THE UNIVERSITY OF CHICAGO

HIGH-RESOLUTION MAPPING OF MRNA MODIFICATIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

LISHENG ZHANG

CHICAGO, ILLINOIS

JUNE 2019

For my advisor Prof. Chuan He who has constantly supported me and encouraged me to pursue science and truthfulness, who is the most powerful man I have ever met in my life

For my family who stay with me whatever difficulties I am facing and show their love

For all my friends who always stand by my side and never fail me when I need their help

TABLE OF CONTENTS

LIST OF FIGURES

Acknowledgements

First and foremost, I want to express my deep gratitude to my PhD advisor, Professor Chuan He. I have been admiring Prof. He and his outstanding research work since I was an undergraduate student in Peking University. And I will never forget the first time I read Prof. He's papers and decided to devote my rest life to studying DNA/RNA biology. In January 2013, I was excited to be admitted to the PhD program at the University of Chicago and joined the lab in September. I thank Prof. He for all the opportunities, resources and support he provides. In all times of my graduate research, Prof. He's distinguished scientific insight and extraordinary strength of thoughts guided me to look for challenging and meaningful research topics; his great patience and encouragement lead me to face and overcome the difficulties in pursuing the real science.

I would like to thank Professor Tao Pan for being my thesis committee and for his advice on designing strategies for studying cellular effects of the mitochondrial tRNA demethylation. When I was a first-year graduate student working on demethylase, I talked with Prof. Pan and benefited from his expertise in RNA biology.

I am grateful to Professor Guangbin Dong for being my committee member. When I was an undergraduate student, Prof. Dong's research papers on palladium catalysis guided me on discovering new reactions. And this thesis is also majorly focusing on developing chemistry-based methodology for sequencing RNA modifications.

I would also like to thank my colleagues and collaborators, especially Chang Liu for all help that she gives to me. I thank Dr. Honghui Ma, Dr. Xingyu Lu, Dr. Xiao Wang, Dr. Boxuan Zhao, Dr. Fange Liu, Dr. Kai Chen, Dr. Guanzheng Luo and Dr. Lulu Hu, Zijie Zhang and Jun Liu for their help on experiments and data analysis.

I thank the University of Chicago facilities, especially the Mass Spectrometry Facility. I thank Dr. Pieter W. Faber and Genomics Facility of the University of Chicago for generous help with high-throughput sequencing.

All the work mentioned here cannot be done without all other Prof. He Group members. I want to express my gratitude to all the current members and alumni in Prof. He Group. I have learned a lot from them, and I have enjoyed all the fruitful discussion with them. I thank Dr. Jordi Taylor for editing the manuscripts.

Finally, I would like to thank my family and all my friends for all their love and company. It is them who make me stronger and support me to go through all the hard time in life.

Abstract

Transcriptome carries a wide range of different chemical modifications. Among these modifications, RNA methylation is the most abundant, exerting important functions in multiple biological processes in eukaryotes. Especially, the recent discoveries of N^6-methyladenosine (m^6A) in transcriptome with reversible dynamics and regulatory roles have drawn considerable attentions from world-wide scientists. While N^6-methyladenosine (m^6A) is widely observed to regulate gene expression in mammals, our lab has also reasoned that positively-charged mRNA modifications such as N^1-methyladenosine (m^1A) could tune RNA secondary structures or protein-RNA interactions, which could impact key biological functions.

N^7-methylguanosine (m^7G) is the positively-charged, well-known and essential modification at the 5' cap of eukaryotic messenger RNA (mRNA), regulating mRNA export, translation, and splicing. m^7G also occurs internally within transfer RNA (tRNA) and ribosomal RNA (rRNA), but whether it exists within eukaryotic mRNA remains to be investigated. Here, we show the presence of internal m^7G sites within mammalian mRNA *via* mass spectrometry quantification; then illustrate the transcriptome-wide profile of internal m^7G methylome with m^7G-MeRIP-seq. To map this modification at base resolution, we developed a chemical-assisted sequencing approach that selectively convert internal m^7G sites into abasic sites, which finally leads to misincorporation at these m^7G sites during reverse transcription. With accurately detecting internal m^7G methylation in tRNA and mRNA, the high resolution of this misincorporation-based method enables us to reveal key features of internal m^7G methylome in human cells. We also identified METTL1 as a methyltransferase that installs a subset of m^7G within mRNA and showed that internal m^7G methylation may promote mRNA translation. Collectively, our ap-

proaches reveal internal m^7G methylome in mammalian mRNA and suggest potential functional roles of m^7G-mediated epitranscriptomic regulation.

After the characterization of the first tRNA demethylase by our group, ALKBH1, which removes the ubiquitous m^1A methylation in a set of tRNAs, leading to translation regulation. In this thesis, we identified ALKBH7, a protein that significantly affects mammalian energy metabolism by exhibiting the overt obesity phenotype in *Alkbh7* knockout mice, is a mitochondrial protein and mediates mitochondrial tRNA G26 N^2,N^2-dimethylguanine ($m^2{}_2G$) demethylation. And we found ALKBH7 regulates mRNA levels of 13 proteins encoded by mitochondrial genome, which indicates its possible role in affecting mitochondrial pre-mRNAs. The misincorporation-based high-resolution RNA-seq strategy will unveil whether mt-tRNA G26 demethylation exists in mature mt-tRNAs or junctions within mitochondrial polycistronic RNA (closed related to mitochondrial pre-mRNAs). The base-resolution mapping can showcase how mt-tRNA methylation/demethylation regulate each stage of mitochondrial polycistronic RNA processing, illustrate the timeline how these methyl groups are installed/removed co-transcriptionally or post-transcriptionally.

PUBLICATIONS DURING PHD STAGE RELATED TO THE WORK PRESENTED IN THIS THESIS

1. **Li-Sheng Zhang**[†], Chang Liu[†], Honghui Ma, Hui-Lung Sun, Qing Dai, Guanzheng Luo, Zi-jie Scott Zhang, Linda Zhang, Lulu Hu, Xueyang Dong, Chuan He. Transcriptome-wide Mapping of Internal N^7-methylguanosine Methylome in Mammalian mRNA.
   Molecular Cell https://doi.org/10.1016/j.molcel.2019.03.036 (2019).

[†]These authors contributed equally to the work.

PUBLICATIONS DURING PHD STAGE UNRELATED TO THE WORK PRESENTED IN THIS THESIS

2. Xingyu Lu, Dali Han, Boxuan Simen Zhao, Chun-Xiao Song, **Li-Sheng Zhang**, Louis C Dore, Chuan He. Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics.
   Cell Res. **25**, 386–389 (2015).
3. Isao Masuda, Ryuma Matsubara, Thomas Christian, Enrique R. Rojas, Srujana S. Yadavalli, **Li-Sheng Zhang**, Mark Goulian, Leonard Foster, Kerwyn Casey Huang, Ya-Ming Hou. tRNA Methylation Is a Global Determinant of Bacterial Multi-drug Resistance.
   Cell Systems https://doi.org/10.1016/j.cels.2019.03.008 (2019).
4. **Li-Sheng Zhang**, Guihua Chen, Xin Wang, Qing-Yun Guo, Xi-Sha Zhang, Fei Pan, Kang Chen, Zhang-Jie Shi. Direct Borylation of Primary C–H Bonds in Functionalized Molecules by Palladium Catalysis.
   Angew. Chem. Int. Ed. **53**, 3899–3903 (2014).

CHAPTER 1

Introduction: RNA Epitranscriptomics

1.1 The emerging fields of epigenetics and epitranscriptomics

Nucleic acids, the overall name for DNA (deoxyribonucleic acid) and RNA (ribonucleic acid), are the biomolecules essential and indispensable to all known forms of life. Nucleic acids exist abundantly in every living cell of every life-form organism, where they are given a mission to encode, store, duplicate, transmit and express genetic information in terms of their programmed sequences consisted of four canonical bases (A/T/C/G for DNA and A/U/C/G for RNA). Based on the central dogma, genetic information from genes flows from DNA to RNA and finally to protein, in which genetic variations occurred in any links of gene expression can have an impact on the changes in heredity and phenotypes.

Dated back to 1940s, the concept of epigenetics was first proposed by Dr. Waddington when studying the whole complex of development processes between genotype and phenotype, which led to the birth to a prototype of the general characteristics of an epigenotype [1,2]. The definition of ‘epigenetics’ was then gradually evolved as "the study of the mechanisms that impart temporal and spatial control on the activities of all those genes required for the development of a complex organism from the zygote to the adult" [3]. The ‘epigenetics’ could be more precisely explicated as "stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence" [4], and also “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” [5], involving all factors that alter chromatin template and trigger variation of gene expression. The field of epigenetics focuses on these stable and remarkable phenotype changes that happened without changes

in the primary genome sequence of genes, which demonstrate one new area of biology for our understanding of chromatin and gene regulatory mechanisms.

Nucleic acids, both DNA and RNA, carry significantly diverse chemical modifications. Instead of the previous view on the role of these modifications, which might be static and contributing minor roles in 'fine-tuning' structural and functional properties of nucleic acids, recent findings strongly indicate their dynamic regulatory roles. No matter in genome and in transcriptome, these diverse chemical modifications serve as functional marks to control gene regulation and cellular processes, in terms of "epigenetics" and "epitranscriptomics". Generally, the major research directions of epigenetics might be summarized as three emerging areas, including histone modifications, DNA modifications, and RNA modifications.

1.2 Epigenetic marks on DNA

Various chemical modifications on DNA are employed as a significant way to control chromatin structures along with histone modifications. Cytosine methylation (5mC) in genomic DNA was first discovered by Wyatt group in 1951[6]. People initiated the study on the regulatory maintenance of the 5mC pattern across cell divisions twenty years later[7,8]. Although the activity of potential mammalian methyltransferases machinery was found to exist in cellular extracts[9,10], Dnmt1 was not identified as the first DNA methyltransferase for maintenance of 5mC until 1983[11], in which its loss induces genome-wide depletion of the CpG methylation in mouse embryonic stem cells (mESCs). Research on Dnmt1 homologs in mouse led to the characterization of Dnmt3a and Dnmt3b, as the methyltransferases for *de novo* 5mC methylation[12-15]. A series of subsequent functional studies proved that DNA 5mC methylation is closely correlated to repres-

sion of gene expression, in which loss of methylation triggered reactivation of some naturally inactive genes in *dnmt1* knockout mice[16].

The histone modifications are controlled by chromatin-associated enzymatic systems that further dynamically regulate the structure of histones and associating chromatins, where there exist 'writers' to install these modifications and the 'erasers' to remove them[17-19]. DNA 5mC methylation was considered to be dynamic and reversible prior to the discovery of 5mC 'eraser' enzymes[20-22]. At the time point several hours after fertilization but before the first-round DNA replication, people observed the rapid decrease of 5mC level on the paternal genome in the zygote, indicating the potential demethylation process on DNA and the existence of DNA demethylases[23,24]. Although the 'writer' enzymes of 5mC were identified in 1980s, the demethylases that remove DNA 5mC mark remained as an unclear question for decades until the discovery of 5-hydroxymethylcytosine (5hmC) in the mammalian genome[25,26] and the identification of TET (ten-eleven translocation) proteins[26,27]. TET proteins are methylcytosine dioxygenase that utilize dioxygen to oxidize 5mC to 5hmC, demonstrating the key chemical basis for active DNA 5mC demethylation. Subsequent studies proved that the TET enzymes can further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[28-30]. Both 5fC and 5caC can be recognized and excised by human thymine DNA glycosylase (TDG), followed by base excision repair (BER) to replace the modified cytosine with a normal cytosine, completing the active demethylation process[29,31]. Additionally, the 5mC oxidation derivatives of 5hmC, 5fC, and 5caC may also be passively diluted to the unmethylated stage through cell division[32].

For deep functional investigation of 5mC and its oxidative derivatives in genomic DNA involves the study of the proteins that recognize and binds 5mC (or 5hmC/5fC/5caC), and then recruit other protein machineries and initiate subsequent actions, in terms of DNA 5mC 'readers'.

The first DNA 5mC 'reader' protein was characterized as methyl-CpG binding protein complex MeCP1 by Bird group[33]. The following studies precisely discovered four 5mC readers comprising the methyl-CpG binding domain (MBD) family[34], including MeCP2, MBD1, MBD2, and MBD4 (MBD3 in this family is not a 5mC reader). Interestingly, MeCP2, MBD1, and MBD2 showcase the functional role in 5mC-dependent transcriptional repression[35]. An unrelated p120 catenin partner protein Kaiso was also found to be a specific 5mC reader and function as a methylation dependent transcriptional repressor[36]. Furthermore, among the oxidative derivatives of 5mC, 5hmC is much more abundant than 5fC and 5caC and appears to be postulated to a stable and distinct epigenetic modification instead of just an intermediate of active demethylation, termed DNA hydroxymethylation. Methyl-CpG-binding protein 2 (MeCP2) was identified as the major 5hmC-binding protein in the brain, showing the similar high affinities on binding 5hmC- and 5mC-containing DNA[37]. Recent studies also unveiled UHRF2 as a *bona fide* 5hmC reader with a role for 5hmC in neuronal function[38].

Besides DNA 5mC and its oxidative derivatives, N^6-methyladenine (6mA), also exists in the eukaryotic genomic DNA. In 2015, three groups reported the presence of 6mA in three different eukaryotes independently, giving an open on the functional study of this methylation modification in eukaryotes[39-41]. Although 6mA in alga genomic DNA has been verified for a long time, the methylation distribution and biological function was not unveiled until recently developed high-throughput sequencing approaches to map 6mA in *Chlamydomonas* genomic DNA. 6mA exhibited distribution enriched at ApT dinucleotides around transcription start sites (TSS), and also labels the active transcribed genes with marking the linker DNA regions between adjacent nucleosomes. These findings implied the potential gene activation function of adenine methylation in the *Chlamydomonas* genome.

1.3 Chemical modifications on mRNA and emergence of RNA epigenetics

Considering genomic DNA being installed by different chemical modifications as epigenetic marks, RNA biomolecules also carry around 109 identified RNA base modifications[42], among which many are known to exist in eukaryotes for decades including tRNA and rRNA modifications, as well as mRNA cap methylations. However, N^6-methyladenosine (m^6A) was only recently identified to be an internal mRNA modification that can be dynamically controlled and reversible. The existence of N^6-methyladenosine (m^6A) in RNA, previously was discovered in tRNA and rRNA, was proved to be present within both eukaryotic and viral mRNAs in 1974[43-48], demonstrating m^6A as the most prevalent internal modification in mRNAs and long non-coding RNAs (lncRNAs) in higher eukaryotes[49]. In the mammalian transcriptome, the abundance of m^6A was revealed to be approximately 3~4 marks per mRNA, with a consensus motif of G(m^6A)C (70%) or A(m^6A)C (30%) and varied methylation status at each site[50-56]. The discovery of the first internal mRNA m^6A demethylase FTO in 2011 demonstrated this abundant mRNA modification to be dynamic and reversible[58], opening a new field on 'RNA epigenetics'[57]. The following work depicted the transcriptome-wide distribution of m^6A (independently reported by two groups) with m^6A-seq based on antibody-mediated capture and massively parallel sequencing. People identified over ten thousand m^6A sites in the transcripts of more than 7,000 human genes[59,60]. The internal m^6A sites majorly accumulate in two distinct regions: around stop codons and within long internal exons, which are highly conserved between human and mouse. These findings therefore attracted tremendous interests in the biological relevance of m^6A and suggested that RNA decoration by m^6A has a fundamental role in regulation of gene expression.

Like the previously mentioned DNA 5mC methyltransferases, the methyl group at N^6 position of adenosine is established by mammalian m^6A 'writers' acting as a protein complex

with four identified components: methyltransferase-like 3 (METTL3), METTL14, Wilms tumor 1-associated protein (WTAP) and KIAA1429. In 1978 people partially purified an m^{6}A-MTase from HeLa cells to carry out the methylation in the region of cap structure instead of internal area[61]. Later internal methyltransferase activity on m^{6}A was detected by in vitro methylation assay, with HeLa cell nuclear extracts[62]. The following efforts further demonstrated the multicomponent nature of m^{6}A methyltransferase system and the complexity of the enzyme complex[63-65]. Core components of m^{6}A 'writer' machinery are characterized in 2014 as METTL3, METTL14, and WTAP[66-69]. Depletion of m^{6}A methyltransferase complex causes a series of dramatic phenotypes such as inhibiting subsequent differentiation in mESCs, early-stage mouse embryo development deficiency, smaller head/eyes/brain ventricle and curved notochord in zebrafish, or defects in gametogenesis in yeast, flies, and plants[70-75].

Recently, people figured out the full m^{6}A methyltransferase complex in HeLa cells identifying METTL3/METTL14/WTAP/VIRMA/HAKAI/ZC3H13 as the key components, in which VIRMA mediates preferential mRNA methylation in 3' UTR and near stop codon. Biochemical studies further unveil that VIRMA recruits the catalytic core components METTL3/METTL14/WTAP to guide region-selective methylations[76].

AlkB family proteins use iron, α-ketoglurate, and dioxygen to catalyze oxidative demethylation of a range of nucleic acid substrates[77,78]. The *E. coli* AlkB, a prototype of the AlkB family proteins, mediates oxidative repair of 1mA and 3mC DNA base methylation lesions; two human AlkB homologues, ALKBH2 and ALKBH3, are responsible for similar DNA repair functions in human cells[79,80]. In recent years, He lab made a well-known and significant discovery that FTO and ALKBH5[58,81], both homologues of *E. coli* AlkB and human ALKBH2 and ALKBH3, serve as m^{6}A 'eraser' and perform quite effective oxidative demethylation of the *N*^{6}-

methyladenosine (m^6A) in eukaryotic messenger RNA and other nuclear non-coding RNAs. This discovery leads to recent rapid development of a new field of biology on investigating gene expression regulation by dynamic RNA modifications[83,84]. For the significant phenotypes, FTO has effects on development while ALKBH5 affects spermatogenesis, suggesting the effects of m^6A on multiple biological phenomena. The subsequent study on FTO demethylation mechanism revealed two key intermediates, N^6-hydroxymethyladenosine (hm^6A) and N^6-formyladenosine (f^6A)[82], generated through the step-by-step oxidation of m^6A, which appears similar to oxidative intermediates of 5mC.

Importantly, for further in-depth functional study on internal m^6A in mRNA requires the characterization of proteins that recognize and binds m^6A, followed by recruiting other protein machineries and biological mechanisms, in terms of mRNA m^6A 'readers' that translate the distribution of m6A into biological functions with the two major modes (direct or indirect) of m^6A-dependent RNA recognition. In 2013, for the direct mode, He group demonstrated the first 'reader' protein YTHDF2 for m^6A[85], which affects the stability of m^6A-containing mRNA. Another direct m^6A 'reader' YTHDF1 was also identified by He group as a key factor that promotes translation efficiency of m^6A-modified mRNA[86]. Besides, in an indirect manner, m^6A-recognition can also be fulfilled by RNA structure change depended on the existence of m^6A, which tunes the accessibility of RNA binding proteins. HNRNPC, as the first reported indirect m^6A 'reader'[87], displays strong interaction with m^6A-containing RNA in terms of m^6A-controlled RNA structural remodeling. Meanwhile, RNA binding by HuR was discovered to be interfered by the presence of m^6A methylation[88]. hnRNPA2B1 has also been shown to selectively recognize methylated pri-microRNA in order to promote microRNA maturation[89,90]. In addition, m^6A depletion prolongs nuclear retention and delays the nuclear exit export of mature

mRNAs of clock genes *per2* and *arntl*, which connect m^6A to the pace of the circadian cycle and the clock speed and stability[91]. Coming back to members of the YT521-B homology (YTH) domain-containing proteins, people later discovered YTHDF3 displayed the property on promoting translation[92] and playing an important role in the initial stages of translation. The nuclear m^6A binding protein YTHDC1 was depicted to function as the factor enhancing exon inclusion in targeted mRNAs through recruiting pre-mRNA splicing factor SRSF3 (SRp20) while blocking SRSF10 (SRp38) mRNA binding[94]. YTHDC2 plays critical roles during spermatogenesis: *Ythdc2* knockout mice are infertile; males have significantly smaller testes and females have significantly smaller ovaries compared to those of littermates[93]. Future efforts on new reader proteins and their related biomechanisms is the critical point for understanding the regulatory roles of m^6A in a myriad of biological events and inspiring the research on 'readers' of other mRNA modifications.

1.4 Scope of this dissertation

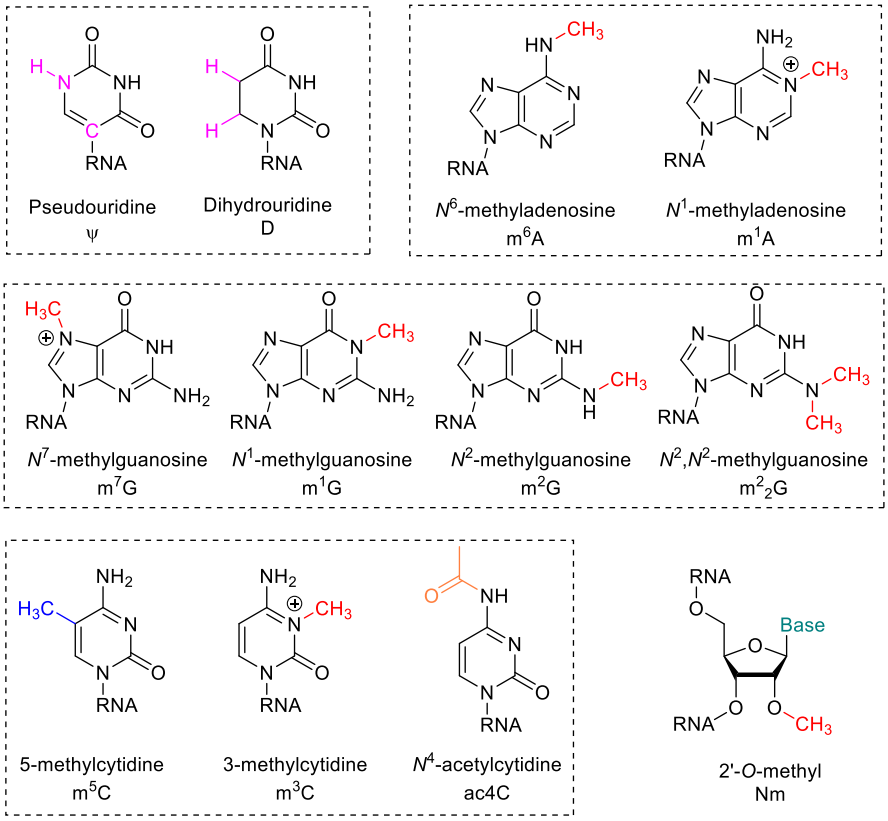

Figure 1. Diverse internal mRNA modifications.
Many of these chemical modifications exist across different RNA species and contribute to epigenetic regulation.

While N^6-methyladenosine (m^6A) is widely observed to regulate gene expression in mammals, we has also speculated that positively charged base modifications such as N^1-methyladenosine (m^1A) could tune RNA secondary structures or protein-RNA interactions, which could impact key biological functions. In 2016, our group developed the base-resolution sequencing method for transcriptome-wide mapping of internal m^1A in mammalian mRNA[95-98]. Other internal mRNA modifications are reported in succession (**Figure 1**) including pseudouridine (Ψ)[99-102], 5-methylcytidine (m^5C)[103-106], 5-hydroxymethylcytidine (hm^5C)[107], N^4-

acetylcytidine (ac4C)[108], 2'-O-methyl (Nm)[109]. This thesis will firstly focus on internal N^7-methylguanosine (m^7G) in mammalian messenger RNA, demonstrating base-resolution sequencing methodology for transcriptome-wide mapping of internal N^7-methylguanosine (m^7G) in mammalian messenger RNA with the characterization of its methyltransferase METTL1 and functional study on its effect on translation. The reason why internal N^7-methylguanosine (m^7G) in mRNA deserves in-depth study could be attributed to two major aspects: (i) Natural internal m^7G sites are unable to generate misincorporation after reverse transcription due to the 7-methyl group does not affect base pair. It could be valuable to develop a chemical-assisted misincorporation-based sequencing method to unveil the precise information on this methylome within mRNA at base resolution and set the basis for further functional investigation or similar sequencing strategies; (ii) m^7G at the cap region has been studied for decades and was proved to be the key factor recruiting a series of 'reader' proteins that recognize and bind cap structure, including eIF4E, CBP20/CBP80 complex and other cap binding proteins. All these identified cap m^7G binding proteins might be the candidates for potential 'readers' for internal m^7G sites within mRNA. The final section on ALKBH7, a potential tRNA G26 N^2,N^2-dimethylguanine ($m^2{}_2G$) demethylase, will confirm the strength of misincorporation-based sequencing strategy for future study on mitochondrial polycistronic RNA.

Chapter 2 presents base-resolution sequencing technology for transcriptome-wide mapping of internal N^7-methylguanosine (m^7G) in mammalian messenger RNA.

Chapter 3 presents the characterization of methyltransferase machinery for a subset of internal m^7G in mRNA, and the functional study on its effect on translation.

Chapter 4 presents ALKBH7, a protein that dramatically affects mammalian energy metabolism, as a mitochondrial protein that mediates mitochondrial tRNA G26 N^2,N^2-dimethylguanine ($m^2{}_2G$) demethylation and possibly regulates mitochondrial pre-mRNAs.

Chapter 5 summarizes the understanding of the broader impact of misincorporation-based base-resolution sequencing methodology for future investigation of diverse mRNA modifications and the directions how it assists RNA epigenetics research.

CHAPTER 2

Base-Resolution Mapping of Internal N^7-Methylguanosine Methylome in Mammalian Messenger RNA

2.1 Introduction: N^7-methylguanosine (m^7G) in RNA

Modifications on messenger RNA contribute to post-transcriptional regulation of mRNA fate[110-113]. While N^6-methyladenosine (m^6A) is widely observed to impact mRNA metabolism in mammals[114], we have also reasoned that positively charged base modifications such as N^1-methyladenosine (m^1A) could tune RNA secondary structures or protein-RNA interactions within mRNA through a combination of electrostatic and steric effects[95,96]. Another positively charged modification is the ubiquitous mRNA cap N^7-methylguanosine (m^7G)[115] **(Figure 2.2A)**. m^7G is installed at the 5' cap co-transcriptionally during transcription initiation[116]. This essential cap modification stabilizes transcripts against exonucleolytic degradation[117,118] and modulates nearly every stage of the mRNA life cycle, including transcription elongation[119], pre-mRNA splicing[120,121], polyadenylation[122], nuclear export[123], and translation[124].

Besides as a part of the cap structure, m^7G also exists internally within transfer RNA[125] and ribosomal RNA[126]. m^7G is found at position 46 of yeast and human cytoplasmic tRNAs, installed by the heterodimers Trm8-Trm82 and METTL1-WDR4, in yeast and mammals respectively; it also occurs at position 1639 of human 18S rRNA, installed by human WBSCR22. These internal m^7G modifications impacts RNA processing and function and have been suggested associated with human diseases. WBSCR22 is involved in late nuclear pre-ribosomal RNA processing[127] and the biosynthesis of the 40S ribosomal subunit[128,129]. Mutations of the m^7G_{46}

methyltransferase complex (Trm8-Trm82) in yeast resulted in rapid degradation of tRNAs[130,131], followed by growth deficiency[132]; mutations in the human homologs (METTL1-WDR4) are associated with microcephalic primordial dwarfism[133].

In this chapter, we demonstrate the presence of internal m^7G sites within mammalian mRNA *via* mass spectrometry quantification and m^7G-MeRIP-seq, high-throughput sequencing based on antibody-based immunoprecipitation. To more accurately map this modification transcriptome-wide, we developed a chemical-assisted base-resolution sequencing approach (m^7G-seq) that takes advantage of the unique chemical reactivity of m^7G, which finally leads to misincorporation at m^7G sites during reverse transcription and thus enables us to reveal key features of the m^7G methylome in human cells at base resolution. We found internal sites on human mRNA that appear to be highly m^7G modified. We also identified METTL1 as a methyltransferase that installs a subset of m^7G within mRNA and showed that internal m^7G methylation may promote mRNA translation. Collectively, we present a method to precisely map the m^7G methylomes in RNA and base-resolution m^7G maps in human mRNA, suggesting potential functional roles of internal m^7G within the mammalian mRNA.

2.2 Quantitative detection of internal N^7-methylguanosine within mRNA

To assess the possible presence of m^7G in mRNA internally, we employed a liquid chromatography-tandem mass (LC-MS/MS) method to quantify the modification level of thoroughly purified poly(A)-tailed RNA (with two rounds of poly(A) enrichment followed by rRNA depletion). All the samples for mRNA (in purpose of both LC-MS/MS and sequencing) were prepared from total RNA with two rounds of $polyA^+$ enrichment where tRNAs were depleted and rRNAs were eliminated to below 0.1% of the original abundance (**Figure 2.2B**). Aiming to monitor m^7G signals from both internal and 5'-cap regions, we applied a 3-step method for RNA digestion, which includes the endonuclease cleavage (nuclease S1) followed by a treatment with phosphodiesterase I to completely expose m^7G signal from the cap structure (**Figure 2.2C and Figure 2.2F**). We observed an m^7G/G level of approximately 0.45% in HEK293T cells ("$polyA^+$ RNA" on the left in **Figure 2.2D**), including m^7G sites at both the 5'-cap and internal regions. We then removed the cap using tobacco decapping enzyme to specifically quantify the internal m^7G level. The m^7G level decreased dramatically after the first round of decapping treatment but remained at a stable level around 0.03% consistently, which did not diminish after two more rounds of cap removal (**Figure 2.2D**). We also performed RNA digestion in a two-step manner without phosphodiesterase I digestion, and found the m^7G level detected was comparable to that after two to three rounds of decapping treatments ("$polyA^+$ mRNA" on the right in **Figure 2.2D**), which suggests the digestion of mRNA with only endonuclease cleavage (nuclease S1) would not affect cap structure and can ensure reliable exposure of internal m^7G level in mRNA. We measured m^7G/G levels ranging from approximately 0.02% - 0.05% across several human and mouse cell lines (**Figure 2.2E**), which represent roughly 5-10% of the m^6A/A ratios in the same cells. Thus, m^7G is present internally within mammalian mRNA.

