THE UNIVERSITY OF CHICAGO

THE EXUBERANT VINE OF EPITRANSCRIPTOME: RNA METHYLATION AS A NEW EPIGENETIC REGULATORY MECHANISM

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

BOXUAN ZHAO

CHICAGO, ILLINOIS

AUGUST 2017

Table of Contents

List of Figures v

Acknowledgement viii

Abstract xi

List of Publications xii

Chapter 1

Introduction: RNA Modifications and Epitranscriptomics 1

1.1 Genetics and epigenetics: beyond the primary sequence 1

1.2 Epigenetic regulation of chromatin structure: histone and DNA modifications 2

1.3 Emergence of RNA epigenetics: chemical modifications on RNA 4

1.4 N^6-methyladenosine (m^6A): the protagonist of epitranscriptomics 7

1.5 Scope of this dissertation 10

Chapter 2

The Molecular Functions of m^6A 11

2.1 Introduction: the spectrum of m^6A readers 11

2.2 Results: diverse functional aspects of m^6A readers 12

2.2.1 Specialized function of individual m^6A reader: YTHDF1 promotes translation 12

2.2.1.1 YTHDF1 interacts with translation machinery and improves ribosome loading of its targets 12

2.2.1.2 YTHDF1 binding is sufficient to promote mRNA translation 17

2.2.1.3 YTHDF1 acting in dual roles to modulate the translation dynamics of m^6A-modified mRNA 19

2.2.1.4 YTHDF1 interacts with initiation factors to promote translation.....24
2.2.2 Co-functioning of m^{6}A readers: YTHDF1, YTHDF2, and YTHDF3.....28
2.2.2.1 YTHDF1 and YTHDF2 co-regulate the translation and decay of common targets.....28
2.2.2.2 YTHDF3 impacts the binding specificity of YTHDF1 and YTHDF2.....32
2.2.3 Switch of m^{6}A reader functions under certain circumstances.....35
2.3 Conclusion and Discussion: m^{6}A marks fast turnover of mRNA.....35
2.4 Materials and Methods.....38

Chapter 3

The Cellular Functions of m^{6}A.....52

3.1 Introduction: m^{6}A marks functional cohorts of transcripts in cells.....52
3.2 Results: m^{6}A regulates zebrafish reproduction and development process.....54
3.2.1 Transcriptome and methylome dynamics during zebrafish embryogenesis.....54
3.2.2 Ythdf2 facilitates the rapid clearance of maternal transcripts.....58
3.2.3 Co-regulation of mRNA degradation by Ythdf2 and miR430 pathways.....65
3.2.4 Ythdf2 affects zebrafish spermatogenesis.....69
3.3 Conclusion and Discussion: m^{6}A impacts cellular state transition.....72
3.4 Materials and Methods.....74

Chapter 4

The Functions of m^{6}A in Human Diseases.....83

4.1 Introduction: m^{6}A regulators associate with disease-related RNAs.....83
4.2 Results: Regulatory roles of m^{6}A in non-infectious and infectious diseases.....84
4.2.1 m^{6}A in cancer: demethylase facilitates cancer cell proliferation.....84

4.2.1.1 Locate the cancer-related m^6A regulator: ALKBH5 associates with high self-renewal of GSCs 85
4.2.1.2 Search for the key target of ALKBH5: FOXM1 is subjected to demethylation and gene expression control by ALKBH5 89
4.2.1.3 Reveal the mechanism of ALKBH5 controlling FOXM1 abundance: HuR and FOXM1-AS 96
4.2.2 m^6A in viral infection: m^6A marks diverse types of viral transcripts 99
4.2.2.1 m^6A in HIV 99
4.2.2.2 m^6A in Zika virus 109
4.2.2.3 m^6A in respiratory syncytial virus and influenza virus 117
4.3 Conclusion and Discussion: m^6A as a key regulator in disease development 124
4.4 Materials and Methods 126

Chapter 5

Summary and Perspectives 142

5.1 The emergence of m^6A: an advanced precision control system of gene regulation 142
5.2 The network of epigenetics: crosstalk between m^6A and other epigenetic mechanisms 145
5.3 The family of RNA modifications: a spectrum of functional decorations 147
5.4 The future of epitranscriptomics study 149
5.4.1 Toward a finer portrait of the multifaceted m^6A modification 149
5.4.2 Uncover the involvement of m^6A in epigenetic inheritance 150
5.4.3 Seek other unknown members of the family of functional mRNA modifications 152

List of References 154

List of Figures

Figure 1.1 Spectrum of diverse RNA chemical modifications 6

Figure 1.2 The writer, eraser and reader proteins of m^6A.....8

Figure 2.1 YTHDF1 interacts with translation machinery.....13

Figure 2.2 Transcriptome-wide identification of YTHDF1 mRNA targets 15

Figure 2.3 Knockdown of YTHDF1 leads to reduced translation of its mRNA targets 16

Figure 2.4 The N-terminal domain of YTHDF1 promotes protein production in a tethering assay.....18

Figure 2.5 YTHDF1 alters the translation dynamics of m^6A-modified mRNA 21

Figure 2.6 YTHDF1 enhances the translation of m^6A-modified mRNAs by two mechanisms 23

Figure 2.7 YTHDF1 associates with translation machinery and protein markers 25

Figure 2.8 Translation promotion mechanism of YTHDF1 revealed by IRES tethering reporter assay.....27

Figure 2.9 Translation efficiency and stability of the common targets of YTHDF1 and YTHDF2 is affected by both m^6A readers.....30

Figure 2.10 Temporal order of YTHDF1 and YTHDF2 binding to common targets.....31

Figure 2.11 YTHDF proteins form an interconnected network in the cytoplasm.....34

Figure 2.12 m^6A-dependent regulatory pathways accelerate mRNA metabolism.....37

Figure 3.1 Characterization of gene expression and m^6A modification change in the early embryonic transcriptome of zebrafish 56

Figure 3.2 The deficiency of m^6A-binding protein Ythdf2 in zebrafish embryo led to a developmental delay 60

Figure 3.3 Identification and characterization of stringent targets of Ythdf2 63
Figure 3.4 Ythdf2 is required for m^{6}A-dependent RNA decay.....65
Figure 3.5 Target overlap of Ythdf2 and miR-430 regulation pathways 68
Figure 3.6 The paternal effect of Ythdf2 affecting sperm function 71
Figure 3.7 Mechanism of m^{6}A impacting cellular state transition events.....73
Figure 4.1 ALKBH5 predicts poor survival of GBM patients and is associated with cancer stem cell niches 86
Figure 4.2 Knockdown of ALKBH5 impairs GSC proliferation and tumorigenicity.....88
Figure 4.3 Characterization of gene expression and m^{6}A modification changes in the ALKBH5-deficient GSCs 91
Figure 4.4 ALKBH5 modulates FOXM1 expression by demethylating FOXM1 nascent transcripts 94
Figure 4.5 HuR and FOXM1-AS function in the regulation of FOXM1 nascent transcripts by ALKBH5 98
Figure 4.6 HIV-1 RNA contains m^{6}A modifications and affects host methylome 101
Figure 4.7 YTHDF proteins bind HIV-1 RNA and negatively regulate post-entry HIV-1 infection 104
Figure 4.8 YTHDF proteins inhibit HIV-1 infection by sequestering viral RNA and blocking viral reverse transcription 107
Figure 4.9 m^{6}A writers and erasers affect HIV-1 protein translation and release.....109
Figure 4.10 ZIKV RNA contains extensive modifications, including m^{6}A regulated by host enzymes.....112
Figure 4.11 m^{6}A regulators modulate ZIKV life cycle 114

Figure 4.12 ZIKV infection influences RNA methylation of host transcripts 116

Figure 4.13 RSV RNA contains m^6A and induces the methylome change in host transcripts 118

Figure 4.14 RSV protein expression and replication are impacted by host m^6A regulators...120

Figure 4.15 The genomes of influenza are methylated and its infection is affected by m^6A regulators.....121

Figure 4.16 YTHDF proteins relocalize into the nucleus after influenza infection 123

Figure 5.1 m^6A synchronizes mRNA processing in response to various internal and external stimuli 144

Acknowledgement

First and foremost, I want to express my sincere gratitude to my doctorate advisor Professor Chuan He. This dissertation would not be possible without his insightful mentorship and unrestricted supports. We first came to know each other back during my undergraduate years at Peking University. After being introduced to the field of chemical biology by my undergraduate advisor Professor Peng Chen, who was one of the first graduate students mentored by Chuan, I was very much inspired by the excellent work done in Chuan's lab and envisioned working with him. I was thrilled to be admitted to the PhD program at the University of Chicago and immediately joined the lab since the start of my graduate study. Before I became focused on the content of this dissertation, Chuan gave me opportunities to fully explore different directions in the lab and to collaborate with several senior group members on distinct projects ranging from protein engineering, DNA modification sequencing, and RNA epigenetics. I benefitted tremendously from these early experiences as they familiarized me with the lab and the research fields, and equipped me with necessary skills to fully explore other research fronts. I eventually put my focus on the newly emerged field of RNA epigenetics, and Chuan gave me adequate patience and ideas to test with. His enthusiasm for science and optimistic attitude towards experimental results have very positively influenced my research style. His ability to eloquently communicate research and efficiently seek partnership have yielded many productive collaborations with other labs. I especially appreciated his generous supports for my fellowship applications and my presentations at several academic conferences throughout my graduate career, which were crucial for me to become an independent and communicative scientist. Shaped by Chuan's passion and warmth, our group maintained a very cooperative and friendly atmosphere, which was very beneficial for carrying out efficient scientific studies.

I felt extremely fortunate to work with this amazing group for the past five years of my life. Dr. Quanjiang Ji collaborated with me during my first year of Ph.D. and familiarized me with the lab. Dr. Chuanxiao Song introduced me to the study of DNA modifications and trained me with nucleic acids chemistry and sequencing techniques. Dr. Ye Fu taught me the basic skills in RNA biology and assisted me to get started with my independent project. Dr. Xingyu Lu provided me with constructive suggestions for my study in DNA sequencing technologies. Dr. Dan Dominissini helped me with numerous explorations in RNA epigenetic projects. I sincerely thank Zhike Lu for his irreplaceable work of bioinformatics analysis in most of my studies. I am especially grateful to Dr. Xiao Wang, as the collaboration between us have resulted in very productive teamwork and great studies being published. I also thank Hailing Shi, Tong Wu, and Chang Liu who have closely worked with me. I am glad to have Ian Roundtree to be the same year with me and making continuous progress toward our Ph.D. degrees together. I sincerely thank all the other group members for their help and discussions on the daily basis.

I would also like to thank our research collaborators. Professor Robert Ho, Dr. Alana Beadell, and Adam Kuuspalu in the Department of Organismal Biology provided crucial help on zebrafish studies. Professor Suyun Huang and Sicong Zhang from the University of Texas MD Anderson Cancer Center kindly collaborated with us in the glioblastoma study. For our collaborators at the Ohio State University, Professor Li Wu and Nagaraja Tirumuru provided essential expertise in the HIV study, while Professor Jianrong Li, Miaoge Xue, and Yu Yang cooperated with us in the RSV and influenza study. Professor Tariq Rana and Gianluigi Lichinchi at the University of California San Diego worked with us on the ZIKV project. Professor Jianjun Chen at the University of Cincinnati has been our long-time collaborator in several cell biology and cancer biology projects. Professor David Allis and Dr. Bryce Carey from the Rockefeller

University along with Professor Craig Thompson and Dr. Lydia Finley from Memorial Sloan Kettering Cancer Center worked with us in the study of α-KG. Professor Edith Heard and Dr. Ines Pinheiro at the Curie Institute collaborated with our lab in the study of X chromosome inactivation.

I am wholeheartedly grateful to the help of Professor Viresh Rawal, Professor Tao Pan, and Professor Robert Ho, for taking the time to be my dissertation committee members. I have benefited a lot from their help over the course of my Ph.D. study. It is their guidance that facilitated the achievement of my dissertation.

Finally, the supports from my friends and family are essential for me to go through the highs and lows in my graduate life and finally reach this stage. I must thank my parents who have always cared for me and supported me throughout my years as a student. The early exposure to science they provided and the constant positive influence they had on me have been the determining factors of my achievements in life. Their continuous, unconditional love has always inspired me and motivated me to try my best during my path to the Ph.D.

Abstract

Nucleic acids, one of the most essential biological macromolecules, store and transfer genetic information of all living organisms. They comprise of four canonical units: A, T, C, and G for DNA, and A, U, C, and G for RNA. Some of these units are chemically modified to record an extra dimension of information on top of the primary sequence. The existence of these modifications has been known for decades, yet the regulation and the functions of them remained largely elusive. My doctoral work is focused on the most prevalent chemical modification on human messenger RNA, N^6-methyladenosine (m^6A). Understanding the regulatory mechanism and cognate functions of m^6A in various biological events is crucial for the clarification of how cells encode epigenetic information into chemical modifications for global gene regulation. The characterization of m^6A metabolism may also provide valuable implications for the studies of complex biological processes and human diseases.

List of Publications Containing Work Presented in This Dissertation *

1. Sicong Zhang, **Boxuan Simen Zhao**, Aidong Zhou, Kangyu Lin, Yaohui Chen, Keping Xie, Oliver Bögler, Sadhan Majumder, Zhike Lu, Chuan He, and Suyun Huang. The m^6A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-Like Cells by Sustaining FOXM1 Expression and Cell Proliferation Programing, *Cancer Cell*, 31(4), 591-606 (2017).
2. **Boxuan Simen Zhao***, Xiao Wang*, Alana C. Beadell*, Zhike Lu, Hailing Shi, Robert K. Ho, and Chuan He, m^6A-dependent maternal mRNA clearance facilitates maternal-to-zygotic transition in zebrafish, *Nature*, 542 (7642), 475-478 (2017).
3. Lijia Ma, **Boxuan Simen Zhao**, Kai Chen, Amber Thomas, Jigyasa H. Tuteja, Xin He, Chuan He, and Kevin P. White, Evolution of transcript modification by N^6-methyladenosine in primates, *Genome Res.*, 27 (3), 385-392 (2017).
4. Hailing Shi, Xiao Wang, Zhike Lu, **Boxuan Simen Zhao**, Honghui Ma, Phillip Hsu, Chang Liu, and Chuan He. YTHDF3 facilitates translation and decay of the N^6-methyladenosine-modified RNA, *Cell Res.*, 27 (3), 315-328 (2017).
5. **Boxuan Simen Zhao**, Ian A. Roundtree, and Chuan He. Post-transcriptional gene regulation by the messenger RNA modification marks, *Nat. Rev. Mol. Cell Biol.*, 18, 31-42 (2017).
6. Huilin Huang*, Hengyou Weng*, Xi Qin*, **Boxuan Simen Zhao**, Lou Dore, Jennifer Strong, Rui Su, Kyle Ferchen, Chenying Li, Lei Dong, Xi Jiang, Chuan He, and Jianjun Chen. The N^6-Adenine Methyltransferase METTL14 Plays an Oncogenic Role in Acute Myeloid Leukemia, *Blood*, 128 (22), 1536-1536 (2016).

* The following chapters of this dissertation contain sections and figures adapted from listed publications with modifications. Chapter 1: publications 5 and 10; Chapter 2: publications 4, 5, and 11; Chapter 3: publications 2 and 5; Chapter 4: publications 1, 7, and 8; Chapter 5: publications 5 and 10.

7. Gianluigi Lichinchi*, **Boxuan Simen Zhao***, Yinga Wu, Zhike Lu, Yue Qin, Chuan He, and Tariq M. Rana. Dynamics of Human and Viral RNA Methylation during ZIKA Virus Infection, *Cell Host Microbe*, 20(5), 666-673 (2016).
8. Nagaraja Tirumuru*, **Boxuan Simen Zhao***, Zhike Lu, Chuan He, and Li Wu, YTHDF Proteins Bind to N^6-methyladenosines of HIV-1 RNA and Regulate Viral Infection, *Elife.*, 5:e15528 (2016).
9. Osama Zahid, **Boxuan Simen Zhao**, Chuan He, and Adam Hall, Quantifying mammalian genomic DNA hydroxymethylcytosine content using solid-state nanopores, *Sci. Rep.*, 6:29565 (2016).
10. Kai Chen, **Boxuan Simen Zhao**, and Chuan He, Nucleic Acid Modifications in Regulation of Gene Expression, *Cell Chem. Biol.*, 23(1), 74-85 (2016).
11. Xiao Wang*, **Boxuan Simen Zhao***, Ian A. Roundtree, Zhike Lu, Dali Han, Honghui Ma, Xiaocheng Weng, Kai Chen, Hailing Shi, and Chuan He. N^6-methyladenosine Modulates Messenger RNA Translation Efficiency, *Cell*, 161(6), 1388–1399 (2015).
12. **Boxuan Simen Zhao** and Chuan He, Fate by RNA Methylation: m^6A Steers Stem Cell Pluripotency, *Genome Biol.*, 16, 43 (2015).
13. Xingyu Lu, **Boxuan Simen Zhao**, and Chuan He, TET Family Proteins: Oxidation Activity, Interacting Molecules, and Functions in Diseases, *Chem. Rev.*, 115 (6), 2225–2239 (2015).
14. Xingyu Lu*, Dali Han*, **Boxuan Simen Zhao***, Chun-Xiao Song, Li-Sheng Zhang, Louis C Doré, and Chuan He, Base-Resolution Maps of 5-Formylcytosine and 5-Carboxylcytosine Reveal Genome-Wide DNA Demethylation Dynamics, *Cell Res.*, 25, 386–389 (2015).
15. **Boxuan Simen Zhao**, and Chuan He, Pseudouridine in a New Era of RNA Modifications, *Cell Res.*, 25, 153–154 (2015).

16. Gary C. Hon, Chun-Xiao Song, Tingting Du, Fulai Jin, Siddarth Selvaraj, Ah Young Lee, Chia-an Yen, Zhen Ye, Shi-Qing Mao, Bang-An Wang, Samantha Kuan, Lee E. Edsall, **Boxuan Simen Zhao**, Guoliang Xu, Chuan He, and Bing Ren, 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming During Differentiation. *Mol. Cell*, 56(2), 286-297 (2014).

(*co-authors contributed equally)

CHAPTER 1

Introduction: RNA Modifications and Epitranscriptomics

1.1 Genetics and epigenetics: beyond the primary sequence

Nucleic acids, one of the essential biological macromolecules, store and transfer genetic information of all living organisms. They carry genes, the molecular unit of heredity, within their primary sequence made of four canonical bases (A/T/C/G for DNA and A/U/C/G for RNA). The field of genetics studies nucleic acids and associating proteins, linking genes and genetic variations to the changes in heredity and phenotypes (heritable traits). However, during the course of genetics research, certain stable phenotype changes were observed without changes in the primary nucleic acid sequence of genes, which brought forth the birth of epigenetics to study these non-genome sequence based heritable factors.

The concept of epigenetics originally dated back to 1942, as Waddington first coined the term ‘epigenetics’ and described ‘epigenetic landscape’ in his work as the cellular phenotypic alterations occur during the development of multicellular organisms[1,2]. The term ‘epigenetics’ was later defined by Holliday 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]. Recently, a consensus definition of the epigenetic trait was proposed at the 2008 Cold Spring Harbor meeting, as "stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence"[4], which includes all alterations to the chromatin template that lead to various patterns of gene expression or silencing.

The study of the molecular mechanisms of epigenetics can be roughly categorized into three major areas: histone modifications, DNA modifications, and RNA regulatory factors (including RNA modifications). Covalent modifications on DNA and histones both exert major

impacts on the structure of chromatin they form, which in turn affects the expression of genes within the associated regions; whereas RNA regulatory factors are more diverse in types and control gene expression through distinct mechanisms.

1.2 Epigenetic regulation of chromatin structure: histone and DNA modifications

The eukaryotic genome is compacted into chromatin templates composed of repeating nucleosomal units containing DNA and histone proteins (H2A, H2B, H3, and H4)[5]. These core histone proteins have a globular domain and flexible histone tails that carry many post translational modifications (PTMs) termed histone modifications, which include acetylation, methylation, ubiquitination, phosphorylation, sumoylation, ribosylation, and citrullination[6]. These modifications are very dynamic and critical for chromatin structure and gene expression. Histone acetylation and methylation were among the first to be described and indicated with correlation to transcriptional changes[7,8]. These modifications occur at specific sites and are regulated by chromatin-associated enzymatic systems, including the ones that establish these modifications (writers) and the ones remove them (erasers). These dynamically controlled modifications impact the structure of histones and associating chromatins, which are crucial for the maintenance of specific gene transcription programs.

Covalent modifications on DNA is another way to control chromatin structures. 5-methylcytosine (5mC) with its oxidative derivatives and N^6-methyladenine (6mA) in DNA are the well-studied examples of these DNA modifications. The existence of cytosine methylation (5mC) in genomic DNA was first reported in 1951[9]. More than two decades later, the regulatory maintenance of the 5mC pattern across cell divisions was proposed[10,11], while the first DNA methyltransferase, Dnmt1, was identified later[12]. Subsequent studies connected the functional

outcomes of DNA methylation to the repression of gene expression[13]. It should be noted that the transcription regulation roles of DNA methylation typically synergize with various histone marks as the methyltransferases, demethylases, and readers of DNA methylation interact with various histone marks or histone modification enzymes.

The demethylation process on DNA was well-characterized prior to the discovery of cognate eraser enzymes, and two major forms of demethylation were described. The replication-dependent methylation dilution, i.e. passive demethylation, was first reported on the maternal genome during mammalian preimplantation development which arose from the absence of Dnmt1-dependent maintenance[14,15]. Haaf group and Walter group later each reported the rapid decrease of 5mC on the paternal genome in the zygote occurred hours after fertilization and before the first round of DNA replication, i. e. active demethylation, which strongly suggested the existence of active DNA demethylases[16,17]. The discovery of the existence of 5-hydroxymethylcytosine (5hmC) in the mammalian genome[18,19] and the identification of Tet proteins that utilize dioxygen to oxidize 5mC to 5hmC[19,20] marked the groundbreaking steps towards the elucidation of the enzyme system regulating the active demethylation of 5mC. The following studies further demonstrated the Tet enzymes can further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[21-23]. 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[22,24]. Additionally, the 5mC oxidation derivatives of 5hmC, 5fC, and 5caC may also be passively diluted to the unmethylated stage through cell division[25].

Yet a more direct approach for investigation of 5mC functions involves the study of the proteins that recognize 5mC and carry out subsequent actions, i.e. 5mC readers. The first 5mC

reader to be characterized was methyl-CpG binding protein complex MeCP1, which was identified by Bird group[26]. The subsequent studies eventually discovered four 5mC readers comprising the methyl-CpG binding domain (MBD) family, including MeCP2, MBD1, MBD2, and MBD4 (MBD3 in this family is not a 5mC reader)[27]. Among them, MeCP2, MBD1, and MBD2 have been shown to be involved in 5mC-dependent transcriptional repression[28]. An unrelated p120 catenin partner protein Kaiso was also found to be a specific 5mC reader and function as a methylation-dependent transcriptional repressor[29]. Distinct from these 5mC readers, a number of proteins preferentially bind to unmethylated cytosines and thus can be referred to as 5mC "anti-readers", such as some transcription factors[30], CTCF protein[31], and CXXC domain-containing proteins[32]. These anti-readers may directly or indirectly impact the transcription of specific genes based on the methylation states of their cognate binding sites. Finally, considering the putative regulatory functions of 5hmC, 5fC, and 5caC, the reader proteins of these oxidative intermediates may also exist to translate their presence into biological functions[33].

Another methylation modification, N^6-methyladenine (6mA), also exists in the eukaryotic genomic DNA[34-37]. Yet its function remained elusive until very recently, three groups reported the presence of 6mA in three different eukaryotes (*Chlamydomonas*, *C. elegans*, and *Drosophila*) independently, shedding light on the function of this methylation modification in eukaryotes[38-40]. The results suggested a potential gene activation role for 6mA and its involvement in nucleosome labeling and histone modifications.

1.3 Emergence of RNA epigenetics: chemical modifications on RNA

Compared with histone and DNA modifications, epigenetic regulation through RNA factors is more diverse and interconnected. Unlike genomic DNA, RNA has more complicated post-transcriptional processing: RNA alternative splicing significantly increases the complexity of

gene expression by alternatively joining exons and removing introns; RNA editing alters the nucleoside sequence of a specific transcript, which may or may not change protein coding regions or potential splicing sites to further diversify the transcriptome; and RNA chemical modifications, most of which do not affect nucleoside sequence, are much more diverse and functionally versatile, suggesting broader functional impacts[41,42]. Many functional non-coding RNAs (ncRNA), including miRNA, siRNA, piRNA, and lncRNA, also have distinct influences on gene expression through functioning in chromatin remodeling, transcription control, or direct regulation of mRNA stability and translation processes. These ncRNA species can also be subjected to various post transcriptional modifications, demonstrating the complex crosstalk between RNA epigenetic regulatory pathways.

Among the 109 identified RNA base modifications[43], many are known to exist in eukaryotes for decades; tRNA and rRNA modifications, as well as mRNA cap methylations, have also been studied extensively in the past, but it was only recently uncovered that an internal modification in mRNA can be dynamically controlled and reversible[41,44-47]. This modification, N^6-methyladenosine (m^6A), has triggered the discovery of a new realm of biological regulation at the post-transcriptional level. A series of studies following this breakthrough have identified many other types of chemical modifications in mRNA and other RNA species with potential functions through developing novel approaches to characterize these modifications in a transcriptome-wide manner (Figure 1.1). These re-discovered modifications include ones occurring on the ribose (2′-*O*-methylation (2′-OMe or N_m)), on adenosine (m^6A, N^1-methyladenosine (m^1A), N^6,N^6-dimethyladenosine ($m^6{}_2A$), 2′-*O*-methylated m^6A (m^6A_m)), on guanosine (N^1-methylguanosine (m^1G), N^7-methylguanosine (m^7G), N^2,N^2-dimethylguanosine ($m^2{}_2G$)), on cytosine (5-methylcytosine (m^5C), 5-hydroxymethylcytosine (hm^5C)), and through isomerization of uridine

(pseudouridine (Ψ)). These functional RNA modifications collectively form the basis of 'epitranscriptomics', i.e. the functionally relevant changes to the transcriptome without altering the primary RNA sequence.

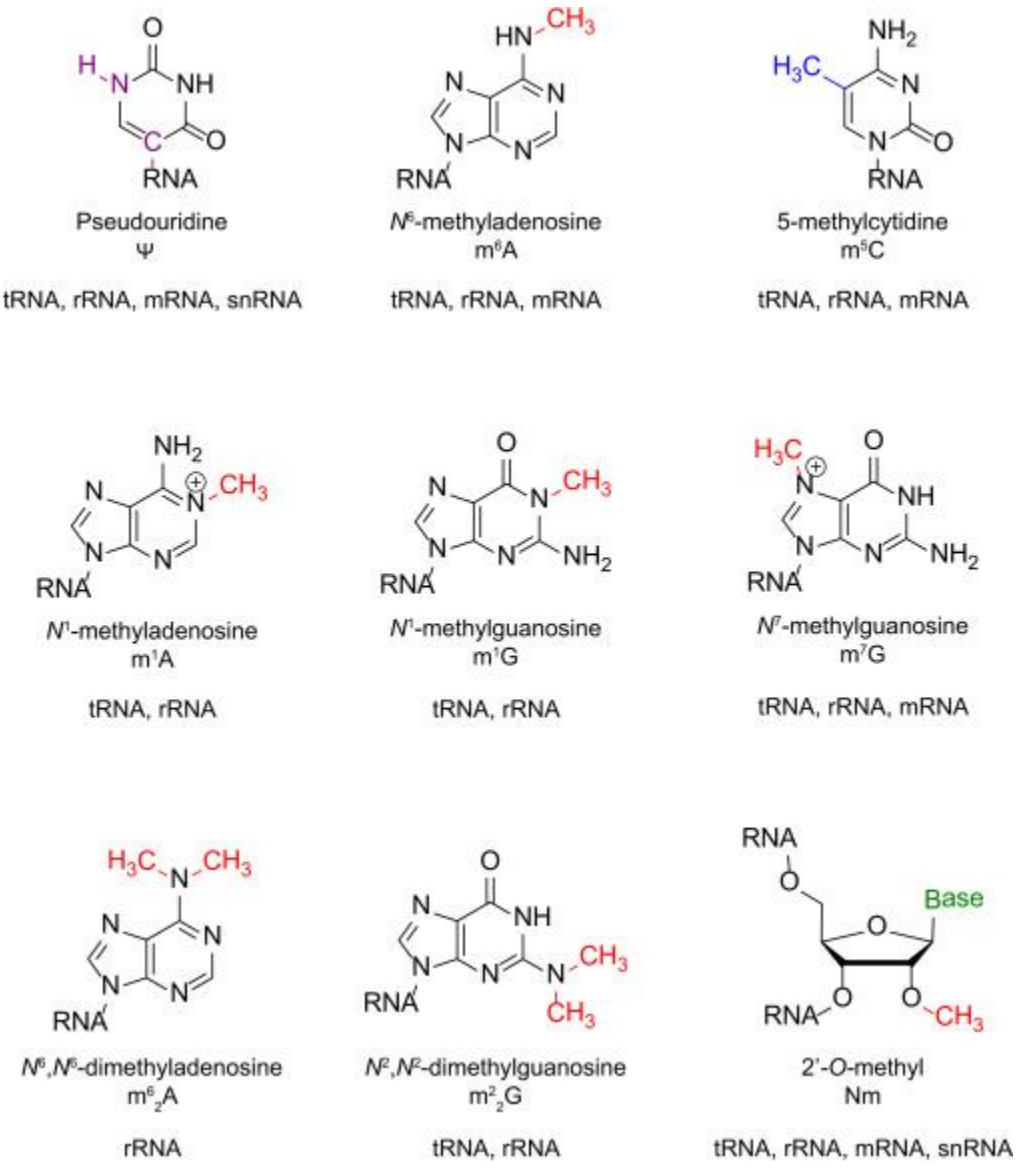


Figure 1.1 Spectrum of diverse RNA chemical modifications
Many chemical modifications have been identified in different RNA species, potentially carrying functional information and subject to epigenetic regulation.

1.4 N^6-methyladenosine (m^6A): the protagonist of epitranscriptomics

The existence of N^6-methyladenosine (m^6A) in RNA was discovered in tRNA[48] and rRNA[49] and was originally thought to be unique to them. In 1974, a series of studies uncovered the presence of m^6A in both eukaryotic[50] and viral mRNAs[51,52], demonstrating m^6A as the most prevalent internal modification in polyadenylated mRNAs and long non-coding RNAs (lncRNAs) in higher eukaryotes[53]. It was revealed that every mRNA of human HeLa cell contains three to five m^6A within a consensus motif of G(m^6A)C (70%) or A(m^6A)C (30%)[54,55], but the methylation percentage at each site varies substantially[56,57]. However, owing to the low abundance of m^6A in mRNA and the lack of effective techniques, functional characterizations of m^6A have been largely absent over the past few decades. This hiatus was ended in 2011 with the discovery of the first m^6A demethylase[45], discovering that m^6A in mRNA is dynamic and reversible. This work was followed by the characterization of the transcriptome-wide distribution of m^6A independently reported by two groups by high-throughput sequencing, which demonstrated that m^6A is enriched around stop codons, in 3′ untranslated regions (3′UTRs) and within internal long exons; and that m^6A peaks are well conserved from mouse to human and some are shown to be dynamic under various stress conditions[58,59]. These discoveries rekindled the tremendous interests in the biological relevance of m^6A, and a large amount of m^6A work has been published and highlighted in recent years.

The deposition, removal, and recognition of m^6A are carried out by cognate factors termed m^6A writers, erasers, and readers, respectively[60] (Figure 1.2). Mammalian m^6A writers function as a protein complex with four identified components so far: methyltransferase-like 3 (METTL3), METTL14, Wilms tumor 1-associated protein (WTAP) and KIAA1429. Two m^6A erasers have been reported: fat mass and obesity-associated protein (FTO) and alkB homolog 5 (ALKBH5).

The function of m^{6}A is mediated partly by reader proteins, which have been identified in members of the YT521-B homology (YTH) domain-containing protein and the heterogeneous nuclear ribonucleoprotein (HNRNP) protein families. Each class of regulators was discovered and studied over the entire course of m^{6}A research as briefly summarized below.

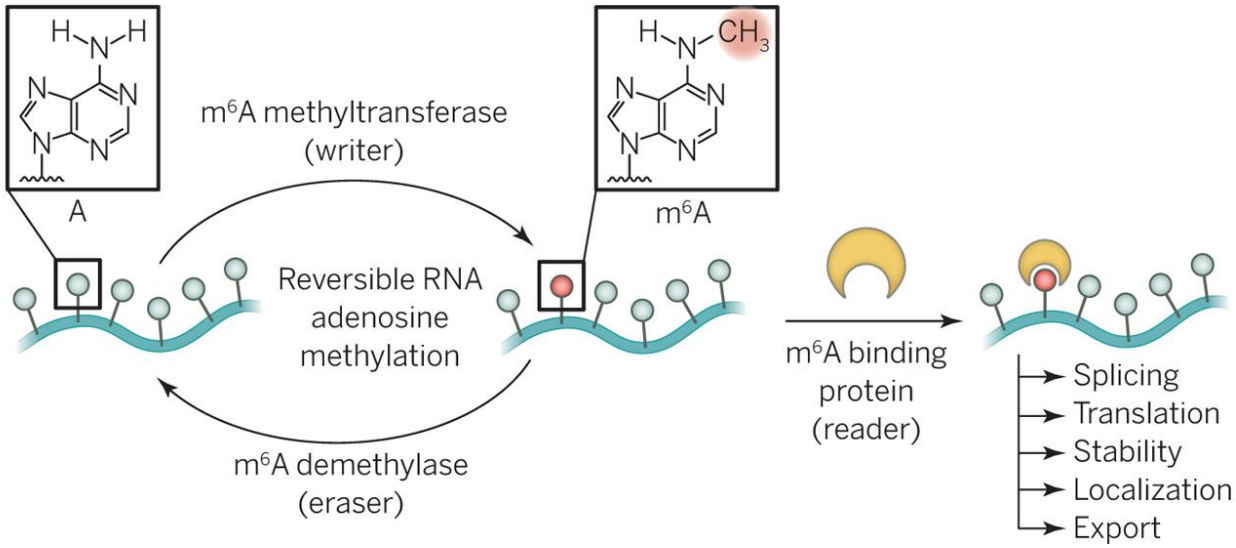


Figure 1.2 The writer, eraser and reader proteins of m^{6}A
m^{6}A is installed by m^{6}A methyltransferase complex (writer), removed by m^{6}A demethylases (eraser), and recognized by m^{6}A-binding proteins (reader) to exert biological functions.

The search for the writer protein of m^{6}A, mRNA m^{6}A methyltransferase (m^{6}A-MTase), dated back to 1978 when an m^{6}A-MTase was partially purified from HeLa cells but was shown to be only responsible for methylation in the cap structure instead of internal ones[61]. Later Rottman group developed an in vitro methylation assay and detected the internal methyltransferase activity of HeLa cell nuclear extracts[62]. Subsequent efforts further led to the discovery of the multicomponent nature of m^{6}A-MTase[63-65]. Due to the complexity of the purification of this enzyme complex, m^{6}A-MTase was not fully characterized until 2014, when three groups independently reported the core components of m^{6}A-MTase, i.e. METTL3, METTL14, and WTAP[66-68]. Another component KIAA1429 was also discovered later[69]. In contrast, the study of the erasers for m^{6}A came much later. The first mRNA m^{6}A demethylase, FTO, was discovered by

He group in 2011[45], which was quickly followed by the identification of another demethylase, ALKBH5[46]. Further study of FTO also revealed two intermediates, N^6-hydroxymethyladenosine (hm^6A) and N^6-formyladenosine (f^6A), generated through the oxidation of m^6A by FTO[70]. Similar to oxidative intermediates of 5mC, these modifications are metastable under physiological conditions and may carry regulatory functions.

With the discoveries made in the characterization of enzymatic system regulating m^6A, came the studies of the reader proteins for m^6A that translate the existence of m^6A into biological functions. Currently, two modes of m^6A-dependent RNA recognition are known, direct or indirect m^6A-reading. The first direct reader for m^6A, YTHDF2, was discovered in 2013 by He group[71], which impacts the stability of m^6A-modified mRNA. Later, another direct m^6A reader, YTHDF1, was also identified to be responsible for the promotion of translation efficiency of m^6A-modified mRNA[72]. Alternatively, m^6A-reading can also be achieved indirectly through the m^6A-dependent RNA structure change that affects the accessibility of RNA binding proteins (m^6A-switch). Pan group reported the first indirect m^6A reader, HNRNPC, whose interaction with m^6A-modified RNA is much enhanced by the m^6A-induced RNA structural remodeling[73]. Additionally, similar to 5mC, certain anti-readers repelling m^6A-modified RNA may exist. For instance, Toll-like receptors TLR3, TLR7, and TLR8 recognize and remove unmodified RNA as invasive RNA species, while the incorporation of m^6A and other RNA modifications ablates this activity[74]. Also, RNA binding by HuR was discovered to be interfered by the presence of m^6A methylation[68]. Both observations can be potentially explained by an anti-reading mechanism of certain RNA binding proteins. Future study of new reader proteins and their associated recognition mechanisms is crucial for understanding the roles of m^6A in various biological events.

1.5 Scope of this dissertation

My dissertation will focus on the central question in the field of m^6A RNA epigenetics: **What are the biologically relevant functions of m^6A on RNA?** This question can be answered through studies at different levels and in various biological systems.

Chapter 2 discusses the functions of m^6A on the molecular level, mainly focusing on the study of m^6A readers and their effects on methylated target mRNAs.

Chapter 3 discusses the functions of m^6A on the cellular level, expanding the investigation to the dynamics of m^6A-regulated mRNA transcripts during cellular events, with zebrafish embryogenesis presented as the major example.

Chapter 4 discusses the functions of m^6A in different disease settings, highlighting the roles of m^6A played during the development of various infectious and non-infectious diseases.

Chapter 5 summarizes the current understanding of the broader impact of m^6A and other RNA modifications, also envisions the directions of future study of RNA epigenetics.

CHAPTER 2

The Molecular Functions of m^6A *

2.1 Introduction: the spectrum of m^6A readers

Considering the essential roles played by m^6A readers to recognize and exert influence on m^6A-modified RNAs, identification and functional characterization of m^6A readers are vital for the understanding of the molecular functions of m^6A. Our lab has reported the discovery of first two m^6A readers: YTHDF1, which binds m^6A-modified RNAs and recruits translation initiation factors, causing m^6A to positively correlate with translation efficiency[72]; and YTHDF2, which recognizes and shuttles m^6A-modified RNAs to degradation, making m^6A negatively correlated with mRNA stability[71]. Our current investigation has identified a few more m^6A readers with unique molecular functions: YTHDF3, which affects translation and decay of methylated mRNAs through forming complexes and cooperating with YTHDF1 and YTHDF2, and may function in the long-distance transport and delivery of mRNA in neurons[75]; YTHDC1, which resides in the nucleus and interacts with multiple nucleoplasm proteins, may impact the splicing and export of m^6A-modified RNAs; heterogeneous nuclear ribonucleoprotein A2/B1 (HNRNPA2B1), one of the most abundant shuttling proteins between nucleus and cytoplasm, may facilitate the packaging and transport of m^6A-modified RNA across nuclear pores. Thus, the influence of m^6A exists at every step of mRNA metabolism, manifested by the spectrum of m^6A readers with diverse functions.

* The author's contribution to the studies presented in this chapter: YTHDF1 study: The author contributed equally to this work with X.W., designed experiments with X.W. and C.H., performed experiments and analyzed data with X.W., and wrote the manuscript with X.W. and C.H. YTHDF3 study: The author assisted with performing experiments with H.S. and X.W., provided valuable discussion during experiments design and suggestions during the manuscript writing by H.S., X.W., and C.H.

2.2 Results: diverse functional aspects of m^{6}A readers

The existence of numerous m^{6}A readers co-regulating groups of m^{6}A-modified RNAs implies that, aside from the specialized function of each reader, the interaction and coordination among different readers also impact the final outcomes of m^{6}A regulation. Additionally, individual readers may change their direct function under different circumstances. These aspects are discussed below.

2.2.1 Specialized function of individual m^{6}A reader: YTHDF1 promotes translation

Our study of YTHDF1 is a typical example showcasing the process of investigating the specialized function of an m^{6}A reader. YTHDF1 is one of the five known members of YTH protein family, and its specificity towards m^{6}A-modified RNA probes has been demonstrated *in vitro* previously[71]. As an abundant cytoplasmic m^{6}A reader like YTHDF2, we tried to pinpoint YTHDF1's function to see if it is similar or distinct from its homolog.

2.2.1.1 YTHDF1 interacts with translation machinery and improves ribosome loading of its targets

In order to hypothesize the function of YTHDF1, we need to first identify the protein partners it associates with and the mRNA targets it binds to. We constructed a HeLa cell line stably expressing an epitope-tagged YTHDF1 (N-terminal Flag and HA tags in tandem) near the endogenous level. Using polysome profiling to dissect different ribosome fractions, we found an enrichment of YTHDF1 co-existing with ribosomal subunits in the 40S portion (Figure 2.1A). Under mild formaldehyde fixation, YTHDF1 was also observed in the 80S portion, but not 60S (Figure 2.1A). We next studied the composition of the YTHDF1-containing complex using tandem-affinity purification of the epitope-tagged YTHDF1 and protein mass spectrometry. A control sample stably expressing Flag-HA peptide without YTHDF1 was processed in parallel.

The result revealed translation as the main theme of the protein interactome of YTHDF1 (Figure 2.1B). Compared to the control group, 119 unique proteins were copurified with YTHDF1, 62 of which involve translation, including 27 out of 33 subunits of 40S, 23 out of 47 subunits of 60S, and 6 out of 13 subunits of translation initiation factor complex 3 (eIF3). These results strongly suggest that YTHDF1 may be involved in the regulation of translation process of its target mRNAs.

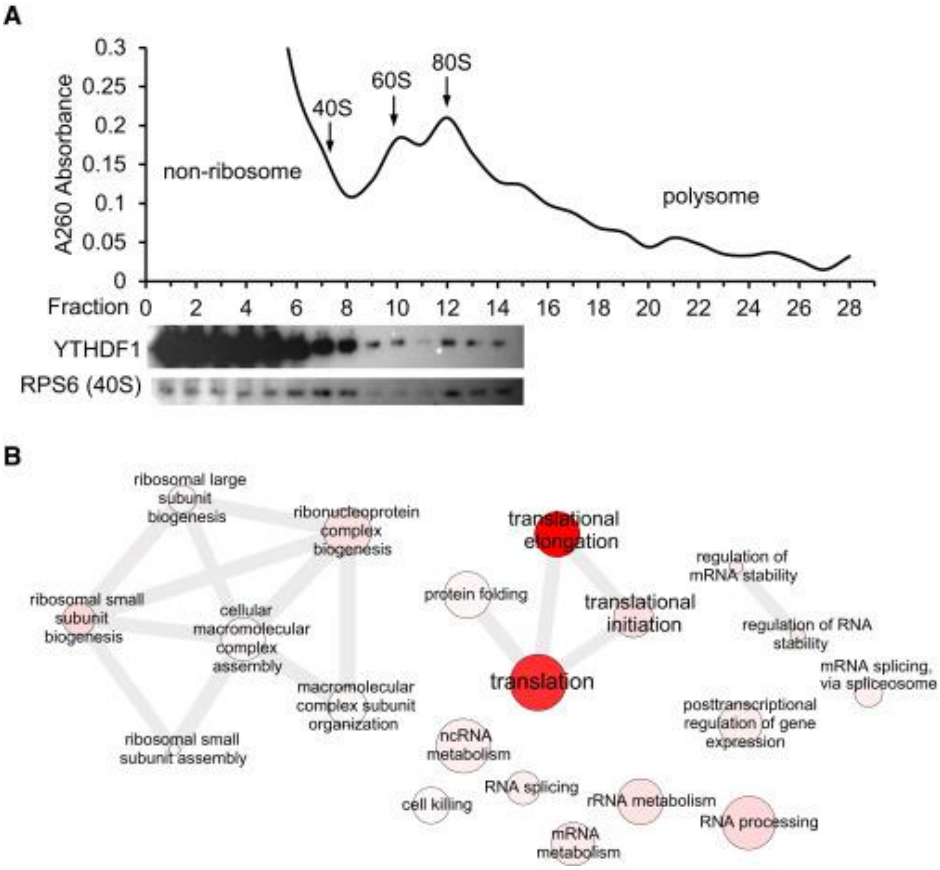


Figure 2.1 YTHDF1 interacts with translation machinery
(A) Western blotting of Flag-tagged YTHDF1 on each fraction of 10%–50% sucrose gradient showing that YTHDF1 associates with 40S and 80S ribosome. The fractions were grouped to non-ribosomal mRNPs, 40S–80S, and polysome. RPS6 is a protein subunit of 40S ribosome. The YTHDF1 HeLa stable line was treated with 0.5% formaldehyde for 7 min before harvesting. (B) Gene ontology analysis of the YTHDF1 protein interactome obtained by co-immunoprecipitation and mass spectrometry identification.

To characterize the mRNA targets and the binding sites of YTHDF1, we applied photoactivatable ribonucleoside cross-linking and immunoprecipitation (PAR-CLIP), which

utilizes a photoreactive nucleoside such as 4-thiouridine (4SU) to cross-link RNA with protein and produce a detectable T-to-C mutation[76]. This experiment identified 4,951 mRNA transcripts as potential targets of YTHDF1. A GRAC (R is G or A) motif, coinciding with the m^6A consensus motif, was repeatedly identified from two biological replicates of PAR-CLIP samples (Figure 2.2A), covering over 50% of all PAR-CLIP clusters in each sample. It should be noted that the cross-linking of YTHDF1 to target RNA requires an appropriately spaced 4SU residue near the binding site. The distribution of the distances between GGAC and T-to-C mutation sites shows that the optimum crosslinking positions are +2, +3, and −3 (setting A as position 0, Figure 2.2B), suggesting that YTHDF1 binds at exact GRAC sites instead of other motifs that co-exist in the proximity of m^6A sites. In addition, 62% of the PAR-CLIP peaks of YTHDF1 overlap with m^6A peaks as determined by antibody-based m^6A profiling (Figure 2.2C), all of which indicate that YTHDF1 recognizes and specifically binds to m^6A sites on its target mRNAs inside cells. In addition, we sequenced RNA obtained from the immuno-purified ribonucleoprotein complex of YTHDF1 as a complementary method (RIP-seq) to reveal YTHDF1-bound RNA. Among 1,714 transcripts identified with enrichment greater than 2-fold, 74% overlap with PAR-CLIP targets (Figure 2.2D). We eventually defined the 1261 shared genes (CLIP+IP) as high-confident targets of YTHDF1.

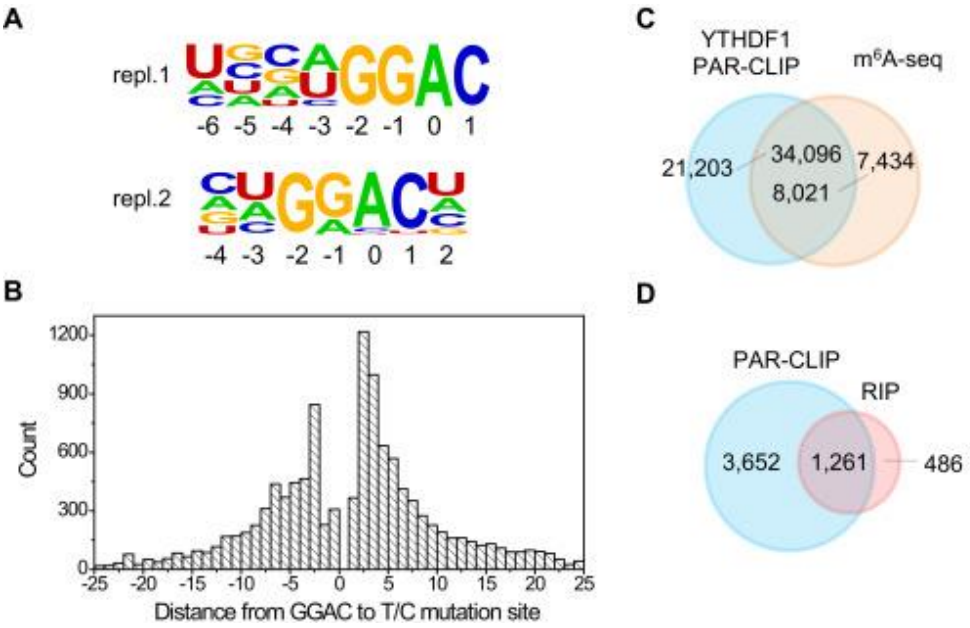


Figure 2.2 Transcriptome-wide identification of YTHDF1 mRNA targets
(A) YTHDF1-binding motifs identified by HOMER from PAR-CLIP peaks of two biological replicates. Motif length was restricted to 6–8 nucleotides. The motif with the lowest p value of replicate 1 (repl.1) was found in 53.8% of 24,753 sites ($p = 1 \times 10^{-703}$), and that of replicate 2 (repl.2) was found in 54.5% of 58,549 sites ($p = 1 \times 10^{-1093}$). (B) The distribution of the distance from GGAC to T-to-C mutation sites. The optimum crosslinking sites are at the −3, +2, and +3 positions. (C) Overlap of YTHDF1 PAR-CLIP peaks and m^{6}A-seq peaks in HeLa cells. (D) Overlap of target genes identified by PAR-CLIP and RIP-seq for YTHDF1.

With YTHDF1 targets identified, we next used ribosome profiling to assess ribosome density of each transcript mRNA with or without perturbation of YTHDF1. HeLa cells transfected with YTHDF1 or control siRNA were subjected to ribosome profiling and mRNA sequencing. Two biological replicates with the use of different YTHDF1 siRNA sequences were studied to avoid off-target effects. Transcripts presented (reads per kilobase per million reads [RPKM] > 1) in both ribosome profiling and mRNA sequencing samples were analyzed in parallel. These transcripts were then categorized as non-targets (absent from PAR-CLIP and RIP), PAR-CLIP only targets, and common targets of PAR-CLIP and RIP. A significant decrease in ribosome-bound mRNA reads for YTHDF1 targets was observed in the YTHDF1 knockdown samples compared to the controls ($p < 0.001$, Mann-Whitney U test) (Figure 2.3A). In contrast, the differences of

mRNA inputs between groups in the knockdown and the control samples were small (Figure 2.3B). Overall, the knockdown of YTHDF1 led to reduced translation efficiency of its target transcripts (Figure 2.3C). The extent of the ribosome occupancy reduction caused by YTHDF1 knockdown correlates with the number of YTHDF1-binding sites on the transcripts, as targets with more CLIP sites were more severely affected by YTHDF1 knockdown (Figure 2.3D). Collectively, these data suggest a promotional role of YTHDF1 in mRNA translation.

