THE UNIVERSITY OF CHICAGO

DECODING THE EPITRANSCRIPTOME: READER PROTEINS TAILORING REGULATORY FUNCTIONS OF mRNA MODIFICATIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

HAILING SHI

CHICAGO, ILLINOIS

AUGUST 2019

Table of Contents

List of Figures

Acknowledgements

First and foremost, I would like to express very great appreciation to my doctorate advisor Prof. Chuan He. I have always been encouraged and motivated by his appealing confidence and ambition, his continuous enthusiasm for fundamental scientific questions, all his hard work devoted to the laboratory, and his genuine care for students. Chuan gives me great freedom and trust to test out a wide range of projects. He has provided me with many opportunities to collaborate with labs of various expertise as well as to present and network in academic conferences. My research would have been impossible without his insightful guidance and steady support.

I also want to thank members in the He laboratory. It has been a great experience studying and working here around so many great people and brilliant minds. I would like to offer my special thanks to Dr. Xiao Wang, who introduced me to all kinds of techniques in RNA biology when I first joint the lab. Although that was only a one-year overlap, we two have been friends since then and it was very kind of her to continue providing advice for both my research and career development even after she left the lab. Dr. Boxuan S. Zhao and Dr. Phillip J. Hsu have worked closely with me on multiple projects and I really enjoyed their company. Zhike Lu, Scott Zhang, and Dr. Lijia Ma helped in bioinformatics analyses. Dr. Honghui Ma gave me valuable guidance in studying protein interactome. Thanks also go to Chang Liu, Zhongyu Zou, Yuxi Ai, Boyang Gao, and Dr. Yuru Wang for their help with experiments. Dr. Kathy Liu, Dr. Jun Liu, Jiangbo Wei, Tong Wu, and Dr. Bryan Harada provided valuable discussion. Dr. Jordi Tauler and Allen C. Zhu provided valuable help on text editing and was greatly appreciated. I sincerely thank all other members.

I also wish to acknowledge my research collaborators. Prof. Tao Zhou and Dr. Xuliang Zhang from ShanghaiTech University in China worked very closely with me on the neuron project.

Thanks to the time-zone difference between China and the U.S., our collaboration doubled the efficiency and yielded fruitful results. Prof. Hongjun Song and Dr. Yi-Lan Weng from the University of Pennsylvania and Prof. Xiaoxi Zhuang and Dr. Jary Y. Delgado from the Department of Neurobiology also provided crucial help on this study. Prof. Michiel Vermeulen from Radboud University in the Netherlands collaborated with us on systematically discovering proteins recruited or repelled by RNA modifications. Prof. Jianjun Chen at City of Hope worked with us on characterizing an additional class of RNA methylation binding proteins. I have also been working on several cancer biology projects with Prof. Yanhong Shi from City of Hope and Prof. Suyun Huang from MD Anderson Cancer Center of the University of Texas. I also thank Prof. Jingyi Fei and Jeff Zhang for their help in super-resolution microscope imaging. My special thanks are extended to the staff in facilities of mass spectrometry and NGS sequencing at the university.

Heartfelt thanks go to members of my dissertation committee, Prof. Tao Pan, Prof. Bryan C. Dickinson, and Prof. Xiaoxi Zhuang, for their time and valuable suggestions. Graduate classes taught by Tao and Bryan benefited me a lot in understanding basics in RNA and chemical biology. Advices given by Tao and Xiaoxi have been a great help to get me prepared for the career after graduate school.

Finally, I am profoundly grateful to supports from my friends and family. I am more than glad to have six classmates from the same undergraduate institute and department as me pursuing Chemistry Ph.D. degree together here. Life in graduate school without them would be much less vibrant. Thank you to my dear parents for their unconditionally support and love during every moment of my life. Their continuous care, encouragement, and understanding have always been a great source of comfort and motivation of mine.

Abstract

Eukaryotic cells utilize a wide range of mechanisms to regulate gene expression, generating divergent expression profiles for distinct cellular functions. In addition to the primary sequences of DNA and RNA that code for protein, modifications on the nucleic acids serve as additional "codes" for precise control of gene product level. "Epigenetics" studies how DNA and histone modifications alter gene expression without changing DNA sequences. As an analogy, a newly-emerging field called "epitranscriptomics" focuses on chemical marks on RNA species. N^6-methyladenosine (m^6A) is the most abundant internal modification on eukaryotic messenger RNA (mRNA) and impacts various aspects of RNA function and metabolism. m^6A methylation can be recognized by assorted binding proteins called "readers", and each reader may group a subset of m^6A-modified transcripts for particular downstream processing. In this thesis, we report the translational promotion function of a reader YT521-B homology domain family 3 (YTHDF3). We then take the physiological example of learning and memory in mice to demonstrate the biological significance of m^6A-mediated translational regulation through another reader YTHDF1. We also report the discovery and characterization of additional classes of m^6A readers. Studies of readers proteins may provide valuable mechanistic insights into the multifunctionality as well as the specificity of functions of RNA modifications, not only in proper cell fate transition but also in optimal cellular response.

List of Publications Based on Work Presented in this Thesis[†]

1. **H Shi***, J Wei*, C He. Where, when and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Molecular Cell*, **2019**, 74 (4):640-650. [Review]
2. **H Shi***, X Zhang*, Y-L Weng*, Z Lu, Y Liu, Z Lu, J Li, P Hao, Y Zhang, F Zhang, Y Wu, JY Delgado, Y Su, MJ Patel, X Cao, B Shen, X Huang, G-l Ming, X Zhuang, H Song[§], C He[§], T Zhou[§]. m^6A facilitates hippocampus-dependent learning and memory through YTHDF1. *Nature*, **2018**, 563:249-253.
3. M Li*, X Zhao*, W Wang, **H Shi**, Q Pan, Z Lu, SP Perez, R Suganthan, C He, M Bjørås, A Klungland. Ythdf2-mediated m^6A mRNA clearance modulates neural development in mice. *Genome Biology*, **2018**, 19(1):69.
4. J Wei*, F Liu*, Z Lu, Q Fei, Y Ai, PC He, **H Shi**, X Cui, R Su, A Klungland, G Jia, J Chen, C He. Differential m^6A, m^6Am, and m^1A demethylation mediated by FTO in the cell nucleus and cytoplasm. *Molecular Cell*, **2018**, 71(6): 973-985.e5.
5. JL Koranda, L Dore, **H Shi**, MJ Patel, LO Vaasjo, MN Rao, K Chen, Z Lu, Y Yi, W Chi, C He[§], X Zhuang[§]. Mettl14 is essential for epitranscriptomic regulation of striatal function and learning. *Neuron*, **2018**, 99: 283-292.e5
6. H Huang*****, H Weng*****, W Sun*****, X Qin*****, **H Shi***, H Wu, BS Zhao, A Mesquita, C Liu, CL Yuan, Y-C Hu, S Hüttelmaier, JR Skibbe, R Su, X Deng, L Dong, M Sun, C Li, S Nachtergaele, Y Wang, C Hu, K Ferchen, KD Greis, X Jiang, M Wei, L Qu, J-L Guan, C He[§], J Yang[§], J Chen[§]. Recognition of RNA N^6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nature Cell Biology*, **2018**, 20: 285–295.

[†] The following chapters of this dissertation contain sections and figures adopted from the listed publications with modifications. Chapter 1: publication 1; Chapter 2: publication 11; Chapter 3: publications 2 & 5; Chapter 4: publications 6 & 9; Chapter 5: publication 1.

7. **H Shi**, C He. Phasing gene expression: mRNA N^6-methyladenosine regulates temporal progression of mammalian cortical neurogenesis. *Biochemistry*, **2018**, 57 (7): 1055–1056. [Viewpoints]
8. PJ Hsu, **H Shi**, C He. Epitranscriptomic influences on development and disease. *Genome Biol*. **2017**,18: 197. [Review]
9. RR Edupuganti*****, S Geiger*****, RGH Lindeboom*****, **H Shi***, PJ Hsu, Z Lu, SY Wang, MPA Baltissen, PWTC Jansen, M Rossa, M Müller, HG Stunnenberg, C He[§], T Carell[§], M Vermeulen[§]. N^6-methyladenosine (m^6A) recruits and repels proteins to regulate mRNA homeostasis. *Nature Structural & Molecular Biology*. **2017**, 24: 870–878.
10. PJ Hsu*****, Y Zhu*****, H Ma*****, Y Guo*****, X Shi, Y Liu, M Qi, Z Lu, **H Shi**, J Wang, Y Cheng, G Luo, Q Dai, M Liu, X Guo, J Sha, B Shen[§], C He[§]. Ythdc2 is an N^6-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Research*, **2017**, 27: 1115–1127.
11. **H Shi***, X Wang*, BS Zhao, Z Lu, H Ma, PJ Hsu, C Liu, C He. YTHDF3 facilitates translation and decay of the N^6-methyladenosine-modified RNA. *Cell Research*, **2017**, 27:315–328.
12. C Qi, **H Shi**, P Ye, L Li, Q Qu, G Sun, G Su, Z Lu, Y Huang, C-G Yang, AD Riggs, C He[§], Y Shi[§]. m^6A RNA methylation regulates the self-renewal and tumorigenesis of glioblastoma stem cells. *Cell Reports*, **2017**, 18(11): 2622-2634.
13. BS Zhao*, X Wang*, AC Beadell*, Z Lu, **H Shi**, A Kuuspalu, RK Ho[§], C He[§], m^6A-dependent maternal mRNA clearance facilitates maternal-to-zygotic transition in zebrafish, *Nature*, **2017**, 542: 475–478.
14. **H Shi**, C He, A glance at N^6-methyladenosine in transcript isoforms. *Nature Methods*, **2016**, 13: 624–625. [News and Views]
15. X Wang*, BS Zhao*, IA Roundtree, Z Lu, D Han, H Ma, X Weng, K Chen, **H Shi**, C He. N^6-methyladenosine modulates messenger RNA translation efficiency, *Cell*, **2015**, 161 (6): 1388–1399.

*Co-authors contributed equally.
[§]These authors jointly supervised the work.

Chapter 1

Introduction: Epitranscriptome and Its Effectors

1.1 Epigenetics: modifications above the primary sequences

The central dogma of molecular biology describes the flow of genetic information, from deoxyribonucleic acid (DNA) to ribonucleic acid (RNA) to protein. Genes are coded in segments of DNA, transcribed into the intermediate carrier messenger RNAs, and then translated into amino acid sequences of proteins that execute all sorts of cellular functions. In multicellular organisms, alongside this information flow is a wide range of complex mechanisms modulating final gene product level and function, accounting for cell type- and tissue-specific gene expression from the same DNA as well as divergent phenotype from identical genotype (**Figure 1.1**).

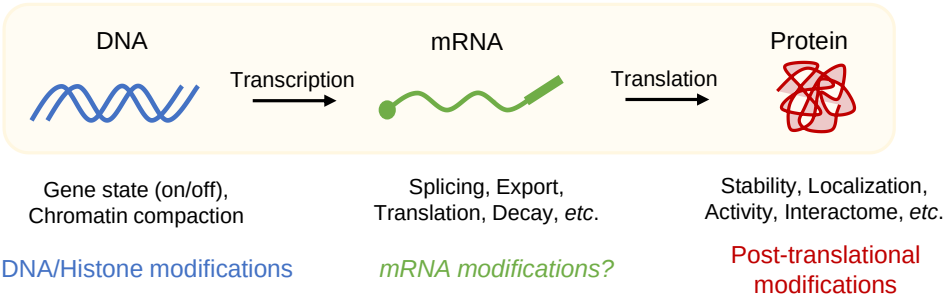

Figure 1.1 Regulatory mechanisms surrounding the central dogma

Key elements and steps in the central dogma are regulated to control gene product level and function. DNA and histones can be modified to modulate gene state—a major research subject of epigenetics (see below), while post-translational modifications on proteins fine-tune their functions. Each step in the life-cycle of mRNA is also precisely controlled, where mRNA modifications may come into play.

In 1942, Conrad Hal Waddington coined the term "epigenetics" and used "epigenotype" to refer to factors besides genotype and environment that dictate context-dependent phenotype[1]. This revolutionary view nurtured and later evolved into a way profound field—the study of stable and heritable changes in gene expression or cellular phenotype resulting from changes in a

chromosome but not in DNA sequences[2,3]. Nowadays, well-established epigenetic mechanisms include DNA modifications, histone modifications, chromatin remodelers, and non-coding RNAs.

5-methylcytosine (5-mC) is the most well studied DNA modification. It is enriched at symmetric CpG dinucleotides in mammals, a key feature that allows modification heredity during cell division[4]. 5-mC is established and maintained by DNA methyltransferases (DNMT), and is subject to active demethylation by ten-eleven translocation enzymes (TET). The regulatory functions of 5-mC is region-dependent. For instance, it is a repressive mark on promoters and enhancers, while gene-body 5-mC is linked to active transcription[4,5]. Other modified DNA bases, including 5-hydroxymethylcytosine (5-hmC) and N^6-methyladenosine (6-mA), are new candidate epigenetic marks being actively studied in model organisms beyond mammals[5].

Histone proteins, the other major components of eukaryotic nucleosomes, also bear assorted post-translational modifications on their *N*-terminal tails and central globular domains[6]. Each type of modification (e.g. methylation, acetylation, phosphorylation) that occurs at a specific histone residue has a distinct distribution pattern and is linked to a unique gene state. Marks associated with active genes include H3K4me3 at promoter regions, H3K9ac and H3K27ac at enhancers and promoters, and H3K36me3 enriched towards the end of the gene body; whereas those linked to gene silencing include H3K9me3 and H3K27me3 in heterochromatin regions[6,7]. Dynamic regulation has been observed for many of the histone modifications in development and diseases, highlighting their biological significance.

1.2 Epitranscriptomics: chemical modifications on RNAs

In contrast, whether similar “epigenetic” mechanisms exist at the mRNA level had remained unexplored until conceptually proposed less than a decade ago[8]. More than 150 distinct

chemical marks on cellular RNA have been identified to date[9,10] since the discovery the first structurally modified nucleoside pseudo-uridine in the 1950s[11]. Among various RNA species, transfer RNA (tRNA) is most extensively modified (11 to 13 modified sites in 76 to 90 nucleotides of length per molecule in human) followed by ribosomal RNA (rRNA) (~100 pseudo-uridine sites and ~100 2'-*O*-ribose methylation sites). Their biogenesis, metabolism, and functions are heavily modulated by those modifications[12–14].

However, it was only until several years ago that broad interest in RNA modification biology resurged as a field named "epitranscriptomics", prompted by recognition of the prevalence and functional significance of internal mRNA modifications, most prominently N^6-methyladenosine (m^6A, **Figure 1.2**). m^6A is the most well-characterized mRNA modification so far. It was first discovered in 1974 as the major form of internal methylation on mammalian mRNA[15,16] and early work showed that it occurs in a sequence context as (G/A)(m^6A)C at the level of 0.1-0.4% of all mRNA adenosine[17–19]. More recent transcriptome-wide m^6A site mapping has given greater detail on its localization and prominence, revealing its prevalence in thousands of transcripts and its unique and conserved distribution preferentially centered around stop codons and 3' untranslated regions (3'UTRs) in the transcriptomes of humans and mice[20–22].

Using chemical/enzyme-based or antibody-based high-throughput sequencing methods, researchers have characterized the distribution of a handful of additional mRNA modifications on the eukaryotic transcriptome (**Figure 1.2**), including pseudo-uridine (Ψ)[23,24], N^1-methyladenosine (m^1A)[25–28], 5-methylcytosine (m^5C)[29,30], 5-hydroxymethylcytosine (hm^5C)[31], 2'-*O*-ribose methylation (Nm)[32], N^6,2'-*O*-dimethyladenosine (m^6A_m)[21], and N^4-acetylcytidine (ac^4C)[33]. Although functional characterization of some of these modifications is still in a very early stage and their low abundance may cause technical challenges, this newfound diversity of mRNA

modifications and perhaps their crosstalk may suggest a complex regulation of mRNA beyond our current understanding.

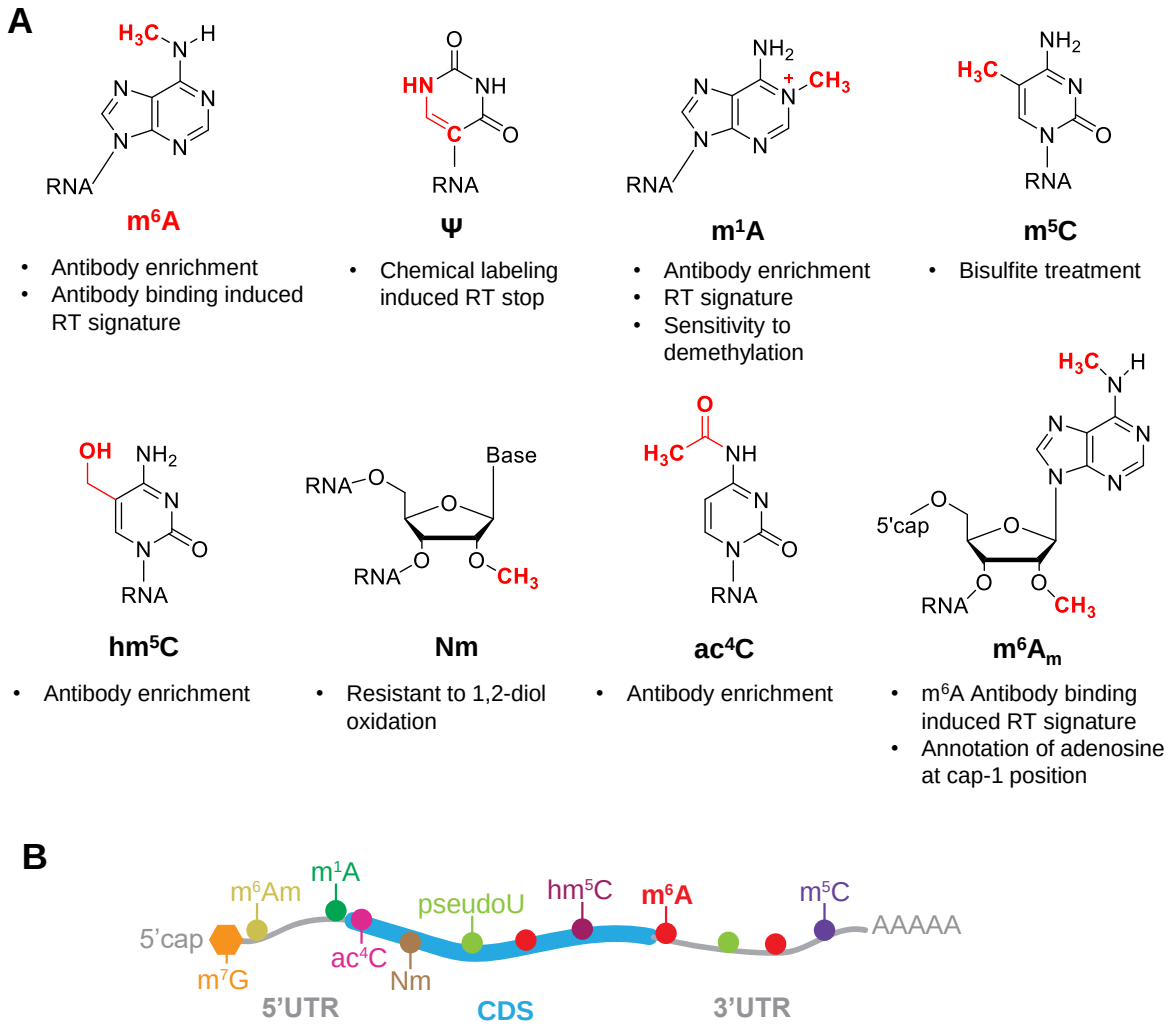

Figure 1.2 Structures and distributions of major mRNA modifications

(**A**) Chemical structures of known mRNA modifications with the chemical modification highlighted in red. Principles of reported strategies for transcriptome-wide modification site mapping are listed under each structure. (**B**) Illustration of the distribution pattern of each type of modification along the 5'UTR, CDS, and 3'UTR of a transcript.

1.3 Effectors of the epitranscriptome: writers, erasers, and readers

The function of an RNA modification is intrinsically linked to its effector proteins, including enzymes ("writers" and "erasers") that alter the modification level and binding proteins

("readers") that recognize the chemical marks. m^{6}A methylation is the best studied modification in this aspect (**Figure 1.3**).

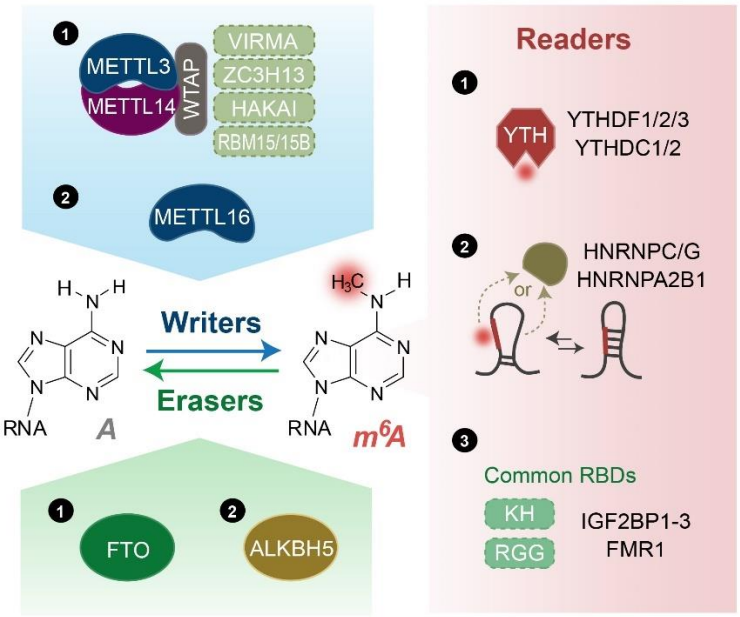

Figure 1.3 Effectors of mammalian m^{6}A pathway: writers, erasers, and readers.

Writers: The majority of m^{6}A methylation on mRNA is installed by a writer complex (①) composed of core subunits METTL3 and METTL14 and additional adaptors proteins. The other known writer METTL16 (②) installs m^{6}A on a hairpin structure of transcript *MAT2A*. **Erasers**: two erasers, FTO and ALKBH5, have been characterized for m^{6}A methylation on mRNA. **Readers**: Three classes of reader proteins utilize different mechanisms to preferentially bind to m^{6}A-containing RNAs. ①YTHDF1-3 and YTHDC1-2 utilize a well-characterized YTH domain to directly recognize m^{6}A. ② A local structure disrupted by m^{6}A could favor RNA-binding events of HNRNPC/G and HNRNPA2B1. ③ RNA binding proteins including IGF2BP1-3 and FMR1 prefer m^{6}A-containing RNAs through tandem common RNA binding domains (RBDs) via a mechanism yet to be characterized.

***m^{6}A writers*.** The majority of the m^{6}A on mRNA is installed co-transcriptionally in a sequence context of RRACH (R = A or G; H = A, C, or U), through a complex composed of multiple subunits with a stable core complex formed between methyltransferase-like 3 (METTL3) and methyltransferase-like 14 (METTL14)[34,35]— the former as the catalytic subunit and the latter as an essential component to facilitate RNA binding[36–38]. Further studies characterized a handful

of additional subunits and revealed how they contribute to the activity and specificity of the writer complex. Wilms tumor 1-associating protein (WTAP) binds to METTL3/14 and is required for optimal substrate recruitment and METTL3/14 localization[39,40]; Vir like m^{6}A methyltransferase associated (VIRMA) is critical for deposition of m^{6}A favorably to 3'UTRs of transcripts[41]; Zinc finger CCCH-type containing 13 (ZC3H13) facilitates nuclear localization of the writer complex and 3'UTR m^{6}A occurence[42]; and RNA binding motif protein 15/15B (RBM15/15B) is reported to bind U-riched regions and may facilitate methylation of certain RNAs[43]. In fruit flies, Zc3h13/Flacc is shown to stabilize the interaction between Wtap/Fl(2)d and Rbm15/Nito[44].

Another m^{6}A writer, methyltransferase-like 16 (METTL16), depends on a different sequence motif combined with a distinct structure context for targeting its substrates[45]. Two validated substrates of METTL16 are U6 small nuclear RNA (snRNA) and a hairpin (hp1) in the 3'UTR of human methionine adenosyltransferase 2A (*MAT2A*) mRNA that encodes the *S*-adenosylmethionine (SAM) synthetase[45]. Truncation/mutation tests, *in vitro* methylation selection assays, and crystal structures of METTL16 with its RNA substrates together suggested that METTL16 prefers a sequence of UAC(m^{6}A)GAGAA in the bulge of a stem loop-structured RNA[45–47]. An important functional implication for METTL16-mediated *MAT2A* methylation is to establish a negative feedback loop for SAM homeostasis, likely through mechanisms involving alternative splicing and m^{6}A-mediated mRNA decay[45,48].

m^{6}A erasers. The identification of enzymes capable of reversing m^{6}A especially sparked a revival of interest and new insights on this long known RNA base modification[8,49,50]. In 2010, my advisor Prof. Chuan He pioneered the concept of dynamic RNA modification for biological regulation[8]. Shortly after, our group reported fat mass and obesity-associated protein (FTO) and AlkB homolog 5 (ALKBH5) functioned as mRNA m^{6}A demethylases[49,50]. FTO is a versatile

demethylase, showing differential preferences for substrates—both RNA species and types of base methylation—in the cell nucleus versus the cytoplasm[51]. On mRNAs, FTO works on both internal m^6A and the m^6A moiety of the cap-m^6A_m under physiological conditions[51,52], although the functional relevance of the latter remains to be clarified. In contrast, mRNA m^6A has been known as the sole substrate of ALKBH5 so far[50].

m^6A readers. One class of direct and robust m^6A readers are proteins containing the YT521-B homology (YTH) domain, including YTH domain family 1-3 (YTHDF1-3) and YTH domain containing 1-2 (YTHDC1-2) in humans. After the initial characterization of YTH domain as a new RNA binding domain[53,54], YTHDF2 and YTHDF3 were reported to bind strongly to m^6A RNA probes[20] in affinity pull down assays. Cytoplasmic YTHDF2 promotes degradation of its target transcripts by recruiting the CCR4-NOT deadenylase complex and the endoribo-nuclease complex RNase P/MRP[55–57]. The other two cytoplasmic m^6A readers, YTHDF1 and YTHDF3, are suggested to promote translation of target transcripts by recruiting translation initiation factors[58–61]. Part of my doctoral work focuses on cellular roles of YTHDF3[59]. Nuclear reader YTHDC1 has been suggested to play multiple roles including regulating mRNA splicing by preferably recruiting a certain splicing factor[62], expediting mRNA export[63], and accelerating decay of certain transcripts[48]. YTHDC2 mediates both mRNA stability and translation and regulates spermatogenesis[64], and its additional RNA binding domains and helicase domains render detailed mechanistic studies enigmatic.

Another type of m^6A reader achieves preferential binding of m^6A-containing RNAs through common RNA binding domains (RBDs) such as K homology (KH) domains, RNA recognition motif (RRM) domains, and arginine/glycine-rich (RGG) domains, where RNA structure may play a role. The presence of m^6A can remodel local RNA structure and consequently

modulates RNA-protein interactions around or nearby, termed an "m^{6}A switch"[65]. Several heterogeneous nuclear ribonucleoproteins (HNRNPs) fall into this category, including HNRNPC, HNRNPG, and HNRNPA2B1, which regulate alternative splicing or processing of target transcripts[65–68]. Fragile X mental retardation 1 (FMR1) contains three KH domains and one RGG domain, and has been shown to prefer m^{6}A-containing RNA, impacting both RNA translation and RNA stability likely through interplay with YTHDF1 and YTHDF2, respectively[69,70]. Insulin-like growth factor 2 mRNA-binding proteins 1-3 (IGF2BP1-3), has been reported to stabilize target mRNA in an m^{6}A-dependent manner[71], in contrast to the role of YTHDF2 in decreasing mRNA stability. Mutation and truncation assays identified domains KH3-4 as indispensable for *in vitro* binding to an m^{6}A-modified RNA probe. A recent study added proline rich coiled-coil 2A (Prrc2a) to the list of m^{6}A readers[72], showing that recombinant Prrc2a prefers to bind a methylated probe and Prrc2a stabilizes a critical m^{6}A-modified transcript required for myelination. The exact domains of Prrc2a that recognize m^{6}A remain to be determined.

In summary, biochemical characterization of m^{6}A effectors, transcriptome-wide identification of their RNA substrates, and investigation on how the fate of substrates changes in the absence of effectors, have altogether provided valuable insights into the molecular mechanisms of m^{6}A-mediated RNA function and metabolism. The expanding list of m^{6}A effectors (especially readers) and their potential to crosstalk with one another suggest great complexities of RNA modification biology.

1.4 *N*^{6}-methyladenosine (m^{6}A) effectors as new biological controllers

Building on the understanding of the m^{6}A prevalence and molecular insights on its effectors, recent work has uncovered that m^{6}A plays an important role in gene expression regulation, development and differentiation, human diseases, and viral infection[14,73–75].

In terms of *in vivo* physiological significance of m^{6}A, studies in cell lineage differentiation and animal development have led the way. One of common theme there is that m^{6}A- and YTHDF2-mediated mRNA decay facilitates timely transcriptome clearance for proper cell fate transition. In naïve mouse embryonic stem cells (mESCs), depletion of METTL3 delays the decay of self-renewal factors that are prominent, preventing the cells from leaving the naïve state[76,77]; while knockdown of METTL3 in primed mESCs further accumulates lineage-commitment transcripts that are dominant, facilitating differentiation[78]. More recent *in vivo* examples include: (i) YTHDF2 facilitates maternal-to-zygotic transition in zebrafish and mice by regulating maternal transcript dosage[79,80]; (ii) METTL3 is critical for T cell homeostasis and differentiation in mice[81]; (iii) generation of blood cells progenitors in zebrafish embryogenesis[82]; and (iv) mouse spermatogenesis[83]. Interestingly, in the case of mouse cortical neurogenesis, m^{6}A-mediated decay is found to be critical for not only timely clearance of early stage marker but also "pre-patterning" of late stage markers—transcriptions of these genes are turned on slightly but transcripts are timely decayed to prevent protein production[84].

Besides mRNA decay, m^{6}A-mediated alternative splicing is another common mechanism underlying the its physiological importance. The reader YTHDC1 modulates splicing of the transcript *Sxl* and thus sex determination in fruit flies[85,86]. The eraser ALKBH5 impacts mouse fertility[50] and is critical for proper splicing and selective degradation of the transcripts with long 3'UTRs[87]. FTO has been reported to affect adipogenesis dependent on its demethylase activity[88] and its regulation of mRNA splicing[89]. Loss of FTO also impedes myogenesis[90] and adult neurogenesis[91] in mice, likely through complex mechanisms beyond splicing regulation.

In contrast, to what extent m^{6}A-mediated translational regulation matters for physiological processes had been largely unknown. In the heat shock response, m^{6}A installation was observed

at 5' untranslated regions (5'UTRs) of response genes, mediating to cap-independent translation[92]. In the dorsal root ganglion model of injury-induced axon regeneration, regeneration-associated genes (RAGs) are heavily m^{6}A-methylated. The loss of the writer component METTL14 or the reader YTHDF1 causes defects in the recovery process, likely due to attenuated *de novo* RAG protein synthesis[93]. Since reader proteins directly link m^{6}A-modified transcripts to downstream functional or processing machineries, investigating reader proteins holds the potential to elucidate physiological functions of m^{6}A at the molecular level. A major part of my work will be devoted to studying roles of m^{6}A in learning and memory from this perspective.

1.5 Scope of this thesis

m^{6}A methylation regulates multiple steps of RNA lifecycle, modulating RNA function, RNA metabolism, and gene expression. In this thesis, I will be focusing on how m^{6}A achieves its multifaceted regulatory roles through a collection of reader proteins.

Chapter 2 presents molecular characterization of a cytoplasmic m^{6}A reader YTHDF3 followed by discussion on the roles of m^{6}A on cytoplasmic mRNA metabolism.

Chapter 3 presents a physiological example showing how m^{6}A-mediated translational regulation through the reader YTHDF1 contributes to learning and memory formation in mice.

Chapter 4 presents the discovery and characterization of additional m^{6}A readers proteins that do not contain YTH domains.

Chapter 5 summarizes, on a broader perspective, recent advances in the understanding of context-dependent roles of m^{6}A. This chapter also discusses potential future directions in the field of epitranscriptomics.

Chapter 2

Characterization of cytoplasmic m^{6}A reader YTHDF3

2.1 Introduction: YTH domain family reader proteins

As introduced in section **1.3**, the YTH domain directly binds m^{6}A in single-stranded RNAs. In humans, there are three proteins under YTH domain family named YTHDF1-3. They share over 65% protein sequence identify and a highly similar domain organization. YTHDF proteins are composed of a low-complexity *N*-terminus and a YTH domain at the *C*-terminus, therefore functioning as an "adaptor" to bring proper cellular machineries towards the m^{6}A-containing target transcripts (**Figure 2.1A**).

Our lab has reported the regulatory functions of YTHDF1 and YTHDF2 in HeLa cells. In the cytosol, YTHDF1 enhances translation of its targets by interacting with initiation factors and facilitating ribosome loading[58], and YTHDF2 promotes mRNA degradation by localizing m^{6}A-modified mRNA to processing bodies[55]. The two proteins share a set of common target mRNAs, to which YTHDF1 binds earlier during the mRNA life cycle than YTHDF2[58]. Together, YTHDF1 and YTHDF2 may contribute to generating a burst of protein synthesis, with increased protein production from temporally controlled mRNAs[58]. Further studies have identified two downstream mRNA decay pathways linked to YTHDF2. The *N*-terminal 100 amino acids (aa) residues of YTHDF2 binds to a small adaptor protein HRSP12 which recruits POP1, a component of the endoribonuclease complex RNase P/MRP[57]. A different region of YTHDF2, 101-200 aa, can recruit CCR4-NOT de-adenylation complex[56] (**Figure 2.1B**).

YTHDF3 is also predominately cytoplasmic, and its structural similarity with YTHDF1 and YTHDF2 indicates potential crosstalk among them. We aim to study YTHDF3 for a more complete picture of m^{6}A-mediated mRNA function and metabolism in the cytosol.

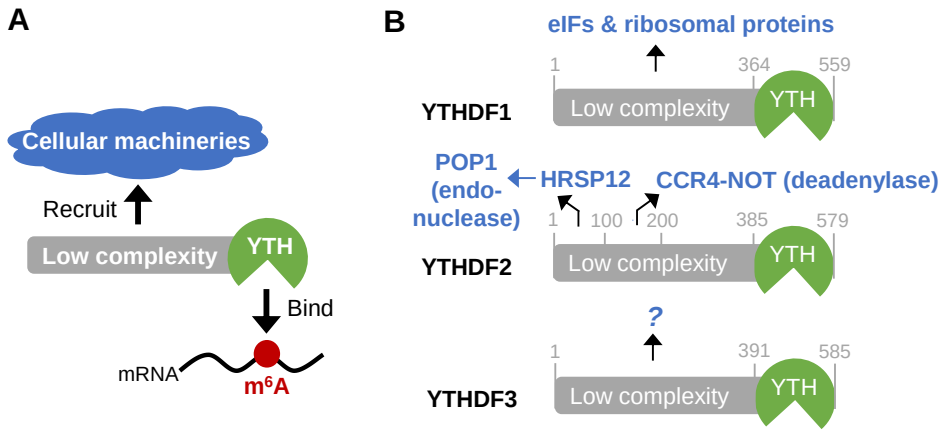

Figure 2.1 The domain structure of YTH domain family (YTHDF) proteins

(**A**) Modularized functions of YTHDF proteins. The *N*-terminal low complexity domain could recruit cellular machineries, while the *C*-terminal YTH domain specifically binds m^{6}A. (**B**) Illustration of functional proteins partners recruited by YTHDF1 and YTHDF2.

2.2 Results

2.2.1 YTHDF3 binds m^{6}A inside cells

Previous studies have confirmed that YTHDF3 selectively binds m^{6}A-modified mRNA *in vitro*[55]. To examine this selective binding activity inside cells, we performed RNA immunoprecipitation-based experiments using cell lysates.

We started by quantifying the m^{6}A level of mRNA purified from the YTHDF3-RNA complex using liquid chromatography-tandem mass spectrometry (LC-MS/MS). Compared to input mRNA, we detected over 2-fold enrichment of m^{6}A in the YTHDF3-bound portion and a slight m^{6}A depletion in the flow-through portion (**Figure 2.2A**), indicating that YTHDF3 binds to m^{6}A-modifed mRNA.

Next, we mapped binding sites of YTHDF3 in HeLa cells using photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP). Three biological replicates of PAR-CLIP identified 11,750 common peaks as YTHDF3 binding sites and 4,459 mRNAs as target transcripts (**Figure 2.2B**). A "GGACH" motif, similar to the m^{6}A consensus motif in the transcriptome, was identified in over 65% of the PAR-CLIP peaks (**Figure 2.2C**). The majority of YTHDF3 binding sites locate at the 3'UTR and the CDS of transcripts and cluster around stop codon (**Figure 2.2D-E**), coinciding with the distribution pattern of m^{6}A along the length of mRNA. Moreover, 59% of the PAR-CLIP peaks overlapped with m^{6}A peaks in HeLa cells profiled using specific anti-m^{6}A antibody (**Figure 2.2F**). Therefore, we conclude that YTHDF3 recognizes the m^{6}A modification on transcripts inside cells.

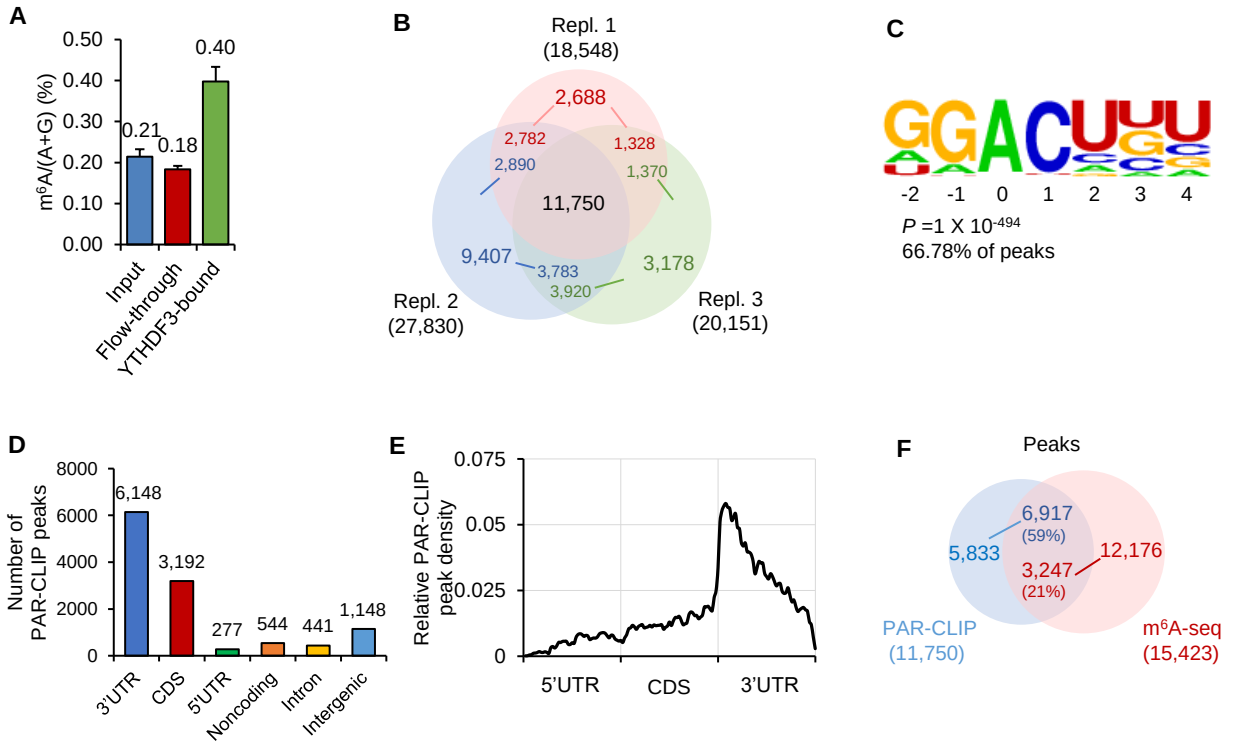

Figure 2.2 YTHDF3 selectively binds m^{6}A-modified transcripts inside cells

(**A**) LC-MS/MS quantification showing m^{6}A is enriched in RNAs pulled down with YTHDF3 from the cell lysate. Error bars, mean ± s.d., *n* = 2, technical replicates. (**B**) Overlap of YTHDF3

(**Figure 2.2, continued**) PAR-CLIP peaks in biological triplicates. (**C**) YTHDF3-binding motif identified by HOMER from the sequences of common PAR-CLIP peaks. (**D**) Region distribution of common PAR-CLIP peaks. (**E**) Distribution of common PAR-CLIP peaks along the length of mRNA. (**F**) Overlap of common PAR-CLIP peaks and m^6A-seq peaks from HeLa cells. PAR-CLIP peak numbers, blue; m^6A peak numbers, red.

2.2.2 YTHDF3 shares target transcripts with YTHDF1 and YTHDF2

To obtain high-confidence YTHDF3 target transcripts for further functional characterizations, we also sequenced the RNA that co-immunoprecipitated (RIP-seq) with YTHDF3. In two biological replicates of RIP-seq, 2,264 different transcripts were identified with over 2-fold enrichment. 1,239 of them overlap with transcripts identified with PAR-CLIP (**Figure 2.3A**) and were defined as high-confidence targets for YTHDF3 ("CLIP+IP"). Among these targets, genes related to "regulation of transcription" are most enriched (**Figure 2.3B**). Comparing with the published target transcript lists of YTHDF1 and YTHDF2[55,58], we found that YTHDFs share a large number of common targets: 58% of YTHDF3 targets are also recognized by YTHDF1, and 60% by YTHDF2 (**Figure 2.3C**). These results suggest a potential coordinated action among YTHDF proteins on common target transcripts.

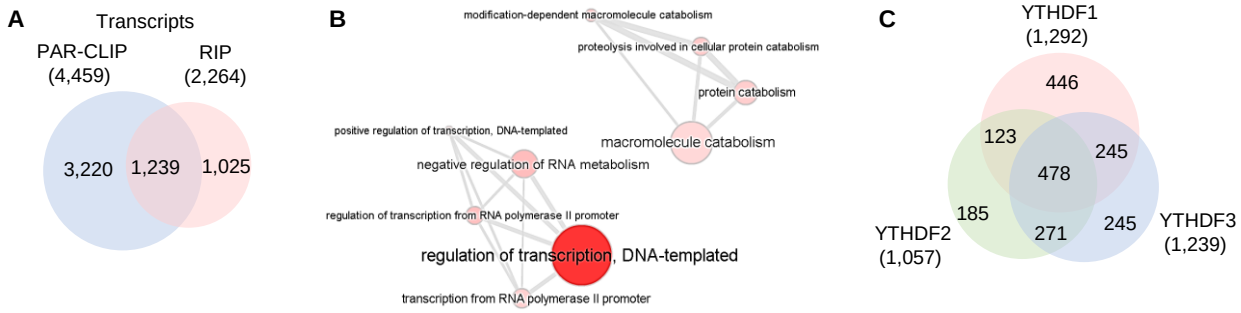

Figure 2.3 High-confidence target transcripts of YTHDF3

(**A**) Overlap of YTHDF3 target transcripts identified in the PAR-CLIP triplicates and the RIP replicates. (**B**) Gene Ontology (GO) analysis of top 500 target genes shared in PAR-CLIP and RIP-seq of YTHDF3 with DAVID 6.7. GO terms with $p < 0.001$ were visualized with REVIGO. (**C**) Overlap of target transcripts among YTHDF1-3.

