METTL3	0.06 (-0.34 – 0.47)	0.92	0.05 (-0.55 – 0.66)	0.98
	WTAP	0.11 (-0.36 – 0.6)	0.92	0 (-0.73 – 0.75)	0.98
	YTHDF2	0.01 (-0.34 – 0.38)	0.92	-0.03 (-0.61 – 0.55)	0.98

METTL3 with BC exposure, however, none reached statistical significance (**Table 5.8**). As above, large confidence intervals limit the interpretations of these findings.

Discussion

We present the first study to investigate the association between m^6A in peripheral blood RNA with cigarette smoking or air pollution. We detected a negative association between global m^6A and long-term smoking status, as evidenced by an apparent dose–response with pack years, and a potential positive association with BC exposure. However, we detected no correlation between acute smoking exposure (smoking on the day of blood collection) with global m^6A, nor with smoking and air pollutants with RWE gene expression.

Primarily, we observed that blood m^6A content was lower in men that smoked in comparison to men that never smoked. m^6A is critical for the inflammatory response and immune system activation,[41-43] both of which may be activated in chronic smokers. Lower total blood RNA m^6A content was also reported in two separate populations with type 2 diabetes.[14, 15] Furthermore, single nucleotide polymorphisms in m^6A consensus sequences were associated with blood pressure in a Chinese population[44] and m^6A is important for cancer progression and metastasis.[8] However, few human population studies have examined m^6A in relation to health, and further research is necessary to unravel the implications of our findings.

BC is formed from incomplete combustion of biomass and fossil fuels and contains some of the same chemical constituents present in cigarette smoke, such as metals and

Table 5.5 Results from sensitivity analyses examining the effects of air smoking status on global m^6A levels in whole blood RNA.

ETS, environmental tobacco smoke

[a]All models adjusted for digestion batch

[b]Smoking models adjusted for: age, sex, digestion batch; Air pollutant models adjusted for: age, sex, smoking status, smoking on day of exposure, average daily temperature, average dew point, day of week and digestion batch

[c]Beta values are presented as the relative percent change in m^6A per 10 $\mu g/m^3$ increase in air pollutant or change from reference to smoking status.Values are on the multiplicative scale and were obtained by exponentiating the regression coeffienient from the log-transformed model.

Type	Exposure	Crude Model[a]		Adjusted Model[b]	
		β (95% CI)[c]	P-Value	β (95% CI)[c]	P-Value
Smoking Status – Full Population	Ever Smoker	-7.79 (-14.5 – -0.57)	0.04	-11.0 (-18.2 – -3.15)	<0.01
	Smoked While Wearing	-4.16 (-13.8 – 6.51)	0.43	-4.0 (-13.7 – 6.86)	0.45
	≤ 3.8 pack years	-7.04 (-15.0 – 1.63)	0.11	-10.4 (-18.7 – -1.28)	0.03
	> 3.8 pack years	-12.5 (-20.2 – -3.92)	<0.01	-14.3 (-22.8 – -5.87)	<0.01
	ETS	-3.04 (-9.30 – 3.65)	0.36	-4.24 (-10.5 – 2.41)	0.20
Air Pollutant – Individuals that did not smoke on exposure day	Black Carbon	7.65 (0.75 – 15.03)	0.03	6.29 (-2.18 – 15.5)	0.15
	PM_{10}	0.58 (-0.03 – 1.2)	0.06	0.18 (-0.81 – 1.19)	0.72
	$PM_{2.5}$	0.48 (-0.14 – 1.11)	0.13	0.3 (-0.59 – 1.2)	0.5

ammonia.[23, 45] In the present study, long term smoking and BC presented opposing associations with global blood m^6A. The association with BC was observed in the population as a whole and when stratified by sex; however, as we only were able to enroll

male smokers in this analysis, our negative associations with smoking were only observed in men. This difference may result from distinct the toxic compounds affecting m^6A between BC and cigarette smoke. Although tobacco smoke is a contributor to fine particulate air pollution, concentrations of some toxicants are present at much higher levels in cigarette smoke, which has more organic carbon than black carbon. Also, the exposure duration may have differential effects on m^6A, as long-term smoking was associated with global m^6A while smoking on the day of blood collection was not. Acute spikes in daily average BC may affect m^6A differently in comparison to long term smoking. Previous *in vitro* studies have suggested that the m^6A response to external stressors is rapid,[46] triaging transcripts into stress granules to regulate translation.[18] Nonetheless, this analysis was conducted in a small population. We cannot discount that these findings may be spurious and require replication in a larger sample size. Future research should also address any differences in exposure duration and timing on m^6A levels, as well as sex-specific effects of smoking, which we were unable to evaluate here.

We also observed no associations between smoking in males and air pollutants with RWE expression in the full population, or when stratified by sex. Previous analyses on 24-hour PM2.5 exposure and RWE gene expression in a population of individuals from the Czech Republic, detected a positive association with METTL3, WTAP, FTO, and ALKBH5 expression.[20] However, this study investigated RNA in leukocytes and examined a different exposure duration than the present study.[47] Interestingly, we observed no correlation between global m^6A levels in blood with RWE gene expression in our population. Similarly, a study on total blood m^6A levels in gastric cancer patients and

Figure 5.5 Heat map of effect estimates as relative percent change in gene expression in sensitivity analyses on smoking status. Values are on the multiplicative scale and were obtained by exponentiating the regression coefficient from the log-transformed model. P-values are indicated on tiles with significant effect estimates following FDR adjustment for multiple comparisons.

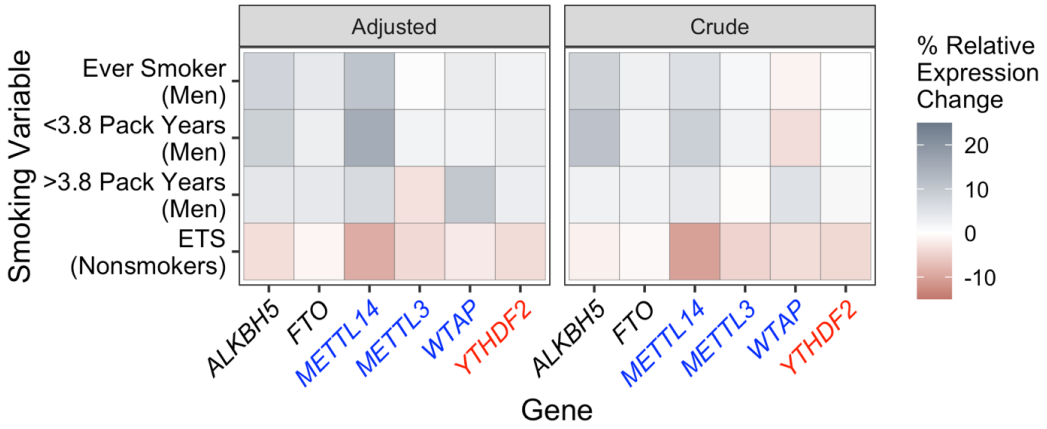
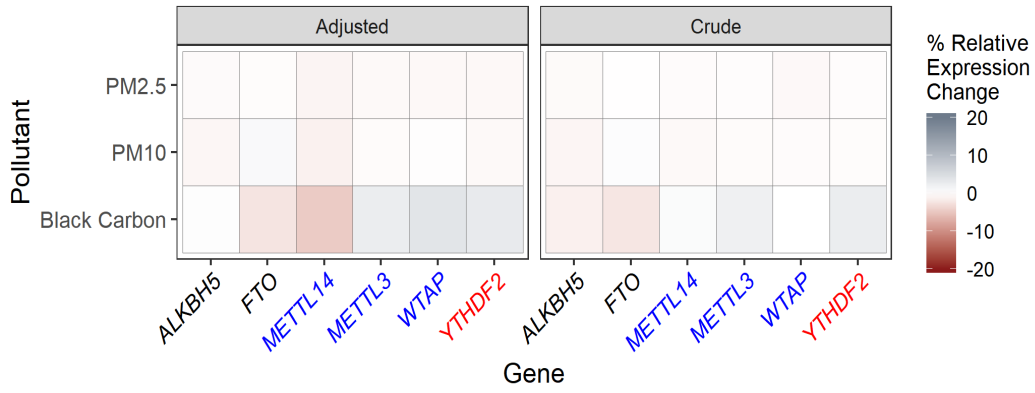


Figure 5.6 Heat map of effect estimates from crude and adjusted models as relative percent change in gene expression following a 6-hour exposure to air pollutants with a 10-unit change in air quality measure in models excluding individuals that smoked on the day of m^{6}A measurement. Values are on the multiplicative scale and were obtained by exponentiating the regression coefficient from the log-transformed model. FDR adjusted P values ≤ 0.05 are indicated.



Exposure	Truck Driver		Office Workers	
	β (95% CI)[a]	P Value	β (95% CI)[a]	P Value
Smoking Status				
Ever Smoker	-12.51 (-21.56 – -2.43)	0.02	-5.28 (-15.25 – 5.87)	0.33
Smoked While Wearing[b]	-8.81 (-21.02 – 5.29)	0.20	7.16 (-11.06 – 29.11)	0.46
≤ 3.8 pack years	-7.84 (-21.61 – 8.36)	0.32	-7.65 (-18.21 – 4.28)	0.19
> 3.8 pack years	-13.85 (-23.21 – -3.36)	0.01	-12.41 (-28.19 – 6.85)	0.19
ETS[b]	-5.82 (-13.96 – 3.08)	0.19	-0.59 (-11.75 – 11.99)	0.92
Air Pollutant				
Black Carbon	4.5 (-3.24 – 12.85)	0.26	10.43 (-4.13 – 27.2)	0.16
PM_{10}	-0.13 (-1.00 – 0.76)	0.78	0.75 (-0.15 – 1.66)	0.10
$PM_{2.5}$	-0.60 (-1.26 – 0.07)	0.08	0.82 (-0.13 – 1.77)	0.09

Table 5.6. Effect estimates from crude linear models stratified by occupation examining the effects of air pollutants and smoking status on global m^6A levels in whole blood RNA from the full study population.

Note: ETS, environmental tobacco smoke; PM, particulate matter. All models adjusted for digest batch.

[a]Beta values are presented as the relative percent change in m^6A per 10 $\mu g/m^3$ increase in air pollutant or change from reference to smoking status.Values are on the multiplicative scale and were obtained by exponentiating the regression coeffienient from the log-transformed model.

[b]Models co-adjusted for smoking status.

Exposure	Males		Females	
	β (95% CI)[a]	P Value	β (95% CI)[a]	P Value
Smoking Status				
ETS (Nonsmokers)	-10.8 (-24.2 – 5.02)	0.16	-0.85 (-11.5 – 11.1)	0.87
Air Pollutant				
Black Carbon	6.15 (-1.17 – 14.0)	0.10	10.2 (-1.3 – 23.1)	0.08
PM_{10}	0.39 (-0.35 – 1.13)	0.30	0.90 (-0.75 – 2.58)	0.27
$PM_{2.5}$	0.01 (-0.57 – 0.60)	0.97	1.53 (-0.15 – 3.24)	0.07

Table 5.7. Effect estimates from crude linear models stratified by sex examining the effects of air pollutants and environmental tobacco smoke on global m^6A levels in whole blood RNA.

ETS, environmental tobacco smoke; PM, particulate matter. All models adjusted for digest batch.

[a]Beta values are presented as the relative percent change in m^6A per 10 $\mu g/m^3$ increase in air pollutant or change from reference to smoking status.Values are on the multiplicative scale and were obtained by exponentiating the regression coeffienient from the log-transformed model.

