THE UNIVERSITY OF CHICAGO

POST-TRANSCRIPTIONAL GENE REGULATION BY N^6-METHYLADENOSINE AND SELECTIVE RNA BINDING PROTEINS

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:
BIOCHEMISTRY AND MOLECULAR BIOLOGY

BY

IAN A. ROUNDTREE

CHICAGO, ILLINOIS

DECEMBER 2017

Table of Contents

Acknowledgement

I would like to begin by thanking Prof. Chuan He for giving me the opportunity to join his lab as a graduate student in 2013. I came to the University of Chicago with very little knowledge or experience in biology, and his lab offered me a chance to grow and learn as an individual and as a scientist. Throughout my time in the lab I have found his guidance to be relentlessly encouraging and optimistic. It is no surprise that the members of the lab enjoy his leadership, and are truly appreciative of his hard work.

I would also like to thank the members of the He lab with whom I have worked closely in the past few years. The group is composed of thoughtful, hard-working, and intelligent scientists. More importantly, the members of the lab are helpful and caring people, which was a feature of the lab well before I joined and will remain one long after I have left. I would specifically like to acknowledge Dr. Xiao Wang, with whom I worked closely with during my first year and learned the skills that would be the foundation of my graduate work. I would also like to thank Dr. Guanzheng Luo, who generously offered his time and help in bioinformatics. I would like to thank Dr. Qili Fei for helping me with numerous experiments and bioinformatics questions. I would also like to thank Dr. Kai Chen and Zijie Zhang for their helpful discussions and companionship. Without the help of the entire group, my training in the lab would have been impossible, and I am thankful to have made so many friends in my time here.

Thank you to the members of my thesis committee, Prof. Tao Pan, Prof. Joe Piccirilli, and Prof. Alex Ruthenburg, who took the time to meet with and advise me throughout the course of my graduate student career.

Thank you to the University of Chicago MSTP for the opportunity to take part in the program, and for the freedom to pursue my research interests in pursuit of my graduate training. I am grateful for the Frank Family for their generous donation to our program and for their continued interest in the training of physician scientists. Thank you to the NIH for pre-doctoral funding.

I would like to thank my MSTP classmates for their friendship and comradery throughout the past 5 years. I am excited to see what the future brings for all of your pursuits, both academic and otherwise.

I am thankful for Sylvia, who has been incomparable in her patience and encouragement in our time together. I am lucky to be by your side as we continue our future together.

Lastly, I would like to thank my family for their support. Since I first began research during my undergraduate years, they have been extremely supportive of my interests while setting a stellar example of pursuing higher education.

Abstract

Regulation of genetic output is a complex biological phenomenon that allows cells with identical genetic material to enable diverse phenotypic outcomes. Epigenetics is the study of factors that change gene expression without perturbing the underlying genetic sequence. Mechanisms of epigenetics feature a multitude of chemical modifications to DNA as well as accompanying histone proteins, which tune the transcriptional output in a reversible manner. Recently, the concept of reversible RNA epigenetics has emerged as a mechanism to regulate gene expression post-transcriptionally. N^6-methylasdenosein (m^6A) is the most common internal modification in eukaryotic messenger RNA (mRNA). This RNA modification is recognized by m^6A-selective RNA binding protein of the YTH family, which incorporate modified RNAs into canonical pathways for mRNA metabolism. In this work, we describe the function of the only nuclear YTH protein YTHDC1 in promoting the nuclear export of methylated mRNAs by interaction with the splicing and export adaptor protein SRSF3. We then turn to a biological example of rapid mRNA processing during progression through the cell cycle. We propose that m^6A modifications serve roles in the dynamic regulation of the cellular transcriptome.

List of publications based on work described in this thesis

1. Roundtree, I.A., Luo, G. Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., He, E., Shen, B., He, C. N^6-methyladenosine and YTHDC1 promote Nuclear Messenger RNA Export. *Submitted.*

2. Roundtree, I.A., Evans, M.E., Pan, T., He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**, 1187-1200 (2017) [review].

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

4. Roundtree, I.A., He, C. RNA epigenetics – chemical messages for posttranscriptional gene regulation. *Curr. Opin. Chem. Biol.* **30**, 46-51 (2016) [review].

5. Roundtree, I.A., He, C. Nuclear m^6A Reader YTHDC1 Regulated mRNA Splicing. *Trends in Genetics* **32**, 320-321 [spotlight].

6. Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., He, C. N^6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **163**, 1388-1399 (2015).

7. Zhu, T., Roundtree, I. A., Wang, P., Wang, X., Wang, L, Sun, C., Tian, Y., Li, J., He, C., Xu, Y., Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N^6-methyladenosine. *Cell Research* **24**, 1493-1496 (2014).

8. Xu, C., Wang, X., Liu, K., Roundtree, I.A., Tempel, W., Li, Y., Lu, Z., He, C., Min, J., Structural basis for selective binding of m^6A RNA by the YTHDC1 YTH domain. *Nat Chem. Biol.* **10**, 927-929 (2014).

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendix 1

Chapter 1 – Introduction

1.1 Mechanistic Concepts in Gene Regulation

The central dogma of molecular biology describes the flow of genetic information from deoxyribonucleic acid (DNA) to ribonucleic acid (RNA) to protein - the ultimate determinant of cellular function (**Figure 1.1**). Decades of biological research have uncovered mechanisms that regulate each step of the central dogma, allowing organisms to achieve complexity beyond our current understanding.

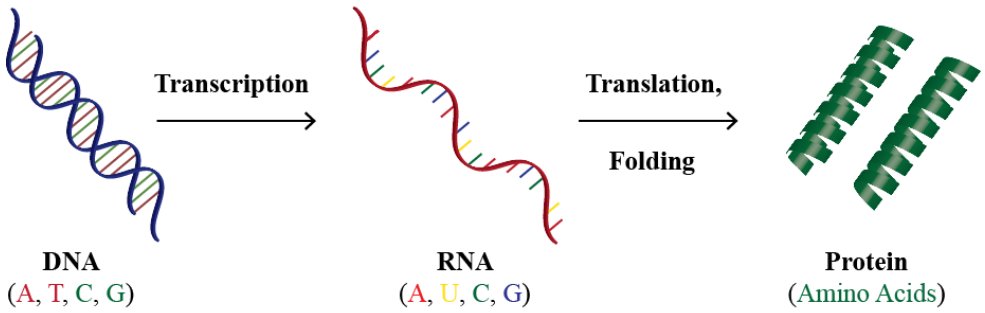

Figure 1.1 The central dogma of molecular biology

DNA, composed of bases A, T, C, and G is transcribed into coding RNA in which T is replaced with U. Eventually this RNA is translated into functional polypeptides.

Transcriptional control of gene expression relies on a combination of DNA structural conformation and DNA binding proteins to determine the extent to which DNA sequences are transcribed into their messages. By controlling the copy number of any given message, transcriptional control enacts widespread programs of genetic information. Some of the best examples of transcriptional control depend on adjacent sequences contorting DNA structure to allow for transcription factor binding and RNA production from downstream loci.

Just as elements for transcriptional control are encoded within the primary sequence of DNA, mechanisms for post-transcriptional regulation can be embedded within a given RNA molecule. RNA structural motifs can both recruit and repel RNA-binding proteins that ultimately

determine transcript fate, as can sequence-specific effector proteins with affinity for RNA.

Finally, coding transcripts are translated into their protein products, which too have their function encoded largely in their primary sequence. Common to each step in this information transfer is the opportunity for intervention. Cells manipulate each step in order to tune their gene expression profiles in response to developmental and environmental queues.

However, the methods for gene regulation outlined above cannot account for the complexity observed in living systems. Indeed, additional mechanisms exist to ensure the timely and accurate expression of genetic information which lay on top of traditional genetics, thusly termed “epigenetics”.

1.2 Reversible Chemical Modifications of DNA and Proteins - Epigenetics

The study of epigenetics (literally, “on top of” or “above” genetics) addresses the potential for cells housing identical genetic material to experience divergent developmental outcomes. Early studies by CH Waddington concerned how genetically identical cells interact with their environment to determine their “epigenotype” and thus their differentiation pathway[1]. More recently, the field has concerned “a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence"[2]. The most common mechanisms for epigenetic manipulation of gene expression include DNA methylation[3], histone modification[4], and silencing by long non-coding RNAs, each of which modify the transcriptional output of their respective regulons.

Epigenetic regulation by definition does not change the underlying genetic code, but instead applies additional layers of information. Most notably, this occurs as a reversible chemical modification of genetic material, such as DNA eukaryotic DNA methylation to form 5-methylcytosine (5mC). Although typically considered a silencing mark in promoter and enhancer regions, the function of m5C is variable and dependent on localization with respect to gene bodies[5]. Similarly, the modification of histone proteins can be diverse in location as well as structure [4]. Again, methylation exerts a regulatory function, this time via modification of the N-terminal lysine residues on the histone tail (**Figure 1.2**).

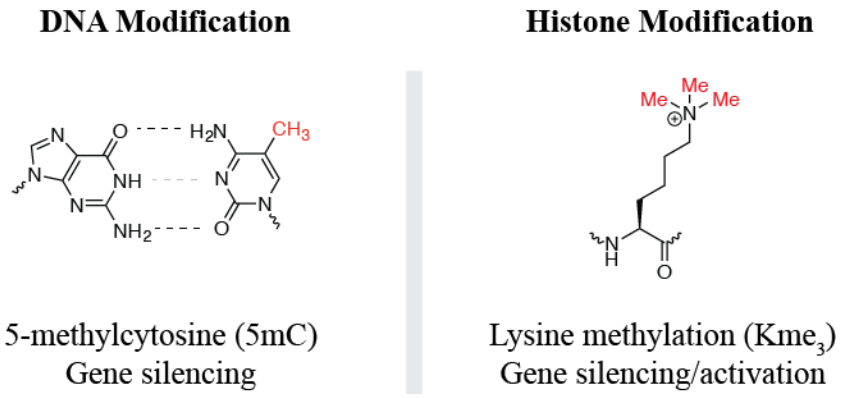

Figure 1.2 Chemical modifications of DNA and histone proteins

DNA modifications such as 5-methylcytosine (left) are commonly found at promoters and enhancer of silenced genes. Lysine trimethylation (right) is one of several potential lysine modifications found on histone tails.

Unlike byproducts of cellular metabolism or oxidative damage[6], epigenetic marks on DNA and proteins are installed and removed by enzymatic cellular machinery. This potential for dynamic modification at the hands of methyltransferases, demethylases grants the eukaryotic cell access to increased chemical space in which regulate gene expression. The diversity of additional chemical moieties not mentioned here extend the potential for DNA and histone modifications to determine genetic output.

1.3 N^6-methyladenosine modification of eukaryotic RNA

Eukaryotic RNAs bear considerable amounts of chemical modifications, currently represented by over 150 distinct structures[7]. Abundant cellular RNAs such as ribosomal RNAs (rRNA) and transfer RNA (tRNA) are frequently modified in order to confer resistance to degradation and maintain structure[8,9]. Messenger RNA (mRNA) is modified during splicing events, which notably change the primary sequence of the message, and at the 5' and 3' ends of the molecule by cap and poly(A) structures, respectively. The 5' cap consists of an unusual 5'-5' linkage which protects transcripts from exonuclease activity, and is recognized specifically by the cap-binding-complex to initiate translation initiation[10,11]. The poly(A) tail is bound by poly(A)

binding proteins, which regulate translation and decay of mRNAs, as well as additional roles in both the nucleus and the cytoplasm[12].

mRNA is also decorated by a plethora of internal modifications, the most common of which is methylation at the N-6 position of adenosine, N^6-methyladenosine (m^6A) (**Figure 1.3**).

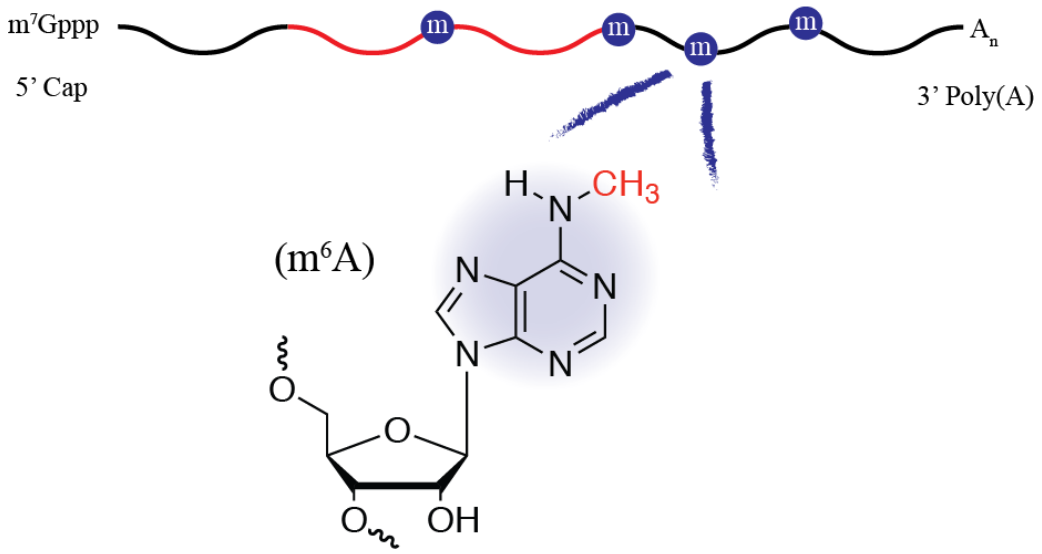

Figure 1.3 Common modifications of mammalian messenger RNA

Mammalian mRNA is modified with a 5' cap structure and 3' poly(A) tail. Internally, m^6A is the most common chemical modification.

Discovered along with the 5' cap, studies into the biological function of internal methylations have remained elusive largely due to their low abundance [13]. Early estimates of m^6A predict that each mRNA contains between three and five such modified bases per transcript, the majority of which occur within a G(m^6A)C (70%) or A(m^6A)C (30%) motif[14,15]. The degeneracy of this motif throughout the transcriptome indicates that the vast majority of GAC and AAC sequences lack detectable methylation, while sites of modification vary significantly in modification fraction[16,17]. Much of this heterogeneity can likely be attributed to the methyltransferase complex which catalyzes the transfer of a methyl group from S-adenosylmethionine (SAM) to the consensus m^6A sequence[18,19].

Investigations into the biological role of m^6A was revitalized by two recent discoveries in RNA modification biology. The first was the discovery of the RNA demethylase activity of the fat mass and obesity-associated protein (FTO), which catalyzes the removal of m^6A in nuclear

transcripts[20]. This study, along with the characterization of a second demethylase Alkbh5, helped to establish the paradigm of reversible RNA methylation as a post-transcriptional regulatory element[21,22] (**Figure 1.4**).

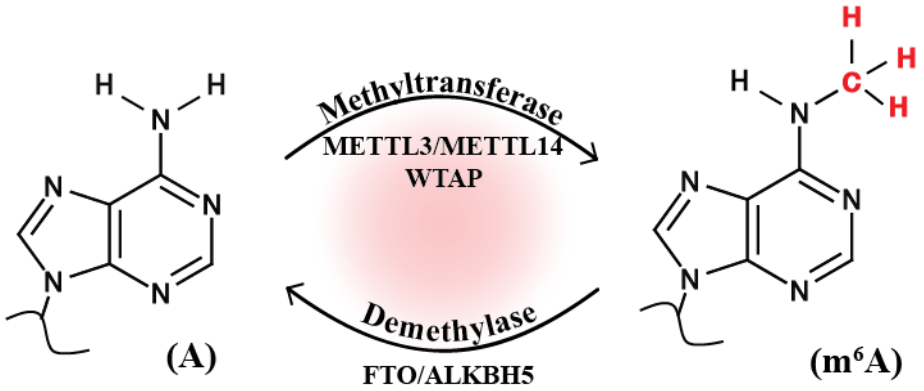

Figure 1.4 Reversible m^{6}A methylation of RNA

The catalytic core of the m^{6}A methyltransferase consists of METTL3, METTL14, and WTAP. Demethylases FTO and ALKBH5 actively remove m^{6}A methylation in RNA.

A second advance in m^{6}A biology came with the rise of high-throughput sequencing technology. Using modification-specific antibody capture of methylated RNA, researchers mapped m^{6}A in the mammalian transcriptome for the first time, revealing signatures of methylation that are highly suggestive of a conserved feature in mRNA[23,24].

1.4 Effects of m^{6}A on mRNA metabolism[1]

Selective RNA metabolism allows cells to functionally group transcriptome elements for selective regulation and metabolism. Recent studies have shown that mRNA methylation can serve as a post-transcriptional regulatory feature by facilitating the recruitment of modified transcripts into canonical pathways. Prominent examples of such discrimination include m^{6}A-dependent translation initiation[25–29] and regulation of RNA stability[30–32]. Methylation has also been implicated in the processing of pre-mRNA within the nucleus, regulating alternative splicing and

[1] See Appendix 1 for a review of RNA modifications in regulating gene expression.

polyadenylation pattens[33–37].

The majority of these functions can be attributed to the presence of m^{6}A-specific protein-RNA interactions. In particular, the YTH Domain Family (YTHDF) proteins have been identified as a selective class of so called 'reader' proteins, and shown to facilitate several properties associated with m^{6}A methylated mRNAs in mammalian cells[23,30]. The YTH Domain Containing (YTHDC) proteins similarly share a conserved YTH domain, and likely mediate properties of methylated mRNAs as well (**Figure 1.5**).

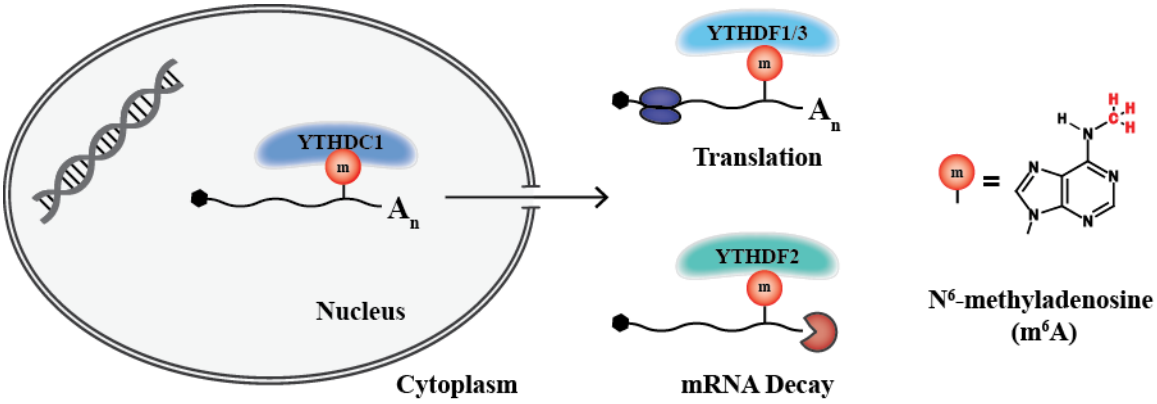

Figure 1.5 Selective mRNA metabolism by YTH proteins

Cytoplasmic proteins of the YTH Domain Family mediate selective translation initiation and mRNA decay. The nuclear protein YTH Domain Containing 1 regulates the fate of nuclear mRNA.

Work contained herein concerns the role for the nuclear protein YTHDC1 in the post-transcriptional regulation of mRNAs containing *N*^{6}-methyladenosine.

1.5 Scope of this thesis

In this thesis, I am to decipher roles for m^{6}A in post-transcriptional regulation of gene expression through functional analysis of m^{6}A-specific 'reader' proteins.

Chapter 2 presents the biochemical, structural, and cellular characterization of the nuclear protein YTHDC1 as a selective binding protein for m^{6}A in RNA.

Chapter 3 presents an investigation into novel roles for selective processing and export of

methylated mRNAs within the nucleus

Chapter 4 presents the potential for m^6A to facilitate the rapid turnover of the cellular transcriptome during progression through the cell cycle

Chapter 5 summarizes recent advances in m^6A biology. This chapter will discuss broad implication of reversible RNA modification as a regulatory element, and present concepts that may shape future research in the field.

References

---

1. Waddington, C. H. The epigenotype. *Endeavour* **1,** 18–20 (1942).
2. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics An operational definition of epigenetics. *Genes Dev.* **23,** 781–783 (2009).
3. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–220 (2013).
4. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–395 (2011).
5. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–492 (2012).
6. Ciccia, A. & Elledge, S. J. The DNA Damage Response: Making It Safe to Play with Knives. *Mol. Cell* **40,** 179–204 (2010).
7. Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways — 2013 update. *Nucleic Acids Res.* **41,** 262–267 (2013).
8. Sharma, S. & Lafontaine, D. L. J. 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. *Trends Biochem. Sci.* **40,** 560–575 (2015).
9. Yacoubi, B. El, Bailly, M. & de Crecy-Lagard, V. Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. *Annu. Rev. Genet.* 69–95 (2012).
10. Cowling, V. H. Regulation of mRNA cap methylation. *Biochem J* **302,** 295–302 (2010).
11. Topisirovic, I., Svitkin, Y. A., Sonenberg, N. & Shatkin, A. J. Cap and cap-binding

proteins in the control of gene expression. *WIREs RNA* 277–298 (2011).

12. Ku, U. & Wahle, E. Structure and function of poly(A) binding proteins. *Biochim. Biophys. Acta* **1678,** 67–84 (2004).
13. Desrosiers, R. C., Friderici, H. & Rottman, F. M. Characterization of Novikoff Hepatoma mRNA Methylation and Heterogeneity in the Methylated 5' Terminus. *Biochemistry* **14,** 4367–4374 (1975).
14. Wei, C. M., Gershowitz, A. & Moss, B. 5'-Terminal and Internal Methylated Nucleotide Sequences in HeLa Cell mRNA. *Biochemistry* **15,** 397–401 (1976).
15. Wei, C. & Moss, B. Nucleotide Sequences at the N6-Methyladenosine Sites of HeLa Cell Messenger Ribonucleic Acid. *Biochemistry* **16,** 1672–1676 (1977).
16. Kane, S. E. & Beemon, K. Precise Localization of m6A in Rous Sarcoma Virus RNA Reveals Clustering of Methylation Sites: Implications for RNA Processing. *Mol. Cell. Biol.* **5,** 2298–2306 (1985).
17. Liu, N. *et al.* Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* 1848–1856 (2013). doi:10.1261/rna.041178.113.4
18. Bokar, J. A., Rath-Shambaugh, M. E., Ludwiczak, R., Narayann, P. & Rottman, F. Characterization and Partial Purification of mRNA N6-Adenosine Methyltransferase from HeLa Cell Nuclei. *J. Biol. Chem.* **269,** 17697–17704 (1994).
19. Liu, J. *et al.* A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10,** 93–95 (2014).
20. Jia, G. *et al.* N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7,** 6–9 (2011).
21. Zheng, G. *et al.* ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* 18–29 (2012). doi:10.1016/j.molcel.2012.10.015
22. He, C. RNA epigenetics? *Nat. Chem. Biol.* **6,** 863–865 (2010).
23. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–206 (2012).
24. Meyer, K. D. *et al.* Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. *Cell* **149,** 1635–1646 (2012).

25. Wang, X. *et al.* N6-methyladenosine Modulated Messenger RNA Translation Efficiency. *Cell* **161,** 1388–1399 (2015).
26. Meyer, K. D. *et al.* 5' UTR m6A Promotes Cap-Independent Translation. *Cell* 999–1010 (2015). doi:10.1016/j.cell.2015.10.012
27. Li, A. *et al.* Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. *Cell Res.* **27,** 444–447 (2017).
28. Shi, H. *et al.* YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Res.* **27,** 315–328 (2017).
29. 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).
30. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 117–120 (2014). doi:10.1038/nature12730
31. Wang, Y. *et al.* N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* **16,** 191–198 (2014).
32. Du, H. *et al.* YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat. Commun.* **7,** 1–11 (2016).
33. Ke, S. *et al.* A majority of m6A residues are in the last exons, allowing the potential for 3′ UTR regulation. *Genes Dev.* 1–17 (2015). doi:10.1101/gad.269415.115.9
34. Fustin, J. *et al.* RNA-Methylation-Dependent RNA Processing Controls the Speed of the Circadian Clock. *Cell* **155,** 793–806 (2013).
35. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* **61,** 507–519 (2016).
36. Alarcon, C. R. *et al.* HNRNPA2B1 Is a Mediator of m6A-Dependent Nuclear RNA Processing Events. *Cell* 1299–1308 (2015). doi:10.1016/j.cell.2015.08.011
37. Liu, N. *et al.* N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 560–564 (2015). doi:10.1038/nature14234

Chapter 2 – Molecular characterization of YTHDC1

2.1 Identification of the YTH domain

YTHDC1 (also known at YT521-B) is the founding member of the YTH (YT Homolog) protein family, with annotation of the YTH domain based on the conservation of 14 invariant and 19 highly conserved residues[1]. Initial structural analysis of the YTH domain suggested that is it related to the pseudouridine synthase and archaeosine transglycosylase (PUA) domain, which utilizes aromatic stacking to recognize single stranded nucleic acids[2]. Indeed, the YTH domain of YTHDC1 was found to bind short, degenerate RNA sequences by SELEX resembling a KVVURC motif (K = G or U, V = G or C or A, R = G or A)[3].

In light of the identification of YTH Domain Family proteins as methyl-selective RNA binding proteins by RNA affinity chromatography[4] we chose to examine the RNA binding properties of YTHDC1.

2.2 Results

Although the YTH domains of YTHDF1, YTHDF2, and YTHDC1 are highly conserved (**Figure 2.1**), the only YTH proteins isolated by m^6A-affinity chromatography were YTHDF1 and YTHDF2.

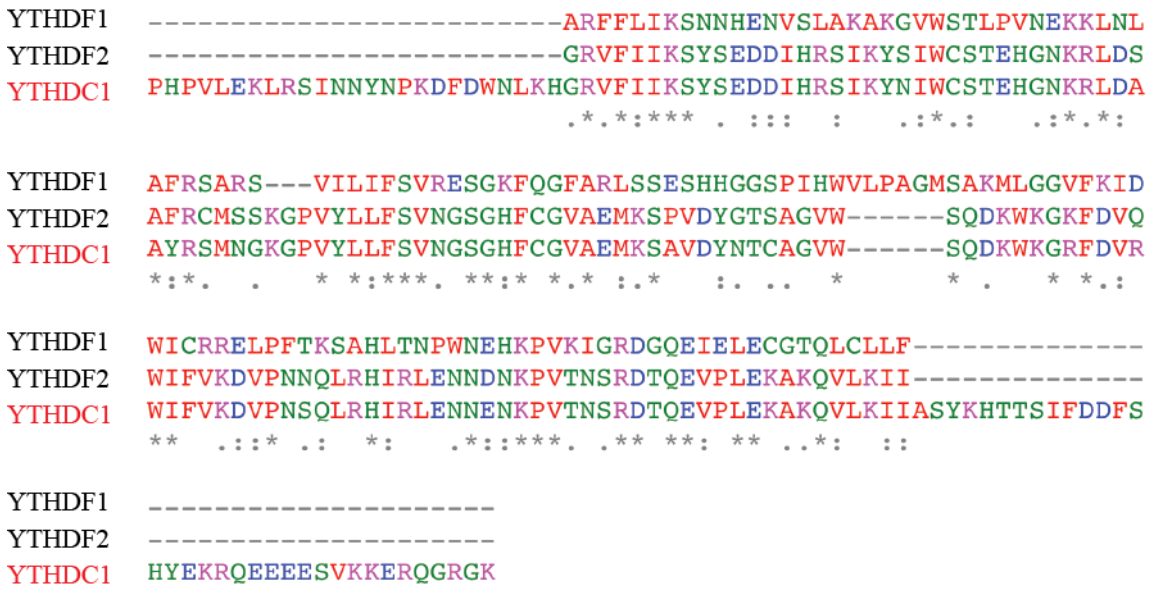

Figure 2.1 Alignment of the YTH domains of YTHDF1, YTHDF2, and YTHDC1[5]

2.2.1 YTHDC1 preferentially binds m^{6}A in RNA

The YTH domain of YTHDC1 was thus cloned and purified in order to investigate RNA binding specificity with respect to RNA methylation (**Figure 2.2a**). It is worth noting that the full length YTHDC1 was not sufficiently expressed in *E. coli*. We purified mRNA from HeLa cells by two rounds of poly(A) selection, and incubated it with limiting amounts of the YTHDC1 YTH domain. The mixture was allowed to complex, and the YTH domain was isolated using His-affinity beads. After isolation of input, bound, and flow-through RNA, nucleoside composition of the digested samples was determined by high performance liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS)[6]. Enrichment for m^{6}A in the IP portion, combined with depletion in the flow-through sample, suggested that the YTH domain of YTHDC1 is an m^{6}A-specific RNA binding domain (**Figure 2.2b**). Immunoprecipitation of endogenous YTHDC1 from HeLa cells further indicated that YTHDC1 binds to m^{6}A modified RNAs (**Figure 2.2c,d**).

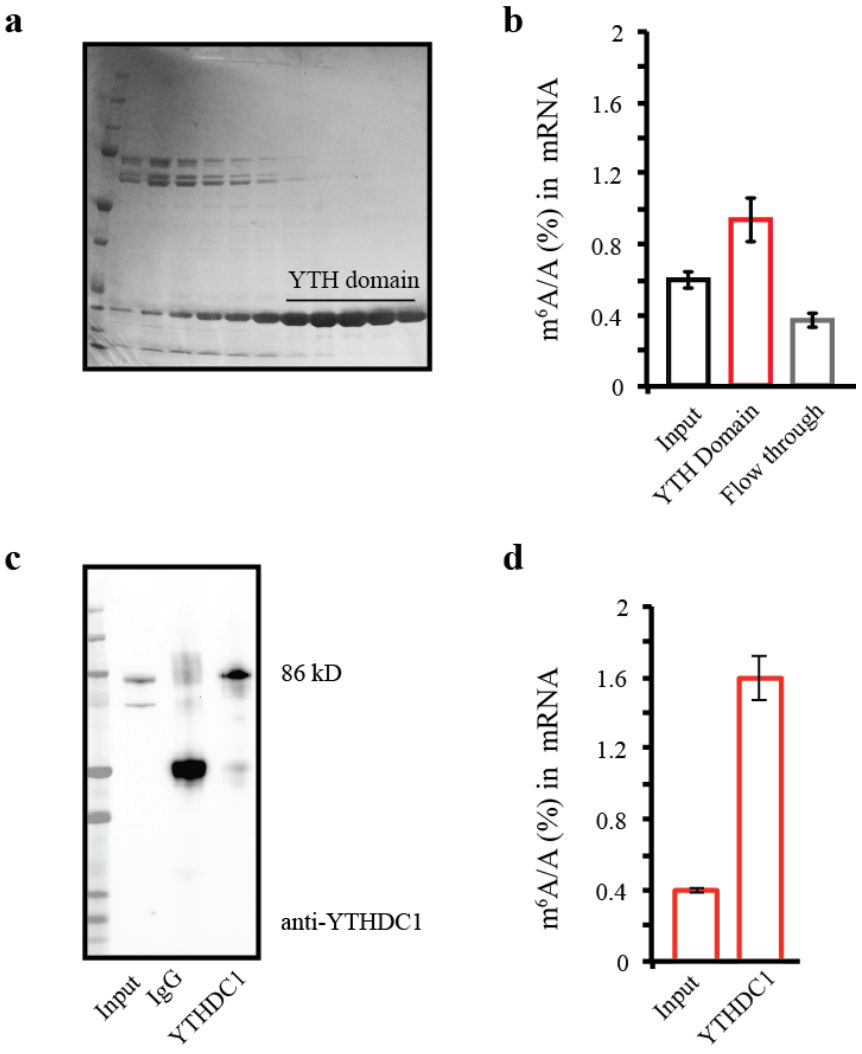

Figure 2.2 LC-MS/MS of YTHDC1-bound RNA

(a,b) The YTH domain of YTHDC1 is selective for m^6A in mRNA *in vitro*. (c,d) mRNA immunoprecipitated by endogenous YTHDC1 is enriched for m^6A.

We further evaluated the affinity of the YTH domain of YTHDC1 in a gel mobility shift assay. RNA probes were designed with an embedded consensus m^6A sequence GG(X)CU within a 42-mer, with X=A and X=m^6A. The YTH domain of YTHDC1 is able to shift the methylated probe in the gel shift assay, but does not show significant binding to the probe lacking m^6A (**Figure 2.3**).

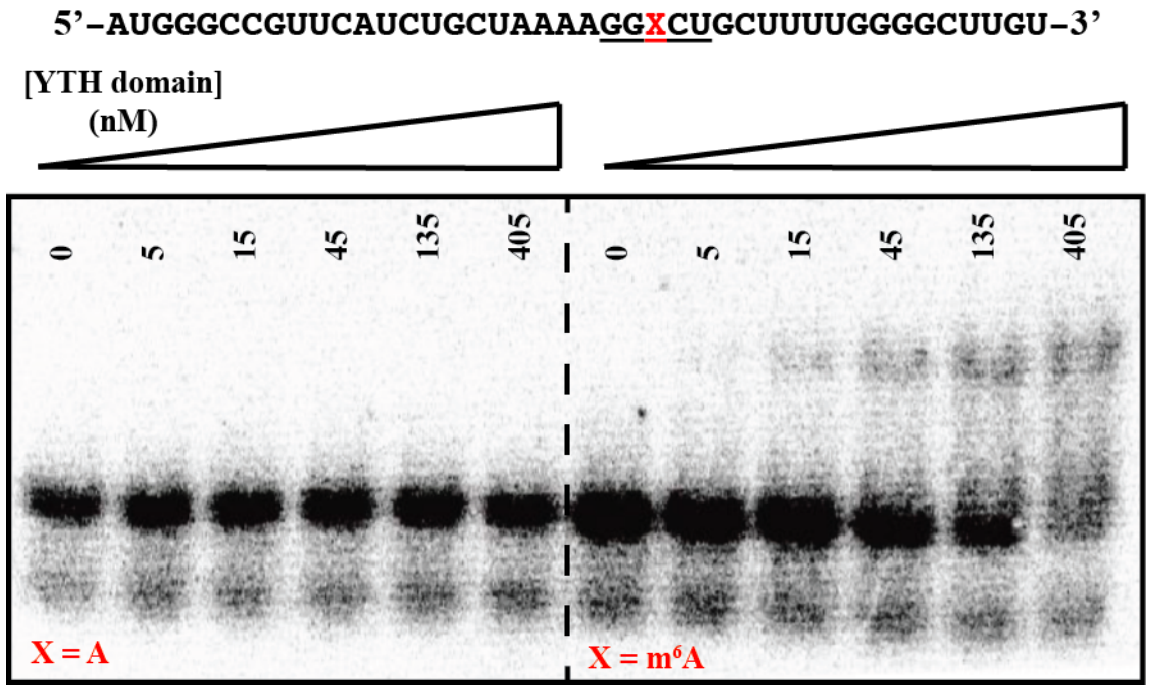

Figure 2.3 Gel shift assay of the YTH domain of YTHDC1

Gel image of His-YTH domain of YTHDC1. At concentrations of 405 nM and above, the YTH domain cannot enter the gel completely. At this concentration, the intensity of the upper band (RNA-protein complex) and the lower band (free probe) appear equal. In each lane, 4 nM RNA probe was labeled with ^{32}P.

These data further validated our conclusion that the YTH domain of YTHDC1, like that of both YTHDF1 and YTHDF2, is a methyl-specific RNA binding domain.

2.2.2 Analysis of YTHDC1 binding specificity *in* vivo

Having established the binding specificity of YTHDC1 *in vitro*, we sought methods with which to analyze the native RNA targets of the nuclear protein. To characterize the direct binding sites, we utilized photoactivatable ribonucleoside crosslinking and immunoprecipitation (PAR-CLIP)[7] of YTHDC1 in HeLa cells. PAR-CLIP utilizes the photoreactive nucleoside 4-thiouridine to facilitate crosslinking to adjacent aromatic residues of RNA binding proteins. Upon immunoprecipitation and treatment with Protease K, crosslinked RNA is tagged with remnants of the digested protein. This residue commonly causes T to C mutations during reverse transcription (the crosslinked residue pairs with G rather than A), leaving a chemical footprint of RNA binding.

Analysis of sequencing reads generated by PAR-CLIP of YTHDC1 immediately suggested

binding to mRNA, with reproducible peak regions mapping to 3' UTRs (38%) and exon (61%) (**Figure 2.4**).

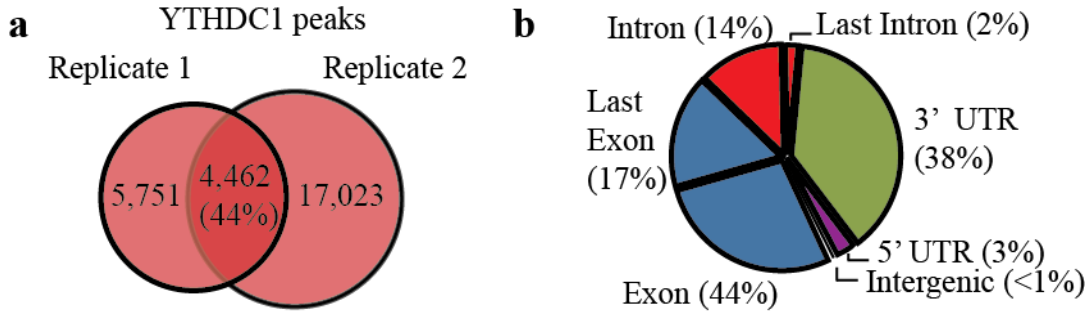

Figure 2.4 Analysis of YTHDC1 binding regions by PAR-CLIP

(a) Replicate experiments of YTHDC1 PAR-CLIP produced 4,462 binding regions (b) YTHDC1 binding regions occur mostly within 3' UTRs and coding regions of mRNAs

We searched YTHDC1 binding regions for a consensus binding motif, finding that the canonical GGAC motif is the most enriched sequence in our experiments. Other motifs may reflect binding profiles of interacting proteins, such as the GAAGAA motif predicted for SRSF1[8] (**Figure 2.5a**). Routine analysis of PAR-CLIP data requires incorporation of a T to C mutation in order to define a binding region, as well as sufficient coverage of that region to help eliminate signal generated by highly abundant cellular RNAs. We therefore analyzed the distance from observed T to C mutations to the nearest GGACU sequence. We find that most mutations occur within 20 nucleotides of a consensus motif, suggesting that YTHDC1 binds directly to sites of m^6A modification (**Figure 2.5b**).