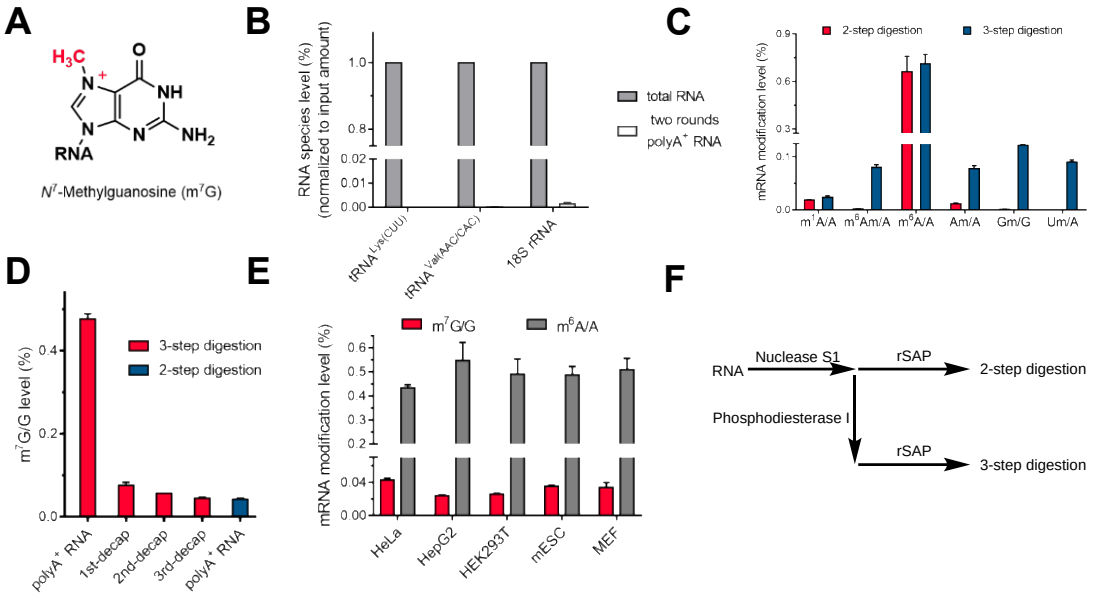

Figure 2.2. Quantification of internal N^7-methylguanosine in mRNA by LC-MS/MS. (A) Chemical structure of the positively charged RNA modification N^7-methylguanosine (m^7G). (B) 18S ribosome RNA and tRNA depletion tests with two rounds of polyA$^+$ RNA enrichment. Mean values ± s.d. are shown, n = 2. (C) Phosphodiesterase I exposes Nm signals at the cap structure. (D) LC-MS/MS quantification of m^7G/G levels in polyA+ RNA (HEK293T cells) using a three-step RNA digestion protocol (red) or a two-step digestion protocol (blue). Mean values ± s.d. are shown, n = 3. (E) LC-MS/MS quantification of m^7G/G and m^6A/A levels in cap-depleted polyA+ mRNA isolated from human and mouse cells. Mean values ± s.d. are shown, n = 3. (F) Schematic outline of 2-step digestion and 3-step digestion, without or with Phosphodiesterase I.

2.3 Transcriptome-wide profile of internal N^7-methylguanosine by antibody-based m^{7}G-MeRIP-seq

To study the distribution of internal m^{7}G methylome, we performed m^{7}G-MeRIP-seq to map the transcriptome-wide distributions of internal m^{7}G sites. We first conducted dot blot to

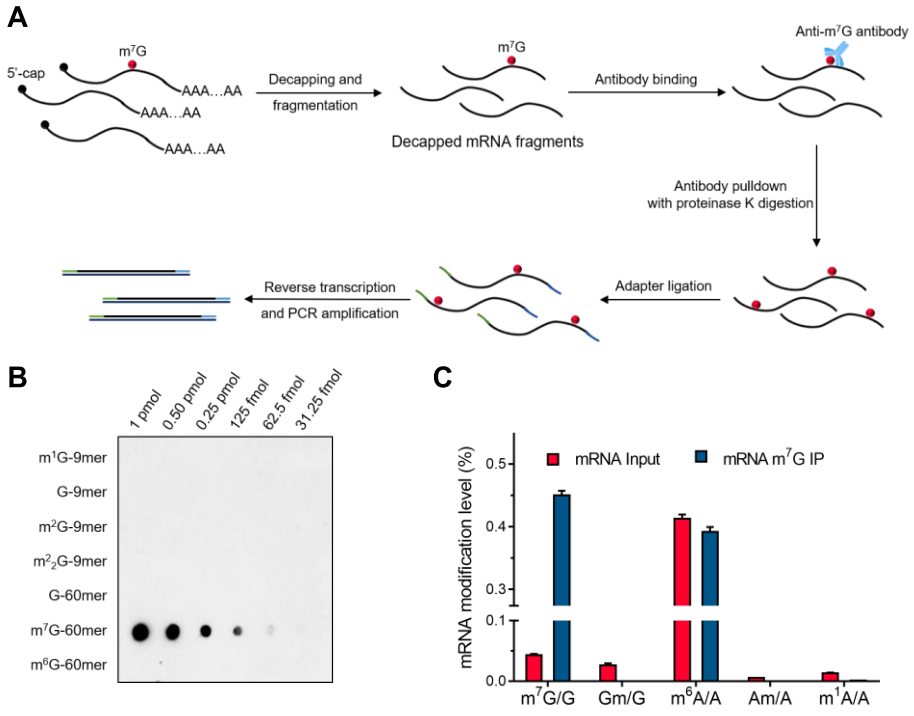

Figure 2.3.1 An antibody-based sequencing approach for m^{7}G-MeRIP-seq. (A) A schematic outline of m^{7}G-MeRIP-seq. (B) Dot blots demonstrating high specificity of the anti-m^{7}G antibody. Increasing amounts (indicated above the top blot) of synthetic RNA oligonucleotides containing m^{7}G, m^{2}G, m$^2{}_2$G, m^{1}G, m^{6}G or unmodified G residues were spotted onto a membrane and probed with anti-m^{7}G antibodies. Anti-m^{7}G antibody detected m^{7}G and did not exhibit cross-reactivity with other G modifications or unmodified G. (C) LC-MS/MS quantification of m^{7}G/G, Gm/G, m^{6}A/A, Am/A, m^{1}A/A levels showing enrichment of m^{7}G after immunoprecipitation (IP) using anti-m^{7}G specific antibody (MBL), without enriching other modifications. Cap-depleted polyA$^+$ mRNA (purified from HEK293T cells) was used as the input control. Mean values ± s.d. are shown, n = 3.

characterize that the selected antibody (MBL) binds to m^{7}G specifically over all known methylated G (including m^{1}G, m^{2}G, m$^2{}_2$G, and m^{6}G) or unmethylated G (**Figure 2.3.1B**). Then we

briefly immunoprecipitated methylated RNA after cap removal (with two rounds of tobacco decapping treatments) using the anti-m^7G specific antibody (MBL), followed by high-throughput sequencing (**Figure 2.3.1A**). Based on m^7G-MeRIP-seq, we found that m^7G was enriched around 12 folds in the anti-m^7G antibody-bound fraction, starting with cap-depleted and fragmented poly(A)-tailed RNA from HEK293T cells. In contrast, other known mRNA modifications including m^6A, Nm (Gm and Am) or m^1A, were not enriched (**Figure 2.3.1C**).

By applying m^7G-MeRIP-seq to poly(A)-tailed RNA (cap- and rRNA-depleted) from human and mouse cells in two replicates, we identified 7,469, 7,911, 9,243, 6,184 and 6,743 potential internal m^7G peaks in HeLa, HEK293T, HepG2, MEF and mES cells respectively (fold change (FC) ≥ 3, false discovery rate (FDR) ≤ 0.05), in adequately expressed transcripts (above FPKM 1.0) that were shared in replicates. We detected 3,823 m^7G peaks that overlapped in the three human cell lines, indicating a relatively high confidence of antibody enrichment, and 2,217 m^7G peaks overlapped in the two mouse cell lines (**Figure 2.3.2A**). About 50% of the methylated transcripts are methylated once, and each of them carries around 1.9 peaks on average (**Figure 2.3.2B**). A statistically significant GA-rich motif was found in the MeRIP-seq peak regions (**Figure 2.3.2C**).

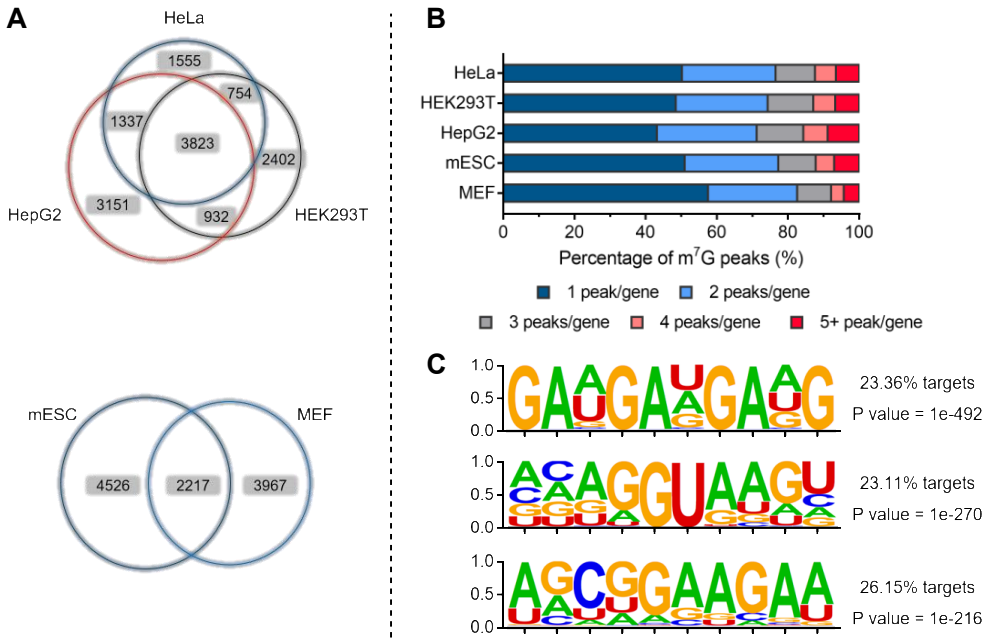

Figure 2.3.2. Applying m^{7}G-MeRIP-seq to human and mouse cell lines. (A) The overlap of m^{7}G-MeRIP-seq m^{7}G peaks among three human cell lines (left) and two mouse cell lines (right) (fold change (FC) ≥ 3, false discovery rate (FDR) ≤ 0.05, FPKM ≥ 1.0). (B) The percentage of methylated genes with 1, 2, 3, 4 or 5+ peaks per gene in the indicated human and mouse cell types. Half of the methylated transcripts carry only one m^{7}G peak. (C) Top three motifs identified with m^{7}G-MeRIP-seq by MEME.

We detected internal m^{7}G peaks in all three transcript segments of 5′ untranslated regions (UTR), coding sequences (CDS), and 3′ UTRs. Similar distribution patterns were observed in all three human cell lines and two mouse cell lines (**Figure 2.3.3A**). Meta-gene profiles of internal m^{7}G peaks along mRNA demonstrated a major accumulation area located at 3' UTR and a minor one within 5' UTR (2.3.3B**)**. Although m^{7}G is enriched at the 3' UTR, there is nearly no overlap between m^{6}A and m^{7}G methylated peaks in HeLa cell line. Furthermore, we investigated the peak conservation between human and mouse, and found that 2,557 identified m^{7}G peaks in mESC could be mapped faithfully to adequately expressed orthologous in human HepG2 cells; 1,016 of them had an m^{7}G peak at the orthologous positions, representing 39.7% overall conser-

vation. The CDS regions exhibited the highest degree of conservation with around 41-43% conserved peaks, compared to other gene segments with nearly 15–20% conservation (**Figure 2.3.3C**). Gene ontology (GO) analysis revealed that the methylated genes are enriched under GO terms related to RNA splicing, mRNA processing, cell-cell adhesion and cell division (**Figure 2.3.3D**).

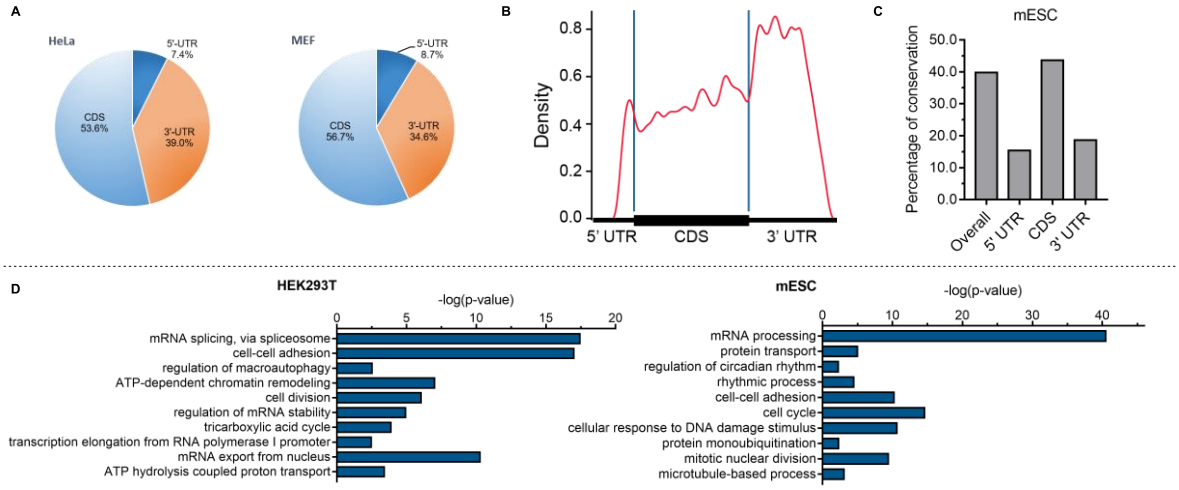

Figure 2.3.3. m^{7}G-MeRIP-seq mapped transcriptome-wide distributions of internal m^{7}G sites in mammalian mRNA. (A) Pie charts presenting the fraction of m^{7}G peaks in each of three transcript segments (5’ UTR, CDS, and 3’ UTR) in HeLa (left) and MEF (right) cells. (B) Meta-gene profiles of m^{7}G peak distribution along a normalized transcript composed of three rescaled non-overlapping segments (5’ UTR, CDS, and 3’ UTR), based on antibody-enriched m^{7}G peaks in HeLa cells (fold change (FC) ≥ 4, false discovery rate (FDR) ≤ 0.05, FPKM ≥ 1.0). (C) Human–mouse m^{7}G conservation shown as percent of orthologous positions *versus* shared m^{7}G peaks according to their locations in the transcript (between HepG2 cells and mESC cells). (D) Gene ontology (GO) analysis of internal m^{7}G-methylated transcripts relative to all adequately expressed genes in HEK293T (left) and mESC (right) cells (FPKM ≥ 1.0).

We investigated the distribution of methylated genes (1≤FPKM≤50) in ten averagely divided expression bins. When calculating the percentage ratio of methylated genes *versus* all genes in each expression decile, the methylated transcripts exhibit a progressively larger fraction as gene expression level increases in both human and mouse cell lines **(Figure 2.3.4B)**; we thus

excluded the possibility that the increased fractions of the methylated genes are likely due to increased detection of highly expressed genes (**Figure 2.3.4A**). We also plotted the fractions of genes with m^7G peaks in each of the segments relative to expression level and found that the transcripts of relatively high expressed genes were more likely to be methylated at CDS and 3' UTR segments than 5' UTR (**Figure 2.3.4C**).

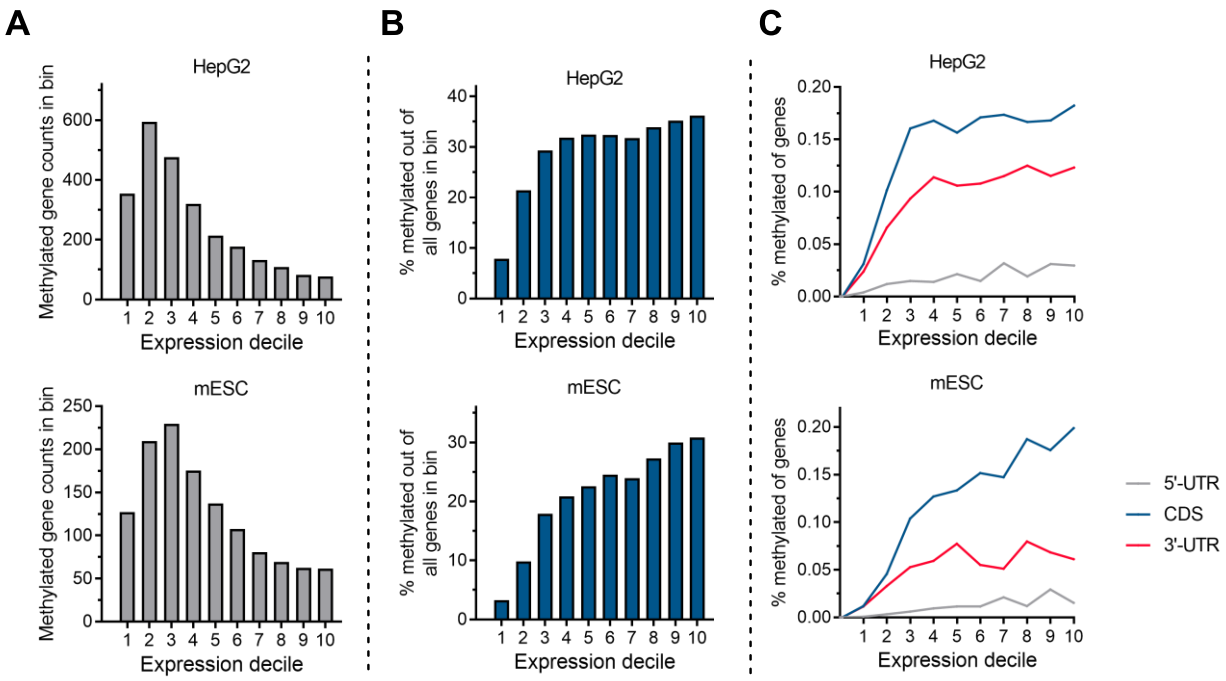

Figure 2.3.4. Distribution of methylated genes in averagely divided expression bins (A) The counts of methylated genes in HepG2 (left), and mESC (right) cells in each expression bin are plotted. (B) The percentages of methylated genes out of all genes (within one expression bin) in HepG2 (left) and mESC (right) cells exhibit a progressively larger fraction as gene expression level increases (the expression level is equally divided into ten bins in a range of $1 \leq FPKM \leq 50$). (C) Fraction of genes with m^7G peaks in each of the segments (5' UTR, CDS, and 3' UTR) related to the gene expression level.

2.4 Development of m^7G-seq at base-resolution established on m^7G chemical property

To achieve more accurate mapping of m^7G at base-resolution with an orthogonal approach, as has been done for other modifications[97], we developed m^7G-seq by taking advantage of the unique chemical reactivity of m^7G in a reduction-induced depurination reaction (**Figure 2.4.1**). The positive charge on the five-membered ring makes m^7G (3a-1 in **Figure 2.4.1**) especially susceptible to $NaBH_4$-mediated reduction, which eliminates the aromaticity of the five-membered ring attached to the ribose but not affecting unmodified G[134]. Subsequent heating (55 °C) at pH 4.5 induces depurination of the reduced m^7G alone (3a-2 in **Figure 2.4.1**) and generates an RNA abasic site (3a-3), which can be further captured by biotin-ligated hydrazide to yield a biotinylated RNA in a one-pot reaction (3a-4). After pull-down and reverse transcription using HIV reverse transcriptase, the biotinylated sites derived from the abasic sites would be mutated to predominantly T as well as other bases[135,136], and thus we can identify the exact internal m^7G sites based on these mutations at single-base resolution.

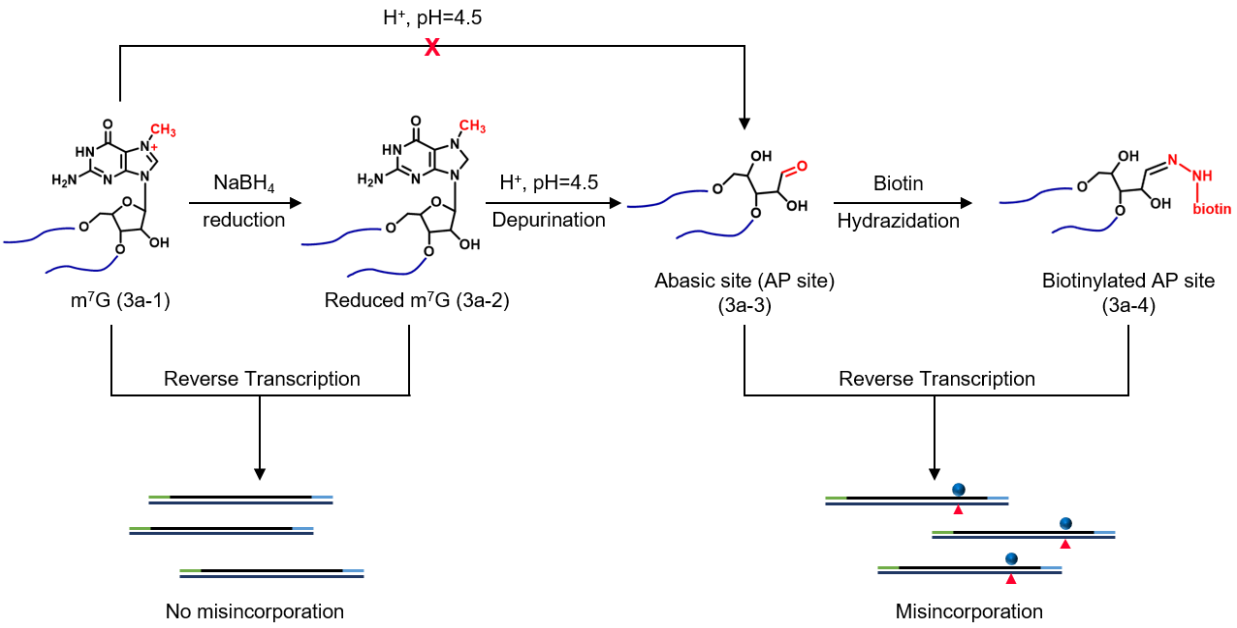

Figure 2.4.1. Schematic diagram showing the chemical reactivity of m^{7}G under reduction and biotin labeling conditions in m^{7}G-seq. Only the reduced form of m^{7}G can generate biotinylated AP sites with biotin hydrazide under mildly acidic conditions. Biotinylated AP sites could induce misincorporation when using HIV reverse transcriptase.

We validated the m^{7}G-seq assay including the reduction and depurination reactions as well as the pull-down enrichment step using model RNAs (**Figure 2.4.2A**). 125-mer RNA probes with a single G (unmodified or m^{7}G modified) incorporated site-specifically were synthesized and subjected to dephosphorylation, reduction and biotinylation consecutively. Dephosphorylation was necessary to avoid potential side reactions of free phosphate groups (at the 5' end of the RNA) with biotin-ligated reagents. After biotin pull-down and reverse transcription, the oligos with m^{7}G modification presented a highly enriched signal compared with the unmodified ones at full length (**Figure 2.4.2B**), which indicated that the reverse transcriptase could read through the biotinylated abasic sites with misincorporation at those sites as demonstrated by Sanger sequencing of the further PCR-amplified products (**Figure 2.4.2C**).

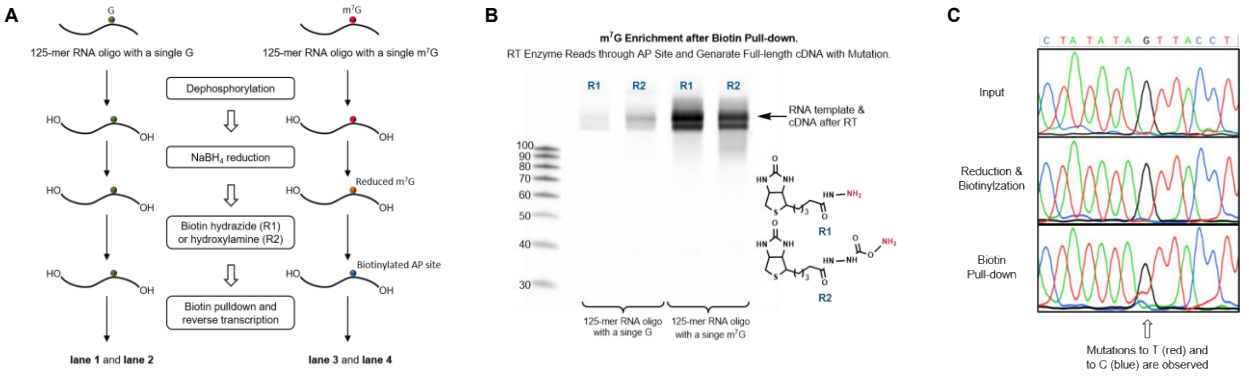

Figure 2.4.2. Verification of chemical reduction/depurination and pull-down enrichment in m^{7}G-seq. (A) m^{7}G reduction, depurination and labeling using synthetic 125-mer RNA probes with a single m^{7}G site or an unmodified G as negative control. Biotin hydrazide or biotin hydroxylamine was used as R1 and R2, respectively. After biotin pull-down, the RNA probes were eluted by Proteinase K digestion followed by reverse transcription using a specific primer. (B) Gel image results of reverse transcription. The biotin labeling of the m^{7}G-containing probe led to a significant RNA pull-down enrichment compared with the control probe. (C) Sanger sequencing results showing mutations (T as red and C as blue) generated in the biotin pull-down sample.

We then optimized the chemical assay to be compatible with library construction and sequencing (**Figure 2.4.3**). mRNA fragments (around 60-80 mer, shorter fragments could lead to more efficient m^{7}G reduction, depurination and labeling) were subject to cap removal and PNK end repair to leave the hydroxyl group at the 3' end. Then dephosphorylation further ensured the depletion of phosphate groups at both ends. After the 3' adapter installation to fragmented poly(A)-tailed RNA, the ligated mRNA fragments underwent a two-step chemical treatment before further enrichment by the biotin tag. We then performed the reverse transcription followed by cDNA 3'-adaptor ligation and PCR amplification. In this way, as m^{7}G sites were converted to biotinylated abasic sites, we can observe misincorporation at the specific sites that were further enriched in the pull-down samples compared with input samples and those before biotin enrichment. To evaluate this method before applying it to mRNA, we designed a 60-mer RNA oligo sequence whose size is similar to mRNA fragment and bears one single fully methylated m^{7}G in

the AAGAU sequence. After applying the aforementioned protocol, we observed substantial mutations at the modified site after generation of the abasic site (G→T 11.4%, G→C 16.4%; ~28% overall mutation rate), and higher mutation rates after further pull-down enrichment (G→T 28.9%, G→C 22.1%), whereas the adjacent bases did not exhibit any misincorporation above background signals (**Figure 2.4.3**). Note that the chemical reduction and depurination reactions are not stoichiometric with fractions of m^7G escaping reduction and depurination. The pull-down step further enriches modified abasic sites, leading to increased mutation rates.

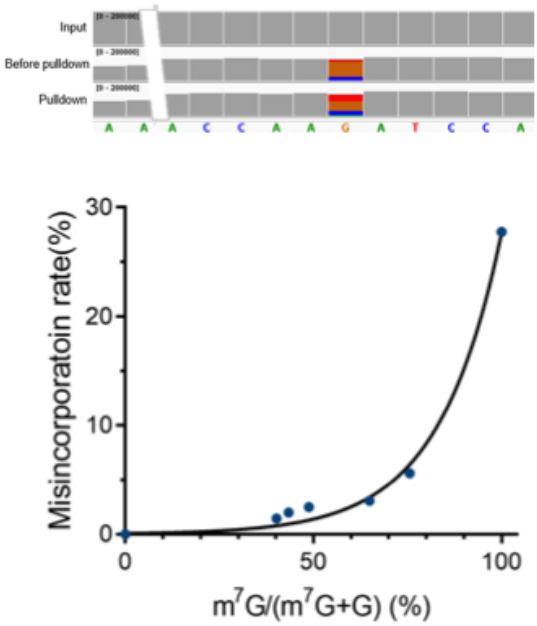

Figure 2.4.3. Verification of m^7G-seq by high-throughput sequencing. Above: IGV (Integrative Genomics Viewer, Broad Institute) plot of synthetic 60-mer probe with fully methylated G in the middle of an AA(m^7G)AU motif in the input, before pulldown and pulldown samples. Below: A fitting curve of the correlation between the estimated methylation level and misincorporation rate in m^7G-seq using the synthetic 60-mer RNA probe with m^7G at different modification fractions in an AA(m^7G)AU motif.

While the biotin-based enrichment allows confident assignments of the modified m^7G sites, the modification fraction at the modified site could in principle be measured using the mis-incorporation rate before enrichment and correlate with a calibration curve generated using RNA probes that bear m^7G at different modification fractions and undergo the same chemical treatment and high-throughput sequencing. A series of probes with the same AAGAU sequence but different m^7G/G ratios was prepared from *in vitro* transcription and then subjected to the base-resolution m^7G-seq protocol (**Figure 2.4.4**). A calibration curve was generated that can normalize the misincorporation rate (before enrichment) to the methylation status, which can then be used to estimate internal m^7G methylation fraction in biological RNA samples.

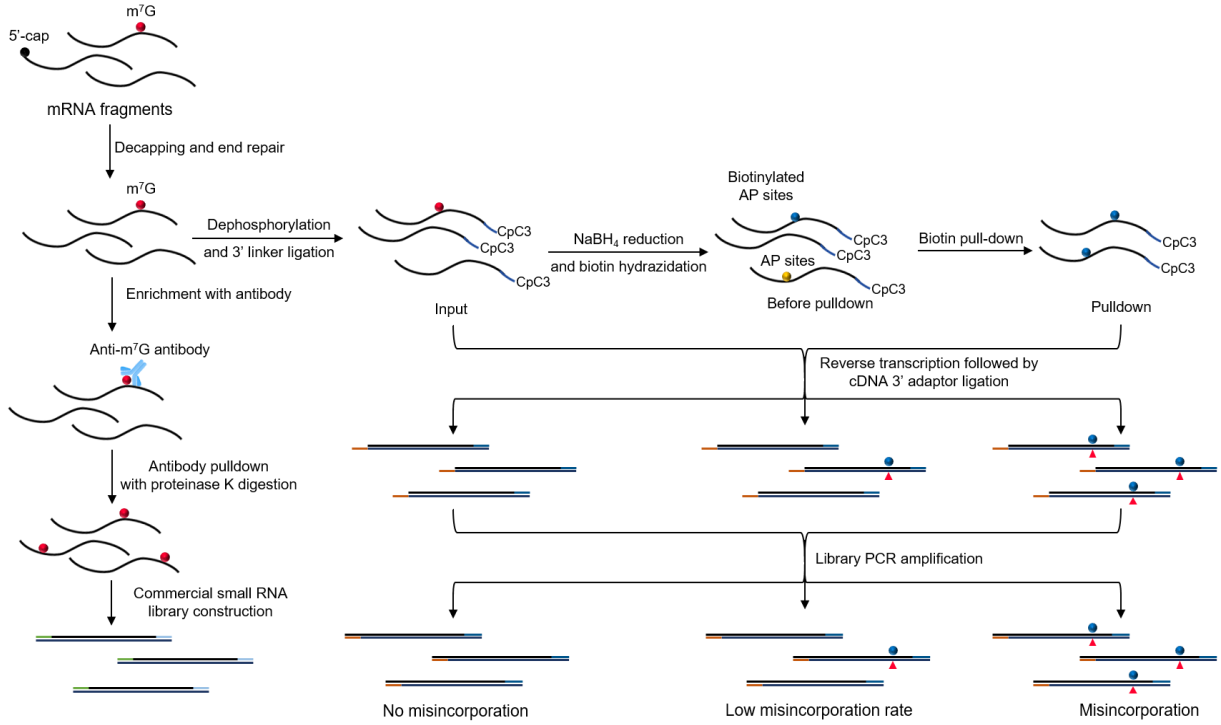

Figure 2.4.4. Schematic outline of m^7G-seq library construction. RNA fragments are firstly ligated at 3' end before the chemical reactions. They, whether undergo biotin pulldown or not, are then subject to reverse transcription before 3' adaptor ligation to the resulting cDNA. Only those fragments with internal m^7G sites are expected to generate misincorporation during reverse transcription that can be further detected using high-throughput sequencing.