Table 5.8. Effect estimates from crude linear models stratified by sex examining the effects of ETS air pollutants on reader, writer and eraser mRNA expression in whole blood RNA.

CI: Confidence Interval; ETS, environmental tobacco smoke; FDR: False discovery rate; PM, particulate matter

[a]All models adjusted for RNA concentration and quality

		Males		Females	
Exposure	Gene	β (95% CI)[c]	FDR P-Value	β (95% CI)[c]	FDR P-Value
ETS (Nonsmokers)	*ALKBH5*	-8.86 (-27.5 – 14.47)	0.61	-8.5 (-18.9 – 3.26)	0.43
	FTO	-0.72 (-17.8 – 19.9)	0.99	-3.42 (-14.8 – 9.47)	0.69
	METTL14	-13.9 (-35.5 – 14.7)	0.57	9.4 (-9.64 – 32.5)	0.59
	METTL3	-13.8 (-25.1 – -0.71)	0.24	5.75 (-4.38 – 17)	0.57
	WTAP	-16.2 (-31.2 – 2.05)	0.3	-3.67 (-15.5 – 9.86)	0.69
	YTHDF2	-19.4 (-29.0 – -8.4)	0.03	0.05 (-10.4 – 11.8)	0.99
Black Carbon	*ALKBH5*	-6.09 (-15.2 – 4.03)	0.69	6.61 (-5.74 – 20.6)	0.68
	FTO	1.68 (-5.17 – 9.04)	0.69	-7.57 (-18.3 – 4.53)	0.68
	METTL14	3.21 (-6.54 – 14.0)	0.69	-2.39 (-19.7 – 18.6)	0.8
	METTL3	4.59 (-2.5 – 12.2)	0.69	-4.26 (-13.8 – 6.28)	0.68
	WTAP	1.93 (-5.75 – 10.3)	0.69	4.76 (-8.31 – 19.7)	0.68
	YTHDF2	2.46 (-3.57 – 8.89)	0.69	3.05 (-7.78 – 15.2)	0.68
PM_{10}	*ALKBH5*	-0.89 (-1.89 – 0.11)	0.49	1.38 (-0.77 – 3.59)	0.81
	FTO	0.62 (-0.07 – 1.33)	0.49	1.07 (-1.12 – 3.31)	0.85
	METTL14	-0.39 (-1.37 – 0.6)	0.78	-0.03 (-3.41 – 3.45)	0.98
	METTL3	-0.14 (-0.86 – 0.56)	0.95	0.14 (-1.7 – 2.03)	0.97
	WTAP	-0.06 (-0.85 – 0.72)	0.95	0.87 (-1.47 – 3.27)	0.85
	YTHDF2	-0.08 (-0.7 – 0.53)	0.95	0.76 (-1.18 – 2.75)	0.85
$PM_{2.5}$	*ALKBH5*	-0.07 (-0.87 – 0.73)	0.89	-0.16 (-2.32 – 2.03)	0.96
	FTO	0.28 (-0.26 – 0.84)	0.89	-1.46 (-3.55 – 0.67)	0.89
	METTL14	-0.13 (-0.91 – 0.63)	0.89	-0.39 (-3.72 – 3.05)	0.96
	METTL3	0.13 (-0.42 – 0.68)	0.89	-0.67 (-2.47 – 1.15)	0.96
	WTAP	0.09 (-0.51 – 0.71)	0.89	-0.15 (-2.47 – 2.21)	0.96
	YTHDF2	0.15 (-0.32 – 0.63)	0.89	0.22 (-1.71 – 2.19)	0.96

controls found no correlation between m^6A and ALKBH5 and FTO gene expression.[48] Other previous human studies have observed associations between global m^6A and RWE gene expression in whole blood, these studies included populations of type 2 diabetics,[14, 15] which may present a greater transcriptional response in RWEs than individuals without

clinical pathologies. Furthermore, the participants in those previous studies were older than the present study population (mean age approximately 55 years in comparison to 33 years). However, our findings may be the result of a stronger relationship between m^6A and RWE activity or protein abundance than with gene expression. Alternatively, the lack of relationship between RWE gene expression and global m^6A in our population may be driven by a large proportion of m^6A in rRNA, which is the predominant form of RNA with m^6A and is added by a different set of methyltransferases.[49-51] Due to limitations in RNA sample concentrations, we were only able to assess expression of the most well-known m^6A RWEs in this study and were not able to examine those responsible for rRNA methylation. Further research is necessary to investigate these hypotheses.

Finally, this is one of the first observational human population studies to evaluate blood m^6A levels and little information is available on how m^6A levels vary according to population characteristics. In our population of office workers and truck drivers from Beijing, China, m^6A levels did not differ by age, sex, or BMI. However, it appears that some associations with RWEs may be sex specific. Similarly, a study of blood m^6A levels in gastric cancer patients found no differences in blood RNA m^6A levels between sexes.[48] Clinical studies on diabetes demonstrated a correlation between m^6A and BMI, however, this population contained diabetics and they did not investigate correlations with sex.[14] A study of total m^6A in blood from healthy patients and those with major depressive disorder indicated no interaction between sex, diagnosis, and dexamethasone use.[37]

Although this study is highly novel, it had several limitations. Firstly, our information on smoking status was self-reported, which is subject to bias. Serum cotinine

would be a more reliable measure of recent smoke exposure, but these data were unavailable. Nonetheless, measurement error would have likely biased our results towards the null due to social desirability bias, as smokers may disproportionately fail to report their smoking habit. Furthermore, we only enrolled male smokers in this analysis and were unable to address any sex-specific associations of smoking with m^6A. Although we were able to confirm that our associations between m^6A and air pollutants were not sex specific, there may be some differences in RWE expression between sexes that merit future investigation. Future research on epitranscriptomics in human populations will need to focus on biological differences between sexes. Additionally, no information on the cell type in whole blood was available, which may mediate the relationship between exposures and m^6A. Previous research has demonstrated that total m^6A levels can vary between tissues and that tissues can be identified by specific m^6A peaks, although immune cells were not evaluated.[52] Furthermore, total m^6A levels were shown to increase following dendritic cell activation by LPS[42] and m^6A plays a critical role in the immune system.[53] Thus, as we were unable to adjust for blood cell type composition in our analysis, our measured associations may be reflective of cell type alterations, rather than specific m^6A modifications. This is particularly true for smoking, which has been shown to be associated with blood cell type composition and other epigenetic modifications.[54, 55] Future research will investigate the impact on cell type on associations of environmental pollutants with m^6A. Also, mRNA generally makes up only 1% of total RNA in a cell. Due to limitations on the sample quantity, we were unable to perform m^6A RNA immunoprecipitation sequencing or examine m^6A levels in mRNA. However, with technological advances and

increasing number of cohort studies designed specifically for investigations on m^6A, the impact of these issues will diminish.

Despite these limitations, this is the first human study reporting relationships between m^6A and smoking or environmental exposures. Tobacco smoking and ambient air pollution are both major contributors to cancer and cardiovascular disease mortality.[56] We find that global m^6A is downregulated in long-term smokers and may be upregulated following acute BC exposure. Coupled with the emerging body of literature on m^6A in stress in experimental model systems and those showing a critical role for m^6A in human health and disease, our study demonstrates the potential for m^6A for use as a biomarker in environmental health research. Future studies will profile m^6A on specific mRNA transcripts to determine which pathways regulate the response to cigarette smoke and particulate air pollution at the transcriptional level and investigate these effects in relation to cardiovascular disease and lung function.

References

1. Zaccara, S.; Ries, R. J.; Jaffrey, S. R., Reading, writing and erasing mRNA methylation. *Nat Rev Mol Cell Biol* **2019,** *20* (10), 608-624.

2. Zhao, B. S.; Roundtree, I. A.; He, C., Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* **2017,** *18* (1), 31-42.

3. Wang, X.; Lu, Z.; Gomez, A.; Hon, G. C.; Yue, Y.; Han, D.; Fu, Y.; Parisien, M.; Dai, Q.; Jia, G.; Ren, B.; Pan, T.; He, C., N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014,** *505* (7481), 117-20.

4. Aguilo, F.; Zhang, F.; Sancho, A.; Fidalgo, M.; Di Cecilia, S.; Vashisht, A.; Lee, D. F.; Chen, C. H.; Rengasamy, M.; Andino, B.; Jahouh, F.; Roman, A.; Krig, S. R.; Wang, R.; Zhang, W.; Wohlschlegel, J. A.; Wang, J.; Walsh, M. J., Coordination of m(6)A mRNA Methylation and Gene Transcription by ZFP217 Regulates Pluripotency and Reprogramming. *Cell Stem Cell* **2015,** *17* (6), 689-704.

5. Batista, P. J.; Molinie, B.; Wang, J.; Qu, K.; Zhang, J.; Li, L.; Bouley, D. M.; Lujan, E.; Haddad, B.; Daneshvar, K.; Carter, A. C.; Flynn, R. A.; Zhou, C.; Lim, K. S.; Dedon, P.; Wernig, M.; Mullen, A. C.; Xing, Y.; Giallourakis, C. C.; Chang, H. Y., M6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **2014,** *15* (6), 707-719.

6. Wang, Y.; Li, Y.; Yue, M.; Wang, J.; Kumar, S.; Wechsler-Reya, R. J.; Zhang, Z.; Ogawa, Y.; Kellis, M.; Duester, G.; Zhao, J. C., N(6)-methyladenosine RNA modification regulates embryonic neural stem cell self-renewal through histone modifications. *Nat Neurosci* **2018,** *21* (2), 195-206.

7. Weng, H.; Huang, H.; Wu, H.; Qin, X.; Zhao, B. S.; Dong, L.; Shi, H.; Skibbe, J.; Shen, C.; Hu, C.; Sheng, Y.; Wang, Y.; Wunderlich, M.; Zhang, B.; Dore, L. C.; Su, R.; Deng, X.; Ferchen, K.; Li, C.; Sun, M.; Lu, Z.; Jiang, X.; Marcucci, G.; Mulloy, J. C.; Yang, J.; Qian, Z.; Wei, M.; He, C.; Chen, J., METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes Leukemogenesis via mRNA m(6)A Modification. *Cell Stem Cell* **2018,** *22* (2), 191-205 e9.

8. Dai, D.; Wang, H.; Zhu, L.; Jin, H.; Wang, X., N6-methyladenosine links RNA metabolism to cancer progression. *Cell Death Dis* **2018,** *9* (2), 124.

9. Lichinchi, G.; Zhao, Boxuan S.; Wu, Y.; Lu, Z.; Qin, Y.; He, C.; Rana, Tariq M., Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. 2016; Vol. 20, pp 666-673.

10. Zhang, C.; Fu, J.; Zhou, Y., A Review in Research Progress Concerning m6A Methylation and Immunoregulation. *Front Immunol* **2019,** *10*, 922.

11. Kaspi, A.; Khurana, I.; Ziemann, M.; Connor, T.; Spolding, B.; Zimmet, P.; Walder, K.; El-Osta, A., Diet during Pregnancy is Implicated in the Regulation of Hypothalamic RNA Methylation and Risk of Obesity in Offspring. *Molecular Nutrition & Food Research* **2018,** *62* (14), 1613-4125.

12. Song, T.; Yang, Y.; Wei, H.; Xie, X.; Lu, J.; Zeng, Q.; Peng, J.; Zhou, Y.; Jiang, S.; Peng, J., Zfp217 mediates m6A mRNA methylation to orchestrate transcriptional and post-transcriptional regulation to promote adipogenic differentiation. *Nucleic Acids Res* **2019,** *47* (12), 6130-6144.