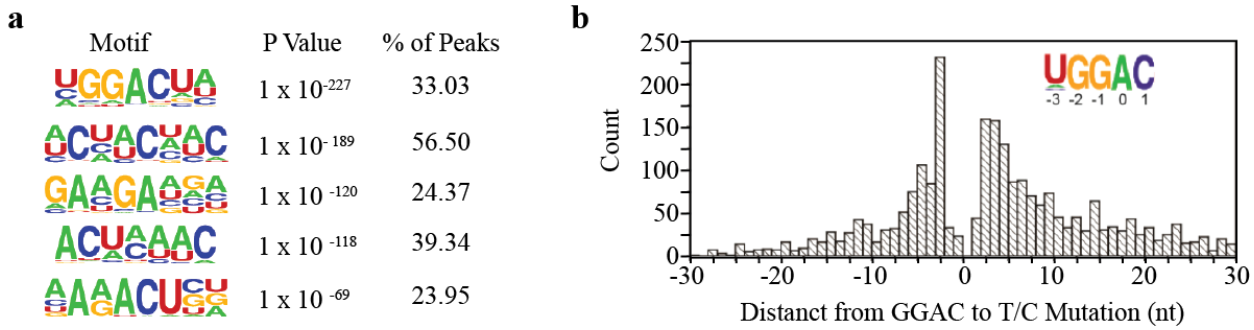

Figure 2.5 YTHDC1 motif analysis

Figure 2.5 continued

(a) Motif analysis of YTHDC1 binding sites by HOMER[9] (b) Distance between T to C mutations and GGAC motifs within YTHDC1 binding regions

2.2.3 Structural basis for YTH-domain specificity for m^{6}A-modified RNA

Concomitant with our work in characterizing the specificity of YTHDC1, the crystal structure of the YTH domain in complex with a 5-mer of m^{6}A RNA sequence (GGm^{6}ACU) was solved in collaboration with the lab of Dr. Jinrong Min[10]. The YTH domain adopts a familiar α/β fold, consistent with a previously reported solution structure of the apo form[11]. The two structures are nearly identical, different significantly only in the resolution of L_{45}, the flexible loop spanning β_4 and β_5 which is visible in the complex structure only. It is this loop which forms a deep, hydrophobic pocket in conjunction with residues from β_1, β_2 and the intervening residues that form α_1 between them (**Figure 2.6a**). Surface projections a highly charged binding cleft (**Figure 2.6b**), which features a deep pocket of hydrophobic, aromatic side chains to accommodate the methylated adenosine (**Figure 2.6c**).

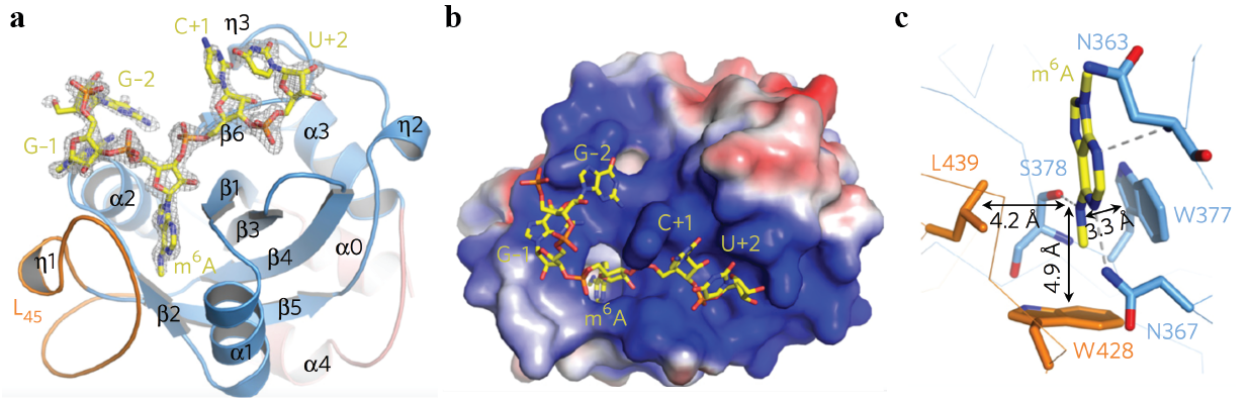

Figure 2.6 Structural characterization of the YTH domain – m^{6}A RNA complex

(a) Structure of the YTH domain of YTHDC1 in complex with GG(m^{6}A)CU RNA. The loop L_{45} is shown in orange, with helix and beta sheet structures in cyan. 5-mer RNA is shown in yellow.

Two tryptophan side chains within the m^{6}A binding pocket, W377 and W428, are completely conserved throughout the YTH family and required for RNA binding. Other structural studies reveal a similar pocket in the YTH domains of YTHDF1 and YTHDF3, as well as the yeast protein

MRB1 of *Zygosaccharomyces rouxii*[12–15]. These binding modes are reminiscent of those of methyllysine and methylarginine modifications of histone proteins[16,17]. Selectivity for the -1 and +1 positions in the YTH domain of YTHDC1 are mainly derived from sequence-specific hydrogen bonding which favors G at the -1 position and cation-π and π-π interactions with C at the +1 position. The YTH domain of YTHDC1 thus favors the GGACU consensus motif of m^6A methylation in transcripts which contain and lack base modification. This distinguished the protein from other members of the mammalian YTH family, which largely derive selectivity from m^6A methylation, and are otherwise less sequence specific despite a high level of conservation in primary sequence within the RNA binding domain.

2.3 Conclusions and Discussion: YTH domain proteins as m^6A effectors

Extensive biochemical and structural studies have identified the YTH proteins as a family of RNA binding proteins with selectivity for m^6A in RNA. The presence of the YTH domain throughout eukaryotes, as well as the conservation of preference for m^6A suggest that reversible RNA methylation has evolved as a form of post-transcriptional gene regulation. The existence of bona fide, m^6A-selective RNA binding proteins (m^6A effector proteins or m^6A 'readers') establishes a paradigm of reversible chemical modification of RNA in which m^6A methyltranferases (m^6A 'writers') and m^6A demethylases (m^6A 'erasers') shape the RNA binding profiles and thus the fate of methylated mRNAs.

Mammals have five YTH proteins (YTHDF1-2, YTHDC1-2). While each displays selectivity for m^6A, only YTHDC1 has a strong preference for GG(m^6A)CU and resides within the cell nucleus. These proteins suggest a 'division of labor' amongst YTH proteins in mediating the biological impact of m^6A methylation, which will undoubtedly remain a focus of future research. Currently, there is little evidence that the proteins are able to distinguish between subsets of methylated mRNAs, which suggests that all m^6A methylated mRNAs are subject to similar routes of mRNA metabolism. However, varied sequence context may affect the activity of the both the methyltransferase and demethylase enzymes, altering the presence of transcripts in selective pathways for RNA metabolism. Additionally, the presence of m^6A 'reader' proteins of another

family may use alternate 'reading' mechanisms and in turn distinguish the presence of methylation within and outside common motif sequences. An extension of this principle may allow for separate utilization of modification throughout the transcript, such as unique roles for 5' UTR and 3' UTR methylation.

Reversible RNA methylation can be dynamically tuned by several mechanisms, few of which are understood. Regulation of methyltransferase, demethylase, or effector protein abundance or activity is perhaps the simplest model. More precise regulation may occur within the sequence-context of m^6A and a variety of 'reader' proteins, as methylation is not confined to merely GG(m^6A)CU motifs. Subtle changes YTH domain sequence and structure can have profound effects on RNA affinity, highlighted by the 10-fold decrease in affinity for observed for the YTH domain of YTHDC2 versus other mammalian YTH domains[12] and alternative binding grooves and lack of m^6A specificity for the YTH domain of Mmi1[18]. These factors lead to the widely heterogeneous nature of m^6A-based regulation, but suggest a role for RNA modification in a range of fundamental steps in RNA metabolism and biological processes.

2.4 Methods

2.4.1 Expression of the YTH domain of YTHDC1

The YTH domain of YTHDC1 was subcloned from commercial cDNA (Open Biosystems) into the pet-28(a) vector.

(YTH)DC1-F: CGTCACATATGCAAACCAGTAAACTCAAATATGTGC
(YTH)DC1-R: GGCATCTCGAGTCAGTGACGCATTTTATGAATGACCTG

Resultant clones were transformed into *E. coli.* strain BL21 (New England Biolabs) and cultured in suspension at 37° C. Expression was induced with 1 mM IPTG (GoldBio) at 16° C overnight. The bacteria was pelleted and resuspended in lysis buffer (20 mM Tris, pH 7.5, 200 mM NaCl, and 0.1% (v/v) Triton X-100 (Fisher). Lysates were sonicated on ice for 10 minutes (25% amplitude, 30 seconds on, 30 seconds off), and centrifuged at 13,000 rpm for 30 minutes at 4° C. The soluble portion was filtered through a 0.45 μm syringe filter and applied to a Ni-NTA cartridge

according to manufacturer's protocol (Qiagen). Crude products were concentrated and further purified by gel filtration (10 mM Tris, pH 7.5, 150 mM NaCl, 1 mM DTT). Yield og purified YTH domain was ~10 mg/L of bacterial culture.

2.4.2 RNA isolation and purification of mRNA

Total RNA was isolated from HeLa cells using TriZol reagent (Invitrogen) according to manufacturer's instructions. RNA was pelleted form the aqueous portion using an equal volume of isopropanol and 1 μg glycogen as a co-precipitant (Ambion). The pellet was washed with 75% ethanol and resuspended in RNase-free water.

Messenger RNA was purified by two rounds of poly(A) selection using the mRNA DIRECT kit (Invitrogen) from total RNA samples. The final elution was performed in RNase-free water.

2.4.3 *In vitro* RNA IP

1 μg of mRNA was fragmented in water by sonication and used in the *in vitro* IP. 200 ng was saved as input, and 800 ng was incubated with the YTH domain of YTHDC1 (500 nM final concentration) diluted to in 200 uL IPP buffer (10 mM Tris, pH 7.4, 150 mM NaCl, 0.1% NP-40, 40 U/mL RNase inhibitor, 0.5 mM DTT). The solution was incubated at 4° C for 2 hours. Anti-His-tag beads were washed with IPP buffer, resuspended in 50 μL IPP buffer and added to the IP mixture for an additional 2 hours. The aqueous phase was collected as the flow-through portion. Beads were washed with 4° C IPP buffer 4 x 300 μL. TriZol was added to the washed beads and flow-through portion for RNA isolation and LC-MS/MS.

2.4.4 Analysis of m^6A abundance by LC-MS/MS

100 ng of mRNA was digested by Nuclease P1 (1U) for 2 hours at 37° C in 25 μL Nuclease P1 digestion buffer (25 mM NaCl, 2.5 mM $ZnCl_2$). Following digestion by Nuclease P1, 3 μL of 10X FastAP buffer and 1U FastAP were added and incubated at 37° C overnight. The final digestion was diluted to 100 μL with RNase-free water and subjected to LC-MS/MS. Nucleosides were

separated by reverse phase ultra-performance liquid chromatography on a C-18 column (Agilent) and detected using an Agilent 6410 QQQ triple-quadrupole mass spectrometer. Nucleosides were quantified in positive ionization mode with nucleoside to nucleobase transition masses of 282 > 150 (m^6A) and 268 > 136 (A). The ratio of m^6A/A was calculated according to standard curves for each nucleoside.

2.4.5 Cross-linking and immunoprecipitation of endogenous YTHDC1

Crosslinking was performed using Stratalinker 2400 (Stratagene) autocrosslink option on untreated HeLa cells.

HeLa cells were washed with PBS two times, then collected using a cell lifter and spun at 2,000 x g for 5 minutes. Pellets were suspended in 2 volumes (compared to pellet) Buffer A (10 mM HEPES, pH 7.5, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT), vortexed briefly, and incubated on ice for 15 minutes. NP-40 was added to a final concentration of 0.25%, vortexed, and left on ice for 5 minutes. Suspensions were spun at 2,000 x g for 3 minutes at 4 °C. The supernatant (cytoplasmic extract) was removed and saved. Nuclei were suspended in 2 volumes Buffer B (20 mM HEPES, pH 7.5, 0.42 M KCl, 4 mM $MgCl_2$, 1 mM EDTA, 0.5 mM DTT, 10% glycerol), vortexed briefly, and incubated on ice for 30 minutes. The suspension was spun at 15,000 x g for 15 minutes at 4° C, and supernatant combined with the cytoplasmic extract. This was incubated on ice for 15 minutes and spun again at 15,000 x g for 15 minutes at 4° C. The supernatant was removed and used as the HeLa lysate.

Lysate was treated with rabbit IgG control or primary antibody against YTHDC1 (ab122340, Abcam, 5 μg/IP) and incubated at 4 °C for two hours. After two hours, washed Dynabeads Protein A (Thermo Fisher) were added, and the solution was incubated for an additional hour, followed by washing with 4 x 500 μL with Buffer C (wash buffer) (50 mM Tris-HCl, pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.2 mM EDTA, 0.1% NP-40, 10 mM β-ME, 10% glycerol). Washed beads were resuspended in 100 μL wash buffer. 100 μL 2X Proteinase K buffer (100 mM Tris-HCl, pH 7.5, 200 mM NaCl, 2 mM EDTA, 1% SDS) was added, followed by Proteinase K (Thermo Fisher) to

a final concentration of 500 μg/mL and heated for 1 hour at 65 °C. RNA was isolated using TriZol. Recovered RNA was purified using two rounds of poly(A)-selection and subjected to LC-MS/MS.

2.4.6 Electrophoretic mobility shift assay for YTHDC1

RNA probes were synthesized using the Expedite Nucleic Acid Synthesis System: 5'- AUGGGCCGUUCAUCUGCUAAAAGGXCUGCUUUUGGGGCUUGU-3', X = m^6A or A. The probes were labeled with ^{32}P using T4 PNK (Fermentas) at 37° C for 1 hour in a reaction mixture containing 2 μL RNA probe (1 μM), 1 μL T4 PNK, 1 μL ^{32}P-ATP (PerkinElmer), 5 μL T4 PNK Buffer (Fermentas) and 41 μL RNase-free water (final [RNA] was 40 nM). The Reaction was purified using micro bio-spin columns with bio-gel P3- in Tris buffer (BioRad 732-6250) to remove excess ATP. The probe was diluted with 2.5 μL 20X SSC (Promega) and heated to 65° C for 10 minutes, followed by slow cooling to room temperature.

Protein dilution were prepared in binding buffer (10 mM HEPES, pH 8.0, 50 mM KCl, 1 mM EDTA, 0.05% Triton X-100, 5% glycerol, 1 mM DTT, 40 U/L RNase inhibitor with 10 μg/mL Salmon sperm DNA as a blocking agent) to concentrations ten times those shown in Figure 2.3. For each well 1 μL protein solution and 1 μL RNA probe (final concentration 4 nM) were incubated in 10 μL on ice for 30 minutes. Protein-RNA complexes were electrophoresed with a-4-20% TBE Gel (Novex) at 4° C at 160 V. The gel was vacuum dried at 80° C for 2 hours, and exposed to a phosphor screen (K-screen, Fuji film) over night before reading by a BioRad Molecular Imager FX (BioRad).

2.4.7 Photoactivatable ribonucleoside crosslinking and immunoprecipation (PAR-CLIP) of YTHDC1

PAR-CLIP was performed based on a previously reported procedure[7] with the following modifications. 4 × 15 cm plates of cells were transfected with Lipofectamine 2000 according to manufacturer's protocol. After 6 h, the medium was replaced, and cells were cultured in fresh medium supplemented with 200 μM 4-thiouridine (Sigma) overnight. The first RNase T1 digestion

was conducted at 1 U/μl RNase T1 for 8 min. For the second digestion, the concentration of RNase T1 was reduced from 100 U/μl to 20 U/μl. Following dephosphorylation, one-tenth of the sample was partitioned for ^{32}P labeling. The remaining volume was treated with 1 U/μl T4 PNK at 37 °C for 10 min, followed by addition of ATP to 1 μM for 30 min at 37 °C. This sample was then washed and digested with Proteinase K. RNA was purified using Zymo RNA Clean and Concentrator before library construction using the Tru-seq small RNA sample preparation kit (Illumina). The cDNA library was sequenced by Illumina Hiseq2000 with a single-end 50-bp read length.

2.4.8 Data analysis of YTHDC1 PAR-CLIP

Treatment of raw data: Reads were trimmed using Cutadapt v.1.4.1[19] and aligned to hg19 using TopHat v.2.0.11[20]. PAR-CLIP peaks were called using PARalyzer v.1.5[21] with default parameters. Binding motifs were determined by HOMER v.4.7[22]. The distances between GGAC and T to C mutation sites were calculated by setting the A of GGAC as position zero.

References

1. Stoilov, P., Refalska, I. & Stamm, S. YTH : a new domain in nuclear proteins. *TRENDS Biochem. Sci.* **27,** 495–497 (2002).
2. Theobald, D. L., Mitton-Fry, R. M. & Wuttke, D. S. NUCLEIC ACID RECOGNITION BY OB-FOLD PROTEINS. *Annu. Rev. Biophys. Biomol. Struct.* 115–133 (2003). doi:10.1146/annurev.biophys.32.110601.142506
3. Zhang, Z. *et al.* The YTH Domain Is a Novel RNA Binding Domain. *J. Biol. Chem.* **285,** 14701–14710 (2010).
4. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–206 (2012).
5. Mcwilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41,** 597–600 (2013).
6. Jia, G. *et al.* N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7,** 6–9 (2011).

7. Hafner, M. *et al.* Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* **141,** 129–141 (2010).
8. Wang, X. *et al.* Predicting sequence and structural specificities of RNA binding regions recognized by splicing factor SRSF1. *BMC Genomics* **12,** 1–11 (2011).
9. 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).
10. Xu, C. *et al.* Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* **10,** 927–929 (2014).
11. He, F. *et al.* Solution structure of the YTH domain in YTH domain-containing protein 1 (putative splicing factor YT521). *To be Publ.* (2009). doi:10.2210/PDB2YUD/PDB
12. Xu, C. *et al.* Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. *J. Biol. Chem.* **290,** 24902–24913 (2015).
13. Zhu, T. *et al.* Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res.* **24,** 1493–1496 (2014).
14. Theler, D., Dominguez, C., Blatter, M., Boudet, J. & Allain, H. Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. *Nucleic Acids Res.* **42,** 13911–13919 (2014).
15. Luo, S. & Tong, L. Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. **111,** (2014).
16. Wang, G. G. *et al.* Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459,** 847–851 (2009).
17. Jacobs, S. A. & Khoransanizahed, S. Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail. *Science.* **295,** 2080–2083 (2002).
18. Wang, C. *et al.* A novel RNA-binding mode of the YTH domain reveals the mechanism for recognition of determinant of selective removal by Mmi1. **44,** 969–982 (2016).
19. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).
20. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).

21. Corcoran, D. L. *et al.* PARalyzer : definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12,** R79 (2011).
22. 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).

Chapter 3 Nuclear roles for mRNA m^{6}A methylation

3.1 The fate of nuclear RNA

Nearly all of the RNA transcribed in the nucleus is eventually destroyed before making it to the cytoplasm. Only 50% of nucleotides incorporated by RNA polymerase I (Pol I) end up in mature rRNA, while the rest is trimmed away in within the nucleolus by pre-rRNA processing machinery. The outcome is bleaker for Pol II-derived transcripts, in which less than 5% of a nascent transcript survives to become a mature mRNA product in the cytoplasm, largely due to the removal of introns during splicing and other quality control measures[1]. Thus, the largest pool of RNA turnover exists in the nucleus rather than the cytoplasm, and yet the factors that determine the nuclear fate of RNA, in particular pre-mRNA, remain largely unexplored.

The bridge between nuclear pools of RNA and cytoplasmic transcripts available for translation is the transport of RNA. What little RNA that survives nuclear processing must be shuttled to the cytoplasm where ribosomes are poised to translate their messages. In order to facilitate this process, eukaryotic cells have evolved intricate systems to couple transcription by Pol II with maturation processes such as 3' polyadenylation, pre-mRNA splicing, 5' capping and nuclear export[2]. Regulation of 3' end formation[3], alternative splicing[4], and nuclear decay have emerged as points of intervention in determining gene expression. Mechanisms which help evade the widespread RNA decay in the nucleus may serve as invaluable opportunities to preserve products of transcription. Recently, examples of selective mRNA export have suggested that well conserved components of export pathways may be serve additional roles in regulating the availability of cytoplasmic transcripts[5].

3.2 Internal mRNA modifications and RNA processing

Post-transcriptional modifications of RNA are essential for their cellular function. Pre-mRNAs depend on the 5' cap for transcript stability[6,7], 3' end formation[8], and pre-mRNA splicing[8–10], as well as export[11] translation initiation in the cytoplasm[12]. The poly(A) tail recruits members of the poly(A) binding protein family to facilitate several aspects of RNA

metabilosm[13,14]. Pre-mRNA splicing results in a drastic alteration of the nascent RNA sequence, producing a functional mRNA in conjunction with capping and polyadenylation.

Internal modifications of RNA nucleosides are linked to several stages of RNA maturation in the nucleus. Early studies suggested that m^{6}A methylation is essential for efficient processing and transport of nuclear mRNA[15,16], which immediately suggests that m^{6}A-methylation of mRNAs may serve a role in selecting marked transcripts for differential processing. The presence of several nuclear 'reader' proteins has validated this hypothesis, highlighting a role for m^{6}A in the regulation of alternative splicing[17–20] and alternative polyadenylation[21]. YTHDC1, the only nuclear member of the YTH family in mammals, has also been shown to regulate alternative exon usage, and perhaps does so in an m^{6}A-dependent manner[22–26]. As with m^{6}A in the cytoplasm, we hypothesized that methylation of nuclear mRNAs facilitates selective incorporation into canonical pathways of RNA metabolism, a multitude of which may contribute to shortened half-life times observed of modified mRNAs (**Figure 3.1**).

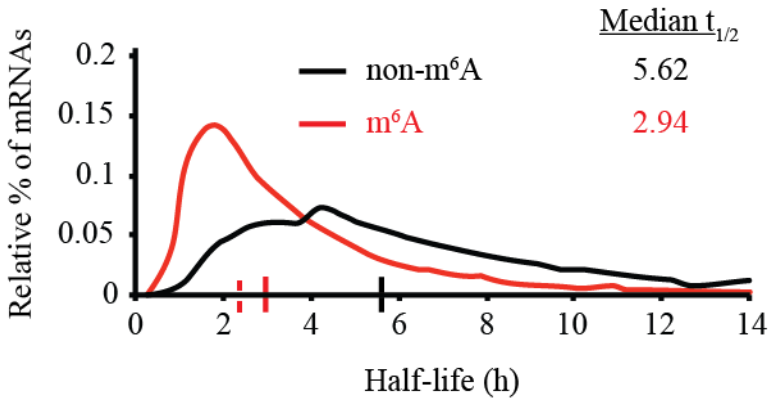

Figure 3.1 m^{6}A methylated mRNAs experience selective metabolism

mRNAs marked with m^{6}A experience significantly reduced half-life times in HeLa cells. YTHDF2-facilitated mRNA decay of m^{6}A-containing transcripts may be one of many examples of selective mRNA processing that facilitates expedited clearance.

In this chapter, we discuss a novel role for m^{6}A and YTHDC1 in promoting the processing and export of methylated mRNAs form the nucleus to the cytoplasm.

3.3 Results

3.3.1 YTHDC1 facilitates nuclear to cytoplasmic transport of m^{6}A methylated mRNAs

m^{6}A promotes translation initiation and cytoplasmic mRNA decay via recognition by YTHDF1 and YTHDF2, respectively[27,28]. We hypothesized that YTHDC1 may promote nuclear processes in an m^{6}A-dependent manner, as the protein itself is strictly nuclear (**Figure 3.2**).

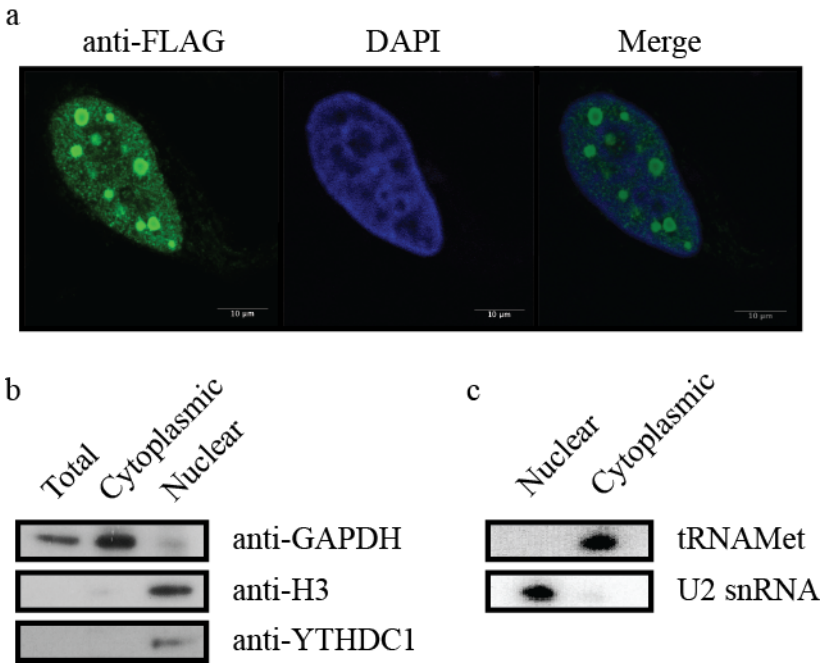

Figure 3.2 Cellular localization of YTHDC1

(a) Immunofluorescence of YTHDC1-FLAG in HeLa cells (b) Western blot analysis of cellular fractionation of HeLa cells (c) Northern blot analysis of cellular fractionation of HeLa cells.

We first asked if methylated transcripts in the nucleus experience shorter half-lives in the nucleus compared to non-methylated mRNAs, as is the case with cytoplasmic mRNA. We tested this by treating cells with Actinomycin D and monitoring m^{6}A levels in nuclear and cytoplasmic mRNA over six hours[29]. We found that cytoplasmic m^{6}A levels diminish over the time-course, and that this behavior is dependent on the presence of YTHDF2, consistent with previous models of m^{6}A-dependent mRNA decay (**Figure 3.3a**). We observe a similar trend in the nucleus,

suggesting that m^{6}A-dependent clearance of mRNA from the nucleus is a function of the nuclear 'reader' YTHDC1 (**Figure 3.3b**). We confirmed the observations occur post-transcriptionally by comparing native transcription rates of methylated versus non-methylated transcripts, which do not differ significantly in HeLa cells (**Figure 3.3c**).

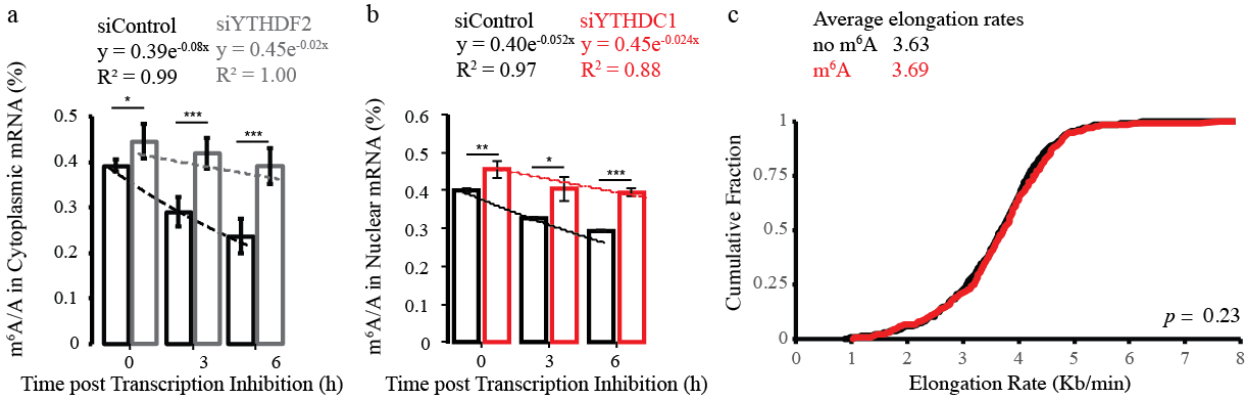

Figure 3.3 Clearance of m^{6}A methylated mRNAs for the nucleus and cytoplasm, transcription rates of m^{6}A methylated and non-methylated genes

(a) m^{6}A methylation in cytoplasmic mRNA analyzed at 0, 3, and 6 hours after transcription inhibition. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance. Curves fit to exponential decay. (b) m^{6}A methylation in nuclear mRNA analyzed at 0, 3, and 6 hours after transcription inhibition. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance. Curves fit to exponential decay. (c) Native transcription elongation rates of methylated versus non-methylated mRNAs in HeLa cells[30], *P*-value calculated using the Mann-Whitney-Wilcoxon Test.

We confirmed that both YTHDF2 and YTHDC1 act independently by measuring nuclear and cytoplasmic m^{6}A levels following knockdown of the two proteins, respectively. While we found that YTHDF2 does not affect the nuclear clearance of m^{6}A-containing mRNAs (**Figure 3.4a**), YTHDC1 limits their initial cytoplasmic availability, and relative abundance of methylation normalizes over the six-hour time-course (**Figure 3.4b**). Overexpression of recombinant YTHDC1 resulted in a decreased m^{6}A/A ratio in the nuclear fraction on mRNA, but did not significantly affect cytoplasmic levels (**Figure 3.4c**).

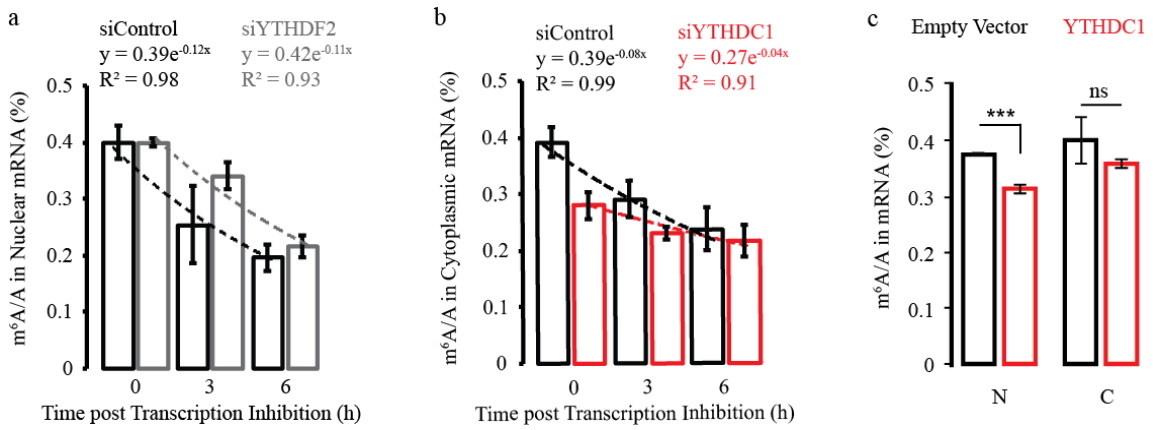

Figure 3.4 Effects of YTHDF2 and YTHDC1 knockdown on nuclear and cytoplasmic mRNA levels

(a) m^{6}A methylation in nuclear mRNA analyzed at 0, 3, and 6 hours after transcription inhibition. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance. Curves fit to exponential decay. (b) m^{6}A methylation in cytoplasmic mRNA analyzed at 0, 3, and 6 hours after transcription inhibition. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance. Curves fit to exponential decay (c) Effect of YTHDC1 overexpression on nuclear and cytoplasmic m^{6}A/A ratios in mRNA. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance.

We next sought to decipher by what mechanism both m^{6}A and YTHDC1 facilitate nuclear clearance of mRNAs. Nuclear accumulation of transcripts following knockdown of YTHDC1 suggests a function in either nuclear export or decay. In order to help distinguish the two mechanisms, we analyzed subcellular m^{6}A levels following siRNA treatment against YTHDC1. Compared to control, knockdown of YTHDC1 results not only in nuclear accumulation of m^{6}A in mRNA, but an accompanying depletion of the cytoplasmic m^{6}A/A ratio (**Figure 3.5a**). These data are suggestive of nuclear retention of m^{6}A-methylated mRNAs in the absence of YTHDC1, as a defect in nuclear decay would not likely result in depleted cytoplasmic transcript levels. It is worth noting that following knockdown of YTHDC1, neither levels of YTHDF2 nor components of the mRNA export machinery NXF1 or ALYREF are significantly depleted (**Figure 3.5b**).

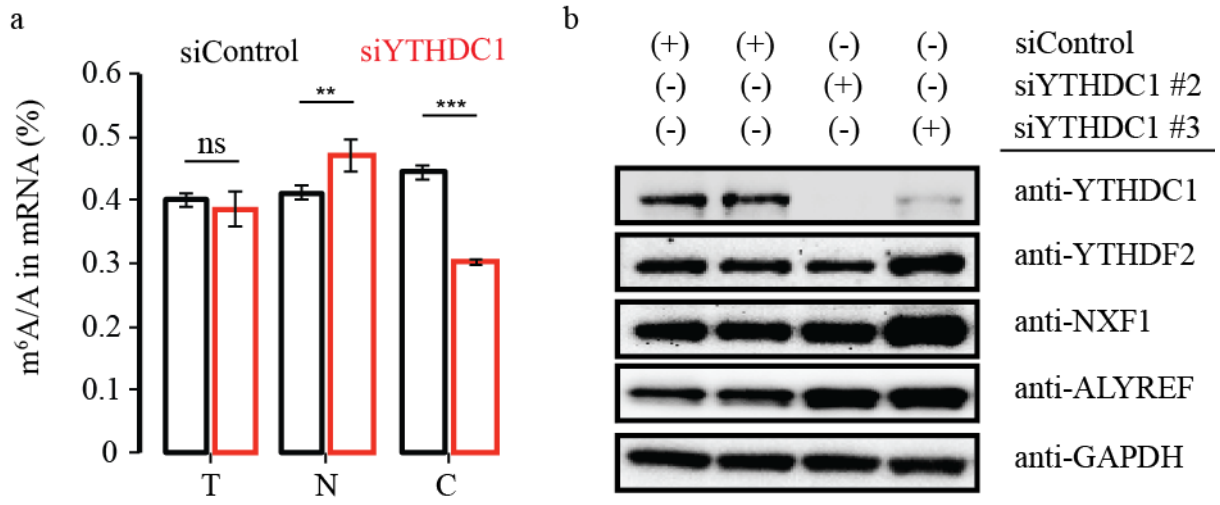

Figure 3.5 Knockdown of YTHDC1 affects the subcellular abundance of methylated mRNAs

(a) Quantification of m^{6}A methylation in total, nuclear, and cytoplasmic mRNA. (T=total, N=nuclear, C=cytoplasmic), Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance (b) Expression of mRNA export machinery following knockdown of YTHDC1 by Western Blot

We next sought a genome-wide understanding of the effects of YTHDC1 knockdown on the subcellular abundance of mRNAs. To do so, we combined RNA immunoprecipitation followed by high-throughput sequencing (RIP-Seq)[31] with our previously generated PAR-CLIP data to generate a high-confidence list of YTHDC1 target transcripts. We performed RIP-seq in biological replicates, using two methods of mRNA purification; poly(A) selection and rRNA depletion. Comparison of the two conditions indicate that the experiment was highly reproducible, with major variations in enrichment coming from transcripts lacking poly(A) tails such as RMRP and RPPH1, two RNA components of cellular RNases (**Figure 3.6a**). We required target transcripts to demonstrate direct YTHDC1 binding by PAR-CLIP across two replicates, and have enrichment fold of at least two-fold in RIP-seq. These criteria resulted in 737 high-confidence targets of YTHDC1 in HeLa cells (**Figure 3.6b**), though this analysis likely represents a conservative estimate of actual protein-RNA interactions between YTHDC1 and nuclear RNA. As expected, YTHDC1 target transcripts are highly enriched for with annotated m^{6}A methylation (**Figure 3.6c**) and code for protein products with broad roles in gene regulation (**Figure 3.6d**).

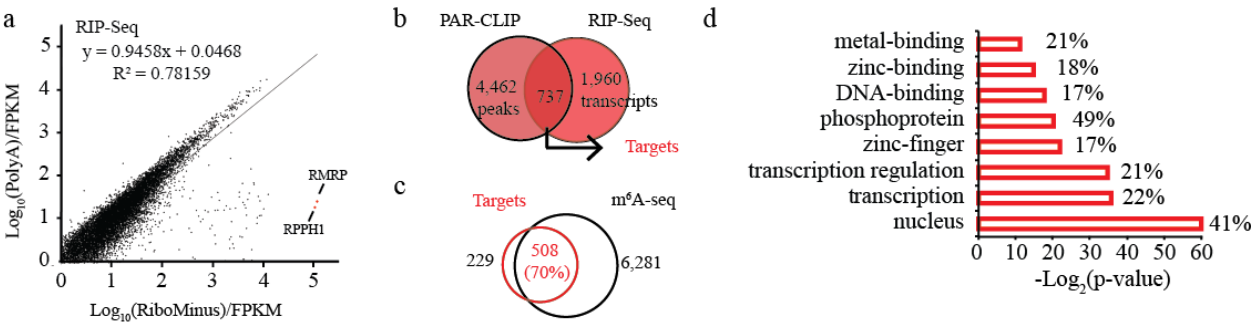

Figure 3.6 Analysis of YTHDC1 target transcripts

(a) Comparison of YTHDC1 RIP-seq replicates. Transcripts enriched in the rRNA-depleted sample but not enriched in the poly(A) selected sample are transcripts lacking poly(A) tails such as RMRP and RPPH1. (b) Selection of YTHDC1 target transcripts from PAR-CLIP and RIP-seq analysis (c) Comparison of YTHDC1 target transcripts and transcripts with known m^6A methylation in HeLa cells (d) Functional annotation of YTHDC1 target transcripts[32].

With an understanding of YTHDC1 targets, we proceeded to analyze the distribution of targets versus non-targets (no PAR-CLIP peaks or enrichment in RIP-seq) following knockdown of YTHDC1 by subcellular RNA-sequencing. The data from RNA sequencing are consistent with LC-MS/MS experiments: knockdown of YTHDC1 does not affect target abundance, but results in the nuclear accumulation of YTHDC1 targets compared to non-targets, and is accompanied by a relative depletion in the cytoplasm under knockdown conditions (**Figure 3.7a-c**).