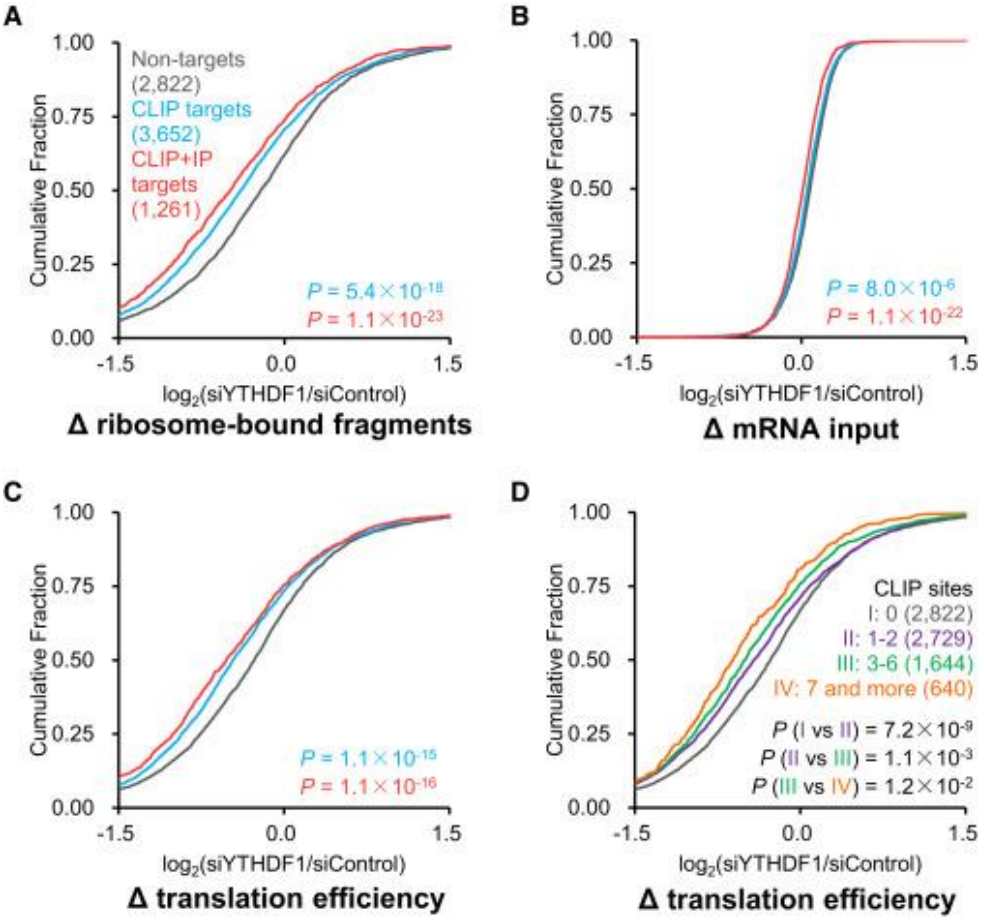


Figure 2.3 Knockdown of YTHDF1 leads to reduced translation of its mRNA targets
(A–C) Cumulative distribution log_2-fold changes of ribosome-bound fragments (A), mRNA input (B), and translation efficiency (C, the ratio of ribosome-bound fragments and mRNA input) between siYTHDF1 and siControl for non-targets (gray), PAR-CLIP-only targets (blue), and common targets of PAR-CLIP and RIP (red). *p* values were calculated using a two-sided Mann-Whitney test. (D) The mRNA lifetime log_2-fold changes were further grouped and analyzed on the basis of the number of CLIP sites on each transcript. The extent of translation reduction caused by YTHDF1 knockdown correlates with the number of YTHDF1-binding sites for mRNA targets of YTHDF1. *p* values were calculated using a Kruskal-Wallis test.

2.2.1.2 YTHDF1 binding is sufficient to promote mRNA translation

The observed increase in ribosome loading mediated by YTHDF1 on its target transcripts could be attributed to either increased translation or potentially hindered elongation. To probe the exact role of YTHDF1 on translation, we combined a switchable gene-expression system with a luciferase-based tethered reporter assay[77]. As the C-terminal YTH domain of all YTH family proteins is primarily engaged in m^6A binding, we studied the function of the N-terminal domain of YTHDF1 (N_YTHDF1). N_YTHDF1 was fused with λ peptide (N_YTHDF1_λ), which specifically and tightly binds F-luc-5BoxB (five Box B sequence inserted into the 3′ UTR of the luciferase reporter). A Tet-Off inducible promoter was installed onto the F-luc-5BoxB construct, which blocks its transcription in the presence of doxycycline (DOX), thus enabling the evaluation of protein-expression dynamics upon DOX removal (Figure 2.4A). We first compared luciferase expression with and without the N_YTHDF1 tether after an 8-hour induction. The result showed a ∼42% increase in translation with a slight decrease in the mRNA abundance when YTHDF1 was tethered to the luciferase mRNA (Figures 2.4B and C). The overall translation efficiency of the YTHDF1-tethered transcript had a ∼72% increase over the control, showing a major effect of YTHDF1 in promoting translation efficiency (Figure 2.4D). In order to reveal how YTHDF1 dynamically promotes translation, we treated cells with a 2-hour pulse induction of transcription (removing DOX for 2 hours before adding back) and monitored the dynamics of the reporter luciferase protein expression. Compared to the control group, cells with the N_YTHDF1-tethered reporter showed a slightly elevated expression rate of luciferase at the beginning but a noticeably increased translation rate after the pulse induction period (Figure 2.4E). This observation confirms that YTHDF1 binding directly elevates mRNA translation efficiency.

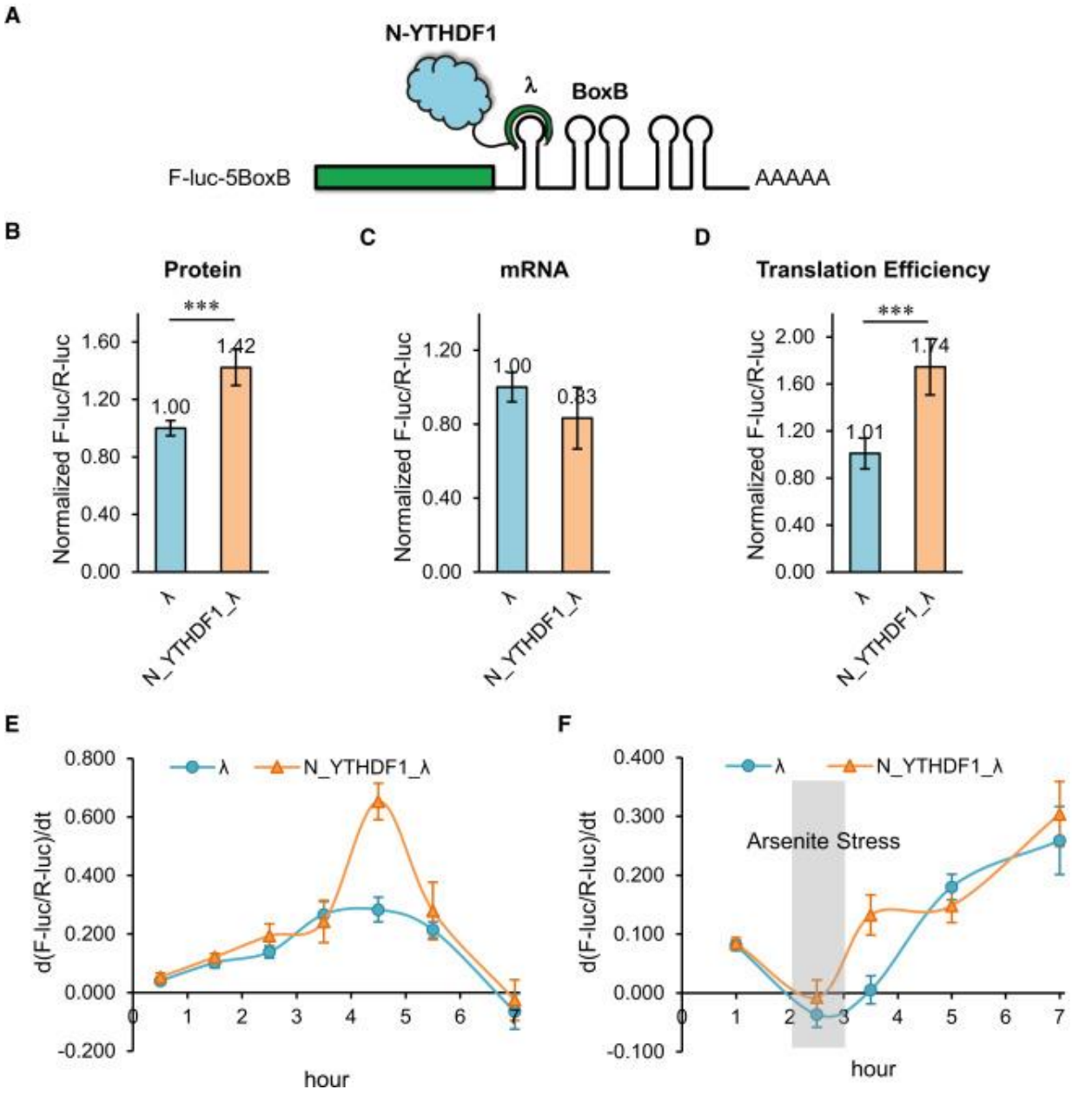


Figure 2.4 The N-terminal domain of YTHDF1 promotes protein production in a tethering assay

(A) Construct of the tethering reporter assay. The mRNA reporter consists of an inducible promoter, firefly luciferase as the coding region, and five Box B sequence at 3′ UTR (F-luc-5BoxB). The N-terminal domain of YTHDF1 (N_YTHDF1) was fused with λ peptide (N_YTHDF1_λ), which recognizes Box B RNA with a high affinity. R-luc lacks the inducible promoter and was used as an internal control to normalize the F-luc signal. (B) Under constant induction, the tethering of N_YTHDF1_λ to F-luc-5BoxB led to an on-average 42% increased translation in comparison with the control. The translation outcome was determined as a relative signal of F-luc divided by R-luc. Error bars, mean ± SD, $p = 5.9 \times 10^{-4}$ (two-sided Student's t test for paired samples), $n = 6$ (three biological replicates × two technical replicates). (C) Under constant induction, the mRNA abundance decreased slightly in the N_YTHDF1_λ-tethered group

(**Figure 2.4, continued**) compared with the control. The mRNA abundance was determined by qRT-PCR of F-luc and R-luc. Error bars, mean × SD, $p = 0.031$, $n = 6$. (D) The translation efficiency of the reporter mRNA increased by ∼72% in the N_YTHDF1_λ-tethered group compared with the control. The translation efficiency is defined as the quotient of reporter protein production (F-luc/R-luc) divided by mRNA abundance. Error bars, mean ± SD, $p = 3.2 \times 10^{-5}$, $n = 6$. (E) F-luc-5BoxB was induced with a pulse expression for 2 hr. The mRNA reporter showed higher translation when tethered with N_YTHDF1_λ compared with the control. y axis, d(F-luc/R-luc)/dt, indicating the changing rate of protein production. Error bars, mean ± SD, $n = 4$. (F) After a 2-hour pulse expression and a 1-hour arsenite (1 mM) stress treatment, translation of the reporter protein was largely diminished. The translation recovery was assessed after the stress was released. The result showed that the N_YTHDF1_λ-tethered group exhibited faster translation recovery than the control group. Error bars, mean ± SD, $n = 4$.

Additionally, we also investigated the effects of YTHDF1 on the luciferase reporter expression during the stress response. The tethered cells were given a 2-hour pulse induction followed by arsenite treatment for 1 hour, after which protein expression dynamics were monitored for another 5 hours. The result showed that although arsenite stress largely diminished protein expression, cells could gradually restore translation when the stress was lifted (Figure 2.4F). During the recovery stage, the N_YTHDF1-tethered transcript exhibited a faster restoration rate, indicating that YTHDF1 could facilitate the post stress response of cells by promoting translation recovery.

2.2.1.3 YTHDF1 acting in dual roles to modulate the translation dynamics of m^{6}A-modified mRNA

With YTHDF1's function of translation promotion confirmed, the next question to ask is whether this effect is m^{6}A dependent. By knocking down m^{6}A methyltransferase (METTL3) and performing ribosome profiling, we observed an overall decreased translation efficiency of YTHDF1 target transcripts with the reduction of m^{6}A [66], compared to non-targets (Figure 2.5A). The results confirm that YTHDF1 promotes translation efficiency in an m^{6}A-dependent manner.

The translation status of cytoplasmic mRNAs can be generally categorized into non-ribosome mRNPs (mRNA–protein particles), translatable mRNA pool (mRNPs associated with

translation factors but not being actively translated), and actively translating. To further uncover the effects of YTHDF1 on m^6A-modified RNA, sucrose gradient was utilized to fractionate cytoplasmic mRNAs into fractions with sedimentation coefficients of 20–35S that are exclusively non-ribosome mRNPs; 40–80S fractions containing the translatable mRNA pool and other large mRNPs; and the rest polysome fractions (Figure 2.5B). The knockdown of YTHDF1 does not change the m^6A/A ratio of total mRNA (Figure 2.5C). Hence, YTHDF1 does not alter the overall methylated mRNA level but affects the association of its target mRNA with the ribosome. If this is the case, upon the knockdown of YTHDF1, the amount of methylated mRNA should increase in mRNPs and decrease within the translating and translatable pools. Indeed, we observed a ~71% increase in the m^6A/A ratio of mRNA isolated from non-ribosome mRNPs, a ~44% decrease from 40–80S fractions, and a 25% decrease from polysome fractions (Figure 2.5B). Thus, YTHDF1 modulates the subcellular distribution and translation status of the m^6A-modified mRNA. It is worth noting that the role of YTHDF1 in mRNA stability cannot be completely excluded because a ~24% increase of m^6A/A ratio in mRNA was observed after 24-hour overexpression of YTHDF1 (Figure 2.5D). Because mRNA translation and degradation often closely correlate, an elevated level of YTHDF1 might retain mRNA in translation and slow down decay as a secondary effect. These results indicate that YTHDF1 is functionally distinct from YTHDF2 and may not be directly involved in mRNA decay.

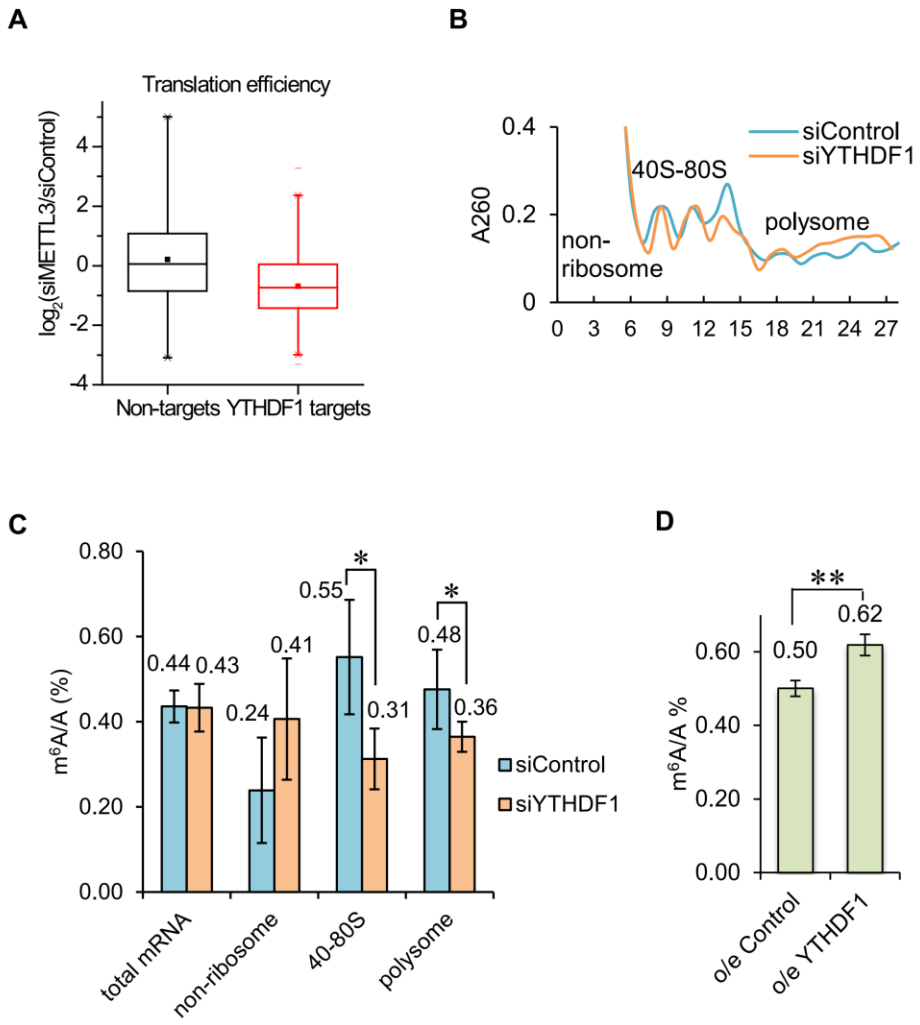


Figure 2.5 YTHDF1 alters the translation dynamics of m^{6}A-modified mRNA

(A) Knockdown of the m^{6}A methyltransferase (METTL3) reduced the translation efficiency of YTHDF1 target transcripts. Cumulative distribution *log*$_2$-fold changes of the translation efficiency between siMETTL3 and siControl for non-targets (black) and YTHDF1 RNA targets (red). $p \approx 0$, two-sided Mann-Whitney test. (B) Polysome profiles of YTHDF1 knockdown and control samples. (C) Quantification of the m^{6}A/A ratio of total mRNA, the non-ribosome portion, 40S–80S, and polysome determined by LC-MS/MS for the YTHDF1 knockdown samples compared to controls after 48 hours. *p* values were determined using a two-sided Student's t test for paired samples. Error bars represent mean ± SD. For total mRNA, $n = 8$ (four biological replicates × two technical replicates), $p = 0.71$. For the rest, $n = 5$ (two biological replicates, two technical replicates + three technical replicates), $p = 0.083$, 0.035, 0.049 for non-ribosome, 40S–80S, and polysome fractions, respectively. (B) Quantification of the m^{6}A/A ratio of poly(A)-selected total mRNA by LC-MS/MS for the YTHDF1 overexpression sample compared to control after 24 hours. *p* values were determined using two-sided Student's t test for paired samples. Error bars represent mean ± SD, $p = 0.0015$, $n = 6$ (three biological replicates × two technical replicates).

The promotion of translation may be attributed to two different mechanisms: either by recruiting more mRNAs to translation machinery or accelerating the translation initiation rate (Figure 2.6A). The above results already confirmed an active role of YTHDF1 in trafficking more transcripts to translation machinery, thus we next investigated whether YTHDF1 could also accelerate translation initiation. We performed quantitative translation initiation sequencing (qTI-seq, Figures 2.6B and C), a method that measures the ribosome occupancy at translation initiation sites (TIS) to probe initiation as a rate-limiting step in translation [78]. We observed an increase of the TIS occupancy in YTHDF1 knockdown samples compared to control samples (Figure 2.6D). An increase of the ribosome occupancy at TIS could be caused by either increased mRNA population in the translation initiation step or slowed translation-initiation rate or both. If YTHDF1 promotes mRNA translation solely by delivering more cellular mRNAs to translation machinery (Role A, Figure 2.6A), a decrease of TIS occupancy is expected under the YTHDF1 knockdown conditions. The increased TIS occupancy observed upon YTHDF1 knockdown suggests that YTHDF1 also directly accelerates the translation initiation rate of the ribosome-bound mRNA (Role B, Figure 3B); and this effect might be more dominant than recruiting mRNA to the translation machinery in this system.

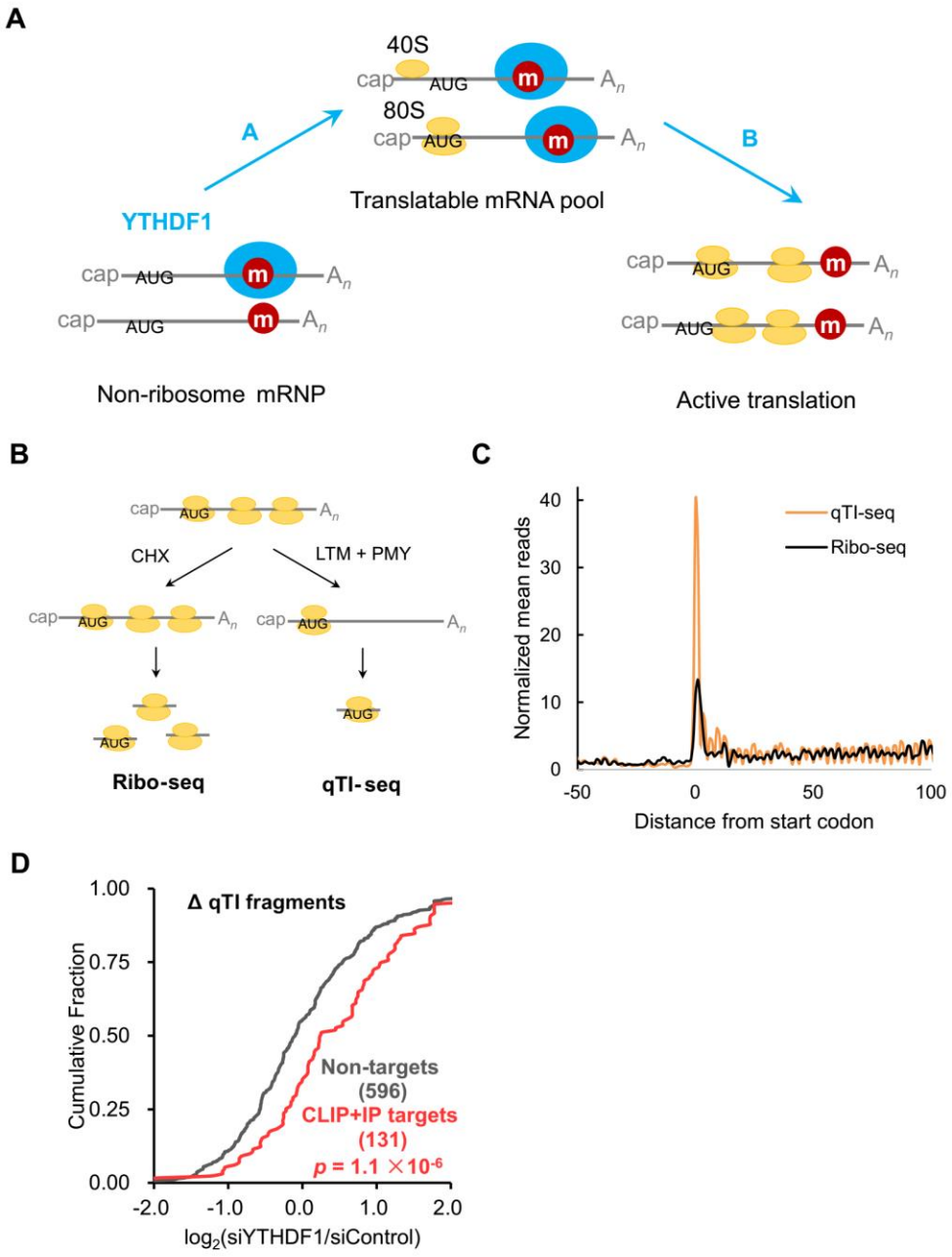


Figure 2.6 YTHDF1 enhances the translation of m^{6}A-modified mRNAs by two mechanisms
(A) A diagram illustrating that YTHDF1 plays two potential roles in the translation of m^{6}A-modified RNAs: in Role A, YTHDF1 shuttles more mRNAs to translation machinery; in Role B, YTHDF1 accelerates the translation initiation rate of methylated mRNAs. (B) Scheme of ribosome profiling (Ribo-seq) and quantitative translation initiation sequencing (qTI-seq). CHX, cycloheximide; LTX, lactimidomycin; PMY, puromycin. (C) Metagene analysis of Ribo-seq and qTI-seq. Normalized ribosome-protected fragment (RPF) reads were averaged across the entire transcriptome and aligned at the annotated start codon. (D) Cumulative distribution of the *log$_2$*-fold changes of qTI-seq reads between the YTHDF1 knockdown and knockdown control samples for non-targets (gray) and common targets based on PAR-CLIP and RIP (red).

2.2.1.4 YTHDF1 interacts with initiation factors to promote translation

From the previous polysome profiling and protein mass spectrometry experiments, we have already revealed that YTHDF1 enriches with ribosomal subunits and translation initiation factors in the 40S portion (Figure 2.7A) and interacts with translation initiation factor complex 3 (eIF3). This interaction between YTHDF1 and eIF3 was further confirmed by YTHDF1 immunoprecipitation and western blotting (Figure 2.7B). Several RNA-binding proteins with known roles in translational control were also identified by mass spectrometry, e.g., YBX1, IGF2BP1, G3BP1, and PCBP2. These *trans*-acting factors may function collectively with YTHDF1 to affect translation of methylated mRNA. Interestingly, the interaction between YTHDF1 and stress granule marker G3BP1[79] is RNA dependent, whereas the interaction between YTHDF1 with eIF3 is not (Figure 2.7B). This observation suggests that the association of YTHDF1 with translation machinery could be a direct binding, whereas its involvement with stress granules is more contingent on the presence of bound mRNA.

The differential interaction of YTHDF1 with the aforementioned two classes of protein partners was verified with colocalization patterns under fluorescence immunostaining. We treated cells with arsenite to induce stress and formation of stress granules. The co-localization of YTHDF1 with eIF3 is significant under both normal and stress conditions; while the colocalization of YTHDF1 with G3BP1 increased by more than 75% when stimulated with arsenite stress (Figure 2.7C and D). These data suggest that the association between YTHDF1 and stress granules is conditional and likely driven by the dynamics of bound mRNA, whereas the interaction between eIF3 and YTHDF1 is independent of external factors.

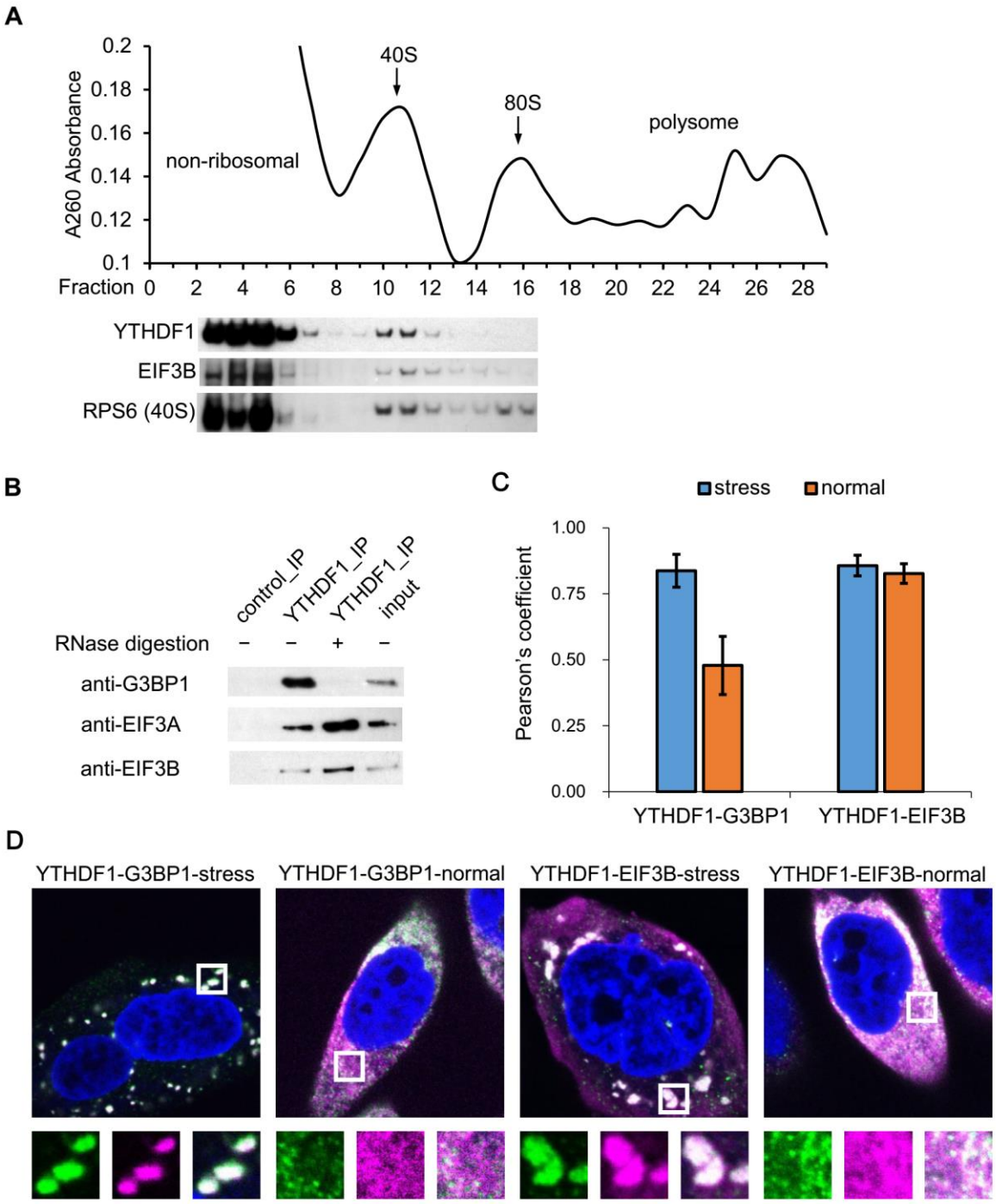


Figure 2.7 YTHDF1 interacts with translation machinery and protein markers
(A) Western blotting of Flag-tagged YTHDF1 on each fraction of 10%–50% sucrose gradient showing that YTHDF1 interacts with the ribosome. The fractions were grouped to non-ribosomal mRNPs, 40S–80S, and polysome. RPS6 is a protein subunit of 40S ribosome, and eIF3B is a component of the translation initiation complex (each lane is aligned to the corresponding fraction on the upper plot). (B) Western blotting showing that eIF3A, eIF3B, and G3BP1 co-immunoprecipitated with YTHDF1. The association of YTHDF1 with G3BP1 requires RNA, and

(**Figure 2.7, continued**) those with eIF3A/eIF3B are independent of RNA. (C) Co-localization between YTHDF1 and different protein markers were characterized by the Pearson's coefficient, for each pair, $n = 4$. The Pearson's coefficient is used to quantify the degree of colocalization between two fluorophores used to label the corresponding proteins. YTHDF1 co-localizes better with stress granules after arsenite treatment (1 mM, 1 hour) compared to the normal growth condition. It showed high colocalization with translation machinery under both normal and stress conditions, with the latter slightly better. (D) Fluorescence immunostaining of Flag-tagged YTHDF1 (green, anti-Flag, Alexa 488) and protein markers (G3BP1 for stress granules and EIF3B for translation machinery; magenta of Alexa 647 is the color for the marker, green + magenta = white for the colocalization spot). The scale of the magnified region (while frame) is 1.7 μm × 1.7 μm.

Next, we employed IRES reporters to further understand the mechanism of YTHDF1-dependent translation promotion. The translation of IRES reporters bypasses the requirement for the cap and cap-binding factor eIF4E [80]. The EMCV IRES directly binds the eIF4G subunit of the eIF4 complex. The HCV IRES bypasses the eIF4 complex and eIF4G-induced loop formation by directly recruiting 40S and eIF3 [81]. The CrPV IRES recruits the ribosome completely independent of initiation factors (eIFs). We constructed reporters with the N-terminal domain of YTHDF1 (N_YTHDF1) tethered to the 3′ UTR of these IRES reporters (Figure 2.8A). Translation assays revealed that YTHDF1 can promote the translation of both cistron-encoding cap (*Renilla* luciferase) and IRES reporters (Firefly luciferase, Figure 2.8B), supporting the role of YTHDF1 in trafficking mRNA to active translation. The role of YTHDF1 in translation initiation could be evaluated by pairwise comparisons of IRES- versus cap-dependent translation. Tethering of N_YTHDF1 enhanced the translation of EMCV IRES reporter by ~27% in comparison to cap-dependent translation (Figure 2.8C). In contrast, the CrPV and HCV IRES reporters afforded 12%–26% less protein production increases compared with the corresponding cap-dependent translation in the presence of N_YTHDF1 tethering (Figure 2.8C). These results indicate that YTHDF1-dependent translation requires eIFs and likely relies on the eIF4G-dependent loop formation (Figure 2.8D).

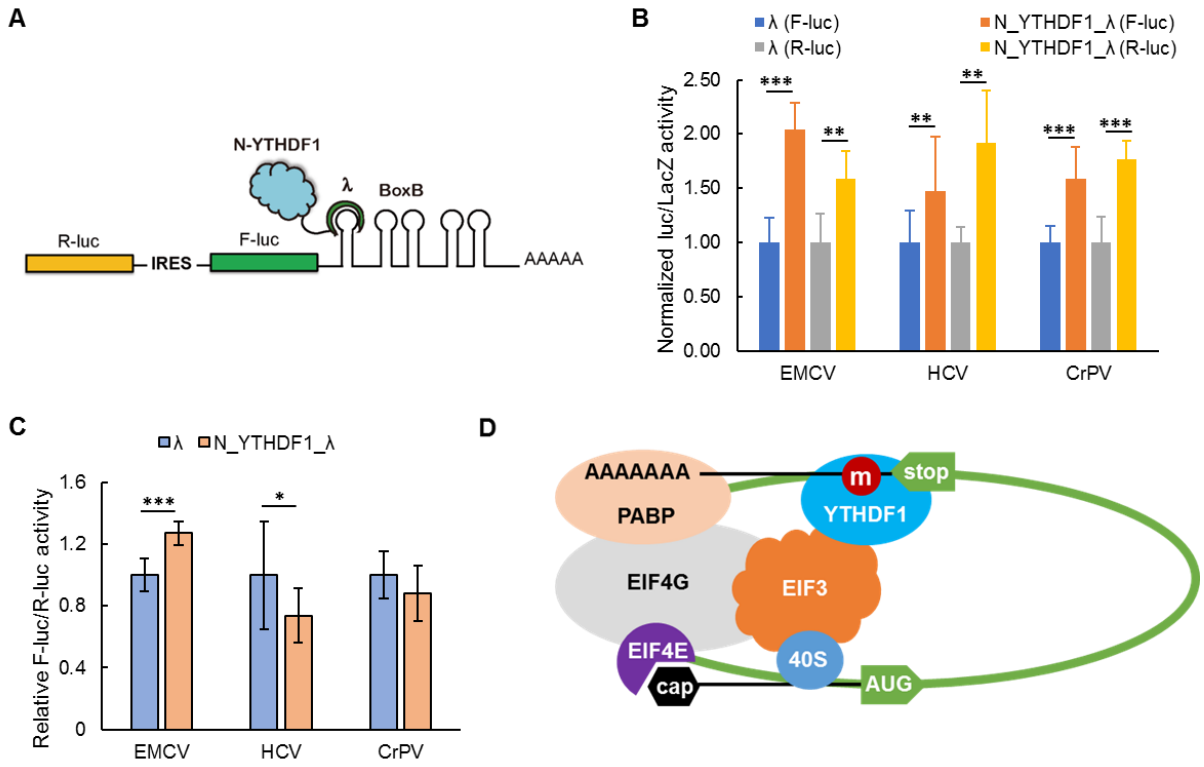


Figure 2.8 Translation promotion mechanism of YTHDF1 revealed by IRES tethering reporter assay

(A) Construct of the bicistronic IRES reporter assay. As the first coding region, *Renilla* luciferase (R-luc) reports cap-dependent translation of the mRNA and serves as a control. The second coding region encodes Firefly luciferase (F-luc) whose translation is controlled by different types of upstream IRES elements. Five Box B sequence was inserted at the 3' UTR as the tethering site for the N-terminal domain of YTHDF1. (B) Tethering assay showing the result of both cap-dependent translation (R-luc) and IRES-dependent translation (F-luc) normalized with LacZ activity as a transfection control. The effects of N_YTHDF1_λ tethering on cap dependent translation are comparable for all three groups. EMCV-dependent translation showed significantly more increase while the HCV and CrPV data revealed slightly less increase compared to the cap-dependent translation. Error bars, mean ± s.d., *P* (EMCV-F-luc) = 2.5×10^{-4}, *P* (EMCV-R-luc) = 7.4×10^{-3}, *P* (HCV-F-luc) = 1.5×10^{-3}, *P* (HCV-R-luc) = 3.0×10^{-3}, *P* (CrPV-F luc) = 7.4×10^{-3}, *P* (CrPV-R-luc) = 3.7×10^{-3}, two-sided Student's *t*-test for paired samples, *n* = 8 (biological replicates). (C) Tethering N_YTHDF1_λ to the EMCV IRES reporter led to an on average 32% increased translation of Firefly luciferase while N_YTHDF1_λ had little effect on the HCV or CrPV IRES reporter. Error bars, mean ± s.d., *P* (EMCV) = 9.5×10^{-5}, *P* (HCV) = 0.037, *P* (CrPV) = 0.050, two-sided Student's *t*-test for paired samples, *n* = 8 (biological replicates). (D) A Proposed Model of Translation Promotion by YTHDF1: YTHDF1 recruits m^6A-modified transcripts to facilitate translation initiation. The association of YTHDF1 with translation initiation machinery may be dependent on the loop structure mediated by eIF4G and the interaction of YTHDF1 with eIF3.

To sum up, our experimental evidence supports the dual roles of YTHDF1 in delivering cellular mRNAs to translation machinery and directly facilitating translation initiation (Figure 2.6A). Initiation is typically the rate-limiting step of translation [82], during which eIF4G binds both cap-binding protein eIF4E and poly(A)-binding protein to form a "closed loop". YTHDF1 could be spatially located in proximity to translation initiation sites bridged by eIF4G, forming extra bridging interaction through binding with eIF3 and 40S, thus notably accelerates the formation of loop structure and translation initiation process (Figure 2.8D).

2.2.2 Co-functioning of m^{6}A readers: YTHDF1, YTHDF2, and YTHDF3

Since many m^{6}A readers share their localization and may regulate the same groups of methylated transcripts, it is essential to investigate their interactions and potential cooperation mechanisms. We started by studying the relationship between m^{6}A readers in the cytoplasm, namely YTHDF1, YTHDF2, and YTHDF3.

2.2.2.1 YTHDF1 and YTHDF2 co-regulate the translation and decay of common targets

It has been previously reported that m^{6}A reader protein YTHDF2 decreases the stability of its m^{6}A-modified target mRNA [71], which seems to contradict the findings of the translation-promotion role of m^{6}A by YTHDF1. YTHDF1 and YTHDF2 may regulate their own subsets of mRNA targets independently; however, they also share ~50% common target transcripts (Figure 2.9A). We investigated the effects of perturbing YTHDF1 and YTHDF2 on the translation and the stability of their common targets from ribosome profiling and mRNA lifetime profiling. The transcripts were categorized into non-targets and shared targets of YTHDF1 and YTHDF2 (other transcripts were omitted for clarity). We assessed both the lifetime and translation efficiency of mRNA and drew a two-dimensional plot (Figure 2.9B and C). The result showed that the shared targets behaved quite differently compared to non-targets. With YTHDF1 knockdown, the

translation efficiency of shared targets decreased notably, whereas the average lifetime did not change. The knockdown of YTHDF2 resulted in a substantial increase in the lifetime of the shared targets but only a slight difference in the translation efficiency. These data verified the mRNA-destabilizing role of YTHDF2 and the translation-promotion role of YTHDF1. Whereas YTHDF2 controls the lifetime of the methylated transcripts, YTHDF1 ensures the efficient protein expression of these shared transcripts.

In order to provide direct evidence that YTHDF1 and YTHDF2 function by binding directly to their target mRNA, we used the same tethered luciferase reporter assay to guide N_YTHDF1_λ or N_YTHDF2_λ to the reporter transcript and assessed the impact on the cognate mRNA abundance and protein production (Figure 2.9D–F). Similar to the ribosome profiling results, N_YTHDF1 tethering led to an increased protein production, N_YTHDF2 tethering gave a decreased mRNA abundance, and combining N_YTHDF1_λ and N_YTHDF2_λ resulted in both elevated protein expression and reduced mRNA abundance. These results validated that YTHDF1 and YTHDF2 possess distinct functions: YTHDF1 promotes mRNA translation, whereas YTHDF2 facilitates degradation. When these two proteins cooperate to regulate the shared target transcripts, a relatively shallow transcript signal could be converted to a sharp protein production profile.

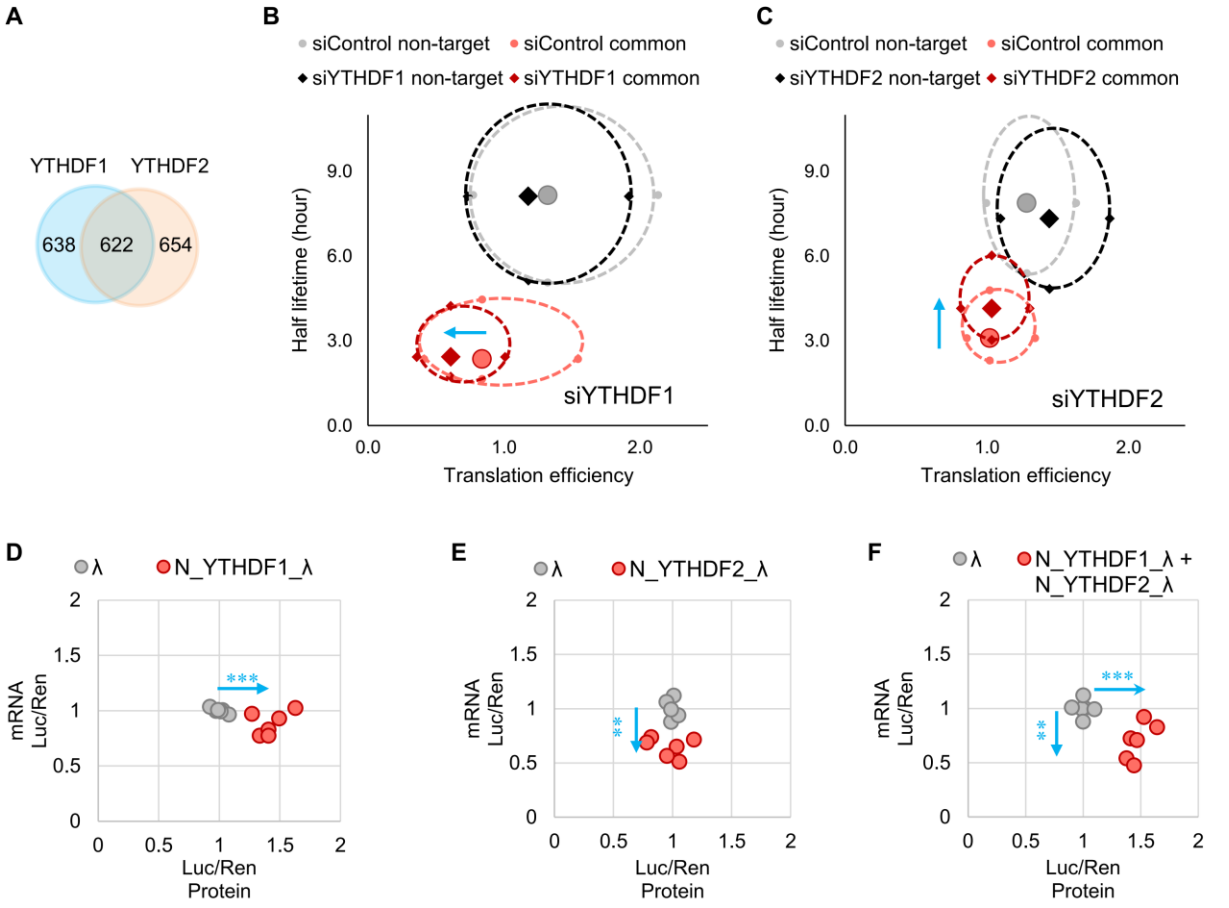


Figure 2.9 Translation efficiency and stability of the common targets of YTHDF1 and YTHDF2 is affected by both m^{6}A readers

(A) Overlap of YTHDF1 and YTHDF2 mRNA Targets. Approximately 50% of YTHDF1 targets are also YTHDF2 targets. (B and C) Evaluations of translation efficiency and mRNA half-life times of shared targets or non-targets of YTHDF1 and YTHDF2 with and without perturbation. Ribosome profiling and mRNA-seq data were collected under YTHDF1 knockdown (B) or YTHDF2 knockdown (C) conditions. Transcripts were categorized into common targets of YTHDF1 and YTHDF2 (red) or non-targets (black) under the knockdown conditions, and each compared with their corresponding control (pink and gray, respectively). The solid diamonds and circles represent median of the translation efficiency and half-life time. The four periphery dots surrounding each median are data quartiles (25% and 75% of each variance) and connected by dashed lines. The blue arrows denote the directions of changes compared to the control. (D–F) Evaluations of both mRNA abundance and protein production of the reporter transcripts using the tethering assay. Cells transfected with inducible luciferase genes were tethered by YTHDF1 (D), YTHDF2 (E), or both (F). Protein production was calculated by normalized luciferase signal (F-luc/R-luc). mRNA abundance was quantified by qRT-PCR with normalization. Data points representing the tethered group (red dots) were compared to those of the control group (tethered with control λ peptide, gray dots), and the directions of changes were shown by blue arrows.

Next, to probe the temporal order of their binding to common RNA targets, we performed metabolic labeling of nascent mRNA in pulse-chase experiments and evaluated the association of nascent mRNA with YTHDF1 or YTHDF2, respectively (Figure 2.10). The results showed that YTHDF1 binds to nascent mRNA transcripts earlier than YTHDF2, which is consistent with the expectation that translation of most of these mRNAs should occur before degradation under normal growth conditions.

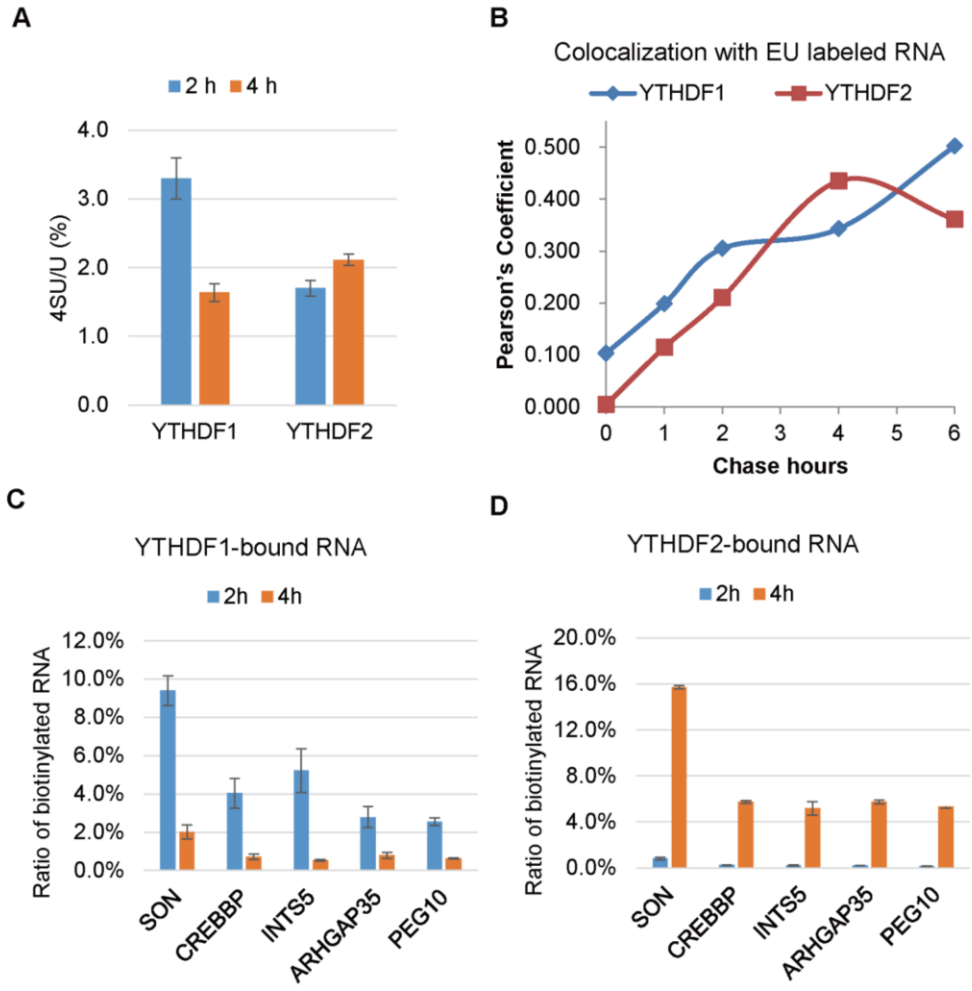


Figure 2.10 Temporal order of YTHDF1 and YTHDF2 binding to common targets
(A) Quantification of the 4-thiouridine (4SU) /U ratio of YTHDF1- and YTHDF2-bound RNAs. Nascent transcribed mRNAs were labeled by 4SU for 1 hour. RNAs bound by YTHDF1 or YTHDF2 were isolated at 2 or 4 hours post-labeling, then analyzed by LC-MS/MS. Error bars represent mean ± SD, n = 2. The results indicate that YTHDF1 binds nascent RNA before YTHDF2. (B) Nascently transcribed RNAs were labeled by ethynyl-uridine (EU) for 1 hour and

(**Figure 2.10, continued**) the association of YTHDF1 or YTHDF2 with the nascent RNA was characterized by colocalization analysis at indicated time points. The Pearson's coefficient is used to quantify the degree of colocalization between the two fluorophores that were used to label the YTH protein and nascent RNA. (C and D) After one hour pulse of EU labeling, RNAs that associate with YTHDF1 or YTHDF2 were isolated at 2 or 4 hours post-labeling. YTHDF1- or YTHDF2-bound RNAs (input) were further biotinylated via Click reaction. Nascent (biotinylated) RNAs were captured by Streptavidin-conjugated magnetic beads. For the five YTHDF1 and YTHDF2 common target genes, the amounts of their nascent RNAs in the YTHDF-bound portions were determined by qPCR, and the ratios of their captured amounts to input were calculated and shown in the figure. The results show that at 2 hours post labeling YTHDF1 binds nascent RNAs at a higher percentage than YTHDF2 (C), whereas at 4 hours post-labeling YTHDF2 binds more nascent RNAs than YTHDF1 (D). Therefore, these transcripts encounter with YTHDF1 earlier than YTHDF2. Error bars represent mean ± SD, $n = 2$ (two technical replicates).

2.2.2.2 YTHDF3 impacts the binding specificity of YTHDF1 and YTHDF2

Our recent study of another cytoplasmic m^6A reader, YTHDF3, further revealed the complexity of the interactions between m^6A readers. Forming a complex with YTHDF1 and YTHDF2, YTHDF3 shares many associating protein partners with YTHDF1 revealed by protein mass spectrometry[75]. Depletion of YTHDF3 affects target binding activity of YTHDF1 and YTHDF2 as demonstrated in PAR-CLIP experiments of YTHDF1 and YTHDF2 in cells depleted of YTHDF3, with pulled-down RNAs semi-quantified with 5′-^{32}P labeling. Knockdown of YTHDF3 led to increased amount of RNA pulled down with YTHDF1 and YTHDF2, and this increased binding correlated with YTHDF3 knockdown efficiency (Figure 2.11A). In parallel, transcriptome-wide YTHDF1 and YTHDF2 RIP-seq in control and cells YTHDF3-depleted cells revealed that: in control samples, for YTHDF1 or YTHDF2, their common targets with YTHDF3 both showed greater enrichment than their unique targets (Figure 2.11B); while in the absence of YTHDF3, RIP enrichment of YTHDF1 or YTHDF2 targets was reduced significantly (Figure 2.11C and D). These results indicate that the specific binding of YTHDF1 and YTHDF2 toward their target mRNAs is facilitated by the presence of YTHDF3.