13. Zhao, X.; Yang, Y.; Sun, B.-F.; Shi, Y.; Yang, X.; Xiao, W.; Hao, Y.-J.; Ping, X.-L.; Chen, Y.-S.; Wang, W.-J.; Jin, K.-X.; Wang, X.; Huang, C.-M.; Fu, Y.; Ge, X.-M.; Song, S.-H.; Jeong, H. S.; Yanagisawa, H.; Niu, Y.; Jia, G.-F.; Wu, W.; Tong, W.-M.; Okamoto, A.; He, C.; Danielsen, J. M. R.; Wang, X.-J.; Yang, Y.-G., FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Research* **2014,** *24* (12), 1403-1419.

14. Shen, F.; Huang, W.; Huang, J. T.; Xiong, J.; Yang, Y.; Wu, K.; Jia, G. F.; Chen, J.; Feng, Y. Q.; Yuan, B. F.; Liu, S. M., Decreased N(6)-methyladenosine in peripheral blood RNA from diabetic patients is associated with FTO expression rather than ALKBH5. *J Clin Endocrinol Metab* **2015,** *100* (1), E148-54.

15. Yang, Y.; Shen, F.; Huang, W.; Qin, S.; Huang, J. T.; Sergi, C.; Yuan, B. F.; Liu, S. M., Glucose Is Involved in the Dynamic Regulation of m6A in Patients With Type 2 Diabetes. *J Clin Endocrinol Metab* **2019,** *104* (3), 665-673.

16. Xu, K.; Yang, Y.; Feng, G. H.; Sun, B. F.; Chen, J. Q.; Li, Y. F.; Chen, Y. S.; Zhang, X. X.; Wang, C. X.; Jiang, L. Y.; Liu, C.; Zhang, Z. Y.; Wang, X. J.; Zhou, Q.; Yang, Y. G.; Li, W., Mettl3-mediated m(6)A regulates spermatogonial differentiation and meiosis initiation. *Cell Res* **2017,** *27* (9), 1100-1114.

17. Yang, Y.; Huang, W.; Huang, J. T.; Shen, F.; Xiong, J.; Yuan, E. F.; Qin, S. S.; Zhang, M.; Feng, Y. Q.; Yuan, B. F.; Liu, S. M., Increased N6-methyladenosine in Human Sperm RNA as a Risk Factor for Asthenozoospermia. *Sci Rep* **2016,** *6*, 24345.

18. Anders, M.; Chelysheva, I.; Goebel, I.; Trenkner, T.; Zhou, J.; Mao, Y.; Verzini, S.; Qian, S. B.; Ignatova, Z., Dynamic m(6)A methylation facilitates mRNA triaging to stress granules. *Life Sci Alliance* **2018,** *1* (4), e201800113.

19. Zhou, J.; Wan, J.; Gao, X.; Zhang, X.; Jaffrey, S. R.; Qian, S. B., Dynamic m^6A mRNA methylation directs translational control of heat shock response. *Nature* **2015,** *526* (7574), 591-4.

20. Cayir, A.; Barrow, T. M.; Guo, L.; Byun, H.-M., Exposure to environmental toxicants reduces global N6-methyladenosine RNA methylation and alters expression of RNA methylation modulator genes. *Environmental Research* **2019,** *175*, 228-234.

21. Bai, L.; Tang, Q.; Zou, Z.; Meng, P.; Tu, B.; Xia, Y.; Cheng, S.; Zhang, L.; Yang, K.; Mu, S.; Wang, X.; Qin, X.; Lv, B.; Cao, X.; Qin, Q.; Jiang, X.; Chen, C., m6A Demethylase FTO Regulates Dopaminergic Neurotransmission Deficits Caused by Arsenite. *Toxicol Sci* **2018,** *165* (2), 431-446.

22. Brunekreef, B.; Holgate, S. T., Air Pollution and Health. *The Lanclet* **2002,** *360* (9341), P1233-1242.

23. Centers for Disease Control and Prevention (US), National Center for Chronic Disease Prevention and Health Promotion (US), Office on Smoking and Health (US), 2010. How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking- Attributable Disease: A Report of the Surgeon General, Publications and Reports of the Surgeon General. Centers. *Centers for Disease Control and Prevention (US), Atlanta (GA)*. **2010**.

24. Pope, C. A., 3rd; Burnett, R. T.; Turner, M. C.; Cohen, A.; Krewski, D.; Jerrett, M.; Gapstur, S. M.; Thun, M. J., Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure-response relationships. *Environ Health Perspect* **2011,** *119* (11), 1616-21.

25. Ferrari, L.; Carugno, M.; Bollati, V., Particulate matter exposure shapes DNA methylation through the lifespan. *Clin Epigenetics* **2019,** *11* (1), 129.

26. Lee, K. W.; Pausova, Z., Cigarette smoking and DNA methylation. *Front Genet* **2013,** *4*, 132.

27. Hou, L.; Wang, S.; Dou, C.; Zhang, X.; Yu, Y.; Zheng, Y.; Avula, U.; Hoxha, M.; Díaz, A.; McCracken, J.; Barretta, F.; Marinelli, B.; Bertazzi, P. A.; Schwartz, J.; Baccarelli, A. A., Air pollution exposure and telomere length in highly exposed subjects in Beijing, China: A repeated-measure study. *Environment International* **2012,** *48*, 71-77.

28. Baccarelli, A.; Barretta, F.; Dou, C.; Zhang, X.; McCracken, J. P.; Díaz, A.; Bertazzi, P.; Schwartz, J.; Wang, S.; Hou, L., Effects of particulate air pollution on blood pressure in a highly exposed population in Beijing, China: A repeated-measure study. *Environmental Health: A Global Access Science Source* **2011,** *10* (1), 1-10.

29. ISO, 1993. International Organization for Standardization 9835. Ambient Air-Determination of a Black Smoke Index, International Organisation for Standardisation, Standardisation, Geneva.

30. Janssen, N. A. H.; van Vliet, P. H. N.; Aarts, F.; Harssema, H.; Brunekreef, B., Assessment of exposure to traffic related air pollution of children attending schools near motorways. *Atmospheric Environment* **2001,** *35* (22), 3875-3884.

31. Kinney, P. L.; Aggarwal, M.; Northridge, M. E.; Janssen, N. A. H.; Shepard, P., Airborne concentrations of PM2.5 and diesel exhaust particles on Harlem sidewalks: A community-based pilot study. *Environmental Health Perspectives* **2000,** *108* (3), 213-218.

32. Fu, L.; Amato, N. J.; Wang, P.; McGowan, S. J.; Niedernhofer, L. J.; Wang, Y., Simultaneous Quantification of Methylated Cytidine and Adenosine in Cellular and Tissue RNA by Nano-Flow Liquid Chromatography–Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal. Chem.* **2015,** *87ha* (15), 7653-7659.

33. Bustin, S. A.; Benes, V.; Garson, J. A.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M. W.; Shipley, G. L.; Vandesompele, J.; Wittwer, C. T., The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **2009,** *55* (4), 611-22.

34. Ledderose, C.; Heyn, J.; Limbeck, E.; Kreth, S., Selection of reliable reference genes for quantitative real-time PCR in human T cells and neutrophils. *BMC Research Notes* **2011,** *4* (27).

35. Vaiphei, S. T.; Keppen, J.; Nongrum, S.; Chaubey, R. C.; Kma, L.; Sharan, R. N., Evaluation of endogenous control gene(s) for gene expression studies in human blood exposed to 60Co gamma-rays ex vivo. *J Radiat Res* **2015,** *56* (1), 177-85.

36. Vega-Sanchez, R.; Arenas-Hernandez, M.; Vazquez-Perez, J. A.; Moreno-Valencia, Y.; Gomez-Lopez, N., Evaluation of reference genes for expression studies in leukocytes from term human pregnancy. *Placenta* **2015,** *36* (2), 240-245.

37. Engel, M.; Eggert, C.; Kaplick, P. M.; Eder, M.; Roh, S.; Tietze, L.; Namendorf, C.; Arloth, J.; Weber, P.; Rex-Haffner, M.; Geula, S.; Jakovcevski, M.; Hanna, J. H.; Leshkowitz, D.; Uhr, M.; Wotjak, C. T.; Schmidt, M. V.; Deussing, J. M.; Binder, E. B.; Chen, A., The Role of m(6)A/m-RNA Methylation in Stress Response Regulation. *Neuron* **2018,** *99* (2), 389-403 e9.

38. Rydbirk, R.; Folke, J.; Winge, K.; Aznar, S.; Pakkenberg, B.; Brudek, T., Assessment of brain reference genes for RT-qPCR studies in neurodegenerative diseases. *Sci Rep* **2016,** *6*, 37116.

39. Vandesompele, J.; De Preter, K.; Pattyn, F.; Poppe, B.; Van Roy, N.; De Paepe, A.; Speleman, F., Accurate normalization of real-time quantitative RT-PCR data

by geometric averaging of multiple internal control genes. *Genome biology* **2002,** *3* (7), 1-12.

40. R Core Team, A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna, Austria*.

41. Li, H.-B.; Tong, J.; Zhu, S.; Batista, P. J.; Duffy, E. E.; Zhao, J.; Bailis, W.; Cao, G.; Kroehling, L.; Chen, Y.; Wang, G.; Broughton, J. P.; Chen, Y. G.; Kluger, Y.; Simon, M. D.; Chang, H. Y.; Yin, Z.; Flavell, R. A., m6A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. *Nat.* **2017,** *548* (7667), 338-342.

42. Wang, H.; Hu, X.; Huang, M.; Liu, J.; Gu, Y.; Ma, L.; Zhou, Q.; Cao, X., Mettl3-mediated mRNA m(6)A methylation promotes dendritic cell activation. *Nat Commun* **2019,** *10* (1), 1898.

43. Yu, R.; Li, Q.; Feng, Z.; Cai, L.; Xu, Q., m6A Reader YTHDF2 Regulates LPS-Induced Inflammatory Response. *Int J Mol Sci* **2019,** *20* (6).

44. Mo, X. B.; Lei, S. F.; Zhang, Y. H.; Zhang, H., Examination of the associations between m(6)A-associated single-nucleotide polymorphisms and blood pressure. *Hypertens Res* **2019,** *42* (10), 1582-1589.

45. Wang, J.; Liu, D.; Ge, X.; Wu, Y.; Shen, F.; Chen, M.; Zhao, J.; Xie, C.; Wang, Q.; Xu, W.; Zhang, J.; Hu, J.; Allan, J.; Joshi, R.; Fu, P.; Coe, H.; Sun, Y., Characterization of black carbon-containing fine particles in Beijing during wintertime. *Atmospheric Chemistry and Physics* **2019,** *19* (1), 447-458.

46. Zhao, B. S.; Nachtergaele, S.; Roundtree, I. A.; He, C., Our views of dynamic N 6 -methyladenosine RNA methylation. *RNA* **2018,** *24* (3), 268-272.

47. Rossner, P.; Tulupova, E.; Rossnerova, A.; Libalova, H.; Honkova, K.; Gmuender, H.; Pastorkova, A.; Svecova, V.; Topinka, J.; Sram, R. J., Reduced gene expression levels after chronic exposure to high concentrations of air pollutants. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **2015,** *780*, 60-70.