2.5 Single-nucleotide resolution m^7G maps in human tRNAs: verification of m^7G-seq

We first applied m^7G-seq to ribosomal RNA and transfer RNAs isolated from HeLa, HepG2, and HEK293T cells and examined the known m^7G_{1639} site in human rRNA[127]. We observed highly increased mutation only at the known m^7G site. With reduction and biotinylation, we could obtain a misincorporation pattern of 18.0% G→T, 9.9% G→C, and 2.8% G→A at m^7G_{1639} site in HeLa rRNA; after pull-down enrichment of biotinyalted fragments, we detected a distribution of misincorporation with 38.2% G→T, 13.0% G→C and 3.7% G→A (**Figure 2.5.1C**), implying a predominant G→T and G→C mutation pattern induced by the chemical treatment followed by RT using HIV reverse transcriptase. The overall mutation rate is around 25-28% at the m^7G_{1639} site before enrichment across all three human cell lines (**Figure 2.5.1A**), suggesting nearly 100% modification at this site (**Figure 2.4.3**).

Our method also identified m^7G sites in 22 human tRNAs at the position 46 (**Figure 2.5.1A**). These tRNAs, with different sequence contexts around m^7G_{46}, varied in the misincorporation rate in a range of 5-31% before biotin enrichment (18-60% after enrichment, **Figure 2.5.1B**), suggesting an estimated high methylation status ranging from 60%-100% (**Figure 2.4.3**). Our observation of m^7G modification in tRNA is also consistent with a recent report based on a cleavage assay[137]. Especially, the misincorporation levels before pulldown enrichment, which indicate the methylation status, are correlated to sequence motifs around the methylated sites (**Figure 2.5.1D**). The methyltransferase at the position 46 displays diverse methylation efficiency in the presence of various motifs. These motifs are majorly related to AG(m^7G)U and AG(m^7G)C, while we also identified two new motifs as GG(m^7G)U and GA(m^7G)U.

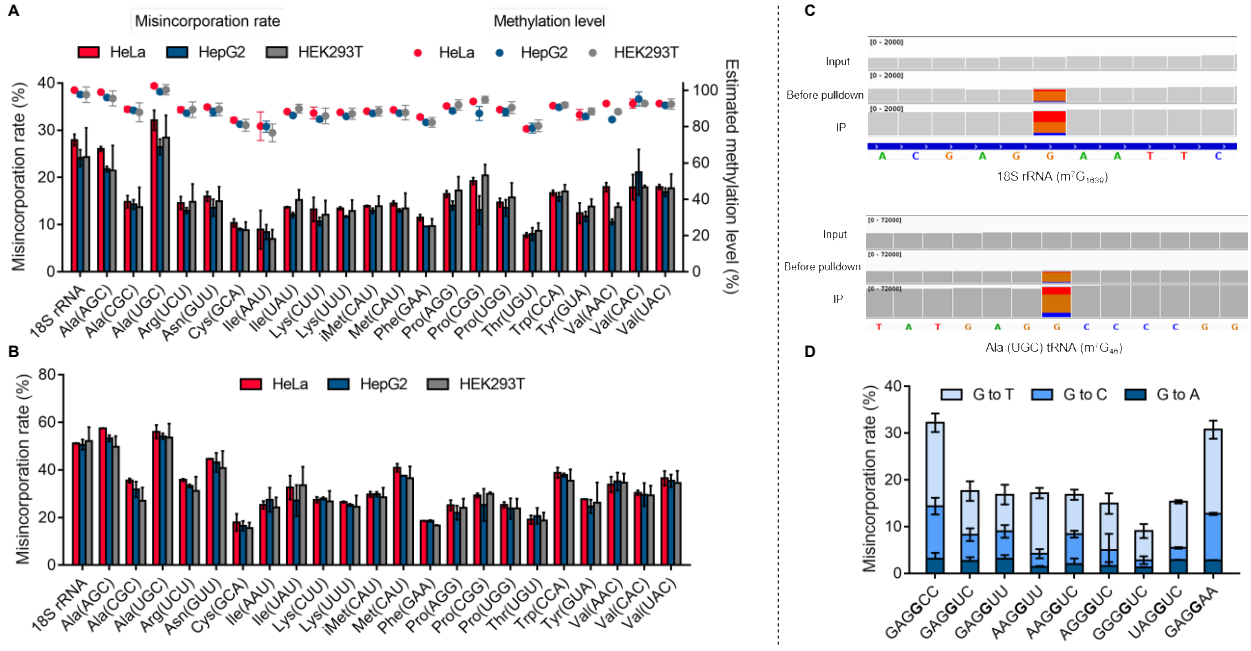

Figure 2.5.1. Base-resolution mappings of m^{7}G in human rRNA and tRNAs by m^{7}G-seq. (A) The misincorporation rates (without enrichment) for m^{7}G$_{1639}$ in human 18S ribosomal RNA and 22 cytoplasmic tRNAs containing internal m^{7}G$_{46}$ are shown in bars for three human cell lines. The corresponding estimated methylation levels are shown in dot. Mean values ± s.d. are shown, n = 2. (B) The misincorporation rate (after pulldown) for m^{7}G$_{1639}$ in human 18S ribosomal RNA and 22 tRNAs containing internal m^{7}G$_{46}$ are shown in bar for three human cell lines. Mean values ± s.d. are shown, n = 2. (C) IGV (Integrative Genomics Viewer, Broad Institute) plot of 18S rRNA m^{7}G$_{1639}$ (NR_038958) and tRNA(Ala) (UGC) m^{7}G$_{46}$ in HeLa cells (bottom two rows) showing the misincorporation only after chemical reduction, depurination and labeling before and after enrichment but not in the input samples before chemical treatment. (D) The misincorporation distribution pattern (without enrichment) shown for different internal m^{7}G motifs in tRNAs and 18S rRNA. Mean values ± s.d. are shown, n = 2.

As tRNAs tend to be heavily modified, we also characterized various mutation patterns of other G modifications to distinguish m^{7}G from other G modifications in order to avoid potential mis-assignments (**Figure 2.5.2A**). The dimethylated m$^2{}_2$G$_{26}$, which blocks Watson-Crick base pairing, showed high misincorporation rates throughout all three samples with a majority of G→T mutation up to around 90% as expected (**Figure 2.5.3**); however, the monomethylated m^{2}G$_{10}$ showed relatively low mutation rates below 5%.

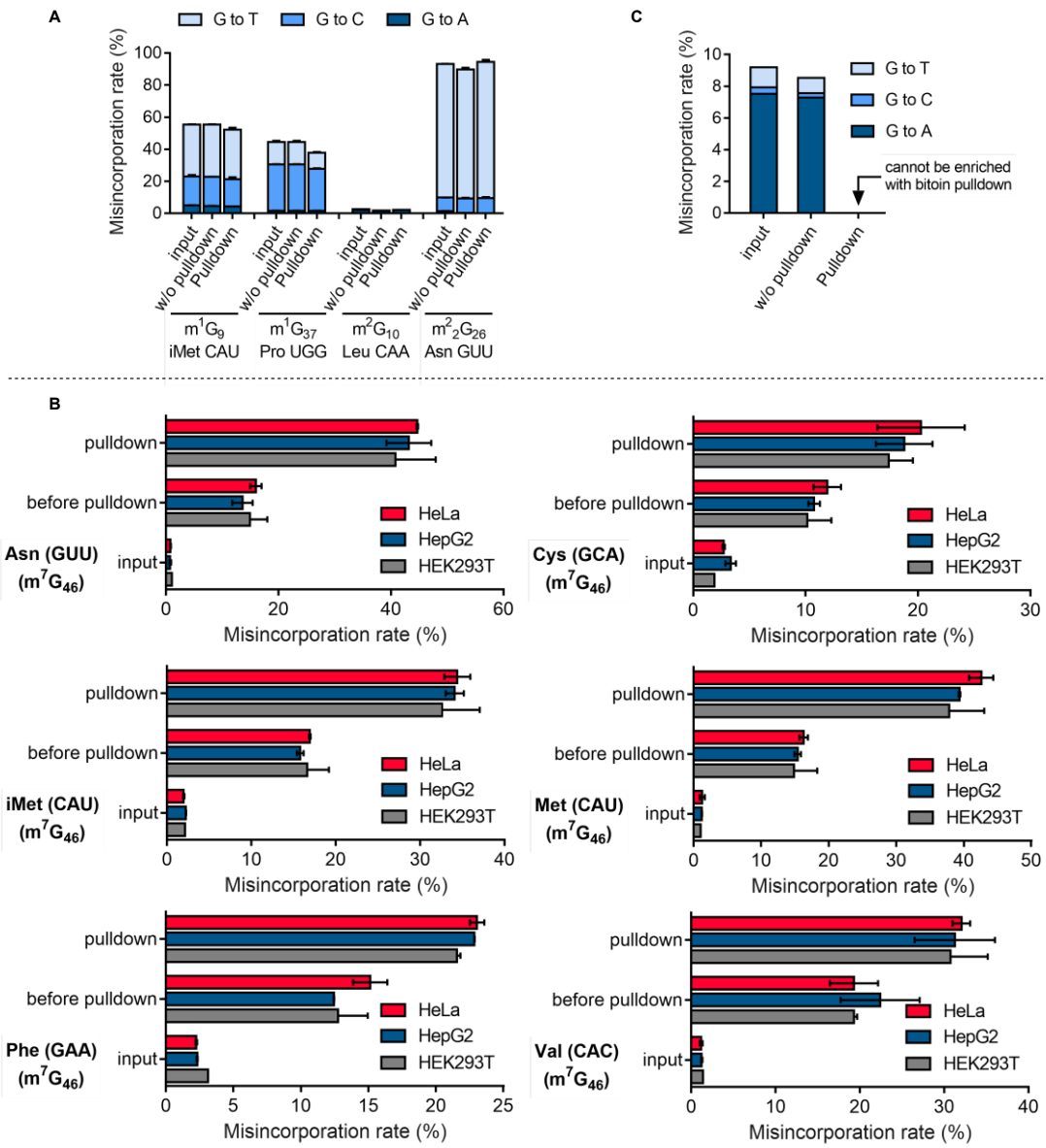

Figure 2.5.2. The unique misincorporation distribution pattern as a criterion to identify internal m^{7}G. (A) The misincorporation distribution patterns (input, before pulldown and after pulldown) for other G modifications in tRNAs. Mean values ± s.d. are shown, n = 2. (B) The overall misincorporation level (HIV as the reverse transcriptase) for known internal m^{7}G$_{46}$ in tRNAs (two replicates for HeLa cells (red), HepG2 cells (blue), and HEK293T cells (grey) in RNA input, before pulldown, and after pulldown). Mean values ± s.d. are shown. (C) The misincorporation distribution pattern (input, before pulldown and after pulldown) for m^{6}G modification in the synthetic probe containing a single m^{6}G in the middle. Mean values ± s.d. are shown, n = 2.

When we examined m^1G, the misincorporation patterns differed based on the modification locations. While both m^1G_9 and m^1G_{37} displayed the overall mutation rate around 40-60%, m^1G_9 tended to give higher G→T mutation but m^1G_{37} caused more G→C mutation in our assay (**Figure 2.5.4**). Importantly, none of these G methylations showed an increased misincorporation after reduction and depurination treatment compared with m^7G; we observed nearly no mutation in input samples, but high mutation in samples before enrichment and after pull-down only for m^7G. Therefore, we compare mutation rates of Gs before and after our chemical manipulation as a criterion to identify m^7G sites in our later analysis to exclude other guanosine modifications that could cause misincorporation during reverse transcription (**Figure 2.5.2B**). As O^6-methylguanosine (m^6G) was also observed in both DNA and RNA as a highly mutagenic base[138], we constructed a 60-mer probe with one O^6-methylguanosine (m^6G) in the middle by *in vitro*

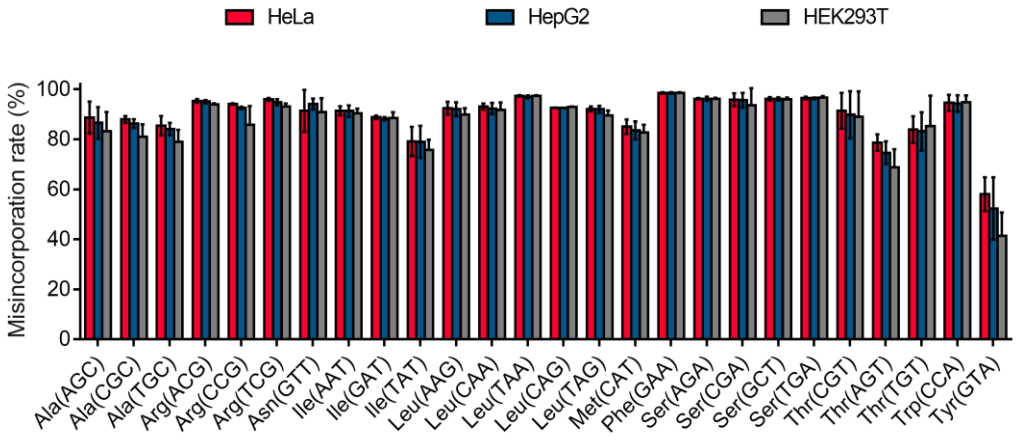

Figure 2.5.3. Misincorporation-based methodology recognizes $m^2{}_2G_{26}$ in human tRNAs. The misincorporation rates (RNA input) by HIV transcriptase for $m^2{}_2G_{26}$ in human cytoplasmic tRNAs containing internal G26 are shown in bars for three human cell lines. The corresponding estimated methylation levels are shown in dot. Mean values ± s.d. are shown, n = 2.

transcription. We subjected this probe to m^7G-seq treatment and high-throughput sequencing. The m^6G sited exhibited a mutation pattern with mainly G→A mutation (~7.5%) in both input and before pulldown samples but could not be enriched with biotin pulldown (**Figure 2.5.2C**). Therefore, we can further exclude m^6G in m^7G-seq.

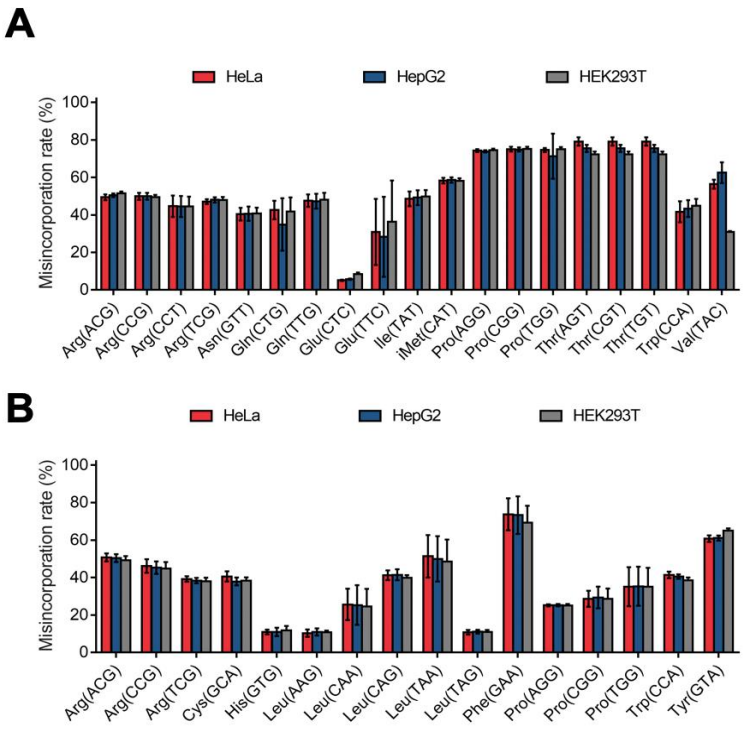

Figure 2.5.4. Misincorporation-based methodology recognizes m^1G_9 and m^1G_{37} in human tRNAs. (A) The misincorporation rates (RNA input) by HIV transcriptase for m^1G_9 in human cytoplasmic tRNAs containing internal G9 are shown in bars for three human cell lines. The corresponding estimated methylation levels are shown in dot. Mean values ± s.d. are shown, n = 2. (B) The misincorporation rates (RNA input) by HIV transcriptase for m^1G_{37} in human cytoplasmic tRNAs containing internal G37 are shown in bars for three human cell lines. The corresponding estimated methylation levels are shown in dot. Mean values ± s.d. are shown, n = 2.

Because mammalian m^7G_{46} sites in tRNAs are known to be installed by the METTL1/WDR4 complex[137,140]. We applied m^7G-seq to METTL1 knockdown cells with both transient knockdown (siRNA) in HeLa cells and stable knockdown (shRNA) in HepG2 cells to further validate the applicability of our method. Both siMETTL1 and shMETTL1 cells showed

obvious decreases in misincorporation rate at all m^7G_{46} sites (**Figure 2.5.5A** and **Figure 2.5.5B**), showing reduced m^7G modification fraction at these sites with METTL1 knockdown. Interestingly, the extent of reduction of m^7G is not universal with METTL1 knockdown: while $tRNA^{Ala(UGC)}$ and $tRNA^{Thr(UGU)}$ showed nearly 30% decreases in the methylation level, $tRNA^{Pro(HGG)}$ and $tRNA^{Val(AAC)}$ only showed a few percent m^7G decreases. Either METTL1/WDR4 has preferential tRNA targets or there could be additional m^7G methyltransferase(s) that could compensate m^7G methylation by METTL1, which should be investigated in the future.

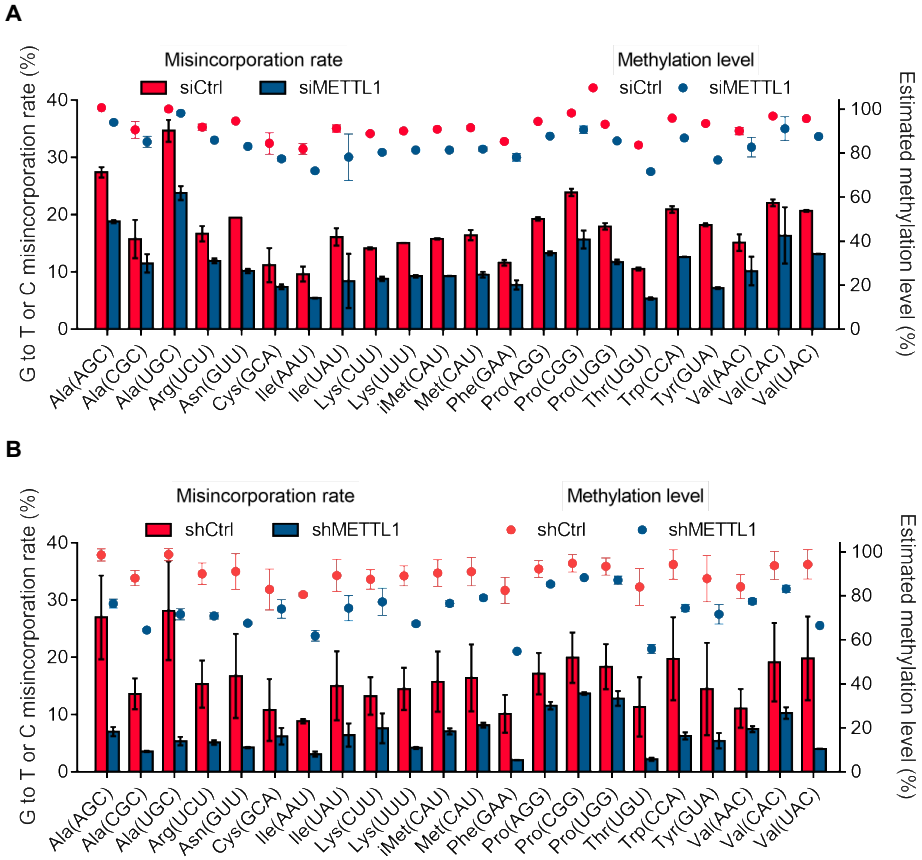

Figure 2.5.5. METTL1 serves as a methyltransferase for m^7G_{46} in human tRNAs. (A) The misincorporation rates and estimated methylation levels for m^7G_{46} in 22 human tRNAs upon METTL1 knockdown *versus* control in HeLa cells. Mean values ± s.d. are shown, n = 2. (B) The misincorporation rates and estimated methylation levels for m^7G_{46} in 22 human tRNAs upon METTL1 stable knockdown *versus* control in HepG2 cells. Mean values ± s.d. are shown, n = 2.

2.6 Single-nucleotide resolution internal m^7G methylome in mRNA

We next applied m^7G-seq without pull-down enrichment to mRNAs purified from HeLa and HepG2 cells, as an attempt to identify frequently methylated sites. We used 4% mutation rate above background input mutation as a cutoff and computationally identified over 90 sites in mRNA without the enrichment step; these sites potentially possess estimated methylation levels around or above 20% based on estimations using our calibration curve (**Figure 2.4.3**) and comparison to tRNAs. We further identified ~30 sites manually with modifications estimated to be around or above 50% (Table 2.6). These sites showed misincorporation rates and possess m^7G methylation levels approaching those in tRNAs (**Figure 2.5.1** and **Figure 2.5.2**). Representative mRNA internal m^7G sites conserved in HeLa and HepG2 cells are shown in **Figure 2.6.1**. The AG(m^7G)A motif shown in CNOT2 (5' UTR) and EIF2S2 (CDS) resembles the same motif for rRNA m^7G_{1639}. Our m^7G-MeRIP-seq found dominant GA-enriched motif, which was also discovered in a number of purine-enriched region, with GG(m^7G)AA in 3' UTR of GPR107 and UA(m^7G)AA in 3' UTR of RTN4 as examples. We also found motifs such as AUCG(m^7G)A and UU(m^7G)AU with high modification fractions conserved in both cell lines, suggesting presence of multiple methyltransferases that specifically install these sites. The IGV (Integrative Genomics Viewer, Broad Institute) plot of raw data depicts the misincorporation image of mRNA internal m^7G (**Figure 2.6.2**).

Coordinates	Gene	Location	M1 (%) *	M2(%) **	Sequence
chr9: 132897621 (+)	GPR107	3'UTR	80	80	GGCGGGAAUGG
chr2: 55199948 (-)	RTN4	3'UTR	80	75	AGCUAGAAAAA
chr2: 110373764 (+)	SOWAHC	3'UTR	80	80	GAUCGGAAUGG
chr12: 70671991 (+)	CNOT2	5'UTR	80	70	GUGAGGAAAGG
chr5: 150429427 (-)	TNIP1	CDS	75	70	CCGAAGAUGGA
chr1: 120839457 (+)	FAM72B	5'UTR	75	60	GAUCGGAGAAG
chr6: 114292322 (-)	HDAC2	5'UTR	70	75	AGGUGGUUUGG
chr7: 128098281 (+)	HILPDA	3'UTR	70	75	GGUGGGCUGUU
chr20: 32677658 (-)	EIF2S2	CDS	70	65	AGAAGGACACA
chr9: 96327971 (+)	FAM120A	3'UTR	60	55	CCUUUGAUCUU

Figure 2.6.1. Representative highly methylated internal m^{7}G sites in human mRNA. These sites are shared in both HeLa and HepG2 cells identified using the base-resolution m^{7}G-seq in replicates (without enrichment). Sequence motifs at base resolution and the estimated methylation levels in two human cell lines are also shown.

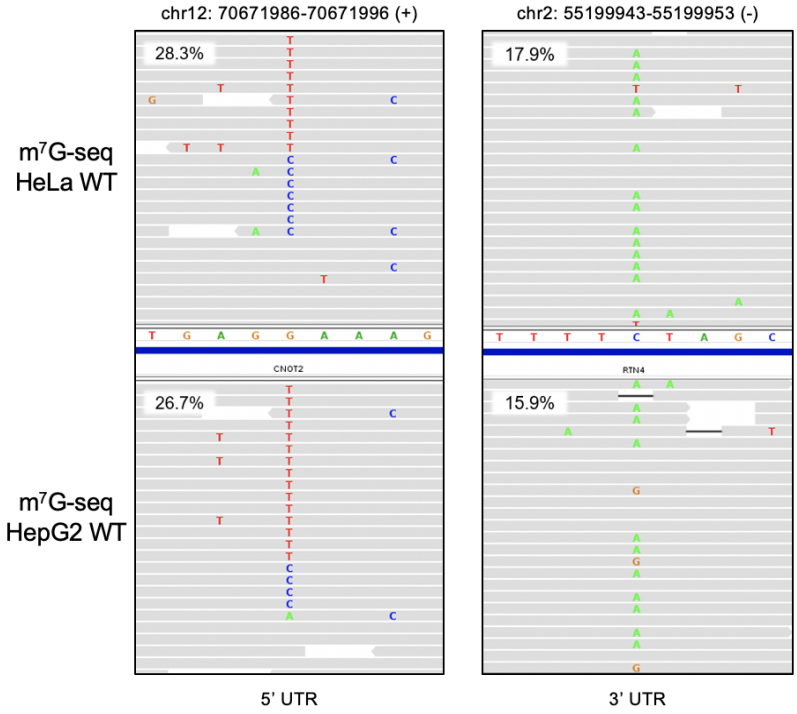

Figure 2.6.2. IGV plot of representative internal m^{7}G sites in mRNA. The sites are shared in HeLa cells (upper row) and HepG2 cells (bottom row) showing specific misincorporation in pulldown samples.

Considering that m^7G may play functions at specific cellular locations (i.e. chromatin associated RNA or nascent RNA), we also wanted to identify lowly to moderately modified m^7G sites from the whole-cell mRNA and compare with m^7G-MeRIP-seq results. We then performed enrichment of the biotin-tagged mRNA fragments obtained from m^7G reduction and depurination, followed by reverse transcription using HIV RT enzyme. We used 8.0% mutation rate increase over the background input control as our cutoff because the modified m^7G sites show higher mutation rate after enrichment. We identified around 3500 internal mRNA m^7G sites in HeLa cells from overlap of two replicates. We then overlaid HeLa m^7G sites to those shared in HepG2 replicates and identified 801 overlapped ones. Nearly 70% of these 801 single sites overlap with targets revealed from m^7G-MeRIP-seq results, showing consistency between two orthogonal approaches. It should be noted that m^7G-MeRIP-seq can enrich m^7G sites ranging from low modification fraction to high modification fraction, as well as potential non-specific binding. The base resolution m^7G-seq utilizes mutation over background input to detect m^7G with much higher accuracy, but it may miss most lowly modified sites due to inadequate sequencing depth and our stringent mutation cutoff criterion. HIV reverse transcriptase itself is known to generate baseline mutation rate, thus applying a stringent mutation rate cutoff is important to ensure reliable assignments of m^7G sites despite potential loss of lowly modified sites.

We reasoned that some of the m^7G sites identified by m^7G-seq without overlap with m^7G-MeRIP-seq peaks may still be genuine m^7G sites. They may exist in structured RNA which hindered the antibody-based pulldown. Nevertheless, to obtain more accurate assignments we decided to assign m^7G sites not only shared in two human cell lines but also overlapped with m^7G-MeRIP-seq peaks as confident sites. Gene ontology (GO) analysis uncovered the similar enriched GO terms of RNA processing and cell-cell adhesion from these 807 sites, which corre-

lates well with the m^7G-MeRIP-seq data (**Figure 2.6.3A**). Half of the more confident m^7G sites exist in CDS and the rest show almost equal distribution at 5'UTR and 3'UTR (**Figure 2.6.3B**). On average we observed mostly one methylated peak per gene and this is true to all three transcript segments (**Figure 2.6.3F**). The general misincorporation level ranged from 8%-23% after pull-down enrichment (**Figure 2.6.3D**). We then analyzed the sequence motifs of the base-resolution mRNA internal m^7G sites. The main motifs were found to be G(m^7G)A, A(m^7G)A, C(m^7G)C, G(m^7G)G, C(m^7G)U and G(m^7G)C, which showed wider variations than those in tRNAs (**Figure 2.6.3C**). We picked up six representative motifs and plotted them along 5' UTR, CDS, and 3' UTR. The G(m^7G)G motif appears to be more likely at 3' UTR, while the C(m^7G)C and G(m^7G)C motif accumulates more at 5' UTRs and barely appears at 3' UTR (**Figure. 2.6.3E**).

