THE UNIVERSITY OF CHICAGO

N^6-METHYLADENOSINE-DEPENDENT RNA STRUCTURAL SWITCHES MODULATE RNA-PROTEIN INTERACTIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

NIAN LIU

CHICAGO, ILLINOIS

DECEMBER 2015

TABLE OF CONTENTS

Page

LIST OF FIGURES. viii

LIST OF TABLES. x

ACKNOWLEDGEMENTS. xi

ABSTRACT. xii

LIST OF PUBLICATIONS. xiii

CHAPTER

1. Introduction 1

1.1 Epigenetics 1

1.2 Post-transcriptional RNA modification 1

1.3 N^6-methyladenosine (m^6A) RNA modification 3

1.4 Scope of this thesis 4

1.5 References 5

2. Probing the m^6A Sites and Methylation Fraction at Nucleotide-Resolution 8

2.1 Introduction 8

2.2 Results and discussion 9

2.2.1 A method to detect m^6A status in mRNA/lncRNA 9

2.2.2 m^{6}A status in lncRNA and mRNA 15

2.2.3 An m^{6}A-containing RNA structural motif 17

2.2.4 Concluding remarks 19

2.3 Experimental section 20

2.3.1 Cell culture 20

2.3.2 SCARTLET oligonucleotide sequences 20

2.3.3 Synthesis of m^{6}A-RNA oligo 22

2.3.4 SCARLET 22

2.3.5 RNA structural mapping 23

2.3.6 m^{6}A/MeRIP-seq and its data analysis 24

2.4 References 25

3. Widespread *N*^{6}-methyladenosine-dependent RNA Structural Switches Regulate RNA-Protein Interactions 27

3.1 Introduction 27

3.2 Results and discussion 27

3.2.1 m^{6}A facilitates RNA-HNRNPC interactions *in vitro* and *in vivo* 28

3.2.2 m^{6}A alters RNA structure to increase the accessibility of U-tracts 30

3.2.3 Increased accessibility of U-tract enhances HNRNPC binding 31

3.2.4 PAR-CLIP–MeRIP identifies m^6A-modified HNRNPC binding sites across the transcriptome 35

3.2.5 Transcriptome-wide identification of m^6A-switches for HNRNPC binding 38

3.2.6 *METTL3/L14* knockdown decreases HNRNPC binding activities 40

3.2.7 Molecular features of high-confidence m^6A-switches 42

3.2.8 HNRNPC m^6A-switches affect cell proliferation by regulating gene expression 45

3.2.9 HNRNPC m^6A-switches regulate alternative splicing 45

3.2.10 Conclusion and discussion 47

3.3 Experimental section 52

3.3.1 Mammalian Cell Culture, siRNA KD and Western Blot 52

3.3.2 Native Gel Shift Assay 52

3.3.3 In vitro Pull Down and Mass Spectrometry 52

3.3.4 Expression of Recombinant HNRNPC1 53

3.3.5 In vitro Gel Shift assay and UV Crosslinking 53

3.3.6 Filter-binding Assay 53

3.3.7 CLIP-2dTLC 54

3.3.8 V1/S1 RNA Structural Probing 55

3.3.9 CMCT RNA Structural Probing 55

3.3.10 RNA Terminal Truncation 55

3.3.11 PAR-CLIP 55

3.3.12 PAR-CLIP–MeRIP 56

3.3.13 RNA Deep Sequencing and Genomic Mapping 57

3.3.14 Normalization of the Mapping Reads for Subsequent Analysis 57

3.3.15 Identification of HNRNPC binding sites 58

3.3.16 Detection of PAR-CLIP–MeRIP peaks 58

3.3.17 Detection of decreased HNRNPC binding peaks upon METTL KD 58

3.3.18 Identification of enriched motifs 59

3.3.19 Identification of HNRNPC m^{6}A-switches 59

3.3.20 Distribution of HNRNPC m^{6}A-switches 59

3.3.21 RNA-seq Analysis 60

3.3.22 Gene Ontology Analysis 60

3.3.23 Cell Proliferation Analysis 61

3.3.24 Evolutionary Conservation Analysis 61

3.3.25 RT-PCR Quantitation 61

3.3.26 Graphic and Statistical Analysis 62

3.4 References 64

4. *N*^{6}-methyladenosine-dependent RNA Structural Switches Recruit HNRNPG for Alternative Splicing Regulation 67

4.1 Introduction 67

4.2 Results and discussion.....69

4.2.1 HNRNPG selectively binds m^6A methylated RNA 69

4.2.2 m^6A alters RNA structure to expose HNRNPG binding consensus motif 73

4.2.3 Transcriptome-wide identification of m^6A-switches regulating HNRNPG binding.....75

4.2.4 m^6A-selective HNRNPG binding events regulate alternative splicing 78

4.2.5 Conclusion and discussion 82

4.3 Experimental section 84

4.3.1 Mammalian cell culture, siRNA knockdown and Western blot 85

4.3.2 Gel shift and RNA pull down assays 85

4.3.3 Protein expression and *in vitro* UV cross-linking assays 86

4.3.4 RNA structural probing 87

4.3.5 PARCLIP and PARCLIP-MeRIP 87

4.3.6 Detection and distribution analysis of m^6A sites within HNRNPG binding sites 88

4.3.7 RNA-sequencing 89

4.3.8 Evolutionary conservation, graphic and statistical analysis 90

4.3.9 RT-PCR quantitation 90

4.4 References 92

5. Summary and perspective 96

5.1 Introduction 96

5.2 Detection of m^6A status 96

5.3 The selective m^6A methylation by m^6A 'writers', 'erasers' and other machineries 97

5.4 Mechanisms for m^6A function 101

5.4.1 Direct m^6A readers: YTH-domain proteins 102

5.4.2 m^6A affects RNA structure 103

5.4.3 Indirect m^6A readers: m^6A-switches 104

5.4.4 The association between low-complexity domains and m^6A readers 105

5.4.5 m^6A-dependent microRNA maturation 106

5.5 Conclusions 107

5.6 References 107

LIST OF FIGURES

Figure 1.1. Chemical structure of internal mRNA/lncRNA modifications. 2

Figure 1.2. RNA epigenetic marking and control requires three groups of proteins. 3

Figure 2.1. Schematic diagram of SCARLET. 10

Figure 2.2. SCARLET detects m^6A status. 11

Figure 2.3. Determination of m^6A status of rRNA and an mRNA from a transfected plasmid. ..13

Figure 2.4. SCARLET determination of m^6A status in the MALAT1 lncRNA 14

Figure 2.5. m^6A/MeRIP-seq analysis of four RNAs tested for m^6A status in this work. 18

Figure 2.6. Prediction of secondary structure motifs around two MALAT1 m^6A residues displaying highest modification fraction. 19

Figure 3.1. HNRNPC selectively binds m^6A methylated RNA. 28

Figure 3.2. m^6A destabilizes the hairpin structure to expose the U-tract accessibility. 30

Figure 3.3. Destabilized hairpin oligos increase HNRNPC binding. 32

Figure 3.4. The increased accessibility of U-tracts increased HNRNPC binding. 33

Figure 3.5. m^6A is enriched in the vicinity of HNRNPC binding sites. 34

Figure 3.6. PAR-CLIP–MeRIP identifies m^6A-switches transcriptome-wide. 36

Figure 3.7. Validation of two identified m^6A-switches. 39

Figure 3.8. Global m^6A reduction decreases HNRNPC binding at m^6A-switches. 41

Figure 3.9. Molecular features of high-confidence m^6A-switches. 43

Figure 3.10. m^6A-switches regulate the cell proliferation rates. 44

Figure 3.11. m^{6}A-switches regulate the abundance of target mRNAs. 46

Figure 3.12. m^{6}A-switches regulate mRNA alternative splicing. 48

Figure 3.13. Summary of the sequencing samples.....51

Figure 4.1. HNRNPG selectively binds m^{6}A modified RNAs. 70

Figure 4.2. The C-terminal RNA binding domain of HNRNPG is m^{6}A responsive. 71

Figure 4.3. The RNA-binding properties of HNRNPG revealed by PAR-CLIP. 72

Figure 4.4. m^{6}A destabilizes the hairpin-stem structure to recruit HNRNPG. 73

Figure 4.5. Transcriptome-wide identification of m^{6}A-switches recruiting HNRNPG.....76

Figure 4.6. Properties of high-confidence HNRNPG m^{6}A-switches. 76

Figure 4.7. HNRNPG and m^{6}A-writers co-regulate the expression of m^{6}A-switch-containing genes. 77

Figure 4.8. HNRNPG loss and m^{6}A writers depletion cause similar alternative splicing changes on m^{6}A-switch-containing transcripts.....78

Figure 4.9. m^{6}A-switches recruit HNRNPG for alternative splicing regulation. 79

Figure 4.10. The m^{6}A-selective HNRNPG binding regulates alternative splicing. 80

Figure 4.11. Proposed model: m^{6}A-switches recruit HNRNPG via its C-terminal low-complexity domain to perform alternative splicing steps. 81

Figure 4.12. HNRNPG exist in the form of granules *in vitro* and *in vivo*. 83

Figure 5.1. m^{6}A-coded RNA epigenetics: writers, erasers and readers. 98

Figure 5.2. m^{6}A-regulated RNA metabolism in mammalian cells. 101

LIST OF TABLES

Table 2.1. m^6A status in the seven RRACH consensus sites under the largest m^6A/MeRIP-seq peak of the MALAT1 lncRNA. 16

Table 2.2. m^6A status in the lncRNA TUG1 and 3 mRNAs (ACTB, TPT1 and BSG). 17

Table 3.1. Information on oligonucleotides. 62

Table 5.1. The phenotypes of m^6A perturbation in cells and organisms. 97

ACKNOWLEDGEMENTS

I am deeply grateful to my advisors, Professor Tao Pan and Professor Chuan He, for giving the opportunity to be their joined student and work on the fantastic projects. Their scientific insight and guidance helps me enter the Molecular Biology discipline from a totally different background. Thank them for their patience and inspiration.

I thank Professor Tao Pan for giving me detailed guidance throughout my PhD projects. His broad vision and deep interest in science, organized working habits as well as optimistic attitude towards work and life deeply influenced me and will surely continue to influence me in the future. Besides, I am grateful that he gives me opportunities to present my work at many international conferences, which I think promotes my learning process a lot.

I am grateful to Professor Joseph Piccirilli for serving on my thesis committee, and for his support and sharing his lab equipments, which are quite helpful for the completion of the thesis.

I want to thank my two long-term co-workers. Dr. Marc Parisien has done lots of bioinformatic work for my projects. Dr. Qing Dai provides all the RNA oligo synthesis. Both their collaborations are essential for the completion of my thesis.

I thank Katherine Zhou, Dr. Guanqun Zheng, Dr. Chengqi Yi, Dr. George Perdrizet and Dr. Paula Lopez for their collaborations. And I want to express my gratitude to all the current and past members in both Pan and He group. I have learned a lot from them, and I have enjoyed all the discussion with them.

Finally, I would like to thank my family and friends for all their love and encouragement. It is impossible to make the completion of this thesis without their support.

ABSTRACT

Various chemical modifications have been found in cellular RNAs, but their functions remain an uncharted territory. N^6-methyladenosine (m^6A), the most abundant and dynamic internal modification in eukaryotic messenger RNA (mRNAs), is indispensable for cell viability, pluripotency and human health, but how m^6A achieves such wide-ranging biological functions remains unclear. The m^6A functional studies have been hindered by the lack of methods for its precise detection. To address this problem, I developed one method to detect m^6A modification status at nucleotide-resolution, and identified exact m^6A sites in human mRNAs and long non-coding RNAs (lncRNAs). Afterwards, I discovered that m^6A alters the local RNA structure to control the RNA-structure-dependent accessibility of RNA binding sites, thus affecting RNA-protein interactions; I termed this mechanism 'm^6A-switch'. Two members of heterogeneous nuclear ribonucleoproteins (HNRNPs), HNRNPC and HNRNPG, are found to be regulated by m^6A-switches. These m^6A-switch-regulated HNRNPC/G binding activities affect the RNA abundance and alternative splicing events. These findings illustrate how RNA-binding proteins gain regulated access to their RNA binding sites through m^6A-dependent RNA structural remodeling, and provides a new direction for investigating RNA-modification-coded cellular biology.

LIST OF PUBLICATIONS

Chapter 1

Liu, N., Pan, T. RNA epigenetics. ***Transl. Res.*** 165, 28–35 (2015).

Chapter 2

Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. Probing N^6-methyl-adenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. ***RNA*** 19, 1848-1856 (2013).

Liu, N., Pan, T. Probing RNA modification status at single nucleotide resolution in messenger and long non-coding RNA. ***Methods in Enzymology*** (Available online 2 June 2015).

Liu, N., Pan, T. Probing m^6A RNA methylation with SCARLET. In ***Methods in Molecular Biology: Post-transcriptional Gene Regulation*** 1358, 285-292 (2015).

Chapter 3

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., Pan, T. N^6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. ***Nature*** 518, 560-564 (2015).

Zhou, K., Parisien, M., Dai, Q., **Liu, N.**, Diatchenko, L., Sachleben, J., Pan, T. N^6-methyl-adenosine modification in a long non-coding RNA hairpin predisposes its conformation to protein binding. ***J. Mol. Biol.*** (Available online 3 Sep 2015).

Chapter 4

Liu, N., Parisien, M., Dai, Q., Zhou, K., and Pan, T. N^6-methyladenosine-dependent RNA structural switches recruits HNRNPG for alternative splicing regulation. ***Mol. Cell*** (*under review*).

Chapter 5

Liu, N., Pan, T. N^6-methyladenosine-coded RNA Epigenetics. ***Nat. Struct. Mol. Biol.*** (*in revision*).

Liu, N., Pan, T. RNA epigenetics. ***Translating Epigenetics to the Clinic.*** (*invited and in preparation*).

Chapter 1 - Introduction

N^6-methyladenosine (m^6A) is the most abundant internal modification in eukaryotic messenger RNA. m^6A can be reversible, has dynamic patterns in the cell, and can play significant epigenetic roles analogous to DNA methylation and histone modifications. Recent discoveries on the location, mechanism and functions of m^6A have shed light onto a new layer of gene regulation at the RNA level, thus opening the field of m^6A-coded RNA epigenetics or epitranscriptomics.

1.1 Epigenetics

Epigenetics refers to the inheritable changes of gene expression and cellular identity caused by mechanisms other than alterations that occur to the underlying genetic sequence[1]. DNA modifications and histone modifications represent two well-known categories of chemical modifications that epigenetically control cell lineage and cell fate[2]. For instance, 5-methylcytosine (5mC) is able to control pluripotency status and cell fate[3]. Active demethylation of 5mC occurs after fertilization to retrieve the pluripotent state of embryonic stem cells through as yet undefined mechanisms[4,5]. Besides 5mC, 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) have also been discovered to open a new paradigm of active DNA demethylation through 5mC oxidation[6,7,8,9]. These modifications are highly regulated and can potentially modulate the epigenetic landscape of a cell by affecting gene expression[10,11]. Besides DNA modifications and histone modifications, RNA is also subject to various chemical modifications[12]. Whether and how RNA modifications play similar epigenetic roles as DNA modifications and histone modifications is largely unknown.

1.2 Post-transcriptional RNA modifications

Over 100 types of post-transcriptional modifications have been identified in cellular RNA starting in the 1950s (http://mods.rna.albany.edu/). For example, the human ribosomal RNA (rRNA) contains over 200 modifications consisting of three major types: ~100 2'-*O*-methylated nucleotides (Nm), ~100 pseudouridines (Ψ), and ~10 base methylations (e.g. 5-methyl-cytosine, m^5C). Each human transfer RNA (tRNA) contains on average 14 modifications consisting of various base methylations, Ψ, Nm and chemically elaborate, modified wobble bases that require catalysis by multiple enzymes[13,14]. rRNA modifications are generally used as quality control checkpoints in ribosome assembly[15]. tRNA modifications outside the anticodon loop are generally used to maintain tRNA stability or modulate tRNA folding, whereas modifications in the anticodon loop are generally used to tune decoding capacity and to control decoding accuracy[16].

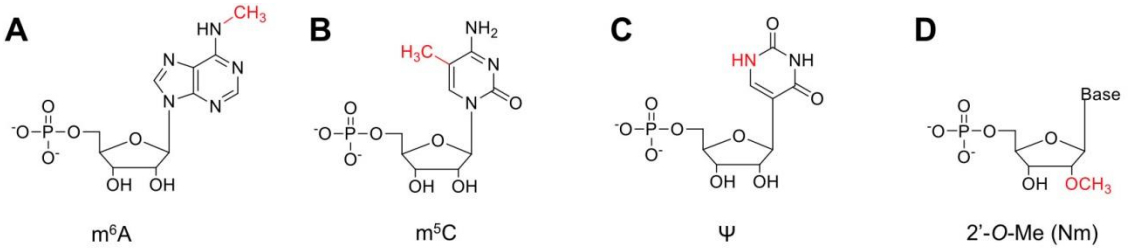


Figure 1.1. Chemical structure of internal mRNA/lncRNA modifications. A) m^6A; **B**) m^5C; **C**) Ψ; **D**) Nm.

However, despite their biological importance, mRNA and long non-coding RNA (lncRNA) modifications have not been intensively studied until recently. As of today, several mRNA modifications have been identified, they are N^6-methyladenosine (m^6A), 5-methylcytosine (m^5C), pseudouridine (Ψ), and 2'-*O*-methylated nucleosides (Nm)[12,17-21] (**Figure 1.1**). The molecular functions of these modifications, yet to be determined, would serve as a new layer of gene regulation at the RNA level.

1.3 N^6-methyladenosine (m^6A) RNA modification

Discovered in the 1970s[22-25], the most abundant internal mRNA/lncRNA modification is made of N^6-methyl adenosine (m^6A), present on average in over 3 sites per mRNA molecule[26-29]. The m^6A presence cannot be detected by the commonly used reverse transcriptases in cDNA synthesis. It was therefore extremely difficult to map m^6A at single nucleotide resolution. Global m^6A modification was shown to be functionally important as siRNA knockdown of a known human m^6A-methyltransferase (METTL3) led to apoptosis in cell culture[30]. Suggested functions for m^6A modification include effects on mRNA splicing, transport, stability, and immune tolerance[30,31].

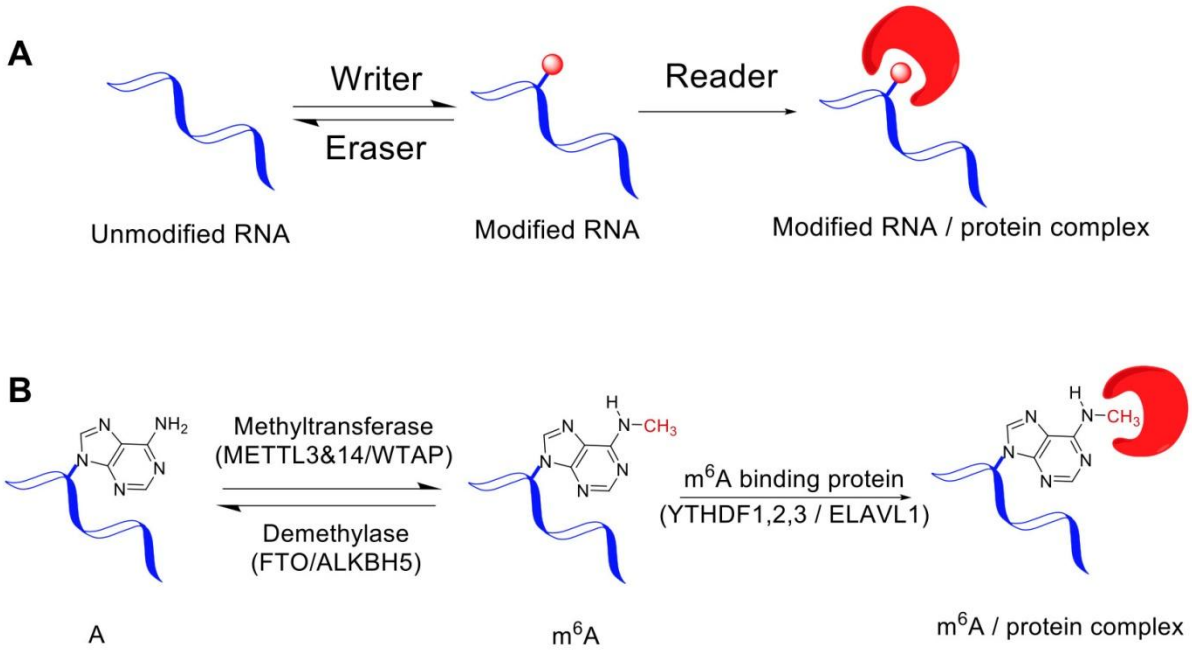


Figure 1.2. RNA epigenetic marking and control requires three groups of proteins. A) Writer = modification enzymes; eraser = demodification enzymes; reader = RNA binding proteins that recognize specific sites in the modified mRNA/lncRNA. **B**) m^6A writers, erasers and readers.

Interest in mRNA/lncRNA modification was revived in 2011 upon the discovery that m^6A modification is the cellular substrate for the human enzyme FTO[32]. FTO belongs to a family of human genes that are homologous to the *E. coli* AlkB protein which catalyzes oxidative reversal of methylated DNA and RNA bases[33,34]. In genome-wide association studies, the human FTO gene is highly associated with diabetes and obesity in the human population[35,36]. FTO knockout mice are much leaner than the wild-type mice, presumably due to perturbations in controlling cellular metabolism[37]. The discovery of FTO acting on m^6A in mRNA/lncRNA indicates that m^6A modification is subject to sophisticated cellular control.

The discovery of this first RNA demodification enzyme also highlights the idea that RNA modification may act as epigenetic markers and controls akin to DNA methylation and histone modification[38,39]. Three groups of proteins are needed for epigenetic control that maintains specific modification patterns in cell type and cell state dependent manners. 'Writers' catalyze chemical modification at specific sites; 'erasers' remove modification at specific sites; and 'readers' recognize the modified sites in DNA or histones (**Figure 1.2A**). For m^6A in mRNA/lncRNA, members in all three groups of proteins have now been found in mammalian cells (**Figure 1.2B**). However, the current list of these proteins likely represents just the beginning. In particular, the number of reader proteins that recognize m^6A modified mRNA/lncRNA sites will certainly expand greatly in the coming years.

1.4 Scope of the thesis

Despite its physiological importance, the mechanisms of m^6A functions remain unclear. And the m^6A function studies have been hindered by lack of methods for its precise detection.

Chapter 2 presents one novel method for nucleotide-resolution detection of m^6A status in mammalian mRNAs and lncRNAs.

Chapter 3 presents the discovery of m^6A-dependent RNA structural switches (m^6A-switch), which regulate RNA-protein interactions in the cell.

Chapter 4 presents that m^6A-switches recruit HNRNPG for alternative splicing regulation.

Chapter 5 summarizes current progress regarding m^6A-dependent RNA epigenetics, including all these findings presented in Chapter 2-4, and discusses future directions in the field.

1.5 References:

1. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396-8 (2007).
2. Fuks, F. DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev* **15**, 490-5 (2005).
3. Sasaki, H. & Matsui, Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* **9**, 129-40 (2008).
4. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293**, 1089-93 (2001).
5. Ooi, S.K. & Bestor, T.H. The colorful history of active DNA demethylation. *Cell* **133**, 1145-8 (2008).
6. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-30 (2009).
7. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-5 (2009).
8. He, Y.F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-7 (2011).
9. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-3 (2011).
10. Ficz, G. et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398-402 (2011).
11. Pastor, W.A. et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394-7 (2011).
12. Bokar, J.A. The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. *Fine-tuning of RNA functions by modification and editing* 141-177 (Springer, 2005).
13. El Yacoubi, B., Bailly, M. & de Crecy-Lagard, V. Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. *Annu Rev Genet* **46**, 69-95 (2012).
14. Grosjean, H., de Crecy-Lagard, V. & Marck, C. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett* **584**, 252-64 (2009).
15. Song, X. & Nazar, R.N. Modification of rRNA as a 'quality control mechanism' in ribosome biogenesis. *FEBS Lett* **523**, 182-6 (2002).

16. Agris, P.F. Decoding the genome: a modified view. *Nucleic Acids Res* **32**, 223-38 (2004).
17. Carlile, T.M. et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **advance online publication**(2014).
18. Dominissini, D. et al. Topology of the human and mouse m^6A RNA methylomes revealed by m^6A-seq. *Nature* **485**, 201-6 (2012).
19. Meyer, K. et al. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. *Cell* **149**, 1635-1646 (2012).
20. Schwartz, S. et al. Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-162 (2014).
21. Squires, J.E. et al. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Research* **40**, 5023-5033 (2012).
22. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc Natl Acad Sci U S A* **71**, 3971-5 (1974).
23. Adams, J.M. & Cory, S. Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA. *Nature* **255**, 28-33 (1975).
24. Wei, C.M., Gershowitz, A. & Moss, B. 5'-Terminal and internal methylated nucleotide sequences in HeLa cell mRNA. *Biochemistry* **15**, 397-401 (1976).
25. 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**, 387-94 (1975).
26. Wei, C., Gershowitz, A. & Moss, B. N6, O2'-dimethyladenosine a novel methylated ribonucleoside next to the 5' terminal of animal cell and virus mRNAs. *Nature* **257**, 251-3 (1975).
27. Narayan, P. & Rottman, F.M. An in vitro system for accurate methylation of internal adenosine residues in messenger RNA. *Science* **242**, 1159-62 (1988).
28. Horowitz, S., Horowitz, A., Nilsen, T.W., Munns, T.W. & Rottman, F.M. Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc Natl Acad Sci U S A* **81**, 5667-71 (1984).
29. Harper, J.E., Miceli, S.M., Roberts, R.J. & Manley, J.L. Sequence specificity of the human mRNA N6-adenosine methylase in vitro. *Nucleic Acids Res* **18**, 5735-41 (1990).
30. Bokar, J.A. The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. in *Fine-tuning of RNA functions by modification and editing* (ed. Grosjean, H.) 141-178 (Springer-Verlag, Berlin, Heidelberg, New York, 2005).
31. Kariko, K., Buckstein, M., Ni, H. & Weissman, D. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23**, 165-75 (2005).
32. Jia, G. et al. N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **7**, 885-7 (2011).
33. Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **419**, 174-8 (2002).
34. Delaney, J.C. & Essigmann, J.M. Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proc Natl Acad Sci U S A* **101**, 14051-6 (2004).