2.2.3 Effects of YTHDF3 on translation of target transcripts

2.2.3.1 YTHDF3 facilitates translation promotion

We next monitored mRNA targets of YTHDF3 upon depletion of YTHDF3 or the m^{6}A methyltransferase components in HeLa cells, respectively. To evaluate whether YTHDF3 is involved in translation, we performed ribosome profiling in HeLa cells transfected with either siYTHDF3 or control siRNAs. We investigated the ribosome density of each transcript by sequencing in parallel input mRNA and the ribosome footprint on mRNAs in order to estimate translation efficiency. Considering the established function of YTHDF1 in promoting translation[58], we catalogued the evaluated transcripts into three groups: non-targets, YTHDF3 targets (YTHDF3 CLIP+IP), and YTHDF1 targets that are not YTHDF3 targets (YTHDF1 unique).

Compared to non-targets, a noticeable decrease of translation efficiency was observed in siYTHDF3 cells for both YTHDF3 targets ($p < 2\times10^{-8}$, Mann-Whitney *U* test) and YTHDF1 unique targets ($p < 3\times10^{-7}$, Mann-Whitney *U* test) (**Figure 2.4A-F**). This result suggests that, YTHDF3 facilitates the translation of its RNA targets, and such effect interplays with the role of YTHDF1. Of interest, the accumulation of the input RNA (**Figure 2.4C&F**) in the absence of YTHDF3 largely contributes to the decrease of target translation efficiency. This may indicate a crosstalk between YTHDF3 and YTHDF2—the reader mediating RNA decay (discussed below).

We next tested whether this observation depends on m^{6}A. Based on the published ribosome profiling data in HeLa cells depleted of the m^{6}A methyltransferase complex component METTL3[35], we observed that upon a global loss of m^{6}A, YTHDF3 targets experienced a significant decrease in translation efficiency compared to non-targets ($p < 4\times10^{-4}$, Mann-Whitney *U* test, **Figure 2.4G-I**), supporting that YTHDF3 functions in an m^{6}A-dependent manner.

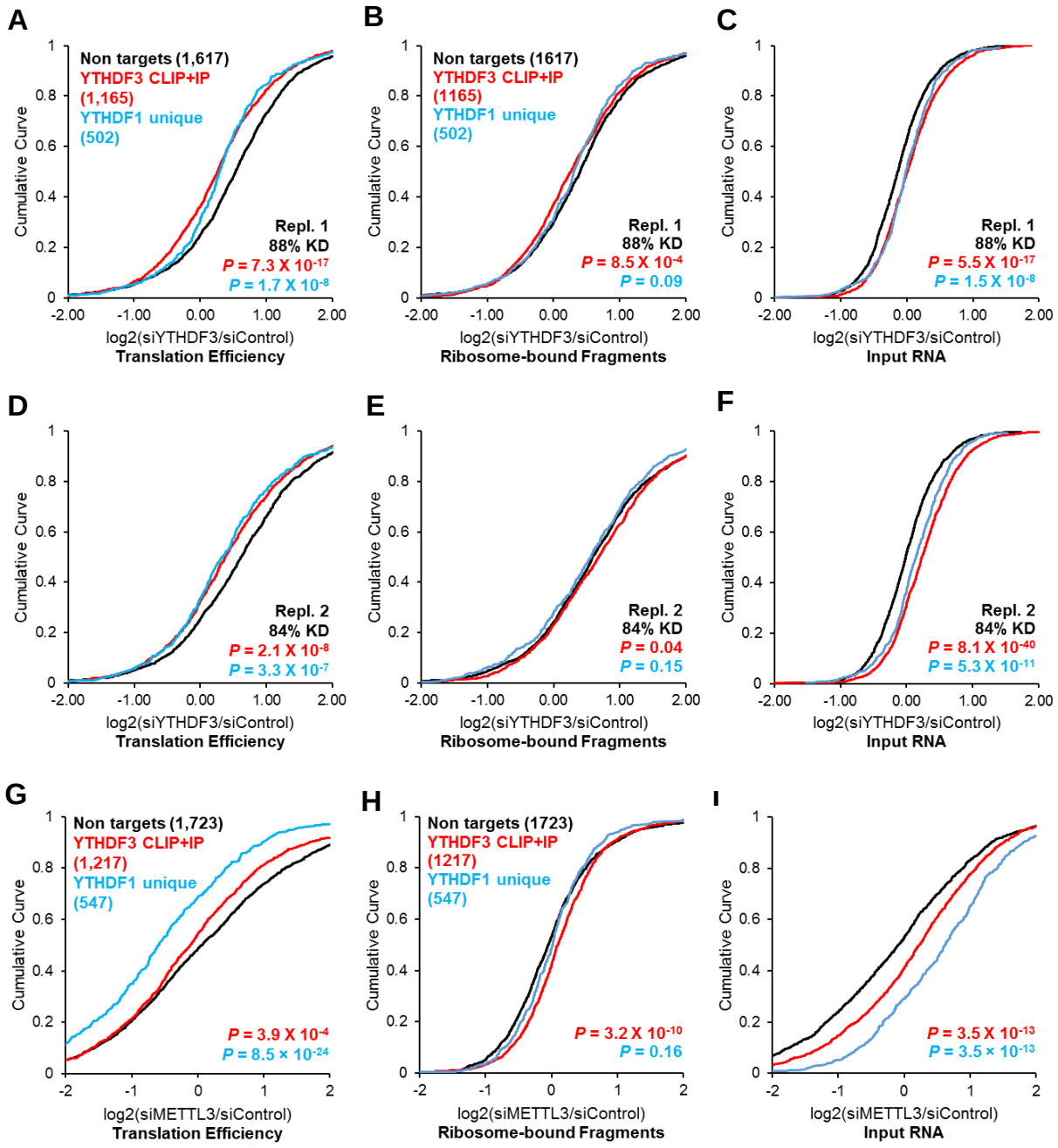

Figure 2.4 Translation regulation effect of YTHDF3 shown in ribosome profiling

Cumulative distributions of *log*$_2$-fold changes of translation efficiency (**A**, **D**, **G**), ribosome-bound fragments (**B**, **E**, **H**), and input RNA (**C**, **F**, **I**), between siControl and siYTHDF3-replicate 1 (**A**-**C**), siYTHDF3-replicate 2 (**D**-**F**), and siMETTL3 (**G**-**I**). Non-targets, black; YTHDF3 CLIP+IP targets, red; YTHDF1 unique targets, blue. Number of genes in each group was indicated in parentheses. *P* values were calculated from a two-sided Mann-Whitney *U* test compared to non-targets.

We then performed polysome profiling of YTHDF3-deficient cells to verify the observed trend. A different YTHDF3 siRNA oligo was used to eliminate any potential siRNA off-targeting. The three groups of mRNA-protein particles (mRNPs) were separated (**Figure 2.5A**): non-ribosome (mRNPs without any ribosome), 40S-80S (mRNPs associated with ribosome but not being translated), and polysome (actively translating pool). Changes in the distribution of specific mRNAs among non-ribosome and polysome fractions were quantified via RT-qPCR. Noticeable decreases in the polysome fraction were observed for the selected YTHDF3 target (*APC* mRNA) and YTHDF1 unique targets (*TSC1* mRNA and *DST* mRNA), but not for a non-target (*RPL30* mRNA) (**Figure 2.5B**). This result validates the function of YTHDF3 in promoting translation of its mRNA targets and facilitating the function of YTHDF1.

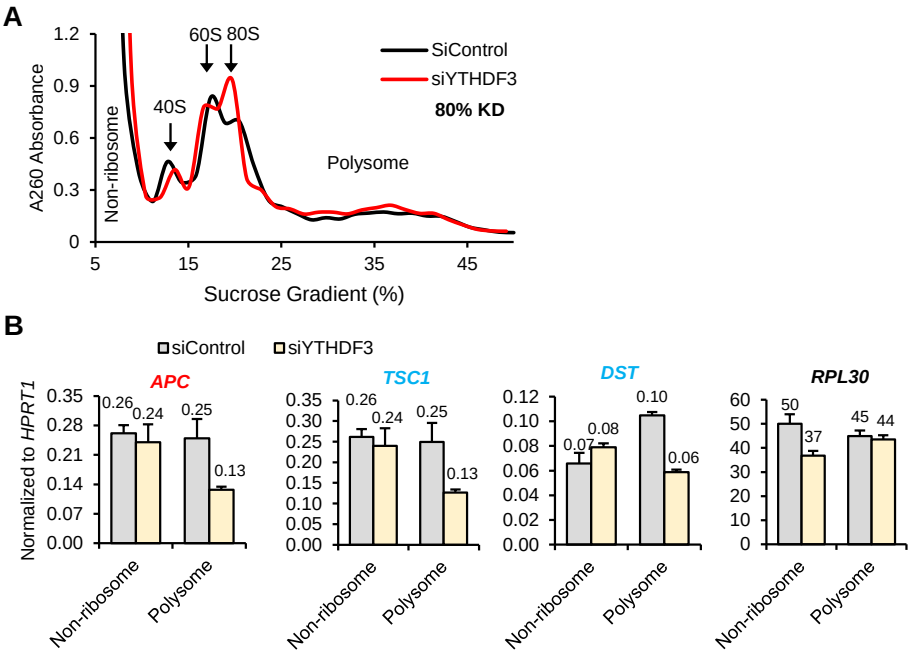

Figure 2.5 Polysome profiling validation of decreased target translation

(**A**) Polysome profiling of HeLa cells treated with siControl and siYTHDF3. (**B**) Redistribution of representative targets in non-ribosome and polysome portions of mRNPs upon depletion of YTHDF3 measured by RT-qPCR. *APC*, a YTHDF3 target; *TSC1* and *DST*, YTHDF1 unique targets; and *RPL30*, a non-target. Error bars, mean ± s.d., $n = 2$, technical replicates.

2.2.3.2 YTHDF3 and YTHDF1 recruit similar protein partners

There are two mechanisms through which YTHDF3 could promote translation: it may directly recruit translation machineries such as initiation factors (eIFs), or it could interact with YTHDF1 and affect the function of YTHDF1.

To investigate these possibilities, we studied protein partners of YTHDF3 using a HeLa stable cell line expressing YTHDF3 with a Flag-HA tandem epitope tag at the N-terminus. After polysome profiling of the stable cell line, western blotting of each fraction showed that YTHDF3 is distributed mainly in 40S mRNPs, suggesting YTHDF3 might regulate translation initiation (**Figure 2.6A**). However, neither eIF3A nor eIF3B was detected by western blotting in proteins specifically co-immunoprecipitated with YTHDF3 (**Figure 2.6B**). Unlike YTHDF1[58], which interacts with eIF3, no direct interaction may exist between YTHDF3 and eIF3 components.

We next compared protein partners of YTHDF3 and YTHDF1. Components of the YTHDF3-containing protein complex from tandem-affinity purification were subject to protein mass spectrometry analysis. Protein partners reported for YTHDF1 using the same method were used for comparison[58]. YTHDF3 and YTHDF1 shared 86 out of 154 proteins identified to interact with YTHDF3, with 39 of the shared partners being 40S and 60S ribosomal proteins (**Figure 2.6C**). These results suggest that YTHDF3 promotes translation through interacting with YTHDF1.

In addition, immunoprecipitation and western blotting identified additional RNA binding proteins as shared protein partners of YTHDF1 and YTHDF3 (**Figure 2.6D**). They both interact with YBX1, which is known to impact mRNA splicing, localization, and translation[94]; and eIF4A3, a component of exon junction complex that affects mRNA translation before being displaced by the first round of translating ribosomes[95]. These proteins may shuttle mRNAs to YTHDF3 and YTHDF1 for accelerated translation.

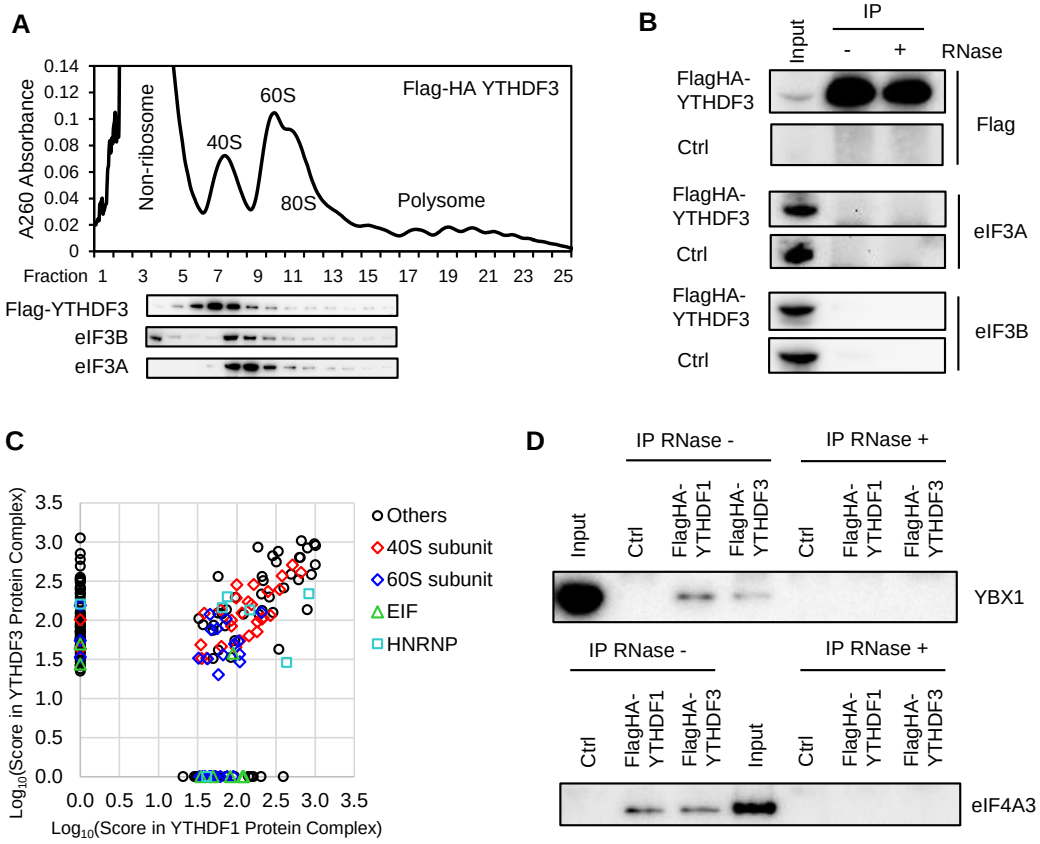

Figure 2.6 Protein partners of YTHDF3

(**A**) Polysome profiling of HeLa cells stably overexpressing Flag-HA tagged YTHDF3, and western blotting of Flag, eIF3A, and eIF3B in each fraction. (**B**) Western blotting showed no visible presence of eIF3A nor eIF3B in the proteins co-immunoprecipitated (IP) with YTHDF3. (**C**) Comparison of components of protein complex co-immunoprecipitated with Flag-HA-YTHDF3 and Flag-HA-YTHDF1. (**D**) Western blotting showing YBX1 and eIF4A3 were co-immunoprecipitated (IP) with both YTHDF1 and YTHDF3 in an RNA-dependent manner.

2.2.3.3 Additional factors exist to modulate translational roles of YTHDF3

It should be noted that, while the majority of YTHDF3 targets shows decreased translation upon YTHDF3 depletion, there is still a group of transcripts behaving differently (**Figure 2.4A&D**), indicating secondary effects and/or that there are additional factors modulating the translational roles of YTHDF3. To test this, we compared the characteristics of transcripts that showing consistent increased translation in the absence of YTHDF3 in the two replicates of

ribosome profiling with those showing consistent decreased translation. 319 transcripts showed more positive change in translation efficiency (TE) than the median of non-targets (TE > median) when YTHDF3 expression was disturbed; while 614 transcripts showed more negative change in translation efficiency (TE < median).

The transcripts in "TE > median" group tend to code proteins with different functions and cellular colocalization compared to those with reduced translation (**Figure 2.7A**). It is possible that YTHDF3 works as a chaperon protein and these two groups of transcripts are processed separately through distinct protein partners of YTHDF3 in specific biological contexts. In addition, the "TE > median" transcripts have much lower base translation activity as measured in the knockdown control sample (**Figure 2.7B**), and they are stronger YTHDF3 targets in terms of RIP enrichment (**Figure 2.7C**). In fact, the top 40% YTHDF3 RIP targets were shown to be less actively translated in general compared to the low 20% ones (**Figure 2.7D**). These differences suggest that YTHDF3 may serve as an RNA storage site, forming an untranslatable mRNA-protein complexes (mRNPs) preferentially with its high affinity targets.

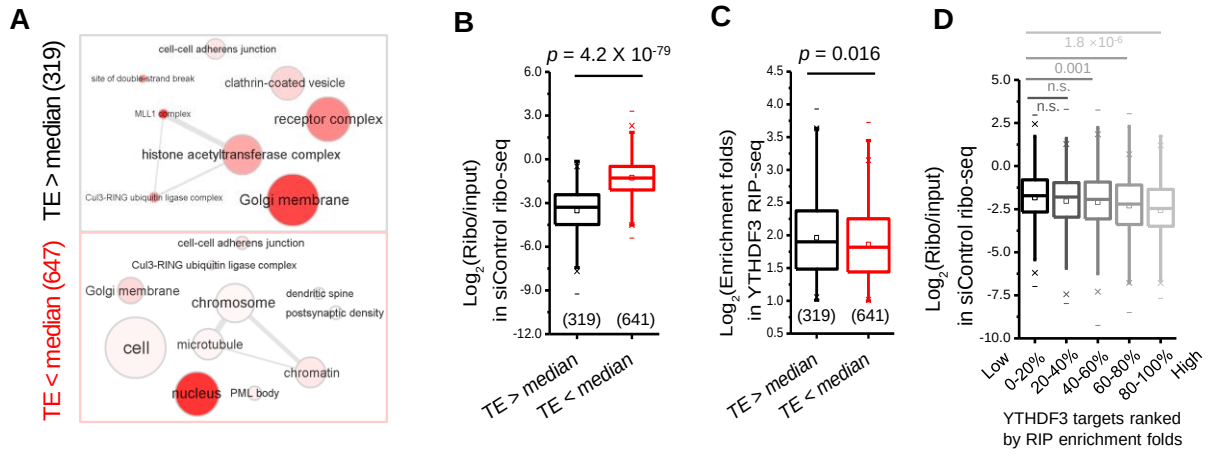

Figure 2.7 Analyses of factors contributing to the differential TE outcome

(**A**) Gene Ontology (GO) analysis with DAVID 6.7 of the group of genes with increased translation efficiency (TE > median) and that with decreased translation efficiency (TE < median) in absence of YTHDF3. GO terms with $p < 0.05$ were visualized with REVIGO. (**B**-**C**) Box plots of translation

(**Figure 2.7, continued**) efficiency (Log_2(Ribo/input)) in siControl samples (**B**) and YTHDF3 RIP enrichment folds (**C**) of these two groups of genes. (**D**) Box plot of the correlation between YTHDF3 RIP enrichment folds and translation efficiency in siControl samples. Box, 25-75%; "–", max and min; "×", 1% and 99%; "□", median. *P* values, a two-sided Mann-Whitney *U* test (**B-D**).

2.2.4 Crosstalk among YTHDF proteins: interconnected m^{6}A regulation in the cytosol

2.2.4.1 YTHDF3 modulates the functions of YTHDF1 and YTHDF2

We examined the crosstalk among YTHDFs based on the observation that YTHDFs share hundreds of common targets. The direct binding between YTHDFs was tested with an *in vitro* binding assay using purified GST tagged YTHDF1-3 and cell lysate containing Flag-HA tagged YTHDF1-3. Flag blotting after GST pull down showed that YTHDFs directly interacted with each other (**Figure 2.8A**).

We observed a 40% increase of the m^{6}A level in mRNAs when all of the YTHDFs were depleted (**Figure 2.8B**); the additional knockdown of YTHDF3 further enhances the m^{6}A accumulation comparing to the double knockdown of YTHDF1 and YTHDF2. These results indicate that all three YTHDFs contribute collectively to accelerating the metabolism of m^{6}A-modified mRNAs in the cytosol.

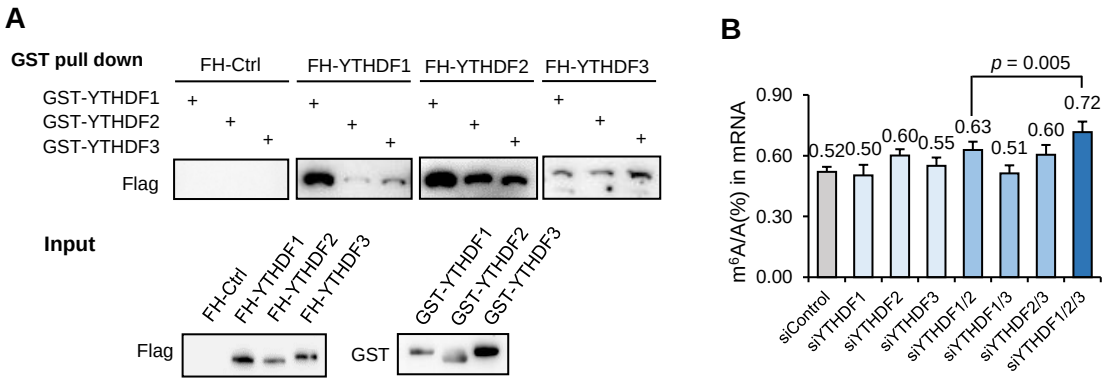

Figure 2.8 YTHDF proteins interact with each other and co-regulate mRNA m^{6}A level

(**A**) *In vitro* binding assay showing that YTHDFs co-immunoprecipitated with one another. Purified recombinant GST tagged YTHDF1-3 were incubated with cell lysates from Flag-HA (FH) tagged YTHDF1-3 as indicated. FH tagged proteins were western blotted in the eluent after GST

(**Figure 2.8, continued**) affinity pull down. (**B**) LC-MS/MS quantification of m^6A levels of HeLa cells treated with siControl, siYTHDF1, siYTHDF2, siYTHDF3, and combinations of those oligoes. Error bars, mean ± s.d., n = 4 (two biological replicates × two technical replicates).

We then utilized tethering reporter assays to evaluate the function of YTHDF proteins on their common targets. In the tethering assays, the 3'UTR region of the firefly luciferase (*F-Luc*) was modified with hairpin RNA elements such as *BoxB* and *MS2* sequences. The *C*-terminal YTH domain is replaced with an affinity tag that specifically recognizes one of those RNA elements, such as *BoxB* binding protein λ peptide and *MS2* binding protein MS2; therefore it mimics the interaction between YTH domain and the 3'UTR m^6A methylation on target transcripts. The renilla luciferase (*R-Luc*) was co-transfected as an internal control. (**Figure 2.9A&C**).

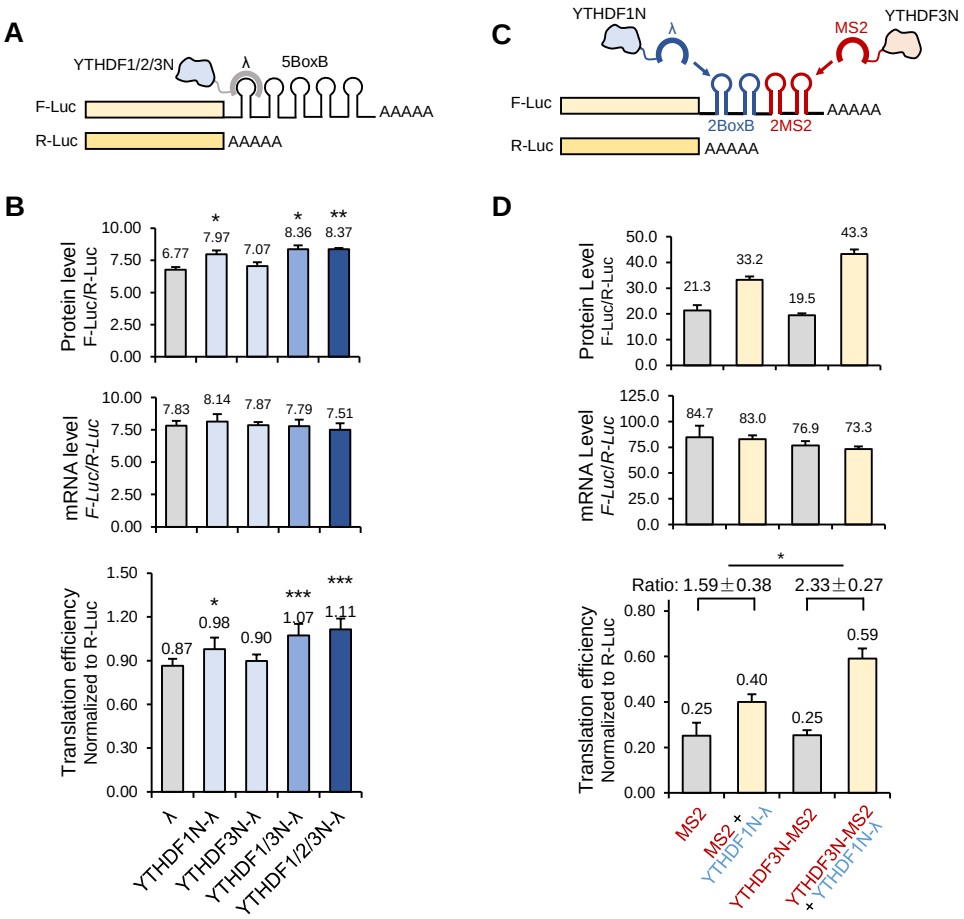

Figure 2.9 Luciferase reporter assays showing cooperative functions of YTHDF proteins

(**Figure 2.9, continued**) (**A**, **C**) Constructs of the single (**A**) and double (**C**) tethering assays. The *C*-terminal YTH domain of YTHDF proteins was replaced with affinity tags towards RNA elements in the 3'UTR of firefly luciferase (F-Luc). Renilla luciferase (R-Luc) was co-transfected as an internal control for normalization. (**B**, **D**) Normalized protein level, mRNA level, and translation efficiency of F-Luc reporters under different tethering conditions in single (**B**) and double (**D**) tethering assays.

In the single tethering assay, YTHDF1 tethering was sufficient to promote reporter translation, and the promotion effect could be enhanced by the presence of YTHDF3 (**Figure 2.9A-B**). Tethering of YTHDF3 itself did not result in increased reporter production (**Figure 2.9B**), consistent with the lack of direct interactions between YTHDF3 and eIFs (**Figure 2.6B**). To further confirm this observation, we designed a double tethering assay with YTHDF1N-λ and YTHDF3N-MS2 to reflect the bindings of YTHDF1 and YTHDF3 to multiple m^6A sites on a same mRNA transcript (**Figure 2.9C**). The presence of YTHDF3N-MS2 magnified the translational promotion effect of YTHDF1N-λ on F-Luc (**Figure 2.9D**). The protein level of F-Luc was largely increased while its mRNA level remained unchanged (**Figure 2.9D**). These results further support the role of YTHDF3 in cooperatively promoting translation with YTHDF1.

2.2.4.2 YTHDF3 modulates the RNA binding activities of YTHDF1 and YTHDF2

We then explored the molecular mechanisms underlying the integrative functions of YTHDF proteins. First, we observed an increase in YTHDF1 and YTHDF2 protein level when YTHDF3 was depleted (**Figure 2.10A**). This compensation effect indicates their relevance in functions. Second, we checked how depletion of YTHDF3 affects target binding activities of YTHDF1 and YTHDF2. PAR-CLIP was performed for YTHDF1 and YTHDF2 in cells depleted of YTHDF3, and RNAs pulled down were semi-quantified with 5'-^{32}P labelling. Knockdown of YTHDF3 led to increased amount of RNA pulled down with YTHDF1 and YTHDF2, and the extent of the increase correlated with the YTHDF3 knockdown efficiency (**Figure 2.10B**). To

investigate how YTHDF3 affects RNA binding specificity of YTHDF1 and YTHDF2, we performed RIP-seq for YTHDF1 and YTHDF2 in control cells and cells without YTHDF3. In control samples, YTHDF1/3 common targets showed greater enrichment than YTHDF1 unique targets, as was the case for YTHDF2 (**Figure 2.10C**, left). In the absence of YTHDF3, the RIP enrichment of YTHDF1 or YTHDF2 targets was reduced significantly (**Figure 2.10C**, middle and left). RIP followed by RT-qPCR of a number of target transcripts confirmed that, when expression of YTHDF3 was disturbed, YTHDF1 and YTHDF2 showed a decreased binding specificity, with reduced binding towards their target transcripts but increased binding towards non-target transcripts (**Figure 2.10D**), indicating that the binding specificity of YTHDF1 and YTHDF2 towards their target mRNAs could be modulated by YTHDF3.

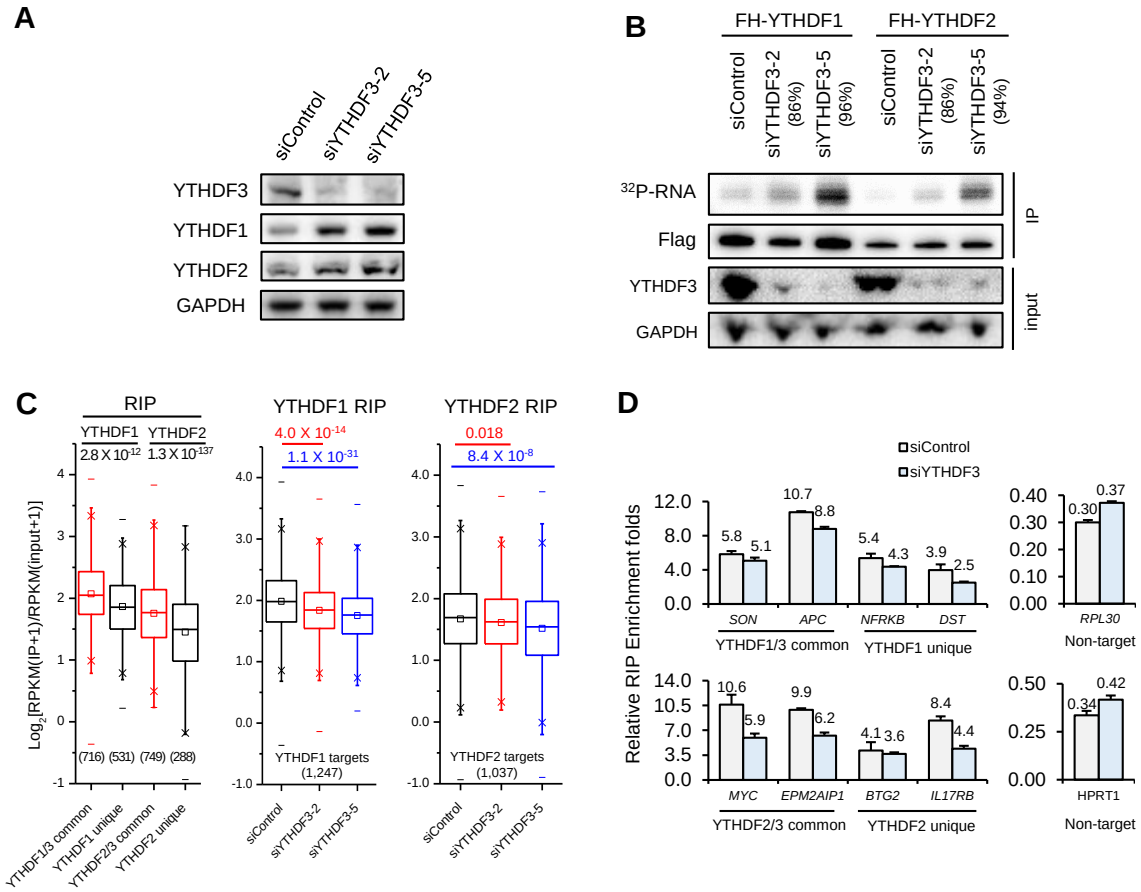

Figure 2.10 YTHDF3 regulates binding capacity and specificity of YTHDF1-2

(**Figure 2.10, continued**) (**A**) Western blotting showing that knockdown of YTHDF3 leads to upregulated protein levels of YTHDF1 and YTHDF2 in wild-type HeLa cells. (**B**) Total RNAs bound by YTHDF1 and YTHDF2 quantified with PAR-CLIP followed by 5'-^{32}P labelling in the control HeLa cells and cells depleted of YTHDF3 using two different siYTHDF3 oligos. The knockdown efficiency was labelled. (**C**) Box plot of RIP enrichment of different groups of YTHDF targets in siControl samples, and that of YTHDF1 targets or of YTHDF2 targets in siControl and siYTHDF3 samples. Box, 25-75%; "–", max and min; "×", 1% and 99%; "□", median. (**D**) Representative target transcripts co-immunoprecipitated with Flag-HA-YTHDF1 or Flag-HA-YTHDF2 upon YTHDF3 depletion were quantified with RT-qPCR and normalized to their amounts in input RNA. Error bars, mean ± s.d., n = 2, technical replicates.

2.2.4.3 Temporal order in the action of YTHDF proteins

By metabolically labelling nascent RNA with nucleoside-analogues such as 4-thio-uridine (4SU) and 5-ethynyl-uridine (EU), the occupancy of YTHDFs on nascent RNA could be quantified over time (**Figure 2.11A&C**).

It is observed that both YTHDF1 and YTHDF3 bound mRNA targets prior to YTHDF2, as quantified by the 4SU incorporation percentage in the YTHDF-mRNA complex 2 hours and 4 hours after 4SU labeling (**Figure 2.11B**). Besides 4SU labelling, EU-mediated Click chemistry allows imaging and purification of nascent RNA[96]. We co-stained YTHDF proteins and the EU-labelled RNA that clicked with a fluorophore. YTHDF3 appeared to co-localize better than YTHDF1 with nascent RNAs (**Figure 2.11D**). In the other assay, we first purified YTHDF1- and YTHDF3-bound RNAs and then captured the nascent RNAs therein by biotinylating EU-labelled RNAs and streptavidin pull down. The percentage of nascent RNAs bound by YTHDFs were quantified by RT-qPCR of representative common targets. YTHDF3 bound to a higher percentage of the EU-labelled transcripts than YTHDF1 at both time points (**Figure 2.11E**), suggesting that YTHDF3 binds more nascent RNA than YTHDF1.

Therefore, the m^{6}A-modified transcripts might engage YTHDF3 and then YTHDF1 for translation promotion before being decayed through the YTHDF2-mediated pathway in the

cytoplasm. While both YTHDF1 and YTHDF3 facilitate translation of their target mRNAs, they also collectively affect partitioning methylated transcripts to YTHDF2 for accelerated decay.

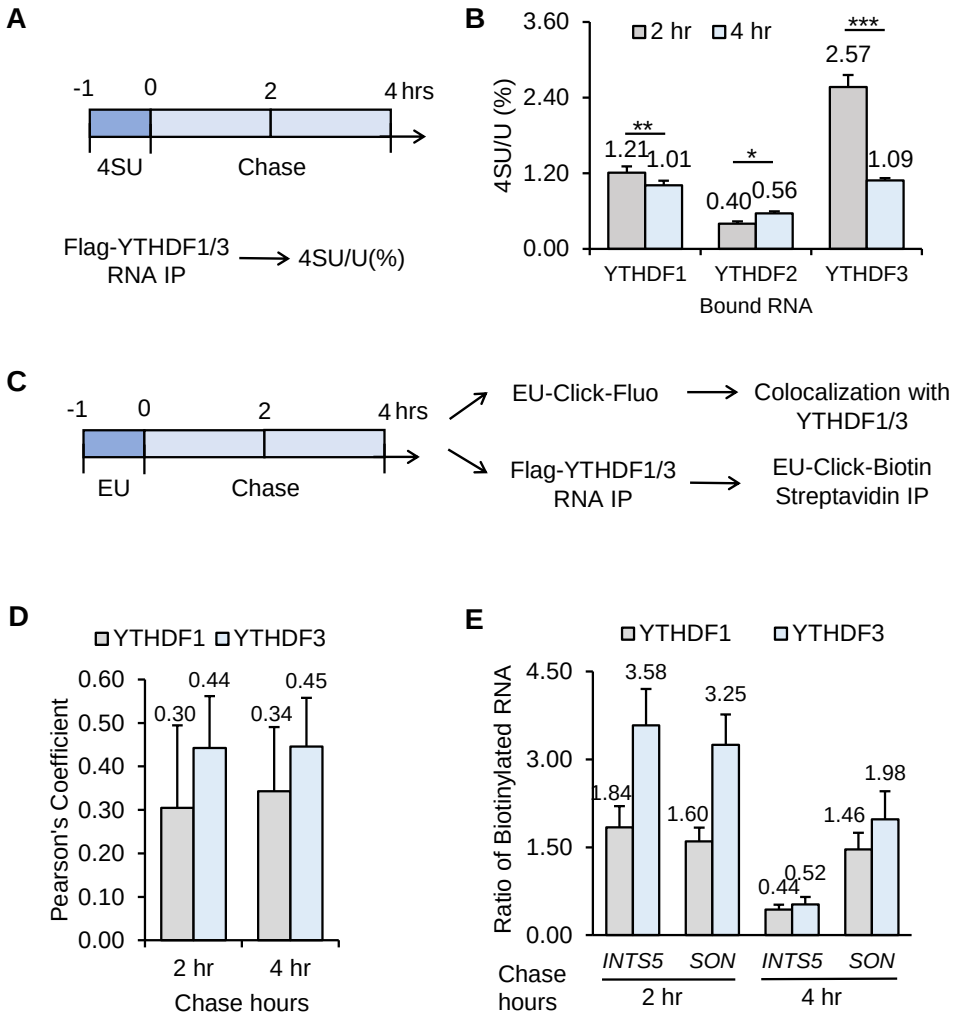

Figure 2.11 Temporal order of YTHDFs' interaction with m^{6}A-modified transcripts

(**A**) Scheme of experimental designs for 4SU (4-thio-uridine) pulse and chase. (**B**) LC-MS/MS quantification of 4SU level in mRNAs pulled down with YTHDF1-3. Error bars, mean ± s.d., n = 3~4. P values, paired two-sided Student's t-test. *, $p < 0.05$; **, $p < 0.005$; ***, $p < 0.0005$. (**C**) Scheme of experimental designs for nascent RNA labelling with EU (5-ethynyl-uridine). HeLa cells were treated with EU for 1 hour and analyzed 2-hour and 4-hour post labelling. (**D**) Pearson's coefficient quantification of co-localization between EU-labelled RNA and YTHDF proteins. Error bars, mean ± s.d., n = 8. (**E**) Percentage of nascent transcripts in YTHDF-mRNA complex, quantified by RT-qPCR for representative YTHDF common targets. Error bars, mean ± s.d., n = 2, technical replicates.

2.3 Conclusion and Discussion

In this work, we present a systematic characterization of YTHDF3 in HeLa cells. Our studies on YTHDF3 binding sites and target mRNAs supported its specific binding to m^6A modifications on cellular mRNAs (**Figure 2.2**) and revealed hundreds of common targets shared by YTHDFs (**Figure 2.3**). YTHDF3 facilitated YTHDF1 in promoting translation in the tethering assay (**Figure 2.9**). The knockdown of YTHDF3 resulted in: (i) reduced translation efficiency of mRNA targets of both YTHDF3 and YTHDF1 (**Figure2.4&5**); (ii) decreased RNA-binding specificity of both YTHDF1 and YTHDF2 (**Figure 2.10**); and (iii) further increased cellular m^6A level in cells with depletion of both YTHDF1 and YTHDF2 (**Figure 2.8B**). Nascent RNA labelling and time-course assays revealed that YTHDF3 functions during the early time point of RNA life cycle comparing to YTHDF1 or YTHDF2 (**Figure 2.11**).

These results led us to propose an interconnected and dynamic model for the regulatory functions of YTHDFs in the cytosol (**Figure 2.12**): 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 activities of the other two, impacting target translation efficiency and stability. YTHDF3 may add robustness to the network. Its presence and potential post-translational modifications could provide additional layers of regulation of methylated target mRNAs. Besides the reversible, dynamic, and non-stoichiometric nature of the m^6A modification itself, a divergent expression pattern of YTHDF

proteins in different cell types or developmental stages could provide temporal-spatial control of protein production from m^{6}A-modified transcripts.

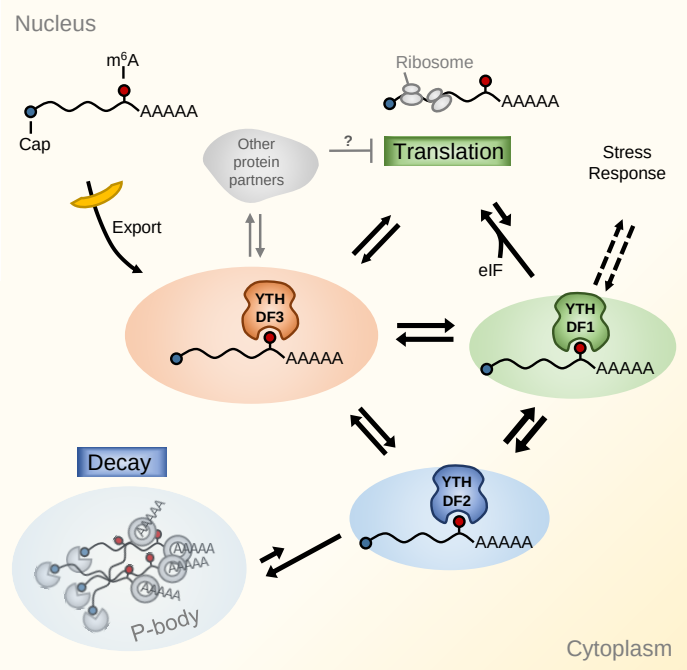

Figure 2.12 A proposed model of an integrated partition network for m^{6}A-methylated transcripts mediated by YTHDFs in the cytosol

While YTHDF1 functions in translation regulation 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 (grey) to negatively impact translation.

Our conclusions were supported by another study published around the same time, which reported that YTHDF3 promotes initial steps of translation by interacting with ribosomal proteins[60]. Besides, there are several follow-up studies demonstrating the roles of YTHDF3 beyond canonical cap-dependent translation. Yang *et. al.* reported that m^{6}A methylation also exists prevalently on circular RNA species, and YTHDF3 is essential for protein translation from the m^{6}A-modified circular RNA likely through recruiting the translation initiation factor eIF4G2[61]. Coots *et. al.*

reported that, for a great deal of transcripts, m^6A-promoted cap-independent translation co-exists in physiological conditions with eIF4F-mediated cap-dependent translation initiation; YTHDF3, but not YTHDF1 nor YTHDF2, contributes to the eIF4F-independent translation process[97].