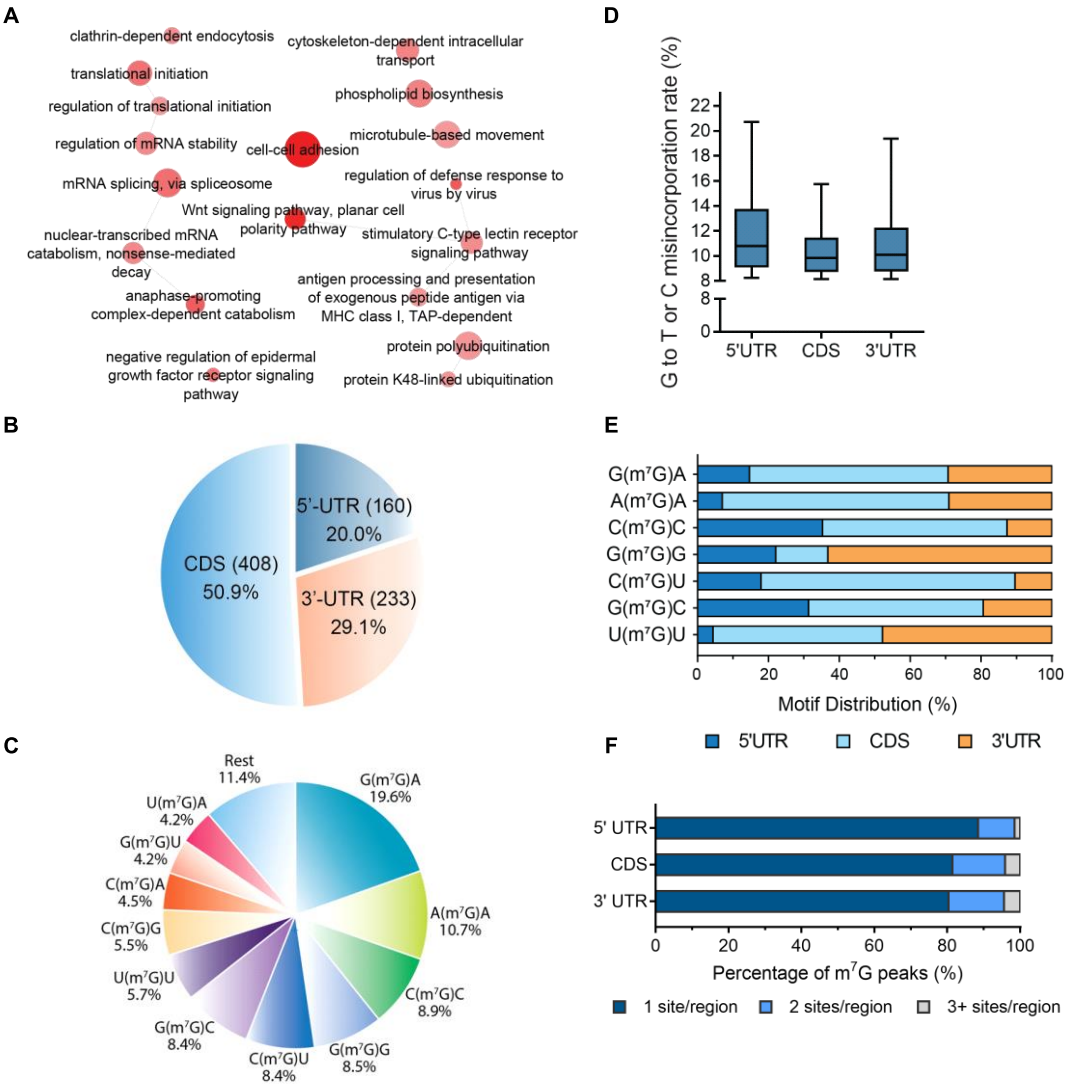

Figure 2.6.3. Single-nucleotide resolution m^{7}G maps in human mRNAs. (A) GO analysis for 803 m^{7}G sites from high confident base-resolution m^{7}G-seq results that are shared between HeLa and HepG2 cells in replicates as well as overlapping with m^{7}G-MeRIP peaks. (B) Pie chart showing the fraction of 803 m^{7}G peaks in each of three non-overlapping transcript segments (5' UTR, CDS and 3' UTR). (C) Pie chart displaying the fraction of main internal m^{7}G motifs in mRNA based on confident targets. (D) The misincorporation rates for internal mRNA m^{7}G sites in three non-overlapping transcript segments. Two sided Mann-Whitney test. (E) m^{7}G peak distribution of the top seven m^{7}G motifs in mRNA along a normalized transcript composed of three rescaled segments. (F) The percentage of methylated genes possessing confident m^{7}G sites with 1, 2, or 3+ sites per region in three transcript segments (5' UTR, CDS and 3' UTR)

2.7 Conclusion and Discussion

Although m^6A is the most prevalent modification in mammalian mRNA[83,114,140,141], other modifications have been shown to exist in mRNA and play functional roles[95-109]. In this study, we developed a decapping/mass-spec protocol to expose signals only from internal m^7G that is resistant against decapping enzymes. We found m^7G/G level of ~0.02% to ~0.05% across human and mouse cell lines as internal N^7-methylguanosine within mammalian mRNA. We then profiled its distribution within mammalian mRNA using m^7G-MeRIP-seq and unveiled GA or GG enriched sequence motifs. The MeRIP sites overlapped well among three human cell lines and between two mouse cell lines. Between human and mouse cell lines, ~40% overall conservation was observed for m^7G peaks at the orthologous positions.

We then developed m^7G-seq to achieve transcriptome-wide mapping of internal m^7G at single-nucleotide resolution in human cells without using the antibody-based enrichment. The chemical-assisted m^7G-seq takes advantage of the unique reactivity of the positively charged five-membered ring of m^7G that is prone to reduction and depurination, leading to the formation of a biotinylated abasic site; the abasic site can then be labeled with biotin, enriched or without enriched and read as a mismatch during reverse transcription, resulting in the identification of the m^7G site. The method offers two noticeable advantages of: (1) converting m^7G sites to abasic sites for biotin-based enrichment followed by RT-mediated mutation to map m^7G sites at base resolution; and (2) also estimating the methylation status of highly methylated sites (without enrichment) to uncover frequently modified m^7G sites in human mRNA. Therefore, this new approach enables us to map the precise location and sequence motif of the internal m^7G modification with high specificity and fidelity. Without enrichment the method also allowed assessment

of the modification fraction at m^7G sites by correlating the misincorporation rate to the methylation level using a calibration curve generated using m^7G-containing RNA standards.

Applying the m^7G-seq approach (without enrichment) to tRNAs, we successfully characterized the base-resolution human cytosolic tRNA m^7G_{46} methylation, which clearly distinguished m^7G from other guanosine modifications (including m^1G, m^2G or $m^2{}_2G$). These tRNA m^7G sites all exist with high modification fractions. We also performed m^7G-seq in METTL1 knockdown cells and confirmed the tRNA m^7G methylation function of METTL1. The ability to monitor the m^7G methylation status by m^7G-seq may aid future functional studies of tRNA and rRNA m^7G methylation and monitor potential dynamics of these modified sites during cellular processes.

Our method also uncovered over 90 internal m^7G sites within human mRNA that may exhibit around to above 20% modification fraction. Approximately 30 of them can be considered as highly methylated internal m^7G sites in human mRNA. The identification of these highly methylated sites, conserved in two different human cell lines, will stimulate further functional characterizations of these modifications and the target mRNAs.

We further conducted biotin pull-down enrichment to unveil more internal m^7G sites and to verify the sites discovered before enrichment. After pull-down enrichment, the misincorporation rate difference (above the input baseline mutation rate potentially introduced by HIV RT) was elevated to 8-25% (abasic sites were further enriched). After applying stringent cutoff, we were able to identify more than 800 sites shared by two human cancer cell lines, from which we performed statistical analysis for features of these confident internal m^7G sites. These sites exhibit additional sequence motifs beyond the known GA- or GG-containing motifs found in

tRNA and uncovered for mRNA using m^7G-MeRIP-seq, suggesting potential presence of multiple m^7G methyltransferases that may install mRNA m^7G.

The current m^7G-seq protocol still has limitations. We noticed that the mild chemical reactions for selective m^7G reduction and depurination could not achieve quantitative yields, with a portion of m^7G sites converted to abasic sites and then a portion of these generating abasic sites that could be labeled with biotin hydrazide. These explain higher mutation rates after enrichment of biotin-labeled RNA than those before enrichment. This limitation will affect accurate measurements of the modification levels of individual m^7G sites. We have performed reactions on calibration and real biological samples under consistent conditions that should normalize variations introduced by the depurination and labeling chemistry. However, further improvement of the reaction efficiency and robustness as well as calibration consistency would further improve the quantitative feature of m^7G-seq.

2.8 Materials and Methods

Cells culture

Human HeLa, HepG2, and HEK293T cell lines and mouse embryonic stem cells (mESC) and mouse embryonic fibroblasts (MEF) used in this study were all purchased from ATCC (the American Type Culture Collection). HeLa cell line was grown in DMEM (Gibco, 11965) media supplemented with 10% FBS and 1% 100X Pen/Strep (Gibco), while HepG2, HEK293T and MEF cell lines were maintained in DMEM (Gibco, 11995), supplemented with 10% FBS and 1% 100X Pen/Strep (Gibco). mESC cells were grown under a typical feeder-free culture condition: cells were cultured in DMEM (Gibco, 11995), supplemented with 15% FBS, 1% Pen/Strep (Gibco), 1X Glutamax (Gibco), 1X non-essential amino acids (Gibco), 1X 2-Mercaptoethanol (Gibco), and 1000 U/ml leukemia inhibitory factor (Millipore, ESG1107), together with two inhibitors: 3 μM CHIR99021 (STEMCELL Technologies, dissolved in DMSO) and 1 μM PD0325901 (STEMCELL Technologies, dissolved in DMSO). mESC cells were distinctively cultured on gelatinized culture plates (0.2% Gelatin). The culture of mESC cells were passaged every 2 days. All cells were cultured at 37 °C under 5.0% CO_2.

Antibodies

The antibodies used in this study are listed below in the format of name (catalog; supplier; dilution fold): Mouse anti-m^7G (RN017M; MBL; 1000). Mouse anti-METTL1 (11525-MM05; Sino Biological; 1000). Rabbit anti-m^6A (E1610S; NEB; 1000). Mouse anti-WDR4 (sc-100894; Santa Cruz; 100). Mouse anti-RPS6 (sc-74459; Santa Cruz; 500). Goat anti-mouse lgG-HRP (sc-2005; Santa Cruz; 3000). Goat anti-GAPDH-HRP (A00192; Genscript; 2000).

RNA isolation

Generally, with harvesting cells as the first step, the media was aspirated, and the cells were washed once with proper volume of ice-cold DPBS buffer for each plate. Then total RNA was isolated from cells with TRIzol reagent (Invitrogen) and then extracted following manufacturer's protocol by isopropanol precipitation. Total RNA isolation for RT-qPCR: total RNA was isolated from wild-type or transiently transfected cells with TRIzol reagent (Invitrogen) and then extracted with Direct-zol RNA MiniPrep (Zymo Research), including an on-column DNase I digestion. mRNA isolation for LC-MS/MS: total RNA was extracted as described previously. mRNA was extracted by two rounds of polyA$^+$ purification with Dynabeads mRNA DIRECT kit (Ambion). mRNA concentration was measured using Qubit RNA HS Assay Kit (Thermo Fisher Scientific) with Qubit 2.0 fluorometer. General procedures for ethanol precipitation: 0.1 volume of 3M NaOAc (pH=5.5), 3 volumes of 100% ethanol and 1 μL GlycoBlue Coprecipitant (15 mg ml^{-1}, Thermo Fisher Scientific) were added to the RNA solution for precipitation. The solution was stored at -80 °C for 1 h or up to overnight, and was then centrifuged at 15,000 g for 15 min. The supernatant was removed, and the pellet was washed with 1 ml 80% ethanol once, air dried and finally dissolved in nuclease-free water.

rRNA and tRNA depletion evaluation

Four-leaf clover qRT-PCR was performed as previously reported (Honda et al., 2015) for RNA after two rounds of polyA$^+$ purification to check the remaining level of tRNA^{LysCUU} and tRNA-Val$^{AAC/CAC}$, two tRNAs known to bear m^{7}G. Mature tRNAs in total RNA were first subjected to deacylation treatment. Then the DNA/RNA hybrid SL-adaptors were specifically hybridized to the deacylated tRNA, followed by ligation with T4 RNA ligase 2 (NEB) to form a "four-leaf

clover" structure. Two biological replicates and three technical replicates were applied. All tRNA samples (total RNA or mRNA) after the ligation were subject to 50 cycles of RT-qPCR amplifications with corresponding pairs of forward and reverse primers. RT-qPCR was performed to test the rRNA depletion. The two RT-qPCR primer pairs for 18S rRNA detection listed as shown in Table S4.

The two replicates for both total RNA and mRNA were subjected to 25-cycle amplification of RT-qPCR with technical triplicates. The Cq value of 18S rRNA was normalized to GAPDH and the depletion level was calculated.

LC-MS/MS

Around 200-300 ng mRNA was digested in a three-step manner with nuclease S1 (1uL, Sigma-Aldrich) in 20 µL reaction buffer containing 10 mM of NH_4OAc (pH= 5.3) at 42 °C for 2 h. To detect m^7G level, including the cap of mRNA, a second digestion with phosphodiesterase I (0.001U, Sigma-Aldrich) was performed, with the addition of 2.5 µl NH_4HCO_3 (1M, freshly prepared in water) at 37 °C for 1 h. Then, 1 µL of shrimp alkaline phosphatase (rSAP, NEB) was added along with 2.5 µL of 10X CutSmart buffer (NEB) and incubated at 37 °C for 2 h. After the incubation, the sample was diluted with additional 35 µL water and filtered with 0.22 µm filters (4mm diameter, Millipore) and 8 µl of the entire solution was injected into LC-MS/MS as one sample. The two-step digestion simply excludes the phosphodiesterase I digestion. For all the quantification, a mock control with only digestion buffers and enzymes was included each time and was later used for subtraction of baseline signals. Nucleosides were separated, by reverse phase ultra-performance liquid chromatography, on a C1 column with on-line mass spectrometry detection by an Agilent 6410 QQQ triple-quadrupole LC mass spectrometer, in positive elec-

trospray ionization mode. The nucleosides were quantified with retention time and the nucleoside-to-base ion mass transition of 284-152 (G), 282.1-150.1 (m^6A), 268-136 (A), 298.1-166.1 (m^7G). Quantification was performed in comparison with the standard curve, obtained from pure nucleoside standards running with the same batch of samples. The m^7G level and m^6A level were calculated as the ratio of m^7G to G and m^6A to A based on calibrated concentration curves.

Decapping of mRNA

Decapping of mRNA was performed with Tobacco Decapping Plus 2 (#94, Enzymax). The reaction was prepared, with a maximum of 6 µg fragmented mRNA in nuclease-free water with 5 µL 10X Decapping Reaction Buffer (100 mM Tris-HCl pH 7.5, 1.0 M NaCl, 20 mM $MgCl_2$, 10 mM DTT), 1 µL 50 mM $MnCl_2$, 2.5 µL SUPERase-In RNase Inhibitor (Thermo Fisher Scientific) and 8 µL Tobacco Decapping Plus 2 enzyme, diluted to a final volume of 50 µL. The reaction was incubated at 37 °C for 2 hours. Decapped RNA was extracted from the solution with Oligo Clean & Concentrator (Zymo Research).

Dot blot assay

Four 9-mer RNA oligos were synthesized in-house with either G, m^1G, m^2G, or $m^2{}_2G$ at a single internal site (5'-GAGCXUUAG-3', X = G, m^1G, m^2G or $m^2{}_2G$). Three 60-mer oligos (5'-CCAATAAAATATTAACCACCAATAAAATATTAACCAA**G**ATCCACCAATAA AA-TATTAACC-3') were synthesized by *in vitro* transcription using MEGAshortscript T7 Transcription Kit (AM1354, Thermo Fisher Scientific) with GTP or m^7GTP (M6133, Sigma Aldrich) or m^6GTP (N-1031, TriLink) as sources for the single G site in the middle. All seven oligos were used to test the specificity of the commercially available m^7G antibodies used in m^7G-MeRIP-

seq. Serial portions, in two-fold dilution of each oligo, were denatured and spotted on the nylon membrane (GE Healthcare) followed by UV-crosslinked. The membranes were then blocked with 5% non-fat milk in 1X PBST for 30 min at room temperature and incubated overnight afterward at 4 °C in 1% non-fat milk with the addition of anti-m^7G antibody (RN017M, MBL). After extensive washing, three times with 1X PBST, the membranes were incubated with anti-mouse IgG HRP conjugated secondary antibody diluted 1:3000 in 1% non-fat milk in 1X PBST for 1 hour at 25 °C. After three more washes and developing with SuperSignal West Dura Extended Duration kit (Thermo Fisher Scientific), the membranes were finally imaged and analyzed using the FluorChem R imager (Protein Simple).

m^7G-MeRIP-seq

Around 4-6 µg human or mouse mRNA (two rounds of $polyA^+$ purification) was fragmented into 50-100 nt using RNA Fragmentation Reagents (AM8740, Invitrongen) following the manufacturer's standard protocol. The fragmented mRNA was further decapped with the standard decapping procedure (Tobacco Decapping Plus 2, Enzymax) and then concentrated with RNA Clean & Concentrator (Zymo Research).

The end structures of these RNA fragments were repaired by T4 Polynucleotide Kinase (EK0032, Thermo Fisher Scientific): (1) 3' de-phosphorylation: RNA was mixed with 5 µL 10X T4 Polynucleotide Kinase Reaction Buffer (B0201S, NEB) and 5 µL T4 PNK, diluted to a final volume of 50 µL and incubated at 37°C for 1 hour; (2) 5'-phosphorylation: to the reaction mixture, 5 µL 10 mM ATP (P0756S, NEB) and 2.5 µL extra T4 PNK were added, and the mixture was kept at 37 °C for 30 min.

The repaired mRNA fragments (~2 μg) was then incubated with 4 μL anti-m^7G antibody (MBL) in 250 μL 1X IPP buffer (10mM Tris-Cl, pH=7.4; 150 mM NaCl; 0.1% NP-40) with freshly added 5% SUPERase-In RNase inhibitor (Thermo Fisher Scientific) at 4 ℃ for 2-4 hours. Then 40 μL Dynabeads Protein G resins (Thermo Fisher Scientific) were washed twice with 1X IPP buffer, resuspended in 20 μL IPP buffer and added into the antibody-RNA mixture for another 2 hours at 4 °C. The resins were then washed with 1X IPP buffer at 4 °C for four times. RNA was finally eluted with Proteinase K (recombinant, PCR grade, EO0491, Thermo Fisher Scientific) digestion. 45 μL 1X Proteinase K digestion buffer (2X recipe: 2% SDS, 12.5 mM EDTA, 100 mM Tris-Cl (pH=7.4), 150 mM NaCl) with 5 μL Proteinase K was used to resuspend the resins, and the solution was incubated at 55 °C for 30 min. The second round of the digestion was prepared in the same way and incubated at 55 °C for another 15 min. Both flow-throughs were combined, and m^7G-containing RNA was recovered with RNA Clean & Concentrator (Zymo Research). Decapped and fragmented $polyA^+$ RNA (as "input") and immunoprecipitated RNA (as "IP") were subjected to small RNA library construction following manufacturer's instructions of NEBNext Small RNA Library Prep Set for Illumina (NEB). All libraries were sequenced on Illumina NextSeq 500 with single-end 80 bp read length.

m^7G-seq

Around 4-6 μg human mRNA (two rounds of $polyA^+$ purification) was fragmented into 50-100 nt using RNA Fragmentation Reagents (AM8740, Invitrongen) following the manufacturer's standard protocol. The fragmented mRNA was further decapped with the standard decapping procedure (Tobacco Decapping Plus 2, Enzymax) and then concentrated with Oligo Clean and Concentrator (Zymo Research). The end repair/3'-dephosphorylation was performed with T4

Polynucleotide Kinase (EK0032, Thermo Fisher Scientific): RNA was mixed with 5 μL 10X T4 Polynucleotide Kinase Reaction Buffer (B0201S, NEB) and 5 μL T4 PNK, diluted to a final volume of 50 μL and incubated at 37°C for 1 hour. Then a further 5'-/3'-dephosphorylation step was conducted at 37 °C for 1.5 hour in 50 μL 1X CutSmart Buffer (NEB) with 5 μL Shrimp Alkaline Phosphatase (rSAP, NEB). Starting with the 3'-adapter ligation, repaired and dephosphorylated RNA fragments (20 μL) were incubated with 1.6 μL 100 μM (2 equivalents) 3' linker (5'rApp-AGATCGGAAGAGCGTCGTG-3SpC3) at 70 °C for 2 min and transferred to ice immediately. 5 μL 10X T4 RNA Ligase Reaction Buffer, 15 μL 50% PEG8000 and 4 μL T4 RNA ligase 2 truncated KQ (NEB) were added accordingly to the RNA-adaptor mixture, then the reaction was diluted to a final volume of 50 μL and incubated at 25 °C for 2 hours followed by 16 °C for 8 hours or more. The reaction was then diluted to 97 μL, and the excessive adaptors were digested with 3 μL 5' Deadenylase (NEB) at 30 °C for 1 h followed by adding 1.5 μL RecJf (NEB) for ssDNA digestion at 37 °C for another hour. 3'-end ligated RNA was extracted using RNA Clean & Concentrator (Zymo Research). Save 50 ng as "input"

To enrich m^7G sites, RNA was reduced, biotinylated, and pulled-down in a row. RNA was subjected to reduction by 1 M $NaBH_4$ (freshly prepared in water) as a 5X reaction buffer and incubated at 25 °C for 45~60 min, with occasional low-speed shaking. RNA was recovered with 7 volumes (to quench the reaction) of RNA Binding Buffer provided by RNA Clean & Concentrator (Zymo Research) and the rest steps remained the same as the protocol. We added 5 μL 1 M MES buffer (pH=4.5) to the eluted RNA in a volume of 35 μL and mixed it immediately followed by adding 10 μL EZ-Link Hydrazide-Biotin (Thermo Fisher Scientific, 50 mM solution in DMSO). The solution was incubated at 55 °C for 1 hour. RNA was purified quickly (Save 50 ng as "before pulldown") and proceeded to pull-down assay with Dynabeads MyOne Streptavidin

C1 beads (Thermo Fisher Scientific). 20 µL resins was added. Respectively, resins were washed twice with 1X B&W buffer (as the manufacturer's protocol), resuspended in 2X B&W buffer (with a volume equal to the sample), and incubated with the sample at 4 °C for 20 min. After pull-down, the resins were washed five times with 1X Wash buffer (50 mM Tris, pH 7.4, 300 mM KCl, 0.05% (v/v) NP-40) and subjected to proteinase K digestion. The digestion reaction was prepared as previously mentioned and incubated at 55 °C for 30 min. The flow-through was saved and RNA was recovered with RNA Clean & Concentrator (Zymo Research).

The 3' ligated RNA (as "input"), the RNA before pull-down assay (as "before pull-down"), and final elution after pull-down (as "pulldown") were subjected to reverse transcription. RNAs were first incubated with 1 µL of 2.0 µM RT primer (5'-ACACGACGCTCTTCCGATCT-3') at 65 °C for 2 min and cooled on ice immediately. The reactions were then prepared with RNA mixture, 2 µL 10 mM dNTP (Deoxynucleotide (dNTP) Solution Mix, NEB), 2 µL 10X AMV Reverse Transcriptase Reaction Buffer (NEB), 1-2 µL Recombinant HIV reverse transcriptase (Worthington), and 1 µL RNaseOUT recombinant ribonuclease inhibitor (Thermo Fisher Scientific) and then incubated the 20 µL reaction mixture at 37 °C for 1 hour. The enzymes were deactivated by incubation at 75 °C for 5 min and stored on ice. The cDNAs were purified with Oligo Clean and Concentrator (Zymo Research). Purified cDNA was then subject to cDNA 3'-adaptor ligation. The cDNA was first denatured with 1 µL of 50 µM cDNA linker (5' Phos-NNNNNNNNNNAGATCGGAAG AGCACACGTCTG-3SpC3) at 75 °C for 2 min. 2 µL 10X T4 RNA Ligase Reaction Buffer, 10 µL 50% PEG8000, 2 µL 10mM ATP and 1 µL T4 RNA ligase 1 (high concentration, NEB) were added accordingly to the cDNA-adaptor mixture, then the 20 µL reaction was incubated at 25 °C for 12 hours. The library was then amplified with the universal and indexed primers from NEBNext Multiplex Oligos for

Illumina (NEB). All libraries were sequenced on Illumina NextSeq 500 with single-end 80 bp read length.

Misincorporation-Methylation calibration curve with synthetic 60-mer probe

A 60-mer RNA probe with the sequence as CCAATAAAATATTAACCACCAATAAAA-TATTAACCAA**G**ATCCACCAATAAAATATTAACC was synthesized with MEGAshortscript T7 Transcription Kit (Thermo Fisher Scientific) by *in vitro* transcription. A series of probes with different methylation levels was prepared by manipulating m^7GTP/GTP ratio during the *in vitro* transcription. All these probes were subjected to m^7G-seq and the misincorporation pattern of each was analyzed and a fitting curve was plotted based on misincorporation rate-methylation level relation. All libraries were sequenced on Illumina NextSeq 500 with single-end 80 bp read length.

Peak calling

Peak calling is adopted in the analysis of MeRIP-seq data. Adaptors and low-quality reads (length shorter than 15) were trimmed from raw sequencing reads using cutadapt software (https://cutadapt.readthedocs.io/en/stable/#). Reads were then aligned to the relevant genome (human-hg19 or mouse-mm10) using Tophat2 (version 2.2.1) without any gaps and allowed for at most two mismatches. Peaks enriched in immunoprecipitation over input experiments were identified using exomePeak and MeTPeak packages (Meng et al., 2013; Cui et al., 2016). For each cell line, only peaks identified (FDR≤0.05) in both replicates were considered as a candi-date target. The further annotation of peaks was made use of PAVIS tool and HOMER software (http://homer.ucsd.edu/homer/ngs/rnaseq/index.html). For the human-mouse conservation, MEF

and mESC mRNA m^7G peaks were converted to the homologous coordinates in the human hg19 genome, using the LiftOver tool of the UCSC genome browser (https://genome.ucsc.edu/cgi-bin/hgLiftOver). The converted peaks were compared with human internal mRNA m^7G peaks with bedtools (https://bedtools.readthedocs.io/en/latest/). METTL1-regulated internal m^7G peaks were analyzed by exomePeak and MeTDiff packages (Meng et al., 2013; Cui et al., 2018), based on the MeRIP-seq data from HeLa siMETTL1 *vs* siControl samples or HepG2 shMETTL1 *vs* siControl samples.

Identification of m^7G-seq induced mutation

Identification of m^7G-seq induced mutation. RNA input samples (without reduction or biotinylation, as "input"), with reduction and biotinylation samples (before biotin pull-down, as "before-pulldown"), and biotin pull-down samples (as "pulldown") were all trimmed with cutadapt tool to remove adapters and low-quality reads (length shorter than 15). Reads were then processed with BBMap tool (https://sourceforge.net/projects/bbmap/) to remove PCR duplicates. The harvested reads were further trimmed the random barcode using cutadapt tool and remove low-quality reads again. Reads were then aligned to the relevant genome (human-hg19) using Tophat2 (version 2.2.1) with at most four mismatches (Arango et al., 2018). Identification of sequence variants was carried out by determining the base composition at each position using fine-tuned bam-readcount (https://github.com/genome/bam-readcount). The generated bam-readcount results were parsed and analyzed by in-house scripts. Variants at base 'G' were initially collected in before-pulldown or pulldown sample if (1) the sum of G→T or G→C misincorporation count is equal to or larger than 2; (2) the sum of G→T or G→C misincorporation rate is no less than two percent; (3) overall coverage at base 'G' was at least 18 reads. Identified guanosine positions

(according to the encoding DNA strand) were then filtered to exclude known genomic polymorphism sites (dbSNP). The generated preliminary mutations sites must appear in two replicates of "before-pulldown" or "pulldown" samples (mutation sites with a G to T misincorporation rate ≥ 40% were removed). 3% misincorporation rate of G→T or G→C (above the background mismatch rate in RNA input samples) was set for "before-pulldown" samples and 8% was set for "pulldown" samples. Variants at base 'G' in "input" samples serve for removing natural mutations or false positives from both "before-pulldown" and "pulldown" samples. The further analysis to uncover high-confident internal m^7G site candidates is based on the comprehensive mutation patterns across input, before-pulldown and pulldown samples.

Gene ontology (GO) enrichment

For GO analysis, the list of RefSeq IDs of target genes was first uploaded into DAVID Bioinformatics Resources (http://david.abcc.ncifcrf.gov) and analyzed with functional annotation clustering. The resulting file was downloaded and extracted with GO terms and corresponding P values. The new list containing GO terms with $P < 0.01$ was imported into REVIGO and visualized with the interactive graph, which was used as the final output figures.

CHAPTER 3

Functional Investigation on Internal N^7-Methylguanosine in Mammalian mRNA

3.1 Introduction: tRNA m^7G methyltransferases

The similarity of some of the m^7G motifs observed within mRNA to the known AG(m^7G)H motifs[137,146] identified in tRNAs suggested that the METTL1-WDR4 heterodimer complex, the known tRNA m^7G methyltransferase[131], might introduce some of these internal m^7G modifications in mammalian mRNAs. N^7-methylguanosine (m^7G) at position 46 is one of the most prevalent tRNA modifications in eukaryotes, prokaryotes, and some archaea. As one of the few known modifications that introduces a positive charge at the variable loop of tRNA, the yeast methyltransferase system consists of heterodimeric complex with Trm8p and its essential co-factor Trm82p, installing m^7G modification of tRNA in the presence of S-adenosylmethionine (SAM)[130,131]. Under heat stress with high temperature, the Trm8 or Trm82 yeast show growth defects due to the rapid decay of tRNA ValAAC[147].

In human, METTL1 (methyltransferase like 1; ortholog of Trm8) was reported to be regulated by the protein kinase B (PKB) and ribosomal S6 kinase (RSK) under growth factor stimulation[140]. METTL1 is also closely correlated to cancer cell viability and sensitivity in the presence of 5-fluorouracil[148,149]. WDR4 (WD repeat domain 4) serves as the ortholog of yeast Trm82p and a component of the human m7G tRNA methyltransferase METTL1-WDR4 complex. WDR4 mutants displayed a distinct form of microcephalic primordial dwarfism in terms of facial dysmorphism, brain malformation, and severe encephalopathy with seizures[133,150]. Over-

expression of Wdr4 in mouse has an impact on learning and memory in a model of Down syndrome[151]. We come up with a hypothesis that such a tRNA methylation machinery might also contributing to establishment of internal mRNA m^7G methylome.

3.2 A subset of internal m^7G sites are installed by METTL1

To test this hypothesis mentioned above, we knocked down METTL1 in HeLa cells and quantified the corresponding internal m^7G level within cap- and rRNA-depleted mRNA. We observed a ~25% reduction of the m^7G/G ratio in siMETTL1 cells compared to controls (**Figure 3.2.1A**). To map changes in m^7G peaks, we applied m^7G-MeRIP-seq to mRNAs isolated from siControl *vs.* siMETTL1 HeLa cells as well as shControl *vs.* shMETTL1 HepG2 cells. The knockdown efficiency in both mRNA and protein levels was validated with RT-qPCR and western blotting, respectively (**Figure 3.2.1B**). We uncovered 1,563 altered peaks in transient knockdown cells and 3,939 in stable knockdown ones; in each case, around 74-77% of peaks became hypo-methylated upon METTL1 knockdown (fold change (FC) $\geq$ 1) (**Figure 3.2.1C**). Several representative hypo-regulated peaks are plotted using raw data in IGV software (**Figure 3.2.1D**).

