THE UNIVERSITY OF CHICAGO

ROLES OF mRNA METHYLATIONS IN BIOLOGY:
PROCESSING AND DEVELOPMENT

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:

COMMITTEE ON IMMUNOLOGY

BY

PHILLIP J. HSU

CHICAGO, ILLINOIS

DECEMBER 2018

Table of Contents

Acknowledgement

To begin, I would like to express my deep gratitude to my doctorate advisor, Prof. Chuan He. Chuan has been a constant source of encouragement, unfailing support, and new ideas. His enthusiasm for exploring novel areas of research and ability to ask the important questions consistently motivated me to learn to do the same. Moreover, his willingness to let me explore my own research interests, and his helpfulness even when results were negative – allowed me to grow as a scientist. Most importantly, his optimism and warmth as a mentor were contagious, creating a lab environment in which work became a joy.

I would like to thank my fellow members of the He lab – a family with whom I cherished doing life together daily. Whether we were collaborating on experiments, discussing puzzling results, or enjoying the great food Chicago has to offer, these many individuals made even the most disappointing data more endurable and the sweetest moments more relishable. I admired how hard-working and talented they were as scientists and how generous and thoughtful they were as people. Specifically, I would like to extend my gratitude to Dr. Honghui Ma, with whom I worked closely in my first year and half, for his caring mentorship and for teaching me the foundational research skills that would shape the rest of my graduate work. I would also like to thank Linda Zhang for working with me through two challenging projects, as well as several collaborations with other groups – making work both more productive and enjoyable. Thank you to Dr. Zhike Lu, Scott Zijie Zhang, and Dr. Guanzheng Luo for helping me time after time with bioinformatics questions. Thank you also to Dr. Zhong Zheng for working together on exciting immunology collaborations. I am grateful to Dr. Ian Roundtree, Dr. Siggy Nachtergaele, and Dr. Thomas Lu for guiding me through grants, conferences, publishing, and the other aspects of academia. My thanks go to Hailing Shi for helping me through experiments, for thoughtful suggestions, and for

being a good friend. Additionally, I would like to thank Tong Wu, Allen Zhu, Cody He, and Caraline Sepich for help in the lab and for companionship. I am grateful for all lab members, both past and present, and will sincerely miss working alongside you.

I would also like to thank our research collaborators. Prof. Bin Shen and Yunfei Zhu worked closely with me on the YTHDC2 project, and Prof. Yongchao Ma and Prof. Yanhong Shi collaborated with us on the FMRP project. Prof. Seungmin Hwang, Prof. Michaela Gack, Prof. Bana Jabri, Dr. Scott Biering, Dr. Konstantin Sparrer, Will Riedl, and Dr. Reinhard Hinterleitner collaborated with virus experiments. These collaborations helped expand the breadth of my learning as a researcher.

Thank you to my thesis committee – Prof. Tao Pan, Prof. Marcus Clark, and Prof. Seungmin Hwang, for the time and effort it took to meet with me and advise me, and for your wisdom through both group and individual meetings.

I am thankful for the University of Chicago MSTP leadership, both past and present, who provided constant guidance and advocated for me to allow me to pursue research that is interdisciplinary in the truest sense. Marcus, Elise, Shay, Sarah, Alison, Shanetha, and Kristin made up a team that selflessly supported me through all of the challenges involved in graduate training. I would also like to thank the NIH for pre-doctoral funding.

Dr. Karen Kim has been an invaluable mentor throughout my medical and graduate training. I am grateful for her wisdom in helping shape my career path, encouragement in research, and the immense amount of advice she provided for pursuits both in and out of the lab.

Thank you to my classmates for your friendship these past four years. I am consistently impressed by what you do, and I can’t wait to see the great things you will achieve in the future. I

would also like to express my thankfulness to all the great people I met off the job, especially my CCUC family and Bridgeport team.

Lastly, thank you to my parents Hawpeng and Shuling Hsu, and my brother Eric, for loving me unceasingly, for supporting me through late nights and long phone calls, and for encouraging me in seasons of doubt and setbacks. Thank you for setting an example in faith, for sharing the wisdom I needed so often, and for paving the way ahead.

Abstract

Biological function relies on the precise control of gene expression. Genetic information is stored and transferred in the nucleic acids DNA and RNA, which are each composed of four canonical bases. While the sequence of bases in DNA and RNA are essential, they are not sufficient to convey genetic information at the level of intricacy needed to support biological function. Epigenetic modifications modify the flow of genetic information without changing nucleic acid sequence. Like epigenetic modifications on DNA and protein, post-transcriptional chemical modifications on RNA play important functional roles, governing RNA function and metabolism to control gene expression. N^6-methyladenosine (m^6A), the most abundant post-transcriptional mRNA modification in eukaryotes, exerts many of its effects on gene regulation through reader proteins that bind specifically to m^6A-containing transcripts. This work characterizes the m^6A binding specificities and biological functions of two proteins: YTH domain containing 2 (YTHDC2) and fragile X mental retardation protein (FMRP). We report that YTHDC2 and FMRP are readers of m^6A that affect translation efficiency and nuclear export, respectively, and are essential for mammalian spermatogenesis and neural development.

List of Publications Based on Work Described in this Thesis

1. Yang Y*, Hsu PJ*, Chen YS, Yang Y. Dynamic transcriptomic m^6A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Research* **28**, 616-624 (2018). [Review]

2. Hsu PJ, He C. Making Changes: N^6-Methyladenosine-Mediated Decay Drives the Endothelial-to-Hematopoietic Transition. *Biochemistry* **56**, 6077-6078 (2017). [Viewpoint]

3. Hsu PJ, Shi H, He C. Epitranscriptomic influences on development and disease. *Genome Biology* **18**, 197 (2017). [Review]

4. Lin Z, Hsu PJ, Xing X, Fang J, Lu Z, Zou Q, Zhang KJ, Zhang X, Zhou Y, Zhang T, Zhang Y, Song W, Jia G, Yang X, He C, Tong MH. Mettl3-/Mettl14-mediated mRNA N^6-methyladenosine modulates murine spermatogenesis. *Cell Research* **27**, 1216-1230 (2017).

5. Edupuganti RR, Geiger S, Lindeboom RGH, Shi H, Hsu PJ, Lu Z, Wang SY, Baltissen MPA, Jansen PWTC, Rossa M, Müller M, Stunnenberg HG, He C, Carell T, Vermeulen M. N^6-methyladenosine (m^6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat Struct Mol Biol* **24**, 870-878 (2017).

6. Hsu PJ, Zhu Y, Ma H, Guo Y, Shi X, Liu Y, Qi M, Lu Z, Shi H, Wang J, Cheng Y, Luo G, Dai Q, Liu M, Guo X, Sha J, Shen B, He C. Ythdc2 is an N^6-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Research* **27**, 1115-1127 (2017).

7. Shi H, Wang X, Lu Z, Zhao BS, Ma H, Hsu PJ, Liu C, He C. YTHDF3 facilitates translation and decay of N^6-methyladenosine-modified RNA. *Cell Research* **27**, 315-328 (2017).

Book Chapters

8. Hsu PJ, He C. High-resolution mapping of N^6-methyladenosine using m^6A crosslinking immunoprecipitation sequencing (m^6A-CLIP-seq). *Methods Mol Biol* (2019) (In press).

9. Hsu PJ, He C. Identifying the m^6A Methylome by Affinity Purification and Sequencing. *Methods Mol Biol* **1649**, 49-57 (2018).

List of Figures

Chapter 1 – Introduction[1]

1.1 Epigenetics: elaborating on a sequence

Biological function is driven by the flow of genetic information, which is stored and transferred in the nucleic acids DNA and RNA. DNA contains segments of genes, the functional units defining heredity, which operate as templates for RNA. RNA may then dictate the expression of proteins, which perform cellular functions and allow organisms to live within the world around them.

Nucleic acids are composed of four canonical bases: A/T/C/G in DNA, and A/U/C/G in RNA. The respective sequences of DNA and RNA, while important, are not sufficient to guide biological function at the level of intricacy present within cells. Instead, another layer of control is necessary, one that elaborates upon the message conveyed by the nucleotide sequence while keeping the sequence unchanged. This layer of control was first named the "epigenotype" in 1942 by English geneticist C. H. Waddington, who used it to describe the complex pathways connecting genotype with phenotype[1]. Waddington proposed that this field of "epigenetics" should elucidate mechanisms by which environmental signals heritably affect gene expression and, thus, adult characteristics. A more recent definition of epigenetics is "information heritable during cell division other than the DNA sequence itself"[2]. Together, these two definitions effectively describe a set of mechanisms by which nucleotides may be modified, at times heritably, to enact specific cellular processes.

[1] Parts of this introduction are adapted from Hsu, P.J., Shi, H., He C. Epitranscriptomic influences on development and disease. *Genome Biology* **18**(1):197 (2017). With modifications.

Epigenetic modifications on DNA and histones have been known to play essential roles in controlling gene expression. In DNA, methylations such as 5-methylcytosine (5mC) act prominently as silencing marks on promotors and enhancers, but may also play diverse roles dependent upon cellular context, at times being targeted to the bodies of active genes[3]. Histones contain a wide range of modifications, which influence transcription and affect biological processes including differentiation, neurological disorders, and cancers[4]. Notably, these modifications may be removed by enzymes such as demethylases, allowing the cell greater control of which regions are modified. Similarly, enzymatic cellular machinery is required to install epigenetic modifications, allowing the establishment of an epigenetic landscape specific to each cell and its environment. Altogether, epigenetic modifications provide cells with exquisite control of gene expression, allowing each cell to put into effect a unique set of processes from the same genetic sequence.

1.2 The emerging epitranscriptome

Like DNA and protein, RNA contains epigenetic modifications that govern its function and metabolism. RNA can be chemically modified with over 150 unique chemical forms in different organisms, several of which have begun to be characterized in eukaryotes (**Figure 1.1**). These modifications were discovered beginning in 1951, when ion-exchange analysis of RNA revealed an abundant unknown modification later identified as pseudouridine (Ψ)[5–8]. Discoveries of other abundant modifications using radioactive labeling followed: 2'-*O*-methylation (2'OMe) and N^1-methyladenosine (m^1A) were discovered in tRNA and rRNA, and 2'OMe, N^6-methyladenosine (m^6A), and 5-methylcytidine (m^5C) were found in mRNA and viral RNA[9–12]. As modifications were systematically characterized and catalogued, hints to their functions emerged. m^6A, the most

abundant internal modification of eukaryotic mRNA, was shown in early studies to facilitate processing of pre-mRNA and transport of mRNA[13,14].

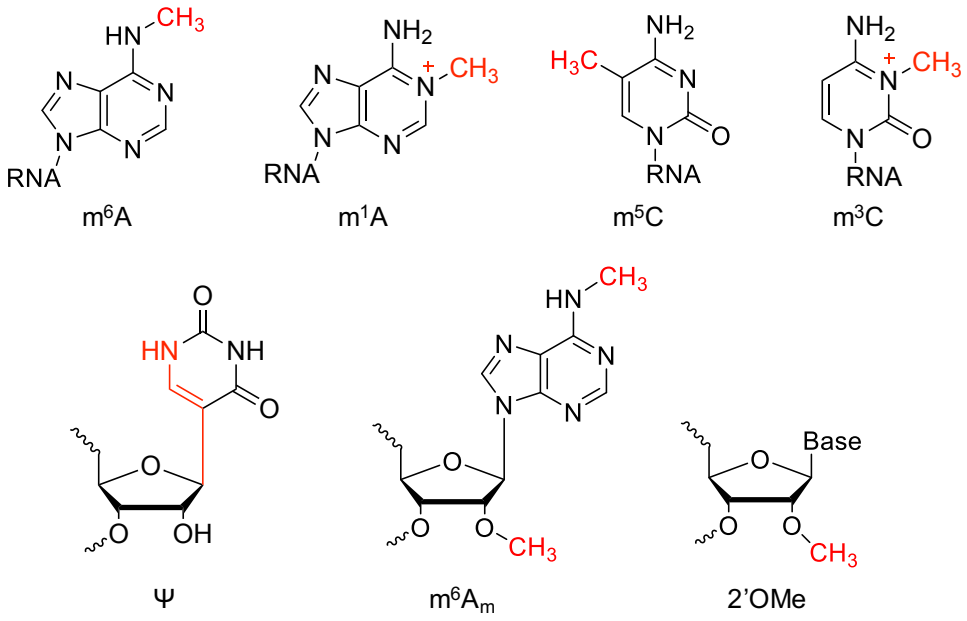

Figure 1.1 Chemical modifications of RNA in eukaryotes
Chemical structures of several modifications that have been characterized on eukaryotic RNA.

Prof. He proposed in 2010 that post-transcriptional RNA modifications could be reversible and may significantly impact gene expression regulation[15]. This hypothesis was confirmed with the discovery of fat-mass and obesity-associated protein (FTO), the first demethylase of m^6A on RNA, soon followed by alkB homologue 5 (ALKBH5), a second m^6A demethylase[16,17]. In 2012, m^6A-specific antibodies were used to profile m^6A sites through immunoprecipitation followed by high-throughput sequencing. Thousands of m^6A sites were identified in human and mouse cell lines, with enrichment around the stop codon and 3'UTR[18,19]. These advances sparked extensive research on RNA post-transcriptional modifications in this new era of epitranscriptomics.

1.3 m^6A writers and readers lead the way

m^6A is installed by a methyltransferase complex including the *S*-adenosyl methionine binding protein methyltransferase-like 3 (METTL3), first identified over two decades ago[20,21] (**Figure 1.2**). Recent experiments have identified that METTL3 and METTL14 are essential components of a writer complex, where METTL3 is catalytically active while METTL14 plays critical structural functions[22,23]. Functional roles of m^6A were discovered through experiments inactivating METTL3, showing that loss of m^6A compromises circadian rhythm, embryonic stem cell fate transition, and naïve pluripotency[24–26]. A new m^6A methyltransferase, METTL16, has been shown to regulate splicing of the human SAM synthetase MAT2A, promoting its expression through enhanced splicing of a retained intron in SAM-depleted conditions, thus acting as a regulation loop[27]. METTL16 was also shown to be the m^6A methyltransferase of the U6 small nuclear RNA.

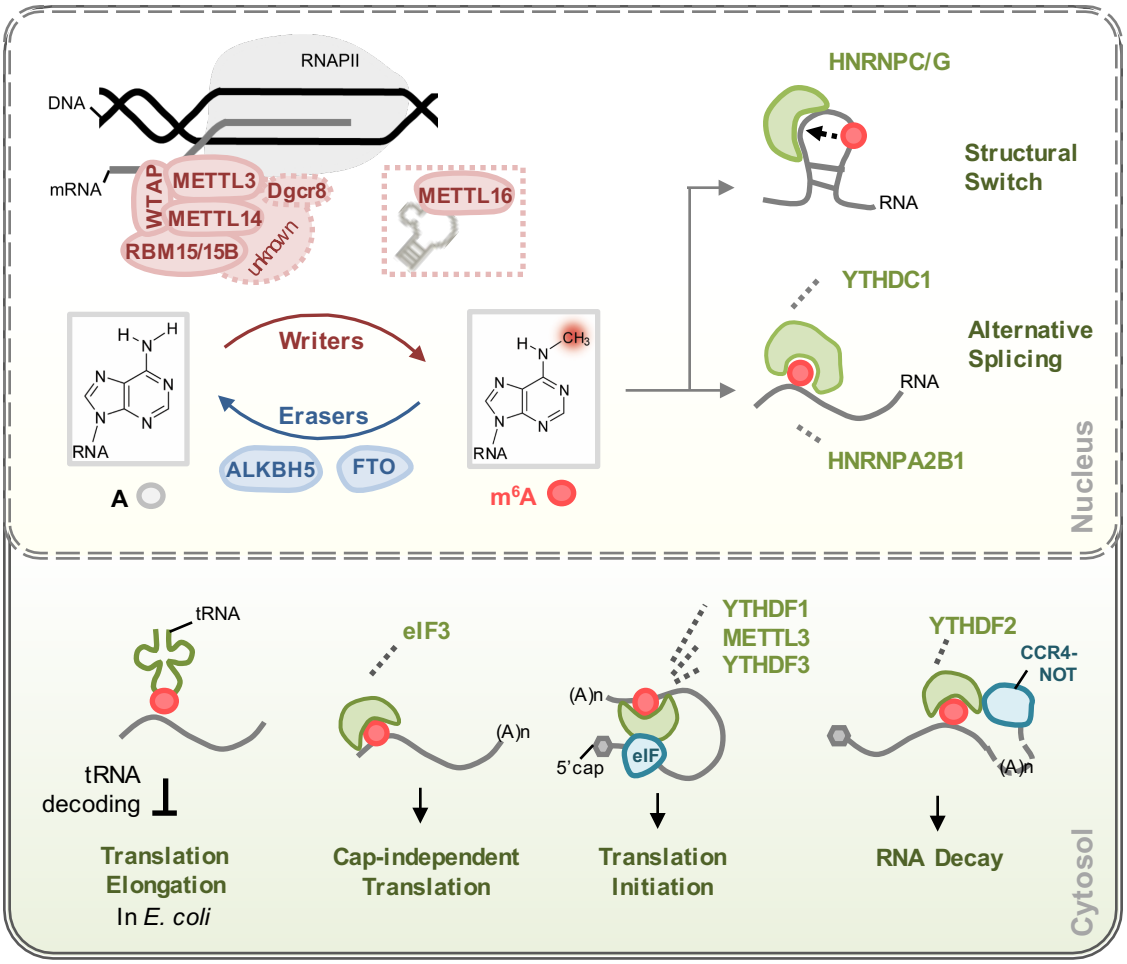

Figure 1.2 The m^{6}A machinery
A schematic representation of the writers, readers, erasers, and cellular components of eukaryotes that interact with m^{6}A and the RNA that contains it

Importantly, m^{6}A regulates gene expression through various m^{6}A-recognition proteins. YTH domain containing 1 (YTHDC1), an m^{6}A "reader," acts in the nucleus to influence mRNA splicing[28], while heterogeneous nuclear ribonucleoprotein C (HNRNPC) and HNRNPG bind to RNAs whose structures have been altered by m^{6}A to promote mRNA processing and alternative splicing[29,30]. In the cytosol, the m^{6}A readers YTH domain family 1 (YTHDF1) and YTHDF3 affect the translation of their targets through ribosome loading in HeLa cells[31–33], and YTHDF2 facilitates mRNA degradation by recruiting the CCR4-NOT deadenylase complex[34,35]. The m^{6}A reader

YTHDC2 also functions in the cytosol, and a part of the work herein describes its function. As research elucidates roles of m^6A readers, it is becoming evident that their roles may be complex. m^6A in the 5'UTR could facilitate cap-independent translation initiation through a process involving eIF3 [36,37]. The exact "reading" mechanism of this process is still unclear. Under heat shock, YTHDF2 shields 5'UTR m^6A from FTO, allowing selective mRNA translation. It will be important to determine functional roles of readers under different biological conditions.

1.4 Effects of m^6A at the molecular level

m^6A appears to influence almost every stage of mRNA metabolism. Three recent studies demonstrated interactions with translation, transcription, and microprocessor machinery (**Figure 1.2**). In an *Escherichia coli* translation system, m^6A on mRNA interferes with tRNA accommodation and translation elongation[38]. Although m^6A does not interfere with the structure of the codon-anticodon interaction, minor steric constraints destabilize base pairing. The magnitude of the resulting delay is affected by the position of m^6A, implicating that m^6A may be an important regulator of tRNA decoding. m^6A was also shown to be correlated with decreased translation efficiency in a study using MCF7 cells[39]. In this experiment, an inducible reporter system was used to demonstrate that transcripts with slower rates of transcription received greater deposition of m^6A, and that m^6A deposition occurs co-transcriptionally. METTL3 interacts with RNAPII under conditions of slower transcription, and methylated transcripts demonstrated decreased efficiency of translation. As m^6A has been shown to promote translation in other studies[31,36,37], the role of m^6A in affecting translation could be transcript- and position-dependent. While the m^6A itself could reduce translation efficiency as shown from the in vitro experiment[38], the YTH domain proteins could promote translation in response to stimuli or signaling. A recent

study showed that Mettl3 binds to RNA co-transcriptionally, and that this interaction is necessary for microprocessor components Dgcr8 and Drosha to physically associate with chromatin to mediate gene silencing[40]. Mettl3 and Dgcr8 relocalize to heat-shock genes under hyperthermia and work in concert to promote the degradation of their targets, allowing timely clearance of heat-shock responsive transcripts after heat-shock has ended. These studies reveal important roles of m^6A in enhancing the dynamic control of gene expression, a function especially important under changing cell conditions.

1.5 Influences of m^6A on development and differentiation

We recently proposed that m^6A shapes the transcriptome to facilitate cell differentiation[41]. Such a role could be critical during development, as is suggested by several recent studies. m^6A is necessary for sex determination in *Drosophila*[42,43]. Depletion of the *Drosophila* METTL3 homologue Ime4 leads to the absence of m^6A on the sex determination factor *Sex lethal* (*Sxl*). Without m^6A, the YTHDC1 homologue YT521-B is unable to properly splice *Sxl*, leading to failure of X inactivation and thus improper sex determination. Moreover, depletion of Ime4 affects neuronal function, causing shortened lifespan and irregularities in flight, locomotion, and grooming. m^6A has also been shown to regulate clearance of maternal mRNA in zebrafish in the maternal-to-zygotic transition[44]. Zebrafish embryos lacking the m^6A reader Ythdf2 become developmentally delayed due to impaired decay of m^6A-modified maternal RNAs. Because maternal RNAs are not properly decayed, activation of the zygotic genome is also impaired.

Previous studies have demonstrated roles of m^6A in the differentiation of mouse and human embryonic stem cells[25,26,45]. More recently, effects of m^6A on differentiation have been shown in mice. Two studies separately showed that the meiosis-specific protein MEIOC, which is necessary

for proper meiotic prophase I during spermatogenesis, interacts with the m^{6}A reader YTHDC2[46,47]. Mice lacking *Meioc* are infertile, lacking germ cells that have reached the pachytene phase of meiotic prophase I. Notably, mice lacking *Mettl3* display similar phenotypes, demonstrating infertility and defects in germ cells[48]. m^{6}A also affects somatic cell differentiation in mice. Knockout of *Mettl3* in mouse T cells caused failure of naïve T cells to proliferate and differentiate; in a lymphopaenic adoptive transfer model, most naïve *Mettl3*-deficient T cells remained naïve, and no signs of colitis were present[49]. The lack of *Mettl3* caused upregulation of SOCS family proteins, which inhibited the IL-7-mediated STAT5 activation necessary for T cell expansion. Two studies of FTO also demonstrated roles of m^{6}A in somatic cell differentiation. FTO expression was shown to increase during myoblast differentiation, and its depletion inhibited differentiation in both mouse primary myoblasts and mouse skeletal muscle[50]. The demethylase activity of FTO is required, as a point mutation of FTO to remove demethylase activity impairs myoblast differentiation. FTO is also dynamically expressed during postnatal neurodevelopment, and its loss impedes the proliferation and differentiation of adult neural stem cells[51].

1.6 Involvement of m^{6}A writers in human cancer

As shown by the previous section, m^{6}A is a critical factor for cell differentiation. Considering that cancer is driven by the misregulation of cell growth and differentiation, cancer cells may hijack aberrant methylation to enhance their survival and progression. Several studies have demonstrated roles of demethylation or lack of methylation in promoting cancer progression. In *MLL*-rearranged acute myeloid leukemia (AML), FTO is highly expressed, promotes oncogene-mediated cell transformation and leukemogenesis, and inhibits all-*trans*-retinoic acid (ATRA)-induced AML cell differentiation[52]. At the molecular level in AML, FTO causes a decrease in m^{6}A

methylation as well as a decrease in the transcript expression of these hypo-methylated genes. *ASB2* and *RARA* are functionally important targets of FTO in *MLL*-rearranged AML; their forced expression rescues ATRA-induced differentiation. The oncogenic role of FTO is not limited to AML; another study showed that inhibition of FTO in glioblastoma stem cells (GSC) suppresses cell growth, self-renewal, and tumorigenesis[53]. This study demonstrated that other components of m^6A machinery impact glioblastoma as well. Knockdown of METTL3 or METTL14 affects the mRNA expression of genes critical to GSC function, and enhances GSC growth, proliferation, and tumorigenesis. In agreement with these findings that lack of methylation tends to promote cancer progression, Zhang et al. showed that ALKBH5 is highly expressed in GSCs, and its knockdown suppresses their proliferation[54]. The protein abundance of the ALKBH5 target *FOXM1* is greatly increased in GSCs due to the demethylation activity of ALKBH5; removal of m^6A at the 3' end of *FOXM1* pre-mRNA promotes *FOXM1* interaction with HuR, which enhances FOXM1 protein expression. A lncRNA antisense to *FOXM1* facilitates the interaction between ALKBH5 and *FOXM1*, and depletion of either ALKBH5 or its antisense lncRNA inhibits GSC tumorigenesis. ALKBH5 also promotes a breast cancer phenotype; under hypoxic conditions, ALKBH5 expression increases, thus decreasing levels of m^6A and upregulating expression of the pluripotency factor NANOG[55].

Together, the aforementioned studies suggest that a decrease in RNA m^6A methylation tends to facilitate cancer progression, and that RNA methylation could notably affect cell growth and proliferation. However, other studies indicate that the role of m^6A may be more complex in different cancers. In hepatocellular carcinoma (HCC), although METTL14 down-regulation is associated with tumor metastasis, METTL3 enhances the invasive ability of HCC cells[56]. Several other studies also point to an oncogenic role of the methyltransferase complex. METTL3 plays an

oncogenic role in cancer cells, promoting the translation of cancer genes through interactions with translation initiation machinery[57]. Interestingly, METTL3 promotes translation independent of its methyltransferase activity or any interaction with the m^{6}A reader YTHDF1. WTAP, a component of the m^{6}A methyltransferase complex, also promotes leukemogenesis, and its levels are increased in primary AML samples[58]. RBM15, another methyltransferase complex component, is altered in acute megakaryoblastic leukemia, undergoing translocation to fuse with *MKL*[59].