There are cases in which YTHDF3 is distinctly regulated among YTHDFs. For instance, it has been described that YTHDF3 is regulated with insulin receptors coherently in aging processes[98]. Recently, it is found that YTHDF3, rather than YTHDF1 or YTHDF2, plays a negative role in antiviral activities by promoting the translation of a transcription corepressor, forkhead box protein O3 (FOXO3), of interferon genes[99]. In collaboration with Prof. Suyun Huang from the University of Texas, we found a high level of YTHDF3 is positively correlated with breast cancer brain metastasis, and YTHDF3 promotes translation of metastasis-related genes (manuscript submitted). Those models could be probed in the future to further understanding functions of YTHDF3 in different biological processes or systems.

2.4 Methods

2.4.1 Antibodies

The antibodies used in this study for western blotting were listed below in the format of name (catalogue; supplier; dilution fold): Rabbit anti-YTHDF1 (17479-1-AP; Proteintech; 2,000). Rabbit anti-YTHDF2 (24744-1-AP; Proteintech; 2,000). Mouse anti-YTHDF3 (sc-377119; Santa Cruz Biotech; 500). Goat anti-eIF3B (sc-16377; Santa Cruz Biotech; 2,000). Rabbit anti-eIF4A3 (17504-1-AP; Proteintech; 2,000). Rabbit anti-eIF3A (3411; Cell Signaling Technology; 2,000). Rabbit anti-YBX1 (ab76149; Abcam; 2,000). Mouse anti-Flag HRP conjugate (A8592; Sigma; 5,000). Mouse anti-HA HRP (sc-7392 HRP; Santa Cruz Biotech; 2,000). Donkey anti-Rabbit IgG-HRP (sc-2313; Santa Cruz Biotech; 2,000). Donkey anti-Goat IgG-HRP (sc-2033; Santa Cruz

Biotech; 2,000). Goat anti-GAPDH HRP (A00192-100; GenScript; 5,000). Mouse anti-GST HRP (sc-138 HRP; Santa Cruz Biotech; 500).

2.4.2 Plasmid construction

Flag-tagged YTHDF3 was cloned from commercial cDNA clones (Open Biosystems) into vector pcDNA 3.0 (BamHI, XhoI; forward primer (with coding sequence for Flag-tag), CGTACGGATCCGATTACAAGGACGACGATGACAAGATGTCAGCCACTAGCG; reverse primer, CGTAGCTCGAGTCATTGTTTGTTTCTATTTCTCTCCCTAC).

Double tethering reporter pmirGlo-Ptight-2BoxB-2MS2: 2BoxB-2MS2 sequence was designed as the following: CTCGACTAAGTCCAACTACTAAACTGGGCCCTGAAGAA GGGCCCATATAGGGCCCTGAAGAAGGGCCCTAGCAAGTTCAAATAAGGCTAGTCCG TTATCAACTTGGCCAACATGAGGATCACCCATGTCTGCAGGTCGACTCTAGAAAACA TGAGGATCACCCATGTCTGCAGTATTCCCGGGTTCATTAGATCCTAA[100,101], synthesized, and PCR amplified and cloned to 3'UTR of firefly luciferase in pmirGlo-Ptight vector reported[55] (NheI, SbfI; forward primer, ATACGCTAGCCTCGACTAAGTCCAACTACTAAACTGGG; reverse primer, GTATCCTGCAGGTTAGGATCTAATGAACCCGGGAATACTG).

Tether effector: We first constructed pcDNA-Flag-MS2 vector by inserting Flag-MS2 protein coding sequence into pcDNA 3.0. MS2 coding sequence was amplified from pZS*12-MS2-GFP vector[102], by two rounds of PCR (BamHI, XbaI; forward primer 1 (with EcoRV and XhoI sites), GGAGGTTCGGGGATATCGGGCTCGAGCATGGCTTCTAACTTTACTCAG TTCGTTCTC; reverse primer 1, CATTCTAGACTAGTAGATGCCGGAGTTTGCTGC; forward primer 2 (with coding sequence for Flag tag), CATGGATCCATGGATTACAAG GACGACGATGACAAGGGAGGTTCGGGGATATCGGG; reverse primer 2, CATTCTAGAC TAGTAGATGCCGGAGTTT). The resulting plasmid was subjected to a second round of cloning

by inserting N-terminal of YTHDF3, resulting in plasmid pcDNA-YTHDF3N-MS2 (EcoRV, XhoI; forward primer, CATGATATCGATGTCAGCCACTAGCGTGGATC; reverse primer, CATCTC GAGCCTACACTAGAAGGTGAAGCACTGACAG). The pcDNA-Flag-YTHDF1N-λ construct was reported previously[58]. Finally, insert Flag-MS2, Flag-YTHDF3N-MS2, and Flag-YTHDF1N-λ were cloned to a modified pPB-CAG vector with MfeI and AgeI sites between BglII and XhoI sites, for optimized expression in HeLa cells (MfeI, AgeI; forward primer, ATCGCAATTGATGGATTACAAGGACGACGATGACAAG; Flag-MS2 and Flag-YTHDF3N-MS2: reserve primer, CATGACCGGTCTAGTAGATGCCGGAGTTTGCTG; Flag-YTHDF1N-λ: reverse primer, CATGACCGGT TCAGTTTGCAGCTTTCCATTGAGC).

Flag-HA tandem tagged YTHDF1-3 were cloned by insert their CDS into a modified pPB-CAG vector where there is a Flag-HA-tandem tag inserted upstream of the restriction enzyme site BglII. The vector was digested with BglII and MfeI, and the CDS of YTHDF1-3 were cloned with BglII and MfeI (YTHDF1: forward primer: CGTACAGATCTATGTCGGCCACCAGCG; reverse primer: CCATACTCGAGTCATTGTTTGTTTCGACTCTGCC; YTHDF2: forward primer: CGTACAGATCTATGTCGGCCAGCAGCC; reverse primer: CGATGCTCGAGTTA TTTCCCACGACCTTGACG; YTHDF3: forward primer: CGTACAGATCTATGTCAGCCAC TAGCGTG; reverse primer: CGTAGCTCGAGTTATTGTT TGTTTCTATTTCTCTCCCTAC).

2.4.3 Cell culture, siRNA knockdown, plasmid transfection, and RNA isolation

Human HeLa cell line used in this study was purchased from ATCC (CCL-2) and grown in DMEM (Gibco, 11995) 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) media supplemented with 10% FBS (Tet system approved, Clontech), 1% 100 × Pen/Strep (Gibco) and 200 μg/ml G418 (Sigma).

AllStars negative control siRNA from Qiagen (SI03650318) was used as control siRNA in knockdown experiments. YTHDF3 siRNAs were ordered from Qiagen. (siYTHDF3-2: Hs_YTHDF3_2 with the target sequence ATGGATTAAATCAGTATCTAA; siYTHDF3-5: Hs_YTHDF3_5 with the target sequence TAAGTCAAAGAAGACGTATTA). YTHDF1 and YTHDF2 siRNAs were reported previously[55,58]. Transfection was achieved by using Lipofectamine RNAiMAX (Invitrogen) for siRNAs, Lipofectamine 2000 (Invitrogen) for transfection of one or two plasmids, and Lipofectamine LTX Plus (Invitrogen) for transfection of multiple plasmids in tethering assay following the manufacturer's protocols.

Stable overexpression cell lines: Stable expression HeLa cell line with double-tagged YTHDF1, YTHDF2, and YTHDF3 individually (N terminal Flag and HA in tandem) was created by puromycin selection (2 μg/ml) with the modified pPB-CAG vector. The control cell line with expression of only tandem Flag and HA peptides was created similarly.

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 Dynabeads® mRNA DIRECT kit (Ambion) followed by further removal of remaining rRNA using RiboMinus Eukaryote Kit v2 (Ambion). mRNA concentration was measured by Qubit® RNA HS Assay Kit with Qubit® 2.0 fluorometer.

RNA isolation for RT–qPCR: we followed the instruction of Direct-zol™ RNA Miniprep kit (Zymo Research) or RNeasy Miniprep kit (Qiagen) both with an in-column DNase I digestion step. For RT-qPCR following polysome profiling, RNA was extracted according to the manual of TRizol® reagent (added 1:1), and contaminate DNA was removed using RNA clean and concentrator-25 (Zymo Research) with additional in-column DNase I digestion.

2.4.4 LC-MS/MS

50-100 ng of mRNA were digested by nuclease P1 (1 U, Wako) in 25 μl of buffer containing 20 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 (1 U, sigma). 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 retention time and the nucleoside to base ion mass transitions of 284 to 152 (G), 282.1 to 150.1 (m^6A), 268 to 136 (A), 245 to 113.1 (U), 261.3 to 129 (4SU, 4-thio-uridine), and 244 to 112 (C). Quantification was performed in comparison with the standard curve obtained from pure nucleoside standards running with the same batch of samples. The m^6A level was calculated as the ratio of m^6A to A or the ratio of m^6A to (A+G) based on the calibrated concentrations[49]. For 4SU quantification, 0.25 mM DTT was supplemented in the digestion buffer and each digestion step lasted for only 1 h in order to avoid oxidation of 4SU. The 4SU level was calculated as the ratio of 4SU to U based on the calibrated concentrations.

2.4.5 Tethering assay

The procedure was adapted from the one reported previously[55,58]. 50 ng reporter plasmid (pmirGlo-Ptight-5BoxB) and 450 ng effecter plasmid (λ, Y1N-λ, Y2N-λ, Y3N-λ, combination of equal amount of two of them, or combination of equal amount of the later three, in pcDNA3.0) in the single tethering assay, or 50 ng reporter plasmid (pmirGlo-Ptight-2BoxB-2MS2) and 250 ng of each effecter plasmid (Flag-MS2, Flag-YTHDF3N-MS2, and Flag-YTHDF1N-λ in the

combination indicated) in the double tethering assay, were used to transfect HeLa Tet-Off® Advanced Cell Line (613356, Clontech) in each well of six-well plate at 60~80% confluency under doxycycline (DOX, 100 ng/mL) inhibition. After 6 hours, transfection mixture was replaced with fresh media containing DOX (100 ng/mL). After another 18 hours, each well was trypsin-digested, extensively washed with PBS by suspending and spinning down for three times and re-seeded into 96-well plate (1:30) and 12-well plate (1:3) in media without DOX to induce transcription of Firefly luciferase (*F-Luc*). 4 hours after F-Luc induction, cells in the 96-well plate were assayed by Dual-Glo Luciferase Assay Systems (E2920, Promega). F-Luc activity was normalized by Renilla luciferase (R-Luc) to evaluate protein production from the reporter. At the same time, samples in the 12-well plate were processed to extract total RNA (DNase I digested) by TRIzol® reagent (Invitrogen), followed by RT–qPCR quantification. The amount of *F-Luc* mRNA was also normalized by that of *R-Luc* mRNA. Translation efficiency of *F-Luc* mRNA was calculated as the ratio of normalized F-Luc activity (protein level) to normalized *F-Luc* mRNA level.

2.4.6 PAR-CLIP

PAR-CLIP for sequencing: We followed the previously reported protocol[55] except starting with 150-200 million Flag-HA-tagged YTHDF3 stable overexpression HeLa cells instead of transient overexpression ones. YTHDF3-RNA complex was SDS-PAGE purified with a size-selection range from 60 kDa to 95 kDa, and the RNA fragments was extracted via ethanol precipitation after protease K digestion of the gel slices. The purified RNA pellet was dissolved in 12 µl RNase-free water, of which 6 µl was subjected to small RNA library preparation with NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (E7300S, NEB).

PAR-CLIP for quantification of protein-bound RNAs: 20 million Flag-HA-tagged YTHDF1 or YTHDF2 stable expression HeLa cells treated with either siControl or siYTHDF3

were subjected to the same PAR-CLIP procedure while using γ-^{32}P-ATP in T4 PNK 5' end-repairing. After stringent washing following radioactive-labelling, the samples were subjected to SDS-PAGE and the gel was exposed to a blanked phosphorimager screen overnight. The screen was then imaged with the Molecular Imager FX™ (Bio-Rad).

2.4.7 RIP-LC-MS/MS, RIP-seq and RIP-RT-qPCR

RIP-LC-MS/MS. The RNA immunoprecipitation (RIP) procedure was reported previously[55]. 60 million HeLa cells stably overexpressing Flag-HA tagged YTHDF3 were subjected to RIP procedure. Input, flow-through, and YTHDF3-bound RNA were purified with Trizol reagent. mRNAs of the three portions were further purified by depleting rRNA with RiboMinus Eukaryote Kit v2 (Ambion) followed by depleting tRNA with RNA Clean and Concentrator-5 (Zymo Research, 200 nt cutoff protocol). 50 ng purified mRNA of each sample were subjected to LC-MS/MS quantification of m^{6}A levels as described above.

RIP-seq. The procedure was reported previously[55]. 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). For RIP-seq of YTHDF1 and YTHDF2 with or without YTHDF3 depletion, 60 million HeLa cells stably expressing Flag-HA tagged YTHDF1 or YTHDF2 were subjected to Flag-IP. Input RNAs and IP RNAs were subjected to rRNA removal in parallel before library preparation.

RIP-RT-qPCR. 20 million HeLa cells stably overexpressing Flag-HA YTHDF1 or Flag-HA YTHDF2 were collected 48 hours after the treatment of siControl or siYTHDF3. The RIP procedure was the same as that in RIP-seq till before eluting. 0.5 ml Trizol® reagent was added to

anti-Flag M2 magnetic beads (Sigma) to extract YTHDF1- or YTHDF2-bound (IP) RNAs. Amount of non-target or target transcripts in both the input and IP RNAs were analyzed with RT-qPCR, and IP enrichment ratio of a transcript was calculated as the ratio of its amount in IP to that in the input yielded from same number of cells.

2.4.8 Ribosome and polysome profiling

Ribosome profiling. We followed the procedure reported previously[58]. Ribosome profiling was conducted using TruSeq Ribo Profile (Mammalian) Library Prep Kit (Illumina, RPHMR12126). Oligo siYTHDF3-5 was used in two biological replicates. The sequencing data obtained from ribosome profiling were denoted as "ribosome-bound fragments" and those from input as "input RNA". Translation efficiency was defined as the ratio of "ribosome-bound fragments" and "input RNA", which reflected the relative occupancy of 80S ribosome per mRNA species[103].

Polysome profiling was performed to verify genome-wide analysis results from ribosome profiling. We followed the procedure reported previously[104], with following modifications. (1) Before collection, cycloheximide (CHX) was added to the media at 100 μg/ml for 7 min. (2) The lysis buffer was formulated as 20 mM HEPES, pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, 100 μg/ml CHX, 1% Triton-X-100, freshly add 1:100 protease inhibitor (Roche), 40 U/ml SUPERasin (Ambion). Oligo siYTHDF3-2 was used in this experiment to perturb YTHDF3 expression. siControl, siYTHDF3-2, Flag-HA-tagged YTHDF3 stable expression HeLa cells, and stable expression control HeLa cells were analyzed with polysome profiling.

2.4.9 Sequencing data analysis

General pre-processing of reads. All samples were sequenced by illumine Hiseq2000 or Hiseq4000 with single end 50bp read length. For libraries that generated from small RNA (PAR-CLIP and ribosome profiling), the adapters were trimmed by using FASTX-Toolkit[105]. The deep

sequencing data were mapped to human genome version hg38 by Tophat version 2.0[106] without any gaps and allowed for at most two mismatches. RIP and Ribosome profiling were analyzed by DESeq[107] to generate RPKM (reads per kilobase, per million reads).

Data analysis for each experiment. (1) PAR-CLIP data were analyzed by PARalyzerv1.1[108] with default settings. Binding motif was analyzed by HOMER (v4.7)[109]; (2) for RIP, enrichment fold was calculated as $\log_2$(IP/input); (3) for ribosome profiling, only genes with RPKM >1 in input RNA sample and RPKM >0 in ribosome-bound fragments were used for analysis and the change fold was calculated as $\log_2$(siYTHDF3/siControl) or $\log_2$(siMETTL3/siControl).

Integrative data analysis and statistics. PAR-CLIP targets were defined as gene targets reproducible among three biological replicates (4,459). RIP targets (2,264) were genes with $\log_2$(IP/input) > 1 in both biological replicates. The overlap of PAR-CLIP and RIP targets were defined as high confidence targets (1,239). For **Figure 2.4**, the three groups of transcripts are defined as the following: (1) YTHDF3 targets: high confidence targets of YTHDF3; (2) YTHDF1 unique: high confidence targets of YTHDF1 reported previously[58] as "CLIP+IP", but not high confidence target of YTHDF3; (3) Non-target: non-target for a reader protein is defined as (a) complementary set of PAR-CLIP targets; (b) RIP enrichment fold <0; Non-target in the figure is defined as the intersection of the non-targets of YTHDF1 and that of YTHDF3. For the comparison of PAR-CLIP and m^{6}A peaks, at least 1 nt overlap was applied as the criteria of overlap peaks. Nonparametric Mann–Whitney *U*-test (Wilcoxon rank-sum test, two-sided, significance level = 0.05) was applied in ribosome profiling data analysis as previous reported[110]. Gene Ontology (GO) term analyses were performed by DAVID[111,112]. Top terms were then selected for visualizations by the interactive graph function of REVIGO[113].

Data accession. All the raw and processed data can be accessed under GSE86214 (https://www.ncbi.nlm.nih.gov/geo/). The analysis of **Figure 2.2F** and **Figure 2.4G-I** utilized data from GSE46705. The analysis of **Figure 2.3C** and **Figure 2.4A-F** utilized data from GSE49339 and GSE63591.

2.4.10 Tandem affinity purification of protein interactome

We followed the procedure reported previously[58,114]. The protein mass spectroscopy was performed by Institutes of Biomedical Sciences at Fudan University, Shanghai.

2.4.11 In vitro binding assay using purified protein

GST-tagged YTHDF1-3 were purified as described previously[55]. 60 million of Flag-HA YTHDF1-3 stable expression cell lines and control cell lines were lysed in 1 mL IPP buffer (150 mM NaCl, 0.2% NP-40, 10 mM Tris, pH 7.5, 1:100 protease inhibitor cocktail, 0.5 mM DTT). The cell lysate was aliquoted into three portions, and concentrated GST-tagged YTHDF1-3 were added to the final concentration of 2.5 μM. The mixture was rotated at 4°C for 2 hrs before 50 μL of pre-washed GST-affinity magnetic beads (Pierce) was added for another 2-hr rotation. The GST-affinity beads were then washed with 400 μL IPP buffer for four times. Proteins bound by GST-affinity beads were eluted with 25 μL 2x Laemmli Sample Buffer (Bio-Rad) at 95°C for 5 min and subjected to Flag blotting analysis.

2.4.12 Pulse-chase experiment by metabolic labeling of nascent mRNA

LC-MS/MS test with 4-thiouridine (4SU) labeling, and colocalization and qPCR test with 5-ethinyluridine (EU) labeling were performed with the same procedure described before[58]. Flag-HA-tagged YTHDF1-3 stable expression HeLa cells were used in these experiments.

Chapter 3

YTHDF1-mediated functions of m^{6}A in learning and memory

3.1 Introduction: functions of m^{6}A highlighted in the brain

As introduced in section **1.4**, multiple studies have demonstrated that m^{6}A governs mRNA stability during cell fate transition and animal development. m^{6}A also affects mRNA translation. In HeLa cells, an m^{6}A binding protein YTHDF1 facilitates translation initiation of m^{6}A-modified mRNAs[58]; while in MCF7 cells, the function of YTHDF1 on translation is more complex, with negative effects observed using reporters[115]. Therefore, how m^{6}A modulates mRNA translation and to what extent this function impacts physiological events in intact biological systems, remain to be elucidated.

Previous studies have suggested the functional importance of m^{6}A in the brain. m^{6}A is shown to modulate numerous neural functions, including dopaminergic signaling in the mouse midbrain[116], flight behavior in flies[86], mouse embryonic and adult neurogenesis[84,91], and mouse axon regeneration[93]. Moreover, a positive correlation between m^{6}A accumulation and brain activity has been suggested in recent researches. Upregulation of m^{6}A has been observed with mouse brain maturation[22]. Dynamic m^{6}A removal and installation appeared to occur after memory-inducing experience in the prefrontal cortex and the hippocampus in mice, and disrupting expression of the demethylase FTO led to enhanced performance in certain memory tasks[117,118].

Learning and memory are fundamental functions of brains, and long-term memory formation is believed to require activity-induced protein synthesis[119]. We speculated whether learning and memory processes could be affected by translational effects of m^{6}A and its binding protein YTHDF1.

3.2 Results

3.2.1 YTHDF1 depletion led to defects in hippocampus-dependent learning and memory

3.2.1.1 Construction and general characterization of the *Ythdf1*-KO mice

We constructed *Ythdf1*-knockout (*Ythdf1*-KO) mice using CRISPR/Cas9[120]. Two sgRNAs (red) were designed to target the 4th exon (E4) of *Ythdf1* (**Figure 3.1A**). Founder #4 with a 179-bp deletion (**Figure 3.1B-D**) was crossed with C57BL/6 wild-type mice for at least five generations for further analysis. The complete elimination of YTHDF1 protein was verified in the hippocampus (**Figure 3.1E-F**) and other brain regions (**Figure 3.1G**) of the *Ythdf1*-KO mice.

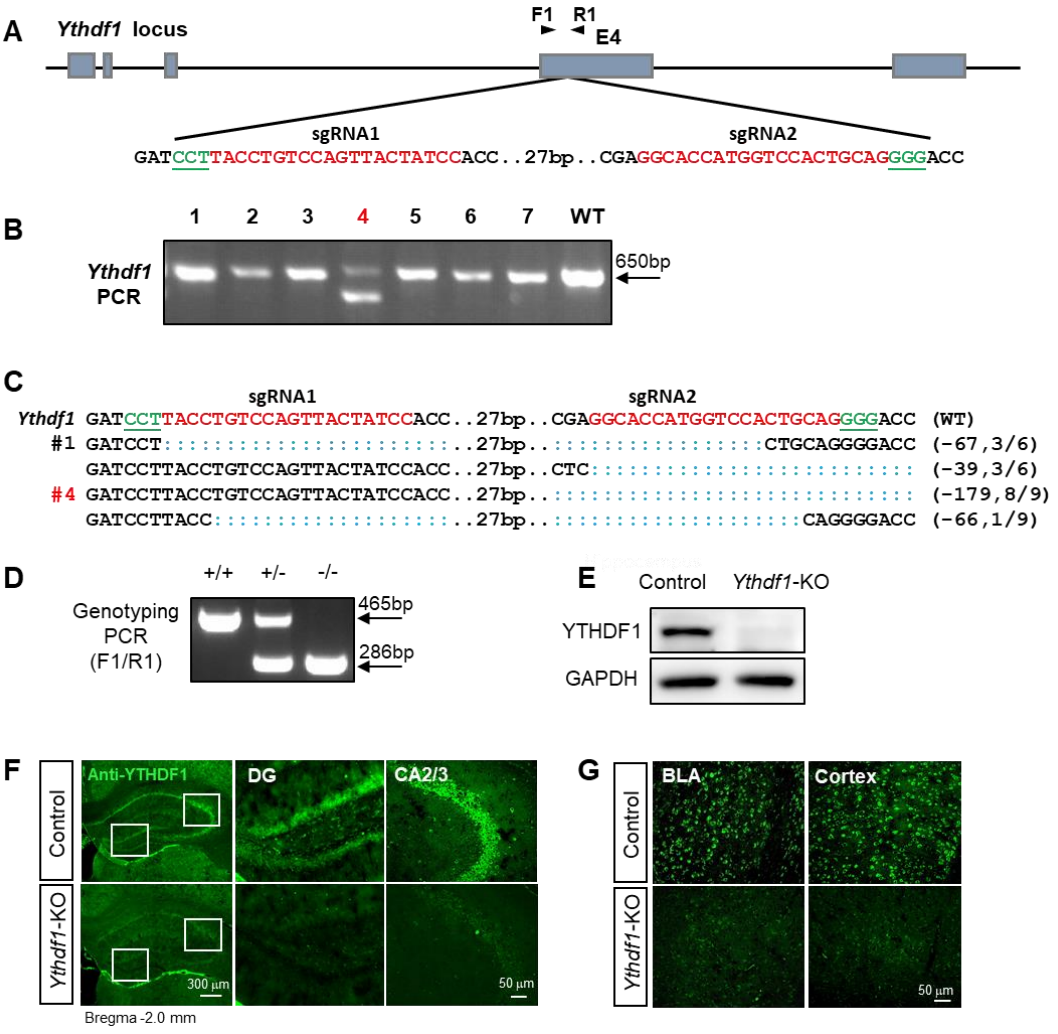

Figure 3.1 CRISRP/Cas9 strategy to generate *Ythdf1*-KO mice

(**Figure 3.1, continued**) (**A**) Schematic diagram of the targeting strategy for generating *Ythdf1*-KO mice using CRISPR/Cas9. PAM sequence, underlined, green; F1 and R1, genotyping primers. (**B-C**) Genotyping PCR product sizes (**B**) and sequences (**C**) of the seven founders co-injected with Cas9 mRNA and the two sgRNAs. (**D**) Representative genotyping PCR products of off-spring mice with different genotypes. (**E**) Validation of *Ythdf1* knockout by western blot using mouse hippocampal tissues. (**F-G**) Representative images of YTHDF1 immunostaining in the mouse hippocampus (**F**) and other brain regions (**G**) in the control and *Ythdf1*-KO mice. DG, dentate gyrus; BLA, basal lateral amygdala.

Compared to wild-type (WT) control littermates, *Ythdf1*-KO mice develop normally before sacrifice (up to four months of age) and appear normal in gross hippocampal (**Figure 3.2A**) and cortical morphology (**Figure 3.2B**). Adult hippocampal neurogenesis also appeared normal as quantified by hippocampal neural stem cell staining (**Figure 3.2C**).

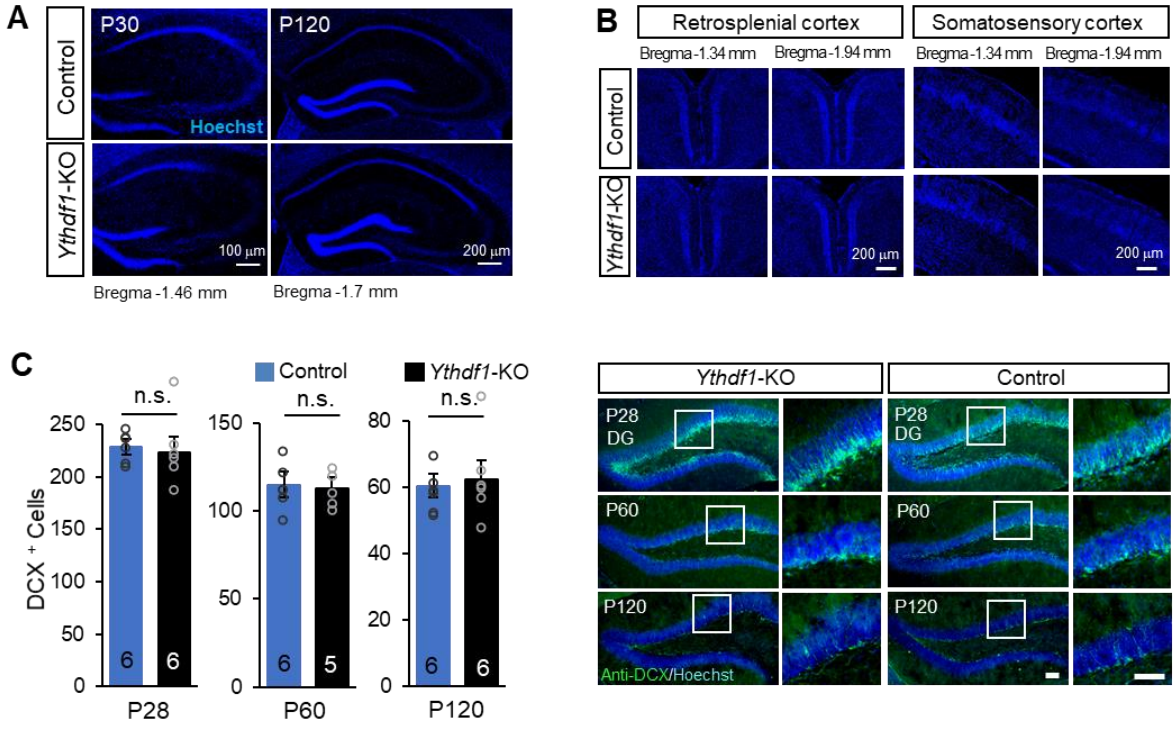

Figure 3.2 Hippocampal and cortical morphology and adult neurogenesis of *Ythdf1*-KO mice

(**A-B**) Representative images of hippocampal (**A**) and cortical (**B**) morphology staining using Hoechst in adult control and *Ythdf1*-KO mice. P30/P120, postnatal day 30/120. (**C**) Quantification of the number of doublecortin (a marker of neurogenesis) positive (DOX^+) cells and representative images of DOX immunostaining in the DG region of *Ythdf1*-KO and wild-type control mice at different postnatal development stages. Scale bar, 100 μm. *P* values, two tailed *t*-test. Numbers in bars, numbers of mice. Error bars, mean ± s.e.m.

Moreover, loss of YTHDF1 does not alter mouse motor abilities in the open-field test (**Figure 3.3A**) nor general emotional states as measured by the light-dark box transition test, the elevated-plus maze test, and the tail-suspension test (**Figure 3.3B-D**).

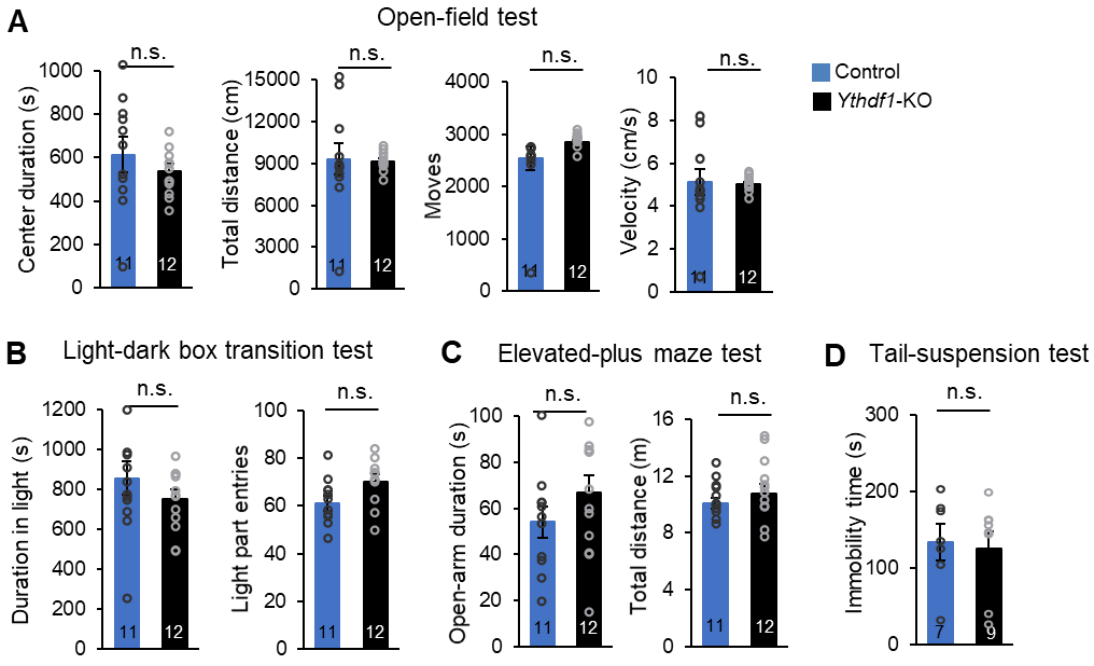

Figure 3.3 Motor activities and emotional states of the *Ythdf1*-KO mice

(**A**) Motor activities measured by various parameters as listed in the open-field test. (**B-C**) Anxiety-like behavior measured by the light-dark box transition test (**B**) and the elevated-plus maze test (**C**). (**D**) Depressive-like behavior measured by the tail suspension test. *P* values, two tailed *t*-test. Numbers in bars, numbers of mice. Error bars, mean ± s.e.m.

3.2.1.2 Hippocampal learning and memory defects in the *Ythdf1*-KO mice

In the mouse brain, *Ythdf1* mRNA is preferentially expressed in the hippocampus[121], a key region participating in spatial learning and memory. We first examined hippocampus-dependent spatial learning and memory of the mice by Morris water maze (MWM) tests[122]. The mice are trained to learn the location of a platform inside a water pool based on visual cues around (**Figure 3.4A**, left). In the visible platform training, *Ythdf1*-KO mice performed as proficiently as WT

control mice (**Figure 3.4B**), indicating normal visual functions and competent acquisition of procedural learning. However, in the hidden platform training, *Ythdf1*-KO mice spent longer time navigating to the platform than control mice (**Figure 3.4C**), suggesting impaired spatial learning.

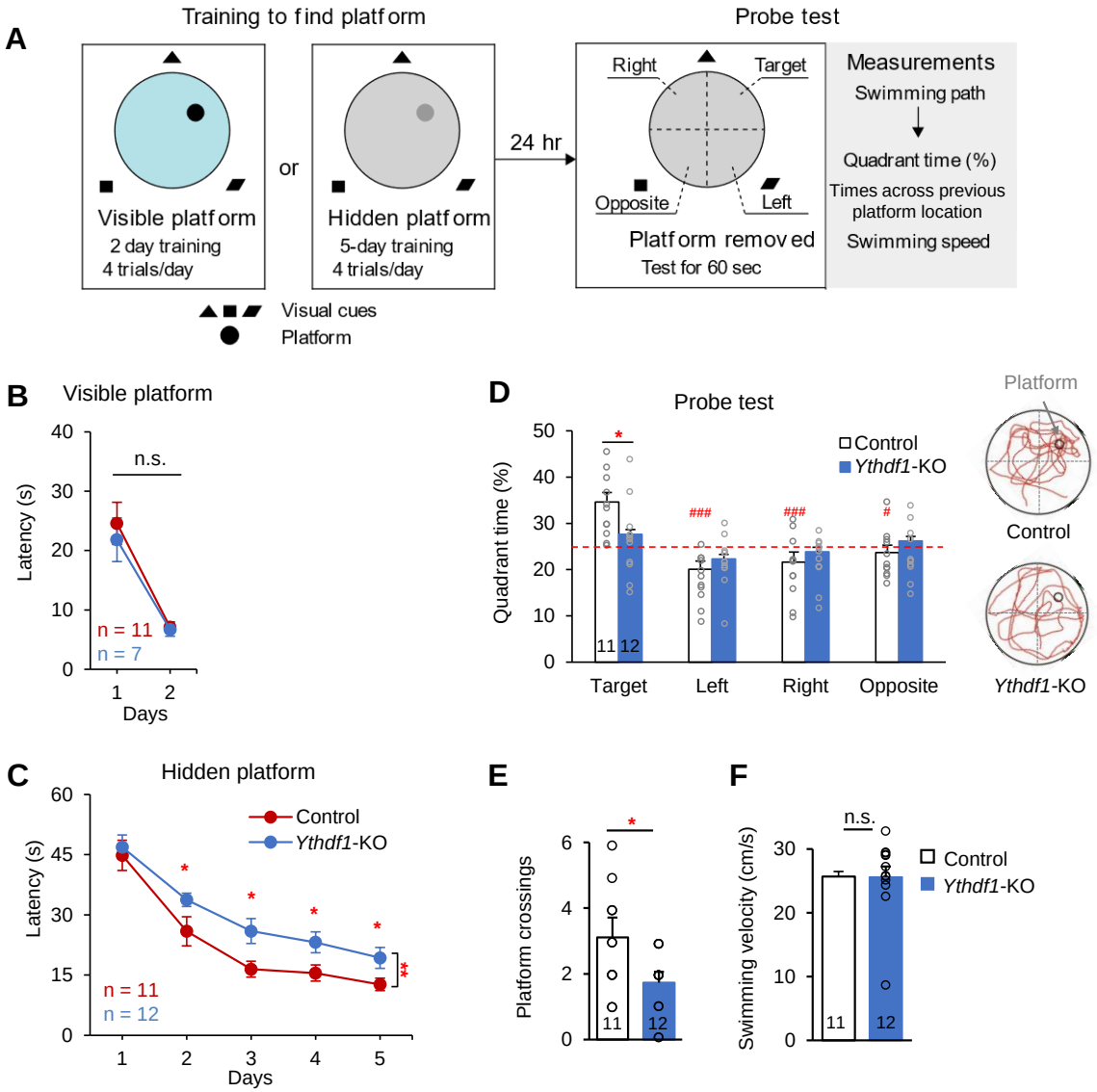

Figure 3.4 Spatial learning and memory defects in the Morris water maze (MWM)

(**A**) Schematics of procedure of MWM trainings and MWM probe tests. (**B-C**) Learning curves of control (red) and *Ythdf1*-KO (blue) mice in Morris water maze (MWM) tests in the visible (**B**) and hidden (**C**) platform trainings. (**D-F**) Quadrant time (%) (**D**, left), representative swimming paths (**D**, right), number of crossings over previous platform location (**E**), and swimming velocity (**F**) of control and *Ythdf1*-KO mice in the MWM probe test. The red dash line represents the chance level (25%). *P* values, two-way repeated measures ANOVA with post hoc test (**B-C**), two-way ANOVA with two tailed *t*-test (relative to "Target" or between genotypes) (**D**), and two-tailed *t*-test (**E-F**). Numbers in bars, numbers of mice. Error bars, mean ± s.e.m.

After the training, the mice were subjected to a probe test, where the platform was removed. Each mouse was allowed a certain time to explore in the water pool and its swimming path was recorded to evaluate whether the mouse remember the previous platform location (**Figure 3.4A**, right). While the control mice spent most of the time in the target quadrant where the platform was located, *Ythdf1*-KO mice failed to do so. They spent similar time in each quadrant (**Figure 3.4D**) and seldomly swam across the previous platform location (**Figure 3.4E**) despite being normal in swimming activity (**Figure 3.4F**), suggesting their defects in spatial memory.

To further confirm the importance of YTHDF1 in hippocampus-dependent learning and memory, we performed classical fear conditioning tests. Each mouse was trained with one or more pairs of tones and foot shocks in a conditioning chamber, and later tested for freezing responses to either: (i) the conditioning chamber itself without the tone (contextual fear memory), or (ii) the tone but inside a different chamber as a context (auditory fear memory) (**Figure 3.5A**). Contextual fear memory is sensitive to hippocampal defects, while auditory fear memory depends on the amygdala[123].

We first titrated the intensity of electric shock to establish fear conditioning protocols that did not saturate fear response (**Figure 3.5B**). Under a moderate training protocol (0.5 mA, 2 s, 1 pair), *Ythdf1*-KO mice showed less freezing response during inter-trial intervals (ITI) but not when the tone sounded (**Figure 3.5C-D**), suggesting that contextual but not auditory learning was impaired. 24 hours after conditioning, *Ythdf1*-KO mice showed deficits in contextual (**Figure 3.5E**) but not auditory fear memory (**Figure 3.5F**). Consistently, under a weaker training protocol (0.5 mA, 1s, 1 pair), 2-hour contextual but not auditory fear memory of the mice was impaired (**Figure 3.5G-H**). Together, these data support that genetic deletion of *Ythdf1* disrupts learning and memory formation in the mouse hippocampus.

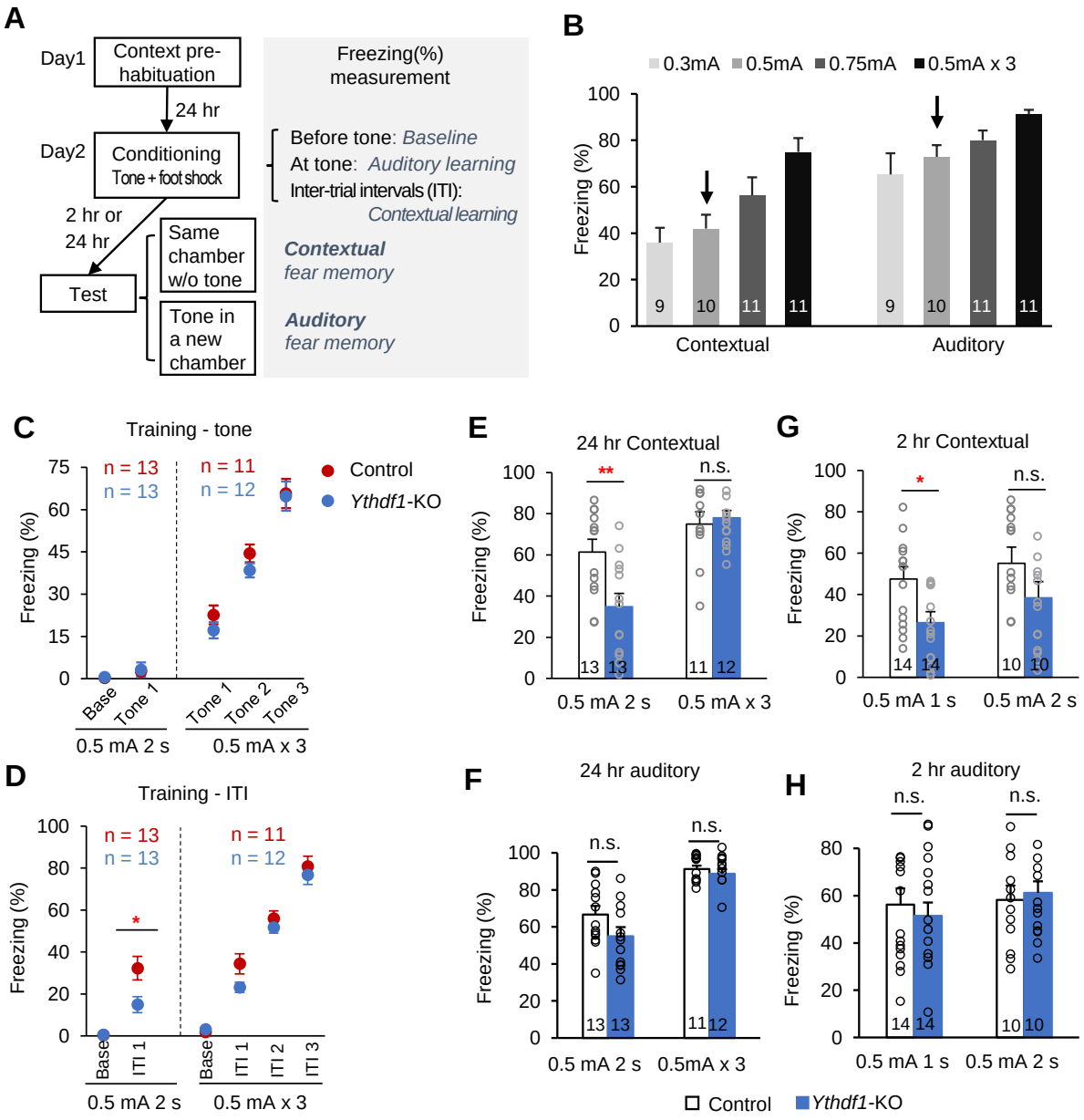

Figure 3.5 Contextual and auditory fear memories of the *Ythdf1*-KO mice

(**A**) Schematics of the fear conditioning procedures (left) and freezing responses measured at different stages (right). (**B**) Titration curves of the freezing level of wild-type mice 24 hours after trainings with different foot shock intensities. The conditioning protocols used in later experiments (moderate protocols) are indicated by arrows. (**C-D**) Learning curves for auditory (**C**) and contextual (**D**) fear conditioning (FC) under moderate (left) or strong (right) training protocols. The training sessions were separated into three parts: baseline (base), tone periods (tone), and inter-trial intervals (ITI). (**E-H**) Contextual (**E&G**) and auditory (**F&H**) fear memory assessed 24 hours (**E-F**) or 2 hours (**G-H**) after the indicated FC. *P* values, two-way repeated measures ANOVA with post hoc test (**C-D**) and two-tailed *t*-test (**E-H**). Numbers in bars, numbers of mice. Error bars, mean ± s.e.m.