48. Ge, L.; Zhang, N.; Chen, Z.; Song, J.; Wu, Y.; Li, Z.; Chen, F.; Wu, J.; Li, D.; Li, J.; Wang, C.; Wang, H.; Wang, J., Level of N6-Methyladenosine in Peripheral Blood RNA: A Novel Predictive Biomarker for Gastric Cancer. *Clinical Chemistry* **2020,** *66* (2), 342-351.

49. Ma, H.; Wang, X.; Cai, J.; Dai, Q.; Natchiar, S. K.; Lv, R.; Chen, K.; Lu, Z.; Chen, H.; Shi, Y. G.; Lan, F.; Fan, J.; Klaholz, B. P.; Pan, T.; Shi, Y.; He, C., N(6-)Methyladenosine methyltransferase ZCCHC4 mediates ribosomal RNA methylation. *Nat Chem Biol* **2019,** *15* (1), 88-94.

50. Pendleton, K. E.; Chen, B.; Liu, K.; Hunter, O. V.; Xie, Y.; Tu, B. P.; Conrad, N. K., The U6 snRNA m(6)A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **2017,** *169* (5), 824-835 e14.

51. van Tran, N.; Ernst, F. G. M.; Hawley, B. R.; Zorbas, C.; Ulryck, N.; Hackert, P.; Bohnsack, K. E.; Bohnsack, M. T.; Jaffrey, S. R.; Graille, M.; Lafontaine, D. L. J., The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic Acids Res* **2019,** *47* (15), 7719-7733.

52. Liu, J.; Li, K.; Cai, J.; Zhang, M.; Zhang, X.; Xiong, X.; Meng, H.; Xu, X.; Huang, Z.; Peng, J.; Fan, J.; Yi, C., Landscape and Regulation of m(6)A and m(6)Am Methylome across Human and Mouse Tissues. *Mol Cell* **2020,** *77* (2), 426-440 e6.

53. Shulman, Z.; Stern-Ginossar, N., The RNA modification N(6)-methyladenosine as a novel regulator of the immune system. *Nat Immunol* **2020,** *21* (5), 501-512.

54. Bauer, M.; Fink, B.; Thurmann, L.; Eszlinger, M.; Herberth, G.; Lehmann, I., Tobacco smoking differently influences cell types of the innate and adaptive immune system-indications from CpG site methylation. *Clin Epigenetics* **2015,** *7*, 83.

55. Su, D.; Wang, X.; Campbell, M. R.; Porter, D. K.; Pittman, G. S.; Bennett, B. D.; Wan, M.; Englert, N. A.; Crowl, C. L.; Gimple, R. N.; Adamski, K. N.; Huang, Z.; Murphy, S. K.; Bell, D. A., Distinct Epigenetic Effects of Tobacco Smoking in Whole Blood and among Leukocyte Subtypes. *PLoS One* **2016,** *11* (12), e0166486.

56. Lelieveld, J.; Klingmuller, K.; Pozzer, A.; Poschl, U.; Fnais, M.; Daiber, A.; Munzel, T., Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions. *Eur Heart J* **2019,** *40* (20), 1590-1596.

Chapter 6: Modified ribonucleosides and HIV-1 and SARS-CoV-2 infection

Introduction

The recent appreciation of the functions of modified ribonucleosides has led to investigations into their roles in disease pathology. Modified ribonucleosides have been identified in high abundance in ribosomal RNA (rRNA) and transfer RNA (tRNA).[1, 2] In mRNA, N^6-methyladenosine's (m^6A) transcriptome-wide mapping and high prevalence have made it one of the most studied among the more than 100 diverse RNA modifications.[3] Furthermore, identification of m^6A RNA regulatory proteins, including readers (RNA binding proteins), writers (methyltransferases), and erasers (demethylases), indicates a dynamic regulation in response to cellular stress.[4-6]

Methyltransferase like 3 (METTL3), methyltransferase like 14 (METTL14), and Wilms tumor-1 associated protein (WTAP) make up the methyltransferase complex responsible for methylating the N^6 position of adenosine of mRNA.[6-8] Depletion of components of the methyltransferase complex, such as METTL3 and METTL14, causes delay in differentiation in embryonic stem cells.[9]

The Fe^{2+} and 2-oxoglutarate-dependent dioxygenases are a family of demethylases responsible for oxidatively removing methylations in DNA, RNA, and proteins.[4, 10, 11] Fat-mass and obesity associated protein (FTO) and ALKBH5 are the known m^6A demethylases for mRNA.[12, 13] In addition, FTO was shown to demethylate N^6,2'-*O*-dimethyladenosine (m^6A_m) commonly found at the first nucleotide after the 5'-cap in mRNA.[14] Depletion and knockout studies for FTO and ALKBH5 have revealed potential functions of m^6A. For example, ALKBH5 depletion resulted in an increase in mRNA nuclear export,[15] and FTO

knockout led to increases in m^6A level and elevated binding of mRNA to a splicing factor, SRSF2.[16]

Our understanding about the roles of m^6A and the other chemical modifications in RNA in disease pathology, especially in host-viral interaction, is limited. However, previous investigations have identified m^6A in viral transcripts affecting replication through the use of host epitranscriptome machinery.[17] For example, HIV-1 viral transcripts require the presence of m^6A for proper packaging and export of viral RNA.[17, 18] Therefore, methylated adenosine in RNA may have a role in disease progression and pathology. Here, we explored the interplay of RNA modifications, especially m^6A and m^6A_m, with HIV-1 and SARS-CoV-2 infections.

Experimental Section

Global profiling of the viral and host epitranscriptomes by LC-MS/MS/MS analysis

Viral RNAs were purified from infected Vero cells. The viral genome was treated with DNaseI to remove residual DNA, concentrated and quantified. mRNA (50 ng) was incubated at 37°C for 2 hrs with 0.5 unit of nuclease P1 in a 25 μL buffer containing 25 mM NaCl and 2.5 mM $ZnCl_2$. To the resulting mixture were subsequently added alkaline phosphatase (0.25 unit) and 3 μL of 1.0 M NH_4HCO_3 buffer. After incubation at 37°C for an additional 2 hr, the digestion mixture was dried by a Speed-vac and reconstituted in 100 μL of ddH_2O.

Profiling and quantifications of modified ribonucleosides were conducted by employing a nano-flow liquid chromatography-nanoelectrospray ionization-multi-stage

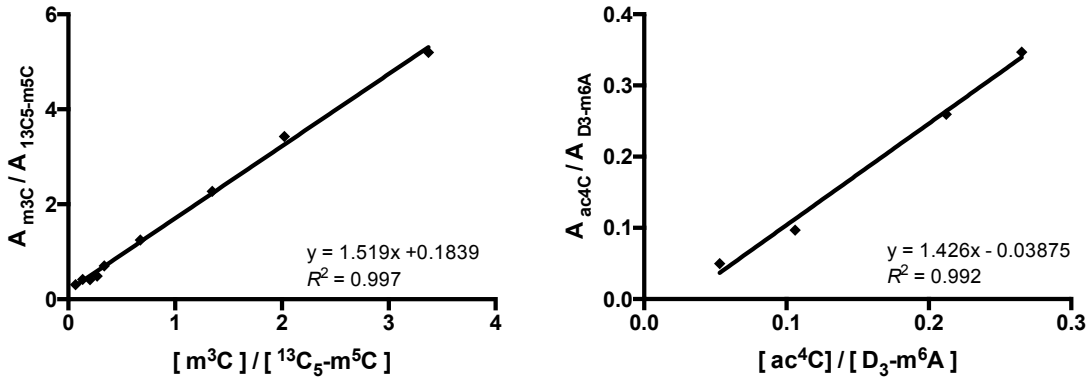


Figure 6.1 Calibration curves for 3-methylcytidine (m^3C) and *N*^{4}-acetylcytidine (ac^4C). Mixture containing 5.16 fmol of $^{13}C_5$-5-methylcytosine isotope labeled standard with unlabeled m^3C standard ranged from 3.48 – 17.4 fmol were analyzed to construct m^3C calibration curve. For the calibration curve of ac^4C, 42.5 fmol of D^3-m^6A internal standard was mixed with unlabeled ac^4C standard ranged from 2.25 – 11.27 fmol for LC-MS/MS/MS analyses.

mass spectrometry (nLC-nESI-MS/MS/MS) method described in Chapter 2.[19] In short, to 10 ng of digested viral RNA were added the following stable isotope-labeled standard nucleosides: 8 pmol [$^{13}C_5$]-adenosine, 42.5 fmol D^3-*N*^{6}- methyladenosine, 17.25 fmol of [$^{13}C_5$]-2'-*O*-methyladenosine, 9.7 fmol [$^{13}C_5$]-2'-*O*- methylcytidine, 3.3 pmol [$^{15}N_3$]-cytidine, 5.5 pmol of [$^{15}N_2$]-uridine, 5.1 fmol [$^{13}C_5$]-5-methylcytidine, and 5.6 fmol [$^{13}C_5$]-5-methyluridine. The enzymes in the digestion mixture were then removed by chloroform extraction. The aqueous layer was subsequently dried by a Speed-vac, and reconstituted in 100 μL of ddH_2O:acetonitrile (1:9 v/v), and then incubated at -20°C for 20 min to precipitate residual salts in the digestion mixture. The mixture was centrifuged, and the supernatant was again collected, dried by Speed-vac, reconstituted in water, and subjected to LC-MS/MS and MS/MS/MS analyses.

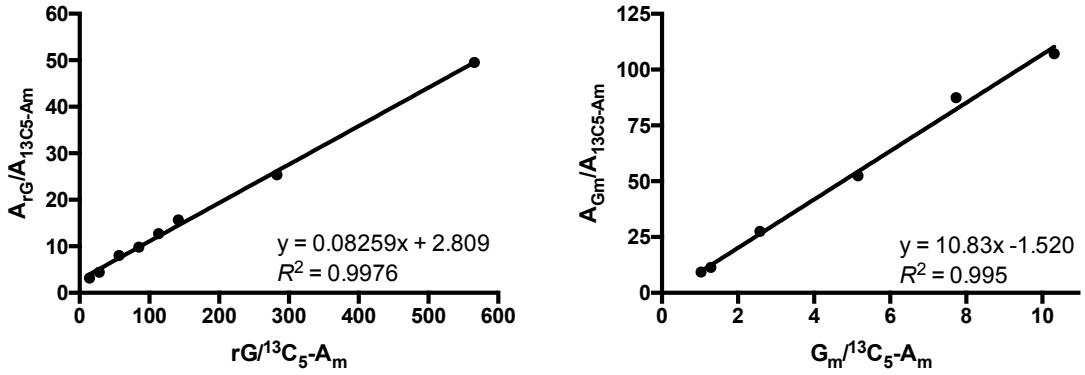


Figure 6.2 Calibration curves for guanosine and 2'-*O*-methylguanosine (G_m). Mixture containing 17.25 fmol of $^{13}C_5$-2'-*O*-methyladenosine ($^{13}C_5$-Am) isotope labeled standard with unlabeled standard ranged from 0.244 – 9.76 pmol and 17.79 – 177.92 fmol of rG and G_m, respectively, were analyzed to construct calibration curves.

nLC-nESI-MS/MS/MS experiments in scheduled in SRM mode were performed on an LTQ-XL linear ion trap mass spectrometer coupled with an EASY-nLC II (Thermo Fisher Scientific, San Jose, CA). The mass spectrometer was operated in the positive-ion mode, with the electrospray, capillary, and tube lens voltages being 2.0 kV, 12 V, and 100 V, respectively. Samples were loaded onto a 5 cm in-house packed porous graphitic carbon (PGC, 5 μm in particle size, Thermo Fisher Scientific) trapping column (150 μm i.d). Analytes were eluted onto a 20-cm Zorbax SB-C18 (5 μm in particle size, 100 Å in pore size, Michrom BioResources) analytical column (75 μm i.d.) at a flow rate of 300 nL/min. The gradient employed was 0%–16% B in 5 min, 16%–22% B in 23 min, 22%–50% B in 17 min, 50%–90% B in 5 min, and finally 30 min at 90% B, using 0.1% (v/v) formic acid in water and 0.1% (v/v) formic acid in acetonitrile as mobile phases A and B, respectively.