Using LC-MS/MS to quantify the m^6A level change in mRNAs from different combinations of YTHDFs knockdown, we first observed a 21% increase of m^6A from cells with

double knockdown of YTHDF1 and YTHDF2 (Figure 2.11E), confirming our previous conclusion that both readers facilitate the fast turn-over of methylated transcripts. Yet we observed a higher 38% increase in m^6A level when all three YTHDFs were depleted (Figure 2.11E), demonstrating that the additional knockdown of YTHDF3 further aggravates the m^6A accumulation. These results indicate that all three YTHDFs contribute collectively to accelerating the metabolism of m^6A-modified mRNAs in the cytoplasm.

Finally, by metabolically labeling nascent RNA with nucleoside-analogues, we quantified the occupancy of labeled RNA by YTHDFs over time. It is observed that both YTHDF1 and YTHDF3 bound mRNA targets prior to YTHDF2, with YTHDF3 binding more nascent RNAs than YTHDF1 (Figure 2.11F), suggesting a flow of m^6A-modified transcripts from YTHDF3 to YTHDF1. Therefore, the m^6A-modified transcripts might engage YTHDF3 and then YTHDF1 for translation promotion before being decayed through the YTHDF2-mediated pathway in the cytoplasm.

These results led us to propose an interconnected and dynamic model for the regulatory functions of YTHDFs in the cytosol (Figure 2.11G): after a target m^6A-modified RNA is exported from the nucleus to the cytoplasm, it might be first recognized by YTHDF3 or a YTHDF3-YTHDF1 complex, which facilitates YTHDF1 binding for enhanced protein translation; the mRNA could be partitioned among all three YTHDF proteins and eventually bound by YTHDF2 for accelerated decay. The expression of YTHDF3 could act as a "buffering agent" for target access to YTHDF1 and YTHDF2. Fluctuation of YTHDF3 expression could affect the RNA-binding activity of the other two, impacting mRNA translation efficiency and stability. YTHDF3 may thus add robustness to the network. Its presence and potential post-translational modifications could provide additional layers of regulation of methylated target mRNAs.

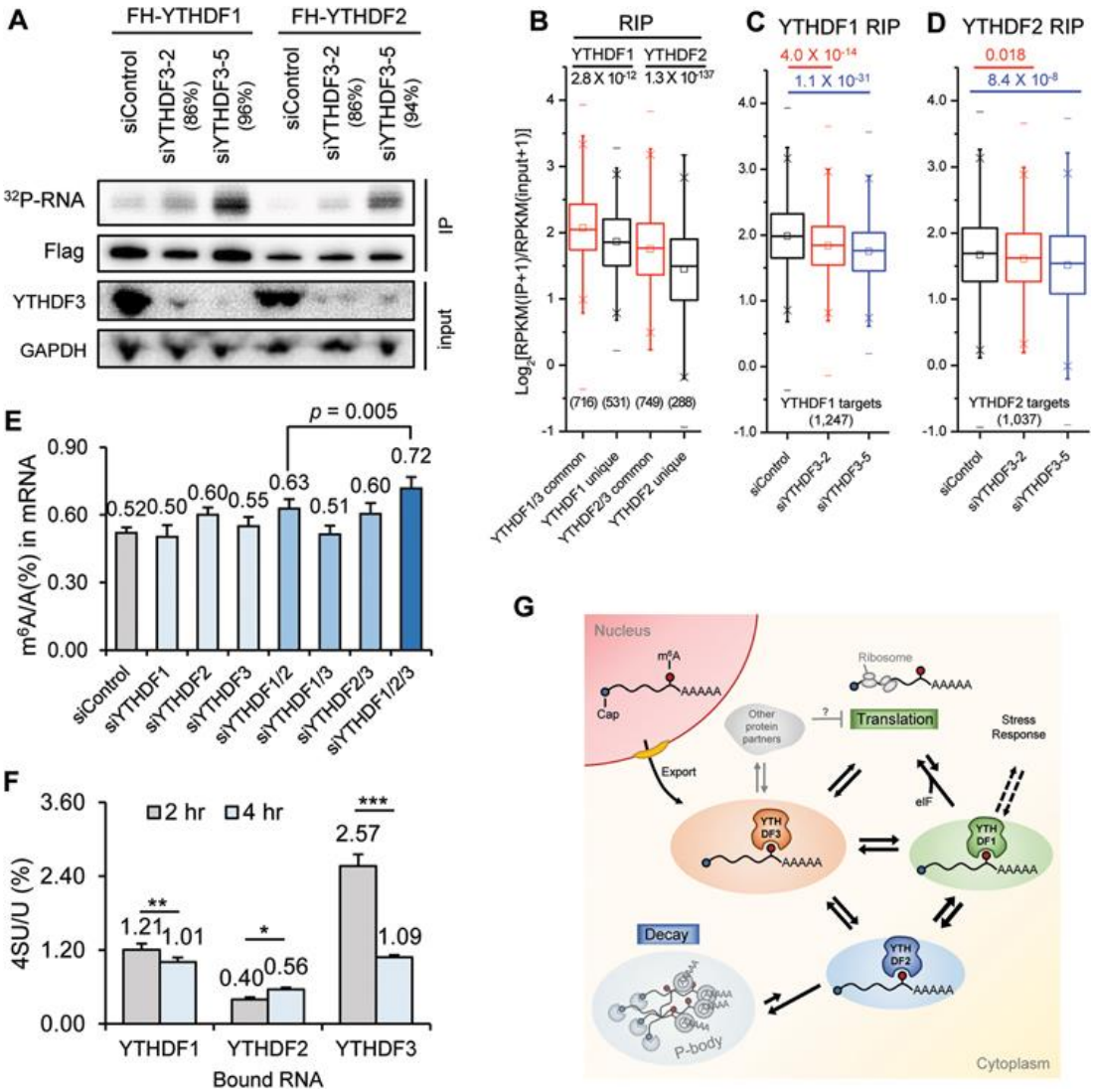


Figure 2.11 YTHDF proteins form an interconnected network in the cytoplasm

(A) Total RNAs bound by YTHDF1 and YTHDF2 quantified with PAR-CLIP followed by 5′-^{32}P labeling in the control HeLa cells and cells depleted of YTHDF3 using two different siYTHDF3 oligos. Knockdown efficiency of the two siYTHDF3 oligos was indicated, respectively. Samples loaded in the radioactivity gel were normalized with immunostaining of the Flag-tagged protein. (B-D) Genome-wide analysis of target affinity of YTHDF1 and YTHDF2 with or without YTHDF3. Box plot of RIP enrichment of different groups of YTHDF targets in siControl samples (B), and that of YTHDF1 targets (C) or of YTHDF2 targets (D) in siControl and siYTHDF3 samples. Box, 25%-75%; "–", max and min; "×", 1% and 99%; "□", median. *P* values were calculated from a two-sided Mann-Whitney test. (E) LC-MS/MS quantification of m^{6}A levels of HeLa cells treated with siControl, siYTHDF1, siYTHDF2, siYTHDF3, and combinations of those oligos. Error bars, mean ± sd, n = 4 (two biological replicates × two technical replicates). (F) LC-MS/MS quantification of 4SU (4-thio-uridine) level in mRNAs pulled down with YTHDF1-3 2-hour and 4-hour post a 1-hour 4SU labeling of nascent RNAs. Error bars, mean ± sd, n = 3~4. *P* values were calculated using paired two-sided Student's *t*-test. $^{*}P < 0.05$; $^{**}P < 0.005$; $^{***}P < 0.0005$; (G) A proposed model for an integrated partition network for m^{6}A-modified transcripts mediated by YTHDFs in the cytosol. While YTHDF1 functions in translation regulation

(**Figure 2.11, continued**) and YTHDF2 dominates in accelerating mRNA decay, YTHDF3 could serve as a hub for fine-tuning the RNA accessibility of YTHDF1-2. These three mRNA pools controlled by YTHDF1-3 could be interchangeable and highly dynamic, resulting in an interconnected and dynamic mRNA modulation through m^{6}A. YTHDF3 might also interact with other protein partners (gray) to negatively impact translation.

2.2.3 Switch of m^{6}A reader functions under certain circumstances

In addition to the primary function of each m^{6}A reader, some readers may alter the function under certain circumstances, typically upon an internal or external stimulation causing drastic changes to the cellular homeostasis. For instance, it is recently reported that during the heat shock response, instead of carrying out its regular function of facilitating cytosolic mRNA decay, YTHDF2 relocalizes to the nucleus and promotes 5′ UTR methylation by inhibiting FTO binding. This process enables selective cap-independent translation under stress conditions despite global translation suppression[83]. Our recent study of regulatory roles of m^{6}A in host-virus interaction also revealed that the functions of m^{6}A readers may switch to viral RNA sequestration and their subcellular localization may change upon viral infection (detailed discussion in chapter 4).

The potential mechanism behind these abrupt switches in reader functions may be attributed to signaling events-induced post-translation modifications (PTM) to m^{6}A readers, or newly established protein-protein interactions between m^{6}A readers and other native or foreign proteins. The understanding of these functional switches will be greatly improved by further investigation through protein mass spectrometry experiments to characterize changes in PTM states and protein interactome of m^{6}A readers under these special conditions.

2.3 Conclusion and Discussion: m^{6}A marks fast turnover of mRNA

The life cycle of mRNAs is regulated by transcriptional and post-transcriptional regulatory processes, including processing, export, translation, and decay. Recent studies have revealed that m^{6}A and its related factors influence each of these steps (Figure 2.12A). As these processes are

generally coupled, we propose that mediators of N^6-adenosine methylation may work in concert to shape the methylation pattern and protein binding of specific transcripts, thereby affecting their metabolism. The co-regulation of translation and decay by YTHDF1 and YTHDF2 of their shared targets is one of such examples. The combined function of the YTHDF1-dependent translation promotion and YTHDF2-dependent decay may result in a spike in protein production within a set period (Figure 2.12B). This effect, along with other m^6A-mediated effects such as accelerated export of certain methylated mRNA, suggests a critical function for m^6A-based gene regulation: writers and erasers dictate the levels of target-specific m^6A. In turn, readers decode these messages and may functionally sort methylated mRNAs into distinct functional groups. During cell differentiation and development, when the translation of groups of transcripts is accomplished within a short time span, methylation could sort these transcript groups into a fast track for processing, translation, and decay. Methylation could be particularly beneficial in grouping and synchronizing the expression of hundreds to thousands of mRNAs that otherwise may possess markedly different properties with varied stabilities and translation efficiencies. Such a mechanism may also help in generating translation 'pulses' to satisfy the need for bursts of protein synthesis as well as rapid decay to regulate cell differentiation during early development.

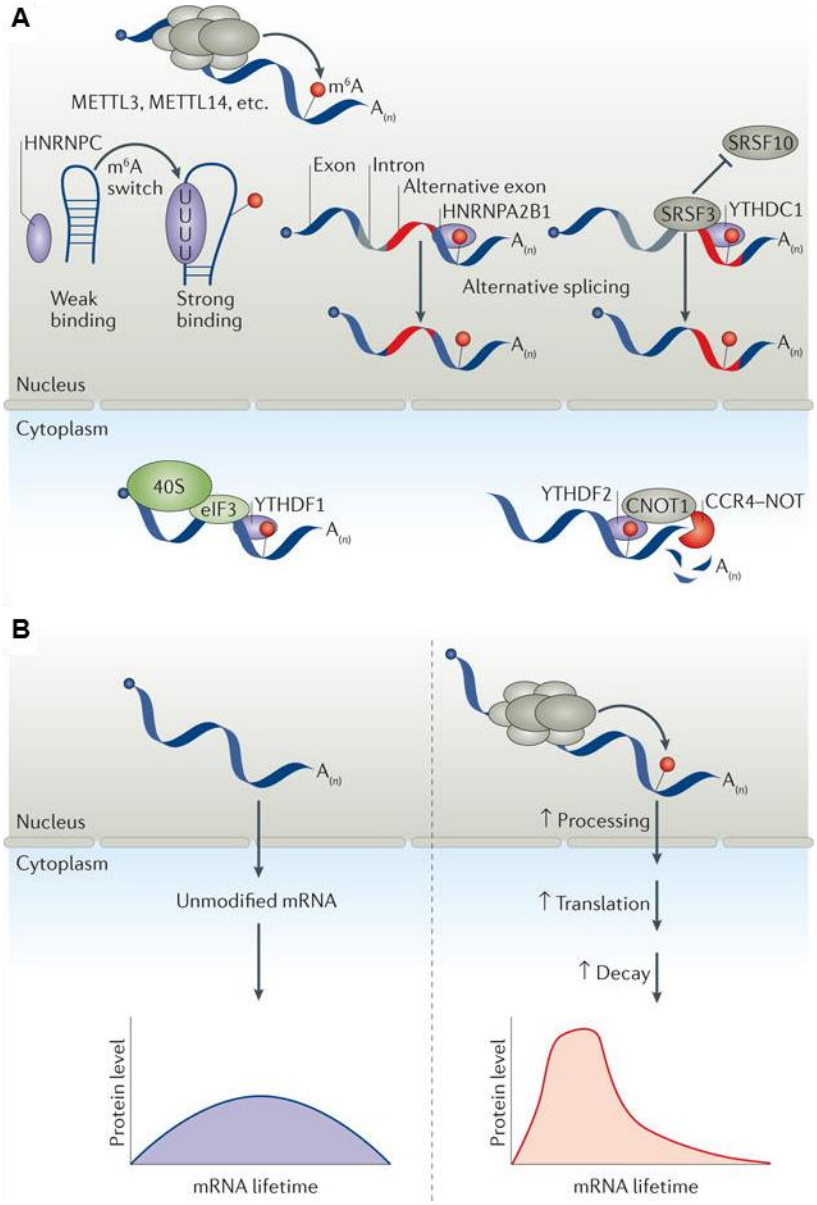


Figure 2.12 m^{6}A-dependent regulatory pathways accelerate mRNA metabolism

(A) After being deposited by the methyltransferase core catalytic components methyltransferase-like 3 (METTL3) and METTL14, *N*^{6}-methyladenosine (m^{6}A) is recognized by various reader proteins. In the nucleus, heterogeneous nuclear ribonucleoprotein C (HNRNPC) functions as an indirect m^{6}A reader by binding unstructured m^{6}A switch regions and regulating splicing, whereas YTHDC1 regulates alternative splicing by binding m^{6}A directly and recruiting the splicing factors serine and arginine-rich splicing factor 3 (SRSF3) while blocking binding by SRSF10. HNRNPA2B1 also mediates alternative splicing in a manner similar to YTHDC1. In the cytoplasm, YTHDF1 mediates translation initiation of m^{6}A-containing transcripts by binding directly to m^{6}A and recruiting eukaryotic initiation factor 3 (eIF3), thereby facilitating the loading of the eukaryotic small ribosomal subunit (40S). YTHDF2 promotes mRNA decay by binding to CCR4–NOT transcription complex subunit 1 (CNOT1), thereby facilitating the recruitment of the CCR4–NOT complex and inducing accelerated deadenylation. (B) Methylated transcripts may be sorted by reader proteins into a fast track (right) for processing, translation, and decay. This fast-

(**Figure 2.12, continued**) tracking effectively groups transcripts with otherwise markedly different properties to ensure their timely and coordinated translation and degradation, possibly generating a sharp 'pulse' of gene expression to satisfy a need for translational bursts and subsequent clearance of these transcripts.

2.4 Materials and Methods

Plasmid construction and protein expression

Flag-tagged YTHDF1 was cloned from commercial cDNA clones (Open Biosystems) into vector pcDNA 3.0 using the following restriction enzymes and primers: EcoRI, XhoI. Forward primer: CGTACGAATTCATGGATTACAAGGACGACGATGACAAGATGTCGGCCACCA-GCG; reverse primer: CCATACTCGAGTCATTGTTTGTTTCGACTCTGCC. Plasmids with high purity for mammalian cell transfection were prepared with a Maxiprep kit (Qiagen). λ peptide sequence (MDAQTRRRERRAEKQAQWKAAN) was fused to the C-terminus of N-YTHDF1 by subcloning N-YTHDF1 into pcDNA 3.0 with forward primer containing Flag-tag sequence and reverse primer containing λ peptide sequence (pcDNA-Flag-Y1Nλ, BamHI, XhoI; forward primer, CGTACGGATCCATGGATTACAAGGACGACGATGACAAGATGTCGGCCACCAGCG; reverse primer,CTATGGCTCGAGTCAGTTTGCAGCTTTCCATTGAGCTTGTTTCTCAGC-GCGACGCTCACGTCGTCGTGTTTGTGCGTCCATACCGACGCTGGGGGCAGAATT). The construction of λ peptide control that contains an N-terminal Flag tag and GGS spacer (pcDNA-Flag-λ), the N_YTHDF2_λ construct and the tether reporter pmirGlo-Ptight-5BoxB were reported previously[71].

IRES-containing tethering reporter plasmids were constructed by inserting the 5BoxB sequence into the 3'UTR of the original IRES reporter plasmids (pRF-HCV, pRF-EMCV, pRF-CrPV. Gifts from Prof. Anne Willis) [81]. 5BoxB insert was amplified using the following primer with single restriction site FseI:

Forward primer: CGATACGGCCGGCCTTCCCTAAGTCCAACTACCAAAC; reverse primer: CTATGGGGCCGGCCATAATATCCTCGATAGGGCCC).

Mammalian cell culture, siRNA knockdown and plasmid transfection

Human HeLa cell line used in this study was purchased from ATCC (CCL-2) and grown in DMEM (Gibco, 11965) media supplemented with 10% FBS and 1% 100 × Pen/Strep (Gibco). HeLa Tet-off cell line was purchased from Clontech and grown in DMEM (Gibco) medium supplemented with 10% FBS (Tet system approved, Clontech), 1% 100 × Pen/Strep (Gibco) and 200 μg ml^{-1} G418 (Clontech). AllStars negative control siRNA from Qiagen (1027281) was used as control siRNA in knockdown experiments. *YTHDF1* siRNAs were ordered from Qiagen. (Hs_YTHDF1_1 with target sequence CCGCGTCTAGTTGTTCATGAA; Hs_YTHDF1_8 with target sequence CAGGCTGGAGAATAACGACAA). *YTHDF2* siRNA was reported previously [71]. METTL3 siRNA was reported previously [66]. Transfection was achieved by using Lipofectamine RNAiMAX (Invitrogen) for siRNA, and Lipofectamine LTX Plus (Invitrogen) for transfection of one or multiple types of plasmids (tethering assay) following the manufacturer's protocols.

RNA isolation

mRNA isolation for LC-MS/MS: total RNA was isolated from wild-type or transiently transfected cells with TRIzol reagent (Invitrogen). mRNA was extracted using FastTrack MAG mRNA Isolation Kits (Invitrogen) or Ultracapture mRNA Isolation Kit (Wisegene) followed by further removal of contaminated rRNA using RiboMinus Eukaryote Kit v2 (Ambion). mRNA concentration was measured by NanoDrop. Total RNA isolation for RT–PCR: following the instruction of RNeasy kit (Qiagen) with DNase I digestion step. Ethanol precipitation: to the RNA solution being purified or concentrated, 1/10 volume of 3 M NaOAc, pH 5.5, 1 µl glycogen (10 mg ml^{-1}) and 2.7 volume of 100% ethanol were added, stored at –80°C for 1 h to overnight, and

then centrifuged at 15,000*g* for 15 min. After removing the supernatant, the pellet was washed twice by using 1 ml 75% ethanol, and dissolved in the appropriate amount of RNase-free water as indicated.

PAR-CLIP

We followed the previously reported protocol [71]. Two biological replicates were conducted.

RIP-seq

The procedure was reported previously [71]. Input mRNAs were prepared by either Poly(A) selection (replicate 1, FastTrack MAG Micro mRNA isolation kit, Invitrogen) or rRNA removal (replicate 2, RiboMinus Eukaryote Kit v2, Ambion). Input mRNA and IP with 150-200 ng RNA of each sample were used to generate the library using TruSeq stranded mRNA sample preparation kit (Illumina).

Ribosome profiling

Ribosome profiling was conducted using ARTseq Ribosome Profiling Kit (Mammalian, Epicentre). Two biological replicates with the use of different YTHDF1 siRNA sequences (siYTHDF1_1 and siYTHDF1_8) were studied to avoid potential off-target effects of specific siRNA. The manufacturer's procedure was followed exactly until library construction. For library construction, the end structures of the RNA fragments for ribosome profiling and mRNA input were repaired using T4 PNK: (1) 3′ de-phosphorylation: RNA (20 μl) was mixed with 2.5 μl PNK buffer and 1 μl T4 PNK, and kept at 37°C for 1 h; (2) 5′-phosphorylation: to the reaction mixture, 1 μl 10 mM ATP and 1 μl extra T4 PNK were added, and the mixture was kept at 37°C for 30 min. The RNA was purified by RNA clean and concentrator (Zymo) and finally dissolved in 10 μl water. The library was constructed by NEBNext small RNA sample preparation kit for high-throughput sequencing (NEB). The sequencing data obtained from ribosome profiling were denoted as

ribosome-bound fragments and that from RNA input as mRNA input. Translation efficiency was defined as the ratio of ribosome-protected fragments and mRNA input, which reflected the relative occupancy of 80S ribosome per mRNA species.

RNA-seq for mRNA lifetime measurements

Two 10-cm plates of HeLa cells were transfected with YTHDF1 siRNA or control siRNA at 30% confluency. After 6 h, each 10-cm plate was re-seeded into three 6-cm plates, and each plate was controlled to afford the same amount of cells. After 48 h, actinomycin D was added to 5 μg ml^{-1} at 6 h, 3 h, and 0 h before trypsinization collection. The total RNA was purified by RNeasy kit (Qiagen). Before construction of the library with Tru-seq mRNA sample preparation kit (Illumina), ERCC RNA spike-in control (Ambion) was added to each sample (0.1 μl per sample) proportional to total RNA.

Quantitative translation initiation sequencing (qTI-seq)

We followed the previously reported protocol [78]. Briefly, two 15-cm plates of cells were harvested in 5 ml cold PBS and collected by centrifugation. 400 μL lysis buffer containing lactimidomycin (LTM; 5 μM) was added to the cell pellets. Cells were then incubated on ice for 10 min. After centrifugation at 4 °C and 12,000g for 15 min, the supernatant was saved for puromycin treatment (25 μM, room temperature 15 min). Puromycin-treated samples were then subjected to the same procedure as regular ribosome profiling. Libraries were constructed similarly and the data analysis was carried out according to the protocol provided[78].

Sequencing data analysis

General pre-processing of reads: All samples were sequenced by Illumina Hiseq2000 with single end 100 or 50-bp read length. For libraries that generated from small RNA (PAR-CLIP and ribosome profiling), the adapters were trimmed by using FASTX-Toolkit [84]. The deep sequencing

data were mapped to Human genome version hg19 by Tophat version 2.0 [85] without any gaps and allowed for at most two mismatches. RIP and Ribosome profiling were analyzed by DESeq [86] to generate RPKM (reads per kilobase, per million reads). mRNA lifetime data were analyzed by Cuffdiff version 2.0 [85] to calculate RPKM.

Data analysis for each experiment: (1) for RIP, enrichment fold was calculated as $\log_2$(IP/input); (2) PAR-CLIP data were analyzed by PARalyzerv1.1 [87] with default settings. Binding motif was analyzed by HOMER (v4.7) [88]; (3) for ribosome profiling, only genes with RPKM >1 were used for analysis and the change fold was calculated as $\log_2$(siYTHDF1/siControl); (4) for qTI-seq, mapped reads that fall within a 51bp window centering at annotated translation initiation sites (TIS) were fed to Cuffdiff 2.0 to calculate differential expression of TIS occupancy between siYTHDF1 and siControl samples, and Figure S3H were plotted with the average of two biological replicates using different siYTHDF1 sequences; (5) for mRNA lifetime profiling: RKPM were converted to attomole by linear-fitting of the RNA spike-in.

The degradation rate of RNA *k* was estimated by

$$\ln\left(\frac{A_t}{A_0}\right) = -kt$$

where *t* is transcription inhibition time (h), A_t and A_0 represent mRNA quantity (attomole) at time *t* and time 0. Two *k* values were calculated: time 3 h versus time 0 h, and time 6 h versus time 0 h. The final lifetime was calculated by using the average of k_{3h} and k_{6h}.

$$t_{\frac{1}{2}} = \frac{2\ln 2}{k_{3h} + k_{6h}}$$

Integrative data analysis and statistics: PAR-CLIP targets were defined as reproducible gene targets among two biological replicates (4,913). RIP targets (1,747) were genes with $\log_2$(IP/input) > 1. The overlap of PAR-CLIP and RIP targets were defined as CLIP+IP targets

(1,261). And non-targets (2,822) should meet the conditions of: (1) complementary set of PAR-CLIP targets; (2) RIP enrichment fold <0. For the comparison of PAR-CLIP and m^6A peaks, at least 1 bp overlap was applied as the criteria of overlap peaks. Two biological replicates were conducted for ribosome profiling and mRNA lifetime profiling, respectively. Genes with sufficient expression level (RPKM >1) were subjected to further analysis. The change fold that used in the main text is the average of the two $\log_2$(siYTHDF1/siControl) values. Nonparametric Mann–Whitney *U*-test (Wilcoxon rank-sum test, two sided, significance level = 0.05) was applied in ribosome profiling data analysis as previously reported [89]. Gene Ontology (GO) term analyses were performed by DAVID [90]. Top 30 terms were then selected for visualizations by the interactive graph function of REVIGO [91].

Data accession: all the raw data and processed files have been deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) are accessible under GSE63591.

LC-MS/MS

100-200 ng of mRNA was digested by nuclease P1 (2 U) in 25 µl of buffer containing 10 mM of NH_4OAc (pH=5.3) at 42°C for 2 h, followed by the addition of NH_4HCO_3 (1 M, 3 µl, freshly made) and alkaline phosphatase (0.5 U). After an additional incubation at 37°C for 2 h, the sample was diluted to 50 µl and filtered (0.22 µm pore size, 4 mm diameter, Millipore), and 5 µl of the solution was injected into LC-MS/MS. Nucleosides were separated by reverse phase ultra-performance liquid chromatography on a C18 column with on-line mass spectrometry detection using an Agilent 6410 QQQ triple-quadrupole LC mass spectrometer in positive electrospray ionization mode. The nucleosides were quantified by using the nucleoside to base ion mass transitions of 282 to 150 (m^6A), and 268 to 136 (A). Quantification was performed in comparison with the standard curve obtained from pure nucleoside standards running on the same batch of

samples. The ratio of m^6A to A was calculated based on the calibrated concentrations (Jia et al., 2011).

RT–PCR

Real-time PCR (RT–PCR) was performed to assess the relative abundance of mRNA. All RNA templates used for RT–PCR were pretreated with on column DNase I in the purification step. When the examined gene has more than one isoform, only regions shared by all isoforms were selected for primer design to evaluate the overall expression of that gene. RT–PCR was performed by using Platinum one-step kit (Invitrogen) with 200–400 ng total RNA template or 10–20 ng mRNA template. *HPRT1* was used as an internal control because: (1) *HPRT1* mRNA does not have m^6A peak based on the m^6A profiling data; (2) *HPRT1* mRNA was not bound by YTHDF1 or YTHDF2 from the PAR-CLIP and RIP sequencing data; (3) *HPRT1* showed relatively invariant expression upon YTHDF1 or YTHDF2 knockdown based on the RNA-seq data; (4) *HPRT1* is a house-keeping gene. The primer sequences of *SON*, *CREBBP*, *RPL3*0, *HPRT1*, F-luc, and R-luc have been reported previously [71].

Polysome profiling

Four 15-cm plates of HeLa cells were prepared for each sample (siControl, siYTHDF1 or YTHDF1 stable expression line). Before collection, cycloheximide (CHX) was added to the media at 100 μg ml^{-1} for 2 min. The media was removed, and the cells were collected by cell lifter with 5 ml cold PBS with CHX (100 μg ml^{-1}). The cell suspension was spun at 400g for 2 min and the cell pellet was washed once by 5 ml PBS-CHX per plate. 1 ml lysis buffer (10 mM Tris, pH 7.4, 150 mM KCl, 5 mM $MgCl_2$, 100 μg ml^{-1} CHX, 0.5% Triton-X-100, freshly add 1:100 protease inhibitor, 40 U ml^{-1} SUPERasin) was added to suspend the cells and then kept on ice for 15 min

with occasional pipetting and rotating. After centrifugation at 15,000*g* for 15 min, the supernatant (~1.2 ml) was collected and absorbance tested at 260 nm (150–200 $A_{260\,nm}$ ml^{-1}).

A 10/50% w/v sucrose gradient was prepared in a lysis buffer without Triton-X-100. Supernatant from the previous step was loaded onto the sucrose gradient and centrifuged at 4°C for 4 h at 27,500 r.p.m. (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 resulting from sucrose gradient were used for western blotting or RNA extraction. For western blotting, 10 μl of the sucrose fraction were loaded to each well. For RNA analysis, the fractions were then categorized into three main sub-types: non-ribosome portion, 40S–80S and polysome based on western results. They were pooled to isolate total RNA by TRIzol reagent for RT–PCR and mRNA for LC-MS/MS test of the m^6A/A ratio.

Tandem affinity purification of YTHDF1 protein interactome and mass spectrometry identification

We followed the previously reported protocol [92]. Briefly, stable expression HeLa cell line with double-tagged YTHDF1 (N terminal Flag and HA in tandem) was created by puromycin selection. The control cell line expressing only tandem Flag and HA peptides was created similarly. Thirty 15 cm plates of YTHDF1 or control HeLa stable line cells were collected by cell lifter and suspended in ice-cold PBS (5 mL per plate). The cell pellets were pooled and washed once with 30 mL cold PBS. 45 mL hypotonic buffer (10 mM Tris, pH 7.4, 10 mM KCl, 1.5 mM $MgCl_2$) was added into the tube. The pellet was loosened by inverting the tube and swollen after 15 min incubation on ice. The cells were collected at 3000 rpm for 10 min at 4 ^oC. The pellet volume roughly doubled and the supernatant was carefully removed by pipet. An equal volume of hypotonic buffer with protease inhibitor and SUPERasin was added. The mixture was

homogenized by douncing with pestle 25 to 30 times and then spun down at 3500 rpm at 4 ℃ for 15 min. The resulted pellet (nuclear part) was discarded and the supernatant was pre-S100 (cytoplasmic part). The supernatant was transferred into a new tube with 10X buffer (300 mM Tris, pH 7.4, 1.4M KCl, 30 mM $MgCl_2$) at 10% volume of total supernatant volume added and well mixed. The mixture was cleared by ultra-centrifuge at 27,500 rpm for 60 min at 4 ℃. The resulted supernatant (S100) was dialyzed in 1L BC-100 buffer overnight (20 mM Tris, pH 7.4, 100 mM KCl, 20% glycerol, 0.5 mM DTT). The dialysate was cleared at 14,000 rpm for 20 min at 4 ℃ and the supernatant was transferred to a new tube. 250 μL anti-Flag magnetic beads (Sigma) were washed by washing buffer (50 mM Tris, pH 7.9, 100 mM KCl, 5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.1% NP-40, 10% glycerol, protease inhibitor and SUPERasin) three time, combined with the cleared cytoplasmic extract and incubated at 4 ℃ for 4 h with rotation. The beads were then washed by washing buffer for four times, followed by incubation with 300 μL elution solution containing 3 × Flag peptide (0.5 mg/ml in washing buffer, Sigma) at 4°C for 1 h. During this time, 50 μL anti-HA magnetic beads (Pierce) were washed three times with washing buffer. The eluted samples were incubated with anti-HA beads for 3 h at 4 ℃ followed by four washes with washing buffer (without SUPERasin). The final protein complex was eluted by 200 μL elution buffer (0.5 mg/mL HA peptide in washing buffer, Sigma), purified by TCA precipitation and trypsin digested. The protein mass spectroscopy was performed by Institutes of Biomedical Sciences at Fudan University, Shanghai.

Protein co-immunoprecipitation and western blot

Stable line cells expressing Flag-HA tagged YTHDF1 or Flag-HA peptide were collected by cell lifter (three 15-cm plates for each), and pelleted by centrifuge at 400 *g* for 5 min. The cell pellet was resuspended in 2 volumes of lysis buffer (150 mM KCl, 10 mM HEPES pH 7.6, 2 mM

EDTA, 0.5% NP-40, 0.5 mM DTT, 1:100 protease inhibitor cocktail, 400 U mL^{-1} RNase inhibitor), and incubated on ice for 10 min. To remove the cell debris, the lysate solution was centrifuged at 15,000 *g* for 15 min at 4°C. The lysate was then split into two portions. One of them was subjected to RNase digestion by adding 100 U RNase A and 50 U RNase T1 into the lysate solution and incubating for 20 min at room temperature. The resulting supernatant was further cleared by passing through a 0.45-μm membrane syringe filter. While 50 μl of cell lysate was saved as Input, the rest was incubated with the anti-Flag M2 magnetic beads (Sigma) in ice-cold NT2 buffer (200 mM NaCl, 50 mM HEPES pH 7.6, 2 mM EDTA, 0.05% NP-40, 0.5 mM DTT, 200 U mL^{-1} RNase inhibitor) for 4 h at 4°C. Afterwards, the beads were subject to extensive wash with 8 × 1 ml portions of ice-cold NT2 buffer, followed by incubation with the elution solution containing 3 × Flag peptide (0.5 mg ml^{-1} in NT2 buffer, Sigma) at 4°C for 2 h. The eluted samples, saved as IP, were analyzed by western blotting. For IP samples, each lane was loaded with 2 μg IP portion; and the input lane was loaded with 10 μg input portion which corresponded to ~1% of overall input.

Fluorescence microscopy

Fluorescent immunostaining: the previously reported protocol[79] was followed. Cells were grown in an 8-well chamber (Lab-Tek). After treatment indicated in each experiment, cells were washed once with PBS and then fixed in 4% paraformaldehyde in PBS (Alfa Aesar) at room temperature for 15 min. The fixing solution was removed, and ice-cold methanol was immediately added to each chamber and incubated for 10 min at room temperature. Cells were rinsed once with PBS and incubated with blocking solution (10% FBS with PBST) for 1 h at room temperature under rotation. After that, the blocking solution was replaced with primary antibody (diluted by fold indicated in Antibodies section in blocking solution) and incubated overnight at 4°C. After being washed 4 times with PBST (300 μl, 5–10 min for each wash), secondary antibody (1:300

dilution in PBST) was added to the mixture and incubated at room temperature for 1 h. After washing 4 times with PBST (300 μl, 5–10 min for each wash), Slowfade antifade reagent with DAPI (Invitrogen) was added to the slides for direct imaging.

Image capture and analysis: images were captured by Leica TCS SP2 AOBS Laser Scanning Confocal microscope, analyzed by ImageJ. The colocalization was quantified by JAcoP (ImageJ plug-in) and the Pearson coefficients were gained under Costes' automatic threshold [93].

Tethering assay

Basic setting: 50 ng reporter plasmid (pmirGlo-pTight-5BoxB) and 450 ng effecter plasmid (λ, N-YTHDF1-λ, N-YTHDF2-λ, or both N-YTHDF1-λ and N-YTHDF2-λ in pcDNA3.0) was used to transfect HeLa cells in each well of six-well plate at 60~80% confluency under doxycycline (100 ng/mL) inhibition. After 6 h, the transfection mixture was replaced with fresh media containing doxycycline. After 18 h, each well was trypsin-digested, extensively washed with PBS and reseeded into 96-well plate (1:20) and 12-well plate (1:2) without doxycycline. 8 hours after reseeding and induction, cells in 96-well plate were assayed by Dual-Glo Luciferase Assay Systems (Promega). Firefly luciferase (F-luc) activity was normalized by *Renilla* luciferase (R-luc) to evaluate translation of the reporter. Samples in 12-well plate were processed to extract total RNA (DNase I digested), followed by RT–PCR quantification. The amount of F-luc mRNA was also normalized by that of R-luc mRNA.

For pulse-induction: 2 h after reseeding, 500 ng/mL doxycycline was added into the wells, ending the 2-hour pulse induction of transcription. Then the translation dynamics was monitored by assaying cells each hour from reseeding (set as $t = 0$) to 8 h after reseeding (except 7 h).

For stress-recovery: 2 h after reseeding, 500 ng/mL doxycycline was added into the wells, ending the 2-hour pulse induction of transcription. At the same time, 1 mM sodium arsenite was also

added into the wells, and removed at 3 h, completing the 1-hour stress treatment. The translation dynamics was monitored by assaying cells each hour from reseeding (set as $t = 0$) to 8 h after reseeding (except 7 h).

For translation efficiency evaluation: Protein production was calculated by normalized luciferase signal (F-luc/R-luc) after 8 h post reseeding and induction. mRNA abundance was quantified by qRT-PCR with normalization (F-luc/R-luc) after 8 h post reseeding and induction.
For IRES-containing tethering reporter assay: 5 ng reporter plasmid (pRF-HCV-5BoxB, pRF-EMCV-5BoxB, pRF-CrPV-5BoxB), 45 ng effecter plasmid (λ and N-YTHDF1-λ in pcDNA3.0), and 5 ng control plasmid (pJ7-LacZ) were used to transfect ~ 1.5×10^5 HeLa cells in each well of a 96-well plate. After 24 h half of the cells were assayed by Dual-Glo Luciferase Assay Systems (Promega) to measure F-luc and R-luc activity. The other half of the cells were assayed by Galacto-Light Plus System (Applied Biosystems) to measure the LacZ activity, which was then used to normalize both F-luc and R-luc signals. Finally, F-luc activity was normalized by R-luc to evaluate IRES-dependent expression compared to cap-dependent expression.

Pulse-chase experiment by metabolic labeling of nascent mRNA

Pulse-chase experiment was designed to investigate the sequential binding of YTHDF1 and YTHDF2. Two metabolites, 4-thiouridine (4SU) and 5-ethinyluridine (EU), were each used in the following experiments. They mimic the natural form of uridine and can be incorporated into the newly transcribed nascent mRNA. By following these nascent mRNA and studying the proteins they interact with in temporal order, we could reveal the binding sequence of YTHDF1 and YTHDF2 during the lifetime of an mRNA.

LC-MS/MS test with 4-thiouridine (4SU) labeling: one 15-cm plate of stable line HeLa cells expressing Flag-HA tagged YTHDF1 or Flag-HA tagged YTHDF2 was cultured with 0.5

mM 4SU containing media for 4 h. Then 4SU was removed and media were replaced with fresh ones. Cells were collected after chasing for 2 h or 4 h. Immunoprecipitation of the YTH-bound fraction was performed using anti-Flag magnetic beads. RNA bound by YTHDF1 or YTHDF2 was extracted from IP fraction and mRNA was purified accordingly. LC-MS/MS was carried out as previously described. The amount of 4SU in each mRNA portion indicates the extent of association between each YTH protein and nascent mRNA.

Colocalization and qPCR test with 5-ethinyluridine (EU) labeling: one 15-cm plate of stable line HeLa cells expressing Flag-HA tagged YTHDF1 or Flag-HA tagged YTHDF2 was cultured with 0.1 mM EU containing media for 1 h. EU was then removed and media were replaced with fresh ones. For colocalization study, cells were fixed after chasing for 2 h, 4 h, 6 h or 8h. Then cells were treated with Click-iT RNA Alexa Fluor 594 Imaging Kit (Life Technologies) and imaged accordingly. Colocalization analysis was carried out as described previously [71,93]. The Pearson's coefficient is used to quantify the degree of colocalization between two fluorophores. More colocalization indicates more association between the YTH protein and nascent mRNA. For qPCR test, cells were collected after chasing for 2 h or 4 h. Total RNA was isolated and subjected to biotinylation by Click reaction using Click-iT Nascent RNA Capture Kit (Life Technologies) following manufacturer's protocol. The cDNA of both input RNA and captured nascent RNA were synthesized and subjected to qPCR test using qPCR primers for common targets of YTHDF1 and YTHDF2.

Antibodies

The antibodies used in this study were listed below in the format of name (application; catalogue; supplier; dilution fold): Rabbit anti-YTHDF1 (Western; ab99080; Abcam; 1,000). Rabbit anti-YTHDF1 (Western, 17479-1-AP; Proteintech; 2,000). Rabbit anti-YTHDF2 (Western,

24744-1-AP; Proteintech; 2 000). Mouse anti-YTHDF3 (Western, sc-377119; Santa Cruz Biotech; 500). Mouse anti-Flag HRP conjugate (Western; A5892; Sigma; 5000). Goat anti-GAPDH HRP conjugate (Western; A00192; GeneScript; 15,000). Rat anti-Flag (IF; 637304; Biolegend; 300). Mouse anti-RPS6 (Western; #2317; Cell signaling; 1,000). Rabbit anti-G3BP1 (Western; A302-033A; Bethyl; 2,000). Mouse anti-G3BP1 (IF; sc-365338; Santa Cruz; 200). Goat anti-EIF3B (IF; sc-16377; Santa Cruz Biotech; 100) (Western; 500). Rabbit anti-EIF3A (Western; A302-002A; Bethyl; 2,000). Donkey anti-rat Alexa 488 (IF; A21208; Molecular Probes; 300). Goat anti-rabbit Alexa 647 (IF; A21446; Molecular Probes; 300). Goat anti-mouse Alexa 647 (IF; A21236; Molecular Probes; 300). Donkey anti-goat Alexa 647 (IF; A21447; Molecular Probes; 300).

CHAPTER 3

The Cellular Functions of m^{6}A *

3.1 Introduction: m^{6}A marks functional cohorts of transcripts in cells

Beyond the above-mentioned molecular functions, m^{6}A may also affect more complex phenotypic events by marking relevant functional cohorts of mRNAs and regulate their metabolism accordingly. Thereby the molecular functions of m^{6}A collectively translate into the control of complex cellular functions. Such controls may be required during the cellular transition between distinct states during differentiation and development when cells rapidly replace their stage-specific transcriptomes to re-establish a new identity. m^{6}A could be important in shaping the levels of mRNAs of various transcription factors and therefore may serve as barriers to or act as facilitators of these transitions.

One example of such impacts of m^{6}A was discovered in the study of the mammalian circadian rhythm. Maintenance of the circadian rhythm (clock) involves a negative feedback loop of gene expression, in which clock proteins downregulate the transcription of clock genes. However, only one-fifth of these rhythmic genes are driven by *de novo* transcription[94], indicating that post-transcriptional regulation has prominent roles in circadian rhythm control. Transcripts of numerous clock genes and clock output genes are modified by m^{6}A[95]. *METTL3* knockdown leads to reduced m^{6}A of two key clock genes, period circadian clock 2 (*PER2*) and aryl hydrocarbon receptor nuclear translocator like (*ARNTL*), which prolongs their nuclear retention and thereby the circadian period[95]. These results demonstrate how changes in mRNA metabolism can have

* The author's contribution to the studies presented in this chapter: Zebrafish Ythdf2 study: The author contributed equally to this work with X.W. and A.V.B., designed experiments with X.W., A.V.B., C.H. and R.K.H., performed biochemistry experiments with X.W., A.V.B. and H.S., performed cell biology experiments with A.V.B., X.W. and A.K., analyzed sequencing data with X.W. and Z.L., and wrote the manuscript with X.W., A.V.B. and C.H.

prolonged effects. Similarly, the cell cycle is an oscillating process that is functionally coupled with the circadian clock[96]. A notable shift in cell cycle duration following perturbation of m^6A in mRNAs of transcription factors was also reported in mouse ES cells[97]. More detailed studies are needed to fully elucidate the mechanism of cell cycle regulation by m^6A.

Cell differentiation is another process being impacted by m^6A factors, as both transcription factors driving stemness or differentiation can be methylated and subjected to m^6A regulation. In-depth mechanistic studies carried out in mouse embryonic stem cells (mESCs) have demonstrated that proper differentiation requires m^6A. Using homologous recombination to completely inactivate *Mettl3*, a study confirmed the crucial role of METTL3 in stem cell differentiation and detailed how m^6A may drive mESCs away from the pluripotent state[98]. Specifically, m^6A supports the timely transition from naive pluripotency to lineage commitment, potentially by facilitating the decay of naive pluripotency-promoting transcripts. Cell reprogramming, the reverse process of cell differentiation, was shown to be affected by METTL3 in the same study, in which the reprogramming of differentiated mouse epiblast stem cells to mESCs was blocked by inactivation of *Mettl3* early during development but facilitated by inactivation late during development. Zinc finger protein 217 (ZFP217) is partially responsible for stabilizing key pluripotency and reprogramming transcripts by inhibiting their METTL3-mediated methylation, thus promoting self-renewal of mouse ES cell and reprogramming of somatic cells[97].

Apart from these studies in cell lines demonstrating the vital functions of m^6A in cellular transitions, there is still a lack of report on m^6A impacting a certain process on the whole organism level. We thus chose zebrafish development as our model system to study the organismal functions of m^6A, as maternal RNAs are rapidly degraded during the early stage of zebrafish embryogenesis, hinting m^6A involvement. We have found that m^6A marks up to one-third of these maternal RNAs

and zebrafish Ythdf2 is responsible for the clearance of these m^{6}A-modified transcripts. Given these roles, we suspect that m^{6}A-dependent RNA decay acts as a general mechanism to initiate and facilitate the switching of cell states: clearing the genes products that govern the old cell state in order to enter the new one.

3.2 Results: m^{6}A regulates zebrafish reproduction and development process

During the early stages of embryogenesis, maternally inherited RNAs govern gene expression before relinquishing control to the zygotic genome. This changeover, termed maternal-to-zygotic transition (MZT), exists in all animal species[99,100]. During MZT, the massive maternal RNA population is rapidly cleared in two modes: by maternally supplied factors and by newly synthesized zygotic gene products[101]. In zebrafish, zygotically expressed *miR-430* regulates the clearance of a few hundred maternal mRNAs[102]. Codon usage has recently been shown to correlate with maternal RNA stability[103]. Yet factors and pathways in charge of the majority of maternal RNA decay remain elusive[104-106].

3.2.1 Transcriptome and methylome dynamics during zebrafish embryogenesis

We first studied the molecular characteristics of MZT by profiling early embryonic transcriptomes of zebrafish using mRNA sequencing (mRNA-seq) at five time points: 0, 2, 4, 6, and 8 hours post fertilization (h.p.f.). Based on their abundance changes over time, these transcripts were subjected to clustering analysis and divided into three superclusters we describe as maternal, semi-stable, and zygotic genes. Each supercluster comprises two subclusters with distinct trends (Figure 3.1A). The maternal supercluster (containing 5,211 genes) exhibited high expression levels at early time points and were degraded to very low levels afterward. These genes formed two subclusters: the first started to decrease early on from 1 h.p.f. (maternal-early; 2,167 genes), while the second declined rapidly from 4 h.p.f. onward (maternal-late; 3,044 genes). The expression of

genes in the zygotic supercluster (containing 2,472 genes) was non-existent initially but started to increase at later stages. The first subcluster was activated at approximately 3 h.p.f. (zygotic-early; 831 genes), and the second started expression after 4 h.p.f. (zygotic-late; 1,641 genes). The semi-stable supercluster (containing 3,948 genes), unlike the other two superclusters, featured relatively stable expression patterns, indicating a mixture of mRNAs from maternal and zygotic portions. One of the subclusters exhibited medium expression levels before further increasing after 4 h.p.f. (semi-stable-1; 1,527 genes), while the other subcluster showed high expression levels initially and slowly decreased in abundance starting from 4 h.p.f. (semi-stable-2; 2,421 genes). Collectively, the above clustering analysis highlights the dynamics of gene expression patterns and RNA methylation levels during zebrafish early development.

We then applied m^6A-seq and m^6A-CLIP-seq that identifies m^6A sites at a higher resolution to obtain m^6A profiles of zebrafish embryonic mRNAs (Figure 3.1B). Two methods identified m^6A sites with high overlapping ratios in more than 40% of gene transcripts (4,600 out of 11,631) during the studied time period (Figure 3.1C). Just as mRNA abundance shifted from maternally derived to zygotically derived at 4 h.p.f., so did the distribution of the m^6A modification shift from maternal RNAs to zygotic RNAs after 4 h.p.f. (Figure 3.1D and E). The measured m^6A sites showed an RRACU (R = G/A) consensus motif (Figure 3.1F) and were enriched near the start and stop codons (Figure 3.1G), resembling similar features previously reported in other organisms.