35. 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**, 889-94 (2007).
36. Gerken, T. et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469-72 (2007).
37. Church, C. et al. A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet* **5**, e1000599 (2009).
38. He, C. Grand challenge commentary: RNA epigenetics? *Nat Chem Biol* **6**, 863-5 (2010).
39. Yi, C. & Pan, T. Cellular dynamics of RNA modification. *Acc Chem Res* **44**, 1380-8 (2011).

Chapter 2. Probing the m^{6}A Sites and Methylation Fraction at Nucleotide-Resolution

2.1 Introduction

Mammalian mRNAs contain ~3 m^{6}A residues per transcript[1]. But how m^{6}A is distributed along the transcript remains obscure until recently. m^{6}A detection is challenging, since it is indistinguishable from A during reverse transcription, and there are no known chemical reagents which can specifically label m^{6}A to facilitate its detection.

The breakthrough in m^{6}A detection came in 2012, when two groups combined the anti-m^{6}A immunoprecipitation and deep sequencing to reveal the m^{6}A topology along the mammalian transcriptome[2,3]. Briefly, polyA-enriched or rRNA-depleted RNAs were first chemically fragmented into average sizes of ~100 nucleotides as the input control sample. Then, the m^{6}A-specific antibody was applied to immuno-precipitate (IP) m^{6}A-containing mRNA fragments as the IP sample. Both the input control and IP samples were subjected to deep sequencing to detect the m^{6}A-containing RNA fragments. This method, called m^{6}A/MeRIP-seq in this perspective, identified tens of thousands of m^{6}A peaks among ~25% of all transcripts in human cells. The m^{6}A/MeRIP-seq peaks are on average ~100-200 nucleotides wide, due to the overlaps of RNA fragments from the initial fragmentation. Most surprisingly, these m^{6}A peaks are not evenly distributed along the transcript. The m^{6}A peaks are enriched near stop codons and in the 3' untranslated region (UTR), although m^{6}A also locates in the coding region and 5'UTR. The physiological meaning of this asymmetric distribution is still unclear. Remarkably, the m^{6}A/MeRIP-seq methods also detected the anti-correlated localization pattern between m^{6}A peaks and miRNA-binding sites within 3'UTRs, and suggested that microRNA levels may control methylation of their target RNAs[3].

Despite powerful, the m^6A/MeRIP-Seq method cannot identify which adenosine residue under the deep sequencing peaks is actually modified, nor can it determine the modification fraction for any modification site. As the dynamic m^6A status could be crucial for its function, new methods that can unambiguously determine the m^6A status at single nucleotide resolution are needed to further understand the biological function of this highly abundant modification. To address this challenge, we developed a method that directly measures the precise location and modification fraction, i.e., the m^6A status in any candidate site in mRNA/lncRNA at single nucleotide resolution.

2.2 Results and Discussion

2.2.1 A method to detect m^6A status in mRNA/lncRNA

Our method combines site specific cleavage of each m^6A containing candidate site followed by radio-labeling and site-specific ligation and complete nuclease digestion to enable isolation of the candidate m^6A containing residue in the mixture of total RNA or $polyA^+$ RNA (**Figure 2.1**). The candidate m^6A containing residue is then analyzed by thin-layer chromatography (TLC) to reveal the presence or absence of m^6A and its modification fraction. Using the site-specific RNase H digestion followed by radio-labeling and TLC analysis of an RNA modification residue was applied previously to investigate pseudouridine (Ψ) modification in U2 small nuclear RNA[4-7]. These highly successful studies were made possible by first purifying the abundant U2 snRNA from total RNA. However, the low abundance of individual mRNA/lncRNA makes it very difficult and even impossible for efficient isolation, therefore, modifications in individual mRNA/lncRNA cannot be studied as those in the abundant U2 snRNA.

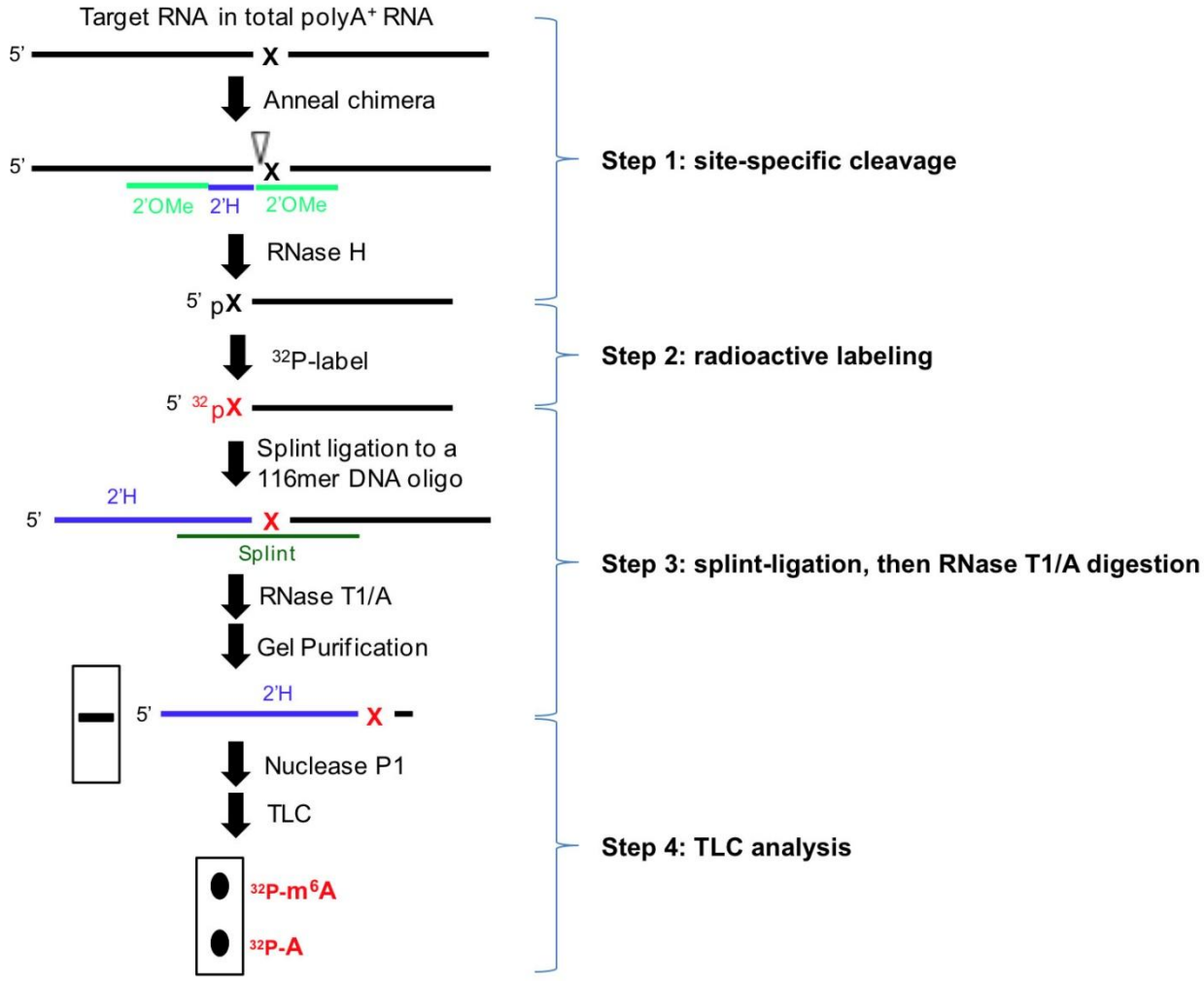


Figure 2.1. Schematic diagram of SCARLET. SCARLET consists of four steps: site-specific cleavage at the target nucleotide site; radioactive labeling of the target nucleotide; splint-assisted ligation followed by RNases T1/A digestion; and thin Layer Chromatography (TLC).

We aimed at developing a method that directly determines m^6A modification status at any mRNA/lncRNA site from the total RNA sample without the need of purifying a specific RNA. To achieve this, we combined RNase H site-specific cleavage, splinted ligation, ribonuclease digestion and TLC to generate a new method, named Site-specific Cleavage And

Radioactive-labeling followed by Ligation-assisted Extraction and TLC (SCARLET, **Figure 2.1**).

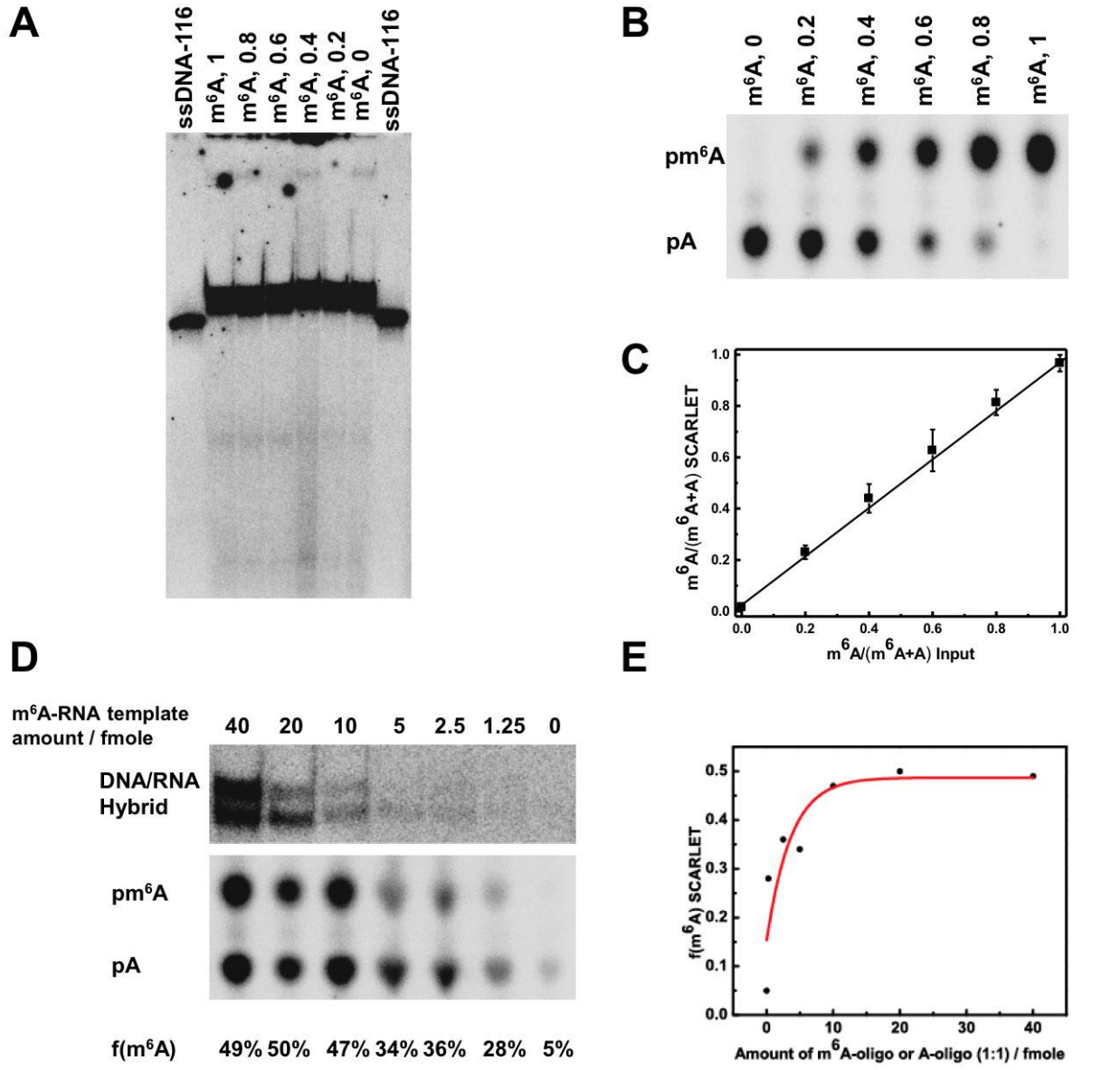


Figure 2.2. SCARLET detects m^{6}A status. A) Denaturing PAGE showing the ^{32}P-labeled bands after ribonucleases T1/A digestion. Lanes m^{6}A-1, 0.8, 0.6, 0.4, 0.2, 0 correspond to 200 fmole total of the synthetic RNA template mixture containing 100%, 80%, 60%, 40%, 20%, 0% m^{6}A in the presence of 1 µg polyA$^+$ RNA from HeLa. ^{32}P-labeled 116 nucleotides single stranded DNA (ssDNA-116) was loaded as control. **B**) TLC result of the purified gel bands from panel A. **C**) Quantitative plot for the TLC result from panel C from three independent replicates. **D**) SCARLET result for varying amount of oligonucleotide mixtures containing equal moles of m^{6}A-modified and unmodified RNA templates in 1 µg HeLa polyA$^+$ RNA. The bands in the

(legend continued on next page)

upper denaturing PAGE figure represent the ^{32}P-labeled DNA/RNA hybrid after RNase T1/A digestion. The lower TLC plate shows the m^{6}A modification fraction for each corresponding mixture. **E**) The quantitative plot for the panel E result. The sequence of the RNA template used is (A = A or m^{6}A): 5'UUAGACUCUGAACAACGCGCGCUAUAGAUAUAAAAAGUGG.

The crucial new feature is the ligation-assisted extraction of the ^{32}P-labeled RNA of interest. SCARLET starts with the total RNA or total polyA$^+$ RNA sample. First, we choose a candidate site in a candidate RNA of interest. We apply RNase H cleavage guided by a complementary 2'OMe/2'H chimeric oligonucleotide to achieve site-specific cleavage 5' to the candidate site[4,8]. The cut site is radiolabeled with ^{32}P and the ^{32}P-labeled RNA fragment splint-ligated to a 116-nucleotide single-stranded DNA oligonucleotide using DNA ligase. The sample is then treated with RNases T1/A to completely digest all RNA, while the ^{32}P-labeled candidate site remains with the DNA oligonucleotide as DNA-^{32}P-(A/m^{6}A)p and DNA-^{32}P-(A/m^{6}A)Cp, which migrate as 117/118-mers on a denaturing gel. This labeled band is excised and eluted from the gel, digested with nuclease P1 into mononucleotides containing 5' phosphate, and the m^{6}A modification status is determined by thin-layer chromatography. One crucial step of SCARLET is the splint-ligation, which attaches the candidate ribonucleotide to one DNA oligo and thus prevents its digestion by RNase T1/A.

We first tested this method using mixtures of two synthetic, 40-mer RNA oligonucleotides of the same sequence, one unmodified and the other containing one defined m^{6}A modification. To mimic the condition for biological site determination, this RNA oligo mixture at varying ratios was added to 1µg HeLa total polyA$^+$ RNA before the start of SCARLET (**Figures 2.2A-B**). The m^{6}A fraction determined by SCARLET accurately reflected the input m^{6}A fraction (**Figure 2.2C**), indicating that SCARLET can quantitatively determine m^{6}A fraction from the total polyA$^+$ RNA pool. Furthermore, SCARLET was completely accurate

when ~10 femtomole of the RNA template was present and still sensitive when as little as ~1 femtomole RNA template was present in 1 µg polyA$^+$ RNA (**Figure 2.2D-E**).

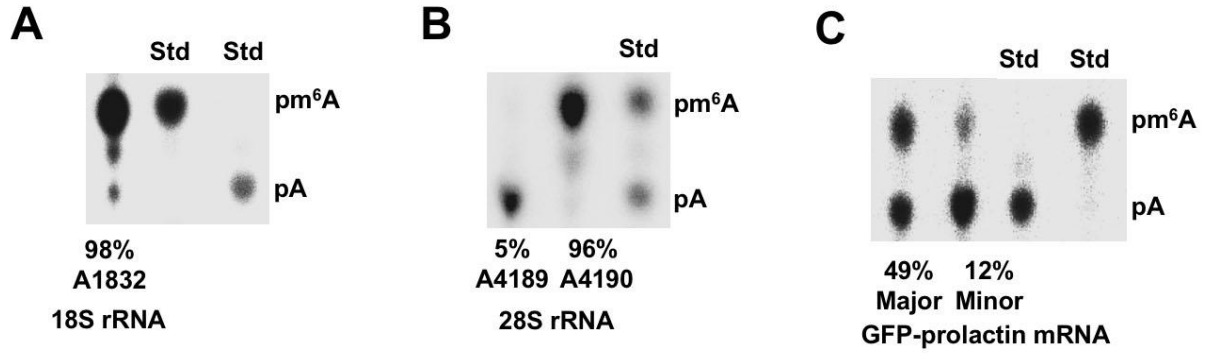


Figure 2.3. Determination of m^{6}A status of rRNA and an mRNA from a transfected plasmid. A) SCARLET result determining the m^{6}A modification fraction at the A1832 site of human 18S rRNA. Standard ^{32}P-labeled m^{6}A and A nucleoside were run as controls. **B**) SCARLET result determining the m^{6}A status at the A4189 and A4190 site of human 28S rRNA. **C**) SCARLET result determining the m^{6}A status at two sites of the bovine prolactin mRNA segment derived from a transfected plasmid in the MDA-MB-231 cell line.

To test the feasibility of SCARLET for biological sites, we first applied the method to determine the modification fraction of the previously reported m^{6}A sites in human rRNA in a total HeLa RNA sample[9]. The A1832 of human 18S rRNA is essentially fully modified at ~98% (**Figure 2.3A**). We also determined the m^{6}A status at A4189 and A4190 of human 28S rRNA; previous methods could not make clear which one of these two sites is modified [9]. We found that A4190 is ~96% modified while A4189 site is minimally modified (**Figure 2.3B**), indicating that SCARLET can easily resolve ambiguity of modification sites.

We also determined the m^{6}A fraction in an mRNA derived from a transfected plasmid in a human breast cancer cell line[10]. This mRNA fuses the coding sequence of GFP with the 3' UTR of bovine prolactin mRNA which is known to have one strong and one cryptic m^{6}A modification site[11]. We found that the modification fraction is 49% at the major site and 12% at

the cryptic site in this breast cancer cell line (**Figure 2.3C**). These results indicate that SCARLET can readily identify m^{6}A sites in a variety of cellular RNAs.

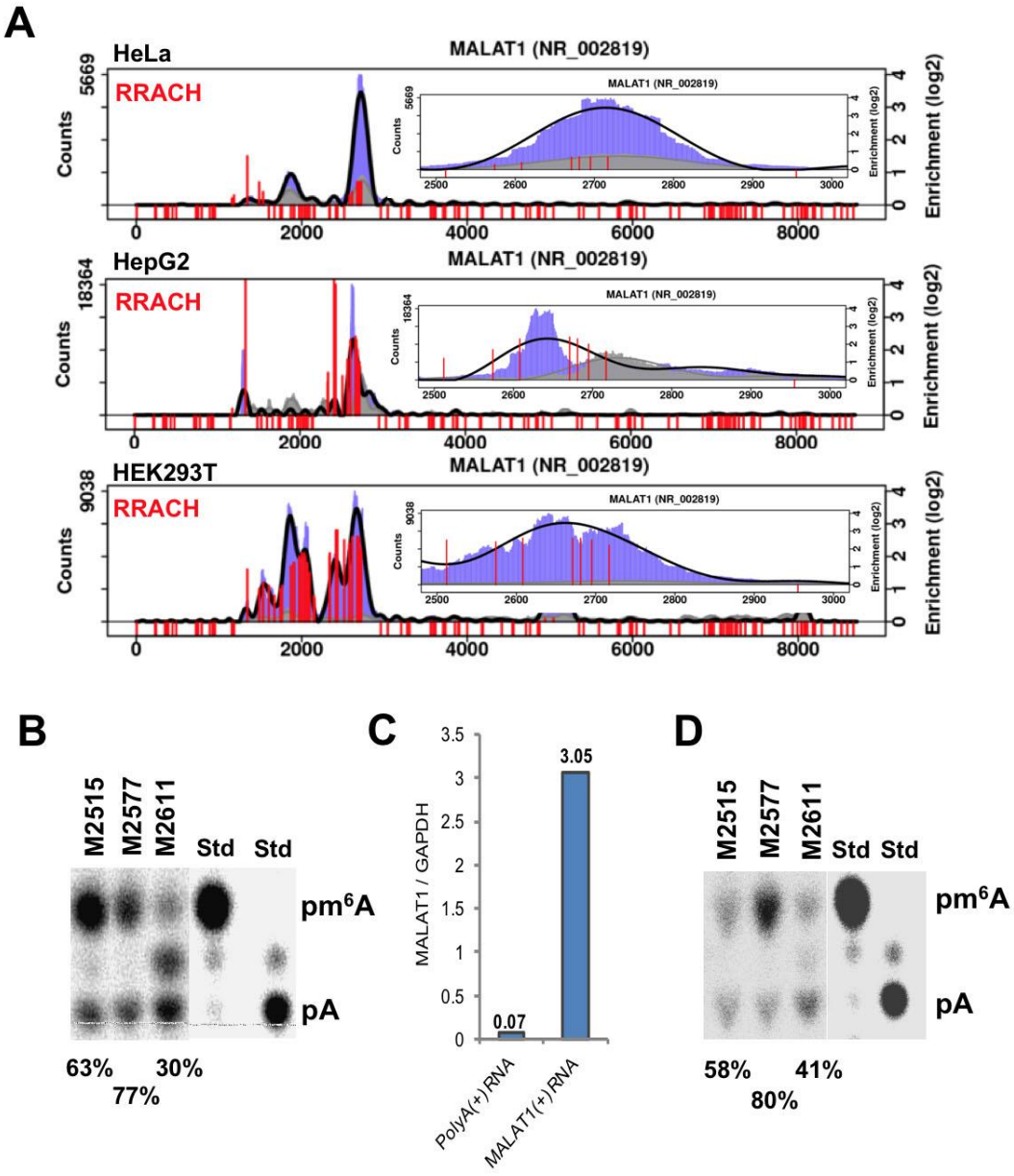


Figure 2.4. SCARLET determination of m^{6}A status in the MALAT1 lncRNA. A) m^{6}A/MeRIP-seq data for the lncRNA MALAT1 in three cell lines. HeLa data is from our own experiment, HepG2 from [17] and HEK293T from [18]. The red dots and lines correspond to all RRACH consensus motifs. The m^{6}A-antibody-IP data is in blue and the input sample is in gray. The RRACH consensus sites under the enriched peaks are shown as extended red lines.

(legend continued on next page)

B) SCARLET result for the three MALAT1 m^6A sites displaying the highest modification fraction, determined from total HeLa $polyA^+$ RNA. **C**) Affinity enrichment of MALAT1 RNA using 3'-biotinylated DNA oligos complementary to MALAT1 residues 2449-2498. MALAT1 is ~43 folds enriched compared with GAPDH-mRNA, as demonstrated by real time quantitative RT-PCR. **D**) SCARLET result for the same three MALAT1 m^6A sites as in panel B, determined from MALAT1 enriched RNA.

2.2.2 m^6A status in lncRNA and mRNA

We next applied SCARLET to determine the m^6A status in the nuclear-localized Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1), a relatively abundant lncRNA of > 6.5 kb and is known to regulate alternative splicing[12-14]. m^6A/MeRIP-seq of the $polyA^+$ RNA from three human cell lines identified three peaks of varying intensities; and seven m^6A-consensus sequences consisting of RRACH[15,16] (R=A, G; H=A, C, U) are present in the largest peak (**Figure 2.4A**). We found that m^6A is present at four of these seven sites (**Figure 2.4B**, **Table 2.1**). Among the four modified sites, the modification fraction varies between 11-77% in HeLa and 7-51% in HEK293T, and the two upstream sites have higher m^6A fractions than the two downstream sites, even though the upstream sites generally showed lower m^6A/MeRIP-seq signal. This result may be explained by the structural context of the two upstream m^6A sites (see below). We also determined the m^6A status at these four MALAT1 sites in a breast cancer line MDA-MB231 and human foreskin fibroblast HFF1. Among these lines, the modification fraction varied by up to ~3-fold, although the rank order of m^6A fraction at these four sites was similar (**Table 2.1**).

To validate that SCARLET accurately reports m^6A status of endogenous RNAs in the complex mixture of $polyA^+$ RNA, we affinity enriched MALAT1 RNA from the total $polyA^+$ RNA using MALAT1 RNA complementary biotinylated oligonucleotides and re-determined the m^6A fraction at three sites. Our purification enriched MALAT1 by > 40-fold (**Figure 2.4C**); and

the m^6A fraction of the enriched MALAT1 was within 1.3-fold of those determined using total $polyA^+$ RNA (**Figure 2.4D**), validating that SCARLET works well in the complex mixture of total $polyA^+$ RNA.