Considering the complex findings, it is likely that different types of cancers can be derived from unique imbalances or misregulation of mRNA methylation. In AML, increased WTAP and RBM15 expression (or writer proteins themselves) could block differentiation, leading to leukemia, while increased eraser expression could cause leukemia through separate pathways. The intricate network of interactions is reminiscent of studies of DNA methylation; just as misregulation of DNMT and TET proteins are both associated with cancer[60–63], misregulation of the m^{6}A machinery can lead to cancer through unique mechanisms. Interestingly, the oncometabolite D-2-hydroxyglutarate (D2-HG), which could act as a nonspecific inhibitor of the iron- and αKG-dependent dioxygenases FTO and ALKBH5, accumulates in ~20% of AMLs[64], and may thus contribute to outcome of these cancers through inhibiting RNA demethylation. Further investigation is necessary to uncover mechanisms by which aberrant methylation affects proliferation of various cancers.

1.7 An extensive family of RNA modifications

Recent work in epitranscriptomics has naturally focused on roles of m^{6}A due to its high abundance and the relatively early discovery of the m^{6}A machinery, which facilitated functional and mechanistic studies. It is important, however, to keep in mind that m^{6}A is one of over 150

known RNA modifications[65], many of which have received also significant attention of their own. Recent advances in high-throughput sequencing and mass spectrometry have revitalized research on the broad spectrum of post-transcriptional modifications, elucidating functions of known modifications as well as newly discovered modifications on mRNA (**Figure 1.1**).

Methylation of the N^1 position of adenosine (m^1A) was recently discovered on mRNA, occurring at levels ~10-30% of that of m^6A on mRNA depending on cell line or tissue[66,67]. m^1A occurs in more structured regions and is enriched near translation initiation sites. The level of m^1A responds dynamically to nutrient starvation and heat shock, and the 5' UTR peaks correlate with translation upregulation. As it is positively charged, it may markedly alter RNA structure as well as interactions with proteins or other RNAs. Zhou et al. demonstrated that m^1A causes A-U Hoogsteen base pairs in RNA to be strongly disfavored, and that RNA containing m^1A tends to adopt an unpaired anti conformation[68]. m^1A was also shown to affect translation; its presence at the first or second codon position, but not the third codon, blocks translation in both *E. coli* and wheat germ extract systems[69]. Additionally, m^1A is present in early coding regions of transcripts without 5'UTR introns, which are associated with low translation efficiency and facilitate noncanonical binding by the exon junction complex[70]. These studies point to a main role of m^1A in translation and RNA-RNA interactions. The exact functional roles of 5' UTR m^1A sites require future studies; however, there are also other m^1A sites in mRNA which could play distinct roles. Methods to map low abundance m^1A sites in mRNA will be critical to understanding its biological roles[71]. Interestingly, YTHDF1-3 and YTHDC1 were shown to selectively bind m^1A[72]; however, this finding must be taken cautiously, as m^1A undergoes conversion to m^6A by the by the Dimroth rearrangement. Identification of readers of m^1A will also prove critical for future studies.

3-methylcytosine (m^3C) was recently identified as a modification in mRNA, present at a rate of around 0.004% of cytosines in human cell culture[73]. It is installed by METTL8, and its function and localization have yet to be identified.

Pseudouridine (Ψ), which is generated by isomerization of uridine, is the most abundant RNA modification in total RNA[7]. It was recently identified on mRNA and mapped by several groups using similar techniques (PseudoU-seq, Ψ-seq, PSI-seq, and CeU-seq), which use the water-soluble diimide CMCT to generate strong RT stops at ψ sites[74–77]. PseudoU-seq and Ψ-seq identified >200 and >300 sites, respectively, on human and yeast mRNAs, and Ψ/U in mRNA has been quantified to be around 0.2-0.7% in mammalian cell lines. Direct evidence of biological functions of Ψ on mRNA has yet to be identified. However, several findings point to potential biological roles. Ψ affects the secondary structure of RNA and alters stop codon readthrough[78,79]. Depletion of the pseudouridine synthase PUS7 decreases the abundance of mRNAs containing Ψ, suggesting that Ψ may also affect transcript stability[75]. Moreover, pseudouridinylation on transcripts is affected by stresses such as heat shock and nutrient deprivation, suggesting that Ψ may be a response to various stresses[74,75,77].

1.8 Modifications on transfer RNAs and other RNAs

tRNAs contain more modifications than any other RNA species, with each tRNA containing on average 14 modifications[65]. Recent studies have identified tRNA demethylases and methyltransferases, as well as the functions of their modifications.

Liu et al. recently identified the first tRNA demethylase; ALKBH1 demethylates m^1A58 in $tRNA^{iMet}$ and several other tRNA species[80]. m^1A58 increases $tRNA^{iMet}$ stability, and its demethylation by ALKBH1 decreases the rate of protein synthesis. A related demethylase,

ALKBH3, removes m^6A from tRNA and increases translation efficiency *in vitro*, though its cellular targets and functions have yet to be identified[81].

m^5C on tRNA can influence translation as well, particularly affecting stress responses. Deletion of the tRNA m^5C methyltransferase NSUN2 reduces tRNA m^5C levels and promotes cleavage of unmethylated tRNAs into fragments, which decrease protein translation rates and induce stress response pathways[82]. Lack of *Nsun2* in mice leads to an increase in undifferentiated tumor stem cells due to decreased global translation, which increases the self-renewal potential of the tumor-initiating cells[83]. Interestingly, lack of *Nsun2* also prevents cells from activating survival pathways when treated with cytotoxic agents, suggesting that the combination of m^5C inhibitors and chemotherapeutic agents may effectively treat certain cancers.

m^5C also plays an important role in the translation of the mitochondrial transfer RNA for methionine (mt-$tRNA^{Met}$). m^5C is deposited onto cytosine 34 of mt-$tRNA^{Met}$ by the methyltransferase NSUN3[84–86]. Lack of NSUN3 leads to deficiencies such as reduced mitochondrial protein synthesis, reduced oxygen consumption, and defects in energy metabolism. Mutation of NSUN3 is also associated with several diseases, including maternally inherited hypertension and combined mitochondrial respiratory chain complex deficiency. Mechanistically, m^5C is oxidized by ALKBH1/ABH1 into 5-formylcytidine, which is necessary for reading the AUA codon during protein synthesis.

Methylation and editing of tRNA may require intricate mechanisms and conditions. NSun6, which installs m^5C72 onto tRNA, recognizes both the sequence and shape of tRNA[87]. Without a folded, full-length tRNA, NSun6 does not methylate m^5C72. C-to-U deamination of C32 in *Trypanosoma brucei* $tRNA^{Thr}$ also depends on multiple factors[88]. Methylation of C32 to m^3C by two enzymes, the m^3C methyltransferase TRM140 and the deaminase ADAT2/3, is a required step

in the deamination process. m^3C must then be deaminated to 3-methyluridine (m^3U) by the same mechanism, which is then demethylated to become U.

The recent discoveries of the first tRNA demethylases, effects on translation and differentiation, and complex mechanisms of tRNA methylation and editing will undoubtedly inspire investigations to elucidate functions of tRNA modifications and the biological processes to which they respond.

Ribosomal RNA (rRNA) is also marked by abundant modifications; the >200 modified sites in human rRNAs make up around 2% of rRNA nucleotides. Most modifications on rRNA are Ψ or 2'OMe, although rRNA also contains around 10 base modifications[65]. Functions of rRNA modifications are largely unknown, but studies of 2'OMe on rRNA are beginning to provide hints to their functions. The C/D box snoRNAs SNORD14D and SNORD35A, which are necessary to install 2'OMe onto rRNA, are necessary for proper leukemogenesis, and are upregulated by leukemia oncogenes[89]. C/D box snoRNA expression in leukemic cells correlated with protein synthesis and cell size, suggesting a potential role of 2'OMe on rRNA in translation.

The processing and functions of other non-coding RNA species have recently been shown to undergo regulation by m^6A. Alarcón et al. demonstrated that pri-microRNAs contain m^6A, which is installed by METTL3 and promotes recognition and processing into mature microRNA by DGCR8[90]. m^6A is also present on the lncRNA *XIST*, and is necessary for *XIST* to mediate transcriptional silencing on the X chromosome during female mammalian development[91]. Finally, m^6A is present on human box C/D snoRNA species; it impedes the formation of *trans* Hoogsteen-sugar A-G base pairs, thus affecting snoRNA structure, and also blocks binding by human 15.5-kDa protein[92].

1.9 Scope of this thesis

In this thesis, I aim to elucidate roles of m^6A and its reader proteins in regulating gene expression and affecting biological function.

Chapter 2 presents characterization of the m^6A binding specificity and biological function of the probable ATP-dependent RNA helicase YTHDC2.

Chapter 3 presents investigation of the function of FMRP as a facilitator of nuclear export of m^6A-containing mRNAs.

Chapter 4 summarizes recent findings in epitranscriptomics, discussing broader impacts and looking ahead toward unexplored frontiers in the field.

References

1. Waddington, C. H. The epigenotype. 1942. *Endeavor* **1,** 10–13 (1942).
2. Feinberg, A. P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447,** 433–440 (2007).
3. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9,** 465–76 (2008).
4. Shi, Y. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nat. Rev. Genet.* **8,** 829–33 (2007).
5. Volkin, E. & Cohn, W. E. Nucleoside-5′-Phosphates from Ribonucleic Acid. *Nature* **167,** 483–484 (1951).
6. Cohn, W. E. 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. *Biochim. Biophys. Acta* **32,** 569–571 (1959).
7. Cohn, W. E. Pseudouridine, a Carbon-Carbon Linked Ribonucleoside in Ribonucleic Acids:

Isolation, Structure, and Chemical Characteristics. *J. Biol. Chem.* **235,** 1488–1498 (1960).

8. Davis, F. F. & Allen, F. W. Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* **227,** 907–915 (1957).

9. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **71,** 3971–5 (1974).

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

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

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

13. Camper, S. a, Albers, R. J., Coward, J. K. & Rottman, F. M. Effect of undermethylation on mRNA cytoplasmic appearance and half-life. *Mol. Cell. Biol.* **4,** 538–543 (1984).

14. Finkel, D. & Groner, Y. Methylations of adenosine residues (m6A) in pre-mRNA are important for formation of late simian virus 40 mRNAs. *Virology* **131,** 409–425 (1983).

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

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

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

18. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed

by m6A-seq. *Nature* **485,** 201–206 (2012).

19. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149,** 1635–1646 (2012).

20. Bokar, J. A., Rath-Shambaugh, M. E., Ludwiczak, R., Narayan, P. & Rottman, F. Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei: Internal mRNA methylation requires a multisubunit complex. *J. Biol. Chem.* **269,** 17697–17704 (1994).

21. 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,** 1233–47 (1997).

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

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

24. Fustin, J. M. *et al.* RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell* **155,** 793–806 (2013).

25. Batista, P. J. *et al.* m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells. *Cell Stem Cell* **15,** 707–19 (2014).

26. Geula, S. *et al.* m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science (80-.).* **347,** 1002–1006 (2015).

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

28. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell*

61, 507–519 (2016).

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

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

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

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

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

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

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

36. Meyer, K. D. *et al.* 5' UTR m6A Promotes Cap-Independent Translation. *Cell* 1–12 (2015). doi:10.1038/cr.2014.162

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

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

39. Slobodin, B. *et al.* Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Article Transcription Impacts the Efficiency of mRNA

Translation via Co-transcriptional. *Cell* **169,** 326–337.e12 (2017).

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

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

42. Haussmann, I. U. *et al.* m6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* **540,** 301–304 (2016).

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

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

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

46. Abby, E. *et al.* Implementation of meiosis prophase I programme requires a conserved retinoid-independent stabilizer of meiotic transcripts. *Nat. Commun.* **7,** 10324 (2016).

47. Soh, Y. Q. S. *et al.* Meioc maintains an extended meiotic prophase I in mice. *PLoS Genet.* **13,** e1006704 (2017).

48. Lin, Z. *et al.* Mettl3-/Mettl14-mediated mRNA N 6 -methyladenosine mod- ulates murine spermatogenesis. *Cell Res.* **27,** 1–15 (2017).

49. 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).

50. 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).

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

52. 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).

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

54. Zhang, S. *et al.* m6A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program. *Cancer Cell* **31,** 591–606 (2017).

55. Zhang, C. *et al.* Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m6A-demethylation of NANOG mRNA. *Proc. Natl. Acad. Sci.* E2407–E2056 (2016). doi:10.1073/pnas.1602883113

56. Ma, J. *et al.* METTL14 suppresses the metastatic potential of HCC by modulating m ^{6} A-dependent primary miRNA processing. *Hepatology* 1–39 (2016). doi:10.1002/hep.28885

57. Lin, S., Choe, J., Du, P., Triboulet, R. & Gregory, R. I. The m6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol. Cell* **62,** 335–345 (2016).

58. Bansal, H. *et al.* WTAP is a novel oncogenic protein in acute myeloid leukemia. *Leukemia* **28,** 1171–4 (2014).

59. Ma, Z. *et al.* Fusion of two novel genes, RBM15 and MKL1, in the t(1;22)(p13;q13) of acute megakaryoblastic leukemia. *Nat. Genet.* **28,** 220–1 (2001).

60. Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363,** 2424–2433 (2010).

61. Zhang, W. & Xu, J. DNA methyltransferases and their roles in tumorigenesis. *Biomark. Res.*

5, 1 (2017).

62. Kanai, Y., Ushijima, S., Nakanishi, Y., Sakamoto, M. & Hirohashi, S. Mutation of the DNA methyltransferase (DNMT) 1 gene in human colorectal cancers. *Cancer Lett.* **192,** 75–82 (2003).

63. Jones, P. A., Issa, J.-P. J. & Baylin, S. Targeting the cancer epigenome for therapy. *Nat. Rev. Genet.* **17,** 630–41 (2016).

64. Elkashef, S. M. *et al.* IDH Mutation, Competitive Inhibition of FTO, and RNA Methylation. *Cancer Cell* **31,** 619–620 (2017).

65. Machnicka, M. A. *et al.* MODOMICS: A database of RNA modification pathways - 2013 update. *Nucleic Acids Res.* **41,** 262–267 (2013).

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

67. Li, X. *et al.* Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat. Chem. Biol.* **12,** 311–6 (2016).

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

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

70. Cenik, C. *et al.* A Common Class of Transcripts with 5'-Intron Depletion, Distinct Early Coding Sequence Features, and N1-Methyladenosine Modification. *RNA* **23,** 270–283 (2017).

71. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and

validate. *Nat. Rev. Genet.* **18,** 275–291 (2017).

72. 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).

73. Xu, L. *et al.* Three distinct 3-methylcytidine (m3C) methyltransferases modify tRNA and mRNA in mice and humans. *J. Biol. Chem.* jbc.M117.798298 (2017). doi:10.1074/jbc.M117.798298

74. Carlile, T. M. *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515,** 143–6 (2014).

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

76. Lovejoy, A. F., Riordan, D. P. & Brown, P. O. Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine Synthases Modify Specific mRNAs in S. cerevisiae. *PLoS One* **9,** e110799 (2014).

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

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

79. Karijolich, J., Kantartzis, A. & Yu, Y.-T. in *Methods Mol Biol* **629,** 141–158 (2010).

80. Liu, F. *et al.* ALKBH1-Mediated tRNA Demethylation Regulates Translation. *Cell* **167,** 816–828 e16 (2016).

81. Ueda, Y. *et al.* AlkB homolog 3-mediated tRNA demethylation promotes protein synthesis in cancer cells. *Sci. Rep.* **7,** 42271 (2017).

82. Blanco, S. *et al.* Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. *EMBO J.* **33,** 2020–2039 (2014).

83. Blanco, S. *et al.* Stem cell function and stress response are controlled by protein synthesis. *Nature* **534,** 335–340 (2016).

84. Van Haute, L. *et al.* Deficient methylation and formylation of mt-tRNAMet wobble cytosine in a patient carrying mutations in NSUN3. *Nat. Commun.* **7,** 12039 (2016).

85. Nakano, S. *et al.* NSUN3 methylase initiates 5-formylcytidine biogenesis in human mitochondrial tRNAMet. *Nat. Chem. Biol.* **12,** 546–551 (2016).

86. Haag, S. *et al.* NSUN3 and ABH1 modify the wobble position of mt-tRNAMet to expand codon recognition in mitochondrial translation. *Embo J.* **35,** DOI: 10.15252/embj.201694885 (2016).

87. Long, T. *et al.* Sequence-specific and shape-selective RNA recognition by the human RNA 5-methylcytosine methyltransferase NSun. *J. Biol. Chem.* **291,** 24293–24303 (2016).

88. Rubio, M. A. T. *et al.* Editing and methylation at a single site by functionally interdependent activities. *Nature* **542,** 494–497 (2017).

89. Zhou, F. *et al.* AML1-ETO requires enhanced C/D box snoRNA/RNP formation to induce self-renewal and leukaemia. *Nat. Cell Biol.* **19,** 844–855 (2017).

90. Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S. F. N6-methyladenosine marks primary microRNAs for processing. *Nature* **519,** 482–5 (2015).

91. Patil, D. P. *et al.* m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537,** 369–373 (2016).

92. Huang, L., Ashraf, S., Wang, J. & Lilley, D. M. Control of box C/D snoRNP assembly by N6-methylation of adenine. *EMBO Rep.* e201743967 (2017).

doi:10.15252/embr.201743967

Chapter 2 – Characterization of YTHDC2

2.1 One remaining uncharacterized YTH protein

m^6A exerts many of its functions through "reader" proteins in the cytoplasm and nucleus that selectively bind to m^6A-containing transcripts [1–9]. In particular, the YT521-B homology (YTH) family of proteins possess the evolutionarily conserved YTH domain that selectively recognizes m^6A [5]. In the cytosol, YTH domain family 1 (YTHDF1) and YTHDF3 act in concert to affect the translation of their targets by facilitating ribosome loading in HeLa cells [4,10,11], while YTHDF2 decreases mRNA stability by recruiting the CCR4-NOT deadenylase complex [2,12]. In the nucleus, YTH domain containing 1 (YTHDC1) acts to influence mRNA splicing [6]. Working in concert, YTHDF1, YTHDF2, YTHDF3, and YTHDC1 may provide a fast track to quickly process mRNAs in the nucleus, facilitate their translation, and then degrade the translated mRNAs.

The fifth member of the YTH protein family, YTHDC2, is annotated as a probable ATP-dependent RNA helicase. To gain a more complete understanding of the YTH protein family, we set out to characterize the m^6A binding specificity and function of YTHDC2.

2.2 Results

Although it contains the highly evolutionarily conserved YTH domain found in the other YTH family proteins, YTHDC2 possesses multiple helicase domains that may also act on RNA, prompting the question of whether it binds m^6A (**Figure 2.1**). Ythdc2 is highly expressed in mouse testes, and was found to interact with the meiosis-specific protein MEIOC, which is required for meiotic prophase I[13–15], pointing to a role in spermatogenesis.

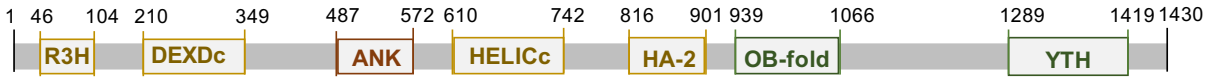

Figure 2.1 YTHDC2 contains multiple helicases domains
Domains of YTHDC2 (https://www.ncbi.nlm.nih.gov/gene/64848).

2.2.1 YTHDC2 preferentially binds m^{6}A

To determine whether YTHDC2 binds m^{6}A, we performed an *in vitro* pulldown assay using lysate from HeLa cells stably overexpressing FLAG-tagged YTHDC2. YTHDC2 bound preferentially to an RNA probe containing m^{6}A compared to a control RNA probe containing A (~3.2-fold difference) (**Figure 2.2A**). Moreover, YTHDC2 bound preferentially to an RNA probe containing the m^{6}A consensus motif (GGACU) compared to a control RNA probe containing a random sequence (ACAGA). YTHDC2 preferred binding to m^{6}A rather than A regardless of the sequence of the probe (**Figure 2.2A**). We further confirmed that YTHDC2 binds selectively to m^{6}A-containing RNA using gel shift (**Figure 2.2B**). As YTHDC2 may play a role in meiosis, we repeated our experiments using mouse testes lysate, and saw that Ythdc2 in mouse testes lysate also bound preferentially to m^{6}A within the GGACU motif (**Figure 2.2C**).

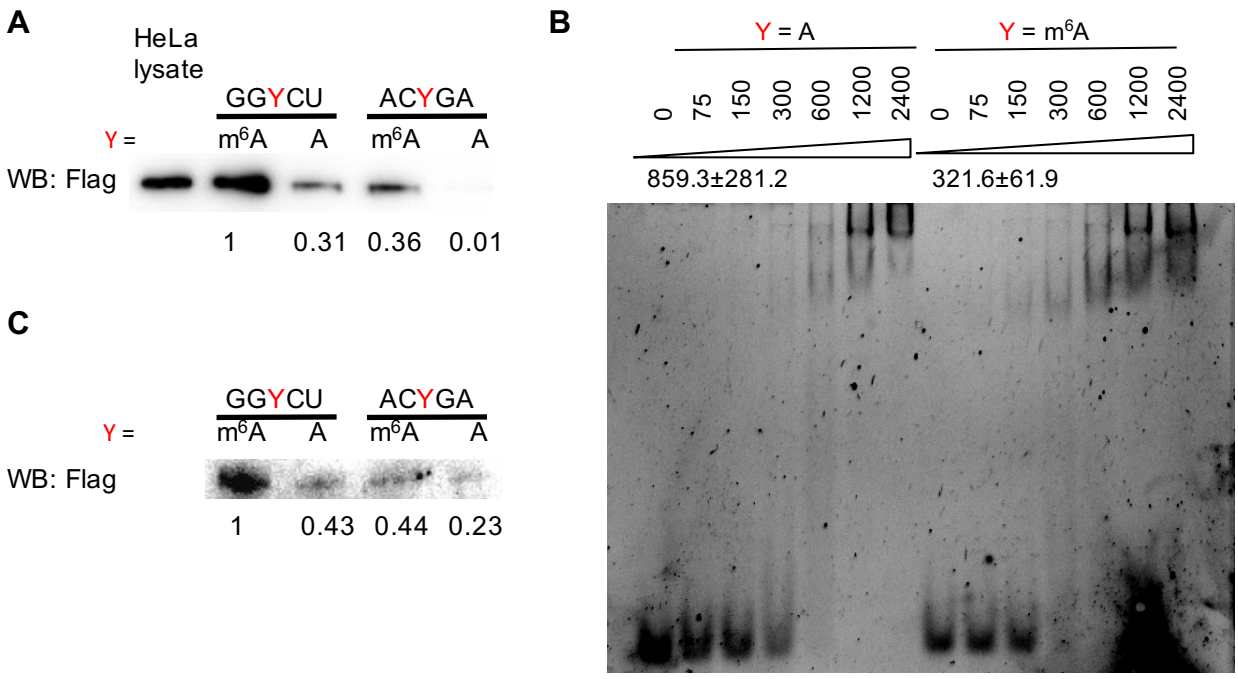

Figure 2.2 YTHDC2 binds preferentially to m^{6}A-marked RNA transcripts *in vitro*
In vitro probe pulldown assay in (A) HeLa cells overexpressing FLAG-YTHDC2 or (C) mouse testes showing YTHDC2 binds preferentially to probe with m^{6}A on GGACU. GG**Y**CU Probe sequence: 5'- CGUGG**Y**CUGGCU-**B**-3' (**Y**=m^{6}A or A, **B**=biotin) AC**Y**GA Probe sequence: 5'-GAUACYGAGAAG-B-3' (**Y**=m^{6}A or A, **B**=biotin). Labels: relative protein expression. (B) Gel shift assay measuring the dissociation (K_d, nM, indicated at the upper left corner of the gel) of FLAG-tagged YTHDC2 with methylated and unmethylated RNA probes. Protein concentrations are listed in nM. 4 nmol RNA probe was used.

To further confirm the consensus motif, we performed crosslink immunoprecipitation sequencing (CLIP-seq) in adult mouse testes, and observed that Ythdc2-bound fractions enrich a consensus motif of GGACU (**Figure 2.3A**). Moreover, the binding sites of YTHDC2 cluster around the stop codon, mirroring the locations of m^{6}A along the mRNA transcript (**Figure 2.3B**). These results indicate that YTHDC2 can preferentially recognize m^{6}A-containing transcripts.