A full list of transitions are found in **Table 2.2**. Isotopically labeled standards were not available for some of the modified nucleosides. Hence, we employed stable isotope-

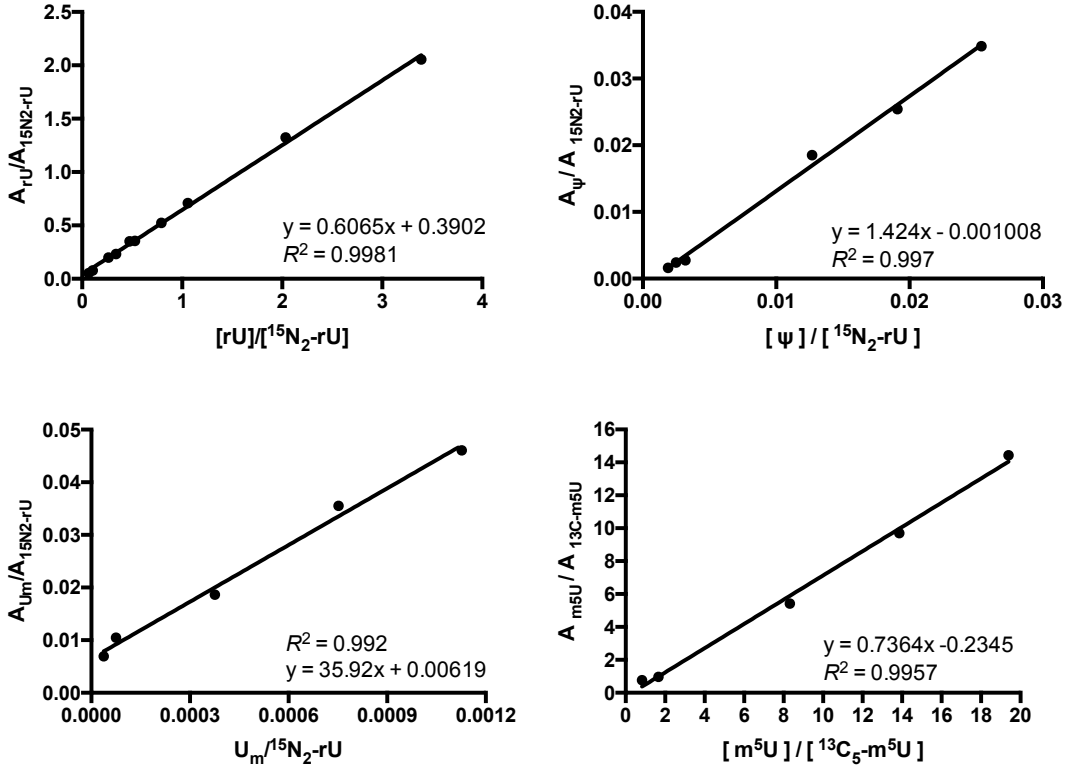


Figure 6.3 Calibration curves for uridine, 5-methylcytosine (m^5U) and pseudouridine (ψ). A mixture containing 5590 pmol of $^{15}N_2$-uridine isotope labeled standard with unlabeled standard ranged from 0.37 – 18.95 pmol, 10.67 – 142.24 fmol, 0.21 – 4.20 fmol for uridine ψ, and U_m respectively, were analyzed to construct calibration curves. For the calibration curve of m^5U, 5.6 fmol of $^{13}C_5$-m^5U internal standard were mixed with unlabeled standard ranged from 4.62 – 13.40 fmol for LC-MS/MS/MS analyses.

labeled forms of other nucleosides, which exhibited similar elution times as the modified nucleosides, as surrogate standards for the quantification (**Figure 6.1-6.5**). The calibration curves for rA, A_m, m^6A, rC, C_m, and m^5C were reported previously.[20]

LC-MS/MS/MS quantification for rA and m^6A

Enzymatic digestion was preformed using previous method.[21] Briefly, 50 ng of mRNA was digested by 0.5 unit of nuclease P1 (NP1) in 25 μL solution of 25 mM NaCl and 2.5 mM $ZnCl_2$, and incubated at 37°C for 2 hr. To the resulting mixture were added

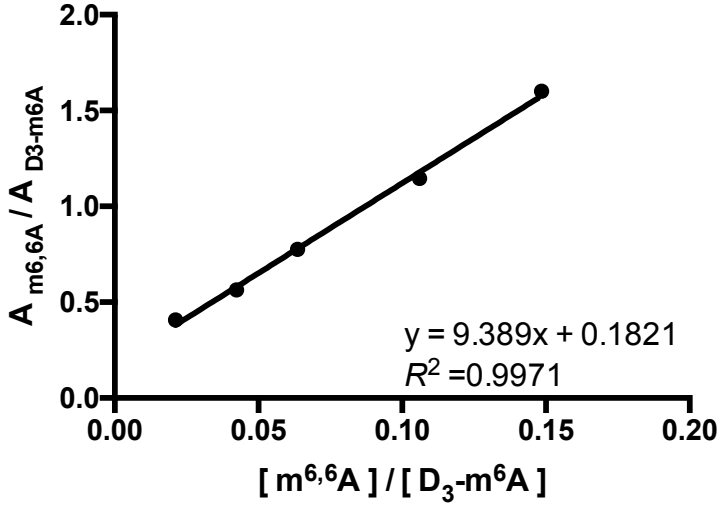


Figure 6.4 Calibration curves for *N*^{6,6}-dimethyladenosine (m^{6,6}A). For the calibration curve of m^{6,6}A, 42.5 fmol of D^3-m^{6}A internal standard were mixted with unlabeled m^{6,6}A standard ranged from 0.90 – 4.51 fmol were analyzed for LC-MS/MS/MS analyses.

	Transitions	Isolation Width		NCE	
		MS^2	MS^3	MS^2	MS^3
rA (MS/MS)	268 → 136	3	2	35	
[$^{13}C_5$]-rA	273 → 136	3	2	35	
m^{6}A	282 → 150 → 94	3	2	43	37
[D_3]-m^{6}A	285 → 150 → 94	3	2	43	37
m^6A_m	296 → 150 → 94	3	2	37	35

Table 6.1. A list of precursor ions for ribonucleosides monitored in MS^3, isolation width, and normalized collision energy (NCE).

0.25 units alkaline phosphatase (CIP) and 3 μL of 1M NH_4HCO_3 buffer. After incubating at 37°C for an additional 2 hr, the digestion mixture was dried by a Speed-vac and reconstituted in 100 μL of ddH_2O.

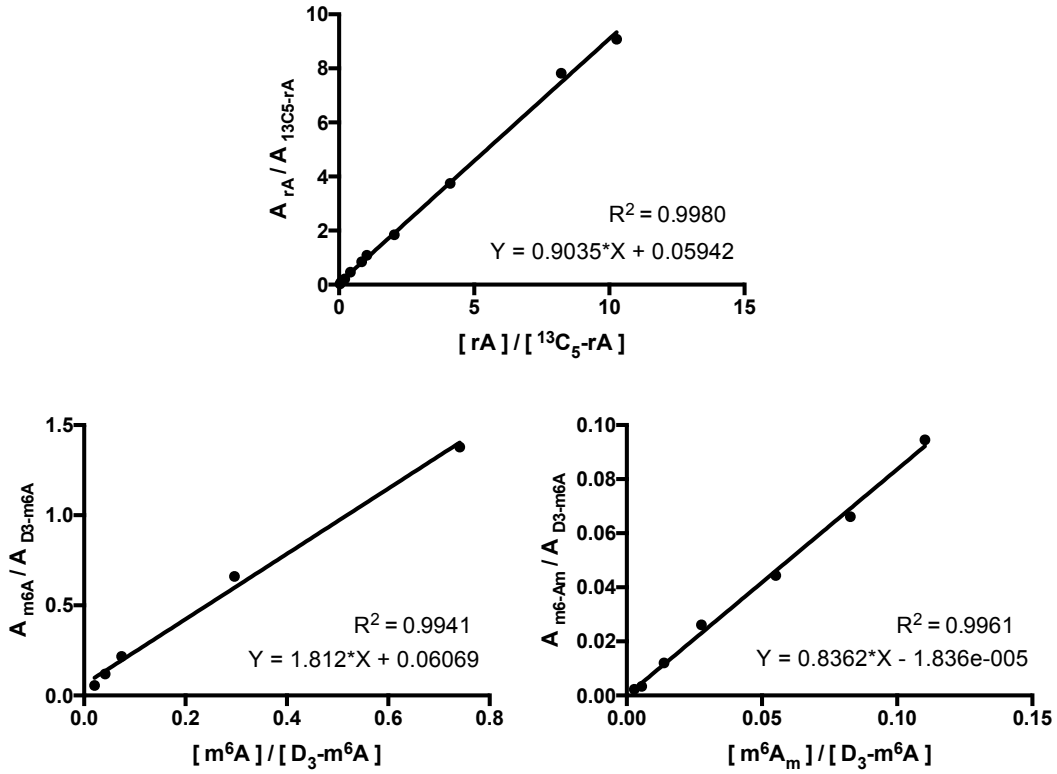


Figure 6.5 Calibration curves for Adenosine, N^6-methyl-2'-O-methyladenosine (m^6A_m) and N^6-methyladenosine (m^6A). For the calibration curve of adenosine, 8 pmol of $^{13}C_5$-labeled adenosine was added unlabeled adenosine standard ranging from 105.8 fmol - 32 pmol were analyzed. For the calibration of m^6A_m and m^6A, 85 fmol of D_3-N^6-methyladenosine (D^3-m^6A) internal standard were mixed with m^6A_m and m^6A unlabeled standards ranging from 0.45 – 9.4 fmol and 1.8 – 63 fmol, respectively, were analyzed.