3.2.1.3 Long-term-potentiation (LTP) defects in the *Ythdf1*-KO acute hippocampal slices

We next used electrophysiological characterization to study how YTHDF1 depletion affects hippocampal synaptic functions. Basal synaptic properties of hippocampal CA1 neurons were examined by whole-cell patch-clamp (see Methods **3.4.5**).

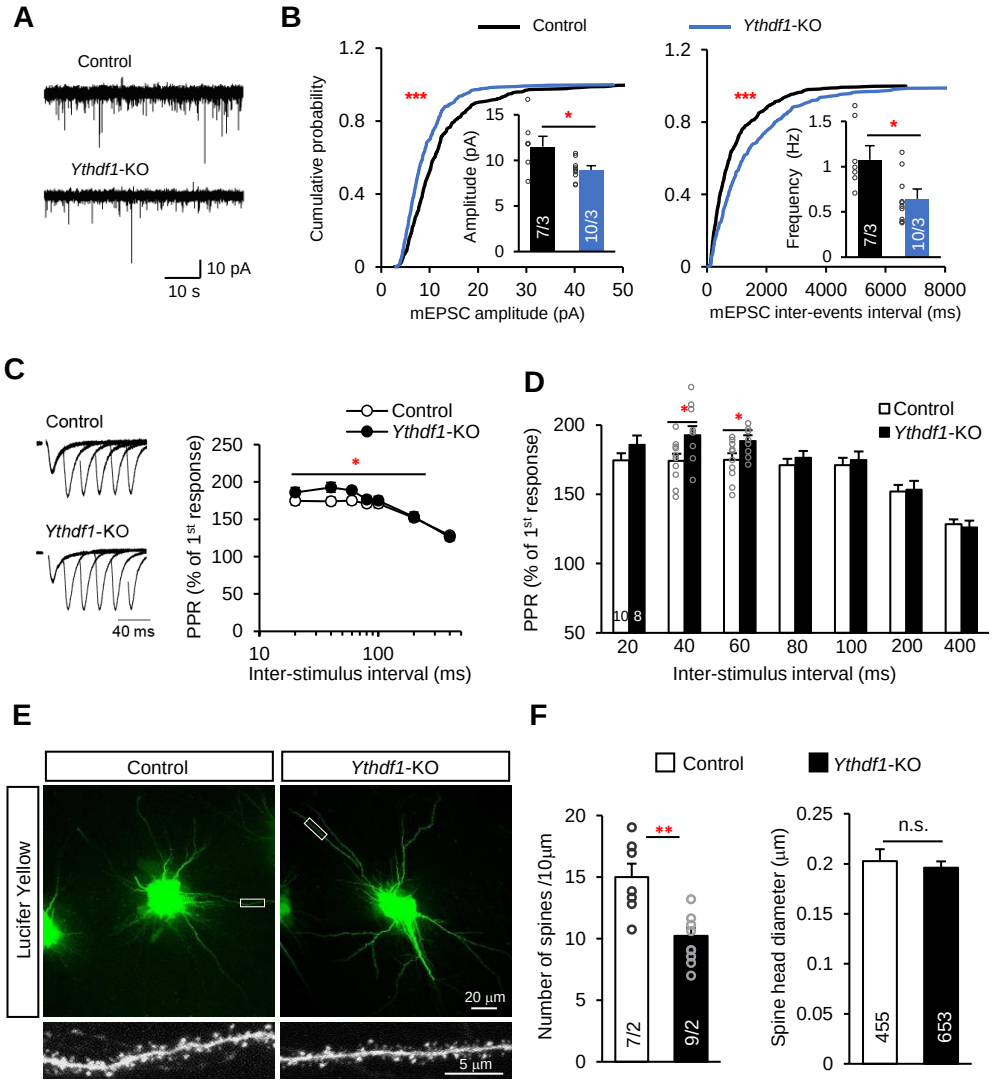

Figure 3.6 Basal synaptic properties of the *Ythdf1*-KO hippocampal CA1 neurons

(**A-B**) Representative traces (**A**) and quantification of amplitude (**B**, left) and frequency (**B**, right) of mEPSCs in control and *Ythdf1*-KO hippocampal CA1 neurons. (**C-D**) PPR with different inter-stimulus intervals in CA1 neurons from wild-type control and *Ythdf1*-KO mice. (**E-F**) Representative images of Lucifer Yellow staining (**E**) and statistical analyses of spine density (**F**, left) and spine size (**F**, right) in CA1 neurons from adult control and *Ythdf1*-KO brain. *P* values, Kolmogorov-Smirnov test for cumulative distributions followed by comparisons with Mann–

(**Figure 3.6, continued**) Whitney *U* test (**B**), two-way repeated measures ANOVA with post hoc two-tailed *t*-test (**C**), and two-tailed *t*-test (**F**). Numbers in bars, numbers of neurons/mice (**B**), neurons/mice (**F**, left), spines (**F**, right). Error bars, mean ± s.e.m.

In *Ythdf1*-KO CA1 neurons, spontaneous miniature excitatory postsynaptic currents (mEPSCs) were significantly decreased in amplitude and frequency, compared to control (**Figure 3.6A-B**), suggesting that loss of YTHDF1 might affect presynaptic glutamate release and/or postsynaptic number of AMPA receptors. Paired-pulse ratio (PPR) analyses also indicate reduced presynaptic release probability in *Ythdf1*-KO CA1 neurons (**Figure 3.6C-D**), confirming the defects in basal synaptic transmission. Morphologically, *Ythdf1*-KO CA1 neurons have reduced dendritic spine density yet unaltered spine size (**Figure 3.6E-F**).

Long-term potentiation (LTP) is a major cellular model for explaining learning and memory. To test whether YTHDF1 modulates long-term synaptic plasticity, we recorded LTP induced by high frequency stimulation (HFS) in the CA1 region of hippocampal slices. Compared to WT controls, *Ythdf1*-KO slices failed to generate normal levels of field excitatory postsynaptic potential (fEPSP) after 2 × HFS induction (**Figure 3.7A-B**). The initial potentiation following HFS was similar between control and *Ythdf1*-KO slices (**Figure 3.7A**); therefore, it is less likely that the decreased LTP was due only to impairments in basal synaptic transmission. *Ythdf1*-KO slices were also defective in late phase LTP (induced by 4 × HFS) (**Figure 3.7C-D**), a process requiring activity-induced synaptic protein synthesis. Indeed, depletion of YTHDF1 noticeably reduced the abundance of key proteins involved in LTP in the postsynaptic density (PSD) fraction of hippocampal neurons (**Figure 3.7E-F**), although such decreases were not observed for those proteins in the whole hippocampal tissue (**Figure 3.7G-H**). Together, these results suggest that YTHDF1 depletion in mice impairs basal synaptic transmission and LTP in hippocampal neurons, contributing to the observed defects in learning and memory.

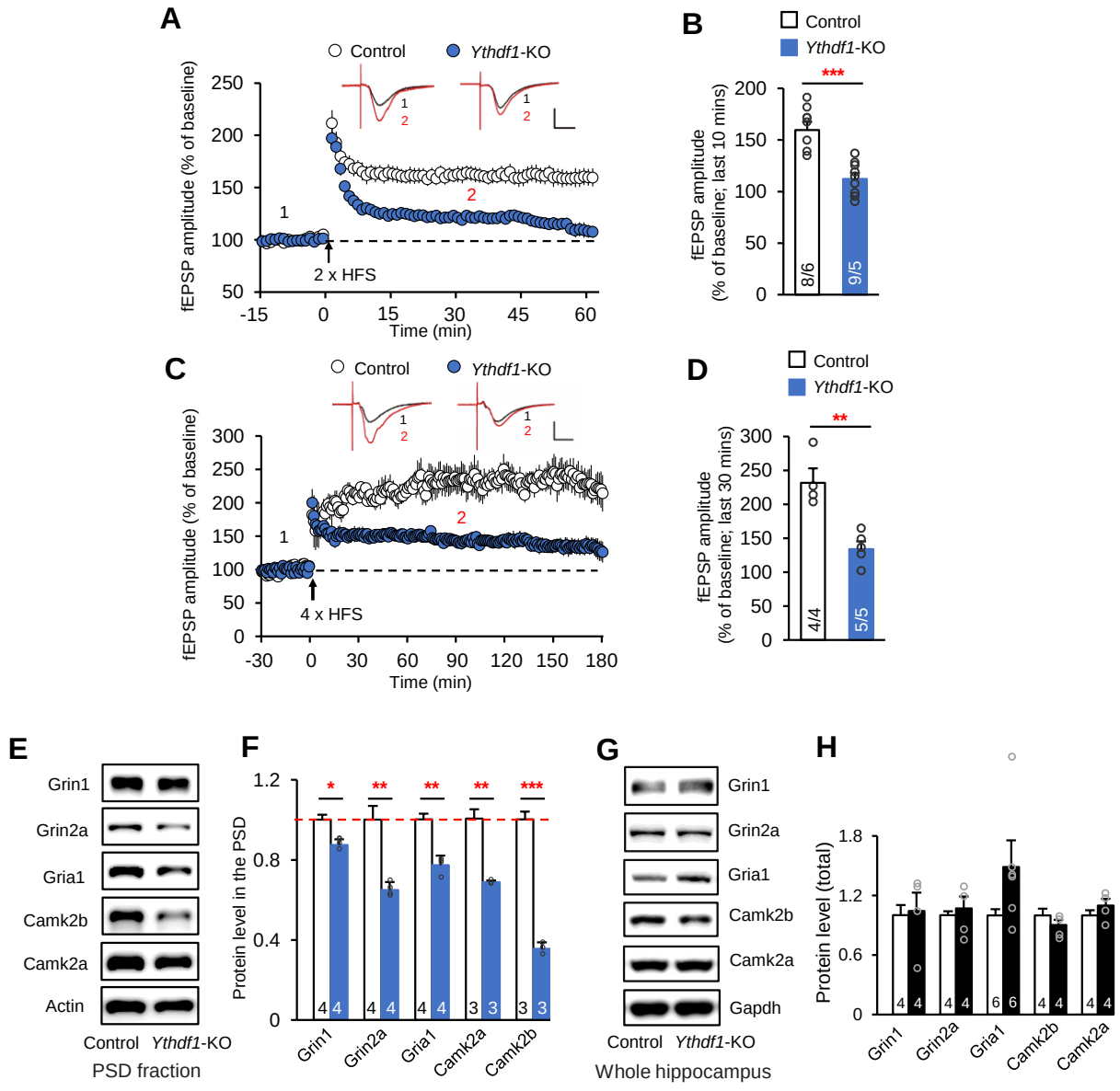

Figure 3.7 Long-term-potentiation (LTP) in the *Ythdf1*-KO hippocampal CA1 slices

(**A-B**) Summary plots (**A**) and average amplitude (**B**) of long-term potentiation (LTP) induced by 2 × high frequency stimulation (HFS) in the CA1 region of control and *Ythdf1*-KO acute slices. fEPSP, field excitatory postsynaptic potential. (**C-D**) Summary plots (**C**) and average amplitude (**D**) of late phase LTP induced by 4 × HFS. Top panels, sample traces taken at time points 1 and 2 indicated above the summary plots; scale bars, 10 ms (horizontal) and 0.2 mV (vertical) (**A&C**). (**E-H**) Representative western blots (**E**, **G**) and quantification (**F**, **H**) of a number of LTP-related proteins in the control and *Ythdf1*-KO hippocampal postsynaptic density (PSD) fraction (**E-F**) and the whole hippocampus (**G-H**). *P* values, two-tailed *t*-test. Numbers in bars, numbers of slices/mice (**B**, **D**) and number sof mice (**F**, **H**). Error bars, mean ± s.e.m.

3.2.1.4 Hippocampal-specific YTHDF1 re-expression rescues the defects

To confirm that the observed defects were derived from the loss of YTHDF1 specifically in the hippocampus, we investigated whether re-expressing YTHDF1 in the hippocampus of adult KO mice would be sufficient to rescue the phenotypes. We delivered the adeno-associated virus (AAV) expressing either mouse YTHDF1 (AAV-YTHDF1) or a control fluorescent protein mCherry (AAV-control) (**Figure 3.8A**) specifically to the hippocampus by bilateral stereotactic injection (**Figure 3.8B**), resulting in selective re-expression in injected regions (**Figure 3.8C-E**).

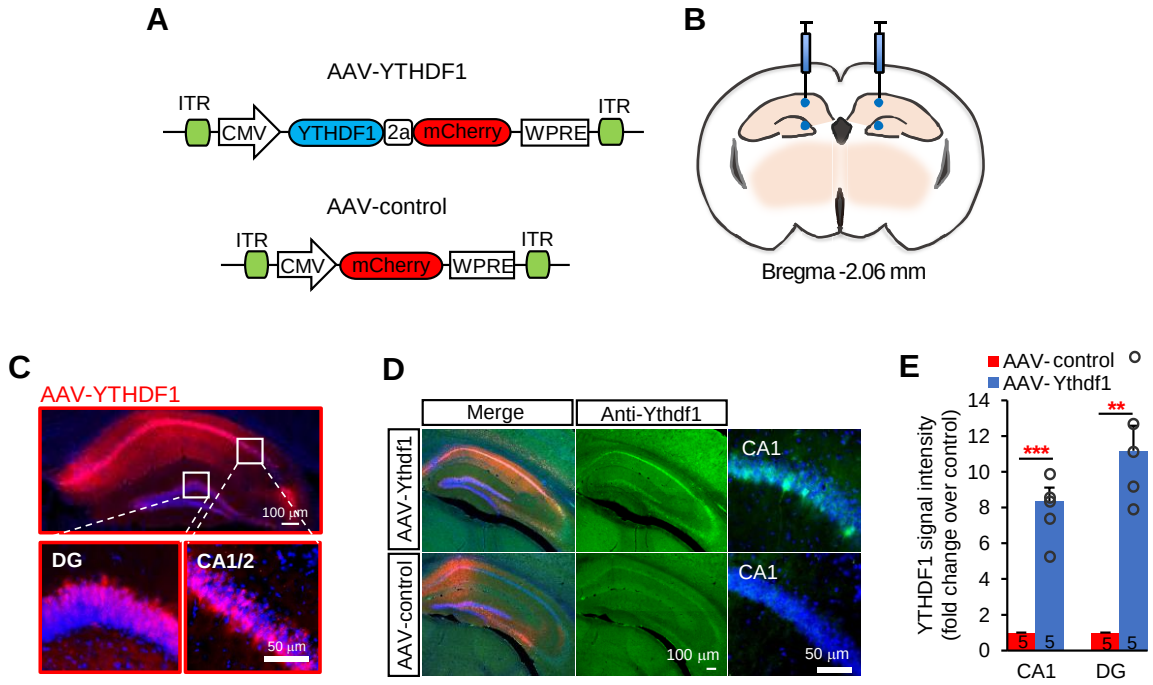

Figure 3.8 Hippocampal-specific re-expression of mouse YTHDF1

(**A**) Schematics of AAV constructs overexpressing mouse YTHDF1 (AAV-YTHDF1) or control (AAV-control). ITR, inverted terminal repeats; CMV, cytomegalovirus promoter; WPRE, Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element. (**B**) Illustration of bilateral viral injections into the mouse hippocampus. (**C-D**) Representative fluorescence images (**C**) and YTHDF1 immunostaining (green) (**D**) in the mouse hippocampus after virus infection. Hoechst, blue; YTHDF1 co-expressed with mCherry, red. (**E**) YTHDF1 protein overexpression level indicated by immunofluorescent signal intensity in the CA1 and DG regions. *P* values, two tailed *t*-test. Numbers in bars, numbers of mice. Error bars, mean ± s.e.m.

Hippocampal YTHDF1 re-expression in the KO mice substantially enhanced their learning (**Figure 3.9A**) and memory (**Figure 3.9B-C**) performances in MWM tests and rescued contextual fear memory to normal levels (**Figure 3.9D**), with no obvious effect on auditory fear memory (**Figure 3.9E**), anxiety-like behavior (**Figure 3.9F**), nor motor activity (**Figure 3.9G**); it also reversed hippocampal LTP deficiency in *Ythdf1*-KO mouse (**Figure 3.10**).

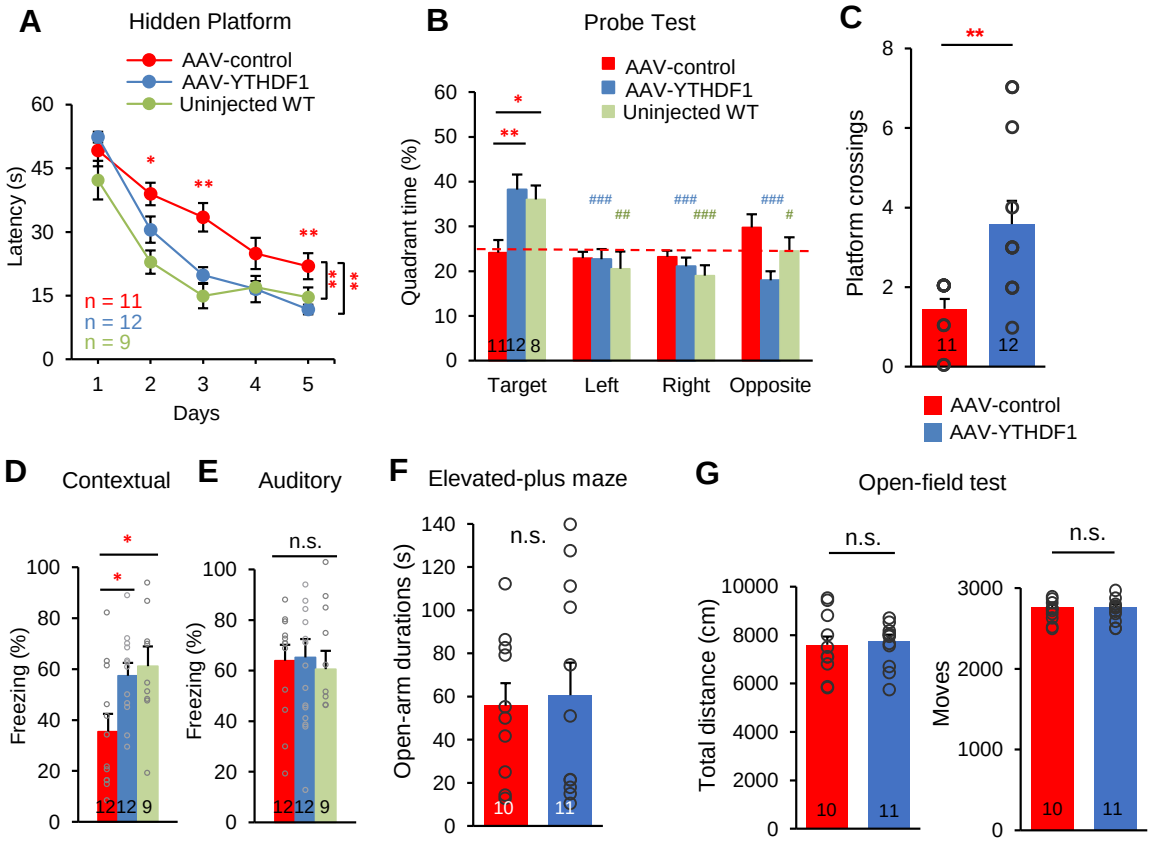

Figure 3.9 YTHDF1 re-expression rescues hippocampal learning and memory defects

(**A-C**) Learning curves in MWM hidden-platform trainings (**A**), quadrant time (%) in MWM probe tests (**B**), and the number of crossings over the previous platform location (**C**) for *Ythdf1*-KO mice injected with AAV-control (red) or AAV-YTHDF1 (blue), compared to uninjected wild-type (WT, green) in MWM probe tests. (**D-E**) Contextual (**D**) and auditory (**E**) fear memories assessed 24 hours after fear conditioning. (**F**), Anxiety-like behavior of the injected mice measured as open-arm durations in elevated-plus maze. (**G**) Motor activities of the injected mice measured as total distance and number of moves in the open-field test. *P* values, two-way repeated measures ANOVA with post hoc two-tailed *t*-test (AAV-YTHDF1 relative to AAV-control or comparisons between curves) (**A**), two-way ANOVA with post hoc two-tailed *t*-test (comparison within group or with "Target") (**B**), one-way ANOVA with post hoc Fisher test (**D-E**), and two-tailed *t*-test (**F-G**). Numbers in bars, numbers of mice. Error bars, mean ± s.e.m.

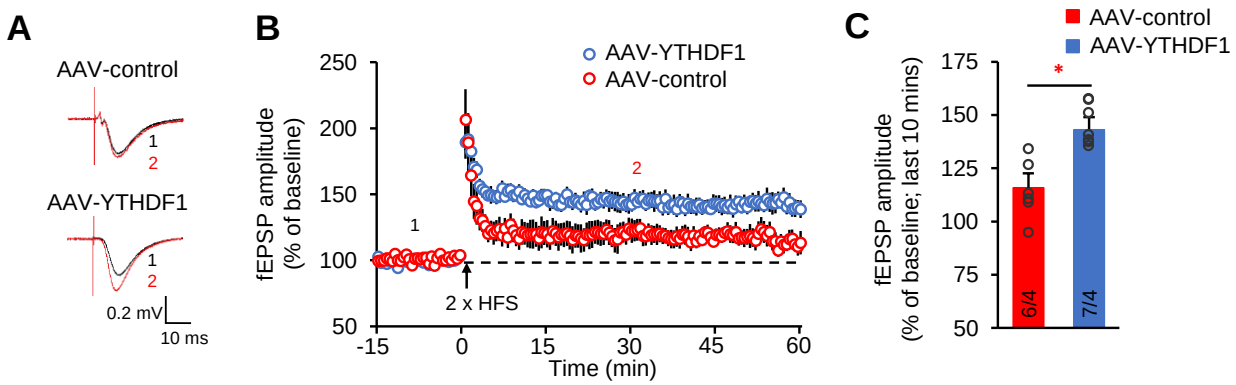

Figure 3.10 Hippocampal YTHDF1 re-expression rescues the LTP defects

Representative traces (**A**), summary plots (**B**), and average amplitude (**C**) of LTP induced by 2 × HFS in acute slices from *Ythdf1*-KO mice injected with AAV-YTHDF1 or AAV-control. Sample traces (**A**) were taken at the time points 1 and 2 indicated the summary plots (**B**). *P* values, two-tailed *t*-test. Numbers in bars, numbers slices/mice. Error bars, mean ± s.e.m.

3.2.1.5 Hippocampal-specific YTHDF1 knockdown phenocopies the *Ythdf1*-KO mice

To test whether acute loss of YTHDF1 in the hippocampus was sufficient to induce the phenotypes of *Ythdf1*-KO mice, we injected AAV expressing a short hairpin RNA targeting *Ythdf1* transcripts (AAV-RNAi) to the hippocampus of adult WT mice (**Figure 3.11A**). Specific knockdown of YTHDF1 but not other YTH proteins was verified in N2A cells (**Figure 3.11B**). In mice injected with AAV-RNAi, learning and memory performances in MWM tests were dramatically impaired (**Figure 3.11C-F**), as was contextual fear memory (**Figure 3.11G**) but not emotional states (**Figure 3.11H**) nor auditory fear memory (**Figure 3.11I**).

Moreover, hippocampal-specific knockdown of *Mettl3* (**Figure 3.12A**) also phenocopied the effects of YTHDF1 depletion, leading to defects in spatial memory (**Figure 3.12B-D**) and contextual fear memory (**Figure 3.12E**) without affecting auditory fear memory (**Figure 3.12F**) nor locomotor activities (**Figure 3.12G**). These knockdown results further support that the observed phenotypes in *Ythdf1*-KO mice come from direct depletion of YTHDF1 instead of potential developmental defects caused by lack of YTHDF1 and are m^6A-dependent.

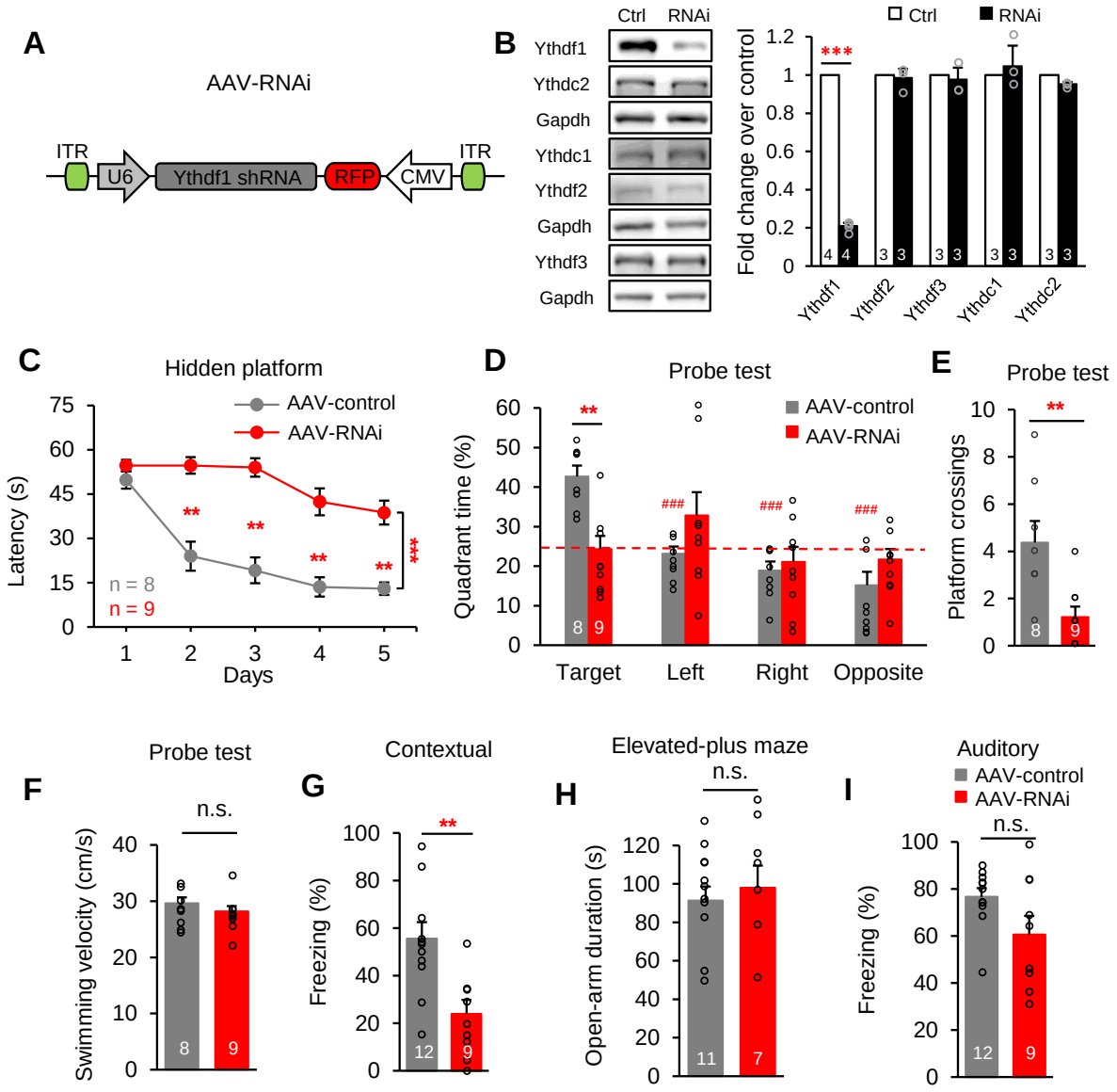

Figure 3.11 Impaired spatial learning and memory after hippocampal knockdown of Ythdf1

(**A**) Schematics of the AAV construct expressing Ythdf1 shRNA. (**B**) Western blot and quantification of protein expression level of YTH proteins in N2A cells after Ythdf1-shRNA (RNAi) or control vector (Ctrl) transfection. (**C**) Spatial learning curves in the hidden-platform MWM training sessions for RNAi (red) and control (gray) mice. (**D-F**) Spatial memory performances measured by quadrant time (%) (**D**) and number of platform crossings (**E**), and motor activities (**F**) of RNAi (red) and control (gray) mice in MWM probe tests. (**G**, **I**) Contextual (**G**) and auditory (**I**) fear memories assessed 24 hours after fear conditioning in RNAi and control mice. (**H**) Anxiety level of mice assessed by open-arm durations in elevated-plus maze. *P* values, two-way repeated measures ANOVA with post hoc two-tailed *t*-test (**C**), two-way ANOVA with two-tailed *t*-test (comparison between group or to "Target") (**D**), and two tailed *t*-test (**B**, **E-I**). Numbers in bars, numbers of biologically independent samples (**B**) and mice (**D-I**). Error bars, mean ± s.e.m.

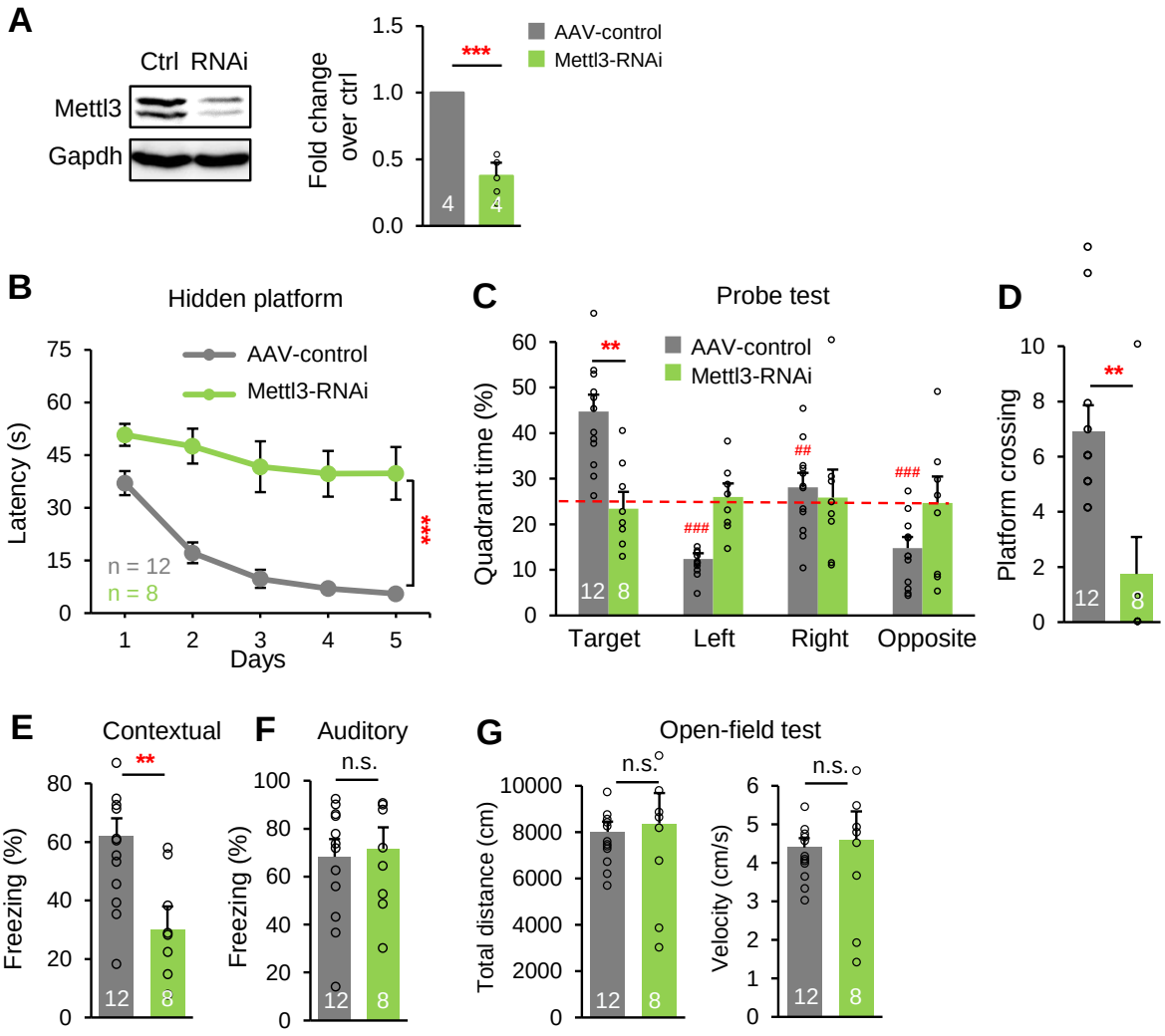

Figure 3.12 Impaired spatial learning and memory after hippocampal knockdown of Mettl3

(**A**) Representative western blot (left) and quantification (right) of Mettl3 protein level in N2A cells transfected with Mettl3-shRNA (RNAi) or control vector (Ctrl). (**B-D**) Spatial learning curves in the hidden-platform MWM training sessions (**B**), and spatial memory performances measured by quadrant time (%) (**C**) and the number of platform crossings (**D**) in MWM probe tests, for Mettl3-RNAi and control mice. (**E-F**) Contextual (**E**) and auditory (**F**) fear memories measured by freezing levels 24 hours after fear conditioning in Mettl3-RNAi and control mice. (**G**) Motor activities of mice accessed in the open-field test. *P* values, two-way repeated measures ANOVA with post hoc two-tailed *t*-test (**B**), two-way ANOVA with two-tailed *t*-test (comparison between groups or to "Target") (**C**), and two tailed *t*-test (**A**, **D-G**), Numbers in bars, numbers of biologically independent samples (**A**) and mice (**C-G**). Error bars, mean ± s.e.m.

3.2.2 m^{6}A methylation and YTHDF1-binding enrich functional transcripts in neurons

3.2.2.1 Transcriptome-wide characterization of YTHDF1 targets in the mouse hippocampus

We next proceeded to investigate underlying molecular mechanisms. We mapped YTHDF1 binding sites and m^{6}A sites on hippocampal mRNAs using crosslinking-and-immunoprecipitation-based sequencing methods (CLIP-seq). Biological triplicates of YTHDF1-CLIP-seq identified 3,552 common peaks as high-confidence peaks (**Figure 3.13A**) on 1,042 transcripts (defined as "YTHDF1-CLIP targets"), with validated pull-down efficiency of YTHDF1 (**Figure 3.13B**). About two-thirds of the high-confidence YTHDF1-CLIP peaks are mapped to mature mRNAs (**Figure 3.13C**) and enriched near the stop codon and 3'UTR (**Figure 3.13D**). Functional annotation of YTHDF1-CLIP targets reveals substantial enrichment for synaptic transmission and LTP (**Figure 3.13E**), consistent with the neuronal phenotypes observed in *Ythdf1*-KO mice.

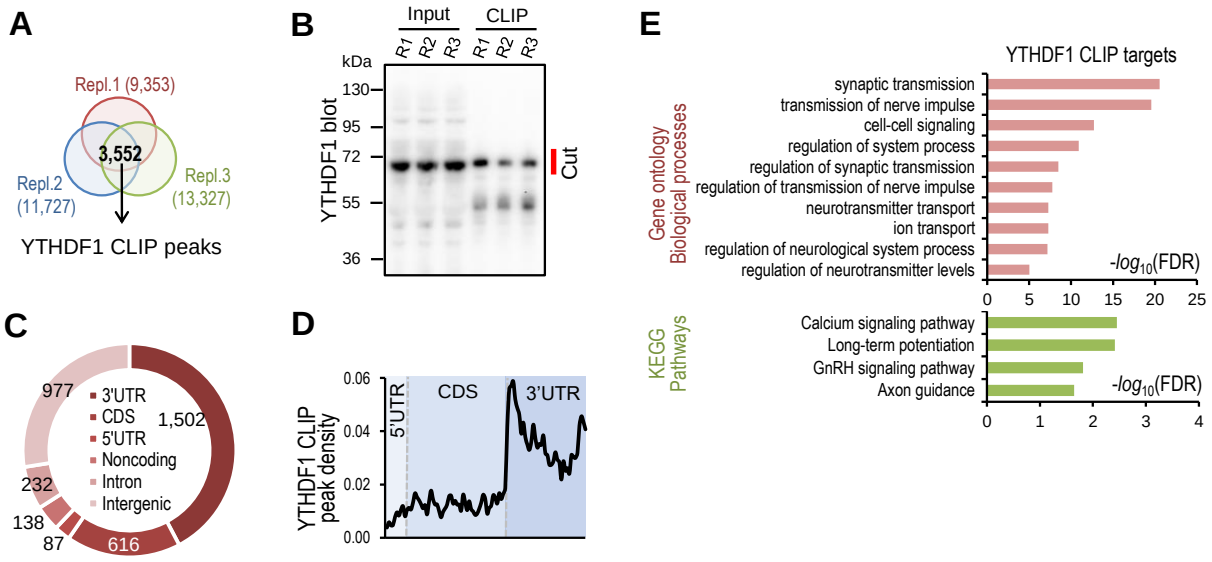

Figure 3.13 YTHDF1 binding sites and target transcripts in the mouse hippocampus

(**A**) Peak overlap among three biological replicates of YTHDF1-CLIP-seq. (**B**) Validation of immunoprecipitation efficiency for YTHDF1-CLIP-seq. The position of the gel slice cut during the step of protein-RNA complex size selection was indicated in red (see Methods **3.4.16**). (**C-D**) Distributions of high-confidence YTHDF1-CLIP peaks in different regions of genome (**C**) and transcripts (**D**). (**E**) Functional annotation of YTHDF1-CLIP targets (n = 1,032) in the adult mouse hippocampus.

3.2.2.2 m^{6}A profiling in the mouse hippocampus

Biological triplicates of m^{6}A-CLIP-seq using purified hippocampal poly(A)+ RNA yielded ~11,000 common peaks, with the GGACU consensus sequence (**Figure 3.14A**) and enrichment at CDS and 3'UTR (**Figure 3.14B-C**) on 3,460 transcripts (defined as "m^{6}A-modified transcripts"). Similarly, genes mediating neuronal biological processes are overrepresented in m^{6}A-modified transcripts (**Figure 3.14D**).

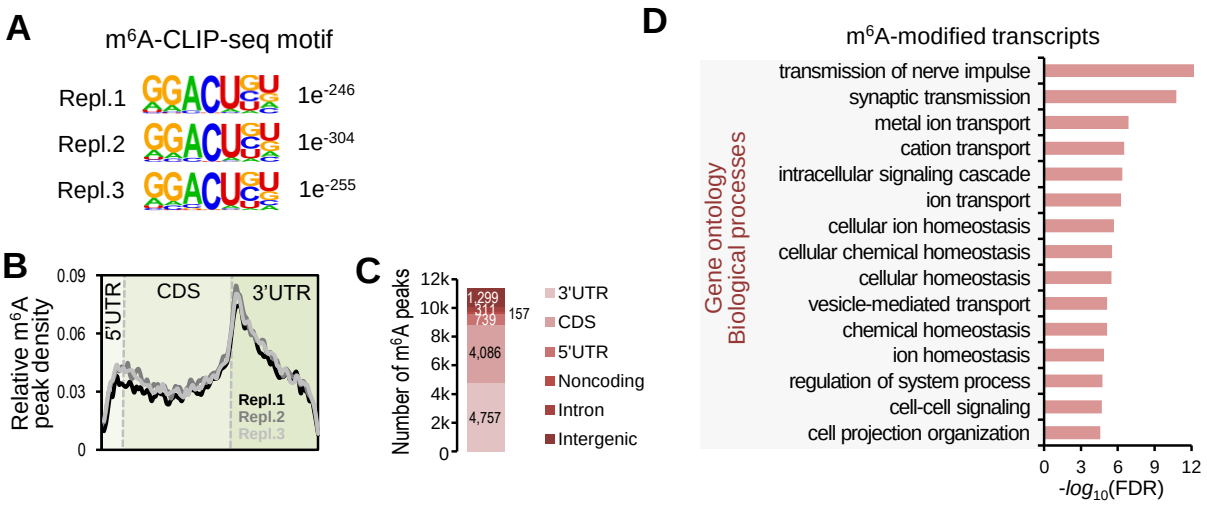

Figure 3.14 m^{6}A methylation sites on the mouse hippocampal transcriptome

(**A**) Consensus motif and its *P* value generated by HOMER of the three sets of hippocampal m^{6}A sites from biological triplicates of m^{6}A-CLIP-seq. (**B-C**) Distribution of m^{6}A-CLIP peaks along the different regions of transcripts (**B**) and genome (**C**). (**D**) Functional annotation of m^{6}A-modified transcripts in the adult mouse hippocampus (number of mutations in m^{6}A-CLIP-seq >= 5, *n* = 2,922).

At the transcript level, YTHDF1-CLIP targets on average bear higher numbers of m^{6}A-CLIP peaks (**Figure 3.15A**) and crosslinking-induced mutations (**Figure 3.15B**) detected in the m^{6}A-CLIP-seq, either compared to transcripts without YTHDF1-CLIP peaks (defined as "non-YTHDF1-CLIP transcripts"), or compared to m^{6}A-modified transcripts; at the peak level, 30% of YTHDF1-CLIP peaks overlap (> 1 nt) with m^{6}A-CLIP peaks (**Figure 3.15C**; in comparison, 0.65-0.72% of background peaks overlap with m^{6}A-CLIP peaks). These results indicate that YTHDF1

preferentially recognizes m^{6}A sites in the adult mouse hippocampus. Key synaptic plasticity transcripts, including *Gria1*, *Grin1*, and *Camk2a*, contain one or multiple overlapped YTHDF1-CLIP peaks and m^{6}A-CLIP peaks (**Figure 3.15D**).