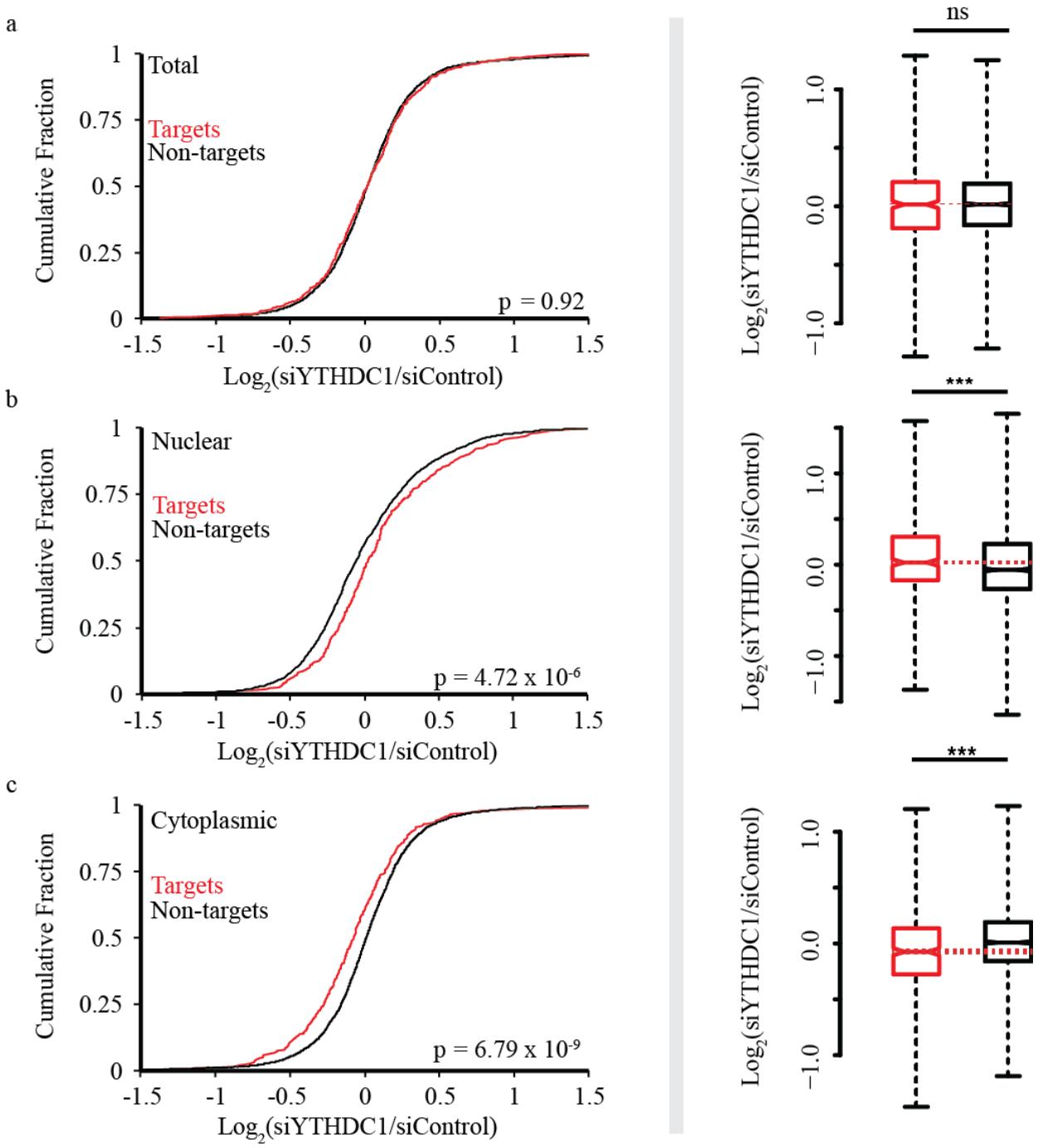

Figure 3.7 Subcellular abundance of YTHDC1 target transcripts by RNA-sequencing

(a) Left: Cumulative distribution of $\log_2$ fold changes in RNA expression following knockdown of *YTHDC1* in total mRNA. *P*-values calculated using the Mann-Whitney-Wilcoxon Test. Right: Boxplot representation of RNA-seq fold changes. Whiskers represent three times the interquartile range. *P*-values calculated using Welch's *T*-test (b) Left: Cumulative distribution of $\log_2$ fold changes in RNA expression following knockdown of *YTHDC1* in nuclear mRNA. *P*-values

Figure 3.7 continued calculated using the Mann-Whitney-Wilcoxon Test. Right: Boxplot representation of RNA-seq fold changes. Whiskers represent three times the interquartile range. *P*-values calculated using Welch's *T*-test (c) Left: Cumulative distribution of $\log_2$ fold changes in RNA expression following knockdown of *YTHDC1* in cytoplasmic mRNA. *P*-values calculated using the Mann-Whitney-Wilcoxon Test. Right: Boxplot representation of RNA-seq fold changes. Whiskers represent three times the interquartile range. *P*-values calculated using Welch's *T*-test, *** = $p < 0.001$. Data represent biological replicates using unique siRNAs against *YTHDC1*.

These population shifts are not due to changes in target abundance upon knockdown of YTHDC1, suggesting a discriminatory effect of YTHDC1 on targets versus non-targets (**Figure 3.8**).

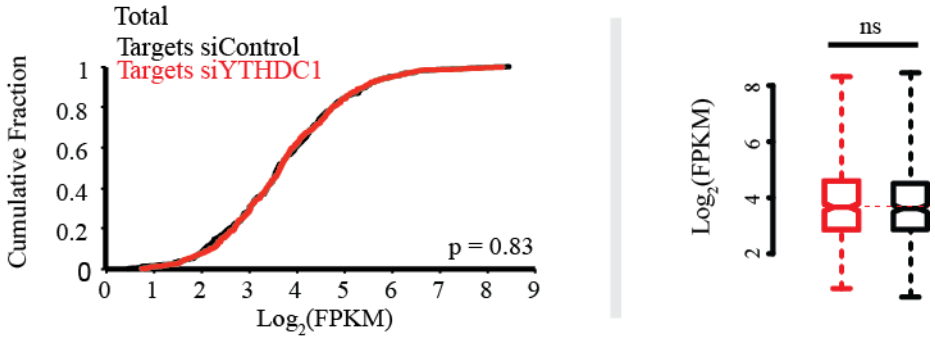

Figure 3.8 YTHDC1 target abundance upon knockdown of YTHDC1

(a) Effect of YTHDC1 knockdown on expression of YTHDC1 targets. *P*-value calculated using the Mann-Whitney-Wilcoxon Test (left) and Welch's *T*-Test (right).

In order to directly observe nuclear accumulation of poly(A)-containing RNAs, we performed fluorescence *in situ* hybridization (FISH) using a poly(dT) probe. Knockdown of YTHDC1 leads to an accumulation of poly(A) signal in the nucleus and decrease cytoplasmic availability of poly(A) RNAs in the cytoplasm (**Figure 3.9**).

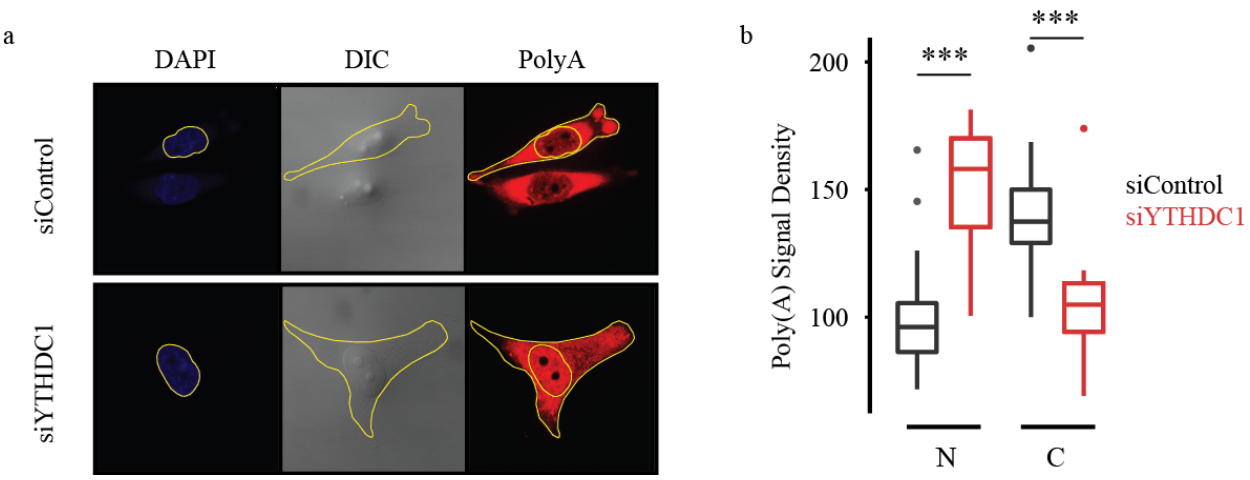

Figure 3.9 Representative poly(A) imaging following knockdown of YTHDC1

(a) Representative analysis of poly(A) imaging. Nuclei were defined using DAPI signal. Cytoplasmic regions were defined by subtracting nuclear signal from total signal, as defined by DIC imaging. (b) Quantification of poly(A) signal density. (N=nuclear, C=cytoplasmic). n = 25, *** = $p < 0.001$, two-sided *t*-test.

Taken together, these data suggest that YTHDC1 functions in mediating the cytoplasmic availability of mRNA, and likely does so through methyl-specific association with RNA. However, the observed nuclear accumulation may result from several defects in nuclear processing which to this point have remained unexplored. As previously discussed, a shift in distribution is not indicative of a defect in nuclear decay. However, a recent study identified YTHDC1 as an interacting protein of the NEXT complex, which indeed mediates the nuclear degradation of promoter upstream transcripts and pre-mRNAs[33,34]. Although we ewre able to confirm the reported interaction between YTHDC1 and the NEXT core component ZCCHC8 (**Figure 3.10a**), the knockdown of YTHDC1, nor the ZCCHC8 or MTR4 components of the NEXT complex resulted in an increase in m^6A/A in nuclear RNA (**Figure 3.10b**).

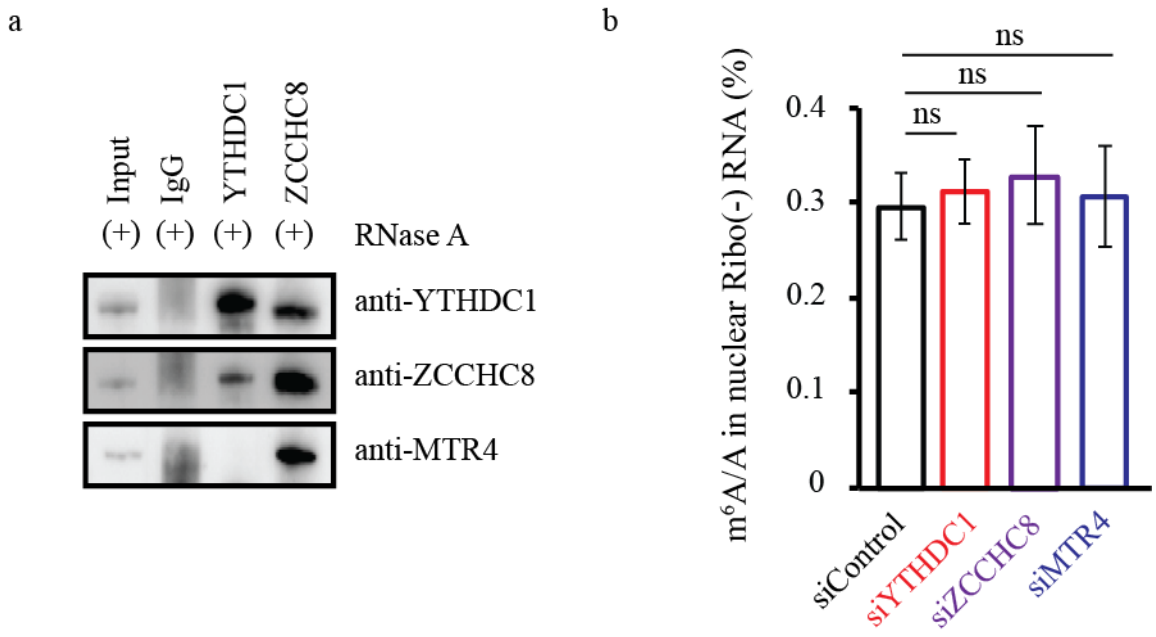

Figure 3.10 Interactions between YTHDC1 and components of the NEXT complex

(a) Western Blot of endogenous IP of YTHDC1 and ZCCHC8 (b) m^6A/A ratio in nuclear, ribosomal depleted pre-mRNA following knockdown of YTHDC1, ZCCHC8, or MTR4. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates).

3.3.2 YTHDC1 facilitates mRNA transport independent of pre-mRNA splicing

In order to understand the role of YTHDC1 in nuclear processing, we analyzed changes in alternative splicing upon knockdown of YTHDC1 genome wide. We ensured sufficient depth and exon-exon junction coverage by sequencing samples using 150 base pair reads in paired-end mode. We utilized a probabilistic model of isoform usage based on data sampling and Bayesian hypothesis testing called Mixture of Isoforms (MISO)[35]. The method takes advantage of junction reads to estimate the likelihood of exon usage, and reports this value of the percent spliced in (PSI or Ψ value) for each exon. By comparing probabilities to annotated alternative splicing events, MISO allows for comparison between samples by computing a change in Ψ (ΔΨ) for each exon. For our analysis, we set the threshold for alternative splicing events at a ΔΨ value of ± 0.2, or a 20% likelihood that the data represent a change in isoform usage. With these parameters, MISO identified 250 exon skipping events upon knockdown of YTHDC1, suggesting that YTHDC1 favors exon inclusion (**Figure 3.11a**). By comparison, knockdown of YTHDC1 produced only176 exon inclusion events, and fewer than 40 other alternative splicing events such as mutually

exclusive exons or alternative 5' or 3' splice site usage (data not shown).

It is possible that YTHDC1 plays multiple roles in pre-mRNA processing and mRNA export. We therefore asked if transcripts with annotated alternative exon usage upon knockdown of YTHDC1 were direct targets of YTHDC1 or not. In doing so, we found that of the 426 alternative splicing events observes, only 85 of them occur in YTHDC1 targets (**Figure 3.11b**). This represents only 11.5% of our defined YTHDC1-targets, and suggests to us that YTHDC1 plays unique roles in pre-mRNA splicing and mRNA export. Furthermore, comparing YTHDC1-dependent skipped exon events to those observed upon knockdown of the methyltransferase METTL3, we observe only weakly negative correlations (**Figure 3.11c**).

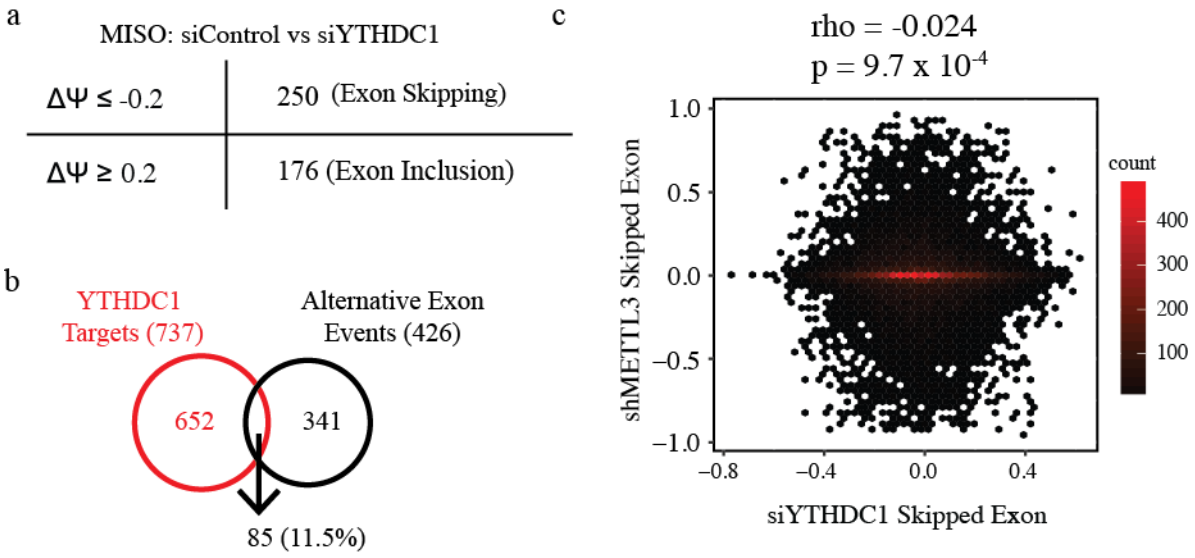

Figure 3.11 YTHDC1-dependent alternative splicing

(a) Analysis of alternative exon usage by MISO (b) Comparison of YTHDC1 targets with transcripts that undergo exon skipping or exon inclusion upon knockdown of YTHDC1 (c) Correlation of YTHDC1-dependent skipped exons with METTL3-dependent exons[17].

We further utilized the MISO software to generate schematics of alternative splicing events in a low-throughput manner. Analysis of reads for two prominent YTHDC1 targets further confirmed that these transcripts do not undergo changes in splicing upon knockdown of YTHDC1 (**Figure 3.12a,b**).

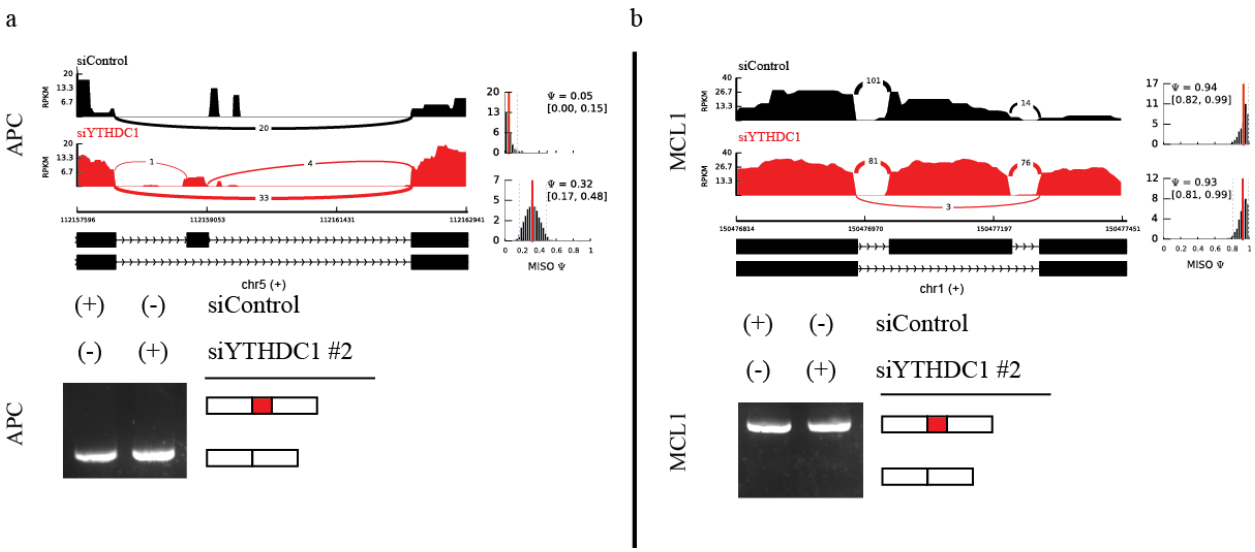

Figure 3.12 Sashimi plots and RTPCR for YTHDC1 targets APC and MCL1

Sashimi plots for (a) APC and (b) MCL1 show that exon usage is not significantly perturbed upon knockdown of YTHDC1 (red plot). RTPCR of the relevant exons confirms the presence of the same dominant isoform under control and knockdown conditions.

Alternative splicing and mRNA export are linked by several mechanisms, most of them involving machinery that participates in both processes. Although we observed limited alternative splicing upon knockdown of YTHDC1, we hypothesized that nuclear trafficking and transport to the cytoplasm requires YTHDC1. We analyzed several YTHDC1 targets (**Figure 3.13a**), including APC and MCL1 using a metabolic labeling experiment. In short, we labeled nascent RNA using 4-ethynyl-urudine for four hours, at which time we isolated nuclear and cytoplasmic RNA. Nascent RNA was enriched, reverse transcribed, and used for RT-PCR (**Figure 3.13b**).

a

Gene	CLIP Peaks	RIP Enrichment
APC	1 (Last Exon)	4.45 ± 0.45
MCL1	4 (Exon 3, 3' UTR (3))	3.85 ± 1.85
SOX12	1 (3' UTR)	2.92 ± 0.80

b

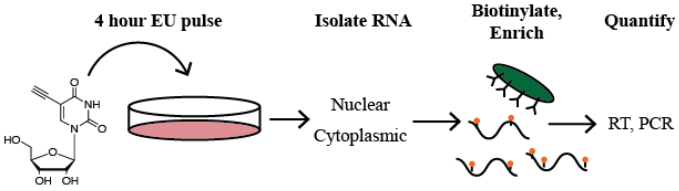

Figure 3.13 Schematic of nascent RNA labeling

We consistently observe that nascent RNA of YTHDC1 targets accumulate within the nucleus over these four hours, and fail to populate the cytoplasmic pool of RNA compared to the control

sample (**Figure 3.14**). Of note, primers were designed to hybridize specifically to exon-exon junction regions. Therefore this experiment gives information about the trafficking of mature mRNA species.

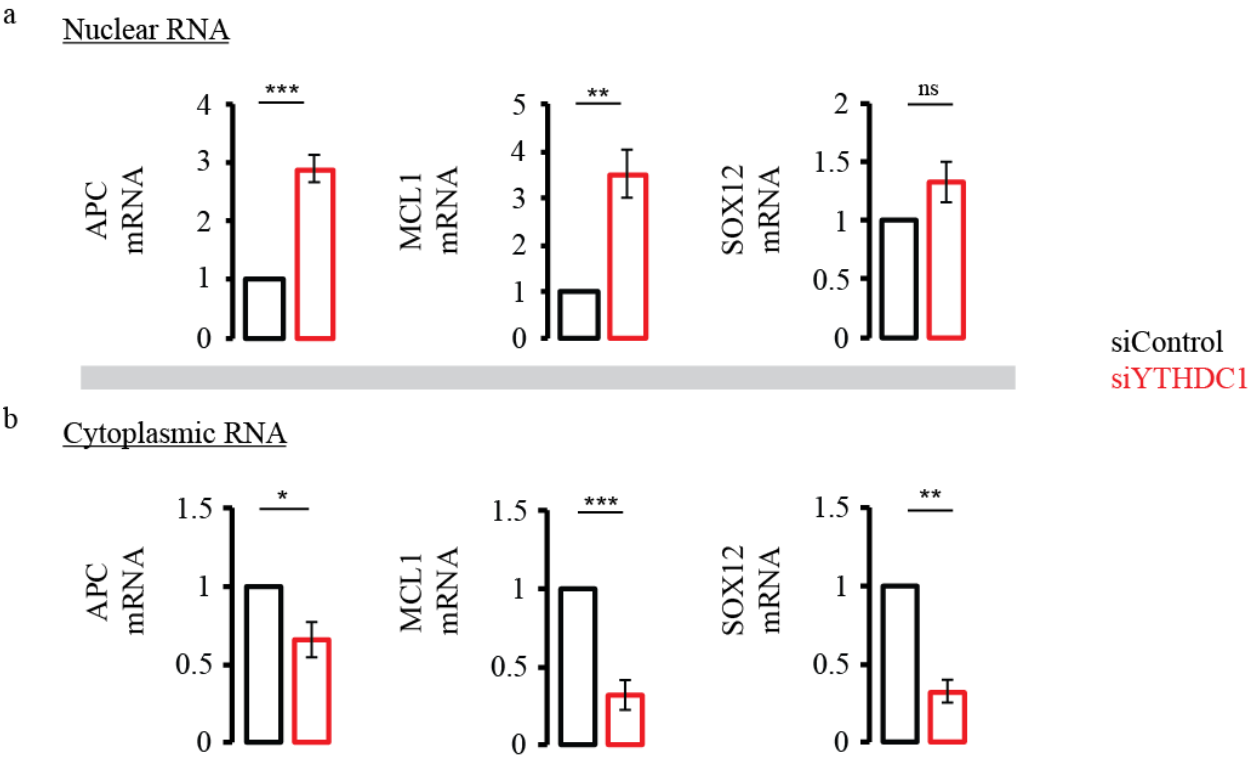

Figure 3.14 Nascent RNA labeling of YTHDC1 targets

(a,b) RT-qPCR of nascent mRNA. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates), * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance.

In contrast, non-target mRNAs of YTHDC1 did not show a significant defect in nuclear to cytoplasmic trafficking (**Figure 3.15**).

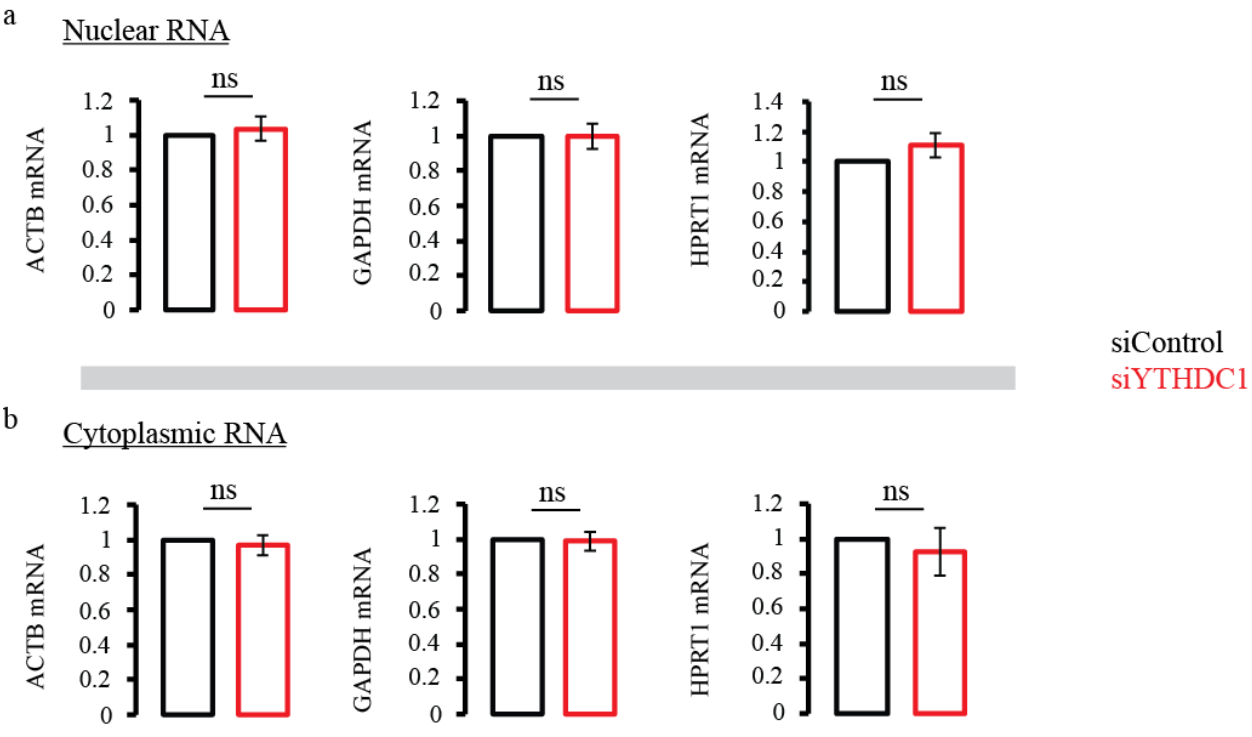

Figure 3.15 Nascent RNA labeling of YTHDC1 non-targets

(a,b) RT-qPCR of nascent mRNA. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates), * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance.

We next asked if the nuclear accumulation of YTHDC1 target APC was due to increased nuclear half-life time in the absence of YTHDC1, as we first observed with m^6A levels following Actinomycin D treatment. Indeed, APC mRNA shows a reduced rate of nuclear clearance upon knockdown of YTHDC1 (**Figure 3.16**)

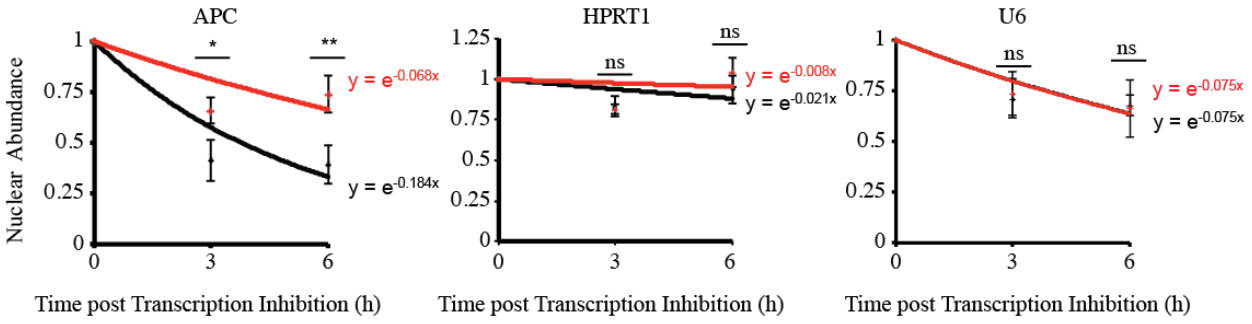

Figure 3.16 Nuclear half-life of YTHDC1 target APC and non-targets HPRT1, U6

Figure 3.16 continued

RT-qPCR of nuclear RNA following treatment with Actinomycin D. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates), * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance. Curves fit to exponential decay.

In order to directly assess the effect of YTHDC1 binding, we constructed a reporter system to measure both RNA localization and protein production of a single transcript (**Figure 3.17a**). YTHDC1 consists of a central YTH domain flanked by regions of low complexity in both the N- and C-termini, from which we derived a series of effector constructs to bind to our Firefly luciferase mRNA (**Figure 3.17b**).

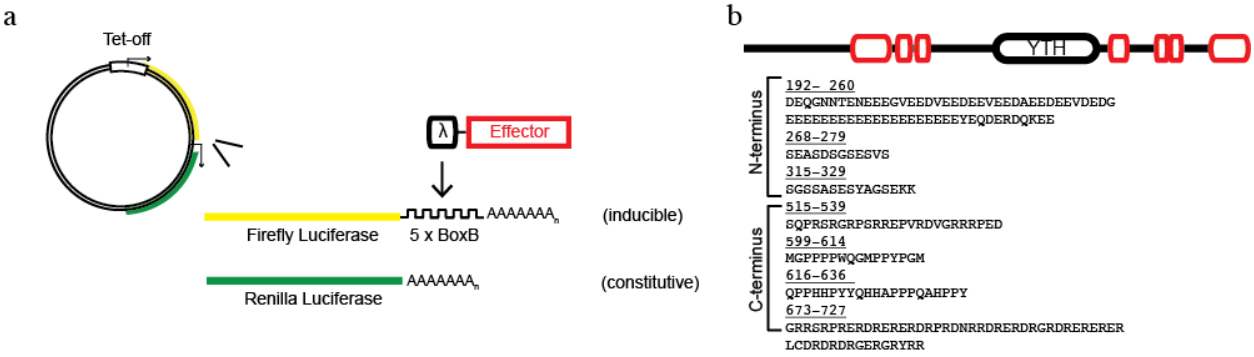

Figure 3.17 Tethered reporter assay and domain architecture of YTHDC1 effectors

(a) Schematic representation of the YTHDC1 reporter assay. Firefly luciferase is under the control of tetracycline, while Renilla luciferase is constitutively expressed. (b) Domain architecture of YTHDC1. Red areas are regions of predicted low complexity[36].

Tethering of the full-length protein to a luciferase reporter increases luciferase levels of the inducible mRNA over the experimental time course (**Figure 3.18a**), and favors cytoplasmic redistribution of the reporter transcript (**Figure 3.18b**). The C-terminus, which contains several alternating SR sequences reminiscent of SR-family splicing and adaptor proteins[22], was alone capable of increasing luciferase signal and re-localizing the Firefly luciferase mRNA (**Figure 3.18c,d**), suggesting that this region is sufficient to promote nuclear to cytoplasmic transport of mRNA. Interestingly, both of these constructs are strictly nuclear in our assay (**Figure 3.18e**).

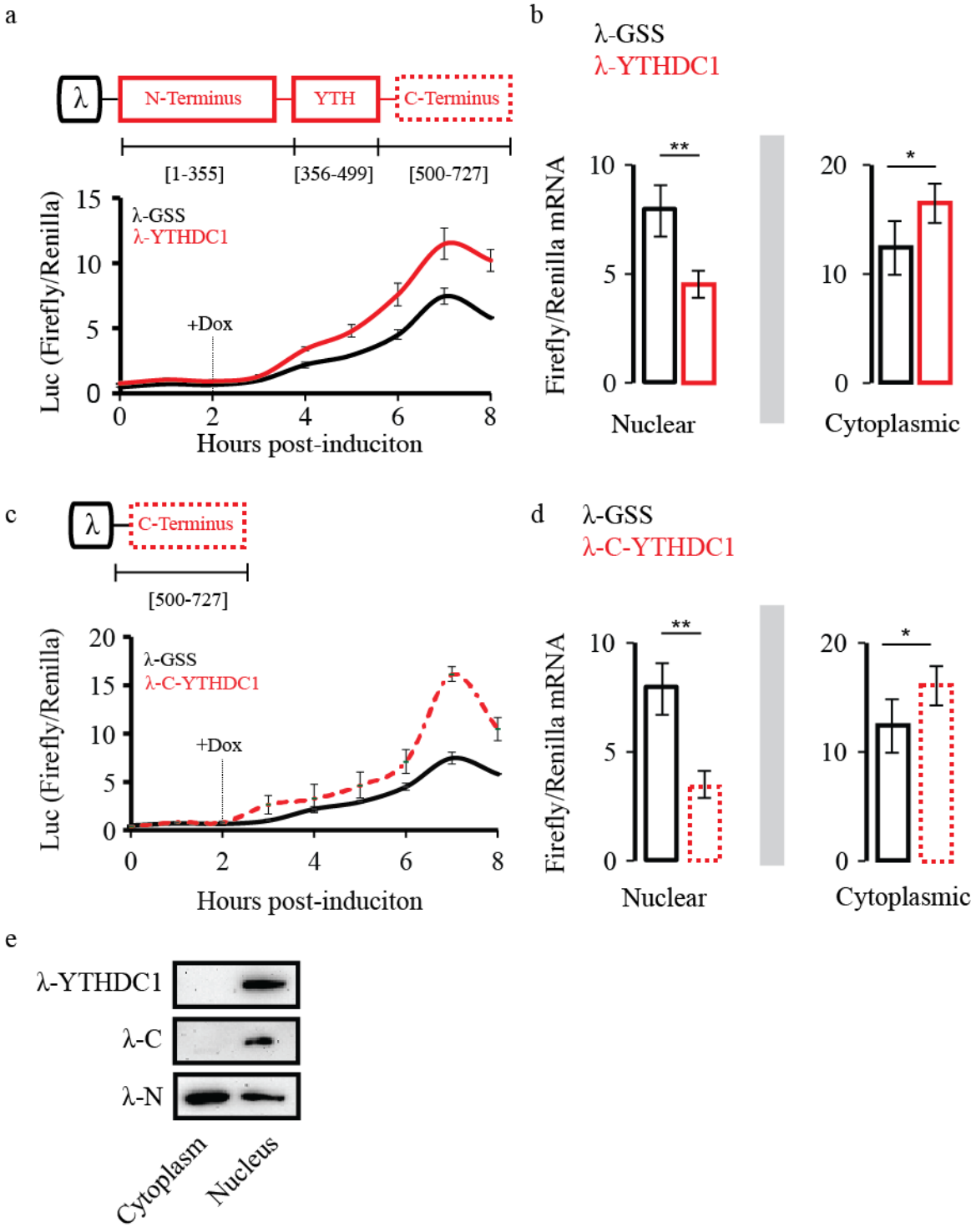

Figure 3.18 Tethering of full-length and C-terminus of YTHDC1

(a) Translation of reporter mRNA in response to binding by the full-length YTHDC1. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates). (b) Nuclear and cytoplasmic RT-qPCR of reporter mRNA 4 hours after induction of reporter transcript. Error bars represent mean ± standard deviation, n = 4, (2 biological replicates x 2

Figure 3.18 continued technical replicates), two sided *t*-test with equal variance. (c) Translation of reporter mRNA in response to binding by the YTHDC1 C-terminus. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates). (d) Nuclear and cytoplasmic RT-qPCR of reporter mRNA 4 hours after induction of reporter transcript. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates), two-sided *t*-test with equal variance. (e) Subcellular localization of effector constructs by anti-FLAG Western blot.

The N-terminal region of YTHDC1 was found in both nuclear and cytoplasmic lysate, but was not capable of promoting mRNA localization to the cytoplasm or increasing luciferase levels (**Figure 3.19**).

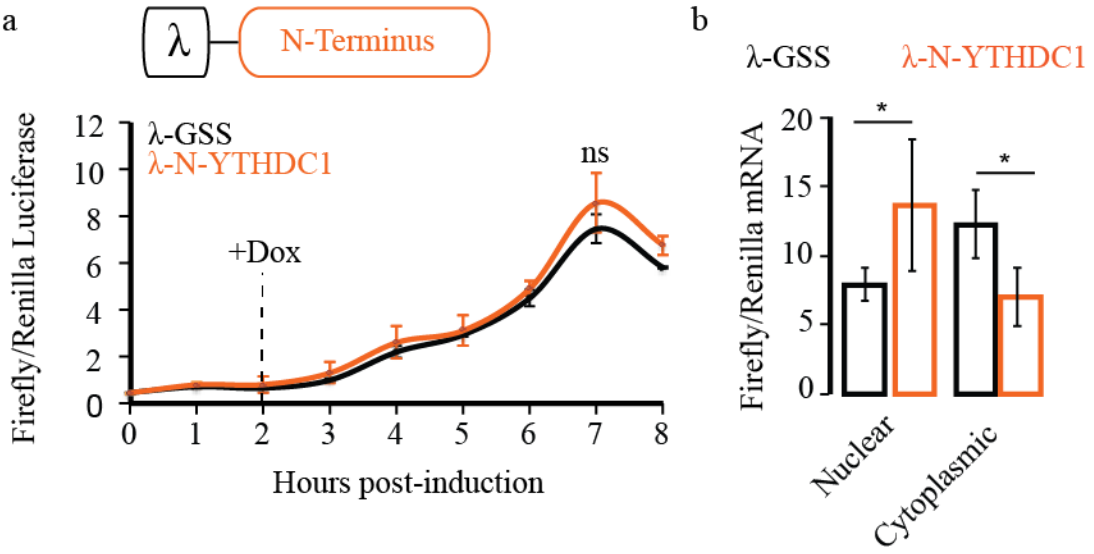

Figure 3.19 Tethering of the YTHDC1 N-terminus

(a) Translation of reporter mRNA in response to binding by the YTHDC1 C-terminus. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates). (b) Nuclear and cytoplasmic RT-qPCR of reporter mRNA 4 hours after induction of reporter transcript. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates), two-sided *t*-test with equal variance.

This result is in agreement with previous reports suggesting that the N-terminus of YTHDC1 plays roles in pre-mRNA splicing as well as XIST-mediated transcription silencing [37,38], and represents a novel role for the C-terminal low-complexity region of YTHDC1, which resembles SR-proteins known to function in mRNA export. We thus conclude that the C-terminal, SR-like region of YTHDC1 functions to accelerate mRNA export, promoting increased levels of cytoplasmic YTHDC1-target transcript and protein levels.

3.3.3 YTHDC1 interacts with pre-mRNA splicing and mRNA export machinery

YTHDC1 interacts with several SR-proteins, which belong to a family of nuclear proteins with multiple roles in pre-mRNA processing[26,39]. Recently, these proteins have emerged as definitive adaptor proteins with roles in mRNA export via their interaction with the bulk mRNA export receptor NXF1[40–43].

We confirmed the RNA-independent interaction between YTHDC1 and several members of the SR-protein family **(Figure 3.20a**). We knocked down each of these pre-mRNA processing factors in order to determine if any of them play a role on the trafficking of m^6A methylated mRNA. We found that both SRSF3 and SRSF10 show an accumulation of m^6A in nuclear mRNA when knocked down (**Figure 3.20b**). Due to the large nuclear accumulation upon SRSF3 knockdown, and its string interaction with NXF1[43], we chose to pursue this protein as a potential adaptor for YTHDC1-mediated mRNA export. We confirmed that indeed SRSF3, but not YTHDC1, interacts directly with NXF1, implying that the terminal protein-protein interaction in this pathway of mRNA export occurs without YTHDC1 (**Figure 3.20c**). The high shuttling rate of SRSF3 in HeLa cells also suggests that this protein mediated the terminal interaction with NXF1 to facilitate export of m^6A methylated mRNAs[44].