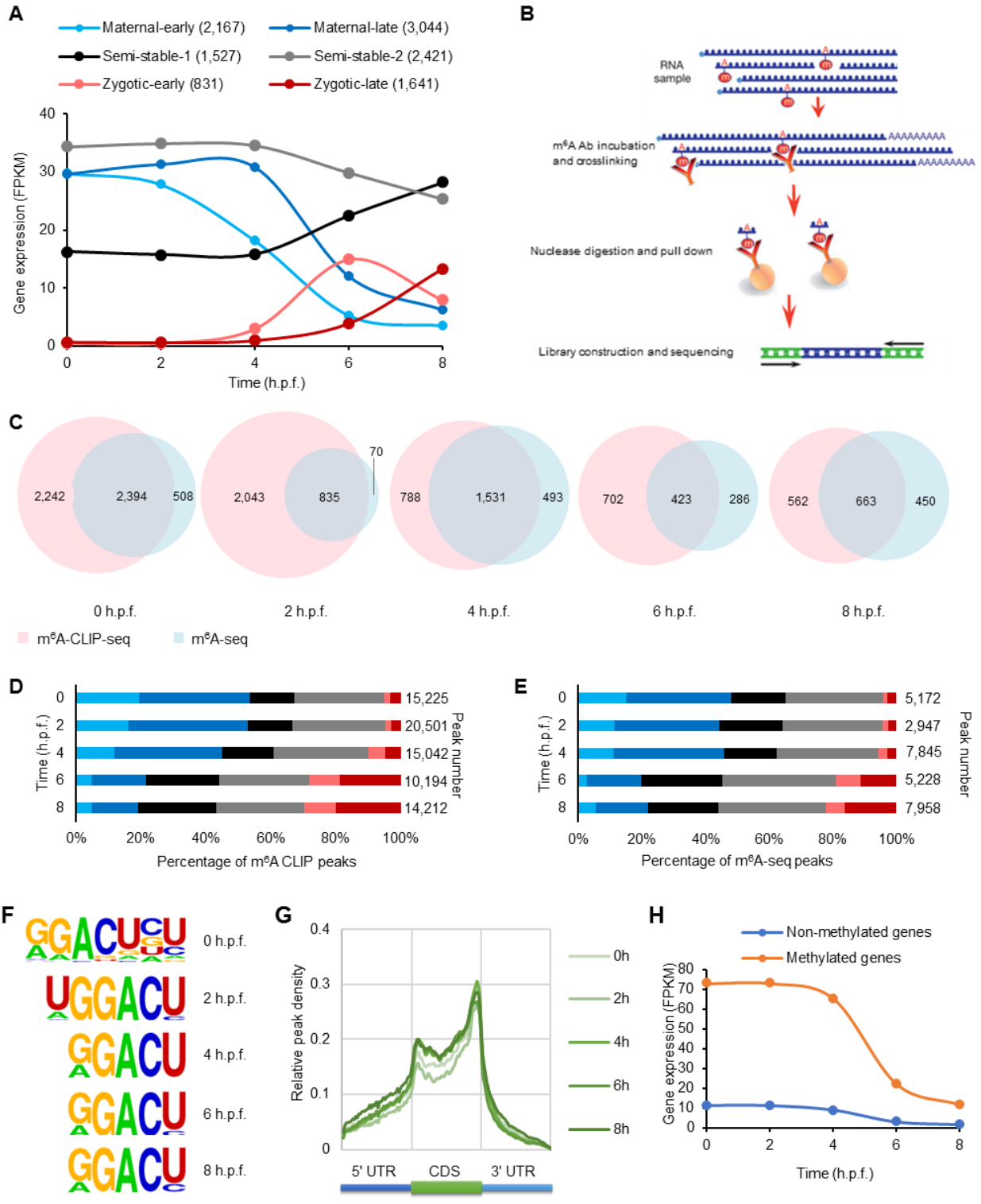


Figure 3.1 Characterization of gene expression and m^{6}A modification change in the early embryonic transcriptome of zebrafish

(**Figure 3.1, continued**)

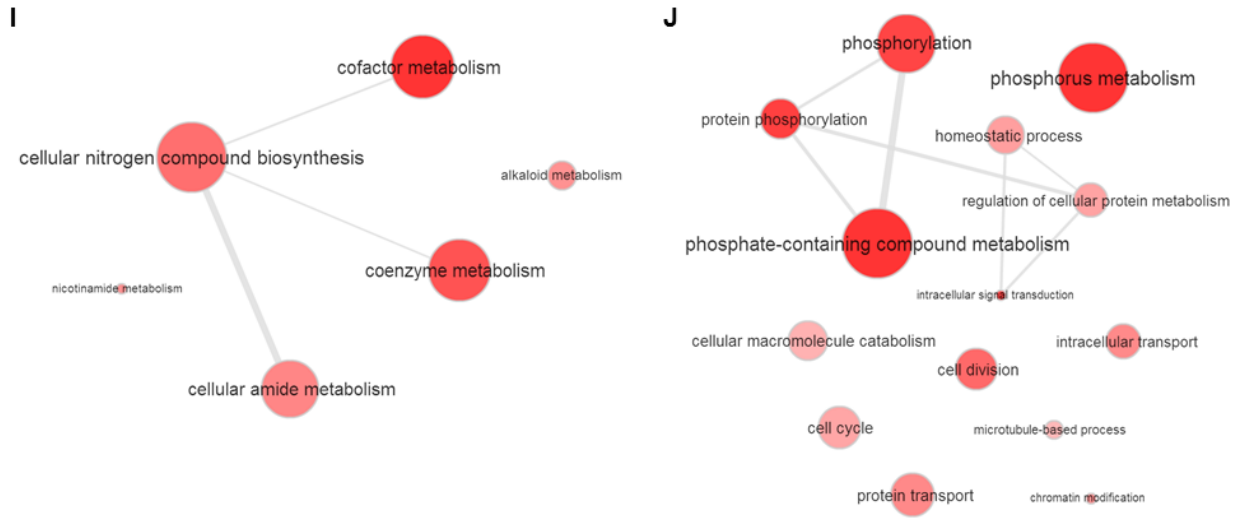


Figure 3.1 Characterization of gene expression and m^{6}A modification change in the early embryonic transcriptome of zebrafish
(A) Expression profiles of six gene groups clustered by their RNA abundance over time (0–8 h.p.f.) determined by RNA sequencing: maternal-early (degradation starting at 2 h.p.f), maternal-late (decrease after 4 h.p.f.), zygotic-early (onset at 4 h.p.f. and peak at 6 h.p.f.), zygotic-late (onset at 6 h.p.f.), and two semi-stable gene groups (relatively stable expression over time). (B) Workflow scheme of m^{6}A-CLIP-seq. (C) Venn diagrams show the overlap of two replicates of m^{6}A-modified transcripts determined by m^{6}A-seq and m^{6}A-CLIP-seq at five time points. (D, E) Distribution of m^{6}A sites quantified by the number of m^{6}A-CLIP-seq peaks (D) and m^{6}A-seq peaks (E) in the six gene groups over time (the same color code as A). (F) Consensus motif identified by HOMER with m^{6}A-CLIP peaks at five time points. (G) Metagene profiles depict the subtranscript distribution pattern of m^{6}A-sites (from m^{6}A-seq) within the zebrafish transcriptome. m^{6}A-seq peak signals are enriched after the start codon and before the stop codon. (H) Expression profiles of m^{6}A-modified and non-modified maternal transcripts over time. (I, J) Gene Ontology (GO) analysis of non-methylated maternal transcripts (I) and methylated ones (J).

Notably, more than 36% of maternal genes were m^{6}A methylated. The methylation ratio of maternal-late genes is 43.6%, higher than that of maternal-early ones at 25.9%. While the methylation ratios of both zygotic subclusters are similar at around 31%. We then divided the maternal supercluster genes into methylated and non-methylated groups. Compared to non-methylated maternal transcripts, methylated ones are present at a higher abundance with more marked decay at 4 h.p.f. (Figure 3.1H), and are enriched in phosphorus metabolism (for example,

nucleotide and ATP binding) and cell cycle regulation functions, whereas non-methylated ones featuring nitrogen metabolism (Figure 3.1I and J), suggesting distinctive kinetics and functional roles of methylated maternal transcripts.

3.2.2 Ythdf2 facilitates the rapid clearance of maternal transcripts

The fast-degrading profiles of methylated maternal transcripts strongly suggest the involvement of m^6A in the regulation of maternal clearance through the decay promotion function of Ythdf2. Zebrafish *ythdf2* is expressed ubiquitously throughout early embryogenesis (Figure 3.2A). We investigated the role of m^6A-dependent mRNA decay in embryogenesis by examining *ythdf2* loss of function at early stages of zebrafish development. Mating heterozygous mutants (*ythdf2*$^{+/-}$), we obtained the first generation of *ythdf2* homozygous mutant fish (*ythdf2*$^{-/-}$ F_1) harboring no apparent disabilities as adults. To isolate the maternal function of Ythdf2, we crossed female *ythdf2*$^{-/-}$ F_1 fish to male wild-type fish to produce maternal *ythdf2*$^{-/-}$ embryos (Figure 3.2B, red embryos) with the loss of *ythdf2* mRNA (Figure 3.2A).

We next closely monitored the developmental process of maternal *ythdf2*$^{-/-}$ embryos and wild-type ones and compared their differences over time (Figure 3.2C). We observed that both genotypes completed cell cycle 10 (during which the mid-blastula transition (MBT) is initiated) at 3 h.p.f. to reach the 1,000 (1 k)-cell stage and cell cycle 11 at 3.3 h.p.f. to reach the high stage. However, whereas wild-type embryos complete cell cycle 12 within the next 20 min to reach the oblong stage at 3.7 h.p.f., maternal *ythdf2*$^{-/-}$ embryos arrest at the high stage for around 45 min before completing cell cycle 12 and continuing cell divisions. Mutant embryos lose around another 30 min of developmental time between their pseudo-sphere and pseudo-50% epiboly stages, before eventually initiating gastrulation at approximately 7 h.p.f. (versus wild-type embryos at

approximately 5.5 h.p.f.). The pre-gastrulation developmental delay of maternal *ythdf2*$^{-/-}$embryos is maintained at least throughout the first 30 h of development, marked by the delayed onset of involuntary tail movement, delayed cessation of somitogenesis, and delayed pigmentation. This delay can be partially rescued by the injection of *ythdf2* mRNA, suggesting that these phenotypes are derived from the deletion of *ythdf2* (Figure 3.2D).

Quantifying the total amount of DNA pre- and post-MBT in the wild-type and mutant embryos, we found both a delay in DNA accumulation and a decrease in the amount of total DNA in mutant embryos, indicating fewer cell divisions within the studied time frame (Figure 3.2E). These *ythdf2* mutants also lack the recognizable, condensed phospho-Histone 3 (pH3)-positive nuclei characteristic of metaphase and anaphase, though the overall number of pH3-positive as well as proliferating cell nuclear antigen (PCNA)-positive (to mark interphase) cells is the same for each group (Figure 3.2F). This defect can be partially rescued by the injection of *ythdf2* mRNA, suggesting a block in late G2 phase or early M phase of the cell cycle in *ythdf2* mutants. These observations reveal that Ythdf2 is necessary for the normal progression of the cell cycle during MZT.

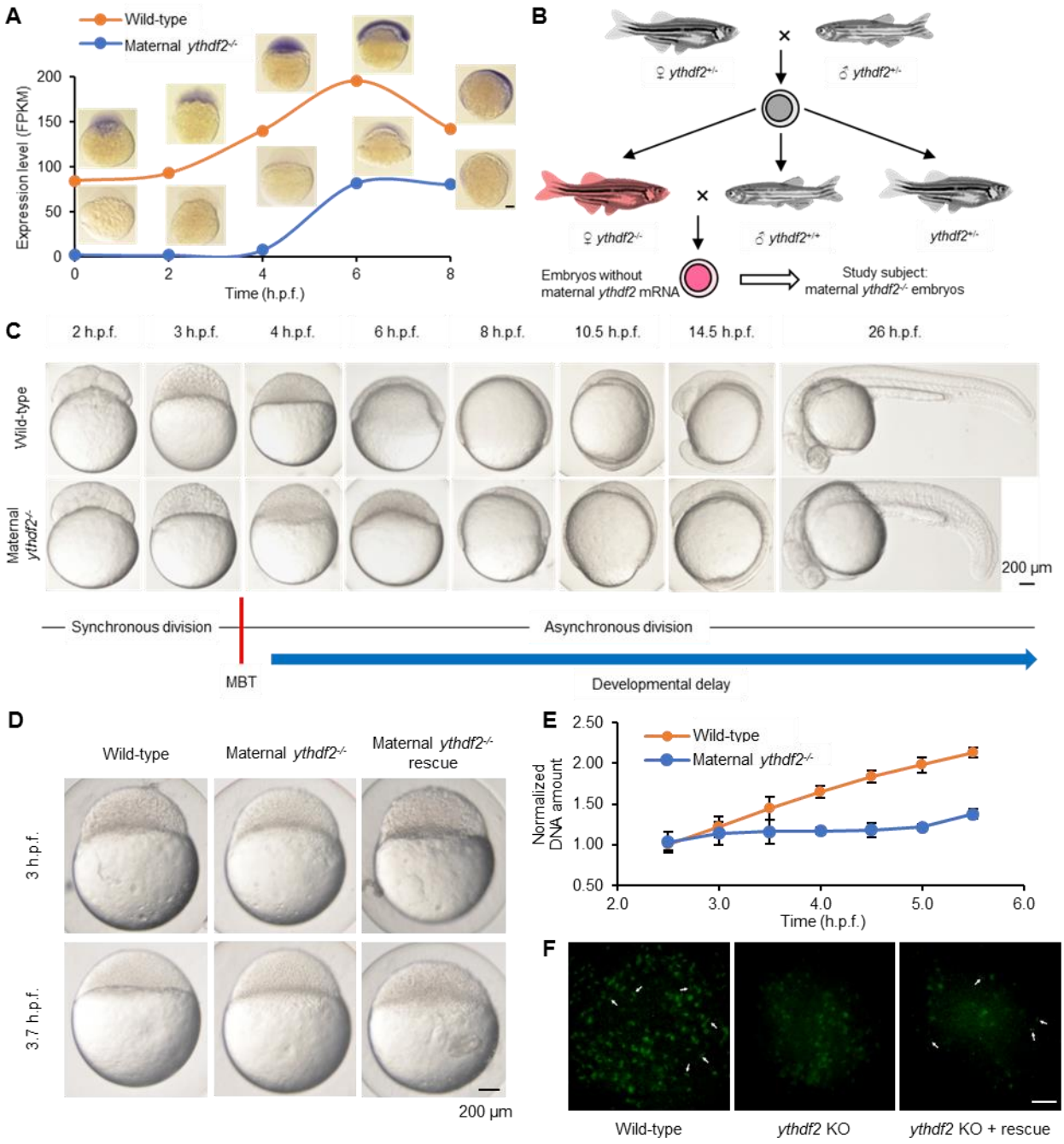


Figure 3.2 The deficiency of m^{6}A-binding protein Ythdf2 in zebrafish embryo led to a developmental delay

(**Figure 3.2, continued**)

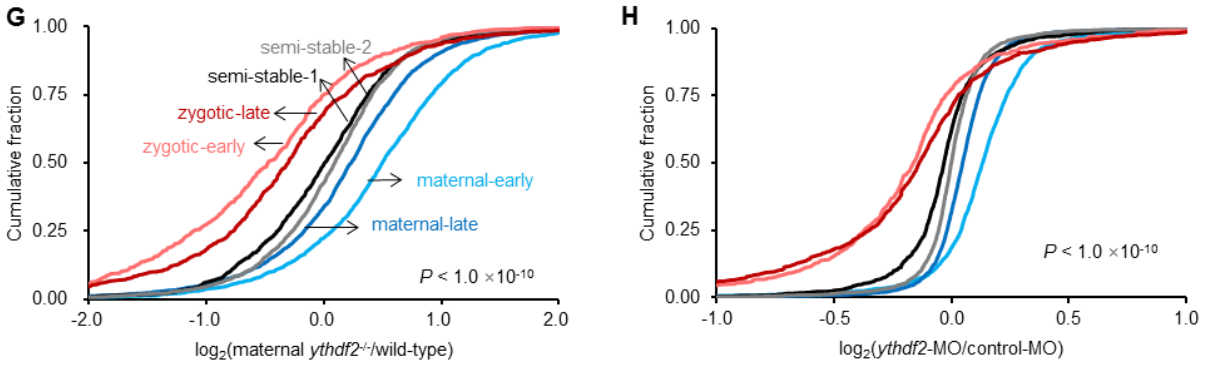


Figure 3.2 The deficiency of m^{6}A-binding protein Ythdf2 in zebrafish embryo led to a developmental delay

(A) Expression profile of *ythdf2* mRNA in wild-type and maternal *ythdf2*$^{-/-}$ embryos from crossing *ythdf2*$^{-/-}$ female and wild-type male fish. mRNA-seq and *in situ* hybridization revealed the mutation eliminated *ythdf2* mRNA before 4 h.p.f. (B) Crossing scheme to show the origin of maternal *ythdf2*$^{-/-}$embryos used in this study. (C) Time-matched bright field images of embryos showing that maternal *ythdf2*$^{-/-}$ embryos experience an early-stage delay maintained throughout early development. (D) Injection of *ythdf2* mRNA allows partial rescue of the *ythdf2* mutant phenotype. Maternal *ythdf2*$^{-/-}$ mutant embryos injected with 150 pg *ythdf2* mRNA showed an overall reduction in the cell cycle 12 delay compared to uninjected embryos of the same genotype. In these partially rescued embryos, an appreciable number of cells during cell cycle 12 divided at the same time as wild-type cells did. This partial rescue is nearly 100% penetrant ($n = 69$, four experiments). (E) Quantification of DNA amounts in wild-type and mutant embryos pre- and post-MBT (2.5–5.5 h.p.f.). Error bars, mean ± s.e.m., $n = 3$ (biological replicates). (F) Immunofluorescent microscopy of wild-type, *ythdf2* mutant, and *ythdf2* mRNA-injected mutant embryos fixed at 3 h 50 min post-fertilization and stained with anti-pH3 antibody to label nuclei in late G2 phase through M-phase of the cell cycle. The results showed that unlike wild-type embryos, maternal *ythdf2*$^{-/-}$ mutants are lacking in the tight, condensed pH3 nuclei staining characteristic of metaphase/anaphase, which can be partially rescued by the injection of *ythdf2* mRNA. Scale bar, 200 μm. (G and H) Cumulative distribution of the $\log_2$ fold changes of RNA expression at 4 h.p.f. for the six gene groups between maternal *ythdf2*$^{-/-}$ and wild-type (G), or *ythdf2*-MO-treated and control (H). Compared to the relatively unchanged curves of semi-stable gene groups (black and gray), the right shift of blue curves indicates the increase of maternal RNAs, and the left shift of red curves indicates the decrease of zygotic RNAs in *ythdf2* loss-of-function samples versus control. *P* values were calculated using two-sided Kruskal–Wallis test. Representative data shown from two independent experiments.

To further characterize the changes at the molecular level, we subsequently performed mRNA-seq on embryos spawned by either the wild-type or *ythdf2*$^{-/-}$female fish at the same five time points. Differential analysis revealed that the most notable change occurs around MBT. At

4 h.p.f., maternal transcripts were upregulated in embryos devoid of maternal *ythdf2* compared to wild-type, accompanied by downregulation of zygotic transcripts (Figure 3.2G). We also employed a translation-blocking morpholino (MO) specific for zebrafish *ythdf2* as an alternative approach to induce *ythdf2* loss of function. mRNA-seq results showed that, similar to the observation in maternal *ythdf2*$^{-/-}$embryos, a smaller yet still notable change occurred at 4 h.p.f. with upregulated maternal transcripts and downregulated zygotic transcripts in *ythdf2*-MO treated embryos (Figure 3.2H). These results indicate that Ythdf2 facilitates the clearance of maternal transcripts and that the failure to remove maternal transcripts hinders the activation of zygotic genes.

Ythdf2 appears to affect primarily methylated maternal mRNA at 4 h.p.f. because the total mRNA m^{6}A abundance increases by over 50% in *ythdf2*$^{-/-}$embryos compared to wild-type (Figure 3.3A), and over 60% of upregulated methylated transcripts are maternal genes at 4 h.p.f. (Figure 3.3B). We defined the most stringent Ythdf2 targets as the top 20% of upregulated genes differentially expressed in both maternal *ythdf2*$^{-/-}$ mutants and MO-treated samples and also present in both m^{6}A-seq and m^{6}A-CLIP-seq at 4 h.p.f. (Figure 3.3C). 135 genes met these stringent criteria and showed clustering of phosphorus metabolism, cell cycle process, and reproduction (Figure 3.3D). Among these genes, *mtus1a*, *mylipa*, *setdb1a*, *vps26a*, and *zgc:162879* have been reported as regulators of cell cycle, whereas *buc* and *tdrd1* are involved in gamete generation. Quantitative reverse-transcription PCR (RT–qPCR) of these gene transcripts showed pronounced retention in mutant embryos compared to wild-type (Figure 3.3E). Thereby the retention of methylated maternal transcripts (examples of methylation pattern shown in Figure 3.3F) in the absence of functional Ythdf2 may lead to the observed defects during early zebrafish development.

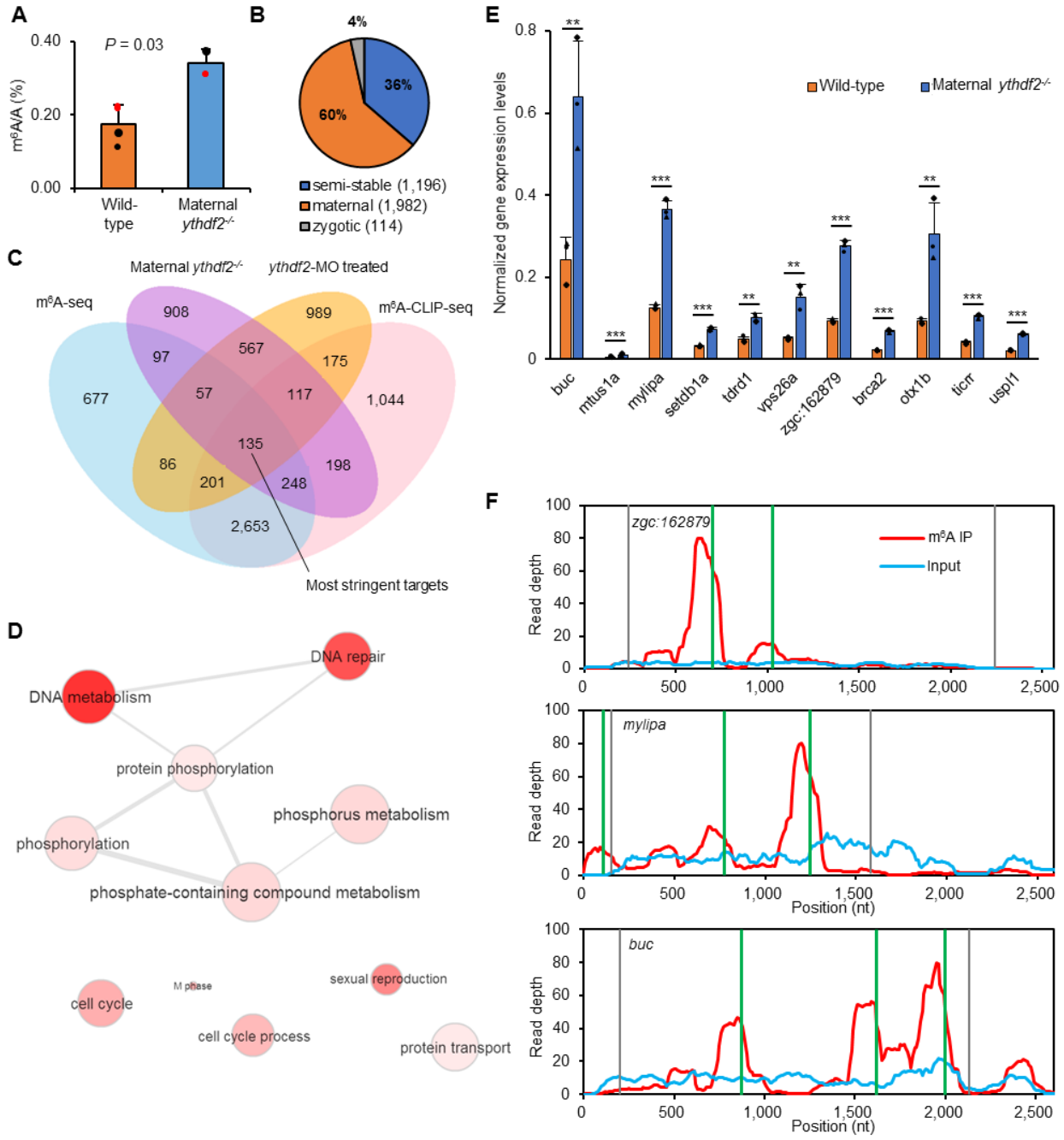


Figure 3.3 Identification and characterization of stringent targets of Ythdf2

(A) Quantification of the m^{6}A/A ratio of the total mRNA purified from maternal *ythdf2*$^{-/-}$and wild-type embryos by LC-MS/MS. *P* values were determined using two-sided Student's *t*-test for paired samples. Error bars, mean ± s.d. of two technical replicates (data points with the same color) from two biological experiments. (B) The composition of upregulated, methylated genes in maternal *ythdf2*$^{-/-}$ samples at 4 h.p.f. A total of 3,292 genes are both methylated and upregulated upon *ythdf2* loss of function, over 60% of which are maternal genes. (C) Venn diagram depicts the overlapping of m^{6}A-modified transcripts determined by m^{6}A-seq, m^{6}A-CLIP-seq, and differentially expressed transcripts between *ythdf2* loss of function samples and controls at 4 h.p.f.

(**Figure 3.3, continued**) The 135 intersected genes were defined as the most stringent Ythdf2 RNA targets. Without including the MO-treated result (but with *ythdf2*$^{-/-}$ result), we can estimate about 383 Ythdf2 target mRNAs. (D) GO analysis of the most stringent Ythdf2 target genes. (E) RT–qPCR validation of the relative abundance of representative Ythdf2 targets increased at 4 h.p.f. in maternal *ythdf2*$^{-/-}$ samples compared with wild-type. (F) Examples of Ythdf2 RNA targets harboring both m^{6}A-seq peaks and m^{6}A-CLIP-seq peaks. Coverage of m^{6}A immunoprecipitation (IP) and input fragments are indicated in red and blue, respectively. m^{6}A-CLIP-seq peaks are highlighted in green. Grey lines signify CDS borders.

To determine whether m^{6}A methylation is sufficient to trigger mRNA degradation in a Ythdf2-dependent manner, we produced green fluorescent protein (GFP) reporter mRNAs with or without m^{6}A (Figure 3.4A). These reporter mRNAs were injected into wild-type or maternal *ythdf2*$^{-/-}$ embryos and their decay kinetics were monitored by RT–qPCR. The results showed that degradation of the m^{6}A-modified reporter mRNA is faster than that of the unmethylated control during MZT in wild-type embryos, whereas in maternal *ythdf2*$^{-/-}$ mutant embryos, the clearance of the methylated reporter mRNA is noticeably hampered during MZT when compared to the unmethylated control (Figure 3.4B). Similar results were also observed in embryos co-injected with control or *ythdf2*-MO (Figure 3.4C). Further, fluorescence microscopy showed that the m^{6}A-modified GFP reporter produced consistently brighter fluorescence in maternal *ythdf2*$^{-/-}$ embryos compared to wild-type ones collected during the oblong stage (cell cycle 13 for most cells; 4.2 h.p.f. for mutant embryos and 3.8 h.p.f. for wild-type embryos) (Figure 3.4D). Maternally deposited *ythdf2* is present in the embryo before any zygotic transcription has taken place (Figure 3.2A), and loss-of-function maternal mutant alleles exhibit the observed cell cycle pause during MBT and delayed clearance of methylated maternal transcripts. Together, these results further indicate that m^{6}A methylation negatively affects the stability of target mRNAs and subsequent protein synthesis in zebrafish early embryos and that Ythdf2 is a maternally supplied factor to direct maternal mRNA clearance.

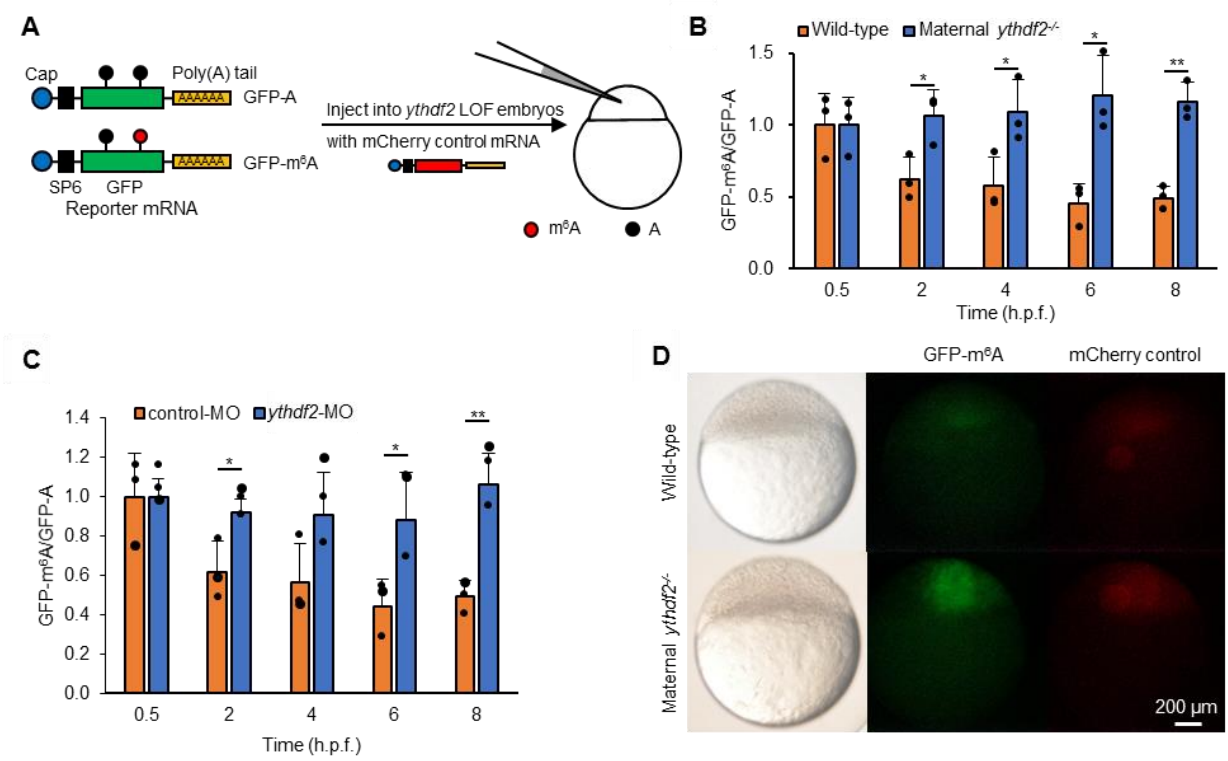


Figure 3.4 Ythdf2 is required for m^{6}A-dependent RNA decay
(A) The design of the reporter RNA. m^{6}A-modified (GFP-m^{6}A) or non-modified GFP mRNAs (GFP-A) were prepared and co-injected with mCherry control mRNA into one-cell stage embryos. LOF, loss of function. (B) The *in vivo* degradation of GFP-m^{6}A mRNAs was faster in wild-type embryos but slightly slower in maternal *ythdf2*$^{-/-}$ embryos than GFP-A mRNAs. mRNA abundance was determined by RT–qPCR and normalized to 0.5 h.p.f. values. (C) The *in vivo* degradation of GFP-m^{6}A mRNAs was faster than GFP-A mRNAs when co-injected with control MO. By contrast, GFP-m^{6}A mRNAs displayed an indistinguishable decay rate from GFP-A mRNAs when co-injected with *ythdf2* MO. The abundances of GFP-m^{6}A and GFP-A mRNAs were determined by RT–qPCR. Error bars, mean ± s.d., $n = 3$ (technical replicates). *P* values were determined using two-sided Student's *t*-test for two samples with equal variance. (D) Comparison of GFP-m^{6}A expression levels in wild-type and maternal *ythdf2*$^{-/-}$ embryos at the oblong stage of development (stage-matched embryos, 3.8 h.p.f. for wild-type and 4.2 h.p.f. for mutants). The m^{6}A-modified GFP reporter RNA produced consistently higher GFP protein levels in maternal *ythdf2*$^{-/-}$embryos versus wild-type embryos. Quantification of the overall phenomena: 61 out of 72 embryos (85%) for maternal *ythdf2*$^{-/-}$ and 59 out of 85 (70%) for wild-type in two experiments. Error bars, mean ± s.d., $n = 3$ (technical replicates). *P* values were determined using two-sided Student's *t*-test for two samples with equal variance.

3.2.3 Co-regulation of mRNA degradation by Ythdf2 and miR430 pathways

The existence of both maternally-driven and zygotically-driven modes of maternal mRNA clearance within animal species is a fascinating but cryptic phenomenon. In zebrafish, the miR-

430 pathway is known to facilitate maternal transcripts degradation and may overlap with the Ythdf2-based regulation. miR-430 is responsible for the translational inhibition and clearance of a few hundred zebrafish maternal mRNAs[1,2] but is itself transcribed zygotically and dependent on the maternally provided pluripotency-inducing factors, Nanog, Pou5f3, and Sox19b for its transcription[3].

We investigated potential overlap between the maternal Ythdf2- and the zygotic miR-430-based pathways and found that: i) The targets of miR-430[1] and Ythdf2 exhibit a significant overlap (37%, $p = 2.2\times10^{-22}$, Fisher's Exact Test, Figure 3.5A), and the common targets are mainly clustered in functions related to cell cycle regulation and intracellular transport (Figure 3.5B). ii) The specific targets of miR-430 decayed later than both the specific targets of Ythdf2 and targets in common between Ythdf2 and miR-430 in wild-type zebrafish embryos, and these specific targets of miR-430 were also less affected upon *ythdf2* knockout (Figure 3.5C–F). The decay of common targets occurs the earliest, suggesting the necessity of timely removal of these critical developmental factors with parallel and/or redundant pathways. Further, the expression of these common targets doesn't "rebound" as much at 8 h.p.f. compared to 6 h.p.f. as the other two groups, which may indicate they are not pleiotropic and do not need to be transcribed appreciably at future times. iii) Maternal *ythdf2*$^{-/-}$ embryos showed similar delays in zygotic genome activation to Nanog+Sox19b+Pou5f3 LOF embryos at 4 h.p.f. (Figure 3.5G) and the disruption of the two pathways exhibits similar developmental arrest phenotypes[3]. According to our aforementioned definition (the most stringent Ythdf2 targets are the top 20% of upregulated genes differentially expressed in both maternal *ythdf2*$^{-/-}$ mutants and MO-injected samples and also present in both m^{6}A-seq and m^{6}A-CLIP-seq at 4 h.p.f.), *nanog*, *pou5f3*, and *sox19b* transcripts are not stringent Ythdf2 RNA targets. By checking the correlation of mRNA-seq data upon *ythdf2* LOF and the

published mRNA-seq data with combined loss of all three factors[3], we found that zygotic gene expression is significantly decreased upon the removal of the respective regulators in both datasets (Figure 3.5G). In addition, we observed that *nanog* and *pou5f3* transcripts were underrepresented in *ythdf2*$^{-/-}$ samples at 4 h.p.f. These data suggest that both Ythdf2- and miR430-mediated pathways regulate the clearance of critical maternal transcripts and may function to facilitate activation of the zygotic genome through promoting the maternal factors required for miR-430 expression.

Such co-existing mechanisms for the clearance of maternal transcripts have also been indicated in other species such as pathways mediated by pumilio[107] and brain tumor[108] in *Drosophila* and poly(C)-binding proteins in *C. elegans*[109]. Together, these factors suggest the importance of timely, robust removal of maternal transcripts by overlapping yet temporally distinct mechanisms for proper development. Both pathways account for only a portion of total maternal mRNA clearance in zebrafish, with other pathways and mechanisms remain to be discovered.

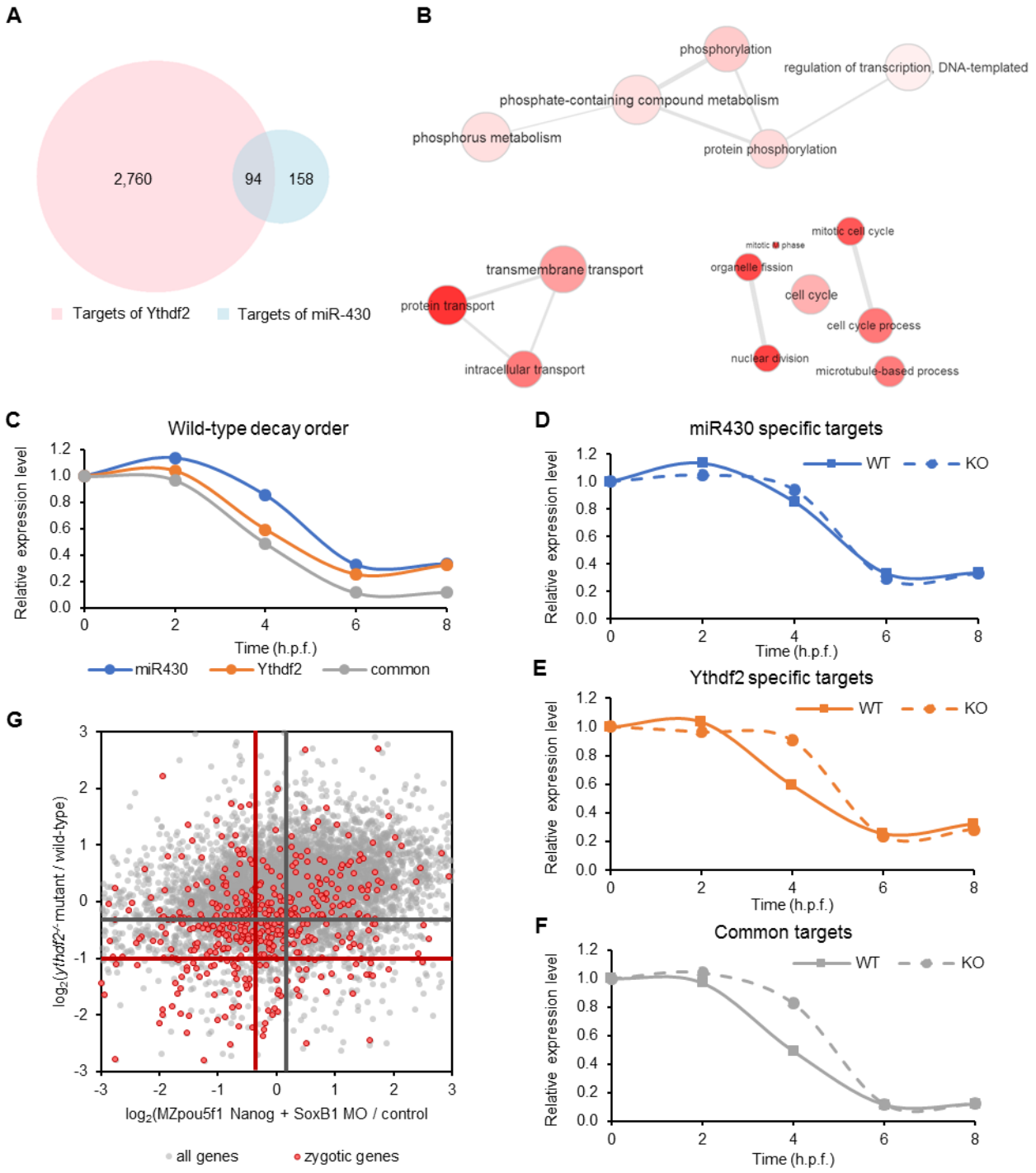


Figure 3.5 Target overlap of Ythdf2 and miR-430 regulation pathways

(A) Target overlap between Ythdf2 and miR-430. Genes with more than 1.5-fold upregulation upon removal of maternal Ythdf2 or miR-430 were used as corresponding targets in the analysis. (B) GO analysis of the common target genes of Ythdf2 and miR-430. (C–F) Differential temporal regulation of the targets of miR430 and Ythdf2. In wild-type embryos, common targets are decayed first, followed by Ythdf2-specific targets, and then miR430-specific targets (C). Upon maternal *ythdf2* knockout, the decay of the miR430-specific targets showed minimal delay (D),

(**Figure 3.5, continued**) whereas the Ythdf2-specific targets and the common targets show more noticeable decay delays; the latter finding is consistent with Ythdf2 as a maternally supplied factor able to degrade maternal mRNAs earlier in time (E and F). Relative expression level is calculated by the median of the expression levels of all genes and normalized to the value at 0 h.p.f., respectively. (G) Maternal *ythdf2*$^{-/-}$ embryos show a similar zygotic genome activation delay as Nanog, Sox19b, and Pou5f3 loss-of-function embryos at 4 h.p.f. The $\log_2$ fold change of the mRNA-seq data of maternal *ythdf2*$^{-/-}$ mutants and the published mRNA-seq data with combined loss of *nanog*, *sox19b* (*soxB1*), and *pou5f3* (*oct4* homolog) were plotted in the same diagram. Zygotic genes (red dots) were separately plotted out from all genes (gray dots). Intersections of crosslines mark the median points of each group of genes.

3.2.4 Ythdf2 affects zebrafish spermatogenesis

From our study of the function of Ythdf2 during zebrafish embryogenesis, we discovered that in addition to its role in maternal clearance, Ythdf2 seemed to also have a separate paternal function. We crossed fish with different *ythdf2* genotypes and counted embryos with either wild-type or developmentally delayed phenotypes (Figure 3.6A). The results showed that maternal genotype is the sole determinant for developmental delay phenotypes in the offspring: wild-type females produce no delayed embryos while *ythdf2*$^{-/-}$ females produce ~95% developmentally delayed embryos, regardless of paternal genotype, confirming our previous conclusion of Ythdf2's maternal effect. However, we have also observed marked paternally-associated embryonic lethality and/or sterility from *ythdf2* homozygous mutant males (Figure 3.6A and B): *ythdf2*$^{-/-}$ males produce on average 70% progeny whose development never progresses past the 1-cell stage. This phenotype occurs regardless of whether *ythdf2*$^{-/-}$ males are mated to wild type, *ythdf2*$^{+/-}$, or *ythdf2*$^{-/-}$ females. Further, the percentage of embryos suffering from cell cycle 12 delay during the maternal-zygotic transition is identical between offspring from m(*ythdf2*$^{-/-}$) × p(*ythdf2*$^{+/+}$) or m(*ythdf2*$^{-/-}$) × p(*ythdf2*$^{-/-}$) matings (Figure 3.6A). Thus, we infer that Ythdf2 must be pleiotropic, acting both to degrade maternal transcripts and to produce normal sperm. Importantly, the surviving embryos from matings in which the paternal genotype is *ythdf2*$^{-/-}$ (regardless of the maternal genotype) appear normal, survive to adulthood with no obviously disabilities, and harbor

a transcriptome that is essentially indistinguishable from wild-type after the maternal phase of Ythdf2 action has ended. We, therefore, consider the maternal effect and the paternal effect of Ythdf2 separate gene functions, ones that arise in different animals (fathers versus offspring) in different biological processes (spermatogenesis or fertilization versus embryonic development) and can segregate independently of one another.

We then further characterized the defects suffered by *ythdf2*$^{-/-}$ sperms from both phenotypic and molecular levels. From the examination of the undeveloped oocytes/embryos from matings with *ythdf2*$^{-/-}$ male fish, we found that the paternal nuclear contribution to the oocyte appears aberrant. As opposed to the compact collection of DAPI-stained chromosomes found in the oocyte pronucleus, the sperm pronucleus appears to contain many small chromosomal fragments, and full pronuclear fusion does not seem to occur. Such defects suggest that true fertilization is not achieved by most sperm from *ythdf2*$^{-/-}$ males, despite the fact that mutant sperm is able to penetrate the oocyte. To further identify the differences in molecular contents of sperms from wild-type and *ythdf2*$^{-/-}$ male fish, we harvested mature sperms from both genotypes and performed high-throughput sequencing analysis of their chromatin accessibility by ATAC-seq[110] and RNA content by RNA-seq. ATAC-seq results showed that the mutant sperm has a more closed chromatin structure with fewer ATAC peaks detected. Both mRNA-seq and sRNA-seq revealed that, opposing to the notion that sperms do not carry much RNA contents, a large number of fragmented and intact mRNA, rRNA, and tRNA, as well as primary and mature miRNA, are detected in sperms. The expression profiles of these RNA species from both groups are notably different (Figure 3.6C and D). These results indicate that the defects of *ythdf2*$^{-/-}$ sperms may be originated from the misregulation of the chromatin structure of sperm DNA or the aberrant expression and incorporation of functional RNA species into the mature sperm during spermatogenesis. Future

work investigating the molecular targets of Ythdf2 in the testes of *ythdf2*$^{-/-}$ male fish will hopefully provide more information about the exact mechanism of m^{6}A and Ythdf2 regulating zebrafish spermatogenesis.

A

Female x male genotype	MBT-period delay (of non-arrested ones)	1-cell arrest lethality
WT x WT	0.0%	5.5%
WT x KO	2.0%	70.4%
KO x WT	93.3%	5.0%
KO x KO	97.8%	73.5%

B

Female x male genotype	1-cell arrest lethality	number of embryos scored
WT x HT	26.7%	770
HT x HT	31.3%	276
KO x HT	32.7%	363
HT x KO	70.3%	598

WT, wild-type; HT, heterozygous; KO, knockout

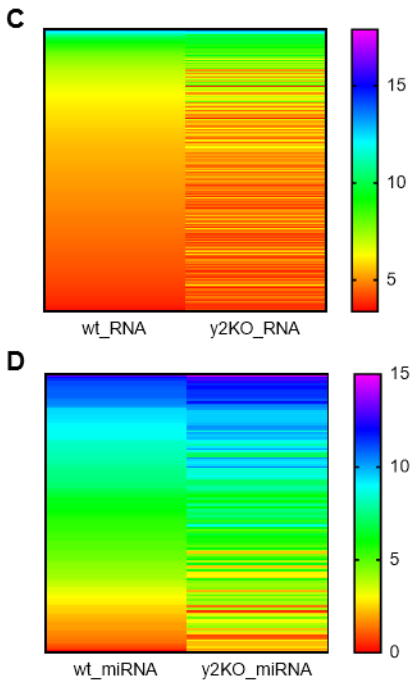


Figure 3.6 The paternal effect of Ythdf2 affecting sperm function
(A) Table showing the percentage of embryos suffering from MBT-period developmental delay and lethality for crosses between different homozygous genotypes. Homozygous mutant females produce ~95% embryos that suffer from MBT-period developmental delay, regardless of paternal genotype. Homozygous mutant males produce on average ~70% embryos that do not develop past the one-cell stage, regardless of maternal genotype. These two phenotypes can segregate independently from one another and reveal that Ythdf2 is pleiotropic in zebrafish development. Each row is tabulated from 6 to 18 crosses, scoring between 365 and 1,026 embryos, and using randomly selected males and females from each genotype from at least 4 fish of each sex. (B) Table showing additional crosses to characterize the paternal-associated lethality and/or sterility of *ythdf2* loss of function. Heterozygous males produce about half the number one-cell arrested offspring as homozygous mutant males, perhaps indicating a function for Ythdf2 in later gametogenesis; this arrest occurs without respect to maternal genotype. Data for each row is tabulated from 5 crosses each, using randomly selected males and females from each genotype from at least 4 fish of each sex. (C and D) Heatmaps of gene expression profiles of mRNA and primary miRNA (C) and mature miRNA (D). The log_2 fold of normalized reads density of each gene corresponding to a given color on the heatmap can be elucidated from the color bar. Data calculated from the mean values of two biological replicates.

3.3 Conclusion and Discussion: m^{6}A impacts cellular state transition

In conclusion, our work in zebrafish revealed that m^{6}A mRNA methylation and the m^{6}A reader protein Ythdf2 act in a previously hypothesized but unidentified maternal mode to regulate zebrafish maternal mRNA clearance in order to control development through MZT (Figure 3.7A). This study supports a vital function of m^{6}A in controlling the dynamics of global gene expression during development, in which hundreds to thousands of transcripts, each with their own distinct metabolism, need to be grouped and coordinated for translation and decay.

The collective influence of m^{6}A affects cell state transition most probably through the regulation of a subset of key transcription factors that determine cell fate. m^{6}A has been found on transcripts of numerous transcription factors that control cellular state and lineage commitment. Such transcription factors include the core pluripotency factors NANOG and Yamanaka reprogramming factors[111] — the transcription factors POU domain, class 5, transcription factor 1 (POU5F1; also known as OCT4), SOX2, MYC and Krueppel-like factor 4 (KLF4) — that are necessary and sufficient to induce the formation of pluripotent stem cells[97,112,113]. The cellular composition of expressed transcription factors can either maintain cell state or promote cell differentiation. By facilitating the downregulation of transcripts encoding such dominant transcription factors, the barrier to cell-state transition could be tuned by changes in mRNA m^{6}A levels (Figure 3.7B). Thus, inside cells in which methylated transcripts maintain pluripotency (or any other cell state), reducing m^{6}A levels may increase the barrier for differentiation by reducing the decay and prolonging the lifetime of these mRNAs. Conversely, in cells in which methylated transcripts drive a transition to a new cell state, reducing m^{6}A levels could promote cell differentiation, whereas increased overall methylation could induce stemness and suppress cell differentiation. In addition to shaping cell states during cell differentiation and development, the

mechanism enabled by m^{6}A and its regulators may be essential for transcriptome switching in other highly orchestrated transition events from diverse biological systems.

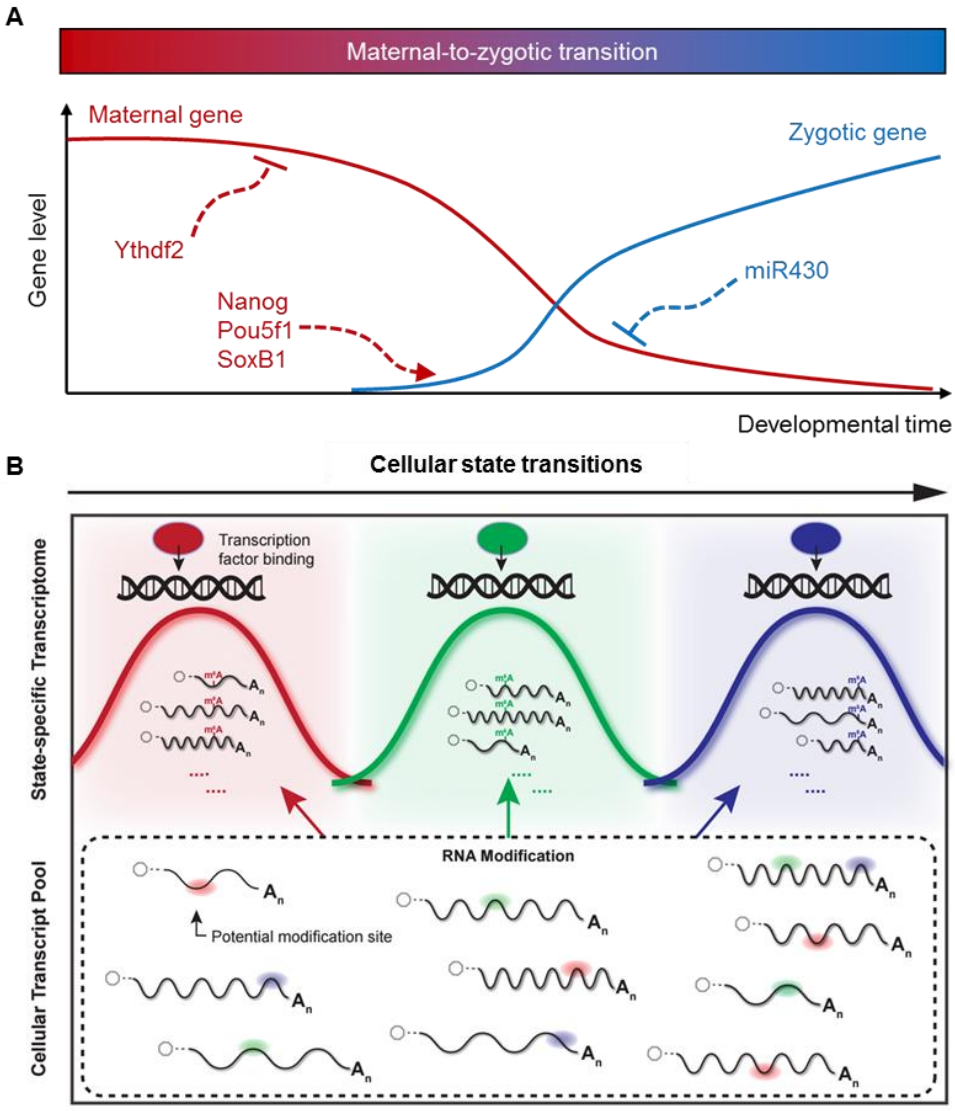


Figure 3.7 Mechanism of m^{6}A impacting cellular state transition events
(A) Proposed mechanism of Ythdf2 regulating zebrafish MZT: Ythdf2-mediated RNA decay ensures the timely clearance of maternal RNA and the subsequent activation of zygotic gene expression, then coordinates with miR430 to ensure the complete removal of maternal transcripts. (B) Proposed model of m^{6}A driving the transition between cellular states. Developmental programs require timely switching of the cellular transcriptome to bring about phenotypic changes. Master transcription factors (TFs) largely define the cellular pool of mRNA. These TFs activate transcription of tens to hundreds of transcripts at different stages of development. mRNA m^{6}A methylation is a mechanism to group these distinct sets of transcripts together for coordinated translation and decay, thereby facilitating the turnover of transcriptomes and triggering the transition of cellular states.

3.4 Materials and Methods

Zebrafish maintenance

Male heterozygous *ythdf2*$^{+/-}$ fish in the *AB background was custom made by ZGeneBio. TALEN mutagenesis was performed to mutate *ythdf2* (Ensembl ENSDART00000127043) with L1 recognition sequence 5′-GGACCTGGCCAATCCCC-3′, R1 recognition sequence 5′-GGCAC AGTAATGCCACC-3′, and spacer sequence 5′-TCCCAATTCAGGAATG-3′. Purchased fish were outcrossed to in-house wild-type *AB fish. Embryos were obtained from natural crosses, were raised under standard conditions, and were staged according to literature[114]. Embryos were reared at 28.5 °C and all experiments and observations were performed as close to this temperature as possible. Fish lines were maintained in accordance with AAALAC research guidelines, under a protocol approved by the University of Chicago IACUC (Institutional Animal Care & Use Committee).