We also determined the m^6A status in many RRACH sites in another lncRNA (TUG1) and three mRNAs (ACTB, TPT1 and BSG), chosen on the basis of the m^6A/MeRIP-seq results (**Figure 2.5**). Among the sixteen sites tested, only four showed m^6A modification above 5% (**Table 2.2**). This result shows that the previously reported 20% m^6A modification in bovine prolactin mRNA[19] is not abnormal, suggests that many sites in mRNA carry incomplete m^6A modification, and highlights that not every RRACH consensus sequence sites under the m^6A/MeRIP-seq peaks are modified.

Table 2.1. m^6A status in the seven RRACH consensus sites under the largest m^6A/MeRIP-seq peak of the MALAT1 lncRNA. a. m^6A sites with >5% modification are shown in bold. b. Not determined. c. f(m^6A) for HeLa have been tested multiple times, with the standard deviation shown at the upper right of the average value.

MALAT1 Site	Consensus motif	f(m^6A) [a]			
		HeLa [c]	HEK293T	MDA-MB231	HFF1
2515	**GGACU**	**0.61** $^{\pm 0.03}$	**0.41**	**0.39**	**0.67**
2577	**GGACU**	**0.80** $^{\pm 0.03}$	**0.51**	**0.58**	**0.88**
2611	**GGACU**	**0.38** $^{\pm 0.07}$	**0.13**	**0.20**	**0.49**
2674	AGACA	0.03	0.03	--[b]	--
2684	AGACA	0.02	0.02	--	--
2698	GAACC	0.02	0.03	--	--
2720	**GGACU**	**0.10** $^{\pm 0.02}$	**0.07**	**0.08**	**0.14**

Table 2.2. m^{6}A status in the lncRNA TUG1 and 3 mRNAs (ACTB, TPT1 and BSG).
a. m^{6}A sites with >5% modification are shown in bold. b. Not determined.

				fraction (m^{6}A)[a]	
ID	Name	Site	Motif	HeLa	MDA-MB-231
NR_002323	**TUG1**	**1114**	**GGACU**	**0.22**	**0.13**
NR_002323	TUG1	2497	GGACC	0.01	--[b]
NR_002323	TUG1	2564	GGACC	0.01	--
NR_002323	TUG1	2601	GAACA	0.01	--
NR_002323	TUG1	2953	AAACU	0	--
NR_002323	TUG1	3071	GGACU	0.04	--
NR_002323	TUG1	5125	AGACU	0.01	--
NR_002323	TUG1	5133	AAACA	0.01	--
NR_002323	TUG1	5138	GAACC	0.01	--
NR_002323	TUG1	5910	AGACU	0	--
NM_001101	**ACTB**	**1216**	**GGACU**	**0.22**	**0.28**
NM_003295	**TPT1**	**687**	**GGACU**	**0.15**	**0.19**
NM_003295	TPT1	694	AGACA	0.04	--
NM_003295	TPT1	703	GGACU	0.01	--
NM_198591	**BSG**	**1335**	**GGACU**	**0.06**	**0.08**
NM_198591	BSG	1442	GAACU	0.01	--

2.2.3 An m^{6}A-containing RNA structural motif

The precise mapping of m^{6}A sites enabled the determination of RNA secondary structural context in which m^{6}A occurs. According to the RNA secondary structure prediction, the two m^{6}A residues in MALAT1 lncRNA (M2515, M2577 site) displaying the highest modification fractions are located in two hairpin stems (**Figure 2.6A**). We carried out phylogenic comparison for the MALAT1 RNA among primates (**Figure 2.6B-C**). Phylogenetic co-variations of two base pairs in the stem are present that support the hairpin structure of the M2515 site. M2577 site is

only conserved among great apes and human and no co-variation is present within this very limited data set.

The location of both m^{6}A sites in the stem of a hairpin loop, however, may hinder the access of the m^{6}A antibody. The structural context of these two m^{6}A sites is therefore consistent with the low-intensity m^{6}A/MeRIP-seq reads for these two sites.

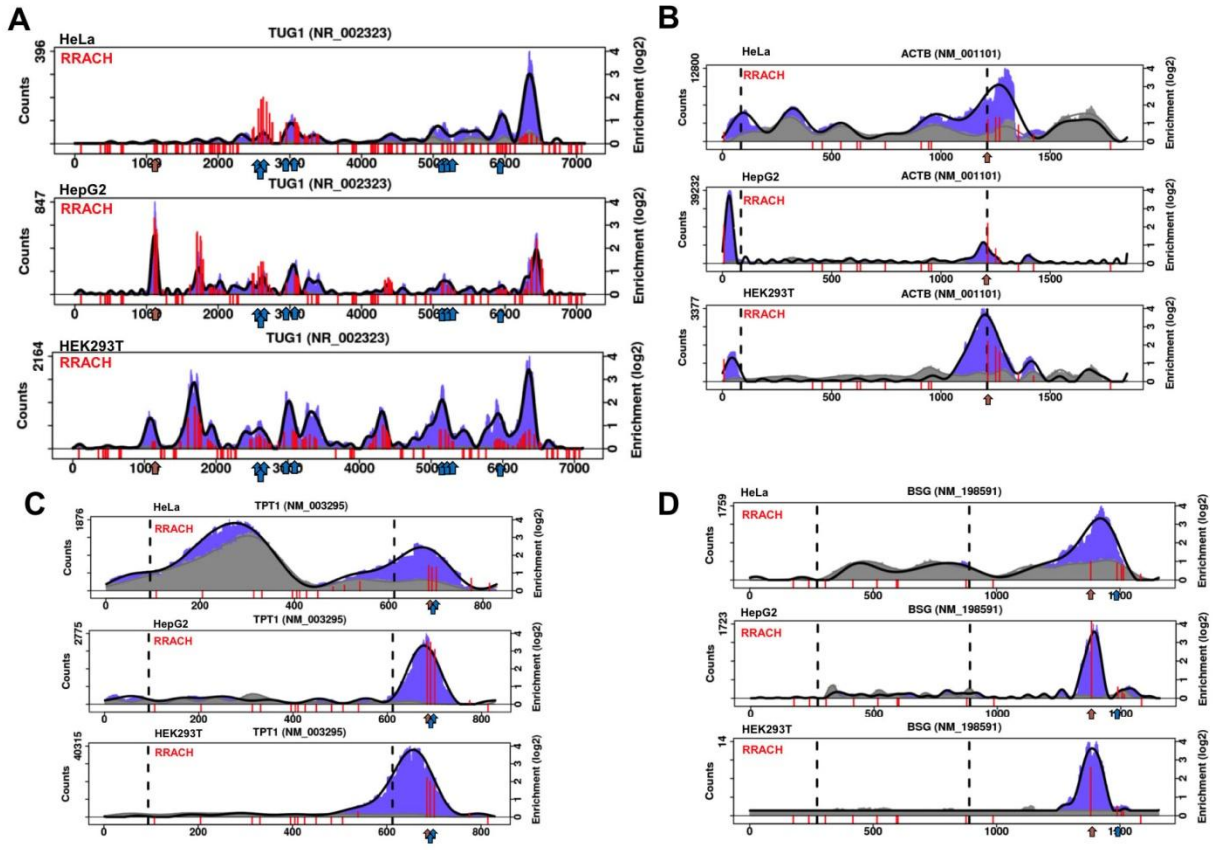


Figure 2.5. m^{6}A/MeRIP-seq analysis of four RNAs tested for m^{6}A status in this work. They include one lncRNA TUG1 (**A**), ACTB mRNA (**B**), TPT1 mRNA (**C**) and BSG mRNA (**D**) in three different human cell lines (HeLa, HepG2 and HEK293T). SCARLET tested sites are labeled with bold arrows. The sites tested positive for m^{6}A are indicated by purple arrows, whereas the sites tested negative for m^{6}A are indicated by blue arrows.

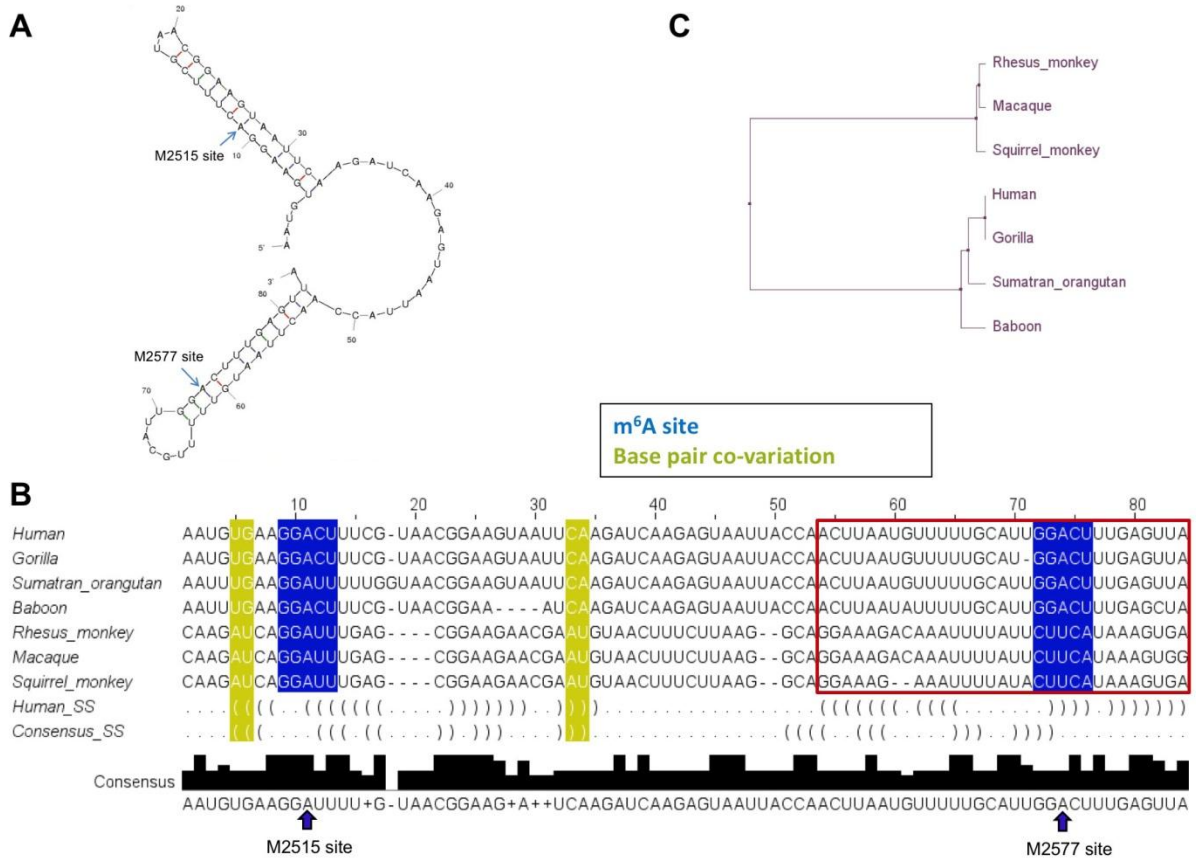


Figure 2.6. Prediction of secondary structure motifs around two MALAT1 m^{6}A residues displaying highest modification fraction. A) Mfold prediction of residues 2505-2587. The two m^{6}A modification sites are labeled with blue arrows and predicted to be located in two separate hairpin stems. **B**) Sequence alignment of MALAT1 homologs among primates. The M2515 site is conserved and the hairpin structure supported by two base pair co-variations. The M2577 site is conserved only among human, gorilla, orangutan and baboon, and no sequence change is present to provide information on structural conservation. **C**) Phylogenetic tree of the MALAT1 homologs.

2.2.4 Concluding remarks

In conclusion, we report the first nucleotide-resolution method to determine m^{6}A location and modification fraction in mammalian mRNA/lncRNA. Starting from total polyA$^+$ RNA, SCARLET determined the m^{6}A status in several sites in two human lncRNAs and three human mRNAs, and we found that m^{6}A fraction varies between 6-80% among these sites. We also

found that many consensus RRACH sequences under the m^6A/MeRIP-seq peaks were not modified. One crucial advantage of SCARLET is its ability to directly determine the m^6A modification fraction at any site, an important consideration in elucidating the biological function of m^6A dynamics which is currently not accessible using existing techniques.

In principle, SCARLET can also be used to investigate RNA modifications other than m^6A, such as 5-methyl-C (m^5C), pseudo-uridine (Ψ), 2'-*O*-methyl ribonucleosides (Nm), whose modification fractions or locations in mammalian mRNA/lncRNAs still remain obscure. SCARLET requires only common and readily available lab equipment and material and should be readily applicable to investigate the dynamics and biology of RNA modifications.

2.3 Experimental section

2.3.1 Cell culture

Human cervical cancer cell line HeLa (CCL-2), embryonic kidney cell line HEK293T (CRL-11268), foreskin fibroblasts cell line HFF-1 (SCRC-1041) and breast cancer cell line MDA-MB-231 (HTB-26) were obtained from American Type Culture Collection (ATCC) and were cultured under standard conditions. Cells of ~70% confluency were harvested for RNA extraction.

2.3.2 SCARTLET oligonucleotide sequences

PB-A/m^6A Chimera: Nm = 2'Ome-modified nucleotide.
5'mGmCmGmUmUmGmUTCAGmAmGmUmCmUmAmA3'
PB-A/m^6A Splint:
5'CCACTTTTTATATCTATAGCGCGCGTTGTCTATTAACTCACAGGACCGGCGATGGCTG3'
M2515 Chimera: 5'mAmCmGmAmAmAmGmUCCTTmCmAmCmAmUmU3'
M2515 Splint:
5'TGATCTTGAATTACTTCCGTTACGAAAGTCTATTAACTCACAGGACCGGCGATGGCTG3'
M2577 Chimera: 5'mUmCmAmAmAmGmUCCAAmUmGmCmAmAmAmA3'
M2577 Splint:
5'GGATTTAAAAAATAATCTTAACTCAAAGTCTATTAACTCACAGGACCGGCGATGGCTG3'
M2611 Chimera: 5'mUmGmCmUmAmGmUCCTCmAmGmGmAmUmUmU3'
M2611 Splint:

5’ACCTGGGTCAGCTGTCAATTAATGCTAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
M2674 Chimera: 5’mCmUmGmCmUmGmUCTTCmCmUmAmGmAmUmU3’
M2674 Splint:
5’CTGGTTCCTGGAATCCTGTCTGCTGCTGTCTATTAACTCACAGGACCGGCGATGGCTG3’
M2684 Chimera: 5’mAmUmCmCmUmGmUCTGCmUmGmCmUmGmUmC3’
M2684 Splint:
5’TCATCAAACACTGGTTCCTGGAATCCTGTCTATTAACTCACAGGACCGGCGATGGCTG3’
M2698 Chimera: 5’mCmAmCmUmGmGmUTCCTmGmGmAmAmUmCmC3’
M2698 Splint:
5’CCTCAGTCCTAGCTTCATCAAACACTGGTCTATTAACTCACAGGACCGGCGATGGCTG3’
M2720 Chimera: 5’mCmCmUmCmAmGmUCCTAmGmCmUmUmCmAmU3’
M2720 Splint:
5’AACTGCTGCTTGCTCGCTTGCTCCTCAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
ATCB1214 Chimera: 5’mCmAmUmAmGmUCCGCmCmUmAmGmAmAmG3'
ATCB1214 Splint:
5’GAAAGGGTGTAACGCAACTAAGTCATAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
TPT1 687 Chimera: 5’ mCmUmUmAmAmGmUCCTGmGmUmGmUmUmGmUmG 3’
TPT1 687 Splint:
5’AGATGACATCAGTCCCATTTGTCTTAAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
TPT1 694 Chimera: 5’ mAmUmUmUmGmUCTTAmAmGmUmCmCmUmGmGmU 3’
TPT1 694 Splint:
5’GAGCTCAAGATGACATCAGTCCCATTTGTCTATTAACTCACAGGACCGGCGATGGCTG3’
TPT1 703 Chimera: 5’ mAmCmAmUmCmAmGmUCCCAmUmUmUmGmUmCmU 3’
TPT1 703 Splint:
5’ATAAATGAAGAGCTCAAGATGACATCAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
18S rRNA Chimera:
5’mAmCmCmUmUmGmUTACGmAmCmUmUmUmUmA3'
18S rRNA Splint:
5’TTCCGCAGGTTcACCTACGGAAACCTTGTCTATTAACTCACAGGACCGGCGATGGCTG3’
GFP-Prolactin Major Chimera:
5’mAmGmUmCmUmGmUTTTTmAmUmUmUmAmAmG3’
GFP-Prolactin Major Splint:
5’GCAGATTTTGACATCGCTACAGAGTCTGTCTATTAACTCACAGGACCGGCGATGGCTG3’
GFP-Prolactin Minor Chimera:
5’mAmCmAmGmAmGmUCTGTmUmUmUmUmAmUmU3’
GFP-Prolactin Minor Splint:
5’GCGCGCAGATTTTGACATCGCTACAGAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
T-1114 Chimera: 5’mUmAmAmAmAmGmUCCACmGmUmGmCmAmCmC3’
T-1114 Splint:
5’CTCTTCCAGTGAGCCCGCTTGCTAAAAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
T-2497 Chimera: 5’mCmCmCmUmGmGmUCCAAmAmUmUmUmAmUmUmGmU3’
T-2497 Splint:
5’GAGCATTAATAACTAAAAAATCCCCTGGTCTATTAACTCACAGGACCGGCGATGGCTG3’
T-2564 Chimera: 5’mGmUCCAAmUmGmCmUmGmCmUmCmU3’
T-2564 Splint:
5’TTCAATCATTTGAGATTGTGGGGTGTGGTCTATTAACTCACAGGACCGGCGATGGCTG3’
T-2601 Chimera: 5’mUmCmCmUmAmGmAmUmGmUTCATmGmAmAmUmUmUmC3’
T-2601 Splint:
5’TCCAGTGACCTTCACGGGATCCTAGATGTCTATTAACTCACAGGACCGGCGATGGCTG3’
T-2953 Chimera: 5’mGmCmAmCmAmGmUTTTCmAmCmUmCmAmAmA3’

T-2953 Splint:
5’TTCACTTTTACTCTGGGCTTCTGCACAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
T-3071 Chimera: 5’mAmGmGmAmAmGmUCCCCmUmUmAmGmGmUmG3’
T-3071 Splint:
5’CGGAAGAGTTCAAGGTGTTGTAAGGAAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
T-5125 Chimera: 5’mGmAmGmUCTATmGmCmUmAmCmAmUmCmU3’
T-5125 Splint:
5’CAATCAGACTTGAGGTTCTGTTTAGGAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
T-5133 Chimera:5’mGmGmUmUmCmUmGmUTTAGmGmAmGmUmCmUmA3’
T-5133 Splint:
5’CTTATCCTCAATCAGACTTGAGGTTCTGTCTATTAACTCACAGGACCGGCGATGGCTG3’
T-5138 Chimera: 5’mUmUmGmAmGmGmUTCTGmUmUmUmAmGmGmA3’
T-5138 Splint:
5’AAGGCCTTATCCTCAATCAGACTTGAGGTCTATTAACTCACAGGACCGGCGATGGCTG3’
T-5910 Chimera: 5’mGmUCTGAmGmAmGmCmCmCmAmA3’
T-5910 Splint:
5’CAACACAAGCAAACTCAGTGGTCATGAGTCTATTAACTCACAGGACCGGCGATGGCTG3’
ssDNA-116:
5’GGAGAGACAACTTAAAGAGACTTAAAAGATTAATTTAAAATTTATCAAAAAGAGTATTG
ACTTAAAGTCTAACCTATAGGATACTTACAGCCATCGCCGGTCCTGTGAGTTAATAG3’

2.3.3 Synthesis of m^6A-RNA oligo

m^6A-containing RNA oligonucleotides were synthesized according to a method reported previously[20].

2.3.4 SCARLET

Total RNA was isolated using PerfectPure RNA cultured cell kit (5 Prime). Polyadenylated RNA ($polyA^+$ RNA) was enriched via the GenElute mRNA miniprep kit (Sigma-Aldrich). 1 µg $polyA^+$ RNA was annealed with 3 pmole corresponding Chimeric oligo in 3 µl 30 mM TrisHCl, pH 7.5 by heating at 95 °C for 1 min. This sample was incubated with 1 µl 6x RNase H mix (2x T4 polynucleotide kinase buffer (T4 PNK, USB), 1 U/µl RNase H (Epicentre #R0601K)) and 1 U thermosensitive alkaline phosphatase (TAP, Thermo Scientific) at 44 °C for 1 h. RNase H and TAP were inactivated by heating at 75 °C for 5 min. The RNA fragment was radioactively labeled by adding 1 µl 6x T4 PNK mix (1x T4 PNK buffer, 6 U/µl T4 PNK, 28 µCi/µl [γ-^{32}P] ATP) and incubation at 37 °C for 1 h. T4 PNK was inactivated at 75 °C for 5 min.

The above mix was annealed with 4 pmole corresponding Splint and 5 pmole 116 mer DNA oligo (ssDNA-116) by heating at 75 °C for 3 min . Ligation reaction was performed by incubating with 2.5 µl 4x ligation mix (1.4x T4 PNK buffer, 0.27 mM ATP, 57% DMSO, 1.9 U/µl T4 DNA ligase) at 37 °C for 3.5 h, and stopped by mixing with equal volume of 2x RNA loading buffer (9 M urea, 100 mM EDTA). RNA was digested by incubating with 1 µl RNase T1/A mixture (160 U/µl RNase T1, 0.16 mg/ml RNase A) at 37 °C overnight. The ligation product was isolated by denaturing-PAGE followed by crush&soak of corresponding gel slices. The RNA pellet was digested into individual nucleotides by incubation with 3 µl nuclease P1 mix (0.33 U/µl nuclease P1 in 30 mM sodium acetate/acetic acid, pH 4.8) at 37 °C for 2 h. The reaction mix was spotted on cellulose TLC plate (20 x 20 cm; Merck), and TLC was run with isopropanol:HCl:water (70:15:15, v/v/v). After that, dry the TLC plate at room temperature for 1 h, wrap the plate in plastic film and expose it to a blanked phosphorimager screen. The result can be visualized on a phosphorimager.

Detailed procedures for SCARLET were described at recent publications [21,22].

2.3.5 RNA structural mapping

The synthetic MALAT1 RNA oligos were 5'-labeled with [γ-^{32}P] ATP by T4 PNK and gel purified. Mix ^{32}P-labeled RNA, 0.25 µg/ µl E. coli tRNA, 62.5 mM Tris-HCl and RNase free water in a total volume of 36.8 µl. Heat at 90 °C for 2 min, then cool down at room temperature for 3 mins. Add 4.6 µl of 24 mM $MgCl_2$ and 4.6 µl of 1M KCl, then place at room temperature for 5 mins. Split into 4 separate aliquots, and add 1 µl freshly made nuclease stock (nuclease V1, T1, S1 or RNase free water as CNTL) to each aliquots. Incubate at room temperature for 15 mins. Add 5 µl RNA loading buffer, and immediately place on ice until loading on a 20% urea-denaturing gel. For the alkaline hydrolysis group, mix end-labeled RNA, 1 µl 1 µg/ µl E. coli

tRNA, 1 µl 5x BH buffer, and RNase free water in a total volume of 5 µl. Heat at 95 °C for 1 min. Quick spin down to bottom. Add 5 µl of urea loading buffer. Immediately place on ice until loading on the gel. For the G-ladder group, mix end-labeled RNA, 1 µl 1 µg/ µl E. coli tRNA, 1 µl Nuclease T1 (1U/µl), and RNase free water in 5 µl volume. Heat the sample at 65 °C for 1 min and then spin down to bottom. Add 5 µl of urea loading buffer. Immediately put on ice until loading on the gel.

2.3.6 m^{6}A/MeRIP-seq and its data analysis

The m^{6}A/MeRIP-seq for HeLa RNA was performed as previously described[23].

For m^{6}A/MeRIP-seq data analysis, we used the deep sequencing data from GEO sets GSE37005 (HepG2) and GSE29714 (HEK293T). The deep sequencing data was mapped to a *Homo Sapiens* genome provided in a GenBank flat file (retrieved April 1st, 2013). The mapping program is an in-house mapper allowing for at most one mismatch for an alignment of at least 22 nucleotides. When a read aligns to many locations, the longest alignment is retained. For enrichment computations, both the control and the IP'ed data were mapped and compared for each gene (coding RNAs with IDs of the form NM_xx and non-coding RNAs of the form NR_xx). The raw data aligned on a gene makes for a "mountain" plot. Because the plot is edgy we smoothed out the plot using low-frequency fast Fourier transforms (FFT), yielding a smoothed mountain plot. From the smoothed mountain plots for both the control and IP'ed data we compute the enrichment values for any given site on a gene (e.g. an RRACH position): we first find the peaks of the IP'ed data on the smoothed mountain plot. At these peak positions, their enrichment values, Ep, is the log ratio of the IP'ed data (ip) to the control (ck), and to the height of the peak, Hp, to the average height across the whole gene, Hg: Ep = log((Hp/Hg)_ip / (Hp/Hg)_ck). The enrichment value at a given gene location El, separate by a distance d to a

peak is thus: El = sum over all peaks p of Ep * exp(-d). On the mountain plots, the raw IP'ed data is shown in blue, the control data in grey, and the FFT-smoothed curves in black. Red vertical bars indicate the RRACH sites with the height of the bar indicating the enrichment value at that genomic location.