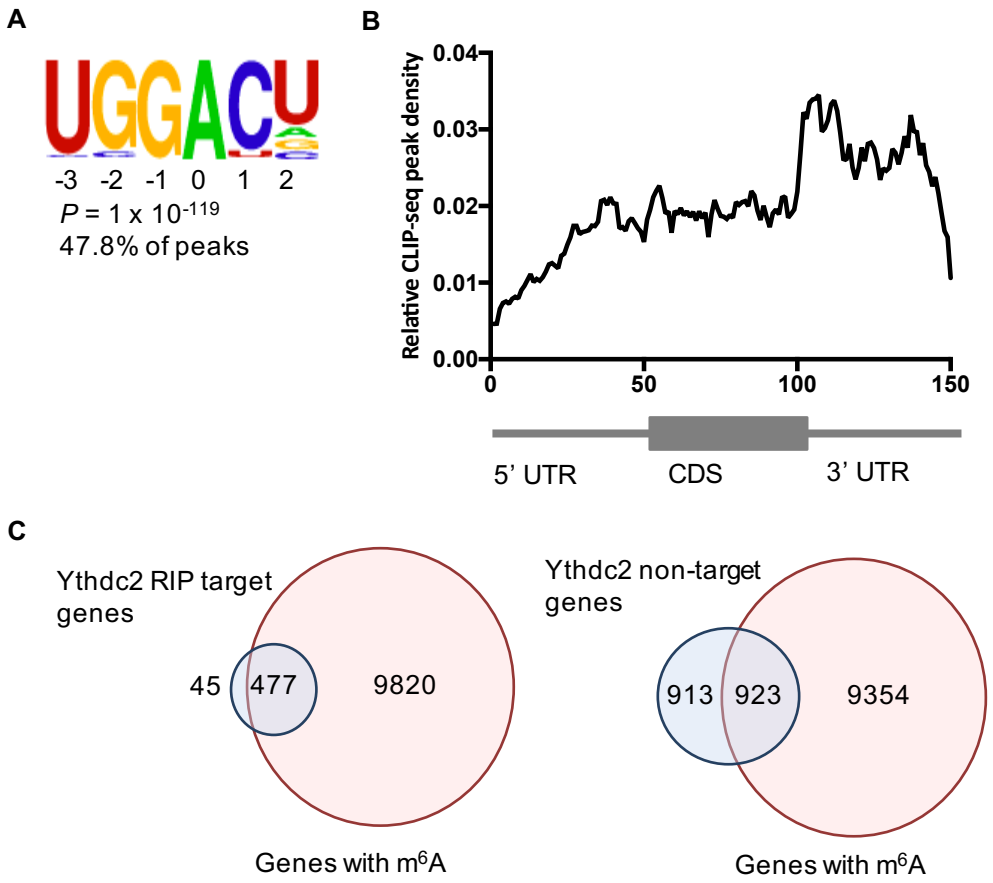

Figure 2.3 YTHDC2 binds preferentially to m^{6}A-marked RNA transcripts *in vivo*
(A) Consensus motif of Ythdc2 binding identified by HOMER of CLIP-seq of Ythdc2 in mouse testes. (B) Distribution of peak densities in CLIP-seq of Ythdc2 in mouse testes. (C) Overlap of Ythdc2 RIP-seq genes and genes containing m^{6}A in young mouse testes. Targets of YTHDC2 are defined as genes enriched in the RIP: Log2(RIP/input) >= 1. Non-targets are defined as genes depleted in the RIP: Log2(RIP/input) <= -1. Genes with m^{6}A are defined as genes enriched in the m^{6}A IP: Log2(m^{6}A IP/input) >= 2.

Because previous work indicated that Ythdc2 may play a role in early spermatogenesis, we performed RNA immunoprecipitation sequencing (RIP-seq) of Ythdc2 in testes of mice at 16.5 days post-partum (d.p.p.), which are enriched in meiotic spermatocytes, to identify gene targets of Ythdc2[14]. Two biological replicates of RIP-seq assays identified 522 target genes of Ythdc2. Of these target genes, 477 (91.4%) contain m^{6}A, compared to 50.3% of genes depleted in the IP

fraction ("Ythdc2 non-targets") (**Figure 2.3C**). These results strengthen our conclusion that Ythdc2 is an m^6A "reader" protein.

2.2.2 Ythdc2 affects mouse spermatogenesis

GO term	*P* (FDR)	Genes
meiotic nuclear division	0.036	*Mei1, Tubgcp6, Mki67, Smc1a, Smc3, Smc1b*
lateral element	0.011	*Sycp2, Sycp1, Brca1, Smc3, Smc1b*

Figure 2.4 YTHDC2 targets genes associated with spermatogenesis
GO analysis of Ythdc2 RIP-seq genes; meiosis-related GO terms.

Gene ontology (GO) analysis of Ythdc2 RIP-seq targets in mouse testes identified meiosis-related categories among the most enriched pathways (**Figure 2.4**). In agreement with a previous report, we also identified that Ythdc2 is highly expressed in mouse testes (**Figure 2.5A**). Our results pointed us toward examining the function of Ythdc2 in spermatogenesis.

To investigate biological functions of Ythdc2, we generated two *Ythdc2*$^{-/-}$ founder mice by microinjection of CRISPR/Cas9 targeting exon 5 of the *Ythdc2* gene (**Figure 2.5B**). *Ythdc2*$^{-/-}$ mice with 37-bp deletion were viable and reached adulthood. However, both male and female mice were infertile; male mice had smaller testes, and female mice had smaller ovaries with progressive loss of germ cells (**Figure 2.5C-D, Figure 2.6A-C**). Comparative histological staining of the testes, caput epididymis, and cauda epididymis of adult mice showed that *Ythdc2*$^{-/-}$ mice lacked spermatozoon in the seminiferous tubules and epididymis (**Figure 2.5E**). *Ythdc2*$^{-/-}$ mice with 4-bp deletion showed the same phenotype as those with the 37-bp deletion (data not shown), eliminating the possibility of off-target effects induced by CRISPR/Cas9.

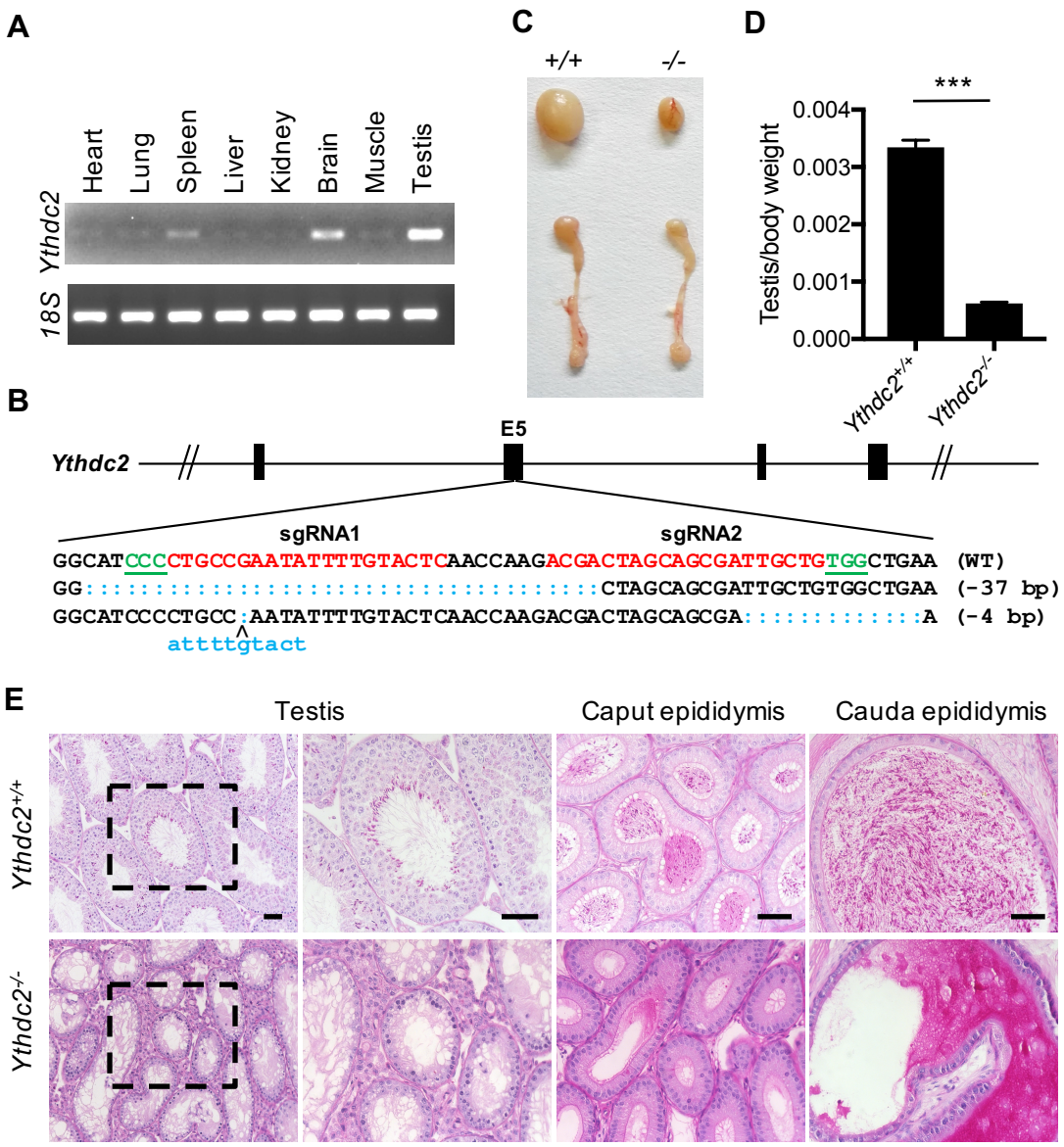

Figure 2.5 *Ythdc2*-deficient mice demonstrate defects in testes

(A) RT-PCR analysis of *Ythdc2* mRNA levels in various organs of adult mice. (B) Schematic diagram of sgRNAs targeting at *Ythdc*2 locus. Two founders were generated by micro-injection of sgRNA/Cas9 mRNA into one-cell embryos. PAM sequences are underlined and highlighted in green. sgRNA targeting sites are in red. Mutant sequence are in blue. (C) Testes and epididymis from adult *Ythdc2* $^{-/-}$ mice are smaller in size than wild type. (D) The testis/body weight ratio was calculated for adult *Ythdc2* $^{+/+}$ and *Ythdc2* $^{-/-}$ mice. Error bars, mean±sd, n=4, biological replicates ***P<0.001 (Student's t-test) (E) Periodic Acid Schiff (PAS) staining of adult testes and epididymis. Scale bar, 50 μm.

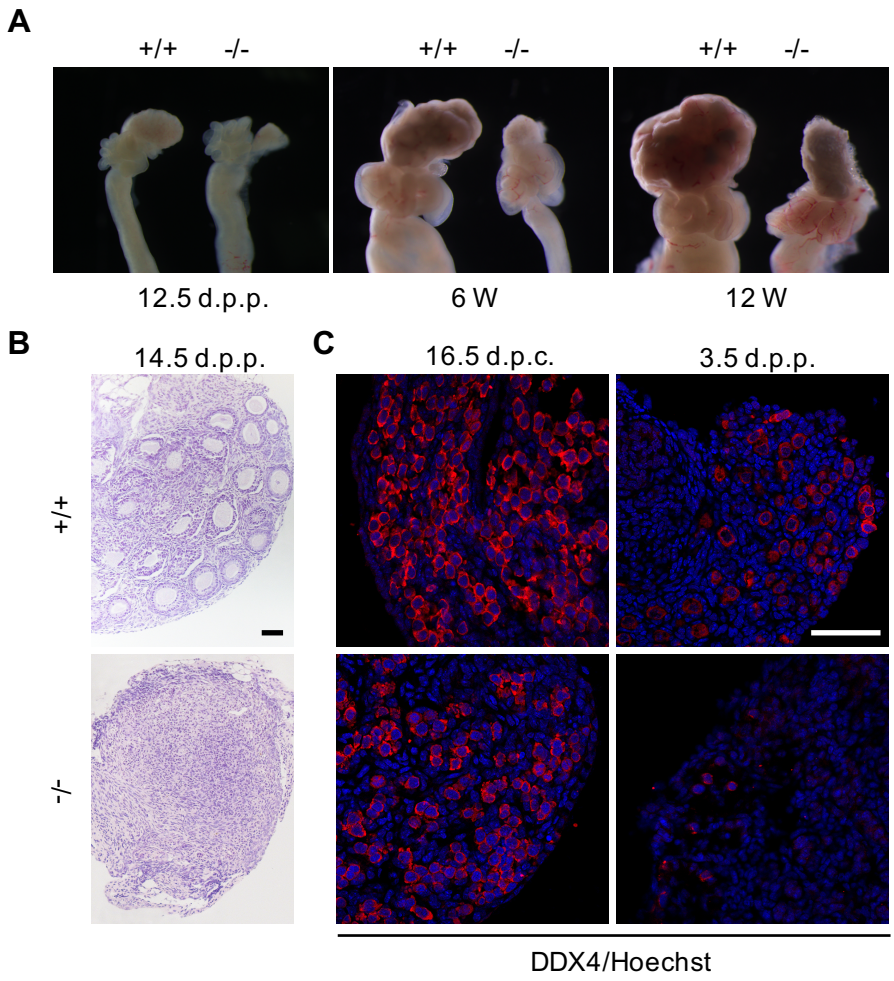

Figure 2.6 Female *Ythdc2*-deficient mice demonstrate defects in ovaries
A) Ovaries from *Ythdc* $^{-/-}$ mice are smaller in size than wild type at 12.5 d.p.p., 6 week-old and 12 week-old. (B) H&E staining of 14.5 d.p.p. ovaries sections. (C) Histological sections of 16.5-d.p.c. and 3.5 d.p.p.

We examined expression levels of *Ythdc2* mRNA and protein over time in testes, and found that *Ythdc2* expression rises between 7 and 12 d.p.p. and remains steady through adulthood (**Figure 2.7A,** data not shown). As meiosis in male mice is initiated at 8 d.p.p., we hypothesized that meiotic prophase I may be affected in *Ythdc2*$^{-/-}$ mice [16]. Immunofluorescence staining showed abnormalities in meiotic prophase I in *Ythdc2*$^{-/-}$ mice. Spermatocytes in *Ythdc2*$^{-/-}$ mice seemed to undergo leptotene normally. However, unlike control spermatocytes, spermatocytes in *Ythdc2*$^{-/-}$

mice failed to undergo pachytene, instead reaching a terminal zygotene-like stage, displaying fragmented axial elements and partially synapsed lateral elements (**Figure 2.7B,** data not shown). We also observed a similar failure to reach pachytene in oocytes in female *Ythdc2*$^{-/-}$ mice (**Figure 2.8A-B**).

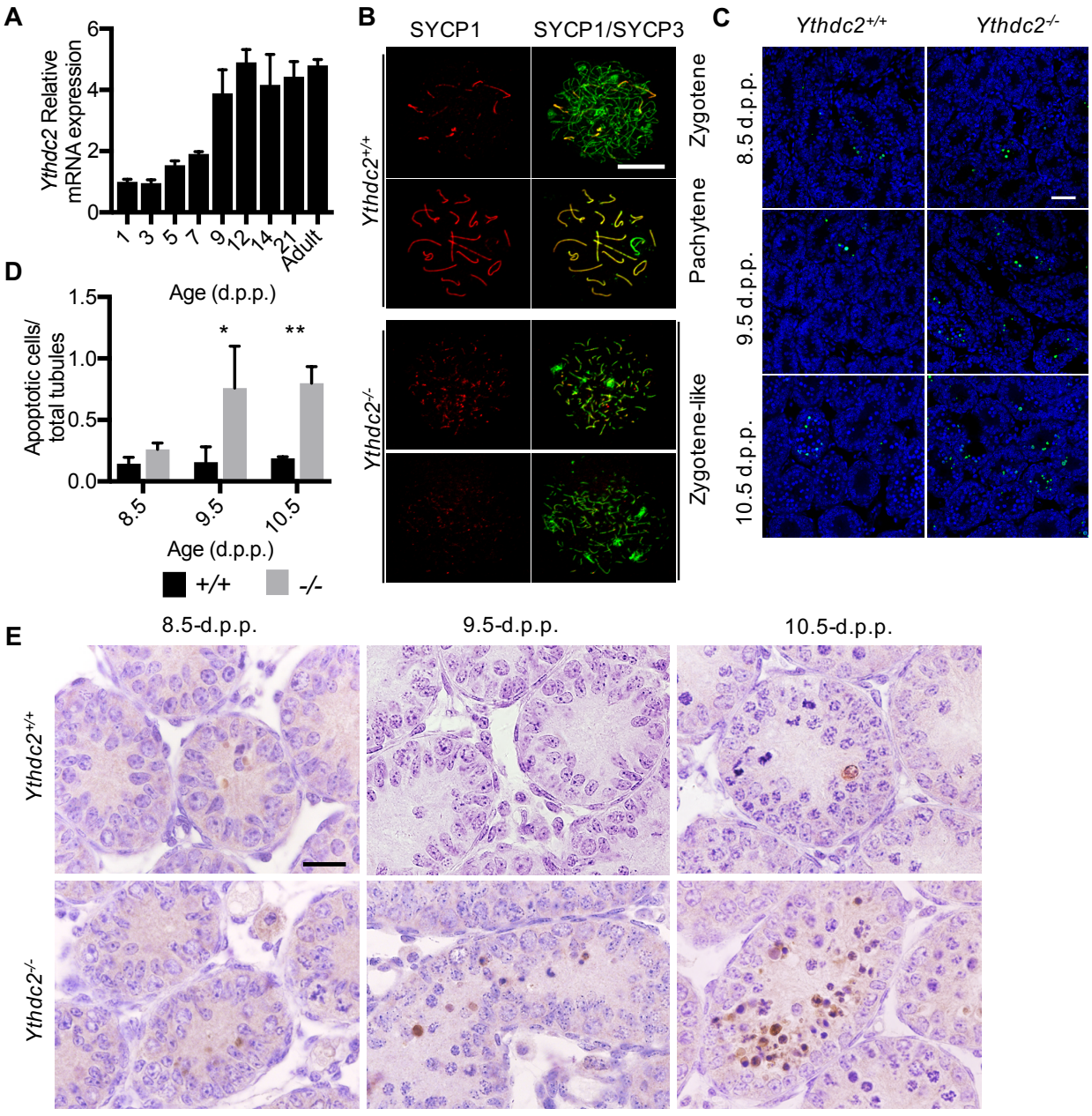

Figure 2.7 *Ythdc2*-deficient mice demonstrate defects in meiotic prophase I
(A) RT-qPCR analysis of *Ythdc2* mRNA levels in wild type testes over time, normalized to Rn18s as an internal control. Error bars, mean±sd, n=3, technical replicates (Student's t-test). (B) *Ythdc2*$^{+/+}$ and *Ythdc2*$^{-/-}$ spermatocyte chromosome spreads at 15.0 d.p.p. stained for SYCP1 (red) and SYCP3 (green). Scale bar, 10 µm. (C) TUNEL assay of testes sections from 8.5 d.p.p to 10.5 d.p.p.. DNA was stained with Hoechst (blue). Scale bar, 20 µm. (D) Quantification of apoptotic cells in *Ythdc2*$^{+/+}$ (black columns) and *Ythdc2*$^{-/-}$ (grey columns) testes at 8.5 to 10.5 d.p.p.. Error bars, mean±sd, n=3, biological replicates **P*<0.05, ***P*<0.01 (Student's t-test). (E) Hematoxylin and eosin (H&E) staining and cleaved Caspase-3 staining of testes sections at 8.5 -10.5 d.p.p. Scale bar, 20 µm.

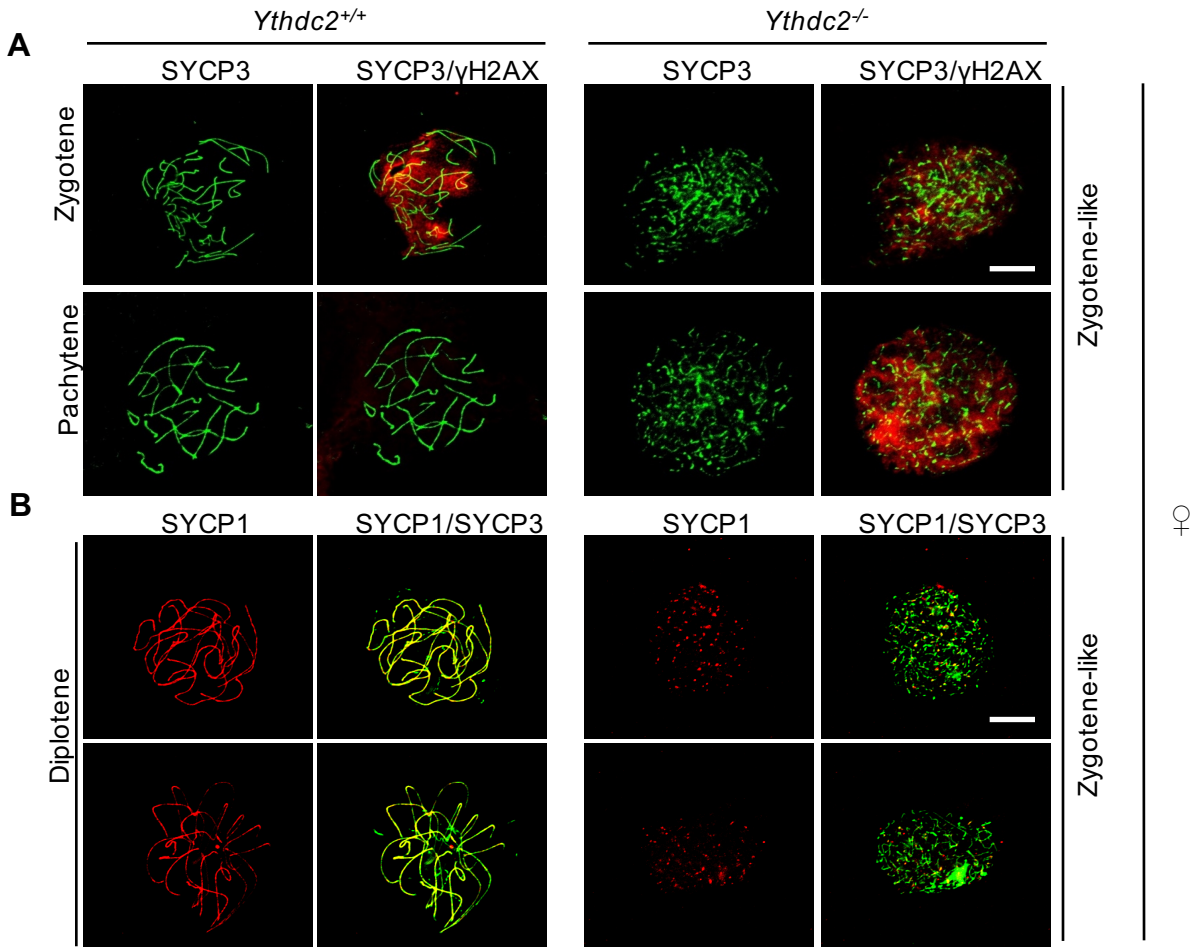

Figure 2.8 Meiotic defects in female *Ythdc2*-dependent meiocytes
(A) *Ythdc2*$^{+/+}$ and *Ythdc2*$^{-/-}$ oocyte chromosome spreads at 15.5 d.p.c. stained for SYCP3 (green) and γH2AX (red). Scale bar, 10 μm. (B) *Ythdc2*$^{+/+}$ and *Ythdc2*$^{-/-}$ oocyte chromosome spreads at 2.0 d.p.p. stained for SYCP3 (green) and SYCP1 (red). Scale bar, 10 μm.

We reasoned that the abnormal spermatocytes may undergo apoptosis, and thus performed TUNEL assays on testes sections from mice at the age of 8.5-10.5 d.p.p., around the time of meiotic initiation. We found substantially more apoptosis in the testes of *Ythdc2*$^{-/-}$ mice than those of the wild-type after 9.5-10.5 d.p.p. (**Figure 2.7C-D**). Cleaved Caspase-3 staining of young mouse testes also revealed that spermatocytes in *Ythdc2*$^{-/-}$ mice at 9.5-10.5 d.p.p. undergo apoptosis (**Figure 2.7E**). The number of spermatocytes in *Ythdc2*$^{-/-}$ mice is drastically reduced after 9.5 d.p.p. (**Figure 2.9**). The defect in spermatogenesis in *Ythdc2*$^{-/-}$ mice appeared to be limited to the spermatocyte

stage, as *Ythdc2*$^{-/-}$ mice displayed normal spermatogonia and sertoli cells (**Figure 2.10, Figure 2.11**). Thus, Ythdc2 is required for proper spermatocyte development.

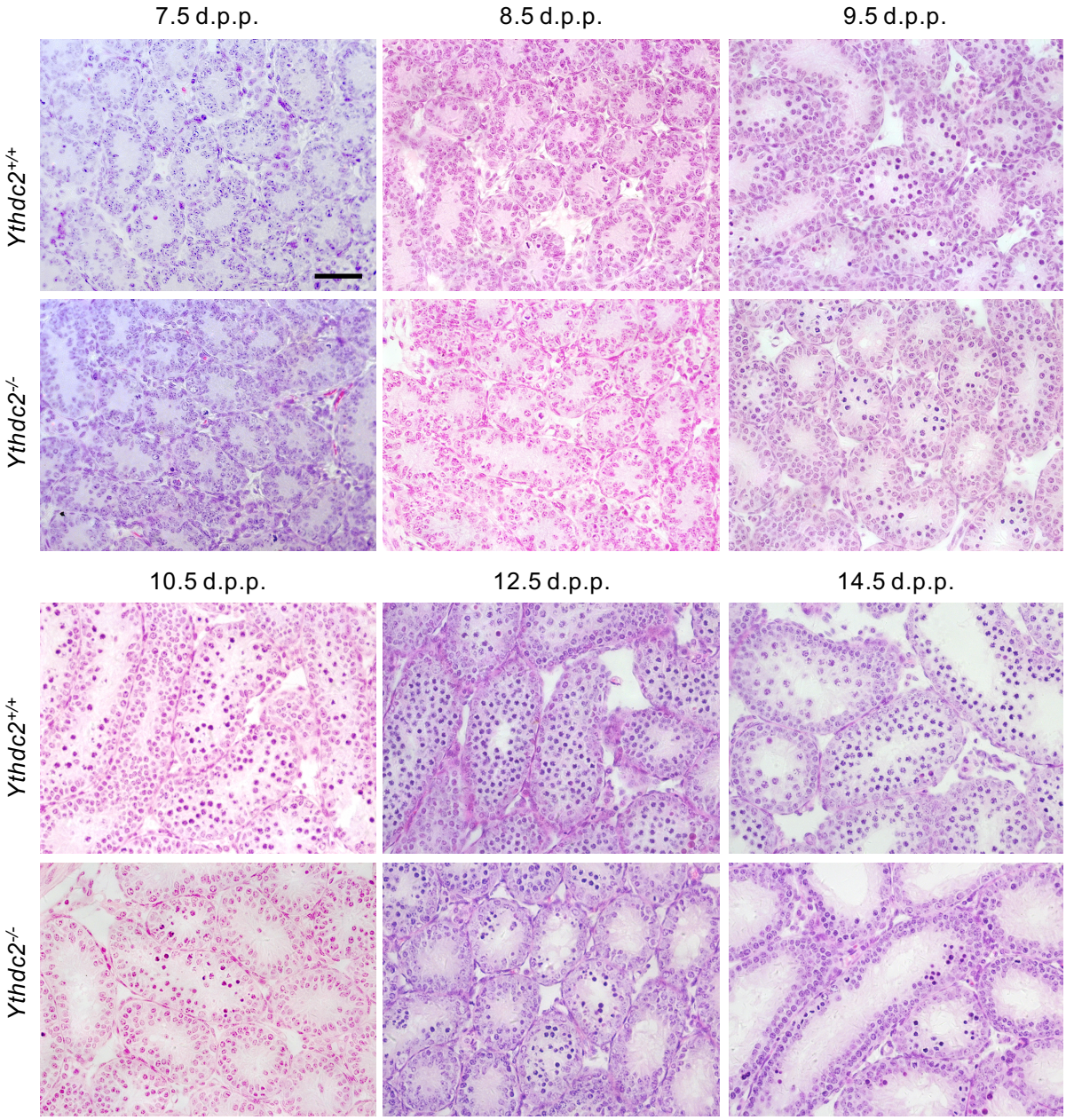

Figure 2.9 Comparative histological analysis of testes of young *Ythdc2*$^{+/+}$ and *Ythdc2*$^{-/-}$ mice
H&E staining of testes sections at 7.5-14.5 d.p.p. Scale bar 50 μm.