In situ hybridization

The open reading frame of zebrafish *ythdf2* was purchased from Open Biosystems (clone 5601005) and subcloned into a pCS2+ vector using restriction enzyme sites of BamHI and XhoI. The resulting vector was linearized by HindIII and used as a template for *ythdf2* probe preparation. Antisense digoxigenin (DIG) RNA probes were generated by *in vitro* transcription using standard reagents and methods. *In situ* hybridization protocol was followed essentially as previously reported[115]. All experiments were repeated at least once from biological samples.

Morpholino and mRNA injection

Control and *ythdf2* morpholinos (5′-TGGCTGACATTTCTCACTCCCCGGT-3′) were obtained from Gene Tools (Oregon). 3 ng of either control or gene-specific morpholino was injected into *AB wild-type embryos at the one-cell stage.

GFP and mCherry were subcloned into pCS2+ vectors and linearized by NotI. GFP-m^{6}A, GFP-A, and mCherry-capped and polyadenylated mRNA were generated by *in vitro* transcription using mMessage mMachine SP6 kit (Thermo Fisher) and Poly(A) tailing kit (Thermo Fisher) according to the manufacturer's protocol. Products were purified with the MEGAclear transcription clean-up kit (Thermo Fisher) and used for injections directly. For GFP-m^{6}A, we spiked 6 nmol m^{6}ATP into the 100 nmol ATP supplied in the transcription reaction, in order to ensure that less than 0.3% of GFP mRNAs are without m^{6}A on average. (GFP mRNA is 942 nt; each mRNA has 1.89 m^{6}A on average.) 35 pg of either GFP reporter mRNA and 10 pg of mCherry mRNA were injected together in 1.25 nl into embryos at the one-cell stage. *ythdf2* mRNA containing the *ythdf2* 5′ UTR and a 3′ Flag tag, which was used to rescue the mutant phenotype and validate the knockdown efficiency of *ythdf2* MO, was constructed in pCS2+ vector (forward primer: 5′-CGTACGGATCCTGTCTGATCTGCAGCTGTAG-3′; reverse primer: 5′-CGATGCT CGAGTTACTTGTCATCGTCGTCCTTGTAATCTATTCCAGATGGAGCAAGGC-3′) and prepared in the same way as mCherry mRNAs.

Antibodies

Antibodies used in this study are listed below in the format of name (application; catalogue number; supplier): mouse anti-Flag HRP conjugate (Western; A5892; Sigma), rabbit anti-m^{6}A (m^{6}A-seq and m^{6}A-CLIP-seq; 202003; Synaptic Systems), rabbit anti-histone H3 (IF; ab5176; Abcam), and anti-rabbit Alexa Fluor 488 (IF; ab150077; Abcam).

Microscopy

All images were observed with a Leica MZFLIII microscope and captured with a Nikon D5000 digital camera using Camera Control Pro (Nikon) software. For fluorescent microscopy, standard ET-GFP and TXR LP filters (Leica) were used. For bright field imaging of live embryos,

only saturation was adjusted and was adjusted identically for all images. For fluorescent imaging of live embryos, no image processing was performed. For fluorescent imaging of fixed embryos, contrast and exposure were adjusted for all to obtain the lowest amount of background while preserving the morphology of all visible nuclei. All experiments were repeated at least once from biological samples.

DNA quantification

To compare the total amount of DNA in wild-type and mutant embryos at different time points during the MZT, 10 embryos per time point per condition were dechorionated and pipetted into standard DNA lysis buffer. The number of embryos in each tube was counted twice to ensure uniformity. Proteinase K was added to 100 μg ml^{-1} and the embryos were incubated for 4 h at ~55 °C with occasional mixing. Proteinase K was inactivated by a 10-min incubation at 95 °C and the DNA was then phenol-chloroform-extracted, ethanol-precipitated, and resuspended in 100 μl Tris (pH 8.5) and 1 mM EDTA using standard procedures. Double-stranded DNA content was measured with NanoDrop. Three biological replicates (comprised of the offspring of three different fish mating pairs of the appropriate genotype) were measured for each time point for both the control and experimental samples. Biological replicates were averaged together to determine the average DNA amount per time point per genotype and to compute standard errors of the mean. All DNA values were normalized to that of wild-type embryos at 2.5 h.p.f.

Western blot

Embryos were collected into standard 2× protein sample buffer with added β-mercaptoethanol and protease inhibitors and immediately put on ice for a few minutes. The embryo mixtures were carefully but thoroughly pipetted up and down to dissolve and homogenize the embryos, and then samples were heated at 95 °C for 5 min and frozen at −80 °C. Before use,

samples were again heated for 5 min and then centrifuged at 12,000 r.p.m. to remove debris. Supernatants were loaded into a 10-well, 1.5 mm Novex 4–20% Tris-Glycine Mini Protein Gel (Thermo Fisher) with 6 embryos per well. The gel was transferred onto a nitrocellulose membrane using iBlot2 gel transfer system (Thermo Fisher) set to P3 for 7 min with iBlot2 mini gel transfer stacks (Thermo Fisher). Membranes were blocked in 5% BSA, 0.05% Tween-20 in PBS for 1 h, and then incubated overnight at 4 °C with anti-Flag–HRP conjugate (Sigma) diluted 1:10,000 in 3% BSA. Proteins were visualized using the SuperSignal West Pico Luminol/Enhancer solution (Thermo Fisher) in FluorChem M system (ProteinSimple).

RNA isolation

mRNA isolation for LC-MS/MS: total RNA was isolated from zebrafish embryos with TRIzol reagent (Invitrogen) and Direct-zol RNA MiniPrep kit (Zymo). mRNA was extracted by removal of contaminating rRNA using RiboMinus Eukaryote Kit v2 (Thermo Fisher) for two rounds. Total RNA isolation for RT–qPCR: we followed the instruction of Direct-zol RNA MiniPrep kit (Zymo) with DNase I digestion step. Total RNA was eluted with RNase-free water and used for RT–qPCR directly.

LC-MS/MS

100–200 ng of mRNA was digested by nuclease P1 (2 U) in 25 μl of buffer containing 10 mM of NH_4OAc (pH 5.3) at 42 °C for 2 h, followed by the addition of NH_4HCO_3 (1 M, 3 μl, freshly made) and alkaline phosphatase (0.5 U). After an additional incubation at 37 °C for 2 h, the sample was diluted to 50 μl and filtered (0.22 μm pore size, 4 mm diameter, Millipore), and 5 μl of the solution was injected into LC-MS/MS. Nucleosides were separated by reverse-phase ultra-performance liquid chromatography on a C18 column with on-line mass spectrometry detection using an Agilent 6410 QQQ triple-quadrupole LC mass spectrometer in positive electrospray

ionization mode. The nucleosides were quantified by using the nucleoside to base ion mass transitions of 282 to 150 (m^6A), and 268 to 136 (A). Quantification was performed in comparison with the standard curve obtained from pure nucleoside standards running on the same batch of samples. The ratio of m^6A to A was calculated on the basis of the calibrated concentrations[9].

m^6A-seq

Total RNA was isolated from fish embryos collected at different time points with TRIzol reagent and Direct-zol RNA MiniPrep kit. For each time point, ~200 embryos were collected to ensure RNA yield and that samples were representative. mRNA was further purified using RiboMinus Eukaryote Kit v2. RNA fragmentation was performed by sonication at 10 ng μl^{-1} in 100 µl RNase-free water using Bioruptor Pico (Diagenode) with 30 s on/off for 30 cycles. m^6A-immunoprecipitation (IP) and library preparation were performed according to the previous protocol[17]. Sequencing was carried out on Illumina HiSeq 2000 according to the manufacturer's instructions.

m^6A-CLIP-seq

Additional high-throughput sequencing of zebrafish methylome was carried out using a modified m^6A-seq method, which is similar to previously reported methods[116,117]. Briefly, total RNA and mRNA were purified as previously described for m^6A-seq. Purified mRNA (1 µg) was mixed with 2.5 µg of affinity purified anti-m^6A polyclonal antibody (Synaptic Systems) in IPP buffer (150 mM NaCl, 0.1% NP-40, 10 mM Tris-HCl (pH 7.4)) and incubated for 2 h at 4 °C. The mixture was subjected to UV-crosslinking in a clear flat-bottom 96-well plate (Nalgene) on ice at 254 nm with 0.15 J for 3 times. The mixture was then digested with 1 U μl^{-1} RNase T1 at 22 °C for 6 min followed by quenching on ice. Next, the mixture was immunoprecipitated by incubation with protein-A beads (Invitrogen) at 4 °C for 1 h. After extensive washing, the mixture was

digested again with 10 U μl^{-1} RNase T1 at 22 °C for 6 min followed by quenching on ice. After additional washing and on-bead end-repair, the bound RNA fragments were eluted from the beads by proteinase K digestion twice at 55 °C for 20 and 10 min, respectively. The eluate was further purified using RNA clean and concentrator kit (Zymo Research). RNA was used for library generation with NEBNext multiplex small RNA library prep kit (NEB). Sequencing was carried out on Illumina HiSeq 2000 according to the manufacturer's instructions.

mRNA-seq

Total RNA was isolated from wild-type and mutant fish embryos collected at different time points with TRIzol reagent and Direct-zol RNA MiniPrep kit. For each time points, ~20 embryos were collected to ensure RNA yield and that samples were representative. mRNA was further purified using RiboMinus Eukaryote Kit v2. RNA fragmentation was performed using Bioruptor Pico as described previously. Fragmented mRNA was used for library construction using TruSeq stranded mRNA library prep kit (Illumina) according to manufacturer's protocol. Sequencing was carried out on Illumina HiSeq 2000 according to the manufacturer's instructions.

Data analysis of high-throughput sequencing data

All samples were sequenced by Illumina Hiseq 2000 with single-end 50-bp read length. The deep-sequencing data were mapped to zebrafish genome version 10 (GRCz10). (1) For m^6A-seq, reads were aligned to the reference genome (danRer10) using Tophat v2.0.14[118] with parameter -g 1–library-type = fr-firststrand. RefSeq Gene structure annotations were downloaded from UCSC Table Browser. The longest isoform was used if the gene had multiple isoforms. Aligned reads were extended to 150 bp (average fragments size) and converted from genome-based coordinates to isoform-based coordinates, in order to eliminate the interference from introns in peak calling. The peak-calling method was modified from published work[58]. To call m^6A peaks,

the longest isoform of each gene was scanned using a 100 bp sliding window with 10 bp step. To reduce bias from potential inaccurate gene structure annotation and the arbitrary usage of the longest isoform, windows with read counts less than 1 out of 20 of the top window in both m^{6}A-IP and input sample were excluded. For each gene, the read counts in each window were normalized by the median count of all windows of that gene. A Fisher exact test was used to identify the differential windows between IP and input samples. The window was called as positive if the FDR < 0.01 and $\log_2$(enrichment score) ≥ 1. Overlapping positive windows were merged. The following four numbers were calculated to obtain the enrichment score of each peak (or window): (a) reads count of the IP samples in the current peak or window, (b) median read counts of the IP sample in all 100 bp windows on the current mRNA, (c) reads count of the input sample in the current peak/window, and (d) median read counts of the input sample in all 100 bp windows on the current mRNA. The enrichment score of each window was calculated as $(a \times d)/(b \times c)$. (2) For m^{6}A-CLIP-seq, after removing the adaptor sequence, the reads were mapped to the reference genome (danRer10) using Bowtie2. Peak calling method was similar to the previous study[116]. Briefly, mutations were considered as signal and all mapped reads were treated as background. A Gaussian Kernel density estimation was used to identify the binding regions. The motif analysis was performed using HOMER[88] to search motifs in each set of m^{6}A peaks. The longest isoform of all genes was used as background. (3) For mRNA-seq, reads were mapped with Tophat and Cufflink (v2.2.1) was used to calculate the FPKM of each gene to represent their mRNA expression level[85]. (4) For fish gene group categorization, we used the input mRNA-seq data from m^{6}A-seq. FPKM of all genes were first normalized to the highest value of five time points, with only genes with FPKM >1 analyzed. Then Cluster3.0[119] was used to divide all genes into six clusters, with the parameters: adjust data – normalize genes; *k*-means cluster – organize genes, 6

clusters, 100 number; *k*-means – Euclidean distance. The result clustered file with clustered number was merged with original FPKM values, imported and processed in R, and plotted in Excel. (5) For GO analysis, the list of target genes was first uploaded into DAVID[90,120] and analyzed with functional annotation clustering. The resulting file was downloaded and extracted with GO terms and corresponding *P* values. The new list (contains GO terms with $P < 0.01$) was imported into REVIGO[91] and visualized with the interactive graph, which was used as the final output figures.

Integrative data analysis and statistics

Methylated genes (at each time point) were defined as overlapped gene targets between m^6A-seq and m^6A-CLIP-seq. Ythdf2-regulated genes were defined as overlapped gene targets between the lists of the top 20% upregulated genes in both *ythdf2* knockout and MO-injected samples. The most stringent Ythdf2 target genes at 4 h.p.f. (135) were defined in the main text, as overlapped genes of methylated genes at 4 h.p.f. (3,237) and Ythdf2-regulated genes at 4 h.p.f. (876).

RT–qPCR

Quantitative reverse-transcription PCR (RT–qPCR) was performed to assess the relative abundance of mRNA. All RNA templates used for RT–qPCR were pretreated with on-column DNase I digestion in the purification step. RT–qPCR primers were designed to span exon-exon junctions to only detect mature mRNA. RT–qPCR was performed by using SuperScript III one-step RT–PCR system (Thermo Fisher) with 50–100 ng total RNA template. *Actb1* was used as an internal control as it showed relative invariant expression during the studied time period according to pilot RT–qPCR data. *P* values were determined using two-sided Student's *t*-test for two samples with equal variance. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. The sequences of primers used in this

study are listed below: *actb1*: forward 5′-CGAGCAGGAGATGGGAACC-3′, reverse 5′-CAACG GAAACGCTCATTGC-3′; *buc*: forward 5′-CAAGTTACTGGACCTCAGGATC-3′, reverse 5′-G GCAGTAGGTAAATTCGGTCTC-3′; *zgc:162879*: forward 5′-TCCTGAATGTCCGTGAATGG -3′, reverse 5′-CCCTCAGATCCACCTTGTTC-3′; *mylipa*: forward 5′-CCAAACCAGACAACC ATCAAC-3′, reverse 5′-CACTCCACCCCATAATGCTC-3′; *vps26a*: forward 5′-AAATGACAG GAATAGGGCCG-3′, reverse 5′-CAGCCAGGAAAAGTCGGATAG-3′; *tdrd1*: forward 5′-TAC TTCAACACCCGACACTG-3′, reverse 5′-TCACAAGCAGGAGAACCAAC-3′; *setdb1a*: forward 5′-CTTCTCAACCCAAAACACTGC-3′, reverse 5′-CTATCTGAAGAGACGGGTGAA AC-3′; *mtus1a*: forward 5′-TGGAGTATTACAAGGCTCAGTG-3′, reverse 5′-TTATGACCACA GCGACAGC-3′; *GFP*: forward 5′-TGACATTCTCACCACCGTGT-3′, reverse 5′-AGTCGTCC ACACCCTTCATC-3′.

Data presentation

All the raw data and processed files have been deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) and accessible under GSE79213.

CHAPTER 4

The Functions of m^{6}A in Human Diseases [*]

4.1 Introduction: m^{6}A regulators associate with disease-related RNAs

Since m^{6}A marks functional cohorts of mRNA, it is not surprising that m^{6}A is linked to numerous types of human diseases which usually are caused by aberrant gene expression. For instance, many cancer types have been linked to m^{6}A, although the connections are mostly indirect[121-124]. Several recent reports, however, provided molecular evidence of direct involvement of m^{6}A during cancer development: The depletion of METTL3 was reported to cause apoptosis and reduced invasiveness of cancer cells[125,126]; while the activation of ALKBH5 by hypoxia induces m^{6}A demethylation of the key pluripotency factor NANOG, which increases transcript stability and promotes proliferation and enrichment of human breast cancer stem cells[113].

In addition to cancers, m^{6}A has also been implicated in other types of diseases and health conditions such as metabolism misregulation and obesity: FTO was suggested to influence pre-adipocyte differentiation[127-129], and SNPs in *FTO* associate with body mass index in human populations and the occurrence of obesity and diabetes[130,131]. The connection between m^{6}A and neuronal disorders has also been documented. For instance, dopamine signaling is dependent on FTO and on m^{6}A methylation of key signaling transcripts[132], and mutations in the prion-like domain of the reader protein HNRNPA2B1 are known to cause neurodegeneration through

[*] The author's contribution to the studies presented in this chapter: Glioblastoma study: The author performed biochemistry experiments with S.Z., analyzed data with S.Z. and Z.L., and contributed to the manuscript writing by S.Z. and S.H. HIV study: The author contributed equally to this work with N.T., acquired data with N.T., W.L, Z.L. and L.W., analyzed and interpreted data with N.T., W.L, Z.L., C.H. and L.W., and wrote the manuscript with N.T., W.L, Z.L., C.H. and L.W. ZIKV study: The author contributed equally to this work with G.L., performed experiments with G.L. and Y.W., analyzed and interpreted data with G.L., Y.W., Z.L., Y.Q., C.H. and T.M.R., and wrote the manuscript with G.L., C.H. and T.M.R. RSV and influenza study: The author worked with M.X. and Y.Y. in the collaborative study between J.L. and C.H. groups, performed experiments with M.X. and Y.Y., analyzed and interpreted data with M.X., Y.Y., Z.L., C.H. and J.L., and drafted manuscripts with M.X., Y.Y., C.H. and J.L.

dysregulated protein polymerization[133]. FTO and ALKBH5 have been associated with the developments of depressive disorders[134-137], and addiction, epilepsy, attention deficit disorder and other neurological disorders have also been associated with m^{6}A regulators[138-140]. Reproductive disorders, inflammation, and viral infection are also among the diseases influenced by m^{6}A[141-146].

Speaking of viral infection, the existence of m^{6}A in viruses that integrate their genome into the host genome has been known since the discovery of m^{6}A[147,148]. These viruses include SV40, adenovirus, herpes virus, Rous sarcoma virus, influenza virus[149], and more recently, human immunodeficiency virus (HIV) [142,143,146]. The virus-host interaction has been under our investigation as viral RNAs are produced from virus-hijacked host machineries and thus carry internal m^{6}A deposited by host enzymes, which could be utilized by viruses to enhance their survival in mammalian host cells[146]. Our experiments have identified multiple m^{6}A sites on the short viral RNA genome of HIV, RSV, influenza, and the recently outbroken Zika virus[150]. To our surprise, the effects of modulating m^{6}A regulators vary with different types of virus species investigated, which are discussed in the following section.

4.2 Results: Regulatory roles of m^{6}A in non-infectious and infectious diseases

To evaluate the broad impacts of m^{6}A in different disease settings, we have chosen representative examples of both non-infectious and infectious diseases, namely human glioblastoma and several types of RNA virus infection, as our model systems for disease-related studies.

4.2.1 m^{6}A in cancer: demethylase facilitates cancer cell proliferation

Despite the numerous studies suggesting the connection between m^{6}A and cancer, the exact molecular mechanism of how m^{6}A exerts regulatory functions during the development of cancer remains elusive. To address this issue, we investigated the involvement of m^{6}A regulators in

human glioblastoma (GBM) and during the tumorigenesis process of glioblastoma stem-like cells (GSCs). The results showed the importance of correct deposition of m^6A on vital genes and the profound influence it may have in cancer development.

4.2.1.1 Locate the cancer-related m^6A regulator: ALKBH5 associates with high self-renewal of GSCs

To study the m^6A modulators that may result in a poor clinical outcome in GBM patients, we first queried The Cancer Genome Atlas (TCGA; http://www.cbioportal.org)[151] and found a significant correlation between elevated expression of ALKBH5 and poor GBM patient prognosis (Figure 4.1A). Similar observations were made in another data base as well. We then examined ALKBH5 expression in established and primary glioma cell lines representing different stages of malignancy. ALKBH5 was weakly or moderately expressed in immortalized normal human astrocytes (NHAs), Hs683 and SW1783 cell lines (lower-grade gliomas), and U87MG and LN229 GBM cell lines, but was highly expressed in U251MG GBM cells and in patient-derived primary GBM cultures that enrich for cells with self-renewal and tumor propagation potential, i.e., GSCs (Figure 4.1B). The elevated ALKBH5 expression in the GSCs led us to hypothesize that ALKBH5 is associated with cancer stem cell (CSC) niches. Then, using immunofluorescence analysis, we found that ALKBH5 co-expressed with SOX2, a transcription factor indicative of CSC self-renewal, and Nestin, an intermediate filament protein recognized as the marker of neural stem cells, in bulk GBM samples (Figure 4.1C). Moreover, gene expression analysis using the TCGA GBM dataset validated the positive correlation between ALKBH5 and SOX2 or Nestin expression (Figure 4.1D). Furthermore, comparing ALKBH5 levels between GSCs and their matching bulk tumors indicated a significant increase in ALKBH5 expression in the GSC population (Figure 4.1E). Therefore, the high ALKBH5 expression in the GBM subpopulation characterized by a CSC

niche suggests an association of ALKBH5 with CSCs within the bulk tumor, thereby ALKBH5 became the focus of our investigation.

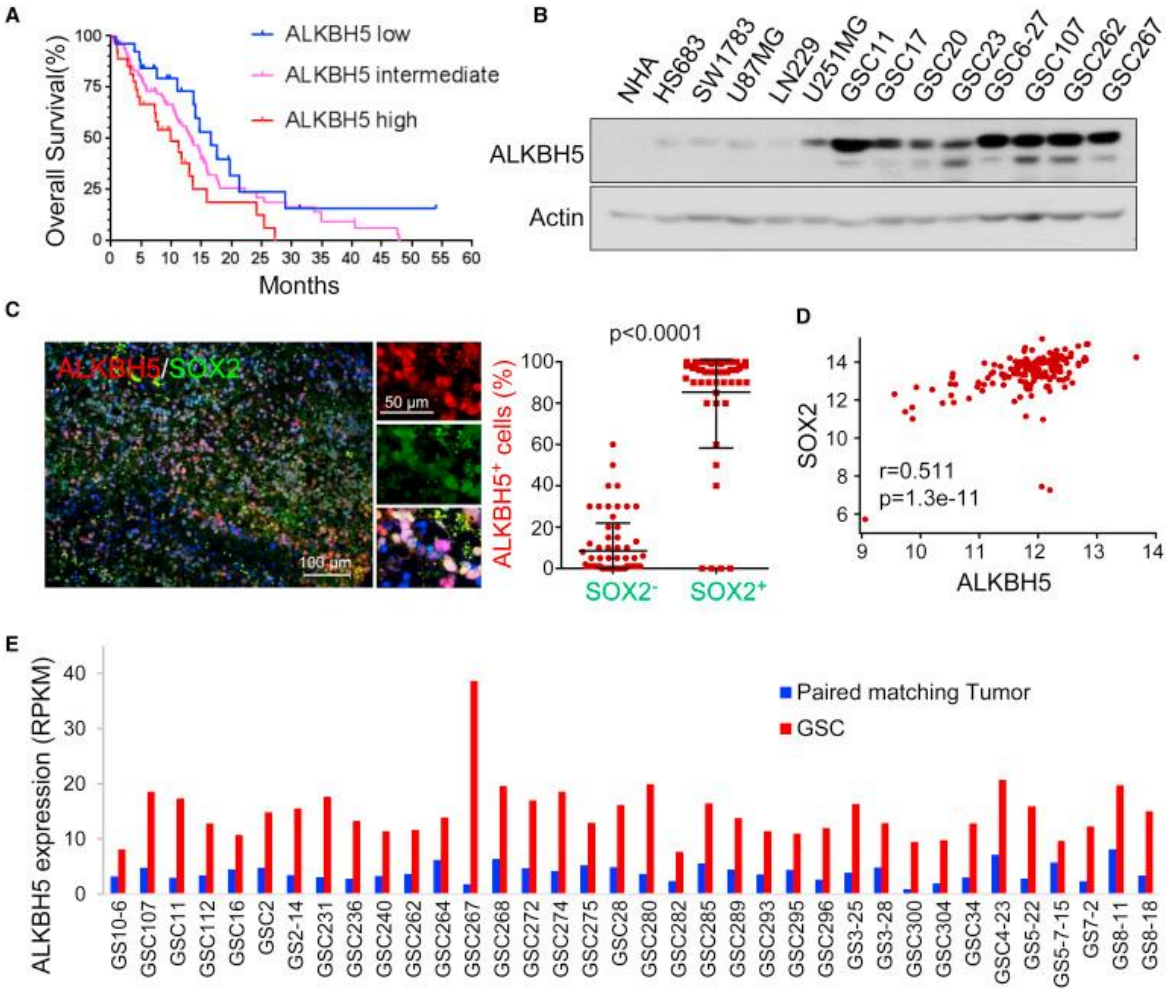


Figure 4.1 ALKBH5 predicts poor survival of GBM patients and is associated with cancer stem cell niches

(A) Correlation between ALKBH5 mRNA expression and survival of GBM patients in the TCGA dataset. Overall patient survival in groups of high, intermediate, and low expression was analyzed by Kaplan-Meier survival curve. The median overall survival duration of patients with high ALKBH5 expression (9.9 months) versus low ALKBH5 expression (16.6 months) was compared by log-rank test (p = 0.037). (B) Western blotting of ALKBH5 in NHAs, glioma cells, and GSCs. Actin served as a loading control. (C) Correlation between ALKBH5 and SOX2 protein expression in GBM specimens. Tumor sections from 15 GBM specimens were immunofluorescence (IF)-stained with anti-ALKBH5 and anti-SOX2 antibodies. Left: representative images. Right: in five random selected microscope fields of each tumor, the percentage of ALKBH5-positive cells among SOX2-positive versus SOX2-negative cells was compared by t test. Lines show mean and s.d. (D) Pearson correlation between ALKBH5 and SOX2 mRNA expression (RNA-seq V2 RSEM [log_2]) in the TCGA GBM dataset. (E) ALKBH5 mRNA expression in GSCs and paired matching tumors by RNA-seq analysis.

To determine whether ALKBH5 is directly affecting GSC self-renewal, we used two distinct short hairpin RNAs (shRNA1 and shRNA2; hereafter referred to collectively as shALKBH5) to ablate ALKBH5 expression. Results showed that in GSC11, GSC17, and GSC23 cells, loss of ALKBH5 inhibited cell growth (Figure 4.2A) and decreased DNA replication, as revealed by 5-bromo-2′-deoxyuridine (BrdU) incorporation assay (Figure 4.2B and C). We then examined the effect of ALKBH5 depletion on GSC tumorigenicity by intracranial injection into mice with 10,000 viable GSC11 or GSC17 cells transduced with shCtrl or shALKBH5. Compared with the mice injected with shCtrl-GSC cells, those injected with shALKBH5-GSC cells displayed extended survival with a lower rate of tumor formation (Figure 4.2D). Moreover, the inhibition of brain tumor formation by shALKBH5 could be effectively rescued by overexpression of the wild-type ALKBH5 but not the reported catalytic inactive mutant ALKBH5 H204A (Figure 4.2E). These results strongly indicate that the high expression of ALKBH5 is critical to GSC proliferation and tumorigenicity.

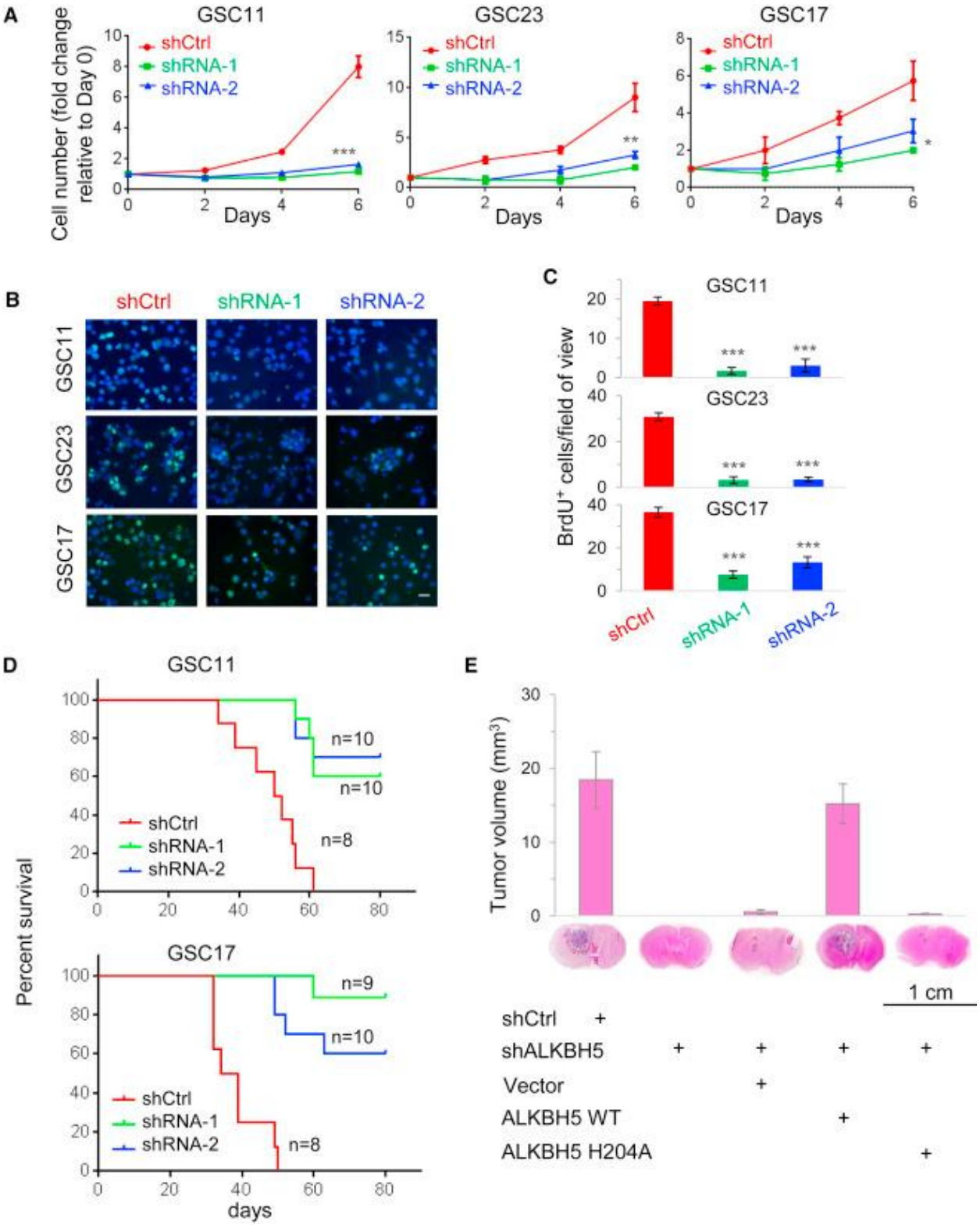


Figure 4.2 Knockdown of ALKBH5 impairs GSC proliferation and tumorigenicity
(A) The proliferation of GSCs with or without ALKBH5 knockdown as determined by cell counting. $n = 3$. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. (B) The proliferation of GSCs with or without

(**Figure 4.2, continued**) ALKBH5 knockdown as assessed by BrdU (5-bromo-2′-deoxyuridine) incorporation for 3 hours. Scale bar, 20 μm. (C) Randomly selected microscope fields of (B) were quantified for BrdU-positive GSCs. $n = 5$. $^{***}p < 0.001$. (D) Survival analysis of mice intracranially implanted with GSCs with or without ALKBH5 knockdown. GSCs were implanted intracranially into nude mice, and tumor formation was determined by histology. (E) Top: Bar graph shows tumor volumes formed by indicated GSC17 cells in mouse brains. Bottom: representative images of H&E staining for tumor formation. $n = 8$. All bar plot data are means ± s.e.m.

4.2.1.2 Search for the key target of ALKBH5: FOXM1 is subjected to demethylation and gene expression control by ALKBH5

To identify the direct RNA targets and understand the regulatory role of ALKBH5 in comprehensive GSCs gene expression, we employed microarray analysis to compare the gene expression profile following ALKBH5 knockdown in GSC11 and GSC17 cells. A total of 206 genes that were differentially expressed by at least 2-fold in both cell lines were selected for ingenuity pathway analysis (IPA) (Figure 4.3A). We found that the largest subset of genes is primarily involved with "Cell Cycle," followed by "Cellular Assembly and Organization" and "DNA Replication, Recombination, and Repair".

To test whether the altered gene expression could be a consequence of ALKBH5-mediated m^6A demethylation, we compared the m^6A distribution in control and ALKBH5 knockdown GSCs by m^6A-seq. The GGACU motif was identified to be highly enriched within m^6A sites in the GSC11 cells (Figure 4.3B). In total, m^6A-seq identified 39,288 and 34,415 m^6A peaks from 12,358 to 12,196 m^6A-modified transcripts in control and ALKBH5-deficient cells, respectively (Figure 4.3C and D). In the GSC11 cells with ALKBH5 stable knockdown by shRNA, 5,053 new peaks appeared with the disappearance of 9,650 peaks. The other 29,362 peaks were found in both knockdown and control cells (Figure 4.3C). Since ALKBH5 is an m^6A demethylase, the 5,053 unique peaks are expected to contain genuine targets of ALKBH5. We further investigated the m^6A distribution patterns within both total and unique peaks. A similar pattern of total and common m^6A distribution in control and ALKBH5-deficient cells was observed when the RNA

species were divided into 5′ UTR, coding sequence (CDS), 3′ UTR regions of mRNAs, and non-coding RNAs (Figures 4.3E). Interestingly, in contrast to the 9,650 lost peaks that displayed identical distribution to the total peaks, the 5,053 unique ALKBH5-dependent peaks showed a distinct pattern in which a relative increase of m^6A deposit appeared in 5′ UTR and non-coding RNAs (Figure 4.3E). These unique peaks included 1,302 peaks from 5′ UTR, 2,724 from CDS, and 678 from 3′ UTR of mRNAs, and 349 peaks from 281 non-coding RNAs.

We then asked whether these peaks were associated with differentially expressed genes in the microarray analysis. Among the 206 differentially regulated transcripts identified by microarray, 148 genes were found in the RNA-seq dataset (m^6A-seq input library), including 85 genes repeatedly showing over 2-fold quantitative difference in the same direction. Filtering the 5,053 unique m^6A peaks with the 85 genes with 2-fold expression changes resulted in the identification of 13 peaks harbored by 12 genes, including ABHD4, ANP32E, CENPF, FANCI, FOXM1, MKI67, BCL2, UBE2C, NT5E, IGFBP5, TSPYL2, and CDCA2 (Figure 4.3F). It is likely that a few key candidate genes are responsible for the observed phenotype and altered expression of the majority of proliferation-related genes. Next, we performed Ingenuity's upstream regulator analysis for the 206 genes as well as the 85-gene subset, each resulting in a similar list of regulators predicted to be activated or inhibited following ALKBH5 knockdown. Among these potential regulators, only FOXM1 was one of the 12 candidate genes and had a high predicted activation *Z* score, suggesting that FOXM1 downregulation might explain the loss of the GSC proliferation signature (Figure 4.3F).

FOXM1 is a pivotal transcription factor in cell-cycle regulation and plays critical roles in the self-renewal and tumorigenesis of GSCs. The IPA-defined and other reported FOXM1 transcriptional targets were all decreased following ALKBH5 knockdown in both our microarray

and RNA-seq dataset (Figure 4.3G). Given that FOXM1 could mediate the expression of a number of ALKBH5-responsive pro-proliferative genes in GSCs, our subsequent research was focused on this potentially important target gene.

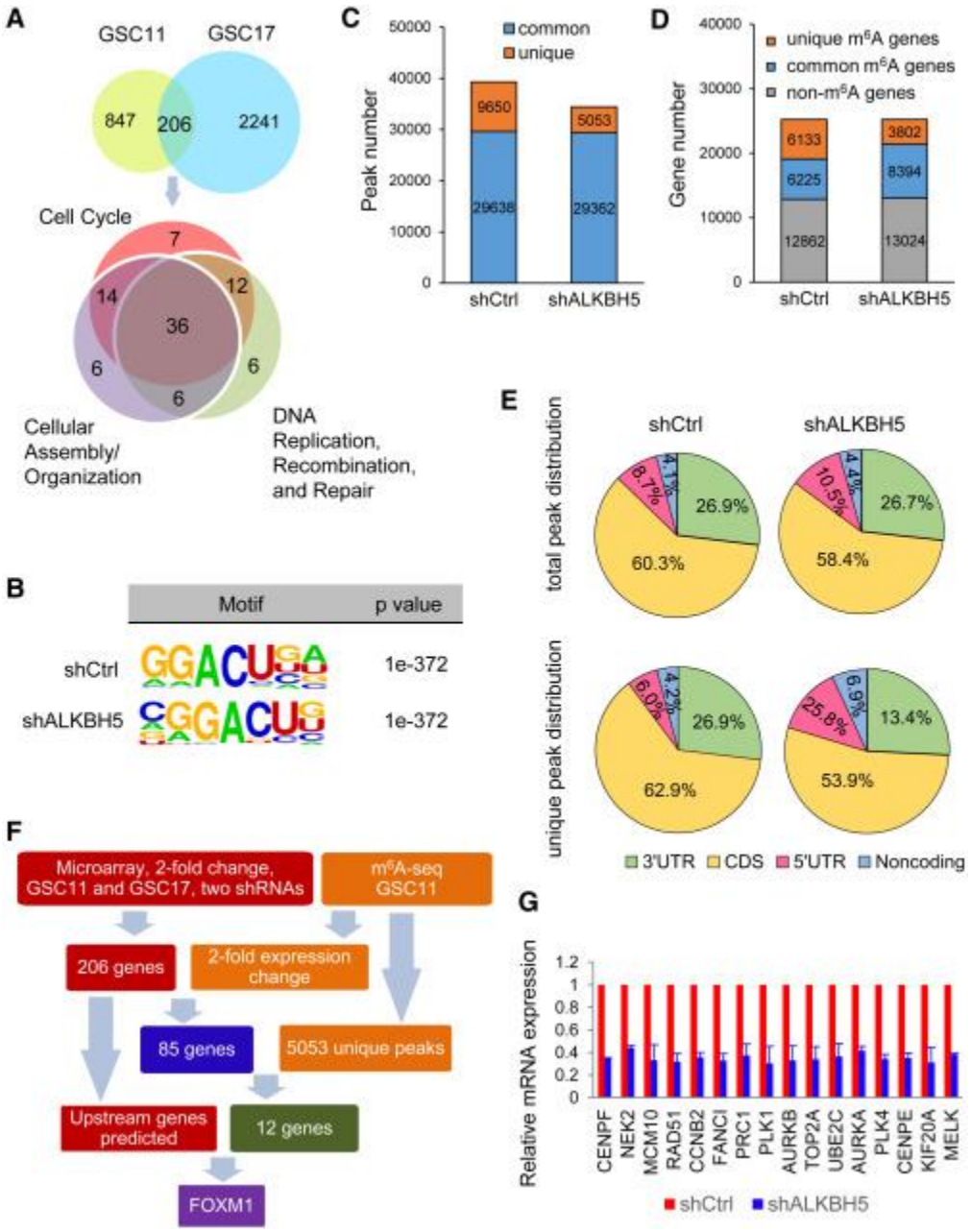


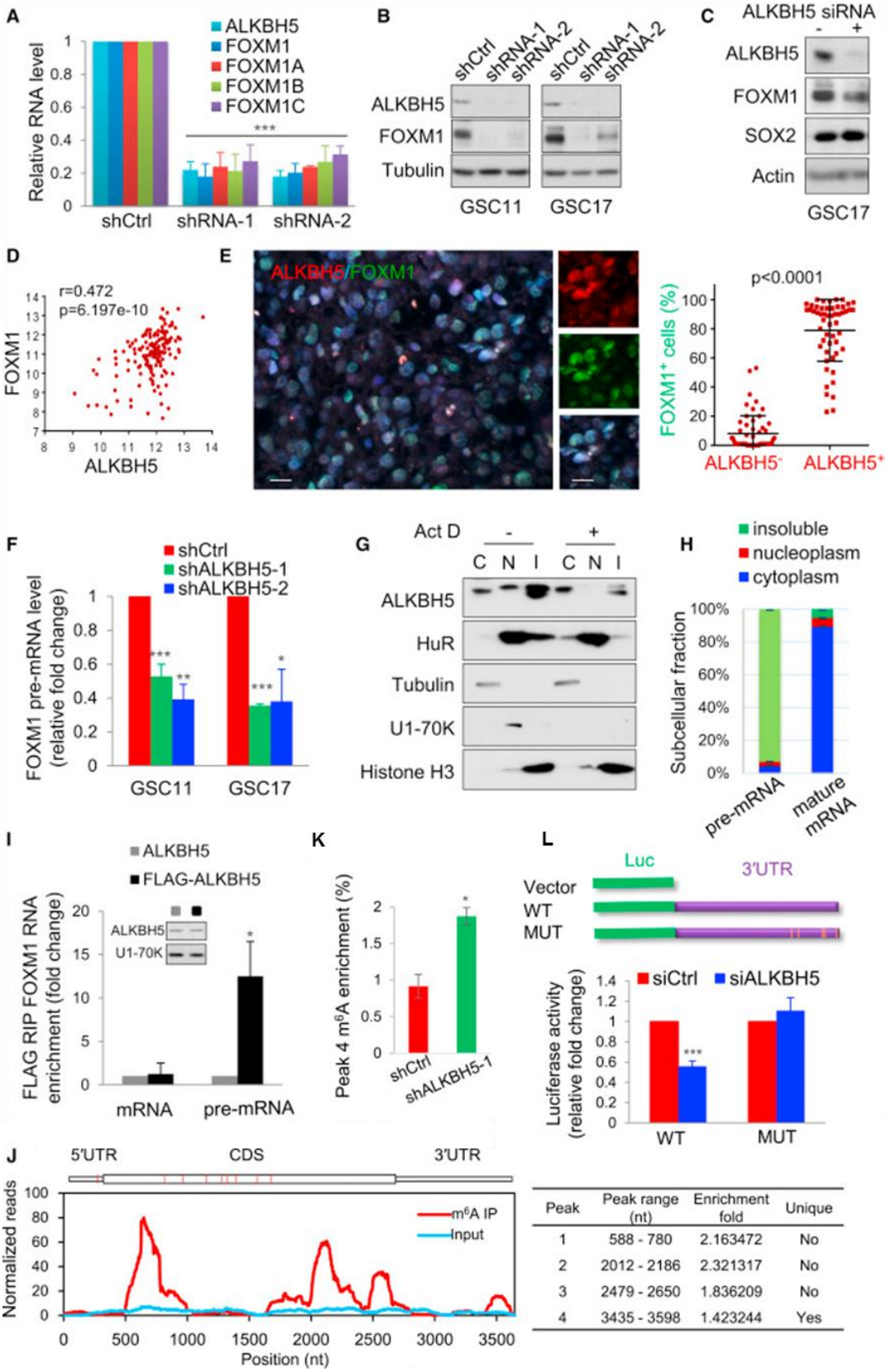
Figure 4.3 Characterization of gene expression and m^{6}A modification changes in the ALKBH5-deficient GSCs

(A) Venn diagrams show 206 genes with over 2-fold expression change in shALKBH5 compared with shCtrl in both GSC11 and GSC17 (top). IPA identified the enriched gene ontology processes of the 206 genes (bottom). (B) Top consensus motif identified by HOMER with m^{6}A-seq peaks in

(**Figure 4.3, continued**) GSC11 cells with or without ALKBH5 knockdown. (C) Number of m^6A peaks identified in m^6A-seq in shCtrl and shALKBH5 GSC11 cells. (D) Number of m^6A-modified genes identified in m^6A-seq. Common m^6A genes contain at least one common m^6A peak, while unique m^6A genes contain no common m^6A peaks. (E) Graphs of m^6A peak distribution showing the proportion of total m^6A peaks in the indicated regions in control and ALKBH5-deficient cells (top) and the appearance of new m^6A peaks (unique peaks in shALKBH5) or loss of existing m^6A peaks (unique peaks in shCtrl) after ALKBH5 knockdown (bottom). (F) Schematic of ALKBH5 downstream analysis pinpointing FOXM1 as the unique target. (G) Relative mRNA levels of FOXM1 downstream targets from microarray analysis. Data represent means ± SEM for GSC11 and GSC17 cells. Combined analysis of two independent biological replicates in (B) to (E).

Since FOXM1 could be a direct substrate of ALKBH5 and a driver for the altered proliferative gene profile, we investigated the regulatory mechanism of FOXM1 expression by ALKBH5. We first measured the change of total mRNA expression and different isoforms (FOXM1A, -B, and -C) levels of FOXM1 upon ALKBH5 knockdown. Consistent with the gene expression data, both GSC11 and GSC17 with ALKBH5 knockdown displayed approximately 70%–80% lower expression of total and isoforms of FOXM1 mRNA (Figure 4.4A). We also measured the protein expression of FOXM1 and SOX2 following ALKBH5 knockdown by small interfering RNA (siRNA) treatment or shRNA lentiviral transduction. Both transient and stable knockdown reduced FOXM1 protein abundance (Figure 4.4B and C). In contrast, SOX2 expression was only down-regulated by stable ALKBH5 knockdown but almost unaffected by transient ALKBH5 knockdown, suggesting that SOX2 may not be a direct target of ALKBH5 in GSCs (Figure 4.4C). Next, we examined the correlation between ALKBH5 and FOXM1 in GBM patient samples. We queried the TCGA dataset and found a significant positive correlation (Pearson's $r = 0.4723914$, $p = 6.197 \times 10^{-10}$) between ALKBH5 and FOXM1 mRNA expression (Figure 4.4D). This result was confirmed by immunofluorescence analysis of the expression and correlation of ALKBH5 and FOXM1 proteins in primary GBM samples (Figure 4.4E). Collectively, these results indicate that FOXM1 is a direct downstream target of ALKBH5 in GBM.

Since ALKBH5 primarily locates in the nucleus, it is possible that ALKBH5 functions on FOXM1 nascent transcripts and then regulates the expression of mature RNA or protein. Indeed, FOXM1 precursor mRNA (pre-mRNA) expression was decreased by 40% following ALKBH5 knockdown in GSCs (Figure 4.4F), suggesting ALKBH5 affects FOXM1 pre-mRNA abundance. To further probe the exact effects, we first separated nuclei from the cytoplasmic fraction of GSC11 cells and fractionated them into a soluble nucleoplasmic fraction containing nuclear speckles and an insoluble fraction containing nascent RNAs and chromatin-associated components. Western blotting for the compartment-specific proteins α-tubulin, U1-70K, and histone H3 confirmed organelle subfractionation (Figure 4.4G). We found that ALKBH5 was present in both the nucleoplasmic and insoluble fractions, but was only enriched in the insoluble fraction in a transcription-dependent manner, as its nuclear abundance decreased after actinomycin D treatment blocking transcription (Figure 4.4G). In addition, FOXM1 pre-mRNA was predominantly in the insoluble fractions (Figure 4.4H), which can be explained by the widespread co-transcriptional splicing model in human brain[152]. Moreover, the results of native RNA immunoprecipitation (RIP) assay revealed that FOXM1 pre-mRNA interacts with FLAG-ALKBH5 (Figure 4.4I). Together, these results indicate that ALKBH5 directly interacts with FOXM1 nascent transcripts and regulates its abundance.



Peak	Peak range (nt)	Enrichment fold	Unique
1	588 - 780	2.163472	No
2	2012 - 2186	2.321317	No
3	2479 - 2650	1.836209	No
4	3435 - 3598	1.423244	Yes

Figure 4.4 ALKBH5 modulates FOXM1 expression by demethylating FOXM1 nascent transcripts

(**Figure 4.4, continued**) (A) qPCR analysis of FOXM1 mRNA expression in GSC11 with or without ALKBH5 knockdown. Samples were normalized to GAPDH mRNA. (B) Western blotting of FOXM1 in GSCs with or without ALKBH5 knockdown. (C) Western blotting of FOXM1 and SOX2 in GSC17 cells treated with control or ALKBH5 siRNAs. (D) Correlation between FOXM1 and ALKBH5 mRNA expression (RNA-seq V2 RSEM [log_2]) in the TCGA GBM dataset. (E) Correlation between ALKBH5 and FOXM1 protein expression in GBM specimens. Left: Representative IF images of 15 GBM specimens. Scale bar, 20 μm. Right: in five random selected microscope fields of each tumor, the percentage of FOXM1-positive cells among ALKBH5-positive versus ALKBH5-negative cells was compared by t test. Lines show mean and s.d. (F) qPCR analysis of FOXM1 pre-mRNA in GSCs with or without ALKBH5 knockdown. (G) Cytoplasmic (C), nucleoplasmic (N), and insoluble (I) fractions of GSC11 cells treated with or without 5 μg/mL actinomycin D (Act D) for 2 hours were subjected to western blotting for indicated proteins. (H) Distribution of FOXM1 transcripts in subcellular fractions assessed by qPCR. (I) RNA immunoprecipitation (RIP) analysis of transcripts from the nuclear extracts of GSC17 cells expressing exogenous ALKBH5 with or without FLAG tag. Enrichment of FOXM1 mature and pre-mRNA with FLAG was measured by qPCR and normalized to input. Western blotting of ALKBH5 showing equal expression of tagged or untagged proteins. (J) Left: peaks show the relative abundance of m^6A sites along FOXM1 mRNA in GSC11-shALKBH5 cells. The blue and red reads are from the non-immunoprecipitate control input and MeRIP of GSC11-shALKBH5, respectively. Right: identification of a unique peak at 3′ UTR. (K) MeRIP-qPCR analysis of fragmented FOXM1 RNA in GSC17 with or without ALKBH5 knockdown. (L) Relative activity of the wild-type or mutant FOXM1 3′ UTR firefly luciferase reporter in GSC17 cells treated with control or ALKBH5 siRNAs.

From our m^6A-seq data of shALKBH5 cells, we have identified four statistically significant m^6A peaks in FOXM1 mRNA, with one peak near the stop codon and one unique peak on 3′ UTR (Figure 4.4J). This unique peak was specifically demethylated by ALKBH5, suggesting functional relevance. The m^6A status of each peak was further measured by MeRIP of fragmented RNA. Using specific primers to detect the unique peak region, we found m^6A level was consistently elevated by ~2-fold on this region in shALKBH5 cells (Figure 4.4K). To test the function of this unique 3′ UTR m^6A peak, we performed the FOXM1 3′ UTR-reporter luciferase assay and found that ALKBH5 knockdown decreased the activity of the luciferase construct containing the FOXM1 3′ UTR, while mutating all the m^6A sites within this unique peak rendered resistance to the effect of ALKBH5 knockdown (Figure 4.4L). These data indicate ALKBH5 exerts regulatory functions on FOXM1 through demethylating this unique 3′ UTR m^6A peak.