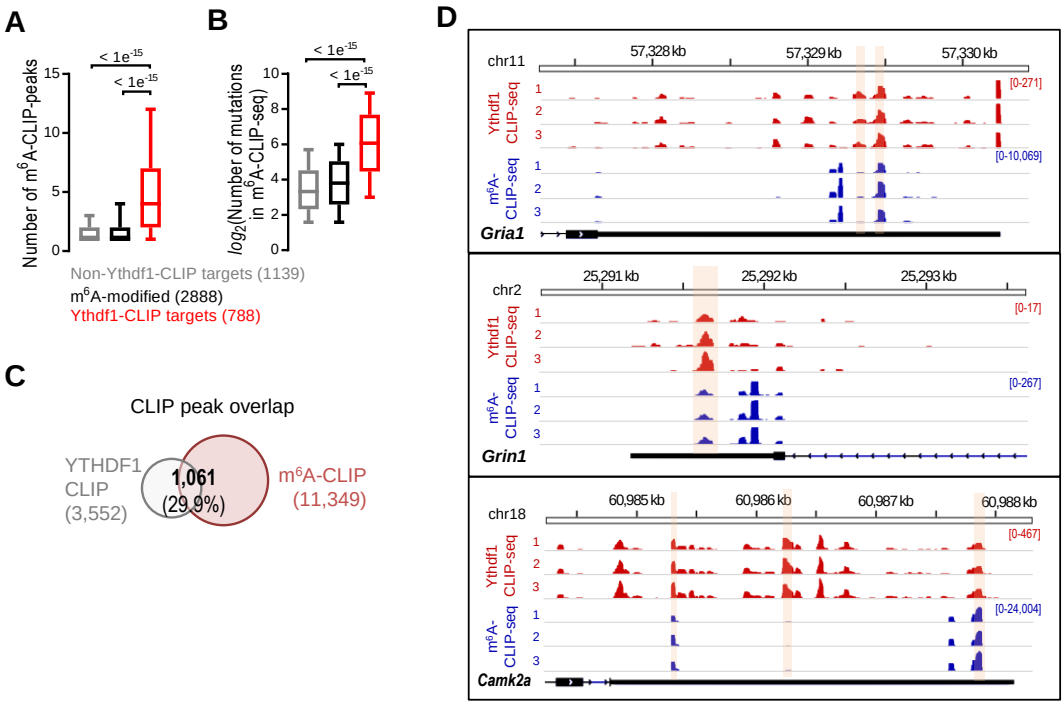

Figure 3.15 Overlap of YTHDF1 binding sites and m^{6}A sites on the hippocampal mRNAs

(**A-B**) Box-plots of the number of m^{6}A-CLIP peaks (**A**) and the log$_2$ number of m^{6}A-CLIP-seq mutations (**B**) on m^{6}A-modified transcripts, non-YTHDF1-CLIP transcripts, and YTHDF1-CLIP targets. Box-plot elements: center line, median; box limits, upper and lower quartiles; whiskers, 1-99%; error bars; number in parentheses, number of genes. *P* values, two-sided unpaired Kolmogorov-Smirnov test. (**C**) Peak overlap between high-confidence YTHDF1-CLIP peaks and high-confidence m^{6}A-CLIP peaks. The percentage of YTHDF1-CLIP peaks overlapped is indicated. (**D**) IGV screenshots of the piled mutated reads for the each of the biological triplicates of YTHDF1-CLIP-seq (red) and m^{6}A-CLIP-seq (blue). Three examples of synaptic plasticity transcripts were presented; the overlapped peak regions are highlighted in light orange.

3.2.3 YTHDF1 promotes translation in mouse neurons in response to stimulus

3.2.3.1 The effects of m^{6}A and YTHDF1 on the basal mRNA and protein levels

We then profiled mRNA and protein abundance in the hippocampus of *Ythdf1*-KO and WT control mice. We noticed that YTHDF1-CLIP targets and m^{6}A-modified transcripts exhibit a very

slight decrease in mRNA abundance (**Figure 3.16A**) and no global protein level changes (**Figure 3.16B**) in *Ythdf1*-KO hippocampus compared to control. These mild changes prompted us to investigate whether YTHDF1 functions in a neuronal-stimulus-dependent manner.

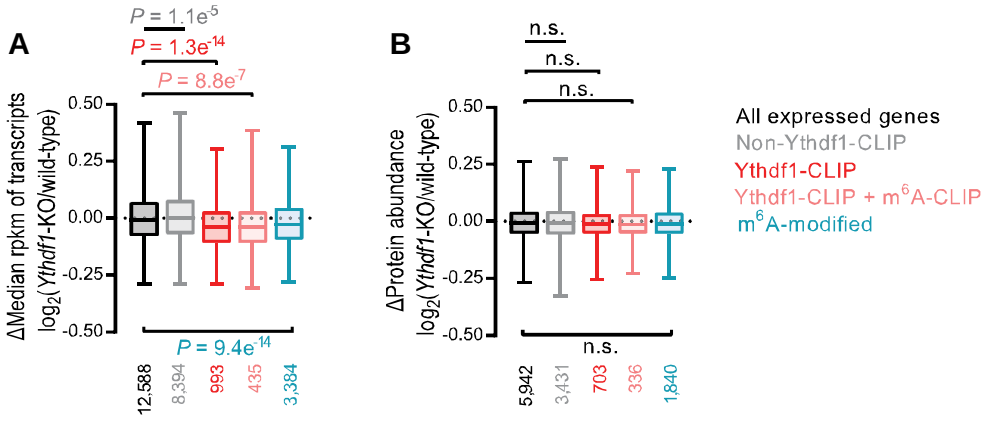

Figure 3.16 Mild roles of YTHDF1 and m^{6}A on global hippocampal mRNA and protein levels

Box-plots of mRNA abundance (**A**) and protein abundance (**B**) $\log_2$ fold changes (Δ) between *Ythdf1*-KO hippocampus and wild-type control for all expressed genes (black), non-YTHDF1-CLIP transcripts (gray), YTHDF1-CLIP targets (red), transcripts with overlapped YTHDF1-CLIP peaks and m^{6}A-CLIP peaks (pink), and m^{6}A-modified transcripts (blue). Box-plot elements: center line, median; box limits, upper and lower quartiles, whiskers, 1-99%; *P* values, two-sided unpaired Kolmogorov-Smirnov test; number of genes is indicated below each box.

3.2.3.2 YTHDF1 promotes protein synthesis in response to stimuli in cultured neurons

To test the hypothesis of stimuli-dependency, we monitored nascent protein synthesis in cultured WT control and *Ythdf1*-KO hippocampal neurons before and a few hours after potassium chloride depolarization (KCl, 50 mM). Neurons were isolated from WT control and *Ythdf1*-KO mouse embryos, and nascent proteins were imaged using an alkyne-puromycin analogue clicked with fluorophore Alexa488[124]. KCl depolarization induced noticeable protein synthesis 2 hours (**Figure 3.17A&C**) and 4 hours (**Figure 3.17B&D**) post stimulation in control neurons but much less in *Ythdf1*-KO ones. In the control neurons, a higher nascent protein signal at 2 hours post KCl compared to the one at 4 hours post KCl may suggest a wave of protein production in response to neuronal stimuli (**Figure 3.17C-D**).

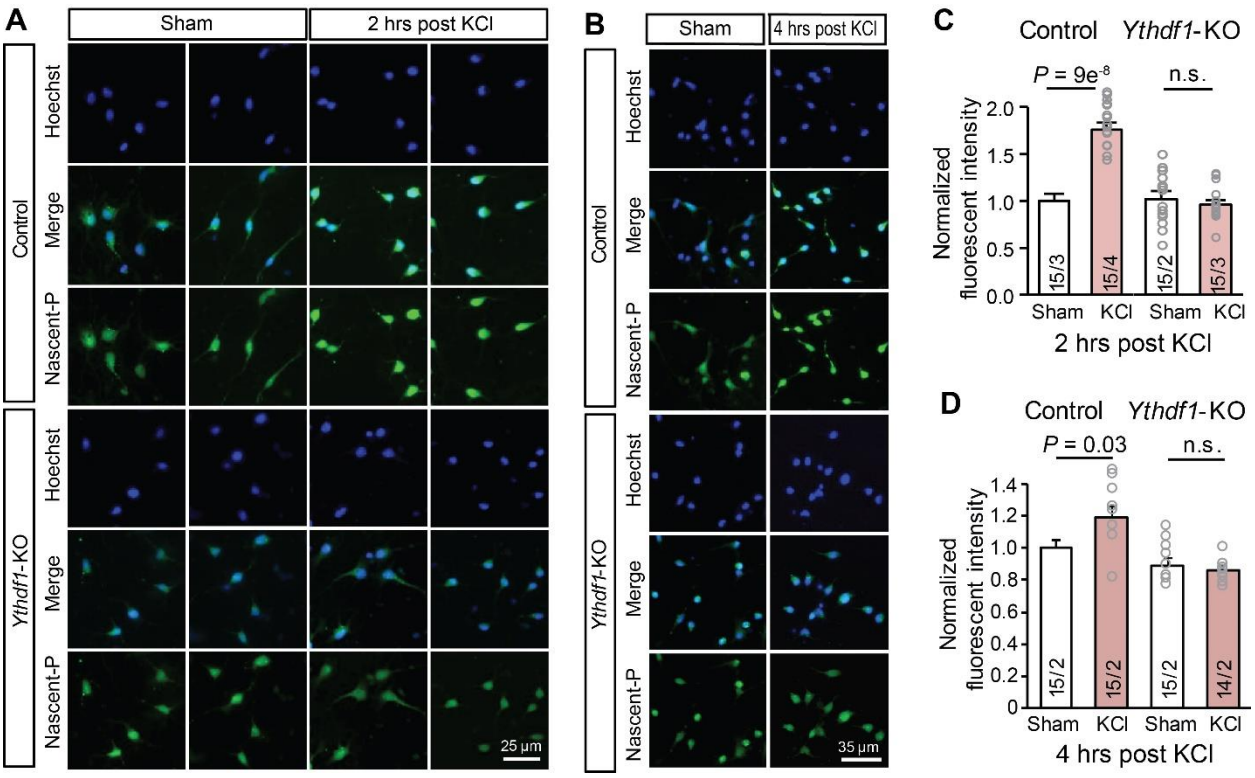

Figure 3.17 Nascent protein quantification in neurons before and after KCl depolarization

Representative images (**A-B**) and quantification (**C-D**) of nascent protein (Nascent-P) synthesis in cultured wild-type control and *Ythdf1*-KO hippocampal neurons before (sham) and 2 hours (**A&C**) or 4 hours (**B&D**) after KCl depolarization. *P* values, two-tailed *t*-test. Numbers in bars, numbers of images/biologically independent samples. Error bars, mean ± s.e.m.

We also constructed a reporter system in which the *N*-terminus of mouse YTHDF1 (YTHDF1-N) was tethered to the 3'UTR of the firefly luciferase (F-Luc) coding sequence, mimicking direct binding of YTHDF1 on m^{6}A-modified transcripts; *Renilla* luciferase was co-transfected for normalization (**Figure 3.18A**)[55]. YTHDF1-N tethering did not affect F-Luc protein level before KCl stimulation (**Figure 3.18B**), consistent with proteomics results (**Figure 3.16B**). However, increased F-Luc production was observed at 2 and 4 hours post KCl stimulus for YTHDF1-N tethering compared to the control (**Figure 3.18C-E**), supporting that YTHDF1 promotes protein synthesis upon neuronal stimulation.

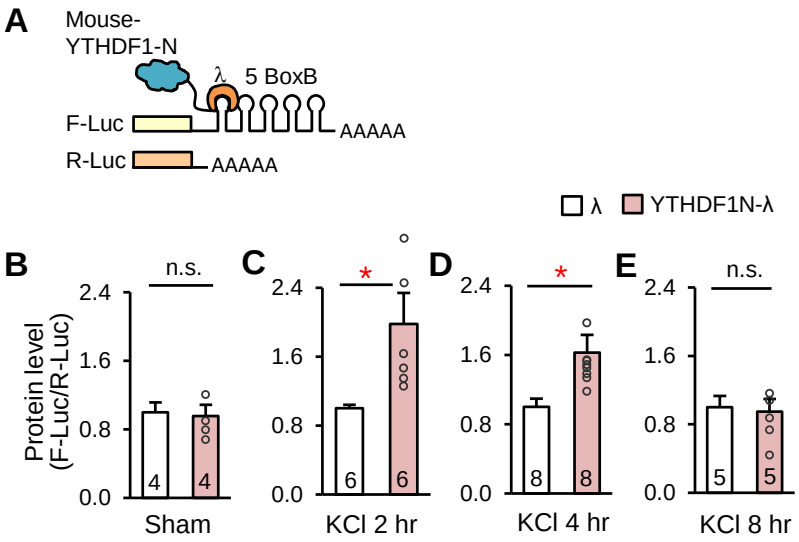

Figure 3.18 YTHDF1 tethering reporter assays in neurons before and after KCl stimuli

(**A**) Schematics of a tether reporter system that mimics the binding between YTHDF1 and 3'UTR m^6A sites of target transcripts. YTHDF1-N, truncated N-terminal mouse YTHDF1 (1-389 aa); F-Luc/R-Luc, firefly/*Renilla* luciferase. (**B-E**) Normalized F-Luc reporter expression in cultured hippocampal neurons tethered with YTHDF1-N or control, before (sham, **B**), 2 hours (**C**), 4 hours (**D**), and 8 hours (**E**) after KCl depolarization. *P* values, two-tailed *t*-test. Numbers in bars, biologically independent samples. Error bars, mean ± s.e.m.

Indeed, *Bsn*, one of the top YTHDF1-CLIP targets, showed attenuated protein expression after fear conditioning (FC) in the *Ythdf1*-KO hippocampus (**Figure 3.19A**), as did *Camk2a* in the *Ythdf1*-KO PSD fraction (**Figure 3.19B**).

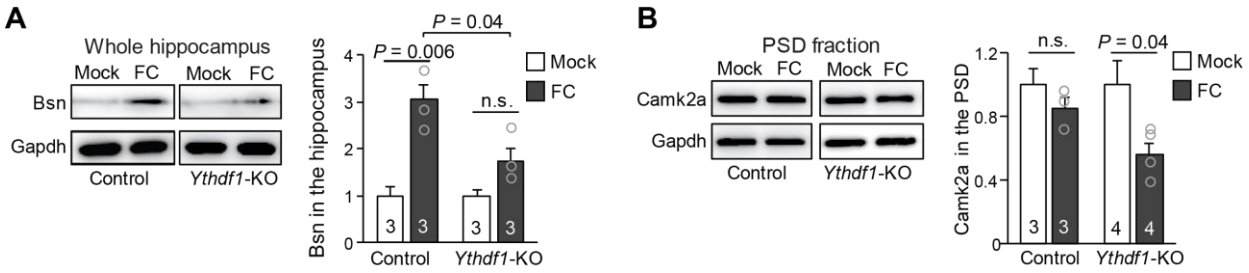

Figure 3.19 Western blots showing deficient protein production after FC in *Ythdf1*-KO mice

(**A-B**) Representative western blot (left) and quantification (right) of Bsn (**A**) and Camk2a (**B**) protein levels, in the whole hippocampus and the PSD fraction, respectively, before (Mock) and 2 hours after fear conditioning (FC). Normalized to the Mock condition for each genotype separately. *P* values, two-tailed *t*-test. Numbers in bars, numbers of replicates. Error bars, mean ± s.e.m.

3.2.3.3 Potential mechanisms underlying stimulus-dependent function of YTHDF1

Finally, we investigated the molecular mechanisms that might drive the observed stimulus-dependent function of YTHDF1. We examined how YTHDF1 and m^{6}A profile in the hippocampus response to neuronal stimuli. After fear conditioning, YTHDF1 protein increased by 30% in the PSD fraction, although no change occurred at the tissue level (**Figure 3.20A-B**). This suggests that YTHDF1 may undergo translocation to PSD in response to stimulus, which could contribute to localized translation in synapses and thus synaptic plasticity.

We also examined potential m^{6}A landscape changes of YTHDF1 targets in the dentate gyrus (DG) in an electroconvulsive treatment (ECT) model, in which dentate granule cells are synchronously activated[125,126]. The anti-m^{6}A-immunoprecipitation followed by sequencing (m^{6}A-RIP-seq) was applied to mRNAs purified from the DGs before and after ECT, and the ECT-induced changes in transcript abundance was evaluated in both the input libraries and the m^{6}A-antibody-enriched RIP libraries (**Figure 3.20C**). The m^{6}A-RIP-seq of DG mRNAs showed that while the level of YTHDF1-CLIP targets was not differentially regulated in response to ECT compared to other transcripts (**Figure 3.20D**), the m^{6}A-methylated copies of YTHDF1-CLIP targets were up-regulated in abundance after ECT (**Figure 3.20E**). This up-regulation effect was also observed, to a mild extent, for YTHDF1-CLIP targets and m^{6}A-modified transcripts in the naïve WT mouse hippocampus compared to *Ythdf1*-KO one (**Figure 3.16A**). With the observation that more YTHDF1 protein is present in the PSD fraction after stimulus (**Figure 3.20A-B**), these results suggest that more binding of YTHDF1 to its m^{6}A-methylated targets may occur upon stimulation to facilitate translation of the targets as well as to help stabilize them.

Briefly, stimulus-induced responses, including (i) synaptic enrichment of YTHDF1 protein (**Figure 3.20A-B**), (ii) up-regulation in the abundance of m^{6}A-methylated copies of YTHDF1 targets (**Figure 3.20D**), and (iii) perhaps a global increase of mRNA m^{6}A level, may together contribute to the stimulus-dependent feature of the function of YTHDF1.

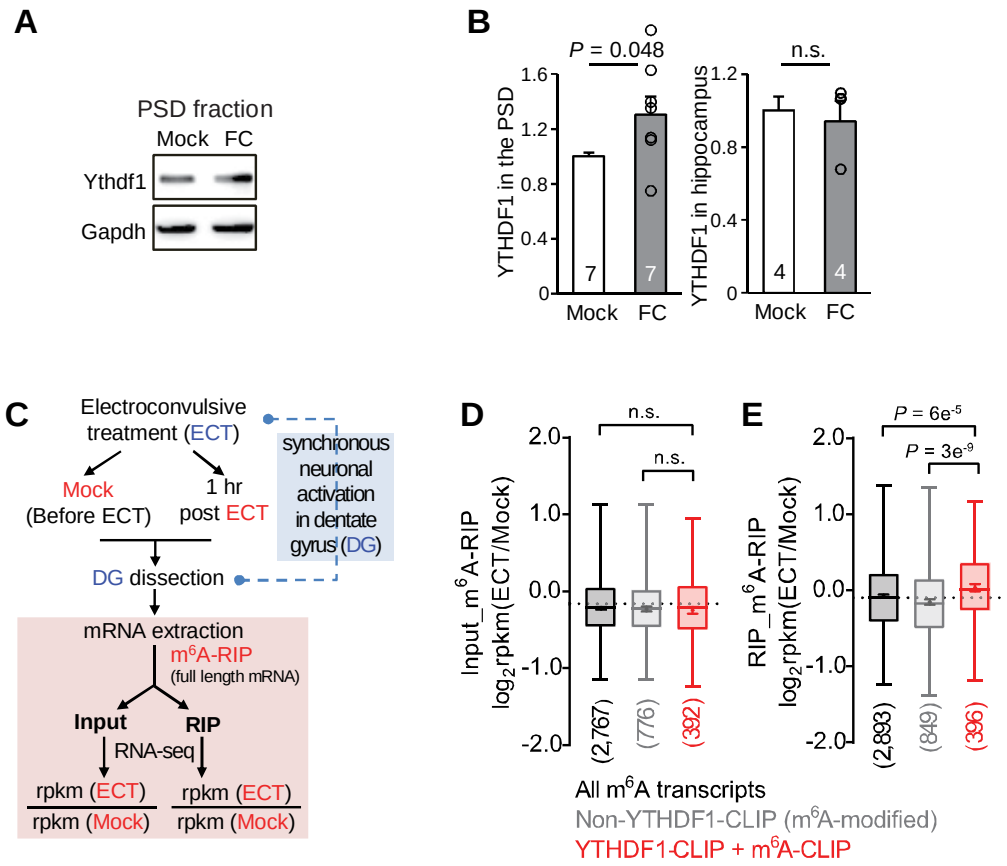

Figure 3.20 Potential mechanisms for neuronal-stimuli-dependent YTHDF1 functions

(**A-B**) Representative western blot (**A**) and quantification of YTHDF1 protein level in the hippocampal postsynaptic density (PSD) fraction (**B**, left) and the whole hippocampus (**B**, right) before (Mock) and 2 hours after fear conditioning (FC). *P* values, two-tailed *t*-test. Numbers in bars, numbers of biologically independent samples. Error bars, mean ± s.e.m. (**C**) Schemes of the experimental design to quantify the change in the extent of m^{6}A methylation for each transcript in the DG region before (Mock) and 1 hour after electroconvulsive treatment (ECT). (**D-E**) Box-plots of transcript abundance $\log_2$ fold change between electroconvulsive treated (ECT) and untreated (Mock) dentate gyrus, for m^{6}A-modified transcripts, m^{6}A-modified non-YTHDF1-CLIP transcripts, and transcripts with overlapped YTHDF1-CLIP and m^{6}A-CLIP peaks, in "Input" (**D**) and m^{6}A-enriched "RIP" (**E**) RNA-seq libraries. Dash lines, median $\log_2$ fold change of all reliably detected transcripts (rpkm > 1). Box-plot elements: center line, median; box limits, upper and lower quartiles; whiskers, 1-99%; number in parentheses, number of genes. *P* values, two-sided unpaired Kolmogorov-Smirnov test.

3.2.4 Ongoing research on YTHDF1 within other parts of the brain

3.2.4.1 m^{6}A and striatum-dependent learning in mice

In collaboration with our group, Koranda *et. al.* reported the role of METTL14 in striatal function and learning[127]. They deleted METTL14 in either striatonigral (dopamine D1 receptor (D1R) positive) or stratopallidal (dopamine D2 receptor (D2R) positive) neurons, which profoundly impaired striatal-mediated behaviors. Take striatonigral-specific *Mettl14*-knockout (D1R-fM14) as an example (littermate controls labelled as "D1Rctrl", KO labelled as "D1Rexp")). *Mettl14* deletion reduced striatal m^{6}A levels (**Figure 3.21A**). m^{6}A-seq and RNA-seq analyses of D1Rctrl and D1Rexp showed that the magnitude of change in RNA abundance positively correlated with the magnitude of decrease in m^{6}A peak enrichment[127].

Interestingly, functional annotation of m^{6}A-modified differentially expressed genes between the genotypes revealed downregulation of neuron- or synapse-specific genes (**Figure 3.21B**) and upregulation of non-neuron-specific genes, such as those related to metabolism, ribosome components, and translation. (**Figure 3.21C**).

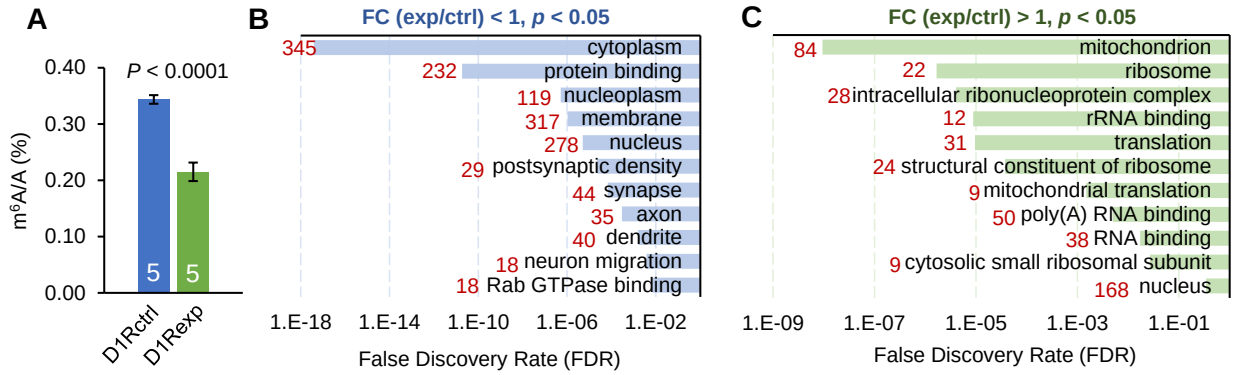

Figure 3.21 Differential expression analyses in D1R-fM14 striatum

(**A**) LC-MS/MS quantification of striatal polyA$^+$ RNA m^{6}A levels from control and D1R-fM14exp mice. Numbers in bars, numbers of mice. (**B-C**) GO analysis of m^{6}A-modified mRNAs with reads per kilobase of transcript, per million mapped reads (rpkm) fold change (FC) < 1 and p < 0.05 (**B**) and FC > 1 and p < 0.05 (**C**) in D1R-fM14exp compared to controls. Number of genes in each category shown in red.

These results suggest that YTHDF2's function on mRNA decay is unlikely to dominate for m^{6}A-modified neuron- and synapse-specific transcripts, since a decrease of m^{6}A level co-occurred with decreased transcript level. One possible explanation is that neuronal transcripts may undergo transport along the neurites as granules and local translation at synapses[128,129], necessitating specialized regulation different from those for non-neuron-specific or housekeeping transcripts. Our findings in the hippocampus suggested binding of YTHDF1 contributes to increased RNA abundance (**Figure 3.16A&20E**). mRNA decay and translation are intrinsically linked[130], and YTHDF1-mediated translation promotion may help stabilize those transcripts as well. It is also likely that there are m^{6}A readers yet to be determined that account for the observation in *Mettl14* KO mice.

The substantial heterogeneity of cell types in the brain causes great challenges in interpreting tissue-wide data. To examine the translation effects of YTHDF1 in a more refined neuronal cell types, we utilized the strategy of translating ribosome affinity purification (TRAP) where EGFP-tagged ribosomal subunit protein L10A is expressed under cell-type-specific promoters[131]. The *Ythdf1*-KO mouse line was crossed with B6;FVB-Tg(Drd1a-EGFP/Rpl10a)CP73Htz/J line, which incorporates the EGFP-L10a ribosomal fusion protein targeted to the transcription start site of the *D1R* (also called Drd1a) gene[132]. The crossed control and *Ythdf1*-KO mice will be treated with the selective D1R agonist SKF-81297, and polysomal RNAs that are being actively translated in D1R neurons will be purified through EGFP pull down. Combined with the list of m^{6}A-modified transcripts regulated by METTL14 in D1R neurons, we expect to achieve a transcriptome-wide characterization of YTHDF1 functions in distinct neuron types in response to stimuli.

3.2.4.2 m^{6}A and synaptosomal transcriptome

Around the same time when our study of YTHDF1 in hippocampal neurons was published, Merkurjev *et. al.* reported a first characterization of synaptic m^{6}A epitranscriptome[133]. They found that transcripts that are hypermethylated in the synaptosome compared to in the whole tissue exhibit strong functional enrichment in synaptic pathways. Disturbing *Ythdf1* expression by shRNAs in dissociated hippocampal neuronal cultures results in 20% yet not statistically significant decrease of *Apc* (adenomatous polyposis coli, m^{6}A-modified in the synapse) mRNA level and a dramatic decrease of APC protein level, consistent with our findings.

Taking advantage of the *Ythdf1*-KO mice model that we have, we tested how YTHDF1 depletion affects the cortical synaptic transcriptome profiles. Around 300 genes are upregulated in the *Ythdf1*-KO cortical synaptosome, related to mitochondrial function, ribosome, and cytoskeleton; while ~200 genes are down-regulated, showing functional enrichment for synaptic processes (**Figure 3.22**). Examining YTHDF1 at a higher resolution in either cell types or subcellular compartments holds the promise of an advanced understanding of m^{6}A biology.

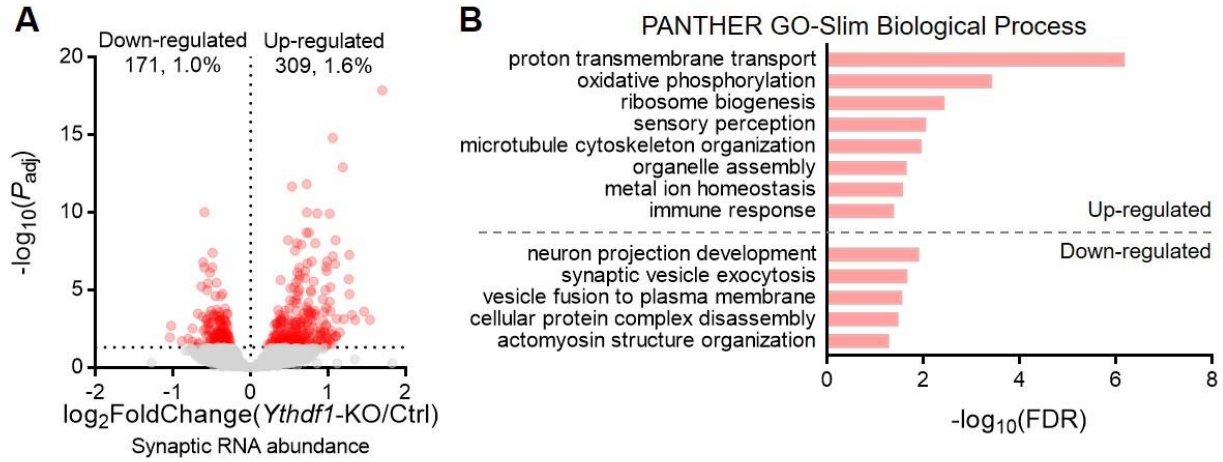

Figure 3.22 Differential expression analyses of *Ythdf1*-KO mouse cortical synaptosome

(**A**) Volcano plot showing differentially expressed genes in *Ythdf1*-KO mouse cortical synaptosome. The dots with an adjusted *P* value (P_{adj}) < 0.05 are highlighted in red. (**B**) Functional annotations of the upregulated transcripts (top) and the downregulated transcripts (FDR < 0.05).

3.3 Conclusion and Discussion

In summary, we show that mRNA m^{6}A methylation facilitates learning and memory formation in the mouse hippocampus mainly by promoting translation from target transcripts upon neuronal stimuli, mediated through its binding protein YTHDF1. Mice with genetic deletion of *Ythdf1* (*Ythdf1*-KO) exhibit learning and memory defects (**Figure 3.4-5**) as well as impaired hippocampal synaptic transmission (**Figure 3.6**) and long-term potentiation (**Figure 3.7**). YTHDF1 re-expression in the hippocampus of adult *Ythdf1*-KO mice rescues behavioral and synaptic defects (**Figure 3.9-10**), while hippocampus-specific acute knockdown of *Ythdf1* or *Mettl3*, the catalytic component of m^{6}A methyltransferase complex, recapitulates the hippocampal deficiency (**Figure 3.11-12**). Transcriptome-wide mapping of YTHDF1 binding sites and m^{6}A sites on hippocampal mRNAs uncovered key neuronal genes (**Figure 3.13-14**). Nascent protein labeling (**Figure 3.17**) and tether reporter assays (**Figure 3.18**) in hippocampal neurons revealed that YTHDF1 enhances protein synthesis in a neuronal-stimulus-dependent manner.

The presence of YTHDF1 could expedite new protein synthesis required for long-lasting changes of synapse plasticity and thus memory formation, whereas, in *Ythdf1*-KO mouse hippocampus, stimulus-dependent protein synthesis is attenuated, resulting in less efficient synaptic strengthening and a lower probability of reaching "thresholds" for memory formation (**Figure 3.23**). The m^{6}A-dependent translation promotion could be stimulation-induced as shown here for YTHDF1, potentially a general theme for RNA methylation-dependent translational regulation.

A recent paper discussed the stress-induced m^{6}A epitranscriptome change in the brain[134]. An acute restraint stress, where the mouse is kept for 15 min in a ventilated 50 mL falcon tube, is shown to induce m^{6}A changes that are brain-region-, time-, and transcript-dependent. Of note,

conditional knockout of *Mettl3* or *Fto* in the excitatory neurons in the neocortex and hippocampus both led to improved fear memory, indicating a rigorous regulation of m^{6}A dynamics during stress response and memory formation.

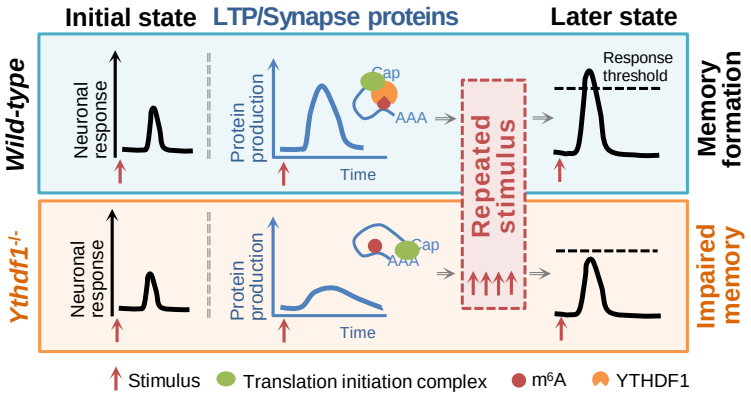

Figure 3.23 A proposed mechanism for how YTHDF1 contributes to memory formation

YTHDF1 promotes translation of m^{6}A-modified target transcripts, including synaptic transmission and LTP-related ones, in response to learning stimulus, thus facilitating synapse strength adequately for a memory to occur.

The *Ythdf1*-KO mice also show interesting phenotypes in immune response. Our group recently reported an increased CD8+ T cell antitumor response in *Ythdf1*-deficient mice[135]. YTHDF1 promotes the translation of lysosomal proteases transcripts it binds to, and loss of YTHDF1 contributes to a lasting antitumor response beneficial to PD-L1 checkpoint blockade immunotherapy. On-going research in the group about YTHDF1 also includes liver damage repair. A common theme in these processes is how cells and organisms achieve optimal response upon external stimuli. The functional significances of m^{6}A and YTHDF1 in these biological systems highlight RNA modifications as a mechanism underlying dynamic yet controllable cellular response.

3.4 Methods

3.4.1 Animals and cell lines

All animals were maintained under a 12/12-hour light/dark cycle with lights on at 7:00 AM, and temperature and humidity were kept at 22 ± 1 °C, 55% ± 5%, with *ad libitum* access to food and water. Male adult (8-16 weeks of age) mice were used for behavioral tests. Animal experiments, except for electroconvulsive treatment (ECT), were carried out in accordance with protocols approved by the Institutional Animal Care and Use Committee of the School of Life Science and Technology of Shanghaitech University and with the Guidance Suggestions for the Care and Use of Laboratory Animals, formulated by Ministry of Science & Technology of the People's Republic of China. Animal procedures used in ECT were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee of Johns Hopkins University School of Medicine and University of Pennsylvania School of Medicine.

The N2A cell line used in *in vitro* transfection experiments was purchased from Cell Bank of Chinese Academy of Sciences and authenticated by the supplier. It is not in the list of commonly misidentified cell lines maintained by the International Cell Line Authentication Committee (ICLAC). Cells were tested negative for mycoplasma contamination before use.

3.4.2 Generation of *Ythdf1*-KO mice

The YTH domain family protein-1 knockout mice (*Ythdf1*-KO) were generated based on CRISPR/Cas9. sgRNA expression plasmids were generated by annealing and cloning oligos that were designed to target Exon 4 of *Ythdf1* into the BsaI sites of pUC57-sgRNA (Addgene 51132).

mYthdf1-E4-1 T7 gRNA up: TAGGATAGTAACTGGACAGGTA;

mYthdf1-E4-1 gRNA down: AAACTACCTGTCCAGTTACTAT;

mYthdf1-E4-2 T7 gRNA up: TAGGCACCATGGTCCACTGCAG;

mYthdf1-E4-2 gRNA down: AAACCTGCAGTGGACCATGGTG.

In vitro transcription and microinjection of CRISPR/Cas9 were performed as previously described[120]. Briefly, Cas9 expression construct pST1374-Cas9-N-NLS-Flag-linker-D10A (Addgene 51130) was linearized with Age I and transcribed using the mMACHINE™ T7 Ultra Kit (Ambion, AM1345). Cas9 mRNA was purified by RNeasy Mini Kit (Qiagen, 74104). pUC57-sgRNA expression vectors were linearized by Dra I and transcribed using the MEGAshortscript Kit (Ambion, AM1354). sgRNAs were purified by MEGAclear Kit (Ambion, AM1908). A mixture of Cas9 mRNA (20 ng/μl) and two sgRNAs (5 ng/μl each) was injected into cytoplasm and male pronucleus of zygotes obtained by mating of CBF1. Injected zygotes were transferred into pseudo-pregnant CD1 female mice. Founder mice used for experiments were backcrossed to C57BL/6 for at least five generations. *Ythdf1*-KO mice used for the experiments were sacrificed at age of 8-16 weeks and did not show obvious development defects before sacrifice.

mYthdf1-E4 C9 For: CACCTGAGTTCAGATCATTAC;

mYthdf1-E4 C9 Rev: GCTCCAGACTGTTCATCC.

Amplicon length: 650 bp. Applicable to genotyping founders and targeted ESC.

3.4.3 Genotyping

Mice were weaned at third postnatal week and genotyped by PCR. *Ythdf1*-KO and wild-type alleles were detected by PCR assays in which primer F1 (5'- GTGTATGAGGTGGTCAG CAT-3') and primer R1 (5'- CTTGTTGAGGGAGTCACTGT-3') amplified a 465-bp fragment (wild-type) and a 286-bp fragment (*Ythdf1*-KO) (**Figure 3.1D**).

3.4.4 Mouse behavioral tests

Open-field test. Mice were exposed to a square open arena (40 cm × 40 cm) with opaque base and walls (40 cm high). Each mouse was allowed 30 min to explore the area and its activity

was recorded and analyzed using the Tru Scan Activity System (Coulbourn Instruments). The surface was cleaned with 70% ethanol after each mouse was tested.

Elevated-plus maze. Elevated-plus maze apparatus consists of two open arms (50 cm × 9 cm), two enclosed arms (50 cm × 9 cm × 39 cm) and a central area (9 cm × 9 cm). The maze is elevated 70 cm above ground in a room with normal light. Mice were placed in the central area individually and allowed 5 min to explore the maze. The time each mouse spent in the open arms during the 5-min exploration was counted by Anymaze software.

Light-dark box transition test. The light and the dark compartments of the light-dark transition box (35 cm × 35 cm × 40 cm) were separated by an opaque plexiglass board with a hole. The light compartment was illuminated by strong light (400 lux). During the test, mice were individually placed at the center of the light compartment facing away from the hole and allowed 30 min to explore freely in the box. Activity of each mice was monitored. The time mice spent in the light compartment as well as the number of transitions between the two compartments were automatically calculated by Tru Scan Activity System (Coulbourn Instruments).

Tail-suspension test. Tail-suspension test was used to assess behavioral despair of mice. Each mouse was suspended by its tail with an adhesive tape for 6 min and was video recorded. Total immobility time during the test was scored by independent observers. Mice were considered immobile only when hung passively and motionlessly for at least 2 seconds.

Morris water maze task. Morris water maze test was specifically designed to evaluate spatial reference memory abilities (**Figure 3.4A**)[122]. The Morris water-tank consists of a circular pool (diameter 120 cm, height 50 cm) filled with water maintained at room temperature (23 ± 1ºC) and is made opaque with nontoxic white paint. The pool is located in an experimental room with many extra-maze visual cues and virtually divided into four equal quadrants. A circular platform,

10 cm in diameter, is placed in the middle of one fixed quadrant ("target") of the pool, just above water surface (visible platform) or 1 cm underneath the water surface (hidden platform). For visible platform training, mice were given a training of four trials with 30 min inter-trial intervals each day for two consecutive days, and they were released from each starting point in a random order. For hidden platform training, mice were given a training of four trials each day for five consecutive days. 24 hours after last trial of training (day 6), the platform was removed, and all mice were given one probe trial for 60 sec searching (probe test). The escape latency to visible or hidden platform and the exploring time in each quadrant of the pool were automatically recorded by water maze system (Coulbourn, Inc.). Mice were trained at the same time of a day during their light phase.

Contextual and auditory fear conditioning. The fear conditioning test was performed as previously described (**Figure 3.5A**)[123]. Mice were first handled for 5 min each day for three consecutive days and habituated to the conditioning chamber for 5 min the day before training. On the day of training, after 3 min exploration in the conditioning chamber, each mouse received one pairing of a tone (2,800 Hz, 75 dB, 30 s) with a short co-terminating foot shock (0.5 mA, 1 s) for the weak training protocol, a long foot shock (0.5 mA, 2 s) for the moderate protocol, or three pairings of a tone (2,800 Hz, 75 dB, 30 s) with a long co-terminating foot shock (0.5 mA, 2 s) for the strong protocol, after which they were remained in the chamber for additional 30 s and then returned to home cages. 2 hours and 24 hours after the conditioning, mice were tested for freezing (behavioral immobility) in response to the training context (training chamber) and to the tone (in the training chamber with a new environment and odor). The percentage of freezing time was calculated as an index of fear learning and memory. For contextual fear memory tests, mice were returned to the conditioning chamber for 3 min and freezing behavior was counted using StartFear

Combined system (Panlab). For auditory fear memory tests, mice were placed in a changed chamber and freezing responses were recorded during the last 3 min when the tone was delivered. Tests of contextual and auditory fear memory were done in a counterbalanced manner.

3.4.5 Electrophysiological recording of hippocampal slices

Extracellular field recordings and whole-cell miniature excitatory postsynaptic current (mEPSC) recordings in the hippocampal CA1 region were conducted in 380 μm thick acute brain slices from 6-9-week-old wild-type control and *Ythdf1*-KO mice of either sex. Coronal sections that contain hippocampal formations were prepared according to the previously described method[136]. Briefly, mice were anaesthetized with sodium pentobarbital and were sacrificed by decapitation. Transverse slices of the hippocampus (380 μm) were cut using the vibratome at 4 °C in a modified artificial cerebrospinal fluid (mACSF) consisting of 110 mM Choline chloride, 2.5 mM KCl, 0.5 mM $CaCl_2$, 7 mM $MgSO_4$, 25 mM $NaHCO_3$, 1.25 mM NaH_2PO_4, 25 mM D-glucose, and 3.1 mM sodium pyruvate, which was saturated with 95% O_2 and 5% CO_2. Slices were transferred to an incubating chamber with oxygenated (95% O_2 and 5% CO_2) normal ACSF containing 120 mM NaCl, 2.5 mM KCl, 2.5 mM $CaCl_2$, 1.3 mM $MgSO_4$, 26 mM $NaHCO_3$, 1 mM NaH_2PO_4, 10 mM D-glucose (pH 7.3–7.4) and incubated at 30 °C for at least 2 hours before recording. Data were collected with a MultiClamp 700B (Molecular Devices), digitized using Digidata 1440A and pClamp 10.1 data acquisition system (Molecular Devices). Frequency, duration and magnitude of extracellular stimulus were controlled with a Master 8 pulse stimulator (A-M Systems). Evoked synaptic responses were triggered with a bipolar electrode.

LTP recording. To record the extracellular field excitatory postsynaptic potentials (fEPSPs), a glass micro-electrode (4–8 MΩ, filled with 0.5 M sodium acetate) was placed in the stratum radiatum of the CA1 region, and a bipolar tungsten stimulating electrode was placed along

the Schaffer collateral fibers and was 100-150 μm away from the recording pipette. Intensity of the stimulation was adjusted to produce an fEPSP with an amplitude of 30-40% of the maximum response. Test stimulation was delivered once per 30 seconds (0.033 Hz) or per minute (0.017 Hz). After recording a stable baseline for at least 30 min, early-phase LTP or later-phase LTP was induced by two (100 Hz for 1 s, 30 s interval) or four trains (100 Hz for 1 s, 5 min interval) high-frequency stimulation (HFS), respectively. Magnitudes of LTP and L-LTP were calculated based on the averaged fEPSP values during the last 10 min and 30 min of summary plots, respectively.