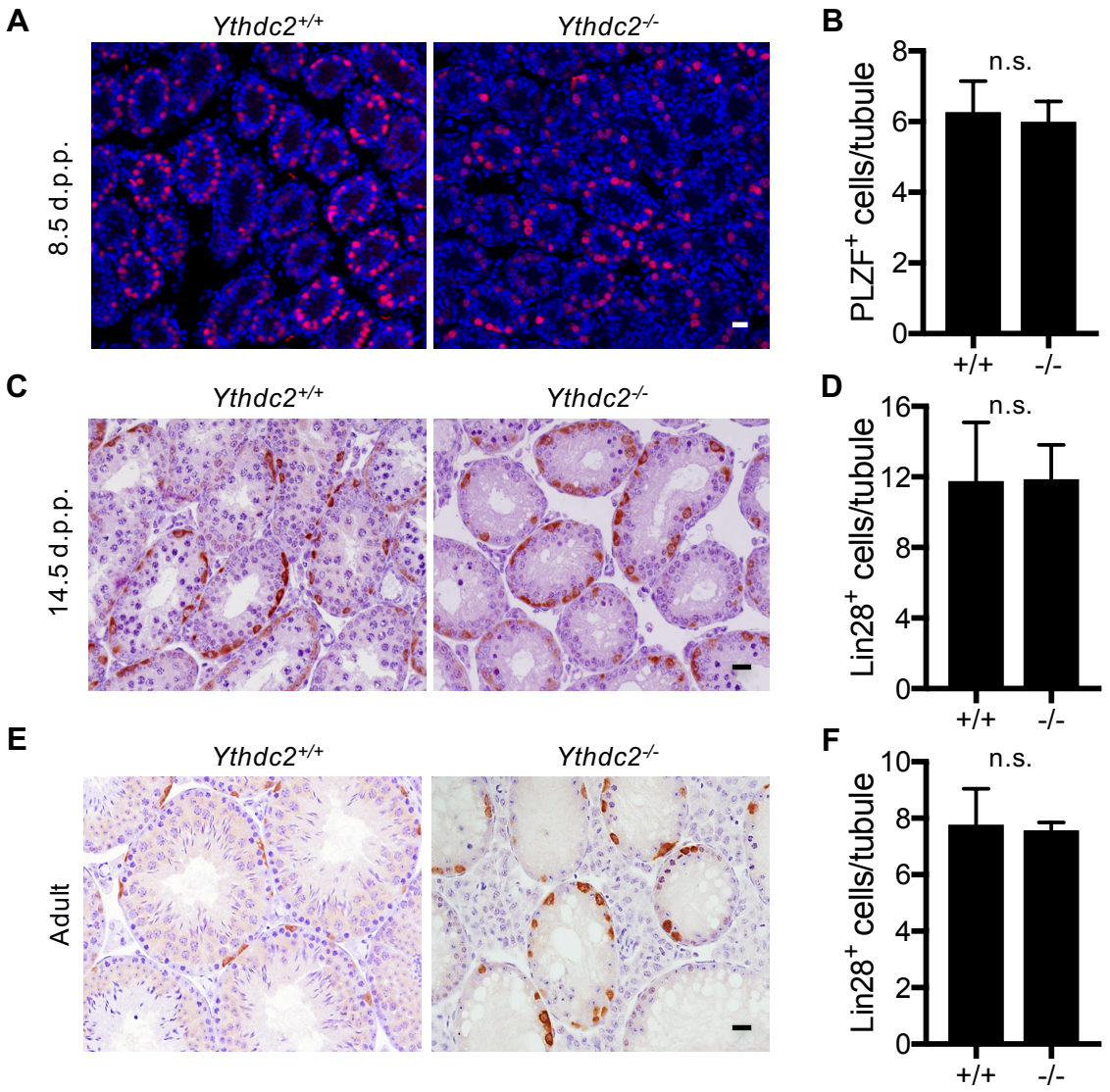

Figure 2.10 *Ythdc2*-deficient mice show normal spermatogonia
(A) PLZF staining (red) in 8.5 d.p.p. *Ythdc2*$^{+/+}$ and *Ythdc2*$^{-/-}$ testes. DNA was stained with Hoechst (blue). (B) Quantification of PLZF positive cells for (A). Error bars, mean±sd, n=3, biological replicates, n.s., non-significant (Student's t-test). (C) and (E), Lin28 staining (red) in 14.5 d.p.p. and adult Ythdc2$^{+/+}$ and Ythdc2$^{-/-}$ testes, respectively. Sections were counterstained with hematoxylin. (D) and (F), Quantification of Lin28 positive cells for (C) and (E). Error bars, mean±sd, n=3, biological replicates, n.s., non-significant (Student's t-test). Scale bar, 20 μm.

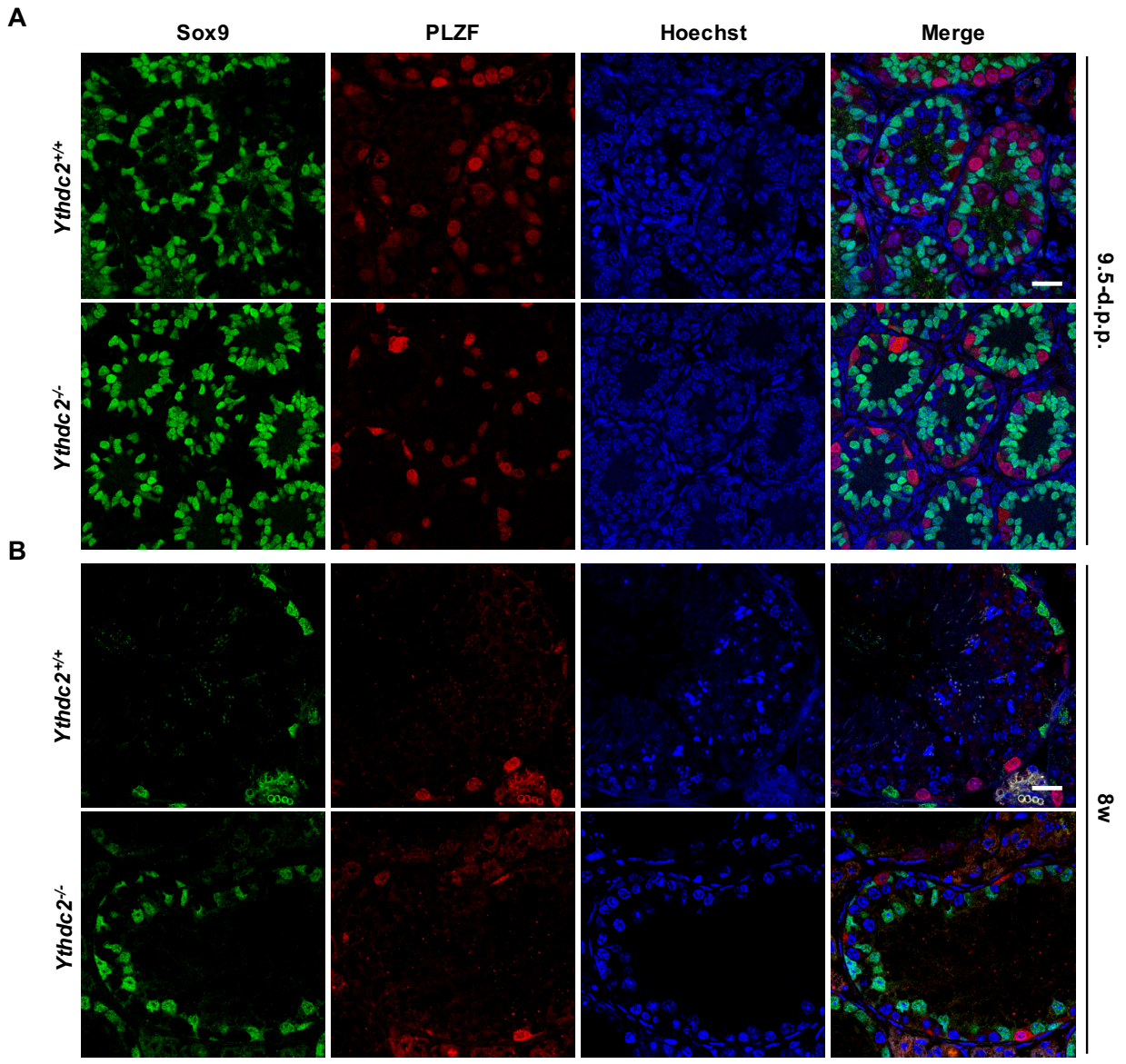

Figure 2.11 *Ythdc2*-deficient mice show normal sertoli cells
(A-B) Sox9 (green) and PLZF (red) staining in (A) 9.5 d.p.p. and (B) adult *Ythdc2*$^{+/+}$ and *Ythdc2*$^{-/-}$ testes. DNA was stained with Hoechst (blue). Scale bar, 20 μm.

2.2.3 Ythdc2 affects translation efficiency and mRNA abundance of its targets

To investigate the function of Ythdc2 in mouse testes, we evaluated the expression of its targets in 8.5 d.p.p. mouse testes. We focused on Smc3, a top RIP-seq target of Ythdc2 involved in spermatogenesis. Our m^{6}A-seq data showed that Smc3 contains m^{6}A, and we confirmed that it

contains m^{6}A by qRT-PCR (**Figure 2.12A**). Western blot of Smc3 in 8.5 d.p.p. mouse testes showed decreased expression in *Ythdc2*$^{-/-}$ mice (**Figure 2.12B**). Conversely, qRT-PCR of Smc3 in 8.5 d.p.p. mouse testes showed an increase in its mRNA abundance in *Ythdc2*$^{-/-}$ mice (**Figure 2.12C**). Overall, this resulted in a 37% decrease in translation efficiency in *Ythdc2*$^{-/-}$ mice (**Figure 2.12D**). We also saw similar changes of translation efficiency and mRNA abundance in Cep76, another top RIP-seq target of Ythdc2, which affects cell cycle (**Figure 2.12E-H**) Thus, Ythdc2 may function in affecting the translation efficiency and mRNA abundance of its targets in mouse testes.

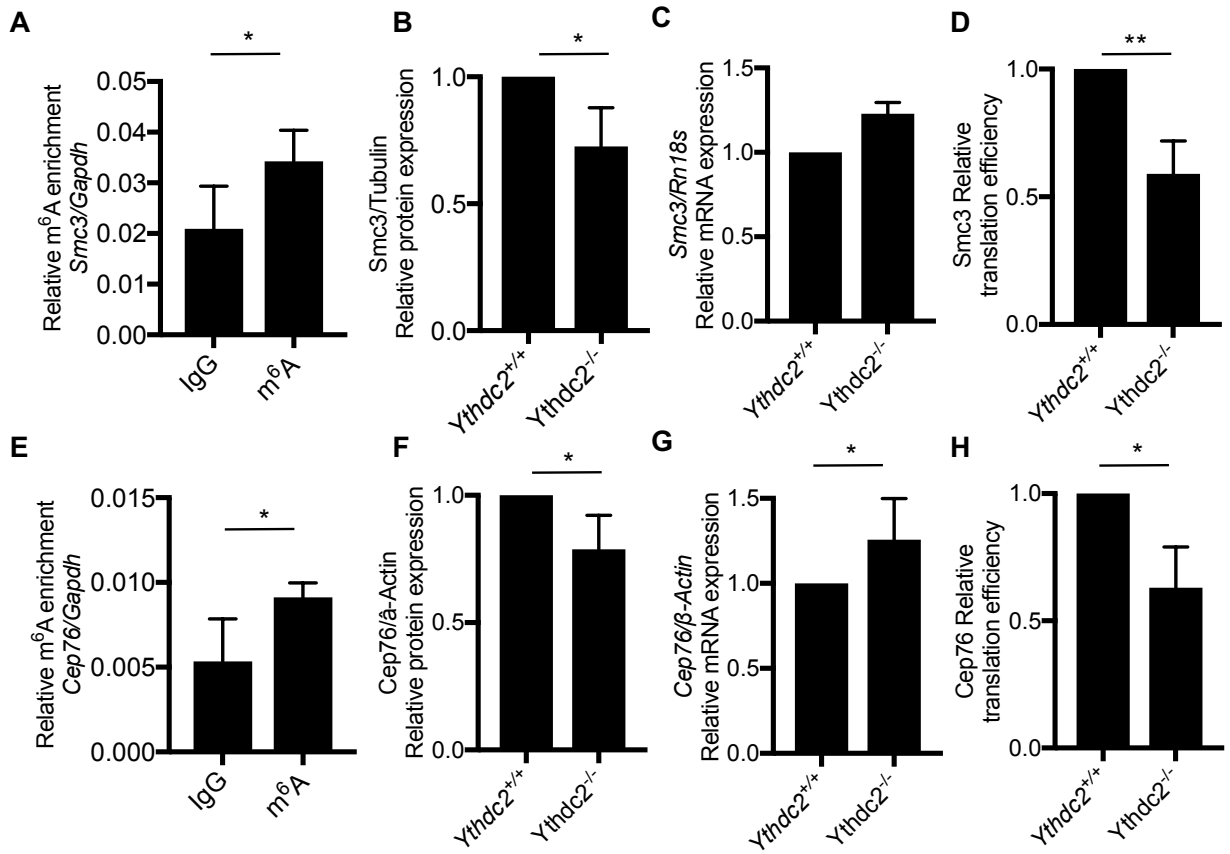

Figure 2.13 YTHDC2 affects the translation efficiency and stability of cellular targets
(A, E) Enrichment of m^{6}A in (A) Smc3 or (B) Cep76, determined by qRT-PCR. Figure shows m^{6}A enrichment normalized to Gapdh, an unmethylated gene. Error bars, mean ± sd, n = 4. *P<0.05 (Student's t-test). (B-D, F-H) protein level (n = 3), mRNA level (n = 6), and translation efficiency (protein level / mRNA level) of (B-D) Smc3 or (F-H) Cep76. Error bars, mean ± sd, *P<0.05 (Student's t-test).

We then investigated the direct role of YTHDC2 in affecting translation efficiency and mRNA abundance using a luciferase-based tethered reporter assay in HeLa cells as a model (**Figure 2.13A**) [3]. We replaced the YTHDC2 C-terminal YTH domain, which engages in m^6A binding in YTH family proteins [17,18], with λ peptide (N-YTHDC2-λ) known to specifically and tightly bind F-Luc-5boxB (five Box B RNA sequences inserted into the 3' UTR of a firefly luciferase reporter). We installed a Tet-Off inducible promotor, which prevents transcription in the presence of doxycycline (DOX), onto F-Luc-5BoxB, thus allowing inducible expression of the luciferase reporter. We compared the expression of firefly luciferase (normalized to renilla luciferase) with and without the presence of N-YTHDC2-λ, and saw a 30% increase in translation accompanied by a 15% decrease in mRNA abundance (**Figure 2.14B**). Overall, this resulted in a 52% increase in translation efficiency, indicating that YTHDC2 can enhance translation efficiency and decrease the stability of its targets in HeLa cells (**Figure 2.14C**).

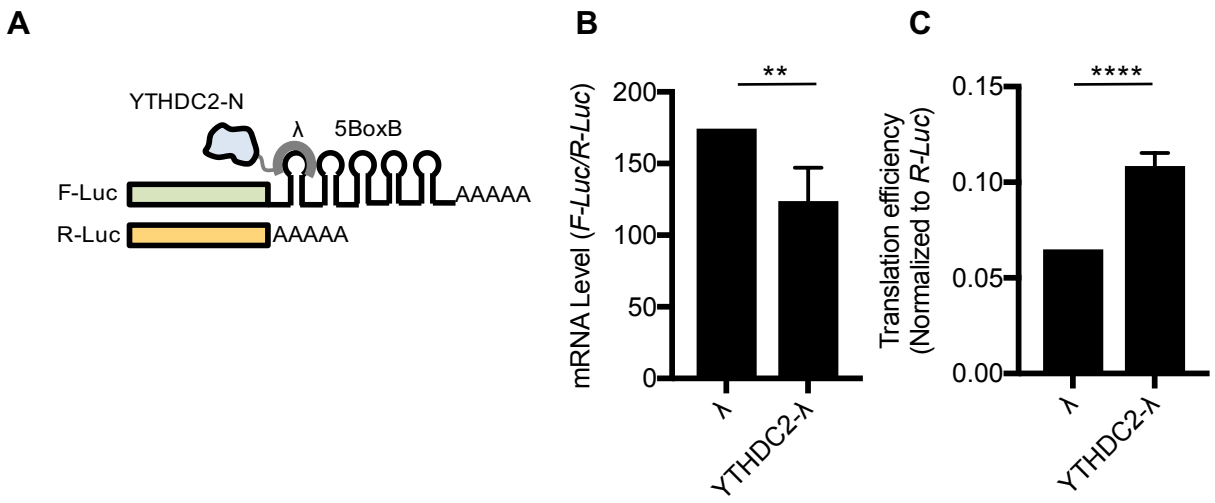

Figure 2.14 YTHDC2 affects the translation efficiency and stability *in vitro*
(A) Constructs in the tethering reporter assay. Effector: the YTH domain of YTHDC2 was replaced with λ peptide (N-YTHDC2-λ), which binds with high affinity to the BoxB RNA motif. Reporter: 5 BoxB domains were fused to the 3' UTR of firefly luciferase mRNA (F-Luc-5BoxB). Renilla luciferase (R-Luc) mRNA was used as an internal control to normalize luciferase signals from different samples. (B) mRNA level of F-Luc normalized to R-Luc 4 hours post F-Luc induction. Error bars, mean ± sd, $n = 4$. $^{**}P<0.01$ (Student's t-test). (C) Translation efficiency of F-Luc normalized to R-Luc 4 hours post F-Luc induction. Error bars, mean ± sd, $n = 4$. $^{****}P<0.0001$; (Student's t-test).

To investigate whether Ythdc2 also affects the stability and m^{6}A profile of its targets, we performed m^{6}A-seq on the testes of 8.5 d.p.p. *Ythdc2*$^{-/-}$ mice. m^{6}A was located on its usual consensus motif of GGACU in both control and *Ythdc2*$^{-/-}$ testes (**Figure 2.15A**). The CDS and 3' UTR harbored the majority of m^{6}A, as expected, but showed no substantial changes in m^{6}A upon *Ythdc2* depletion (**Figure 2.15B**). There was a slight increase of m^{6}A in the 5' UTR of mRNAs from *Ythdc2*$^{-/-}$ testes compared to control mice (**Figure 2.15B**). We then compared the RNA expression of genes in *Ythdc2*$^{+/+}$ and *Ythdc2*$^{-/-}$ testes. We observed a slight increase in the expression of genes with m^{6}A in *Ythdc2*$^{-/-}$ testes (**Figure 2.15C**). This difference was not present when comparing all genes, regardless of whether m^{6}A is present, in *Ythdc2*$^{+/+}$ and *Ythdc2*$^{-/-}$ testes (**Figure 2.15C**). We also observed a modest increase in the expression of Ythdc2 targets in *Ythdc2*$^{-}$

$^{-/-}$ mice compared to *Ythdc2*$^{+/+}$ mice (**Figure 2.15D**). Our results indicate a role of YTHDC2 in regulating target RNA stability.

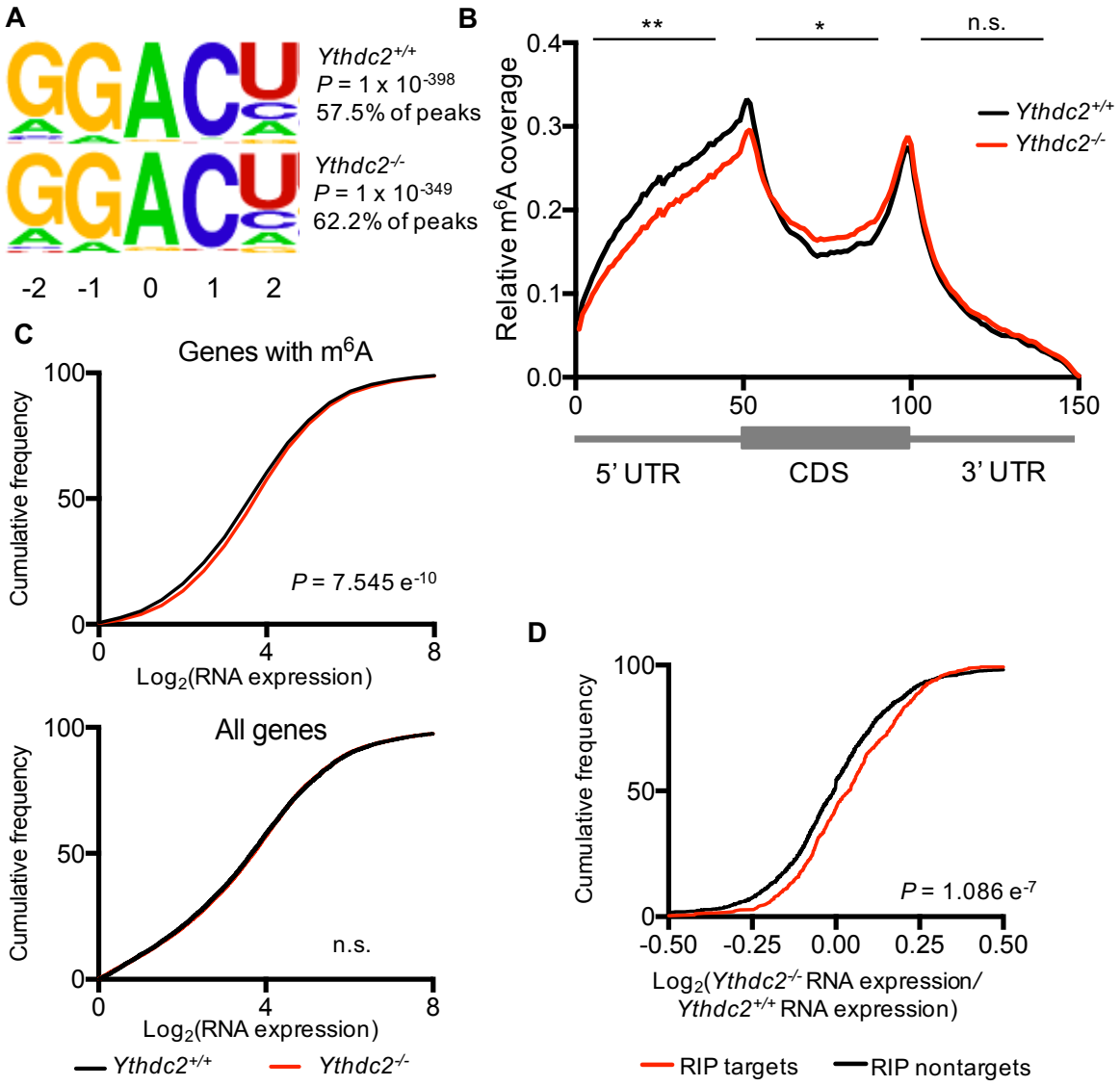

Figure 2.15 YTHDC2 affects mRNA stability
(A) Consensus motif of m^{6}A as identified by HOMER. (B) Metagene distribution of m^{6}A peaks. *P<0.05, **P<0.01 (Mann-Whitney U-test). (C) mRNA expression of genes with m^{6}A in *Ythdc2*$^{+/+}$ and *Ythdc2*$^{-/-}$ mice (Mann-Whitney U-test) (top); mRNA expression of all genes in *Ythdc2*$^{+/+}$ and *Ythdc2*$^{-/-}$ mice (bottom). (D) Cumulative frequency of log$_2$-fold changes of RNA expression upon *Ythdc2* depletion, in RIP targets and non-targets. *P* value was calculated using two-sided Mann-Whitney U test.

Next, we asked whether translation efficiency and stability of targets of YTHDC2 are regulated through its binding of m^{6}A. We knocked down m^{6}A methyltransferase METTL3 in HeLa cells and performed ribosome profiling [19]. We observed upon reduction of m^{6}A an overall decrease in translation efficiency and increase in mRNA abundance of YTHDC2 target transcripts, as determined by PAR-CLIP in HeLa cells (**Figure 2.16A-C**). This data suggests that the roles of YTHDC2 in enhancing translation efficiency and decreasing mRNA stability could be affected through its binding of m^{6}A, although this protein possesses multiple domains that can also bind RNA and potentially affect target fates.