4.2.1.3 Reveal the mechanism of ALKBH5 controlling FOXM1 abundance: HuR and FOXM1-AS

To further explain the negative correlation between methylation and FOXM1 abundance, we investigated the carrier of the biological consequences of m^6A: RNA binding proteins. The nuclear RNA binding protein HuR has been shown to disfavor binding with m^6A-modified RNA and to exert stabilizing effects on its bound RNA [68]. To examine the role of HuR in our system, we first tested whether HuR interacted with FOXM1 nascent transcripts. RIP-qPCR revealed that HuR interacts with FOXM1 pre-mRNA, and this interaction decreased following ALKBH5 knockdown (Figure 4.5A). Moreover, transient HuR knockdown suppressed FOXM1 expression (Figure 4.5B and C), suggesting that HuR plays an important role in the regulation of FOXM1 by ALKBH5 by promoting expression of FOXM1 nascent transcripts through binding to the unmethylated 3′ UTR.

Regarding ALKBH5's preferential regulation of the FOXM1 3′ UTR m^6A peak, we investigated the local elements that may contribute to its specific binding. Intriguingly, we found a long non-coding RNA (lncRNA), *LOC100507424* (chr12: 2945982-2968961, GRCh37/hg19), which is transcribed in the opposite direction to *FOXM1* and has 457 nucleotides complementary to the last exon of FOXM1 mRNA (Figure 4.5D), suggesting possible *cis* regulation of FOXM1 expression. Biochemical assays performed on LOC100507424 (hereafter referred to as FOXM1-AS) uncovered that FOXM1-AS interacts with ALKBH5 and FOXM1 nascent transcripts and is involved in the ALKBH5-FOXM1 regulation. Native RIP was performed to assess the interaction between FOXM1-AS and ALKBH5 and revealed that FOXM1-AS was enriched in FLAG-ALKBH5 (Figure 4.5E). 4′-aminomethyltrioxalen (AMT)-assisted crosslinking of RNA-RNA hybrids in GSCs showed that FOXM1 pre-mRNA, but not mature mRNA, was retrieved from the

pull-down fractions of AMT-treated GSCs using the biotinylated FOXM1-AS RNA probes (Figure 4.5F). Notably, RNA recovery from RIP of FLAG-ALKBH5-protein complex revealed that knocking down FOXM1-AS resulted in a decrease in ALKBH5-associated FOXM1 nascent transcripts (Figure 4.5G), an increase in the m^6A modification of FOXM1 pre-mRNA (Figure 4.5H), and a decrease of HuR and FOXM1 RNA association (Figure 4.5I), suggesting a positive regulatory role of FOXM1-AS. Finally, we evaluated the biological functions of FOXM1-AS, and found that knockdown of FOXM1-AS in GSCs led to a significant decrease in the expression of FOXM1 RNA (Figure 4.5J) and substantially reduced GSC growth (Figure 4.5K), confirming the positive impacts of FOXM1-AS on FOXM1 expression.

Taken together, the above results pictured a complex regulatory mechanism of ALKBH5 and m^6A controlling the proliferation of GSCs through coordinating with other gene regulation factors (Figure 4.5L): the elevated expression of ALKBH5 in GSCs is accompanied by the association of lncRNA FOXM1-AS to specifically demethylate the FOXM1 3′ UTR m^6A peak, in turn facilitates the binding of HuR and increases the stability of FOXM1 pre-mRNA, finally contributes to the high expression of FOXM1 and subsequently improves the proliferation and tumorigenesis of GSCs.

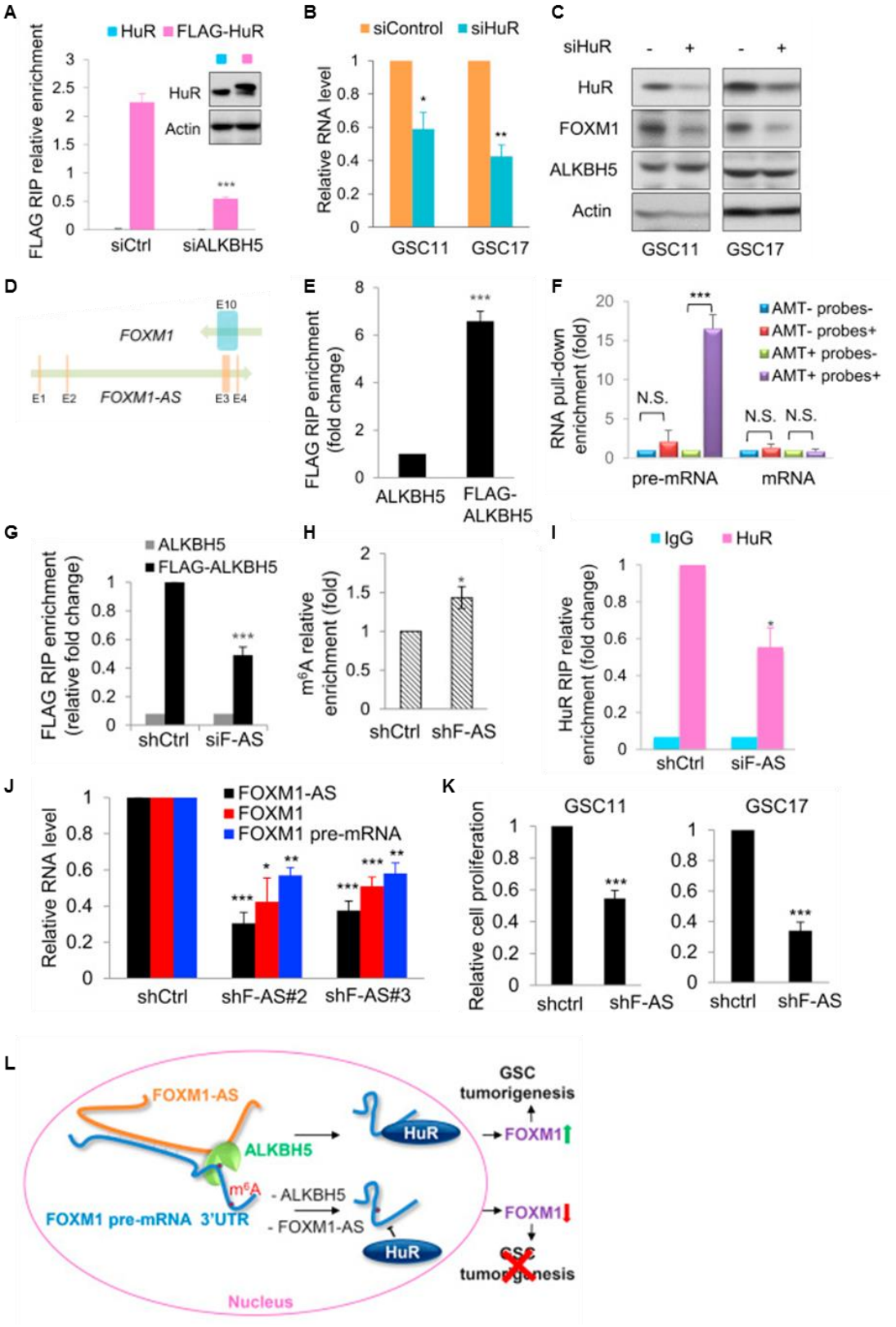


Figure 4.5 HuR and FOXM1-AS function in the regulation of FOXM1 nascent transcripts by ALKBH5

(**Figure 4.5, continued**) (A) RIP analysis of the interaction of FLAG with FOXM1 pre-mRNA using total cell lysates of GSC17 cells expressing exogenous HuR with or without FLAG tag and treated with control or ALKBH5 siRNAs. Enrichment of FOXM1 pre-mRNA with FLAG was measured by qPCR and normalized to input. Western blotting of HuR showed equal expression of tagged or untagged proteins. (B) qPCR analysis of FOXM1 in GSCs treated with control or HuR siRNAs. (C) Western blotting of FOXM1 RNA expression in GSCs treated with control or HuR siRNAs. (D) Physical map of *FOXM1* and *FOXM1-AS* (F-AS) transcription units. Arrows indicate the direction of transcription. Vertical bars indicate exons. (E) Analysis of the interaction between F-AS and ALKBH5 by an RIP assay using anti-FLAG antibodies in the nuclear extracts of GSC17 cells expressing exogenous ALKBH5 with or without FLAG tag. The RNA and protein in the RNA-protein complex were then detected by qPCR. (F) RNA pull-down assay of F-AS-associated RNA from GSC17 cells. Biotinylated F-AS probes or antisense probes were incubated with AMT-crosslinked or untreated RNA and collected with streptavidin beads. Interaction with F-AS was quantified by qPCR. (G) RIP analysis of the interaction of FOXM1 pre-mRNA with FLAG in the nuclear extracts of GSC17 cells expressing exogenous ALKBH5 with or without FLAG tag and treated with control or F-AS siRNAs. Enrichment of FOXM1 pre-mRNA with FLAG was measured by qPCR and normalized to input. (H) MeRIP-qPCR analysis of m^6A levels of FOXM1 pre-mRNA in GSC17 cells with or without F-AS knockdown. (I) RIP analysis of the interaction of FOXM1 pre-mRNA with HuR in GSC17 cells with or without F-AS knockdown. Enrichment of FOXM1 pre-mRNA with HuR was measured by qPCR. (J) qPCR analysis of indicated RNA in GSC17 cells with or without F-AS knockdown. (K) Proliferation of GSCs with or without F-AS knockdown, as determined by cell counting. (L) The proposed model of ALKBH5, HuR, and FOXM1-AS co-regulating the expression of FOXM1 in GSCs. All bar plot data are means ± s.e.m. of three independent experiments. $^{*}p < 0.05$, $^{***}p < 0.001$. N.S., no significance.

4.2.2 m^6A in viral infection: m^6A marks diverse types of viral transcripts

Based on the Baltimore classification system, viruses can be categorized into seven different families based on their genetic materials and their method of replication. Among them, four types use RNA as the viral genome, including double strand RNA (dsRNA), single strand RNA (ssRNA) with either positive or negative sense, and ssRNA reverse transcribed to DNA. We investigated regulatory functions of m^6A in three types of ssRNA viruses and the corresponding changes in the host transcriptome upon viral infection.

4.2.2.1 m^6A in HIV

The human immunodeficiency virus (HIV) uses positive-sense ssRNA as its viral genome, which is reverse transcribed into DNA and incorporated into the host genome, marking HIV as a typical example of a retrovirus. It is known that host proteins interact with HIV nucleic acids or

viral proteins and either enhance or inhibit viral replication in cells; yet it is less clear whether host proteins can post-transcriptionally modify HIV RNA, which may affect interactions between RNA and host or viral proteins, thereby affecting HIV infection.

To investigate the presence of m^6A in HIV RNA and to map the m^6A modification on HIV genome, we isolated RNA samples from $CD4^+$ Jurkat T-cells or primary $CD4^+$ T-cells infected with replication-competent HIV-1_{NL4-3} and performed m^6A-seq with poly(A)-enriched RNA. We identified similar profiles of m^6A peaks in HIV-1 RNA from these two cell types, which are mainly enriched in the 5' and 3' UTRs as well as the *rev* and *gag* genes of the HIV-1 genome (Figure 4.6A and B). These results confirmed the existence of m^6A modifications on HIV-1 genome.

To examine the effect of HIV-1 infection on m^6A modifications of cellular RNAs, we compared the distribution of m^6A peaks in cellular RNAs from HIV-1 infected and uninfected T-cells. In Jurkat and primary $CD4^+$ T-cells, HIV-1 infection did not significantly affect the percentages of total m^6A peaks mapped to the human genome in the 5′ UTR, coding DNA sequence (CDS), 3′ UTR and noncoding regions (Figure 4.6C and D). However, we identified over 4,000 unique m^6A peaks in each cell types with HIV-1 infection and we performed GO analysis on genes with unique methylation peaks in infected samples. The results showed that these genes enrich in functional clusters, such as metabolism, immune system process, multicellular organismal process, and development (Figure 4.6E and F), indicating widespread impacts on host biological systems induced by HIV-1 infection.

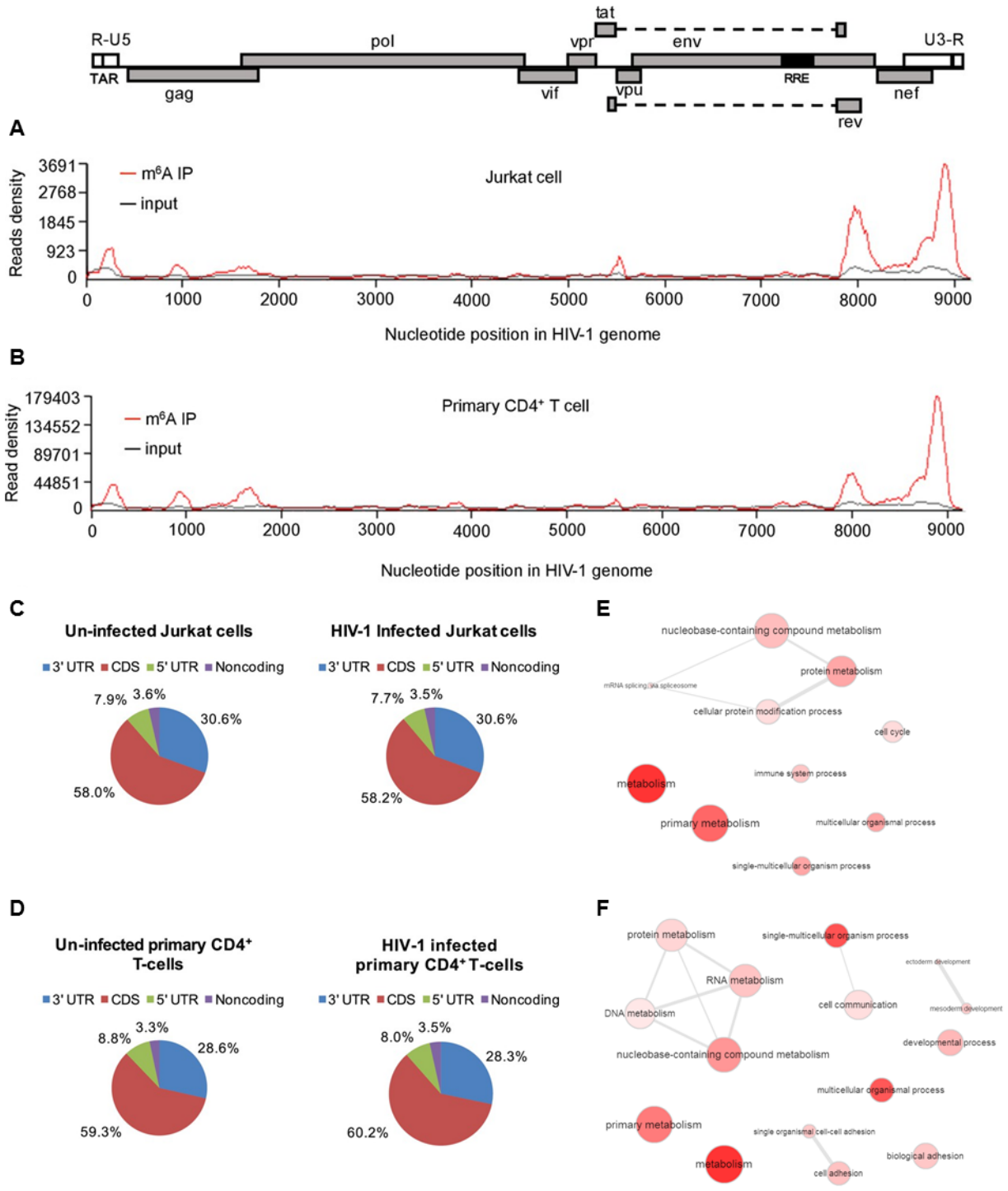


Figure 4.6 HIV-1 RNA contains m^{6}A modifications and affects host methylome
(A and B) The distribution of m^{6}A reads from m^{6}A-seq mapped to HIV-1 genome (red line) in HIV-1 infected Jurkat cells (A) or primary CD4$^+$ T-cells (B). Baseline signal from the RNA-seq of input samples is shown as a black line. A schematic diagram of HIV-1$_{NL4-3}$ genome is shown above. TAR, transacting response element; RRE, Rev response element. Jurkat cells (A) or primary CD4$^+$ T-cells (B) were infected with HIV-1$_{NL4-3}$ and total RNA was extracted for m^{6}A-

(**Figure 4.6, continued**) seq at 72 or 96 hours post-infection (hpi), respectively. (C and D) Pie charts show the distribution of m^6A peaks in the 5′ UTR, coding DNA sequence (CDS), 3′ UTR, and noncoding regions of transcripts from uninfected and HIV-1-infected Jurkat T-cells (C) or primary $CD4^+$ T-cells (D). (E and F) GO graphs showing functional clusters from genes with unique m^6A peaks identified in HIV-1-infected Jurkat cells (E) or primary $CD4^+$ T-cells (F) when compared to uninfected cells. Data presented are the average results of duplicated samples (n = 2).

We next sought to investigate the interaction of m^6A readers with the methylated HIV-1 RNA, as they may have distinct functions on these foreign RNA species. We utilized CLIP-seq to map the binding sites of YTHDF1 on the HIV-1 genome in infected HeLa cells that overexpressed FLAG-tagged YTHDF1 and identified multiple CLIP peaks of YTHDF1-bound HIV-1 RNA, including the transactivation response element (TAR) in the 5' UTR leader sequence, in the *env* gene, in the *rev* gene, and the 3' UTR (Figure 4.7A). Some YTHDF1-binding sites (e.g. in the 3' UTR and the *rev* gene) partially overlap with the identified m^6A-containing regions in the HIV-1 genome. These data demonstrate that m^6A reader binds to m^6A-modified HIV-1 gRNA during viral infection.

Because HIV-1 RNAs are present in the cytoplasm and the nucleus of infected cells at different stages of viral lifecycle[153], we hypothesized that YTHDF1–3 proteins may directly interact with methylated HIV-1 RNAs, thereby affecting the metabolism and/or processing of the viral RNA. To examine the roles of YTHDF1–3 proteins in post-entry HIV-1 infection, we either overexpressed or knocked down the individual YTHDF proteins in human cell lines, and examined the effect on HIV-1 infection using a single-cycle, luciferase reporter HIV-1 pseudotyped with vesicular stomatitis virus G protein (VSV-G) to overcome the requirement of HIV-1 receptors during viral entry[154]. Compared to vector control cells, at 24-hour post-infection (hpi), overexpression of individual FLAG-tagged YTHDF1–3 proteins in HeLa cells (Figure 4.7B) significantly inhibited HIV-1 infection by approximately 10-fold (Figure 4.7C, $p<0.0005$) and drastically reduced the synthesis of full-length viral Gag protein (Pr55) (Figure 4.7B). In contrast,

stable knockdown of individual endogenous YTHDF1–3 proteins in HeLa cells (Figure 4.7D) significantly increased HIV-1 infection by four- to 14-fold ($p<0.05$) relative to control cells (Figure 4.7E). Similar observations were made for replication-competent HIV-1_{NL4-3} infection, as knockdown of YTHDFs notably increased HIV-1 infection in Jurkat cells (Figure 4.7F and G) and activated primary $CD4^+$ T-cells from healthy donors (Figure 4.7H and I). Overall, these data suggest that overexpression of YTHDF1–3 proteins significantly inhibits HIV-1 infection, while knockdown of these proteins efficiently promotes HIV-1 gene expression. Thus, endogenous YTHDF1–3 proteins in $CD4^+$ T-cells act as negative regulators to inhibit post-entry HIV-1 infection.

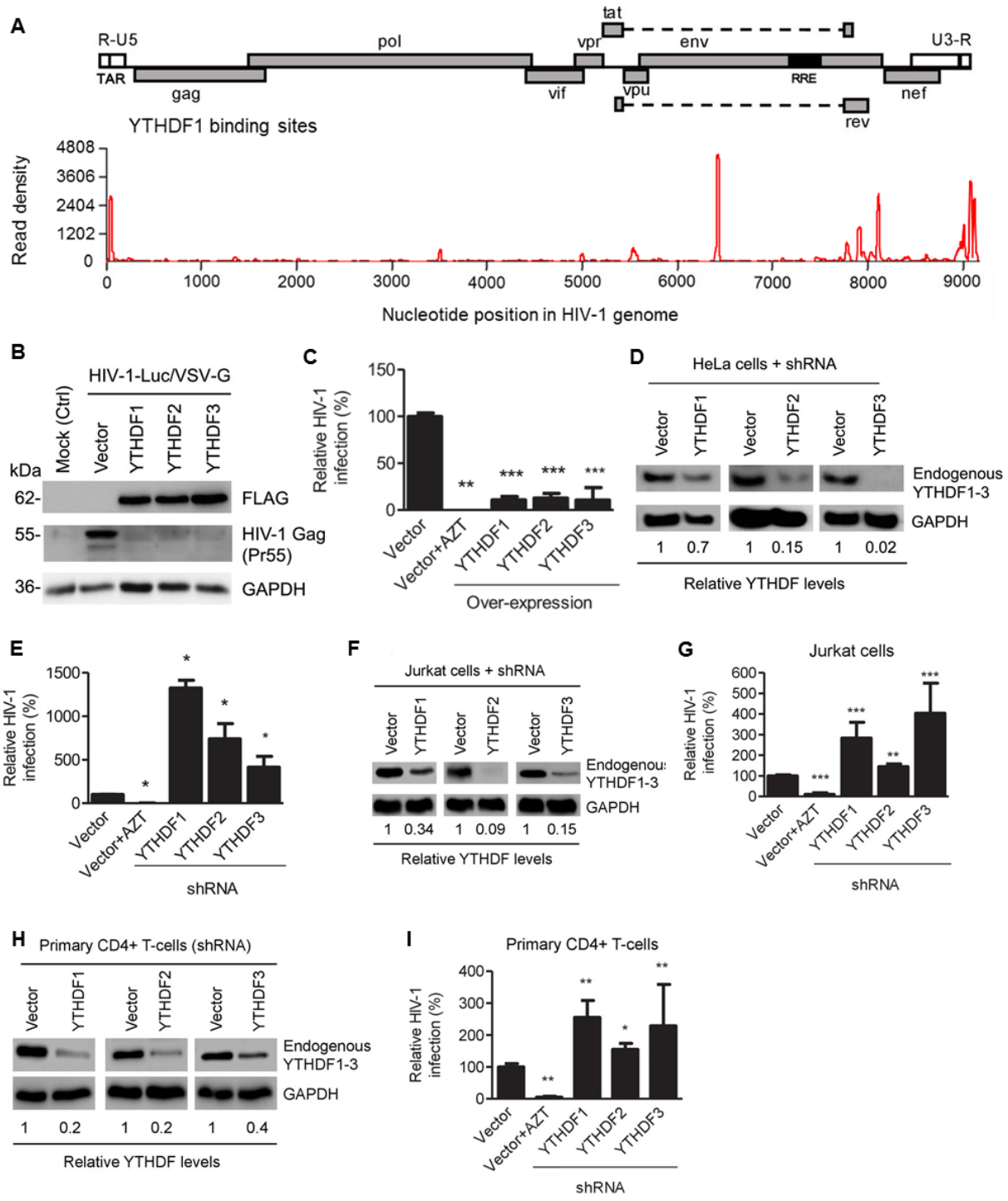


Figure 4.7 YTHDF proteins bind HIV-1 RNA and negatively regulate post-entry HIV-1 infection

(A) YTHDF1 binds to the HIV-1 gRNA. HeLa cells overexpressing individual FLAG-tagged YTHDF1 were infected with HIV-1-Luc/VSV-G at a multiplicity of infection (MOI) of 0.5 for 2 hours. CLIP assay was performed with anti-FLAG. The peaks represent the read density and the HIV-1 genome organization, and corresponding nucleotide positions are shown. (B and C) Overexpression of YTHDF1–3 proteins in HeLa cells significantly inhibits HIV-1 infection

(**Figure 4.7, continued**) compared to vector control cells. Overexpression of YTHDF1–3 proteins in HeLa cells was confirmed by immunoblotting (B). HeLa cells overexpressing YTHDF1–3 proteins were infected with HIV-1 Luc/VSV-G at an MOI of 0.5, with viral infection measured by luciferase activity at 24 hpi (C) and the expression of HIV-1 Gag and YTHDF1–3 proteins (FLAG-tagged) determined using immunoblotting at 24 hpi (B). (D and E) Individual knockdown of endogenous YTHDF1–3 proteins in HeLa cells significantly increases HIV-1 infection compared to vector control cells. HIV-1 infection assays were performed as described for panel B. (F) Individual knockdown of endogenous YTHDF1–3 proteins in Jurkat $CD4^+$ T cells was confirmed by immunoblotting. (G) Knockdown of YTHDF1–3 proteins significantly increases HIV-1 infection compared to vector control cells. (H) Individual knockdown of YTHDF1–3 proteins in activated primary CD4+ T-cells from a healthy donor. (I) Knockdown of YTHDF1–3 proteins notably increases HIV-1 infection compared to vector control cells. For all panels, GAPDH was used as a loading control and mock-infected vector control cells were used as a negative control. The reverse transcriptase inhibitor azidothymidine (AZT) treated cells were used as positive control for effective HIV-1 infection. The vector controls without AZT were set as 100% for relative infection. *p<0.05, **p<0.005, and ***p<0.0005, compared to vector control without AZT treatment. All results are shown as mean ± s.d. (n=3) and data presented are representative of at least three independent experiments.

To investigate the mechanisms of HIV-1 inhibition by YTHDF1–3 proteins, we assessed the stage of HIV-1 life cycle affected by the YTHDF1–3 proteins. We first measured levels of HIV-1 late reverse transcription (RT) products in infected cells, which represent the levels of the full-length viral cDNA[155]. Overexpression of each of the YTHDF1–3 proteins in HeLa cells significantly reduced the level of HIV-1 late RT products by four- to ten-fold (p<0.005) compared to the vector control cells at 24 hpi (Figure 4.8A), suggesting that the inhibition of viral reverse transcription contributes to the impairment of post-entry HIV-1 infection. In contrast, the knockdown of individual YTHDF1–3 proteins in HeLa cells elevated the levels of HIV-1 late RT products by two- to three-fold compared to vector control cells (Figure 4.8B). Furthermore, the level of HIV-1 2-LTR circles in infected HeLa cells, a surrogate marker of nuclear import of viral cDNA[156], was also significantly reduced over 10-fold (p<0.05) by overexpression of YTHDF1–3 proteins (Figure 4.8C), corresponding to the reduced late RT products observed in this experiment. Using our established Jurkat cell lines with stable knockdown of individual YTHDF1–3 proteins, we found that the levels of HIV-1 late RT products were significantly increased in YTHDF1 down-

regulated cells by 2.7-fold ($p<0.05$) compared to control cells, while the knockdown of YTHDF2 or YTHDF3 only increased late RT products by 20–30% (Figure 4.8D). These data suggest that YTHDF1–3 proteins could negatively regulate HIV-1 replication by inhibiting viral reverse transcription in HIV-1 infected cells.

We then hypothesized that YTHDF1–3 proteins could inhibit the reverse transcription of HIV-1 gRNA through sequestering HIV-1 gRNA. To test this hypothesis, we used a single-cycle, VSV-G-pseudotyped HIV-1 to infect HeLa cells overexpressing individual YTHDF1–3 proteins or vector control cells, immunoprecipitated YTHDF proteins from the infected cells at 3 hpi, and then measured HIV-1 gRNA levels in the IP samples. The quantification of the HIV-1 gRNA by qRT-PCR revealed a strong and specific association ($p<0.005$) of HIV-1 gRNA with YTHDF proteins in HIV-1-infected YTHDF1–3-expressing cells compared to control cells (Figure 4.8E). To examine the impact of YTHDF1–3 on HIV-1 *gag* RNA kinetics, we quantified HIV-1 *gag* RNA levels in YTHDF1–3-expressing HeLa cells and vector control cells over a time course of 6–24 hpi. The relative levels of *gag* RNA in HIV-1 infected cells were normalized to that of the vector control cells at 6 hpi. In the control cells, compared to 6 hpi (set as 100%), the level of *gag* RNA was reduced to 40% at 12 hpi and then increased to 80% at 24 hpi (Figure 4.8F), suggesting degradation of HIV-1 gRNA at 12 hpi during the reverse transcription and then increased *gag* mRNA at 24 hpi during viral gene transcription. In contrast, the levels of *gag* RNA in YTHDF1–3-expressing cells were reduced to 40% at 12 hpi and to 13–25% at 24 hpi ($p<0.0005$) compared to that of the vector control cells at 6 hpi (Figure 4.8F). These data suggest that YTHDF1–3 proteins can sequester and degrade HIV-1 RNA in infected cells, thereby leading to inhibition of HIV-1 reverse transcription.

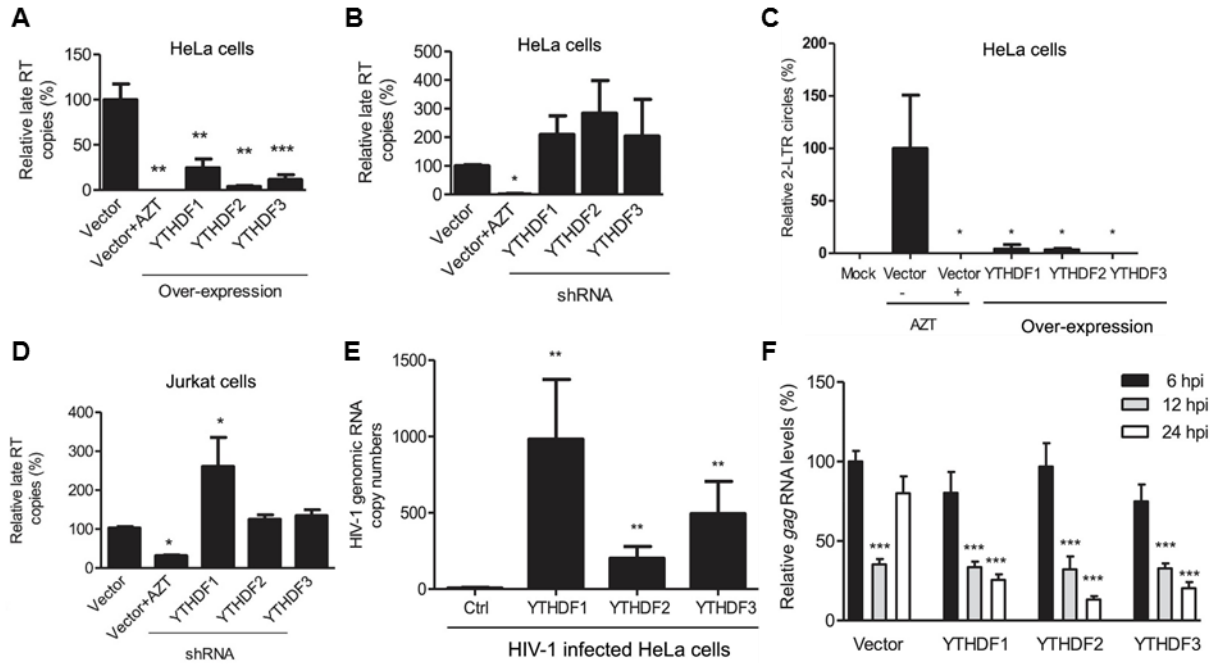


Figure 4.8 YTHDF proteins inhibit HIV-1 infection by sequestering viral RNA and blocking viral reverse transcription

(A and B) HeLa cells over-expressing (A) or knocking-down (B) individual YTHDF1–3 proteins were infected with HIV-1-Luc/VSV-G at an MOI of 0.5. Genomic DNA was isolated from the cells 24-hour post-infection and HIV-1 late reverse transcription (RT) products were quantified by qPCR. (C) YTHDF proteins reduce the formation of HIV-1 2-LTR circles in infected HeLa cells. At 24-hour post-infection, DNA was isolated from the cells and the 2-LTR circles were analyzed by qPCR and normalized to GAPDH levels. (D) Knocking-down individual YTHDF1–3 proteins in Jurkat cells and qPCR quantifying HIV-1 late RT products revealed similar trends in RT copy change. AZT treated vector control cells were used as a positive control for effective HIV-1 inhibition. (E) HIV-1 gRNA is bound by YTHDF1–3 proteins expressed in HeLa cells. HeLa cells stably overexpressing FLAG-tagged YTHDF1–3 proteins or empty vector control cells (Ctrl) were infected with HIV-1-Luc/VSV-G at an MOI of 5 for 3 hours. Cell lysates were immunoprecipitated with anti-FLAG, RNA was extracted and HIV-1 *gag* RNA levels were measured. (F) YTHDF1–3 affect HIV-1 *gag* RNA kinetics. HIV-1 *gag* RNA levels in YTHDF1–3-expressing HeLa cells were quantified by qRT-PCR. The relative levels of *gag* RNA in infected cells were normalized to that of the vector control cells at 6 hpi. *$p<0.05$, **$p<0.005$, and ***$p<0.0005$, compared to vector control cells. All results are shown as mean ± s.d. ($n=3$) and data presented are representative of at least three independent experiments.

In addition to the functions of m^6A readers, we also examined the role of the m^6A writers and erasers in HIV-1 protein expression and viral release in virus-producing cells. We knocked down endogenous METTL3, METTL14, or both, in HEK293T cells using siRNA, and then transfected the cells with an HIV-1 proviral DNA plasmid (pNL4-3). We determined the levels of

HIV-1 Gag protein expression in the cells and the capsid p24 protein released in the supernatants. Interestingly, we found that partial knockdown of METTL3, METTL14, or both, inhibited HIV-1 Gag expression in the cells by 60–70% (Figure 4.9A), and reduced the levels of HIV-1 p24 release by 30–50% compared to control cells (Figure 4.9B). Similarly, we knocked down endogenous ALKBH5, FTO, or both, in HEK293T cells using siRNA, and then transfected the cells with the pNL4-3 plasmid. We determined the levels of HIV-1 Gag protein expression in the cells and the capsid p24 protein released in the supernatants. Interestingly, we found that partial knockdown of FTO significantly promoted HIV-1 Gag synthesis in the cells by 2.5- to 6.5-fold (Figure 4.9C), and increased the levels of HIV-1 p24 release by two- to three-fold compared to control cells (Figure 4.9D). Thus, the m^6A modification of HIV-1 RNA can enhance HIV-1 protein synthesis. Our results are in agreement with a recent report[143] showing that silencing of the m^6A writers (METTL3 and METTL14) or the eraser ALKBH5 decreases or increases HIV-1 p24 expression in the infected MT4 cells, respectively. These results suggest that modulation of m^6A writers and erasers affect the efficiency of HIV-1 protein synthesis and that elevated m^6A levels on HIV-1 RNA could facilitate of viral protein translation and release.

In summary, we found that HIV-1 RNA is m^6A-methylated in infected cells, and that binding of YTHDF1–3 proteins to methylated HIV-1 RNA inhibits viral reverse transcription and gene expression. In contrast, partial knockdown of the m^6A writers decreased HIV-1 Gag synthesis and viral release, while partial knockdown of FTO had the opposite effects, suggesting that m^6A modification of HIV-1 RNA could enhance HIV-1 protein synthesis and viral release. Based on our results, we proposed a working model demonstrating the two aspects of m^6A regulation in HIV infection (Figure 4.9E): YTHDF proteins inhibit post-entry HIV-1 infection by sequestering viral gRNA and blocking viral reverse transcription; yet after viral genome incorporation into the host

genome, newly produced viral RNAs resemble host mRNAs and benefit from m^{6}A modification, resulting in the promotion of viral protein translation and release.

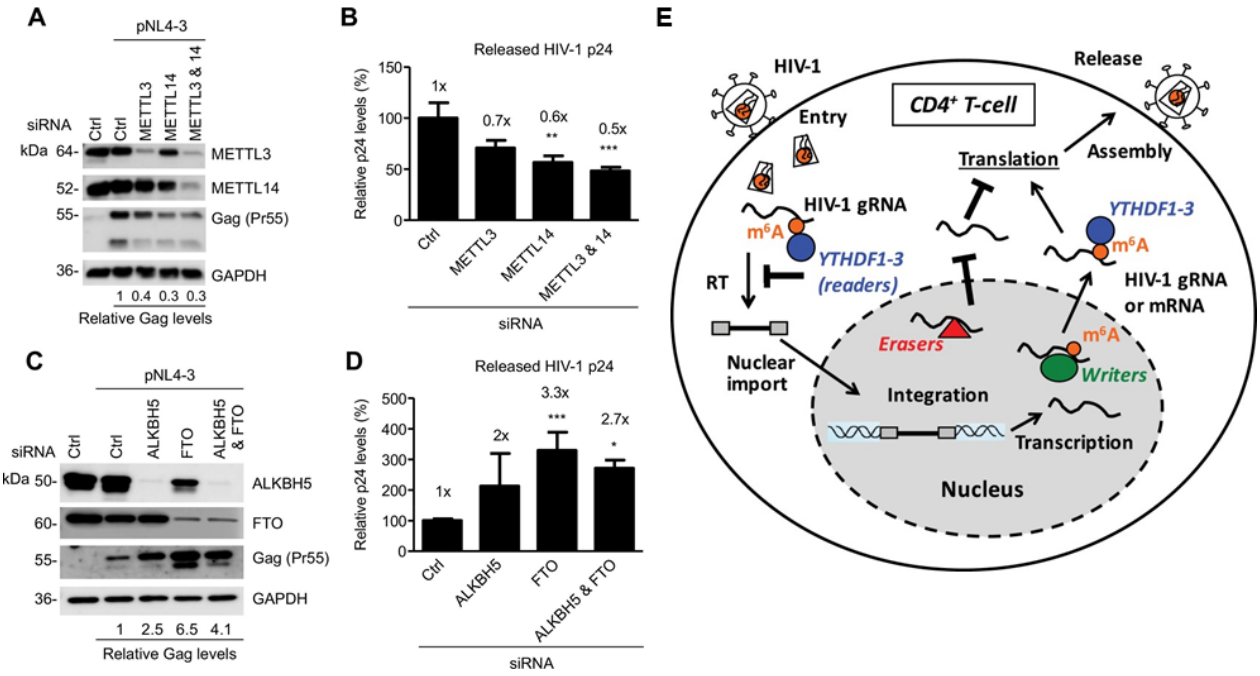


Figure 4.9 m^{6}A writers and erasers affect HIV-1 protein translation and release
(A and B) Individual or combined knockdown of endogenous METTL3 and METTL14 inhibits HIV-1 Gag protein expression (A) and release (B). HEK293T cells were transfected with indicated siRNA, and then with an HIV-1 proviral DNA plasmid (pNL4-3). Cells and supernatants were collected for analyses at 36 hours post-transfection. Expression of METTL3, METTL14 and HIV-1 Gag proteins in the transfected HEK293T cells were detected by immunoblotting. (C and D) Knockdown of endogenous ALKBH5, FTO, or both promotes HIV-1 Gag protein expression (C) and release (D). HEK293T cells were transfected with indicated siRNA, and then with pNL4-3. Cells and supernatants were collected at 36 hours post-transfection. Expression of ALKBH5, FTO and HIV-1 Gag proteins in the cells was detected by immunoblotting. GAPDH was used as a loading control. Relative levels of Gag expression were normalized to GAPDH levels. HIV-1 capsid p24 levels in supernatants were measured by ELISA. The relative levels (%) are also shown. *p<0.05, **p<0.005, and ***p<0.0005, compared to the siRNA control. The results are shown as mean ± s.d. (n=3) and data presented are representative of three independent experiments. (E) Proposed mechanisms and dynamics of m^{6}A modification on HIV-1 RNA regulating viral infection in cells.

4.2.2.2 m^{6}A in Zika virus

Zika virus (ZIKV) is a mosquito-borne flavivirus first discovered in 1947, and has spread rapidly since the 2015 outbreak in Brazil and caused major damage globally. ZIKV uses a positive

(+) single-stranded RNA as its genome, which encodes a single polyprotein precursor that is cleaved by viral and host proteases to produce three structural and seven nonstructural proteins. Although our understanding of the molecular mechanisms involved in ZIKV infection of human cells has increased dramatically in the past few years, key determinants of ZIKV pathogenicity, such as cell-type specificity, mode of entry, and host factors essential for replication, are still largely unknown. In particular, there is a large gap in our understanding of the genetic and epigenetic regulatory mechanisms governing the ZIKV viral life cycle and viral interactions with host cells.

To investigate whether ZIKV viral RNA is post-transcriptionally methylated, human HEK293T cells were infected with MR766 ZIKV for 48 hours and viral RNA was purified from virions released into the supernatant. The presence of internal m^6A and 2′-*O*-Me nucleosides (m^6A, Am, Cm, Um, and Gm) in ZIKV RNA was quantified by LC-MS/MS. Interestingly, ZIKV RNA contained a high level of methylation analyzed (Figure 4.10A). All four nucleotides contained 2′-*O*-Me groups, with U and C showing the most extensive modification, followed by A and G. In particular, the abundance of m^6A (about 3% of total adenosine) was strikingly high compared to mammalian mRNAs, where m^6A accounts for only ∼0.4%–0.5% of adenosines[66]. We next examined the m^6A patterns on ZIKV RNA by m^6A-seq and identified twelve discrete m^6A peaks spanning the full length of ZIKV RNA (Figures 4.10B). We used a stringent peak calling method (false discovery rate [FDR] < 0.01) and sequenced the methylome of ZIKV with very high depth, leading to the identification of statistically significant peaks. Notably, a cluster of m^6A peaks was observed in the NS5 coding region and one peak was present in the 3′ UTR region. These data revealed that ZIKV viral RNA is indeed extensively decorated with modifications.

As there have been no reports of flaviviral enzymes with internal m^6A methyltransferase activity, it is possible that ZIKV RNA adenosines are modulated by host m^6A writers and erasers. To investigate this, we transduced 293T cells with shRNAs to knock down m^6A writers (METTL3 and METTL14) and erasers (ALKBH5 and FTO), followed by examining m^6A abundance in ZIKV RNA by MeRIP-qPCR. The density of m^6A on viral RNA was decreased by knocking down m^6A writers and increased by depletion of m^6A erasers, compared with the abundance in cells expressing the control shRNA (Figure 4.10C). These results provide evidence that deposition and removal of m^6A on ZIKV RNA is mediated by host enzymes.

Since the replication of positive single-stranded flavivirus takes place in the cytoplasm[157], we examined the subcellular localization of METTL3, METTL14, and ALKBH5 proteins in mock- and ZIKV-infected 293T cells. Although METTL14 and ALKBH5 were more abundant in the nucleus, all three enzymes were readily detectable in both the nuclear and cytoplasmic fractions (Figure 4.10D). Furthermore, there was no apparent redistribution of the enzymes upon ZIKV infection. Together, these results show that methylation and demethylation of viral RNA adenosine occur in the cytoplasm of the host cell.

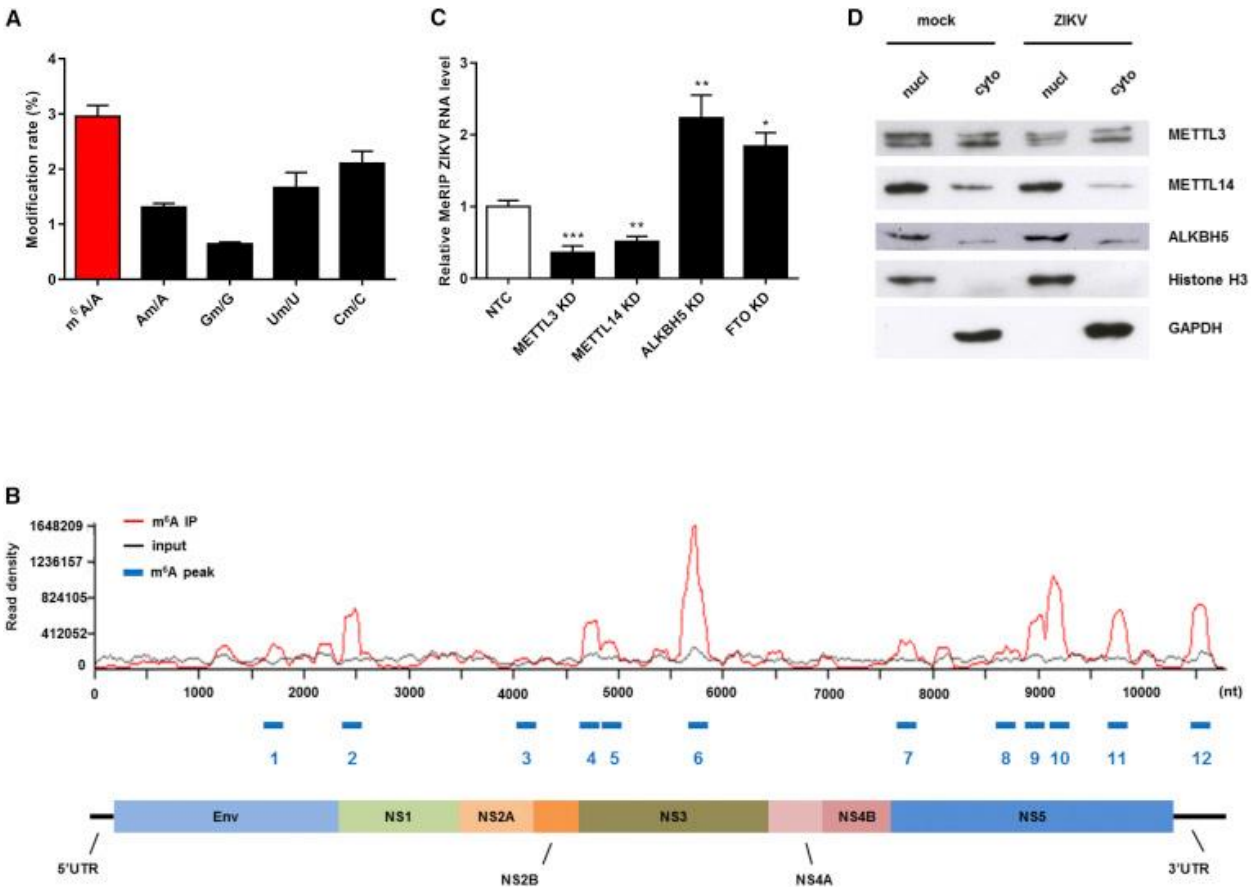


Figure 4.10 ZIKV RNA contains extensive modifications, including m^{6}A regulated by host enzymes

(A) LC-MS/MS quantification of m^{6}A and 2′-*O*-Me modifications on all four bases of ZIKV genomic RNA (RNA, 50 ng/sample). Data are expressed as the ratio of modified to unmodified bases (m^{6}A/A, Am/A, Gm/G, Um/U, and Cm/C). $n = 3$. (B) m^{6}A-seq of ZIKV RNA showing the distribution of m^{6}A reads mapped to the ZIKV genome (red line). The baseline signal from input samples is shown as a black line, and the m^{6}A peaks are shown as blue rectangles along the x axis. A schematic diagram of the ZIKV genome is shown below to indicate the location of the m^{6}A-enriched sequences. Data are representative of $n = 3$ determinations. (C) Modulation of ZIKV RNA methylation by METTL3/METTL14 and ALKBH5. 293T cells were transfected with a non-targeting control shRNA (NTC) or shRNAs targeting METTL3, METTL14, ALKBH5, or FTO (knockdown, KD). RNA was isolated by Me-RIP and quantified by qPCR. $n = 3$. (D) Localization of METTL3, METTL14, and ALKBH5 in the nucleus and cytoplasm of ZIKV-infected cells. Nuclear and cytoplasmic fractions of mock- or ZIKV-infected 293T cells were subjected to western blot analysis using antibodies against METTL3, METTL14, and ALKBH5 enzymes. Histone H3 and GAPDH were probed as controls for each fraction. Data are the mean ± s.e.m. of the indicated number of replicates. Student's t test, $^{*}p < 0.05$, $^{**}p < 0.005$, $^{***}p < 0.0005$.

We next investigated the effects of modulating different m^{6}A regulators on ZIKV viral processes. To evaluate whether perturbation of ZIKV m^{6}A directly or indirectly affects ZIKV

replication efficiency, we transduced 293T cells with METTL3, METTL14, ALKBH5, FTO, or control shRNAs and found that the viral titer, ZIKV RNA levels in cell supernatants, and expression of ZIKV envelope protein were significantly increased by METTL3 and METTL14 knockdown and decreased by ALKBH5 knockdown, respectively (Figure 4.11A–C). In the meantime, we also analyzed the effects of RNAi knocking down YTHDF1–3 on ZIKV replication. We observed an increase in ZIKV replication when YTHDF proteins were depleted (Figure 4.11D). Notably, YTHDF2 knockdown resulted in the largest increase in ZIKV replication as compared to YTHDF1 and YTHDF3. These results indicate that expression levels of m^6A writers and readers are negatively correlated with ZIKV viral replication and that modulation of the m^6A RNA landscape by host enzymes profoundly influences viral processes.

We then sought to reveal the molecular mechanism of the observed negative functions of m^6A regulators during ZIKV replication. We first studied whether YTHDF proteins could bind to ZIKV RNA and whether they affect RNA metabolism and viral replication. We overexpressed FLAG-tagged YTHDF proteins in 293T cells and examined viral RNA abundance and YTHDF protein binding to methylated RNA. Overexpression of all three YTHDF proteins decreased ZIKV RNA expression, with YTHDF2 having the greatest effect (Figure 4.11E). Furthermore, immunoprecipitation of the YTHDF proteins with anti-FLAG antibodies followed by qRT-PCR revealed that YTHDF proteins indeed bound ZIKV RNA, with YTHDF2 IP fraction containing the greatest amount of viral RNA (Figure 4.11F), suggesting that host m^6A readers may bind and sequester ZIKV viral RNA to interfere with ZIKV infection. To determine whether modulation of m^6A abundance on ZIKV RNA affects its binding to YTHDF2, we examined cells expressing control, METTL3, METTL14, or ALKBH5 shRNA. Notably, the association of ZIKV RNA with YTHDF2 was significantly reduced by METTL3 and METTL14 silencing and, conversely,

significantly increased by ALKBH5 knockdown (Figure 4.11G). All results taken together, we concluded that m^{6}A readers may sequester ZIKV viral RNA in an m^{6}A-dependent way and reduce the efficiency of viral replication.