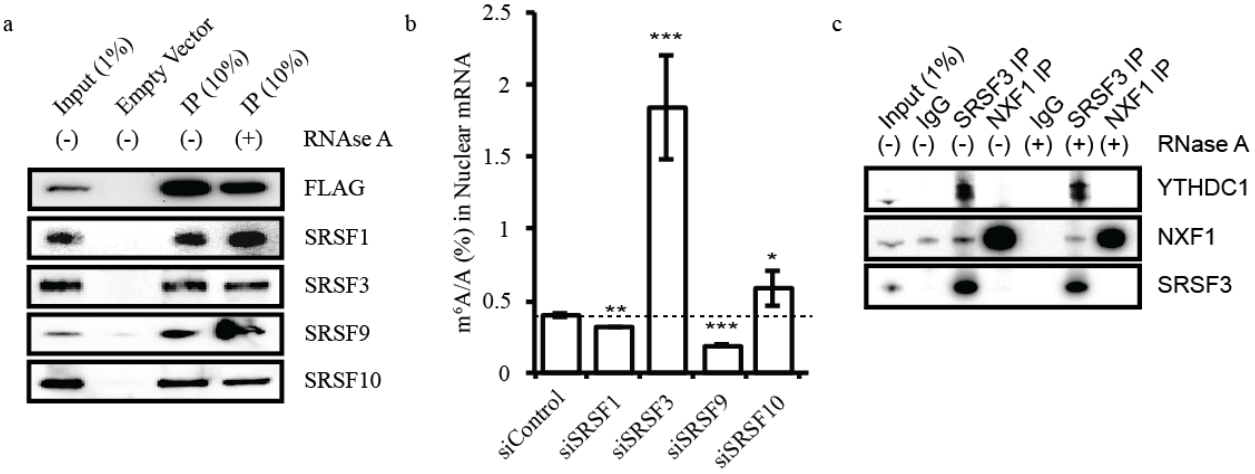

Figure 3.20 Interaction between YTHDC1 and SR-family proteins, nuclear m^6A abundance upon SR-protein knockdown

Figure 3.20 Continued

(a) Co-IP of FLAG-tagged YTHDC1 (b) Nuclear m^6A/A levels upon knockdown of SR-family proteins. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates), * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance.

The function of SR-proteins is tuned by a myriad of post-translational modifications, including methylation, acetylation, and phosphorylation. Phosphorylated SR-proteins recruit components of the spliceosome, and stably interact with the catalytic machinery. They require dephosphorylation for release and recruitment of the mRNA export receptor[45]. We sought to determine of YTHDC1 interacts with SRSF3 in the context of splicing or mRNA export by probing for phosphorylation in SRSF3 that is immunoprecipitated by YTHDC1. IP of endogenous YTHDC1 is able to pull down SRSF3 as expected. However, we did not observe phosphorylation at the expected molecular weight of SRSF3 (**Figure 3.21a**). As the C-terminus of YTHDC1 mediated export of reporter mRNA in our tethering assay, we hypothesized that this protein fragment interacts directly with SRSF3 in its dephosphorylated form. IP of this fragment was capable of pulling down the adaptor protein as expected (**Figure 3.21b**), while the N-terminal did not show any interaction with SRSF3 (**Figure 3.21c**).

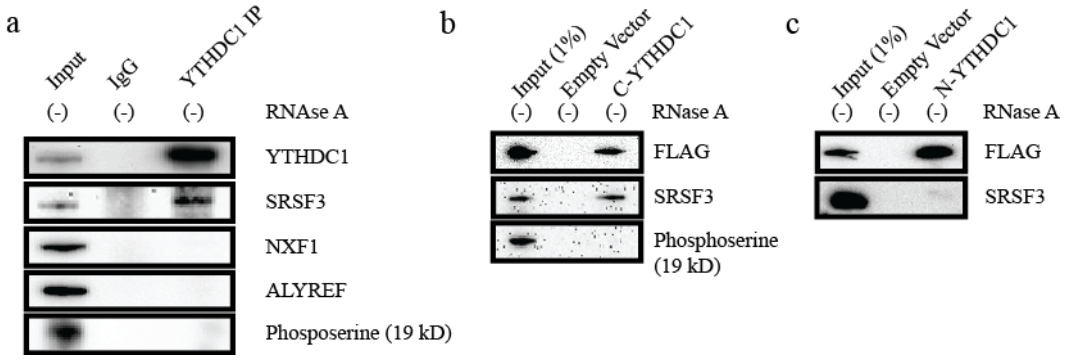

Figure 3.21 Interaction between YTHDC1 and an export-competent SRSF3

(a) Endogenous YTHDC1 interacts with SRSF3 in its dephosphorylated form, but does not directly interact with NXF1 or ALYREF. SRSF3 interacts with the C-terminus of YTHDC1 (panel b), but not the N-terminus (panel c) of YTHDC1.

Although YTHDC1 has been shown to confer m^{6}A specificity in RNA binding[46], we examined the propertied of SRSF3 to recruit methylated mRNAs. *In vivo*, SRSF3 enriched m^{6}A in mRNA, as does the export receptor NXF1 (**Figure 3.22a**). This may be the result of transient protein-protein interactions with YTHDC1 rather than properties of the SRSF3 RRM. Therefore, we cloned and expressed the RRM of SRSF3 for use in an *in vitro* IP with HeLa mRNA (**Figure 3.22b**). Analysis of bound RNA by LC-MS/MS showed no enrichment *in vitro* (**Figure 3.22c**), suggesting that SRSF3 associates with methylated mRNAs through interactions with YTHDC1 or other components of the m^{6}A regulatory machinery such as METTL3 or METTL14[47].

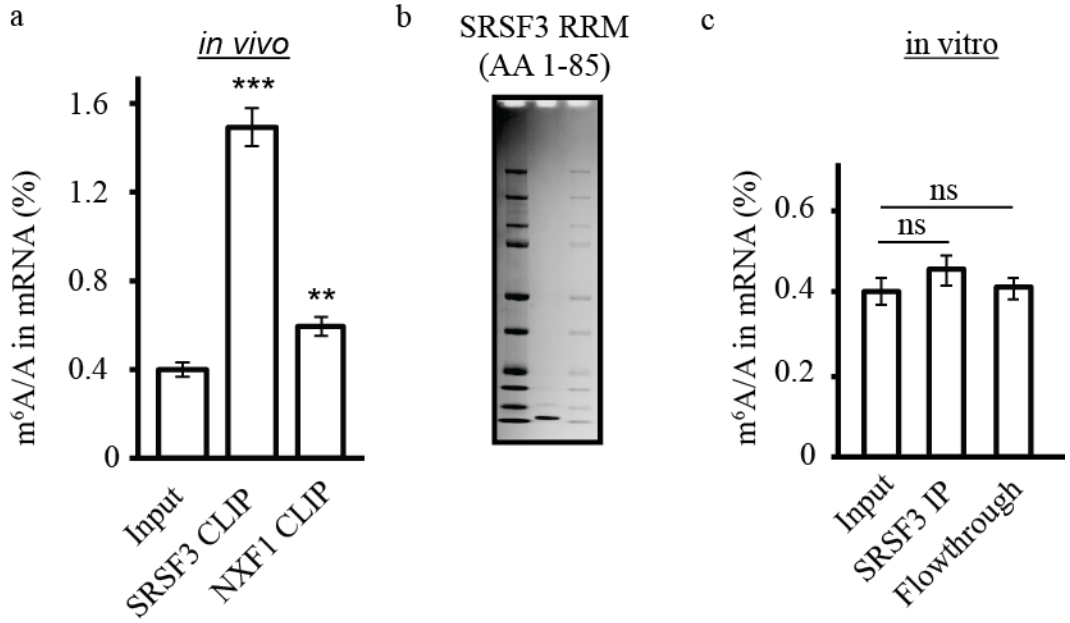

Figure 3.22 SRSF3 specificity for m^{6}A in RNA

(a) CLIP of endogenous SRSF3 and NXF1 from HeLa cells enriches m^{6}A in mRNA (b) SDS PAGE of recombinant SRSF3 RRM (amino acids 1-85) (c) *In vitro* IP of the SRSF3 RRM. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates), * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, two-sided *t*-test with equal variance.

Selectivity for m^{6}A methylation in mRNA export is likely mediated by YTHDC1 binding within the nucleus. This model suggests that initial RNA binding occurs by YTHDC1, followed by delivery to both SRSF3 and NXF1. In order to test this, we analyzed RNA binding of both SRSF3 and NXF1 under control and YTHDC1 knockdown conditions. Cross-linking and immunoprecipitation of endogenous SRSF3 shows enrichment under control conditions, which is

partially diminished by knockdown of YTHDC1 (**Figure 3.23a**). Sequencing of mRNA associated with the SRSF3 also shows reduced enrichment upon siRNA treatment against *YTHDC1*, suggesting that YTHDC1 contributes to the RNA binding capacity of SRSF3 (**Figure 3.23b**). Under the same conditions we observe a complete loss of m^6A enrichment for NXF1 (**Figure 3.23c**), as well as lack of target enrichment of the export receptor (**Figure 3.23d**).

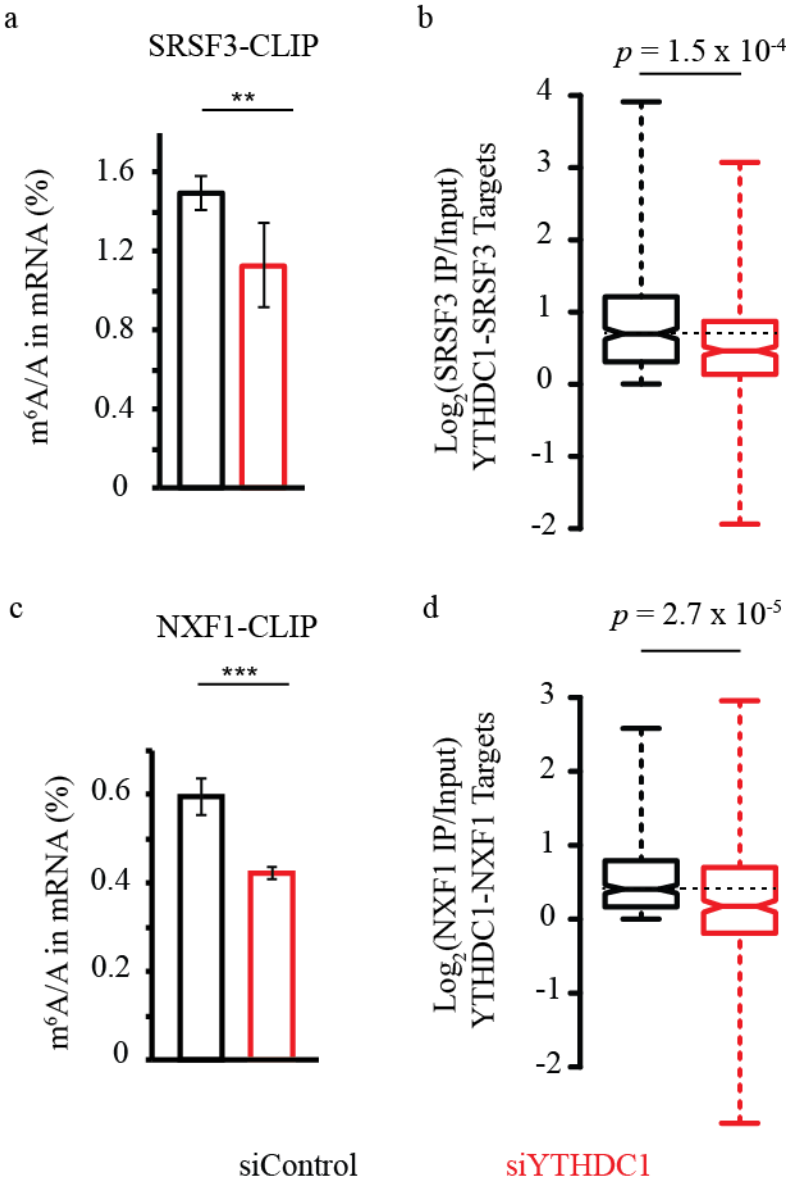

Figure 3.23 RNA binding of SRSF3 and NXF1 as a function of YTHDC1

(a) CLIP LC-MS/MS of mRNA bound by SRSF3. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates), * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$,

Figure 3.23 continued two-sided *t*-test with equal variance. (b) RIP-seq enrichment of YTHDC1-SRSF3 co-targets. *P*-values calculated using Welch's *T*-test (c) CLIP LC-MS/MS and (d) RIP-seq of NXF1.

These results predict that subcellular abundance of YTHDC1 targets should depend on their association with SRSF3 for delivery to NXF1. We therefore segregated YTHDC1 targets into those that are enriched by SRSF3 in RIP-seq versus those that show no enrichment (**Figure 3.24a**). Upon knockdown of YTHDC1, only those targets that are bound by SRSF3 exhibit a shift in subcellular distribution (group IV), while YTHDC1 targets that are not bound by SRSF3 are relatively unaffected (group II) (**Figure 3.24b,c**).

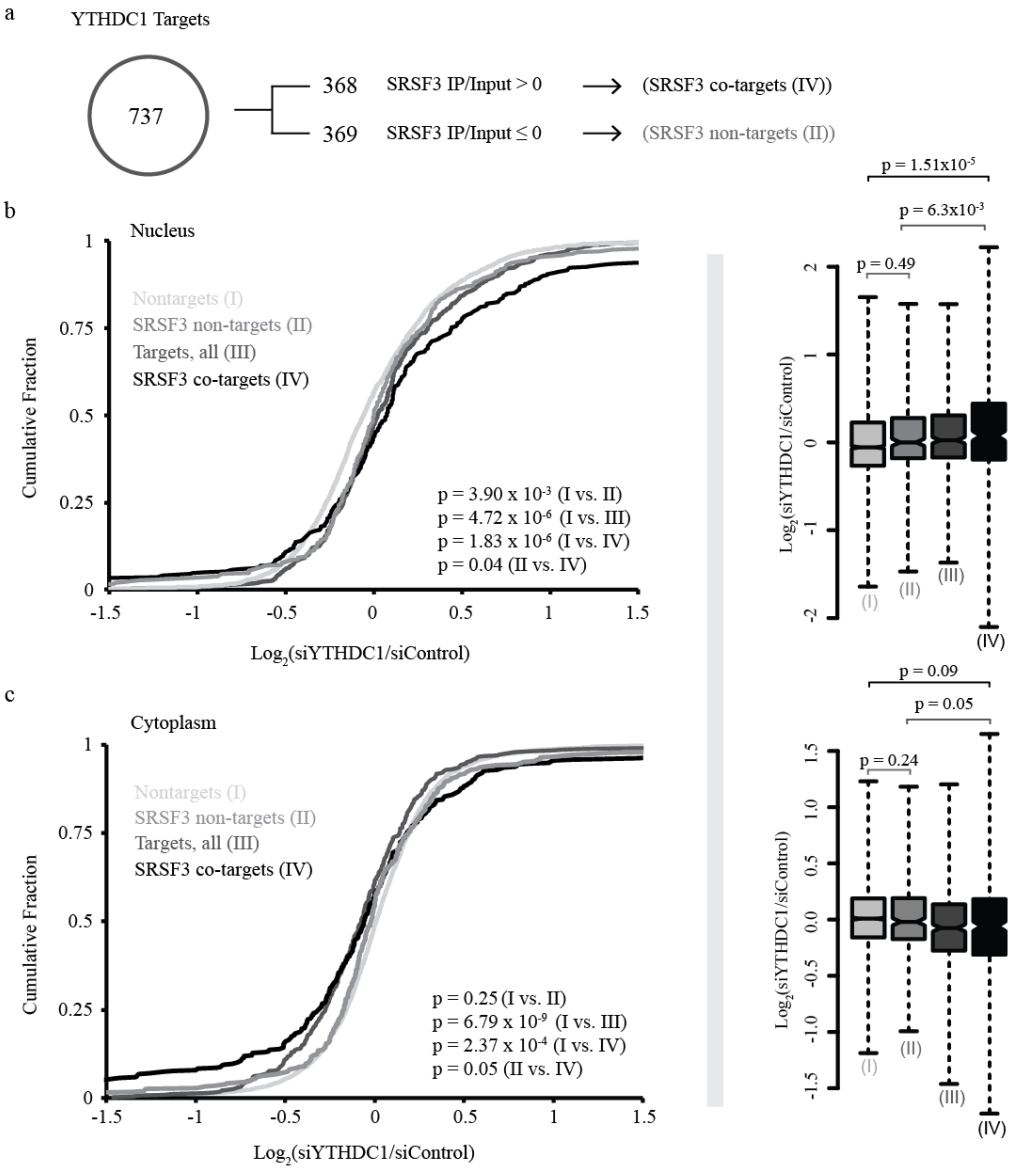

Figure 3.24 Nuclear export of YTHDC1 targets depends on association with SRSF3

Figure 3.24 continued (a) Definition of YTHDC1 SRSF3 co-targets. Of 737 YTHDC1 targets, 368 are enriched by SRSF3 (IP/Input > 0 in two replicates of RIP-seq) and termed YTHDC1-SRSF3 co-targets (Group IV). 369 are bound by YTHDC1 but not enriched by SRSF3 and termed SRSF3 non-targets (Group II). (b,c) Subcellular abundance of transcripts based on binding by YTHDC1 and/or SRSF3 following siControl or siYTHDC1 treatment. Cumulative fractions: Mann-Whitney-Wilcoxon Test. Box plot: Welch's *T*-Test. Whiskers represent three times the interquartile range.

In light of these results, we propose a model in which nuclear mRNA bearing m^6A is selectively bound by YTHDC1. Through protein-protein interaction with SRSF3, these mRNAs are incorporated into export-competent mRNPs and delivered to the cytoplasm via NXF1 (**Figure 3.25**).

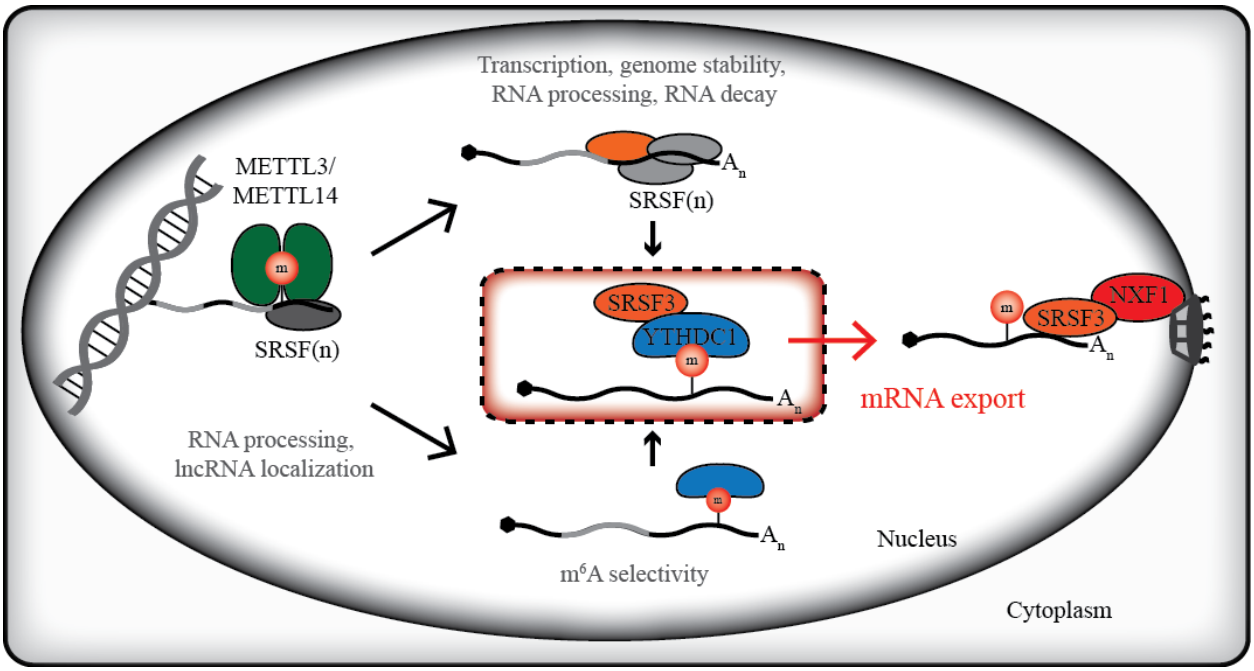

Figure 3.25 Model for m^6A-dependent mRNA export

The core methyltransferase complex consisting of catalytic subunits METTL3 and METTL14 deposits m^6A on nascent mRNA. METTL14 can recruit components of RNA processing machinery, including splicing and adaptor proteins of the SR-family (shown as SRSF(n)). In addition to other nuclear roles, YTHDC1 enriches m^6A methylated mRNA, coupling the processing of these transcripts with SR-proteins for both splicing and export via direct interaction

with SRSF3. YHDC1 delivers mRNAs to export-competent mRNPs, and eventually to NXF1 via interaction between SRSF3 and NXF1.

3.3.4 m^{6}A in pre-mRNA and introns

m^{6}A directs pre-RNA processing reactions, namely the intron removal that occurs during splicing. This requires that methylation occurs prior to pre-mRNA splicing, and is in fact a co-transcriptional event for many transcripts[48,49]. Despite this, m^{6}A methylation in pre-mRNA, and in particular introns has not been well characterized, although the majority of METTL3 and METTL14 binding sites reside within introns[50,51]. It is believed that the majority of m^{6}A methylations is conserved throughout the splicing reaction, which suggests that while co-transcriptional methylation is somehow directed to exons. Given the vastness of intronic RNA sequences, we hypothesized that introns contain a significant amount of m^{6}A methylation, and that this methylation may serve roles in both transcription and pre-mRNA splicing.

We first sought to quantify m^{6}A in pre-mRNA by isolating a population of RNAs that have yet to be spliced. After exploring pre-mRNA splicing inhibitors such as Isoginkgetin[52], we relied simply on rRNA depletion of nuclear RNA. Sequencing of these RNAs revealed that the majority of reads mapped either to introns or inton-exon junctions (**Figure 3.26a**). LC-MS/MS of these RNAs revealed a methylation ratio of m^{6}A/A of nearly 0.25%, indicating that the presence of introns does not dilute the m^{6}A/A ratio significantly (**Figure 3.26b**).

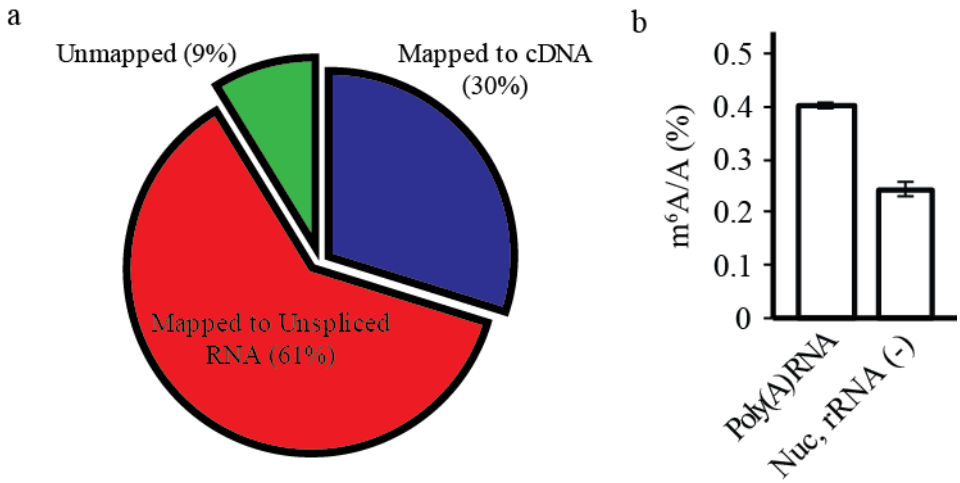

Figure 3.26 Evidence for m^{6}A methylation in introns of pre-mRNA

Figure 3.26 continued (a) Mapping of nuclear, rRNA-depleted RNA (b) LC-MS/MS of poly(A) and nuclear, rRNA-depleted RNA. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates).

These data suggest that while m^6A in introns sequences is relatively depleted compared to exons and mature mRNA, the relative concentration within introns is on the order of 0.5 times that of exons.

To obtain a map of m^6A in pre-mRNA, we performed m^6A-sequencing of this RNA population. In order to obtain sufficient coverage of intron and intergenic regions, we sequenced these libraries roughly 10-times deeper than traditional RNA-sequencing requires. To analyze m^6A peaks in introns, we utilized a mapping strategy that required mapping first to the transcriptome, followed by mapping to the human genome. This strategy allowed us to analyze peaks from exons and introns independently. In two biological replicates we found more than 25,000 m^6A peaks, over 23,000 of which were common to our two samples (**Figure 3.27a**). Unlike our exon peaks in these samples which showed enrichment for the expected GGAC motif (data not shown), we find a GGAUAG hexanucleotide motif enriched in our intron peaks (**Figure 3.27b**). Since this motif itself contains two potential sites of modification, it is not possible at this point to determine the sequence context for $m^6{}_A$ methylation in this experiment.'

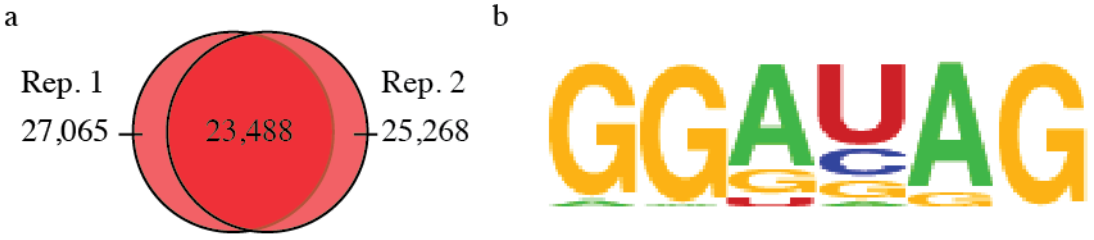

Figure 3.27 m^6A-seq of pre-mRNA

(a) Comparison of replicates from pre-mRNA m^6A-seq. 23,488 regions were common to both data sets. (b) Most enriched motif for m^6A sites within introns.

If m^6A methylation is co-transcriptional, transcript selectivity may be driven in part by recruitment of the methyltransferase to the active polymerase. In this case, we would expect that genes with intron methylation are the same genes as those with exon methylation, perhaps due to a promiscuous enzyme that is granted access to the nascent sequence. However, of the 4,538 unique transcripts with reproducible m^6A methylation in their introns, only 951 have known methylation within their mature transcript (**Figure 3.28a**). We next asked if intron m^6A peaks

displayed a biased distribution. While m^{6}A in mature mRNA favors last exons and the 3' UTR, intron methylation is biased towards the 5' introns of a nascent mRNA (**Figure 3.28b**). This may result from the extended length of the first intron in mammals, or an alternative mechanism of methyltransferase recruitment. Potential mechanisms for dichotomous methylation are discussed in the last section of this chapter.

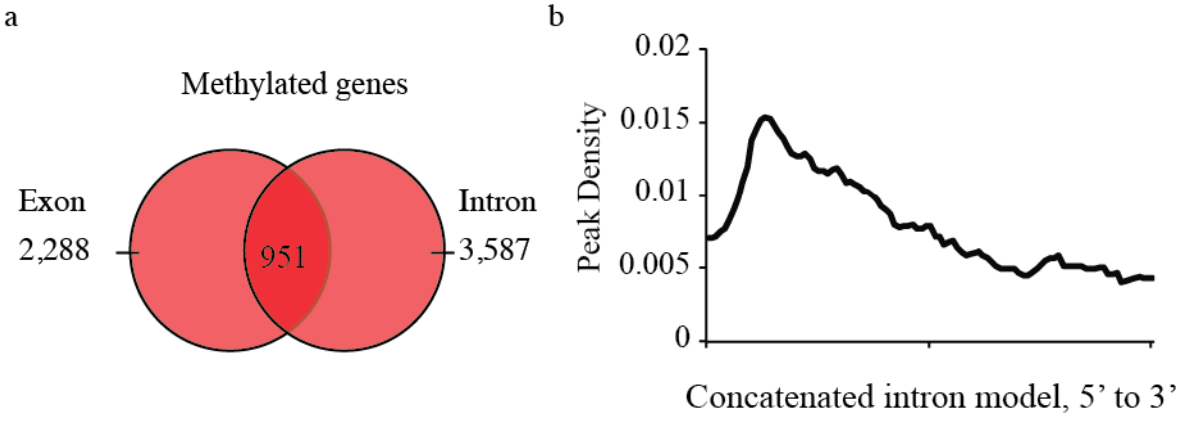

Figure 3.28 Comparison of intron and exon m^{6}A methylation

(a) Nascent transcripts with m^{6}A in introns are largely distinct from mature transcripts with m^{6}A methylation (b) Intron m^{6}A favors 5' introns.

3.4 Conclusions and discussion: unique roles for YTHDC1 in nuclear RNA

Currently, the function of m^{6}A in the nucleus is heterogeneous. A number of direct and indirect reader proteins play roles in alternative splice site selection, including YTHDC1 as well as members of the SR and HNRNP. Previous roles for SR and HNRNP proteins in alternative splicing have proven antagonistic[53,54], further clouding our understanding of how these proteins may combine to mediate the role of m^{6}A in the nucleus.

In our studies, we identify a unique role for m^{6}A methylation in facilitating nuclear export of mRNA. YTHDC1, the only nuclear member of the YTH protein family, facilitates the incorporation of methylated mRNAs into canonical pathways for mRNA transport. This mechanism depends on protein-protein interaction with SRSF3, a potent adaptor protein to the

mRNA export receptor NXF1. How multifunctional adaptor protein of the SR-family function in both m^6A-dependent and m^6A-independent pathways remains unknown. However, the tunable nature of these proteins suggests that their interaction with methylation machinery is highly context dependent.

Currently, we do not understand if and how several unique pre-mRNA processing mechanisms may be united by post-transcriptional modifications. It is likely that alternative pre-mRNA processing is highly dependent on cellular or developmental context, and we are beginning to understand how m^6A methylation is utilized during these processes. Examples of this include sex determination in the *Drosophila* embryo, in which m^6A directs requisite splicing patters[19,20].

The biochemical coupling of all nuclear processing reactions nearly guarantees that m^6A affects genetic output in every stage of the mRNA lifecycle. We have seen how cytoplasmic events are modulated by m^6A at the hands of YTHDF1 and YTHDF2. Within the nucleus, YTHDC1, HNRNPA2B1, and HNRNPC facilitate processing and export of methylated transcripts. We have yet to observe, however, how changes in downstream processing at the hands of RNA methylation feed back to the transcriptional apparatus. In this scenario, it may be most useful to investigate the earliest stages of m^6A deposition – methylated introns proximal to the 5' end of nascent mRNA. With transcriptional control in mind, the potential relationship between post-transcriptional RNA modification and transcriptional output will likely be topic of future studies.

3.5 Methods

3.5.1 Plasmid construction

λ peptide sequence MDAQTRRRERRAEKQAQWKAAN was fused to the N-terminus of YTHDC1-Flag[46] by subcloning into pcDNA3.0. YTHDC1 C-Terminal and YTHDC1 N-Terminal regions were similarly subcloned into pcDNA3.0 with C-terminal Flag tags.

λ-YTHDC1 (AA 1-727, EcoRI, XhoI)
F:
CAGCTTGAATTCATGGACGCACAAACACGACGACGTGAGCGTCGCGCTGAGAAACA
AGCTCAATGGAA AGCTGCAAACGGTGGTAGCGCGGCTGACAGTC

R:
GCATGCCTCGAGTTACTTGTCATCGTCATCCTTGTAATCTCTCCCCCCTCTTCTATATC
GACCTCTCC

λ-N-Terminus (AA 1-355, EcoRI, XhoI)
F:
CAGCTTGAATTCATGGACGCACAAACACGACGACGTGAGCGTCGCGCTGAGAAACA
AGCTCAATGGAA AGCTGCAAACGGTGGTAGCGCGGCTGACAGTC
R:
GCATGCCTCGAGTTACTTGTCATCGTCATCCTTGTAATCTCTCCCCCCTGCATCTTGA
AGCACATATTTG AG

λ-C-Terminus (AA 500-727, EcoRI, XhoI)
F:
CAGCTTGAATTCATGGACGCACAAACACGACGACGTGAGCGTCGCGCTGAGAAACA
AGCTCAATGGA AAGCTGCAAACGGTGGTAGCTTGTATCAGGTCATTCATAAAATGC
R:
GCATGCCTCGAGTTACTTGTCATCGTCATCCTTGTAATCTCTCCCCCCTCTTCTATATC
GACCTCTCC

λ-YTHDC1-W377A
Point mutations were introduced using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) according to manufacturer's protocol.

F: GGATCTCCTATACACGCGGTGCTTCCAGCAGGA

R: TCCTGCTGGAAGCACCGCGTGTATAGGAGATCC

Construction of the λ-control peptide and pmirGlo-Ptight-5BoxB was previously reported[55].

His-SRSF3 (AA1-85) was subcloned from pcDNA3.2 V5-DEST 3XFlag (Addgene Plasmid #46736)[56] into pet28a(+) using the following primers:

F:
ATCGCCATGGATGCATCATCATCATCATCATGGCAGCAGCCATCGTGATTCCTGTCC
ATTG (NcoI)
R: GCTACTCGAGTTATTTTTCACCATTCGACAGTTCCA (XhoI)

Plasmids with high purity for mammalian cell transfection were prepared using Highspeed Maxiprep or Midiprep Kits (Qiagen).

3.5.2 Immunofluorescence

HeLa cells were seeded in 8 chamber microscope slides (Nunc 155409) and treated with siRNA as previously described. After 72 h, cells were washed with PBS, then fixed with 4% paraformaldehyde (Sigma) in PBST (PBT + 0.05% Tween-20), freshly prepared by heating at 65° C until clear, at RT for 15 minutes. Fixative was removed and chilled (-20° C) methanol was added dropwise and incubated at RT for 15 minutes. Cells were washed with PBS and blocked with 10% FBS in PBST at RT for 1 h. Primary antibody incubation was performed in 10% FBS in PBST at 4 °C overnight. After 4 x 5 minute washes with PBST secondary antibody incubation was performed at RT for 1 h. After 4 x 5 minute washes with PBST cells were treated with SlowFade Gold antifade reagent with DAPI (Molecular Probes) and imaged using a Leica TCS SP5 II STED laser scanning confocal microscope (Leica Microsystems, Inc.).

3.5.3 Cellular Fractionation

HeLa cells were fractionated using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific) according to manufacturer protocol with the following modifications: following cytoplasmic isolation, nuclei were washed extensively (4 x 200 μL) with PBS.

3.5.4 RNA isolation

RNA from subcellular compartments was collected using Directzol RNA miniprep (Zymo Research) and treated with DNaseI prior to elution in water.

RNA was further purified for RNA-sequencing by RiboMinus Eukaryotic Kit v2 (Ambion) unless otherwise noted, and concentrated using RNA Clean and Concentrator (Zymo Research) for RNA ≥ 200 nt. Sequencing libraries were prepared using the TruSeq Stranded mRNA LT Library Construction Kit (Illumina) according to the manufacturers protocol and sequenced with 50 bp single end reads.

3.5.5 Western blotting protocol

Protein electrophoresis was performed with NuPAGE Novex 4-12% Bis-Tris Protein Gels (Novex) in MOPS SDS running buffer. Blots were transferred to a nitrocellulose membrane and blocked with 5% milk in TBST (0.05% Tween-20) for 30 minutes at room temperature. Primary antibodies were incubated overnight in 5% milk in TBST at 4°C over night. Secondary antibodies were incubated at room temperature for one hour, washed and developed using chemiluminescence with SuperSignal West Pico Luminol/Enhancer solution (Thermo Fisher) in the FluorChem M system (ProteinSimple).

For blots probing for phosphorylation, membranes were blocked using 5% Bovine Serum Albumin (BSA) (EMD Millipore), and primary and secondary dilutions were prepared using 5% BSA in TBST.

Primary Antibodies (Target, source, use, dilution, supplier)

FLAG: mouse monoclonal [M2] – HRP conjugated, Western Blot, 1:10,000, Sigma (A8592)
GAPDH: Goat polyclonal – HRP conjugated, Western Blot, 1:10,000, GenScript (A00192-100)
H3: Rabbit polyclonal, Western Blot, 1:3,000, Abcam (ab1791)

YTHDC1: Rabbit polyclonal, Western Blot, IP, 1:1,000, Abcam (ab122340)
SRSF1: Rabbit monoclonal EPR8239, Western Blot, 1:10,000, Abcam (ab129108)
SRSF3: Rabbit polyclonal, Western Blot, 1:1,000, Abcam (ab73891)
SRSF3 (SRp20): Rabbit polyclonal, IP, MBL (Code No. RN080PW)
SRSF9: Rabbit polyclonal, Western Blot, 1:1,000, Abcam (ab155484)
SRSF10 (FUSIP1): Mouse monoclonal, Western Blot, 1:1000, Abcam (ab77209)
NXF1: Rabbit monoclonal [EPR8009], Western Blot, IP 1:5,000, Abcam (ab129160)
YTHDF1: Rabbit polyclonal, Western Blot, 1:1,000, Abcam (ab99080)
HNRNPC: Rabbit monoclonal [EPNCIR152], Western Blot, 1:10,000, Abcam (ab133607)
HNRNPA2B1: Rabbit polyclonal, Western Blot, 1:1,000, Abcam (ab31645)
Digoxigenin: Mouse monoclonal [21H8], IF, 1:300, Abcam (ab420)
ALYREF: Rabbit monoclonal [EPR17942], Western Blot, 1:5,000, Abcam (ab202894)
ZCCHC8: Rabbit monoclonal [EPR13612], IP, Western Blot 1:1,000, Abcam (ab181152)
MTR4: Rabbit polyclonal, IP, Abcam (ab70552)
MTR4: Rabbit polyclonal, Western Blot, 1:1,000, Abcam (ab93337)
Phosphoserine: Rabbit polyclonal, Western Blot, 1:100, Abcam (ab9332)

Secondary Antibodies

Rabbit IgG: Goat polyclonal, Western Blot, 1:5,000, Bethyl (A120-101P)
Mouse IgG: Goat polyclonal, Western Blot, 1:5,000, Bethyl (A90-116P)
Mouse IgG: Goat polyclonal – TexasRed, IF, 1:300, Molecular Probes (T-6390)
Rat IgG: Donkey polyclonal – Alexa488, IF, 1:300, Molecular Probes (A-21208)

3.5.6 Northern blotting protocol

RNA from nuclear and cytoplasmic extracts were isolated using Trizol according to manufacturer protocol, then further purified using RNA Clean and Concentrator with DNase treatment (Zymo Research). 300 ng of total RNA was diluted 1:2 in 2X TBE-Urea Sample Buffer (Thermo Fisher Scientific), heated to 65 °C for 10 minutes, and separated on a 6% TBE-Urea gel (180 V, 1 hour) run at 4 °C. RNA was transferred at 4 °C to an Amersham HyBond-N+ membrane in 0.5X TBE Buffer. The membrane was crosslinked using the Stratalinker 2400 (Stratagene) autocrosslink

option. The membrane was blocked in pre-hybridization buffer (10X final concentration Denhardt's Solution (Thermo Fisher), 6X final concentration SSC Buffer (Ambion), 0.1% SDS, 10 µg/mL salmon sperm DNA (Thermo Fisher)) at 42 °C for 2 hours.

DNA probes (10µM stock, 2 uL probe per reaction) were labeled with P^{32} using T4 PNK (Thermo Fisher) and purified using Oligo Clean and Concentrator (Zymo Research). Labeled oligos were incubated with membrane at 42 °C while rotating overnight. Membranes were washed 4 x 1 hour with pre-hybridization buffer at 42 °C, then dried at 80 °C for 30 minutes before radioisotope exposure.