The profiling and quantification of modified ribonucleoside by tandem mass spectrometry were conducted following a previously published procedure with slight modifications.[2] In short, to 10 ng of digested RNA were added to stable isotope-labeled standards of 8 pmol of $^{13}C_5$-labeled adenosine, 42.5 fmol of D_3-N^6-methyladenosine. All enzymes from the digestion mixture was removed by chloroform extraction. The aqueous layer was dried and reconstituted in 10 μL of doubly distilled H_2O and 90 μL of acetonitrile to remove precipitates in that may block the nano-flow columns. This mixture was then

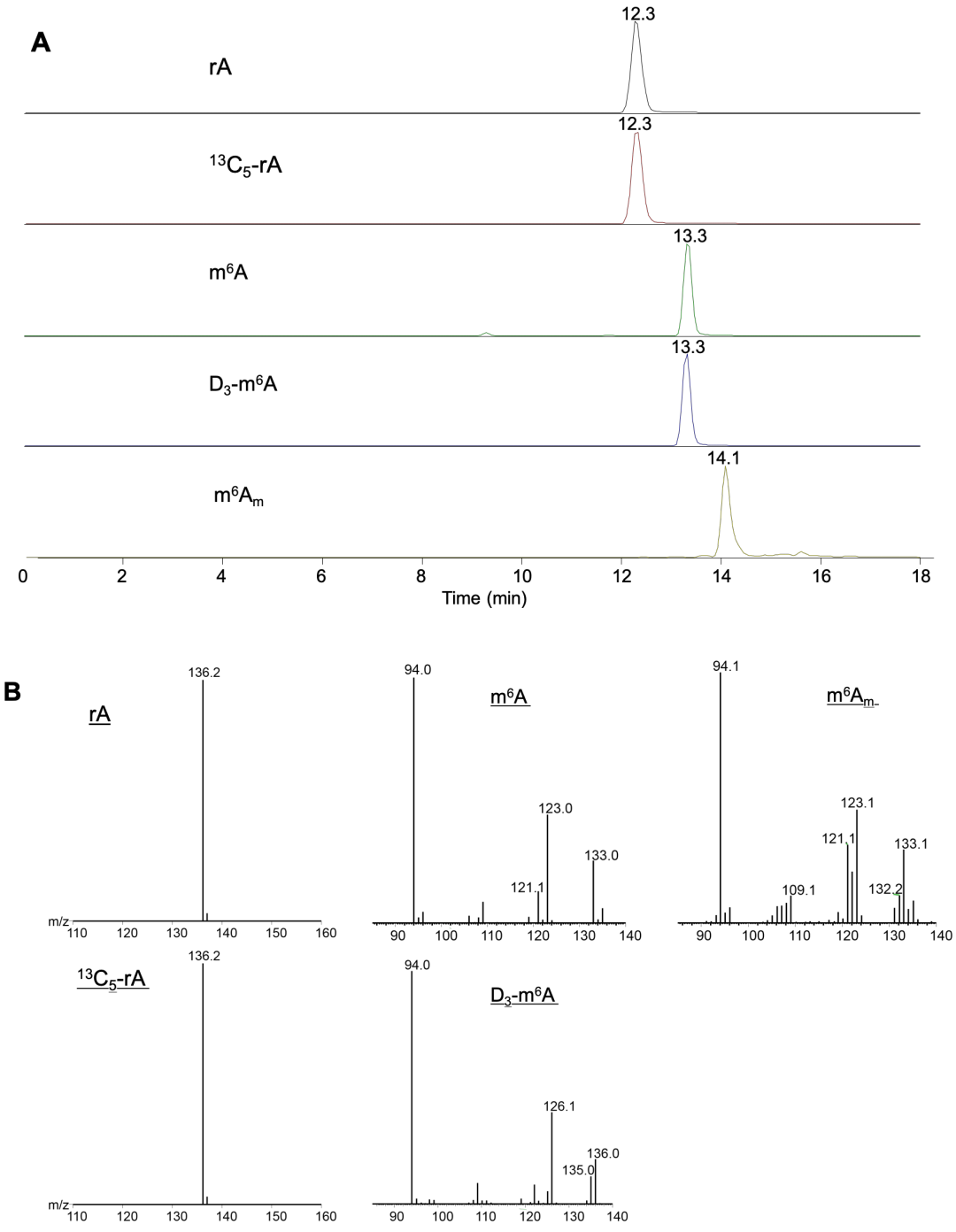


Figure 6.6: Representative A) SICs and B) MS/MS for rA and $^{13}C_5$-rA, and MS/MS/MS for m^{6}A, m^{6}A$_m$ and D^3-m^{6}A.

incubated at -20°C for 20 min, centrifuged and the supernatant was collected and dried using a speed vac.

SRM experiments were performed on an LTQ-Orbitrap Velos mass spectrometer coupled with an EASY-nLC II (Thermo Fisher Scientific, San Jose, CA). The mass spectrometer was operated in the positive-ion mode with the electrospray, capillary, and tube lens voltages being 2.0 kV, 12 V, and 100 V, respectively. SRM transitions are found in **Table 6.1**. Samples were loaded onto a 5 cm in-house packed Magic C18-AQ pre-column (150 μm i.d, 5 μm beads, 100 Å in pore size; Michrom BioResources, Auburn, CA). Analytes were resolved onto a 20-cm Magic C18-AQ analytical column (75 μm i.d.) at a flow rate of 300 nL/min. The gradient employed was 0-40% B in 20 minutes, 40-90% B in 5 min, and finally 5 min at 90% B using 0.1% (v/v) formic acid in water and 0.1% (v/v) formic acid in acetonitrile as mobile phases A and B, respectively.

For calibration curve construction, we employed stable isotope-labeled standards that eluted at similar times as the nucleoside of interest. rA, m^6A and m^6A_m calibration curves employed in this analysis are found in **Figure 6.5**. Representative fragmentation pattern and chromatogram for m^6A_m is found in **Figure 6.6**.

Results and Discussion

HIV-1 viral genome contains modified ribonucleosides and infection results in aberrant epitranscriptome level.

Recent evidence has suggested that RNA modifications play a role in host-viral interactions.[17, 18] LC-MS/MS/MS analysis was conducted to identify if modified ribonucleoside levels are altered upon HIV-LA1 (HIV-1) infection in the mRNA of MT4

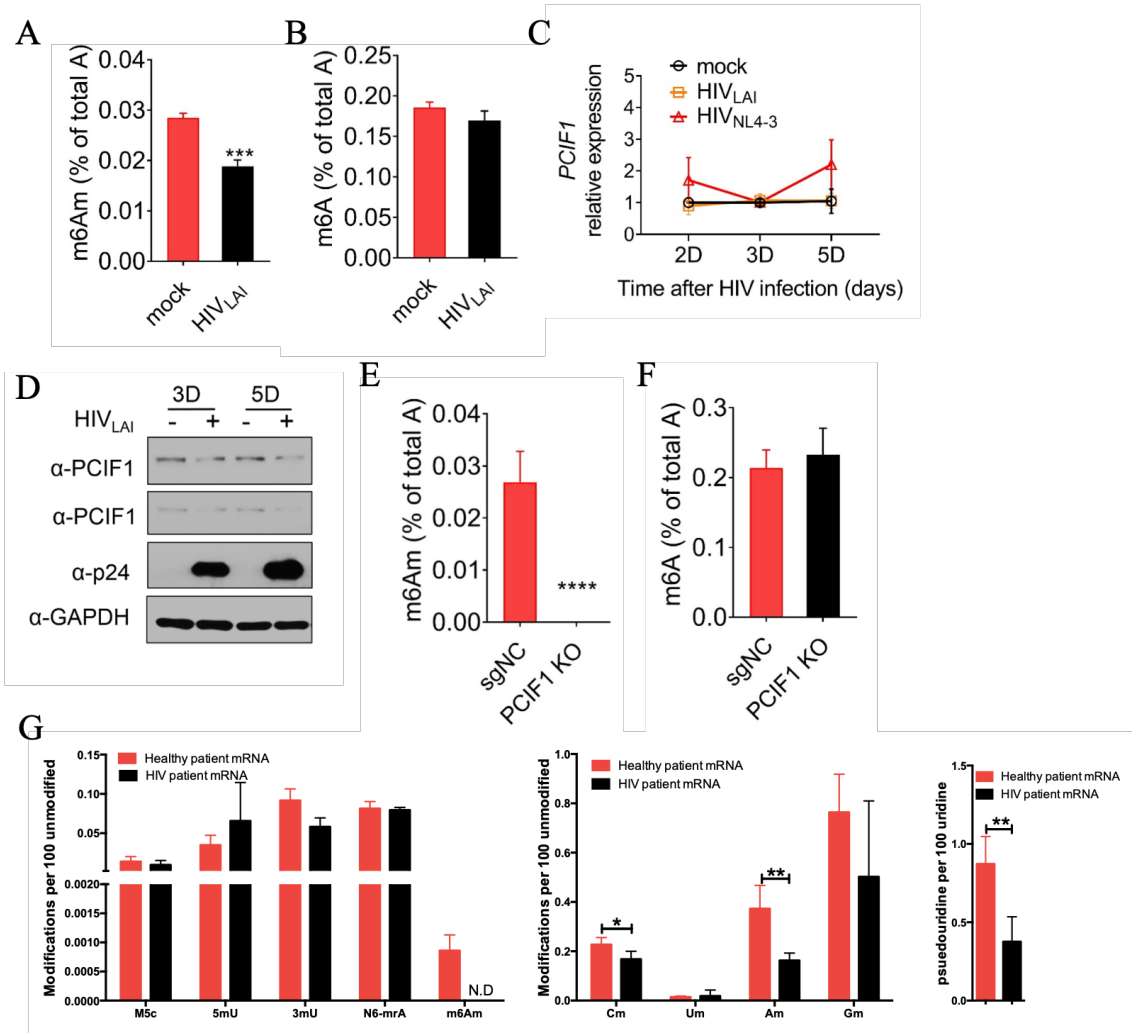


Figure 6.7 HIV infection results in diminished m^6A_m host by PCIF1 and host epitranscriptome level A) and B) represents LC-MS/MS/MS analysis of mock and HIV-LAI infected MT4 cells at a dose of 0.4 MOI for 3 days. Mean ± SD of 3 biological replicates. Statistical analysis was conducted using an unpaired student t test. PCIF1 mRNA (C) and protein (D) levels were taken on day 2, 3, and 5 after infection (n = 3). Figure D) top represents high exposure, while D) bottom represents PCIF at low exposure. E) and F) represents the m^6A_m and m^6A level in PCIF1 knockout in Jurkat T cells. G) represents epitranscriptome analysis of PBMC cells extracted from human healthy (n = 4) and HIV infected (n = 5) patients. N.D means not detectable. Statistical analysis was conducted using 3 biological replicates with unpaired student t test.

human T cells. The analysis revealed that m^6A_m level in HIV-1-infected MT4 cells was lower than in the mock infection (**Figure 6.7A**). In addition, N^6-methyladenosine (m^6A) level was not affected by HIV-1 infection (**Figure 6.7B**). The results suggest that m^6A_m regulatory proteins may be responsible for aberrant m^6A_m level. Phosphorylated CTD Interacting Factor 1 (PCIF1) is the known N^6 methyltransferase for capped m^6A_m.[22] However, PCIF1 does not influence internal mRNA m^6A level.[22] Moreover, the results corroborate that of Boulias (2019), in that PCIF1 CRISPR knock-out in Jurkat T cells resulted in undetectable m^6A_m level but m^6A level was unaltered (**Figure 6.7E and 6.7F**). In that vein, the m^6A_m and m^6A results suggest that PCIF1 activity is attenuated upon HIV-1 infection. Q-PCR analysis was then conducted on MT4 cells infected with either HIV-LAI and HIV-NL4-3 strains to determine if PCIF1 transcript level was altered. However, PCIF1 mRNA levels are unaffected upon HIV-LAI or HIV-NL4-3 strain infection (**Figure 6.7C**). Nonetheless, Western blot experiments suggest a decrease in PCIF1 level in HIV-LAI in a dose-dependent manner (**Figure 6.7D**). Viral infection was confirmed with p24, a known HIV antigen.[17] Taken together, the results indicate that viral infection by HIV-1 influences host m^6A_m level by altering PCIF1 protein level.