2.4 References

1. Bokar, J.A. The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. *Fine-tuning of RNA functions by modification and editing* 141-177 (Springer, 2005).
2. Dominissini, D. et al. Topology of the human and mouse m^6A RNA methylomes revealed by m^6A-seq. *Nature* **485**, 201-206 (2012).
3. Meyer, K. et al. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. *Cell* **149**, 1635-1646 (2012).
4. Zhao, X. & Yu, Y.T. Detection and quantitation of RNA base modifications. *Rna* **10**, 996-1002 (2004).
5. Ma, X. et al. Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *Embo J* **24**, 2403-13 (2005).
6. Wu, G., Xiao, M., Yang, C. & Yu, Y.T. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J* **30**, 79-89 (2011).
7. Wang, X. et al. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388-99 (2015).
8. Yu, Y.T., Shu, M.D. & Steitz, J.A. A new method for detecting sites of 2'-O-methylation in RNA molecules. *Rna* **3**, 324-31 (1997).
9. Piekna-Przybylska, D., Decatur, W.A. & Fournier, M.J. The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res* **36**, D178-83 (2008).
10. Vilfan, I.D. et al. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J Nanobiotechnology* **11**, 8 (2013).
11. Horowitz, S., Horowitz, A., Nilsen, T.W., Munns, T.W. & Rottman, F.M. Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc Natl Acad Sci U S A* **81**, 5667-71 (1984).
12. Wilusz, J.E., Freier, S.M. & Spector, D.L. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919-32 (2008).
13. Bernard, D. et al. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J* **29**, 3082-93 (2010).
14. Tripathi, V. et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **39**, 925-38 (2010).
15. Csepany, T., Lin, A., Baldick, C.J., Jr. & Beemon, K. Sequence specificity of mRNA N6-adenosine methyltransferase. *J Biol Chem* **265**, 20117-22 (1990).

16. Harper, J.E., Miceli, S.M., Roberts, R.J. & Manley, J.L. Sequence specificity of the human mRNA N6-adenosine methylase in vitro. *Nucleic Acids Res* **18**, 5735-41 (1990).
17. Dominissini, D. et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-6 (2012).
18. Meyer, K.D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635-46 (2012).
19. Narayan, P. & Rottman, F.M. An in vitro system for accurate methylation of internal adenosine residues in messenger RNA. *Science* **242**, 1159-62 (1988).
20. Dai, Q. et al. Identification of recognition residues for ligation-based detection and quantitation of pseudouridine and N6-methyladenosine. *Nucleic Acids Res* **35**, 6322-9 (2007).
21. Liu, N., Pan, T. Probing RNA modification status at single nucleotide resolution in messenger and long non-coding RNA. *Methods in Enzymology* (Available online 2 June 2015).
22. Liu, N., Pan, T. Probing m^6A RNA methylation with SCARLET. In *Methods in Molecular Biology: Post-transcriptional Gene Regulation* **1358**, 285-292 (2015).
23. Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N. & Rechavi, G. Transcriptome-wide mapping of N(6)-methyladenosine by m(6)A-seq based on immunocapturing and massively parallel sequencing. *Nat Protoc* **8**, 176-89 (2013).

Chapter 3. Widespread N^6-methyladenosine-dependent RNA Structural Switches Regulate RNA-Protein Interactions

3.1 Introduction

RNA-protein interactions determine many aspects of cellular biology, and related defects lead to numerous human diseases, including muscular atrophies, neuropathies, metabolic disorders, and cancers[1]. So far, over 1,000 RNA-binding proteins (RBPs) have been identified in humans, and the majority of them bind messenger RNA (mRNA) and long non-coding RNA (lncRNA)[2,3]. Most mRNA/lncRNA RBPs bind single-stranded RNA binding motifs (RBM) that are often buried in local RNA secondary structures[4,5]. Although mRNA/lncRNAs are less structured than ribosomal RNA (rRNA) or transfer RNA (tRNA), appreciable amounts of secondary structures have been found in several transcriptome-wide studies[6-9]. A major goal of understanding mRNA/lncRNA-RBP interactions is to identify how RBPs gain regulated access to their binding sites.

Here, we report the widespread m^6A-dependent regulation of RNA-protein interactions through RNA structural remodeling, termed m^6A-switches. We found that the m^6A residue can alter RNA structure to increase the accessibility of RBMs for heterogeneous nuclear ribonucleoprotein C (HNRNPC), thus significantly enhancing RNA-HNRNPC interactions. Through combining PAR-CLIP with Me-RIP, we identified 39,060 m^6A-switches in HNRNPC RNA binding sites. Among them, HNRNPC binding activities at 2,798 m^6A-switches significantly decreased upon m^6A-methyltransferases (*METTL3/L14*) knockdowns. Further analysis of these m^6A-switches revealed their enrichment in introns and their influences on cell proliferation and alternative splicing.

3.2 Results and Discussion

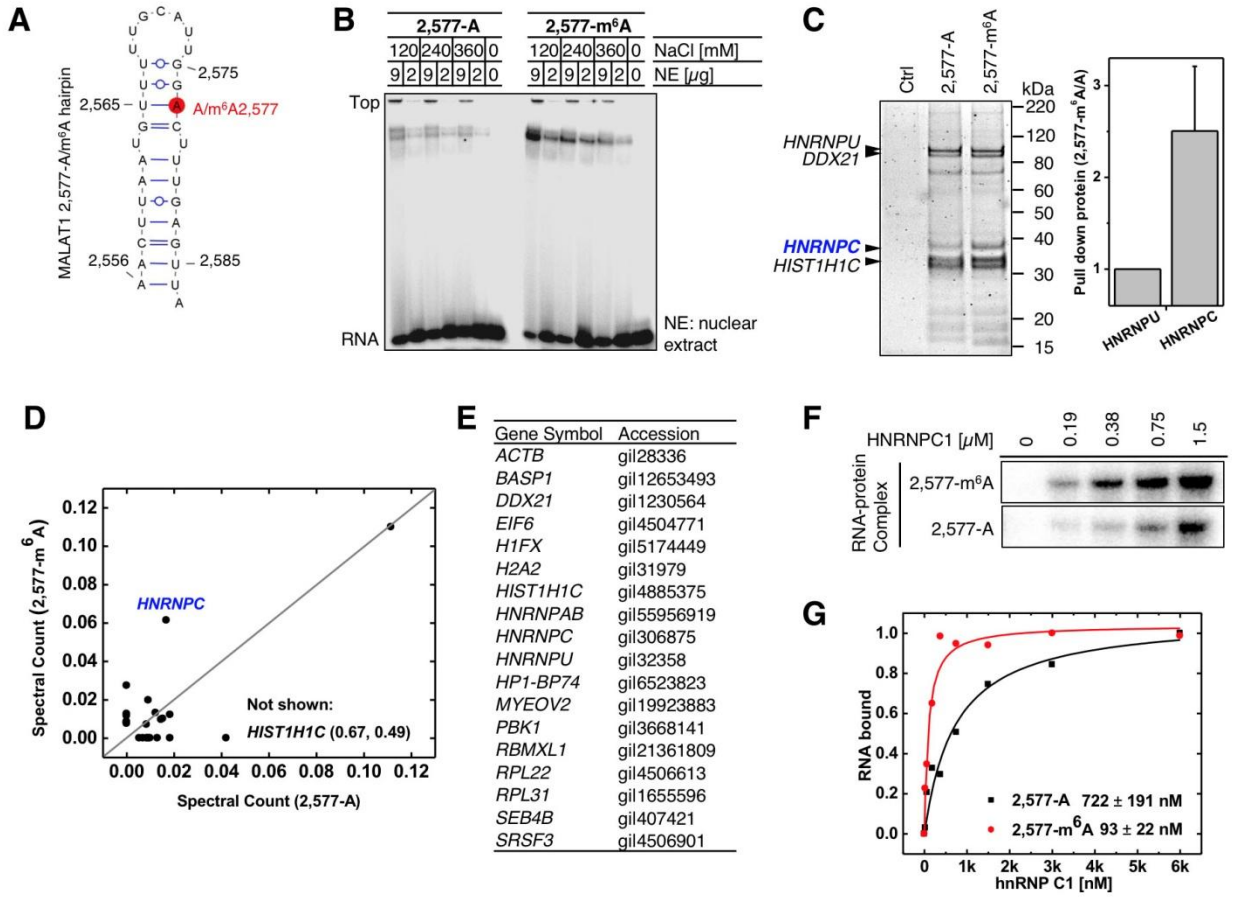


Figure 3.1. HNRNPC selectively binds m^{6}A methylated RNA. A) Secondary structure of the MALAT1 hairpin with m^{6}A methylation at 2,577 site shown in red[13]. Nucleotide position numbers correspond to their locations along the human MALAT1 transcript (NR_002819). **B**) m^{6}A increases binding with nuclear proteins. **C**) RNA pull down showing HNRNPC preferably binds methylated RNA. n = 4, ± s.d., biological replicates. **D**) RNA pull down showing HNRNPC preferably binds methylated RNA. **E**) The list of proteins with identified peptides by mass spectrometry in **D**. **F**) Recombinant HNRNPC1 binds stronger with MALAT1 2,577-m^{6}A hairpin compared to the unmethylated hairpin, as determined by *in vitro* UV crosslinking assay[14]. **G**) Filter binding showing m^{6}A increases HNRNPC1 binding with respective apparent dissociation constant (K_d) indicated at lower right; n = 3, ± s.d., technical replicates.

3.2.1 m^{6}A facilitates RNA-HNRNPC interactions *in vitro* and *in vivo*

We previously identified one specific m^{6}A site at A2,577 of the human long non-coding RNA MALAT1 (Metastasis Associated Lung Adenocarcinoma Transcript)[10]. The RNA region around this m^{6}A residue folds into a hairpin with the m^{6}A residue located in the stem[10] (**Figure 3.1A**). To examine the effect of the m^{6}A residue on RNA-protein interactions, we performed

native gel shift assay with the methylated/unmethylated MALAT1 hairpin and HeLa nuclear extract, since the MALAT1 RNA is localized in the nucleus[11,12]. Both hairpins formed two distinct complexes; however, the methylated hairpin bound more proteins under all conditions tested (**Figure 3.1B**), indicating that m^6A2,577 enhanced RNA-protein interactions.

To identify the protein components binding the MALAT1 hairpin, we performed RNA pull down assay under the same conditions as the native gel shift assay. We found most proteins bound the hairpins with similar amounts regardless of the methylation status, except one protein bound significantly stronger in the presence of m^6A (**Figure 3.1C**). Mass spectrometry of the gel slices identified this protein as the heterogeneous nuclear ribonucleoprotein C1/C2 (referred to as HNRNPC). Quantitative mass spectrometry of the entire pulled down protein samples showed that the 2,577-m^6A hairpin bound ~3-fold more HNRNPC protein as the unmethylated hairpin (**Figure 3.1D-E**). *In vitro* UV crosslinking assay[15] validated that recombinant HNRNPC1 protein binds the hairpin, and m^6A2,577 increases the binding affinity significantly (**Figure 3.1F**). Quantitative filter-binding assays[16] revealed that recombinant HNRNPC1 bound the methylated MALAT1 hairpin ~8-fold stronger than the unmethylated one (**Figure 3.1G**).

HNRNPC belongs to the large family of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (HNRNPs), which bind nascent transcripts to influence pre-mRNA processing and other aspects of mRNA metabolism[17]. HNRNPC is abundant in the nucleus, acts as a tetramer and regulates pre-mRNA splicing, export and translation[15-19]. Recently, iCLIP identified 438,360 HNRNPC binding sites transcriptome-wide at high resolution[15,19]; one binding site was in close proximity to A2,577 in MALAT1 which supports HNRNPC binding to this MALAT1 hairpin *in vivo*.

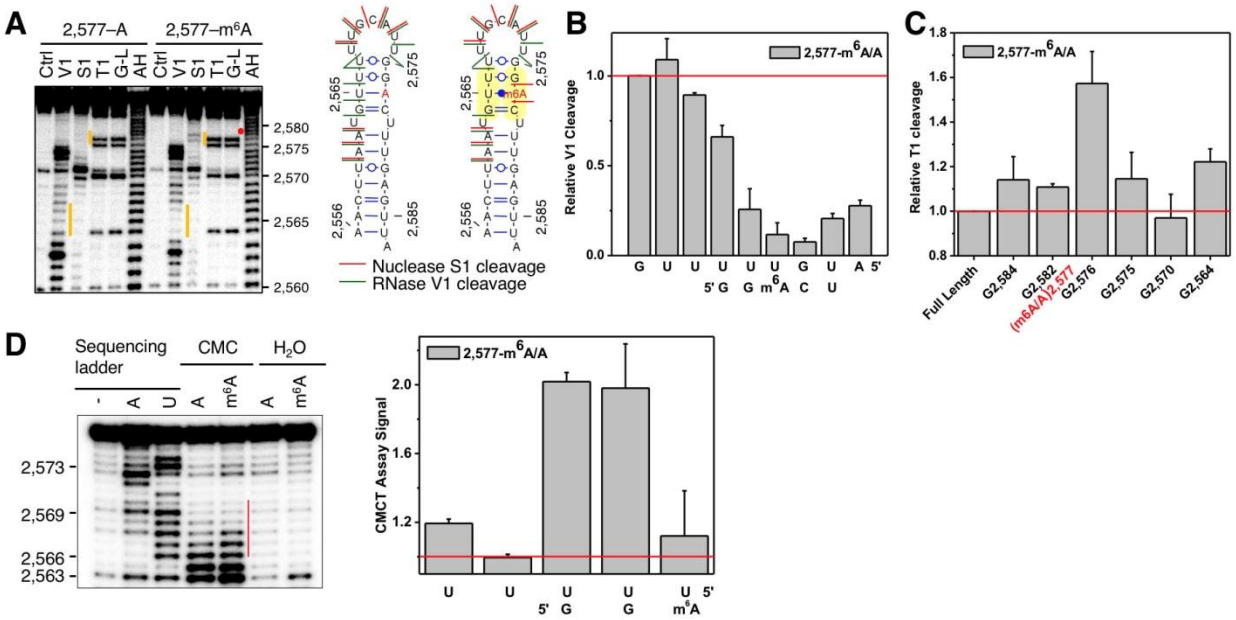


Figure 3.2. m^{6}A destabilizes the hairpin structure to expose the U-tract accessibility. A) RNA structural probing showing m^{6}A disrupts local RNA structure highlighted in yellow. Ctrl, no nuclease added; V1, RNase V1 digestion; S1, nuclease S1 digestion; T1, RNase T1 digestion; G-L, G-ladder; AH, alkaline hydrolysis. The orange bars mark the structurally altered RNA regions in the presence of m^{6}A (red dot). **B**) Quantification of the RNase V1 cleavage signal for the U-tract region from RNA structural mapping assay in **A**. n = 3, ± s.d., technical replicates. **C**) Quantitative of the RNase T1 cleavage signal from RNA structural mapping assay in **A**). Increased RNase T1 cleavage signal (single-stranded specific & cleavage after guanosines) was observed due to the surrounding m^{6}A residue. To correct for sample loading difference, the ratio for each band signal among all bands in each lane was calculated. The y-axis value Relative T1 cleavage = $(m^6A_{native}/m^6A_{denature})/(A_{native}/A_{denature})$. n = 2, technical replicates. **D**) Quantitative CMCT mapping showing increased signals for the U-tract bases around the U base-pairing with m^{6}A. Quantitation of band signals within the U-tract region is shown on the right. n = 4, ± s.d., technical replicates.

3.2.2 m^{6}A alters RNA structure to increase the accessibility of U-tracts

Next, we investigated the molecular mechanism for the m^{6}A-facilitated RNA-HNRNPC interactions. It has been reported multiple times that HNRNPC preferably binds single-stranded uridine tracts (U-tracts) composed of at least 5 continuous uridine residues[16,19,20]. Interestingly, a U_5-tract is half buried in the MALAT1 hairpin-stem, opposing the A/m^{6}A2,577 site. Considering that m^{6}A residues (tandem and separated) within RNA stems could destabilize the thermo-stability of RNA duplexes[21], we hypothesized that the m^{6}A2,577 residue destabilizes the

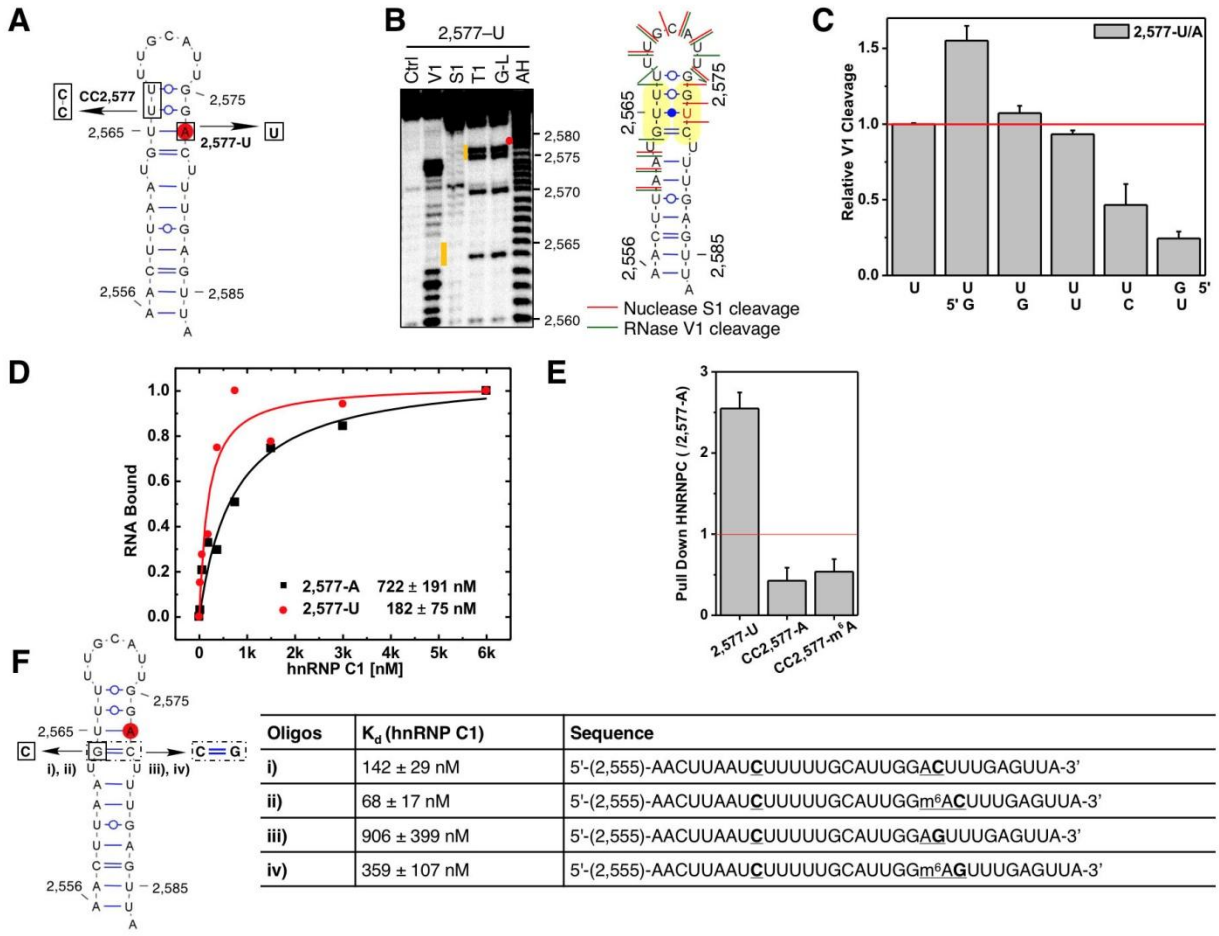
MALAT1 hairpin-stem to make its opposed U_5-tract more single-stranded/accessible, thus enhancing its interaction with HNRNPC.

To determine the effect of the m^6A2,577 residue on modulating the hairpin structure, we performed RNA structural probing assays using nuclease S1 (single-stranded RNA substrate specific) and RNase V1 (double-stranded/stacking RNA substrate specific) digestion (**Figure 3.2A**). Both the methylated and unmethylated RNA oligos formed the hairpin structure. However, obvious structural differences were observed in the U-tract and the m^6A consensus motif GGACU. Compared to the unmethylated hairpin, nuclease S1 digestion in the GAC (A= m^6A) motif significantly increased, whereas RNase V1 digestion markedly decreased at the U-tract opposing the GAC motif. Clearly, the m^6A residue destabilizes the stacking properties of the region centered around the U-residue that pairs with A/m^6A2,577 (**Figure 3.2B-C**).

We also performed quantitative CMCT assay[22] to monitor RNA structural changes at the U-tract in the presence of m^6A modification (**Figure 3.2D**). Consistent with S1/V1 mapping assay, the U-tract region showed increased reactivity with CMCT in the presence of m^6A, indicating that m^6A increases the U-tract's accessibility.

3.2.3 Increased accessibility of U-tract enhances HNRNPC binding

To confirm that enhanced HNRNPC binding is indeed due to the increased accessibility of the U-tract, we performed the following experiments. First, the A2,577-to-U mutation, which increased the accessibility of U-tract due to the mutation from 2,577A-U to U-U non-Watson-Crick base-pair (**Figures 3.3A-C**), enhanced HNRNPC1 binding by ~4-fold compared with the 2,577-A hairpin (**Figure 3.3D**). Besides, RNA pull down assays indicated that the A2,577-to-U mutation pulled down similar amount of HNRNPC as m^6A2,577 modification (**Figure 3.3E**).



Oligos	K_d (hnRNP C1)	Sequence
i)	142 ± 29 nM	5'-(2,555)-AACUUAAU**C**UUUUUGCAUUGGA**C**UUUGAGUUA-3'
ii)	68 ± 17 nM	5'-(2,555)-AACUUAAU**C**UUUUUGCAUUGGm^6A**C**UUUGAGUUA-3'
iii)	906 ± 399 nM	5'-(2,555)-AACUUAAU**C**UUUUUGCAUUGGA**G**UUUGAGUUA-3'
iv)	359 ± 107 nM	5'-(2,555)-AACUUAAU**C**UUUUUGCAUUGGm^6A**G**UUUGAGUUA-3'

Figure 3.3. Destabilized hairpin oligos increase HNRNPC binding. A) Illustration of mutated RNA oligos. **B**) Structure probing of the 2,577A-to-U mutated MALAT1 hairpin (2,577-U), same annotation as in Figure 3.2**A**). **C**) Quantification of the RNase V1 cleavage signal for the U-tract region from RNA structural mapping assays as in **B**). To correct for sample loading difference, each band signal was normalized to the band signal of the 3' most U of the U-tract. n = 2, technical replicates. **D**) Filter-binding curves displaying the binding affinities between recombinant HNRNPC1 and 2,577-U/A oligos. n = 3, ± s.d., technical replicates. **E**) RNA pull down with mutated oligos. n = 3, ± s.d., technical replicates. **F**) Filter-binding results showing the binding affinities between recombinant HNRNPC1 and four mutated MALAT1 oligos. (i) Mutate G-C to C-C, A2,577: predicted to weaken the hairpin stem and increase HNRNPC binding. Results: binding improved from 722 nM K_d to 142 nM (5-fold); (ii) Mutate G-C to C-C, m^6A2,577: in this context of weaker stem, m^6A is predicted to confer a smaller effect compared to wild-type hairpin. Result: improved binding only 2-fold instead of 8-fold; (iii) Restore C-C to C-G, A2,577: predicted to restore the hairpin stem and decrease HNRNPC binding compared to C-C mutant. Result: binding decreased by 6.4-fold; (iv) Restore C-C to C-G, m^6A2,577: in this context of restored stem, m^6A is again predicted to confer increased binding compared to A2,577 hairpin. Result: improved binding by 2.5-fold. n = 3 each, ± s.d., technical replicates.