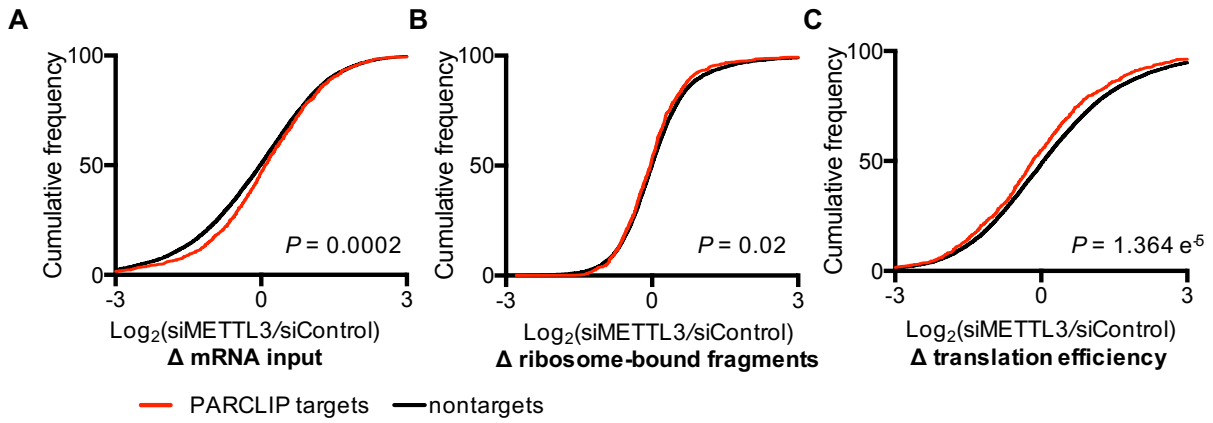

Figure 2.16 Effect of YTHDC2 on mRNA abundance and translation efficiency may be m^{6}A dependent
Cumulative frequency of $\log_2$-fold changes of (A) mRNA input, (B) ribosome-bound fragments, and (C) translation efficiency (ratio of ribosome-bound fragments / mRNA input) between siMettl3 and siControl for YTHDC2 PARCLIP targets (red) vs nontargets (black) in HeLa cells. *P* values were calculated using two-sided Mann-Whitney U test.

To explore how YTHDC2 may affect translation efficiency and mRNA stability, we performed polysome profiling of HeLa cells, separating cellular fractions along a 10-50% sucrose gradient. Upon assaying the fractions for expression of YTHDC2 with western blot, we observed YTHDC2 in the 40-80S ribosome fraction, suggesting that YTHDC2 may interact with cellular machinery involved in translation initiation (**Figure 2.17A**). We then immunoprecipitated the

YTHDC2-containing protein complexes from HeLa cells using tandem affinity purification and subjected the isolated peptides to mass spectrometry (**Figure 2.18B**). GO analysis of the top binding partner proteins revealed functions related to translation and translation initiation (**Figure 2.18C**). Moreover, the third most significant GO term of top binding partner proteins of YTHDC2 was nonsense-mediated decay, a process coupled with translation that eliminates mRNAs with premature translation-termination codons [20]. YTHDC2 partner proteins with known roles in nonsense-mediated decay included the exoribonuclease XRN1 and helicases UPF1 and MOV10 [21,22]. Thus, YTHDC2 may recruit decay machineries to accelerate the degradation of mRNA targets during or after translation in mammalian cells. Although the effect on stability observed in testes appears to be very modest, these experiments were done on heterogeneous testes. In purified sperm cells effects could be more significant, which could be studied in the future. Altogether, our results suggest that YTHDC2 may interact with translation and decay machineries to enhance the translation efficiency and decrease the mRNA abundance of its targets.

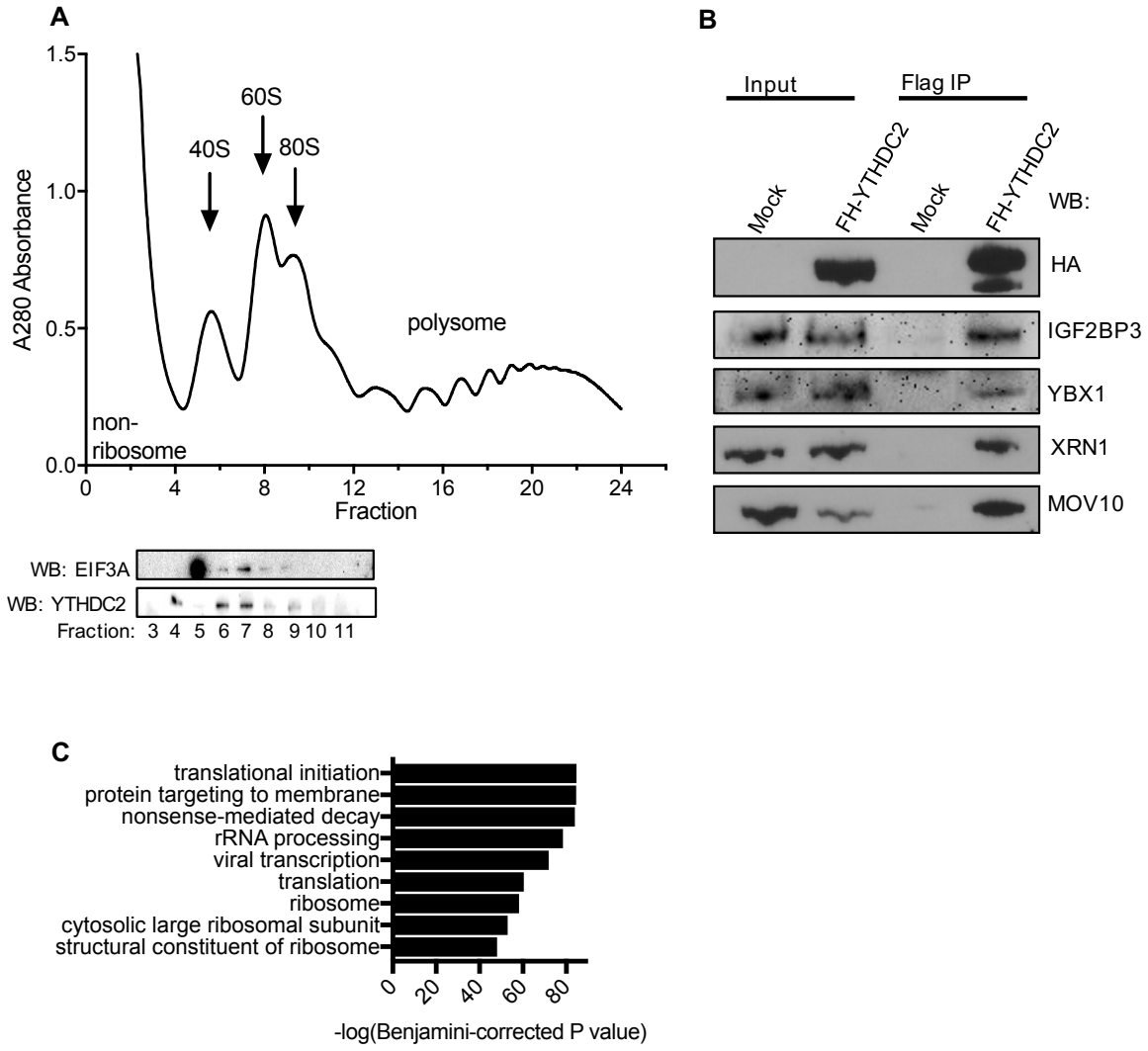

Figure 2.17 YTHDC2 may interact with translation and decay machineries
(A) Polysome profiling of HeLa cells, and western blotting of YTHDC2 and EIF3A in each fraction. (B) Confirmation of components of protein complex containing YTHDC2 obtained from tandem affinity purification followed by western blot. (C) Top GO terms in GO analysis of components of protein complex containing YTHDC2 obtained from tandem affinity purification followed by mass spectrometry.

2.3 Conclusions and Discussion: YTHDC2 belongs to the family, but plays unique roles

YTHDF1, YTHDF2, YTHDF3, and YTHDC1 are members of the YTH protein family with characterized roles in the cytoplasm and nucleus. The role of the final member of the YTH protein family, YTHDC2, remains uncertain. Here, we have systematically characterized the function of Ythdc2 and its role in spermatogenesis, describing it using a knockout mouse model that has yet to be reported for the YTH family proteins. Our results show that YTHDC2 preferentially binds m^6A within its consensus RNA motif of GGACU. YTHDC2 may increase the translation efficiency of its targets, as we showed in a tethering reporter assay, as well as decrease their stability. Ythdc2's role is essential to fertility, as both male and female mice lacking *Ythdc2* are infertile and do not contain germ cells developed past the zygotene stage of meiotic prophase I.

The YTH family proteins hold prominent effects on m^6A-containing mRNA transcripts. YTHDF1-3 function in the cytoplasm; YTHDF1 and YTHDF3 affect the translation efficiency of their targets, and YTHDF2 decreases the stability of its targets. YTHDC1 acts in the nucleus, affecting mRNA processing. Like its family members, YTHDC2 contains the highly evolutionarily conserved YTH domain. Recent studies crystallizing the YTH domains of YTHDF1, YTHDF2, and YTHDC1 have identified the structural basis by which they bind m^6A[5,17,23]. Thus, it is very likely that the YTH domain confers m^6A binding specificity to YTHDC2, However, YTHDC2 is otherwise distinct from its other family members. It is by far the largest protein (~160 kDa as compared to ~60-85 kDa), and it is the only to contain helicase domains. These helicase domains may contribute to RNA binding, or may have effects on secondary structure. YTHDC2 also contains two Ankyrin repeats between its helicase domains; these repeats mediate a diverse range

of protein-protein interactions. As the YTHDC2-containing protein complex includes proteins involved in translation and translation initiation (**Figure 3E**), the Ankyrin repeats may be important for the recruitment or binding of other protein complex members.

Because YTHDC2 contains multiple RNA-binding domains, it is possible that it may be a protein with multiple roles. We have elucidated roles pertaining to affecting translation efficiency and stability, and our data suggest that these roles may depend on the binding of YTHDC2 to m^6A. We have recently proposed that m^6A may be deposited onto transcripts to mark them for coordinated translation and decay[24]. This coordination may allow efficient transcriptome utilization and turnover in cells undergoing differentiation. As little transcription happens after germ cells enter meiotic prophase I[25,26], transcripts that have been translated and are no longer necessary in a later meiotic stage may need to be cleared so that the translation machinery can more effectively target the remaining transcripts. It thus makes sense that Ythdc2 might indeed work on transcripts to accelerate both translation and decay.

It is quite possible that the various RNA-binding domains of YTHDC2 could have functions independent of its m^6A-binding, which should be further studied in the future. Experiments introducing individual mutations to various domains could elucidate the roles of the individual domains and the mechanism underlying YTHDC2 function. Moreover, although we identified over 8,000 genes containing m^6A in mouse testes, RIP-seq of Ythdc2 in mouse testes revealed binding only to a subset of these genes. Future studies are necessary to determine the overarching role of m^6A throughout spermatogenesis, as well as Ythdc2 selectivity to its targets rather than other methylated transcripts.

Two recent articles have pointed toward a role of Ythdc2 in spermatogenesis. Abby et al. suggested that the meiosis-specific protein MEIOC interacts with YTHDC2 in an RNA-

independent manner to stabilize transcripts during meiosis prophase I, allowing proper induction of the meiotic program[14]. Soh et al. also found that MEIOC interacts with Ythdc2, but propose that MEIOC may destabilize its targets and is required to sustain the prolonged length of meiotic prophase I[15]. Our results indicate that the role of YTHDC2 can be complex. Its interactions with XRN1, UPF1, and MOV10 may suggest a role in destabilizing its target RNAs. Our tethering reporter assays and sequencing data also reveal that YTHDC2 could destabilize its target RNAs. It is possible that MEIOC and YTHDC2 work in tandem to respectively influence the stability and translation efficiency of their targets. MEIOC may work to stabilize meiotic transcripts transcribed at the start of meiotic prophase I. YTHDC2 may enhance the translation of transcripts, and then destabilize the transcripts after translation is complete, allowing proper progression of meiotic prophase I.

The m^6A demethylase Alkbh5 has also been shown to be most highly expressed in mouse testes, and its depletion leads to compromised spermatogenesis[27]. The m^6A mark may thus be an important regulator of spermatogenesis. Interestingly, unlike germ cells in *Ythdc2*$^{-/-}$ mice, germ cells in *Alkbh5*$^{-/-}$ mice develop into spermatozoa, albeit at a dramatically decreased rate and with morphological abnormalities. Thus, the m^6A mark may hold multiple roles throughout the various stages of spermatogenesis, and its misregulation may lead to defects in multiple pathways.

Our studies here are limited to functions of Ythdc2 in early spermatogenesis. It will be important to determine whether Ythdc2 holds functions past meiotic prophase I, which is possible through inducible knockout of Ythdc2. YTHDC2 may play important roles in other systems. Our results showed that Ythdc2 is also highly expressed in the spleen and brain; YTHDC2 may thus have potential roles in the immune and nervous systems. YTHDC2 has also been identified as a frequently mutated gene in pancreatic adenocarcinoma patients, and plays a role in cancer

metastasis by increasing the translation efficiency of HIF-1α[28,29]. As a protein with multiple domains, YTHDC2 can possess complex functions in different cellular contexts, potentially involving regulating both translation efficiency and transcript stability. Contributions of each domain to different functions will need to be studied in more details in the future.

2.4 Methods

2.4.1 Animals

All mice were housed under specific pathogen-free (SPF) conditions on a 12-h light–12 h dark cycle. All animal protocols were approved by the Animal Care and Use Committee of Nanjing Medical University.

2.4.2 In vitro Transcription and Microinjection of CRISPR/Cas9

Two sgRNAs were designed to target Exon 5 of the *Ythdc2* gene. Oligos for the generation of sgRNA expression plasmids were annealed and cloned into the Bsa I sites of pUC57-sgRNA (Addgene 51132) [30]. Oligo sequences are as follows: sgRNA1_up: TAGGAGTACAAAATATTCGGCAG; sgRNA1_down: AAACctgccgaatattttgtact; sgRNA2_up: TAGGacgactagcagcgattgctg; sgRNA2_down: AAACCAGCAATCGCTGCTAGTCGT. In vitro transcription and microinjection of CRISPR/Cas9 were performed as described [30]. Briefly, the Cas9 expression construct pST1374-Cas9-N-NLS-Flag-linker (Addgene 44758) was linearized with Age I and transcribed using the T7 Ultra Kit (Ambion), followed by purification using RNeasy Mini Kit (Qiagen). pUC57-sgRNA expression vectors were linearized by Dra I and transcribed using the MEGAshortscript Kit in vitro (Ambion). sgRNAs were purified by MEGAclear Kit (Ambion). Mixture of Cas9 mRNA (20 ng/μl) and two sgRNAs (5 ng/μl each)

were injected into the cytoplasm and male pronucleus of the zygote, obtained by CBF1 mating. Injected zygotes were transferred into pseudo-pregnant CD1 female mice. Founder mice were backcrossed to C57BL/6J. Mice used for experiments are mixed C57BL/6J and CBA genetic background.

Primers used for genotyping of *Ythdc2* mice: GATTCCTCAGTTCCTGTTAGA and GACCAATTCTTTCTCTTCTCTC.

2.4.3 Antibodies

The antibodies used in this study are listed as follows in the format of name (catalogue; supplier; application/amount used): Goat anti-YTHDC2 (Sc-249370; Santa-Cruz; IP/20μg; Western Blot/1:1,000); Rabbit anti-Ythdc2 (A303-026A; BETHYL; IF/1:200; Western Blot/1:2,000); Rabbit anti-m^6A (202003; Synaptic Systems; IP/2.5μg); Rabbit anti-eIF3a (3411; Cell Signaling Technology; Western Blot/1:2,000); Mouse anti-FLAG M2-peroxidase HRP (A8592; Sigma; Western Blot/1:10,000); Goat anti-GAPDH HRP (A00192-100; GenScript; Western Blot/1:5,000); Mouse anti-DDX4 (ab27591; Abcam; IF/1:200); Mouse anti-γ-H2A.X (ab26350; Abcam; IF/1:100); Mouse anti-SYCP3 (ab97672; Abcam; IF/1:100); Rabbit anti-SYCP3 (ab15093; Abcam; IF/1:100); Rabbit anti SYCP1 (ab15090; Abcam; IF/1:100); Mouse anti-SMC3 (ab128919; Abcam; Western Blot/1:2,000, IF/1:200); Rabbit anti-cleaved caspase3 (9661; CST; IHC/1:200); Rabbit anti-Cep76 (gift from Dr. Xueliang Zhu, Institute of Biochemistry and Cell Biology, SIBS, CAS, Western Blot/1:2,000); Rabbit anti-WT1 (sc-192; Santa-Cruz; Western Blot/1:2,000); Mouse anti-PLZF (sc-28319; Santa-Cruz; IF/1:200; Western Blot/1:2,000); Rabbit anti-Lin28 (13456-1-AP; Proteintech; IHC/1:200); Rabbit anti-SOX9 (ab5535; Chemicon; IF/1:200); Rabbit anti-IMP3 (IGF2BP3) (A303-426A; Bethyl; Western Blot/1:1,000); Rabbit

anti-YBX1 (ab76149; Abcam; Western Blot/1: 1,000); Rabbit anti-MOV10 (A301-571A; Bethyl; Western Blot/1:1,000); Rabbit anti-XRN1 (A300-443A; Bethyl; Western Blot/1:1,000); Mouse anti-Tubulin(AT819; Beyotime; Western Blot/1:10,000); Donkey anti-Rabbit IgG-HRP (sc-2313; Santa-Cruz; Western Blot/1:2,000); Donkey anti-Goat IgG-HRP (sc-2033; Santa-Cruz; Western Blot/1:2,000). Alexa Fluor ® 555 donkey anti-mouse IgG (A31570; Life Technologies; IF/1:500); Alexa Fluor ® 555 donkey anti-goat IgG (A21432; Life Technologies; IF/1:500); Alexa Fluor® 488 donkey anti-rabbit IgG (A21206; Life Technologies; IF/1:500); Alexa Fluor ® 488 donkey anti-mouse IgG (A21202; Life Technologies; IF/1:500); Alexa Fluor ® 555 donkey anti-rabbit IgG (A31572; Life Technologies; IF/1:500); and Alexa Fluor ® 633 donkey anti-goat IgG (A21082; Life Technologies; IF 1:500).

2.4.4 Histological analysis and immunofluorescence

Testes were fixed in Modified Davidson's Fluid (MDF), embedded in paraffin, sectioned at 5 μm, and stained with hematoxylin and eosin (H&E). For immunofluorescence, after dewaxing and hydration, paraffin sections were subjected to antigen retrieval by boiled in citrate buffer (1.8 mM citric acid, 8.2 mM sodium citrate, pH6.0), washed with PBS three times, blocked in 5% bovine serum albumin (BSA) for 2 h, and incubated with primary antibodies overnight at 4 °C. After washing with PBS three times, secondary antibodies were added to the samples for 2 h at room temperature (RT). The sections were then washed in PBS three times and incubated with Hoechst 33258 (Sigma) for 5 min at RT. Periodic acid-Schiff (PAS) staining of testis sections was performed according to the manufacturer's protocol (Sigma-Aldrich Cat. No. 395B-1KT).

For immunohistochemistry, after dewaxing and hydration, the sections were incubated with 10% hydrogen peroxide for 10 min to eliminate endogenous peroxidase activity at 37 °C. After

antigen retrieval with citrate buffer and washing, the sections were blocked in 5% BSA and incubated with primary antibodies and secondary antibodies. DAB staining kit (ZSGB-BIO, ZLI-9018) was used to detect the signal according to the manufacturer's instructions. The sections were counterstained with hematoxylin. For spermatocyte chromosome spreads, testes and ovaries were dissected and placed in hypotonic extraction buffer (30 mM Tris-HCl pH8.5, 50 mM sucrose, 17 mM citric acid, 5 mM EDTA, 2.5 mM DTT, 1 mM PMSF) for 45 min at RT. Tissues were separated and minced in 100 mM sucrose. Cell suspensions were pipetted onto slides and fixed (1% PFA, 0.15% TritonX-100, 10 mM sodium borate) for 3 h. After washing with TBS (150 mM NaCl, 20 mM Tris-HCl pH 7.6) four times, slides were blocked in ADB (1% Normal donkey serum, 0.3% BSA, 0.05% TritonX-100) for 1 h at RT, incubated with primary antibodies overnight at 37 °C, blocked in ADB for 5 h, and finally incubated with secondary antibodies for 1.5 h at 37 °C and washed with TBS.

H&E images were taken using an Axioskop 2 plus (ZEISS). Immunofluorescence images were taken with a LSM710 laser scanning confocal microscope (ZEISS).

2.4.5 Plasmid construction and protein expression

Flag-tagged YTHDC2 was cloned from commercial cDNA clones (Dharmacon) into vector pFastBac (*Bam*HI, *Xho*I; forward primer with coding sequence for Flag-tag, 5’ – GCGGATCCATGGACTACAAAGACGATGACGACAAGTCCAGGCCGAGCAGCGTCTC – 3’; reverse primer, 5’ – GCCTCGAGTCAATCAGTTGTGTTTTTTTCTCCCAAG – 3’). Flag-tagged YTHDC2 was purified using the Bac-to-Bac Baculovirus Expression System (Life Technologies) following the manufacturer’s protocols.

Tether effector (N-YTHDC2-λ) was cloned from commercial cDNA clones into vector pcDNA3.0 (*Bam*HI, *Xho*I; forward primer, 5' – GCGGATCCATGTCCAGGCCGAGCAG – 3'; reverse primer with λ peptide transcript 5' – GCCTCGAGTCAGTTTGCAGCTTTCCATTGAGCTTGTTTCTCAGCGCGACGCTCACGTC GTCGTGTTTGTGCGTCCATAGGCATGTTTGGTCTTGG – 3'). Tether control effector and reporter constructed have been previously described [3].

2.4.6 Mammalian cell culture, shRNA knockdown, and plasmid transfection

The 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 x Pen/Strep (Gibco). The 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 x Pen/Strep (Gibco), and 200μg/ml G418 (Sigma).

YTHDC2 shRNA were purchased from GE Dharmacon (RHS4533-EG64848); YTHDC2 shRNA-3 target sequence: ATATAAGAGATGTGACGAGGG; YTHDC2 shRNA-5 target sequence: CTTTAGTCGAAGTTCTGACTA. Lentiviruses containing the YTHDC2 shRNAs were produced in 293T cells with packaging plasmids. After 48 hours, viruses were collected and used to infect Hela cells. Stable YTHDC2 knock down cell lines were selected through 3 days of puromycin selection (2ug/ml).

YTHDC2 with an N terminal FLAG tag and HA tag in tandem was stably overexpressed in HeLa cells by puromycin selection (2 μg/ml) with a modified pPB-CAG vector. The control cell line expressing only FLAG and HA peptides in tandem was created similarly.

2.4.7 Tethering reporter assay

The tethering reporter assay was performed as previously reported, with slight modifications [4]. 30 ng of the reporter plasmid (pmirGlo-Ptight-5BoxB) and 270 ng of the effector plasmid (N-YTHDC2-λ or FLAG-GGS-λ as a control in pcDNA3.0) were used to transfect HeLa Tet-Off Advanced Cells (613356, Clontech) in each well of a six-well plate at 60-80% confluence under doxycycline (Dox, 100 ng/ml) inhibition. After six hours, the transfection mixture was replaced with fresh media containing DOX (100ng/ml). Eighteen hours later, each well was digested with trypsin, washed extensively with PBS by suspending and spinning down three times. Cells were re-seeded into six wells of a 96-well plate (1:10 in each well) and one plate of a 12-well plate (4:10) in media without DOX to induce transcription of *F-luc*. Four hours later, cells in the 96-well plate were analyzed by Dual-Glo Luciferase Assay System (E2920, Promega), and F-luc activity was normalized to R-Luc to evaluate protein production from the reporter. Concurrently, the RNA was extracted from the cells in the 12-well plate with Direct-zol RNA Microprep Kit (Zymo Research), following the manufacturer's protocol and including DNase I digestion. mRNA expression of *F-luc* and *R-luc* were quantified by RT-qPCR, and the expression of *F-luc* was normalized to that of *R-luc*. Translation efficiency of *F-luc* mRNA was calculated as the ratio of normalized F-luc protein level to normalized *F-luc* mRNA level.

2.4.8 RIP-seq

RIP-seq was performed as previously described, with minor modifications [2]. Testes were isolated from 2 16.5 d.p.p. male mice per sample; tunica vaginalis was removed and discarded. The remaining seminiferous tubules were lysed in high salt lysis buffer (300mM NaCl, 0.2% NP-40, 20mM Tris-HCl pH 7.6, 0.5mM DTT, 1:100 cOmplete EDTA-free protease inhibitor (Roche),

1:100 SUPERase In RNase Inhibitor (Ambion)) for 30 minutes with rotation at 4°C, followed by centrifugation at 16,000*g* for 20 minutes. Supernatant was diluted 1:1 in 20mM Tris-HCl pH 7.6, and filtered through a 0.22μm filter. Sample was precleared with 33μl Dynabeads Protein G beads (Life technologies) for 1 hour at 4°C. 100μl Dynabeads Protein G beads were coated with 20μg of Goat anti-YTHDC2 or normal goat IgG (Santa-Cruz) at 4°C for 1 hour following the manufacturer's instructions. 1/10 of the sample was saved as RNA input; the rest was incubated with the pre-coated beads for 4 hours at 4°C. The beads were washed six times with NT2 wash buffer (200mM NaCl, 2mM EDTA, 0.05% NP-40, 50mM Tris-HCl pH 7.6, 0.5μM DTT, 1:1,000 protease inhibitor, 1:1,000 SUPERase In RNAse Inhibitor) to remove excess RNA, and RNA was isolated with Direct-zol RNA Microprep Kit (Zymo Research). The sample immunoprecipitated using normal goat IgG did not contain any detectable RNA as detected by Nanodrop. rRNA was depleted in parallel from both the Input and IP samples using RiboMinus Eukaryote Kit v2 (Ambion). 5 ng of Input and IP RNA, 2 biological replicates each, were used to generate the library using TruSeq stranded mRNA sample preparation kit (Illumina).

2.4.9 CLIP-seq

CLIP-seq was performed as described for PAR-CLIP [31], without using 4SU and with the following modifications. Testes were isolated from 5 adult male mice per sample; tunica vaginalis was removed and discarded. The remaining seminiferous tubules were separated to single cells using a tissue grinder (Fisher), filtered through a 40μm cell strainer in 5ml PBS, then crosslinked with 254nm UV light three times using a Stratalinker 2400 at 150mJ/cm^2.