I also evaluated the epitranscriptomic modifications in mRNA samples of human patients infected with HIV. The LC-MS/MS/MS results revealed a significant decrease in ψ level in PBMC cells of HIV infected patients compared to healthy controls (**Figure 6.7G**). In addition, 2'-*O*-methylations (C_m and A_m) were statistically significantly reduced in HIV patient RNA and a decreasing trend was observed G_m (**Figure 6.7G**). The results suggest that HIV infection influences PCIF1 activity. In addition, another protein worth

investigating is CMTr1 and CMTr2, human 2'-*O* methyltransferases for the first nucleotide and internal 2'-*O* methylations, respectively.[23] CMTr1 activity may also be altered, resulting in reduced 2'-*O* methylations at the first nucleotide after the 5'-cap, and PCIF1 is unable to react on unmethylated 2'-*O*-ribose.[22, 24] Therefore, HIV infection results in undetectable levels of m^6A_m to hinder cap-dependent translation of host mRNA. Another possibility is that the activity of CMTr2, the internal mRNA 2'-*O*-methyltransferase, is altered, resulting in reduced global 2'-*O*-methylations. However, CMTr2 is not responsible for methylation of the first nucleotide after the cap. Further investigations are needed for the elucidation of the mechanism by measuring PCIF1 and CMTr1/2 transcript and protein levels in host cells upon HIV infection.

HIV-1 and SARS-CoV-2 viral genome contains modified ribonucleosides

Previous reports suggest that HIV genomic RNA utilizes host methyltransferases and demethylases to mediate their own modified ribonucleoside level.[17] Therefore, I monitored possible modified ribonucleosides that may be present in HIV and SARS-CoV-2 transcripts. Our LC-MS/MS/MS results showed that, among the 31 ribonucleosides monitored, the 4 canonical ribonucleosides and 9 modified ribonucleosides were identified in HIV genomic RNA, including 2'-*O*-methylations (A_m, C_m, and U_m), m^6A, 5-methylcytidine (m^5C), *N*3-methylcytidine (m^3C), N^4-acetylcytidine, m^6A, pseudouridine (ψ), and 5-methyluridine (m^5U) (**Figure 6.8A**).

The recent coronavirus pandemic has revealed a need to understand better host-viral interactions, especially in contagious viruses. To understand if viruses in general utilize the host epitranscriptome regulators to modify its own transcripts, I preformed LC-

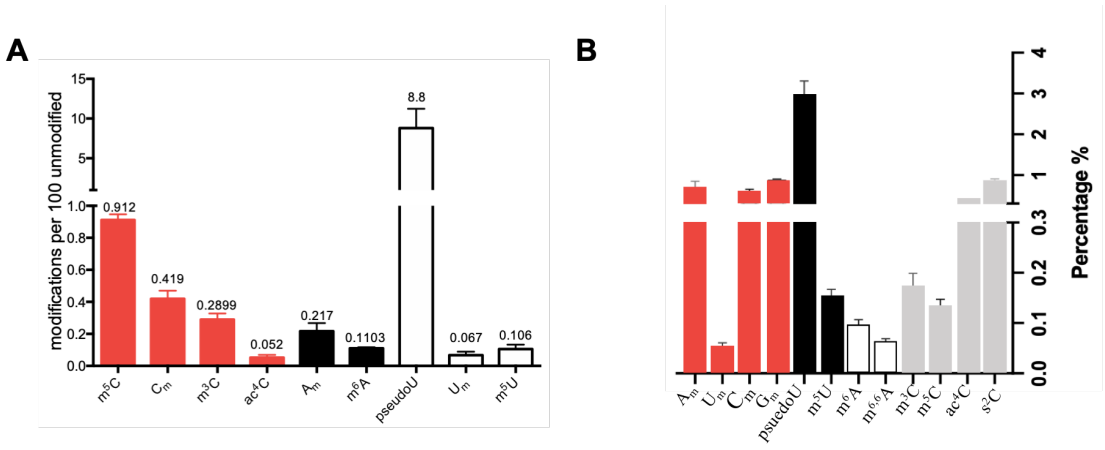


Figure 6.8 LC-MS/MS/MS analysis of HIV-1 and SARS-CoV-2 viral genome. The data represents the global modified ribonucleoside profile in A) HIV and B) SARS-CoV-2 genomic RNA from three independent digestions.

MS/MS/MS analysis on SARS-CoV-2 (COVID-19) viral RNA genome (**Figure 6.8B**). The analysis revealed the presence of 12 modified ribonucleosides on the viral genome, including 2'-O-methylations (A_m, C_m, U_m, G_m), ψ, m^5U, m^6A, N^6-dimethyladenosine ($m^{6,6}A$), m^3C, m^5C, ac^4C, and 2-thiocytidine (s^2C) (**Figure 6.8B**). All the modifications identified in HIV-1 viral genome were also identified in SARS-CoV-2 transcripts (**Figure 6.8A and 6.8B**). Some similar trends appeared in the two viral RNA analysis, e.g., ψ is the most abundant modification in both HIV-1 and SARS-CoV-2 genomes (**Figure 6.8A and 6.8B**). Second, m^6A_m was not detected in either virus. In addition, m^6A level was very similar between the two viruses at a level of 0.110 and 0.096 modifications per 100 adenosines for HIV-1 and SARS-CoV-2, respectively. In contrast, m^5C level was significantly higher in the HIV-1 genome compared to SARS-CoV-2 genome at a level of 0.912 and 0.134 modifications per 100 cytidines, respectively. The results may indicate a

different mechanism of replication and manipulation of host epitranscriptomic regulatory proteins.

Conclusions

Upon viral infection of Zika, West Nile, and Dengue, the viral genome must evade detection from host viral defense machinery.[25] Viruses utilize host cell machinery to replicate, package and expel viral components vital to its'spread.[25] RNA modifications on viral genomes play important role in evading host viral defense.[26] Therefore, it is necessary for the virus to manipulate the host epitranscriptome machinery to disguise viral transcripts as host transcripts.[17, 18, 26, 27] Previous experiments using Nanopore direct sequencing suggests the presence of modified RNA in SARS-CoV-2, but no specific modifications were identified.[28] In this study, we identified 9 and 12 modified ribonucleosides in HIV-1 and SARS-CoV-2 viral RNA genomes, respectively. Among those identified only two, m^6A and ac^4C, are linked HIV virus propagations and stability, respectively.[17, 29] Future investigations are necessary to understand the functional roles of the other modifications, including pseudouridine which occurs at high levels in both HIV-1 and SARS-CoV-2 viral genomes.

In addition, we also identified a potential shared pathway between HIV-1 and SARS-CoV-2 through modulation of m^6A_m levels of viral RNA transcripts. In the case of HIV-1 infection, PCIF1 protein level is attenuated, thereby reducing m^6A_m required for efficient cap-dependent translation. This would suggest that other protein synthesis mechanisms are utilized, including cap-independent translation that requires m^6A. Overall,

the evidence presented suggests epitranscriptome regulator proteins, such as methyltransferases, play an important role in viral-host interactions.

References

1. Su, D.; Chan, C.; Gu, C.; Lim, K.; Chionh, Y.; McBee, M.; Russell, B.; Babu, I.; Begley, T.; Dedon, P., Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nat. Prot.* **2014,** *9* (4), 828-841.

2. Basanta-Sanchez, M.; Temple, S.; Ansari, S. A.; D'Amico, A.; Agris, P. F., Attomole quantification and global profile of RNA modifications: Epitranscriptome of human neural stem cells. *Nucleic Acids Res.* **2015**, gkv971-gkv971.

3. Meyer, K. D.; Saletore, Y.; Zumbo, P.; Elemento, O.; Mason, C. E.; Jaffrey, S. R., Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **2012,** *149* (7), 1635-1646.

4. Fu, Y.; Jia, G.; Pang, X.; Wang, R. N.; Wang, X.; Li, C. J.; Smemo, S.; Dai, Q.; Bailey, K. A.; Nobrega, M. A.; Han, K.-L.; Cui, Q.; He, C., FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nature Communications* **2013,** *4*, 1798-1798.

5. Li, F.; Zhao, D.; Wu, J.; Shi, Y., Structure of the YTH domain of human YTHDF2 in complex with an m6A mononucleotide reveals an aromatic cage for m6A recognition. *Cell Research* **2014,** *24*, 1490-1492.

6. Liu, J.; Yue, Y.; Han, D.; Wang, X.; Fu, Y.; Zhang, L.; Jia, G.; Yu, M.; Lu, Z.; Deng, X.; Dai, Q.; Chen, W.; He, C., A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol* **2014,** *10* (2), 93-5.

7. Ping, X.-L.; Sun, B.-F.; Wang, L.; Xiao, W.; Yang, X.; Wang, W.-J.; Adhikari, S.; Shi, Y.; Lv, Y.; Chen, Y.-S.; Zhao, X.; Li, A.; Yang, Y.; Dahal, U.; Lou, X.-M.; Liu, X.; Huang, J.; Yuan, W.-P.; Zhu, X.-F.; Cheng, T.; Zhao, Y.-L.; Wang, X.; Rendtlew Danielsen, J. M.; Liu, F.; Yang, Y.-G., Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell research* **2014,** *24* (2), 177-89.

8. Wang, P.; Doxtader, K. A.; Nam, Y., Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. *Molecular Cell* **2016,** *63* (2), 306-317.

9. Gerula, M. A. A.; Kol, N.; Hershkovitz, V.; Peer, E.; Mor, N.; Manor, Y. S.; Ben-haim, M. S.; Eyal, E.; Yunger, S.; Pinto, Y.; Jaitin, D. A.; Viukov, S.; Rais, Y.; Krupalnik, V.; Chomsky, E.; Zerbib, M.; Maza, I.; Rechavi, Y.; Massarwa, R.; Hanna, S.; Amit, I.; Levanon, E. Y.; Amariglio, N.; Stern-ginossar, N.; Novershtern, N.; Rechavi, G.; Hanna, J. H., m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Sci.* **2015,** *347* (6225), 1002-1006.

10. Alemu, E.; He, C.; Klungland, A., ALKBHs-facilitated RNA modifications and de-modifications. *DNA Repair* **2016,** *44*, 87-91.

11. Duncan, T.; Trewick, S. C.; Koivisto, P.; Bates, P. A.; Lindahl, T.; Sedgwick, B., Reversal of DNA alkylation damage by two human dioxygenases. *Proceedings of the National Academy of Sciences* **2002,** *99* (26), 16660-16665.

12. Fu, Y.; Jia, G.; Pang, X.; Wang, R. N.; Wang, X.; Li, C. J.; Smemo, S.; Dai, Q.; Bailey, K. a.; Nobrega, M. a.; Han, K.-l., FTO-Mediated Formation of N6-Hydorxylmethyladensoine and N6-formyladenosine in mammalian RNA. *Nat. Comm.* **2013,** (Md), 1-14.

13. Zheng, G.; He, C., RNA demethylation by FTO and ALKBH5. *2-Oxoglutarate-dependent oxygenases* **2015,** (3), 263-274.

14. Mauer, J.; Luo, X.; Blanjoie, A.; Jiao, X.; Grozhik, A. V.; Patil, D. P.; Linder, B.; Pickering, B. F.; Vasseur, J.-J.; Chen, Q.; Gross, S. S.; Elemento, O.; Debart, F.; Kiledjian, M.; Jaffrey, S. R., Reversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nature* **2017,** *541* (371-375).

15. Zheng, G.; Dahl, J. A.; Niu, Y.; Fedorcsak, P.; Huang, C. M.; Li, C. J.; Vågbø, C. B.; Shi, Y.; Wang, W. L.; Song, S. H.; Lu, Z.; Bosmans, R. P. G.; Dai, Q.; Hao, Y. J.; Yang, X.; Zhao, W. M.; Tong, W. M.; Wang, X. J.; Bogdan, F.; Furu, K.; Fu, Y.; Jia, G.; Zhao, X.; Liu, J.; Krokan, H. E.; Klungland, A.; Yang, Y. G.; He, C., ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Molecular Cell* **2013,** *49*, 18-29.