Northern Probe Sequences (5'-3')

tRNA Initiator Methionine:
TGGTAGCAGAGGATGGTTTCGATCCATCGACCTCTGGGTTATGGGCCCAGCACGCTT
CCGCT GCGCCACTCTGCT

U2 snRNA: GAACAGATACTACACTTGATCTTAGCCAA

3.5.7 Actinomycin D Treatment

Actinomycin D (Sigma) was dissolved in DMSO to a final concentration of 5 mg/mL. This stock was diluted 1:1000 (5 µg/mL final concentration) in fresh media and applied to control and YTHDC1 knockdown HeLa cells in culture. Cells were collected by cell lifter at designated time points.

3.5.8 Analysis of High-Throughput Sequencing Data

YTHDC1 PAR-CLIP was performed as previously described[57]. YTHDC1-RIP was performed according to literature procedure[31] using polyA selected RNA and ribosomal RNA depleted RNA as biological replicates. mRNA for subcellular RNA-seq was isolated by ribosomal depletion.

Treatment of raw data: Reads were trimmed using Cutadapt v.1.4.1[58] and aligned to hg19 using TopHat v.2.0.11 [59]. PAR-CLIP peaks were called using PARalyzer v.1.5[60] with default parameters. Binding motifs were determined by HOMER v.4.7[61]. Differential expression between replicate data sets was calculated using Cufflinks v.2.2.1[62]. Overlapping genomic elements were assessed using BedTools v.2.2.1[63] and analyzed using PeakAnnotator v.1.4[64]. Differential splicing was analyzed using MISO v.0.5.3[35]. Gene ontology (GO) analysis was conducted using DAVID Functional Annotation Tools v.6.7[32].

Statistical comparisons were calculated using the R Statistics package.

Raw and processed data files have been deposited in the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) and accessible under GSE74397.

3.5.9 Isoform Analysis by RTPCR

RNA was collected as previously described, including treatment with DNase. Reverse transcription was conducted with SuperScript VILO according to manufacturers protocol, and PCR was conducted using primer sequences designed to span the region in question using Pfusion High Fidelity PCR Master Mix (NEB). Fragments were separated on 3.0% agarose and visualized using ethidium bromide (Sigma).

Primer Sequences:
APC
F: TTCCTTACAAACAGATATGACCAGA
R: TTACCAGAAGTTGCCATGTTG
Product observed: 210 bp
MCL1
F: GAGGAGGACGAGTTGTACCG
R:ACCAGCTCCTACTCCAGCAA
Product observed: 525 bp

3.5.10 PolyA Imaging

HeLa cells were fixed in PFA and MeOH as above. Following methanol treatment, cells were washed with PBST. PBST was removed and RISH positive control DIG probe (PanPath Q152P.9900) (37-mer oligonucleotide complementary to PolyA) was added as a 1:3 dilution in PBST and incubated at 37° C for 2 h. Cells were washed 4 x 5 minutes with PBST at RT and subjected to IF staining protocol. Image quantification was performed using Fiji[65].

3.5.11 Nascent RNA-labeling

Nascent RNA labeling was conducted using the Click-iT Nascent RNA Capture Kit (Molecular Probes) according to the manufacturers protocol. Briefly, HeLa cells treated with siRNA were fed 5-ethynyl-urudine (EU) at a concentration of 200 μM for 4 hours. After 4 hours, RNA from isolated from whole cells (10%) and nuclear and cytoplasmic portions. Click chemistry was performed on RNA, ethanol precipitated and enriched using provided streptavidin beads. cDNA synthesis was performed on the beads using SuperScript VILO (Invitrogen). RTPCR was performed using FastStart Essential DNA Green Master in a LightCycler 96 (Roche).

RT-PCR Primer Sequences:

APC
F: TAGGGGGACTACAGGCCATT
R: TTTAGTTGGGCCACAAGTGC

MCL1
F: CGGACTCAACCTCTACTGTGG
R: CTTGGAAGGCCGTCTCGT

SOX12
F: GCTGAGGAAGGTGAAGAGGA
R: GCGATCATCTCGGTAACCTC

ACTB

F: ACAGAGCCTCGCCTTTGCC

R: GATATCATCATCCATGGTGAGCTGG

GAPDH

F: AGAAGGCTGGGGCTCATTTG

R: AGGGGCCATCCACAGTCTTC

HPRT1

F: TGACACTGGCAAAACAATGCA

R:GGTCCTTTTCACCAGCAAGCT

3.5.12 Luciferase reporter assay

All reporter experiments were conducted in HeLa Tet-off cells (Clontech) in cultured in DMEM with 1% 100 x Pen Strep (Gibco) supplemented with 10% Tet-System approved FBS (Clontech). Cultures were maintained in 200 μg/mL G418 (Takara).

In a 6 well plate, ~$1x10^6$ cells were cultured with 100 ng/mL doxycycline (Dox). Cells were transfected with 50 ng reporter plasmid and 450 ng effector plasmid using Lipofectamine LTX according to manufacturers protocol. The following day cells were washed with PBS, trypsinized and washed thoroughly (4 x 15 mL PBS) to induce luciferase expression. Washed pellets were suspended in 3 mL culture media and seeded in 96-well plates. After 2 hours of induction, Dox was added to a final concentration of 500 ng/mL to halt firefly transcription. Luciferase was quantified using the Dual-Glo Luciferase Assay System (Promega) in a Synergy HTX Multi-Mode Reader (BioTek). Each sample was normalized to Renilla luciferase signal form the same well.

3.5.13 Protein Co-Immunoprecipitation

HeLa cells transfected with empty plasmid or YTHDC1-Flag were washed with PBS two times, then collected using a cell lifter and spun at 2,000 x g for 5 minutes. Pellets were suspended in 2 volumes (compared to pellet) Buffer A (10 mM HEPES, pH 7.5, 1.5 mM $MgCl_2$, 10 mM KCl, 0,5 mM DTT), vortexed briefly, and incubated on ice for 15 minutes. NP-40 was added to a final concentration of 0.25%, vortexed, and left on ice for 5 minutes. Suspensions were spun at 2,000 x g for 3 minutes at 4 °C. The supernatant (cytoplasmic extract) was removed and saved. Nuclei were suspended in 2 volumes Buffer B (20 mM HEPES, pH 7.5, 0.42 M KCl, 4 mM $MgCl_2$, 1 mM EDTA, 0.5 mM DTT, 10% glycerol), vortexed briefly, and incubated on ice for 30 minutes. The suspension was spun at 15,000 x g for 15 minutes at 4° C, and supernatant combined with the cytoplasmic extract. This was incubated on ice for 15 minutes and spun again at 15,000 x g for 15 minutes at 4° C. The supernatant was removed and used as the HeLa lysate.

FLAG IP: Lysate from mock and YTHDC1-transfected HeLa cells was incubated with M2-anti-FLAG beads (Sigma) and rotated at 4° C for two hours. After two hours the beads were washed 4 x 500 uL with Buffer C (wash buffer) (50 mM Tris-HCl, pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.2 mM EDTA, 0.1% NP-40, 10 mM β-ME, 10% glycerol). The beads were resuspended and half were used for Western blot analysis. The remaining half was warmed to 37° C and treated with RNase A (Thermo Fisher) at a final concentration of 1 ug/mL for 30 minutes. The beads were washed 4 x 500 uL wash buffer and used for western blot analysis.

Endogenous IP: Lysate from untreated HeLa cells was treated with rabbit IgG control or primary antibody against SRSF3 or NXF1 (5 ug/IP) and incubated at 4 °C for two hours. After two hours, washed Dynabeads Protein A (Thermo Fisher) were added, and the solution was incubated for an additional hour, followed by washing and RNase treatment for western blotting as described.

SRSF3 and NXF1 RIP-seq in the presence and absence of YTHDC1 was conducted following immunoprecipitation under these conditions and subsequent Ribo(-) treatment of isolated RNA.

3.5.14 Crosslinking immunoprecipitation (CLIP) LC-MS/MS

Crosslinking was performed using Stratalinker 2400 (Stratagene) autocrosslink option on untreated HeLa cells. Protein co-IP was performed as described, and washed beads were resuspended in 100 uL wash buffer. 100 uL 2X Proteinase K buffer (100 mM Tris-HCl, pH 7.5, 200 mM NaCl, 2 mM EDTA, 1% SDS) was added, followed by Proteinase K (Thermo Fisher) to a final concentration of 500 ug/mL and heated for 1 hour at 65 °C. RNA was isolated using TriZol. Recovered RNA was purified using two rounds of poly(A)-selection, digested, and subjected to LC-MS/MS.

3.5.15 m^6A-sequencing

RNA for m^6A-sequencing was prepared y ribosomal depletion of nuclear RNA. 1 μg of RNA was diluted into 100 μL of RNase-free water and sonicated for 30 cycles (30 seconds on, 30 seconds off) at 4 °C (Bioruptor, Diagenode). 5 μL of this RNA was saved as input – the remaining volume was used for m^6A IP.

For m^6A IP, 95 μL fragmented RNA was mixed with 100 μL 5X IP buffer (50 mM Tris-HCl 7.4; 750 mM NaCl, 0.5% NP-40), 12.5 μL rabbit polyclonal anti-m^6A antibody (SYSY, 0.5 mg/mL), 5 μL SUPERNase inhibitor (Ambion) and diluted to a final volume of 500 μL with 287.5 μL of RNase-free warter. The mixture was incubated at 4 °C for 2 hours with mixing in the KingFisher (Thermo).

After 2 hours, 60 μL of washed Protein A beads (Thermo) were transferred to the IP and incubated for an additional 2 hours. The beads were then extensively washed (4 X 1 mL, IP buffer), and then incubated with 50 μL elution buffer (IP buffer + 20 mM m^6A nucleoside) for 1 hour. After saving the supernatant, the elution was repeated a for a second hour with an additional 50 μL of elution buffer. Elution volumes were combined, and RNA was purified using the RNA Clean and

Concentrator (Zymo), eluting in 6 μL of RNase-free water. RNA was used for library construction as previously described.

References

1. Moore, M. J. Nuclear RNA Turnover. *Cell* **108,** 431–434 (2002).
2. Moore, M. J. & Proudfoot, N. J. Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation. *Cell* **136,** 688–700 (2009).
3. Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* **18,** 18–30 (2016).
4. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18,** 437–451 (2017).
5. Wickramasinghe, V. O. & Laskey, R. A. Control of mammalian gene expression by selective mRNA export. *Nat. Rev. Mol. Cell Biol.* **16,** (2015).
6. Furuichi, Y., LaFiandra, A. & Shatkin, A. J. 5'-Terminal structure and mRNA stability. *Nature* **266,** 235–239 (1977).
7. Shimotohno, K., Kodama, Y., Hashimoto, J. & Miura, K.-I. Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. *Proc. Natl. Acad. Sci.* **74,** 2734–2738 (1977).
8. Cooke, C. & Alwine, J. C. The Cap and the 3' Splice Site Similarly Affect Polyadenylation Efficiency. *Mol. Cell. Biol.* **16,** 2579–2584 (1996).
9. Edery, I. & Sonenberg, N. Cap-dependent RNA splicing in a HeLa nuclear extract. *Proc. Natl. Acad. Sci.* **82,** 7590–7594 (1985).
10. Izaurralde, E. *et al.* A Nuclear Cap Binding Protein Complex Involved in Pre-mRNA Splicing. *Cell* **78,** 657–668 (1994).
11. Hamm, J. & Mattaj, I. W. Monomethylated Cap Structures Facilitate RNA Export from the Nucleus. *Cell* **63,** 109–116 (1990).
12. Richter, J. D. & Sonenberg, N. Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* **433,** 477–480 (2005).

13. Ku, U. & Wahle, E. Structure and function of poly(A) binding proteins. *Biochim. Biophys. Acta* **1678,** 67–84 (2004).
14. Mangus, D. A., Evans, M. C. & Jacobson, A. Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol.* **4,** 223 (2003).
15. 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).
16. 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).
17. Alarcon, C. R. *et al.* HNRNPA2B1 Is a Mediator of m6A-Dependent Nuclear RNA Processing Events. *Cell* **162,** 1299–1308 (2015).
18. Liu, N. *et al.* N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 560–564 (2015). doi:10.1038/nature14234
19. Lence, T. *et al.* m6A modulates neuronal functions and sex determination in Drosophila. *Nature* **540,** 242–247 (2016).
20. Haussmann, I. U. *et al.* m6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* **540,** 301–304 (2016).
21. Ke, S. *et al.* A majority of m6A residues are in the last exons, allowing the potential for 3′ UTR regulation. *Genes Dev.* 1–17 (2015). doi:10.1101/gad.269415.115.9
22. Hartmann, A. M., Nayler, O., Schwaiger, F. W. & Stamm, S. The Interaction and Colocalization of Sam68 with the Splicing-associated Factor YT521-B in Nuclear Dots Is Regulated by the Src Family Kinase p59 fyn. *Mol. Biol. Cell* **10,** 3909–3926 (1999).
23. Nayler, O., Hartmann, A. M. & Stamm, S. The ER Repeat Protein YT521-B Localizes to a Novel Subnuclear Compartment. *J. Cell Biol.* **150,** 949–961 (2000).
24. Rafalska, I. *et al.* The intranuclear localization and function of YT521-B is regulated by tyrosine phosphorylation. *Hum. Mol. Genet.* **13,** 1535–1549 (2004).
25. Zhang, Z. *et al.* The YTH Domain Is a Novel RNA Binding Domain. *J. Biol. Chem.* **285,** 14701–14710 (2010).
26. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* **61,** 507–519 (2016).
27. Wang, X. *et al.* N6-methyladenosine Modulates Messenger RNA Translation Efficiency.

Cell **161,** 1388–1399 (2015).

28. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 117–120 (2014). doi:10.1038/nature12730
29. Perry, R. P. & Kelley, D. E. Inhibition of RNA Synthesis by Actinomycin D: Characteristic Dose-response of Different RNA Species. *J. Cell Physiol.* **76,** 127–140 (1970).
30. Fuchs, G. *et al.* 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol.* **15,** R69 (2014).
31. Peritz, T. *et al.* Immunoprecipitation of mRNA-protein complexes. *Nat. Protoc.* **1,** 577–580 (2006).
32. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2008).
33. Lubas, M. *et al.* Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex. *Mol. Cell* **43,** 624–637 (2011).
34. Lubas, M. *et al.* The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis. *Cell Rep.* **10,** 178–192 (2015).
35. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Protoc.* **7,** 1009–1015 (2010).
36. Letunic, I., Doerks, T. & Bork, P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* **43,** D257–D260 (2015).
37. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* **61,** 507–519 (2016).
38. Patil, D. P. *et al.* m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537,** 369–373 (2016).
39. Zhong, X., Wang, P., Han, J., Rosenfeld, M. G. & Fu, X. SR Proteins in Vertical Integration of Gene Expression from Transcription to RNA Processing to Translation. *Mol. Cell* **35,** 1–10 (2009).
40. Huang, Y. & Steitz, J. A. Splicing Factors SRp20 and 9G8 Promote the Nucleocytoplasmic Export of mRNA. *Mol. Cell* **7,** 899–905 (2001).
41. Huang, Y., Gattoni, R., Stevenin, J. & Steitz, J. A. SR Splicing Factors Serve as Adapter

Proteins for TAP-Dependent mRNA Export. *Mol. Cell* **11,** 837–843 (2003).

42. Huang, Y., Yario, T. A. & Steitz, J. A. A molecular link between SR protein dephosphorylation and mRNA export. *Proc. Natl. Acad. Sci.* **101,** 9666–9670 (2004).
43. Müller-McNicoll, M. *et al.* SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev.* **30,** 553–566 (2016).
44. Botti, V. *et al.* Cellular differentiation state modulates the mRNA export activity of SR proteins. *J. Cell Biol.* **216,** 1–17 (2017).
45. Zhou, Z. & Fu, X.-D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* **122,** 191–207 (2013).
46. Xu, C. *et al.* Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* **10,** 927–929 (2014).
47. Schwartz, S. *et al.* Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5' Sites. *Cell Rep.* **8,** 284–296 (2014).
48. Slobodin, B. *et al.* Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell* **169,** 326–337 (2017).
49. Ke, S. *et al.* m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev.* **31,** 990–1006 (2017).
50. Ping, X.-L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* **24,** 177–189 (2014).
51. Liu, J. *et al.* A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10,** 93–95 (2014).
52. O'Brien, K., Matlin, A. J., Lowell, A. M. & Moore, M. J. The Biflavonoid Isoginkgetin Is a General Inhibitor of Pre-mRNA Splicing. *J. Biol. Chem.* **283,** 33147–33154 (2008).
53. Oordt, W. V. D. H. Van, Newton, K., Screaton, G. R. & Cáceres, J. F. Role of SR protein modular domains in alternative splicing specificity in vivo. *Nucleic Acids Res.* **28,** 4822–4831 (2000).
54. Busch, A. & Hertel, K. J. Evolution of SR Protein and HnRNP Splicing Regulatory Factors. *Wiley Interdiscip Rev RNA* **3,** 1–12 (2013).
55. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–120 (2014).

56. Auyeung, V. C., Ulitsky, I., Mcgeary, S. E. & Bartel, D. P. Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing. *Cell* **152,** 844–858 (2013).
57. Xu, C. *et al.* Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* **10,** 927–929 (2014).
58. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).
59. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
60. Corcoran, D. L. *et al.* PARalyzer : definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12,** R79 (2011).
61. 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).
62. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7,** 562–578 (2012).
63. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
64. Salmon-Divon, M., Dvinge, H., Tammoja, K. & Bertone, P. PeakAnalyzer: Genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* **11,** 415 (2010).
65. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).

Chapter 4 Emerging biological roles for m^{6}A methylation

4.1 m^{6}A in rapid transcriptome turnover

Decades have passed between the discovery of m^{6}A and a mechanistic understanding of its effects on mRNA metabolism. As the field continues to appreciate the molecular processes that depend on RNA metabolism, the biological utility of m^{6}A is still largely unknown. To date, m^{6}A is most strongly correlated with short mRNA life-times via direct recruitment of deadenylase complex[1–3]. Prolonged expression of transient developmental genes in the absence of m^{6}A methylation causes a variety of biological defects, namely a failure to emerge from pluripotent states[4,5] (**Figure 4.1**)[2].

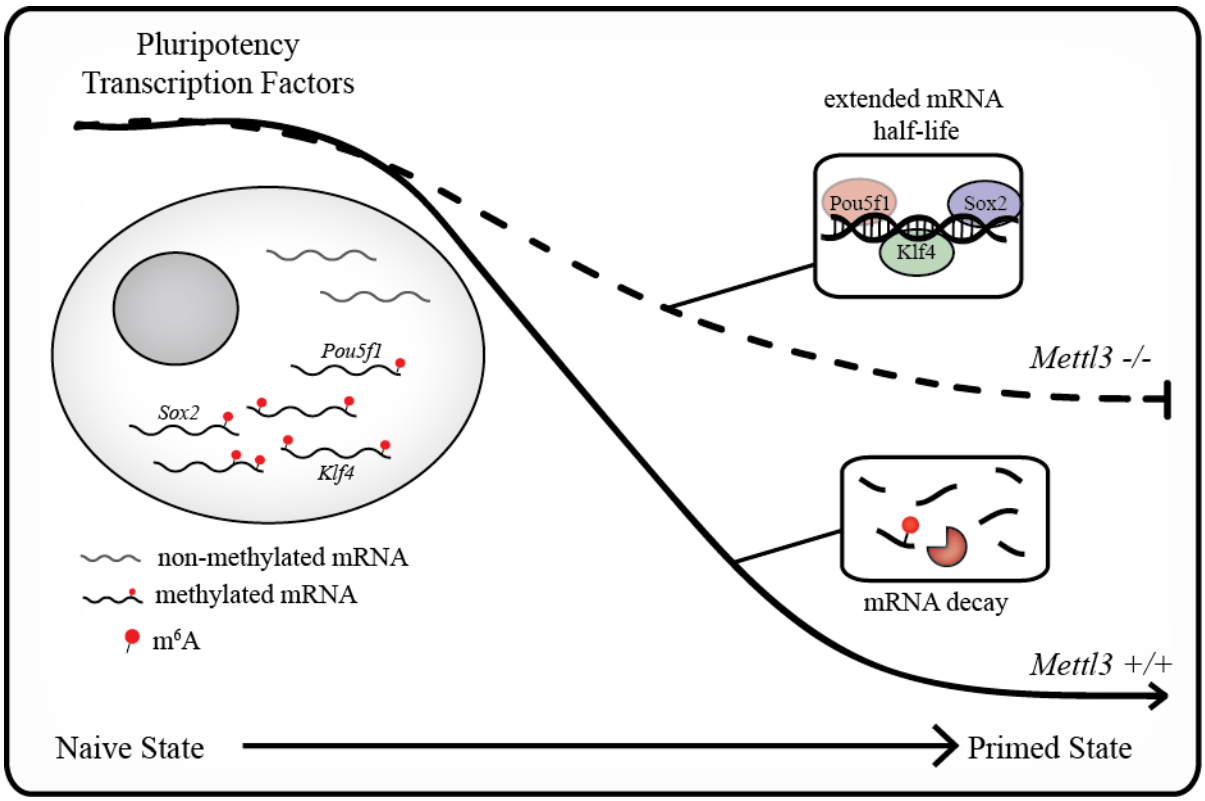

Figure 4.1 m^{6}A affects mouse embryonic stem cell differentiation

[2] Figure 4.1 and its legend are adapted from Zhao, B. S., Roundtree, I. A., He, C. Post-transcriptional gene regulation by RNA modifications. *Nat. Rev. Mol. Biol.* **18**, 31-42 with slight modification.

Figure 4.1 continued

m^6A methyltransferase Mettl3 is required for the transition of mouse embryonic stem cells (mouse ES cells) from a naïve to a more differentiated primed state. During this process, key pluripotency factor transcripts (Pou5f1, Sox2, Klf4 and others) must be rapidly cleared. This clearance is facilitated (presumably) by the effector protein Ythdf2. In cells lacking Mettl3, clearance is defective, preventing the establishment of a differentiated transcriptome required to achieve a primed mouse ES cell state.

A recent study of the maternal to zygotic transition in zebrafish has directly implicated a role for Ythdf2 in timely clearance of maternal transcripts upon activation of the zygotic genome[6]. It is becoming clear that m^6A may serve as a mark for rapid activation of mRNA metabolism, both in response to developmental queues and cellular stresses[7,8].

We sought to study the role of m^6A in a fundamental cellular process that requires rapid turnover of the transcriptome: the cell cycle[9]. We hypothesized that m^6A methylated genes are highly enriched for those that control the cycle, and may be dynamically regulated throughout cell cycle progression. Functional clustering by DAVID[10] confirmed that this is indeed that case for readily available sets of genes with known m^6A modification[11]. We therefore hypothesized that mRNA methylation is required for timely progression throughout the cell cycle.

4.2 Results

4.1.2 Characterization of m^6A machinery throughout the cell cycle

We began our study of RNA methylation in the cell cycle by characterizing the machinery that governs m^6A dynamics throughout stages of the HeLa cell cycle. In order to do so, we utilized fluorescence-assisted cell sorting (FACS) to isolate pure populations of cells from G_0, G_1, S, and G_2/M phases. We defined cell populations based on both DNA and RNA content, where cells with low DNA and RNA content represent G_o, and cells with low DNA and high RNA content represent G_1. Cells in S or G_2/M phase were separated based on DNA content alone (**Figure 4.2a**). These samples were collected and used for analysis for LC-MS/M. By LC-MS/MS we observe that m^6A accumulates in mRNA as cells progress from G_0 to G_2/M (**Figure 4.2b**), immediately suggesting that rapid clearance or dilution of methylated RNA occurs during cell division. Western blotting

of several proteins involved in m^{6}A regulation suggest that both YTHDF2 and YTHDC1 vary in abundance depending on the state of the cell (**Figure 4.2c**).

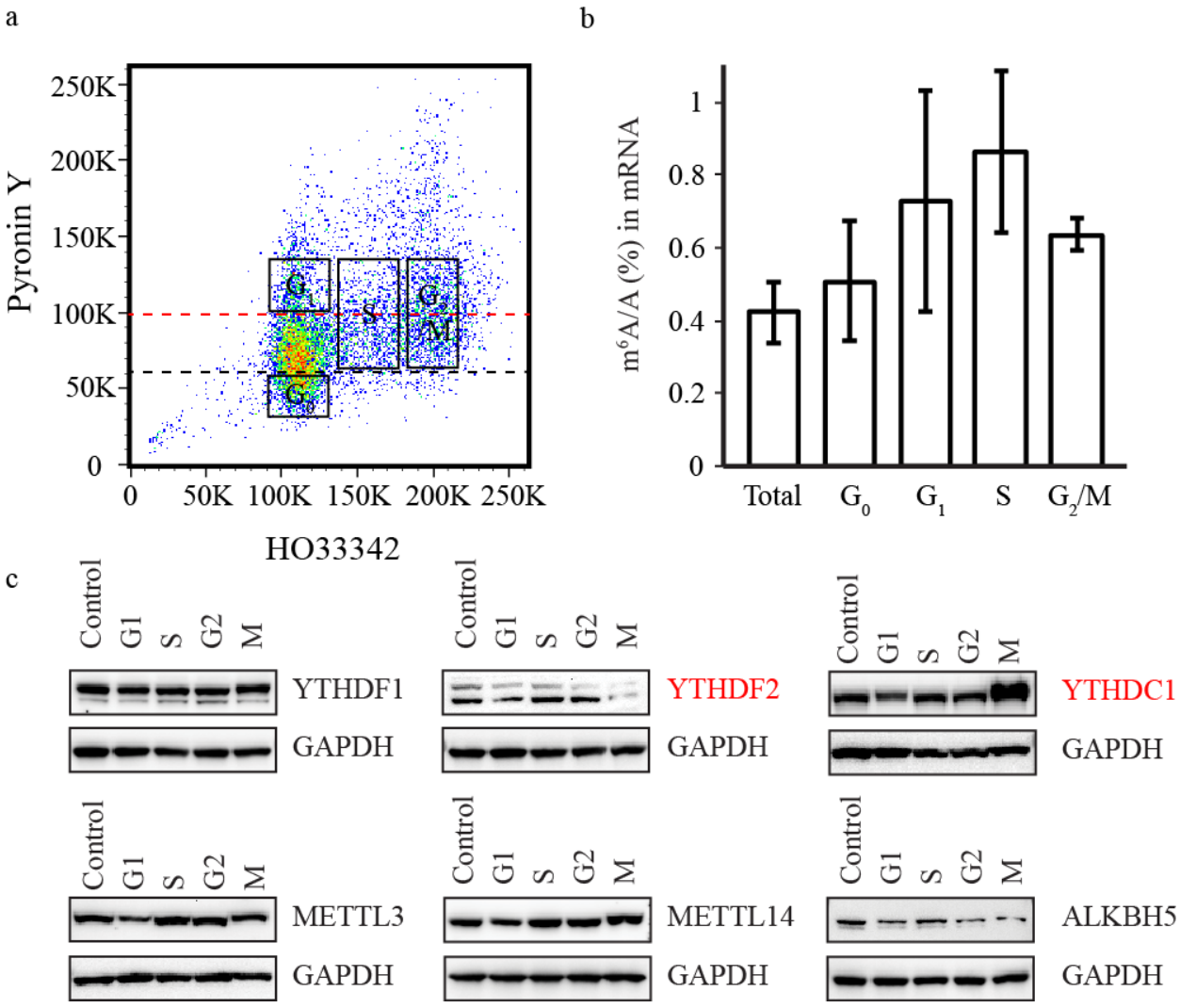

Figure 4.2 Characterization of m^{6}A during cell cycle progression

(a) Cell sorting scheme for analysis of the HeLa cell cycle (b) LC-MS/MS of sorted HeLa cells. Error bars represent mean ± standard deviation, n = 4 (2 biological replicates x 2 technical replicates). (c) Abundance of key m^{6}A regulators throughout the cell cycle. Samples from synchronized HeLa cells[12].

4.2.2 Functions for YTHDF2 in mediating mitosis in HeLa cells

We conceived this project on the notion that YTHDF2 maintains RNA clearance that is required for transcriptome turnover during the cell cycle. In order to pursue this, we constructed knockout HeLa cells using CRISPR/Cas9. Analysis of cell proliferation suggested that while viable, YTHDC1 knockout cells have a slightly reduced growth rate compared to wild type cells (**Figure 4.3a**). We then attempted to synchronize these cells in order to track them through the cell cycle progression by treating with the compound lovastatin, an inhibitor of HMG-CoA reductase. To our surprise, we observed that YTHDF2 knockout cells cannot be synchronized to the G_1 stage (**Figure 4.3b**).

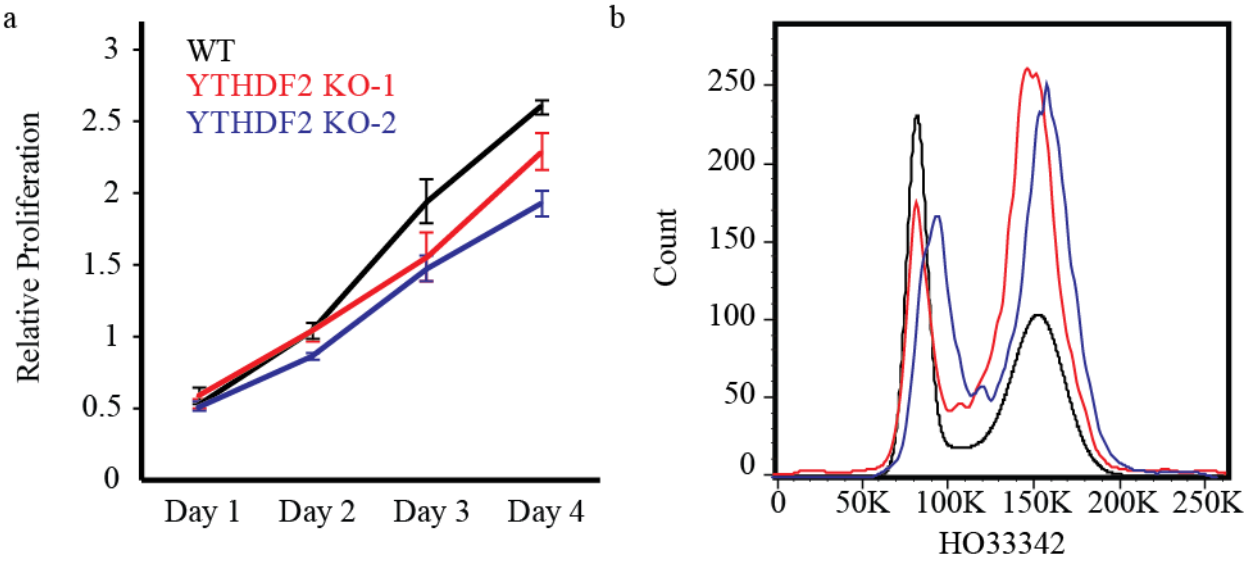

Figure 4.3 Proliferation and synchronization of HeLa cells to G_1 phase

(a) Proliferation of wild type and YTHDF2 knockout cells. Error bars represent mean ± standard deviation, n = 12 (4 biological replicates x 3 technical replicates). (b) G_1 synchronization of wild type and YTHDF2 knockout cells by flow cytometry.

Although this defect represents an obvious inability to respond to the lovastatin block, it did not present a starting point for dynamic analysis of progression.

We turned to synchronization to S phase by the double thymidine block, a blockade of DNA synthesis that should apply more universally to proliferating cells. Wild type and YTHDF2 knockout cells were synchronized to early S phase and released, allowing them to proceed with

DNA synthesis and mitosis over 12 hours. We observe that YTHDF2 knockout cells proceed normally through S phase, but are mitotically delayed (**Figure 4.4**).

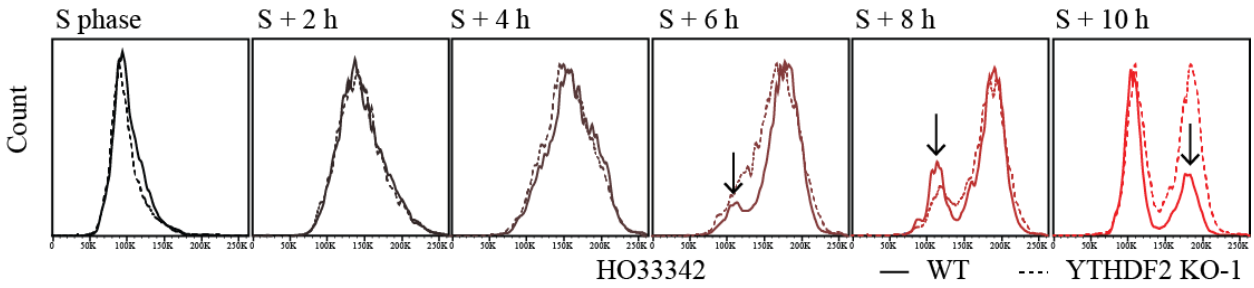

Figure 4.4 Progression of HeLa cells from double thymidine block

Progression of HeLa cells from S phase through mitosis. Wild type profiles are shown with a solid line, knockout profiles are shown with a dashed line. Black arrows indicate deviations in the profile that indicate deviations in the duration of mitosis.

The gross defect observed in mitosis prompted us to consider if the cellular transcriptome of YTHDF2 knockout cells was similarly delayed in its progression.

We turned to RNA-sequencing to for a genome-wide understanding of how wild type and knockout cells evolve their transcriptome during the transition from S to M phases. We find that methylated genes with stage-specific expression are overrepresented in their operative stage, i.e. S-phase genes are more abundant during the presumptive S-phase in our experiment (**Figure 4.5**). The clearance of these transcripts appears to be particularly critical during late stages of the cell cycle, 6-10 hours after release (**Figure 4.5c,d**).

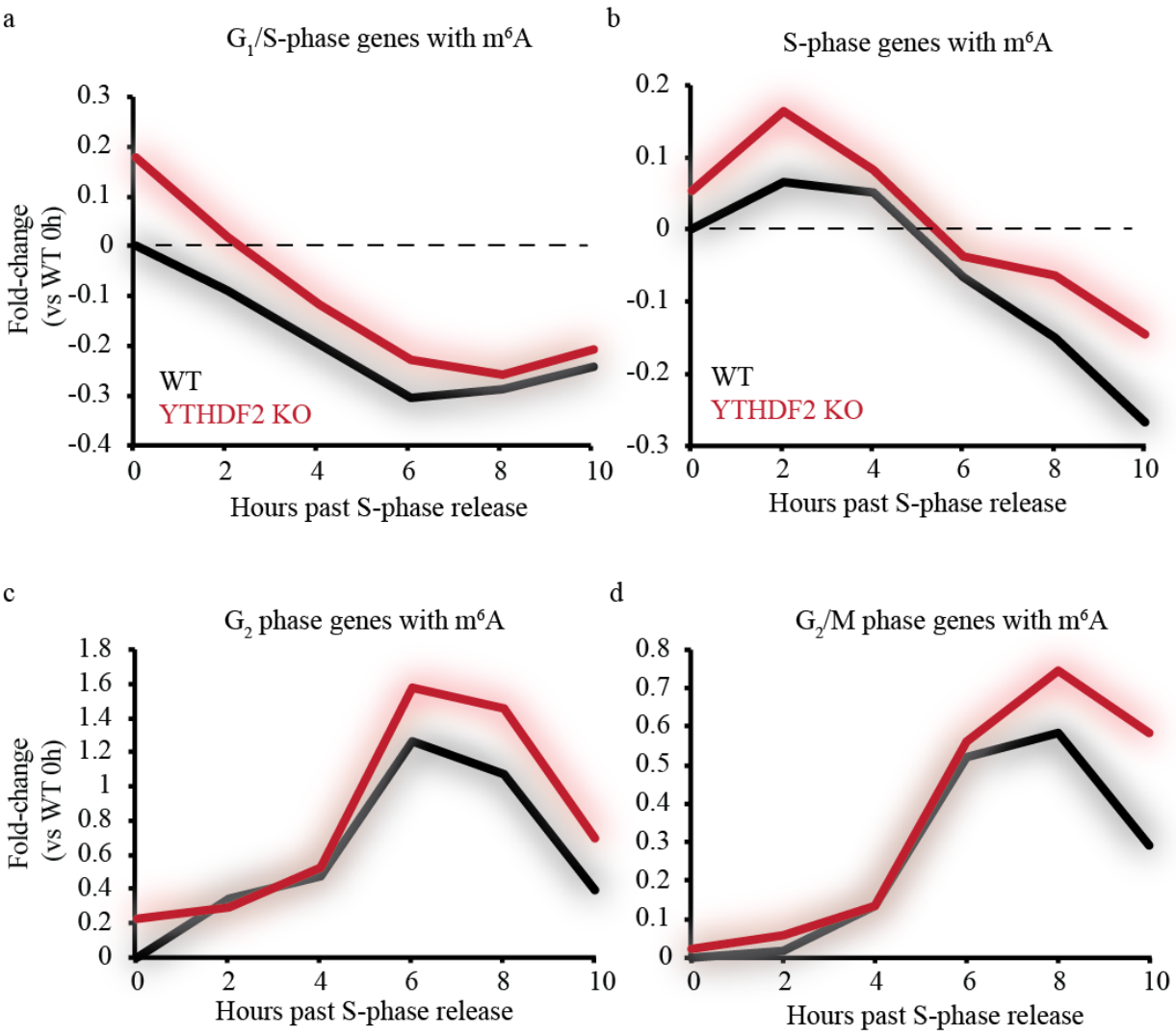

Figure 4.5 Expression of phase-specific, m^{6}A methylated mRNAs following release from S-phase

Expression of (a) G_1/S (b) S (c) G_2 and (d) G_2/M-specific transcripts bearing m^{6}A methylation following release from a double thymidine block. Data are presented as fold chances versus wild type cells at S-phase. Average expression was obtained from normalized FKPM values from 3 biological replicates.

Analyzing these data for defects in single gene expression provided a list of candidate genes whose delayed clearance may be contributory to a mitotic delay (**Figure 4.6a-d**).