Mild enzyme digestion [32]: The first round of T1 digestion was carried out under 0.2U ml^{-1} for 15 minutes. The second round of T1 digestion was conducted under 10U ml^{-1} for 8 minutes.

Library construction: the final recovered RNA sample was further cleaned by RNA Clean & Concentrator (Zymo Research) before library construction by Truseq small RNA sample preparation kit (Illumina).

2.4.10 PAR-CLIP

PAR-CLIP was performed as previously described [2,31] with the following modifications. Sample preparation: we started with 150-200 million HeLa cells stably expressing Flag-HA-tagged YTHDC2. YTHDC2-RNA complex was SDS-PAGE purified with size selection from 130-200 kDa, and RNA fragments were extracted via ethanol precipitation after proteinase 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 library construction by Truseq small RNA sample preparation kit (Illumina).

2.4.11 m^6A-seq

m^6A-seq was performed as previously described [1]. Briefly, 100µg total cellular RNA was extracted from the seminiferous tubules of 8.5 d.p.p. mice using TRIzol following the manufacturer's protocol. PolyA mRNA was enriched using Dynabeads mRNA DIRECT Kit (Ambion) following the manufacturer's protocols. mRNA was sonicated to ~100nt, mixed with 2.5 mg affinity-purified anti-m^6A polyclonal antibody (Synaptic Systems) in IP buffer (150 mM NaCl, 0.1% NP-40, and 10mM Tris-HCl, pH 7.4), and incubated for 2 hours at 4°C. The antibody-RNA complex was isolated by incubation with protein A beads (Invitrogen) at 4°C for 2 hours. The beads were washed three times and eluted competitively with an m^6A monophosphate solution. RNA in the eluate was isolated using RNA Clean and Concentrator (Zymo Research) and used for library preparation with TruSeq stranded mRNA sample preparation kit (Illumina).

2.4.12 Polysome profiling

Polysome profiling of HeLa Control or shYTHDC2 knockdown cells was performed as previously described[10] with the 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 using 20 mM HEPES, pH 7.6, 100 mM KCl, 5 mM MgCl2 , 100 μg/ ml CHX, 1% Triton-X-100, 1:100 protease inhibitor (Roche), and 40 U/ml SUPERasin (Ambion).

2.4.13 Data analysis

Gene structure annotations were downloaded from UCSC mm10 RefSeq. Data analyses were performed as previously described [1,3]. Briefly, after removing adapters, sequencing reads were aligned to the reference genome (mm10) using TopHat (v2.0.14) [33]. For RIPSeq, RPKM were calculated by Cuffnorm [34]. RIP-seq targets were defined as genes enriched in the IP (log2(RIP/input) >= 1), and RIP-seq non-targets were defined as genes depleted in the IP (log2(RIP/input) <= -1). For m^6A-seq, the longest isoform was used if multiple isoforms were detected. Aligned reads were extended to 100 nt (average fragment size) and converted from genome-based coordinates to isoform-based coordinates to eliminate interference from introns in peak calling. To call m^6A peaks, the longest isoform of each human gene was scanned using a 100 nt sliding window with 10 nt steps. To reduce bias from potentially inaccurate gene structure annotation and the arbitrary usage of the longest isoform, windows with reads counts less than 1/20 of the top window in both m^6A IP and input sample were excluded. For each gene, the reads count in each window was normalized by the median count of all windows of that gene. A negative binomial model was used to identify the differential windows between IP and input samples by

using the edgeR package [35]. The window was called positive if FDR < 1% and log2(enrichment score) ≥ 1. Overlapping positive windows were merged. The following four numbers were calculated to obtain the enrichment score of each peak (or window): read count of the IP sample in the current peak/window (a), median read count of the IP sample in all 100 nt windows on the current mRNA (b), read count of the input sample in the current peak/window (c), and median read count of the input sample in all 100 nt windows on the current mRNA (d). The enrichment score of each window was calculated as (a × d)/(b × c). For CLIP-seq, we used a modified version of PARalyzer [36] to allow for all mutations, rather than just T to C mutations, for peak calling, while filtering out all SNPs and potential sequencing errors. Consensus motif was determined using HOMER [37]. For PAR-CLIP, we analyzed data using PARalyzer with default settings. Nonparametric Mann-Whitney U-test (Wilcoxon rank-sum test, two sided, significance level = 0.05) was applied to RNA-seq and ribosome profiling data analysis, as previously reported [3,38].

2.4.14 RT-qPCR

All RNA templates used in RT-qPCR were digested with DNase I during purification with Direct-zol RNA Microprep Kit (Zymo Research) to avoid genomic DNA contamination. All RT-qPCR primers covered exon-exon junctions when possible. 500ng of RNA was reverse-transcribed into cDNA with PrimeScript RT (Takara), then subjected to qPCR analysis with FastStart SYBR Green Master Mix (Roche) in a Roche LightCycler 96. Primer sequences are as follows: mYthdc2_F GGTCCGATCAATCATCTGT; mYthdc2_R GAAGTAACGAATAGGCATGT; m18s_F CATTCGAACGTCTGCCCTATC; m18s_R CCTGCTGCCTTCCTTGGA; mSmc3_F AACCAGGAGCTCAACGAGAC; mSmc3_R TGGAAACTCAGCAGCAAGAGA; mCep76_F

CCTCGGTCACCAGCAATGAA; mCep76_R GCTTGTTGTTCGCTCAAACTCA. The primer sequences of *F-luc* and *R-luc* have been reported [3].

2.4.15 Western blotting

Proteins were extracted with 8 M urea lysis buffer (8 M urea, 75 μM NaCl, 50 μM Tris-Cl pH 8.2) containing 1 mM PMSF, and quantified using the Bradford Protein Assay Kit (Beyotime). Protein samples were separated in a 7% SDS-PAGE gel and transferred to PVDF membranes (Millipore). Appropriate primary antibodies were incubated overnight at 4 °C after blocking in 5% non-fat milk. After washed in TBST three times, secondary antibodies were used at appropriate dilution for 2 h at RT. The signals were detected by the addition of Ncm ECL Ultra® (New Cell & Molecular Biotech Co., Ltd).

2.4.16 TUNEL assay

Detection of apoptotic cells was performed using TUNEL Apoptosis Detection Kit (Vazyme) according to the manufacturer's instructions. Images were taken with a LSM710 laser scanning confocal microscope (ZEISS).

2.4.17 In vitro probe pulldown

40 million HeLa cells stably overexpressing FLAG-tagged YTHDC2 or the seminiferous tubules of 3 16.5 d.p.p. mice were lysed at 4°C for 30 minutes in 3.5 ml lysis buffer (250 mM NaCl, 0.5% NP40, 10% glycerol, 20mM Tris pH 7.5, 1:100 cOmplete EDTA-free protease inhibitor (Roche), 1:100 SUPERase In RNase Inhibitor (Ambion)) or high salt lysis buffer (300mM NaCl, 0.2% NP-40, 20mM Tris-HCl pH 7.6, 0.5mM DTT, 1:100 cOmplete EDTA-free

protease inhibitor (Roche), 1:100 SUPERase In RNase Inhibitor (Ambion)), respectively, followed by centrifugation at 16,000*g* for 20 minutes. Supernatant of seminiferous tubules were diluted 1:1 in 20mM Tris-HCl pH 7.6. Both were filtered through a 0.22μm filter. 80 μl of the filtered supernatant was saved as input. The remaining supernatant was divided evenly among 4 tubes and incubated with 1.5 μg of biotinylated RNA probes**.** The lysate-probe mixture was incubated at 4°C for 2 hours with rotation. 15 μl Dynabeads MyOne Streptavidin C1 beads (Invitrogen) were washed 3x with lysis buffer, and added to each sample. The samples were then incubated at 4°C for 1.5 hours with rotation. The beads were isolated and washed four times with lysis buffer, and protein was eluted with 50 μl 2x Laemmli sample buffer (Bio-rad) at 95°C for 5 minutes. Protein was subjected to FLAG or YTHDC2 blotting analysis.

2.4.18 EMSA (electrophoretic mobility shift assay/gel shift assay).

The RNA probe was synthesized with a sequence of 5' - ACCGGXCUGUUACCAACACCCACACCCC-FAM 3' (X = m^6A or A; FAM = 6-carboxyfluorescein). FLAG-YTHDC2 was diluted to concentration series of 24 μM, 12 μM, 6 μM, 3 μM, 1.5 μM, and 0.75 μM in binding buffer (10 mM HEPES, pH 8.0, 50 mM KCl, 1 mM EDTA, 0.05% Triton-X-100, 5% glycerol, 10 μg ml^{-1} salmon DNA, 1 mM DTT and 40 U ml^{-1} RNasein). 1 μl RNA probe (4 nM final concentration) and 1 μl protein (2 400 nM, 1 200 nM, 600 nM, 300 nM, 150 nM, 75 nM final concentration) were added to 8 μl binding buffer, and the solution was incubated on ice for 30 minutes. The entire 10 μl RNA-protein mixture was loaded onto a Novex 6% TBE gel and run at 4°C for 90 minutes at 90V, and fluorescence was quantified. The K_d (dissociation constant) was calculated with nonlinear curve fitting (log(inhibitor) vs. response)

using Prism 7 software, where X = Log(protein concentration) and Y is the ratio of [RNA–protein]/([free RNA]+[RNA-protein]).

2.4.19 Tandem affinity purification of YTHDF3 protein interactome

We followed the previously reported procedure [10,39]. Protein complex components were identified by mass spectrometry and validated by western blot.

References

1. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–206 (2012).
2. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–20 (2014).
3. Wang, X. *et al.* N^6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161,** 1388–1399 (2015).
4. Shi, H. *et al.* YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Res.* **27,** 315–328 (2017).
5. Xu, C. *et al.* Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* **10,** 927–9 (2014).
6. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell* **61,** 507–519 (2016).
7. Liu, N. *et al.* N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518,** 560–564 (2015).
8. Zhou, J. *et al.* Dynamic m(6)A mRNA methylation directs translational control of heat

shock response. *Nature* **526,** 591–594 (2015).

9. Zhao, B. S. *et al.* m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **542,** 475–478 (2017).
10. Wang, X. *et al.* N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161,** 1388–1399 (2015).
11. Li, A. *et al.* Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. *Cell Res.* **27,** 444–447 (2017).
12. Du, H. *et al.* YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. *Nat. Commun.* **7,** 12626 (2016).
13. Morohashi, K. *et al.* Cyclosporin A associated helicase-like protein facilitates the association of hepatitis C virus RNA polymerase with its cellular cyclophilin B. *PLoS One* **6,** e18285 (2011).
14. Abby, E. *et al.* Implementation of meiosis prophase I programme requires a conserved retinoid-independent stabilizer of meiotic transcripts. *Nat. Commun.* **7,** 10324 (2016).
15. Soh, Y. Q. S. *et al.* Meioc maintains an extended meiotic prophase I in mice. *PLoS Genet.* **13,** e1006704 (2017).
16. Kerr, J. B., Loveland, K. L., O'Bryan, M. K. & De Kretzker, D. M. in *Knobil and Neill's Physiology of Reproduction* 827–947 (2006).
17. Zhu, T. *et al.* Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res.* **24,** 1493–6 (2014).
18. Wang, X. & He, C. Dynamic RNA Modifications in Posttranscriptional Regulation. *Mol. Cell* **56,** 5–12 (2014).
19. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-

adenosine methylation. *Nat. Chem. Biol.* **10,** 93–5 (2014).

20. Brogna, S. & Wen, J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat. Struct. Mol. Biol.* **16,** 107–113 (2009).

21. Gregersen, L. H. *et al.* MOV10 Is a 5′ to 3′ RNA Helicase Contributing to UPF1 mRNA Target Degradation by Translocation along 3′ UTRs. *Mol. Cell* **54,** 573–585 (2014).

22. Nagarajan, V. K., Jones, C. I., Newbury, S. F. & Green, P. J. XRN 5'→3' exoribonucleases: Structure, mechanisms and functions. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829,** 590–603 (2013).

23. Li, F., Zhao, D., Wu, J. & Shi, Y. Structure of the YTH domain of human YTHDF2 in complex with an m6A mononucleotide reveals an aromatic cage for m6A recognition. *Cell Res.* **24,** 1490–1492 (2014).

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

25. Monesi, V. Ribonucleic acid synthesis during mitosis and meiosis in the mouse testis. *J. Cell Biol.* **22,** 521–532 (1964).

26. Page, J. *et al.* Inactivation or non-reactivation: what accounts better for the silence of sex chromsoomes during mammalian male meiosis? *Chromosoma* **121,** 307–326 (2012).

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

28. Fanale, D. *et al.* Germline copy number variation in the YTHDC2 gene: does it have a role in finding a novel potential molecular target involved in pancreatic adenocarcinoma susceptibility? *Expert Opin. Ther. Targets* **18,** 841–850 (2014).

29. Tanabe, A. *et al.* RNA helicase YTHDC2 promotes cancer metastasis via the enhancement

of the efficiency by which HIF-1α mRNA is translated. *Cancer Lett.* **376,** 34–42 (2016).

30. Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11,** 399–402 (2014).
31. Hafner, M. *et al.* PAR-CliP - A Method to Identify Transcriptome-wide the Binding Sites of RNA Binding Proteins. *J. Vis. Exp.* (2010).
32. Kishore, S. *et al.* A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat. Methods* **8,** 559–64 (2011).
33. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
34. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7,** 562–78 (2012).
35. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2009).
36. Corcoran, D. L. *et al.* PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12,** R79 (2011).
37. 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).
38. Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science (80-.).* **336,** 233–237 (2012).
39. Shi, Y. *et al.* Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422,** 735–738 (2003).

Chapter 3 – Characterization of FMRP

3.1 FMRP as a potential m^{6}A reader

Fragile X syndrome (FXS) is the most commonly occurring form of inherited mental retardation in humans, affecting approximately 1 in 4,000 males and 1 in 8,000 females[1]. FXS is characterized by mental retardation, autistic behaviors, and seizures, and results from loss of function of the fragile X mental retardation protein (FMRP)[2]. FMRP is an mRNA binding protein that represses the translation of its target transcripts by stalling ribosome translocation[3–5]. FMRP also plays roles in mRNA localization, trafficking FMRP-associated mRNA granules in neurons from the soma to neurites[6]. Although FMRP is predominantly localized in the cytoplasm, it contains nuclear localization and nuclear export signals, which allow FMRP to enter the nucleus[7]. As localization of FMRP within nuclear pores has been observed, it has been suggested that FMRP shuttles between the cytoplasm and nucleus[8]. Whether it plays any role in the export of mRNA targets, as well as for which mRNA targets it may facilitate nuclear export, are unknown.

Our collaborators found in a global mass-spectrometry-based affinity screening for m^{6}A readers that FMRP may preferentially bind m^{6}A-containing transcripts[9]. We thus chose to further characterize the m^{6}A binding affinity of FMRP. As reader proteins may play multiple roles[10], we also wondered whether the shuttling activity of FMRP also affects m^{6}A-modified transcripts.

3.2 Results

3.2.1 FMRP preferentially binds m^{6}A *in vivo*

To determine whether FMRP binds directly to m^{6}A on mRNA rather than indirectly to unmethylated regions of m^{6}A-containing transcripts, we analyzed levels of m^{6}A bound to Flag-

FMRP stably expressed in HEK293T cells using cross-linking immunoprecipitation followed by quantitative ultra-high-performance liquid chromatography–tandem mass spectrometry (CLIP-LC-MS/MS QQQ). Mass spectrometry of immunopurified mRNA samples showed more than two-fold enrichment of m^6A in mRNAs pulled down by FMRP compared to input (**Figure 3.1A**). This enrichment was not present in Flag-FMRP-I304N, a mutant with decreased mRNA binding capacity, nor in immunoprecipitated RNA without RiboMinus treatment (**Figure 3.1B**), showing that m^6A binding by FMRP is specific and limited to mRNA.

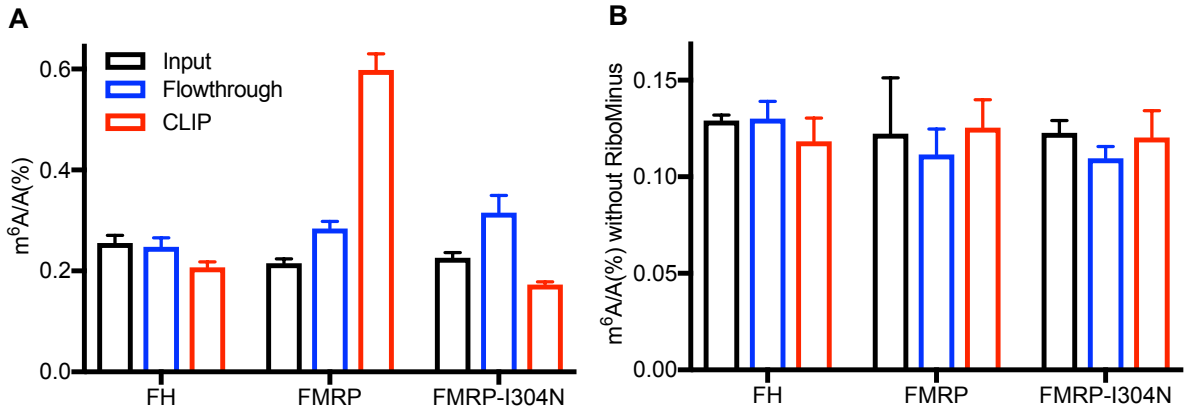

Figure 3.1 FMRP pulls down m^6A *in vivo*
LC-MS/MS QQQ quantification of (A) mRNA or (B) total RNA without RiboMinus in input, flowthrough, and CLIP fractions of HEK293T cells stably expressing Flag-HA (FH), FMRP, or FMRP-I304N.

To the same end, we also combined photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) and anti-m^6A immunoprecipitation (m^6A-IP) approaches. PAR-CLIP showed that the top two binding motifs of FMRP are similar to the consensus motif of m^6A (DRACH, where D=A,G,U; R=A,G; H=A,C,U) (**Figure 3.2A**). m^6A-IP performed on RNA fragments bound by FMRP in the PAR-CLIP experiment produced RNA libraries that could be subjected to sequencing, demonstrating that regions of RNA bound by FMRP contain m^6A. The top consensus motif of RNA isolated by m^6A-IP following FMRP PAR-CLIP matched the DRACH motif exactly (**Figure 3.2A**). Metagene analysis of FMRP PAR-CLIP

peaks and PAR-CLIP-m^{6}A-IP peaks showed that m^{6}A-modified binding sites of FMRP tend to cluster in the CDS and 3'UTR regions of transcripts, and FMRP PAR-CLIP targets contained more m^{6}A than non-targets (**Figure 3.2B-C**). We also saw that Fmrp PAR-CLIP targets in C57BL/6 mouse cortices contained more m^{6}A than non-targets (**Figure 3.2D**). Thus, FMRP directly targets m^{6}A sites on mRNA.

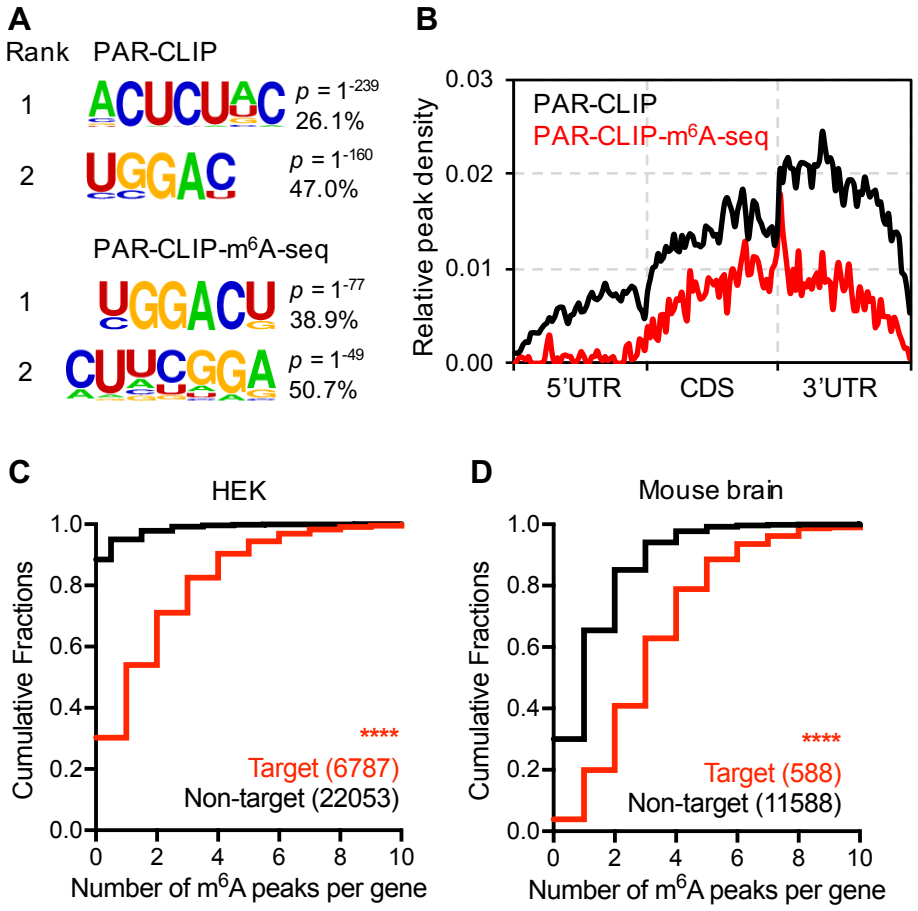

Figure 3.2 FMRP binds m^{6}A directly
(A) Top motifs for PAR-CLIP and PAR-CLIP-m^{6}A-seq of FMRP. (B) Distribution of peak densities in PAR-CLIP and PAR-CLIP-m^{6}A-seq of FMRP in HEK cells. (C) Cumulative frequency of m^{6}A peaks per gene in FMRP targets and nontargets in HEK cells ****$P < 0.0001$ (Mann-Whitney U test) ($n = 2$). (D) Cumulative frequency of m^{6}A peaks per gene in Fmrp targets and nontargets in P11 C57BL/6 mouse cortices ****$P < 0.0001$ (Mann-Whitney U test) ($n = 2$).

3.2.2 FMRP binding to m^6A involves the KH domain and is affected by salt concentration

The I304N mutant of FMRP has decreased RNA binding capacity, and a patient harboring the mutation presents with physical and behavioral characteristics matching those of severe Fragile X syndrome[9,11]. The I304N mutation is located in the nucleic acid binding region of the second KH domain of FMRP (**Figure 3.3A**). We wondered whether binding to m^6A requires the KH domain. To test the necessity of the KH domain for m^6A recognition, we performed an *in vitro* probe pulldown experiment using methylated and unmethylated 12-mer RNA probes. Wild-type FMRP bound methylated probe with greater preference than unmethylated probe, the I304N did not display any noticeable binding to the probe when we at standard salt concentration for this experiment (250 mM NaCl) (**Figure 3.3B**). This finding reflected others' findings that the I304 mutant has decreased RNA binding capacity. Thus, to determine whether the I304N mutant has any preference for methylated or unmethylated RNA, we performed the *in vitro* probe pulldown assay using lower salt concentrations. The I304N mutant demonstrated low but detectable RNA binding capacity in binding buffers containing 100 mM or 150 mM NaCl, and did not show a preference for a methylated probe (**Figure 3.3B**). On the contrary, the I304N mutant preferred the unmethylated probe at both low salt concentrations. Thus, the KH2 domain of FMRP is involved in m^6A recognition by FMRP. Our finding does not exclude the possibility that other domains are necessary for m^6A selectivity as well, as the KH1 and RGG domains both interact with RNA. Moreover, the I304N mutant does not affect only one residue involved in RNA binding, but rather destabilizes the entire KH domain[12], thus potentially also affecting other domains of FMRP that may contribute to binding of m^6A.

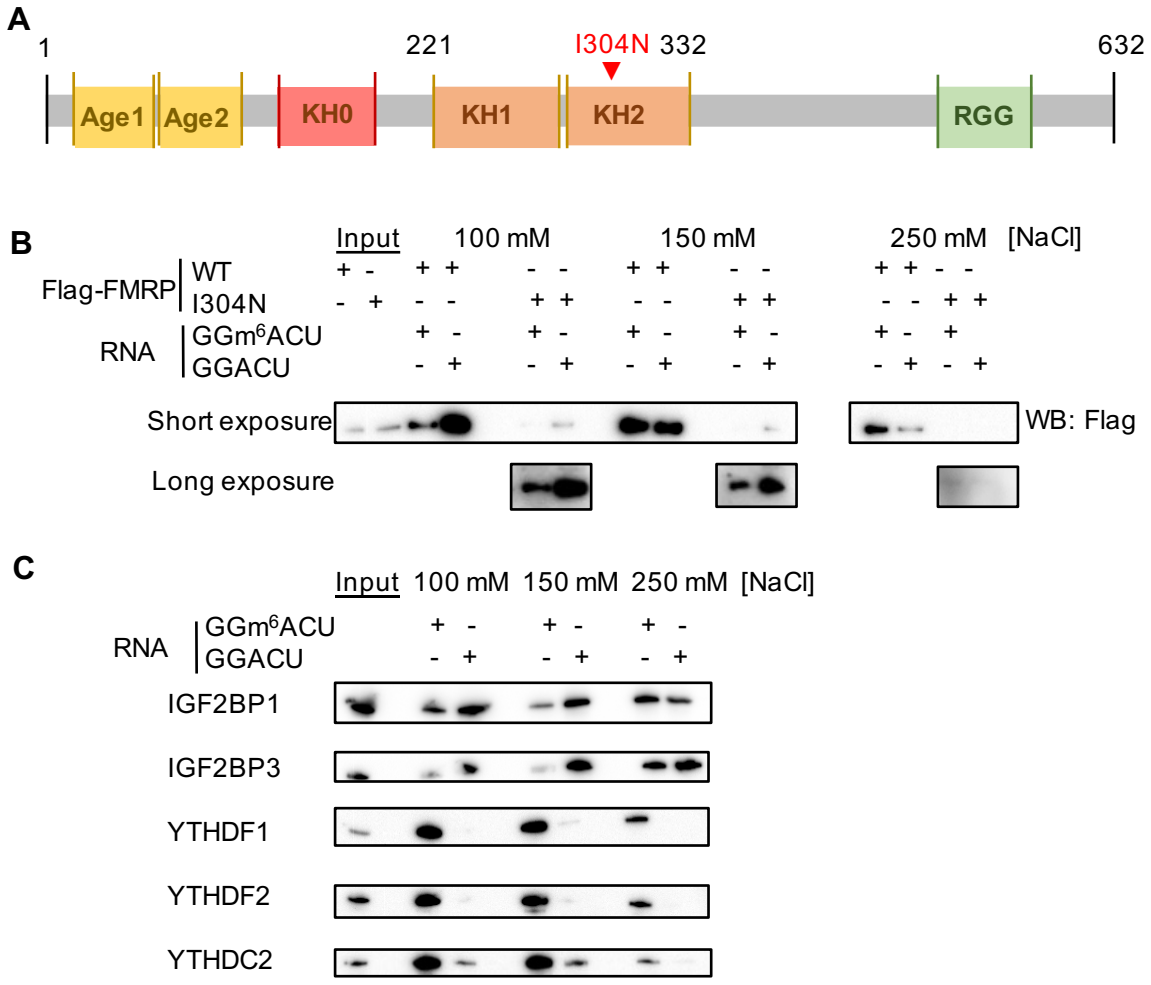

Figure 3.3 The FMRP KH domain and surrounding salt concentration affect binding to m^{6}A (A) Domains of FMRP (B) *In vitro* probe pulldown assay in HEK cells overexpressing FLAG-FMRP or FLAG-FMRP-I304N. Probe sequence: 5'- CGUGG**Y**CUGGCU-**B**-3' (**Y**=m^{6}A or A, **B**=biotin). Top, concentration of NaCl in lysis/binding buffer. (C) *In vitro* probe pulldown assay of IGF2BP and YTH proteins, using the same probes as in (B).