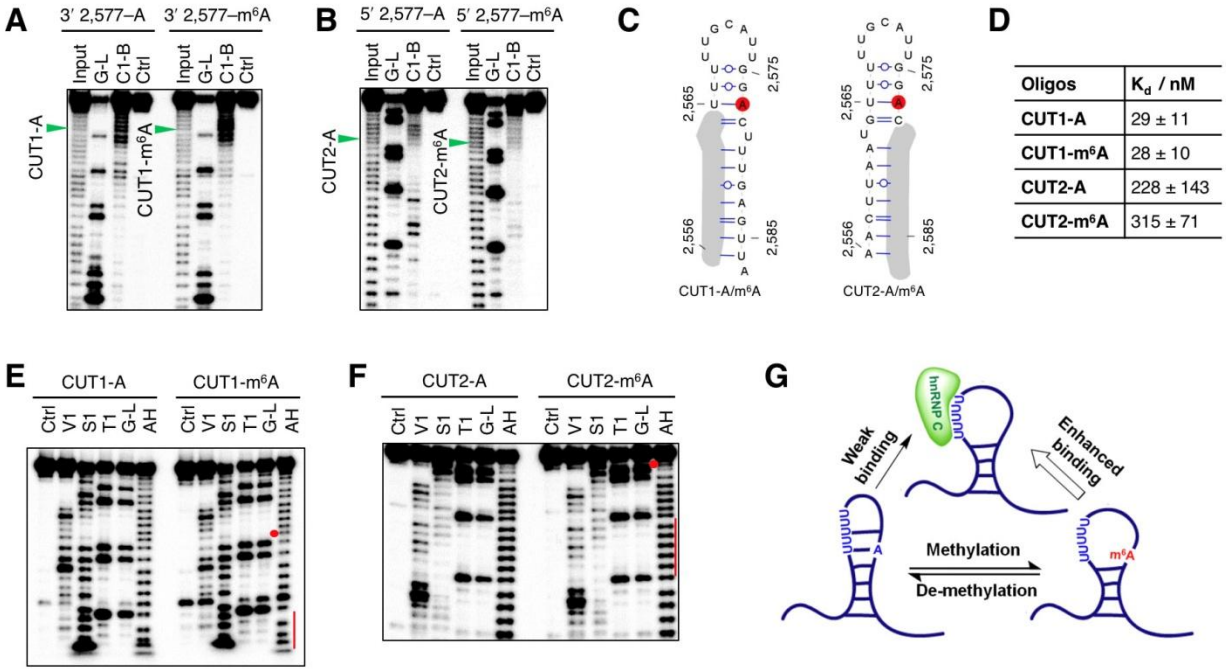


Figure 3.4. The increased accessibility of U-tracts increased HNRNPC binding. A) RNA alkaline hydrolysis terminal truncation assay showing recombinant HNRNPC1 binding to terminal truncated MALAT1 hairpin oligos (2,577 site m^{6}A methylated or unmethylated). In this assay, 3′ radiolabeled MALAT1 2,577 hairpin oligos were terminal truncated by alkaline hydrolysis into RNA fragments which were then incubated with HNRNPC1 protein followed by filter binding wash steps. The remaining RNA on the filter paper was isolated and analyzed by denaturing gel electrophoresis, as indicated in the lane "C1-bound or C1-B". "Input" refers to alkaline hydrolysis truncated RNA oligos used for incubation with HNRNPC1; "G-L or G-ladder" was generated from RNase T1 digestion; "Ctrl" refers to the intact MALAT1 hairpin without alkaline hydrolysis truncation. One pair of methylated/unmethylated truncated oligos (CUT1, marked by green arrows) was selected for subsequent biochemical analysis, due to their strong interaction with HNRNPC1. **B**) RNA terminal truncation assay as in **A** except 5′ ^{32}P-labeled oligos were used. One pair of methylated/unmethylated truncated oligos (CUT2, marked by green arrows) was selected for subsequent biochemical analysis. **C**) Sequence of CUT1 and CUT2 oligos. **D**) Truncated oligos with exposed U-tracts increased HNRNPC binding regardless of m^{6}A. n = 3, ± s.d., technical replicates. **E**) Structure probing of the CUT1 oligos using RNase V1 and nuclease S1 digestion, same annotation as in **Figure 3.2A**. The red dot marks the m^{6}A site and the red line marks the U-tract region. **F**) Structure probing of the CUT2 oligos using RNase V1 and nuclease S1 digestion, same annotation as in **E**. **G**) Illustration of the m^{6}A-switch model.

These results indicate that increased accessibility of U-tract by U2,577 mutation was sufficient to enhance its interaction with HNRNPC. Second, mutation of the U-tract, by replacing the two G-U pairs above the A/m^{6}A-U with G-C pairs, significantly reduced the RNA pull down

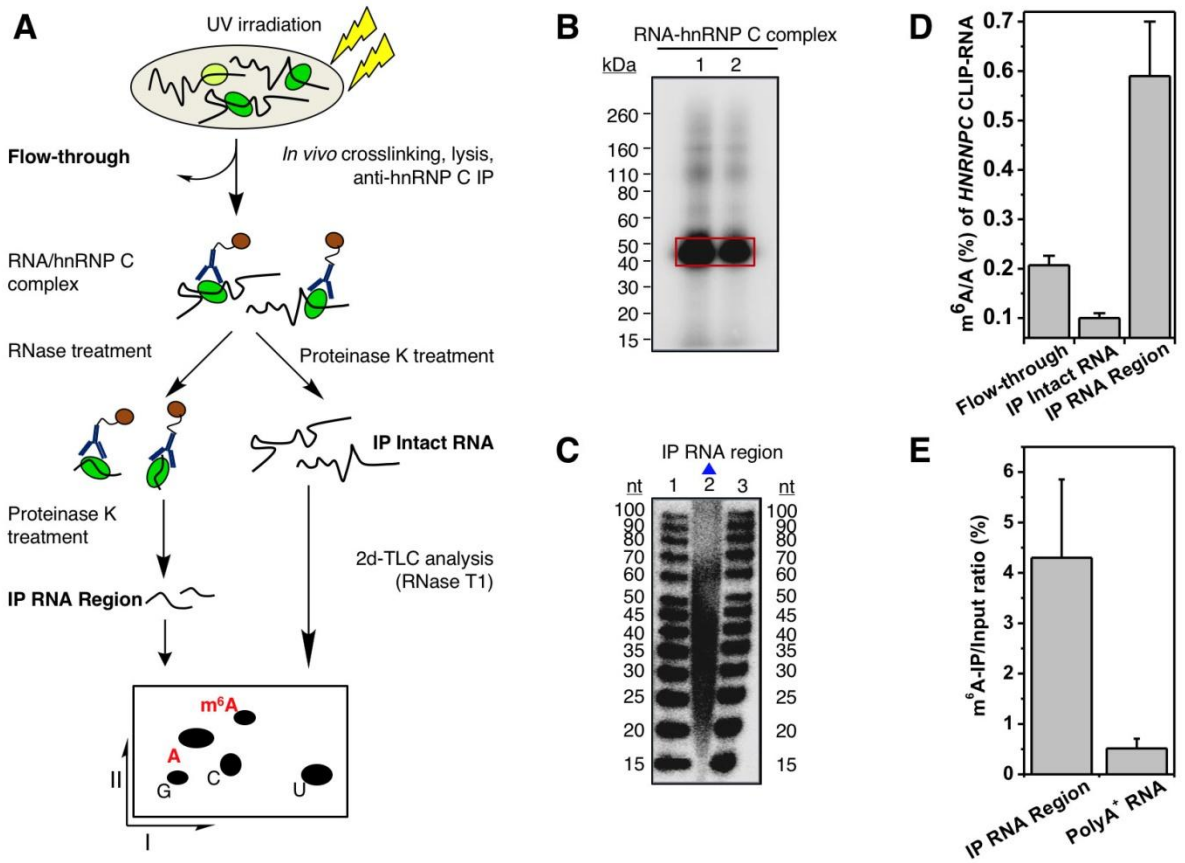


Figure 3.5. m^{6}A is enriched in the vicinity of HNRNPC binding sites. A) Schematic diagram of the CLIP-2dTLC protocol. IP, immunoprecipitation; nt, nucleotide. The RNase T1 used in our 2dTLC assay cleaves single-stranded RNA after guanosines, so the m^{6}A/A ratio determined here represents the m^{6}A fraction of all adenosines following guanosines. **B)** Analysis of crosslinked RNA-HNRNPC complexes (CLIP RNP) using denaturing gel electrophoresis (lanes 1 and 2). Positions of the protein size standards are shown on the left. HNRNPC IP RNA region (RNA samples within RNA-HNRNPC crosslinked complexes) were extracted from the gel slices marked by the red rectangle. **C)** Denaturing gel analyzing the size distribution for the HNRNPC PAR-CLIP RNA samples (lane 2). The RNA size standards were loaded in lanes 1 and 3. **D**) CLIP-2dTLC showing the m^{6}A enrichment in HNRNPC bound RNA regions. n = 3, ± s.d., biological replicates. **E)** HNRNPC bound RNA regions had higher anti-m^{6}A pull down yield than polyA$^+$ RNA. n = 3, ± s.d., biological replicates.

amount of HNRNPC protein regardless of m^{6}A2,577 modification (**Figure 3.3E**), validating that HNRNPC binding occurs at the U-tract. Moreover, binding studies with 4 other mutated MALAT1-A/m^{6}A2,577 oligos also validated our hypothesis that the increased accessibility of U-

tract, rather than the m^6A residue itself, is primarily responsible for the enhanced interaction with HNRNPC (**Figure 3.3F**).

Third, we performed alkaline hydrolysis terminal truncation[23] to identify interactions between truncated 2,577-A/m^6A hairpin fragments and recombinant HNRNPC1. We identified two pairs of truncated hairpins with highly accessible U-tracts (CUT1-A/m^6A and CUT2-A/m^6A, **Figures 3.4A-C**) that conferred efficient interactions with HNRNPC1. Compared to the original full-length 32-mer 2,577-A hairpin, the 5′ truncated 23-mers (CUT1-A/m^6A) and the 3′ truncated 23-mers (CUT2-A/m^6A) improved HNRNPC binding by ~25-fold and ~3-fold, respectively (**Figure 3.4D**). Most importantly, the m^6A modification did not increase the accessibility of U-tract in either truncated pair (**Figures 3.4E-F**), and thus exhibited little effect on HNRNPC1 binding (**Figure 3.4D**).

Based on these results, we propose that m^6A modification can alter its local RNA structure and enhance the accessibility of base-paired residues or nearby regions to modulate protein binding (**Figure 3.4G**). We use the term m^6A-switch to describe this new mechanism regulating RNA-protein interactions through m^6A-dependent RNA structural switches.

3.2.4 PAR-CLIP–MeRIP identifies m^6A-modified HNRNPC binding sites across the transcriptome

To determine whether m^6A facilitates RNA-HNRNPC interactions generally *in vivo*, we performed Cross-Linking followed by Immuno-Precipitation and 2-Dimensional Thin-Layer Chromatography (CLIP-2dTLC, **Figure 3.5A-C**) and detected the m^6A/A ratio of HNRNPC target RNAs from HEK293T cells. We found the HNRNPC bound RNA region has ~5.9-fold

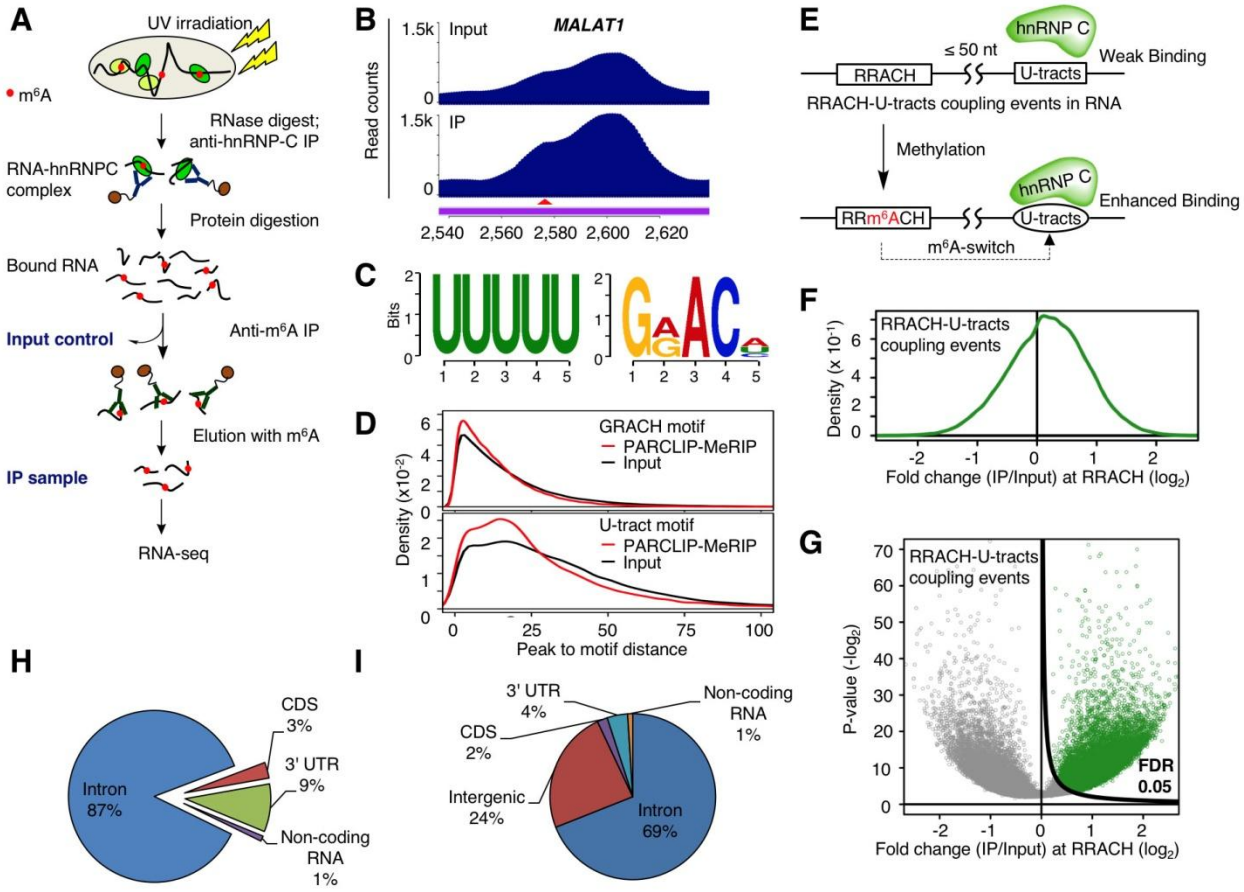


Figure 3.6. PAR-CLIP–MeRIP identifies m^{6}A-switches transcriptome-wide. A) Illustration of the PAR-CLIP–MeRIP protocol. IP, immunoprecipitation. **B**) PAR-CLIP–MeRIP identified m^{6}A residue around the MALAT1 2,577 site. **C**) Binding motifs identified by FIRE with PAR-CLIP–MeRIP peaks. **D**) Density plots illustrating the distribution of distance between the PAR-CLIP–MeRIP/Input peaks and the nearest GRACH motif (top) or the nearest U-tracts in (bottom). E) Definition and identification of HNRNPC m^{6}A-switches based on the PAR-CLIP–MeRIP analysis. Approximately 89% of PAR-CLIP–MeRIP peaks harboring both the U-tract and RRACH motif have "RRACH-U-tract" inter-motif distance within 50 nucleotides, significantly higher than the 64% of such coupling within the genomes. HNRNPC m^{6}A-switches are identified as m^{6}A methylated RRACH-U-tracts coupling events. **F**) Density plot showing the positive enrichment at RRACH sites. **G**) Volcano plot depicting all coupling events (unfilled circles) as defined in **b**, according to their p-values[28] (P, y-axis) and fold change values at RRACH sites (E, x-axis). To identify HNRNPC m^{6}A-switches, we generated the π-value, $\pi = E \cdot (-\log_{10} P)$, as one comprehensive parameter to pick meaningful genomic loci[29]. HNRNPC m^{6}A-switches identified from PAR-CLIP–MeRIP experiments should fulfill the following requirements: (i) read counts at both the control and IP sample ≥ 5; (ii) π-value ≥ 0.627, corresponding to False Discovery Rate (FDR) ≤ 5%. **H**) Pie chart depicting the region distribution of HNRNPC m^{6}A-switches identified by PAR-CLIP–MeRIP. **I)** Pie chart depicting HNRNPC PAR-CLIP peaks. These are enriched in introns, consistent with previous reports that HNRNPC binds mainly nascent transcripts[19,23,25].

higher m^6A level than the HNRNPC bound intact RNA, and ~3-fold higher m^6A level than the flow through RNA (**Figure 3.5D**). Besides, the HNRNPC-bound RNA regions had much higher anti- m^6A pull down yield (4.3%) than the polyA+ RNA samples (**Figure 3.5E**). This result suggests the enrichment of m^6A residues in the immediate vicinity of HNRNPC binding sites as well as a widespread presence of m^6A-facilitated RNA-protein interactions *in vivo*.

Our next goal was to identify transcriptome-wide m^6A-modified RNA binding sites of HNRNPC with high resolution, in accordance with our CLIP-2dTLC result showing that HNRNPC binding regions are m^6A-enriched. To this end, we developed a method that combined Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP, [24]) and m^6A-antibody immunoprecipitation (MeRIP, [25,26]), termed PAR-CLIP–MeRIP (**Figure 3.6A**). HEK293T cells were treated following normal PAR-CLIP procedures, and HNRNPC bound RNA fragments (Input control) were immunoprecipitated using the previously established anti-m^6A antibody to enrich m^6A-containing RNA fragments (IP). Libraries were prepared from both the Input control and IP samples, and subjected to next-generation sequencing. The RNA-seq data were mapped to human genome version hg19, and then analyzed by PARalyzer [27] with default settings. Through comparing the normalized read counts between the IP sample and Input control (Experimental sections), we could identify enriched PAR-CLIP–MeRIP peaks within HNRNPC binding sites, such as the MALAT1-2,577 peak (**Figure 3.6B**).

To determine whether HNRNPC PAR-CLIP–MeRIP peaks shared common sequence elements, we performed an unbiased search for enriched motifs under PAR-CLIP–MeRIP peaks (IP/Input ≥ 2). Remarkably, the top two consensus motifs are the previously established m^6A consensus motif GRACH (a subset of RRACH) and the HNRNPC RNA binding motif (U-tracts) (**Figure 3.6C**). In addition, the median distances between PAR-CLIP–MeRIP peaks and both

identified motifs were shorter than those of the Input control peaks (HNRNPC PAR-CLIP peaks, **Figure 3.6D**), indicating that both motifs became enriched during the Me-RIP process. These findings reinforce the authenticity of our HNRNPC PAR-CLIP–MeRIP peaks.

3.2.5 Transcriptome-wide identification of m^{6}A-switches for HNRNPC binding

Our findings that both the m^{6}A consensus motif and U-tracts motif are mostly within 50 nucleotides (nts) apart from PAR-CLIP–MeRIP peaks (**Figure 3.6D**) suggest a transcriptome-wide association between m^{6}A residues and U-tracts within the HNRNPC binding sites. Indeed, ~89% of PAR-CLIP–MeRIP peaks harboring both the U-tract and RRACH motif have "RRACH-U-tract" inter-motif distance within 50 nts, significantly higher than 64% of genomic distributions. We termed such transcriptome-wide associations as "RRACH-U-tracts" coupling events (**Figure 3.6E**). To identify such "RRACH-U-tracts" coupling events, we searched for all U-tracts within HNRNPC PAR-CLIP sequences whose distance from the nearest RRACH motif is less than 50 nts. Among 125,844 RRACH-U-tracts coupling events identified in total, 78,232 events (62%) showed positive enrichment (IP/Input > 1) at the RRACH motif (**Figures 3.6F-G**), significantly higher than the enrichment fractions at RRACH sites from the m^{6}A/MeRIP-seq experiments in previous studies [25,26]. This result indicates that the majority of RRACH-U-tracts coupling events were enriched by the PAR-CLIP–MeRIP process from HEK293T cells. Using a site-detection algorithm with an estimated FDR ≤ 5% [30], we identified 39,060 m^{6}A-modified RRACH-U-tracts coupling events. We consider the m^{6}A residues within RRACH-U-tracts coupling events serving as m^{6}A-switches to increase the access of U-tracts and increase RNA-HNRNPC interactions. Therefore, our PAR-CLIP–MeRIP approach identified 39,060 HNRNPC m^{6}A-switches, accounting for ~7% of 592,477 HNRNPC binding sites identified by PAR-CLIP in HEK293T cells.

We examined the distribution of HNRNPC m^{6}A-switches along transcripts relative to functional regions in mRNA. The majority (87%) of m^{6}A-switches occur within introns (**Figure 3.6H**). This distribution pattern is consistent with HNRNPC binding patterns identified by our PAR-CLIP approach (**Figure 3.6I**) and previous reports that HNRNPC mainly binds nascent transcripts and shows widespread binding across introns[15,19].

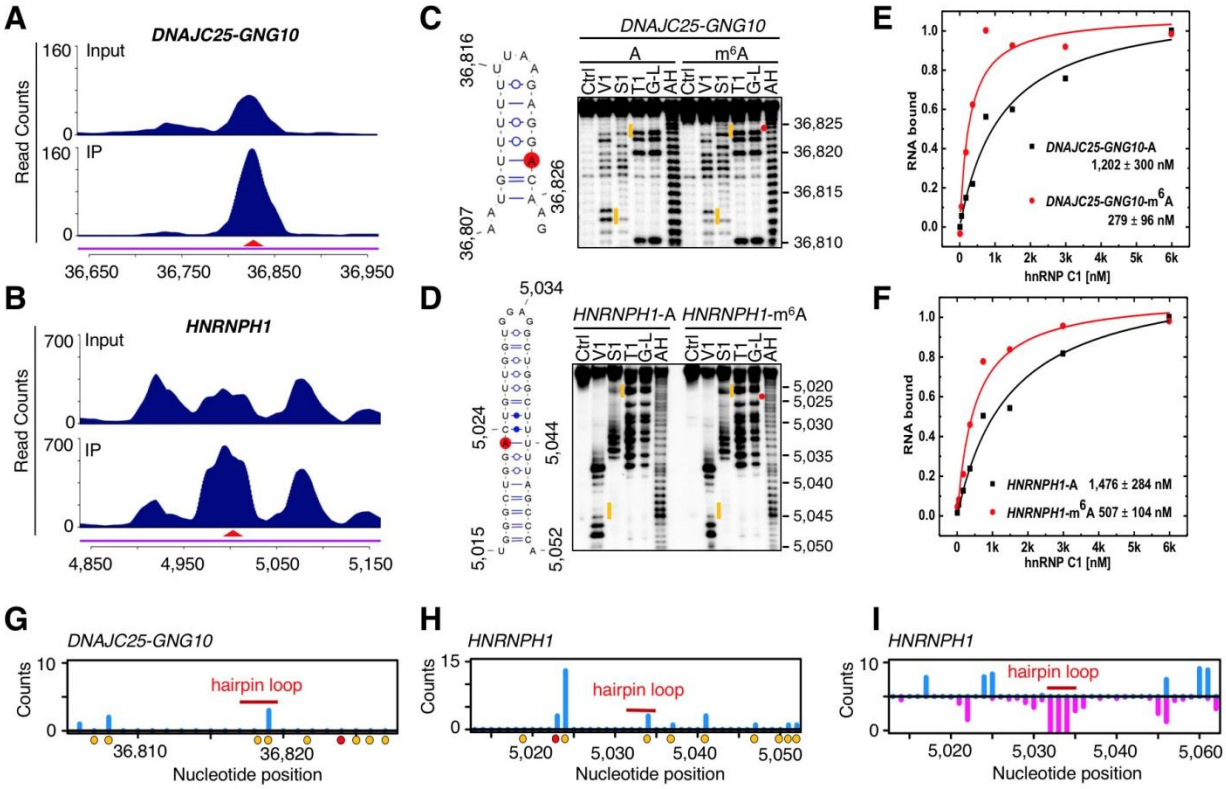


Figure 3.7. Validation of two identified m^{6}A-switches. A-B) PAR-CLIP–MeRIP data detected positive IP/Input enrichment at the RRACH sites (red arrowheads) on the *DNAJC25-GNG10* gene (**A**) and *HNRNPH1* gene (**B**) in HEK293T cells. **C-F)** Validation of two m^{6}A-switches by S1/V1 structural probing and filter binding. n = 4, ± s.d., technical replicates. Same annotation as in Figure 3.2**A**. **G-H**) *In vivo* DMS mapping of the *DNAJC25-GNG10* hairpin (**G**) and *HNRNPH1* (**H**); data are from[31]. A and C residues are marked with orange dots and the m^{6}A residue is marked with a red dot. The hairpin loops are indicated by red bars. **I,** Transcriptome-wide S1/V1 mapping around the *HNRNPH1* m^{6}A-switch site. Blue bars represent V1 signal; magenta bars represent S1 signal. The hairpin loop is indicated by a red bar; data are from[32]. Not enough reads could be collected to make a plot for the *DNAJC25-GNG10* m^{6}A-switch region.

Next, we selected two m^{6}A-switches in introns to examine their effects on RNA-HNRNPC interactions (**Figures 3.7A-B**). Both m^{6}A-switches form hairpin structures with the RRACH motif and U_{5-7}-tracts located in the hairpin stem base-paired with each other. S1/V1 RNA probing assays identified similar RNA structural patterns as the MALAT1 hairpin, with increased nuclease S1 cleavage around the m^{6}A residue and decreased RNase V1 cleavage in the U-tracts paired with m^{6}A (**Figures 3.7C-D**). Binding assays showed that these two m^{6}A-switches (in the *DNAJC25-GNG10* and *HNRNPH1* transcripts) increased their interactions with recombinant HNRNPC1 by ~4 and ~3-fold, respectively (**Figures 3.7E-F**). These results validated the effect of intronic m^{6}A-switches on regulating RNA-HNRNPC interactions.

3.2.6 *METTL3/L14* knockdown decreases HNRNPC binding activities

Our results so far suggest that numerous cellular RNA-HNRNPC interactions are regulated by m^{6}A-RNA methylation. To access the effect of reducing global m^{6}A level on RNA-HNRNPC interactions, we performed HNRNPC PAR-CLIP experiments in *METTL3* and *METTL14* knockdown (KD) HEK293T cells. METTL3 and METTL14 are the only identified m^{6}A-methyltransferases so far, and their knockdowns have been reported to reduce global m^{6}A-RNA methylation levels by ~30-40% [33,34].