PPF recording. For paired-pulse facilitation (PPF) recording, a second stimulus was delivered following the first one with different time intervals. The two stimuli were separated by 20, 40, 60, 80, 100, 200, 400, 600, 800, and 1,000 ms interval. The amplitude of the population response to the second stimulus was compared with that to the first one to obtain the PPF ratio.

Miniature EPSCs. Voltage clamp recordings were obtained from neurons in hippocampus slices equilibrated for at least 1 hour in the recording chamber. Micropipettes (tip diameter: 1.5 to 2.0 μm; resistance: 4 to 6 MΩ) were pulled from borosilicate capillaries (P-97; Sutter Instruments) and filled with an internal solution (adjusted to pH 7.2) composed of 145 mM potassium gluconate, 5 mM NaCl, 1 mM $MgCl_2$, 0.2 mM EGTA, 10 mM HEPES, 2 mM magnesium adenosine triphosphate, 0.1 mM sodium-guanosine-5'-triphosphate, and 10 mM phosphocreatine disodium. For mEPSC recordings, 0.5 μM tetrodotoxin was added to the perfusion solution. Picrotoxin (100 μM) was present in all of experiments to block γ-aminobutyric acid (GABA) type A receptor–mediated inhibitory synaptic currents. Neuron signals, which were amplified using an Axoclamp-700B amplifier (bandwidth filter set at 1 kHz for voltage clamp recordings) were digitized (Digidata 1440A, pClamp 10.1; Molecular Devices). The series resistance was 12 to 20 MΩ and

was monitored throughout the experiment. Data were discarded when access resistance changed by more than 15% during the experiment.

3.4.6 Plasmid constructs and viruses

For reporter assay, pPB-CAG-Flag-YTHDF1-N-λ and pPB-CAG-Flag-λ were constructed by inserting the Flag- YTHDF1-N-λ (YTHDF1-N: N-terminus of mouse YTHDF1, 1-389 aa) and Flag-λ fragments into pPB-CAG backbone vector between BglII and XhoI restriction sites, respectively. For AAV vectors, pAAV-CMV-mouse-YTHDF1-2a-mCherry-WPRE, pAAV-CMV-mCherry-WPRE, pAAV-CMV-RFP-U6-Ythdf1-shRNA, and pAAV-CAG-eGFP-H1-Mettl3-shRNA were all designed and constructed by standard methods. The following oligonucleotide sequences were used for knockdown:

Ythdf1-shRNA: 5'-GATCCTTACCTGTCCAGTTAC-3';

Mettl3-shRNA: 5'-GCACACTGATGAATCTTTAGG-3';

Scramble control: 5'-AACAGTCGCGTTTGCGACTGG-3'.

AAV viruses were prepared by Taitool Biotech (Shanghai).

3.4.7 In vivo stereotactic injections

For viral injection, male mice (8-10-weeks of age) were anesthetized with 5% chloral hydrate (100 μl/10 g body weight) by intra-peritoneal (i.p.) injection and placed on a stereotaxic apparatus. Small bilateral holes were drilled into the skull at -1.7 mm posterior and -1.5 mm lateral to bregma for injections into the hippocampal CA1 and dentate gyrus (DG) regions. Glass cannula filled with a virus solution was lowered to CA1 (-1.5 mm) and DG (-2.0 mm), and the virus solution (0.6 μl) were injected using a Nanoject II (Drummond) system at a rate of 0.1 μl per min sequentially into each side of hippocampus. Injection cannula was slowly withdrawn 5 min after the virus infusion. The scalp was then sealed, and injected mice were monitored as they recovered

from anesthesia. Behavioral experiments or electrophysiological recordings were performed at least 10 days after virus injection. Virus infection was examined at the end of all the behavioral tests.

3.4.8 Immunohistochemistry

Ythdf1-KO and wild-type male mice (from P28 to 16-week old) were perfused with phosphate-buffered saline (PBS) followed by 4% paraformaldehyde in PBS. After post-fixation in 4% PFA for 12 hours at 4 °C and dehydration in 30% sucrose-PBS solution for another 24 hours, the brains were frozen-sectioned into coronal slices (35 μm) for next step use. For anti-YTHDF1 and anti-DCX staining, slices were incubated in diluted antibody solution at 4 °C overnight then detected by Alexa Fluor conjugated second antibodies. All slices were counterstained with Hoechst in the final step incubation. Fluorescent image acquisition was performed using Axioimager Z2 microscope or LSM 510 confocal microscope (Zeiss). Images were analyzed with Image-Pro Plus and ImageJ software. Brain slices from mice injected with AAV-CMV-mouse-YTHDF1-2a-mCherry-WPRE (AAV-YTHDF1) and AAV-CMV-mCherry-WPRE (AAV-control) were used for YTHDF1 over-expression quantification assay (**Figure 3.8**).

3.4.9 Western blot

Samples were homogenized in RIPA buffer (Beyotime) containing 1 mM PMSF, 1×protease inhibitor cocktail and 1×phosphatase inhibitor cocktail (Sigma). Lysates were boiled at 100 °C with 6×loading buffer (Beyotime) for 8 min and then stored at -80 °C for next step use. A total of 30 μg protein per sample was resolved on SDS-PAGE (10%) at 80 V for 20 min and then 110 V for 110 min. Proteins on the gel were transferred onto PVDF membranes (Millipore) and blocked in 5% milk blocking solution for 1 hour at room temperature, incubated in a diluted primary antibody solution at 4 °C overnight, and incubated in a dilution of secondary antibody

conjugated to HRP for 2 hours at room temperature (dilution folds indicated in the section of antibodies). Protein bands were detected using ECL western blotting detection reagents (Millipore) and Amersham Imager 600 system (GE).

3.4.10 Dissociated neuron culture and tethering reporter assays

Hippocampal neurons from E18 C57BL/6 mice embryos of either sex were cultured at a density of 200,000 cells per well on poly-D-lysine pre-coated 6-well plates. Neuron cultures were maintained in complete medium (Neurobasal medium supplemented with 0.5mM GlutaMAX-I and 2% B-27). Plasmid transfection was conducted using 4D-Nucleofector System (Lonza) immediately after neuron dissociation.

The reporter plasmid (pmirGlo-5BoxB) and the effecter plasmid (Flag-λ, or Flag-mouse-YTHDF1-N-λ in pPB-CAG vector) were used to transfect neuron cultures at a ratio of 1:9 as previously reported[55]. After transfection, neurons were plated in plating medium (Neurobasal medium supplied with 0.5 mM GlutaMAX-I, 2% B-27 and 5% FBS) for 6 hours, then changed to complete medium for further culturing. 3 days after transfection, neurons were treated with KCl at the final concentration of 50 mM for 2, 4, and 8 hours. Then neurons were collected and assayed by Dual-Glo Luciferase Assay Systems (Promega) to test protein production.

3.4.11 Nascent protein synthesis assay

WT control and *Ythdf1*-KO mouse hippocampal neurons were cultured on pre-coated glass cover slides. Twelve days later, protein synthesis assay was conducted using Click-iT Plus OPP Alexa Fluor™ 488 Protein Synthesis Assay Kit (Invitrogen, C10456) following the manufacturer's protocol. Briefly, the neurons were treated with 50 mM KCl for 10 min before the complete culture medium was changed back. Click-iT OPP (Component A) was diluted 1:1000 in pre-warmed culture medium as a 20 μM working solution. 2 hours or 4 hours after the KCl treatment, the

culture medium was replaced with the working solution for another 30-min incubation under culturing conditions. The medium was then removed, and the neurons were washed once with PBS before being fixed and permeabilized with 4% PFA and 0.5% Triton X-100 (in PBS) for 15 min at room temperature, respectively. After two more rounds of wash with PBS, the neurons were incubated with a freshly prepared Click reaction cocktail for 30 min at room temperature in the dark and rinsed once with the reaction rinse buffer. Finally, neurons were counter stained with NuclearMask Blue Staining working solution and washed twice with PBS prior to imaging and analysis.

3.4.12 Quantitative mass spectrometry

Hippocampal samples from WT control and *Ythdf1*-KO mice (8-12 weeks, male) were freshly isolated and snap frozen in liquid nitrogen. For each mouse, 100 mg tissues were used for further preparation. Tissues were ground into fine powder in liquid nitrogen then lysed with 500 µl freshly prepared lysis buffer (20 mM triethylammonium bicarbonate (TEAB, pH 8.5), 8 M urea, protease inhibitor cocktail (Roche), and 1 mM DTT). The yielded lysate was treated with ultrasonication at 4 °C for 30 s to shear DNA, followed by centrifuge at 16000×g for 10 min at 4 °C. The resultant supernatant was carefully separated and transferred into a new tube. Protein concentrations were measured using BCA Protein Assay Kit (Thermo Scientific). For each condition, a total of 100 µg protein was reduced with 5 mM tris(2-carboxyethyl)phosphine (TCEP) for 3 hours at 30 °C, then alkylated with 10 mM methyl methanethiosulfonate (MMTS) for 45 min at room temperature (protected from light). Samples were then diluted with 20 mM TEAB to obtain a final concentration of 1 M urea prior to digestion with 2.5 µg trypsin overnight at 37 °C. Resultant tryptic peptides were finally labeled with TMT10plex Mass Tag Labeling Kit (Thermo Scientific) according to manufacturer's protocol and followed with LC−MS/MS analysis.

3.4.13 PSD preparation

Hippocampal tissues from WT and *Ythdf1*-KO mice (8-12 weeks, male) were isolated, snap frozen in liquid nitrogen, and stored at -80 °C before use. PSD fraction preparation was prepared as previously described[136], hippocampal tissues were homogenized in homogenization buffer (320 mM sucrose, 5 mM sodium pyrophosphate, 1 mM EDTA, 10 mM HEPES (pH 7.4), 1×protease inhibitor cocktail, and 1×phosphatase inhibitor cocktail (Sigma)). Resultant homogenate was centrifuged at 800 × g for 10 min at 4 °C to yield post-nuclear pelleted fraction 1 (P1) and supernatant fraction 1 (S1). S1 was further centrifuged at 15,000×g for 20 min at 4 °C. Then pellet P2 (which contains synaptosome) was resuspended in 4 mM HEPES (pH 7.4) and incubated with agitation at 4 °C for 30 min. Suspended P2 was centrifuged at 25,000 × g for 20 min at 4 °C. Resulting pellet was resuspended in 50 mM HEPES (pH 7.4), mixed with an equal volume of 1% Triton X-100, and incubated with agitation at 4 °C for 15 min. PSD fraction was generated by centrifugation at 32,000 × g for 20 min at 4 °C. Final PSD pellet was resuspended in 50 mM HEPES followed by protein quantification and then boiled with 6×loading buffer for western blot.

3.4.14 Lucifer yellow labelling by intracellular injection

WT control and *Ythdf1*-KO mice (8-12 weeks, male) were perfused with 4% PFA in PBS and their brains were removed to perform the intracellular injection of the fluorescent dye Lucifer Yellow (LY). The brains were post-fixed for 24 h in 4% PFA in PBS, and coronal sections were obtained (200 µm). Intracellular injections were performed as previously described[137]. In brief, sections were placed under differential interference contract (DIC) microscope to find health CA1 neurons and a continuous current (5-10 nA) was used to inject cells with LY. At least five CA1 pyramidal cells per mouse were injected individually with LY, with the current applied till the distal tips of each dendrite fluoresced brightly (5-10 min). Images (z-stacks) for spine density

counting were acquired using LSM 510 confocal microscope with a 63X oil objective. Spine counting and spine morphology analyses were performed using Neurostudio software.

3.4.15 YTHDF1-CLIP-seq in mouse hippocampus

15 µg rabbit anti-YTHDF1 antibody (Proteintech, 17479-1-AP), 50 µl Protein A beads (ThermoFisher), 50 µl Protein G beads (ThermoFisher), and four pairs of hippocampi from C57BL/6 mice (9 weeks, male) were used for each biological replicate. Three biological replicates were performed.

For each replicate, Protein A/G beads were washed three times with PBST (PBS with 0.05% Tween-20) and resuspended with anti-YTHDF1 antibody in 250 µl PBST prior to overnight rotating at 4 °C. On the following day, mouse hippocampal tissues were dissected, homogenized in 500 µl HBSS buffer, and crosslinked in 6-well plates on ice for four times with 254 nm UV light, 0.15 J/cm^2 per time (UV stratalinker 2400, Stratagene). 1.5 ml lysis buffer (150 mM NaCl, 0.5% NP-40, 50 mM Tris-HCl (pH 7.5), 2 mM EDTA, 1% protease inhibitor cocktail (Roche), 0.5 mM DTT) was added to the crosslinked tissue pellet for a 40-min rotating at 4 °C. After being cleared by maximum-speed centrifuge, the lysate underwent a first round of RNA digestion with 0.2 U/µl RNase T1 (ThermoFisher, EN0541) for 15 min at room temperature followed by a five-min quenching on ice. A 100 µl aliquot of the resultant lysate was saved as "Input", and the remaining lysate was incubated with the Protein A/G beads conjugated with anti-YTHDF1 antibody. After three hours of rotating at 4°C, the beads were washed three times with 1 ml IP wash buffer (50 mM Tris-HCl (pH 7.5), 300 mM KCl, 0.05% NP-40, 0.5 mM DTT, 1% protease inhibitor cocktail), and resuspend in 200 µl IP wash buffer supplemented with 10 U/µl RNase T1 for a second round of RNA digestion for 8 min at room temperature. The previously saved "Input" was digested in parallel, and immediately supplemented with 4× Laemmli sample buffer (Bio-

Rad). After a 5-min quenching on ice, the beads were washed three times with high-salt wash buffer (50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 0.05% NP-40, 0.5 mM DTT, 1% protease inhibitor cocktail, 1% SUPERase• In™) and another three washes with PNK buffer (50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$). The RNA fragments co-immunoprecipitated with Protein A/G beads ("CLIP") were subject to end-repair by: (1) 1 U/µl T4 PNK (ThermoFisher, EK0031) in 100 µl 1× PNK buffer A (ThermoFisher) at 37°C for 20 min with vigorous shaking; and then (2) 1 mM ATP (final concentration) with another 0.5 U/µl T4 PNK at 37°C for another 20 min. The beads were washed with PNK buffer for another five times and then resuspended in 100 µl 2× Laemmli sample buffer. The YTHDF1-RNA complex was size-selected by SDS-PAGE (size indicated in Extended Data Fig. 8b), and the gel slice at the same molecular weight was cut for "Input" samples in parallel. To extract RNA, the gel slices were mashed and digested with 2 mg/ml protease K (ThermoFisher, RNA-grade, 25530049) at 55°C for 1 hour. Then gel particles were filter out, and the RNA were purified by Acid-Phenol:Chloroform extraction and overnight ethanol precipitation. The "Input" RNA fragments were end-repaired using T4 PNK and further cleaned up using RNA Clean & Concentrator-5 (Zymo Research). RNA libraries were generated using NEBNext multiplex small RNA library preparation kit (NEB, E7300S) for both "Input" and "CLIP" samples.

3.4.16 m^{6}A-CLIP-seq in mouse hippocampus

RNA extraction: Total RNA was extracted from hippocampal tissue dissected from C57BL/6 mice (8-16 weeks, male) using Trizol (Invitrogen) and isopropanol precipitation. Poly(A)+ RNA was purified using Dynabeads™ mRNA DIRECT™ Purification Kit (Invitrogen) following the manufacturer's instructions. For the m^{6}A-CLIP-seq, we followed the protocol reported[79] with a smaller amount of starting material: 300 ng purified poly(A)+ RNA, with 2.5 µg

anti-m^{6}A antibody (Synaptic System, 202003) and 25 µl Protein A/G beads. Three biological replicates were performed, and the pair of hippocampi from one mouse were pooled for each replicate.

3.4.17 RNA-seq of mouse hippocampus

Poly(A)+ RNA from WT littermate control and *Ythdf1*-KO mouse (8-16 weeks, male) hippocampus was purified as described in **3.4.16**. The RNA libraries were prepared using Truseq stranded mRNA sample preparation kit (Illumina) according to manufacturer's protocol. Three biological replicates were performed for each genotype, and two hippocampi from one mouse were pooled for each replicate.

3.4.18 ECT and m^{6}A-RIP-seq of dentate gyrus

Adult male, 6~8-week old C57BL/6 mice were used (Charles River) and housed in a standard facility. Electroconvulsive treatment (ECT) was achieved with pulses consisted of 1.0 s, 100 Hz, 16-18 mA stimulus of 0.3 ms delivered using the Ugo Basile ECT unit (Model 57800) as previously described[138]. Mock mice were handled in parallel without the electrical current delivery.

m^{6}A-RIP-seq. Total RNA from adult mouse DG was isolated using the TRIzol reagent according to the manufacturer's instructions (Invitrogen). mRNA purification was performed with poly(A)+ RNA selection twice using Dynabeads Oligo (dT)25 (Thermo Fisher; 61006). A total of 150 ng of mRNA was subjected to m^{6}A-SMART-seq using anti-m^{6}A rabbit polyclonal antibody (Synaptic Systems, 202003) as previously described[93]. Briefly, 5 µg of anti-m^{6}A polyclonal antibody was conjugated to Dynabeads Protein A (Thermo Fisher; 10001D) and used for each affinity pull-down. The m^{6}A RNA was eluted twice with 6.7 mM *N*^{6}-methyladenosine (Sigma-Aldrich; M2780) in 1 × IP buffer (10 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 0.1% (vol/vol) Igepal CA-630) and recovered by RNA Clean and Concentrator-5 (Zymo Research). Libraries

were generated using the SMART-seq protocol as described[139]. Three biological replicates for each condition were sequenced using Illumina NextSeq 550 from a single end for 75 bases.

3.4.19 Sequencing data analysis

General processing (for all sequencing samples unless specified). Sequencing was carried out on NextSeq500 with single end 80-bp read length or NextSeq550 with single end 75-bp according to manufacturer's instructions. Sequencing data were mapped to mouse genome version mm10 downloaded from UCSC using Tophat v2.0.14[140]. For RNA-seq analysis, RPKM were calculated by Cuffnorm[141]. For CLIP-seq experiments, after removing the adapter by Cutadapt[142], the reads were aligned to the mouse genome (mm10) by Bowtie 2[143].

Peak calling in YTHDF1-CLIP-seq. All mapped reads were treated as background and mutations were treated as signals for peak calling. PARalyzer[108] was use for peak calling in CLIP-seq sample with a few modifications: (1) mutations in both "CLIP" and "Input" were removed from "CLIP"; (2) sites with 100% mutations rate were also removed. The remaining mutations were used for peak calling. At least two mutation sites are need in each peak (MINIMUM_CONVERSION_LOCATIONS_FOR_CLUSTER =2).

Peak calling in m^{6}A-CLIP-seq. We followed the same peak calling method reported previously[79]. Consensus motif was determined using HOMER[109] for the m^{6}A-CLIP peaks identified in each replicate.

m^{6}A-RIP-seq analysis. Low-quality bases and adaptor sequences from original reads were removed using Trimmomatic[144]. The remaining reads were then mapped to the mouse genome (mm10) using STAR aligner[145]. Mapped reads between samples were normalized using DESeq2[146]. The "Input" and "RIP" libraries were normalized respectively.

Integrative analysis. (1) Definitions for groups of transcripts (**Figure 3.15-16&20**): (i) The common peaks (peaks from replicate 1 with > 1 nt overlap in peak location with those from both replicate 2 and 3) of CLIP-seq are defined as high-confidence CLIP peaks; (ii) Transcripts with high-confidence YTHDF1-CLIP peaks are defined as "YTHDF1-CLIP targets"; (iii) Transcripts without YTHDF1-CLIP peaks in any of the three YTHDF1-CLIP-seq replicates are defined as "non-YTHDF1-CLIP targets", and they were used as a control group for analyzing gene expression change in absence of YTHDF1; (iv) Transcripts with high-confidence m^6A-CLIP peaks are defined as "m^6A-modified transcripts"; (v) Transcripts with overlapped high-confidence YTHDF1-CLIP peaks and high-confidence m^6A-CLIP peaks (> 1 nt in peak location) are defined as "YTHDF1-CLIP + m^6A-CLIP" transcripts. (2) Functional annotation of a list of genes were generated by DAVID[111,112], for YTHDF1-CLIP targets (**Figure 3.13E**) and m^6A-modified transcripts with no less than five mutations in m^6A-CLIP-seq (**Figure 3.14D**). (3) Only genes with sufficient expression (rpkm > 1 in RNA-seq of WT triplicates; rpkm > 1 in m^6A-RIP-seq input or RIP libraries respectively) were kept and subject to further analyses. The median rpkm value of the sequencing triplicates was used for differential analyses.

High-throughput sequencing data can be accessed under GSE106607.

3.4.20 Antibodies

The antibodies used in this study are listed below in the format of name (application; catalog; supplier; dilution fold): Rabbit anti-Ythdf1 (western blot, 17479-1-AP, Proteintech, 500-1,000; IF, 200). Rabbit anti-Ythdf2 (western blot, 24744-1-AP, Proteintech, 500). Rabbit anti-Ythdf3 (western blot, 25537-1-AP, Proteintech, 500). Rabbit anti-Ythdc1 (western blot, 14392-1-AP, Proteintech, 500). Rabbit anti-Ythdc2 (western blot, ab176846, Abcam, 1,000). Rabbit anti-Mettl3 (western blot, ab195352, Abcam, 1,000). Mouse anti-GAPDH (western blot, G8795, Sigma,

3,000). Rabbit anti-DCX (IF, ab18723, Abcam, 1,000). Mouse anti-Actin (western blot, A4700, Sigma, 1,000). Rabbit anti-GluR1 (western blot, AB1504, Merck, 1,000). Rabbit anti-CaMKII (western blot, 4436S, Cell Signaling, 1,000). Rabbit anti-Grin1 (western blot, 5704S, Cell Signaling, 1,000). Rabbit anti-Grin2a (western blot, 4205S, Cell Signaling, 1,000). Mouse anti-Bsn (western blot, ab82958, Abcam, 500). Goat anti-Mouse IgG HRP conjugated (western blot, AP308P, Merck, 5,000). Goat anti-Rabbit IgG HRP conjugated (western blot, AP307P, Merck, 5,000). Alexa Fluor 488 Goat anti-Mouse IgG (IF, A11029, ThermoFisher, 1,000). Biotin-SP-conjugated Goat Anti-Rabbit IgG (IF, 111-065-003, Jackson ImmunoResearch, 500). Cy2-conjugated Streptavidin (IF, 016-220-084, Jackson ImmunoResearch, 1,000).

Chapter 4

An expanding spectrum of m^{6}A reader proteins

4.1 Introduction: complexity in the binding modes of RBPs

The RNA recognition modes of RNA binding proteins (RBPs) are intrinsically complicated. While many RBPs bind motifs with similar primary sequences, they may differ in preferences for certain structural features (such as loops, stems, and bulged stems), flanking base composition, and spaced "bipartite" motifs depending on the RBP's RNA binding domain (RBD) components[147].

It is likely that m^{6}A-modified transcripts can be specifically recognized by proteins rather than through direct binding of the m^{6}A moiety. Researchers have made great efforts in discovering additional m^{6}A readers via mainly two complementary systematic strategies: firstly, quantitative proteomics of interactomes of m^{6}A-modified and unmodified RNA probes in various cell lysates[69]; secondly, data mining of published transcriptome-wide RBP binding sites to see how well they overlap with known m^{6}A sites[71]. In this chapter, I present two collaborative projects that identified and characterized additional classes of m^{6}A readers using these two approaches.

4.2 Results

4.2.1 Quantitative protein pull down discovered additional m^{6}A interactors

4.2.1.1 Quantitative proteomics of m^{6}A RNA probe interactome

To identify proteins that favorably bind m^{6}A-modified RNA, we carried out RNA pull down with mammalian cell lysate followed by quantitative protein mass spectrometry (**Figure 4.1A**). A probe with four m^{6}A modifications within the GGACU consensus motif was used together with a unmethylated control. The probes were incubated the heavy-SILAC (stable isotope labeling with amino acids in cell culture)-labeled (H) and light-SILAC labelled (L) cell lysates

respectively. A "forward" and a label-swap "reverse" pull down were performed. A typical RNA-pull down interactome plot (**Figure 4.1A**, right) shows m^6A readers at the top right corner (high forward H/L ratio; low reverse H/L ratio) while the m^6A-repelled proteins at the bottom left corner. Background proteins are clustered together around the origin of the plot and show a ~1:1 ratio in both experiments.

We used nuclear and cytoplasmic lysates from HeLa and mESC cells to identify respective human and mouse proteins that are m^6A readers or repelled proteins. As a positive control, proteins containing YTH domains were robustly identified as the most prominent m^6A readers in all sets of experiments (**Figure 4.1B-C&E-F**). Other readers include FMRP proteins (FMR1, FXR1, and FXR2) and HNRNP proteins (HNRNPF, HNRNPH1, and HNRNPH2). These experiments also identified robust m^6A repelled proteins G3BP1 and G3BP2, which are markers of stress granules[148]. Analyses of common RNA binding domains in m^6A readers or repelled proteins revealed that besides YTH domain, the RNA binding capacity of KH, RRM, and RBD domains may also be regulated by m^6A (**Figure 4.1D&G**).

To address whether the identified readers or repelled proteins depend on the RNA sequence flanking m^6A, we performed the pull-down experiment with another two sets of 5-nt long degenerate probes in mESC whole cell extracts (**Figure 4.2**). YTHDF1-3 remained the strongest m^6A readers even with the degenerate sequence context, while YTHDC1 became less prominent compared to in GGACU context (**Figure 4.1F&4.2B**), consistent with the report that only the YTH domain of YTHDC1 harbors a distinctly selective binding pocket for the nucleotide preceding the m^6A nucleotide[149,150]. In contrast, readers such as FMR1 recognize m^6A in an RNA-sequence-dependent manner (**Figure 4.2**). There was limited overlap in the repelled proteins of the two degenerate sequence probes except G3BP1/2 (**Figure 4.2**).

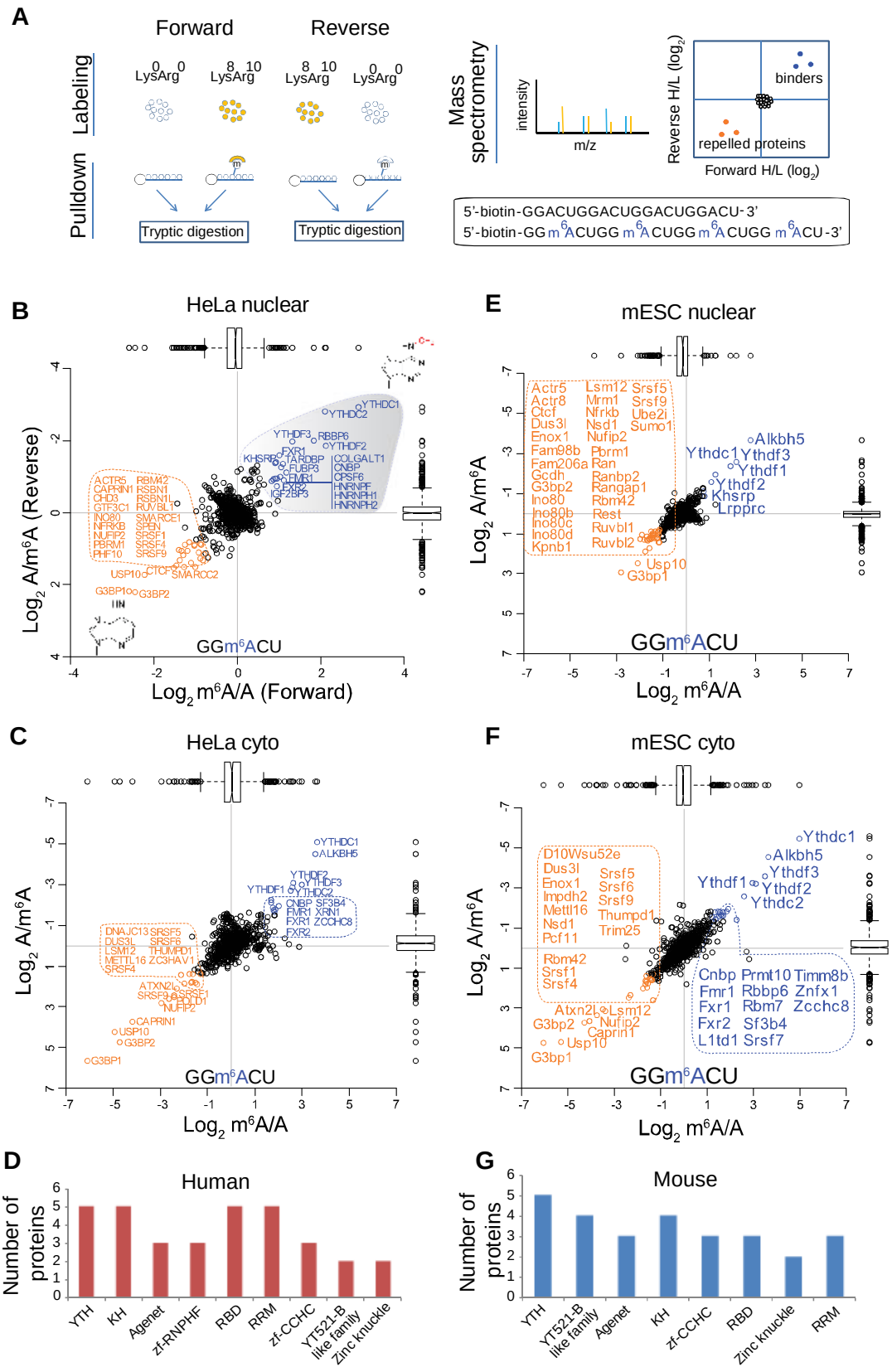

Figure 4.1 Quantitative proteomics of m^{6}A readers and repelling proteins in humans in mice

(**Figure 4.1, continued**) (**A**) Schematic representation of the workflow of SILAC-based quantitative proteomics characterization of m^{6}A readers and repelled proteins. (**B-C**, **E-F**) Results of SILAC-based m^{6}A RNA pulldowns in HeLa nuclear extract (**B**), HeLa cytoplasmic lysates (**C**), mESC nuclear extract (**E**), and mESC cytoplasmic lysates (**F**). m^{6}A-interacting proteins are depicted in blue, repelled proteins are in orange, and background proteins are in black. Box-plots: center lines, medians; box edges, interquartile range (IQR), whiskers, ±1.5×IQR (the threshold to identify significant interactors); outlier values, black circles. (**D&G**) Protein domains present in human (**D**) and mouse (**G**) m^{6}A interactors.

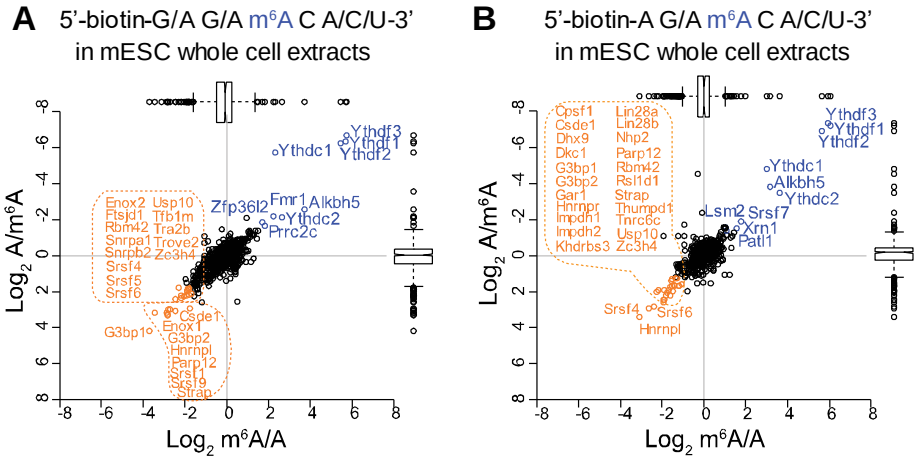

Figure 4.2 m^{6}A interactors in degenerated sequences

Scatter plots of SILAC based m^{6}A pull downs in mESC whole cell extract using degenerate sequence probes 1 (**A**) and 2 (**B**). The sequence context of the probe is depicted at the top.

4.2.1.2 G3BP1 is an m^{6}A repelled protein regulating mRNA stability

G3BP1 and G3BP2 share similar domain structures, and their *C*-termini comprise canonical RNA binding domains RRM and RGG (**Figure 4.3A**)[151]. For the rest of the section we focus on G3BP1. We tested the RNA-sequence-context preference of G3BP1 using recombinant proteins and RNA pull down. m^{6}A in the GGACU motif (methylated "A" is underlined) clearly abolished the binding of GST-G3BP1 while the trend was less apparent in the GAACU context (**Figure 4.3B**). PAR-CLIP of Flag-tagged G3BP1 in HEK293T cells identified ~3,800 binding sites of G3BP1 on mRNA, which are enriched in the 3'UTR and in the context of "AAC" (**Figure 4.3C-D**).

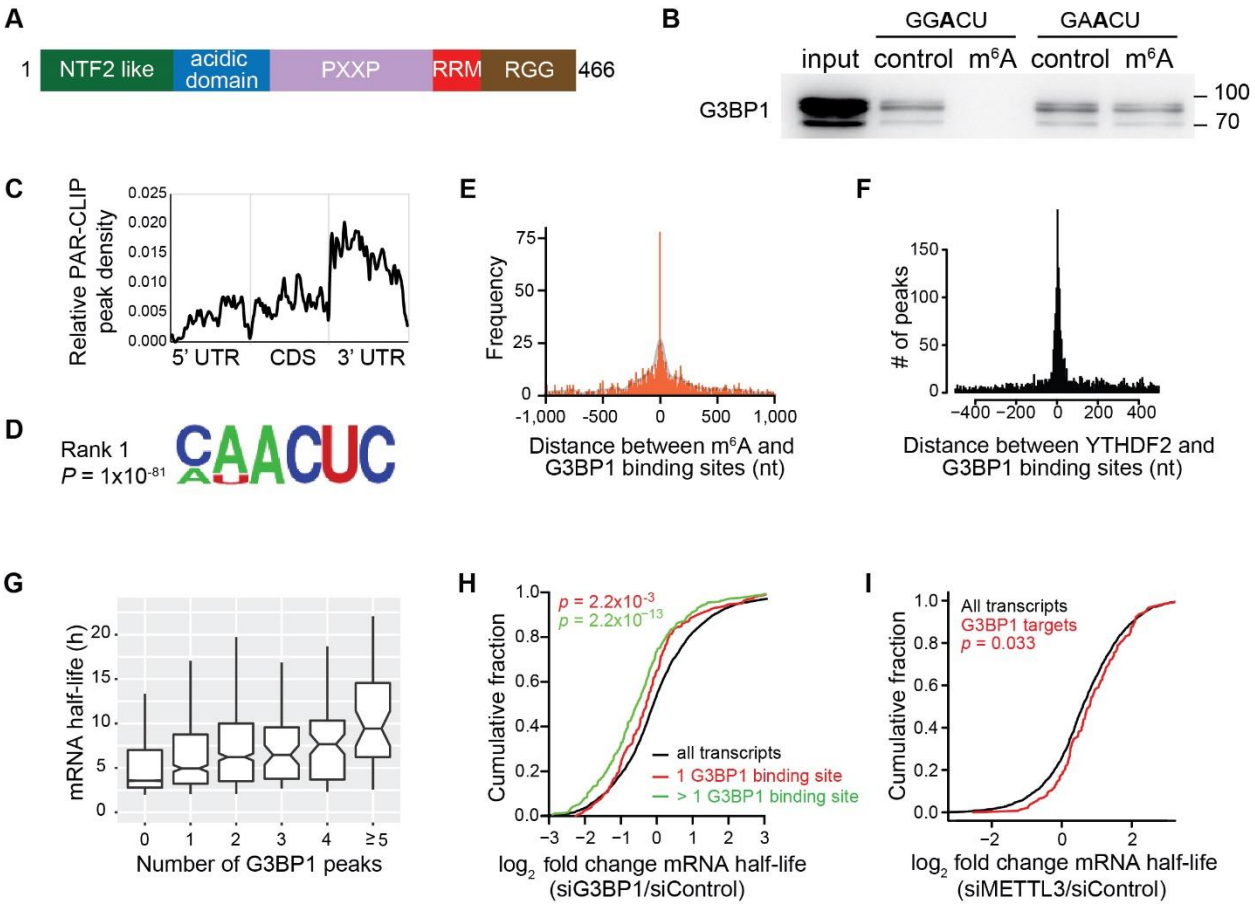

Figure 4.3 G3BP1 is an m^{6}A repelled protein regulating mRNA stability

(**A**) The G3BP1 domain structure. (**B**) GST western blot showing RNA-sequence-context-dependent inhibition of GST-G3BP1 binding to RNA by m^{6}A. (**C**) Metagene profiles of G3BP1 binding site distribution across the transcriptome. (**D**) The top consensus sequences of G3BP1 on mRNA. (**E**) The distribution of G3BP1 relative to m^{6}A sites. The distance between PAR-CLIP peaks and m^{6}A sites was counted in 10-nucleotide bins. A Gaussian kernel smoothing over the histogram is plotted as a transparent black line. (**F**) Distribution of G3BP1 peaks relative to YTHDF2 over the transcriptome. (**G**) The correlation between G3BP1 mRNA binding and target gene transcript stability. Box-plot elements: center lines, medians; box edges, interquartile range (IQR), whiskers, ±1.5×IQR; outlier values, not shown. (**H**) Cumulative distribution showing the effect of G3BP1 knockdown on target genes. The $\log_2$ fold changes in mRNA half-life were grouped and analyzed on the basis of the number of G3BP1-binding sites on each transcript. *P* values were calculated by two-sided Mann–Whitney *U* test. (**I**) A cumulative distribution showing the effect of METTL3 knockdown on the life-time of G3BP1 target genes.

We next compared the G3BP1 binding sites with published single-base-resolution m^{6}A sites in HEK293T cells[21]. The binding sites of G3BP1 cluster around m^{6}A sites (**Figure 4.3E**), suggesting that the binding between G3BP1 and a good portion of the target sites is subject to the

presence of m^6A methylation. We also observed substantial overlap between the binding sites of YTHDF2[55] and G3BP1 (**Figure 4.3F**), indicating their potential crosstalk on regulating mRNA stability. Indeed, we observed a positive correlation between the G3BP1 target half-life[152] and the number of G3BP1 binding sites on the target (**Figure 4.3G**). The depletion of G3BP1 in HEK293T cells led to a global decrease of G3BP1 target life-time, especially for those with multiple G3BP1 binding sites (**Figure 4.3H**). Moreover, reducing the m^6A level by knockdown of the writer component METTL3 stabilized G3BP1 targets (**Figure 4.3I**), suggesting the stabilization effect of G3BP1 is in part regulated by m^6A, potentially through the competition for binding sites with YTHDF2. In summary, these results reveal the intricate interplay among m^6A, m^6A readers, and m^6A repelled proteins to regulate mRNA turnover.

4.2.1.3 FMR1 is an m^6A reader modulating translation of m^6A-modified transcripts

FMR1 contains three KH domains, an RGG domain, and two tandem Agenet (Age) domains (**Figure 4.4A**)[153]. The lack of FMRP causes Fragile X mental retardation, and a missense mutation I304N results in a variant phenotype of the severe syndrome[154]. FMR1 and its analogue proteins FXR1 and FXR2 were consistently enriched by the m^6A RNA probe in the SILAC experiments (**Figure 4.1**). We further verified this preference for m^6A modified RNA by purified recombinant protein GST-FMR1 and RNA pull down (**Figure 4.4B**). The mutant GST-FMR1 I304N, in contrast, barely bound the RNA probes. Using cross-linking immunoprecipitation followed by ultra-high-performance liquid chromatography-tandem mass spectrometry (CLIP-UPLC-MS/MS), we validated that Flag-HA-tagged FMR1, but not the control nor I304N mutant constructs, enriches m^6A-modified transcripts inside cells (**Figure 4.4C**). We also confirmed the overlap between FMRP binding sites and m^6A modified sites in HEK293T cells from published datasets[22,155] (**Figure 4.4D**).

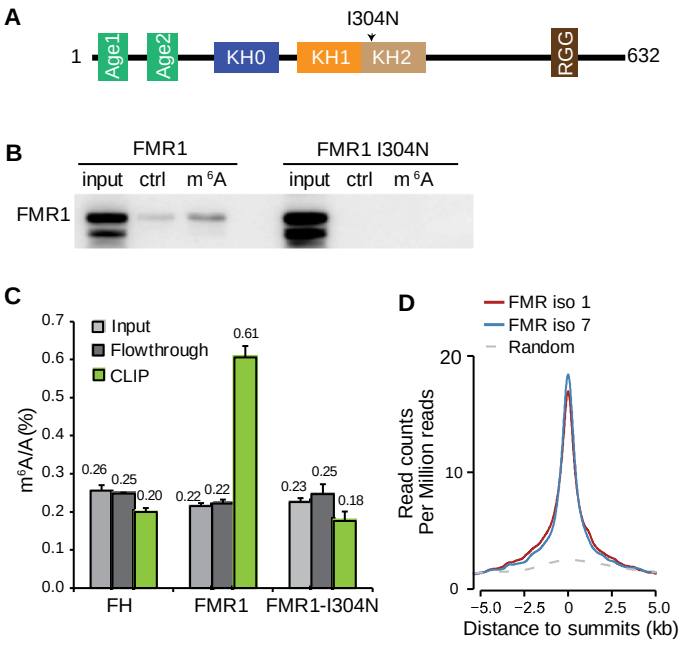

Figure 4.4 FMR1 favorably binds m^{6}A-modified transcripts

(**A**) The domain structure of FMR1. (**B**) Western blots showing preferential m^{6}A binding by recombinant GST-FMR1 *in vitro*. (**C**) Representative LC-MS quantification showing enrichment of m^{6}A in Flag-HA-tagged FMR1-bound mRNA in HEK293T cells. Error bars represent the range; $n = 2$ independent experiments. (**D**) Visualization of binding of FMR1 isoforms 1 and 7 to mRNA relative to known m^{6}A sites on mRNA. Randomized peaks were generated from the transcriptome and used as a control.

Given that FMRP has been suggested to repress translation by stalling ribosomes on the transcripts[156], we examined whether FMR1 would modulate the m^{6}A- and YTHDF1-mediated regulation on translation. From the published FMRP and YTHDF1 PAR-CLIP datasets[58,155], we observed a significant overlap of their target transcripts (**Figure 4.5A**). We performed pulse SILAC experiments (**Figure 4.5B**) to gain a proteome-wide insight into (i) how FMR1 regulates the translation of its targets and the common targets with YTHDF1 and (ii) how this is modulated by global m^{6}A level. The overexpression of FMR1 or the I304N mutant was under control of an inducible promoter responding to doxycycline, and the cells were treated with either siControl or siMETTL3 to evaluate the contribution of m^{6}A. The time till half of the protein is labelled with stable isotope reflects the translation rate—the longer the time, the slower the translation.