Interestingly, at the lowest salt concentration we used (100 mM NaCl), wild-type FMRP also demonstrated a preference for the unmethylated probe over the methylated probe (**Figure 3.3B**). This preference may arise due to the nature of the interaction between the nucleic acid binding region of the FMRP KH domains and adenosine in the RNA transcript. At lower salt concentrations, hydrogen bonds may form between the protein backbone of the FMRP KH domain and the N^6 and N^1 positions of the adenine[13]. These hydrogen bonds would stabilize an otherwise

disfavored RNA-protein interaction, as the nucleic acid binding region of the FMRP KH domains is hydrophobic due to the presence of an isoleucine or leucine in close proximity to the N^6 position of adenine. However, at higher salt concentrations, the hydrogen bonds would not form, and the hydrophobic binding pocket would prefer m^6A, which does not have the capacity to form a hydrogen bond at the N^6 position of adenine. To determine whether this preference for unmethylated probe at lower salt concentrations is not limited to FMRP, we tested the binding of IGF2BP proteins to methylated or unmethylated probes at various salt concentrations. The IGF2BP proteins also contain KH domains, and were recently shown to bind m^6A-marked transcripts[14]. As expected, the IGF2BP proteins preferred the unmethylated probe at lower salt concentrations, but preferred the methylated probe at higher salt concentrations (**Figure 3.3C**). This preference for unmethylated probe at lower salt concentrations seemed to be limited to KH domain proteins, as the YTH-family proteins YTHDF1, YTHDF2, and YTHDC2, which do not contain KH domains, demonstrated a strong selectivity for the methylated probe at all tested salt concentrations (**Figure 3.3C**). Thus, the hydrophobicity of the nucleic acid binding region of the KH domain may contribute to m^6A selectivity of FMRP.

3.2.3 FMRP facilitates the nuclear export of mRNA targets

To explore whether FMRP affects metabolism of methylated mRNAs, we depleted FMRP from HEK cells and quantified mRNA m^6A levels from total, cytoplasmic, and nuclear lysate by LC-MS/MS. Depletion of FMRP caused a substantial increase of m^6A levels in the nucleus, as well as a slight increase in the total and cytoplasmic fractions (**Figure 3.4A**). We also saw similar results in total, cytoplasmic, and nuclear fractions of cortices of postnatal day 11 (P11) C57BL/6 mice lacking Fmrp (*Fmr1*KO mice; Jackson Laboratory #003025) (**Figure 3.4B**).

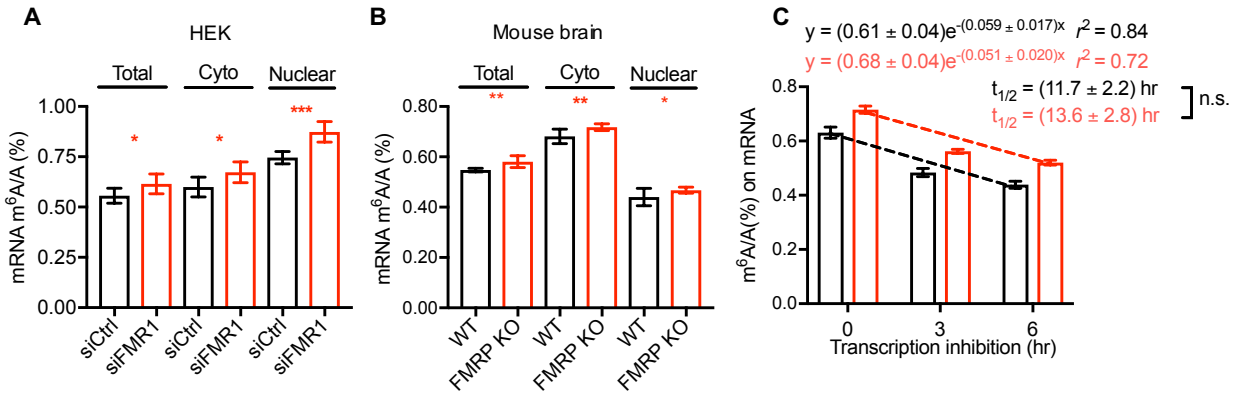

Figure 3.4 Depletion of FMRP leads to accumulation of m^{6}A in the nucleus
(A) HEK cells (n = 5, biological replicates) or (B) P11 C57BL/6 mouse cortices (n = 7, biological replicates) depleted of FMRP. Error bars, mean ± standard deviation *P < 0.05, **P < 0.01, ***P < 0.001 (Student's t-test). (C) LC-MS/MS QQQ quantification of mRNA m6A levels in HEK cells 0, 3, or 6 hours after transcription inhibition. Error bars, mean ± standard deviation, n = 2, two-sided t test with equal variance. Curves are fit to exponential decay.

To address the possibility that the increase in m^{6}A level may be due to RNA decay rather than RNA export, we treated cells HEK depleted of FMRP, as well as control cells, with Actinomycin D, and determined the level of m^{6}A on transcripts at 3 hours and 6 hours after the halt of transcription. We saw approximately similar decay of methylated mRNA in cells depleted of FMRP compared to control cells (**Figure 3.4C**)**.** Moreover, to our knowledge, FMRP has not been shown to interact with any cellular decay machinery.

We then performed RNA-seq on cytoplasmic and nuclear fractions of cortices of P11 WT and *Fmr1*KO mice, as well as m^{6}A sequencing (m^{6}A-seq) of P11 WT mice. In *Fmr1*KO mice, there was a decrease in the abundance of targets of FMRP[5] containing m^{6}A in the cytoplasm relative to the nucleus, but no difference in overall cytoplasmic or overall nuclear abundance (**Figure 3.5A-B**). We observed no difference in the abundance of non-targets of Fmrp in the cytoplasmic relative to the nucleus, nor overall cytoplasmic or overall nuclear abundance (**Figure 3.5B**). Interestingly, we also saw that there was a smaller but noticeable difference in the

abundance of unmethylated Fmrp targets in the cytoplasm relative to the nucleus (**Figure 3.5A**). Thus, FMRP may affect nuclear export indirectly, or FMRP may target more transcripts than were previously determined[5].

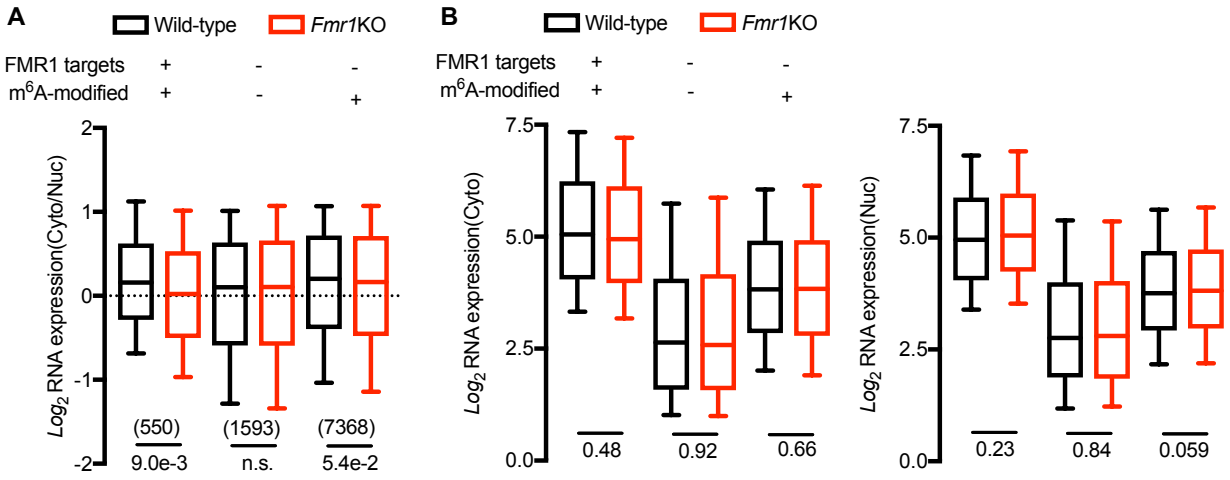

Figure 3.5 Depletion of Fmrp inhibits export of Fmrp targets
Abundance of Fmrp targets or nontargets with or without m^{6}A in (A) the cytoplasm relative to the nucleus or (B) the cytoplasm or nucleus alone of P11 C57BL/6 mouse cortices. *P* values calculated using two-sided Mann-Whitney U test shown below (n = 4, biological replicates).

The difference in the abundance of transcripts in the cytoplasm relative to the nucleus was especially significant for transcripts that were heavily methylated (containing 3 or more m^{6}A peaks) (**Figure 3.6A-B**). Additionally, a separate analysis of the top 40 targets of Fmrp associated with phenotypes[15] showed that these targets demonstrated a decrease in cytoplasmic level relative to nuclear level (**Figure 3.6C**). Overall, these results suggest that FMRP facilitates the nuclear export of its targets.

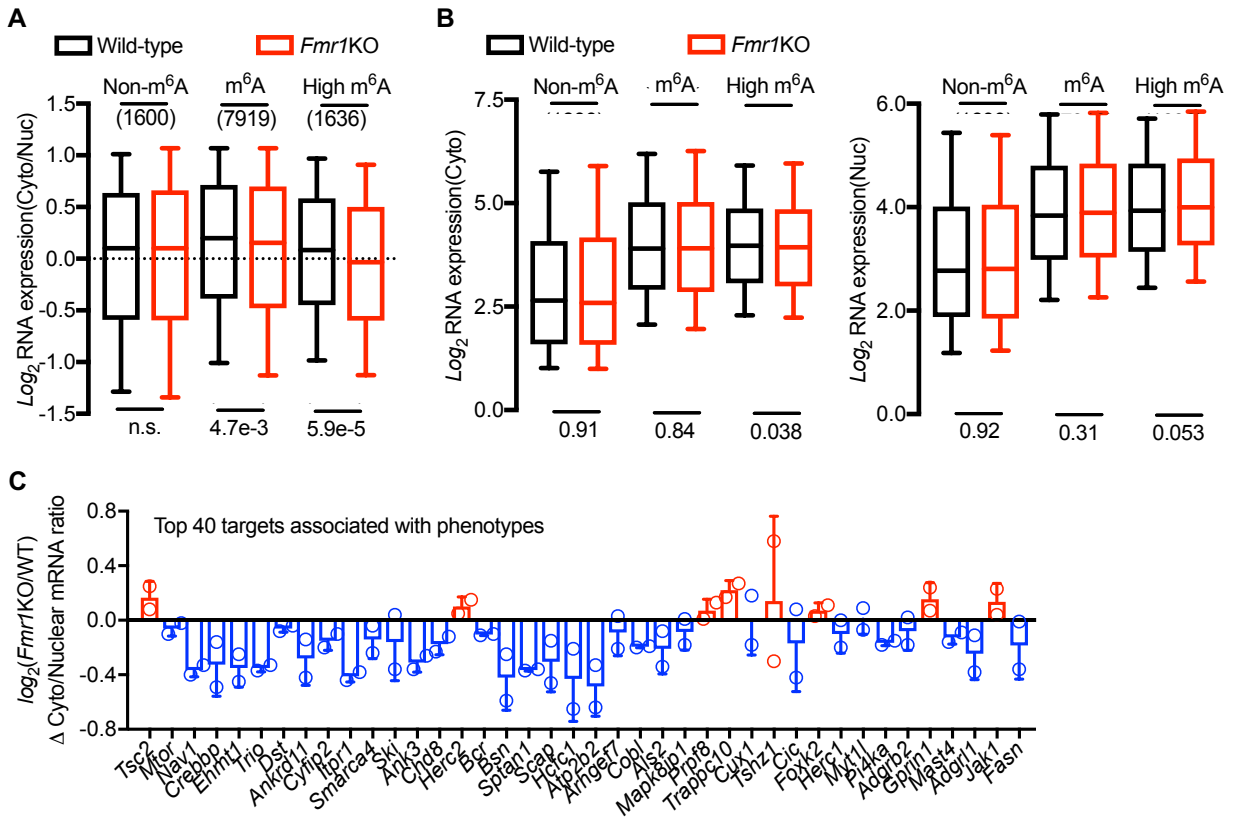

Figure 3.6 Depletion of Fmrp causes greater inhibition of export in heavily methylated transcripts

Abundance of unmethylated, methylated, and highly methylated transcripts (containing 3 or more m^{6}A peaks) in (A) the cytoplasm relative to the nucleus or (B) the cytoplasm or nucleus alone of P11 C57BL/6 mouse cortices. *P* values calculated using two-sided Mann-Whitney U test shown below (n = 4, biological replicates). (C) Abundance of top 40 targets of FMRP in the cytoplasm relative to the nucleus of P11 C57BL/6 mouse cortices (n = 4, biological replicates).

Finally, to explore the subcellular localization of m^{6}A-modified targets of FMRP in a dynamic state rather than at a steady state, we metabolically labelled nascent RNA transcripts in HEK cells with 5-Ethynyl Uridine (EU), and quantified their nuclear expression using qRT-PCR. Depletion of FMRP caused the accumulation of m^{6}A-modified targets of FMRP in the nucleus, but not transcripts that are non-targets of FMRP (**Figure 3.7A**). All targets of FMRP that we tested were pulled down in the PAR-CLIP-m^{6}A-IP, demonstrating that FMRP is bound directly to an m^{6}A site on the transcript. This data supported our findings that FMRP facilitates the nuclear

export of RNA transcripts, but the results could also be interpreted as an effect of FMRP in repressing the transcription of its targets. To exclude this possibility, we performed the same experiment using nascent RNA from the whole cell rather than nuclear or cytoplasmic fractions. Depletion of FMRP did not increase the transcription of m^{6}A-modified targets of FMRP; if anything, transcription decreased in some m^{6}A-modified targets of FMRP upon FMRP depletion (**Figure 3.7B**). Our data suggest that depletion of FMRP attenuates nuclear export of its target transcripts.

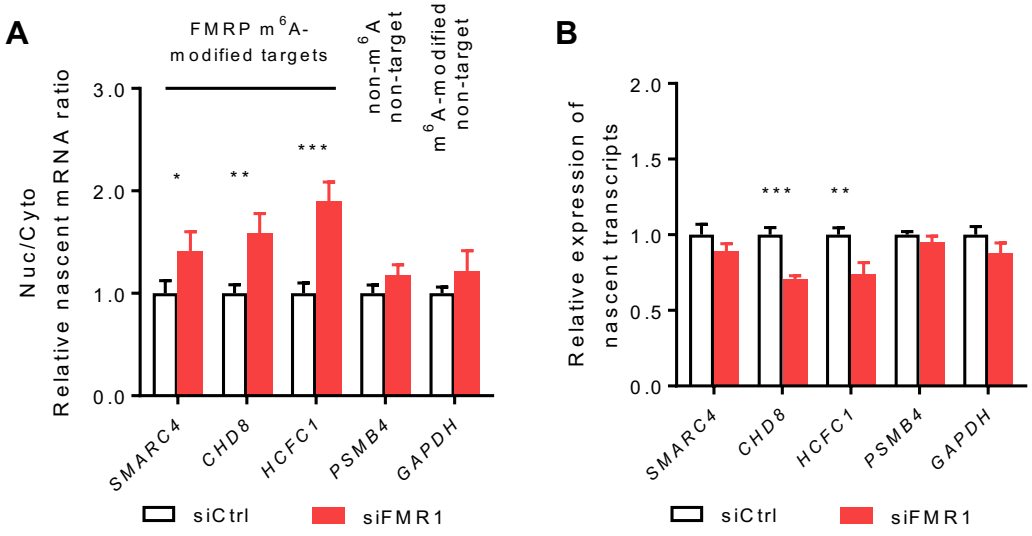

Figure 3.7 Analysis of FMRP export in a dynamic state
Abundance of EU-labeled nascent FMRP targets or nontargets with or without m^{6}A in (A) the nucleus relative to the cytoplasm or (B) the whole cell, quantified using qRT-PCR. Error bars, mean ± standard deviation, n = 2, biological replicates $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ (Student's t-test). 0.2mM EU, 1 hour labeling time.).

3.3 Conclusions and Discussion:

Studies using RNA affinity chromatography followed by mass spectrometry have identified several m^{6}A reader proteins, including a number of potential readers that do not belong to the prominent YTH family. Whether these proteins bind m^{6}A, whether any binding is direct,

and what role they play require investigation. FMRP is an mRNA binding protein with established roles in affecting translation and trafficking of its targets. Here we demonstrate an additional role – the direct binding of m^6A on transcripts and the nuclear export of these m^6A-containing transcripts.

This work adds to the expanding body of evidence that m^6A reader proteins may play multiple roles. For example, YTHDC1 affects both the splicing and export of its targets, and YTHDC2 affects the translation and abundance of its targets[10,16,17]. The roles of FMRP may be dependent on the functions of its protein partners as well as cellular context. Given that different m^6A reader proteins appear to have similar roles, it will be important to determine how these proteins target different transcripts and how they may complement or hinder the function of similar reader proteins.

Both FMRP and YTHDC1 play roles in the nuclear export of mRNAs[17]. Notably, the effect of FMRP on export of m^6A-containing transcripts appears to be weaker than the effect of YTHDC1; LC-MS/MS QQQ data of cells depleted of YTHDC1 demonstrate both a decrease in m^6A level in the cytoplasm as well as an increase in m^6A level in the nucleus, whereas LC-MS/MS QQQ data of cells depleted of FMRP demonstrate no decrease in m^6A level in the cytoplasm. This difference may be due to one or more of several reasons. YTHDC1 may bind more transcripts overall than FMRP; in the study of the effect of YTHDC1 on export, over 700 targets of YTHDC1 in HeLa cells were identified using a stringent cutoff that took into account both PAR-CLIP and RIP-seq data. In comparison, FMRP in mouse cortices had around 550 targets using a cutoff that took into account only PAR-CLIP. While there may be more "true" targets of YTHDC1 and FMRP than identified due to these cutoffs, it is within reason that YTHDC1 has a broader range of targets, and thus show a stronger effect in LC-MS/MS QQQ measures. Indeed, we found over 7,000

methylated genes in mouse cortices, of which only a small subset was targeted by FMRP. Moreover, in mouse cortices, FMRP may have different roles in different cell types, thus only affecting the export of a subset of cells rather than the whole brain.

Another reason behind the difference in effects of YTHDC1 or FMRP knockdown on m^6A levels may be due to the multiple roles played by FMRP in the cell. We noticed that, in fact, cells depleted of FMRP appear to have slightly increased m^6A levels in the cytoplasm, as well as in RNA extracted from the whole cell. FMRP may facilitate the export of only a small subset of transcripts, and the increase in m^6A levels in the cytoplasm may be caused by roles of FMRP unrelated to export. For example, FMRP may interact with or compete with other m^6A reader proteins in the cytoplasm that affect RNA stability.

Additionally, FMRP's role in ribosome stalling may affect the processing and eventual decay of a large number of transcripts, including ones not directly targeted by FMRP. It has been shown that FMRP lacking an RGG motif can still associate with polyribosomes[18]. Given that this mutant may not bind RNA, FMRP association with polyribosomes may be RNA-independent, thus affecting ribosome stalling on a large number of transcripts. Transcripts on which ribosomes are stalled undergo decay[19], thus allowing FMRP to potentially alter the stability of cytoplasmic transcripts without direct binding activity.

Finally, any given transcript target of FMRP may not always be methylated. We found 558 target transcripts of FMRP, 550 of which were methylated. However, our m^6A-seq data does not tell us whether each of the 550 methylated transcripts always have methylations. FMRP may still bind onto its targets despite a lack of m^6A. Indeed, the binding of FMRP to the m^6A probe in a recent affinity chromatography-mass spectrometry appeared weaker than the binding of many other proteins, including the YTH family of proteins. *In vitro* probe pulldown experiments also

showed that FMRP tolerates unmethylated transcripts, and in certain conditions perhaps even prefers them.

Our studies here are limited to the function of FMRP in affecting the export of methylated transcripts. It will be important to determine whether other roles of FMRP, including ribosome stalling and mRNA localization within the cytoplasm, are carried out with respect to the methylation status of transcripts. These studies would shed light not only on the function of FMRP, but on the comprehensive function of m^6A in a physiological setting.

3.4 Methods

3.4.1 Animals

The *Fmr1*KO mice (Jackson Laboratory #003025 or *Fmr1*KO1) were obtained from Dr. Anis Contractor's laboratory. This mouse line was created and initially characterized by the Dutch-Belgian Fragile X Consortium[20] and backcrossed onto the C57BL/6 background and distributed by Dr. David Nelson's laboratory at Baylor College of Medicine.

3.4.2 RNA isolation and RNA-seq

The nuclear and cytoplasmic fractions of postnatal day 11 (P11) Fmr1KO and wild type mouse cortices, as well as HEK cells, were isolated using the NE-PER kit (ThermoFisher #78835) following manufacturer's instructions. Total RNA was extracted using Trizol, mRNA was further purified using the mRNA DIRECT kit (ThermoFisher #61012) and RiboMinus Eukaryotic kit v2 (ThermoFisher #A15015). RNA-seq library preparation was performed with TruSeq stranded mRNA sample preparation kit (Illumina)

3.4.3 Antibodies

The antibodies used in this study are listed as follows in the format of name (catalogue; supplier; application/amount used): Rabbit anti-FMRP (ab17722; Abcam; Western Blot/1:1 000); Mouse anti-FLAG M2-peroxidase HRP (A8592; Sigma; Western Blot/1:10 000).

3.4.4 Plasmid construction and protein expression

FMRP isoform 1 (pFRT-TODestFLAGHAhFMRPiso1; 48690) and FMRP-I304N isoform 1 (pFRT-TODestFLAGHAhFMRPiso1I304N; 48692) vectors were obtained from Addgene.

Flag-tagged FMRP and FMRP-I304N were cloned from the above vectors into a modified pPB-CAG vector containing a Flag-HA tag (*Age*I, *Sal*I on insert, *Age*I, *Xho*I on vector; forward primer 5' – atcgACCGGTATGGAGGAGCTGGTGGTGGAA – 3'; reverse primer, 5' – atcgGTCGACTTAGGGTACTCCATTCACGAGACC – 3').

3.4.5 Mammalian cell culture and plasmid transfection

HEK cells were grown in DMEM (Gibco, 11995) media supplemented with 10% FBS and 1% 100 x Pen/Strep (Gibco). FMRP and FMRP-I304N with an N terminal FLAG tag and HA tag in tandem were stably overexpressed in HEK cells by puromycin selection (5 μg/ml) with a modified pPB-CAG vector. The control cell line expressing only FLAG and HA peptides in tandem was created similarly.

3.4.6 CLIP-LC-MS/MS QQQ

2.4×10^8 HEK293T cells stably expressing Flag-HA, Flag-HA-FMRP, or Flag-HA-FMRP-I304N were UV-cross-linked twice with 150 mJ/cm^2 at 254 nm and harvested with a cell

scraper. Cells were lysed in two volumes of lysis buffer containing 150 mM KCl, 2 mM EDTA, 10 mM HEPES, 0.5% NP-40, 0.5 mM DTT, 1:100 cOmplete protease inhibitor (Roche; 04693116001), and 1:100 SUPERase In RNase inhibitor (20 U/μL; Ambion; AM2694) for 30 min in a rotation wheel at 4 °C. Lysate was cleared by centrifugation at 14,000 r.p.m. for 15 min at 4 °C followed by syringe filtration through a 0.45-μm filter. 50 μL of lysate was saved as input. 200 μL of anti-Flag magnetic beads (Sigma; M8823) were washed with 500 μL of ice-cold wash buffer four times (200 mM NaCl, 2 mM EDTA, 0.05% NP-40, 50 mM Tris-HCl, pH 7.6, 0.5 mM DTT, 1:1,000 SUPERase In RNase inhibitor, 1:1,000 protease inhibitor). Lysate was added to the beads and incubated for 2 h on a rotation wheel at 4 °C. Supernatant was removed and saved as flow-through, and beads were washed six times with 1 mL of ice-cold wash buffer. 250 μL of elution solution (500 ng μl−1 3× Flag peptide (Sigma; F4799) in wash buffer) was added to each sample, and the mixture was rotated for 1 h at 4 °C to elute. We made proteinase K solution by adding 10% proteinase K solution (20 mg/mL; Invitrogen; 25530049) to 2× proteinase K buffer (100 mM Tris, pH 7.7, 150 mM NaCl, 12.5 mM EDTA, 2% (w/v) SDS). The eluent, 50 μl of input, and 50 μl of flow-through were each incubated in one volume of proteinase K solution for 45 min at 55 °C to release RNA. Three volumes of TRIzol were added to each sample, and total RNA was extracted on a Direct-zol microprep spin column according to the manufacturer's protocol for RNA extraction and DNase I digestion. rRNA was removed with the RiboMinus Eukaryote Kit v2. 50 ng of input, flow-through, and immunoprecipitated mRNA were analyzed by LC-MS/MS as previously reported[21].