METTL3 and *METTL14* KD did not change the cellular abundance of HNRNPC (**Figure 3.8A**), but had a dramatic impact on HNRNPC binding transcriptome-wide. We focused on HNRNPC binding among the previously defined RRACH-U-tracts coupling events, and found the majority (69%) of 52,107 coupling events showed decreased fold change (KD/Ctrl) on their U-tracts sites (**Figure 3.8B-C**), indicating decreased HNRNPC binding at U-tracts of most coupling events. In total, we identified 16,582 coupling events with decreased U-tracts-

HNRNPC interactions upon *METTL3/L14* KD, using a site-detection algorithm with an FDR ≤5% [30] (**Figure 3.8C**). Significant overlaps were observed between *METTL3* and *METTL14* KD samples (**Figure 3.8D**), consistent with previous reports that these two m^{6}A-methyltransferases work together as a cellular complex and have numerous common target sites throughout the transcriptome [33,34].

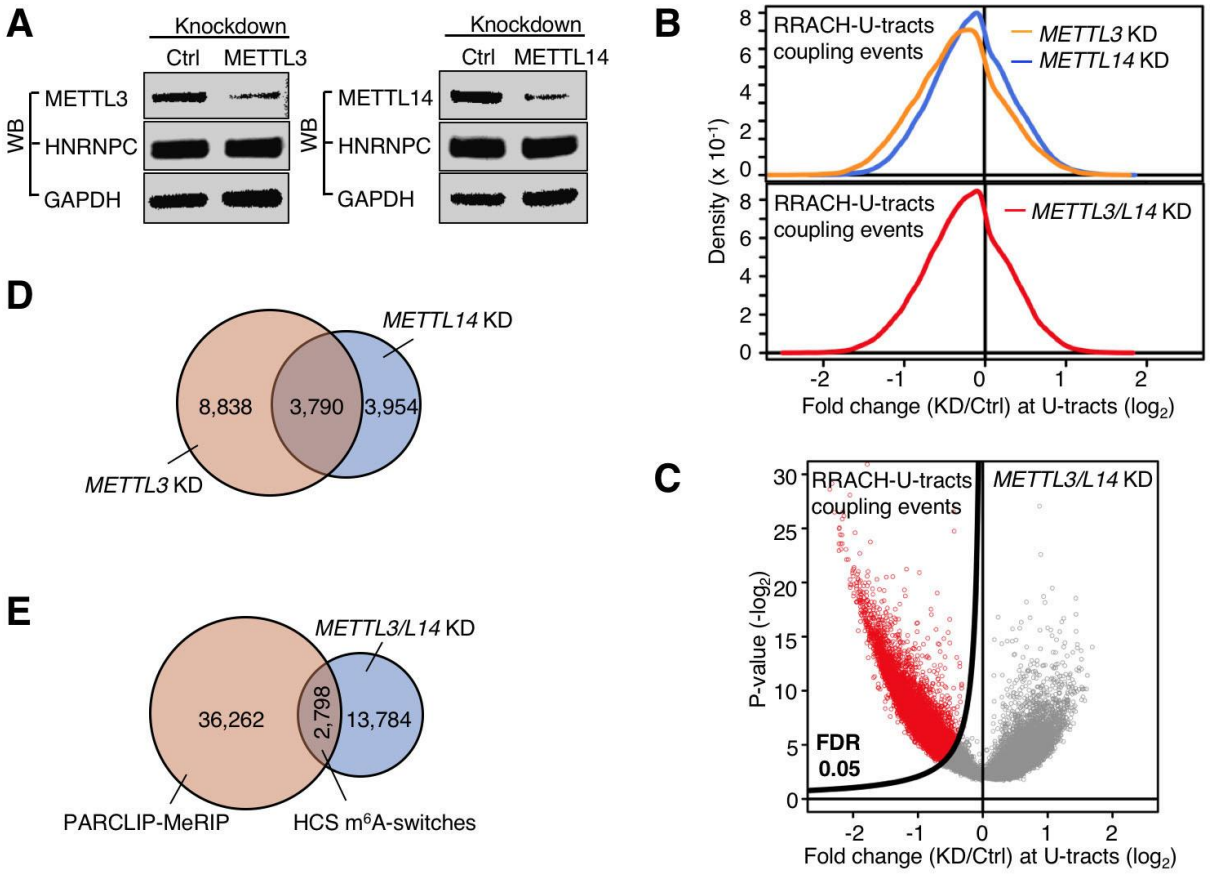


Figure 3.8. Global m^{6}A reduction decreases HNRNPC binding at m^{6}A-switches. A) Western blot showing stable HNRNPC protein abundance upon *METTL3/L14* KD. **B**) Density plot showing negative enrichment at the U-tracts. **C**) Volcano plot of the *METTL3/L14* KD data depicting RRACH-U-tracts coupling events (unfilled red circles) as defined in Extended Data Fig. 4b, according to their p-values[28] (P, y-axis) and fold change values at the U-tracts (E, x-axis). **D**) Overlap of RRACH-U-tract coupling events with decreased HNRNPC binding by *METTL3* and *METTL14* KD. **E**) Identification of HCS m^{6}A-switches.

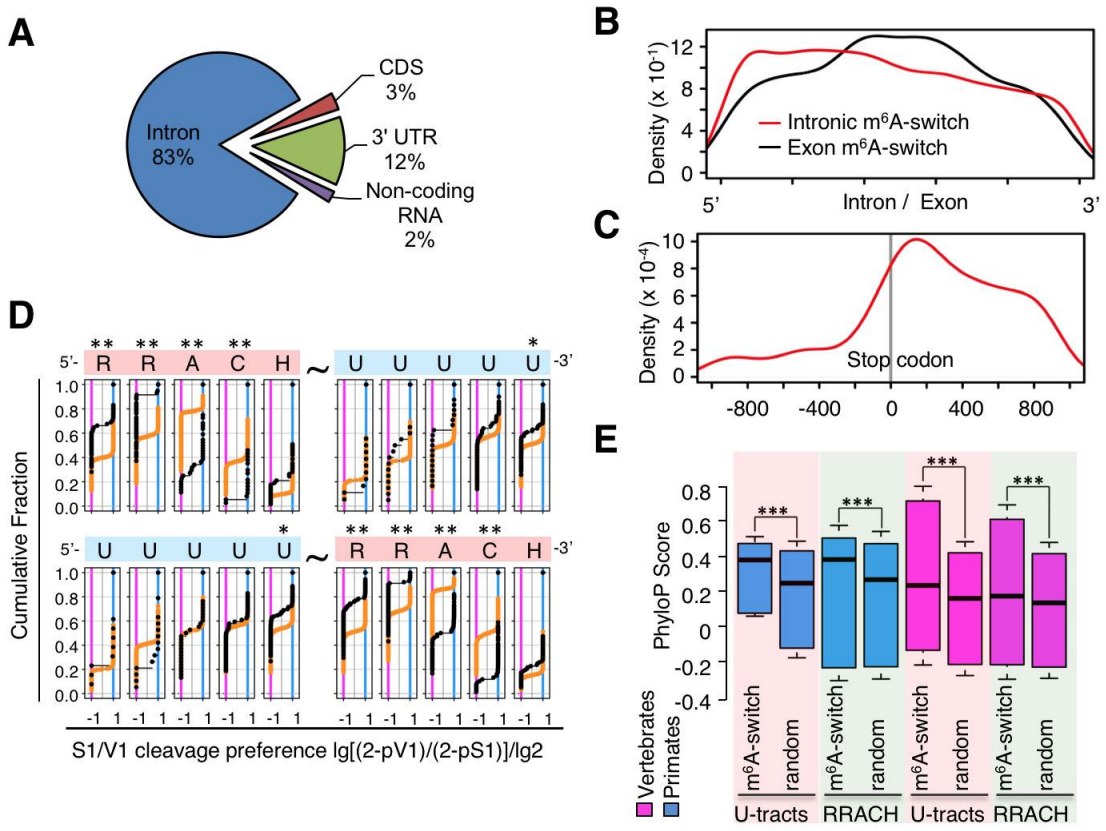
It is noteworthy that either *METTL3* or *METTL14* KD cannot reduce the m^{6}A-RNA methylation level at every m^{6}A sites [33,34]. Therefore, the m^{6}A-switch sites showing decreased HNRNPC binding represented only a subset of the total cellular HNRNPC m^{6}A-switches. In total, 2,798 m^{6}A-switches identified by PAR-CLIP–MeRIP experiments showed decreased HNRNPC binding upon *METTL3/L14* KD (**Figure 3.8E**). These sites composed the High Confidence Set (HCS) of m^{6}A-switches and were used for subsequent analysis.

3.2.7 Molecular features of high-confidence m^{6}A-switches

HCS m^{6}A-switches are enriched in introns (**Figure 3.9A**), similar to the distribution pattern of the 39,060 HNRNPC m^{6}A-switches identified by the PAR-CLIP–MeRIP approach (**Figure 3.6H**). HCS m^{6}A-switches showed slight enrichment at the 5′ end of introns (**Figure 3.9B**). HCS m^{6}A-switches within exons were enriched near the stop codon and in the 3′UTR (**Figure 3.9C**), consistent with previous results on the topology of human m^{6}A methylomes in mRNAs [25,26].

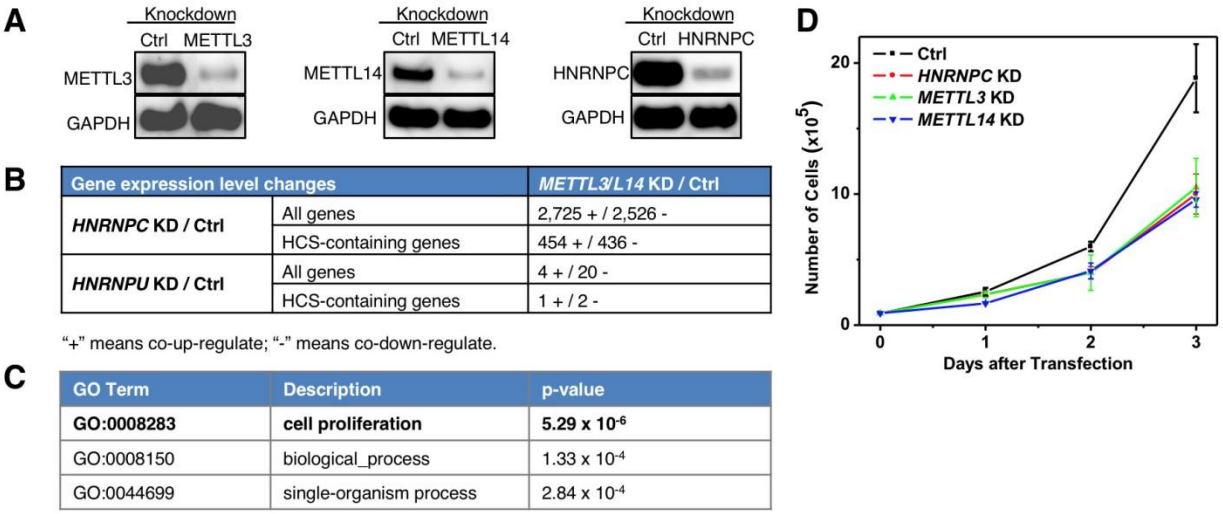
We analyzed the currently available transcriptome-wide RNA structural mapping data [6-9] to look for structural patterns among the 2,798 HCS m^{6}A-switches. We found that HCS m^{6}A-switches display a distinct structural pattern at both the RRACH motif and U-tract motif compared to a control group from the top 1,000 highly expressed genes (**Figure 3.9D**). For HCS m^{6}A-switches, “AC” in the “RRACH” motif is more structured while “RR” is less structured. In the U-tract motif, U residues become increasingly single-stranded from 5’ to 3’ direction. This pattern, while counter the classical expectation from RNA structural perspective that “RR” are expected to base pair more frequently than “AC”, is consistent with our m^{6}A-switch model that the purine residues within the RRACH motif and 3’ U residues in the U-tracts motif increase

structural melting/dynamics, as validated in our three demonstrated m^{6}A-switch hairpins (**Figures 3.7G-I**).



EVO Fraction (%)	Genome	HCS m^{6}A-switch	Fold Change
Intron	0.030	0.34	11.60
Exon	0.46	2.11	4.58
Total	0.06	0.64	10.65

Figure 3.9. Molecular features of HCS m^{6}A-switches. A) Region distribution of HCS m^{6}A-switches. **B**) Density plot showing m^{6}A-switches distribution relative to exon/intron boundaries. **C**) m^{6}A-switches in coding RNA were enriched in the 3'UTR and near the stop codon. **D**) Cumulative distribution of HCS m^{6}A-switches (black) and control (orange) regarding the S1/V1 cleavage preference (data from[32]) at U-tracts and RRACH motif. U-tract can be 3' (upper) or 5' (lower) of the RRACH motif. *: $p < 0.05$, **: $p < 10^{-4}$, Kolmogorov-Smirnov test. **E**) Phylogenetic conservation of HCS m^{6}A-switches among primates and vertebrates. ***: $p < 10^{-16}$, Mann-Whitney-Wilcoxon test. **F**) EVOfold analysis for the 2,798 HCS m^{6}A-switches. The chances for HCS m^{6}A-switches to have EVOfold records are significantly higher than random genomic sequences. We first calculated the number of HCS sites in the EVO database if occurring in random to be ~1.7 HCS sites. We found 18 HCS sites are present in EVO database, resulting in ~11x enrichment. This result is further divided into intronic and exonic regions.



B

Gene expression level changes		*METTL3/L14* KD / Ctrl
***HNRNPC* KD / Ctrl**	All genes	2,725 + / 2,526 -
	HCS-containing genes	454 + / 436 -
***HNRNPU* KD / Ctrl**	All genes	4 + / 20 -
	HCS-containing genes	1 + / 2 -

"+" means co-up-regulate; "-" means co-down-regulate.

C

GO Term	Description	p-value
GO:0008283	**cell proliferation**	**5.29 x 10^{-6}**
GO:0008150	biological_process	1.33 x 10^{-4}
GO:0044699	single-organism process	2.84 x 10^{-4}

Figure 3.10. m^{6}A-switches regulate the cell proliferation rates. A) *HNRNPC*, *METTL3/L14* KD confirmed by Western blots. **B**) *HNRNPC* KD and *METTL3*/*L14* KD co-regulated the expression of a large number of genes. Gene expression changes between Ctrl and *HNRNPC*, *HNRNPU*, *METTL3*/*L14* KD HEK293T cells were analyzed by Cuffdiff2 [36,37], and the absolute numbers of differential expressed genes are shown. HCS-containing genes refer to the 1815 genes containing high confident m^{6}A-switches. The RNA-seq data from *HNRNPU* KD HEK293T cells (GEO34995 dataset[38]) was analyzed for comparison with a different mRNA binding protein. HnRNPU did not show preferential interaction with the 2,577-m^{6}A modified MALAT1 hairpin. **C**) Gene ontology analysis of the m^{6}A-switch-containing genes whose expression level were co-differentially-regulated by *HNRNPC* and *METTL3/L14* KD, against all m^{6}A-switch-containing genes as background. **D**) *HNRNPC*, *METTL3/L14* KD decreased the proliferation rates of HEK293T cells to a similar extent. n = 4, ± s.d., biological replicates.

We also investigated whether HNRNPC m^{6}A-switches are conserved across species. The PhyloP scores of HCS m^{6}A-switch sites across vertebrates and primates are much higher than those of random genomic regions (**Figure 3.9E**). Besides, we found that HCS m^{6}A-switches are over 10 times more likely to have EVOfold records than random genomic sequences (**Figure 3.9F**). This analysis indicates that some HCS m^{6}A-switches evolved from common ancestors of human and rodents, since EVOfold elements are identified using sequence co-variations for base pairing across mammalian lineages [35].

3.2.8 HNRNPC m^{6}A-Switches affect cell proliferation by regulating gene expression

To reveal the function of m^{6}A-switches on RNA biology, we performed polyA$^+$ RNA-seq from *HNRNPC*, *METTL3*, *METTL14* KD and control cells (**Figure 3.10A**). *METTL3*/*L14* KD, which has been shown to decrease HNRNPC binding transcriptome-wide, co-regulated the expression of 5,251 genes with *HNRNPC* KD. In comparison, *METTL3*/*L14* KD co-regulated only 24 genes with KD of another mRNA binding protein hnRNP U (**Figure 3.10B**), which was not enriched in our m^{6}A-hairpin pull down (**Figure 3.1C**). Approximately 45% of 1,815 HCS m^{6}A-switch-containing genes were co-regulated by *HNRNPC*, *METTL3*/*L14* KD, indicating that m^{6}A-switch-regulated HNRNPC binding affects the abundance of target mRNAs. Gene Ontology (GO) analysis suggests that m^{6}A-switch-regulated gene expression may influence "cell proliferation" and other biological processes (**Figure 3.10C**). We found that *HNRNPC*, *METTL3* and *METTL14* KD decreased cell proliferation rate to similar extents (**Figure 3.10D**).

The m^{6}A-switch-regulated expression of genes within these GO categories was validated by qPCR (**Figure 3.11A**). One example of HNRNPC m^{6}A-switch regulating expression levels of "cell proliferation" related genes can be seen for the *ARHGAP5* gene. *METTL3* and *METTL14* KD leads to a strong decrease in HNRNPC binding around one *ARHGAP5* m^{6}A-switch site (**Figure 3.11B-D**), accompanied by the significantly increased *ARHGAP5* expression level similar to *HNRNPC* KD (**Figure 3.11E**). These results validate the role of HNRNPC m^{6}A-switches in regulating the expression of "cell proliferation" related genes.

3.2.9 HNRNPC m^{6}A-switches regulate alternative splicing

We also analyzed splicing pattern changes between control and *HNRNPC*, *METTL3*,

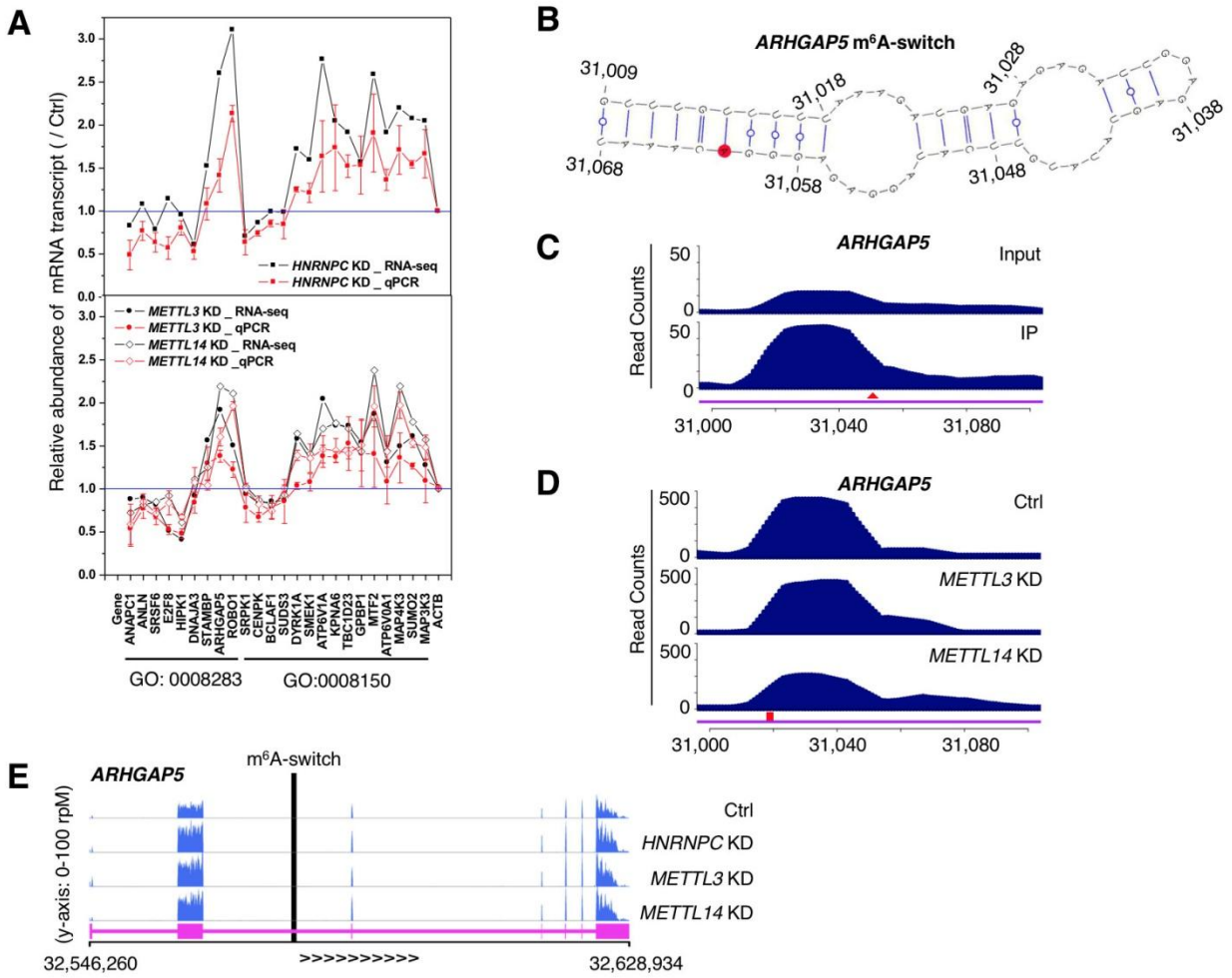


Figure 3.11. m^{6}A-switches regulate the abundance of target mRNAs. A) *HNRNPC*, *METTL3/L14* KD co-regulated the abundance of m^{6}A-switch-containing transcripts by RNA-seq and qPCR. **B**) An example of m^{6}A-switch among co-regulated transcript is the *ARHGAP5* transcript (NM_001030055). Its proposed secondary structure with m^{6}A methylation site in red is shown with the opposing the U-tract in a stem. **C-D**) PAR-CLIP–MeRIP detected positive IP/Input enrichment at the RRACH site (red arrowhead) of the *ARHGAP5* m^{6}A-switch (**C**), while *METTL3/L14* KD decreased HNRNPC binding at the U-tract (red square) of this m^{6}A-switch (**D**). **E**) The expression level of *ARHGAP5* gene was co-upregulated by *HNRNPC*, *METTL3/L14* KD, as shown by the RNA-seq data from HEK293T cells. The vertical black line represents the m^{6}A-switch site.

METTL14 KD samples using the DEXSeq software (**Figure 3.12A**; [39]. Significantly Differentially Expressed Exons (SDEE) between control and *HNRNPC* KD called by DEXSeq occurred more frequently in the immediate vicinity of m^{6}A-switches (**Figure 3.12B-C**), and

METTL3/L14 regulated these alternative splicing events in a similar manner as HNRNPC (**Figure 3.12A**), suggesting the role of HNRNPC m^{6}A-switches in alternative splicing regulation.

Within the HCS-containing genes, 109 exons and 159 exons were co-downregulated and co-upregulated by *HNRNPC* and *METTL3* KD, and 47 and 67 exons were co-downregulated and co-upregulated by *HNRNPC* and *METTL14* KD. An example of a m^{6}A-switch-regulated splicing event can be seen at the alternative exon of *CDS2* gene: *METTL3* and *METTL14* KD led to strong decrease in HNRNPC binding at one m^{6}A-switch site close to this exon (**Figures 3.12 D-F**), accompanied by significantly decreased exon inclusion level similar to *HNRNPC* KD (**Figure 3.12G**). These findings suggest that HNRNPC m^{6}A-switches could regulate RNA alternative splicing through manipulating HNRNPC binding activities.

3.2.10. Conclusion and discussion

Here, we show that the m^{6}A mRNA/lncRNA modification can weaken its local RNA structure to increase the accessibility of U-tracts for binding HNRNPC. We developed the PAR-CLIP–MeRIP approach to identify transcriptome-wide m^{6}A-modified RNA binding sites for individual RBPs. HNRNPC PAR-CLIP–MeRIP identified 39,060 m^{6}A-switches, and HNRNPC binding at 2,798 of these U-tracts could be decreased by m^{6}A-methyltransferases (*METTL3* or *METTL14*) KD. Thus, we identified 2,798 HNRNPC m^{6}A-switches with high confidence.