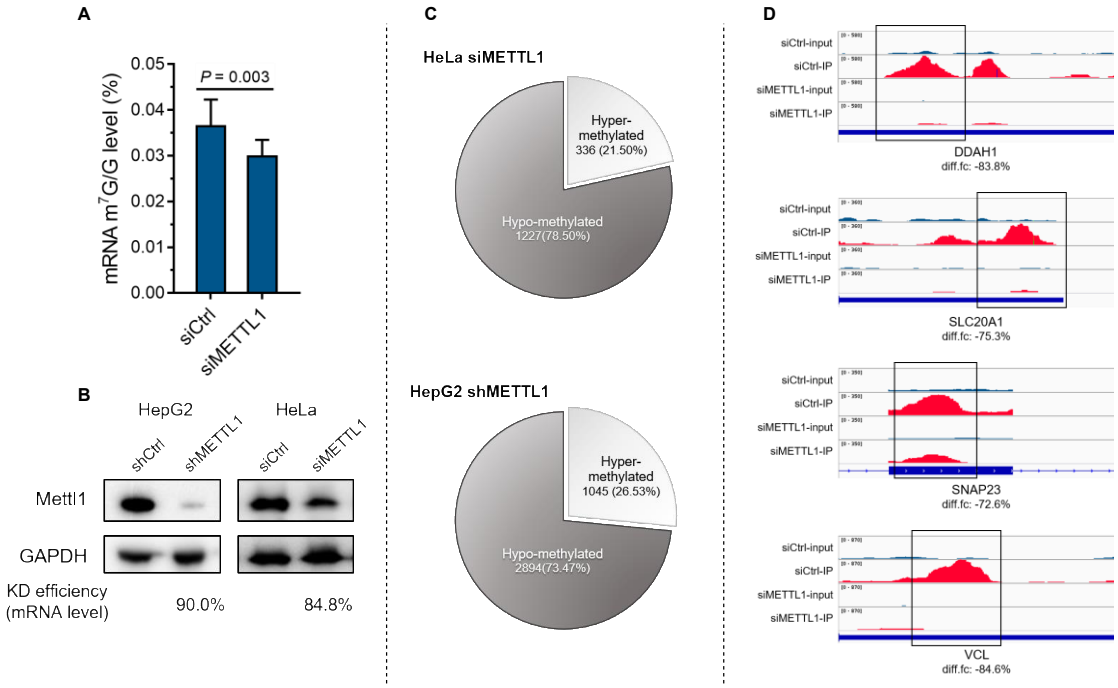

Figure 3.2.1. METTL1 knockdown affects a subset of internal m^{7}G sites in mRNA. (A) The knockdown of METTL1 led to over 20% decrease of the m^{7}G/G level in cap-depleted mRNA from HeLa cells. Mean values ± s.d. are shown, n = 3. (B) Knockdown efficiency tests for siMETTL1 HeLa cells and shMETTL1 HepG2 cells based on RT-qPCR to quantify mRNA level and western blotting to measure protein level. (C) The percentage of hyper- or hypo-methylated peaks with METTL1 knockdown in siMETTL1 HeLa cells and shMETTL1 HepG2 cells (false discovery rate (FDR) ≤ 0.05). (D) Representative internal m^{7}G peaks that were hypo-methylated after METTL1 knockdown in siMETTL1 HeLa cells.

Among these 1,227 potential METTL1 targets in HeLa cells, a majority was enriched in CDS and 3' UTR (**Figure 3.2.2A**), while METTL1 stable knockdown HepG2 cells showed a high enrichment of METTL1 targets in CDS. The metagene profile of these hypo-regulated targets also demonstrated a broad cumulation near stop codon both in CDS and 3' UTR (**Figure 3.2.2B**). m^{7}G peaks acquired from both knockdown cell lines exhibited quite similar GA-enrich motifs (**Figure 3.2.2C**). The METTL1 knockdown resulted in ~54% global reduction of these 1,227 m^{7}G peaks in HeLa cells, and ~61% reduction in HepG2 cells (**Figure 3.2.2D**), with the reduction occurring consistently in all three segments of 5' UTR, CDS, and 3' UTR.

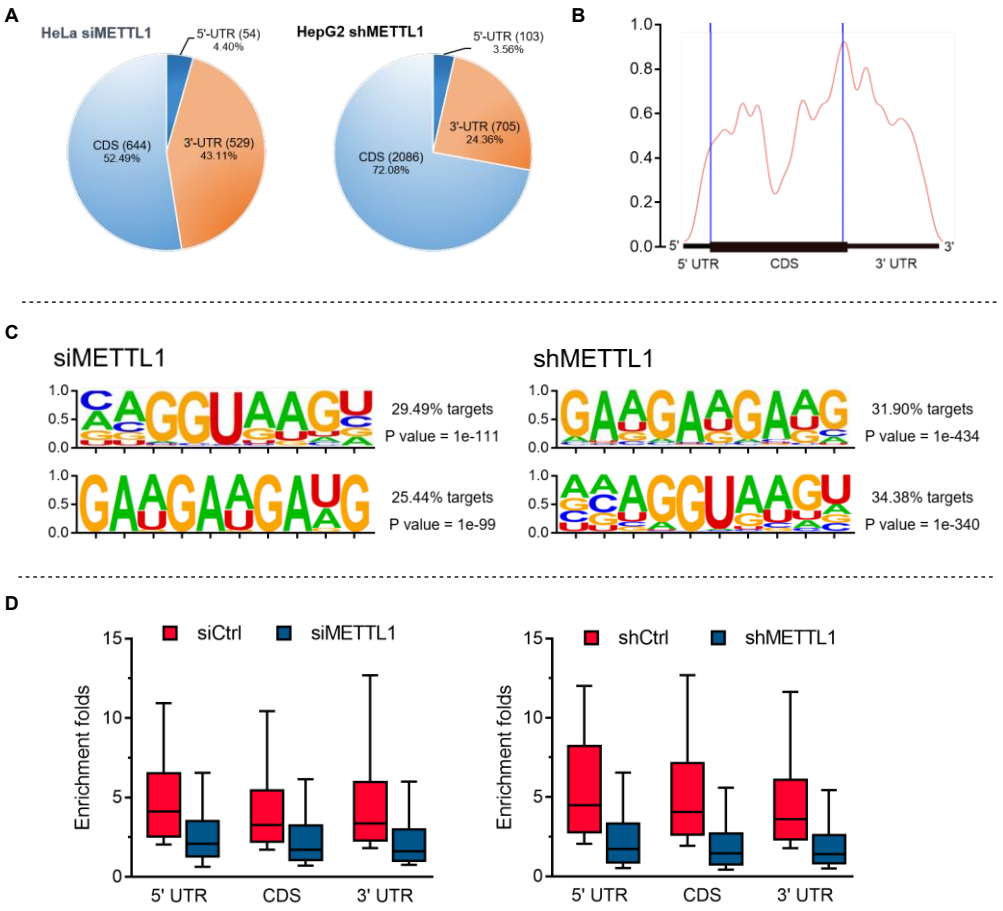

Figure 3.2.2. Statistic features of METTL1 hypo-methylated internal m^{7}G sites. (A) Pie chart showing the fraction of hypo-methylated m^{7}G peaks from m^{7}G-MeRIP-seq of siMETTL1 HeLa (left) and shMETTL1 HepG2 (right) cells in each of three non-overlapping transcript segments (5' UTR, CDS and 3' UTR). (B) Metagene profiles of hypo-methylated (upon METTL1 knockdown) m^{7}G peak distribution along a normalized transcript composed of three rescaled segments. (C) Top two motifs identified from hypo-methylated m^{7}G peaks for siMETTL1 HeLa cells (left) and shMETTL1 HepG2 cells (right). (D) Hypo-methylated peaks of both siMETTL1 cells (left) and shMETTL1 HepG2 cells (right) were classified to three non-overlapping transcript segments (5' UTR, CDS and 3' UTR), which showed a consistent decrease in enrichment fold change upon METTL1 knockdown. Two sided Mann-Whitney test.

The METTL1 knockdown stable cell lines were also subjected to m^{7}G-seq for identifying METTL1 hypomethylated targets by seeking a misincorporation rate decrease at m^{7}G sites. After biotin pulldown, we identified a number of sites with m^{7}G misincorporation rate decrease overlapping with the hypomethylated MeRIP-seq peaks (**Figure 3.2.3**). Most of these sites with

decreased misincorporation rate also showed motifs consistent with our previously characterized GA/GGenriched motifs, which is common METTL1 motifs in tRNAs. In the after-pulldown case, the misincorporation rate change could enable us to detect hundreds of METTL1-affected candidate sites with reduced mismatch rates ranging at 5-20% (**Figure 3.2.3**). However, with the further analysis of samples without enrichment, the more quantitative changes could be observed but with high variations. The differences tend to be small. We think several factors could contribute: i) lower methylation status in many of these sites: ii) potential compensatory effects from other m^7G methylase (see above); iii) inadequate sequencing coverage depth (we need to sequence much deeper in order to have more statistic confidence in assigning mutation differences). We thus still relied on MeRIP to identify hypomethylated peaks and chose those overlap with a)

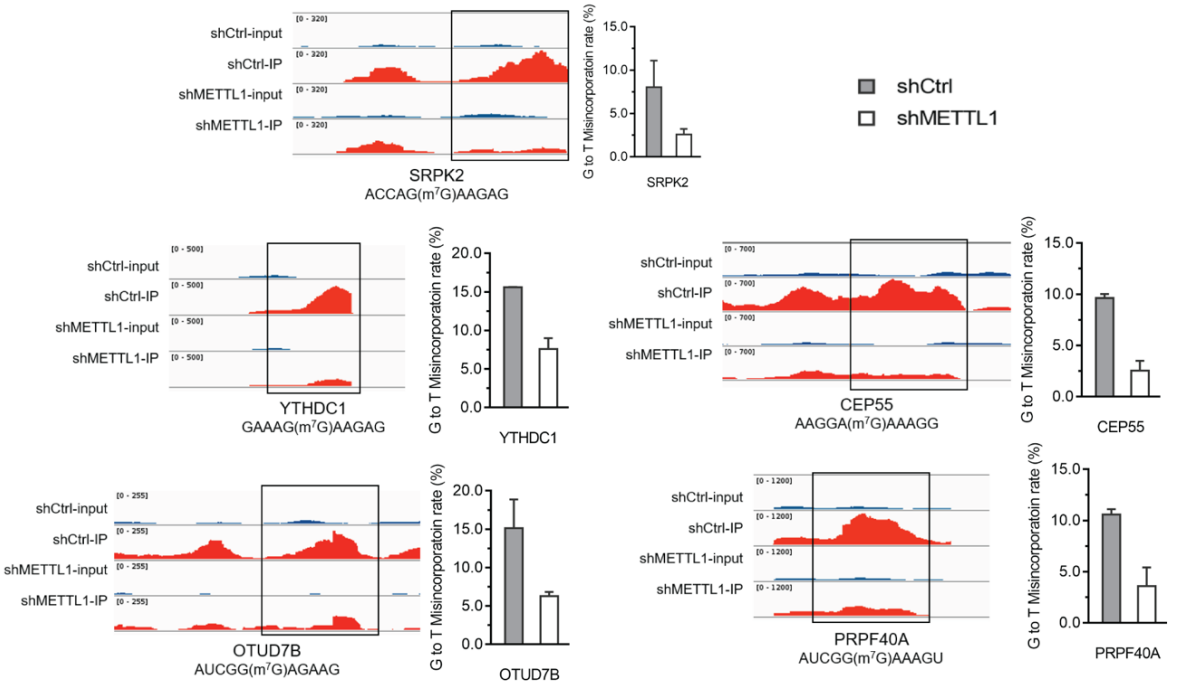

Figure 3.2.3. Representative METTL1 hypo-methylated peaks identified by both MeRIP-seq and m^7G-seq.

base-resolution m^7G sites; b) overlap with METTL1-WDR4 PAR-CLIP sites as confident METTL1 target sites.

Gene Ontology (GO) analysis of the 1,227 hypo-methylated transcripts (upon siMETTL1 knockdown) uncovered a potential correlation of m^7G methylation with RNA splicing as well as cell adhesion (**Figure 3.2.4A**). To further investigate whether the METTL1-WDR4 complex installs m^7G onto mRNA, we performed photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) of both proteins. We found that both proteins bind mRNA, with over 35% and around 4% of METTL1-dependent m^7G-MeRIP-seq peaks overlapping with WDR4 or METTL1 PAR-CLIP peaks, respectively (**Figure 3.2.4B**). The observed hypo-methylation upon METTL1 knockdown and the correlation between METTL1 and WDR4 targets suggest that METTL1, together with WDR4, act as a "writer" complex of m^7G within mRNA, mainly to introduce methylation around the stop codon and CDS. WDR4 might facilitate RNA binding, with a stronger or direct interaction with RNA as suggested by the larger number of overlapped targets between its PAR-CLIP sites with the METTL1-dependent m^7G-MeRIP-seq peaks, whereas METTL1 could serve as the catalytic component for methylation.

We then further conducted biochemistry to treat small RNA fraction (containing tRNA) isolated from shMETTL1 and mRNA isolated from siMETTL1 cells, respectively, using the recombinant protein complex of METTL1/WDR4 prepared with the transfection of both METTL1 and WDR4 plasmids in human HEK293T cell line. Substantially elevated m^7G/G levels were observed after 12 h incubation at 16 °C in the presence of SAM. The m^7G/G level was recovered to 0.8 % when small RNAs (<200 nt; mostly tRNAs) from shMETTL1 HepG2 cell were treated with the complex (**Figure 3.2.4C**), confirming tRNA m^7G methylation activity of METTL1/WDR4. The mRNA m^7G/G ratio went up to ~0.22% compared with the original

0.03% in the input mRNA, indicating *in vitro* mRNA m^7G methylation activity of the same complex. When we treated the same METTL1/WDR4 complex with 12–17 mer RNA probes bearing the consensus motif for m^7G we failed to observe any activity (in vitro methylation assay part Supplementary Methods). We then constructed a 144-mer RNA oligo probe that mimic the sequence of a m^7G site in mRNA UTRs, and subjected it to the same biochemistry assay. A m^7G/G level of 0.40% was obtained, which afforded 0.26 equivalent of m^7G in each probe on average (**Figure 3.2.4D**). These results confirmed the m^7G methylation activity of METTL1/WDR4 complex and suggested that METTL1/WDR4 does not work on short RNA probes and may recognize tRNA-like structures in order to mediate m^7G methylation on mRNA.

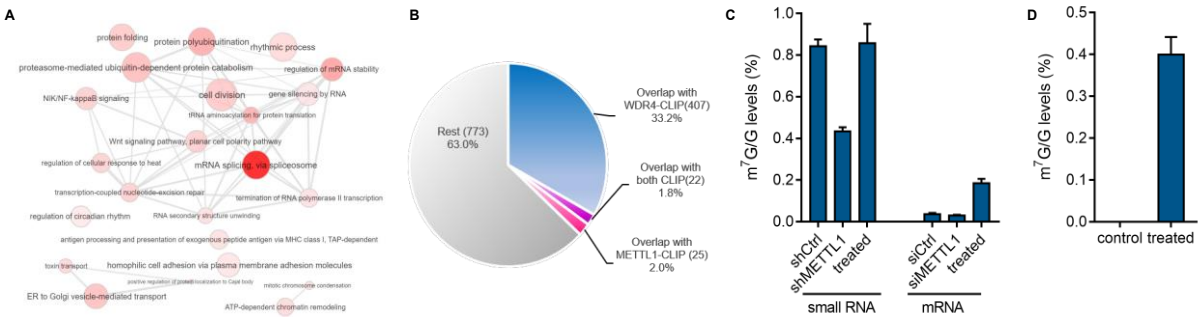

Figure 3.2.4. METTL1-WDR4 complex is responsible for mRNA internal m^7G methylation. (A) GO analysis for potential METTL1 target transcripts that contains hypo-methylated peaks in siMETTL1 HeLa cells. (B) The percentage of hypo-methylated peaks that overlap with WDR4 PAR-CLIP targets, METTL1 PAR-CLIP targets. (C) *In vitro* methylation assay with recombinant the METTL1/WDR4 complex incubated with small RNA (<200 nt) from shMETTL1 HepG2 cells and purified mRNA from siMETTL1 HeLa cells. (D) *In vitro* methylation assay showing m7G methylation of a synthetic 144-mer RNA probe by the recombinant METTL1/WDR4 complex.

3.3 Internal m^7G methylome promotes mRNA translation

To investigate potential functions of METTL1-mediated internal m^7G mRNA methylation, we first assessed mRNA-ribosome association by performing ribosome profiling in HeLa cells transfected by either siMETTL1 or control siRNAs. We sequenced both the input and the ribosome-bound mRNA fragments, and analyzed translation efficiency (ribosome-bound mRNA normalized by the whole cell mRNA level) of five groups of transcripts: 1) 1,437 non-target mRNAs containing m^7G-MeRIP-seq peaks that were unaffected by METTL1 knockdown; 2) 972 METTL1 potential target (PT) mRNAs including m^7G-MeRIP-seq peaks that were hypomethylated upon METTL1 knockdown; 3) 397 PT mRNAs that overlap with either WDR4 or METTL1 PAR-CLIP peaks; and 4) 101 high confidence PT mRNAs that overlap with m^7G base-resolution sites and METTL1/WDR4 PAR-CLIP peaks. We found that METTL1 depletion decreased translation efficiency of all four PT groups, suggesting that loss of METTL1 and thus reduction of internal m^7G may suppress translation of some of METTL1 target transcripts relative to non-targets **(Figure 3.3.1A** and **Figure 3.3.1B**). To exclude that METTL1 knockdown influences the translation of target mRNAs *via* altered tRNA methylation, we examined the enrichment folds of specific tRNA codons related to m^7G-methylated tRNAs within the METTL1 targets (group 2) *vs.* randomly selected non-targets (**Figure 3.3.1C**). We observed no enrichment of these tRNA codons in METTL1 targets *vs.* non-targets. These data suggest that METTL1 knockdown could specifically reduce the translation of METTL1 target mRNAs independently of effects on global translation and tRNA methylation.

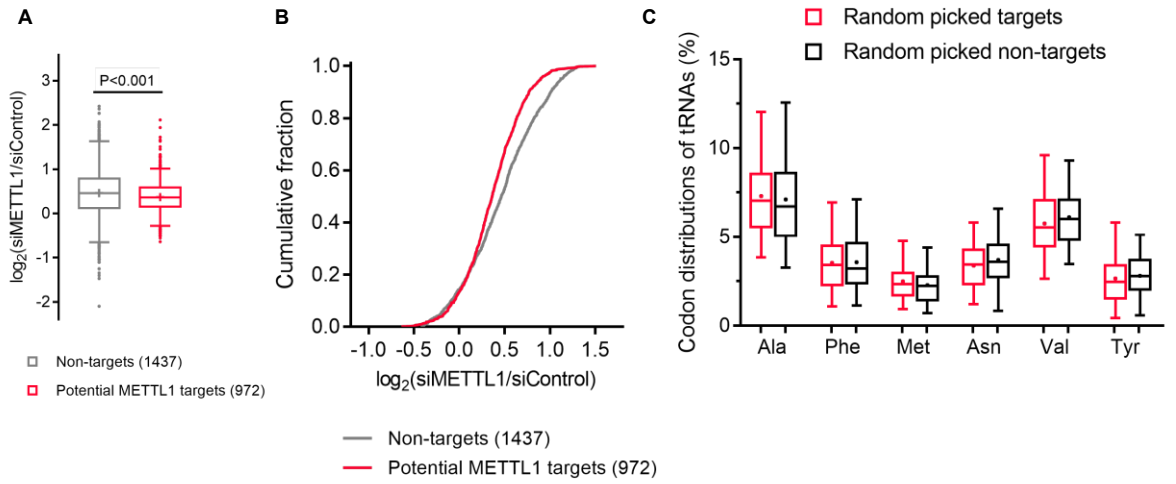

Figure 3.3.1. METTL1 knockdown decreases translation efficiency. (A) Knockdown of METTL1 reduced the translation efficiency of target transcripts. Cumulative distribution log_2-fold changes of the translation efficiency between siMETTL1 and siControl transfection for non-targets (grey), potential targets (METTL1 hypo-methylated targets) (red). Two sided Mann-Whitney test. (B) Cumulative distribution log_2-fold changes of translation efficiency (ratio of ribosome bound fragments to mRNA input) between siMETTL1 and siControl transfection for non-targets (grey), potential targets (red). Two-sided Mann-Whitney test. (C) Codon distribution of tRNAs containing m^7G_{46} in randomly picked non-targets and METTL1-dependent hypo-methylated targets.

We next performed polysome profiling to examine the distribution of translating METTL1 target mRNAs in METTL1-knockdown cells *versus* control cells. Three groups of mRNA-protein particles (mRNPs) were separated: non-ribosome (translation-inactive), 40S-80S (mRNPs associated with ribosome but not being translated), and polysome (active translation pool). Western blotting of each fraction showed that the cytoplasmic METTL1 and WDR4 were distributed mainly in the non-ribosome and 40S mRNPs portions. We quantified changes in the distribution of specific mRNAs among non-ribosome fractions and polysome fractions using RT-qPCR. Upon METTL1 knockdown, the mRNA levels within polysome fractions decreased for the METTL1 targets (DDAH1, SLC20A1, SNAP23, and VCL), but not for a non-target control

(ACTB) (**Figure 3.3.2**). These profiling results provided further support that the METTL1-dependent internal m^{7}G methylation could promote translation of the corresponding transcripts.

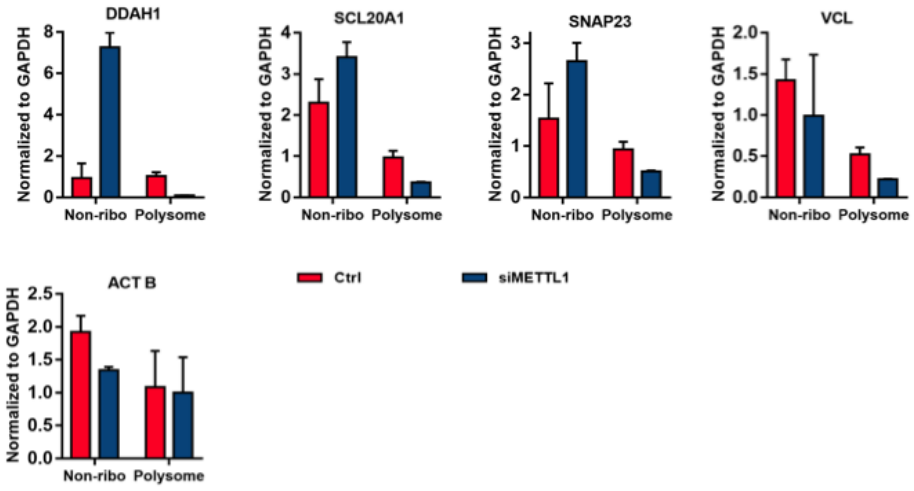

Figure 3.3.2. Redistribution of representative targets in non-ribosome and polysome portions of mRNPs upon depletion of METTL1. DDAH1, SCL20A1, SNAP23, and VCL are potential METTL1 m7G targets, and ACTB is a non-target control. Mean values ± s.d. are shown, n = 2, technical replicates, measured by RT-qPCR.

To further investigate a role for m^{7}G in translation, we performed *in vitro* translation assays in HeLa cell lysates with capped and poly-A tailed luciferase mRNA reporters with both internally nonmethylated reporters containing exclusively G and methylated reporters containing one internal m^{7}G sites located in 3'UTR per transcript (generated with addition of m^{7}GTP during *in vitro* transcription and RNA ligation with the level of m^{7}G quantified using LC-MS/MS, Figure 2.8.3C). We incubated the mRNA reporters in HeLa cell lysates for 30 min at 30 °C, measured luminescence intensity, and normalized intensities to the transcript levels. We found that the methylated reporters displayed elevated (~30-50%) luciferase activities compared to the non-methylated reporter (**Figure 3.3.3A**), suggesting that internal m^{7}G methylation could promote mRNA translation. We added the m^{7}GpppG cap analog to disrupt cap-initiated translation[152,153].

Both reporters displayed substantially reduced and similar translation efficiencies in the presence of the cap analog (**Figure 3.3.3B**), indicating that internal m^7G methylation may promote translation *via* similar cap-dependent pathways.

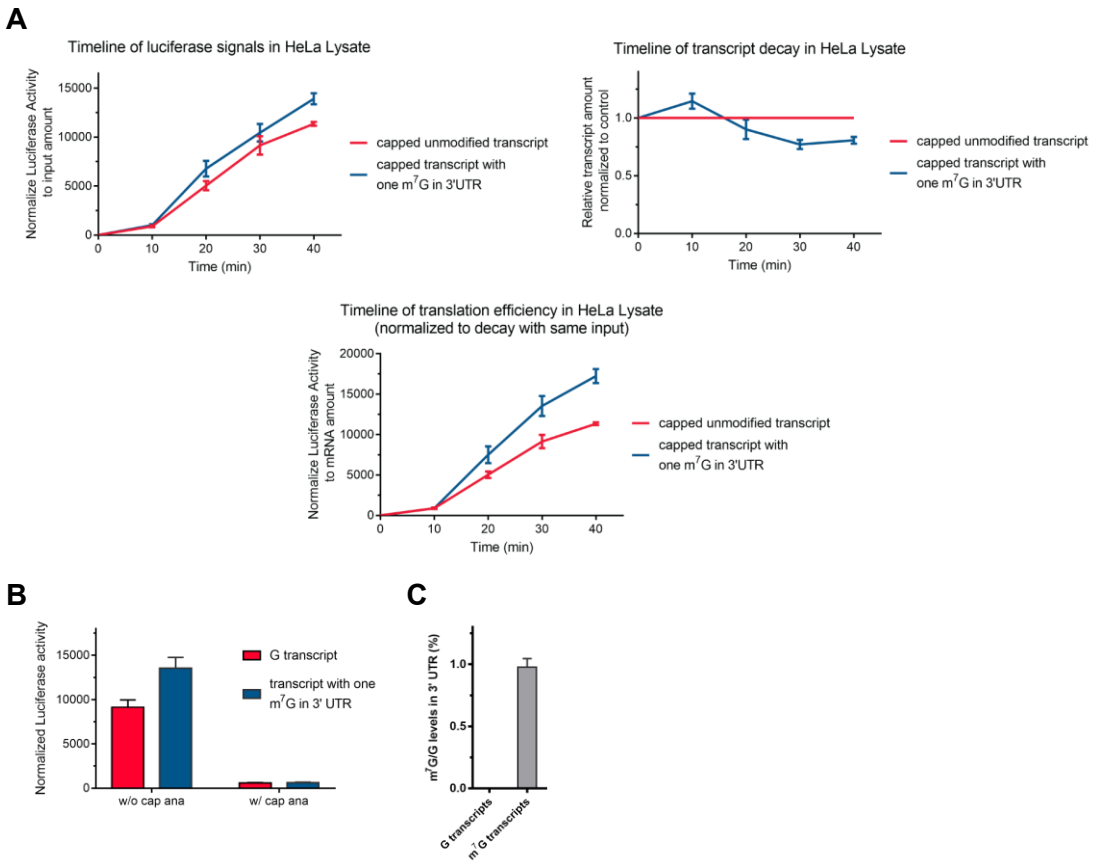

Figure 3.3.3. ***In vitro*** **translation assays with nonmethylated reporters containing exclusively G and methylated reporters containing one internal m^7G site located in 3'UTR.** (A) The kinetics curve of the *in vitro* translation assay with methylated (blue) *vs.* non-methylated (red) reporters over time. Mean values ± s.d. are shown, n = 2. (B) *In vitro* translation of luciferase-encoding, two capped mRNA reporters in HeLa lysate. Reporters are capped and contain either unmodified G (red) or m^7G (blue) within the 3' UTR, incubated in the absence (left) or presence (right) of 1.0 mM free cap analog (m^7GpppG), normalized with transcript levels. Mean values ± s.d. are shown, n = 2. (C) LC-MS/MS quantification of m^7G/G level in probes used for *in vitro* translation. Mean values ± s.d. are shown, n = 2.

3.4 Conclusions and Discussion

We found that the known tRNA m^7G modification complex, METTL1-WDR4[130,131,137], while participate in tRNA methylation as also verified in our knockdown results, also installs a subset of internal m^7G sites in human mRNA. The simultaneous methylation of the position 46 at the extra loop in human tRNAs and human mRNA by the METTL1-WDR4 complex may not be surprising. Other mRNA modification machineries such as eukaryotic ψ synthases PUS1, PUS7 and TRUB1 have been shown to work on both tRNA and mRNA[97-102]. The mRNA m^1A methyltransferase TRMT6/61A and TRMT61B are also known tRNA modification enzymes. NSun family m^5C methyltransferases bind quite extensively to mRNA as well[103-106]. Here, we show that the heterodimer METTL1-WDR4 is responsible for a portion of internal mRNA m^7G methylation. Besides our observation that the knockdown of METTL1 led to reduced m^7G methylation within mRNA, the METTL1-WDR4 PAR-CLIP results also revealed the binding of this complex to mRNA and a substantial overlap of the binding sites with m^7G-MeRIP-seq sites in the same cell line. WDR4 appears to possess more binding sites in mRNA and may contribute to stabilizing the heterodimer complex and RNA binding, whereas METTL1 may mediate m^7G methylation. *In vitro* methylation assay using tRNA and mRNA purified from METTL1-depleted cells revealed good m^7G methylation activity of METTL1-WDR4 towards both tRNA and mRNA. This complex also mediates m^7G methylation of a long synthetic RNA probe that mimics internal mRNA m^7G sites; however, the complex failed to react with several RNA probes ranging from 12-mer to 17-mer, suggesting that this methylase complex may require tRNA-like structures with a stem loop of a minimum length to facilitate substrate recognition. Please note that our m^7G-seq identified diverse sequence motifs around internal m^7G sites in mRNA, suggesting involvements of additional methyltransferase(s) for m^7G installation.

Functional roles of m^7G within mRNA could be diverse. We found that the knockdown of METTL1 decreases the translation efficiency of transcripts that contain METTL1-affected hypomethylated m^7G sites. This trend is consistent with our observation that internal m^7G methylation promotes translation based on *in vitro* translation assays using cell lysates and capped luciferase mRNA reporters containing a 3' UTR bearing m^7G *versus* a unmethylated control. We speculate that internal m^7G sites located at 3' UTR may promote translation *via* a looping model suggested previously[154]. Because of the unique nature of positive charge, internal m^7G could reorganize local RNA structure, modulate protein-RNA interactions, and be better recognized through aromatic stacking. It could be recognized by different reader proteins and modulate local RNA structure to affect protein-RNA interactions. Although we propose a potential translation promotion function, we suspect the presence of m^7G could affect nuclear pre-mRNA processing, export, translation and stability. Perhaps different methyltransferases install m^7G to distinct mRNA sequences and/or secondary structural motifs to further modulate the fate of the target mRNA. The identification of more frequently m^7G methylated mRNAs and the availability of the new m^7G-seq method should allow future functional studies and mechanistic investigations.

In summary of chapter 2 and chapter 3 as a whole story, we discovered and mapped internal m^7G sites within mammalian mRNA. Our single-nucleotide-resolution m^7G sequencing technology allows precise mapping of m^7G transcriptome-wide. This method could be widely used to detect m^7G not only in tRNA and rRNA, but also within mRNA and various non-coding RNAs. We provide the first base-resolution m^7G maps in human cell lines. Importantly, we show that certain internal m^7G sites accumulate relatively high m^7G fractions (above 50%), suggesting functional roles. We identified METTL1 as one writer that can install a subset of internal m^7G within mRNA and show that m^7G could facilitate translation promotion of METTL1-affected

transcripts. We propose that m^7G could exhibit diverse functions on target mRNA, perhaps through modulating structure and protein-RNA interactions. This work provides critical resources for future investigations of m^7G that may impact diverse biological processes and human diseases.

3.5 Material and Methods

Ribosome profiling.