3.4.7 PAR-CLIP-m^{6}A-seq

PAR-CLIP-m^{6}A-seq was performed as previously described[22], with modifications. To begin, PAR-CLIP was performed starting with 600 million HEK293T cells stably expressing Flag-HA-FMRP. FMRP-RNA complex was SDS-PAGE purified with size selection from 130-200 kDa, and RNA fragments were extracted via ethanol precipitation after proteinase K digestion of the gel slices. The purified RNA pellet was dissolved in 12 µl RNase-free water, of which 3 µl were set aside as input.

The remaining 9 µl of the purified RNA was then subjected to m^{6}A seq, which was performed as previously described[23] with modifications. Briefly, the RNA was mixed with 2.5 mg affinity-purified anti-m^{6}A polyclonal antibody (Synaptic Systems) in IP buffer (150 mM NaCl, 0.1% NP-40, and 10mM Tris-HCl, pH 7.4), and incubated for 2 hours at 4°C. The antibody-RNA complex was isolated by incubation with protein A beads (Invitrogen) at 4°C for 2 hours. The beads were washed three times and eluted competitively with an m^{6}A monophosphate solution. RNA in the eluate was isolated using RNA Clean and Concentrator (Zymo Research). Library construction was performed by Truseq small RNA sample preparation kit (Illumina).

3.4.8 *In vitro* probe pulldown

60 million HEK293T cells overexpressing FLAG-tagged FMRP or FMRP-I304N were lysed at 4°C for 30 minutes in 3.5 ml lysis buffer (250 mM NaCl, 0.5% NP40, 10% glycerol, 20mM Tris pH 7.5, 1:100 cOmplete EDTA-free protease inhibitor (Roche), 1:100 SUPERase In RNase Inhibitor (Ambion)), followed by centrifugation at 16,000*g* for 20 minutes. Samples were filtered through a 0.22µm filter. 80 µl of the filtered supernatant was saved as input. The remaining

supernatant was divided evenly and incubated with 1.5 µg of the indicated biotinylated RNA probes**.** The lysate-probe mixture was incubated at 4°C for 2 hours with rotation. 15 µl Dynabeads MyOne Streptavidin C1 beads (Invitrogen) were washed 3x with lysis buffer, and added to each sample. The samples were then incubated at 4°C for 1.5 hours with rotation. The beads were isolated and washed four times with lysis buffer, and protein was eluted with 50 µl 2x Laemmli sample buffer (Bio-rad) at 95°C for 5 minutes. Protein was subjected to FLAG blotting analysis.

References

1. Garber, K., Smith, K. T., Reines, D. & Warren, S. T. Transcription, translation and fragile X syndrome. *Curr. Opin. Genet. Dev.* **16,** 270–275 (2006).
2. Verkerk, A. J. M. H. *et al.* Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65,** 905–914 (1991).
3. Jin, P. *et al.* Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nat. Neurosci.* **7,** 113–117 (2004).
4. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146,** 247–261 (2011).
5. Ascano, M. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492,** 382–386 (2012).
6. De Diego Otero, Y. D. D. *et al.* Transport of fragile X mental retardation protein via granules in neurites of PC12 cells. *Mol. Cell. Biol. Cell. Biol.* **22,** 8332–8341 (2002).
7. Eberhart, D. E., Malter, H. E., Feng, Y. & Warren, S. T. The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export

signals. *Hum. Mol. Genet.* **5,** 1083–1091 (1996).

8. Feng, Y. *et al.* Fragile X mental retardation protein: nucleocytoplasmic shuttling and association with somatodendritic ribosomes. *J. Neurosci.* **17,** 1539–47 (1997).
9. Edupuganti, R. R. *et al.* N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* **24,** 870–878 (2017).
10. Hsu, P. J. *et al.* Ythdc2 is an N6-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Res.* **27,** 1115–1127 (2017).
11. 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).
12. Musco, G. *et al.* The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nature structural biology* **4,** 712–716 (1997).
13. Lewis, H. a *et al.* Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100,** 323–332 (2000).
14. Huang, H. *et al.* Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat. Cell Biol.* **20,** (2018).
15. Suhl, J. A., Chopra, P., Anderson, B. R., Bassell, G. J. & Warren, S. T. Analysis of FMRP mRNA target datasets reveals highly associated mRNAs mediated by G-quadruplex structures formed via clustered WGGA sequences. *Hum. Mol. Genet.* **23,** 5479–5491 (2014).
16. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell* **61,** 507–519 (2016).
17. Roundtree, I. A. *et al.* YTHDC1 Mediates Nuclear Export of N6-methyladenosine Methylated mRNAs. *Elife* **6,** 1–28 (2017).
18. Darnell, J. C., Mostovetsky, O. & Darnell, R. B. FMRP RNA targets: Identification and

validation. *Genes, Brain Behav.* **4,** 341–349 (2005).

19. Joazeiro, C. A. P. Ribosomal Stalling During Translation: Providing Substrates for Ribosome-Associated Protein Quality Control. *Annu. Rev. Cell Dev. Biol.* **33,** annurev-cellbio-111315-125249 (2017).

20. Bakker, C. *et al.* Fmrl Knockout M ice : A Model to Study Fragile X Mental Retardation. *Cell* **78,** 23–33 (1994).

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

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

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

Chapter 4 – Summary and Perspectives[2]

4.1 Toward understanding m^{6}A readers: diverse roles for diverse needs

The field of epitranscriptomics began with the discovery of pseudouridine in 1951, but proceeded placidly over the next six decades, largely due to an inadequacy of methods to detect and characterize modifications in low abundance species such as mRNA. Recent advances in high-throughput sequencing and mass spectrometry have filled this need, causing research on post-transcriptional gene regulation to progress at an accelerated rate. Using a combination of these techniques, we have uncovered several novel roles of m^{6}A, at both the gene regulation and biological levels, which occur through two newly discovered m^{6}A readers.

Up until now, the understanding about effects of m^{6}A at the gene regulation level has predominantly focused on individual roles of reader proteins. Reader proteins in the cytoplasm are known to have one main role; as examples, YTHDF1 affects RNA translation[1], YTHDF2 decreases RNA stability[2], YTHDF3 assists YTHDF1 in promoting translation[3,4], and the IGF2BP proteins increase RNA stability[5]. Our results show, for the first time, that a cytoplasmic reader protein can have multiple roles in RNA processing, as YTHDC2 affects both RNA abundance and translation of transcripts containing m^{6}A. This result joins the finding that the nuclear m^{6}A reader YTHDC1 plays roles in both splicing and export[6,7], demonstrating that reader proteins in the cytoplasm may also have multiple roles. Moreover, readers have not been shown to have direct, distinct roles in both the cytoplasm and nucleus. Although YTHDF2 can localize to the nucleus

[2]Parts of section 4.3 are adapted from Hsu, P.J., Shi, H., He C. Epitranscriptomic influences on development and disease. *Genome Biology* **18**(1):197 (2017) with modifications.

during the heat shock response to inhibit demethylation by FTO[8], its effect on RNA translation is indirect. Here, we demonstrate that FMRP, a protein that affects RNA translation and localization in the cytoplasm, also plays a nuclear role in exporting m^6A-containing RNA transcripts. Thus, our understanding of the functions of reader proteins appears to require much more investigation, as more roles are possible under different contexts.

Important questions then arise – why do individual readers have multiple roles, and under what circumstances are these roles carried out? It is possible that readers' targets need to undergo all processes carried out by one reader, and thus using one reader to fulfill all of the necessary roles would be advantageous to the cell. For example, a transcript spliced by YTHDC1 may then also need to be quickly exported out of the nucleus, and would therefore benefit from the existing interaction with YTHDC1 needed for splicing. However, readers may have different targets in different parts of the cell, and thus may also perform distinct roles on distinct targets. It has been shown that FMRP lacking an RGG motif can still associate with polyribosomes[9]; however, FMRP requires its RGG motif to interact with RNA[10]. Thus, the roles of FMRP in translation and RNA export are likely performed on different targets. The question of why a reader would have evolved to perform two or more distinct roles will need to be further studied.

The distinct roles carried out by readers with multiple roles may also depend on the sequence context and methylation status of a transcript is bound by the reader. Both YTHDC2 and FMRP have multiple conserved RNA binding domains, and thus may use different domains to bind transcripts with distinct motifs. Notably, YTHDC2 and FMRP have much less selectivity for m^6A than YTHDF1-3 and YTHDC1, likely caused by RNA binding by specific protein domains that do not have m^6A selectivity. It is possible that only the YTH domain in YTHDC2 has selectivity for m^6A, while other domains bind RNA indiscriminately. Similarly, the KH domain in

FMRP is important for m^6A binding, but there has been no evidence of m^6A selectivity in the RGG domain. It is therefore possible that sequence context and methylation status affect roles played by a given binding protein.

The roles of reader proteins may also depend on cell- or tissue-specific contexts. Although the depletion of Ythdc2 and Fmrp from our mouse models occurred in all tissues, only certain tissues exhibited noticeable defects. Mice lacking Ythdc2 demonstrated infertility, and mice lacking Fmrp exhibited mental irregularities. Many other tissues, however, had normal gross physical appearance and function, including ones in which the proteins are highly expressed. For example, Fmrp is highly expressed in the lung, and mice depleted of Fmrp exhibit no noticeable lung defects. Some tissues may not require the specific functions of these reader proteins that are essential in tissues such as testes for Ythdc2 and brain for Fmrp. Instead, the high expression of these reader proteins in other tissues may reflect other, yet unknown roles specific to each respective tissue. We have also noted differences firsthand; PAR-CLIP of YTHDC2 in HeLa cells did not enrich for a GGACU motif on HeLa mRNA (data not shown), while CLIP-seq of Ythdc2 in mouse testes strongly enriched for a GGACU motif (**Figure 2.3**). It appears that the targets of YTHDC2 in HeLa cells vary greatly from its targets in testes, including in specificity to m^6A. As both YTHDC2 and FMRP may bind unmethylated transcripts, it is possible that reader proteins may even have non-m^6A-related functions that are specific to certain cell types or tissues. The cell- or tissue-specific roles of reader proteins could also depend on differences in partner proteins with which the readers interact, conferring greater tissue specificity to the readers.

Yet another factor affecting the role of reader proteins may be stress contexts within cells. Several studies have already demonstrated this principle. Heat shock causes YTHDF2 to localize to the nucleus and protect transcripts from demethylation by FTO[8], and infection by Hepatitis C

virus causes YTHDF proteins to localize to lipid droplets, where HCV viral assembly occurs[11]. The RNA of other viruses, including Kaposi's sarcoma-associated herpesvirus (KSHV), Zika virus (ZIKV), and HIV-1, are bound by YTHDF1-3[12–14], despite not being endogenous RNAs. These stresses may cause reader proteins to perform functions that they may not normally perform under baseline conditions.

Finally, cells that are differentiating may rely on specific functions of reader proteins as they undergo "transcriptome switching," a concept we recently proposed[15]. To cause the waves of protein expression that occur during differentiation, cellular machinery must tightly regulate gene expression at the RNA level, a role appropriate for m^6A readers. Notably, both YTHDC2 and FMRP have functions in differentiating cells, as Ythdc2 functions in spermatocytes undergoing meiosis, and lack of FMRP causes defects on the still-developing brain.

As the field of epitranscriptomics continues to evolve, it will be important to answer to answer the questions about the multiple roles of reader proteins and the circumstances under which they are required. Doing so will allow greater characterization of how functions of m^6A are carried out, and will also provide clues to the broader question of how RNA binding proteins achieve their selectivity on transcripts.

4.2 A basis behind selectivity

Research on readers of m^6A have been largely devoted to reader proteins with a strong preference for methylated transcripts over unmethylated transcripts. Our work has shown that, unlike these readers, YTHDC2 and FMRP prefer methylated transcripts, but may also bind unmethylated transcripts, thus making them "weaker" readers. This preference may arise due to the fact that, unlike YTHDF1-3 and YTHDC1, which bind m^6A strongly, YTHDC2 and FMRP

have multiple RNA binding domains, each of which may have a unique specificity for methylated or unmethylated transcripts. It will be important to consider, and eventually determine, what domains provide the structural basis behind preference for m^6A.

Although both YTHDC2 and FMRP have multiple RNA binding domains, it is possible that proper RNA binding requires all of the domains to interact together for proper m^6A binding. Our preliminary studies on YTHDC2 using only a recombinantly expressed YTH domain did not show a noticeable preference for m^6A *in vitro* (data not shown). The other domains of YTHDC2 may affect the ability of the YTH domain to bind m^6A properly, whether by affecting the conformation of the YTH domain or affecting the positioning of the RNA in the YTH binding pocket. It is also possible that the other domains directly bind m^6A themselves.

Similarly, FMRP lacking an RGG domain fails to bind RNA[10], while the KH2 domain is involved in m^6A binding (**Figure 3.3B**). It appears that the interaction between all of the RNA binding domains of FMRP is necessary for proper transcript selectivity. It is possible that the domains each have an additive effect in m^6A selectivity. In particular, the RGG motif could play an important role. Although it was once considered a nonspecific RNA binding domain, studies have shown that it binds RNAs with an intramolecular G-quartet structure[16,17]. It may also have some m^6A selectivity of its own, contributing along with the KH domains to the overall preference for m^6A by FMRP. This additive effect appears to be the case for the IGF2BP proteins, which each have four KH domains. In the IGF2BP proteins, the KH3-4 di-domain is necessary for m^6A recognition, while the KH1-2 di-domain makes the preference for m^6A stronger[5]. Thus, there is likely an interplay between the domains of the "weaker" m^6A binding proteins that allows for their binding selectivity.

Another important consideration regarding reader proteins is that certain proteins may bind more than one chemical modification. Along the spectrum of chemical modifications, certain modifications may contain similarities that allow reader proteins to bind multiple modifications. An early hint to this hypothesis is that YTHDF1-3 and YTHDC1 were shown to bind m^1A[18]. However, the conversion of m^1A to m^6A by the Dimroth rearrangement may be the cause for this specificity. As research progresses and our technologies for studying modifications become more sophisticated, it will become more possible to piece together interactions between different modifications or their machinery. Studying the interplay between modifications will provide greater insight into the complex network of interactions that together allow the precise regulation of gene expression. In addition, as more modifications and their functions are characterized, we will gain a greater understanding of cellular processes requiring modifications, as well as potential avenues to disrupt those processes with the goal of treating diseases.

4.3 Looking forward: Concluding remarks and future directions

It is becoming increasingly clear that the epitranscriptome and its modifying enzymes form a complex constellation that holds widely diverse functions. Post-transcriptional RNA modifications allow additional controls of gene expression, serving as powerful mechanisms to affect eventual protein synthesis. In particular, m^6A provides layers of regulation, offering effects dependent on the localization of its writers, readers, and erasers.

To facilitate certain cellular processes, the m^6A machinery can target multiple substrate mRNAs and non-coding RNAs. As we proposed[15], cellular programs may require a burst of expression of a distinct set of transcripts, followed by expression of a different set of transcripts. m^6A can mark and cause timely expression and turnover of subsets of transcripts. The cellular and

compartmental localizations of the writers, readers, and erasers critically affect their functions. Methylation, together with demethylation of subsets of transcripts in the nucleus, may create a methylation landscape that directs the fate of groups of transcripts as they are processed, exported to the cytoplasm, translated, and degraded. Multiple different readers or their associated proteins may be required to fully actualize the effects of the methylations. While transcript turnover or decay is an accepted role of mRNA m^6A methylation, it should be noted that the *Ythdf2* knockout mouse exhibits a less severe phenotype[19] compared to mice lacking *Mettl3* or *Mettl14* (embryonic lethality), demonstrating that the Ythdf2-dependent pathway constitutes a subset of the functions of methylated transcripts. There are other critical regulatory functions of m^6A RNA methylation that remain to be uncovered.

These observations lead us to perceive that methylation occurs at multiple layers. Methyltransferases set the initial methylation landscape in coordination with the transcription machinery. Demethylases could more efficiently tune the methylation landscape of a subset of methylated transcripts, acting as the second layer of regulation. Indeed, demethylases often target only a subset of genes under certain conditions, as depletion of *Alkbh5* does not lead to embryonic lethality, instead causing defects in spermatogenesis[20], while only a portion of *Fto* knockout mice display embryonic lethality. Finally, reader proteins act as effectors in a third layer of regulation, carrying out specific functions upon methylated transcripts.

The field of epitranscriptomics still remains vastly unexplored. As suggested by the results in this thesis, it will be important to understand the spectrum of roles played by reader proteins and the mechanisms underlying their preference for certain transcripts. Future studies will also need to focus on the mechanisms defining which transcripts receive methylations. Moreover, as methylations are often unevenly distributed along the RNA transcript, identifying the mechanisms

underlying regional specificity of methylation, as well as which individual sites along transcripts are methylated, remain as major challenges. The methylation selectivity on particular transcripts may need to be coupled with transcription regulation. How this selectivity is determined and the interplay between methylation and transcription require further exploration. Questions regarding the effects of methyltransferases and demethylases on nuclear processing, splicing, and export also remain. Nuclear regulation of RNA methylation could play critical roles impacting biological outcomes. In particular, it will be important to determine how and why a subset of RNAs undergoes demethylation inside the nucleus, as well as the functional consequences of this required demethylation on gene expression. Interactions between the writers, readers, and erasers with other cellular components are also necessary to reveal functional roles, especially those in complex biological processes *in vivo*.

The results of this thesis point toward specifically toward four directions of continued work, three of which are direct next steps, and the fourth of which applies lessons learned from this thesis toward a related effort. First, it would be important to determine the effects of YTHDC2 and FMRP on methylated transcripts in settings outside of what was studied in this thesis. Ythdc2 was shown to be necessary during meiosis prophase I of spermatogenesis, as its depletion caused apoptosis of spermatocytes during this stage (**Figure 2.7**). However, whether it plays roles later on in spermatogenesis, as well as whether these roles involve m^6A-methylated transcripts, is unknown. Induced depletion of YTHDC2 at later stages in spermatogenesis, as well as CLIP-seq of YTHDC2 at later stages, may provide answers to its roles in these settings. FMRP has known roles in affecting translation and RNA localization. Whether these roles also affect methylated transcripts in particular would be an important question. The combination of m^6A-seq, ribo-seq, proteomics,

and methyltransferase depletion experiments would begin to provide clues about the interplay between FMRP and m^{6}A within its other known roles.

Second, these findings encourage deeper investigation into the basis underlying m^{6}A selectivity of YTHDC2 and FMPR, as well as their "permissibility" of unmethylated transcript targets. The YTH domain of YTHDC2 is highly evolutionarily conserved; thus, mutation of residues previously shown to be necessary for m^{6}A binding would likely abolish all m^{6}A preference. However, given its weaker affinity for m^{6}A compared to the other YTH proteins, as well as its several other RNA binding domains, it is possible that other domains may play a role. Studies using recombinant protein constructs with mutations in YTH and other domains would shed light on the mechanism of m^{6}A selectivity. Similarly, the KH domains of FMRP form a hydrophobic pocket that could be the mechanism underlying m^{6}A selectivity. Mutation of hydrophobic residues that bind near the N^6 position of adenosines, such as L255 and I318, to hydrophilic ones would determine whether the hydrophobicity of the KH domain binding pocket confers specificity to m^{6}A.

Third, investigating the proteins recruited by YTHDC2 and FMRP would prove insightful. Several other reader proteins are known to interact with partner proteins that carry out the effects of the reader protein on methylated transcripts. YTHDF1 and YTHDF3, which work together to promote mRNA translation, interact with proteins involved in translation initiation and translational control such as YBX1, G3BP1, and eIF3A[1,3,4]. Similarly, YTHDF2, which decreases mRNA stability, performs its role by recruiting the CCR4-NOT deadenylation complex, which causes degradation of the m^{6}A-containing transcripts[21]. YTHDC1 also interacts with several protein partners for its various roles; its role in export depends on interactions with nuclear export proteins SRSF3 and NXF1; its interaction with SRSF3 also confers its effect on splicing; and its

interactions with SRSF3 and SRSF7 are thought to affect 3' end shortening, a role it plays in oocytes[6,7,22]. The experiments in this thesis showed that YTHDC2 interacts with many partner proteins involved in translation initiation and nonsense-mediated decay. Investigating the dependence of YTHDC2 on these partner proteins through tethering reporter assay and knockdown experiments would show which proteins are most crucial for the effect of YTHDC2. Determining the partner proteins with which FMRP interacts to facilitate export of methylated mRNAs would also prove insightful. Previous studies have shown that FMRP interacts with the nuclear export factors NXF1 and NXF2[23–25]; knockdown experiments would show whether they affect the export of methylated mRNA, while mass spectrometry experiments would identify other nuclear proteins interacting with FMRP. Interestingly, it was recently found that FMRP interacts with YTHDF2 in an RNA-independent manner to increase the stability of its target mRNAs[26], bringing to question whether readers may either recruit or compete against each other in different cellular contexts. Our preliminary results have shown that FMRP may also interact with YTHDF1 and YTHDC1 (data not shown). The aforementioned mass spectrometry experiments, in combination with m^6A methyltransferase depletion experiments, would answer questions about the recruitment of other m^6A readers.

Finally, the lessons learned from these studies of two m^6A readers, each with multiple roles, may be applied to identify and characterize readers of other modifications. It is possible that RNA binding proteins with known roles also bind specifically to post-transcriptional modifications other than m^6A, and that their binding to these modified transcripts may affect gene regulation in ways that have not yet been identified. Of note, recent studies suggest that the modifications m^1A and 2'OMe on mRNA affect translation[27,28]; however, readers of m^1A and 2'OMe have not yet been identified. Our preliminary results have identified several potential 2'OMe readers that have

known roles in splicing (data not shown), which may lead to identification of roles of 2'OMe in splicing or may show other effects of these proteins when they bind to 2'OMe. Characterizing readers of these modifications would be a significant effort that would require substantial time and resources. However, methods necessary for studying readers, such as affinity chromatography, PAR-CLIP, CLIP-QQQ, and reporter assays, have been established and enhanced by studies such as ours, and could thus be applied to characterize new readers for other modifications.

The current findings in the burgeoning field of epitranscriptomics point toward the great breadth of processes affected by post-transcriptional RNA modifications. Continued investigation in this field will surely deepen our current understanding of fundamental aspects of cellular biology and physiology, and will also create new fields and opportunities for the development of new research methodologies and technologies. Moreover, as this research progresses, our understanding of how modifications allow or disrupt cellular processes could allow us to uncover new therapeutic avenues, given the number of diseases shown to involve RNA modifications. Epitranscriptomics is still in its early stages. However, the studies thus far deliver a clear picture of the importance it will hold, both scientifically and translationally, as it continues to grow.

References

1. Wang, X. *et al.* N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161,** 1388–1399 (2015).
2. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–20 (2014).
3. Shi, H. *et al.* YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Res.* **27,** 315–328 (2017).

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

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

6. Roundtree, I. A. *et al.* YTHDC1 Mediates Nuclear Export of N6-methyladenosine Methylated mRNAs. *Elife* **6,** 1–28 (2017).

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

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

9. Darnell, J. C., Mostovetsky, O. & Darnell, R. B. FMRP RNA targets: Identification and validation. *Genes, Brain Behav.* **4,** 341–349 (2005).

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

11. Gokhale, N. S. *et al.* N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection. *Cell Host Microbe* 1–12 (2016). doi:10.1016/j.chom.2016.09.015

12. Ye, F., Chen, E. R. & Nilsen, T. W. Kaposi's Sarcoma-Associated Herpesvirus Utilizes and Manipulates RNA N ^{6} -Adenosine Methylation to Promote Lytic Replication. *J. Virol.* JVI.00466-17 (2017). doi:10.1128/JVI.00466-17

13. Lichinchi, G. *et al.* Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. *Cell Host Microbe* 1–8 (2016). doi:10.1016/j.chom.2016.10.002

14. Tirumuru, N. *et al.* N6-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression. *Elife* 1–19 (2016). doi:http://dx.doi.org/10.7554/eLife.15528

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

16. Mandel, J., Ehresmann, B. & Ehresmann, C. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif speci ® cally to its mRNA via a purine quartet motif. *EMBO J.* **4813,** 4803–4813 (2001).

17. Darnell, J. C. *et al.* Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107,** 489–499 (2001).

18. 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).

19. 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 (2017).

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

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

22. Kasowitz, S. D. *et al.* Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLOS Genet.* **14,** e1007412 (2018).

23. Zhang, M., Wang, Q. & Huang, Y. Fragile X mental retardation protein FMRP and the RNA export factor NXF2 associate with and destabilize Nxf1 mRNA in neuronal cells. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 10057–62 (2007).

24. Lai, D., Sakkas, D. & Huang, Y. The fragile X mental retardation protein interacts with a

distinct mRNA nuclear export factor NXF2. *Rna* **12,** 1446–1449 (2006).

25. Kim, M., Bellini, M. & Ceman, S. Fragile X Mental Retardation Protein FMRP Binds mRNAs in the Nucleus. *Mol. Cell. Biol.* **29,** 214–228 (2009).

26. Zhang, F. *et al.* Fragile X mental retardation protein modulates the stability of its m6A-marked messenger RNA targets. *Hum. Mol. Genet.* Online August 2018 (2018). doi:10.1093/HMG/DDY292

27. Dominissini, D. *et al.* The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* (2016). doi:10.1038/nature16998

28. Choi, J. *et al.* 2′-O-methylation in mRNA disrupts tRNA decoding during translation elongation. *Nat. Struct. Mol. Biol.* **25,** (2018).