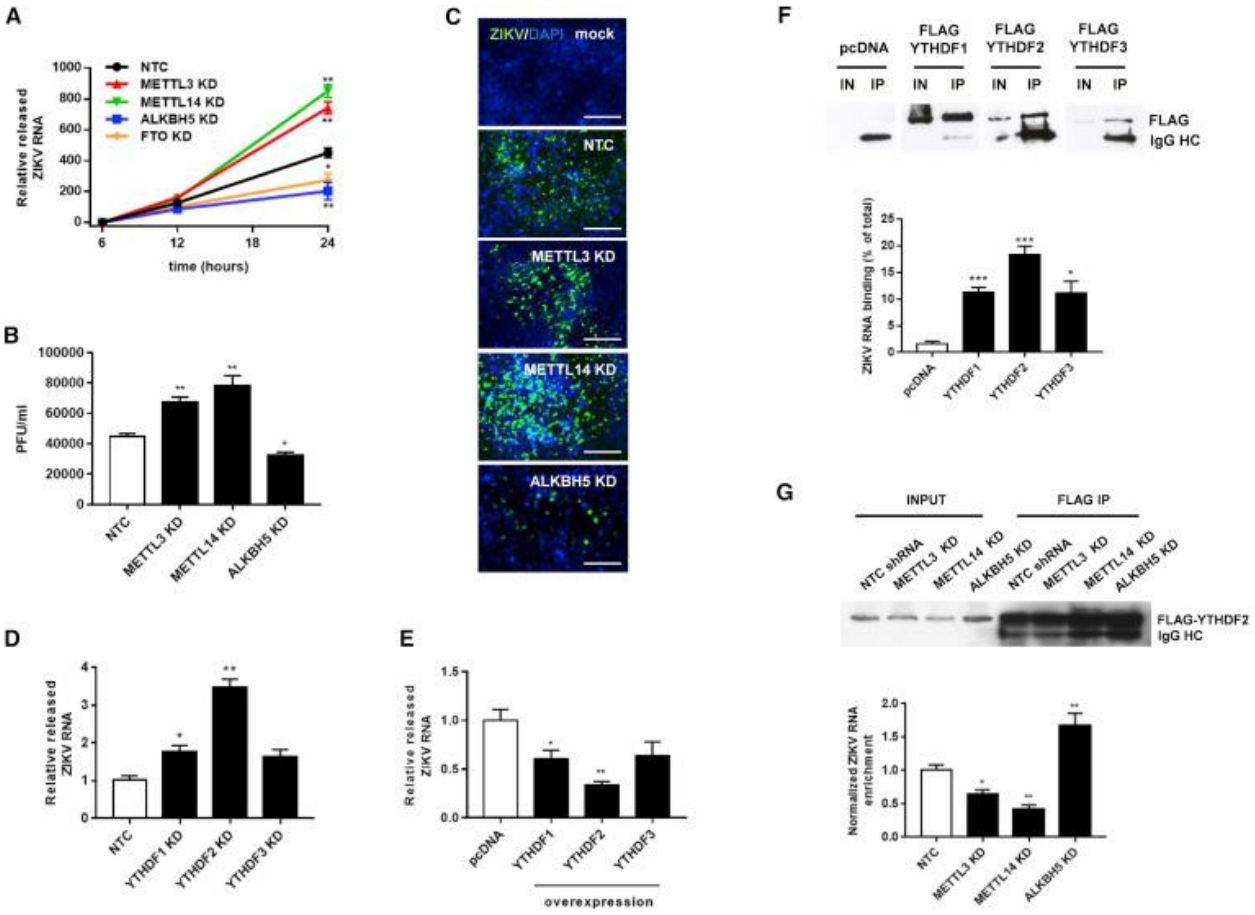


Figure 4.11 m^{6}A regulators modulate ZIKV life cycle

(A) Enhancement of ZIKV replication by METTL3/METTL14 silencing and reduction by ALKBH5/FTO silencing. 293T cells expressing a non-targeting control shRNA (NTC) or shRNAs targeting METTL3, METTL14, ALKBH5, or FTO shRNA (KD) were infected with ZIKV. Supernatants were harvested 6, 12, and 24 hours later for quantification of ZIKV RNA by qRT-PCR. *n* = 3. (B) Viral titers (plaque-forming units [PFU]/mL) at 24-hour post-infection. Cells were treated as described in (A). *n* = 3. (C) Immunostaining of viral envelope protein in cells treated as described in (A). Scale bars, 100 μm. (D) Enhancement of ZIKV RNA expression by YTHDF proteins silencing. 293T cells were transduced with shRNAs targeting YTHDF1–3 or control shRNA. Supernatants were harvested 24 hours later for quantification of ZIKV RNA by qRT-PCR. *n* = 5. (E) Decrease in ZIKV RNA expression by overexpression of YTHDF proteins. 293T cells were transfected with FLAG-tagged YTHDF1–YTHDF3 or control pcDNA vectors. Supernatants were harvested 24 hours later for quantification of ZIKV RNA by qRT-PCR. *n* = 5. (F) YTHDF proteins bind ZIKV RNA. 293T cells transfected as described for (E) were immunoprecipitated with an anti-FLAG antibody and immunoblotted for FLAG proteins (top).

(**Figure 4.11, continued**) "IN" (input) lanes contained 5% of the lysate. ZIKV RNA in the FLAG IP fraction was quantified by qRT-PCR and normalized to the percentage of total intracellular ZIKV RNA (bottom). $n = 3$. (G) Reduction and enhancement of YTHDF2-RNA binding by RNA methylation status. 293T cells were transfected with control or FLAG-YTHDF2 overexpression vector and co-transfected with the indicated shRNAs. Lysates were immunoprecipitated with an anti-FLAG antibody and immunoblotted for FLAG protein (top). Input lanes contained 5% of the lysate. ZIKV RNA in YTHDF2 immunoprecipitates was quantified by qRT-PCR and normalized to the level in cells expressing NTC shRNA (bottom). $n = 3$. All data are the mean ± s.e.m. of the indicated number of replicates. Student's t test, $^{*}p < 0.05$, $^{**}p < 0.005$, $^{***}p < 0.0005$.

Since viral infection has a profound effect on gene and protein expression in the host cell, we examined the abundance and distribution of m^6A on cellular transcripts by performing m^6A-seq on mRNA isolated from uninfected and ZIKV-infected 293T cells. Metagene analysis showed that ZIKV infection increased m^6A levels in the 5′ UTR regions of the transcriptome and correspondingly decreased the levels in the 3′ UTRs (Figure 4.12A). This change may represent a cellular response to the stress of viral infection, similar to the previously described change in m^6A deposition in response to heat shock[83]. To examine the distribution of ZIKV-induced changes in more detail, we compared mRNA m^6A peaks that were either newly gained or lost after ZIKV infection. We observed that the newly emerged m^6A modifications were preferentially deposited in 5′ UTR and CDS and depleted in exon junctions and 3′ UTR regions, compared to the lost m^6A peaks (Figure 4.12B), suggesting that ZIKV infection might affect gene translation, alternative splicing, and mRNA stability as a consequence of differential deposition of m^6A. GO analysis of the genes with changed m^6A peaks identified a number of immune-related categories among the most enriched pathways in both the newly gained and lost m^6A modifications (Figure 4.12C). We have also performed a motif analysis of the newly emerged and lost peaks specifically to determine if there was any change in the top motifs between these two groups. The motif usage changes seem to occur on the overall level, with newly emerged ones preferring "AAC," while lost ones prefer "GAC" (Figure 4.12D). The results showed clear differences in the consensus sequences of m^6A, highlighting the possibility that the substrate specificity of m^6A-related enzymes may change upon

viral infection. Thus, ZIKV infection appears to be sensed by the host methylation and demethylation machineries, which then initiate a directed reprogramming of the post-transcriptional landscape of cellular mRNAs.

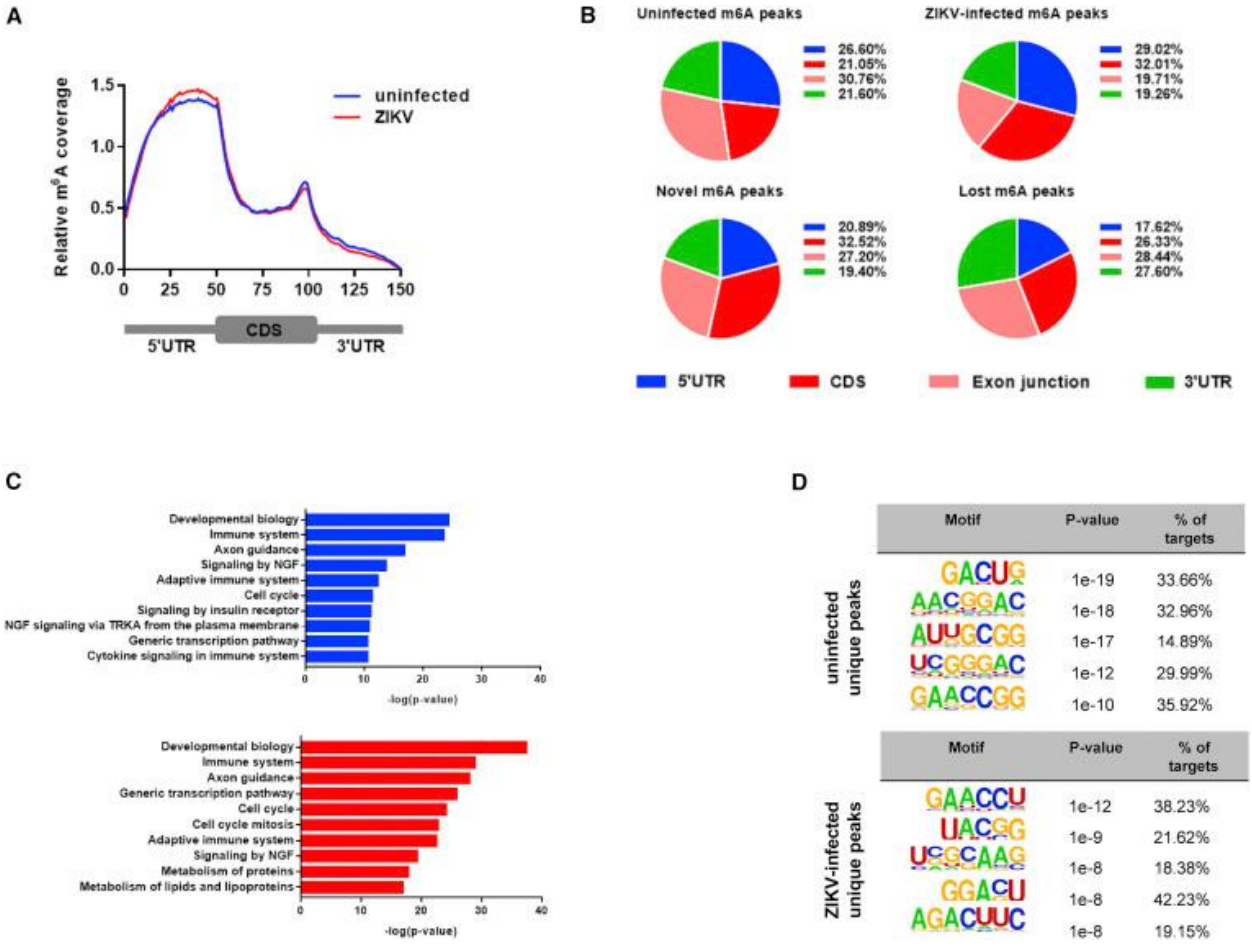


Figure 4.12 ZIKV infection influences RNA methylation of host transcripts
(A) Metagene analysis of normalized m^{6}A peak distribution along a reference mRNA. (B) Distribution of m^{6}A peaks in the 5′ UTR (blue), coding sequence (CDS, red), exon junction (pink), and 3′ UTR (green) of host cell RNA transcripts. 293T cells were mock or ZIKV infected, and m^{6}A peaks in total cellular RNA were analyzed at 24 hours after infection. Charts show the proportion of m^{6}A peaks in the indicated regions in uninfected and ZIKV-infected cells (top) and the appearance of newly emerged m^{6}A peaks or loss of existing m^{6}A peaks after ZIKV infection (bottom). Representative of n = 2 determinations. (C) Gene set enrichment analysis (GSEA) of reactome analysis of pathways associated with newly emerged m^{6}A modifications (top, blue) and loss of existing m^{6}A modifications (bottom, red) at 24 hours after ZIKV infection of 293T cells. The top ten enriched categories for each condition are shown. (D) Motif analysis to identify consensus sequences for m^{6}A methylation in uninfected and ZIKV-infected 293T cells. The top five motifs for each are shown.

In summary, we investigated the role of m^6A during ZIKV infection and found that the depletion or overexpression of the central RNA methylation enzymes impact viral replication, demonstrating that the host RNA methyltransferase machinery acts as a negative post-transcriptional regulator of ZIKV virus. Moreover, ZIKV RNA binding to YTHDF proteins indicates another regulatory layer represented by the m^6A binding proteins. We also explored the m^6A landscape of cellular mRNAs in response to ZIKV infection and identified two classes of transcripts that undergo specific m^6A deposition or loss of the modification. This evidence suggests that m^6A regulators may be able to sense and respond to viral infection and modulate gene expression at the post-transcriptional level. The overall effects of m^6A on viral replication may be due to a combination of direct events regulating viral RNA metabolism and indirect post-transcriptional regulation of host RNAs, which may act as pro- or anti-viral factors.

4.2.2.3 m^6A in respiratory syncytial virus and influenza virus

Respiratory syncytial virus (RSV) and influenza virus are both negative-sense (-) single strand RNA viruses, with the distinction that influenza enters the host nucleus for replication while RSV stays in the cytoplasm for its entire life cycle. We applied the similar strategy previously described and investigated the impacts of m^6A regulators on the replication of these (-) ssRNA viruses.

Starting with the study of RSV, we used m^6A-seq to characterize both the host and viral methylome in RSV-infected cells. The results revealed several m^6A peaks across both the (-) ssRNA genome and (+) ssRNA antigenome of RSV, and interestingly, the peaks are distinct on these two types of viral RNAs, suggesting different methylation mechanism (Figure 4.13A). RSV infection also induced significant changes to the host methylome, with more m^6A peaks emerging at the 5' UTR and 3' UTR of host transcriptome and evident loss of CDS peaks (Figure 4.13B and

C). GO analysis of genes containing emerged or lost m^{6}A peaks showed these two groups of genes cluster very differently in functional terms (Figure 4.13D and E). Genes with emerged m^{6}A peaks cluster in a much more interconnected fashion with major enrichment in metabolism and immune response.

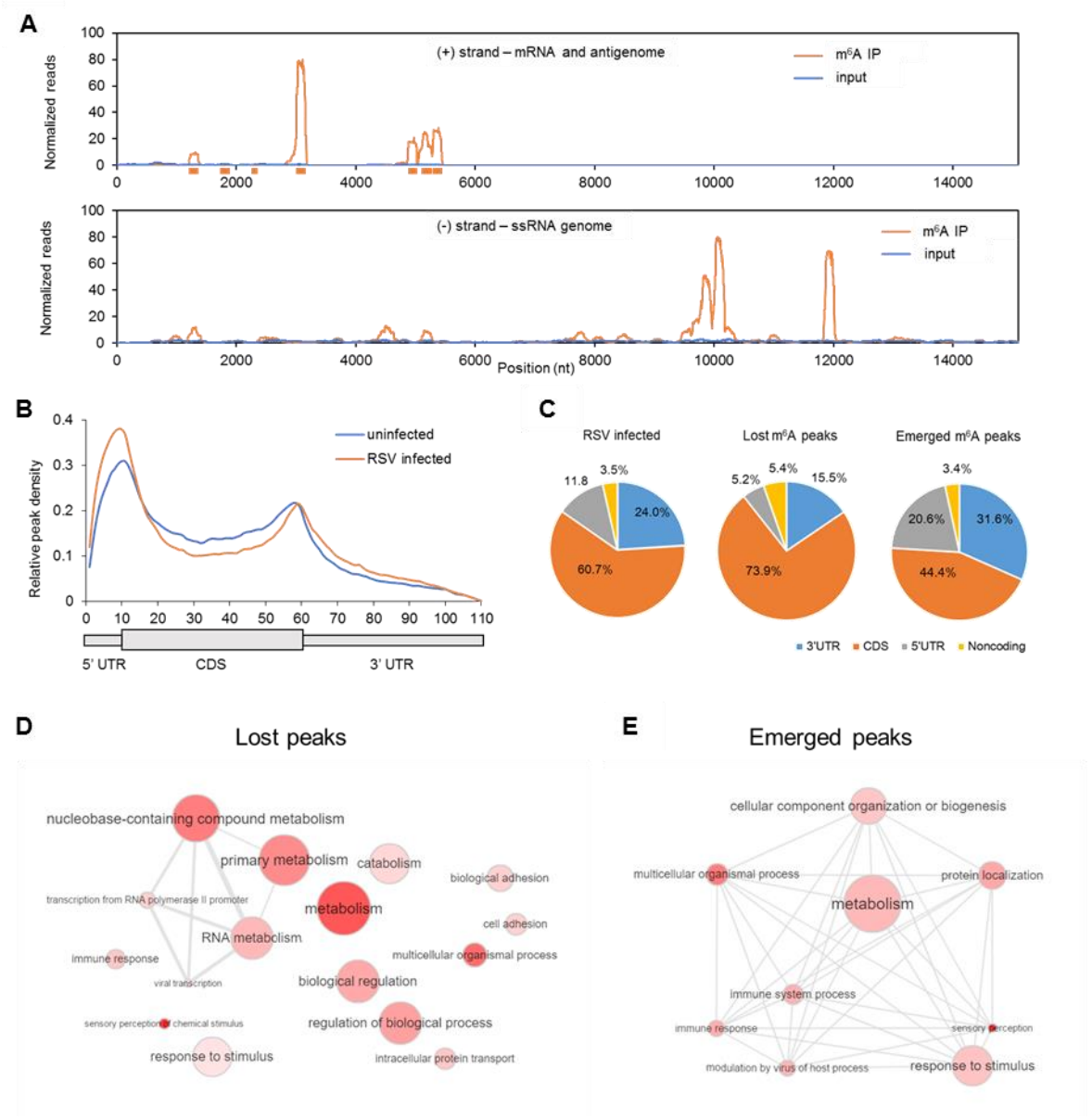


Figure 4.13 RSV RNA contains m^{6}A and induces the methylome change in host transcripts
(A) m^{6}A-seq of RSV RNA showing the distribution of m^{6}A reads (red line) mapped to the RSV antigenome (upper panel) and genome (lower panel). The baseline signal from input samples is shown as a black line. (B) Metagene analysis of normalized m^{6}A peak distribution along the human

(**Figure 4.13, continued**) reference mRNA in control and infected cells. (C) Distribution of m^6A peaks in the 5′ UTR, CDS, and 3′ UTR of host cell RNA transcripts. Charts show the proportion of m^6A peaks in the indicated regions in uninfected and RSV-infected cells and the appearance of newly emerged m^6A peaks or loss of existing m^6A peaks after RSV infection. (D and E) GO graphs showing functional clusters from genes with lost m^6A peaks (D) or emerged m^6A peaks (E) identified in RSV-infected cells. Data presented are the average results of duplicated samples (n = 2).

Subsequent investigation of the effects of modulating m^6A factors on viral infection revealed that m^6A seems to positively regulate the replication of RSV. The overexpression or knockdown of m^6A writers enhances or decreases the viral protein production, respectively (Figure 4.14A and B); while the overexpression or knockdown of m^6A erasers reduces or increases the viral production, respectively (Figure 4.14C and D). Additionally, overexpression of m^6A readers YTHDF proteins improves the viral protein production in RSV-infected cells while knockdown of YTHDFs notably reduces it (Figure 4.14E and F). Using GFP-expressing RSV strain to evaluate the replication rate of RSV in the infected cells by fluorescence microscopy, we observed that overexpression of YTHDFs accelerates RSV replication whereas knockdown slows it down (Figure 4.14G and H). These data suggest that expression levels of m^6A writers and readers are positively correlated with RSV viral replication, while m^6A eraser negatively affects RSV infection.

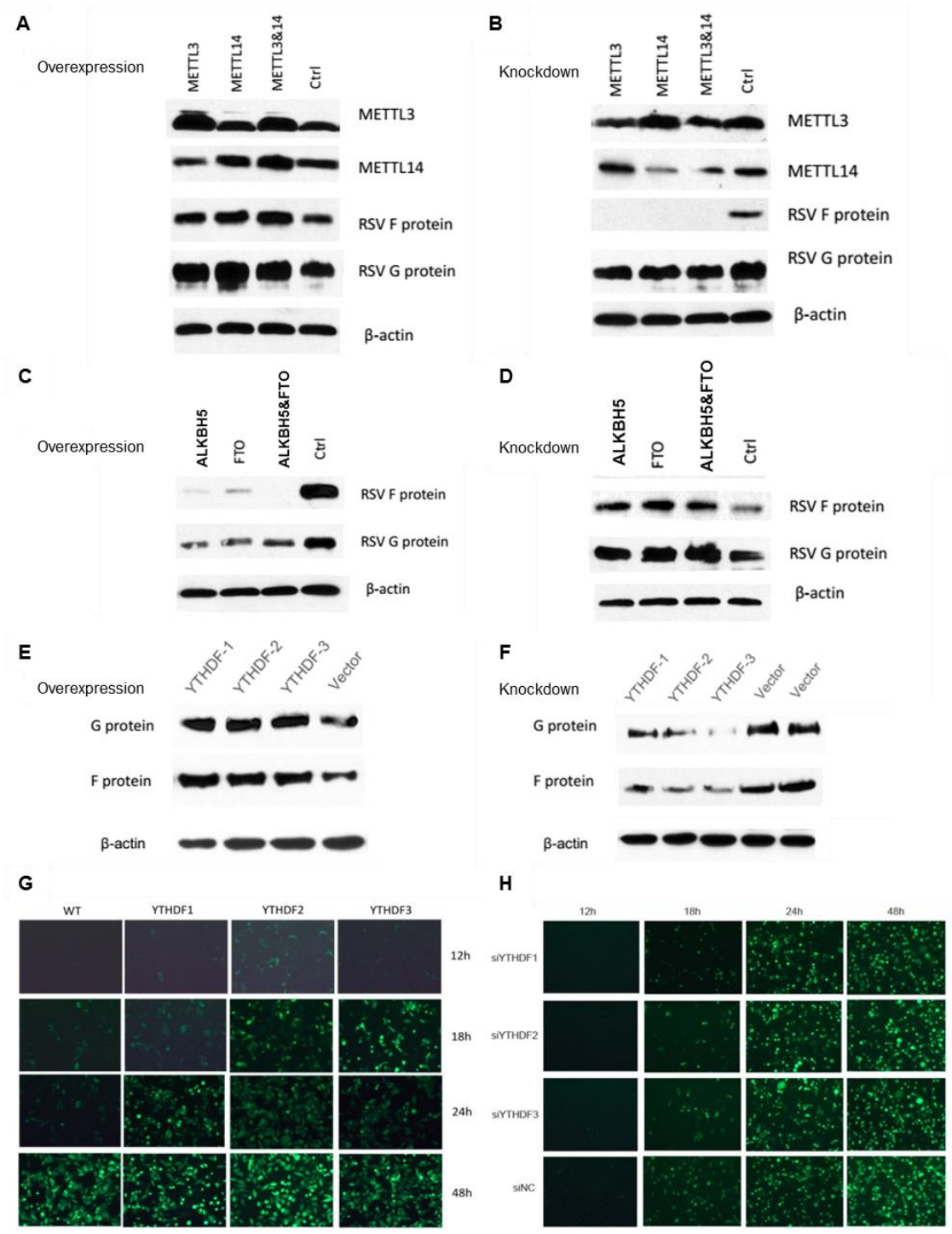
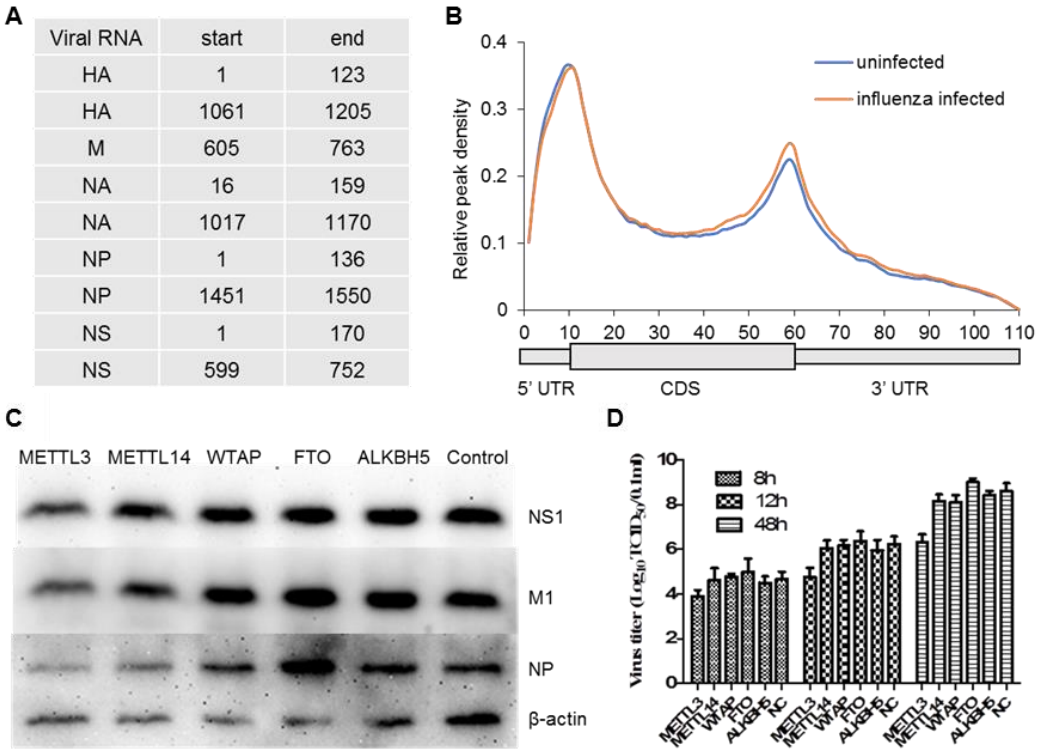


Figure 4.14 RSV protein expression and replication are impacted by host m^{6}A regulators
(A and B) The effects of overexpressing (A) or knocking down (B) m^{6}A writers on RSV protein expression levels. (C and D) The effects of overexpressing (C) or knocking down (D) m^{6}A erasers on RSV protein expression levels. (E and F) The effects of overexpressing (E) or knocking down (F) m^{6}A readers on RSV protein expression levels. (G and H) The effects of overexpressing (G) or knocking down (H) m^{6}A readers on RSV replication rates evaluated by GFP-expressing cell numbers.

We then moved on to the study of influenza virus and similarly applied m^{6}A-seq to study the methylome changes in the host and viral genomes upon infection. Influenza has a unique segmented RNA genome containing eight different strands of genomic ssRNA, and we have identified nine m^{6}A peaks from five strands from them (Figure 4.15A). The host methylome was also changed in the infected cells, with more m^{6}A peaks appearing at the 3' UTR in the transcriptome (Figure 4.15B). Using siRNA to knockdown each m^{6}A regulators, we found that the depletion of m^{6}A writers decreases viral protein production and replication while removing m^{6}A erasers improves both processes (Figure 4.15C and D). These data agree with the previous observations we made in RSV, suggesting similar positive functions of m^{6}A in influenza virus.



Viral RNA	start	end
HA	1	123
HA	1061	1205
M	605	763
NA	16	159
NA	1017	1170
NP	1	136
NP	1451	1550
NS	1	170
NS	599	752

Figure 4.15 The genomes of influenza are methylated and its infection is affected by m^{6}A regulators

(A) m^{6}A-seq of RNA from influenza-infected cells showing the existence of m^{6}A peaks in five out of eight influenza genomic RNAs. (B) Metagene analysis of normalized m^{6}A peak distribution along the human reference mRNA in control and infected cells. (C and D) Knockdown experiments of m^{6}A regulators affect the viral protein expression (C) and replication rate (D) of influenza virus.

To investigate the functions of m^6A readers in the influenza viral processes, we applied immunofluorescence microscopy to study the subcellular location of each m^6A reader upon influenza infection. Intriguingly, we discovered that the invasion of influenza virus induces an evident relocalization of cytoplasmic YTHDF proteins into the nucleus (Figure 4.16A and B). From the IF images in mock infected cells, it was evident that all YTHDF proteins were exclusively localized to the cytoplasmic region (Figure 4.16A). Yet after influenza infection, YTHDF proteins were observed to gradually relocalize into the nucleus, with YTHDF1 and YTHDF2 imported to a larger extent than YTHDF3 (Figure 4.16B). The patterns were very distinct compared to the viral protein HA, which primarily localizes in the cytoplasm. These observations were verified through a separate analysis using Western blotting in the cytoplasmic or nuclear extracts from control and infected cells, which revealed that influenza infection indeed causes YTHDF1 and YTHDF2 proteins to evidently translocate into the nucleus (Figure 4.16C). To explore the molecular mechanism behind this unique relocalization event, we further studied the interactome of YTHDF proteins by IP and protein mass spectrometry. From the results, we discovered that YTHDF2 specifically binds to viral protein NP after influenza infection, which primarily localizes inside the nucleus as a structural protein and encapsidates viral genomic RNAs. The study is ongoing to further elucidate this interaction and how it may affect the regulation of influenza replication by m^6A readers.

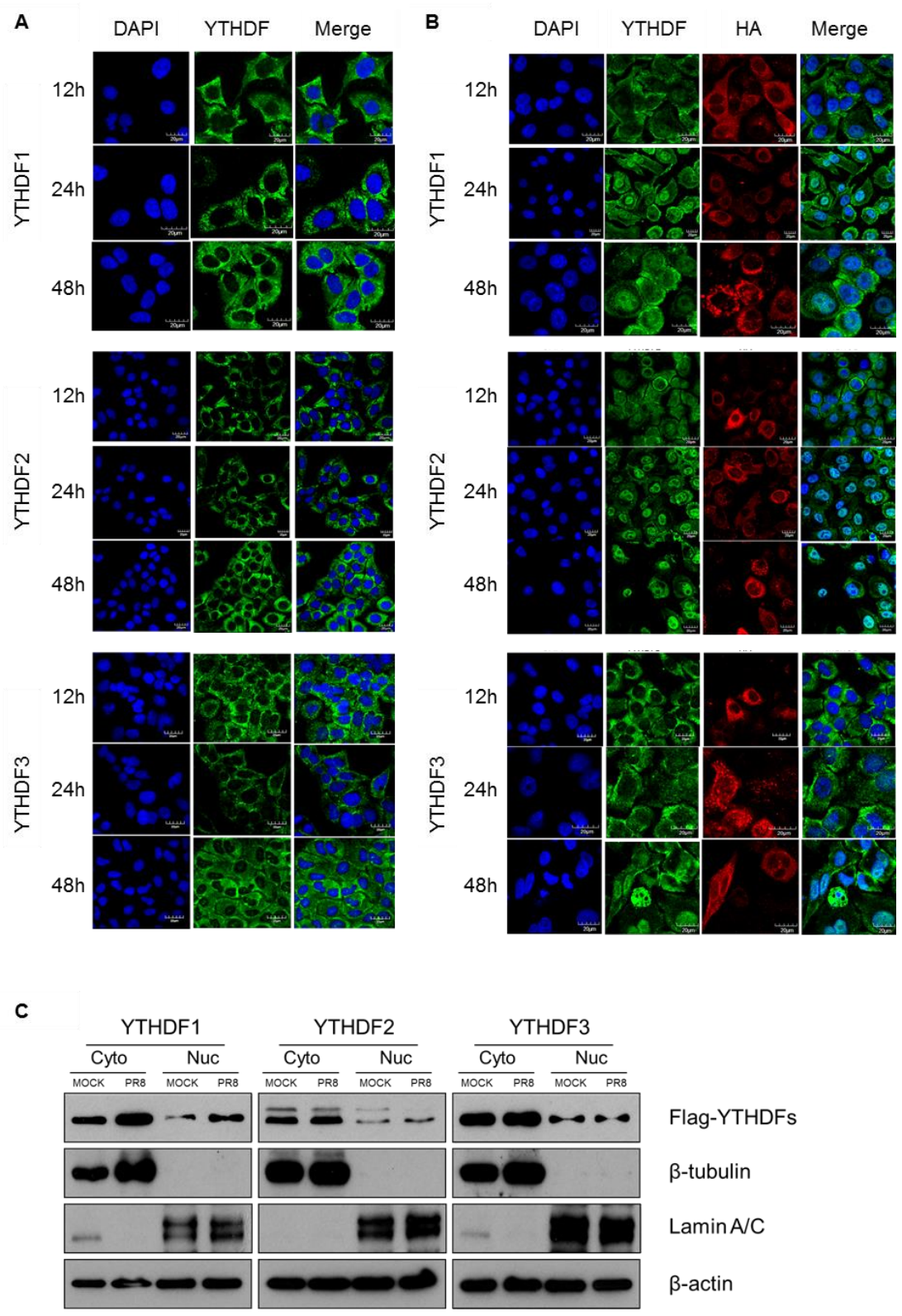


Figure 4.16 YTHDF proteins relocalize into the nucleus after influenza infection

(**Figure 4.16, continued**) (A and B) Immunofluorescence analysis of the subcellular localization of YTHDF proteins in mock infected (A) and influenza infected cells (B). (C) Western blotting analysis of the distribution of YTHDF proteins between nuclear and cytoplasmic regions in mock infected and influenza (PR8) infected cells.

In summary, we investigated the role of m^6A regulators in the life cycles of (-) ssRNA viruses (RSV and influenza virus) and found that the depletion or overexpression of RNA methylation enzymes and readers impacts viral protein expression and replication, demonstrating that the host RNA methyltransferase machinery acts as a positive post-transcriptional regulator of these two types of viruses. We also explored the host methylome changes in response to viral infection and identified two groups of genes with emerged or lost m^6A peaks and distinct functional clusters, confirming our previous finding that m^6A regulators may be able to detect and respond to viral infection by modulating gene expression. The striking finding of the drastic relocalization of m^6A readers after viral infection hints new mechanisms of how m^6A readers may react to external stimulation and potentially change functions as discussed previously. Finally, the opposite effects of m^6A on (-) ssRNA virus and (+) ssRNA virus suggest that m^6A regulators may have different modes of recognition based on the strandness of RNA targets, which invite further investigation.

4.3 Conclusion and Discussion: m^6A as a key regulator in disease development

The study of m^6A in various disease settings is just getting started with emerging advancements of the epitranscriptomics technologies. From our work in glioblastoma, we showed that inhibition of ALKBH5 represses GSC tumorigenesis and that FOXM1 is a key component in mediating ALKBH5-dependent GSC proliferation, highlighting the importance of correct m^6A pattern maintenance on key functional genes. These findings also open up avenues for developing

effective therapeutic strategies in the treatment of GBM. However, modulating methylation patterns through erasers is not the only way m^6A may be affecting cancer development. Our current study in acute myeloid leukemia (AML) suggests that m^6A writer METTL14 plays critical roles in normal myelopoiesis and AML pathogenesis, as featured by blocking myeloid differentiation and promoting self-renewal of normal hematopoietic stem/progenitor cells and leukemia stem cells/leukemia initiating cells. Targeting METTL14, especially in combination with standard differentiation-inducing agents, might be an effective therapeutic strategy to treat METTL14-high AMLs. Therefore, m^6A regulators may have distinct impacts on diverse types of cancers with different molecular mechanisms.

Despite the limited knowledge we have acquired, there are still much more unknown aspects in the broad impacts of m^6A in carcinogenesis and potential treatments. We do not know whether these identified m^6A pathways are ubiquitously conserved in distinct cancer types, and the altered cellular differentiation states of many cancer subtypes may further complicate the characterization of m^6A's functions. From our studies, it seems a systematic analysis of methylated RNAs in cancer cells may uncover novel gene targets for treatment, yet it is unclear whether modulation of m^6A levels will effectively and specifically suppress the self-renewal of cancer cells without any unwanted side-effects. Future research will hopefully answer these questions and better evaluate the potential of m^6A pathways as targets for cancer therapeutics.

Our studies of m^6A and the functions of m^6A regulators in various RNA virus species have generated many interesting results highlighting the complex interplay between host and virus. A recent work reported that dsDNA virus is also affected by m^6A pathway[158]. Depending on the invasion and replication mechanism of the virus, m^6A may have drastically different impacts on each step during the viral life cycle. The m^6A modification may be utilized by viruses as an

immune evasion strategy, escaping detection by innate immunity against infection or taking advantage of the m^6A pathway to enhance viral mRNA and protein production. Yet the host may evolve and utilize the m^6A readers to bind the m^6A-modified viral RNA and inhibit its activities and subsequent viral protein expression. Further details of the interplay between host and multiple viral species will be elucidated by future experiments, which may be crucial for the development of new anti-viral therapies.

Finally, by studying the versatile functions of m^6A regulators in these unique host-virus systems, we may uncover new molecular mechanisms of m^6A that are not present in normal human cells. For instance, the discovery of m^6A readers recognizing both positive- and negative-strand viral RNAs, but have the opposite effects on their downstream activities, suggests an unknown mode of identifying foreign RNAs by m^6A readers. Also, the changes in host transcriptome and methylome, as well as methylation preferences of m^6A writers triggered by viral infection, are also intriguing directions to further investigate, which may yield novel findings of how human cells rapidly respond to the invasion of viruses and new target pathways for antiviral treatments.

4.4 Materials and Methods

Cancer study:

Mice and Animal Housing

Male *Foxn1nu/nu* athymic nude mice at 6-8 week age were purchased from MD Anderson Cancer Center ERO Breeding Core. Mice were grouped by 5 animals in large plastic cages and were maintained under pathogen-free conditions according to the NIH Guide for the Care and Use of Laboratory Animals. All mouse experiments were approved by MD Anderson's Institutional Animal Care and Use Committee.

Intracranial Tumor Assay

Mice were randomly assigned to experimental groups for all the experiments. For the animal survival analysis, mice were intracranially injected with 10,000 GSCs and maintained until moribund or 80 days after injection. For the rescue studies, GSCs with ALKBH5 or FOXM1-AS shRNAs were co-transfected with a FOXM1, an ALKBH5 wild-type, or an ALKBH5 mutant expression construct. A total of 50,000 GSCs were intracranially injected into mice (n = 8 mice per group). At 30 days after GSC11 injection and 20 days after GSC17 injection, the mice were humanely killed, and their brains were harvested. Each mouse brain was fixed in 4% formaldehyde, embedded in paraffin, and examined for tumor formation by histologic analysis of hematoxylin and eosin–stained sections. Tumor volume was calculated by the formula $V = ab^2/2$, where a and b are the tumor's length and width, respectively.

Cell Lines and Primary Cell Cultures

Human glioma Hs683 and SW1783 cell lines and GBM LN229 and U87MG cell lines were from the American Type Culture Collection. U251MG cell line was from Sigma. The immortalized NHA-E6/E7/hTERT cell line has been described previously[159]. These cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum. GSCs were obtained from fresh surgical specimens of human primary and recurrent GBMs and cultured as tumorspheres in DMEM/F12 medium supplemented with B27 supplement (Life Technologies), bFGF and EGF (20 ng/ml each). Only early-passage GSCs were used for the study. All patient-related studies were reviewed and approved by an Institutional Review Board at The University of Texas MD Anderson Cancer Center with informed consent.

Method Details

Immunofluorescence and Immunohistochemistry

For immunohistochemical (IHC) and immunofluorescence (IF) analysis, GBM xenografts or surgical specimens tissue slides were deparaffinized, rehydrated through an alcohol series followed by antigen retrieval with sodium citrate buffer. Tumor sections were blocked with 5% normal goat serum (Vector) with 0.1% Triton X-100 and 3% H_2O_2 in PBS for 60 min at room temperature and then incubated with appropriate primary antibodies 4 °C overnight. IHC staining was performed with horseradish peroxidase (HRP) conjugates using DAB detection. IF staining was performed with appropriate Alexa Fluor 488 or Alexa Fluor 594 secondary antibodies (Invitrogen, dilution 1:1000).

For IF analysis of cultured cells, GSCs were fixed with 4% formaldehyde (Fisher) for 15 min and then blocked with 5% normal goat serum (Vector) with or without 0.1% Triton X-100 in PBS for 60 min at room temperature. Immunostaining was performed using the appropriate primary and secondary antibodies. Nuclei were counterstained with Hoechst. Images were taken with a ZEISS Axio Scope.A1 Upright Microscope.

Specific antibodies against FOXM1 (K19, Western Blotting; G5, IF, IHC), Nestin, HuR, GFAP, Tuj-1, SSEA-1, SOX2 (E4, IF), Nanog, α-Tubulin, β-actin from Santa Cruz; SOX2 (D6D9, immunoblotting), Oct4, Histone H3, CD133 from Cell Signaling; U1-70K, ALKBH5 (ABE547, immunoblotting) from Millipore; FLAG, ALKBH5 (HPA007196) from Sigma were used for the IF, IHC, or immunoblot analyses.

RNA Pull-Down and RNA Immunoprecipitation

A mix of six 90- to 120-nucleotide probes antisense to the FOXM1-AS sequence was designed to capture FOXM1-AS. Antisense probes to the capture probe set omitting the FOXM1–FOXM1-AS overlapping region were also generated and served as the negative control. RNA probes (50 pmol) were transcribed in vitro with the MEGAscript® T7 Transcription Kit and then

labeled with the Pierce RNA 3' End Biotinylation Kit (Life Technologies), treated with TURBO DNase (Life Technologies), and purified with the RNeasy Mini Kit (QIAGEN). GSC17 nuclear pellet was resuspended and homogenized in 1 ml RIP buffer (150 mM KCl, 25 mM Tris pH 7.4, 0.5 mM DTT, 0.5% NP40, 1 mM PMSF and protease Inhibitor). Heat-denatured RNA probes were then incubated with the nuclear extract at RT for one hour and further incubated with 30 μl of Dynabeads MyOne Streptavidin C1 (Life Technologies) at RT for one hour. Associated proteins were detected by Western blotting.

RNA immunoprecipitation was performed with the Magna RIP RNA-Binding Protein Immunoprecipitation Kit (Millipore) according to the manufacturer's instructions. Briefly, magnetic beads coated with 5 μg of normal antibodies against mouse immunoglobulin G (Millipore), HuR (Santa Cruz), or FLAG (Sigma Aldrich) were incubated with pre-frozen cell lysates or nuclear extracts overnight at 4°C. Associated RNA-protein complexes were collected and washed 6 times and then subjected to proteinase K digestion and RNA extraction by TRIzol. The relative interaction between protein and RNA was determined by qPCR and normalized to input. For FLAG RIP, GSCs were transfected with the same amount of either FLAG-tagged or untagged plasmid expressing the same gene (ALKBH5 or HuR). Then both cells were subjected to the same treatment. RNA enriched by FLAG RIP of cells expressing untagged protein was served as negative control. For HuR RIP, treated cells were divided and subjected to RIP by anti-HuR and mouse normal IgG, which was served as negative control.

Subcellular Fractionation Assay

The method for subcellular fractionation assay was adapted from Wuarin and Schibler[160]. Briefly, GSCs were collected by centrifugation and washed with phosphate-buffered saline. Cell pellets were resuspended in cold NP-40 lysis buffer (10 mM Tris-HCl pH 7.5, 0.15% NP40,

150 mM NaCl) for 5 min. The lysate was then transferred onto 2.5 volumes of a chilled sucrose cushion (ice-cold sucrose buffer [10 mM Tris-HCl pH 7.5, 150 mM NaCl, 24% sucrose]), and centrifuged at 13,000 rpm for 10 min at 4°C. The supernatant (the cytoplasmic fraction) was collected for immunoblotting analysis or RNA extraction by TRIzol. The pellet (nuclear fraction) was resuspended in cytoplasmic lysis buffer without NP40 and passed through the sucrose buffer again. The washed nuclear fraction was resuspended in an ice-cold glycerol buffer (20 mM Tris pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol) and then mixed with an equal volume of cold nuclei lysis buffer (10 mM HEPES pH 7.6, 1 mM DTT, 7.5 mM $MgCl_2$, 0.2 mM EDTA, 0.3 M NaCl, 1 M UREA, 1% NP-40) by vortexing twice for 5 seconds each. After incubation for 2 min on ice, the supernatant (nucleoplasm fraction) was collected for immunoblot analysis or TRIzol RNA extraction by centrifuge at 13,000 rpm for 2 min at 4°C. The insoluble pellet was washed with cold phosphate-buffered saline and then dissolved in TRIzol or saved for immunoblotting analysis.

BrdU Incorporation and Cell-Cycle Analysis

For the BrdU (5-bromo-2'-deoxyuridine) incorporation assay, cells were cultured with a BrdU-labeling reagent (Life Technologies) and stained with an anti-BrdU antibody (Cell Signaling) according to the manufacturer's instructions. Five fields of view per slide were examined for BrdU-positive cells. For the cell-cycle analysis, cells were fixed and then stained with PI/RNase Staining Buffer (BD Pharmingen) for 15 min at room temperature. Samples were acquired with a FACScan flow cytometer (BD Biosciences).

Luciferase Reporter Assay

Cells seeded in 12-well plates were transfected with the pMIR-REPORT luciferase vector (Thermo Fisher) fused with or without the wild-type or mutated FOXM1-3′UTR. Mutation of

FOXM1-3′UTR is generated by Geneart (Life Technologies). Transfection efficiency was quantified by co-transfection with an actin promoter-driven Renilla luciferase reporter[161]. The activities of firefly luciferase and Renilla luciferase in each well were calculated by a dual-luciferase reporter assay system (Promega). The ratios between the FOXM1 3′UTR reporter and Renilla control were determined 48 hr after siRNA treatment. The relative luciferase activity of the FOXM1-3′UTR luciferase activities was further normalized to that in cells transfected with the firefly luciferase vector control under the same treating conditions. The *FOXM1* promoter activity was also normalized by co-transfection with the β-actin-Renilla luciferase reporter.

AMT Crosslink Assay

GSCs were suspended in PBS with or without 0.5 mg/mL AMT (4′-Aminomethyltrioxsalen hydrochloride, Sigma) at a concentration of $2x10^7$ cells/ml in 6-well tissue culture plates and incubated on ice for 15 min. Then the plates were covered with a 2-mm-thick glass plate without a lid. The cells were irradiated for 15 min from a distance of 2.5 cm with a handheld 365-nm UV light and mixed every 5 min. After cross-linking, cells were pelleted in 1.5 ml microcentrifuge tubes for RNA isolation by TRIzol. RNA yield was determined with a NanoDrop spectrophotometer. To pull-down the target RNAs, 25 pmol in vitro transcribed and biotin labeled RNA probes were denatured to 90°C for 2 min, transferred immediately on ice. Then probes and 10 μg RNA were mixed in binding buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5% NP40 and 1 mM ribonucleoside vanadyl complexes) and transferred to a 37°C Thermomixer, shaking at 1,200 r.p.m. After 2 hours, 30 μl washed Dynabeads MyOne Streptavidin C1 was added and incubated for 2 hours at room temperature. After six washes, precipitated RNAs were extracted by TRIzol.

RNA Isolation and Quantitative Real Time PCR

RNA was isolated using TRIzol (Life Technologies) following the manufacturer's protocol. mRNA was purified by Dynabeads purification kit (Ambion) or GenElute mRNA Miniprep Kit (Sigma). cDNA was generated using the iScript cDNA Synthesis Kit (BioRad). Quantitative real-time PCR using Powerup SYBR Green PCR Master Mix (Life Technologies) was performed on a 7500 Fast Real-time PCR System (Applied Biosystems). For RNA stability assay, GSCs were plated in a poly-lysine coated 6-cm dish and incubated with actinomycin D (Santa Cruz) at 5 μg/ml for indicated time. Total RNA was isolated for qPCR analysis.

MeRIP-qPCR

MeRIP assay was adapted from a reported protocol[162]. Briefly, intact poly-A-purified RNA was denatured to 70°C for 10 minutes, transferred immediately on ice and then incubated with m^6A antibody in 1ml buffer containing RNasin Plus RNase inhibitor 400 U (Promega), 50 mM Tris-HCl, 750 mM NaCl and 0.5% (vol/vol) Igepal CA-630 (Sigma Aldrich) for 2 hr at 4°C. Dynabeads Protein G (Invitrogen) were washed, added to the mixture and incubated for 2 hr at 4°C with rotation. m^6A RNA was eluted twice with 6.7 mM N^6-methyladenosine 5'-monophosphate sodium salt at 4°C for 1 hour and precipitated with 5 μg glycogen, one-tenth volumes of 3 M sodium acetate in 2.5 volumes of 100% ethanol at -80°C overnight. m^6A enrichment was determined by qPCR analysis. Fragmented mRNA was directly incubated with m^6A antibody containing buffer and treated similarly.

Plasmids and RNA Knockdown

ALKBH5 expression plasmid was generated by cloning the full-length ORF of human ALKBH5 gene (NM_017758.3) into pcDNA3.1-DYK vector (GenScript). ALKBH5 H204A was generated by GenScript. pcDNA3.1-DYK-HuR was generated by GenScript. ALKBH5 and HuR were cloned to pcDNA3.1 vector without tag. Wild-type and H204 mutant of ALKBH5 were

cloned to pLVX (Clontech) for stable expression. Coding region and 3′UTR of FOXM1 were cloned from GSC17 cells into 3xFLAG expression vector (Sigma). Transfections were performed using X-tremeGENE HP DNA Transfection Reagent for plasmid and X-tremeGENE siRNA Transfection Reagent (Roche) for siRNA following the manufacturer's protocols.

Lentiviral Transduction for Stable Cell Lines

Lentiviral vectors expressing non-targeting pLKO.1 control shRNA (SCH002), and two shRNA constructs targeting ALKBH5 (NM_017758), shRNA1 (TRCN0000064783) and shRNA2 (TRCN0000064787) were obtained from Sigma. shRNA for FOXM1-AS was generated according to the pLKO.1 protocol from Addgene. The lentiviral vectors were co-transfected with packaging vectors psPAX2 and pMD2G (Addgene) into 293FT cells for lentivirus production. To establish stable cell lines, GSC cells were transduced by using the above lentiviruses with polybrene (6 μg/ml, Sigma). After 72 hr of transduction, cells were selected with 2 μg/ml puromycin for 4 days. For the ALKBH5 rescue experiment, shRNA targeting 3′UTR of ALKBH5 (shRNA1) was used for knockdown.

Microarray Analysis

Total RNA from GSC11 cells with shControl or shALKBH5-2, GSC17 cells with shControl or shALKBH5-1 was collected on day 7 after transduction for microarray analysis at MD Anderson DNA Core Facility using Affymetrix Human Gene 2.0 ST array, Expression Console Software and Transcriptome Analysis Console v3.0 (Affymetrix). The genes showing altered expression (fold change > 2) compared with the control shRNA in both cell lines were selected and analyzed using Ingenuity Pathway Analysis (IPA, Ingenuity Systems).

Quantification and Statistical Analysis

Data are presented as the mean ± standard error of the means (SEM), or standard deviations (SD). Differences in the mean values between 2 groups were assessed for significance with a 2-tailed Student t-test using GraphPad Prism 6.0. Kaplan-Meier survival data were analyzed using the log-rank test. The Pearson correlation test was used to assess relationships between variables in tumor tissues.

Data and Software Availability

The microarray gene expression and m^6A-seq data are deposited at the GEO under the accession number: GSE93054, GSE87515.

Virus study:

Cell culture

HIV: Human healthy primary $CD4^+$ T-cells were isolated from healthy blood donors' buffy coats (purchased from American Red Cross Blood Service, Columbus, OH) using anti-human CD4-coated magnetic particles according to the manufacturer's instructions (BD Biosciences, San Jose, CA) as described[155]. Isolated $CD4^+$ T cells were maintained in complete RPMI media containing interleukin-2 (20 U/ml, PeproTech, Rocky Hill, NJ) and activated with phytohemagglutinin A (PHA, Sigma-Aldrich, St. Louis, MO) as described[155]. HEK293T cells, Jurkat cells, and the HIV-1 indicator cell line GHOST/X4/R5 were cultured as described[155]. HeLa cells overexpressing the empty vector (pPB-CAG), YTHDF1, YTHDF2, or YTHDF3 were maintained in complete DMEM containing 2 μg/ml of puromycin. All parental cell lines were obtained from the American Type Culture Collection (ATCC, Manassas, VA) and the identity of the cell lines has been authenticated using short tandem repeat profiling or genotyping methods as

described[163]. All the cell lines were tested negative for mycoplasma contamination using a PCR-based universal mycoplasma detection kit (ATCC).

ZIKV: All studies were conducted in accordance with approved IRB protocols by the University of California, San Diego. All animal work was approved by the Institutional Review Board at the University of California, San Diego and was performed in accordance with Institutional Animal Care and Use Committee guidelines. ZIKV MR766 virus was expanded by inoculation of Vero cells at an MOI of 5 as previously described[164].