Ribosome profiling was performed according to the procedure reported previously (Wang et al., 2014). Four 15-cm plates of HeLa cells were prepared for 48 h knockdown (siControl *vs* siMETTL1, two plates each). Cycloheximide (CHX) was added to the media at 100 mg ml^{-1} and incubated at 37 °C for 7 min. After removing the media and washed with cold PBS, the cells were collected by cell lifter with 5 ml ice-cold PBS with CHX (100 mg ml^{-1}). The cell pellet was collected by centrifugation at 500 g for 5 min. 1 ml lysis buffer (10 mM Tris, pH 7.4, 150 mM KCl, 5 mM $MgCl_2$, 100 mg ml^{-1} CHX, 0.5% Triton-X-100, freshly add 1:100 protease inhibitor and 40 U ml^{-1} RNasin® Ribonuclease Inhibitors (Promega)) was added to suspend the cell pellets and kept on ice for 15 min. After centrifugation at 13,000 g for 15 min, the supernatant was collected, and its absorbance was tested at 260 nm. 8 µL TURBO DNase (Thermo Fisher Scientific) was added to the lysate and the lysate was then split by the ratio of 1:4 as Portion I and Portion II. 4 µL SUPERase-In RNase inhibitor was then added to Portion I. 40 µL MNase buffer and 3 µL MNase (6,000 gel units, NEB) was added to Portion II. Both portions were kept at room temperature for 15 min, and 8 µL SUPERase-In RNase inhibitor was then added to Portion II to stop the reaction. Portion I was saved and mixed with 1 ml TRIzol to purify input mRNA. A 10/50% w/v sucrose gradient was prepared in a lysis buffer without Triton-X-100. Portion II was loaded onto the sucrose gradient and centrifuged at 4 °C for 3 h at 28,000 rpm. (Beckman, rotor SW28). The sample was then fractioned and analyzed by Gradient Station (BioCamp) equipped with ECONO UV monitor (BioRad) and fraction collector (FC203B, Gilson). The fractions corresponding to 80S were collected, combined, and mixed with an equal volume of TRIzol to puri-

fy the RNA. The RNA pellet was dissolved in 20 µL water, mixed with 20 µL 2X TBE-urea loading buffer (Invitrogen), and separated on a 10% TBE-urea gel. A 21-nt and a 42-nt ssRNA oligo were used as size markers, and the gel band between 21 and 42 nt was cut. 600 µL extraction buffer (300 mM NaOAc, pH 5.5, 1 mM EDTA, 0.1 U ml^{-1} SUPERase-In RNase inhibitor) was added to the gel. The gel slurry was heated at 65 °C for 10 min with shaking, and then filtered through gel filter. RNA was concentrated by ethanol precipitation and finally dissolved in 10 µL of RNase-free water. Input mRNA: the input RNA was first purified by TRIzol and the input mRNA was prepared as previously described. The mRNA was then fragmented by RNA fragmentation kit (Ambion). All samples were subjected to small RNA library construction following manufacturer's instructions of NEBNext Small RNA Library Prep Set for Illumina (NEB). All libraries were sequenced on Illumina NextSeq 500 with single-end 80 bp read length.

Polysome profiling.

Polysome profiling was performed according to the procedure reported previously (Wang et al., 2014; Gandin et al., 2014). Eight 15-cm plates of HeLa cells were prepared for 48h knockdown (transfected with siControl or siMETTL1, two replicates for each siRNA and two plates for each replicate). Before collection, cycloheximide (CHX) was added to the media at 100 µg ml^{-1} for 7 min. The medium was removed, and the cells were washed with ice-cold PBS once. The cells were then collected with ice-cold PBS with CHX at 100 µg ml^{-1}, followed by centrifugation at 500 g for 5 min. 500 µL lysis buffer (20 mM HEPES, pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, 100 µg ml^{-1} CHX, 1% Triton-X-100, and 1:100 protease inhibitor (Roche); 40 U ml^{-1} RNasin® Ribonuclease Inhibitors (Promega) were added prior to use) was added to resuspend cell pellets for each sample and was kept on ice for 20 min with periodic perturbation. After the incubation, the

lysate was centrifuged at 15,000 g for 15 min. The supernatant was collected. An addition of 4 µL TURBO DNase (Thermo Fisher Scientific) was applied to each sample followed by another incubation at room temperature for 15 min. The treated lysate was centrifuged, and the clear supernatant was saved and tested with A_{260} absorbance to adjust and dilute the samples to the same OD value. Part of the clear lysate was kept as input for western blotting.

A 10/50% w/v sucrose gradient was prepared in lysis buffer without Triton-X-100. The rest lysate was loaded onto the sucrose gradient and centrifuged at 4 °C at 28,000 rpm (Beckman, rotor SW28) for 3 hr. The samples were then fractioned and the A_{260} absorbance of each fraction was analyzed by Nanodrop. Part of each sucrose fraction was used for western blotting. Total RNAs of different fractions (non-ribosome, 40S-80S and polysome) were extracted by TRIzol (Invitrogen) and purified with ethanol precipitation as described before. Total RNAs were then subjected to reverse transcription with PrimeScript RT Reagent Kit (Takara) to check the mRNA levels of each gene of interest with RT-qPCR (also see RT-qPCR at the Experimental Procedure section).

PAR-CLIP.

We followed the previously reported protocol (Wang et al., 2014; Hafner et al., 2010). Ten 15-cm plates of HEK293T cells were seeded for each replicate and grown to 80% confluency before the addition of 4 µL 1 M 4SU to each plate. After 14 hours incubation, the media was aspirated; the cells were washed once with 5 ml ice-cold PBS for each plate and crosslinked by 0.15 J cm^{-2} 365 nm UV light twice when on ice. The crosslinked cells were collected with cell lifters. 3 volumes of the lysis buffer (50 mM HEPES, pH 7.5; 150 mM KCl; 2 mM EDTA; 0.5% (v/v) NP-40, with 1:100 protease inhibitor (Roche) and 40 U ml^{-1} RNasin® Ribonuclease Inhibitors (Promega) added freshly) was added to the cell pellet and incubated on ice for 10 min with periodic pertur-

bation. The cell lysate was then centrifuged at 15,000 g for 15 min and the clear supernatant was collected. RNase T1 (1,000 U μL^{-1}, Thermo Fisher Scientific) was added to the clear lysate to a final concentration of 0.1 U μL^{-1} and an incubation at room temperature was performed for 15 min. The reaction was then quenched on ice. After 5 min, 4.5 µg mouse anti-METTL1 or WDR4 antibody were added to each sample. The antibody and the lysate were incubated at 4 °C for 2 hours under periodic rotation. Protein G beads (Thermo Fisher Scientific) were washed (45 µL for each sample) with IP wash buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 0.05% (v/v) NP-40, with 1:100 protease inhibitor (Roche) and 40 U ml^{-1} RNasin® Ribonuclease Inhibitors (Promega) added freshly) for 2 times. The beads were resuspended in 50 µL lysis buffer for each sample and added to the antibody-lysate mixture subsequently. Another one-hour incubation at 4 °C was performed with low-speed rotation.

After the incubation, the beads were washed three times with IP wash buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 0.05% (v/v) NP-40, with 1:100 protease inhibitor (Roche) added freshly) and then resuspended with 200 µL IP wash buffer per sample. The beads were treated with a second round of RNase T1 digestion under a final concentration of 10 U μL^{-1} for 15 min at room temperature. The reaction was then quenched with the addition of 10 µL SUPERase-In followed by a 5-min incubation on ice. The beads were washed three times with high-salt wash buffer (50 mM HEPES, pH 7.5, 500 mM KCl, 0.05% (v/v) NP-40, with 1:100 protease inhibitor (Roche) added freshly) and twice with 1X PNK buffer (NEB) afterwards.

The beads were resuspended with 200 µL of 1X PNK buffer (NEB) and underwent T4 PNK (Thermo Fisher Scientific) end repair with standard procedures as previously mentioned under 37 °C. After the incubation, the beads were washed once with 1X PNK buffer followed by proteinase K digestion as described before. The RNA was recovered with RNA Clean & Con-

centrator (Zymo Research) before library construction by NEBNext Small RNA Library Prep Set for Illumina (NEB). All libraries were sequenced on Illumina NextSeq 500 with single-end 80 bp read length.

RT-qPCR. All RNA templates used for RT–qPCR were digested with DNase I in the purification step. RT-qPCR primers were designed to cover exon-exon junctions shared in all isoforms of the matured mRNAs. About 500 ng RNAs were first reverse-transcribed into cDNAs with PrimeScript™ RT reagent Kit (Takara), and then the cDNAs were subjected to qPCR analysis with FastStart SYBR Green Master Mix (Roche) in machine LightCycler 96 (Roche). *Actb* or *gapdh* were used as internal controls in different cases. The sequences of primers used in this study are listed in Table S4.

***In vitro* methylation assay.** *In vitro* methylation assay was carried out with mRNA purified from siMETTL1 HeLa cells, and small RNA (<200 nt) purified from shMETTL1 HepG2 cells, and the synthetic probe mimicking the UTR region of certain genes. mRNA was purified as described previously. Small RNA fraction was purified from total RNA with *mir*Vana miRNA Isolation Kit (Thermo Fisher Scientific) following manufacturer's protocol. The synthetic probe mimicking the structure of 5'UTR of MPZL1 was prepared by *in vitro* transcription as previously described. The DNA template was directly ordered from Integrated DNA Tech Inc as two ssDNA with the sequences as shown in Table S4.

The ssDNAs were dissolved in Nuclease-free Duplex Buffer and mixed in equal molar amounts. The annealing was performed with denaturation at first at 94 °C for 2 min, followed by gradual cooling at the speed of -1 °C per min. The annealing products were purified with agarose gel and

recovered with MinElute Gel Extraction kit (Qiagen). The RNA was then prepared with MEGAscript T7 Transcription Kit (Thermo Fisher Scientific) according to manufacturer's protocol including TURBO DNase digestion step at 37 °C for 1 hour. The RNA yielded was later purified with RNA Clean and Concentrator (Zymo Research).

Several short RNA probes were also synthesized for *in vitro* methylation assay with the sequences shown in Table S4.

The recombinant FLAG-tag protein complex was purified from HEK293T cells with FLAG-METTL1/ FLAG-WDR4 double overexpression. The plasmids for Flag-tag METTL1 and Flag-tag WDR4 were acquired directly from Vector Builder with the ID as VB180605-1094fas and VB180605-1181pjz respectively. The overexpression was performed with Lipofectamine 2000 Reagent (Thermo Fisher Scientific) following manufacturer's protocol. Cells were harvested after 24 hr overexpression and the proteins were purified using Anti-FLAG M2 magnetic beads (Sigma Aldrich) following the manufacturer's protocol. The *in vitro* methylation assay was performed according to previous method (Liu et al., 2014) for methyltransferase validation but instead of d_3-SAM, SAM (NEB) was used and m^7G/G level (LC-MS/MS) was detected to show activity of the methyltransferase.

CHAPTER 4

ALKBH7-Mediated Mitochondrial tRNA Demethylation

4.1 Introduction: RNA demethylases

The AlkB family proteins use iron, α-ketoglurate, and dioxygen to catalyze oxidative demethylation of a range of nucleic acid substrates.[77,78] The *E. coli* AlkB, a prototype of the AlkB family proteins, mediates oxidative repair of 1mA and 3mC DNA base methylation lesions; two human AlkB homologues, ALKBH2 and ALKBH3, perform similar DNA repair functions in human cells.[79,80,] In 2010, He lab made a well-known paradigm shifting discovery that FTO, one of the human AlkB family proteins, is responsible for oxidative demethylation of methylated RNA bases.[58,82,158-160] With another later identified mRNA demethylase ALKBH5,[81,161,162] He group found FTO and ALKBH5, both homologues of *E. coli* AlkB and human ALKBH2 and ALKBH3, mediates quite effective oxidative demethylation of the N^6-methyladenosine (m^6A) in eukaryotic messenger RNA and other nuclear non-coding RNAs. Cellular RNAs contain more than a hundred structurally distinct post-transcriptional modifications at thousands of sites, serving versatile coding, structural, and catalytic functions,[164,165] in which N^6-methyladenosine (m^6A) is the prevalent and predominant modification on mRNA/lncRNA. Not only for being the most abundant mRNA modification, occurring at 3~4 residues per mammalian mRNA, but also because it is reversible and dynamically regulated with well-established 'writer', 'eraser', and 'reader' protein systems.[83,166] The discovery of reversible RNA methylation leads to recent rapid development of RNA epitranscriptomics field that investigates in-depth gene expression regulation *via* dynamic RNA modifications. Recent studies also revealed new RNA demethylation activities associated with various physiological processes, tumor progression and human can-

cers.[159,160,162] Except for m^6A demethylases existing in mammalian cells, ALKBH10B was demonstrated to be an N^6-methyladenosine (m^6A) demethylase affecting arabidopsis floral transition. [163]

Based on the findings on AlkB Domain of Mammalian ABH8, which can catalyzes hydroxylation of 5-methoxycarbonylmethyluridine at the Wobble Position of tRNA,[167] in 2016 He group discovered the first tRNA demethylase, ALKBH1, which removes the ubiquitous m^1A methylation in a set of tRNAs, with a role of regulating translation.[168] Focusing on tRNA demethylation, in this chapter preliminary unpublished results will be introduced on ALKBH7, a protein that significantly affects mammalian energy metabolism,[169] is a mitochondrial protein and mediates mitochondrial tRNA N^2,N^2-dimethylguanine ($m^2{}_2G$) demethylation (**Figure 4.1**).

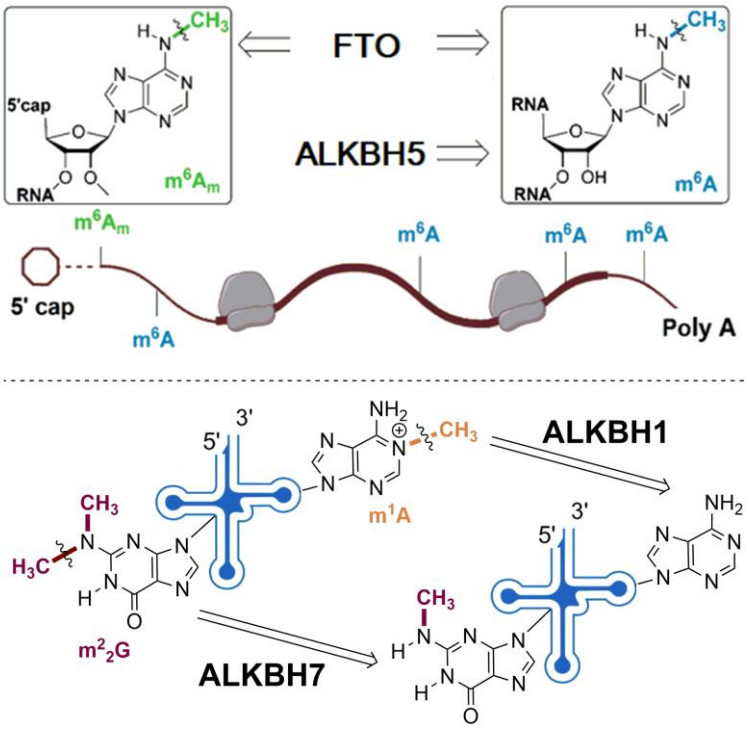

Figure 4.1. mRNA m^6A and m^6A_m demethylation by FTO and ALKBH5, as well as tRNA m^1A and $m^2{}_2G$ demethylation by ALKBH1 and ALKBH7, respectively.

4.2 ALKBH7 demethylates N^2,N^2-dimethylguanine (m^2_2G) at position 26 of some mitochondrial tRNAs

We previously found that the conserved m^1A at nucleotide 58 in the TΨC loop in most eukaryotic tRNAs can be removed by ALKBH1 (**Figure 4.1**).[168] This modification is essential for yeast and mammalian viability, in which one major reason can be attributed to its stabilizing the initiator-$tRNA^{met}$.[170] This discovery unveiled a potential new mechanism of translation regulation through reversible tRNA methylation. Here, we present another discovery of a mitochondrial tRNA demethylase, ALKBH7, which converts the N^2,N^2-dimethylguanine (m^2_2G) modification into N^2-methylguanine (m^2G) in a few mitochondrial tRNAs (**Figure 4.1**).

ALKBH7 belongs to the AlkB family of non-heme Fe(II)-αKG dependent dioxygenase and localizes predominantly in the mitochondria of mammalian cells. ALKBH7 is known to exhibit dramatic metabolic phenotypes in mammals: The *Alkbh7* knockout mice exhibits the overt obesity phenotype, in which the *Alkbh7* deletion dramatically increases body weight and body fat by up to 40%.[169] The sequence and secondary structure alignments show highly conserved secondary structure as well as iron- and αKG -biding motifs of ALKBH7 with other families of AlkB family proteins.[171] Although the function of ALKBH7 has yet to be revealed, we noticed that the active site of ALKBH7 resembles an engineered AlkB protein (D135S/L118V) that works on enzymatic demethylation of m^2_2G (**Figure 4.2.1A**).[172] The m^2_2G modification occurs to cytoplasmic and mitochondrial tRNAs at the G26 position. It blocks Watson-Crick base pairing and therefore also hinders reverse transcription by most RT enzymes. It was thought to be important for the stability of certain tRNAs and has been shown to play significant roles in cellular response to endogenous ROS and exogenous oxidizing agents.[173] Considering the predominant

mitochondrial localization of ALKBH7, we hypothesize that ALKBH7 mediates mitochondrial tRNA m^2_2G demethtylation.

To examine the hypothesis that ALKBH7 may act as a new class of tRNA demethylase, we isolated different fractions of mitochondrial RNA from HepG2 cells and treated with recombinantly expressed human ALKBH7 in the presence of iron, αKG, ascorbic acid, and dioxygen. The RNAs were incubated with 2 nmol of wt-ALKBH7 with or without EDTA (EDTA chelates iron and inhibits the demethylation activity) at pH 7.0 and 37 °C for two hours, followed by complete digestion to single nucleosides and quantification of all known methylated bases using LC-MS/MS. We did not observe notable changes for all modified bases in the mt-RNA fraction with sizes above 200 nucleotides. However, in the fraction of small RNA (size < 200 nt), the result showed that recombinant ALKBH7 could efficiently demethylate m^2_2G under these conditions (**Figure 4.2.1B**), but other modifications such as m^1A and m^1G did not show visible changes.

We then purified mt-tRNA by gel purification, followed by ALKBH7 treatment again. We noticed that the m^2_2G/G ratio is around 1.5% in mitochondrial tRNAs and observed clear demethylation of mt-tRNA. In human mt-tRNA, m^2_2G appears at the 26 position and this modification is highly conserved in several eukaryotic mt-tRNA. m^2_2G26 or m^2G26 both exist in several mt-tRNAs including mt-tRNA-Ile, mt-tRNA-Glu, mt-tRNA-Gln, mt-tRNA-Arg, and mt-tRNA-Ala.[174,175] It was thought that TRMT1 mediates both m^2G26 and m^2_2G26 methylation (**Figure 4.2.2**).[176,177] Our preliminary data suggested that ALKBH7 is more effective in converting m^2_2G to m^2G in tRNA but has less activity in demethylating m^2G because we did not observe noticeable reduction of the m^2G level in the *in vitro* assay so far. In ALKBH7 knockdown HepG2 cells, a ~42% increase of m^2_2G/G ratio in mt-tRNA was observed in several replicates

following a 48-hour knockdown (**Figure 4.2.1C**). We further purified mitochondrial tRNA from wild type MEF cells and *Alkbh7*$^{-/-}$ MEF cells. The knockout of *Alkbh7* doubled the level of m^2_2G/G in mt-tRNA (**Figure 4.2.1D**). We then synthesized DNA probes to pull-down several specific mt-tRNAs known to possess the G26 modification (mt-tRNA-Ile, mt-tRNA-Arg, mt-

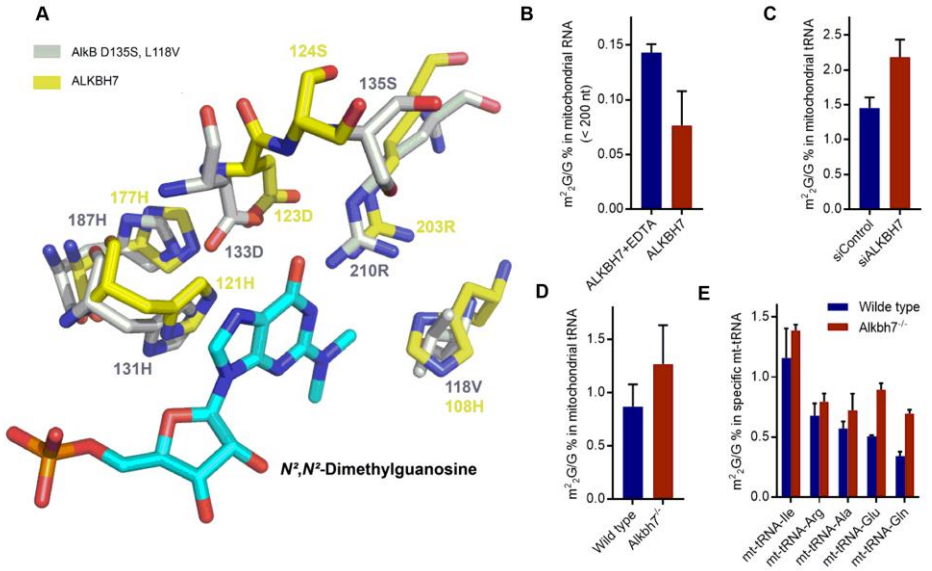

Figure 4.2.1. *In vitro* and *in vivo* evidences for ALKBH7 as a demethylase for some mitochondrial tRNAs. (A) View of active site stereochemistry model in both AlkB (D135S/L118V) mutant (labeled in silver white) and ALKBH7 (labeled in yellow) with m^2_2G coordination based on protein data bank (PDB) ID 3BIE and 4QKD. m^2_2G is labeled in cyan. (B) Demethylation of m^2_2G in mitochondrial RNA (<200 nt) isolated from HepG2 cells by recombinant ALKBH7 in vitro. ALKBH7 can erase around 47% m^2_2G. (C) Quantification of the m^2_2G/G ratio in mitochondrial tRNA purified from HepG2 cells by LC-MS/MS. The transient knockdown of ALKBH7 for 72 hours in HepG2 cells leads to an increase of the m^2_2G/G ratio. (D) Quantification of the m^2_2G/G ratio in mitochondrial tRNA purified from MEF cells by LC-MS/MS. The knockout of Alkbh7 in MEF cells enhances the m^2_2G/G level compared with the wild type. (E) Quantification of the m^2_2G/G ratio in mt-tRNA^{Ile}, mt-tRNA^{Arg}, mt-tRNA^{Ala}, mt-tRNA^{Glu}, and mt-tRNA^{Gln} by LC-MS/MS. Knockout of Alkbh7 in MEF cells leads to a noticeable increase of the m^2_2G/G ratio in these tRNAs compared with the wild type.

tRNA-Ala, mt-tRNA-Glu, mt-tRNA-Gln) from both the wild type and *Alkbh7*$^{-/-}$ MEF cells. Quantifications of the m^2_2G/G ratios by LC-MS/MS revealed that knockout of *Alkbh7* in MEF

cells induced noticeable increases of the m^2_2G/G ratio in these tRNAs (**Figure 4.2.1E**). For mt-tRNA-Asn as another mitochondrial tRNA of G26, the corresponding data has not been included here due to some difficulties in pulldown, which needs to be repeated.

To further confirm the demethylase activity mentioned above, it is critical to examine the demethylation activity of ALKBH7 towards m^2_2G and m^2G using synthetic model RNA probes. RNA probes containing m^2_2G26 or m^2G26 need to be investigated in future study; the demethylation assay will be run under standard conditions for the iron- and αKG-dependent di-oxygenases using recombinant ALKBH7.[172] An inactive mutant ALKBH7 with the iron ligands mutated to Ala can serve as negative controls. To mimic the chemical environment within tRNA, RNA probes are to be designed with m^2_2G or m^2G in both random and core motif sequences (with or without the nearby stems). Results from the demethylation assay can reveal potential sequence and secondary structure preferences of ALKBH7 and prove the existence of this newly characterized reversible RNA methylation system at the position 26 of mt-tRNAs (**Figure 4.2.2**).

Alternatively, it is possible that the entire tRNA is required for effective m^2_2G demethylation. The purified mt-RNAs known to carry m^2_2G of high methylation level (mt-tRNA-Ile) from *Alkbh7*$^{-/-}$ MEF cells can be harvested with biotin-tagged DNA probe pulldown. These isolated individual tRNA can be treated with recombinant ALKBH7 to demonstrate demethylation kinetics monitored by quenching reactions at different time points, and to determine the levels of m^2_2G and m^2G in tRNA using LC-MS/MS. The crystal structure of human ALKBH7 has been reported without bound substrate.[172] To crystallize ALKBH7 with a bound m^2_2G nucleoside and a short RNA probe carrying m^2_2G in its motif sequence can display direct evidence for the recognition mechanism.

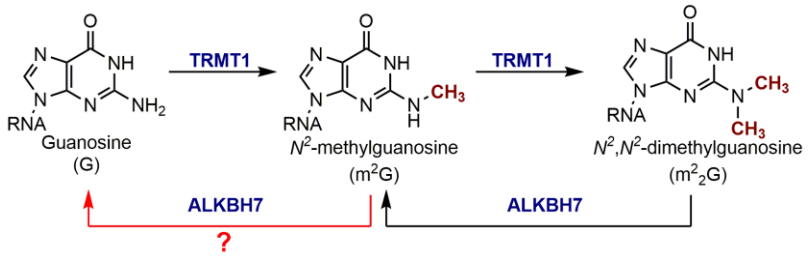

Figure 4.2.2. Proposed reversible RNA methylation at the position 26 of some mt-tRNAs.

After the initial discovery of ALKBH7 demethylation activity towards mt-tRNA methylation, we learned from recent publications on known G26 m^2_2G sites.[174,178] There are 6 mt-tRNAs that contain G26 (**Figure 4.2.3A**); however, people previously identified mt-tRNA-Ile G26 as the m^2_2G site, with G26 in other 5 mt-tRNAs characterized as nonmethylated or unclear. In the previously plotted data (**Figure 4.2.1E**), though we found mt-tRNA-Ile is of the highest m^2_2G/G level in pull-down experiment of individual mt-tRNA, we actually observed lower m^2_2G/G levels in other mt-tRNAs containing G26. We might not exclude the possibility that we may pull-down some cytosolic tRNAs (with abundant m^2_2G) as the contamination, but we do see an increase of m^2_2G/G levels in the case of *Alkbh7* knockout in mice. We further conducted tRNA-seq using HIV reverse transcriptase to read misincorporation information at the position 26 of these 6 mt-tRNAs (**Figure 4.2.3A**) in three human cell lines.

With careful data analysis, we clearly observed a >85% misincorporation rate at G26 of mt-tRNA-Ile, demonstrating the corresponding high methylation status (**Figure 4.2.3B**). To our surprise, there also exist 5-18% misincorporation levels in other 4 mt-tRNAs of G26 (including mt-tRNA-Ala, mt-tRNA-Arg, mt-tRNA-Asn and mt-tRNA-Glu), indicating lower methylation levels (**Figure 4.2.3C**). But mt-tRNA-Gln hardly shows any misincorporation, in which we regard it as the non-methylated G26. We harvested reliable and convincing tRNA-seq data with an

adequate sequencing coverage at each G26 position. Especially, we selected three human cell lines of different ALKBH7 expression levels, with an order as HeLa < HepG2 < HEK 293. Interestingly, within the 4 low-methylated G26 mt-tRNAs, we see mutation rate decreases with the increase of ALKBH7 expression level (**Figure 4.2.3C**).

In cytoplasm, TRMT1 installs all G26 m^2_2G with high methylation levels (**Figure 2.5.2A**); in mitochondria, TRMT1 also produces m^2_2G at G26, but most of them are at low methylation levels, which might illustrate a new biology correlated to ALKBH7 demethylation at these sites. To depict this phenomenon more generally, tRNA modifications are usually of high methylation level; it is rare that mito-tRNA G26 could be of low methylation level in several cases.

When going back to previous data on pull-down of individual mt-tRNAs, we found a controversial case that mt-tRNA-Gln still showcases a m^2_2G/G level at LC-MS/MS (though it seems lowest), but mt-tRNA-Gln was identified as non-methylated G26 in tRNA-seq (**Figure 4.2.3C**). This reminds me that there could contamination from cytosolic tRNAs (and also 5S/5.8S rRNA maybe). On one hand, all individual mt-tRNA pull-down experiments need to be repeated further; on another hand, in the case of studying mt-tRNA m^2_2G, it may not be an excellent way to utilize mass spec all the time. We realized that tRNA-seq followed by misincorporation-based analysis could serve as an efficient methodology to prove ALKBH7 demethylation on all these 5 high-/low-methlated G26 position, with the strategy shown below:

(i) We can overexpress ALKBH7 in HeLa and HAP1 (ALKBH7 low expression), then check how much mt-tRNA-Ile G26 m^2_2G misincorporation rate decreases.

(ii) We then knockdown ALKBH7 in HepG2 and SY5Y (ALKBH7 high expression), then see how much mt-tRNA/Ala/Arg/Asn/Glu G26 m^2_2G misincorporation rate increases.

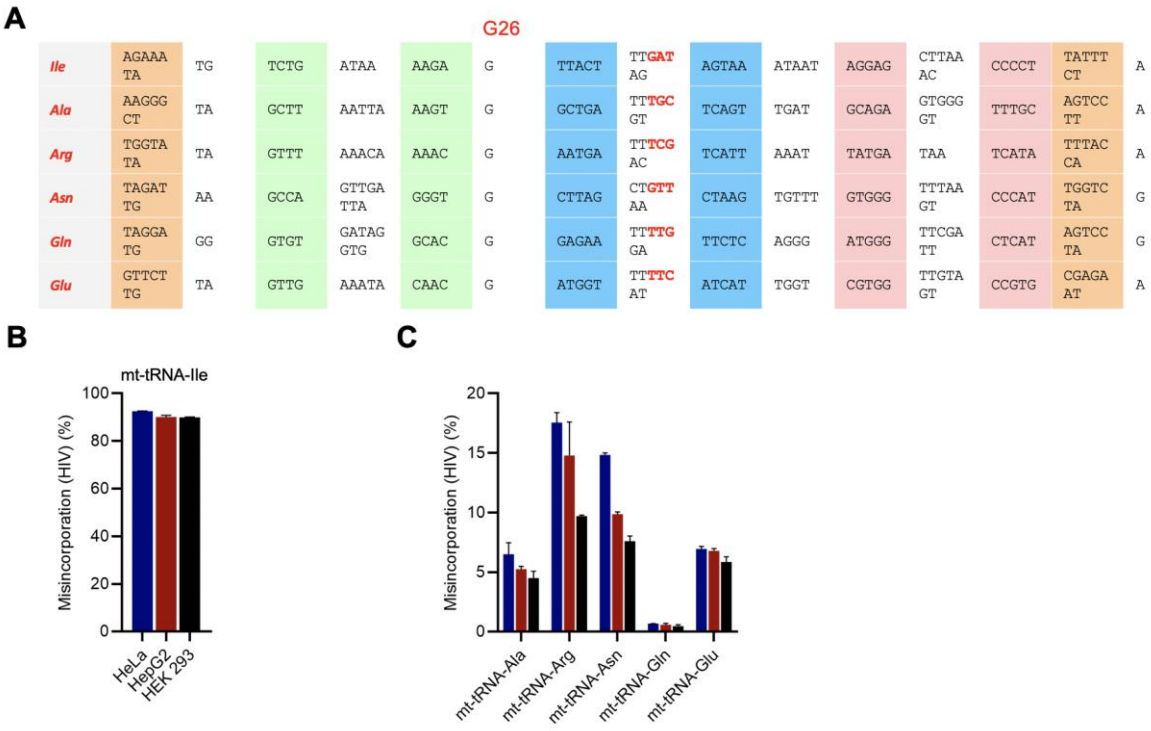

Figure 4.2.3. Mitochondrial tRNA G26 is methylated with diverse methylation level. (A) A Scheme to show all 6 mitochondrial tRNAs that contain G26. (B) Misincorporation levels (by HIV reverse transcriptase) of mt-tRNA-Ile G26 in three human cell lines, which is of a high methylation status. (C) Misincorporation levels (by HIV reverse transcriptase) of other 5 mt-tRNAs containing G26 (mt-tRNA-Ala, mt-tRNA-Arg, mt-tRNA-Asn, mt-tRNA-Gln, mt-tRNA-Glu) in three human cell lines, which are of lower methylation levels except mt-tRNA-Gln.