The majority of HNRNPC m^{6}A-switches are localized in introns, in agreement with the literature that HNRNPC is nucleus-localized and binds nascent transcripts. Consistent with previous reports that the majority binding sites of m^{6}A-methyltransferases are localized in the intronic and intergenic regions[33,34], our findings indicate that numerous intronic sites are m^{6}A methylated. However, previous m^{6}A/MeRIP-seq experiments did not detect substantial amount

of m^{6}A residues in introns[25,26,40]. This is likely due to the differences in the experimental design: previous m^{6}A/MeRIP-seq experiments used the steady state RNA (such as polyA$^+$ RNA) as

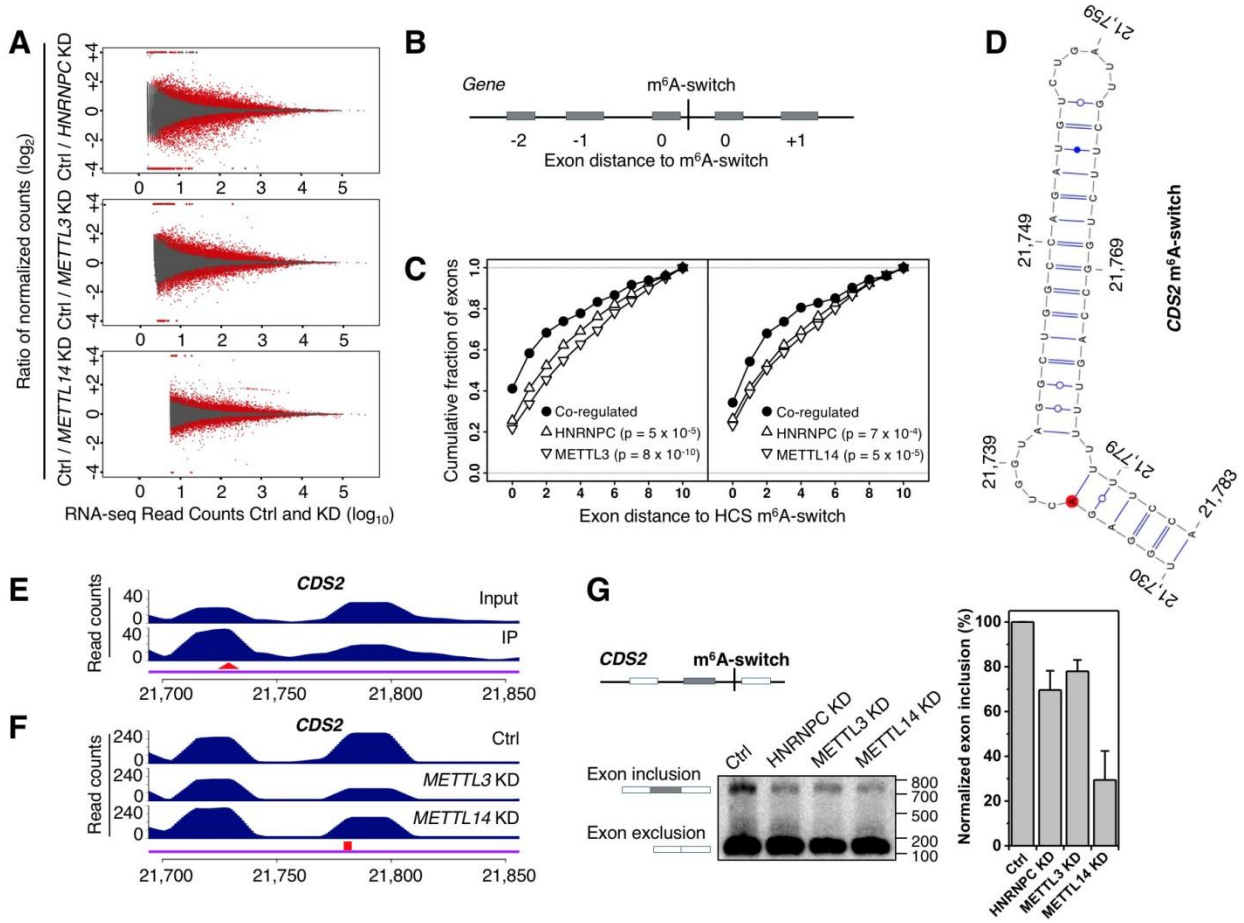


Figure 3.12. m^{6}A-switches regulate mRNA alternative splicing. A) Fold changes (KD/Ctrl, $\log_2$) in normalized exon expression against RNA-seq reads detect the exons in *HNRNPC* KD, *METTL3* KD, *METTL14* KD and control samples. Statistically Significant Differentially Expressed Exons (SSDEE) called by DEXSeq are indicated in red. **B**) Illustration of the relative exon distance to m^{6}A-switches. **C**) Co-regulated exons by *HNRNPC* KD and *METTL3* KD (left) and *METTL14* KD (right) were more enriched around m^{6}A-switch sites than non-co-regulated exons, Kolmogorov-Smirnov test. **D**) Proposed secondary structure of the *CDS2* hairpin with the m^{6}A methylation site shown in red, opposing the U-tract region. Nucleotide position numbers correspond to their locations along the human *CDS2* transcript (NM_003818). **E-G**) Validation of the m^{6}A-switch regulated splicing at one exon neighboring the *CDS2* m^{6}A-switch as shown in PAR-CLIP–MeRIP data (**E**), *METTL3*/*L14* KD data (**F**), and RT-PCR results (**G**). The red triangle and square mark the m^{6}A site and U-tract, respectively. n = 4, ± s.d., biological replicates.

Input RNA samples, which contain a large amount of cytosolic mature mRNAs, and thus the m^6A residues within intron-containing nascent transcripts could not have been adequately sequenced.

The magnitude of m^6A-switches on protein binding should be highly dependent on the context of their local RNA structures. Our structural mapping assays indicate that m^6A can make its opposite U-tracts in a stem more single-stranded/accessible, but m^6A does not completely disrupt the hairpin structure even in these weak RNA hairpins containing several non-canonical base pairs in the stem. The reduced stacking of the U-tracts is likely caused by the steric/hydrophobic effect of m^6A[21] and may serve as a recognition signature to initiate HNRNPC binding (Figure 2F). For the three m^6A switches examined in detail, the enhanced binding affinity for HNRNPC ranges between 3-8 fold. Enhancement at this magnitude can play an important role *in vivo* where the HNRNPC amount is limited and an excess of RBMs at different sites compete for HNRNPC binding[15].

m^6A-switches could influence a multitude of biological processes through regulating RNA-RBP interactions. We show that HNRNPC m^6A-switches can regulate RNA alternative splicing and affect cell proliferation through regulating the expression levels of 'cell proliferation' related genes. Although previous studies have indicated the role of HNRNPC in regulating RNA stability and gene expression levels[41,42] and in RNA splicing regulation[15-19], the detailed mechanism of HNRNPC m^6A-switches regulate these biological processes needs to be further investigated.

Besides HNRNPC, numerous other mRNA/lncRNA binding proteins could be regulated by m^6A-switches. Over 1,000 RBPs have been identified and for many, their single-stranded

RBMs may have propensity to form secondary structures, thus representing a huge challenge for RBP binding. Since m^6A-switches can weaken the RNA structure and increase the accessibility of RBMs, interactions between these RBMs and their RBPs become regulatable by m^6A-switches. Taking the U-tracts as an example, several other RBPs including ELAV1 (HuR) and U2AF65 also interact with this RBM[15,43,44]. It has been reported that the RNA-ELAV1 interaction may also be regulated by m^6A-RNA methylation[33], however, detailed molecular mechanisms are not known. The U2AF65 protein has been shown to directly compete with HNRNPC at U-tracts in mRNA introns to determine the branch site location[15]. M^6A-switches could confer regulation of mRNA binding for both HNRNPC and U2AF65. Thus, we expect m^6A-switches to function as a general mechanism in regulating cellular mRNA/lncRNA-protein interactions.

Aside from m^6A, some other mRNA/lncRNA modifications, such as m^5C, Nm and Ψ, may also remodel RNA structures through their steric or chemical groups in specific RNA structural contexts. For example, widespread dynamic-regulated Ψs have been identified in mRNA/lncRNA[45,46], and some of these Ψ sites could alter their local RNA structures in mRNA/lncRNA, in a similar way as modulating rRNA, tRNA and spliceosomal RNA structures[47]. RNA structure, a second layer of information in parallel to the genetic code, affects every step in RNA biology, including RNA-protein interactions, RNA-RNA interactions and RNA-DNA interactions. Thus, widespread RNA modifications along mRNA/lncRNA could exert their influences in numerous biological aspects through modulating the local RNA structures, like m^6A-switches.

A

Experiment	Sample/replicates	Mapped reads	T-to-C mutation rate
PARCLIP-MeRIP	Input-rep1	28,755,227	0.67
	Input-rep2	34,022,750	0.67
	IP-rep1	22,515,468	0.83
	IP-rep2	48,570,954	0.76
PAR-CLIP	Ctrl-rep1	22,728,884	0.69
	Ctrl-rep2	24,827,034	0.75
	METTL3 KD-rep1	19,697,397	0.67
	METTL3 KD-rep2	31,864,420	0.75
	METTL14 KD-rep1	20,835,499	0.67
	METTL14 KD-rep2	20,102,049	0.71

B

Experiment	Sample/replicates	Total reads	Mapped reads	Mapped rate
RNA-seq	Ctrl-rep1	78,453,301	68,058,093	86.70%
	Ctrl-rep2	78,823,710	70,907,544	90.00%
	*HNRNPC*_KD-rep1	81,933,883	72,840,271	88.90%
	*HNRNPC*_KD-rep2	78,133,137	68,053,888	87.10%
	*METTL3*_KD-rep1	58,140,943	48,123,493	82.80%
	*METTL3*_KD-rep2	58,923,206	53,011,571	90.00%
	*METTL14*_KD-rep1	77,932,017	68,703,305	88.20%
	*METTL14*_KD-rep2	75,442,883	65,812,269	87.20%

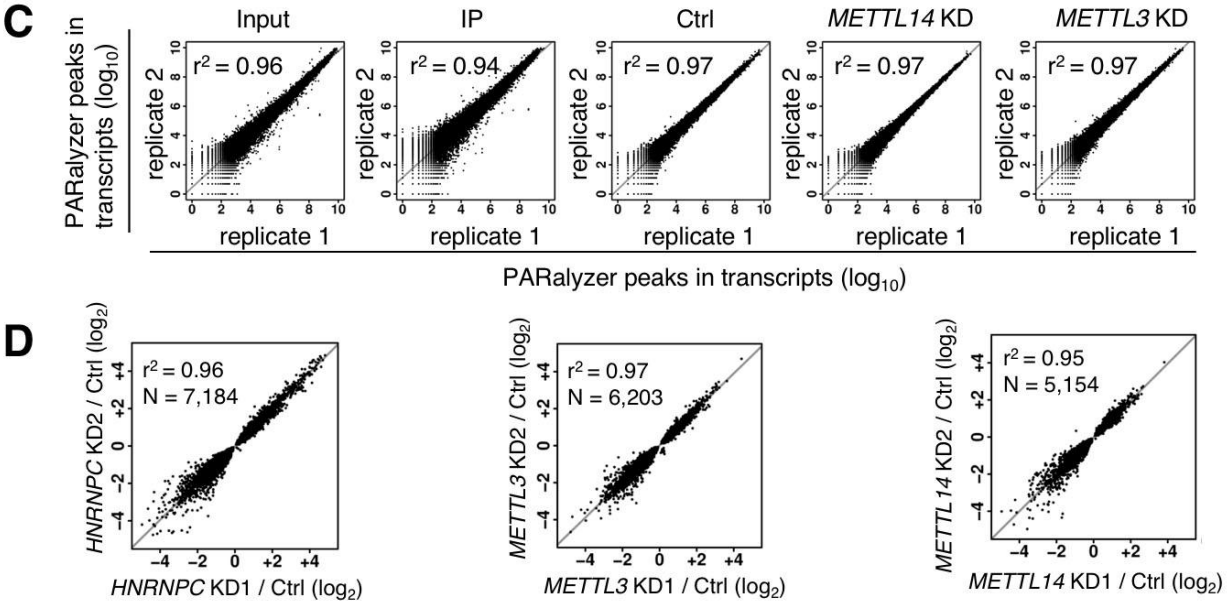


Figure 3.13 Summary of the sequencing samples. A) For PAR-CLIP–MeRIP and PAR-CLIP experiments from HEK293T cells, the number of mapped reads and "T-to-C" mutation rates are given for each replicate. **B**) For RNA-seq experiments from HEK293T cells, the number of total reads, the number of mapped reads as well as the mapping rates is given for each replicate. **C**) Scatter plots comparing transcripts for all PAR-CLIP replicate experiments. The square of Spearman's rank correlation value (r^2) for each pair is shown in the upper left corner of the respective panel. **D**) The detected expression level changes show a strong correlation between gene KD replicates. Scatter plots comparing the fold changes (log_2) in normalized gene expression from replicates of *HNRNPC*, *METTL3* and *METTL14* KD. The square of Spearman's rank correlation value (r^2) for each pair is shown in the upper left corner of respective panels.

In conclusion, we demonstrated a transcriptome-wide regulation of RNA-HNRNPC interactions through m^6A-dependent RNA structural switches. The m^6A-switch model provides an important direction for investigating RNA-modification-regulated cellular biology in the future and also shed light on the intricate interactions within RNA-protein complexes in the cell.

3.3 Experimental Section

3.3.1 Mammalian Cell Culture, siRNA KD and Western Blot

Human cervical cancer cell line HeLa (CCL-2) and embryonic kidney cell line HEK293T (CRL-11268) were obtained from American Type Culture Collection (ATCC) and were cultured under standard conditions. Control siRNA (1027281, Qiagen), METTL3 siRNA (SI04317096, Qiagen), METTL14 siRNA (SI04317096, Qiagen) or HNRNPC siRNA (10620318, Invitrogen) were transfected into HEK293T cells at a concentration of 40 nM using lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's instructions. Cells were collected 48 hours after the transfection, shock-frozen in liquid nitrogen, and stored at -80 °C for further studies. Western blot analysis using METTL3- (HPA038002, Sigma), METTL14- (HPA038002, Sigma), HNRNPC- (sc-32308, Santa Cruz), GAPDH- (A00192-40, Genescript) specific antibodies was performed under standard procedures. Blotting membranes were stained by ECL-prime (RPN2232, GE Healthcare) and visualized by a digital imaging system (G: BOX, SYNGENE).

3.3.2 Native Gel Shift Assay

HeLa nuclear extracts were isolated using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (78833, Thermo Scientific) according to the manufacturer's instructions. The purified radioactively-labeled RNA oligos were refolded by heating at 90 °C for 1 min, followed by incubation at room temperature (RT) for 5 min. 3 µl HeLa nuclear extract and 6 µl refolded RNA were incubated at RT for 30 min and then at 4 °C for 2 hrs. Each sample was mixed with 1 µl 50% glycerol, separated on the 8% native 1x TBE gel, and visualized by phosphorimaging using the Personal Molecular Imager (Bio-Rad).

3.3.3 In vitro Pull Down and Mass Spectrometry

The *in vitro* pull down assay was performed as described [26]. The eluted protein samples were separated on 4-12% polyacrylamide Bis-Tris gels (NP0321BOX, Invitrogen) and stained with SYPRO-Ruby (S12000, Invitrogen) according to the manufacturer's instructions. Protein in gel slices or the entire pulled down protein samples were digested with trypsin and identified using LC-MS/MS by the Donald Danforth Plant Science Center (Washington University, St. Louis. MO).

3.3.4 Expression of Recombinant HNRNPC1

The full-length HNRNPC1 as a GST fusion protein were purified as previously described [15]. The HNRNPC1 cDNA was generated from HeLa cDNA library, confirmed by sequencing, and cloned into vector pGEX-6P-1 (GE Healthcare) using BamHI and Xhol restriction sties. Plasmid DNA was transformed into *E. coli* BL21-CodonPlus (DE3)-RP cells (Stratagene). Freshly transformed cells were cultured in LB medium, containing150 µg/ml ampicillin. Proteins were expressed overnight at 18 °C after induction with 0.25 mM IPTG. Protein purification was carried at 4 °C, and the GST fusion protein was cleaved from the Glutathione-Sepharose beads by GST-tagged Precision Protease (GE Healthcare), leaving vector derived residues Gly-Pro-Leu-Gly-Ser at the N-terminus of the protein.

3.3.5 In vitro Gel Shift assay and UV Crosslinking

All RNA samples were radioactively labeled, and gel purified before incubation with recombinant HNRNPC1 in binding buffer (10 mM Tris pH7.4, 100 mM KCl, 2.5 mM $MgCl_2$) for 30 min at 30 °C. The samples were then UV crosslinked with 150 mJ/cm^2 in a Stratalinker 2400 at 254 nm, separated on NuPAGE Bis-Tris gels (4-12%; MES running buffer; Invitrogen) and exposed to phosphorimaging plates for visualization.

3.3.6 Filter-binding Assay

Filter-binding assays were performed as previously described [16]. Radioactively-labeled RNA oligos and HNRNPC1 were incubated in 10 µl binding buffer (10 mM Tris pH7.4, 100 mM KCl, 2.5 mM $MgCl_2$) at 30 °C for 30 min. Then, 8 µl of the incubated mixture was spotted onto Millipore HAWP 02500 nitrocellulose filters pre-soaked in binding buffer for 10 min, and washed with 15 ml ice-cold binding buffer for three times with gentle negative pressure. The

filter was then dried at room temperature for 20 min, exposed to a phosphorimaging plate and quantified with Bio-Rad Molecular Imager FX and Quantity One software (Bio-Rad). The amount of bound RNA was first normalized as fraction of the RNA bound at saturation, and then plotted against protein concentration using Origin software. Binding curves were fitted using the curve fitting routine and the apparent dissociation constant (K_d) was generated as the concentration of protein required for 50% saturation of binding.

3.3.7 CLIP-2dTLC

HEK293T cells at 70-80% confluency were UV irradiated with 400 mJ/cm^2 at 254 nm, and harvested by centrifuging at 4,000 rpm for 3 min at 4 °C. The pellet of cross-linked cells were resuspended in 1 ml lysis buffer (1x PBS, 0.1% SDS, 1% Nonidet P-40, 0.5% Sodium Deoxycholate, protease inhibitor cocktail and RNase inhibitor) and incubated on ice for 4 hrs. Cell lysate was isolated by centrifuging at 3,000 rpm for 5 min and pre-blocked with 50 µl protein A beads in 300 µl lysis buffer. Another 50 µl protein A beads (Invitrogen) were incubated with 8 µg corresponding antibodies for 4 h at room temperature, and then mixed with the pre-blocked cell lysate at 4 °C overnight. The beads were washed 3 times with 1 ml wash buffer (20 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 0.2% Tween-20), 3 times with 1 ml high salt buffer (5x PBS, 0.1% SDS, 1% Nonidet P-40, 0.5% Sodium Deoxycholate), and 3 times with 1 ml wash buffer. Crosslinked RNA was eluted from beads by incubating with 200 µl RNA elution buffer (100 mM Tris-HCl pH 7.4, 10 mM EDTA, 1% SDS) containing 2 mg/ml proteinase K at 50 °C for 30 min, followed by phenol/chloroform extraction. The beads were washed 3 times with 1 ml wash buffer, and divided into 2x 500 µl in two separate tubes. One tube was incubated with 200 µl RNase T1/A mixture at room temperature for 1 h. The other tube was incubated with 200 µl nuclease-free water at room temperature for 1 h. The beads were washed 3 times with 1 ml high salt buffer, and 3 times with 1 ml wash buffer. Crosslinked RNA was eluted from beads by incubating with 200 µl RNA elution buffer containing 2 mg/ml proteinase K at 50 °C for 30 min followed by phenol/chloroform extraction.

The RNA pellet was dissolved in 7 µl nuclease-free water containing 1 µl RNase T1 (200 U), heated at 65 °C for 2 min, and incubated at 37 °C for 30 min. The T1-digested RNA fragments were labeled upon adding 2 µl T4 PNK mix (4.5 U/µl T4 PNK, 600 Ci/mmol [γ-^{32}P] ATP, 5x PNK buffer) and incubation at 37 °C for 30 min. Unreacted [γ-^{32}P] ATP was removed

using Illustra MicroSpin G-25 columns. The eluted RNA was digested with 1 µl (1U/ µl) nuclease P1 at 37 °C for 1 h. Samples were spotted on cellulose TLC plate and 2D-TLC was run as described [10] using isobutyric acid: 0.5 M NH_4OH (5:3, v/v) as the first dimension and isopropanol:HCl:water (70:15:15, v/v/v) as the second dimension.

3.3.8 V1/S1 RNA Structural Probing

The synthetic RNA oligos were 5′ end-labeled with γ-^{32}P-ATP by T4 PNK (70031, Affymetrix) , gel purified, and re-folded. Structural probing assay with RNase T1, nuclease S1 and RNase V1 was performed as previously described [10]. Note: 3′-end-labeled HNRNPH1 oligos were used for RNA structural probing assay in (Figure 5F).

3.3.9 CMCT RNA Structural Probing

The radioactively labeled RNA oligos were prepared as above and CMCT assay was performed as reported [22]. 3 pmole RNA was annealed in 50 mM potassium borate (pH 8) by heating at 90 °C for 1.5 min then incubation at RT for 3 min. Add CMCT reagent to a final concentration of 12.5 mM, mix well, and incubate the reaction mixture (14 µl) at RT for 20 mins. Isolate the RNA samples by alcohol precipitation. Dissolve the RNA pellet in 1 µl 5x annealing buffer (50 mM Tris-Cl, pH 7.5; 500 mM KCl) and 4 µl primer (3 pmole). Anneal the mixture by heating at 90 °C for 2 mins and then RT for 5 mins. Spin down to bottom and put on ice. Perform the reverse transcription with AMV reverse transcriptase (Promega, M5101) according to the manual. All reactions were then added 10 µl 2x RNA loading buffer, boiled for 2 mins, loaded on 20% urea-gel and exposed to phosphorimaging plates for visualization and quantitation.

3.3.10 RNA Terminal Truncation

RNA terminal truncation assay was carried out as previously reported [23]. RNA samples were first alkaline-hydrolyzed as in the RNA structural probing assay, and then incubated with HNRNPC1 protein in the same conditions as in the filter binding assay. The RNA-Protein complexes were then loaded onto filter papers and washed twice with chilled binding buffer. Air-dry filters and RNA samples were then extracted from the filters and loaded onto denaturing gel as in the RNA structural probing assay.

3.3.11 PAR-CLIP

PAR-CLIP experiments were performed as previously reported [24]. HEK293T cells in 15 cm plates were grown for 14 hours in medium supplemented with 100 μM 4-thiouridine to reach ~80% confluency, washed with 10 ml ice-cold PBS, and irradiated with 365 nm UV light. Immediately after that, cells were scraped off the plates in 2 ml ice-cold PBS and centrifuged at 14,000 rpm for 10 seconds. The cell pellet was then shock-frozen in liquid nitrogen, and stored at -80 °C until use.

The cell pellet was resuspended in 1 ml lysis buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1 m $MgCl_2$, 0.1 m $CaCl_2$, 1% NP-40; 0.1% SDS, 0.5% sodium deoxycholate, 1/100 protease inhibitor cocktail III, Calbiochem), and then digested with a combination of RNase I and Turbo DNases for 3 mins at 37 °C, shaking at 1,100 rpm. The lysate was then immediately cleared by spinning at 14,000 rpm, 4 °C for 30 min, and placed on ice until use.

For each PAR-CLIP experiment, 100 μl protein G Dynabeads (Dynal, 100.02) was washed twice with 1 ml lysis buffer, incubated with 10 μl HNRNPC antibody at room temperature for 1 hour, and washed three times with 1 ml lysis buffer. Then the Dynabeads-antibody complexes were combined with the prepared lysate, followed by incubation at 4 °C for 2 hours. After twice washing with high-salt buffer (50 mM Tris-HCl pH 7.4, 1 M NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate) and three times washing with wash buffer (20 mM Tris-HCl, pH 7.4; 10 mM $MgCl_2$, 0.2% Tween-20), the Dynabeads-protein-RNA complexes were radioactively labeled within 16 μl of hot PNK mix (0.8 μl PNK [NEB], 0.8 μl ^{32}P-γ-ATP, 1.6 μl 10x PNK buffer [NEB], 12.8 μl water at 37 °C for 5 mins followed by another 15 min incubation with the addition of 100 μM ATP. After that, the protein-RNA complexes were eluted from the Dynabeads by boiling the Dynabeads-protein-RNA complexes at 95°C for 5 min, and transfered to one new tube. The RNA-protein complexes were analyzed on NuPAGE Bis-Tris gels (4-12%, MES running buffer, Invitrogen) and visualized by phosphorimaging. The RNA-protein band of the correct size was cut out, and the RNA was isolated upon proteinase K digestion followed by alcohol precipitation. The purified RNA was then used for library construction by Truseq small RNA sample preparation kit (Illumina).

3.3.12 PAR-CLIP–MeRIP

PAR-CLIP–MeRIP experiment applied m^6A-antibody immunoprecipitation [26] to the HNRNPC PAR-CLIP RNA samples, which were prepared as described in the PAR-CLIP experiment section. The HNRNPC PAR-CLIP RNA sample was incubated with m^6A-specific antibody (202003, SYSY), RNase inhibitor (80 units, Sigma-Aldrich), human placental RNase inhibitor (NEB) in 200 μl 1x IP buffer (50 mM Tris-HCl pH 7.4, 750 mM NaCl and 0.5% (vol/vol) Igepal CA-630) at 4 °C for 2 hours under gentle shaking conditions. For each PAR-CLIP–MeRIP experiment, 20 μl protein-A beads (Invitrogen) were washed twice with 1 ml 1x IP buffer, blocked with 2 hours incubation with 100 μl 1× IP buffer supplemented with BSA (0.5 mg/ml), RNasin and Human placental RNas inhibitor, and then washed twice with 100 μl 1x IP buffer. The pre-blocked protein-A beads were then combined with the prepared immuno-reaction mixture and incubated at 4 °C for 2 hours, followed by three washes with 100 μl 1× IP buffer. After that, the RNA was eluted by 1 hour incubation with 20 μl elution buffer (1× IP buffer and 6.7 mM m^6A, Sigma-Aldrich) under gentle shaking conditions, and purified by ethanol precipitation. The purified RNA sample (IP) as well as the input PAR-CLIP RNA sample (Input control) were used for library construction by Truseq small RNA sample preparation kit (Illumina).

3.3.13 RNA Deep Sequencing and Genomic Mapping

Libraries were prepared using TruSeq Small RNA Sample Preparation Kit (RS-200-0012, Illumina) according to the manufacturer's instructions, and then sequenced by Illumina Hiseq2000 with single end 50-bp read length. The control and IP samples from PAR-CLIP–MeRIP experiments (same case for the control and KD samples from *METTL* KD experiments) were sequenced together in one flowcell on two lanes, and the reads from two lanes of each sample were combined for remaining analysis. The raw seq data was trimmed using the Trimmomatic computer program version 0.30 [48] to remove adaptor sequences, and mapped to the Human genome version hg19 by Bowtie 1.0.0 [49] without any gaps and allowed for at most two mismatches (**Figure 3.13**).