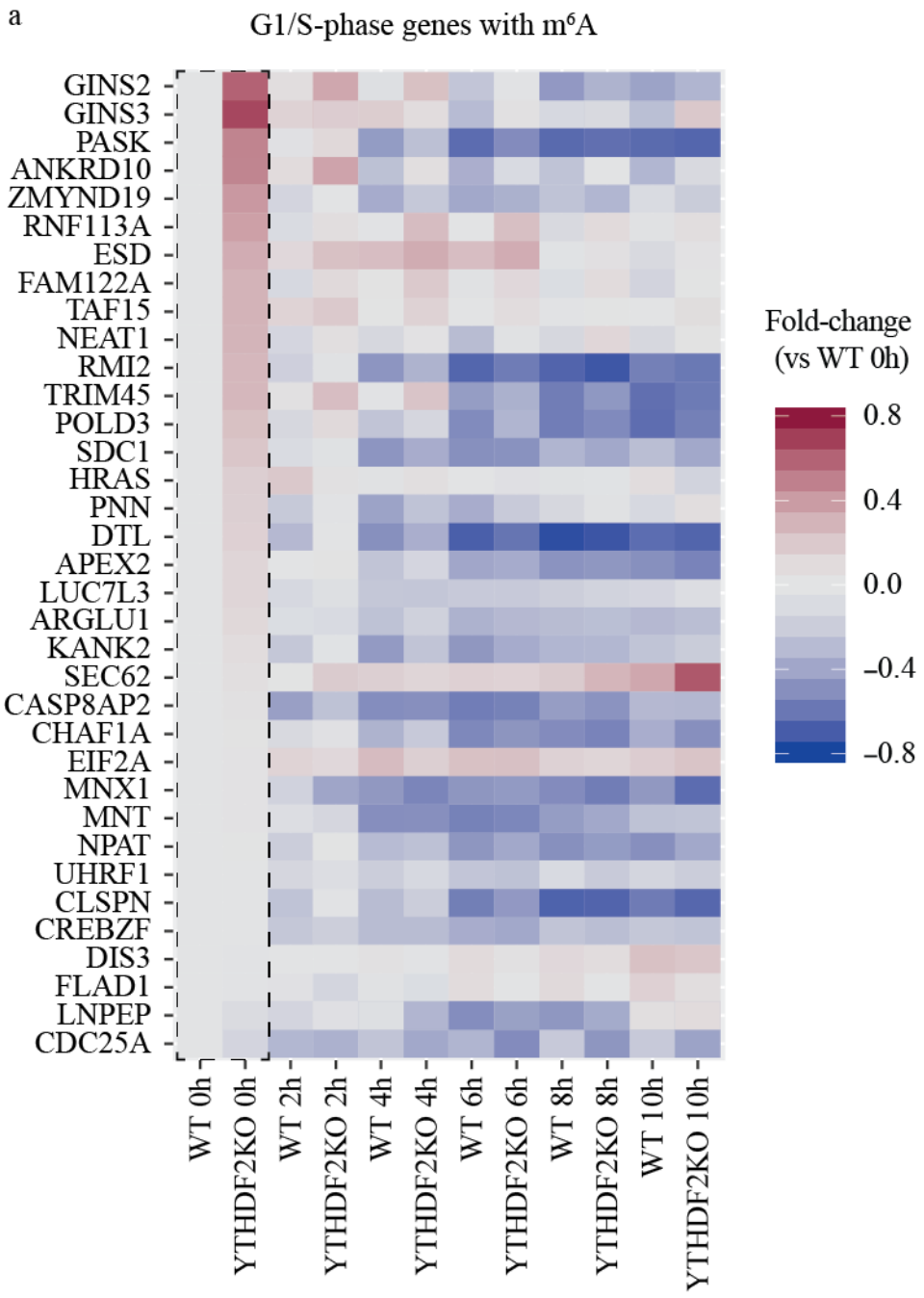

Figure 4.6 Heat map analysis of methylated, phase-specific transcripts following release from S-phase

(a) G_1/S-phase transcripts. Dashed line surrounds time = 0h at which genes were sorted for YTHDF2 KO/WT ratio, greatest to least.

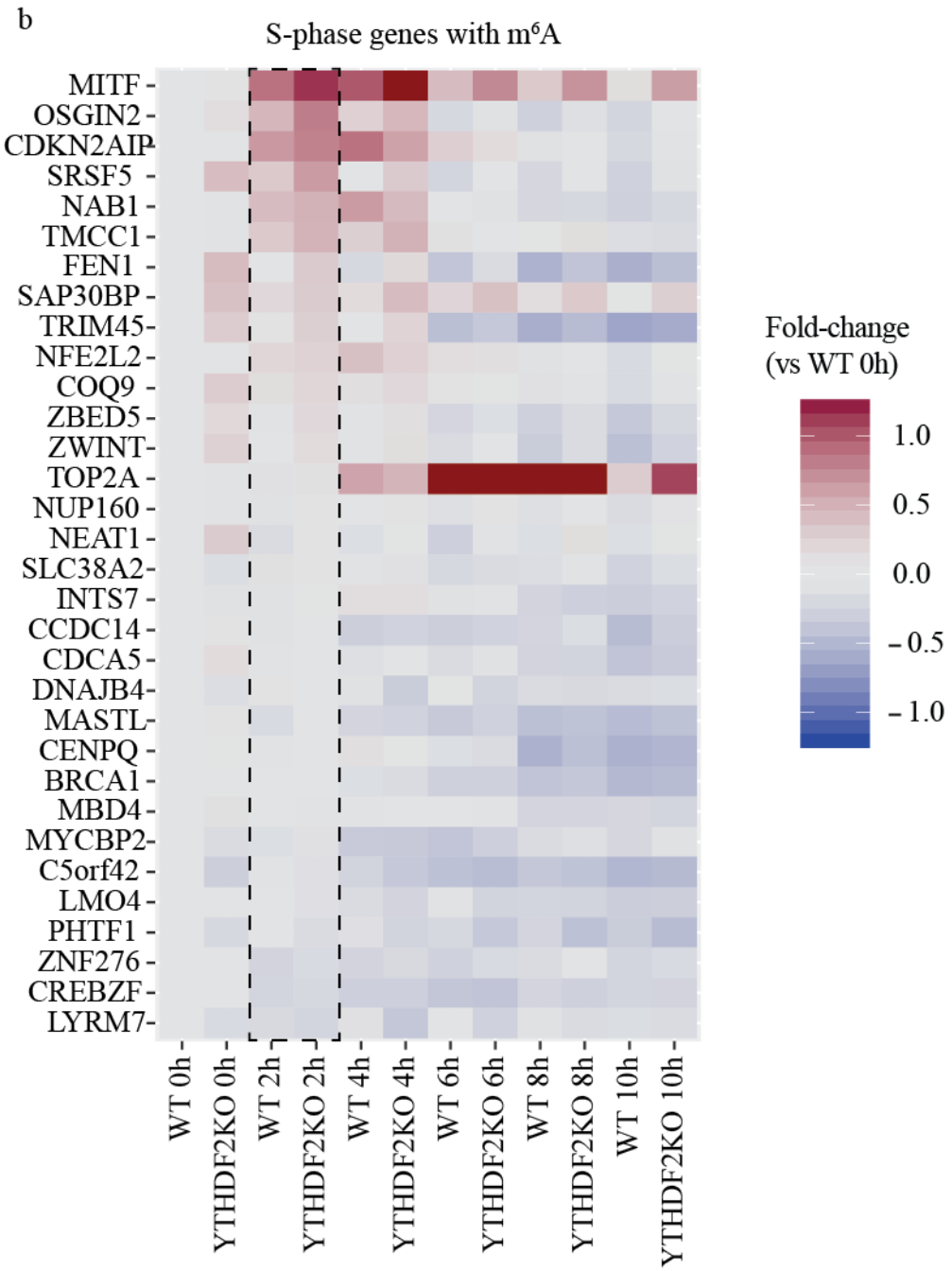

Figure 4.6 continued

(b) S-phase transcripts. Dashed line surrounds time = 2h at which genes were sorted for YTHDF2 KO/WT ratio, greatest to least.

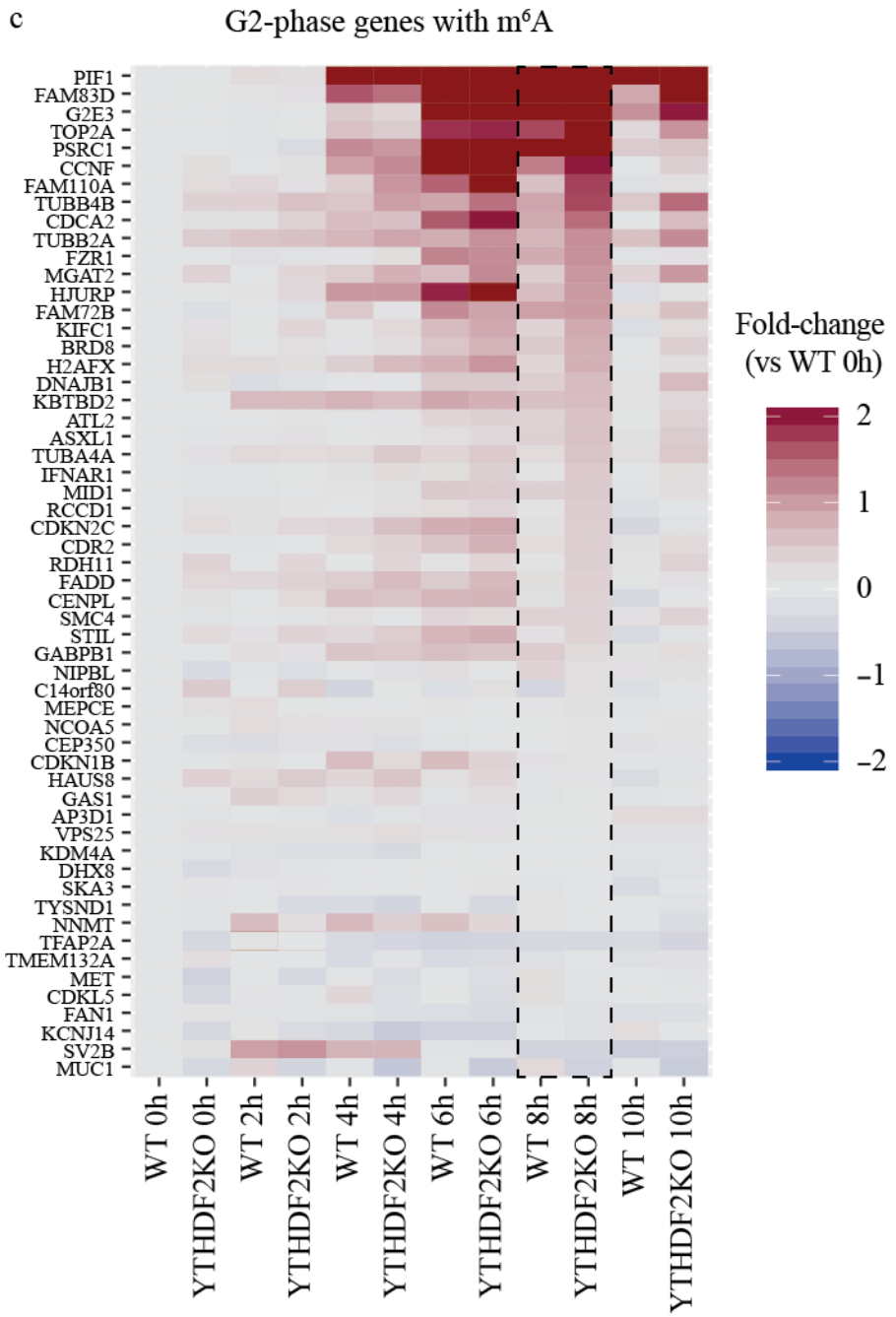

Figure 4.6 continued
(c) G_2-phase transcripts. Dashed line surrounds time = 8h at which genes were sorted for YTHDF2 KO/WT ratio, greatest to least.

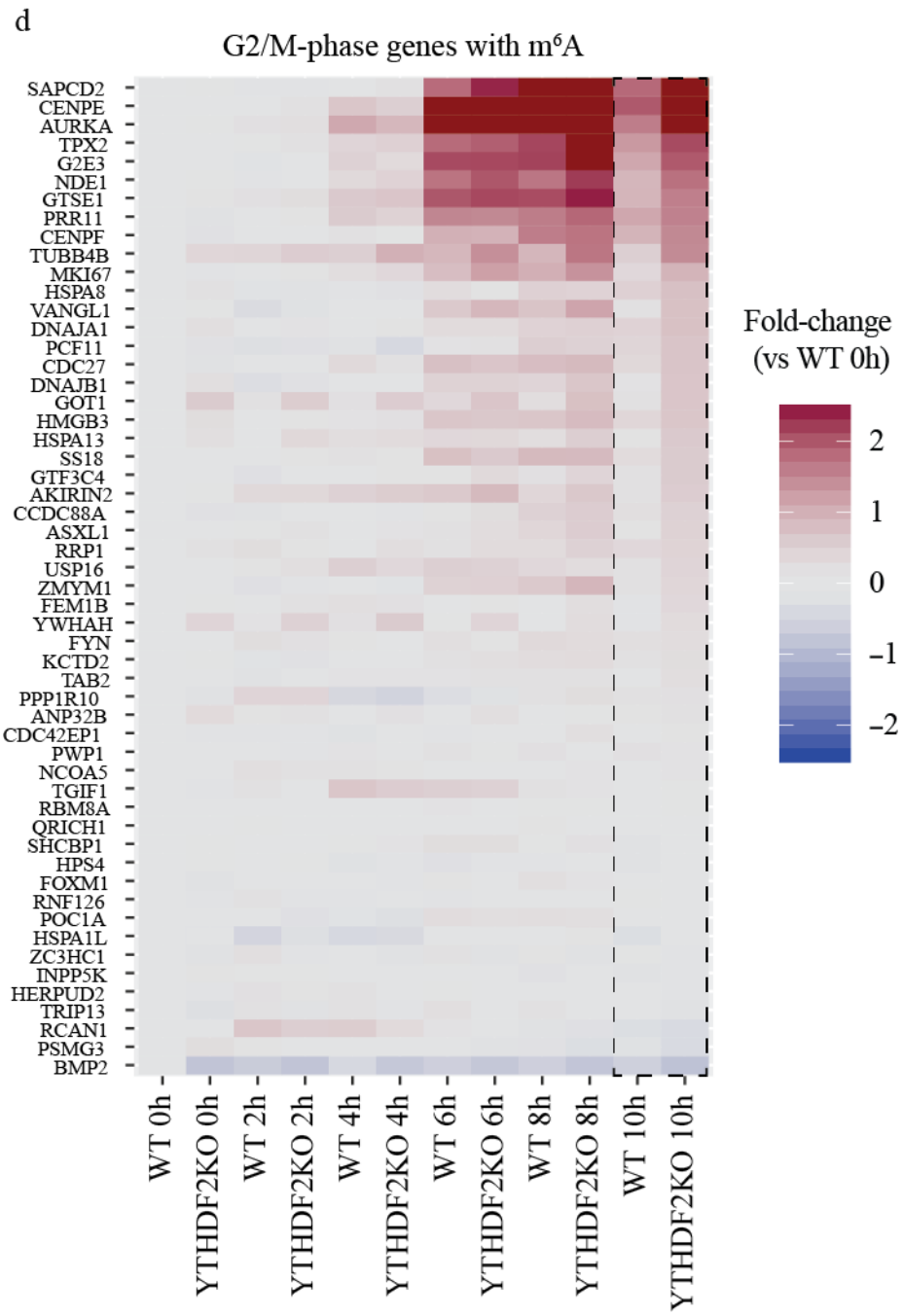

Figure 4.6 continued

(d) G_2-phase transcripts. Dashed line surrounds time = 8h at which genes were sorted for YTHDF2 KO/WT ratio, greatest to least.

The largest deviation between wild type and YTHDF2 knockout HeLa cells occurs between 6-10 hours beyond S-phase release, when transcripts for G_2- and M-phase transcripts are rapidly induced and cleared. We confirmed that the transition from G_2/M phase is delayed by

synchronizing cells to G_2 and tracking their cell division by flow cytometry. As expected, we observe a mitotic delay in knockout cells (**Figure 4.7**)

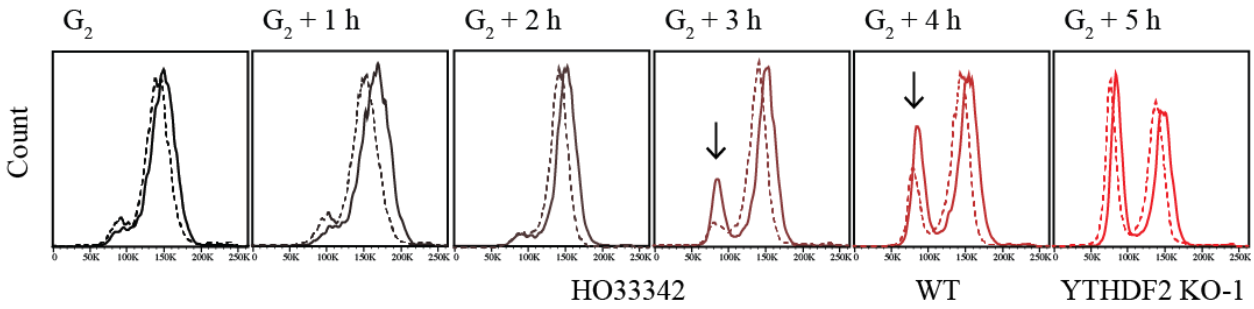

Figure 4.7 Progression of HeLa cells from G_2 synchronization

Progression of HeLa cells from G_2 phase through mitosis. Wild type profiles are shown with a solid line, knockout profiles are shown with a dashed line. Black arrows indicate deviations in the profile that indicate deviations in the duration of mitosis.

4.3 Conclusions and discussion: m^{6}A-mediated transcriptome turnover in progression of the cell cycle

We have shown that cells must rapidly exchange their poll of RNA throughout the cell cycle, as phase-specific genes are cleared within hours of their induction. During mitosis, delayed clearance of G_2- and M-phase transcripts is delayed in cells lacking YTHDF2, suggesting that m^{6}A-dependent mRNA decay is necessary for coordination of mitosis. This represents a novel role for post-transcriptional RNA modifications in regulating an acute cellular process, and implicates m^{6}A modification in one of the most fundamental biological processes to human development and disease.

Our results have led us to conclude that m^{6}A is most likely utilized just prior to or during mitotic events. As this is the stage in which cells spend the least amount if time, it is perhaps unsurprising that selective RNA processing is required. The nature of cell division makes deciphering the molecular defect a challenge, as deregulation of individual proteins can cause unique phenotypes and trigger cell cycle checkpoint arrest. The next stage of this work will aim to identify the causal defects in the YTHDF2 knockout transcriptome that prevent timely progression through these highly-coordinated events. An expanded study into other regulators of the

epitranscriptome interact with the progression of the cell likely will also be valuable contributions to the field.

4.4 Methods

4.4.1 Cell sorting

Cell sorting was based on both DNA content and RNA content. The method requires saturation of DNA by Hoechst, followed by RNA-specific staining by Pyronin Y[13]. HeLa cells were trypsinized and resuspended in media at a density of 4 million cells/mL. Hoechst 33342 (Thermo) was added to a final concentration of 100 μg/mL and incubated at 37 °C for 45 minutes. After 45 minutes, Pyronin Y (Sigma) was added to a final concentration of 1 μg/mL and incubated for an additional 1 hour. Cells were sorted directly from this solution on an AriaII 4-15 (BD) and directly lysed with TriZol for RNA purification.

4.4.2 HeLa cell synchronization

HeLa cell synchronization was performed according to protocol[12].

G_1: HeLa cells were grown in 10 cm plates to 50% confluency. Lovastatin was added to a final concentration of 20 μM and cells were cultured for 24 hours.

S: HeLa cells were grown in 10 cm plates to 40% confluency. Thymidine (Sigma) was dissolved in media to 100 mM in HeLa media, and added to culture to a final concentration of 2 mM. The cells were incubated for 14 hours.

The media was aspirated, and cells were washed thoroughly with PBS (2 x 10 mL).

The cells were given new medium supplemented with 24 μM deoxycytidine (Alfa Aesar) and incubated for 9 hours.

Media was aspirated thoroughly and given media supplemented with 2 mM thymidine for an additional 14 hours.

Cells were washed thoroughly, and incubated with media supplemented with 24 μM deoxycytidine. Cells were collected every 2 hours for analysis.

G_2: Cells were first blocked using the double thymidine procedure. After the second thymidine block, cells were washed thoroughly with PBS and released into fresh growth media for 2 hours. After 2 hours, RO-3306 (Sigma) was added to the media to a final concentration of 10 μM in (stock of 1 mM in DMSO). The cells were incubated for an additional 10 hours.

After 10 hours, the cells were washed thoroughly and given fresh growth medium. Cells were collected every hour for analysis.

4.4.3 Flow cytometry

HeLa cells were collected by cell scraper and resuspended in 250 μL PBS. 250 uL of 4% formaldehyde (Fisher) were added to the resuspensions, and cells were rotated at room temperature for 15 minutes for fixation. After fixing, cells were pelleted at 2,000 x g for 5 minutes, and resuspended in 250 μL of PBS. Staining was performed using Hoechst-33342 at a final concentration of 10 μM at 37 °C for 30 minutes.

Cytometry was performed using the LSRII 3-8 (BD).

4.4.4 Generation of YTHDF2 knockout HeLa cells

YTHDF2 knockout cells were generated using a protocol adapter from New England Biolabs using Cas9 protein from *S. pyogenes* (M0386).

crRNA-1: TGAAGCTGCTTGGTCTACGGGGG
crRNA-2: TGCTGAGAAGTCAATCCCACTGG

crRNAs and tracrRNA were resuspended in nuclease-free IDTE buffer to a final concentration of 100 μM. Aliquots were diluted to a working concentration of 10 μM, final volume of 50 μL. For each crRNA, crRNA and tracrRNA were mixed in equimolar ratios to a final concentration of 1.5 μM (5 μL crRNA + 5 μL tracrRNA + 23.3 μL nuclease-fee Duplex Buffer (IDT). The solution was heated to 95 °C and allowed to cool to room temperature.

Cas9 protein was diluted to a working concentration of 1.5 μM in Cas9 working buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 5% glycerol, 1 mM DTT) (0.49 μL Cap9 protein + 19.51 μL of working buffer)

For each crRNA, 1 μL of the crRNA/tracrRNA mixture was added to 1 μL of diluted Cas9 protein and diluted with 10.5 μL opti-MEM (Invitrogen). The mixture was incubated at room temperature for 5 minutes to assemble RNP complexes.

Following the 5 minute incubation, RNPs were added to 1.2 μL of Liopfectamine RNAiMAX and further diluted with 11.3 μL of opti-MEM. The transfection mixture was incubated at room temperature for 20 minutes.

During this incubation, HeLa cells were diluted to 400,000 cells/mL in media lacking antibiotics. 125 μL of cells (50,000 cells) were added to 96-well tissue culture plates. The transfection mixture was added, and cells were incubated at 37 °C for 48 hours.

After 48 hours, cells were diluted to single cell suspensions and expanded for genotyping by Western blot and PCR.

YTHDF2 KO F: CCTCCCAAAGTGTAGGGATTAC
YTHDF2 KO R: CTACCAACAGCAGAACCTACAA

A single clone from each crRNA was confirmed by Western blot and PCR and used as biological replicates.

References

1. Wang, Y. *et al.* N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* **16,** 191–201 (2014).
2. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 117–120 (2014). doi:10.1038/nature12730
3. Du, H. *et al.* YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. *Nat. Commun.* **7,** 1–11 (2016).
4. Geula, S. *et al.* m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science.* **3,** 1002–1006 (2015).
5. Batista, P. J. *et al.* m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells. *Cell Stem Cell* **15,** 707–719 (2014).
6. Zhao, B. S. *et al.* m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **542,** 475–478 (2017).
7. Zhou, J. *et al.* Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature* **526,** 591–594 (2015).
8. Xiang, Y. *et al.* RNA m6A methylation regulates the ultraviolet-induced DNA damage response. *Nature* 1–22 (2017). doi:10.1038/nature21671
9. Whitfield, M. L. *et al.* Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. *Mol Biol Cell* **13,** 1977–2000 (2002).
10. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2008).
11. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–206 (2012).
12. Ma, H. T. & Poon, R. Y. C. Synchronization of HeLa Cells. *Methods Mol. Biol.* **761,** 151–161 (2011).
13. Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25,** 1860–1872 (2015).

Chapter 5: Summary and perspectives

5.1 Advances and shortcomings in m^{6}A biology

Since the discovery of pseudouridine[1] as the fifth base in RNA, the field of RNA modifications has exploded to encompass over one hundred modifications in every known cellular species. Fueled largely by advances in analytical chemistry and high-throughput sequencing, our ability to identify and profile new modifications has continued to drive research. To date, m^{6}A remains the most well studied internal modification in mammalian mRNA. The discovery of m^{6}A-specific 'reader' proteins revolutionized our understanding of the potential for "RNA epigenetics"[2–4].

In merely the past 4 years, dozens of roles for m^{6}A have been identified. At the molecular level, the modification has been implicated in RNA structure, alternative splicing, translation initiation, mRNA decay, and responses to DNA damage. Much of the work presented in this thesis establishes YTHDC1 as the mediator of yet another function for m^{6}A in mediating mRNA export. Undoubtedly, many more functions have yet to be discovered. However, m^{6}A lacks a broad, unifying role in mediating gene expression, and the multitude of new results have failed to uncover obvious harmonizing concepts. Rather, reversible RNA modification presents great potential for regulatory capacity across several molecular and biological settings. This concept represents a large gap in knowledge that is unlikely to be resolved in the near future.

5.2 Advantages of post-transcriptional regulation over other methods of gene regulation

Reversible mRNA methylation has obvious advantages over sequence-based post-transcriptional regulation. Most simply, methylation of an mRNA can be tuned based on time and place within an organism. The field is only beginning to observe how cells utilize this potential, and what mechanisms enable it. Future studies that identify how methyltransferase, demethylase, and effector components integrate with signaling pathways will further solidify this concept in gene regulation.

5.3 Future challenges in RNA modification research

Much of the data generated from studies of m^6A rely on a binary qualifier; a gene either contains methylation or it does not. This is an obvious misrepresentation of the field, yet is frequently overlooked as a conceptual hurdle in advancing our knowledge, because we currently do not have a deep understanding of what separates the two classes of transcripts. This immediately gives rise to several questions:

1. How is m^6A selectivity achieved?

Methyltransferase selectivity cannot be attributed simply to the presence of a consensus sequence, nor is it strictly dependent on RNA structure. Yet we are able to consistently identify the same set of transcripts as methylated, and often within the same narrow sequence window. Possible sources of selectivity are chromatin components which recruit the methyltransferase, either via RNA Pol II transcription status or through contacts with machinery involved in the regulation of transcription such as state-specific transcription factors. If the latter is the case, we may see extremely different m^6A profiles, and even functions, across diverse cell types.

2. How are methylated sequences differentiated?

As briefly discussed in **Chapter** 2, the five YTH family proteins in mammalian cells each have slightly different sequence selectivity in their recognition of methylated RNA. YTHDC1 features the highest affinity and greatest sequence selectivity for GG(m^6A)CU, while YTHDC2 has relatively poor selectivity for m^6A. Whether these differences amount to meaningful biological discrimination is currently unknown. If these proteins are able to further subdivide m^6A-containing transcripts into distinct groups, it will contribute to greater noise than we can currently decipher. Furthermore, if the function of these proteins is not regulated in unison (as appears to be the case), the effect of m^6A may be increasingly context-dependent.

3. What is the role of m^6A within the nucleus?

Chapter 3 of this thesis details a role for m^6A in facilitating nuclear export of mRNA. Several previous studies have suggested that m^6A mediates alternative splicing at the hands of HNRNP proteins in addition to YTHDC1. Intuitively, it is formally possible that the real function lies between the two, and m^6A methylation is somehow coupled to many components of the nuclear pre-mRNA processing machinery. Disruption of the nuclear architecture may result in deviations in splice patterns among other defects. This would be consistent with the notion of m^6A as a facilitator of canonical mRNA processing. This would predict several outcomes: 1. Constitutive splicing, not alternative, splicing is facilitated by m^6A. Defects in the rate of constitutive splicing may manifest as alternative isoform usage is secondary reactions or pathways are utilized in the absence of m^6A. 2. Other nuclear processes are partial to m^6A, such as, polyadenylation and perhaps transcription. The last two examples here may represent examples where the methyltransferase protein complex may have distinct roles as a scaffold in additional to its catalytic activity.

4. Is m^6A methylation dynamic in a cellular context?

Reversible RNA methylation can be dynamic in that the presence of this mark can be tuned for the appropriate response. Identifying these conditions can be a challenge, as most *in vitro* models do not require rapid responses to environmental stimuli. While potentially useful for understanding molecular processes protein effectors, it is difficult to appreciate the true potential of m^6A modification without studying a model of development or disease. Groundwork laid by the field over the past several years has put future researchers in good position to uncover the true utility of RNA modification, as we have begun to observe in our studies in **Chapter 4** of this thesis.

5.4 Concluding remarks

Functional studies of mRNA modification have expanded rapidly in the last decade. The field has identified multiple 'reader' proteins with various effects on post-transcriptional regulation. As more such proteins are discovered, the roles for m^6A and other chemical modifications in mRNA will be expanded. Whether or not a harmonizing role for mRNA modification will ever be identified is unknown. It is entirely possible that the diversity of

modification and function is a salient feature of post-transcriptional chemical modification, which we will only continue to appreciate as research in the area continues.

References

1. Cohn, W. E. Pseudouridine, a Carbon-Carbon Linked Ribonucleoside in Ribonucleic Acids: Isolation, Structure, and Chemical Characteristics. *J. Biol. Chem.* **235,** 1488–1498 (1960).
2. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–206 (2012).
3. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–120 (2014).
4. He, C. RNA epigenetics? *Nat. Chem. Biol.* **6,** 863–865 (2010).

Appendix 1 Dynamic RNA modifications in gene expression regulation[3]

Appendix 1.1 Abstract

More than one hundred types of chemical modifications have been identified in cellular RNAs. While the 5' cap modification and the poly(A) tail of eukaryotic messenger RNA play key roles in regulation, internal modifications have gained attention of late for their roles in mRNA metabolism. The availability of highly sensitive analytical and sequencing technologies has led to the characterization of several chemical modifications within coding transcripts, spearheaded in large part by studies of the most abundant internal modification, N^6-methyladenosine (m^6A). Identification of proteins that install, recognize, and remove this and other marks have revealed a variety of roles for mRNA modification in nearly every aspect of the mRNA lifecycle, as well as various cellular, developmental, and disease processes. Abundant noncoding RNAs such as transfer RNAs (tRNAs), ribosomal RNAs (rRNAs) and spliceosomal RNAs (snRNAs) are also heavily modified, and the presence of these modifications is required for biogenesis and function of these diverse species. Given the various chemical functionalities that are known to decorate cellular RNAs, our understanding of their biological functions is merely beginning to take shape. However, it has become clear that in both coding and noncoding RNA, dynamic chemical modifications represent a new layer of control of genetic information.

Appendix 1.2 Introduction

Modifications to RNA species have been well documented for over 50 years. mRNAs are known to contain a 5' cap and a poly(A) tail, both deposited on the molecule during or shortly after transcription by RNA Polymerase II. Following elucidation of the 5' cap structure[1,2], functional roles for the modified 5' end were reported, including maintenance of transcript

[3]Appendix 1 is adapted from Roundtree, I.A., Evans, M.E., Pan, T., He, C. Dynamic RNA Modifications in gene expression regulation. *Cell* **169**, 1187-1200 (2017) with slight modifications.

stability[3,4], pre-mRNA splicing[5–7], polyadenylation[8], mRNA export[9], and translation initiation[10]. Similarly, the eukaryotic poly(A) sequence facilitates nuclear export, translation initiation and recycling, and promotes mRNA stability, largely through the association of the poly(A)-binding protein family[11].

Shortly after the discovery of the cap and tail modifications, internal modifications on mRNA were identified, including the most abundant internal modification of mRNA and long-noncoding RNA (lncRNA), N^6-methyladenosine (m^6A)[12–15]. Using P^{32}-labeled cellular RNA and thin layer chromatography, Lavi et al. estimated the abundance of m^6A in poly(A) selected species from both nuclear and cytoplasmic compartments to be about one per 700-800 nucleotides. Non-polyadenylated, non-ribosomal RNA was also found to contain significant amounts of the internal methylation, with m^6A occurring every 1,800-3,000 nucleotides[16]. Digestion of mRNAs with RNases revealed that the modification is largely confined within a G(m^6A)C (~70%) or A(m^6A)C (~30%) sequence, suggesting that the deposition is selective among mRNA sequences[17,18], and that only a portion of consensus sequence motifs bear detectable methylation.

m^6A was found to accelerate pre-mRNA processing and mRNA transport in mammalian cells[19,20], and is essential for mammals. These observations suggested previously unrecognized regulatory roles of the mRNA modification that may impact various cellular processes. Analogous to the diverse chemical marks on histone tails, recent studies also reveal numerous internal modifications within eukaryotic mRNA, including additional methylations of adenosine to form N^1-methyladenosine (m^1A) and N^6,2'-O-dimethyladenosine (m^6Am), as well as cytosine methylation to 5-methylcytosine and its oxidation product 5-hydroxymethylcytosine (hm^5C). Pseudouridine (Ψ) and 2'O-methylation, common modifications of tRNA and rRNA, have also been identified as abundant components of mRNA (**Figure A.1**). In this review, we summarize these and other chemical modifications of coding and noncoding RNA, with a focus on introducing the underlying regulatory mechanisms and their biological consequences. Modifications on the 5' cap and 3' poly(A) tail of mRNA will not be included, which have been extensively reviewed elsewhere[21–23], nor will we discuss the profound modifications of transcripts such as pre-mRNA splicing, circularization or RNA editing, all of which have significant functional implications.

Appendix 1.3 Revealing internal mRNA modifications – The 'epitranscriptome'

The recent advances in studying RNA modifications have benefited tremendously from the improved methods for detection in both analytical chemistry and high-throughput sequencing. Though we aim to provide a conceptual overview of the methods upon which recent progress in the field is based, readers can refer to a more comprehensive review of techniques in studying RNA modifications[24].

Adenosine Methylations

m^{6}A: Recently, two major advances have fueled investigations into the function of internal mRNA modifications. First was the identification of an enzyme, fat-mass and obesity-associated protein (FTO), which catalyzes the oxidative demethylation of m^{6}A in nuclear RNA[25], providing evidence that reversible RNA modifications serve regulatory roles[26]. A second m^{6}A demethylase of the same family, Alkbh5, affects mouse fertility and spermatogenesis[27].

Due to the low cellular abundance of mRNA and the chemically stable nature of N^6-methylation of adenosine, methods to determine the precise modification sites and the modification fraction at these sites hindered biological studies for decades. The second advance in m^{6}A biology came with the use of high-throughput sequencing, which when coupled to modification-specific antibody-based enrichment, provided transcriptome-wide maps of modification sites in both mRNA and lncRNA at ~200 nucleotide resolution[28,29], offering the first view of the m^{6}A 'epitranscriptome'. The implementation of cross-linking has since increased the resolution of this m^{6}A map, allowing for near single-base resolution determination of methylation sites in mRNA, lncRNA, and snoRNA[30,31]. Attempts to determine modification fraction have been addresses by a ligation-based method termed SCARLET, which provides single-base resolution of candidate m^{6}A sites as well as a quantitative modification fraction, albeit in a low-throughput manner[32]. Identification of an m^{6}A-selective reverse transcriptase from *Thermus thermophilus* offered a chance at high-throughput, single-base quantification of m^{6}A modification status, but is still limited by low selectivity in practice[33]. Internal calibration of immunoprecipitation (IP) efficiency

using modified RNA spike-in controls allows for semi-quantitative information of whole transcripts but lacks resolution of traditional m^{6}A-seq[34], highlighting a continued need to develop methods for more quantitative mapping of the m^{6}A epitranscriptome, which accounts for 0.2-0.6% of all adenosines in mammalian mRNA.

m^{1}A: Unlike m^{6}A, methylation at the N^1 position of adenosine occurs on the Watson-Crick interface and generates a positively charged base. Although it is found in human cell lines at levels 10 times lower than m^{6}A, the abundance of N^1-methyladenosine (m^{1}A) is much higher in certain mouse tissues, reaching to about 1/3 of that of m^{6}A. This modification carries a positive charge and thus can dramatically alter protein-RNA interactions and RNA secondary structures through electrostatic effects. m^{1}A maps uniquely to positions near the translation start site and first splice site in coding transcripts, and correlates with upregulation of translation in general[35,36]. This modification could potentially be reversible and is responsive to various types of cellular stress[35,36]. m^{1}A may promote translation through facilitating non-canonical binding of the exon-exon junction complex at 5' UTRs devoid of 5' proximal introns[37].

m^{1}A blocks Watson-Crick base pairing and blocks most reverse transcription (RT) when using standard Illumina sequencing library preparation procedures. The partial read-through of m^{1}A by reverse transcriptases creates mutations that are useful markers for the presence of m^{1}A. However, caution must be taken when performing deep sequencing of sub-stoichiometric m^{1}A on low abundance RNA with traditional RT enzymes and sequencing approaches. The modified RNA fraction can be severely under-represented due to abortive reverse transcription at or adjacent to the site of m^{1}A[38]. Sequencing approaches that ligate both adaptors prior to cDNA synthesis can be particularly problematic in m^{1}A detection, where sequences derived from RT stops are lost during the subsequent PCR steps. Studies of other RNA modifications that block Watson-Crick base pairing on low abundance RNA species such as mRNA may face the same issue, hindering their potential discoveries in mRNA or other low abundance RNA species. More efficient read-through enzymes that can read these modifications with mutation signatures are highly desirable in the future. Additionally, m^{1}A can isomerize to m^{6}A under basic aqueous conditions, and must be handled under mild conditions without prolonged storage.

m^{6}Am: Adjacent to the 5' cap, the second base in many mRNAs can be 2'-O-methylated. A portion of these bases also bear m^{6}A methylation to form N^6,2'-O-dimethyladenosine (m^{6}Am), deposited by a yet unidentified methyltransferase. The presence of this unique modification was

previously described[39], and was further confirmed from transcriptome-wide m^{6}A-seq[31]. The overall abundance of m^{6}Am in mRNA is low, estimated at 1/30-1/10 of the abundance of internal m^{6}A. The m^{6}A portion of this modified nucleoside was known to be a substrate of FTO[40], with a recent study highlighting that m^{6}Am stabilizes mRNA by preventing DCP2-mediated decapping and microRNA-mediated mRNA degradation[41].

Additional modifications of adenosine, such as further base methylation of m^{6}A to N^6,N^6-dimethyladenosine (m^{6,6}A), or the deposition of bigger, more elaborate chemical groups have been identified in eukaryotes but have yet to be characterized within mRNA.

Cytosine Methylation, Oxidation

m^{5}C: Like m^{6}A**,** methylation at the 5 position of cytosine in mRNA was discovered more than 40 years ago[12,14], though in significantly lesser abundance. Capitalizing on bisulfite methodology utilized for 5-methylcytosine identification in DNA[42], m^{5}C sites were mapped in human mRNA and lncRNA species. Distribution of these modified bases appears to favor untranslated regions, particularly the binding sites for Argonaute proteins I-IV[43]. The tRNA m^{5}C methyltransferase NSUN2 has been identified as the methyltransferase responsible for m^{5}C methylation in several mRNAs and lncRNAs using both methylation-iCLIP (miCLIP, Hussain et al., 2013) and 5-azacytidine-mediated RNA immunoprecipitation (Aza-IP)[45]. m^{5}C is recognized by the mRNA export adaptor protein ALYREF, suggesting a role for this modification in nuclear export of m^{5}C-contaiing transcripts[46]. Of note, this study reports a strong bias for m^{5}C sites 100 nucleotides beyond translation initiation sites, unlike the relatively even distribution previously observed using similar sequencing technologies. This study also mapped m^{5}C in mRNA from different mouse tissues.

hm^{5}C: As with 5-methylcytosine in DNA, m^{5}C in RNA can be oxidized by Tet-family enzymes to 5-hydroxymethylcytosine (hm^{5}C)[47]. In *Drosophila melanogaster*, which lacks DNA hydroxymethylation, hm^{5}C is present in greater than 1,500 mRNAs. hMeRIP-seq revealed the presence of this modification largely in exonic and intronic regions of protein-coding transcripts, dependent on the presence of the only known *Drosophila* Tet ortholog[48]. The abundance and potential roles of hm^{5}C in mammals will be interesting to monitor in the future, as will potential precursors and derivatives of this modification in relevant RNA species.