4.3 Investigating cellular functions of the ALKBH7-mediated mt-tRNA demethylation

According to a recently published work by Pan lab,[179] we planned to perform mitochondrial translation assay to show how ALKBH7 affects translation efficiency by regulating a group of mature tRNA. The ALKBH7-mediated tRNA demethylation may either affect tRNA level or its function in translation. We can use Northern blot and tRNA-seq to carefully analyze the levels of the 5 mt-tRNAs that contain G26 m^2_2G: (i) ALKBH7 overexpression in HeLa and HAP1 (ALKBH7 low expression) for the level of mt-tRNA-Ile; (ii) ALKBH7 knockdown in HepG2 and SY5Y (ALKBH7 high expression) for the levels of mt-tRNA/Ala/Arg/Asn/Glu.

In addition, it is necessary to purify total mt-tRNA or mitochondrial lysate from wild-type and *Alkbh7*$^{-/-}$ MEF cells. These tRNAs can be used in *in vitro* translation assays using mitochondrial translation machinery and a reporter mitochondrial mRNA as the template.[180] If m^2_2G in tRNA affects translation we should observe altered protein synthesis using pools of mt-tRNAs that carry different levels of m^2_2G. There are only 13 mitochondrial mRNAs that are translated inside mammalian mitochondria. These proteins are components of the respiration chain complex. We can use western to monitor protein level changes in the presence and absence of ALKBH7.

Importantly, before starting with a pipeline of experiments on mitochondrial translation, we first conducted qRT-PCR to make sure that mitochondrial mRNA levels (13 mitochondrial proteins encoded by mitochondrial DNA) are not affected by ALKBH7 knockdown. However, to our surprise, nearly all 13 mt-mRNA and 2 mt-rRNAs levels increase (some of them dramatically) after ALKBH7 knockdown in HepG2 cells (**Figure 4.3.1A**). These results can be well repeat-

ed and might explain the phenotype of ALKBH7 depletion, which leads to a higher protein level and obesity.

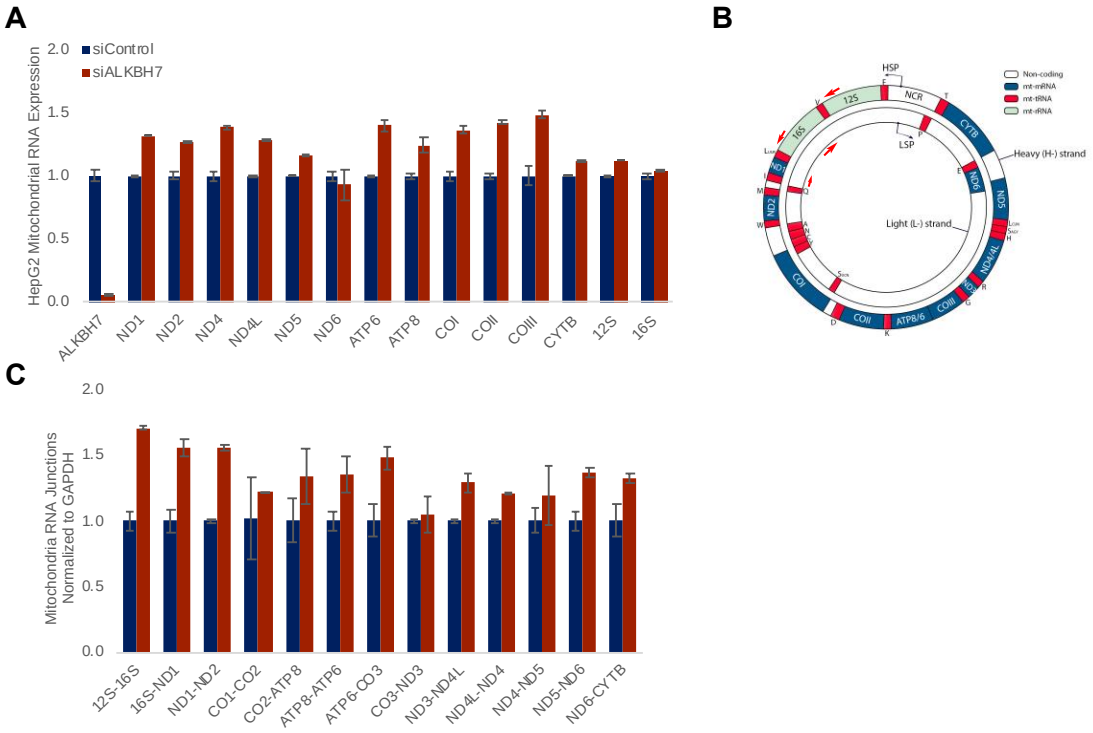

Figure 4.3.1. ALKBH7 depletion affects levels of mitochondrial mature mRNA and pre-mRNA. (A) HepG2 mitochondrial RNA levels (mRNAs and rRNAs) by qRT-PCR normalized to GAPDH, in the cases of siControl vs siALKBH7. (B) A scheme to show mitochondrial DNA and junction primers. (C) HepG2 mitochondrial nascent RNA (by EU-labeling and biotin pull-down) junction levels by qRT-PCR normalized to GAPDH, in the cases of siControl vs si-ALKBH7.

Mitochondrial defects have been linked to a variety of human disorders or diseases. While most mitochondrial proteins are made outside of mitochondria and subsequently been imported into mitochondria, the mitochondrial DNA does encode 2 rRNAs, 22 tRNAs and 13 mRNAs (**Figure 4.3.1B**). These units work together to produce key proteins of respiration chain complex in order to support normal respiration. Mitochondrial transcription leads to long polycistronic RNAs that are subsequently processed to individual rRNAs, tRNAs, and

mRNAs.[181] It is unclear how human mitochondrion fine-tunes its transcription and protein synthesis. One hypothesis focuses on fine-tune of translation through modulating tRNA and rRNA modifications. This discovery and subsequent investigations may uncover a new mechanism on how mitochondrion controls its protein synthesis. Upon the results that ALKBH7, a potential mt-tRNA demethylase, affects the levels of mature mitochondrial mRNA. We ask whether some co-transcriptional mt-tRNA modifications (especially G26 methylation) regulate the long polycistronic RNA processing and the fate of pre-mRNAs in mitochondria. To study the unprocessed RNA or the precursor RNA, it could be one way to check the level change of RNA junctions which contain one tRNA or several tRNAs connected together. These RNA junctions are supposed to exist in mitochondrial unprocessed/precursor RNA but not in mature mRNA or tRNA/rRNA. We designed several primers for each RNA junction and isolated HepG2 mitochondrial nascent RNA (by EU-labeling and biotin pull-down) for qRT-PCR experiments. RNA junction levels were normalized to GAPDH and measured in ALKBH7 knockdown *versus* control. A dramatic increase of RNA junction levels for nearly all junction sites were observed in ALKBH7-depleted cells, compared with the control (**Figure 4.3.1C**). It might indicate mitochondrial polycistronic RNA processing is accelerated and generates more precursor RNA in the working machinery in the absence of ALKBH7, which may demethylate key m^2_2G sites essentially and control the processing speed. This leads to a basic question that when ALKBH7 demethylate G26 m^2_2G: in the mt-pre-RNA stage or in the mature tRNA stage.

Human TRMT1 serves as the methyltransferase for the m^2_2G installation in mt-tRNAs. A recent report (77) suggested that knockdown of TRMT1 resulted in lower m^2_2G levels in tRNA, which caused elevated ROS and reduced cell proliferation. This observation is consistent with our proposal that the ALKBH7-mediated demethylation yields lower levels of m^2_2G in mt-

tRNA, which may lead to decreased mitochondrial translation, balance ROS levels in mitochondria, and affect cell necrosis and proliferation. We will test these hypotheses using the wild-type and *Alkbh7*$^{-/-}$ MEF cells, and study the responses of these cells to exogenous oxidizing agents. The ALKBH7 level could be tuned in response to ROS, as a way to modulate mt-tRNA modification and translation of mt-mRNAs encoding key respiration chain components. If ALKBH7 regulates mature tRNA G26 demethylation post-transcriptionally and control m^2_2G levels in the 5 mt-tRNAs, it is quite possible that ALKBH7 will modulate mitochondrial translation, which means a higher ALKBH7 level causes a lower mt-tRNA m^2_2G level and an attenuated translation efficiency[168,179]. If ALKBH7 is also responsible for mt-tRNA G26 demetylation in the mitochondrial polycistronic RNA stage co-transcriptionally or simultaneously together with the polycistronic RNA processing, it seems also likely that G26 demethylation controls the fate of mitochondrial pre-mRNAs.

4.4 Conclusion and Discussion

In summary, an array of mass spec evidences suggested the new role of human ALKBH7 as a demethylase working on mitochondrial tRNA and might be the first identified natural biological demethylase working on RNA guanosine bases. Especially, the G26 m^2_2G in mt-tRNA-Ile display a high conservation behavior across a number of mammalian species, which demonstrate the key function of this modification. The subsequent study will first focus on mitochondrial translation, to examine the hypothesis whether a lower mt-tRNA m^2_2G level leads to an attenuated translation efficiency in the presence of high-level ALKBH7. More importantly, mt-tRNA G26 methylation/demethylation might modulate mitochondrial polycistronic RNA processing and control the maturation of mitochondrial pre-mRNAs, in which TRMT1 is supposed to be studied along with ALKBH7 in all proposed pathways.

Considering there exist diverse other mitochondrial tRNA modifications like m^1A, m^1G, m^2G, we ask several more universal question (**Figure 4.4.1B**): (i) how RNA methylation/demethylation regulate each stage of mitochondrial polycistronic RNA processing and mt-pre-mRNA biogenesis; (ii) when these methyl groups are installed/removed at mt-tRNAs, co-transcriptionally or post-transcriptionally or co-processing; (iii) how is the difference of methylation status of methylated sites in mt-pre-RNAs and mature mt-tRNAs. Based on the previous sections in this thesis, we will utilize the sequencing methods that quantify misincorporation levels to study all methylation/demethylation behaviors across all stages of mitochondrial RNA, from polycistronic to mature. Actually, we have already performed tRNA-seq using HIV reverse transcriptase towards mature mt-tRNAs and harvested valuable data. We will also isolate mitochondrial polycistronic RNA in several way and conduct the similar sequencing technology relied on observing misincorporation rate change at modified sites. At first, we can use ethynyl ur-

idine (EU) metabolic labeling to pull-down mitochondrial nascent RNA, which is majorly unprocessed/precursor RNA; secondly, considering the mitochondrial DNA bidirectional transcription produces the newly transcribed RNA with a region of dsRNA (**Figure 4.4.1A**), we can use antibody to pull-down the dsRNA according to recently published work;[182,183] thirdly, we can isolate pure unprocessed RNA and precursor RNA from mitochondrial total RNA by cutting gel.[184] The mitochondrial genome is small and we can sequence a bunch of samples with a low depth coverage for each sample, in which misincorporation-based sequencing technology servers as a powerful strategy to monitor mitochondrial RNA methylation/demethylation behaviors accurately and economically.

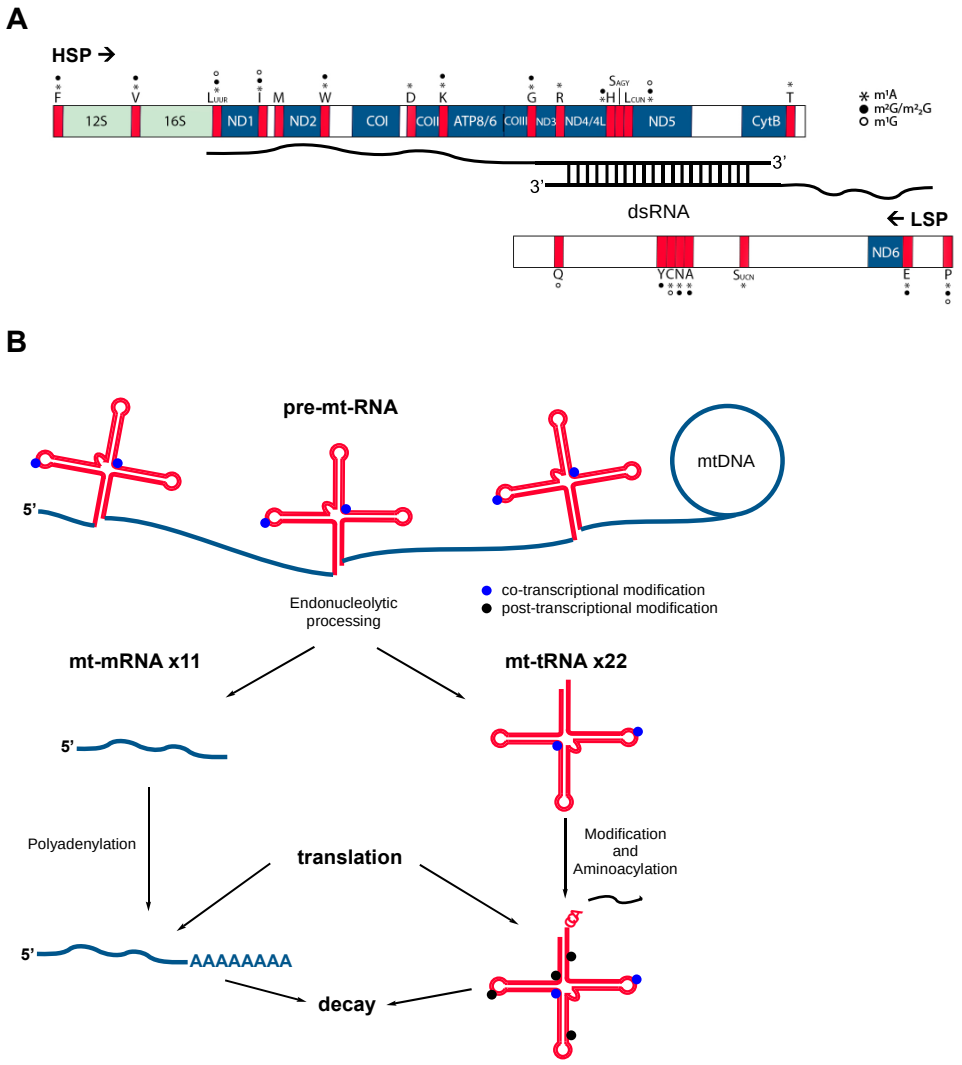

Figure 4.4.1 Mitochondrial tRNA methylation/demethylation might regulate polycistronic RNA processing. (A) Mitochondrial DNA bidirectonal transcription forms unprocessed dsRNA in vivo. (B) An illustration of pre-mt-RNA processing which might be regulated by co-transcriptional modification.

4.5 Materials and Methods

Cells culture

Human HepG2 cell line used in this study was all purchased from ATCC (the American Type Culture Collection). HepG2 cell line was maintained in DMEM (Gibco, 11965), supplemented with 10% FBS and 1% 100X Pen/Strep (Gibco). Cells were cultured at 37 °C under 5.0% CO_2.

siRNA knockdown

AllStars negative control siRNA (Qiagen, SI03650318) was used as control siRNA in knockdown experiments. ALKBH7 siRNAs were ordered from Qiagen (Hs_SPATA11_3 FlexiTube siRNA, SI00730793). Transfection was performed with Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen) following the manufacturer's protocols. Generally, 15-cm plates of HeLa cells were prepared for 48-hour knockdown.

mt-RNA isolation

(i) Mitochondrial total RNA isolation: To harvest the cells at first, the media was aspirated, and the 200~400 million HepG2 cells were washed once with 5 mL ice-cold PBS for each plate. With the cell pellet, mitochondria was further isolated by using Mitochondria Isolation Kit (human, Miltenyi Biotec, based on the renowned MACS Technology), to produce the mitochondria pellet of high purity and high yield. Then total mt-RNA was purified with TRIzol reagent (Invitrogen) and then extracted following manufacturer's protocol by isopropanol precipitation.

(ii) Mitochondrial small RNA (<200 nt) isolation: From the total mt-RNA obtained from step (i), small RNA with a size under 200 nt was further purified with the help of mirVana miRNA Isolation Kit (Thermo Fisher Scientific).

(iii) Mitochondrial tRNA isolation: From the harvested small mt-RNA (<200 nt), several micrograms of small RNA were loaded into Novex TBE-Urea Gels (15%) and then gel-cutting was performed for collecting mt-tRNA bands of desired sizes between 50~80 nt. The mt-tRNA re-

gion cut from gel was further handled by ZR small-RNA™ PAGE Recovery Kit (Zymo Research) to extract RNA with efficiency.

(iv) Mitochondrial nascent RNA isolation: EU-labeling was conducted according to the procedures of Click-iT™ Nascent RNA Capture Kit (Thermo Fisher Scientific) in HepG2 cell line, followed by mitochondrial total RNA isolation in step (i). Then a standard pull-down of biotin-tagged nascent RNA was carried out for isolated mitochondrial total RNA following manufacturer's protocol of Click-iT™ Nascent RNA Capture Kit.

LC-MS/MS

Around 200-300 ng mt-tRNA or mitochondrial small RNA was digested in a two-step manner with nuclease P1 (1uL, Sigma-Aldrich) in 20 ul reaction buffer containing 10 mM of NH_4OAc (pH= 5.3) at 42 °C for 2 h. Then, 1 ul of shrimp alkaline phosphatase (rSAP, NEB) was added along with 2.5 ul of 10X CutSmart buffer (NEB) and incubated at 37 °C for 2 h. After the incubation, the sample was diluted with additional 40 ul water and filtered with 0.22 μm filters (4mm diameter, Millipore) and 8 μl of the entire solution was injected into LC-MS/MS. For all the quantification, a mock control with only digestion buffers and enzymes was included each time and was later used for subtraction of baseline signals. Nucleosides were separated, by reverse phase ultra-performance liquid chromatography, on a C1 column with on-line mass spectrometry detection by an Agilent 6410 QQQ triple-quadrupole LC mass spectrometer, in positive electrospray ionization mode. Quantification was performed in comparison with the standard curve, obtained from pure nucleoside standards running with the same batch of samples. The m^2_2G level was calculated as the ratio of m^2_2G to G based on calibrated concentration curves.

mt-tRNA-seq

4-6 ug mt-tRNA was slightly fragmented using RNA Fragmentation Reagents for 15 minutes at 70°C (AM8740, Invitrongen) following the manufacturer's protocol. Then 3' dephosphorylation/end repair was performed with T4 Polynucleotide Kinase (EK0032, Thermo

Fisher Scientific): RNA was mixed with 5 µL 10X T4 Polynucleotide Kinase Reaction Buffer (B0201S, NEB) and 5 µL T4 PNK, diluted to a final volume of 50 µL and incubated at 37°C for 1 hour. Then a further dephosphorylation step was conducted at 37 °C for 1.5 hour in 50 µL 1X CutSmart Buffer (NEB) with 5 µL Shrimp Alkaline Phosphatase (rSAP, NEB). Starting with the 3'-adapter ligation, repaired and dephosphorylated RNA fragments were incubated with 2 equivalents of 3' adaptor (5'rApp-AGATCGGAAGAGCGTCGTG-3SpC3) at 70 °C for 2 min and transferred to ice immediately. 5 µL 10X T4 RNA Ligase Reaction Buffer, 15 µL 50% PEG8000 and 4 µL T4 RNA ligase 2 truncated KQ (NEB) were added accordingly to the RNA-adaptor mixture, then the reaction was diluted to a final volume of 50 µL and incubated at 25 °C for 2 hours followed by 16 °C for 8 hours. The reaction was then diluted to 100 µL, and the excessive adaptors were digested with 3 µL 5' Deadenylase (NEB) at 30 °C for 1 h followed by adding 1.5 µL RecJf (NEB) for ssDNA digestion at 37 °C for another hour. 3'-end ligated RNA was extracted using RNA Clean & Concentrator (Zymo Research).

Then RNAs were proceeded with reverse transcription step and first incubated with 2 uM RT primer (5'-ACACGACGCTCTTCCGATCT-3') at 65 °C for 2 min and cooled on ice immediately. The reactions were then prepared with RNA mixture, 2 ul 10 mM dNTP (Deoxynucleotide (dNTP) Solution Mix, NEB), 1X AMV Reverse Transcriptase Reaction Buffer (NEB), 1 µL Recombinant HIV reverse transcriptase (Worthington), and 1 µL RNaseOUT recombinant ribonuclease inhibitor (Thermo Fisher Scientific) and then incubated at 37 °C for 1 hour. The enzymes were deactivated by incubation at 75 °C for 5 min and stored on ice. The cDNAs were purified with Oligo Clean and Concentrator (Zymo Research). Purified cDNA was then subject to cDNA 3'-adaptor ligation. The cDNA was first denatured with 50 uM cDNA linker (5' Phos-NNNNNNNNNNAGATCGGAAG AGCACACGTCTG-3SpC3) at 75 °C for 2 min. 2 µL 10X T4 RNA Ligase Reaction Buffer, 10 µL 50% PEG8000, 2 µL 10mM ATP and 1 µL T4 RNA ligase 1 (high concentration, NEB) were added accordingly to the cDNA-adaptor mixture, then

the reaction was incubated at 25 °C for 12 hours. The library was then amplified with the universal and indexed primers from NEBNext Multiplex Oligos for Illumina (NEB). All libraries were sequenced on Illumina NextSeq 500 with single-end 80 bp read length.

RT-qPCR

All RNA templates used for RT–qPCR were digested with DNase I in the purification step. RT-qPCR primers were designed to cover exon-exon junctions shared in all isoforms of the matured mRNAs. About 500 ng RNAs were first reverse-transcribed into cDNAs with PrimeScript™ RT reagent Kit (Takara), and then the cDNAs were subjected to qPCR analysis with FastStart SYBR Green Master Mix (Roche) in machine LightCycler 96 (Roche). *Actb* and *gapdh* were used as internal controls as they showed relative invariant expression according to pilot RT–qPCR data. The sequences of primers used in this study are listed below:

ND1: forward 5′-ATACCCCCGATTCCGCTACGAC-3′,
reverse 5′-GTTTGAGGGGGAATGCTGGAGA-3′;

ND2: forward 5′-GGCCTAGAAATAAACATGCTA-3′,
reverse 5′-GGGCTATTCCTAGTTTTATT-3′;

ND3: forward 5′-ACCCTACAAGCTCTGCACG-3′,
reverse 5′-GCTCATGGTAGTGGAAGTAGAAG-3′;

ND4: forward 5′-AATCAGCCACATAGCCCTCG-3′,
reverse 5′-ATGAGGATGTAAGCCCGTGG-3′;

ND4L: forward 5′-ACTCCCACTAATAGCTTTTTGATG-3′,
reverse 5′-AGGGCTGTGACTAGTATGTTGAG-3′;

ND5: forward 5′-CAAAACCTGCCCCTACTCCT-3′,
reverse 5′-GGGTTGAGGTGATGATGGAG-3′;

ND6: forward 5′-ACACTCACCAAGACCTCAACC-3′,

reverse 5′-TAGTTTTTTTAATTTATTTAGGGGGAAT-3′;

COI: forward 5′-CGATGCATACACCACATGAA-3′,

reverse 5′-AGCGAAGGCTTCTCAAATCA-3′;

COII: forward 5′-GCTGTCCCCACATTAGGCTT-3′,

reverse 5′-ACCGTAGTATACCCCCGGTC-3′;

COIII: forward 5′-AAAAGGCCTTCGATACGGGA-3′,

reverse 5′-ATTTAGCGGGGTGATGCCTG-3′;

ATP8: forward 5′-CCACCTACCTCCCTCACCAAA-3′,

reverse 5′-GATTGTGGGGGCAATGAATGA-3′;

ATP6: forward 5′-CGCCACCCTAGCAATATCAA-3′,

reverse 5′-TTAAGGCGACAGCGATTTCT-3′;

CYTB: forward 5′-AATTCTCCGATCCGTCCCTA-3′,

reverse 5′-GGAGGATGGGGATTATTGCT-3′;

GAPDH: forward 5′-GTCTCCTCTGACTTCAACAGCG-3′,

reverse 5′-ACCACCCTGTTGCTGTAGCCAA-3′;

12S: forward 5′-GGTTGGTCAATTTCGTGCCAGC-3′,

reverse 5′-GGGGTGATCTAAAACACTCTTTACGC-3′;

16S: forward 5′-AGACTTCACCAGTCAAAGCGA-3′,

reverse 5′-ACATCGAGGTCGTAAACCCT-3′;

ALKBH7: forward 5′-AGACGCTGAGCCGAGAACTGG-3′,

reverse 5′-TTCTGACCAGCGCGACTTCTCT-3′;

CHAPTER 5

A New Chapter of mRNA Modifications: Misincorporation-Based Base-Resolution Sequencing

mRNA m^6A is prevalent with regulatory roles but one of many post-transcriptional mRNA modifications that have been unveiled in past several years. The current technologies for m^6A-seq in mRNA are majorly antibody-assisted and enrichment-based, in which people perform peak calling for downstream analysis. However, enrichment-based approach cannot display completed information on mRNA methylation in some highly structured regions, due to weaker binding affinity of antibody; enrichment-based means might also bring false positive peaks in some circumstances. Then enrichment folds of one specific peak is determined by an array of factors and cannot reflect the actual methylation status of one m^6A site. People revealed m^1A methylome in eukaryotic mRNAs with unique distribution pattern in proximity to translation starting sites and the first splice site[95,96]. In this work, m^1A-modified mRNA fragments were enriched by anti-m^1A antibody at the first stage then further confirmed by its unique misincorporation pattern under HIV reverse transcriptase or TGIRT reverse transcriptase. However, the misincorporation level harvested in this way might also not be applied for calculating the accurate methylation level of specific m^1A sites due to the pull-down enrichment process. In our mRNA m^7G work, we showcase our well-developed methodology can expose the reasonable estimated methylation status of highly methylated sites in all tRNAs and in around fifty mRNAs. With this striking progress from misincorporation-based methodology, we become able to select highly methylated internal m^7G sites as representatives for the subsequent functional study, which might express much more dramatic effect caused by this positively-charged guanosine base.

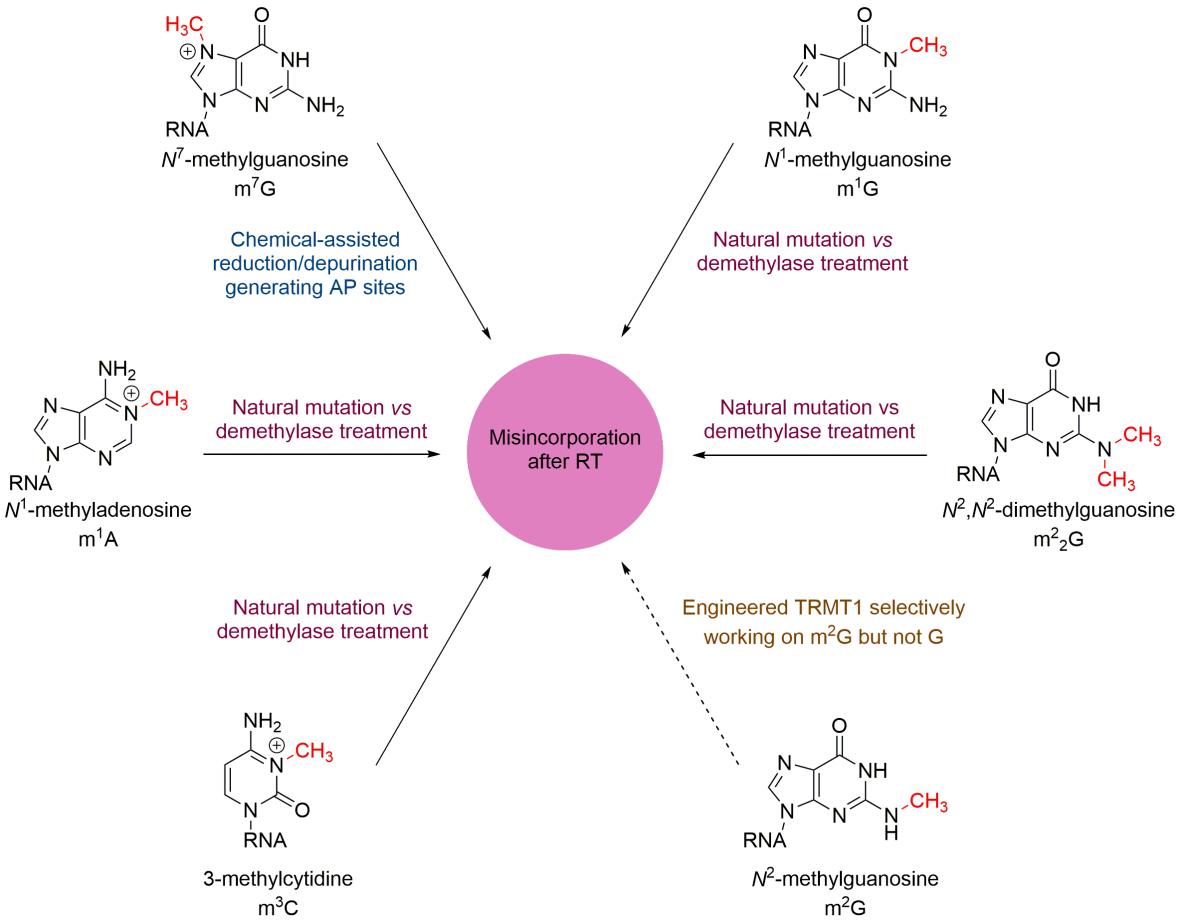

Figure 5.1. Diverse internal mRNA modifications that can be investigated *via* misincorporation-based sequencing at base-resolution.