RSV and influenza: For overexpression experiments, Hela stable cell lines stably overexpressing YTHDFs proteins were infected with 18MOI of influenza strain A/Purto Rico/8/1934(H1N1), abbreviated as PR8. All the cells were collected at 48 h.p.i. for subsequent analysis. For knockdown experiments, 293T cells were transfected with each siRNA. At 48 hours post-siRNA transfection, the cells were infected by influenza virus strain PR8 at an MOI of 5.0. At 8 hours post-virus infection, cell lysates were harvested and subjected to Western blot to detect influenza virus NP, NS1, and M1 protein. Also, the host cell β-actin protein was used as internal control in Western blot.

Cell proliferation assay

Cell proliferation of HeLa cells or Jurkat cells with YTHDF1–3 overexpression or knockdown respectively were determined by the MTS assay (Promega, Madison, WI) as described[155]. Cells (2×10^4) were plated in triplicate in a 96-well plate and cultured for 3 days and the absorbance was read at 490 nm at the indicated times.

Plasmids and shRNA- and siRNA-mediated gene silencing

HIV-1 proviral DNA vector pNL4-3, pNL-Luc-E^-R^+ containing a firefly luciferase reporter gene and the empty vector control were described[165]. The pPB-CAG plasmid vector was

used to overexpress the YTHDF1–3 proteins in HeLa cells. pLenti vectors carrying specific YTHDF1–3 shRNAs were used to knockdown of YTHDF1–3 proteins in different cell types as described[155]. Jurkat cells transduced with lentivirus containing shRNAs specific for YTHDF1, YTHDF2, and YTHDF3 were maintained in puromycin (3 μg/ml) containing complete RPMI media. AlkBH5, FTO, METTL3 and METTL14 gene expression in HEK293T cells was silenced by two rounds of siRNA transfection using specific siRNA (Qiagen, Valencia, CA) transfected with the Lipofectamine RNAiMax reagent (Invitrogen, Waltham, MA) according to the manufacturer protocol (reversible siRNA transfection method). Briefly, HEK293T cells (1.5×10^5) were transfected with specific siRNA or a non-specific control (80 nM). At 24 hr post-transfection, media were replaced and the second round of siRNA transfection was performed using the same siRNA concentration (80 nM). The pNL4-3 construct (0.5 μg) was transfected into the cells (1.5×10^5) 6 hr after of the second round transfection and cells were harvested for immunoblotting 36 hr after the proviral DNA transfection.

HIV-1 infection assays

Single-cycle, luciferase reporter HIV-1 stock (HIV-Luc/VSV-G) was generated by calcium phosphate co-transfection of HEK 293T cells with the pNL-Luc-E^-R^+ and pVSV-G as described[155]. The infectious units of virus stocks were evaluated by limiting dilution on GHOST/X4/R5 cells as described[154]. HIV-1 infection assays using luciferase reporter viruses were performed using a multiplicity of infection (MOI) of 0.5 as described[165]. Cell lysates were obtained 24 hpi and analyzed for luciferase activity using a commercially available kit (Promega) according to the manufacturer's instructions. Total cell protein was quantified using a bicinchoninic acid assay (BCA; Pierce, Waltham, MA) and all luciferase results were normalized to total protein amounts. HIV-1 capsid p24 levels in supernatants were measured by an enzyme-linked

immunosorbent assay (ELISA) using anti-p24-coated plates (AIDS and Cancer Virus Program, NCI-Frederick, MD) as described[154]. Jurkat cells were infected with replication-competent HIV-1_{NL4-3} at an MOI of 0.5 as described[154]. At 72 hpi, cells were washed 3 times and harvested for total RNA extraction using RNAeasy kit (Qiagen). PHA-activated primary $CD4^+$ T cells were infected with HIV-1_{NL4-3} (40 ng p24 equivalent HIV-1 per 10^6 cells) and cells were harvested at 96 hpi for total RNA extraction using RNAeasy kit (Qiagen).

m^{6}A-seq

High-throughput sequencing of viral methylome was carried out using m^{6}A-seq and followed the protocol published previously[58]. In brief, total RNA containing viral RNA was extracted from the cells and purified by poly (dT) beads. Purified polyadenylated RNA was mixed with 2.5 μg of affinity purified anti-m^{6}A polyclonal antibody (202003; Synaptic Systems, Goettingen, Germany) in IPP buffer (150 mM NaCl, 0.1% NP-40, 10 mM Tris-HCl, pH 7.4) and incubated for 2 hours at 4°C. RNA was used for library generation with the small RNA sequencing kit (NEB, Ipswich, MA). Sequencing was carried out on Illumina HiSeq 2000 according to the manufacturer's instructions.

Quantification of viral RNA modification level using liquid chromatography-mass spectrometry (LC-MS/MS)

Virus gRNA (250 μg) was isolated from highly purified virions using RNeasy Mini kit (Invitrogen) and subjected to quantitative analysis of m^{6}A level using LC-MS/MS as described in the previous section.

Quantitative PCR and RT-PCR

Quantitative PCR (qPCR) was performed to assess the relative levels of HIV-1 late reverse transcription (RT) products and 2-LTR circles as described[156,165]. Reverse transcription PCR (RT-

PCR) was used to measure HIV-1 *gag* mRNA as described[156]. To amplify HIV-1 late RT products in cells transduced with pLenti vectors, a different set of primers (LW59 and LW60) were used as described[155]. Sequences of PCR primers and probes are listed below. All HIV-1 stocks used for PCR assays were treated with DNase I (40 U/ml; Ambion, Waltham, MA) prior to infections to avoid plasmid DNA contamination.

List of primers and probes: HIV-1 *gag* forward: CTAGAACGATTCGCAGTTAATCCT, reverse: CTATCCTTTGATGCACACAATAGAG; Unspliced GAPDH forward: GGGAAGCTCAAGGGAGATAAAATTC, reverse: GTAGTTGAGGTCAATGAAGGGGTC; Spliced GAPDH forward: GGAAGGTGAAGGTCGGAGTCAACGG, reverse: CTGTTGTCATACTTCTCATGGTTCAC; MH531 forward (for HIV-1 late reverse transcription (RT) products): TGTGTGCCCGTCTGTTGTGT, reverse: GGATTAACTGCGAATCGTTC; HIV-1 late RT product probe: TCGACGCAGGACTCGGCTTGCT; 2-LTR forward: GCCTGGGAGCTCTCTGGCTAA, reverse: GCCTTGTGTGTGGTAGATCCA; 2-LTR probe: AAGTAGTGTGTGCCCGTCTGTTGTGTGACTC; LW59 (forward, alternative for late RT detection in shRNA vector-transduced cells): GACATAGCAGGAACTACTAGTACCC; LW60 (reverse, alternative for late RT detection in shRNA vector-transduced cells): GGTCCTTGTCTTATGTCCAGAATGC.

Antibodies and immunoblotting

The antibodies used in this study were: anti-GAPDH (clone 4G5, AbD Serotec, Atlanta, GA), anti-FLAG (F-3165, Sigma-Aldrich), anti-FTO (ab124892, Abcam, Cambridge, MA), anti-AlkBH5 (HPA007196, Sigma-Aldrich), anti-METTL3 (15073-1-AP, Proteintech Group, Rosemont, IL), anti-METTL14 (HPA038002, Sigma-Aldrich), anti-YTHDF1 (ab99080; Abcam), anti-YTHDF2 (ABE542, EMD Millipore, Billerica, MA), anti-YTHDF3 (ab103328;

Abcam), and anti-HIV-1 Gag (clone #24–2, the NIH AIDS Reagent Program). Cells were harvested and lysed in cell lysis buffer (Cell Signalling, Beverly, MA) supplemented with protease inhibitor cocktails (Sigma-Aldrich). Immunoblotting was performed as described[155]. Detection of GAPDH (glyceraldehyde-3-phosphate dehydrogenase) expression was used as a loading control.

Immunoprecipitation and RNA isolation

HeLa cells expressing pPB-CAG vector or YTHDF1–3 (2.5×10^6 cells) were seeded in a 60 mm-diameter culture plate. Cells were infected with HIV-Luc/VSV-G at an MOI of 5 for 3 hr. Cells were UV-cross-linked, lysed in cell lysis buffer (Sigma-Aldrich). The cells were incubated in lysis buffer for 10 min on ice with frequent mixing and were sonicated to ensure maximum lysis. The lysed cell suspension was centrifuged for 5 min at 9,300 ×g at 4°C. The supernatant was transferred to fresh tubes and an equal amount of cell lysates were mixed with anti-FLAG-conjugated protein G beads and rotated for 2 hr at 4°C. After the incubation, beads were washed 3 times with cell lysis buffer. Co-immunoprecipitated RNA was isolated from the immunoprecipitates using Trizol (Invitrogen), and RNeasy columns (Qiagen) with an on-column DNase I treatment (Qiagen) and eluted with RNase-free water. Equal volumes of RNA were used as a template for first-strand cDNA synthesis, according to the manufacturer's instructions.

CLIP-seq

We followed the previously reported protocol of the PAR (photoactivatable ribonucleoside-enhanced)-CLIP assay[76] with the following modifications. As HIV-1 infection was inhibited by the addition of 4-thiouridine (data not shown), we omitted that step and performed crosslinking directly. Briefly, six 15-cm plates of HeLa cells stably expressing YTHDF1–3 proteins were seeded at day 1. At day 2, the HeLa cells were infected with the HIV-1-Luc/VSV-G at an MOI of 0.5 for 2 hr, cells were washed to remove cell-free HIV-1. At day 3, the cells were

washed once with 10 mL ice-cold PBS. Cell plates were placed on a tray with ice and irradiated, uncovered, with 0.15 J/cm^2 of 254 nm UV light three times in a Stratalinker 2400 (Stratagene, Santa Clara, CA). Cells were scraped off in PBS and transferred to centrifugation tubes and collected by centrifugation at 500 ×g for 5 min at 4°C. The cell pellets were lysed in 3 volumes of 1% NP40 lysis buffer and incubated on ice for 10 min. The cell lysates were cleared by centrifugation at 13,000 ×g for 15 min at 4°C. Anti-FLAG antibodies were used in the IP as previously described. The recovered RNA samples were further cleaned by RNA Clean & Concentrator (Zymo Research, Irvine, CA) before library construction using the small RNA sample preparation kit (NEB). The first round of RNase T1 digest was carried out at 0.2 U/μl for 15 min and the second round digestion was conducted at 10 U/μl for 8 min.

Measurement of the kinetics of HIV-1 RNA in infected cells

HeLa cells over-expressing individual YTHDF1–3 proteins or the pPB-CAG vector were infected with HIV-1-Luc/VSV-G (MOI of 0.5). Cells were collected at 6, 12 and 24 hpi. Total RNA was isolated from the cells using RNeasy columns (Qiagen) with on-column DNase I treatment (Qiagen) and eluted with RNase-free water. Quantitative RT-PCR was used to measure HIV-1 *gag* RNA levels as described[156]. All HIV-1 stocks used for infection were treated with DNase I (40 U/ml; Ambion) prior to infections to avoid plasmid DNA contamination.

Statistical analysis

Data were analyzed using Mann-Whitney's U test or one-way ANOVA test with Prism software and statistical significance was defined as $p<0.05$.

Immunofluorescence Staining of Flavivirus Envelope Protein

293T cells previously transduced with NTC, METTL3, METTL14, FTO, or ALKBH5 shRNAs were seeded in 24-well plates pretreated with poly-l-lysine (Sigma) 1 day before infection

and then infected as described above. At 24 hours post-infection, cells were washed three times with PBS and fixed in 4% paraformaldehyde in PBS for 20 min at room temperature. Cells were permeabilized in PBS containing 0.2% Triton X-100 for 10 min at room temperature and then blocked with 3% BSA in PBS for 2 hr. Cells were stained with primary antibody (1:1,000) overnight at 4°C, washed three times with PBS, and then stained with Alexa 488-conjugated secondary antibody (1:1,000) for 1 hr at room temperature. The nuclei were stained with DAPI, and cells were analyzed using a Leica microscope.

Data deposition and access

Data accession: all the raw data and processed files from published studies have been deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) and accessible under GSE85724 (HIV) and GSE87516 (ZIKV).

CHAPTER 5

Summary and Perspectives

5.1 The emergence of m^6A: an advanced precision control system of gene regulation

Although the existence of RNA modifications including m^6A has been known for decades, the study of their regulation and function has long been hampered by the issues of low abundance and the lack of detection methods. Applying a combination of high-sensitivity mass spectrometry and newly developed high-throughput sequencing methods, I have overcome these technical difficulties and detailed the molecular functions of m^6A in mammalian cells. To extend the scope of the investigation to more complex processes, I have used various model systems for m^6A characterization, including zebrafish embryonic development, glioblastoma tumorigenesis, and the infection of several types of RNA viruses. Subsequent cell biology and virology experiments revealed the significance of m^6A at the organismal level and in host-pathogen interactions.

From the study of molecular functions of m^6A, I have discovered that m^6A functions as a marker for fast-processing of mRNA, i.e. mRNA with m^6A will get matured faster, translate into proteins more efficiently, and decay faster, each effect dependent on being recognized by m^6A reader proteins with corresponding facilitating properties. m^6A thus may induce spikes of gene expression in the events when a rapid reaction is needed, such as embryogenesis and immune response. In the study of zebrafish early development, m^6A has been found to enable the timely removal of maternal mRNA and activate zygotic gene expression; embryos devoid of a key m^6A reader protein will experience a substantial delay during development. In the studies of several types of viral infection, m^6A has been revealed to either positively or negatively impact the

infection efficiency of different viruses depending on their life cycles, as m^6A reader proteins identify viral RNAs and either sequester them from downstream functions; or assist in viral protein expression and replication. Summarizing what we discovered in these work, we propose two hallmarks to mRNA m^6A methylation: that it serves as a marker to group and synchronize cohorts of transcripts for fast-tracking mRNA processing and metabolism; and that it considerably affects cell-state transition during cell differentiation. How selectivity and transcript grouping are achieved and how writers, erasers, and readers are coordinated in response to different signaling pathways are unknown. We propose that the same stimuli and regulatory processes that tune transcription and translation may also affect these writers, erasers, and readers through different forms of post-translational modifications. For example, when certain transcription factors are activated they may directly affect the accessibility and recruitment of writers. The same signaling pathway may coordinately activate or inactivate erasers and readers through direct recruitment or post-translational modifications (Figure 5.1). This process could be a fundamental mechanism that mammalian cells exploit to coordinate gene expression during development. Defects in such processes may cause or contribute to human diseases.

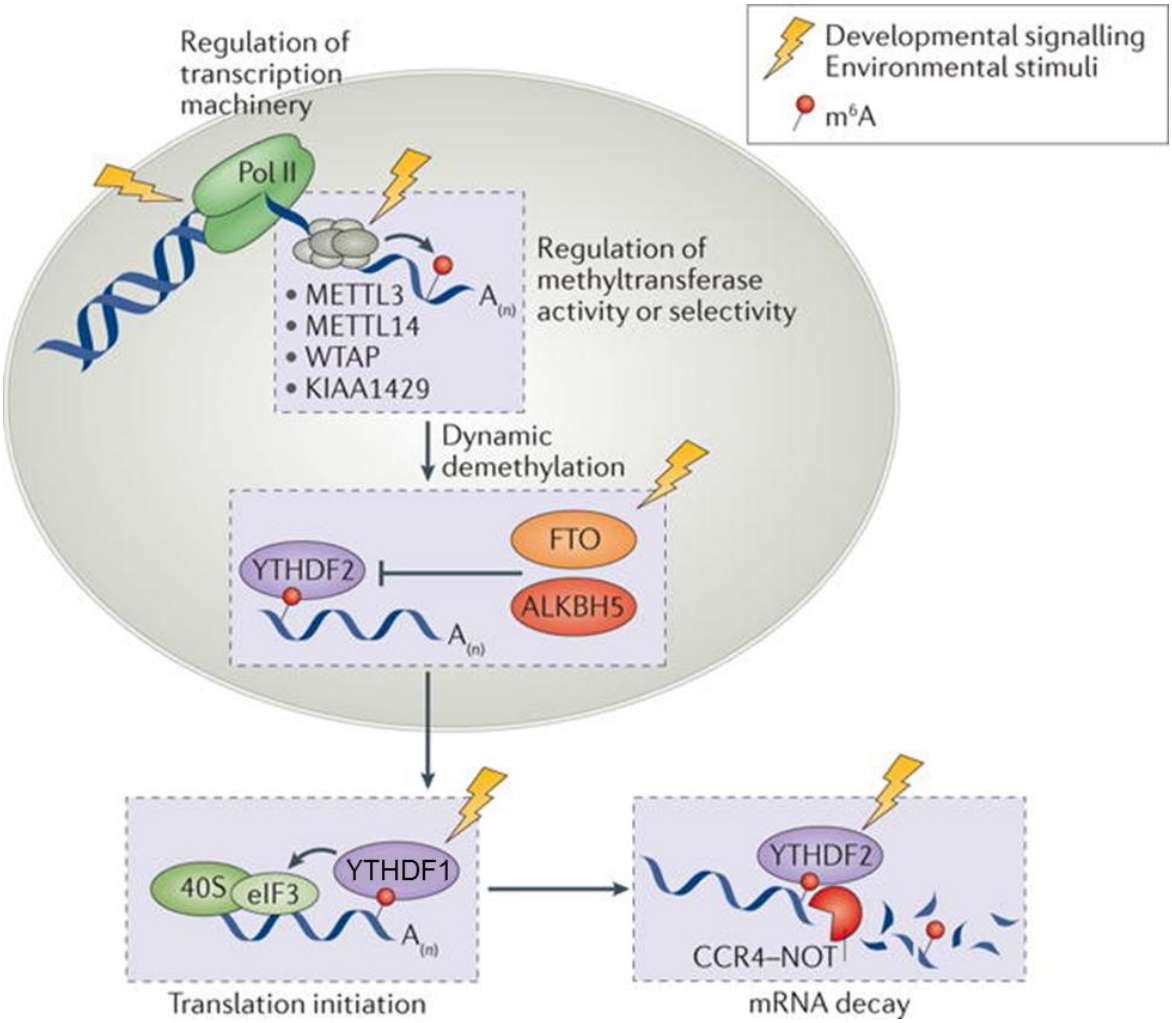


Figure 5.1 m^{6}A synchronizes mRNA processing in response to various internal and external stimuli

The activities of m^{6}A writers, erasers and readers may be regulated by the same signaling pathways and stimuli that tune transcription and translation, potentially through various post-translational modifications on writers, erasers, and readers. This process could constitute an additional mechanism to post-transcriptionally coordinate the expression of large groups of genes in response to internal and external stimuli, which may affect many physiological processes that require rapid responses involving multiple genes.

Collectively, my work on m^{6}A has clarified the uncertainty of whether mRNA modifications convey functional information and demonstrated m^{6}A as a novel epigenetic mechanism for precision control of gene expression. The relative expression of m^{6}A enzymes, the localization and interactome of m^{6}A readers, the potential activity-changing post translational modifications on m^{6}A regulators, the crosstalk between m^{6}A pathway and other epigenetic

mechanisms, all of which can be used to fine tune the exact level of impact of m^6A on gene expression. Further characterization of m^6A metabolism may also provide valuable implications for the studies of complex biological processes as well as the development of new strategies for treating human diseases such as cancer and viral infection.

5.2 The network of epigenetics: crosstalk between m^6A and other epigenetic mechanisms

From our study, m^6A pathway emerges as a key member of the family of regulatory epigenetic mechanisms, which include covalent chemical modifications of DNA (DNA methylation) and of histone proteins associated with DNA (histone modifications), and functional RNA-mediated epigenetic regulation pathways (including miRNA, small interference RNA (siRNA), Piwi-interacting RNA (piRNA), long non-coding RNA (lncRNA), etc.). As we are presenting covalent chemical modification of mRNA, m^6A methylation, as a new epigenetic mechanism of gene regulation, it is important to address its relationship with other established mechanisms, considering that it is highly possible that these epigenetic mechanisms may work in accordance for a coordinated control of global gene activity and expression.

As introduce previously, DNA methylation adds extra epigenetic information to DNA that is not encoded in the sequence. Among all the cytosine derivatives, 5mC is commonly thought to be a repressive mark as methylation of CpG islands in promoter regions can recruit methyl-CpG-binding proteins and contribute to gene repression; 5hmC, on the other hand, could have active functions in the regulation of chromatin structure and gene expression[166]. From our collaborative study with Allis lab and Thompson lab, we have found that α-ketoglutarate (α-KG), a crucial biological compound that aids in many oxidation reactions *in vivo*, may induce global change in the oxidative demethylation process and chromatin structural change in mouse embryonic stem

cells. In parallel, α-KG may also induce transcriptome-wide m^6A pattern shift and abundance change. These preliminary results suggest m^6A may work in accordance with 5mC demethylation pathways to shape chromatin structure and impact gene expression levels with the induction of key metabolites such as α-KG.

Of all the diverse types of histone modifications, histone methylation is one of the most common and well-studied histone modifications, which occurs on different lysine residues and is associated with either transcriptional activation, inactivation, or genomic silencing[167]. H3K36me3 emerges in our collaborative study with Chen lab as a potentially overlapping histone marker with m^6A. On the genome-wide level, m^6A peaks largely correlate with H3K36me3 peaks; and the H3K36me3 methyltransferase, SETD2, impacts the overall cellular m^6A level. These preliminary data suggest potential crosstalk between m^6A and histone methylation.

As for potential interaction between m^6A and other functional RNA pathways, it was reported previously that m^6A affects the processing of miRNA[168] and may impact the function of lncRNA during X chromosome inactivation[169]. X chromosome inactivation (XCI) is a well-studied RNA-mediated epigenetic process, by which one of the two copies of the X chromosome in female mammals is inactivated. This process is mediated by the lncRNA Xist, which is rapidly expressed upon the onset of XCI and progressively coats the X chromosome, causing the silencing of genes[170]. From our collaborative study with Heard lab, we have discovered more detailed evidence of m^6A's involvement in the XCI process than previously reported: m^6A-seq in specially constructed mESCs with inducible XCI revealed that m^6A is present on the transcripts of Xist and many other XCI-related transcripts with XCI stage-dependent patterns; and high-resolution protein mass spectrometry with Xist probes detected more Xist-binding proteins than previously known[171], including m^6A reader Ythdf2. Imaging experiments also revealed m^6A readers associate with XCI

factors in stage-specific fashion. We have also discovered that m^6A has allele-specific preference independent of transcription level, marking the active and inactive X chromosomes differently. Knockdown of m^6A writers or readers impact the onset of XCI profoundly. These evidences strongly indicate multiple regulatory functions of m^6A on both Xist RNA and gene expression control during XCI.

Based on these preliminary findings from our studies, we consider the existence of direct links between m^6A and aforementioned epigenetic pathways highly possible, and that m^6A may also interact with other epigenetic mechanisms given its wide-spread presence on functional RNAs. Our ongoing work and future studies will shed more lights on this complex regulatory network of epigenetic pathways with m^6A as a central player inside.

5.3 The family of RNA modifications: a spectrum of functional decorations

It should be noted that m^6A is but one of many post-transcriptional modifications in mRNA with regulatory roles that have been discovered or re-discovered in recent years. For instance, m^1A has been recently revealed to exist in eukaryotic mRNAs with unique distribution pattern in proximity to translation starting sites and the first splice site[172,173]. In addition to its impacts on RNA structure by weakening base-pairing[174], it has been shown to potentially promote translation[172]. m^5C is known to be present in eukaryotic mRNA at low levels and its high-resolution maps in human mRNA and lncRNA have recently been reported[175]. The oxidative derivatives of m^5C, 5-hydroxymethylcytosine and 5-formylcytosine, have been detected in RNA from *Drosophila* spp. to mammalian cells and brain tissues[176-179], suggesting that it is a dynamic modification with potential regulatory roles. Pseudouridine ψ, as the most abundant modification in cellular RNA[180], has recently been mapped at single-base resolution in yeast and mammalian mRNAs[181-183]. Its ability to alter base-pairing interactions allows it to affect not only RNA

structures but also mRNA coding[184], underscoring its potential as a regulatory element. 2′-*O*-methylation (N_m) residing on the 2′ hydroxyl ribose moiety of all four ribonucleosides has been found in all major classes of eukaryotic RNA[185-187], and its existence in human mRNA was reported at the same time that m^6A was discovered[52]. A recent work has reported the base-resolution map of N_m in mammalian mRNA and discovered its unique enrichment in CDS, near splice sites, and in introns; it is also enriched in codons of three amino acids and unequally distributed between the three codon positions, suggesting potential functional roles in translation elongation dynamics and mRNA splicing[188].

It is now believed that mRNA internal modifications are distributed in unique patterns and affect multiple RNA metabolic processes in order to impact gene expression. While many modifications such as Ψ and m^5C are not yet known to be reversed in mRNA, other methylations on heteroatoms resembling m^6A could be more broadly spread and dynamically/reversibly regulated by specific enzyme systems. They could affect RNA metabolism and function via RNA structure alteration or recognition by specific reader proteins. As new modifications and new functions continue to emerge, these chemical marks on RNA may collectively provide additional tuning that affects biological outcomes at the post-transcriptional level. With the development of new approaches to quantitatively analyze RNA modifications in a transcriptome-wide manner, a quantitative picture of how chemical modifications affect gene expression regulation and their effects in various human diseases will emerge. The pervasiveness of RNAs and RNA modifications suggest their widespread influence in a myriad of other RNA-involved biological events, which will be the focus of future study and is crucial for the clarification of how cells encode epigenetic information into chemical modifications for global gene regulation.

5.4 The future of m^6A and epitranscriptomics study

5.4.1 Toward a finer portrait of the multifaceted m^6A modification

Despite everything we have learned about m^6A so far, there are several key questions about m^6A that we still have no clues about. The most striking characteristics of m^6A, apart from its pervasiveness, are its specificity and dynamics. Although we have some understandings about the enzymatic regulation of m^6A on RNA, many aspects of this process remain unclear. For instance, as the consensus motif of m^6A is rather simple, why is there only a small fraction of potential m^6A sites getting methylated while the rest in majority are not? Why does m^6A have a preferential localization pattern across transcriptome (enrich around stop codon and in UTRs)? Why does a single gene contain several m^6A peaks with different methylation levels (meaning mRNAs from the same gene are being methylated differently at each site)? Why does modulating m^6A writers and erasers have drastically different effects on diverse groups of genes? All the questions essentially boil down to one central mystery: how does one single site on a certain mRNA get targeted for m^6A methylation? The variability of m^6A level on potential methylation sites, within each mRNA transcripts, and inside different cells in a group may all be explained by answering this central question.

Several possible explanations exist for this seemingly stochastic process. As hypothesized previously, cell signaling-controlled post-translational modifications to m^6A writers and erasers may impact the patterns of m^6A deposited during the affected period. It is also possible that protein interactions may alter the binding kinetics of m^6A writers and erasers to different gene regions and transcripts, resulting in differentially regulated m^6A sites depending on their levels of enzyme tethering through protein interactome. Finally, considering the complex relationship between m^6A and other epigenetic regulators, it is equally probable that m^6A writers and erasers are subjected

to the regulation of other epigenetic factors and further diversify the patterns of methylation on relevant gene transcripts.

To test these potential aspects of m^6A regulation, new technologies will need to be developed for better characterizing the dynamics of m^6A. Direct single-molecule sequencing by nanopore devices may differentiate m^6A from other RNA bases and yield m^6A maps for each mRNA molecule, revealing the variability of m^6A patterns on transcripts from the same gene at single-base resolution. Metabolic labeling of RNA may be combined with low-input IP and sequencing method to separate nascent mRNA from background expression and profile actual m^6A changes at higher temporal resolution. Single-cell sequencing techniques applied to epitranscriptome profiling methods may reveal cell-to-cell variation in transcriptome-wide m^6A patterns. The advances in technologies will certainly bring about new horizons in the study of m^6A and provide a complete portrait of the multifaceted m^6A modification.

5.4.2 Uncover the involvement of m^6A in epigenetic inheritance

Another big blank area also exists in the study of m^6A: its involvement in epigenetic inheritance, which refers to epigenetic information being passed down from one generation to the next. Epigenetic inheritance is used to describe all non-DNA sequence-based effects across generations, from cellular levels (such as cell replications) to organismal levels, with many examples reported in microorganisms, plants, and sometimes animals. While DNA methylation, histone modifications, coding and non-coding RNAs, and structural templating have all been implicated in transgenerational epigenetic inheritance[189], there is no study of the potential involvement of m^6A in this process so far.

Cell replication is the most basic form of inheritance, when a parent cell duplicates its genetic materials and then divides to produce two daughter cells. During this process, DNA

methylation and many histone marks are inherited and replicated into the new daughter cells[190]. Yet the dynamics of m^6A across the cell cycles is not known. We have observed an oscillation of m^6A signals by microscopy during the progression of the cell cycle stages, indicating a temporal regulation of m^6A abundance during the cell cycle. In addition, we have also found that Ythdf2 malfunction during zebrafish embryogenesis results in the pausing of the cell cycle at the onset of differentiation. These initial results indicate m^6A is regulated during cell replication and may be required for the normal progression of the cell cycle, while future research is needed to determine the exact mechanism of such regulation.

Epigenetic inheritance on the organismal level is more complex, as sometimes it is difficult to distinguish intergenerational effects (the effects introduced by parental factors) from real transgenerational effects (the effects free from any exposure to the initial trigger signals). For instance, maternal deposition of m^6A-modified RNA into offspring embryos can be viewed as a form of intergenerational inheritance of m^6A. However, there is currently no evidence of RNA modification being inherited through transgenerational mechanisms. To study this question, one potential starting point is to investigate the regulatory functions of m^6A during fertilization events. m^6A regulators may determine the epigenetic characteristics of gametes by impacting gametogenesis, thus affect the fertilization event or the phenotype of the descendants. Another possible study is to induce a change in m^6A methylation using external stimulants (nutrient stress, drug treatment, or other environmental factors) and characterize m^6A in the immediate and distant decedents. The comparison of m^6A pattern changes in the original impacted generation and decedents will reveal if m^6A has intergenerational or transgenerational effects in epigenetic inheritance. By monitoring the changes in m^6A pattern across generations, we will be able to

evaluate the stability of epigenetic inheritance of m^6A and gain valuable insights of long-term impacts of m^6A in populations and its potential influence on evolution.

5.4.3 Seek other unknown members of the family of functional mRNA modifications

Considering the extreme complexity of m^6A regulatory mechanisms, it is daunting to think about that there might be many more functional RNA modifications resembling the roles of m^6A in other aspects of cellular activities or different biological systems. In addition to m^6A, m^1A, m^5C, ψ, and N_m, we believe that additional regulatory chemical modifications will be discovered in mRNA. Each of these modifications may have a dedicated set of writers, erasers, and readers, although some of them might be shared. The potential to decorate distinct parts of the pre-mRNA (5′ UTR, coding sequence, 3′ UTR, splice sites) could be used to modify different groups of transcripts in response to various stimuli. These mRNA chemical modifications could be regulated individually or combinatorially to affect the fate of individual mRNA species. More quantitative technologies will need to be developed to precisely map locations of these mRNA chemical modifications and explore the effects of various mRNA modifications in biological processes such as cell differentiation and development. Further understanding is also needed in how transcript selectivity and site selectivity are achieved, how these mRNA chemical modification processes and their connections with different signaling pathways are regulated, the interplay and potential synergy between modification regulators and other cellular components, and their roles in human physiology.

In summary, the diverse chemical modifications in RNA serve as another layer of information carrier, exponentially expand the complexity of the eukaryotic genome, and precisely regulating almost every aspect of cell physiology. The future of epitranscriptomics study will most certainly be a joint effort among scientists, engineers, and clinical practitioners. With the

development of new technologies (such as single-molecule nanopore sequencing and single-cell epitranscriptome profiling), these pathways will provide new opportunities for researchers to investigate the underlying mechanisms, manipulate the modification status to affect gene expression, and develop small molecules or other means to tune these pathways, pushing the study of epitranscriptomics to progress on both fronts of fundamental research and therapeutic applications in the future.

List of References

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

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

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 Kornberg, R. D. Structure of Chromatin. *Annu Rev Biochem* **46**, 931-954 (1977).

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

7 Allfrey, V. G., Faulkner, R. & Mirsky, A. E. 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).

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

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

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

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

12 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).

13 Li, E., Beard, C. & Jaenisch, R. Role for DNA Methylation in Genomic Imprinting. *Nature* **366**, 362-365 (1993).

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

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

16 Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Embryogenesis - Demethylation of the zygotic paternal genome. *Nature* **403**, 501-502 (2000).

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

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

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

20 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).

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

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

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

24 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).

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

26 Meehan, R. R., Lewis, J. D., Mckay, S., Kleiner, E. L. & Bird, A. P. Identification of a Mammalian Protein That Binds Specifically to DNA Containing Methylated Cpgs. *Cell* **58**, 499-507 (1989).

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

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

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

30 Watt, F. & Molloy, P. L. Cytosine Methylation Prevents Binding to DNA of a Hela-Cell Transcription Factor Required for Optimal Expression of the Adenovirus Major Late Promoter. *Gene Dev* **2**, 1136-1143 (1988).

31 Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-485 (2000).

32 Blackledge, N. P., Thomson, J. P. & Skene, P. J. CpG Island Chromatin Is Shaped by Recruitment of ZF-CxxC Proteins. *Cold Spring Harb Perspect Biol* **5** (2013).

33 Kriaucionis, S. & Tahiliani, M. Expanding the epigenetic landscape: novel modifications of cytosine in genomic DNA. *Cold Spring Harb Perspect Biol* **6**, a018630 (2014).

34 Cummings, D. J., Tait, A. & Goddard, J. M. Methylated Bases in DNA from Paramecium-Aurelia. *Biochim Biophys Acta* **374**, 1-11 (1974).

35 Harrison, G. S., Findly, R. C. & Karrer, K. M. Site-Specific Methylation of Adenine in the Nuclear Genome of a Eukaryote, Tetrahymena-Thermophila. *Mol Cell Biol* **6**, 2364-2370 (1986).

36 Hattman, S., Kenny, C., Berger, L. & Pratt, K. Comparative study of DNA methylation in three unicellular eucaryotes. *J Bacteriol* **135**, 1156-1157 (1978).

37 Rae, P. M. & Spear, B. B. Macronuclear DNA of the hypotrichous ciliate Oxytricha fallax. *Proc Natl Acad Sci U S A* **75**, 4992-4996 (1978).

38 Fu, Y. *et al.* N^6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. *Cell* **161**, 879-892 (2015).

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

40 Zhang, G. *et al.* N^6-methyladenine DNA modification in Drosophila. *Cell* **161**, 893-906 (2015).

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

42 Slotkin, W. & Nishikura, K. Adenosine-to-inosine RNA editing and human disease. *Genome Med* **5**, 105 (2013).

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

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

45 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).

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

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

48 Littlefield, J. W. & Dunn, D. B. Occurrence and Distribution of Thymine and 3 Methylated-Adenine Bases in Ribonucleic Acids from Several Sources. *Biochem J* **70**, 642-651 (1958).

49 Svensson, I., Boman, H. G., Eriksson, K. G. & Kjellin, K. Studies on microbial RNA: I. Transfer of methyl groups from methionine to soluble RNA from Escherichia coli. *J Mol Biol* **7**, 254IN257-271IN258 (1963).

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

51 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).

52 Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc Natl Acad Sci U S A* **71**, 3971-3975 (1974).

53 Wei, C. M., Gershowitz, A. & Moss, B. Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. *Cell* **4**, 379-386 (1975).

54 Wei, C. M., Gershowitz, A. & Moss, B. 5'-Terminal and Internal Methylated Nucleotide-Sequences in Hela-Cell Messenger-RNA. *Biochemistry* **15**, 397-401 (1976).

55 Wei, C. M. & Moss, B. Nucleotide-Sequences at N^6-Methyladenosine Sites of Hela-Cell Messenger Ribonucleic-Acid. *Biochemistry* **16**, 1672-1676 (1977).

56 Kane, S. E. & Beemon, K. Precise Localization of m^6A in Rous-Sarcoma Virus-RNA Reveals Clustering of Methylation Sites - Implications for RNA Processing. *Mol Cell Biol* **5**, 2298-2306 (1985).

57 Carroll, S. M., Narayan, P. & Rottman, F. M. N^6-methyladenosine residues in an intron-specific region of prolactin pre-mRNA. *Mol Cell Biol* **10**, 4456-4465 (1990).

58 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).

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

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

61 Keith, J. M., Ensinger, M. J. & Moss, B. Hela-Cell RNA (2'-O-Methyladenosine-N^6-)-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., Rath-Shambaugh, M. E., Ludwiczak, R., Narayan, P. & Rottman, F. 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., Shambaugh, M. E., Polayes, D., Matera, A. G. & Rottman, F. M. 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 Fu, Y. *et al.* FTO-mediated formation of N-6-hydroxymethyladenosine and N-6-formyladenosine in mammalian RNA. *Nature Communications* **4** (2013).

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

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

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

74 Kariko, K., Buckstein, M., Ni, H. P. & Weissman, D. Suppression of RNA recognition by Toll-like receptors: The impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23**, 165-175 (2005).

75 Shi, H. *et al.* YTHDF3 facilitates translation and decay of N^6-methyladenosine-modified RNA. *Cell Res* **27**, 315-328 (2017).

76 Hafner, M. *et al.* PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J Vis Exp* (2010).

77 Behm-Ansmant, I. *et al.* mRNA degradation by miRNAs and GW182 requires both CCR4 : NOT deadenylase and DCP1 : DCP2 decapping complexes. *Gene Dev* **20**, 1885-1898 (2006).

78 Gao, X. *et al.* Quantitative profiling of initiating ribosomes in vivo. *Nat Methods* **12**, 147-153 (2015).

79 Kedersha, N. & Anderson, P. Mammalian stress granules and processing bodies. *Methods Enzymol* **431**, 61-81 (2007).

80 Fraser, C. S. & Doudna, J. A. Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat Rev Microbiol* **5**, 29-38 (2007).

81 Hertz, M. I., Landry, D. M., Willis, A. E., Luo, G. X. & Thompson, S. R. Ribosomal Protein S25 Dependency Reveals a Common Mechanism for Diverse Internal Ribosome Entry Sites and Ribosome Shunting. *Mol Cell Biol* **33**, 1016-1026 (2013).

82 Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745 (2009).

83 Zhou, J. *et al.* Dynamic m^6A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591-594 (2015).

84 Pearson, W. R., Wood, T., Zhang, Z. & Miller, W. Comparison of DNA sequences with protein sequences. *Genomics* **46**, 24-36 (1997).

85 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515 (2010).

86 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).

87 Corcoran, D. L. *et al.* PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* **12**, R79 (2011).

88 Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**, 576-589 (2010).

89 Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**, 233-237 (2012).

90 Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).

91 Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One* **6**, e21800 (2011).

92 Shi, Y. *et al.* Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422**, 735-738 (2003).

93 Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* **224**, 213-232 (2006).

94 Koike, N. *et al.* Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals. *Science* **338**, 349-354 (2012).

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

96 Feillet, C., van der Horst, G. T. J., Levi, F., Rand, D. A. & Delaunay, F. Coupling between the Circadian Clock and Cell Cycle Oscillators: Implication for Healthy Cells and Malignant Growth. *Front Neurol* **6**, 96 (2015).

97 Aguilo, F. *et al.* Coordination of m^6A mRNA Methylation and Gene Transcription by ZFP217 Regulates Pluripotency and Reprogramming. *Cell Stem Cell* **17**, 689-704 (2015).

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

99 Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033-3042 (2009).

100 Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic Genome Activation During the Maternal-to-Zygotic Transition. *Annu Rev Cell Dev Biol* **30**, 581-613 (2014).

101 Walser, C. B. & Lipshitz, H. D. Transcript clearance during the maternal-to-zygotic transition. *Curr Opin Genet Dev* **21**, 431-443 (2011).

102 Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75-79 (2006).

103 Mishima, Y. & Tomari, Y. Codon Usage and 3'UTR Length Determine Maternal mRNA Stability in Zebrafish. *Mol Cell* **61**, 874-885 (2016).

104 Mathavan, S. *et al.* Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genet* **1**, 260-276 (2005).

105 Aanes, H. *et al.* Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res* **21**, 1328-1338 (2011).

106 Rabani, M. *et al.* High-Resolution Sequencing and Modeling Identifies Distinct Dynamic RNA Regulatory Strategies. *Cell* **159**, 1698-1710 (2014).

107 Gerber, A. P., Luschnig, S., Krasnow, M. A., Brown, P. O. & Herschlag, D. Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **103**, 4487-4492 (2006).

108 Laver, J. D. *et al.* Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the Drosophila maternal-to-zygotic transition. *Genome Biology* **16**, 94 (2015).

109 Stoeckius, M. *et al.* Global characterization of the oocyte-to-embryo transition in Caenorhabditis elegans uncovers a novel mRNA clearance mechanism. *EMBO J* **33**, 1751-1766 (2014).

110 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).

111 Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663-676 (2006).

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

113 Zhang, C. *et al.* Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m^6A-demethylation of NANOG mRNA. *Proc Natl Acad Sci U S A* **113**, E2047–E2056 (2016).

114 Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of Embryonic-Development of the Zebrafish. *Dev Dynam* **203**, 253-310 (1995).

115 Thisse, B. & Thisse, C. In Situ Hybridization on Whole-Mount Zebrafish Embryos and Young Larvae. *Methods Mol Biol* **1211**, 53-67 (2014).

116 Chen, K. *et al.* High-Resolution N-6-Methyladenosine (m(6)A) Map Using Photo-Crosslinking-Assisted m(6)A Sequencing. *Angew Chem Int Edit* **54**, 1587-1590 (2015).

117 Linder, B. *et al.* Single-nucleotide-resolution mapping of m^6A and m^6Am throughout the transcriptome. *Nat Methods* **12**, 767-772 (2015).

118 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36 (2013).

119 de Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453-1454 (2004).

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

121 Reddy, S. M. *et al.* Clinical and genetic predictors of weight gain in patients diagnosed with breast cancer. *Br J Cancer* **109**, 872-881 (2013).

122 Jin, D.-I. *et al.* Expression and roles of Wilms' tumor 1-associating protein in glioblastoma. *Cancer Science* **103**, 2102-2109 (2012).

123 Lin, Y. *et al.* Association between variations in the fat mass and obesity-associated gene and pancreatic cancer risk: a case-control study in Japan. *BMC cancer* **13**, 337 (2013).

124 Machiela, M. J. *et al.* Association of type 2 diabetes susceptibility variants with advanced prostate cancer risk in the Breast and Prostate Cancer Cohort Consortium. *Am J Epidemiol* **176**, 1121-1129 (2012).

125 Bokar, J. A. in *Fine-Tuning of RNA Functions by Modification and Editing* (ed Henri Grosjean) 141-177 (Springer Berlin Heidelberg, 2005).

126 Lin, S., Choe, J., Du, P., Triboulet, R. & Gregory, R. I. The m^6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol cell* **62**, 335-345 (2016).

127 Zhao, X. *et al.* FTO-dependent demethylation of N^6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res* **24**, 1403-1419 (2014).

128 Merkestein, M. *et al.* FTO influences adipogenesis by regulating mitotic clonal expansion. *Nature communications* **6**, 6792 (2015).

129 Zhang, M. *et al.* The Demethylase Activity of FTO (Fat Mass and Obesity Associated Protein) Is Required for Preadipocyte Differentiation. *PLoS One* **10**, e0133788 (2015).

130 Loos, R. J. & Yeo, G. S. The bigger picture of FTO: the first GWAS-identified obesity gene. *Nature reviews Endocrinology* **10**, 51-61 (2014).

131 Frayling, T. M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889-894 (2007).

132 Hess, M. E. *et al.* The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry. *Nat Neurosci* **16**, 1042-1048 (2013).

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

134 Du, T. *et al.* An association study of the m^6A genes with major depressive disorder in Chinese Han population. *J Affect Disord* **183**, 279-286 (2015).

135 Milaneschi, Y. *et al.* The effect of FTO rs9939609 on major depression differs across MDD subtypes. *Mol Psychiatry* **19**, 960-962 (2014).

136 Rivera, M. *et al.* Depressive disorder moderates the effect of the FTO gene on body mass index. *Mol Psychiatry* **17**, 604-611 (2012).

137 Samaan, Z. *et al.* The protective effect of the obesity-associated rs9939609 A variant in fat mass- and obesity-associated gene on depression. *Mol Psychiatry* **18**, 1281-1286 (2013).

138 Choudhry, Z. *et al.* Association between obesity-related gene FTO and ADHD. *Obesity* **21**, E738-E744 (2013).

139 Sobczyk-Kopciol, A. *et al.* Inverse association of the obesity predisposing FTO rs9939609 genotype with alcohol consumption and risk for alcohol dependence. *Addiction* **106**, 739-748 (2011).

140 Rowles, J., Wong, M., Powers, R. & Olsen, M. FTO, RNA epigenetics and epilepsy. *Epigenetics* **7**, 1094-1097 (2012).

141 Casalegno-Garduño, R., Schmitt, A., Wang, X., Xu, X. & Schmitt, M. Wilms' Tumor 1 as a Novel Target for Immunotherapy of Leukemia. *Transplant Proc* **42**, 3309-3311 (2010).

142 Kennedy, Edward M. *et al.* Posttranscriptional m^6A Editing of HIV-1 mRNAs Enhances Viral Gene Expression. *Cell Host Microbe* **19**, 675-685 (2016).

143 Lichinchi, G. *et al.* Dynamics of the human and viral m^6A RNA methylomes during HIV-1 infection of T cells. *Nature Microbiol* **1**, 16011 (2016).

144 Song do, K., Lee, H., Oh, J. Y., Hong, Y. S. & Sung, Y. A. FTO Gene Variants Are Associated with PCOS Susceptibility and Hyperandrogenemia in Young Korean Women. *Diabetes Metab J* **38**, 302-310 (2014).

145 Utsch, B. *et al.* Exclusion of WTAP and HOXA13 as candidate genes for isolated hypospadias. *Scand J Urol Nephrol* **37**, 498-501 (2003).

146 Tirumuru, N. *et al.* N^6-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression. *eLife* **5**, e15528 (2016).

147 Krug, R. M., Morgan, M. A. & Shatkin, A. J. Influenza viral mRNA contains internal N^6-methyladenosine and 5'-terminal 7-methylguanosine in cap structures. *J Virol* **20**, 45-53 (1976).

148 Beemon, K. & Keith, J. Localization of N^6-methyladenosine in the Rous sarcoma virus genome. *J Mol Biol* **113**, 165-179 (1977).

149 Narayan, P. & Rottman, F. M. in *Advances in Enzymology and Related Areas of Molecular Biology* 255-285 (John Wiley & Sons, Inc., 2006).

150 Lichinchi, G. *et al.* Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. *Cell Host Microbe* **20**, 666-673 (2016).

151 Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462-477 (2013).

152 Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* **18**, 1435-1440 (2011).

153 Goff, S. P. Host factors exploited by retroviruses. *Nat Rev Microbiol* **5**, 253-263 (2007).

154 Wang, J. H., Janas, A. M., Olson, W. J., KewalRamani, V. N. & Wu, L. CD4 coexpression regulates DC-SIGN-mediated transmission of human immunodeficiency virus type 1. *J Virol* **81**, 2497-2507 (2007).

155 St Gelais, C., Roger, J. & Wu, L. Non-POU Domain-Containing Octamer-Binding Protein Negatively Regulates HIV-1 Infection in CD4(+) T Cells. *AIDS Res Hum Retroviruses* **31**, 806-816 (2015).

156 Dong, C., Janas, A. M., Wang, J. H., Olson, W. J. & Wu, L. Characterization of human immunodeficiency virus type 1 replication in immature and mature dendritic cells reveals dissociable cis- and trans-infection. *J Virol* **81**, 11352-11362 (2007).

157 Hamel, R. *et al.* Biology of Zika Virus Infection in Human Skin Cells. *J Virol* **89**, 8880-8896 (2015).

158 Ye, F., Chen, E. R. & Nilsen, T. W. Kaposi's Sarcoma-Associated Herpesvirus Utilizes and Manipulates RNA N^6-Adenosine Methylation to Promote Lytic Replication. *J Virol* (2017).

159 Sonoda, Y. *et al.* Formation of intracranial tumors by genetically modified human astrocytes defines four pathways critical in the development of human anaplastic astrocytoma. *Cancer Res* **61**, 4956-4960 (2001).

160 Wuarin, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol Cell Biol* **14**, 7219-7225 (1994).

161 Zhang, N. *et al.* FoxM1 promotes beta-catenin nuclear localization and controls Wnt target-gene expression and glioma tumorigenesis. *Cancer Cell* **20**, 427-442 (2011).

162 Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N. & Rechavi, G. Transcriptome-wide mapping of N(6)-methyladenosine by m(6)A-seq based on immunocapturing and massively parallel sequencing. *Nat Protoc* **8**, 176-189 (2013).

163 Wu, L., Martin, T. D., Carrington, M. & KewalRamani, V. N. Raji B cells, misidentified as THP-1 cells, stimulate DC-SIGN-mediated HIV transmission. *Virology* **318**, 17-23 (2004).

164 Dang, J. *et al.* Zika Virus Depletes Neural Progenitors in Human Cerebral Organoids through Activation of the Innate Immune Receptor TLR3. *Cell Stem Cell* **19**, 258-265 (2016).

165 De Silva, S., Planelles, V. & Wu, L. Differential effects of Vpr on single-cycle and spreading HIV-1 infections in CD4+ T-cells and dendritic cells. *PLoS One* **7**, e35385 (2012).

166 Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell* **151**, 1417-1430 (2012).

167 Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074-1080 (2001).

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

169 Patil, D. P. *et al.* m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369-373 (2016).

170 Hoki, Y. *et al.* A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development* **136**, 139-146 (2009).

171 Chu, C. *et al.* Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161**, 404-416 (2015).

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

173 Li, X. *et al.* Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. *Nat Chem Biol* **12**, 311-316 (2016).

174 Zhou, H. *et al.* m1A and m1G disrupt A-RNA structure through the intrinsic instability of Hoogsteen base pairs. *Nat Struct Mol Biol* **23**, 803-810 (2016).

175 Squires, J. E. *et al.* Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* **40**, 5023-5033 (2012).

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

177 Fu, L. *et al.* Tet-Mediated Formation of 5-Hydroxymethylcytosine in RNA. *J Am Chem Soc* **136**, 11582-11585 (2014).

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

179 Huber, S. M. *et al.* Formation and abundance of 5-hydroxymethylcytosine in RNA. *Chembiochem* **16**, 752-755 (2015).

180 Lane, B. G. in *Modification and Editing of RNA* (ed H. Grosjean, Benne, R.) 1-20 (American Society of Microbiology, 1998).

181 Schwartz, S. *et al.* Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-162 (2014).

182 Carlile, T. M. *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143-146 (2014).

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

184 Fernandez, I. S. *et al.* Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature* **500**, 107-110 (2013).

185 Smith, J. D. & Dunn, D. B. An additional sugar component of ribonucleic acids. *Biochim Biophys Acta* **31**, 573-575 (1959).

186 Hall, R. H. Method for isolation of 2'-O-methylribonucleosides and N1-methyladenosine from ribonucleic acid. *Biochim Biophys Acta* **68**, 278-283 (1963).

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

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

189 Heard, E. & Martienssen, R. A. Transgenerational Epigenetic Inheritance: Myths and Mechanisms. *Cell* **157**, 95-109 (2014).

190 Probst, A. V., Dunleavy, E. & Almouzni, G. Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Bio* **10**, 192-206 (2009).