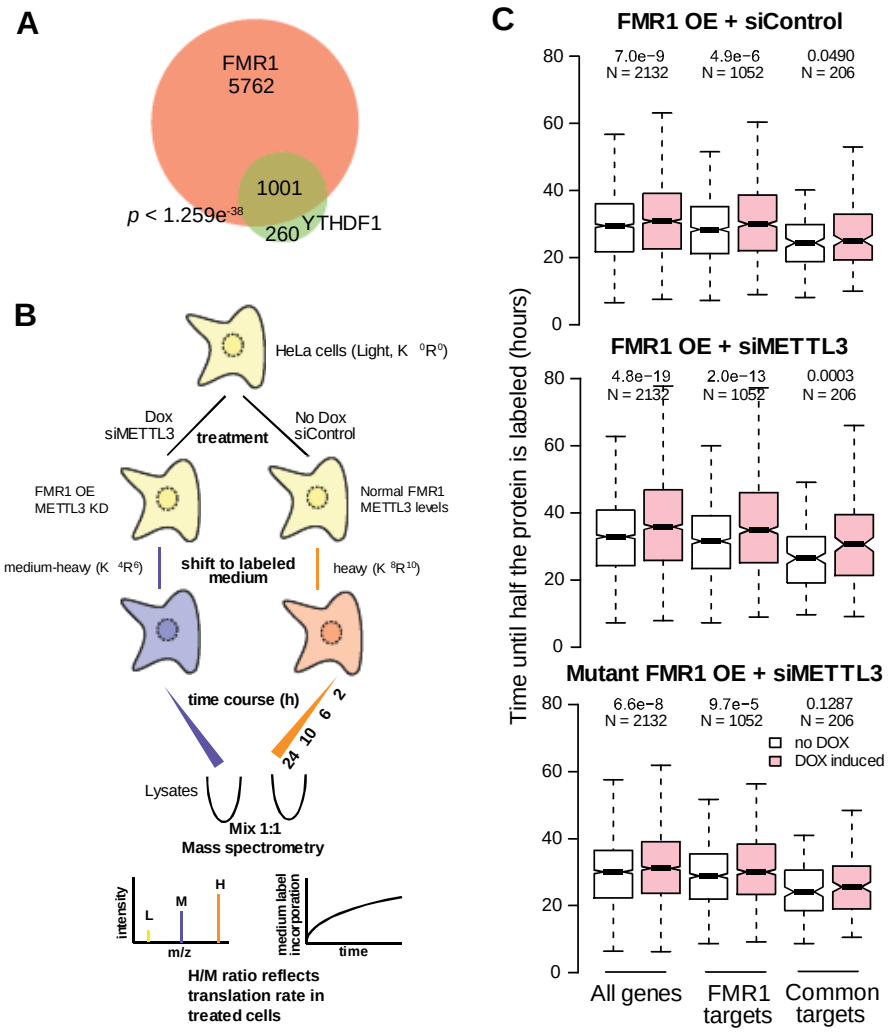

Figure 4.5 Crosstalk between FMR1 and YTHDF1 on translation of m^{6}A-modified mRNA

(**A**) Overlap between FMR1 and YTHDF1 target transcripts. (**B**) A schematic representation of the pulsed SILAC workflow to quantify translation rates at the proteome level. Dox, doxycycline. (**C**) Box-plots depicting the global changes in protein half-lives for indicated group of genes and samples. *N* represents the number of genes. Box-plot elements: center lines, medians; box edges, interquartile range (IQR), whiskers, ±1.5×IQR; outlier values, not shown.

In the siControl-treated cells, FMR1 overexpression significantly decreased translation rate of its targets but to a less extent for its common targets with YTHDF1 (**Figure 4.5C**, top). The translation repression effect of FMR1 (but not FMR1 mutant) on the common targets was apparent only when METTL3 expression was disturbed (**Figure 4.5C**, middle & bottom). On the basis of our results, we propose that YTHDF1 and FMR1 compete for binding to m^{6}A sites on mRNA

while the former depends more strictly on m^{6}A for binding than the latter. After METTL3 knockdown, many putative ACU-containing FMR1-binding sites become available in the absence of YTHDF1, leading to more pronounced translation inhibitory effects of FMR1. These results further illustrate the complex interplay among m^{6}A, m^{6}A readers, and mRNA homeostasis.

4.2.2 IGF2BPs as a new class of m^{6}A readers

4.2.2.1 Identification of IGF2BPs as m^{6}A-binding proteins

In collaboration with Prof. Jianjun Chen's lab, we reported the identification of IGF2BPs as a new class of m^{6}A readers[71]. Besides using the strategy of RNA pull down followed by protein mass spectrometry as discussed in section **4.2.1**, we also analyzed the enrichment of "GGAC" motif in the published RBP binding sites. All three IGF2BP proteins were among the top 15 of 112 RBPs in terms of the significance of m^{6}A motif enrichment (**Figure 4.6**).

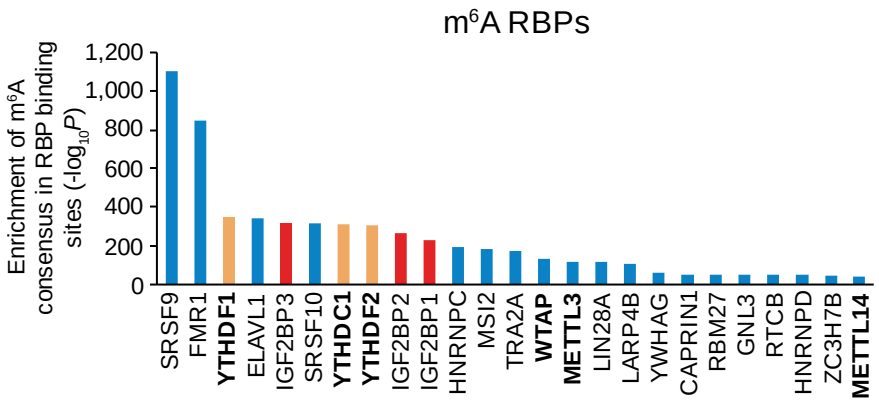

Figure 4.6 Top RBPs with enriched m^{6}A consensus sequence

Enrichment of the m^{6}A consensus sequence 'GGAC' in the binding sites of RBPs. The three IGF2BP paralogues are shown in red, whereas the YTHDFs were shown in orange. *P* values were calculated based on the binomial test by the Homer software.

We then examined the binding between IGF2BPs and m^{6}A inside cells. CLIP followed by LC-MS/MS m^{6}A quantification showed that mRNAs directly bound by Flag-tagged IGF2BPs are preferentially m^{6}A modified in HEK293T cells (**Figure 4.7A**). Analyses of published IGF2BP1-3

PAR-CLIP datasets in HEK293 cells[157] revealed extensive binding at the 3'UTR, similar to the m^{6}A profile[22] (**Figure 4.7B**). RIP-seq of Flag-tagged IGF2BPs were also performed and the common targets from RIP and PAR-CLIP experiments were defined as high-confidence targets, of which over 80% are annotated as m^{6}A-modified (**Figure 4.7C**). IGF2BPs share similar domain structures including two RRM domains and four KH domains. To narrow down the essential RNA binding domains that confer the m^{6}A preference, we made mutations in the GxxG motifs of the KH domains, which abolishes the RNA binding capacity (**Figure 4.7D**). It revealed that KH3-4 domains are indispensable for m^{6}A binding (**Figure 4.7E**).

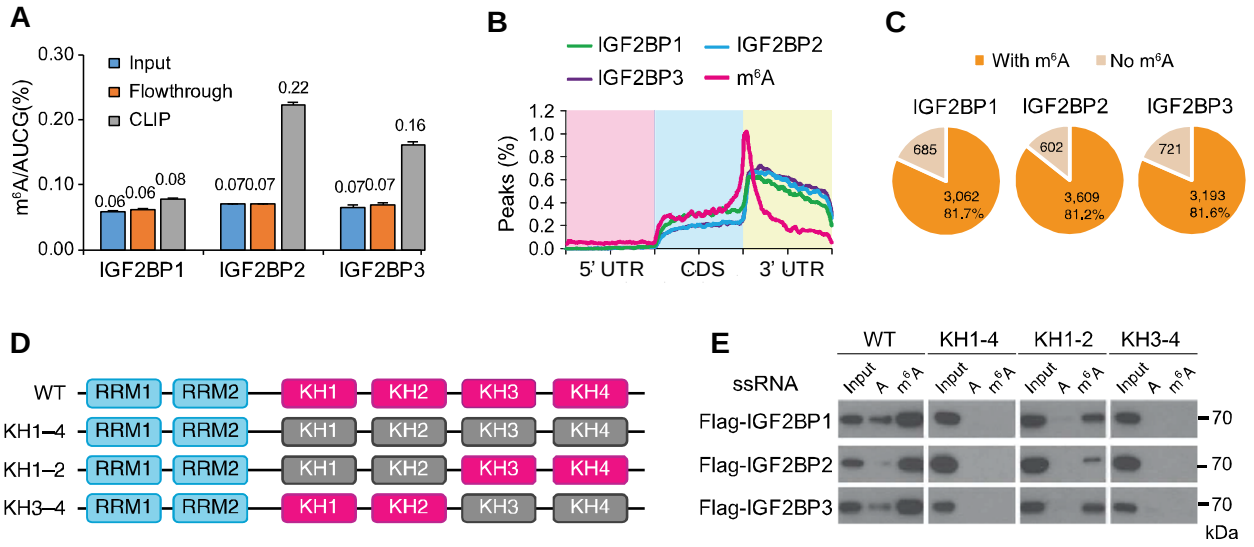

Figure 4.7 IGF2BPs preferentially bind m^{6}A-modified RNAs depending on KH3-4 domains

(A) Quantification of m^{6}A/AGCU ratio by LC-MS/MS in RNAs directly bound by ectopically expressed IGF2BP1 (chicken ZBP1), IGF2BP2 (human) or IGF2BP3 (human). *n* = 2 independent experiments. (**B**) Metagene profiles of enrichment of IGF2BP-binding sites and m^{6}A modifications across mRNA transcriptome. (**C**) Pie charts showing the numbers and percentages of IGF2BP high-confidence target genes that contain m^{6}A peaks. (**D**) Schematic structures showing RNA-binding domains within IGF2BP proteins and a summary of IGF2BP variants used in this study. Blue boxes are RRM domains, red boxes are wild-type KH domains with GxxG core and grey boxes are inactive KH domains with GxxG to GEEG conversions. (**E**) RNA pulldown followed by western blotting showed in vitro binding of single-stranded RNA (ssRNA) baits with wild-type or KH domain-mutated IGF2BP variants, representative of three independent experiments.

4.2.2.2 IGF2BPs regulate mRNA stability

IGF2BPs have been suggested to stabilize its targets in HEK293 cells[157]. Using the target lists identified, we confirmed that in HepG2 cells, knockdown of IGF2BPs also led to decreased abundance of its targets (**Figure 4.8A**, shIGF2BP3 was shown as an example). Moreover, the abundance of their targets was also affected by mRNA m^{6}A level, with decreased expression level when m^{6}A writer component METTL14 was depleted in HepG2 cells (**Figure 4.8B**).

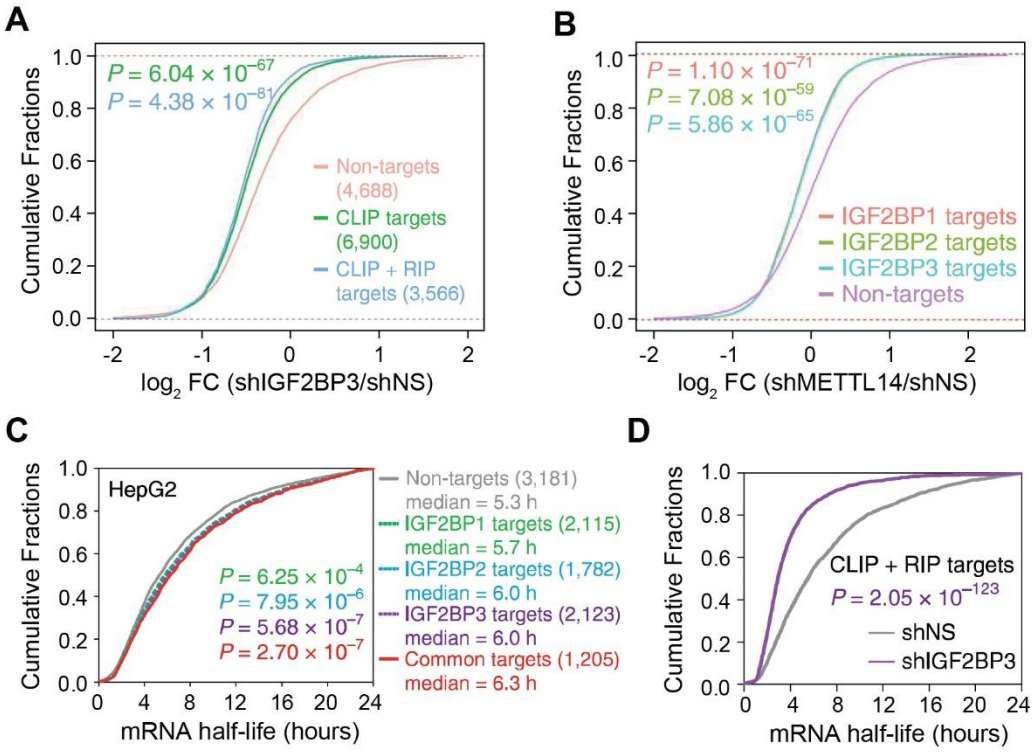

Figure 4.8 IGF2BPs regulate mRNA stability

(**A-B**) Cumulative frequency of mRNA $\log_2$ FC in non-target, CLIP target and CLIP + RIP target genes upon IGF2BP3 silencing (**A**) and in IGF2BP high-confidence target genes in shMETTL14 versus shNS cells (**B**). (**C-D**) Cumulative distribution of mRNA half-lives of non-target or IGF2BP high-confidence target genes in WT HepG2 cells (**C**) and of IGF2BP3 high-confidence targets in shIGF2BP3 or shNS HepG2 cells (**D**). *P* values were calculated using two-sided Wilcoxon and Mann-Whitney test.

To gain more direct evidence in mRNA stability, we performed mRNA half-life time profiling in WT HepG2 cells or cells depleted of IGF2BP3. It was observed that high-confidence

targets of IGF2BPs tend to have longer half-lives ($P < 0.001$) than their non-target counterparts (**Figure 4.8C**). In the shIGF2BP3 cells, the half-lives of IGF2BP3 targets were reduced 50% compared to in control cells (**Figure 4.8D**), confirming its role as an mRNA stabilizer. Further analyses showed that IGF2BPs and YTHDF2 share very few binding sites, and YTHDF2 binding sites tend to have higher GC content[71].

4.3 Conclusions and Discussion

Collaborative efforts from Vermeulen group, Chen group, and our group identified and characterized additional m^{6}A interacting proteins including readers and repelled proteins. Evidence from experimental RNA pull down (**Figure 4.1**) and bioinformatics (**Figure 4.6**) jointly revealed that RBPs with canonical RNA binding domains may also interact with m^{6}A-modified RNA where RNA structure and m^{6}A context sequences are likely to play a role. G3BP1 is repelled by m^{6}A in the "GG<u>A</u>C" RNA context, and G3BP1 stabilizes its targets presumably through competition with YTHDF2 for RNA binding (**Figure 4.3**). FMR1, on the other hand, prefers m^{6}A-modified transcripts (**Figure 4.4**). We propose a model of competitive binding between the "translation activator" YTHDF1 and the "translation repressor" FMR1, where the translation state of their common target transcripts is determined by the predominant binding protein depending on transcript m^{6}A modification state (**Figure 4.5**). Finally, IGF2BP proteins are preferentially recruited by m^{6}A-modified RNAs and stabilizes their RNA targets (**Figure 4.7-8**).

There have been three major types of m^{6}A interactors discovered so far (**Figure 4.9**): (i) direct readers that contain YTH domains; (ii) "m^{6}A switch" readers in which the presence of m^{6}A can remodel local RNA structure and consequently modulates RNA-protein interactions around or nearby[65]; (iii) RNA proteins with canonical RNA binding domains as discussed in this chapter, for which the detailed m^{6}A recognition mechanisms have yet to be explored.

Extensive crosstalk and interplay have been observed within (e.g. Chapter **2**) and between (e.g. Chapter **4**) different types of m^{6}A interactors and there are highly likely more to be discovered (**Figure 4.9**). The expanding list of m^{6}A readers and their interconnecting functions together add to the complexity and multifunctionality of m^{6}A-mediated regulation.

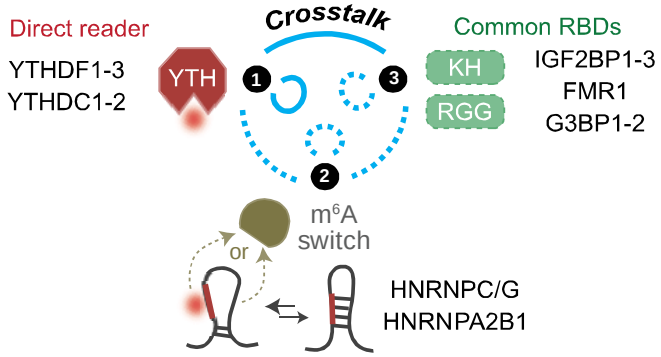

Figure 4.9 Various m^{6}A readers and their crosstalk add to the complexity of m^{6}A regulation
Blue curves indicate crosstalk relationships found (solid) or yet to be examined (dashed).

4.4 Methods

4.4.1 Cell culture, SILAC labeling and cell-extract preparation

Section **4.2.1**: HeLa cells and HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Thermo) supplemented with 10% FBS (Gibco), 1× penicillin–streptomycin (15140-122; Thermo) and 1 mM glutamine (Invitrogen). The cells were regularly tested in-house for the presence of mycoplasma. HeLa cells were SILAC-labeled by culture in SILAC DMEM (88420; Thermo) supplemented with 10% dialyzed FBS (DS-1003; Dundeecell), 1× Glutamax (35050-061; Thermo), 1× penicillin–streptomycin, 73 mg/mL L-lysine (light/K0, Sigma, A6969; or heavy/K8, Sigma, 608041) and 29.4 mg/mL arginine (light/R0, Sigma, A6969; or heavy/R10, Sigma, 608033).

ESCs were grown in serum-free 2i medium. Briefly, we prepared serum-free ESC medium by mixing the following components: Neurobasal medium (Gibco; 21103-049), 250 mL; DMEM/F12 (Gibco; 11320-033), 250 mL; N2-Supplement (Gibco; 17502-048), 2.5 mL; B27 minus RA (Gibco; 12587001), 5 mL; 7.5% BSA (Gibco;15260-037), 3.33 mL; penicillin–streptomycin, 5 mL; 1 mM glutamine (Invitrogen); 1% nonessential amino acids (Invitrogen); 6.3 µl monothioglycerol (M6145; Sigma); LIF (1,000 U/mL); and CHIR99021 and PD0325901 (3 and 1 µM, respectively). For SILAC labeling of ESCs, we used custom-made Neurobasal and DMEM/F12 (Life technologies) without lysine and arginine.

Whole cell lysates for RNA pulldowns were prepared according to the following protocol: Cells were harvested with trypsin and washed twice with chilled PBS. Cells were then resuspended in five pellet volumes of whole cell extract buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% NP-40, 0.5 mM DTT, 10% glycerol, and protease inhibitor cocktail) and incubated for 90 min at 4 °C on a rotating wheel. Crude lysates were then centrifuged for 30 min at 14,000 r.p.m. at 4 °C. Soluble whole cell extracts were finally aliquoted and snap-frozen in liquid nitrogen until further usage. Nuclear and cytoplasmic extracts were prepared as described previously[158].

Section **4.2.2**: the human hepatocellular carcinoma cell line HepG2 (ATCC HB-8065) was maintained in EMEM medium (American Type Culture Collection (ATCC)) supplemented with 10% FBS (Invitrogen), 2 mM L-glutamine and 1% penicillin–streptomycin. HEK293T and HeLa cells were grown in DMEM medium (Invitrogen) containing 10% FBS, 2 mM L-glutamine and 1% penicillin–streptomycin.

4.4.2 SILAC-based RNA pulldowns

All buffer solutions were prepared with high-quality, RNase-free reagents. Each SILAC-based RNA pulldown experiment consisted of four individual pulldowns. 10 µl of streptavidin

Sepharose high-performance beads (GE Healthcare) were used for each pulldown (20 µl of 50% slurry). Beads were first washed twice in 1 mL of RNA binding buffer (50 mM HEPES-HCl, pH 7.5, 150 mM NaCl, 0.5% NP40 (v/v), 10 mM $MgCl_2$). After each wash step, samples were centrifuged for 3 min at 4,000 r.p.m. in a pre-cooled (4 ^oC) tabletop centrifuge to spin down the beads. To inactivate and remove RNases, we incubated the beads with RNase inhibitor RNasin plus (Promega) in RNA binding buffer (100 µl buffer with 0.8 units of RNasin/µl) for 30 min on ice. After centrifugation and removal of RNasin–RNase complexes, beads were preblocked with yeast tRNA (50 µg/mL; AM7119; Life Technologies) in RNA binding buffer overnight at 4 °C on a rotation wheel. The preblocked beads were washed twice with RNA binding buffer and then incubated with 5 µg of biotinylated RNA probe (per pulldown) diluted with RNA binding buffer to a final volume of 600 µl. Beads were incubated for 30 min at 4 °C in a rotation wheel to allow binding of biotinylated probes to the streptavidin beads. The beads were washed once with 1 mL of RNA wash buffer (50 mM HEPES-HCl, pH 7.5, 250 mM NaCl, 0.5% NP-40 and 10 mM $MgCl_2$) and twice with protein incubation buffer (10 mM Tris-HCl, pH 7.5, 150 mM KCl, 1.5 mM $MgCl_2$, 0.1% (v/v) NP-40, 0.5 mM DTT, and complete protease inhibitors without EDTA (Roche)). Beads containing immobilized RNA were then incubated with 600 µg of nuclear extract or 1,000 µg of cytoplasmic extracts in a total volume of 600 µl of protein binding buffer. The incubation reaction also contained 30 µg of yeast tRNA to prevent nonspecific binding, and RNasin. In the forward experiment, the control probe was incubated with light-labeled (R0K0) lysates, whereas the m^6A probe was incubated with heavy (R10K8) lysates. The reverse experiment represented a biological-replicate label swap. The reactions were incubated at room temperature for 30 min and then for 90 min on a rotation wheel at 4 °C. The beads were then washed three times with protein incubation

buffer and twice with ice-cold PBS to remove detergent from the beads. The last PBS wash was used to combine beads from the forward and the reverse pulldowns as follows:

Forward reaction: control probe, R0K0 (light); m^6A probe, R10K8 (heavy)

Reverse reaction: control probe, R10K8 (heavy); m^6A probe, R0K0 (light)

Proteins were on-bead digested with trypsin. Briefly, beads were resuspended in 100 μL of elution buffer (50 mM Tris, pH 8.5, 2 M urea and 10 mM DTT) and then incubated for 20 min at room temperature in a thermoshaker at 1,100 r.p.m. Iodoacetamide was then added to a final concentration of 55 mM, and the mixture was incubated for 10 min in a thermoshaker (1,100 r.p.m.) at room temperature in the dark. Proteins were then partially digested from the beads by the addition of 250 ng of trypsin for 2 h at room temperature in a thermoshaker in the dark. After incubation, the supernatant was collected in a separate tube. The beads were then incubated with 50 μL of elution buffer for 5 min at room temperature in a thermoshaker (1,100 r.p.m.). 100 ng of fresh trypsin was added to the pooled eluates, and proteins were digested overnight at room temperature. Finally, tryptic peptides were acidified to pH <2 with TFA (10%) and desalted with C18 Stage tips before MS analyses.

4.4.3 Mass spectrometry

Tryptic peptides were separated with an Easy-nLC 1000 (Thermo) connected online to a Q Exactive or Orbitrap-Fusion mass spectrometer (Thermo). RNA pulldown samples were separated with a 94-min acetonitrile gradient (7% to 32%) followed by washes at 50% and then 95% acetonitrile for a total of 120 min of data acquisition. For measurements on the Q Exactive, the top ten most abundant peptides were fragmented for every full scan. For RNA pulldowns measured on the LTQ-Orbitrap Fusion Tribrid mass spectrometer (Thermo), a similar LC gradient was used, and samples were measured in top-speed mode. For the whole-cell proteome experiment

measured on the LTQ-Orbitrap Fusion Tribrid mass spectrometer (Thermo), peptides were separated with a 214-min acetonitrile gradient (7% to 30%) followed by washes at 60% and then 95% acetonitrile for a total of 240 min of data acquisition. Scans were collected in data-dependent top-speed mode in cycles of 3 s with dynamic exclusion enabled and set to 60 s.

4.4.4 MaxQuant analysis

Data processing was done with the MaxQuant software. For SILAC-labeling-based experiments, raw data were analyzed with MaxQuant version 1.5.0.1 with the default settings and K0R0, K4R6 or K8R10 (light, medium-heavy or heavy) labels, and matching between runs enabled[159]. We used the UniProt database downloaded on 13 July 2014 for identification. To identify significant interactors, we plotted normalized ratios from MaxQuant output tables for the forward and reverse pulldowns. Outliers were identified independently in the forward and reverse pulldowns by means of box plot statistics (threshold: 1.5× the interquartile range). Proteins were considered significant if they were identified as outliers in both experiments. Proteins identified in MaxQuant output tables were classified into four groups: readers, repelled proteins, background proteins, and contaminants. For whole-cell proteome experiments, we used a Swiss-Prot curated database for identification (downloaded on 7 February 2016). Scatter plots were generated in R.

4.4.5 Plasmids and Constructs

Section **4.2.1**: G3BP1 cDNA was cloned into the HindIII and BamHI sites of pcDNA3-c-Flag vector (Plasmid #20011; Addgene) to create an expression construct with a C-terminal Flag tag. For the creation of N-terminally GST-tagged constructs, G3BP1 cDNA was amplified from HeLa cDNA and cloned into pGEX-5X1 vector (GE Healthcare). We created FMR1 overexpression constructs for the creation of doxycycline-inducible HeLa cell lines by subcloning FMR1 (pFRT-TODestFLAGHAhFMRPiso1; Addgene; 48690) and FMR1-I304N (pFRT-

TODestFLAGHAhFMRPiso1I304N; Addgene; 48692) into AAVS1-TRE3G-EGFP (Addgene; 52343) using *Sal*I and *Mlu*I sites. A Kozac sequence and a Flag-HA tag were added to the *N*-terminus of FMR1 and FMR1-I304N by PCR.

Section **4.2.2**: pcDNA3-based vectors encoding wild-type and KH domain mutant FLAG-tagged chicken ZBP1 (refer to as IGF2BP1), human IGF2BP2 and IGF2BP3 were kindly provided by Dr Hüttelmaier (Martin Luther University, Germany). The RNAi Consortium (TRC) lentiviral vectors encoding shRNAs against IGF2BP1 (TRCN0000075149 and TRCN0000075152), IGF2BP2 (TRCN0000149002 and TRCN0000148565), IGF2BP3(TRCN0000074677 and TRCN0000074673), METTL3 (TRCN0000034715), METTL14 (TRCN0000015933) and their non-specific control (shNS, RHS6848) were purchased from GE Dharmacon; whereas the packing vectors, pMD2.G, pMDLg/pRRE and pRSV-Rev, were obtained from Addgene.

4.4.6 siRNA transfection and lentivirus shRNA infection

Section **4.2.1**: siRNAs targeting human G3BP1 (SR306845) or METTL3 (SR311159) and control siRNAs (SR30004) were obtained from Origene. siRNAs were transfected at a final concentration of 25 nM using Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions.

Section **4.2.2**: lentiviruses were produced in HEK293T cells by co-transfecting individual shRNA construct with packing vectors (pMD2.G, pMDLg/pRRE and pRSV-Rev) into HEK293T cells in a 60-mm cell culture dish using X-tremeGENE HP DNA Transfection Reagent (Roche Diagnostics). The lentivirus particles were harvested at 48 h and 72 h after transfection and directly added into target cells with 4 μg per ml polybrene. After two rounds of infection, cells were selected for at least two passages by adding 1 μg per ml puromycin into growth medium.

4.4.7 Recombinant protein expression and purification

Section **4.2.1**: Rosetta BL-21 bacteria (Novagen) were used for GST-fusion protein expression. Cells were grown in LB medium supplemented with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. IPTG was added to a final concentration of 0.5 mM, after which protein expression was induced for 4 h or overnight at 16 °C. Cells were then harvested, washed with PBS and resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 15% glycerol, 1 mM EDTA, 0.5 mM PMSF, 1 mM DTT, 100 mM NaCl, 1% Triton X-100, 0.25% NP-40, protease inhibitor cocktail, and lysozyme to a final concentration of 0.5 mg/ml). Cells were lysed by repeated (five times) freeze-thaw cycles. Bacterial debris was removed by centrifugation at 14,000 r.p.m. for 20 min at 4 °C, after which soluble extracts were aliquoted and snap-frozen in liquid nitrogen until further usage.

Section **4.2.2**: FLAG-tagged IGF2BP1, IGF2BP2 and IGF2BP3 were expressed in HEK293T cells. For each protein, four 15-cm dishes of cells were prepared and lysed in 4 ml lysis buffer (50 mM Tris-HCl pH 7.5, 300 mM KCl, 0.5% NP-40, 5% glycerol, 5 μg per ml DNase I, 1% RNase T1/A, 1% protease inhibitor, 1 mM dithiothreitol) at 4 °C for 1 h and sonicated (5 s on, 25 s off, for 24 cycles). The lysate was then cleared by centrifuge at 4 °C. The proteins were affinity purified using 40 μl anti-FLAG M2 resin (Sigma-Aldrich) at 4 °C for 2 h. After extensive wash with wash buffer (50 mM Tris-HCl, pH 7.5, 300 mM KCl, 5% glycerol, 1 mM dithiothreitol), proteins were eluted in 500 μl 1× FLAG elution solution (0.5 mg per ml FLAG peptide in wash buffer) at 4 °C for 1 h. Protein purity was verified with SDS–PAGE followed by coomassie staining.

4.4.8 CLIP-LC-MS/MS

Section **4.2.1**: we performed CLIP as described in section **2.4.6** with 10 dishes of confluent HEK293T cells stably overexpressing Flag-HA tags, Flag-HA-tagged FMR1, or I304N mutant except skipping 4SU treatment and using 365 nm UV for crosslinking. After IP and washes, the bound RNA was eluted directly from the beads by Protease K digestion in parallel with "Input" and "Flow-through" samples. Three volumes of TRIzol were added to each sample, and total RNA was extracted on a Direct-zol Microprep spin column (Zymo Research) according to the manufacturer's protocol for RNA extraction and DNase I digestion. rRNA was removed with the RiboMinus Eukaryote Kit v2 (Thermo). 50 ng of input, flow-through, and immunoprecipitated mRNA were analyzed by LC-MS/MS as described in section **2.4.4**.

Section **4.2.2**: we performed CLIP-LC-MS/MS as described above with three 15-cm dishes of confluent HEK293T cells transiently overexpressing FLAG-tagged IGF2BP (1, 2 or 3) except without RNase T1 digestion.

4.4.9 Pulse SILAC

Pulsed SILAC is a variant SILAC method used to measure translation rates in cells[160]. HeLa cells expressing FMR1 under doxycycline control were SILAC-labeled in light medium (K0R0) for a week to allow them to adapt to the culture conditions. Doxycycline was added for 24 hours to induce expression of FMR1 or the FMR1 mutant. Cells without doxycycline induction served as a control. In the case of FMR1 overexpression combined with METTL3 knockdown, METTL3 siRNAs were transfected 24 h before the addition of doxycycline. Control siRNA was transfected into control cells that were not treated with doxycycline. 24 h after doxycycline induction, cells were washed with PBS to remove any light amino acid-containing medium. Cells with FMR1 overexpression alone or in combination with METTL3 knockdown were transferred

to medium-heavy (K4R6) medium. Non-doxycycline-treated control cells were transferred to heavy (K8R10) medium. Cells were subsequently harvested 2, 6, 10 and 24 hours after transfer into medium-heavy or heavy SILAC medium. The harvested cells were then lysed on ice by the addition of 150 µl of lysis buffer (100 mM Tris-HCl, pH 7.6, 4% SDS) and heated for 5 min at 95 °C. The samples were sonicated with a Diagenode Bioruptor to reduce viscosity. Crude lysates were then centrifuged at 14,000 r.p.m. for 20 min, and supernatants were recovered. Lysate protein concentrations were measured with the BCA kit (Thermo). Equal amounts of protein from control and treated samples at each time point were then mixed. DTT was added to a 100 mM final concentration, and the samples were incubated for 15 min. Samples were then further processed via the FASP method[161]. A trypsin-LysC mix (Promega) was used for digestion. Finally, tryptic peptides were acidified to pH <2 with TFA (10%) and desalted with C18 Stage tips before MS analyses.

The time it took for half the protein to be labeled (THPL) was calculated for all identified proteins as described by Visscher *et al*.[162]. In short, we performed a linear regression between the log-transformed fraction of labeled protein and the pulse time. We then extrapolated the THPL by dividing the slope of the regression by log(0.5). In addition to the four pulsed time points, we forced the regression lines though zero as described previously[162]. To ensure high confidence of our individual THPL measurements, we used only the proteins with an identified label ratio in all time points and a Pearson correlation coefficient of 0.95 or higher for downstream analyses.

4.4.10 G3BP1 PAR-CLIP

PAR-CLIP was performed as described in section **2.4.6** with five 15-cm plates of HEK293T cells transfected with Flag-tagged G3BP1 or G3BP2 plasmid at 80% confluency for 24

hours. Two biological replicates were carried out for G3BP1. All samples were sequenced on an Illumina HiSeq 2000 with single-end 50-bp read length.

4.4.11 mRNA lifetime profiling

Section **4.2.1**: 24 hours after transfection with siRNAs, cells were trypsinized and live cells were counted. Cells were then equally divided over three fresh plates. 48 h after transfection, actinomycin D was added to a final concentration of 5 μg/mL (Santa Cruz Biotechnology; sc-200906). Cells were then harvested 0, 3 and 6 hours after the addition of actinomycin D, respectively. The harvested cells were counted once again and snap-frozen. RNA was purified with the Qiagen RNeasy mini kit and spike with ERCC RNA spike-in control (Thermo Fisher Scientific). rRNA was removed with the Ribo-Zero rRNA removal kit (Illumina). RNA fragmentation was performed with the NEBNext Magnesium RNA Fragmentation Module (NEB). The first cDNA strand was synthesized with Superscript III in the presence of actinomycin D and random hexamers. The second strand was synthesized with Escherichia coli DNA polymerase in the presence of dUTP. USER enzyme was added before library amplification to maintain strand identity. NEXTflex 8-bp adaptors were used for adaptor ligation. Size selection was performed with E-gel, and quality was assessed on a Bioanalyzer.

Section **4.2.2**: HepG2 cells with stably expressed shRNAs against IGF2BPs or shNS were seeded into 6-well plates to get 50% confluency after 24 h. Cells were treated with 5 μg per ml actinomycin D and collected at indicated time points. The total RNA was extracted by miRNeasy Kit (Qiagen) and analysed by RT–PCR and RNA-seq. For RNA-seq, an equal amount of ERCC RNA spike-in control (Thermo Fisher Scientific) was added to the total RNA samples as internal controls before library construction. Sequencing libraries were prepared using NEBNext Ultra

Directional RNA Library Prep Kit. RNA stability profiling was generated from two biological replicates.

4.4.12 RIP of Flag-IGF2BPs

Cells seeded in a 10-cm dish at 70–80% confluency were crosslinked by UV and harvested by trypsinization. Nuclear extraction was isolated and sonicated. 1 μg of FLAG (F3165, Sigma-Aldrich), IGF2BP1 (8482, CST) or IGF2BP2 (14672, CST) antibody or a corresponding control IgG (mouse IgG (CS200621, Millipore) for FLAG, rabbit IgG (NI01, Millipore) for IGF2BP1 and IGF2BP2) was conjugated to protein A/G magnetic beads (Thermo Fisher Scientific) by incubation for 4 h at 4 °C, followed by washing three times and incubation with pre-cleared nuclear extraction in RIP buffer (150 mM KCl, 25 mM Tris (pH 7.4), 5 mM EDTA, 0.5 mM DTT, 0.5% NP-40, 1× protease inhibitor) at 4 °C overnight. After washing with RIP buffer for three times, beads were resuspended in 80 μl PBS, followed by DNA digestion at 37 °C for 15 min and incubation with 50 μg of proteinase K (Thermo Fisher) at 37 °C for 15 min. Input and co-immunoprecipitated RNAs were recovered by TRIzol, extraction and analyzed by RNA-seq.

4.4.13 Sequencing data analyses

Section **4.2.1**:

PAR-CLIP: G3BP1 PAR-CLIP peaks were called as described in section **2.4.9**.

mRNA half-life: RNA-seq libraries were quantified with kallisto version 0.43.0[163] by pseudo-alignment to the RefSeq transcript database downloaded 28 June 2016. Absolute transcript concentrations per cell were computed by linear-fitting the RNA spike-in TPM values to attomoles. RNA half-life measures were calculated as described by Wang *et al.*[55]. To ensure accurate RNA half-life estimations, we considered only transcripts with finite RNA half-lives of <24 h for further analyses.

Integral analyses: to assess the distribution of m^6A with respect to PAR-CLIP peaks, we used publicly available single-nucleotide-resolution m^6A data from Linder *et al*.[21] and converted the reported m^6A sites to hg38 with liftOver[164]. For FMR1 and m^6A colocalization analysis, we generated random peaks by randomly permuting the location of m^6A peaks on the transcriptome.

Section **4.2.2**:

RIP–seq: samples were sequenced by Illumina HiSeq 1000 with a single-end 51-base pair (bp) read length. The RIP–seq reads were mapped to human genome version hg19 by Tophat2 version 2.0.13 with default settings[140]. Differential gene expression was calculated by Cuffdiff version v2.2.1[141]. The RIP targets were defined as genes with reads per kilobase, per million reads (RPKM) ≥ 1, immunoprecipitation/input ≥ 2, and $P < 0.05$.

PAR-CLIP: the IGF2BPs PAR-CLIP data were obtained from the public database Gene Expression Omnibus (GEO; accession No. GSE21918)[157]. The adapters were trimmed by using cutadapt version 1.9.1[142]. The processed reads were mapped to human genome version hg19 by bowtie version 1.1.2 with parameters: –v 3 –m 5 --best --strata[165]. The mapped results were analysed by PARalyzer v1.5 with default settings[108]. The results were further filtered by ModeScore ≥ 0.6. For peaks that were larger than 50 nt, we extracted 50 nt centered on the ModeLocation site. GENCODE v24 was used to annotate the filtered peaks, and finally, the PAR-CLIP target genes of IGF2BP1 (7,511), IGF2BP2 (7,974) and IGF2BP3 (9,228) were identified.

RNA-seq: all RNA-seq samples were sequenced by Illumina HiSeq 1000 with single-end 51-bp read length. All reads were mapped to human genome version hg19 by hisat2 v2.0.4 with default settings[166]. Read counts were calculated using HTSeq[167], and was converted to RPKM using our custom Perl script. The average gene expression values of three independent studies were used for the following analysis.

mRNA lifetime profiling: all RNA-seq samples for mRNA lifetime profiling were sequenced and processed as described above. RPKM was converted to attomole by linear fitting of the RNA spike-in as previously described[55]. The degradation rate of RNA and the mRNA half-life were calculated according to the aforementioned formula. The final half-life was calculated by using the average value of 1 h, 3 h and 6 h.

4.4.14 Antibodies

Section **4.2.1**: primary antibodies to the following proteins were used: G3BP1 (Santa Cruz Biotechnology; sc-98561), Flag (Sigma-Aldrich; F1804), GST (Thermo; MA4-004) and METTL3 (Bethyl Laboratories; A301-567-A).

Section **4.2.2**: antibodies used for western blotting were as follows unless otherwise specified: IGF2BP1 (IMP1, clone D33A2, no. 8482), IGF2BP2 (IMP2, clone D4R2F, no. 14672) were from Cell Signaling Technology (CST), IGF2BP3 (IMP3, A303-426A) was from Bethyl Laboratories, and FLAG (M2, F3165) was from Sigma-Aldrich. GAPDH (sc-47724, Santa Cruz) was used as a loading control.

4.4.15 Data availability

Section **4.2.1**: all mass spectrometric data generated in this study have been deposited to the ProteomeXchange Consortium via the PRIDE repository under project identifier PXD006498. RNA-sequencing and PAR-CLIP data have been deposited at Gene Expression Omnibus (GEO) under accession code GSE98856.

Section **4.2.2**: all sequencing data that support the findings of this study have been deposited in NCBI's GEO under accession number GSE90639 (for RIP–seq) and GSE90684 (for RNA-seq).

Chapter 5

Summary and Perspectives

5.1 Finer dissection of m^{6}A functions in examination of m^{6}A readers

m^{6}A reader proteins function as "adaptors" that convert methylation codes into the signals that can be recognized by cellular machineries, thus m^{6}A-modified transcripts could be sorted out for grouped processing and functions based on the binding of a certain reader or a combination of multiple readers (**Figure 5.1**). mRNA stability can be regulated by the destabilizer YTHDF2 and the stabilizer IGF2BPs while mRNA translation may be modulated by the activator YTHDF1/3 and the repressor FMR1. In addition, there have been reports that the m^{6}A methylation itself may interact with translation machineries to modulate translation initiation efficiency[168] or translation elongation dynamics[169].

We currently do not know how readers achieve selectivity towards certain m^{6}A sites or certain m^{6}A-modified transcripts. One likely scenario is that readers may be localized to different regions of mRNA by interacting with other RBPs that recognize distinct features of the RNA. IGF2BP1-3 and YTHDF2, which regulate mRNA stability in opposing directions, were shown to bind distinct sites on mRNA: the former favors 3'UTR in HEK293T cells examined while the latter shows more binding to CDS[71]. Secondly, the density and sequence contexts of m^{6}A sites are likely to matter. Densely populated m^{6}A regions could be more frequently occupied by m^{6}A readers. In the case of FMR1, it binds consensus RNA sequences of GGA and ACU similar to the m^{6}A-containing sequence[155]. Perhaps the protein has evolved to further recognize the methyl group in these particular sequence contexts in order to add an additional layer of regulation. Thirdly, reader proteins might be enriched at specific cellular compartments, and therefore preferentially interact with local RNA species. For instance, reader proteins YTHDF1-3, FMR1, and

HNRNPA2B1, were identified in mammalian stress granule cores[170]. YTHDF1-3, HNRNPK, and IGF2BP2-3 were found to be enriched in the cell protrusion compared to the cell body in breast cancer cells[171]. Finally, it has been observed that readers may act in response to stimuli. The predominantly cytoplasmic YTHDF proteins could redistribute to the nuclei of cells under certain stimuli including heat shock stress[92] and viral infection[172]. In post-mitotic cells, the translation-promoting effect of YTHDF1 becomes especially apparent when active translation of functional proteins is desired, as exemplified in the dorsal root ganglion (DRG) model of injury-induced axon regeneration[93], translation arrest recovery after removal of sodium arsenite stress in HeLa cells[58], and neuronal activity-induced new protein synthesis in the mouse brain[173].

m^{6}A reader proteins form a network of physical or functional interactions. Examining function and specificity of reader proteins promises a finer and deeper understanding of m^{6}A-mediated gene expression regulation.