When we were sequencing tRNAs using m^7G-seq, the sequencing coverage at each base could be deep enough for clear calculation of misincorporation rate without biotin-pulldown enrichment. But in cases of mRNA, the problem of sequencing coverage prevented us from giving clear conclusions of misincorporation level on all potential sites, except the fifty mRNAs. In this way, we need biotin-pulldown results for identifying more internal m^7G sites with lower methylation status. With the well-established model for distinguishing m^7G unique mutation behavior in RNA input *versus* that after pulldown enrichment, we characterized hundreds of reliable internal m^7G sites across human cancer cell lines. Although the misincorporation levels after biotin pull-

down enrichment cannot be utilized directly for methylation fraction estimation of these hundreds of internal m^7G sites, we can design strategies and experiments to study a group of selected ones. For instance, if we pick up around one hundred top-ranked internal m^7G sites (identified after pulldown), we make one hundred biotin-tagged DNA oligo probes to pulldown the one hundred corresponding m^7G-modified regions from mRNA fragments, followed by m^7G-seq without pulldown enrichment. In this way, the sequencing coverage would be enhanced quite a lot at the internal m^7G site and lead to a tRNA-like library quality for harvesting at least thousands of reads accumulated at one specific internal m^7G site (before pulldown). Besides, as for m^1A, the similar enrichment strategy with biotin-tagged DNA oligo probes could be performed for the group of representative m^1A targets prior to m^1A-seq.

As a result, the strategy shifting effect introduced by misincorporation-based methods further impact research on exploring other mRNA modifications with establishing a new platform on misincorporation behaviors of them (**Figure 5.1**). As mentioned in chapter 2, except m^7G or m^1A, an array of other guanosine modifications like m^1G and $m^2{}_2G$ could express a natural misincorporation pattern due to the methyl groups disrupt base-pairing during reverse transcription (**Figure 2.5.3** and **Figure 2.5.4**). The misincorporation pattern of both m^1G and $m^2{}_2G$ attributes to dominant G to T mutation with a high misincorporation percentage. With the help of recently developed engineered AlkB enzymes, people became able to selectively remove the methyl group at m^1G or $m^2{}_2G$ base,[172,185] which paves the path for distinguishing these two methylated guanosine bases in mRNA with different demethylase treatments. Another mammalian tRNA guanosine modification m^2G does not display natural mutation according to our previous data (**Figure 2.5.2**), in which we need to screen for engineered TRMT1 enzyme that can preferentially methylate m^2G rather than unmodified G to expose misincorporation information from

mRNA m^2G sites only.[176,177] In addition, 3-methylcytidine (m^3C) was recently reported to exist in human and mouse mRNA with METTL8 responsible for the mRNA methylation.[186] Considering its property that exhibits natural mutation under reverse transcription with the methyl group at the base-paring side, the similar demethylase-assisted misincorporation-based sequencing methodology could also be utilized for base-resolution m^3C-seq in the future.

For comparing strategies between misincorporation sites and stop sites, pseudouridine ψ as another example, as the most abundant modification in cellular RNA, was reported mapped at single-base resolution in yeast and mammalian mRNAs[99-102], making use of the unique chemical properties of pseudouridine to generate stop sites during reverse transcription. A recent work has also reported the base-resolution map of Nm bases in mammalian mRNA and discovered its unique enrichment in CDS, near splice sites, and in introns, in terms of converting Nm sites into stop sites[109]. Dated back to the initial round of developing base-resolution sequencing techniques for internal m^7G sites, a number of attempts were focusing on aniline cleavage followed by $NaBH_4$ reduction, which also left internal m^7G sites as stop sites. However, the further analysis confirmed that the strategy on handling millions of stop sites for determining m^7G candidate sites are of high false positive rate and especially not adapted to the cases of low methylation status, in which the technical barrier might also exist for ψ-seq and Nm-seq. Pseudouridine ψ exhibits the ability for altering base-pairing interactions that affect RNA structures and mRNA coding[184], indicating the potential regulatory roles; Nm methylome also takes on potential functions in translation elongation dynamics and mRNA splicing. Considering the unique chemical properties of ψ and Nm bases and the significance in their functional study, it might be possible to re-exploit proper approaches for generating RT misincorporation at these bases, which can also expose and

estimate reasonable methylation levels for people to select real top targets for future investigation.

For the completed map of mRNA fate in some top m^7G targets, with the identified representative mRNA targets CNOT2 (5' UTR internal m^7G) and RTN4 (3' UTR internal m^7G) as examples, we could monitor the methylation level of the specific internal m^7G site in different stages of mRNA fate. By conducting ethynyl uridine (EU) metabolic labeling to pull-down chromatin-associated RNA, m^7G -seq without pulldown enrichment can estimate the methylation behavior in nascent RNA stage. The misincorporation-based m^7G -seq makes it possible to monitor methylation level evolvement from pre-mRNA stage to mature mRNA stage, or from nucleus to cytoplasm. Especially, with the previously mentioned strategy utilizing biotin-tagged DNA oligo probes, people are also able to study an array of internal m^7G top targets simultaneously through several mRNA stages, from nascent to mature. Besides, the nuclear internal m^7G could be of great value and importance to be studied for in-depth investigation of internal m^7G function, including aspects of intron methylation and m^7G-dependent pre-mRNA splicing. m^7G at the cap structured has been studied for decades as the key positively-charged factor recruiting a series of 'reader' proteins and processing machineries that recognize and bind cap structure, including eIF4E, CBP20/CBP80 complex and other cap binding proteins. All these existing cap m^7G binding proteins might be promising candidates for potential 'readers' of internal m^7G sites within mRNA.

Overall, diverse mRNA internal modifications display unique distribution patterns and impact multiple RNA metabolic processes to regulate gene expression, through RNA structure alteration or recognition by specific reader proteins. However, our understanding and research on mRNA modifications might still remain at the pretty early stage. The demand for more accurate

and efficient tools to sequence these modifications is urgently needed for all types of samples including cases of single-cell, tiny amount RNA, blood samples, etc. As new modifications and the related new functional pathways continue to be unveiled, this emerging field provides new opportunities for chemical biologists to study these post-transcriptional chemical marks, including developing new high-resolution approaches to map the modifications in a transcriptome-wide manner by high-throughput sequencing, searching for the methylation installation/demethylation uninstallation machineries, investigating the correlation between the modification status and the effect on gene expression, and inventing small molecules or therapeutic means to fine-tune these pathways in various human diseases in the future. The pervasiveness of RNA modifications indicates the widespread influence among a huge among of RNA-involved biological events and pathways, which will be of crucial roles for further understanding of cellular epigenetic information encoding and global gene regulation.

LIST OF REFERENCES

1 Waddington, C. H. The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology. (Allen & Unwin, 1957).

2 Waddington, C. H. The Epigenotype. *Endeavour* **1**, 18-20 (1942).

3 Holliday, R. DNA Methylation and Epigenetic Inheritance. *Philos T Roy Soc B* **326**, 329-338 (1990).

4 Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Gene Dev* **23**, 781-783 (2009).

5 Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21 (2002).

6 Wyatt, G. R. Recognition and Estimation of 5-Methylcytosine in Nucleic Acids. *Biochem J* **48**, 581-584 (1951).

7 Holliday, R. & Pugh, J. E. DNA Modification Mechanisms and Gene Activity during Development. *Science* **187**, 226-232 (1975).

8 Riggs, A. D. X-Inactivation, Differentiation, and DNA Methylation. *Cytogenet Cell Genet* **14**, 9-25 (1975).

9 Kalousek, F. & Morris, N. R. Deoxyribonucleic Acid Methylase Activity in Rat Spleen. *J. Biol. Chem*. **243**, 2440-2442 (1968).

10 Roy, P. H. & Weissbach, A. DNA methylase from Hela cell nuclei. *Nucleic Acids Res*. **2**, 1669-1684 (1975).

11 Bestor, T. H. & Ingram, V. M. 2 DNA Methyltransferases from Murine Erythroleukemia Cells - Purification, Sequence Specificity, and Mode of Interaction with DNA. *Proc Natl Acad Sci U.S.A* **80**, 5559-5563 (1983).

12 Jahner, D. *et al.* De novo methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* **298**, 623-628 (1982).

13 Okano, M. *et al.* DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

14 Okano, M. *et al*. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet*. **19**, 219-220 (1998).

15 Hata, K. *et al*. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129**, 1983-1993 (2002).

16 Li, E. *et al*. Role for DNA Methylation in Genomic Imprinting. *Nature* **366**, 362-365 (1993).

17 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).

18 Allfrey, V. G. *et al*. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci U.S.A* **51**, 786-794 (1964).

19 Kim, S. & Paik, W. K. Studies on the origin of epsilon-N-methyl-L-lysine in protein. *J Biol Chem* **240**, 4629-4634 (1965).

20 Wu, S. C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol*. **11**, 750-750 (2010).

21 Howlett, S. K. & Reik, W. Methylation Levels of Maternal and Paternal Genomes during Preimplantation Development. *Development* **113**, 119-127 (1991).

22 Rougier, N. *et al*. Chromosome methylation patterns during mammalian preimplantation development. *Gene Dev* **12**, 2108-2113 (1998).

23 Mayer, W. *et al*. Embryogenesis: Demethylation of the zygotic paternal genome. *Nature* **403**, 501-502 (2000).

24 Oswald, J. et al. Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* **10**, 475-478 (2000).

25 Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* **324**, 929-930 (2009).

26 Tahiliani, M. *et al*. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324**, 930-935 (2009).

27 Ito, S. *et al*. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133 (2010).

28 Ito, S. *et al*. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **333**, 1300-1303 (2011).

29 He, Y. F. *et al*. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **333**, 1303-1307 (2011).

30 Pfaffeneder, T. *et al*. The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angew Chem Int Ed* **50**, 7008-7012 (2011).

31 Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* **286**, 35334-35338 (2011).

32 Inoue, A. & Zhang, Y. Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos. *Science* **334**, 194-194 (2011).

33 Meehan, R. R. *et al*. Identification of a Mammalian Protein That Binds Specifically to DNA Containing Methylated Cpgs. *Cell* **58**, 499-507 (1989).

34 Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18**, 6538-6547 (1998).

35 Bird, A. P. & Wolffe, A. P. Methylation-induced repression: Belts, braces, and chromatin. *Cell* **99**, 451-454 (1999).

36 Prokhortchouk, A. *et al*. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Gene Dev* **15**, 1613-1618 (2001).

37 Mellén, M. *et al*. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417-1430 (2012).

38 Chen, R. *et al*. The 5-Hydroxymethylcytosine (5hmC) Reader UHRF2 Is Required for Normal Levels of 5hmC in Mouse Adult Brain and Spatial Learning and Memory. *J Biol Chem*. **292**, 4533-4543 (2017).

39 Fu, Y. *et al*. N^6-Methyldeoxyadenosine Marks Active Transcription Start Sites in Chlamydomonas. *Cell* **161**, 879-892 (2015).

40 Greer, Eric L. *et al*. DNA Methylation on N^6-Adenine in C. elegans. *Cell* **161**, 868-878 (2015).

41 Zhang, G. *et al*. N^6-Methyladenine DNA Modification in Drosophila. *Cell* **161**, 893-906 (2015).

42 Cantara, W. A. *et al*. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res* **39**, D195-D201 (2011).

43 Desrosiers, R. *et al*. Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3971-3975 (1974).

44 Desrosiers, R. C. *et al*. Characterization of Novikoff hepatoma mRNA methylation and heterogeneity in the methylated 5' terminus. *Biochemistry* **14**, 4367-4374 (1975).

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

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

47 Wei, C. M. & Moss, B. Methylation of Newly Synthesized Viral Messenger-RNA by an Enzyme in Vaccinia Virus. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3014-3018 (1974).

48 Perry, R. P. *et al*. The methylated constituents of L cell messenger RNA: Evidence for an unusual cluster at the 5' terminus. *Cell* **4**, 387-394 (1975).

49 Wei, C. M. *et al*. Methylated Nucleotides Block 5' Terminus of Hela Cell Messenger RNA. *Cell* **4**, 379-386 (1975).

50 Carroll, S. *et al*. N^6-methyladenosine residues in an intron-specific region of prolactin pre-mRNA. *Mol. Cell. Biol*. **10**, 4456-4465 (1990).

51 Harper, J. E. *et al*. Sequence specificity of the human mRNA N6-adenosine methylase in vitro. *Nucleic Acids Res*. **18**, 5735-5741 (1990).

52 Horowitz, S. *et al*. Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5667-5671 (1984).

53 Kane, S. E. & Beemon, K. Precise Localization of m6A in Rous Sarcoma Virus RNA Reveals Clustering of Methylation Sites: Implications for RNA Processing. *Mol. Cell. Biol*. **5**, 2298-2306 (1985).

54 Schibler, U. *et al*. Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. *J. Mol. Biol.* **115**, 695-714 (1977).

55 Wei, C. *et al*. 5'-Terminal and internal methylated nucleotide sequences in HeLa cell mRNA. *Biochemistry* **15**, 397 - 401 (1976).

56 Wei, C.-M. & Moss, B. Nucleotide sequences at the N^6-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry* **16**, 1672 - 1676 (1977).

57 He, C. Grand challenge commentary: RNA epigenetics? *Nat Chem Biol* **6**, 863-865 (2010).

58 Jia, G. F. *et al*. N^6-Methyladenosine in nuclear RNA is a major substrate of the obesity associated FTO. *Nat Chem Biol* **7**, 885-887 (2011).

59 Dominissini, D. *et al*. Topology of the human and mouse m(6)A RNA methylomes revealed by m(6)A-seq. *Nature* **485**, 201-206 (2012).

60 Meyer, K. D. *et al*. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3 ' UTRs and near Stop Codons. *Cell* **149**, 1635-1646 (2012).

61 Keith, J. M. *et al*. Hela-Cell RNA (2'-O-Methyladenosine-N6-) Methyltransferase Specific for Capped 5'-End of Messenger-RNA. *J Biol Chem* **253**, 5033-5039 (1978).

62 Narayan, P. & Rottman, F. M. An in vitro system for accurate methylation of internal adenosine residues in messenger RNA. *Science* **242**, 1159-1162 (1988).

63 Tuck, M. T. The Formation of Internal 6-Methyladenine Residues in Eukaryotic Messenger-RNA. *Int J Biochem* **24**, 379-386 (1992).

64 Bokar, J. A. *et al*. Characterization and partial purification of mRNA N^6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. *J Biol Chem* **269**, 17697-17704 (1994).

65 Bokar, J. A. *et al*. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N^6-adenosine) methyltransferase. *RNA* **3**, 1233-1247 (1997).

66 Liu, J. Z. *et al*. A METTL3-METTL14 complex mediates mammalian nuclear RNA N^6-adenosine methylation. *Nat Chem Biol* **10**, 93-95 (2014).

67 Ping, X. L. *et al*. Mammalian WTAP is a regulatory subunit of the RNA N^6-methyladenosine methyltransferase. *Cell Res* **24**, 177-189 (2014).

68 Wang, Y. *et al*. N-6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**, 191-198 (2014).

69 Schwartz, S. *et al*. Perturbation of m^6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5' Sites. *Cell Rep* **8**, 284-296 (2014).

70 Batista, Pedro J. *et al*. m^6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells. *Cell Stem Cell* **15**, 707-719 (2014).

71 Bodi, Z. *et al*. Adenosine methylation in Arabidopsis mRNA is associated with the 3' end and reduced levels cause developmental defects. *Front. Plant Sci*. **3** (2012).

72 Clancy, M. J. *et al*. Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N^6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Res*. **30**, 4509-4518 (2002).

73 Geula, S. *et al*. m^6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**, 1002-1006 (2015).

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

75 Schwartz, S. *et al*. High-Resolution Mapping Reveals a Conserved, Widespread, Dynamic mRNA Methylation Program in Yeast Meiosis. *Cell* **155**, 1409-1421 (2013).

76 Yue, Y. *et al*. VIRMA mediates preferential m^6A mRNA methylation in 3' UTR and near stop codon and associates with alternative polyadenylation. *Cell Discovery* **4**:10 (2018)

77 Falnes, P. O. *et al*. AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature* **419**, 178-182 (2002).

78 Trewick, S. C. *et al*. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **419**, 174-178 (2002).

79 Aas, P. A. *et al*. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421**, 859-863 (2003).

80 Yang, C. G. *et al*. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452**, 961-965 (2008).

81 Zheng, G. Q. *et al*. ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol Cell* **49**, 18-29 (2013).

82 Fu, Y. *et al*. FTO-mediated formation of N^6-hydroxymethyladenosine and N^6-formyladenosine in mammalian RNA. *Nature Communications* **4** (2013).

83 Fu, Y. *et al*. Gene expression regulation mediated through reversible m^6A RNA methylation. *Nat Rev Genet* **15**, 293-306 (2014).

84 Pan, T. N^6-methyl-adenosine modification in messenger and long non-coding RNA. *Trends Biochem Sci* **38**, 204-209 (2013).

85 Wang, X. *et al*. N^6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-120 (2014).

86 Wang, X. *et al*. N^6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388-1399 (2015).

87 Liu, N. *et al*. N^6-methyladenosine-dependent RNA structural switches regulate RNA protein interactions. *Nature* **518**, 560-564 (2015).

88 Wang, Y. *et al*. N^6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**, 191-198 (2014).

89 Kim, H. J. *et al*. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467-473 (2013).

90 Alarcon, C. R. *et al*. HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell* **162**, 1299-1308 (2015).

91 Fustin, J.-M. *et al*. RNA-Methylation-Dependent RNA Processing Controls the Speed of the Circadian Clock. *Cell* **155**, 793-806 (2013).

92 Shi, H. *et al*. YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Res* **27**, 315-328 (2017).

93 Hsu, P. J. *et al*. Ythdc2 is an N^6-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Res* **27**, 1115-1127 (2017).

94 Xiao, W. *et al*. Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell* **61**, 507-519 (2016).

95 Dominissini, D. *et al*. The dynamic N^1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441-446 (2016).

96 Li, X. *et al*. Transcriptome-wide mapping reveals reversible and dynamic N^1-methyladenosine methylome. *Nat Chem Biol* **12**, 311-316 (2016).

97 Li, X. *et al*. Base-resolution mapping reveals distinct m1A methylome in nuclear- and mitochondrial-encoded transcripts. *Mol Cell* **68**, 993–1005 (2017).

98 Safra, M. *et al*. The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **551**, 251-255 (2017).

99 Schwartz, S. *et al*. Transcriptome-wide mapping reveals widespread dynamic regulated psedouridylation of ncRNA and mRNA. *Cell* **159**, 148-162 (2014)

100 Li, X. *et al*. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol* **11**, 592-597 (2015).

101 Carlile, T. M. *et al*. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143-146 (2015)

102 Lovejoy, A. F. *et al*. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PLoS One* **9**, e110799 (2014)

103 Yang, X. *et al*. 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell Res* **27**, 606-662 (2017)

104 Cui, X. *et al*. 5-Methylcytosine RNA Methylation in Arabidopsis Thaliana. *Mol Plant* **10**, 1387-1399 (2017)

105 David, R. *et al*. Transcriptome-Wide Mapping of RNA 5-Methylcytosine in Arabidopsis mRNAs and Noncoding RNAs. *Plant Cell* **29**, 445-460 (2017)

106 Hussain, S. *et al*. Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol* **14**, 215-224 (2013).

107 Delatte, B. *et al*. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **351**, 282-285 (2016).

108 Arango, D. *et al*. Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell* 2018 **175**, 1872-1886 (2018).

109 Dai, Q. *et al*. Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat Methods* **14**, 695–698 (2017)

110 Zhao, B. S. *et al*. Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* **18**, 31-42 (2017)

111 Gilbert, W. V. *et al*. Messenger RNA modifications: Form, distribution, and function. *Science* **352**, 1408-1412 (2016)

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

113 Frye, M. *et al*. RNA modifications modulate gene expression during development. *Science* **361**, 1346-1349 (2018).

114 Roundtree, I.A., *et al*. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187-1200 (2017).

115 Furuichi, Y. Discovery of m(7)G-cap in eukaryotic mRNAs. *Proc Jpn Acad Ser B Phys Biol Sci* **91**, 394-409 (2015).

116 Cowling, V. H. Regulation of mRNA cap methylation. *Biochem J* **425**, 295-302 (2009).

117 Shimotohno, K. *et al*. Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. *Proc Natl Acad Sci U S A* **74**, 2734-2738 (1977).

118 Murthy, K. G. *et al*. A nuclear micrococcal-sensitive, ATP-dependent exoribonuclease degrades uncapped but not capped RNA substrate. *Nucleic Acids Res* **19**, 2685-2692 (1991).

119 Pei, Y. & Shuman, S. Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J Biol Chem* **277**, 19639-19648 (2002).

120 Konarska, M. M. *et al*. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell* **38**, 731-736 (1984).

121 Lindstrom, D. L. *et al*. Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol Cell Biol* **23**, 1368-1378 (2003).

122 Drummond, D. R. *et al*. The effect of capping and polyadenylation on the stability, movement and translation of synthetic messenger RNAs in Xenopus oocytes. *Nucleic Acids Res* **13**, 7375-7394 (1985).

123 Lewis, J. D. & Izaurralde, E. The role of the cap structure in RNA processing and nuclear export. *Eur J Biochem* **247**, 461-469 (1997).

124 Muthukrishnan, S. *et al*. 5'-Terminal 7-methylguanosine in eukaryotic mRNA is required for translation. *Nature* **255**, 33-37 (1975)

125 Guy, M. P. & Phizicky, E. M. Two-subunit enzymes involved in eukaryotic post-transcriptional tRNA modification. *RNA Biol* **11**, 1608-1618 (2014).

126 Sloan, K. E. *et al*. Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* **14**, 1138-1152 (2017).

127 Haag, S. *et al*. WBSCR22/Merm1 is required for late nuclear pre-ribosomal RNA processing and mediates N^7-methylation of G1639 in human 18S rRNA. *RNA* **21**, 180-187 (2015).

128 Õunap, K. *et al*. The human WBSCR22 protein is involved in the biogenesis of the 40S ribosomal subunits in mammalian cells. *PLoS One* **8**, e75686 (2013).

129 Létoquart, J. *et al*. Structural and functional studies of Bud23-Trm112 reveal 18S rRNA N^7-G1575 methylation occurs on late 40S precursor ribosomes. *Proc Natl Acad Sci U S A* **111**, E5518-E5526 (2014).

130 Leulliot, N. *et al*. Structure of the yeast tRNA m^7G methylation complex. *Structure* **16**, 52–61 (2008).

131 Alexandrov, A. *et al*. Two proteins that form a complex are required for 7-methylguanosine modification of yeast tRNA. *RNA* **8**, 1253-1266 (2002).

132 Alexandrov, A. *et al*. tRNA m^7G methyltransferase Trm8p/Trm82p: evidence linking activity to a growth phenotype and implicating Trm82p in maintaining levels of active Trm8p. *RNA* **11**, 821-830 (2005).

133 Shaheen, R. *et al*. Mutation in WDR4 impairs tRNA m7G46 methylation and causes a distinct form of microcephalic primordial dwarfism. *Genome Biol* **16**, 210 (2015).

134 Wintermeyer, W. & Zachau, H. G. A specific chemical chain scission of tRNA at 7-methylguanosine. *FEBS Lett* **11**, 160-164 (1970).

135 Küpfer, P. A. & Leumann, C. J. RNA abasic sites: preparation and trans-lesion synthesis by HIV-1 reverse transcriptase. *ChemBioChem* **6**, 1970-1973 (2005).

136 Küpfer, P. A. & Leumann, C. J. The chemical stability of abasic RNA compared to abasic DNA. *Nucleic Acids Res* **35**, 58-68 (2007).

137 Lin, S. *et al*. Mettl1/Wdr4-mediated m^7G tRNA methylome is required for normal mRNA translation and embryonic stem cell self-renewal and differentiation. *Mol Cell* **71**, 244-255 (2018).

138 Hudson, B. H. & Zaher, H. S. O^6-methylguanosine leads to position-dependent effects on ribosome speed and fidelity. *RNA* **21**, 1648-1659 (2015).

140 Cartlidge, R. A. *et al*. The tRNA methylase METTL1 is phosphorylated and inactivated by PKB and RSK in vitro and in cells. *EMBO J* **24**, 1696-1705 (2005).

141 Li, S. & Mason, C. The pivotal regulatory landscape of RNA modifications. *Annu Rev Genomics Hum Genet* **15**, 127-150 (2014).

142 Desrosiers, R. *et al*. Identification of methylated nucleosides in messenger RNA from Novikoff Hepatoma Cells. *Proc Nat Acad Sci* **71**, 3971-3975 (1974).

143 Honda, S. *et al*. Four-leaf clover qRT-PCR: a convenient method for selective quantification of mature tRNA. *RNA Biol* **12**, 501-508 (2015).

144 Meng, J. *et al*. Exome-based analysis for RNA epigenome sequencing data. *Bioinformatics* **12**, 1565-1567 (2013).

145 Cui, X. *et al*. A novel algorithm for calling mRNA m6A peaks by modeling biological variances in MeRIP-seq data. *Bioinformatics* **32**, i378–i385 (2016).

146 Okamoto, H. *et al*. Substrate tRNA recognition mechanism of tRNA (m^7G_{46}) methyltransferase from Aquifex aeolicus. *J Biol Chem* **279**, 49151-49159 (2004).

147 Alexandrov, A. *et al*. Rapid tRNA decay can result from lack of nonessential modifications. *Mol Cell* **21,** 87-96 (2006).

148 Barbieri, I. *et al*. Promoterbound METTL3 maintains myeloid leukaemia by m^6A-dependent translation control. *Nature* **552**, 126-131 (2017).

149 Okamoto, M. *et al*. tRNA modifying enzymes, NSUN2 and METTL1, determine sensitivity to 5-fluorouracil in HeLa cells. *PLoS Genet* **10**, e1004639 (2014).

150 Trimouille, A. *et al*. Further delineation of the phenotype caused by biallelic variants in the WDR4 gene. *Clin Genet* **93**, 374-377 (2018).

151 Pereira, P. L. *et al*. A new mouse model for the trisomy of the Abcg1-U2af1 region reveals the complexity of the combinatorial genetic code of down syndrome. *Hum Mol Genet* **18**, 4756-4769 (2009).

152 Mamane, Y. *et al*. eIF4E--from translation to transformation. *Oncogene* **23**, 3172-3179 (2004).

153 Ray, P. S. *et al*. Two internal ribosome entry sites mediate the translation of p53 isoforms. *EM-BO Rep* **7**, 404-410 (2006).

154 Mazumber, B. *et al*. Translational control by the 3'-UTR: the ends specify the means. *Trends Biochem Sci* **28**, 91-98 (2003).

155 Wang, X. *et al*. N^6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-120 (2014).

156 Gandin, V. *et al*. Polysome fractionation and analysis of mammalian translatomes on a genome-wide scale. *J Vis Exp* **87**, 1-10 (2014).

157 Hafner, M. *et al*. PAR-Clip: a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J Vis Exp* **41**, e2034 (2010).

158 Jia G, *et al*. Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett* **582**, 3313-3319 (2008).

159. Li, Z. *et al*. FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N(6)-Methyladenosine RNA Demethylase. *Cancer Cell* **31**, 127-141 (2017).

160. Su, R. *et al*. R-2HG Exhibits Anti-tumor Activity by Targeting FTO/m(6)A/MYC/CEBPA Signaling. *Cell* **172**, 90-105 (2018).

161. Chen, W. *et al*. Crystal structure of the RNA demethylase ALKBH5 from zebrafish. *FEBS Lett* **588**, 892-898 (2014).

162. Zhang, S. *et al*. m(6)A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program. *Cancer Cell* **31**, 591-606 (2017).

163 Duan, H. C. *et al*. ALKBH10B Is an RNA N(6)-Methyladenosine Demethylase Affecting Arabidopsis Floral Transition. *Plant Cell* **29**, 2995-3011 (2017).

164 Yi, C. & Pan, T. Cellular dynamics of RNA modification. *Acc Chem Res* **44**, 1380-1388 (2011).

165 Boccaletto, P. *et al*. *Nucleic Acids Res* **46**, D303-D307 (2018).

166 Yue, Y. *et al*. RNA N^6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev* **29**, 1343-1355 (2015).

167 Fu, Y. *et al*. *Angew Chem Int Ed Engl* **49**, 8885-8888 (2010).

168 Liu, F. *et al*. ALKBH1-Mediated tRNA Demethylation Regulates Translation. *Cell* **167**, 816-828 (2016).

169 Solberg, A. *et al*. *J Mol Cell Biol* **5**, 194-203 (2013).

170 Anderson, J. *et al*. The essential Gcd10p–Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes Dev* **12**, 3650-3662 (1998).

171 Wang, G. *et al*. The atomic resolution structure of human AlkB homolog 7 (ALKBH7), a key protein for programmed necrosis and fat metabolism. *J Biol Chem* **289**, 27924-27936 (2014).

172 Dai, Q. *et al*. Selective Enzymatic Demethylation of N^2,N^2 -Dimethylguanosine in RNA and Its Application in High-Throughput tRNA Sequencing. *Angew Chem Int Ed* **56**, 5017-5020 (2017).

173 Dewe, J. M. *et al*. TRMT1-Catalyzed tRNA Modifications Are Required for Redox Homeostasis to Ensure Proper Cellular Proliferation and Oxidative Stress Survival. *Mol Cell Biol* **37**, e00214-17 (2017).

174 Suzuki, T. & Suzuki, T. A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs. *Nucleic Acids Res* **42**, 7346-7357 (2014).

175 Clark, W. C. *et al*. tRNA base methylation identification and quantification *via* high-throughput sequencing. *RNA* **22**, 1771-1784 (2016).

176 Liu, J. & Strâby, K. B. The human tRNA(m(2)(2)G(26))dimethyltransferase: functional expression and characterization of a cloned hTRM1 gene. *Nucleic Acids Res* **28**, 3445-3451 (2000).

177 Ellis, S. R. *et al*. Isolation and characterization of the TRM1 locus, a gene essential for the N^2,N^2-dimethylguanosine modification of both mitochondrial and cytoplasmic tRNA in Saccharomyces cerevisiae. *J Biol Chem* **261**, 9703-9709 (1986).

178 Suzuki, T. *et al*. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. *Annu Rev Genet* **45**, 299-329 (2011).

179 Richter, U. *et al*. RNA modification landscape of the human mitochondrial tRNALys regulates protein synthesis. *Nat Commun* **9**, 3966 (2018).

180 Mikami, S. *et al*. An efficient mammalian cell-free translation system supplemented with translation factors. *Protein Expr Purif* **46**, 348-357 (2006).

181 Haute, V. L. *et al*. Mitochondrial transcript maturation and its disorders. *J Inherit Metab Dis* **38**, 655-680 (2015).

182 Dhir, A. *et al*. Mitochondrial double-stranded RNA triggers antiviral signalling in humans. *Nature* **560**, 238-242 (2018).

183 Kim, Y. *et al*. PKR Senses Nuclear and Mitochondrial Signals by Interacting with Endogenous Double-Stranded RNAs. *Mol Cell* **71**, 1051-1063 (2018).

184 Rackham, O. *et al*. Hierarchical RNA Processing Is Required for Mitochondrial Ribosome Assembly. *Cell Rep* **16**, 1874-1890 (2016).

185 Zhang, G. *et al*. Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods.* **12**, 835-837 (2015).

186 Xu, L. *et al*. Three distinct 3-methylcytidine (m3C) methyltransferases modify tRNA and mRNA in mice and humans. *J Biol Chem*. **292**, 14695-14703 (2017).