Isomerization of Uridine

Ψ: Pseudouridine is the most common modification in cellular RNA and an abundant component of rRNA and tRNA[49]. However, its presence in mRNA was largely ignored until recent efforts to map this modification. PseudoU-seq uses selective chemical modification of ψ, which subsequently blocks reverse transcription adjacent to the modified base, identifying and quantifying Ψ in greater than 200 human and yeast mRNAs[50]. Similarly, Ψ-seq identified >300 Ψ-modified mRNAs in human and an additional 41 in yeast[51]. Ψ sites are dynamically regulated by the function of Pus family enzymes, which catalyze the isomerization in response to growth conditions such as heat shock. Ψ is known to affect the secondary structure of RNA, and the function of Ψ in altering stop codon read-through may also be biologically relevant[52,53]. Using a more sensitive chemical labeling followed by pulldown and quantitative mass spec, Li and colleagues estimated Ψ/U at 0.2-0.7% in mammalian cell lines and tissues, and identified over 2,000 sites in human mRNA by CeU-Seq, suggesting this modification is far more prevalent than previously appreciated[54].

Ribose Modification

2'-O-Me: In addition to base methylations, methylation of the 2' hydroxyl is known to exist at the second and third nucleotide in many mRNAs, and as an abundant modification in tRNA and rRNA[39]. The 2' hydroxyl group is a salient feature of RNA; its methylation could have profound impacts on RNA-protein interactions and RNA secondary structures. 2'-O-methylation (2'-OMe or Nm) sites in abundant RNA species such as rRNA could be mapped taking advantage of its higher resistance to alkaline-mediated hydrolysis compared to unmodified nucleosides[55].

A new technology to map 2'-O-Me in less abundant mRNA has been invented. The technique, Nm-seq, is based on ribose sensitivity to periodate cleavage, and leverages the inert nature of 2'OMe to iterative oxidation-elimination-dephosphorylation chemistry. This process allows for enrichment of 2'OMe in low abundant RNA species such as mRNAs, providing a highly sensitive single-base method for detection. Nm-seq uncovered over 7,000 potential methylation sites in human mRNA with a consensus sequence and enrichment within three amino acid codons, indicating extensive ribose methylation in all four bases[56].

Critical to understanding the roles of posttranscriptional modification in mRNA are methods to identify the abundance, sequence context, and cellular dynamics of these distinct entities. Abundant cellular RNAs are heavily modified. Often the use of more than one purification step, such as polyA selection followed by depletion of rRNA as well as size selection are required to purify mRNA for accurate determination of modification status. Quantification of less abundant modifications is best achieved using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) as opposed to analysis by dot-blot or thin-layer chromatography, both of which can suffer from inconsistencies in loading, normalization, and exposure. To this end, methods to provide highly sensitive, quantitative, single-base resolution of RNA modifications remain a crucial goal for the field, and in particular for m^6A, the most abundant and functionally diverse mRNA modification to date.

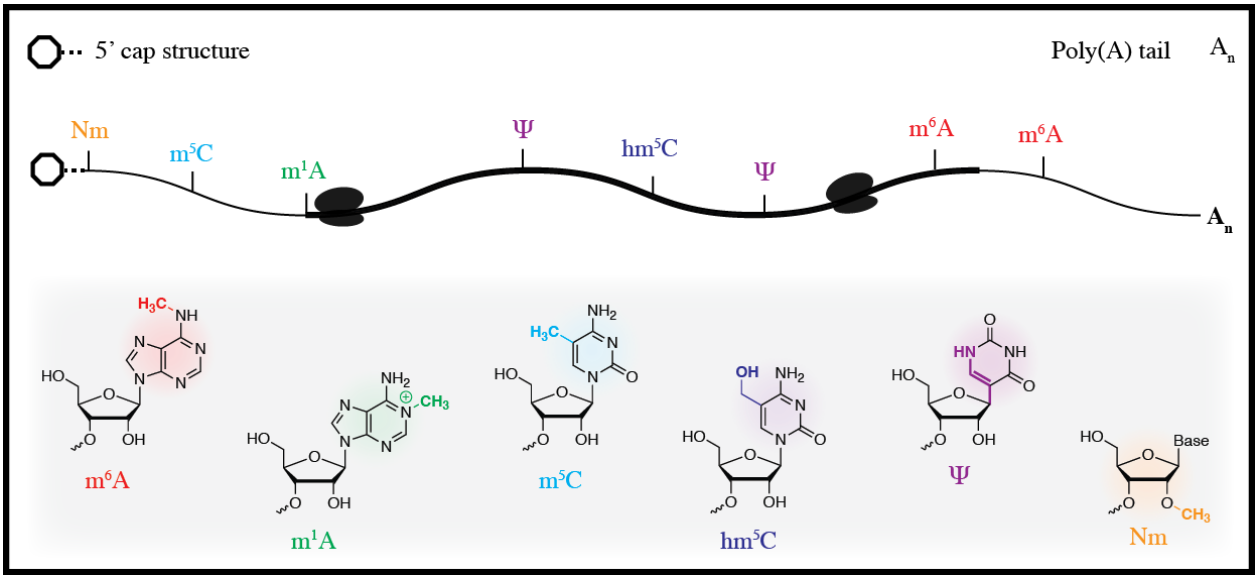

Figure A.1 Chemical modifications in eukaryotic messenger RNA

A schematic representation of common chemical modifications in eukaryotic mRNA transcripts. Several of these modifications map uniquely to the mRNA cap structure, 5' or 3' untranslated regions, or the coding region (bold) of the transcript.

Appendix 1.4 Dynamics of the 'epitranscriptome' – mRNA methyltransferases and demethylases

The epitranscriptome presents a dynamic layer of information, shaped largely by the enzymatic activities of methyltransferases or pseudoU synthases and demethylases. The deposition of m^6A in mammalian mRNA is catalyzed by a heterodimer of METTL3 and METTL14, and regulated by the association of a subunit protein WTAP[57,58]. Recent crystal structures of the METTL3/METTL14 complex have revealed that only METTL3 appears to possess a functional active site within the complex while METTL14 largely functions as a structural scaffold[59–61]. Analysis of the binding sites of this complex suggests that methylation occurs preferentially in coding sequences and 3'UTRs. Additionally, a significant portion of binding sites fall within intronic sequences, suggesting that deposition of m^6A in mRNA takes place co-transcriptionally, perhaps mostly within nuclear speckles[58,62]. Proteomic analysis of these core methyltransferase components by IP-MS revealed an additional factor, KIAA1429, as critical in mediating full activity of the complex[63], the entirety of which is necessary for proper establishment of the cellular m^6A profile (**Figure A.2a**).

m^6A methylation can be removed passively from the transcriptome via degradation of modified RNA or via active demethylation by m^6A demethylases FTO or ALKBH5, both belonging to the AlkB family of dioxygenases known to demethylate N-methylated nucleic acids (**Figure A.2a**). These proteins oxidatively demethylate m^6A *in vitro*, and contribute to m^6A levels in cellular mRNA[25,27]. FTO has also been shown to preferentially demethylate a further modified form of m^6A, m^6Am, adjacent to the 5' cap of mRNA when compared to m^6A[40,41]. Internal m^6A residues are much more abundant within mRNA, and contributions from demethylation of both m^6A and 5' m^6Am by FTO play roles in modulating mRNA metabolism. The m^6A demethylation activity of ALKBH5 critically impacts mRNA nuclear export and spermatogenesis, and both enzymes participate in the various disease mechanisms related to cancer[64–67]. A recent study discovered that the METTL3-METTL14 complex is rapidly recruited to the DNA damage site created by UV irradiation, which mediates local RNA m^6A methylation. This process facilitates recruitment of DNA damage repair polymerase κ, and can be reversed by FTO within a short

period of time[68]. Each of these studies demonstrates that methyltransferase and demethylase actively determine m^6A methylation dynamics in both normal and acute responses to cellular stimuli.

m^1A deposition in tRNA is largely dependent on secondary structure[69]. As m^1A in mRNA occurs in structured, GC-rich regions, a tRNA methylatransferase with moonlighting activity in mRNA may be responsible for this modification in coding transcripts[35]. Another ALKB family protein, ALKBH3, is capable of demethylating m^1A in mRNA, suggesting that this modification may also serve a dynamic regulatory feature[36]. Methyltransferases for both 2'O-methylations at the 5'cap have been identified[70,71], although no enzyme for internal ribose modifications nor an active demethylation process has been reported. The methyltransferase responsible for further methylation of A_m to m^6A_m at the 5' cap is also unknown.

The tRNA methyltransferase NSUN2 has been identified as a mediator of m^5C in nearly 300 mRNAs by miCLIP[44], though fewer coding transcripts were identified as targets using complementary methods[43,45]. m^5C can be oxidized in *Drosophila* by a conserved Tet ortholog CG43444 (dTet) to generate hm^5C in mRNA[48]. The potential of hm^5C for further oxidation and eventual decarboxylation provides m^5C a plausible route to reversibility, though evidence for this has yet to be reported.

Notably, RNA modification enzymes commonly exhibit substrate promiscuity. For example, Ψ in mRNA can be attributed in part to several pseudouridine synthase (PUS) enzymes conserved across eukaryotic genomes[50,51,54]. Perturbations of Ψ sites observed in mRNA in response to environmental stimuli suggest that mRNAs are indeed physiological targets of these enzymes although they are known to also act on tRNA and small nuclear RNAs (snRNAs) substrates in cells. The installation of a carbon-carbon bond between the base and sugar upon isomerization to Ψ however, suggests that this modification is not readily reversible.

Mammalian mRNA carries additional modifications often at low abundance; some of these modifications are observed in rRNA, tRNA and other non-coding RNAs, and could be installed on mRNA by enzymes known to also modify rRNA or tRNA, with the mRNA modification sites perhaps sharing sequence or structural features similar to their non-coding RNA counterparts. A portion of these mRNA modifications could have functional consequences through evolution; a recent study demonstrated that primate mRNA sequences evolve to accommodate modification motifs for m^6A installation in order to gain fitness advantage[72]. However, it is also possible that

some of these modifications are simply low-abundance byproducts derived from off-target activities of tRNA modification enzymes, or incorporated during transcription from pools of modified nucleoside-triphosphates. Therefore, functional characterizations of cellular enzymes that regulate post-transcriptional mRNA modifications are essential to understanding the biology of these enigmatic marks, and will continue to benefit from analysis of RNA modifying enzymes working on non-coding RNA targets.

Appendix 1.5 Properties of mRNA modifications – structure and function

Chemical modifications in RNA affect the defining characteristics of transcripts by altering charge, base pairing, secondary structure and protein-RNA interactions. These properties in turn shape the outcome of gene expression by modulating RNA processing, localization, translation, and eventual decay.

m^6A, the most common modification in mRNA, occupies an exocyclic amine which participates in Watson-Crick base pairing. Watson-Crick base pairing of m^6A with opposite U would force rotation of the carbon-nitrogen bond to display the methyl group at the *anti* conformation, which destabilizes the RNA duplex by on average ~1 kcal/mol per m^6A, thus the presence of m^6A leads to locally unstructured transcripts[73]. This effect also modulates secondary structure *in vivo*[74], and predisposes these unstructured regions for recognition by proteins such as HNRNPC and HNRNPG[75–77]. The m^6A modification directly recruits the binding of m^6A-specific proteins of the YTH domain family, initially identified by RNA affinity chromatography using a methylated RNA probe[28]. These proteins bridge methyl-selective RNA binding with a myriad of cellular processes, and produce m^6A-dependent regulation of pre-mRNA processing, microRNA (miRNA) processing, translation initiation and mRNA decay (**Figure A.2b**).

Starting from transcription, mRNAs are heavily coated with RNA binding proteins. Within the nucleus, several proteins are known to bind precursor RNAs with selectivity for m^6A. YTHDC1 (also known at YT521-B), a member of the mammalian YTH proteins found in the nucleus, promotes inclusion of alternative exons via interactions with members of the splicing related SR-protein family[78]. YTHDC1 also binds to highly expressed lncRNAs such as XIST,

which requires m^6A-specific RNA-protein interactions to mediate X-chromosome silencing[79]. Three members of the HNRNP (Heterogeneous Nuclear RiboNucleoProtein) family also function to regulate the processing of m^6A-modified transcripts. HNRNPA2B1, along with METTL3, co-regulate alternative splicing events as well as the generation of miRNAs from methylated precursors[80], while HNRNPC and HNRNPG mediate splicing outcomes on methylated transcripts by recognizing and binding to m^6A-dependent structural switches[75,77].

Mature mRNAs with m^6A methylation are subject to regulation in the cytoplasm by the remaining YTH family proteins with documented selectivity for m^6A. YTH Domain Family 1 (YTHDF1) associates with initiating ribosomes, delivering its target mRNAs for enhanced translation efficiency in HeLa cells[81]. A second YTH family protein, YTHDF2, directly recruits the CCR4-NOT deadenylase complex and accelerates degradation of methylated transcripts[82,83]. While accelerated decay globally shapes the profile of methylated mRNAs, some transcripts exhibit increased half-life time upon m^6A methylation. This suggests additional pathways for stabilization of these mRNAs, potentially through additional effector proteins[57]. Under heat shock conditions, YTHDF2 localizes to cell nuclei where it directs cap-independent translation of heat shock response transcripts[84]. Under these same conditions, the 43S preinitiation complex is recruited to m^6A-methylated transcripts via interactions with the eIF3 subunit[85]. YTHDF3 mediates translation along with YTHDF1 by interaction with a common set of ribosomal proteins, as well as decay of mRNA targets by associating directly with YTHDF2[86,87], and may play additional cell-type specific roles and functions based on m^6A location within transcripts.

Along with dedicated effector proteins, enzymatic components responsible for establishment of the m^6A epitranscriptome facilitate downstream functions of mRNA methylation. As both the methyltransferase complex and demethylases exist mostly in the nucleus, their presence may affect RNA processing outcomes as a result of nuclear organization and recruitment of accessory processing factors[27,63,78]. Additionally, the core methyltransferase METTL3 enhances translation of bound RNA independent of catalytic activity by direct recruitment of eIF3 to the translation initiation complex[88]. The potential for methyltransferase and demethylases proteins to interact with multifunctional adaptor proteins greatly facilitates transcript sorting through RNA methylation.

m^1A is a unique base methylation because it blocks Watson-Crick base paring and introduces a positive charge. In mRNAs, m^1A exists within highly structured 5' UTRs, suggesting

that it may function to alter predicted secondary structure. Within loop structures, this charge may serve to stabilize interactions with the phosphate backbone of RNA. m^1A methylation in transcripts correlates with increased translation, perhaps due to accessibility or direct recruitment for initiation and elongation factors[35,36]. The positive charge of this modification makes it amenable to specific protein-RNA interactions and unique RNA-RNA interactions, the biological impact of which is currently unknown.

Tandem 2'-O-methylation and N^6-methylation of adenosine adjacent to the 5' cap confers stability in mRNA transcripts by preventing DCP2-dependent decapping and microRNA degradation. This stark contrast to m^6A suggests that location of adenosine methylation is critical in determining function; both m^6Am and m^6A sites in the 5' UTR correlate with increased translation. The role of FTO in regulating both of these marks may depend largely on location and substrate abundance for the enzyme.

The patterns of m^5C distribution on mRNA with respect to cis-acting regulatory motifs and miRNA/RISC binding sites suggest that this modification may be involved in post-transcriptional regulation of mRNA metabolism[43]. NSun2-medited methylation is required for the processing of noncoding vault RNAs (vtRNAs) into small vault RNAs (svRNAs), but downstream consequences in coding transcripts have not emerged as a result of this defect[44]. Recently, ALYREF was shown to recognize m^5C in mRNA via a methyl-specific RNA-binding motif, and regulate the export of bound transcripts in an NSUN2-dependent manner[46]. hm^5C, derived from Tet-dependent oxidation of m^6C, preferentially marks mRNAs within coding regions and favors translation of *Drosophila* transcripts[47,48].

Similar to m^6A methylation, installation of Ψ in mRNA can encode additional information. Compared to U, Ψ can coordinate an additional water molecule allowing for hydrogen bonding to the adjacent phosphate backbone, which could rigidify regions containing Ψ in duplex and single-stranded RNA. This property also enhances the base stacking of Ψ by favoring a 3'-endo conformation of ribose, further restricting the flexibility of the residue[89]. Placement of Ψ within nonsense codons allows for complete read-through, generating a protein product with a C-terminal extension[52,53]. Regarding this stop codon read-through, the effect of Ψ in mRNA is a form of RNA editing, leading to a change of information encoded by the primary RNA sequence. Despite these intriguing effects Ψ could exert on mRNA structure and translation, the function of this modification in mRNA in mediating biological processes has yet to be reported.

2'O-methylation of RNA can have dramatic effects on structure and stability, as the modifications masks the hydrophilic hydroxyl that largely defines the macromolecule. Nm residues function to enhance duplex stability of RNA-RNA hybrids[90], and could thus promote stability and efficacy of RNA-based therapeutics. In *vivo*, 2'OMe is typically found in unstructured CDS regions of mRNA, with half of all sites falling within 50 nucleotides of the nearest splice site, suggesting a link between structural elements and mRNA processing outcomes. Interestingly, Nm sites concentrate to six codons that encode three amino acids: glutamate, lysine, and glutamine, perhaps dependent on the RNA-binding properties of specific methyltransferase[56]. This suggests that Nm in mRNA may affect translation efficiency, a concept that has been demonstrated in modified bacterial mRNA[91].

Research has demonstrated that diversity in mRNA modification can tune nearly every aspect of mRNA function. m^6A, being the most abundant and well-studied mRNA modification, highlights the ability of a small chemical modification to determine fundamental properties as well as metabolic outcomes. Unsurprisingly, these properties are closely connected, as recent work linking Pol II transcription status, m^6A deposition, and translation efficiency has shown[92]. As work on more rare modifications continues, we will likely discover new proteins and mechanisms that amplify chemical changes into more profound biochemical and cellular consequences.

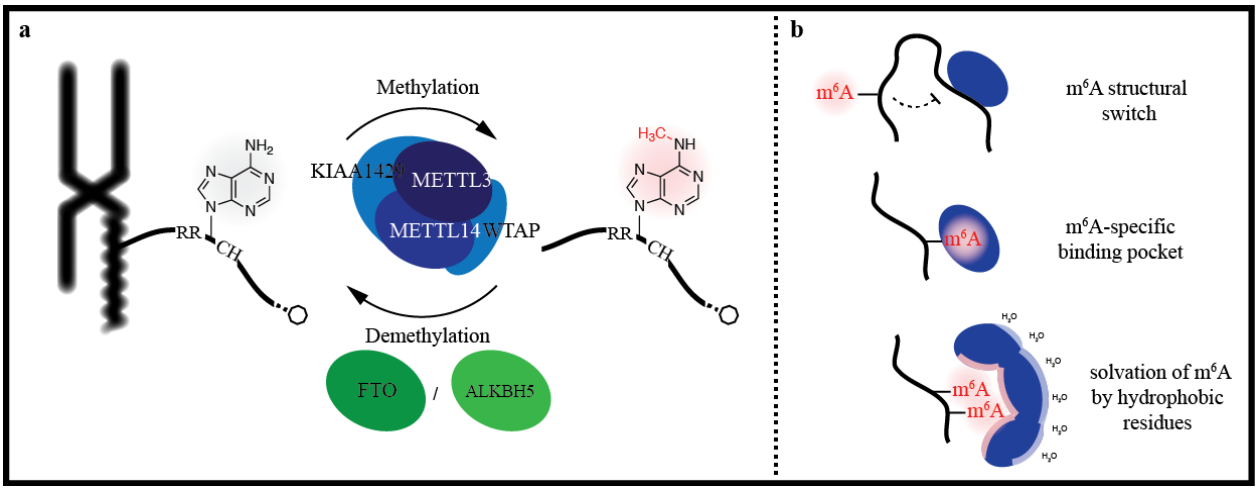

Figure A.2 Active m^6A methylation, demethylation, and downstream consequences for protein-RNA interactions

Figure A.2 continued

(a) m^6A is installed co-transcriptionally by a complex consisting of METTL3, METTL14, WTAP, and KIAA1429. Each of these components binds mRNA and is required for complete methylation, but only METTL3 contributes to the catalytic activity of the complex. (b) m^6A methylation affects protein-RNA interactions through multiple mechanisms. Methylation can perturb the secondary structure of mRNA, exposing or masking potential RNA-binding motifs (top). Selective m^6A-binding proteins exhibit increased affinity for methylated mRNAs, and in turn incorporate these transcripts into various steps of mRNA metabolism (middle). Methylation itself introduces hydrophobic moieties. In the case of m^6A, association with hydrophobic amino acid side chains or low complexity regions of proteins may assist in solvation of the modified base (bottom).

Appendix 1.6 Reading RNA modifications

The identification of cellular factors that mediate the outcomes of modified RNA messages are crucial to our understanding of the biological roles of mRNA modifications, as illustrated in studies of m^6A "reading" processes. The "reading" of a RNA modification can come in several different forms. The modification could be directly recognized by a binding pocket as shown by the binding of m^6A by the YTH domain proteins[93–96] The modification could induce a structural switch in mRNA, leading to binding of RNA-binding proteins that do not directly interact with the modification, as shown in m^6A-switches [75] and likely also in cases involving m^1A, ψ, and 2'-OMe. Modifications such as ψ, m^5C and 2'-OMe could affect the rigidity of RNA structures, while ψ and 2'OMe also significantly alter hydrogen-bonding properties at the modification site; all these factors are expected to contribute to altering protein-RNA interactions of the modified RNA. When methylated RNA probes were used to pull down potential selective m^6A-binding proteins, proteins (i.e. HNRNP family) that do not appear to possess obvious modification-binding pockets were also pulled down. Some of these may involve an RNA structural switch; however, the presence of other "reading" mechanisms also warrant consideration.

One factor neglected in almost all RNA modification studies, regardless on mRNA, rRNA, or tRNA, is the solvation effect (or hydrophobic effect)[97]. Hydrophobic modifications such as a methyl group would induce solvation penalties when exposed in aqueous environment. This same effect partially contributes for m^6A, m^5C, and 2'-OMe to favor stacked conformations in RNA, and binding of m^6A by YTH domains within a hydrophobic pocket. Similarly, the interaction of

these methylations with hydrophobic side chains of a RNA-binding protein lacking a clearly defined pocket could minimize solvation penalties, leading to favored binding involving modified RNAs, albeit with lower selectivity compared to the YTH domain. We propose that some of the m^6A-selective binding may be mediated in part through this effect in the absence of clearly defined base-binding pockets. Such an effect may be enhanced *in vivo* within a protein complex or a compact cellular granule. Certain HNRNP and SRSF type RNA-binding proteins tend to enrich m^6A inside cells; however, the discrete RNA-binding domain of these proteins exhibit low to almost non-existing selectivity to methylated RNA probes *in vitro*. The preferential enrichment of methylated RNA, in particular those with multiple adjacent methylation sites, in more hydrophobic compartments formed through cellular aggregation of these proteins may partially account for the *in vivo* selectivity (**Figure A.2b**). 5'-UTR m^6A methylation has been shown to promote translation initiation although direct reader proteins underpinning this activity have yet to be shown. The interaction of methylated 5' UTR with translation initiation complex to reduce solvation penalties could provide a source of selectivity. Noted that certain tRNAs tend to preferentially associate with translation machinery when modified in the anti-codon loop, in which the reduction of solvation penalty of the hydrophobic adducts may contribute to the preferential ribosome binding[98–100].

The knockdown of METTL3 in mammalian cells has been shown to increase the stability of a majority of target mRNAs, but also reduce cellular levels of a substantial portion of target transcripts[58] This result suggested that the m^6A methylation could reduce transcript stability, likely to be mediated through YTHDF2, as well as stabilize a portion of other transcripts through yet-to-be-identified reader proteins or YTHDF1/YTHDF3. It is possible that the modification site on the transcript and/or potential nuclear pre-sorting of modified transcripts (during transcription) determine their cytoplasmic fates via interactions with different reader proteins. How methylated transcripts gain stability as revealed in a recent study[64], and how specificity is achieved are intriguing questions to be investigated in future studies.

Appendix 1.7 The essential nature of the m^6A methylation in mRNA

Understanding the function of mRNA modification has benefitted immensely from identification and genetic perturbation of methyltransferases, synthases, demethylases, and

effector proteins. The m^6A methylation, for which the greatest number of regulatory factors is known, has been shown to be essential in mammalian development and involved in various human diseases. In *Drosophila*, the methylation is not essential but is critical for sex determination and neuronal functions[101,102].

Methyltransferase components METTL3 and METTL14 are both required for the m^6A methylation in mammalian cells, despite only METTL3 mediating its catalytic activity. In mouse embryonic stem cells (mESCs), depletion of either Mettl3 or Mettl14 reduces m^6A methylation in transcripts of developmental regulators and increases their transcript stability[57]. Loss of *Mettl3* delays turnover of self-renewal factors, preventing differentiation into downstream lineages[103]. While the *Mettl3*$^{-/-}$ mice are not viable, cells derived from early embryos are unable to resolve their naïve pluripotency due to extended transcript lifetime in the complete absence of methylation[104]. In each case, the methylation appears to be critical in embryo development and cell differentiation due to its role in governing stability of key regulatory transcripts. Our own investigations of the YTH family reader proteins uncovered m^6A as a mark that labels groups of transcripts for accelerated metabolism: the nuclear readers appear to promote processing and export of methylated transcripts, whereas cytoplasmic YTHDF1, 2, and 3 readers accelerate translation and decay of their target transcripts. Together, these results suggest a most critical role of m^6A in marking groups of transcripts for coordinated metabolism in response to cellular signaling and/or environmental cues.

During early embryo development waves of master transcriptional factors (TFs) control transcriptome switching and guide cell differentiation programs. Each TF could directly or indirectly affect activation of hundreds or more transcripts that follow the same trend of activation and subsequent clearance when the next TF wave takes over in order for cells to differentiate into new states. While the activation of these cell-state-specific transcripts can be coordinated through transcription activation mediated by master TFs, the coordinated translation, and particularly decay of the same set of transcripts upon cells entering new states require mechanisms for synchronization. Considering the number of TF waves required to satisfy development of hundreds of tissue or cell types in mammals with each wave affecting distinct sets of transcripts, sequence alone will not be able to or enough encode all information for distinct transcripts to participate in diverse transcriptome switching events. We propose that mRNA m^6A methylation offers such a critical mechanism, which also explains its essential nature in mammals. While individual mRNAs

have wide distribution of their respective stability, metabolism, and translation, upon m^{6}A methylation during transcription (the selectivity is presumably endowed by TF recruitment of the methylation machinery) a set of transcripts are synchronized for coordinated metabolism, translation, and decay, allowing timely and coordinated protein synthesis and transcriptome switching during cell differentiation (**Figure A.3**).

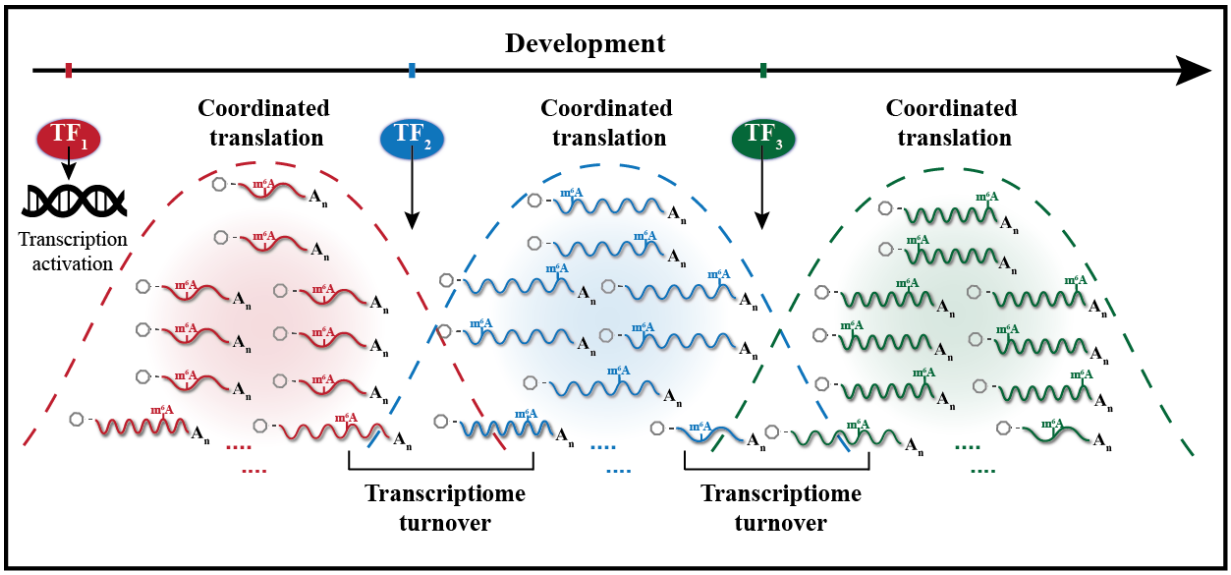

Figure A.3 RNA modification groups transcripts for cellular processes

Developmental programs require timely switching of the cellular transcriptome to bring about phenotypic changes. Master transcription factors (TFs) largely define the cellular pool of mRNA. These TFs activate transcription of tens to hundreds of transcripts at different stages of development. mRNA m6A methylation is a mechanism to group these distinct sets of transcripts together for coordinated translation and decay. The methylation thus provides an additional identity to these transcripts, which can be erased or reset, on top of their sequences for coordinated post-transcription regulation.

To test this hypothesis, we have investigated a prototype of transcriptome switching during early embryo development in vertebrates: the maternal to zygotic transition (MZT) in zebrafish. Maternal transcripts are rapidly degraded and the transcriptome is replaced by newly synthesized zygotic mRNA during MZT[105]. In zebrafish a portion of maternal mRNAs are m^{6}A methylated, and rapidly cleared by Ythdf2. In the absence of Ythdf2 this clearance is delayed, preventing timely initiation of MZT resulting in prolonged developmental delay[106]. This critical role of mRNA m^{6}A methylation may not be limited to early embryo development and could impact a

wide range of cell differentiation and tissue development events, which renders this mark essential to mammalian systems. Similarly, RNA modification provides perhaps the best means to define sets of transcript groups for coordinated response in other signaling or stress events. The diverse chemical marks now known to exist in mRNA offer the possibility to simultaneously mark multiple groups of transcripts in response to different cues in a mammalian cell.

Naturally, methylation-dependent processes can also be controlled via active demethylation. FTO, initially identified by genome-wide association studies for diabetes predisposition, is required for proper splicing in route to adipogenesis[107–110], while ALKBH5, is required for spermatogenesis in mice[27]. More acute cellular events, in addition to developmental cues, require the activity of RNA demethylation. FTO for example, demethylates genes in the 5' UTR under normal conditions, but fails to associate under heat shock conditions because of YTHDF2 binding[84]. Hypoxic conditions affect m^6A-dependent regulation by inducing ALKBH5 expression and promoting metastasis of breast cancer cells[66,111], while FTO plays an oncogenic role in the development of acute myeloid leukemia[64]. Uncoupling the catalytic activity of these enzymes from their roles in cellular organization and protein recruitment will further define the unique roles for m^6A methylation in shaping transcriptomes. Additionally, understanding the selectivity of methyltransferase, demethylases, and reader proteins to transcript groups for coordinated metabolism will be essential for future understanding of their precise contributions to differentiation and development processes and roles in human diseases.

RNA viruses harness the same dynamics that affect cellular RNA outcomes upon entry into mammalian host cells. The RNA genome of HIV-1 contains numerous sites of m^6A methylation, as does the mRNA coding for viral proteins. These sites are regulated by the METTL3/METTL14 methyltransferase as well as m^6A demethylases, and recognized by YTH proteins *in vivo*. YTH binding to HIV-1 genomic and messenger RNA affects reverse transcription, as well as RNA and protein production in HIV-1 infected cells, though the impacts on the viral lifecycle require further investigations[112–114]. m^6A dynamics play a role in Zika virus infection as well as other *Flaviviridae*, with m^6A-sequencing revealing a conserved pattern of modification across several viral genomes[115,116]. Viral utilization of RNA methylation may represent a novel route to prevent infection and replication in host cells, and highlights the range of processes that can be manipulated by reversible modification of coding transcripts.

Because the m^6A methylation on mRNA is a fundamental process critical for cell differentiation, aberrant RNA methylation could contribute to human diseases. Defects in NUDT16-mediated RNA decapping are known to occur in leukemia[117], as are several examples of defects within the internal epitranscriptome of mRNA. For instance, tumor cells could hijack aberrant RNA methylation, either through up- or down-regulation of methyltransferase components, demethylases, or readers, or mutations of these effectors to gain survival and proliferation advantages as indicated in recent studies. For instance, through demethylation of internal m^6A in critical oncogenic transcripts, upregulation of FTO enhances leukemogenesis and inhibits all-trans-retinoic acid (ATRA)-induced AML cell differentiation[64]. Similarly, inhibition of FTO in glioblastoma stem cells suppresses tumor growth and prolongs the lifespan of grafted mice[65]. Expression of ALKBH5 enhances tumor initiation in breast cancer via regulation of key transcription factors[66,67]. It is interesting to note that in a subset of AML and glioblastoma, high levels of m^6A demethylases (downregulation of mRNA m^6A methylation) promote cancer progression. Inhibition of these demethylases with small molecule inhibitors could reduce cancer progression. As α-ketoglutarate-dependent demethylases they could also be inhibited by naturally occurring oncometabolite 2-hydroxyglutarate (2HG), which could lead to benign outcome of these cancers. Interestingly, while 2HG has been shown to promote cancer initiation in subsets of AML and glioblastoma, its presence also correlates with clinical benign outcome in certain glioblastoma and AML cases[118–122], suggesting contributions of m^6A mRNA methylation and its inhibition at the stage of cancer progression. Studies of aberrant RNA methylation in human diseases such as cancer are fast evolving and will further aid our understanding of roles of RNA modifications in human physiology.

Appendix 1.8 Modifications in abundant noncoding RNAs

Thus far, this discussion of RNA modification has focused on messenger RNAs, the regulation of which has obvious and immediate effects on the cellular transcriptomes and eventual proteome. However, as the utility of non-coding RNAs becomes more and more well-studied, the roles their modification have become areas of significant interest. Indeed, well-studied ncRNAs such as let-7 miRNA, XIST, and MALAT1 contain numerous chemical modifications that contribute to their respective roles in cancer[123].

tRNA: tRNAs are the most heavily modified RNA species with regards to both number, density and diversity. Nearly 1 in 5 nucleotides are modified in mammalian tRNA, and over 50 unique modifications have been identified in eukaryotes[124]. The modifications range from simple thiolations and base or sugar methylations to extensive addition of sugars, amino acids, and complex organic adducts. These diverse modifications are catalyzed by a myriad of nuclear and cytoplasmic enzymes, which can act at a single site in a single tRNA or at multiple sites across several tRNA species. Complex modifications often require step-wise installation by a cascade of enzymes (e.g. wybutosine and mcm^5s^2U) or methylation followed by deamination of the same base, a form of RNA editing[125].

The anti-codon loop is one hot-spot of modification. Modifications of the anti-codon loop aid in translation by preventing frameshifting, expanding codon recognition, and strengthening the codon-anticodon interaction (**Figure 4A**). Almost every tRNA is modified either at position 34 or position 37 or both, corresponding to the wobble position and the nucleotide 3' of the anticodon. Position 34 is important for accurate and efficient decoding; modification at this first anticodon (wobble) position can restrict (e.g. mcm^5U34) or expand (e.g. cmo^5U34 or I) the decoding of a tRNA species by affecting the conformational dynamics of the anticodon stem loop or the tRNA-mRNA Watson Crick base-pairing[99]. Position 37 is also heavily modified. Perhaps the best example of this modification is the presence of wybutosine at position 37 of phenylalanine tRNA. Extensive aromatic stacking of yW37 confers conformational stability of the loop, and prevents pairing with U33 to keep the anticodon open. This lack of flexibility can also prevent four-base anticodon pairing and is necessary to prevent frameshifting[126]. While yW37 is present in mammals, it is notably absent in Drosophila, which has led to the suggestion that certain organisms may utilize frameshifting as a mechanism to increase coding diversity[127]. Similar stabilization strategies occur with other tRNA modifications, such as modifications of A37 to i^6A or t^6A to direct codon specific translation and maintain translational accuracy and efficiency.

Outside of the anticodon loop, modifications are known to influence the structure of tRNA. The clearest example is human mitochondrial $tRNA^{Lys}$. tRNAs lacking m^1A9 do not fold into the canonical cloverleaf structure; instead, these hypomodified tRNAs adopt an elongated structure due to A9-U64 base pairing that extends the acceptor stem. The methylation of A9 is sufficient to induce the cloverleaf folding by disrupting this base pairing[128]. Although completely unmodified tRNA has been shown to be less stable than fully modified tRNA, the study of the effects of

individual modifications on tRNA structure is not straightforward in most instances and remains to be elucidated. One recent example highlights the effect of m^5C in tRNA stability, in which $NSUN2^{-/-}$ cells accumulate 5' tRNA fragments and have an impaired translational response to cellular stress[129]. Deficiency in NSUN2 results in microcephaly and other neurological disorders in humans and mice though this tRNA modification based mechanism, in which NSUN2 deficient brains become susceptible to oxidative stress [130].

Besides translation and structure, tRNA modifications have been shown to have a wide variety of functions in many aspects of tRNA biogenesis and function. Modifications can act as quality control in the biosynthesis of tRNAs. For example, yeast $tRNA_i^{Met}$ lacking m^1A58 are targeted for degradation in the nucleus[131], and $tRNA^{Val(AAC)}$ lacking m^7G46 and additional modifications are targeted for rapid tRNA decay as well[132]. Modification at the wobble position in yeast tRNAs has also been shown to affect ribosome A-site loading[133]. In *Leishmania*, a wobble modification can affect the subcellular localization of $tRNA^{Glu}$; tRNAs carrying mcm^5U are imported into the mitochondria whereas tRNAs carrying mcm^5s^2U are not[134].

Until recently, tRNA modification was thought to be stoichiometric and static. However, recent studies in yeast and human tissue culture have shown the opposite; tRNAs can be partially modified and these modifications are dynamic. The application of recently developed sequencing methods revealed that partial modification can occur in several m^1A, N^1-methylguanosine (m^1G), and N^3-methylcytidine (m^3C) sites in tRNA and among different tRNA species in human cell culture[135]. In yeast, stress can modulate the overall levels of modifications such as m^5C, 2'O-methylcytidine (Cm), and N^2,N^2-dimethylguanosine ($m^{2,2}G$) as measured by LC/MS-MS, with different stresses causing different up- or downregulated patterns of modification[136]. For instance, a dynamic m^5C34 modification in $tRNA^{Leu(CAA)}$ was shown to enhance the translation of mRNAs enriched with UUG codons under oxidative stress[137]. Although LC/MS-MS cannot readily identify the specific position in tRNA species that is quantitatively affected, it can simultaneously examine the overall levels of many types of modifications, whereas sequencing is currently restricted to the study of ~10 modification types.

More recently, the first tRNA demethylase, ALKBH1, was identified in human cells[138]. ALKBH1 is responsible for the demethylation of m^1A58 in the TΨC-loop of tRNA; it regulates translational initiation through adjusting the levels of $tRNA_i^{Met}$, as well as translational elongation through adjusting the affinity of a dozen tRNA species to the elongation factor eEF1A. With the

first identified tRNA demethylase, the field has opened to the discovery that more demodification enzymes may exist to regulate modification status in response to cellular signaling or stress in order to reprogram tRNA stability and translation, recently shown for related demethylase ALKBH3[139,140].

rRNA: rRNA is also extensively modified. Approximately 2% of rRNA nucleotides are modified, corresponding to over 100 sites of modification in yeast and over 200 sites in humans. While the number of modifications is large, the diversity of modifications is small; most modifications are 2'-O-methylation of the sugar and Ψ (~50 each in yeast, and ~100 each in humans), although around 10 base modifications have also been identified in both humans and yeast[141,142].