16. Zhao, X.; Yang, Y.; Sun, B.-F.; Shi, Y.; Yang, X.; Xiao, W.; Hao, Y.-J.; Ping, X.-L.; Chen, Y.-S.; Wang, W.-J.; Jin, K.-X.; Wang, X.; Huang, C.-M.; Fu, Y.; Ge, X.-M.; Song, S.-H.; Jeong, H. S.; Yanagisawa, H.; Niu, Y.; Jia, G.-F.; Wu, W.; Tong, W.-M.; Okamoto, A.; He, C.; Danielsen, J. M. R.; Wang, X.-J.; Yang, Y.-G., FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Research* **2014,** *24* (12), 1403-1419.

17. Lichinchi, G.; Zhao, Boxuan S.; Wu, Y.; Lu, Z.; Qin, Y.; He, C.; Rana, Tariq M., Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. 2016; Vol. 20, pp 666-673.

18. Kennedy, E. M.; Bogerd, H. P.; Kornepati, A. V. R.; Kang, D.; Ghoshal, D.; Marshall, J. B.; Poling, B. C.; Tsai, K.; Gokhale, N. S.; Horner, S. M.; Cullen, B. R., Posttranscriptional m6A Editing of HIV-1 mRNAs Enhances Viral Gene Expression. *Cell Host and Microbe* **2016,** *19* (5), 675-685.

19. Gonzalez, G.; Cui, Y.; Wang, P.; Wang, Y., Normalized retention time for scheduled liquid chromatography-multistage mass spectrometry analysis of epitranscriptomic modifications. *Journal of Chromatography A* **2020,** *1623*, 461181.

20. Fu, L.; Amato, N. J.; Wang, P.; McGowan, S. J.; Niedernhofer, L. J.; Wang, Y., Simultaneous Quantification of Methylated Cytidine and Adenosine in Cellular and Tissue RNA by Nano-Flow Liquid Chromatography–Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal. Chem.* **2015,** *87ha* (15), 7653-7659.

21. Wang, X.; Lu, Z.; Gomez, A.; Hon, G. C.; Yue, Y.; Han, D.; Fu, Y.; Parisien, M.; Dai, Q.; Jia, G.; Ren, B.; Pan, T.; He, C., N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014,** *505* (7481), 117-20.

22. Boulias, K.; Tocydlowska-Socha, D.; Hawley, B. R.; Liberman-Isakov, N.; Takashima, K.; Zaccara, S.; Guez, T.; Vasseur, J.-J.; Debart, F.; Aravind, L.; Jaffrey, S. R.; Greer, E. L., Identification of the m6A methyltransferase PCIF1 reveals the location and functions of m6Am in the transcriptome. *Molecular Cell* **2019,** *75* (3), 631-643.

23. Smietanski, M.; Werner, M.; Purta, E.; Kaminska, K. H.; Stepinski, J.; Darzynkiewicz, E.; Nowotny, M.; Bujnicki, J. M., Structural analysis of human 2′-O-ribose methyltransferases involved in mRNA cap structure formation. *Nature Communications* **2014,** *5*, 3004-3004.

24. Akichika, S.; Hirano, S.; Shichino, Y.; Suzuki, T.; Nishimasu, H.; Ishitani, R.; Sugita, A.; Hirose, Y.; Iwasaki, S.; Nureki, O.; Suzuki, T., Cap-specific terminal N 6 -methylation of RNA by an RNA polymerase II–associated methyltransferase. *Science* **2019,** *363* (6423).

25. Whitaker-Dowling, P.; Younger, J. S., Virus-Host Cell Interactions. *Academic Press* **1999**, 1957-1961.

26. Gokhale, N. S.; McIntyre, A. B. R.; McFadden, M. J.; Roder, A. E.; Kennedy, E. M.; Gandara, J. A.; Hopcraft, S. E.; Quicke, K. M.; Vazquez, C.; Willer, J.; Ilkayeva, O. R.; Law, B. A.; Holley, C. L.; Garcia-Blanco, M. A.; Evans, M. J.; Suthar, M. S.; Bradrick, S. S.; Mason, C. E.; Horner, S. M., N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection. *Cell Host and Microbe* **2016,** *20* (5), 654-665.

27. Lv, L.; Wang, Q.; Xu, Y.; Tsao, L. C.; Nakagawa, T.; Guo, H.; Su, L.; Xiong, Y., Vpr Targets TET2 for Degradation by CRL4 VprBP E3 Ligase to Sustain IL-6 Expression and Enhance HIV-1 Replication. *Molecular Cell* **2018,** *70* (5), 961-970.

28. Miladi, M.; Fuchs, J.; Maier, W.; Weigang, S.; Díaz i Pedrosa, N.; Weiss, L.; Lother, A.; Nekrutenko, A.; Ruzsics, Z.; Panning, M.; Kochs, G.; Gilsbach, R.; Grüning, B., The landscape of SARS-CoV-2 RNA modifications. *bioRxiv* **2020**, 1-22.

29. Tsai, K.; Jaguva Vasudevan, A. A.; Martinez Campos, C.; Emery, A.; Swanstrom, R.; Cullen, B. R., Acetylation of Cytidine Residues Boosts HIV-1 Gene Expression by Increasing Viral RNA Stability. *Cell Host and Microbe* **2020,** *28* (2), 306-312.

Chapter 7: Concluding Remarks and Future Directions

Our understanding of RNA biology is limited. However, new technologies have expanded our ability to investigate the epitranscriptome, including sequencing- and mass spectrometry-based techniques. In addition, recent studies have elucidated the regulatory proteins, including readers, writers, and erasers, for a select few modified ribonucleosides, demonstrating their dynamic regulation. Dysregulation and aberrant modified ribonucleoside levels have been linked to disease. Thus, there is a need of sensitive methods to monitor the epitranscriptome and investigations into the dynamic regulation and its influence in disease pathology.

In this dissertation, I started in Chapter 2 by developing a method to monitor 27 modified and 4 canonical ribonucleosides in a single injection using nano-flow liquid chromatography-nanoelectrospray ionization-multi-stage mass spectrometry (nLC-nESI-MS/MS/MS). By employing a commonly used concept in proteomics, i.e., normalized retention time (iRT), I assigned an iRT value to each ribonucleoside based on their respective retention time for a scheduled SRM method. The iRT scores enable accurate predictions of retention times for modified ribonucleosides based on the elution times of unmodified nucleobases. This high-throughput method was compared with a low-throughput method and, even though the method contains more transitions, the precision and accuracy in the quantifications of 5-methycytidine (m^5C) and 2'-*O*-methylcytidine were not compromised. When applied to total RNA samples isolated from HEK293T cells, I identified the presence of 20 modified ribonucleosides with a 2.5 ng injection. Future studies will expand the number of modified ribonucleosides that are monitored in the

method. In this vein, there are over 100 chemically modified ribonucleosides. By purchasing and synthesizing additional modified ribonucleosides, we can expand the method for a better coverage of the epitranscriptome.

In Chapter 3, the binding of YTHDF2 to m^5C was assessed. Reader proteins for a majority of modified ribonucleosides have yet to be identified, including for m^5C. By utilizing a SILAC-based LC-MS/MS experiment together with a biotinylated oligonucleotide probes containing either m^5C and cytidine, we identified potential m^5C binders, including YTHDF2. Pull-down experiments using YTHDF2 recombinant protein incubated with RNA from HEK293T cells revealed high levels of m^5C in the bound fraction compared to the input and flow-through. Mutations of key amino acid residue involved in m^6A regulation, i.e., tryptophan 432, also results in diminished m^5C binding. In addition, diminished m^5C level in 18S rRNA was observed in HEK293T cells upon CRISPR-mediated genetic depletion of *YTHDF2*. For future experiments, there were other candidate reader proteins identified the m^5C SILAC affinity pulldown down experiment. Many of which have not been identified as direct binders of m^5C in RNA. Pulldown experiments using cellular RNA would corroborate the SILAC assay. Another interesting characteristic of the SILAC experiment is the identification of anti-readers, which preferentially bind the unmodified sequence compared to the m^5C-containing sequence. Anti-readers would indicate that the presence of a m^5C actually hinders its binding and function. Therefore, it would be interesting to characterize their functions in RNA biology.

In Chapter 4, I identified a novel RNA modification in mammalian and Drosophila RNA, 5-hydroxymethyluridine (5-hmrU). I developed a method relying on off-line HPLC

enrichment and nLC-MS/MS on a quadrupole-Orbitrap mass spectrometer for the sensitive detection of 5-hmrU. Our method enabled the detection of 5-hmrU in total RNA isolated from mouse tissues, cultured cancer cells, and Drosophila. Drosophila RNA contained relatively high levels of 5-hmrU. In addition, I determined if ten-eleven translocation (Tet) 1-3 proteins contribute to the formation of 5-hmrU in RNA by overexpressing catalytic domains and full length proteins of Tet1-Tet3 in HEK293T cells. Overexpression of catalytic domains of Tet1 and Tet3 resulted in increases in 5-hmrU. In addition, ectopic overexpression of full-length Tet3, but not Tet1 or Tet2, led to an increase in 5-hmrU level. In addition, genetic ablation of Tet in Drosophila head resulted in diminished 5-hmrU level. Together, Tet mediates the formation of 5-hmrU in mammalian and Drosophila RNA. To follow-up this study, 5-hmrU's role in RNA is still unknown. It will be important to explore, in the future, whether and how 5-hmrU may modulate mRNA stability and translation.

In Chapter 5, LC-MS/MS/MS analysis revealed altered N^6-methyladenosine (m^6A) level in human blood samples from Beijing truck drivers exposed to air pollution or tobacco smoke. Long-term smoking was associated with lower global m^6A level and black carbon exposure was positively associated with m^6A level. However, air pollution and tobacco smoke was not associated with known mRNA m^6A reader, writer, and eraser (RWE) proteins. The analysis was conducted using total RNA extracted from peripheral blood; therefore, evaluating the m^6A regulatory proteins in ribosomal RNA may better reflect the results obtained in the m^6A analysis. However, due to limited sample quantities, we were unable to reprocess the Q-PCR results to evaluate this hypothesis. Therefore, future

experiments would target ribosomal RNA m^6A RWE proteins, such as METTL5, an 18S rRNA methyltransferase.

In Chapter 6, global modified ribonucleoside profiling with LC-MS/MS/MS of the viral RNA genomes, including HIV-1 and SARS-CoV-2, revealed the presence of 9 and 12 post-transcriptional modifications, respectively. HIV-1 utilizes host cell machinery to disguise itself from the host innate immune response. Upon HIV-1 infection, PCIF1 protein level is reduced. Therefore, host mRNA exhibits lower level of m^6A_m in cell culture and human patient models. Potentially, HIV-1 inhibits PCIF1 expression to disguise itself as host transcripts. Global modified ribonucleoside profiles showed similarities among the identified ribonucleosides. Therefore, it is possible the two viral infections may alter host epitranscriptome regulators to allow for more efficient transcription and translation of viral RNA genomes. However, the results need further validation by measuring PCIF1, CMTr1-2, PUS1, and other post-transcriptional modification regulators for transcript and protein levels upon SARS-CoV-2 infection.

Overall, this dissertation sought to develop and apply a high-throughput method for assessing the epitranscriptome and its function in health and disease. As more information and new technologies are revealed, so will our understanding into intricacies of RNA biology.