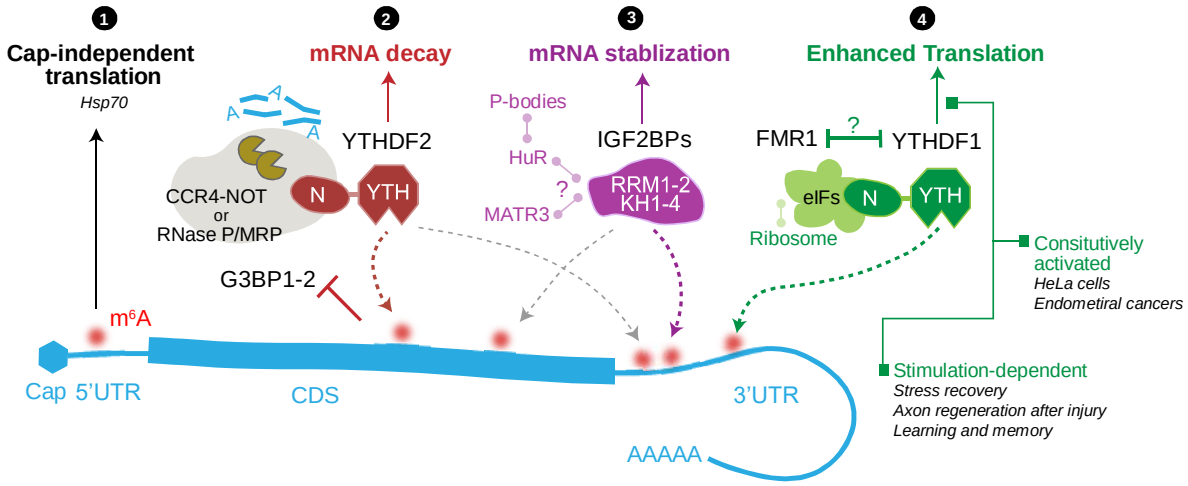

Figure 5.1 Region- and stimuli-dependent roles of m^{6}A and its readers

Multiple layers of contexts could substantially affect how m^{6}A methylation regulates the fate of modified mRNA. Region-dependent regulation: 5'UTR m^{6}A is linked to cap-independent translation, especially during stress response when cap-dependent translation is repressed (①). Reader-dependent regulation: YTHDF2 and IGF2BPs can affect mRNA stability in opposite directions. YTHDF2 recruits CCR4-NOT complex via its *N*-terminal domain to promote mRNA decay (②) while IGF2BPs stabilizes mRNA likely through co-factors HuR and MATR3 (③). They may recognize distinct m^{6}A sites and exhibit differential binding density in CDS and 3'UTR.

(**Figure 5.1, continued**) Stimulation-dependent regulation exemplified by YTHDF1 (④): YTHDF1 preferentially binds 3'UTR m^6A and promotes translation through interaction with translation initiation factors. While it constitutively promotes translation in cancer cells including HeLa cells and endometrial cancer cells, in post-mitotic cells the effect only significantly manifests when induced by stimulations such as recovery/repair processes or learning signals.

5.2 Context-dependent functions of m^6A effectors

Besides the complexity derived from reader proteins, the function of m^6A writers and erasers may also vary depending on specific biological contexts, leading to dynamic and assorted m^6A profiles and thus m^6A functions.

5.2.1 Multifaceted METTL3: Cellular Localization, PTM, and Functions

Cellular distribution of METTL3 varies among cell lines, and there are cases where cellular stress induces its redistribution[174,175]. When interacting with WTAP in the form of a stable dimer with METTL14, METTL3 localizes to the nuclear speckle in HeLa cells[39]. Functional nuclear localization signals (NLS) have been identified in both METTL3 and WTAP, of which key residue mutations abolished preferential nuclear localization of ectopic METTL3 and WTAP in HeLa cells[176]. A fraction of METTL3 protein is also detected in the cytoplasm in multiple human cancer cell lines at various proportions, including HeLa cells[177–179], breast cancer cells (MDA-MB-231 cells)[180], and acute myeloid leukemia cells (MOLM13 cells)[181]. In a rare instance, in mouse cortical neurons METTL14 rather than METTL3 localizes to both cytoplasm and nuclei[133]. It is not yet clear why METTL3 localization varies. The protein abundance ratios between METTL3 and METTL14 as well as other adaptor subunits of the writer complex could vary among cell lines, which could affect localization of METTL3. It is also possible that post-translational modifications (PTMs) alter interactions between METTL3 and its partner proteins, leading to cytoplasmic presence.

m^{6}A-containing genes are enriched in important cellular processes, and a subset of m^{6}A sites appear dynamic in response to stimuli and stress[20,22]. The transcript specificity of m^{6}A methylation could be due to writer recruitment at desired chromatin loci, likely through transcription factors (TFs) and/or epigenetic marks (**Figure 5.2A**). After heat shock, METTL3 was observed to localize to heat-shock genes in the chromatin, and m^{6}A installation on those heat-shock transcripts was suggested to ensure their timely clearance after the stress[174]. Upon DNA UV damage METTL3/14 localizes within 2 min to UV-induced damage sites, co-occurring with increased m^{6}A intensity[175]. In the TGFβ signaling pathway in human pluripotent stem cells, activated transcription factors SMAD family member 2/3 (SMAD2/3) interact with METTL3/14-WTAP, facilitating co-transcriptional m^{6}A installation on selective transcripts[182]. A recent study in acute myeloid leukemia cells showed that a fraction of METTL3 associates with the promoter regions of ~80 active genes specified by a transcription factor CCAAT enhancer binding protein zeta (CEBPZ) independent of METTL14[181]. Their results suggest that specialized "adaptor" proteins might exist in different cells to target the writer to distinct sets of genes in the chromatin, resulting in transcript-specific m^{6}A methylation. Moreover, a gene-body enriched histone modification Histone H3 trimethylation at lysine 36 (H3K36me3) is suggested to recruit the writer complex through interaction with METTL14 in HepG2 cells, favoring m^{6}A installation on mRNA coding sequences (CDS) and 3'UTRs[183]. In contrast to these nuclear roles, cytoplasmic-localized METTL3 is suggested not to function as an m^{6}A writer, but instead as a potential m^{6}A reader (**Figure 5.2B**). Independent of its catalytic activity, METTL3 in the cytoplasm in lung cancer cells promotes translation of a reporter mRNA when tethered to its 3'UTR[178]. Further studies showed that this translation promotion effect depends on interaction between METTL3 and eIF3h, which are coordinately overexpressed in many types of cancers[179].

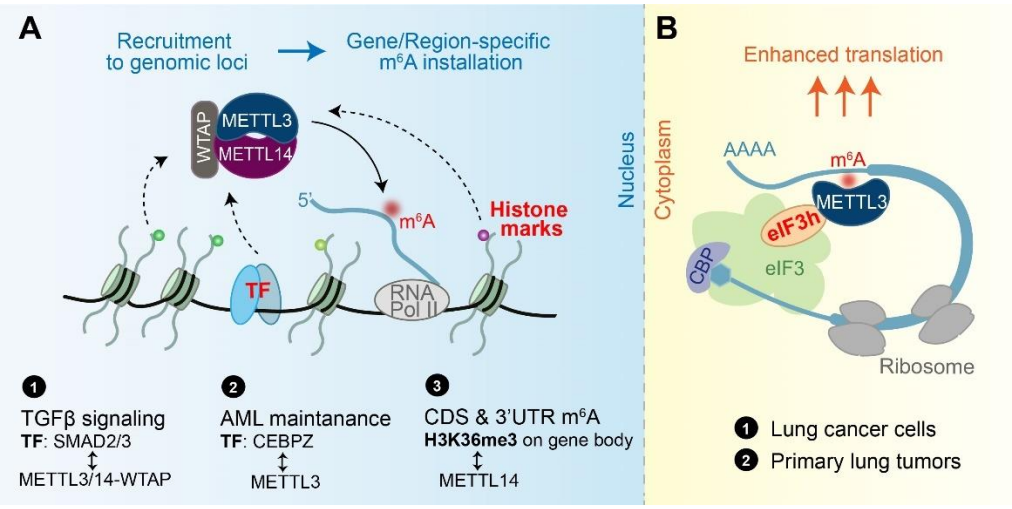

Figure 5.2 A model of cellular-localization dependent functions of m^{6}A writer proteins

(**A**) m^{6}A writer complex installs m^{6}A co-transcriptionally in the nucleus. The recruitment of the writer complex to specific genomic loci by transcription factors (TFs) or histone marks may contribute to the gene- or region-specificity in m^{6}A installation. (**B**) In the cytoplasm, METTL3 itself recognizes 3'UTR m^{6}A sites on mRNA and promotes protein translation from the transcript by facilitating translation loop formation through interaction with eIF3h.

METTL3 could be modulated through PTMs or those of its interacting proteins, which could affect protein stability, localization, writer complex formation, and writer catalytic activity. Human METTL14 has been reported to be phosphorylated at residue Serine399 which lies on the protein-protein interface with METTL3, suggesting regulatory functions[38]. SUMOylation sites have also been detected on four lysine residues of human METTL3 and were shown to reduce activity of METTL3/14 in in vitro methylation assays[184]. This repressive effect was mediated by a currently unknown mechanism that affects neither METTL3 stability, localization, nor interaction with METTL14/WTAP. Further investigations of protein partners of METTL3 in either the nucleus or the cytosol is likely to give more clues for deciphering PTM-mediated regulation of METTL3's functions, especially in a biological system where phosphorylation or SUMOylation pathways are misregulated.

5.2.2 FTO demethylation, cap m^6A_m, and internal m^6A

FTO, the first RNA demethylase identified, was reported to remove the methyl group of m^6A in mRNA both in vitro and inside cells[49]. FTO has also been reported to demethylate N^6, 2-*O*-dimethyladenosine (m^6A_m), a modification that has an identical chemical structure in the base moiety to m^6A and is found on the second base adjacent to the 5' cap (cap-m^6A_m) in a portion of mRNAs[19,185], both *in vitro*[186] and inside cells[52] (**Figure 5.3**). Our group has comprehensively characterized the substrate spectrum of FTO by carefully validating FTO binding targets from CLIP-seq results recently, confirming that FTO possesses effective demethylation activity towards m^1A in specific tRNAs, m^6A_m in some snRNAs, and internal m^6A and cap-m^6A_m in mRNA[51].

Cellular localization of FTO varies among several mammalian cell lines and the spatial regulation of FTO results in the distinct substrate preference in the nucleus *versus* the cytoplasm. FTO mediates mRNA m^6A and cap m^6A_m demethylation in cytoplasm, but mostly mRNA m^6A demethylation in cell nucleus[51], likely because the cap moiety is bound by cap-binding proteins and not accessible for demethylation. It is true that 5-10% of mRNA m^6A is subjected to the FTO-mediated demethylation in common cell lines such as HeLa and HEK293[49,51]—perhaps because nuclear mRNA accounts for a small portion of total cellular mRNA and they may not be adequately accessed by FTO. However, in some AML cells, FTO is highly elevated and mostly localizes in the cytoplasm. In these cells up to ~40% of all mRNA m^6A are subjected to demethylation by FTO[51,187,188]. The highly expressed FTO and its prevalent m^6A demethylation play an oncogenic role in these AML cells; inhibition of FTO by an oncometabolite 2-hydroxyglutarate (2HG) suppresses leukemia progression[187,188]. This effect is context-dependent, as FTO-high AML cells are more sensitive to 2HG inhibition.

Moreover, the abundance of internal m^6A is much higher than m^6A_m in the mRNA, with the ratio m^6A_m/m^6A ratios varying between 1/10 and 1/15 in HeLa, HEK293T, 3T3-L1, MEL624 and brain tissues, as quantified in several recent studies[51,134,189,190]. FTO mediates predominant demethylation of m^6A *versus* m^6A_m inside cells despite m^6A_m being a preferred substrate *in vitro*. So far, all functionally relevant reports of FTO have been consistent with internal m^6A but not cap m^6A_m in mRNA as the relevant substrate, including but not limited to facilitating the 3T3-L1 cell differentiation[88]; responding to heat shock[92] and UV damage[175], regulating the hepatitis C virus infection[191], and promoting the growth and self-renewal of glioblastoma stem cells[192].

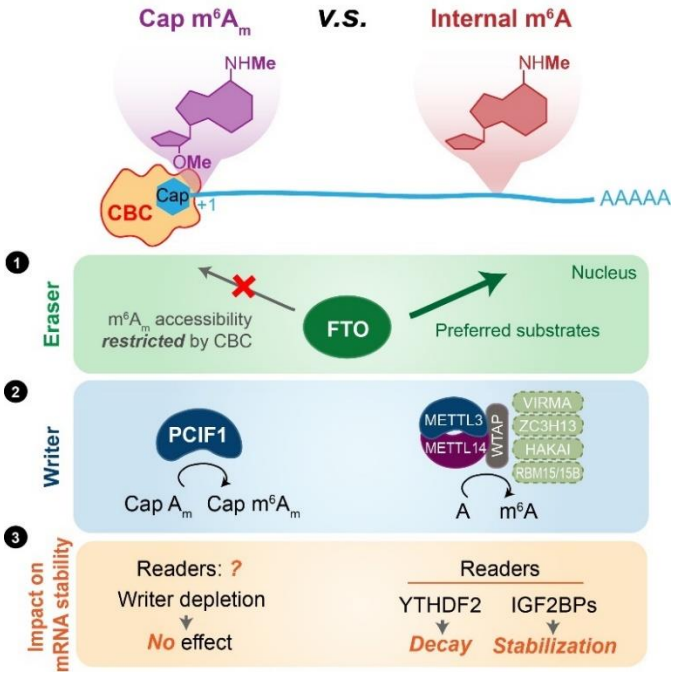

Figure 5.3 Distinct effectors and functions of cap-m^6A_m *versus* internal m^6A in mRNA

While FTO exhibits catalytic demethylation activity towards both cap-m^6A_m and internal m^6A in mRNA, in the cell nucleus, m^6A is the main substrate of FTO since cap-m^6A_m is most likely masked by cap-binding complex (CBC) in mRNA (①). The methyl group on the N^6 position of m^6A_m is installed by a cap-specific methyltransferase PCIF1 while internal m^6A is installed by METTL3-METTL14 in a multi-component writer complex (②). The role of m^6A on mRNA stability through the m^6A readers have been established. However, recent characterization of the m^6A_m writer PCIF1 showed that depletion of the writer had minimal effect on the abundance/stability of the cap-m^6A_m-modified mRNA in cell lines tested (③), arguing against the role of the FTO-mediated mRNA cap m^6A_m demethylation in affecting transcript stability.

The mechanistic study of cap-m^6A_m is still in its early stage. It has been proposed to protect mRNA from DCP2-mediated decapping and microRNA-mediated mRNA degradation[52]. The authors observed substantial changes in the levels of FTO target mRNAs when FTO was depleted, where they assigned the effect to cap-m^6A_m demethylation. This conclusion was not supported by a more careful analysis of the correlation between changes in expression levels of transcripts containing only m^6A or only m^6A_m upon FTO knockdown[51]. Only internal m^6A-modified transcripts showed correlation with the transcript level changes induced by FTO knockdown. The confusion was recently resolved when three groups independently identified PCIF1 as the mRNA cap-m^6A_m methyltransferase[189,193,194]. PCIF1 possess cap-m^6A_m but not internal methylation activity. Two research groups independently showed that the PCIF1-mediated deposition of cap-m^6A_m does not alter gene expression or transcript stability[189,193]. Thus, the transcript-stabilizing effect previously assigned to cap-m^6A_m upon FTO knockdown[52] is mostly derived from the stabilizing role of internal m^6As but not cap-m^6A_m.

Cap-m^6A_m was also proposed to promote translation efficiency of cap-m^6A_m-containing mRNA with the FTO knockdown leading to decreased translation of several target genes[52]. This result was recapitulated in one of the recent cap-m^6A_m methyltransferase studies[193]. However, another two studies reported that cap-m^6A_m has negligible effects on translation under basal conditions[194], and even suppresses translation based on *in vitro* translation assays and *in vivo* reporter assays[189]. The effect of cap-m^6A_m on translation could also be context dependent and has yet to be clarified.

5.3 Future directions on the research of epitranscriptomics

Despite established biological significance of m^6A epitranscriptomics and mechanistic insights on the function of m^6A from readers-oriented studies, there are still key questions remaining to be solved.

1. Specificity of m^6A methylation. What are the molecular determinants for m^6A to be installed or removed on a transcript, or more specifically, on a certain region of the transcript? Is there any information flow between epigenetic marks and epitranscriptomics marks? How do cellular signaling pathways and m^6A coding regulate each other? To answer these questions, previous studies have attempted to examine writers' binding events on the chromatin during cellular response processes or in disease models and to integrate those binding events with distinct transcription factors, epigenetic modifications, or gene expression profiles in those systems (see Section **5.2.1**). The correlation between the gene-body-enriched active histone mark H3K36me3 and m^6A seems to give a decent explanation for the preferential methylation in the CDS and 3'UTR[183]. However, we still do not fully understand why some of the transcripts are methylated and some are not. Even within the methylated transcripts, it is possible that "constitutively" methylated transcripts and "context-specifically" (e.g. cell types, stress response) methylated ones are driven by different writer recruitment mechanisms. Note that in neurons, the writer depletion results in upregulation and downregulation of unique groups of genes (Section **3.2.4**). This difference in downstream processing may already be coded in how methylation occurs.

2. Specificity of target recognition of m^6A readers. Besides the methyl group on the adenosine, what elements on the transcript contribute to the recruitment of a certain reader protein? What if multiple reader proteins share the same binding site—do they bind different copies of a same RNA sequences, do they bind at a different time or location, do they compete for binding, or

do they bind as a complex? The lack of information at high temporal-spatial resolution or single-molecule level could make it challenging to interpret the outcome of reader depletion. An alternative strategy is to bring reader proteins close to endogenous or reporter transcripts. An RNA-targeting platform has been developed to study the precise functions of readers on specific transcripts by fusing the catalytically inactive Cas13b to an m^6A reader[195]. This platform is capable of recapitulating the function of YTHDF2 to promote decay and that of YTHDF1 to promote protein production. Developing additional orthogonal RNA targeting platforms would be quite beneficial for studying crosstalk between the reader proteins.

3. Moonlighting functions of m^6A effectors. Take m^6A writers as an example. (1) The majority of m^6A studies so far has focused on protein-coding transcripts, microRNAs, and a few long-non-coding RNA species, which only account for a small percentage of the genome. Does m^6A methylation occur on other RNA species, and if so, what functions does m^6A play therein? Many of the nuclear and chromatin-associated RNA species are short-lived and their functions have only become appreciated in recent years. It is possible that for those transcripts m^6A is co-transcriptionally installed as well and regulates their functions. (2) Moreover, the non-catalytic functions of the m^6A writers and erasers are also worthy of attention. METTL3 may localize in the cytoplasm to promote translation by interacting with eIF3h[179] (**Figure 5.2B**), calling for more careful investigation of METTL3's interactome for deciphering its potential moonlight functions.

4. Epitranscriptomic marks beyond m^6A. As introduced in Section **1.2** and **Figure 1.2**, mRNAs are decorated with many types of chemical modifications besides m^6A. (1) The writers for some of the mRNA modifications have been indicated from the screening predicted or known non-coding-RNA-modifying enzymes. For example, a subset of nuclear transcripts containing a tRNA-T-loop-like moiety were found to be methylated by tRNA m^1A writer complex

TRMT6/61A in the sequence context of GUUCRA (R = A or G), and m^1A on mitochondrial (mt) coded mRNA were found to be sensitive to both of the known mt-tRNA m^1A writers TRMT61B and TRMT10C[27,28]. The ablation of *N*-acetyltransferase 10 (NAT10), a known tRNA/rRNA ac^4C writer, reduces mRNA ac^4C level[33]. However, the dual-function of the writer makes it difficult to pinpoint the functional contribution from modified sites on mRNA, especially in protein translation where mRNA, tRNA and rRNA all come into play. Identifying specific readers may help clarify this situation. (2) The variety of mRNA modifications also raises the possibility that functional information can be stored in combinations of marks instead of each single one. Techniques that are capable of detecting multiple marks on a transcript at the same time would be highly desirable. Of interest, Dai *et. al.* reported that YTHDF proteins can also be pulled down by an m^1A-containing probe in a CC(m^1A)GG motif and the in vitro affinity for the probe is ~5-20-folds lower than that of its m^6A counterpart[196]. While this observation might be a "side effect" of the Dimroth Rearrangement of m^1A into m^6A, it also proposed a potential model of crosstalk between different mRNA modifications.

5. Towards subcellular epitranscriptomics. As exemplified by FTO (see Section **5.2.2**), the physiologically relevant substrates of an epitranscriptomic enzyme could be largely determined by the enzyme's subcellular localization. Subcellular organelles may enrich certain modifications or modifying enzymes, the functional outcome of which would be much diluted when the whole cell is examined. This topic is also crucial for highly polarized cell types such as neurons. Investigation of m^6A profiles in the synaptosome have revealed significant links between m^6A and transcript functional partitioning[133]. There will be much more to discover by combining current epitranscriptomic methodology with biochemistry-, microfluidics-, or super-resolution-imaging-based techniques that distill signals from desired cellular compartments.

List of References

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

2. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* **128,** 635–638 (2007).

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

4. Jones, P. A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–492 (2012).

5. Breiling, A. & Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics and Chromatin* **8,** 1–9 (2015).

6. Lawrence, M., Daujat, S. & Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* **32,** 42–56 (2016).

7. Audia, J. E. & Campbell, R. M. Histone modifications and cancer markers. *Cold Spring Harb. Perspect. Biol.* **8,** 1–32 (2017).

8. He, C. Grand Challenge Commentary: RNA epigenetics? *Nat. Chem. Biol.* **6,** 863–865 (2010).

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

10. Boccaletto, P. *et al.* MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46(D1),** D303–D307 (2018).

11. Cohn, W. E. & Volkin, E. Nucleoside-5'-Phosphates from Ribonucleic Acid. *Nature* **167,** 483–484 (1951).

12. Pan, T. Modifications and functional genomics of human transfer RNA. *Cell Res.* **28,** 395–404 (2018).

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

14. Frye, M., Harada, B. T., Behm, M. & He, C. RNA modifications modulate gene expression during development. *Science* **361,** 1346–1349 (2018).

15. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *PNAS* **71,** 3971–3975 (1974).

16. Perry, R. P. & Kelley, D. E. Existence of Methylated Messenger RNA in Mouse L Cells. *Cell* **1,** 37–42 (1974).

17. Schibler, U., Kelley, D. E. & Perry, R. P. Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. *J. Mol. Biol.* **115,** 695–714 (1977).

18. Wei, C. & Moss, B. Nucleotide Sequences at the N6-Methyladenosine Sites of HeLa Cell Messenger Ribonucleic Acid. *Biochemistry* **16,** 1672–1676 (1977).

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

20. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–206 (2012).

21. Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12,** 767–772 (2015).

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

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

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

25. Dominissini, D. *et al.* The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530,** 441–446 (2016).

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

27. Li, X. *et al.* Base-Resolution Mapping Reveals Distinct m1A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Mol. Cell* **68,** 993–1005.e9 (2017).

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

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

30. Yang, X. *et al.* 5-methylcytosine promotes mRNA export-NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell Res.* **27,** 606–625 (2017).

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

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

33. Arango, D. *et al.* Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell* **175,** 1872–1886.e24 (2018).

34. Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G. & Rottman, F. M. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA* **3,** 1233–1247 (1997).

35. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10,** 93–5 (2014).

36. Śledź, P. & Jinek, M. Structural insights into the molecular mechanism of the m6A writer complex. *Elife* **5,** 1–16 (2016).

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

38. Wang, X. *et al.* Structural basis of N6-adenosine methylation by the METTL3–METTL14 complex. *Nature* **534,** 575–578 (2016).

39. Ping, X.-L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell research* **24,** 177–89 (2014).

40. Zhong, S. *et al.* MTA Is an Arabidopsis Messenger RNA Adenosine Methylase and Interacts with a Homolog of a Sex-Specific Splicing Factor. *PLANT CELL ONLINE* **20,** 1278–1288 (2008).

41. Yue, Y. *et al.* VIRMA mediates preferential m6A mRNA methylation in 3′UTR and near stop codon and associates with alternative polyadenylation. *Cell Discov.* **4,** 1–17 (2018).

42. Wen, J. *et al.* Zc3h13 Regulates Nuclear RNA m6A Methylation and Mouse Embryonic Stem Cell Self-Renewal. *Mol. Cell* **69,** 1028–1038.e6 (2018).

43. Patil, D. P. *et al.* M6 A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537,** 369–373 (2016).

44. Knuckles, P. *et al.* Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA- binding factor Rbm15/Spenito to the m6A machinery component Wtap/Fl(2)d. *Genes Dev.* **32,** 415–429 (2018).

45. Pendleton, K. E. *et al.* The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **169,** 824–835.e14 (2017).

46. Doxtader, K. A. *et al.* Structural Basis for Regulation of METTL16, an S-Adenosylmethionine Homeostasis Factor. *Mol. Cell* **71,** 1001–1011.e4 (2018).

47. Mendel, M. *et al.* Methylation of Structured RNA by the m6A Writer METTL16 Is Essential for Mouse Embryonic Development. *Mol. Cell* **71,** 986–1000.e11 (2018).

48. Shima, H. *et al.* S-Adenosylmethionine Synthesis Is Regulated by Selective N6-Adenosine Methylation and mRNA Degradation Involving METTL16 and YTHDC1. *Cell Rep.* **21,** 3354–3363 (2017).

49. Jia, G. *et al.* N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7,** 885–887 (2011).

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

51. Wei, J. *et al.* Differential m6A, m6Am, and m1A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Mol. Cell* **71,** 973–985.e5 (2018).

52. Mauer, J. *et al.* Reversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nature* **541,** 371–375 (2017).

53. Stoilov, P., Rafalska, I. & Stamm, S. YTH: A new domain in nuclear proteins. *Trends Biochem. Sci.* **27,** 495–497 (2002).

54. Zhang, Z. *et al.* The YTH domain is a novel RNA binding domain. *J. Biol. Chem.* **285,** 14701–14710 (2010).

55. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–20 (2014).

56. Du, H. *et al.* YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. *Nat. Commun.* **7,** 12626 (2016).

57. Park, O. H. *et al.* Endoribonucleolytic Cleavage of m6A-Containing RNAs by RNase P/MRP Complex. *Mol. Cell* 1–14 (2019). doi:10.1016/j.molcel.2019.02.034

58. Wang, X. *et al.* N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161,** 1388–1399 (2015).

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

60. Li, A. *et al.* Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. *Cell Res.* **27,** 444–447 (2017).

61. Yang, Y. *et al.* Extensive translation of circular RNAs driven by N 6 -methyladenosine. *Cell Res.* **27,** 626–641 (2017).

62. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* **61,** 507–519 (2016).

63. Roundtree, I. A. *et al.* YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *Elife* **6,** 1–28 (2017).

64. Hsu, P. J. *et al.* Ythdc2 is an N6-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Res.* **27,** 1115–1127 (2017).

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

66. Alarcón, C. R. *et al.* HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell* **162,** 1299–308 (2015).

67. Liu, N. *et al.* N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res.* **45,** 6051–6063 (2017).

68. Wu, B. *et al.* Molecular basis for the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1. *Nat. Commun.* **9,** 1–12 (2018).

69. Edupuganti, R. R. *et al.* N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* **24,** 870–878 (2017).

70. Zhang, F. *et al.* Fragile X mental retardation protein modulates the stability of its m6A-marked messenger RNA targets. *Hum. Mol. Genet.* **27,** 3936–3950 (2018).

71. Huang, H. *et al.* Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat. Cell Biol.* **20,** 285–295 (2018).

72. Wu, R. *et al.* A novel m6A reader Prrc2a controls oligodendroglial specification and myelination. *Cell Res.* **29,** 23–41 (2019).

73. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169,** 1187–1200 (2017).

74. Hsu, P. J., Shi, H. & He, C. Epitranscriptomic influences on development and disease. *Genome Biol.* **18,** (2017).

75. Tan, B. & Gao, S. J. RNA epitranscriptomics: Regulation of infection of RNA and DNA viruses by N6-methyladenosine (m6A). *Rev. Med. Virol.* **28,** e1983 (2018).

76. Batista, P. J. *et al.* M6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15,** 707–719 (2014).

77. Geula, S. *et al.* m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347,** 1002–6 (2015).

78. Wang, Y. *et al.* N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* **16,** 191–8 (2014).

79. Zhao, B. S. *et al.* m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **542,** 475–478 (2017).

80. Ivanova, I. *et al.* The RNA m6A Reader YTHDF2 Is Essential for the Post-transcriptional

Regulation of the Maternal Transcriptome and Oocyte Competence. *Mol. Cell* **67,** 1059–1067.e4 (2017).

81. Li, H.-B. *et al.* m6A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. *Nature* **548,** 338–342 (2017).

82. Zhang, C. *et al.* m6A modulates haematopoietic stem and progenitor cell specification. *Nature* **549,** 273–276 (2017).

83. Lin, Z. *et al.* Mettl3-/Mettl14-mediated mRNA N6-methyladenosine modulates murine spermatogenesis. *Cell Res.* **27,** 1–15 (2017).

84. Ki-Jun Yoon, A. *et al.* Temporal Control of Mammalian Cortical Neurogenesis by m6A Methylation. *Cell* **171,** 877–889.e12 (2017).

85. Haussmann, I. U. *et al.* M6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* **540,** 301–304 (2016).

86. Lence, T. *et al.* m6A modulates neuronal functions and sex determination in Drosophila. *Nature* **540,** 242–247 (2016).

87. Klukovich, R. *et al.* ALKBH5-dependent m6A demethylation controls splicing and stability of long 3′-UTR mRNAs in male germ cells. *Proc. Natl. Acad. Sci.* **115,** E325–E333 (2017).

88. Zhang, M. *et al.* The demethylase activity of FTO (Fat mass and obesity associated protein) is required for preadipocyte differentiation. *PLoS One* **10,** 1–15 (2015).

89. Zhao, X. *et al.* FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.* **24,** 1403–19 (2014).

90. Wang, X. *et al.* FTO is required for myogenesis by positively regulating mTOR-PGC-1α pathway-mediated mitochondria biogenesis. *Cell Death Dis.* **8,** e2702 (2017).

91. Li, L. *et al.* Fat mass and obesity-associated (FTO) protein regulates adult neurogenesis. *Hum. Mol. Genet.* **26,** 2398–2411 (2017).

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

93. Weng, Y.-L. *et al.* Epitranscriptomic m 6 A Regulation of Axon Regeneration in the Adult Mammalian Nervous System. *Neuron* **97,** 313–325.e6 (2018).

94. Eliseeva, I. a, Kim, E. R., Guryanov, S. G., Ovchinnikov, L. P. & Lyabin, D. N. Y-Box-Binding Protein 1 and Its Functions. *Biochemistry* **76,** 1402–1433 (2011).

95. Le Hir, H. & Séraphin, B. EJCs at the Heart of Translational Control. *Cell* **133,** 213–216 (2008).

96. Jao, C. Y. & Salic, A. Exploring RNA transcription and turnover in vivo by using click

chemistry. *Proc. Natl. Acad. Sci.* **105,** 15779–15784 (2008).

97. Coots, R. A. *et al.* m6A Facilitates eIF4F-Independent mRNA Translation. *Mol. Cell* **68,** 504–514.e7 (2017).

98. Lu, Y., He, X. & Zhong, S. Cross-species microarray analysis with the OSCAR system suggests an INSR->Pax6->NQO1 neuro-protective pathway in aging and Alzheimer's disease. *Nucleic Acids Res.* **35,** 105–114 (2007).

99. Zhang, Y. *et al.* RNA-binding protein YTHDF3 suppresses interferon-dependent antiviral responses by promoting FOXO3 translation. *Proc. Natl. Acad. Sci.* **116,** 976–981 (2018).

100. Bertrand, E. *et al.* Localization of ASH1 mRNA Particles in Living Yeast. *Mol. Cell* **2,** 437–445 (1998).

101. Gehring, N. H., Neu-Yilik, G., Schell, T., Hentze, M. W. & Kulozik, A. E. Y14 and hUpf3b form an NMD-activating complex. *Mol. Cell* **11,** 939–949 (2003).

102. Guet, C. C. *et al.* Minimally invasive determination of mRNA concentration in single living bacteria. *Nucleic Acids Res.* **36,** (2008).

103. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324,** 218–23 (2009).

104. Gandin, V. *et al.* Polysome fractionation and analysis of mammalian translatomes on a genome-wide scale. *J. Vis. Exp.* **87,** 51455 (2014).

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

106. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).

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

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

109. 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).

110. 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,** 233–7 (2012).

111. Huang, D. W., Lempicki, R. a & Sherman, B. T. Systematic and integrative analysis of large

gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).

112. 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).

113. Fran Supek , Matko Bošnjak, Nives Škunca, T. Š. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6,** (2011).

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

115. Slobodin, B. *et al.* Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine. *Cell* **169,** 326–337 (2017).

116. 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).

117. Widagdo, J. *et al.* Experience-Dependent Accumulation of N6 -Methyladenosine in the Prefrontal Cortex Is Associated with Memory Processes in Mice. *J. Neurosci.* **36,** 6771–6777 (2016).

118. Walters, B. J. *et al.* The Role of The RNA Demethylase FTO (Fat Mass and Obesity-Associated) and mRNA Methylation in Hippocampal Memory Formation. *Neuropsychopharmacology* **42,** 1502–1510 (2017).

119. Sutton, M. A. & Schuman, E. M. Dendritic Protein Synthesis, Synaptic Plasticity, and Memory. *Cell* **127,** 49–58 (2006).

120. Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11,** 399–402 (2014).

121. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445,** 168–176 (2007).

122. Morris, R. G. M., Garrud, P., Rawlins, J. N. P. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297,** 681–683 (1982).

123. LeDoux, J. E. Emotion Circuits in the Brain. *Annu. Rev. Neurosci.* **23,** 155–184 (2000).

124. Schmidt, E. K., Clavarino, G., Ceppi, M. & Pierre, P. SUnSET, a nonradioactive method to monitor protein synthesis. *Nat. Methods* **6,** 275–277 (2009).

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

126. Ma, D. K. *et al.* Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* **323,** 1074–1077 (2009).

127. Koranda, J. L. *et al.* Mettl14 Is Essential for Epitranscriptomic Regulation of Striatal Function and Learning. *Neuron* **99,** (2018).

128. Krichevsky, A. M. & Kosik, K. S. Neuronal RNA granules: A link between RNA localization and stimulation-dependent translation. *Neuron* **32,** 683–696 (2001).

129. Kiebler, M. A. & Bassell, G. J. Neuronal RNA Granules: Movers and Makers. *Neuron* **51,** 685–690 (2006).

130. Hanson, G. & Coller, J. Translation and Protein Quality Control: Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.* **19,** 20–30 (2018).

131. Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P. & Heintz, N. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat. Protoc.* **9,** 1282–1291 (2014).

132. Doyle, J. P. *et al.* Application of a Translational Profiling Approach for the Comparative Analysis of CNS Cell Types. *Cell* **135,** 749–762 (2008).

133. Merkurjev, D. *et al.* Synaptic N6-methyladenosine (m6A) epitranscriptome reveals functional partitioning of localized transcripts. *Nat. Neurosci.* **21,** 1–11 (2018).

134. Engel, M. *et al.* The Role of m6A/m-RNA Methylation in Stress Response Regulation. *Neuron* **99,** 389–403.e9 (2018).

135. Han, D. *et al.* Anti-tumour immunity controlled through mRNA m 6 A methylation and YTHDF1 in dendritic cells. *Nature* **566,** 270–274 (2019).

136. Chiu, S. L. *et al.* GRASP1 Regulates Synaptic Plasticity and Learning through Endosomal Recycling of AMPA Receptors. *Neuron* **93,** 1405–1419.e8 (2017).

137. Shioda, N. *et al.* Aberrant calcium/calmodulin-dependent protein kinase II (CaMKII) activity is associated with abnormal dendritic spine morphology in the ATRX mutant mouse brain. *J Neurosci* **31,** 346–358 (2011).

138. Su, Y. *et al.* Neuronal activity modifies the chromatin accessibility landscape in the adult brain. *Nat. Neurosci.* **20,** 476–483 (2017).

139. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9,** 171–181 (2014).

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

141. 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).

142. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10 (2011).

143. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

144. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

145. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

146. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** (2014).

147. Dominguez, D. *et al.* Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol. Cell* **70,** 854–867.e9 (2018).

148. Tourrière, H. *et al.* The RasGAP-associated endoribonuclease G3BP assembles stress granules. *J. Cell Biol.* **160,** 823–831 (2003).

149. Xu, C. *et al.* Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* **10,** 927–929 (2014).

150. Xu, C. *et al.* Structural basis for the discriminative recognition of N6-Methyladenosine RNA by the human YT521-B homology domain family of proteins. *J. Biol. Chem.* **290,** 24902–24913 (2015).

151. Irvine, K., Stirling, R., Hume, D. & Kennedy, D. Rasputin, more promiscuous than ever: A review of G3BP. *Int. J. Dev. Biol.* **48,** 1065–1077 (2004).

152. Tani, H. *et al.* Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* **22,** 947–956 (2012).

153. Myrick, L. K., Hashimoto, H., Cheng, X. & Warren, S. T. Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. *Hum. Mol. Genet.* **24,** 1733–1740 (2015).

154. De Boulle, K. *et al.* A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat. Genet.* **3,** 31–35 (1993).

155. Ascano, M. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. TL - 492. *Nature* **492,** 382–386 (2012).

156. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146,** 247–261 (2011).

157. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA

target sites by PAR-CLIP. *Cell* **141,** 129–140 (2010).

158. Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159 (2013).

159. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372 (2008).

160. Schwanhäusser, B., Gossen, M., Dittmar, G. & Selbach, M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9,** 205–209 (2009).

161. Wiśniewski, J. R., Zougman, A. & Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* **8,** 5674–5678 (2009).

162. Visscher, M. *et al.* Proteome-wide Changes in Protein Turnover Rates in C. elegans Models of Longevity and Age-Related Disease. *Cell Rep.* **16,** 3041–3051 (2016).

163. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34,** 525–527 (2016).

164. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34,** D590–D598 (2006).

165. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).

166. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12,** 357–360 (2015).

167. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).

168. Meyer, K. D. *et al.* 5′ UTR m6A Promotes Cap-Independent Translation. *Cell* **163,** 999–1010 (2015).

169. Choi, J. *et al.* N6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. *Nat. Struct. Mol. Biol.* **23,** 110–115 (2016).

170. Jain, S. *et al.* ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164,** 487–498 (2016).

171. Mardakheh, F. K. *et al.* Global Analysis of mRNA, Translation, and Protein Localization: Local Translation Is a Key Regulator of Cell Protrusions. *Dev. Cell* **35,** 344–357 (2015).

172. Hao, H. *et al.* N6-methyladenosine modification and METTL3 modulate enterovirus 71 replication. *Nucleic Acids Res.* **47,** 362–374 (2019).

173. Shi, H. *et al.* m6A facilitates hippocampus-dependent learning and memory through YTHDF1. *Nature* **563,** 249–253 (2018).

174. Knuckles, P. *et al.* RNA fate determination through cotranscriptional adenosine methylation and microprocessor binding. *Nat. Struct. Mol. Biol.* **24,** 561–569 (2017).

175. Xiang, Y. *et al.* RNA m6 A methylation regulates the ultraviolet-induced DNA damage response. *Nature* **543,** 573–576 (2017).

176. Schöller, E. *et al.* Interactions, localization and phosphorylation of the m6A generating METTL3-METTL14-WTAP complex. *RNA* **24,** 499–512 (2018).

177. Chen, T. *et al.* M6A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. *Cell Stem Cell* **16,** 289–301 (2015).

178. Lin, S. *et al.* The m6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol. Cell* **62,** 335–345 (2016).

179. Choe, J. *et al.* mRNA circularization by METTL3–eIF3h enhances translation and promotes oncogenesis. *Nature* **561,** 556–561 (2018).

180. Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S. F. N6-methyladenosine marks primary microRNAs for processing. *Nature* **519,** 482–485 (2015).

181. Barbieri, I. *et al.* Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. *Nature* **552,** 126–131 (2017).

182. Bertero, A. *et al.* The SMAD2/3 interactome reveals that TGFβ controls m 6 A mRNA methylation in pluripotency. *Nature* **555,** 256–259 (2018).

183. Huang, H. *et al.* Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-transcriptionally. *Nature* **567,** 414–419 (2019).

184. Du, Y. *et al.* SUMOylation of the m6A-RNA methyltransferase METTL3 modulates its function. *Nucleic Acids Res.* **46,** 5195–5208 (2018).

185. Adams, J. M. & Cory, S. Modified nucleosides and bizarre 5′-termini in mouse myeloma mRNA. *Nature* **255,** 28–33 (1975).

186. Fu, Y. Dynamic Regulation of RNA Modifications by AlkB Family Dioxygenases. In Chemistry Department (Proquest, Order No. 3548228: The University of Chicago). (2012).

187. Li, Z. *et al.* FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N6 -Methyladenosine RNA Demethylase. *Cancer Cell* **31,** 127–141 (2017).

188. Su, R. *et al.* R-2HG Exhibits Anti-tumor Activity by Targeting FTO/m6A/MYC/CEBPA Signaling. *Cell* **172,** 90–105.e123 (2018).

189. Sendinc, E. *et al.* PCIF1 catalyzes m6Am mRNA methylation to regulate gene expression.

bioRxiv **484931,** (2018).

190. Sun, H., Zhang, M., Li, K., Bai, D. & Yi, C. Cap-specific, terminal N6-methylation by a mammalian m6Am methyltransferase. *Cell Res.* **29,** 80–82 (2019).

191. Gokhale, N. S. *et al.* N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection. *Cell Host Microbe* **20,** 654–665 (2016).

192. Cui, Q. *et al.* m6A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. *Cell Rep.* **18,** 2622–2634 (2017).

193. Akichika, S. *et al.* Cap-specific terminal N6-methylation of RNA by an RNA polymerase II–associated methyltransferase. *Science* **363,** eaav0080 (2019).

194. Boulias, K. *et al.* Identification of the m6Am methyltransferase PCIF1 reveals the location and functions of m6Am in the transcriptome. *bioRxiv* **485862,** (2018).

195. Rauch, S., He, C. & Dickinson, B. C. Targeted m6A Reader Proteins to Study Epitranscriptomic Regulation of Single RNAs. *J. Am. Chem. Soc.* **140,** 11974–11981 (2018).

196. Dai, X., Wang, T., Gonzalez, G. & Wang, Y. Identification of YTH Domain-Containing Proteins as the Readers for N1-Methyladenosine in RNA. *Anal. Chem.* **90,** 6380–6384 (2018).