Modifications in eukaryotic rRNA are primarily installed through RNA-dependent mechanisms that rely on a guide RNA to direct protein enzymes to the site of modification. Box C/D snoRNAs direct the installation of 2'-O-methylations while Box H/ACA snoRNAs direct the installation of Ψ (Reviewed in Watkins and Bohnsack, 2012). RNA-independent mechanisms exist in yeast but are responsible for only two modifications. Pus7 installs Ψ at position 50 in 5S RNA and Spb1 installs the 2'O-methylation at position 2922 in 25S RNA[142]. Stand-alone proteins install base modifications throughout the ribosome. SnoRNA-guided pseudouridylations and 2'-O-methylations are thought to occur co-transcriptionally or in early stages of ribosome biogenesis [144,145], while base modifications likely occur later in ribosome biogenesis. However, as most modifications are buried within the ribosome, the modifications must occur before significant folding of the rRNA and the maturation of the ribosome (**Figure 4B**).

Modifications are not randomly distributed throughout rRNA. Instead, the modifications cluster around functional sites in the rRNA, including the decoding site and the peptidyl transfer center (PTC), suggesting their functional relevance[146]. Functional studies of these modifications in yeast have shown that deletion of a single modification rarely has an effect on cell growth or phenotype, although some exceptions exist. Instead, preventing the formation of two or more modifications within a cluster is typically required to observe a measurable phenotype. Cumulatively, rRNA modifications have been shown to affect cell growth and drug sensitivity as well as ribosome biogenesis, abundance, structure, and activity (Reviewed in Sloan et al., 2016)[142].

Certain Ψ and 2'-O-methylation modifications have been shown to occur at substoichiometric levels; that is, specific sites within rRNA are only partially modified. In yeast,

over 10% of sites are modified less than 85%[147] while in human cells, about one-third of 2'-O-methylations are substoichiometric[148]. In one case, the cause of a fractional 2'-O-methylation appears to be the abundance of the snoRNA responsible for its installation[149]; however, for most modifications, the cause of partial modification has not yet been determined. Fractional modifications contribute to ribosome heterogeneity, and ribosomes with different modification status could have distinct functions and serve to translate a subset of mRNAs as potentially "specialized" ribosomes, which may depend on rRNA modification status. This could be useful in response to stress; a cell stress could impact rRNA modifications and therefore function of the ribosome. Indeed, two yeast rRNA Ψ residues have been shown to be induced by post-diauxic growth[50] and heat-shock[51]. While the exact function of these inducible Ψ bases has not been determined, these observations support the idea that rRNA modifications are dynamic and could serve to alter ribosome function.

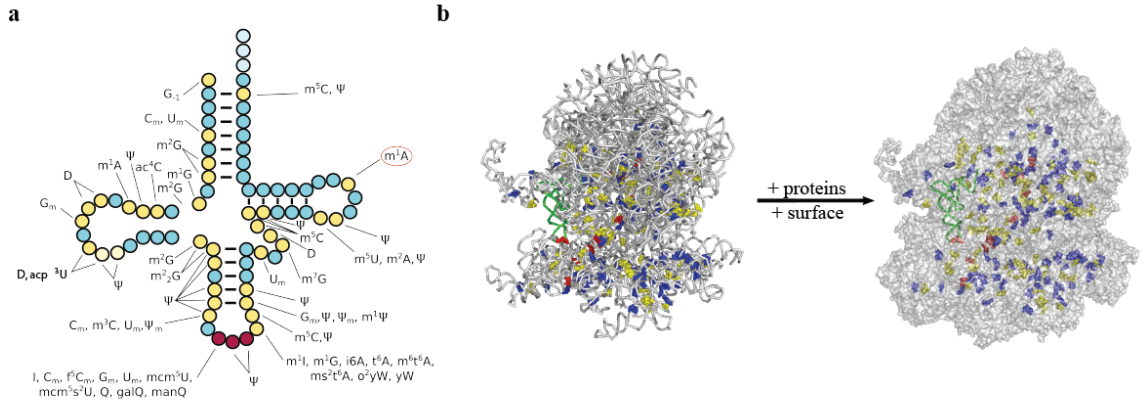

Figure A.4 Landscape of tRNA and rRNA modifications

(a) All annotated mammalian tRNA modifications were pulled from Modomics (http://modomics.genesilico.pl/). Blue residues have no modification annotated while yellow residues are modified in at least one mammalian cytosolic tRNA (*Bos taurus*, *Homo sapiens*, *Mus musculus*, *Oryctolagus cuniculus*, *Ovis aries*, *Rattus norvegicus*). The anticodon, a hot spot of modification, is highlighted in red. m^1A58, the first tRNA modification shown to be reversible, is circled. Modifications that occur in the variable loop were left out for simplicity. m^2G - N^2-methylguanosine; $m^2{}_2G$ - N^2,N^2-dimethylguanosine; m^1G - 1-methylguanosine; ac^4C - N^4-acetylcytidine; m^1A - 1-methyladenosine; D - dihydrouridine; acp^3U - 3-(3-amino-3-

Figure A.4 continued

carboxypropyl)uridine; m^3C - 3-methylcytidine; I - inosine; m^1I - 1-methylinosine; mcm^5U - 5-methoxycarbonylmethyluridine; mcm5s2U - 5-methoxycarbonylmethyl-2-thiouridine; Q - queuosine; galQ - galactosyl-queuosine; manQ - mannosyl-queuosine; f^5Cm - 5-formyl-2'-O-methylcytidine; t^6A- N6-threonylcarbamoyladenosine; ms^2t^6A - 2-methylthio-N6-threonylcarbamoyladenosine; m^6t^6A - N^6-methyl-N^6-threonylcarbamoyladenosine; i^6A - N^6-isopentenyladenosine; o^2yW - peroxywybutosine; yW - wybutosine; $m^1\Psi$ - 1-methylpseudouridine; Ψ_m - 2'-O-methylpseudouridine; m^7G - 7-methylguanosine; m^5C - 5-methylcytidine; m^2A - 2-methyladenosine; m^5U - 5-methyluridine. (b) All annotated *Homo sapiens* rRNA modifications were pulled from Modomics (http://modomics.genesilico.pl/) and mapped onto the cryo-EM structure of the human ribosome (4ug0). An E-site tRNA (green) is shown for orientation. 2'-O-methylations are shown in blue, pseudouridinylations are shown in yellow, and base modifications are shown in red. For simplicity, 2'-O-methyl-pseudouridine is shown in blue. The rRNA ribbon diagram shows the prevalence of modification (~2% of total RNA residues). As the surface rendering incorporating ribosomal proteins shows, the large majority of the modifications are buried within the core of the ribosome.

Despite the existence of partial and inducible modifications, the likelihood that demodification enzymes exist to alter rRNA modification fraction from mature ribosomes seems unlikely under normal growth conditions. As noted above, most of the modifications are buried within the ribosome and would not be easily accessible to a demodification enzyme. It is certainly possible that rRNA modification levels could be dynamically modulated, but any changes in the modification level of rRNA would likely have to occur before ribosome assembly is complete. Changes in the installation machinery or a demodification enzyme that acts early in ribosome biogenesis could serve to modulate rRNA modification levels (and potentially rRNA function). Potential demodification may also be useful to mark damaged ribosomes under cellular stress to facilitate their degradation.

snRNA: Finally, spliceosomal RNAs (snRNAs) are also extensively modified. Like rRNA, the predominant modifications in mammalian snRNA are Ψ and 2'-O-methylations in addition to a few base modifications[150,151]. The number of modifications per snRNA varies, with U2 having the highest number of modifications (the human U2 has over 20 ψ, 2'O-methylations, and a single base methylation, m^6A). Additionally, snRNA components of the minor spliceosome also contain ψ and 2'O-methylations, though much fewer than the major snRNAs[152]. Installation of ψ and 2'-O-methylations has only been shown to occur in an RNA-dependent manner in humans. The

modification enzymes are identical to those responsible for rRNA modification, and rely on common structural components found in the two RNA species. However, as modification typically occurs in Cajal bodies, these snRNAs are referred to as Small Cajal body-specific RNAs or scaRNAs (reviewed in Meier, 2016). Modification in Cajal bodies occurs after export and reimport of the snRNAs. In yeast, both RNA-independent and RNA-dependent mechanisms of modification exist to install ψ.

Similar to rRNA modifications, the modifications cluster to functional regions of the snRNA, especially in base-pairing regions and around the nucleotides responsible for branch-site recognition[151]. As U2 has the most RNA modifications, the functions of these modifications have been studied in depth. A few 2'O-methylations at the 5' end of U2 were shown to be individually required for spliceosome assembly while Ψ at the 5' end exhibit a cumulative positive effect on assembly[154]. Pseudouridines at the branch point pairing region in U2 have also been shown to affect the structure around the branch point adenosine in mRNA[155].

So far, pseudouridylation of human snRNA has not shown to be inducible, but two inducible Ψs have been identified in yeast U2 snRNA. These modifications are not present under normal growth conditions; however, nutrient stress or heat shock can induce installation of these modifications[156]. Ψ93 is installed by a snoRNA-dependent mechanism and is only present under nutritional stress. Ψ56 is installed in an RNA-independent mechanism by the Pus7 enzyme and is present under both nutritional and heat stress. The RNA sequence contexts of these inducible modifications are imperfect, and the imperfect context is strictly required (although not sufficient) for stress-induced pseudouridylation. However, the exact molecular mechanism for induced pseudouridylation has not been determined.

Partial modifications of snRNA have not been reported, although quantitative information on 2'-O-methylations and Ψ in snRNA has yet to be determined. Additionally, no demodification enzyme has been reported to act on snRNA. Despite this, the dynamics of snRNP assembly and disassembly does not rule out the possibility that some of these modifications can be removed in response to a change in cellular conditions.

Appendix 1.9 Concluding remarks

The diverse landscape of RNA modification has revealed itself as a critical entity for post-transcriptional gene regulation. Reversible mRNA methylation offers a tunable mechanism to achieve cellular complexity beyond what can be achieved by primary sequence or secondary structure alone. Most directly, mRNA methylation in the form of N^6-methyladenosine provides a selectivity mark that is decoded by evolutionarily conserved proteins of the YTH family as well as other RNA-binding proteins through different reading mechanisms. The advantage is obvious: the dynamic and sometimes reversible modifications enable grouping different sets of transcripts for rapid, coordinated metabolism in response to cellular and environmental cues, which could be challenging to achieve with other mechanisms.

Fundamental mechanisms that take advantage of m^6A methylation promote incorporation of methylated transcripts into canonical pathways for RNA metabolism. These pathways accelerate processing, translation initiation, and eventual decay of m^6A-modified mRNA during cell differentiation. The result of this selection is increased protein production within limited time frames – an outcome perfectly suited, and indeed required, for developmental and differentiation processes. This hallmark of m^6A-modified transcripts may explain why the mark is essential to mammals. Functions of the m^6A could be diverse in different cell types and different biological processes. Components of the m^6A regulatory network could be mutated or deregulated in certain types of cancer, and the mechanisms that underline these pathologies are current areas of investigation and potential areas of intervention.

Mechanisms for achieving selectivity in m^6A-dependent gene regulation remain a mystery. Components of the methyltransferase targeting system likely exist to limit methylation to a defined, reproducible subset of consensus sequences in response to various signals. Similarly, binding modes of effector proteins must select appropriate RNA substrates and protein binding partners to exert their required function within the cell, and demethylases may execute removal of methylations of target transcripts within specific time windows and cellular locations (**Figure A.5**). Modes of regulation in these areas are likely context-specific and are important areas of future exploration.

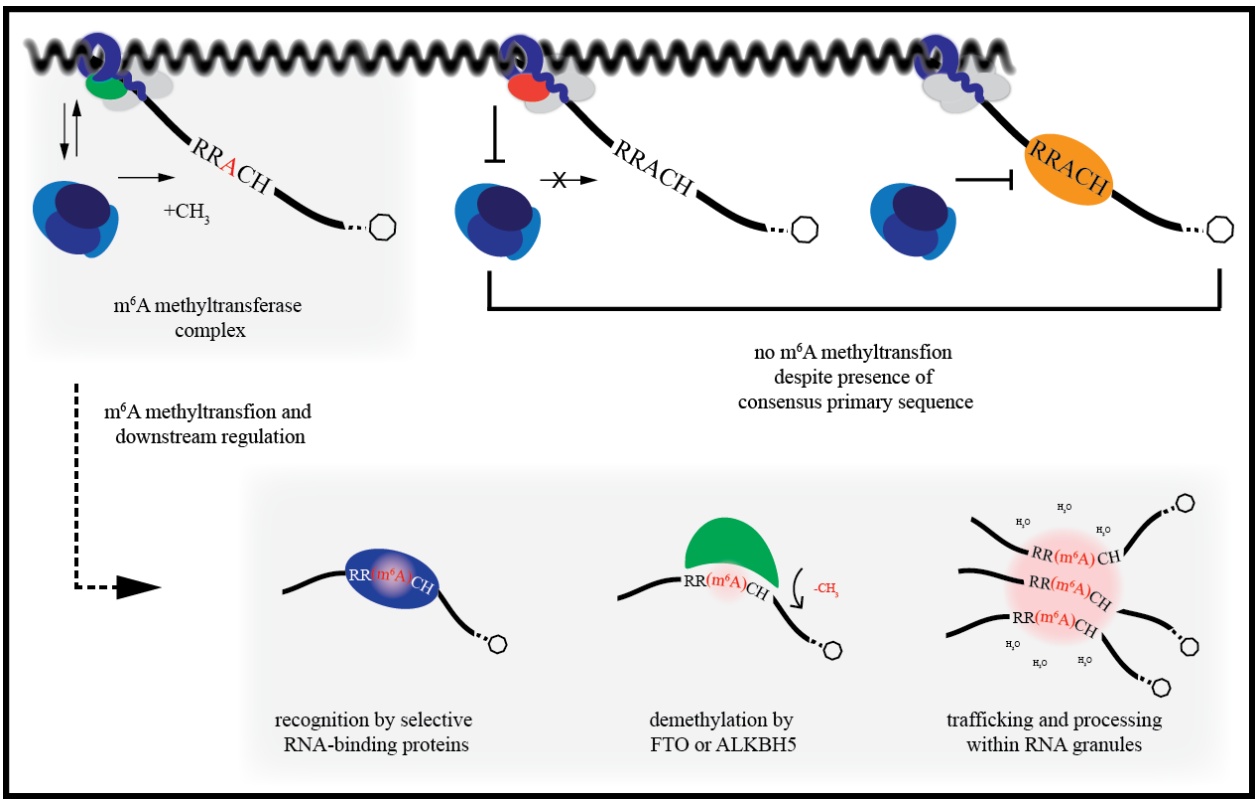

Figure 5. Mechanisms of selectivity in m^{6}A installation and downstream regulation

Mechanisms for selectivity in installation of m^{6}A are largely unknown, and cannot be predicted by primary sequence alone. Potential mechanisms include recruitment of methyltransferase components to nascent RNA by chromatin features associated with RNA Polymerase II (green), or exclusion from the transcription complex (red). RNA-binding proteins that occupy consensus sequences for m^{6}A may also prevent installation of the mark (orange). Once methylated, transcripts can be bound by RNA-binding that recognize modifications or secondary structure changes, or actively demethylated and no longer subject to regulation by m^{6}A-dependent pathways. Transcripts that are heavily decorated with methylations may face large solvation penalties, and benefit from trafficking within RNA granules.

Additional mRNA modifications further define the 'epitranscriptome', increasing potential modes for selectivity in post-transcriptional regulation. The cellular machinery that regulates these modifications, as well as mechanisms that take advantage of Ψ, m^{5}C, hm^{5}C, m^{1}A, 2'O-Me and many uncharacterized modifications are still under investigation.

tRNA modifications have been linked to numerous human diseases, including cancer, neurological disorders, and mitochondrial-linked disorders (reviewed in Torres et al., 2014). Since some tRNA modifications have been shown to be partial, reversible, and responsive to stresses,

these modifications and their dynamic properties will be highly relevant in biological regulation. The knockout of ALKBH1, the first tRNA demethylase discovered, leads to sex distortion and developmental defects in neurons in mouse models, suggesting currently unappreciated regulatory functions. Furthermore, the roles of modification in tRNA fragments, which have been shown to be widely used for regulating gene expression, has yet to be explored [124].

rRNA modification has been linked to dyskeratosis congenital, a human disease that affects pseudouridylation. Some rRNA modifications were identified as inducible and sites of partial modification exist, but no demodification enzymes have been identified. Similarly, snRNA modifications can be inducible but no demodification enzyme has been identified. The dynamics of snRNA, rRNA, and tRNA are likely to play important roles in the regulation of alternative splicing and translation in mammalian cells.

In summary, although RNA modifications are known for decades, recent advances have revealed its critical roles in the regulation of gene expression that impact many fundamental biological processes. Future studies will not only advance our understanding of this layer of biological regulation, but also translate its potentials for the understanding of human health and diseases.

Acknowledgements

I.A.R was supported by the NIH Medical Scientist National Research Service Award T32 GM007281 and is a recipient of the NIH F30 GM117646. M.E.E. was supported by the NIH Chemistry and Biology Training Grant T32 GM008720 and is a recipient of the NSF pre-doctoral fellowship DGE-1144082. We thank NIH HG008935 (C.H. and T.P.) for supporting our research. We thank NIH NHGRI RM1 HG008935 (C.H. and T.P.). C.H. is an investigator of the Howard Hughes Medical Institute.

References

---

1. Shatkin, A. J. Capping of Eucaryotic mRNAs Review. *Cell* **9,** 645–653 (1976).
2. Rottman, F., Shatkin, A. J. & Perry, R. P. Methylated Nucleotides at the 5 ' Termini of

Messenger RNAs: Possible Implications for Processing. *Cell* **3,** 197–199 (1974).

3. Furuichi, Y., LaFiandra, A. & Shatkin, A. J. 5'-Terminal structure and mRNA stability. *Nature* **266,** 235–239 (1977).
4. Shimotohno, K., Kodama, Y., Hashimoto, J. & Miura, K.-I. Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. *Proc. Natl. Acad. Sci.* **74,** 2734–2738 (1977).
5. Konarska, M. M., Padgett, R. A. & Sharp, P. A. Recognition of Cap Structure In Vitro of mRNA Precursors. **38,** 731–736 (1984).
6. Edery, I. & Sonenberg, N. Cap-dependent RNA splicing in a HeLa nuclear extract. *Proc. Natl. Acad. Sci.* **82,** 7590–7594 (1985).
7. Izaurralde, E. *et al.* A Nuclear Cap Binding Protein Complex Involved in Pre-mRNA Splicing. *Cell* **78,** 657–668 (1994).
8. Cooke, C. & Alwine, J. C. The Cap and the 3' Splice Site Similarly Affect Polyadenylation Efficiency. *Mol. Cell. Biol.* **16,** 2579–2584 (1996).
9. Hamm, J. & Mattaj, I. W. Monomethylated Cap Structures Facilitate RNA Export from the Nucleus. *Cell* **63,** 109–116 (1990).
10. Richter, J. D. & Sonenberg, N. Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* **433,** 477–480 (2005).
11. Mangus, D. A., Evans, M. C. & Jacobson, A. Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol.* **4,** 223 (2003).
12. Desrosiers, R., Friderici, K. & Rottman, F. Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells. *Proc. Natl. Acad. Sci.* **71,** 3971–3975 (1974).
13. Perry, R. P., Kelley, D. E., Friderici, K. & Rottman, F. The Methylated Constituents of L Cell Messenger RNA: Evidence for an Unusual Cluster at the 5' Terminus. *Cell* **4,** 387–394 (1975).
14. Dubin, D. T. & Taylor, R. H. The methylation state of polyA-containing-messenger RNA from cultured hamster cells. *Nucleic Acids Res.* **2,** 1653–1668 (1975).
15. Adams, J. M. & Coy, S. Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA. *Nature* **255,** 28–33 (1975).

16. Lavi, U., Fernandez-muhoz, R. & Darnell, J. E. Content of N-6 methyl adenylic acid in heterogenous nuclear and messenger RNA of HeLa cells. *Nucleic Acids Res.* **4,** (1977).
17. Wei, C. M., Gershowitz, A. & Moss, B. 5'-Terminal and Internal Methylated Nucleotide Sequences in HeLa Cell mRNA. *Biochemistry* **20,** 397–401 (1976).
18. Wei, C. & Moss, B. Nucleotide Sequences at the N6-Methyladenosine SIted of HeLa Cell Messenger Ribonucleic Acid. *Biochemistry* **16,** 1672–1676 (1977).
19. 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).
20. 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).
21. Lee, M., Kim, B. & Kim, V. N. Review Emerging Roles of RNA Modification: m6A and U-Tail. *Cell* **158,** 980–987 (2014).
22. Eckmann, C. R., Rammelt, C. & Wahle, E. Control of poly (A) tail length. *WIREs RNA* **2,** 348–361 (2011).
23. Ramanathan, A., Robb, G. B., Chan, S. & Biolabs, N. E. mRNA capping: biological functions and applications. *Nucleic Acids Res.* **44,** 7511–7526 (2016).
24. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* **18,** 275–291 (2017).
25. Jia, G. *et al.* N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7,** 6–9 (2011).
26. He, C. RNA epigenetics? *Nat. Chem. Biol.* **6,** 863–865 (2010).
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. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–206 (2012).
29. Meyer, K. D. *et al.* Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. *Cell* **149,** 1635–1646 (2012).
30. Chen, K. *et al.* High-Resolution N6-Methyladenosine (m6A) Map Using Photo-Crosslinking-Assisted m6A Sequencing. *Angew. Chem. Int. Ed* **54,** 1587–1590 (2015).
31. Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **12,** 767–772 (2015).

32. Liu, N. *et al.* Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19,** 1848–1856 (2013).
33. Harcourt, E. M., Ehrenschwender, T., Batista, P. J., Chang, H. Y. & Kool, E. T. Identification of a Selective Polymerase Enables Detection of N6-Methyladenosine in RNA. *J. Am. Chem. Soc.* **135,** 19079–19082 (2013).
34. Molinie, B. *et al.* m6A-LAIC-seq reveals the census and complexity of the m6A epitranscriptome. *Nat. Methods* **13,** 692–698 (2016).
35. Dominissini, D. *et al.* The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530,** 441–446 (2016).
36. Li, X. *et al.* Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. *Nat. Publ. Gr.* **12,** 311–316 (2016).
37. 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).
38. Hauenschild, R. *et al.* The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence. *Nucleic Acids* **43,** 9950–9964 (2015).
39. Schibler, U. & Perry, R. P. The 5'-termini of hererogeneous nuclear RNA: a comparison among molecules of different sizes and ages. *Nucleic Acids Res.* **4,** 4133–4149 (1977).
40. Fu, Y. Dynamic Regulation of RNA Modifications by ALKB Family Dioxygenases. (The University of Chicago, 2012).
41. Mauer, J. *et al.* Reversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nat. Publ. Gr.* **541,** 371–375 (2017).
42. Schaefer, M., Pollex, T., Hanna, K. & Lyko, F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.* **37,** e12 (2009).
43. Squires, J. E. *et al.* Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. **40,** 5023–5033 (2012).
44. Hussain, S. *et al.* NSun2-Mediated Cytosine-5 Methylation of Vault Noncoding RNA Determines Its Processing into Regulatory Small RNAs. *Cell Rep.* **4,** 255–261 (2013).
45. Khoddami, V. & Cairns, B. R. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* **31,** 458–464 (2013).
46. Yang, X. *et al.* 5-methylcytosine promotes mRNA export — NSUN2 as the

methyltransferase and ALYREF as an m 5 C reader. *Cell Res.* 1–20 (2017). doi:10.1038/cr.2017.55

47. Fu, L. *et al.* Tet-Mediated Formation of 5‐Hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **136,** 11582–11585 (2014).
48. Delatte, B. *et al.* Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **351,** 282–285 (2016).
49. Cohn, W. E. Pseudouridine, a Carbon-Carbon Linked Ribonucleoside in Ribonucleic Acids: Isolation, Structure, and Chemical Characteristics. *J. Biol. Chem.* **235,** 1488–1498 (1960).
50. Carlile, T. M. *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515,** 143–146 (2014).
51. Schwartz, S. *et al.* Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell* **159,** 1–15 (2014).
52. Karijolich, J. & Yu, Y.-T. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **474,** 395–398 (2011).
53. Fernandez, I. S. *et al.* Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature* **500,** 107–111 (2013).
54. Li, X. *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **11,** 592–597 (2015).
55. Marchand, V., Blanloeil-Oillo, F., Helm, M. & Motorin, Y. Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res.* **44,** e135 (2016).
56. Dai, Q. *et al.* Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat. Methods* (2017). doi:10.1038/NMETH.4294
57. Wang, Y. *et al.* N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* **16,** 191–201 (2014).
58. Liu, J. *et al.* A METTL3 METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10,** 93–95 (2014).
59. Sledz, P. & Jinek, M. Structural insights into the molecular mechanism of the m6A writer complex. *Elife* **5,** e18434 (2016).
60. Wang, P., Doxtader, K. A. & Nam, Y. Structural Basis for Cooperative Function of Mettl3 Article Structural Basis for Cooperative Function of Mettl3 and Mettl14

Methyltransferases. *Mol. Cell* **63,** 306–317 (2016).

61. Wang, X. *et al.* Structural basis of N6-adenosine methylation by the METTL3–METTL14 complex. *Nature* **534,** (2016).
62. Ping, X.-L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* **24,** 177–189 (2014).
63. Schwartz, S. *et al.* Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5' Sites. *Cell Rep.* **8,** 284–296 (2014).
64. 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).
65. Cui, Q. *et al.* m6A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. *Cell Rep.* **18,** 2622–2634 (2017).
66. 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.* **113,** 2047–2056 (2016).
67. Zhang, S. *et al.* The m6A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-Like Cells by Sustaining FOXM1 Expression and Cell Proliferation Programing. *Cancer Cell* **31,** 591–606 (2017).
68. Xiang, Y. *et al.* RNA m6A methylation regulates the ultraviolet-induced DNA damage response. *Nature* 1–22 (2017). doi:10.1038/nature21671
69. Takuma, H. *et al.* Substrate tRNA Recognition Mechanism of Eubacterial tRNA (m1A58) Methyltransferase (TrmI). *J. Biol. Chem.* **290,** 5912–5925 (2015).
70. Belanger, F., Stepinski, J., Darzynkiewicz, E. & Pelletier, J. Characterization of hMTr1, a Human Cap1 2'-O-Ribose Methyltransferase. *J. Biol. Chem.* **285,** 33037–33044 (2010).
71. Langberg, S. R. & Moss, B. Post-transcriptional Modifications of mRNA: Purification and Characterization of a Cap I and Cap II RNA (Nucleoside-2'-)-Methyltransferases from HeLa Cells. *J. Biol. Chem.* **256,** 10054–10060 (1981).
72. Ma, L. *et al.* Evolution of transcript modification by N6-methyladenosine in primates. *Genome Biol.* **27,** 1–8 (2017).
73. Roost, C. *et al.* Structure and Thermodynamics of N6‑Methyladenosine in RNA: A Spring-Loaded Base Modification. *J. Am. Chem. Soc.* **137,** 2107–2115 (2015).
74. Spitale, R. C. *et al.* Structural imprints in vivo decode RNA regulatory mechanisms. **519,**

486–490 (2015).

75. Liu, N. *et al.* N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518,** 560–564 (2015).
76. Zhou, K. I. *et al.* N6-Methyladenosine Modification in a Long Noncoding RNA Hairpin Predisposes Its Conformation to Protein Binding. *J. Mol. Biol.* **428,** 822–833 (2016).
77. Liu, N. *et al.* N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein,. *Nucleic Acids Res.* In press (2017). doi:doi: 10.1093/nar/gkx141
78. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* **61,** 507–519 (2016).
79. Patil, D. P. *et al.* m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537,** 369–373 (2016).
80. Alarcon, C. R. *et al.* HNRNPA2B1 Is a Mediator of m6A-Dependent Nuclear RNA Processing Events. *Cell* **162,** 1299–1308 (2015).
81. Wang, X. *et al.* N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161,** 1388–1399 (2015).
82. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–120 (2014).
83. Du, H. *et al.* YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. *Nat. Commun.* **7,** 1–11 (2016).
84. Zhou, J. *et al.* Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature* **526,** 591–594 (2015).
85. Meyer, K. D. *et al.* 5' UTR m6A Promotes Cap-Independent Translation. *Cell* **163,** 999–1010 (2015).
86. Li, A. *et al.* Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. *Cell Res.* **1,** 1–4 (2017).
87. Shi, H. *et al.* YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Res.* **1,** 1–14 (2017).
88. Lin, S. *et al.* The m6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol. Cell* **62,** 335–345 (2016).
89. Charette, M. & Gray, M. W. Pseudouridine in RNA: What, Where, How, and Why. *IUBMB Life* **49,** 341–351 (2000).

90. Yildirim, I., Kierzek, E., Kierzek, R. & Schatz, G. C. Interplay of LNA and 2′‐O‐Methyl RNA in the Structure and Thermodynamics of RNA Hybrid Systems: A Molecular Dynamics Study Using the Revised AMBER Force Field and Comparison with Experimental Results. *J. Phys. Chem. B* 14177–14187 (2014).
91. Hoernes, T. P. *et al.* Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res.* **44,** 852–862 (2016).
92. Slobodin, B. *et al.* Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell* **169,** 326–337 (2017).
93. Xu, C. *et al.* Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* **10,** 927–929 (2014).
94. Xu, C. *et al.* Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. *J. Biol. Chem.* **290,** 24902–24913 (2015).
95. 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).
96. Zhu, T. *et al.* Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res.* **2,** 1493–1496 (2014).
97. Noeske, J. *et al.* High-resolution structure of the Escherichia coli ribosome. *Nat. Struct. Mol. Biol.* **22,** 336–341 (2015).
98. Agris, P. F., Vendeix, F. A. P. & Graham, W. D. tRNA's Wobble Decoding of the Genome: 40 Years of Modification. *J. Mol. Biol.* **366,** 1–13 (2007).
99. Agris, P. F. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep.* **9,** 629–635 (2008).
100. Agris, P. F. The importance of being modified: an unrealized code to RNA structure and function. *RNA* **21,** 552–554 (2015).
101. Lence, T. *et al.* m6A modulates neuronal functions and sex determination in Drosophila. *Nature* **540,** 242–247 (2016).
102. Haussmann, I. U. *et al.* m6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* **540,** 301–304 (2016).

103. Batista, P. J. *et al.* m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells. *Cell Stem Cell* **15,** 707–719 (2014).
104. Geula, S. *et al.* m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* **3,** 1002–1006 (2015).
105. Li, L., Lu, X. & Dean, J. The maternal to zygotic transition in mammals. *Mol. Aspects Med.* **34,** 919–938 (2013).
106. Zhao, B. S. *et al.* m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **542,** 475–478 (2017).
107. Zhang, M. *et al.* The Demethylase Activity of FTO (Fat Mass and Obesity Associated Protein) Is Required for Preadipocyte Differentiation. *PLoS One* **10,** e0133788 (2015).
108. Zhao, X. *et al.* FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.* **24,** 1403–1419 (2014).
109. Frayling, T. M. *et al.* A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science* **316,** 889–895 (2007).
110. Wu, W. *et al.* AMPK regulates lipid accumulation in skeletal muscle cells through FTO-dependent demethylation of N6-methyladenosine. *Sci. Rep.* **7,** 41606 (2017).
111. Zhang, C. *et al.* Hypoxia-inducible factors regulate pluripotency factor expression by ZNF217- and ALKBH5-mediated modulation of RNA methylation in breast cancer cells. *Oncotarget* **7,** 64527–64542 (2016).
112. Lichinchi, G. *et al.* Dynamics of the human and viral m6A RNA methylomes during HIV-1 infection of T cells. *Nat. Microbiol.* **1,** 16011 (2016).
113. Kennedy, E. M. *et al.* Posttranscriptional m6A Editing of HIV-1 mRNAs Enhances Viral Gene Expression. *Cell Host Microbe* **19,** 675–685 (2016).
114. Tirumuru, N. *et al.* N6-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression. *Elife* **5,** e15528 (2016).
115. Lichinchi, G. *et al.* Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. *Cell Host Microbe* **20,** 666–673 (2016).
116. Gokhale, N. S. *et al.* N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection. *Cell Host Microbe* **20,** 654–665 (2016).
117. Anadon, C. *et al.* Epigenetic loss of the RNA decapping enzyme NUDT16 mediates C-MYC activation in T-cell acute lymphoblastic leukemia. *Leukemia* 1–4 (2017).

doi:10.1038/leu.2017.99

118. Brat, D. *et al.* Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* **372,** 2481–2498 (2015).
119. Eckel-Passow, J. *et al.* Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *N. Engl. J. Med.* **372,** 2499–2508 (2015).
120. Yan, H. *et al.* IDH1 and IDH2 Mutations in Gliomas. *N. Engl. J. Med.* **360,** 765–773 (2009).
121. Chou, W. *et al.* The prognostic impact and stability of Isocitrate dehydrogenase 2 mutation in adult patients with acute myeloid leukemia. *Leukemia* 246–253 (2011). doi:10.1038/leu.2010.267
122. Patel, J. *et al.* Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. *N. Engl. J. Med.* **366,** 1079–1089 (2012).
123. Esteller, M. & Pandolfi, P. P. The Epitranscriptome of Noncoding RNAs in Cancer. *Cancer Discov.* 359–369 (2017). doi:10.1158/2159-8290.CD-16-1292
124. Kirchner, S. & Ignatova, Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.* **16,** 98–112 (2015).
125. Rubio, M. A. T. *et al.* Editing and methylation at a single site by functionally interdependent activities. *Nature* **542,** 494–497 (2017).
126. Stuart, J. W., Koshlap, K. M., Guenther, R. & Agris, P. F. Naturally-occurring Modification Restricts the Anticodon Domain Conformational Space of tRNA Phe. *J. Mol. Biol.* 901–918 (2003). doi:10.1016/j.jmb.2003.09.058
127. Waas, W. F., Druzina, Z., Hanan, M. & Schimmel, P. Role of a tRNA Base Modification and Its Precursors in Frameshifting in Eukaryotes. *J. Biol. Chem.* **282,** 26026–26034 (2007).
128. Helm, M., Giege, R. & Florentz, C. A Watson-Crick Base-Pair-Disrupting Methyl Group (m1A9) Is Sufficient for Cloverleaf Folding of Human Mitochondrial tRNA(Lys). *Biochemistry* **38,** 13338–13346 (1999).
129. Blanco, S. *et al.* Stem cell function and stress response are controlled by protein synthesis. *Nature* **534,** 335–340 (2016).
130. Blanco, S. *et al.* Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. *EMBO J.* **33,** 2020–2039 (2014).

131. Kadaba, S., Wang, X. & Anderson, J. T. Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA* **12,** 508–521 (2006).
132. Alexandrov, A. *et al.* Rapid tRNA Decay Can Result from Lack of Nonessential Modifications. *Mol. Cell* **21,** 87–96 (2006).
133. Rezgui, V. A. N. *et al.* tRNA tKUUU, tQUUG, and tEUUC wobble position modifications fine-tune protein translation by promoting ribosome A-site binding. *Proc. Natl. Acad. Sci.* **110,** 12289–12294 (2013).
134. Kaneko, T. *et al.* Wobble modification differences and subcellular localization of tRNAs in Leishmania tarentolae: implication for tRNA sorting mechanism. *EMBO* **22,** 657–667 (2003).
135. Clark, W. C., Evans, M. E., Dominissini, D. A. N., Zheng, G. & Pan, T. tRNA base methylation identification and quantification via high-throughput sequencing. *RNA* **22,** 1771–1784 (2016).
136. Chan, C. T. Y. *et al.* A Quantitative Systems Approach Reveals Dynamic Control of tRNA Modifications during Cellular Stress. *PLoS Genet.* **6,** e1001247 (2010).
137. Chan, C. T. Y. *et al.* Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat. Commun.* **3:937,** 1938 (2012).
138. Liu, F. *et al.* ALKBH1-Mediated tRNA Demethylation Regulates Translation. *Cell* **167,** 816–828 (2016).
139. Ueda, Y. *et al.* AlkB homolog 3-mediated tRNA demethylation promotes protein synthesis in cancer cells. *Sci. Rep.* 1–42271 (2017). doi:10.1038/srep42271
140. Ougland, R. *et al.* AlkB Restores the Biological Function of mRNA and tRNA Inactivated by Chemical Methylation. *Mol. Cell* **16,** 107–116 (2004).
141. Sharma, S. & Lafontaine, D. L. J. 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. *Trends Biochem. Sci.* **40,** 560–575 (2015).
142. Sloan, K. E. *et al.* Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* **0,** 1–16 (2016).
143. Watkins, N. J. & Bohnsack, M. T. The box C/D and H/ACA snoRNPs: key players in the modification , processing and the dynamic folding of ribosomal RNA. *WIREs RNA* **3,** 397–414 (2012).

144. Kos, M. & Tollervey, D. Article Yeast Pre-rRNA Processing and Modification Occur Cotranscriptionally. *Mol. Cell* **37,** 809–820 (2010).
145. Turowski, T. W. & Tollervey, D. Cotranscriptional events in eukaryotic ribosome synthesis. *WIREs RNA* **6,** 129–139 (2015).
146. Decatur, W. A. & Fournier, M. J. rRNA modifications and ribosome function. *Trends Biochem. Sci.* **27,** 344–351 (2002).
147. Taoka, M. *et al.* The complete chemical structure of Saccharomyces cerevisiae rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Res.* **44,** 8951–8961 (2016).
148. Krogh, N. *et al.* Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.* **44,** 7884–7895 (2016).
149. Buchhaupt, M. *et al.* Partial Methylation at Am100 in 18S rRNA of Baker's Yeast Reveals Ribosome Heterogeneity on the Level of Eukaryotic rRNA Modification. *PLoS One* **9,** e89640 (2014).
150. Massenet, S. & Branlant, C. A limited number of pseudouridine residues in the human atac spliceosomal UsnRNAs as compared to human major spliceosomal UsnRNAs. *RNA* **5,** 1495–1503 (1999).
151. Reddy, R. & Busch, H. in *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles* (ed. Birnstiel, M. L.) 1–37 (Springer Berlin Heidelberg, 1988). doi:10.1007/978-3-642-73020-7_1
152. Karijolich, J. & Yu, Y.-T. Spliceosomal snRNA modifications and their function. *RNA Biol.* **7,** 192–204 (2010).
153. Meier, U. T. RNA modification in Cajal bodies. *RNA Biol.* **0,** 1–8 (2016).
154. Dönmez, G., Hartmuth, K. & Lührmann, R. Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA* **10,** 1925–1933 (2004).
155. Lin, Y. & Kielkopf, C. L. X-ray Structures of U2 snRNA - Branchpoint Duplexes Containing Conserved Pseudouridines. *Biochemistry* **47,** 5503–5514 (2008).
156. Wu, G., Xiao, M., Yang, C. & Yu, Y. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO* **30,** 79–89 (2010).

157. Torres, A. G., Batlle, E. & Pouplana, L. R. De. Role of tRNA modifications in human diseases. *Trends Mol. Med.* **20,** 306–314 (2014).