3.3.14 Normalization of the Mapping Reads for Subsequent Analysis

The raw read counts of the biological replicates confirmed the reproducibility between replicates (Figure S3D), and were combined for subsequent analysis. For each genomic site, we

calculated the average read counts within an 11-nt window centered at that site, as the normalized read counts for that site. This normalization smoothed the raw mapping curves, and facilitated identification of peaks within each mapping cluster. To correct for changes in sequencing depth or expression levels between samples, we then normalized the read counts at each genomic site to the total number of read counts on the respective gene. The above defined double-normalization procedures enabled precise identification of changes in the mapping reads at specific genomic locations by directly comparing the normalized read counts between samples. No read counts in the intergenetic region were compared between samples, because the transcription boundaries are not defined at this region and the intergenetic read counts cannot be normalized to correct changes for transcript expression.

3.3.15 Identification of HNRNPC binding sites

HNRNPC binding sites were identified by PARalyzer v1.1 [27] with default settings, since all our RNA samples for deep-seq were incorporated with 4-thiouridine (4-SU) within cell incubation periods, causing T-to-C mutations at the crosslinking sites during sequencing. The sequences identified by PARalyzer v1.1 based on the T-to-C mutations were recognized as HNRNPC binding sites.

3.3.16 Detection of PAR-CLIP–MeRIP peaks

Detection of PAR-CLIP–MeRIP peaks involves comparing the read counts of the IP sample with that of the control (Ctrl) sample as follows: (i) we identified all peaks within HNRNPC binding sites in the IP sample; (ii) we performed transcriptome-wide scanning to compare read counts of each identified peak in (i) with read counts at same genomic locations in the Ctrl sample to calculate the fold change score, score = $\log_2 (H_{IP}/H_{Ctrl})$. The score threshold was set to be 1, corresponding to a twofold increase compared with control.

3.3.17 Detection of decreased HNRNPC binding peaks upon METTL KD

Detection of decreased HNRNPC binding sites involved comparing HNRNPC occupancies in the *METTL* KD (KD) sample with that in control as follows: (i) we identified all peaks within HNRNPC binding sites in the *METTL* KD sample; (ii) we performed

transcriptome-wide scanning to compare read counts of each identified peak in (i) with read counts at the same genomic locations in control to calculate the fold change score, score = $\log_2$ (H_{KD}/H_{Ctrl}). The score threshold was set to be -1, corresponding to a twofold decrease compared with control.

3.3.18 Identification of enriched motifs

To identify enriched motifs, we first sorted the 12,998 HNRNPC PAR-CLIP–MeRIP peaks (with IP/Input enrichment ≥ 2) by the T-to-C mutation frequency. We then chose the top 4,500 peaks with the highest T-to-C mutation frequency for motif analysis using FIRE [50] with default RNA analysis parameters. The top two enriched motifs are the GRACH and the U-tract motif. We also used the top 1,024 and 2,048 peaks for motif analysis, yielding the same motif results as the top 4,500 peaks.

3.3.19 Identification of HNRNPC m^{6}A-switches

To identify transcritpome-wide HNRNPC m^{6}A-switches, we first searched for all coupling events within 50 nucleotides between U_5 and RRACH motif, with the U_5 motif located within HNRNPC binding sites. While 125,844 coupling events were identified in PAR-CLIP–MeRIP samples, 39,498 and 36,456 coupling events were identified in *METTL3* and *METTL14* KD samples, respectively.

For PAR-CLIP–MeRIP samples, the fold change score E at the RRACH motif was calculated for each coupling event. Also, p-value for each coupling event was calculated as described [51]. Then, we generated the π-value, $\pi = E \cdot (-\log_{10} P)$, as one comprehensive parameter to pick meaningful genomic loci [52]. HNRNPC m^{6}A-switches identified from PAR-CLIP–MeRIP experiments should fulfill the following requirements: (i) read counts at both the control and IP sample ≥ 5; (ii) π-value ≥ 0.627, corresponding to FDR ≤ 5%.

For METTL KD samples, the fold change score at the U-tracts motif was calculated for each coupling event. HNRNPC m^{6}A-switches identified from *METTL3/L14* KD samples should fulfill the following requirements: (i) read counts at both the control and KD sample ≥ 5; (ii) π-value ≤ 0.627, corresponding to FDR ≤ 5%.

3.3.20 Distribution of HNRNPC m^{6}A-switches

Pie charts illustrating distribution within each segment were made using the following hierarchy: intron > ncRNA > 3′UTR > 5′UTR > CDS > intergenic.

To plot the distribution of HNRNPC m^6A-switches in their respective localized segments (such as intron, exon, 3′UTR, CDS, 5′UTR), we first identified the distance between each m^6A-switch and the 5′ end of the respective segment. This distance was then divided by the length of that segment to determine a percentile where this m^6A-switch fell, and then this specific percentile bin was incremented. Following this approach, we obtained the distribution pattern of all m^6A-switches within each segment.

3.3.21 RNA-seq Analysis

RNA-seq experiments were performed on two replicate RNA samples from *HNRNPC*, *METTL3*, *METTL14* KD as well as control HEK293T cells (48 hours after transfection). Total RNA samples were extracted according to RNeasy plus kit (Catalog # 74104, Qiagen). Libraries were prepared according to the TruSeq Stranded mRNA LT Sample Prep Kit (Catalog # RS-122-9005DOC). KD and control samples were sequenced together in one flowcell on four lanes, respectively. All samples were sequenced by illumina Hiseq 2000 with pair end 100-bp read length. The reads from the four lanes of each sample were combined for all analysis.

The RNA-seq data was mapped using the splice-aware alignment algorithm TopHat version 1.1.4 [53] based on the following parameters: tophat –num-threads 8 –mate-inner-dist 200 –solexa-quals –min-isoform-fraction 0 –coverage-search-segment-mismatches 1.

Isoform expression level changes were analyzed using cuffdiff [54]. Analysis of differential gene expression were performed using DESeq [55] based on Ensembl gene annotations [56]. Differential splicing was determined using DEXSeq [39] based on Cufflinks-predicted, nonoverlapping exons.

3.3.22 Gene Ontology Analysis

Gene Ontology (GO) enrichment analysis was applied on the 815 co-regulated HCS-containing genes, against all HCS-containing genes as background, using GOrilla [57]. GO hits were filtered by REVIGO [58] to remove redundant ones.

3.3.23 Cell Proliferation Analysis

HEK293T cells were transfected with si-control, si-*HNRNPC*, si-*METTL3* and si-*METTL14* RNAs. After transfection, the numbers of cells were counted at 0, 24, 48 and 72 hrs as described in [59]. Three independent experiments were performed and growth curves were plotted to test the effects on cell proliferation.

3.3.24 Evolutionary Conservation Analysis

Phylogenetic conservation analysis was performed by comparing PhyloP scores at the U-tracts motif and RRACH motif for HNRNPC m^6A-switches to those of randomly selected sequences. The PhyloP scores were accessed from the precompiled phyloP scores [60] (ftp://hgdownload.soe.ucsc.edu/goldenPath/hg19/phyloP46way/) under both primates and vertebrates categories. P-values were evaluated using the Mann-Whitney-Wilcoxon test, ***: $p < 10^{-16}$. For the U-tract motifs, we collected all U-tracts (5x U's) across all chromosomes and randomly selected 10,000 sites among the 38,561,577 sites of our census. The random selection was done separately for primates and for vertebrates. For the RRACH motif, we also collected all RRACH sites across all chromosomes and randomly selected 10,000 sites among the 78,815,225 sites of our census. Here too, the random selection was done separately for primates and vertebrates.

3.3.25 RT-PCR Quantitation

Total RNA samples were extracted from HEK293T cells and reverse transcribed using SuperScript® III First-Strand Synthesis System (Life Technologies, #18080-051). In order to validate the splicing changes identified from our RNA-seq data, we performed RT-PCR measurements using Thermo Scientific™ Taq™ DNA Polymerase under the following conditions: 95 °C for 3 mins, 30 cycles of [95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min] and then finally 72 °C for 10 min. For the target alternate exon, we designed and used primers annealing to both neighboring constitutive exons. The PCR products were separated on 1.2% agarose gel and ethidium bromide stained. In order to validate the gene expression level changes identified from our RNA-seq data, we performed qRT-PCR measurements using Power SYBR® Green PCR Master Mix (Life Technology, # 4367659) under the following conditions: 50 °C for

3 mins followed by 95 °C for 10 mins, 40 cycles of [95 °C for 15 s, 60 °C for 1 min] and then 40 °C for 1 min and 95 °C for 15 s and finally 60 °C for 30 s.

3.3.26 Graphic and Statistical Analysis

Sequence logos were generated using the WebLogo package. R statistical package was used for all statistical analysis (unless stated otherwise).

Table 3.1. Information on oligonucleotides. The oligonucleotides' sequences for each experiment are given in the 5' to 3' direction (from left to right).

RNA pull down experiments	
MALAT1-2,577-A-B	AACUUAAUGUUUUUGCAUUGGACUUUGAGUUA-Biotin
MALAT1-2,577-m^{6}A-B	AACUUAAUGUUUUUGCAUUGGm^{6}ACUUUGAGUUA-Biotin
MALAT1-2,577-U-B	AACUUAAUGUUUUUGCAUUGGUCUUUGAGUUA-Biotin
MALAT1-CC-2,577-A-B	AACUUAAUGUCCUUGCAUUGGACUUUGAGUUA-Biotin
MALAT1-CC-2,577-m^{6}A-B	AACUUAAUGUCCUUGCAUUGGm^{6}ACUUUGAGUUA-Biotin
RNA structural mapping and binding assays	
MALAT1-2,577-A	AACUUAAUGUUUUUGCAUUGGACUUUGAGUUA
MALAT1-2,577-m^{6}A	AACUUAAUGUUUUUGCAUUGGm^{6}ACUUUGAGUUA
MALAT1-2,577-U	AACUUAAUGUUUUUGCAUUGGUCUUUGAGUUA
CUT1-A	UUUUUGCAUUGGACUUUGAGUUA
CUT1-m^{6}A	UUUUUGCAUUGGm^{6}ACUUUGAGUUA
CUT2-A	AACUUAAUGUUUUUGCAUUGGAC
CUT2-m^{6}A	AACUUAAUGUUUUUGCAUUGGm^{6}AC
MALAT1-mutant-i	AACUUAAUCUUUUUUGCAUUGGACUUUGAGUUA
MALAT1-mutant-ii	AACUUAAUCUUUUUUGCAUUGGm^{6}ACUUUGAGUUA
MALAT1-mutant-iii	AACUUAAUCUUUUUUGCAUUGGAGUUUGAGUUA
MALAT1-mutant-iv	AACUUAAUCUUUUUUGCAUUGGm^{6}AGUUUGAGUUA
DNAJC25-GNG10-A	AAUGUUUUUUUAAGAGGACAAG
DNAJC25-GNG10-m^{6}A	AAUGUUUUUUUAAGAGGm^{6}ACAAG
HNRNPH1-A	UGGGCUGGACUGUUGGUGGAGGCUGGCUUUUUAGCCCA
HNRNPH1-m^{6}A	UGGGCUGGm^{6}ACUGUUGGUGGAGGCUGGCUUUUUAGCCCA
MALAT1-2,577-A	AACUUAAUGUUUUUGCAUUGGACUUUGAGUUA
qPCR experiments for expression level changes	
ANAPC1-for	TGCCAAAAGAAATAGCAGTTCAG
ANAPC1-rev	TGCCAAAAGAAATAGCAGTTCAG
ANLN-for	GCCAGGCGAGAGAATCTTCA
ANLN-rev	GGCTGCTGGTTACTTGCTTC
SRSF6-for	ACAAGGAACGAACAAATGAGGG
SRSF6-rev	GCTTCCAGAGTAAGATCGCCTAT
E2F8-for	ACCCAAGCTCAGCCATTGTA
E2F8-rev	GAGTCATAGTTGGTGGCCCT

(Table 3.1 Continued)

HIPK1-for	CCAGTCAGCTTTGTACCCATC
HIPK1-rev	TTGAAACGCAGGTGGACATA
DNAJA3-f	CCCTTTCATTTGTACTGCCTCC
DNAJA3-r	TGATCTCTTTCTGGCTGGCA
STAMBP-for	GTTCTCATCCCCAAGCAAAG
STAMBP-rev	ATCCAGCCCAGTGTGATGA
ARHGAP5-for	GCGGATTCCATTTGACCTCC
ARHGAP5-rev	GCTGCCCTGGTGAAATGAAT
ROBO1-for	TTTGGGCTTCTGCGTAGTTT
ROBO1-rev	GGAGGGTACTGGAGACAGCA
SRPK1-for	CCCTGAGAAGAGAGCCACTG
SRPK1-rev	ACCCTGAAAAGGGAAGAGGA
CENPK-for	AAGGCTAAAAATTCACAAAGCA
CENPK-rev	TCCATATCTTTCCACATTTCTTCA
BCLAF1-for	TCCTGAAAGGTCTGGGTCTG
BCLAF1-rev	AACGGCCCCTAGACTCATCT
SUDS3-for	TGCCTGGGGTTCTGTATTTC
SUDS3-rev	CAGTTCAAGCGAGGGAAGTC
DYRK1A-for	CTTCAGCATGCAAACCTTCA
DYRK1A-rev	GGCAGAAACCTGTTGGTCAC
SMEK1-for	TTGAAGGACTGCACCACTTG
SMEK1-rev	CCTGTGTTTTCGTGGTTGTG
ATP6V1A-for	AAGCATTTCCCCTCTGTCAA
ATP6V1A-rev	CTGCCAGGTCTTCTTCTTCC
KPNA6-for	CCCTGTGTTGATCGAAATCC
KPNA6-rev	GATCTGCTCAGGGGTTCCTC
TBC1D23-for	GGTGAATCTCCTAATGGCTCA
TBC1D23-rev	CGATCCACAGGAGTTGATGT
GPBP1-for	CGTCATTGAATTTTGAGAAGCA
GPBP1-rev	TTAGGACGCCCAATAGCAGA
MTF2-for	GTCTGCATTTGGTTCCTGGT
MTF2-rev	CTGCAGGAAAGGCAACCTTA
ATP6V0A1-for	TCCGTGTCTGGTTCATCAAA
ATP6V0A1-rev	TCTGAGTGCAAACTGGATGG
MAP4K3-for	TCTTCATACCACAGGAAATGC
MAP4K3-rev	AACAGGTTTGTGTGGGGGTA
SUMO2-for	TTCTTTCATTTCCCCCTTCC
SUMO2-rev	TATTTTTCCCCATCCCGTCT
MAP3K3-for	CAGTTCCTCTCCCCACTCTG
MAP3K3-rev	GACAGAGAGGTGCCTGCTTC
RT-PCR experiment for splicing validation	
CDS2-for	CGATTTTCCCAGGATGACAG
CDS2-rev	GAAAGGGCCCTATTGAGGAC

3.4 References

1. Castello, A., Fischer, B., Hentze, M.W. & Preiss, T. RNA-binding proteins in Mendelian disease. *Trends Genet* **29**, 318-27 (2013).
2. Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-406 (2012).
3. Baltz, A.G. et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46**, 674-90 (2012).
4. Antson, A.A. Single-stranded-RNA binding proteins. *Curr Opin Struct Biol* **10**, 87-94 (2000).
5. Dreyfuss, G., Kim, V.N. & Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* **3**, 195-205 (2002).
6. Kertesz, M. et al. Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103-7 (2010).
7. Wan, Y. et al. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505**, 706-9 (2014).
8. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J.S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701-5 (2014).
9. Ding, Y. et al. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696-700 (2014).
10. Liu, N. et al. Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19**, 1848-56 (2013).
11. Wilusz, J.E., Freier, S.M. & Spector, D.L. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919-32 (2008).
12. Tripathi, V. et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **39**, 925-38 (2010).
13. Liu, N. et al. Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19**, 1848-56 (2013).
14. Zarnack, K. et al. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152**, 453-66 (2014).
15. Zarnack, K. et al. Direct Competition between hnRNP C and U2AF65 Protects the Transcriptome from the Exonization of Alu Elements. *Cell* **152**, 453-66 (2014).
16. Gorlach, M., Burd, C.G. & Dreyfuss, G. The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J Biol Chem* **269**, 23074-8 (1994).
17. Krecic, A.M. & Swanson, M.S. hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* **11**, 363-71 (1999).
18. McCloskey, A., Taniguchi, I., Shinmyozu, K. & Ohno, M. hnRNP C tetramer measures RNA length to classify RNA polymerase II transcripts for export. *Science* **335**, 1643-6 (2012).
19. Konig, J. et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**, 909-15 (2010).
20. Ray, D. et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172-7 (2013).

21. Kierzek, E. & Kierzek, R. The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic Acids Res* **31**, 4472-80 (2003).
22. Ehresmann, C. et al. Probing the structure of RNAs in solution. *Nucleic Acids Res* **15**, 9109-28 (1987).
23. Peterson, E.T., Pan, T., Coleman, J. & Uhlenbeck, O.C. In vitro selection of small RNAs that bind to Escherichia coli phenylalanyl-tRNA synthetase. *J Mol Biol* **242**, 186-92 (1994).
24. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-41 (2010).
25. Meyer, K.D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635-46 (2012).
26. Dominissini, D. et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-6 (2012).
27. Corcoran, D.L. et al. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* **12**, R79 (2011).
28. Ouyang, Z., Snyder, M.P. & Chang, H.Y. SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res*, **23**, 377-87 (2013).
29. Xiao, Y. et al. A novel significance score for gene selection and ranking. *Bioinformatics*, **30**, 801-7 (2014).
30. Xiao, Y. et al. A novel significance score for gene selection and ranking. *Bioinformatics* **30**, 801-7 (2014).
31. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J.S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701-5 (2014).
32. Wan, Y. et al. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*, **505**, 706-9 (2014).
33. Wang, Y. et al. N-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**, 191-8 (2014).
34. Liu, J. et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N-adenosine methylation. *Nat Chem Biol* **10**, 93-5 (2014).
35. Pedersen, J.S. et al. Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol* **2**, e33 (2006).
36. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
37. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. in *Nat Protoc*, Vol. 7 562-78 (England).
38. Huelga, S.C. et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* **1**, 167-78.
39. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, 2008-17 (2012).
40. Saletore, Y. et al. The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol* **13**, 175 (2012).

41. Rajagopalan, L.E., Westmark, C.J., Jarzembowski, J.A. & Malter, J.S. hnRNP C increases amyloid precursor protein (APP) production by stabilizing APP mRNA. *Nucleic Acids Res* **26**, 3418-23 (1998).
42. Shetty, S. Regulation of urokinase receptor mRNA stability by hnRNP C in lung epithelial cells. *Mol Cell Biochem* **272**, 107-18 (2005).
43. Lebedeva, S. et al. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol Cell* **43**, 340-52 (2011).
44. Mukherjee, N. et al. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell* **43**, 327-39 (2011).
45. Carlile, T.M. et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* (2014).
46. Schwartz, S. et al. Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-62 (2014).
47. Charette, M. & Gray, M.W. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49**, 341-51 (2000).
48. Lohse, M. et al. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res* **40**, W622-7 (2012).
49. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
50. Elemento, O., Slonim, N. & Tavazoie, S. A universal framework for regulatory element discovery across all genomes and data types. *Mol Cell* **28**, 337-50 (2007).
51. Ouyang, Z., Snyder, M.P. & Chang, H.Y. SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res* **23**, 377-87 (2013).
52. Xiao, Y. et al. A Novel Significance Score for Gene Selection and Ranking. *Bioinformatics* (2012).
53. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
54. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-78 (2012).
55. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
56. Flicek, P. et al. Ensembl 2012. *Nucleic Acids Res* **40**, D84-90 (2012).
57. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
58. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
59. Yang, F., Yi, F., Han, X., Du, Q. & Liang, Z. MALAT-1 interacts with hnRNP C in cell cycle regulation. *FEBS Lett* **587**, 3175-81 (2013).
60. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110-21 (2010).

Chapter 4: N^6-methyladenosine-dependent RNA Structural Switches Recruit HNRNPG for Alternative Splicing Regulation

4.1 Introduction

In eukaryotic cells, the control of alternative splicing is essential for generating precise transcriptomic and proteomic diversity, and mis-regulation of alternative splicing can cause many human diseases[1]. RNA alternative splicing is dictated by multiple splicing factors, which include heterogeneous nuclear ribonucleoproteins (HNRNP)[2]. HNRNP protein families consist of over 20 proteins originally labeled from A1 to U according to their corresponsive 2-dimensional electrophoretic mobilities[3]. HNRNPs are majorly nucleus-localized and associate with nascent RNA transcripts for their processing steps[3].

N^6-methyladenosine (m^6A) is the most abundant and dynamic internal modification in the eukaryotic mRNA [4-9]. It affects almost every step of RNA biology, and displays physiological significance in wide-ranging fields[10-17]. On average, each mammalian mRNA contains over three m^6A sites within the G(m^6A)C (70%) or A(m^6A)C (30%) consensus motif[18,19]. Similar as the dynamic methylations on DNA and histone proteins, m^6A RNA methylation can be installed by m^6A methyltransferases (writers)[20-23], removed by demethylases (erasers)[24,25] and recognized by m^6A-responsive proteins (readers)[6,10,11,15,17]. A large fraction (around 30%) of human m^6A-writers' target RNA sites are located in introns, suggesting that m^6A methylation might occur co-transcriptionally either before or during splicing[21,23]. And various pre-mRNA splicing factors co-localize with the m^6A writers and erasers in the nuclear speckles[21], indicating the m^6A dynamics may add another layer of dynamic control to RNA alternative splicing. Indeed, m^6A writer

depletion impacts alternative splicing changes[6]. But the detailed mechanisms how m^6A manipulate alternative splicing remains unclear.

Besides the direct recognition mode by the YTH-domain proteins[10,11,26], m^6A can alter RNA structures to indirectly modulate HNRNPC binding[15,27]. This mechanism, termed 'm^6A-switch', is supported by biophysical characterizations that the m^6A destabilizes paired RNA regions[28,29], as well as large-scale RNA structural mapping results that m^6A functions as RNA structural remodelers across the transcriptome[28,30]. The m^6A-switched HNRNPC binding events impact the downstream functions of HNRNPC, influencing RNA abundance and alternative splicing of substrate RNAs[15]. Since HNRNPC-mediated splicing changes account only a portion of m^6A-dependent splicing changes, there probably exist other splicing factors regulated by m^6A. Besides HNRNPC, m^6A-switches could affect many other RNA-binding proteins' interactions through manipulating the accessibilities of their RNA binding motifs. What other RNA-binding proteins could be affected by m^6A-switches, what are the detailed mechanisms and their biological consequences remain unaddressed key questions with regard to m^6A-mediated gene regulation.

In this work, we identify HNRNPG as another RNA-binding protein regulated by m^6A-switches. HNRNPG belongs to the large HNRNP protein families, and it has been shown to regulate transcription and alternative splicing on many genes. We find that the m^6A is incorporated into HNRNPG binding consensus sequences to increase their accessibility for HNRNPG recruitment. Furthermore, HNRNPG is regulated by m^6A-switches via its C-terminal disordered low-complexity domain, which is frequently involved in the formation of RNA granules and protein-protein interactions. Finally, m^6A writers depletion causes similar alternative splicing changes as HNRNPG loss at the m^6A modified mRNAs, illustrating the role

of m^6A-switches in recruiting HNRNPG for alternative splicing regulation. Our findings have important implications at dissecting the intricate interactions between nascent RNA transcripts and their binding proteins for pre-mRNA processing events.

4.2 Results and Discussion

4.2.1 HNRNPG selectively binds m^6A methylated RNA

Our previous studies identified one m^6A residue located at the 2,515 site of the human metastasis-associated lung adenocarcinoma transcript (MALAT1) lncRNA, and RNA secondary structure prediction algorithm indicates the m^6A residue is located in one hairpin-stem[8] (**Figure 4.1A**). To investigate this m^6A function, we synthesized RNA hairpin oligos with and without the m^6A at the 2,515 site. Native gel shift assay indicates that m^6A methylation increases nuclear protein loading on the hairpin (**Figure 4.1B**). We then performed RNA pull down assay with biotinylated hairpin oligos, and both denaturing gel electrophoresis and mass spectrometry results identified HNRNPG as the nuclear protein component that preferably binds the m^6A methylated hairpin (**Figures 4.1C-D**). These results suggest that HNRNPG is one m^6A reader in the nucleus.

HNRNPG is coded by *RBMX* (for RNA binding motif gene, X chromosome) in human cells and is expressed ubiquitously[31,32]. HNRNPG belongs to the large family of HNRNPs[3], and has long been ignored for a low nuclear concentration and lack of function until recent demonstrations that it acts as a splicing factor implicated in specific splice site selections[3]. HNRNPG can act positively or negatively on the incorporation of specific exons within many genes[33,34]. Besides, by associating with most transcription units of RNA polymerase II, HNRNPG function as transcription regulator on two genes[35,36]. Furthermore, HNRNPG

influences sister chromatids co hesion[37], DNA damage repair[38], as well as proper neural development of zebrafish and frog embryos[39,40].

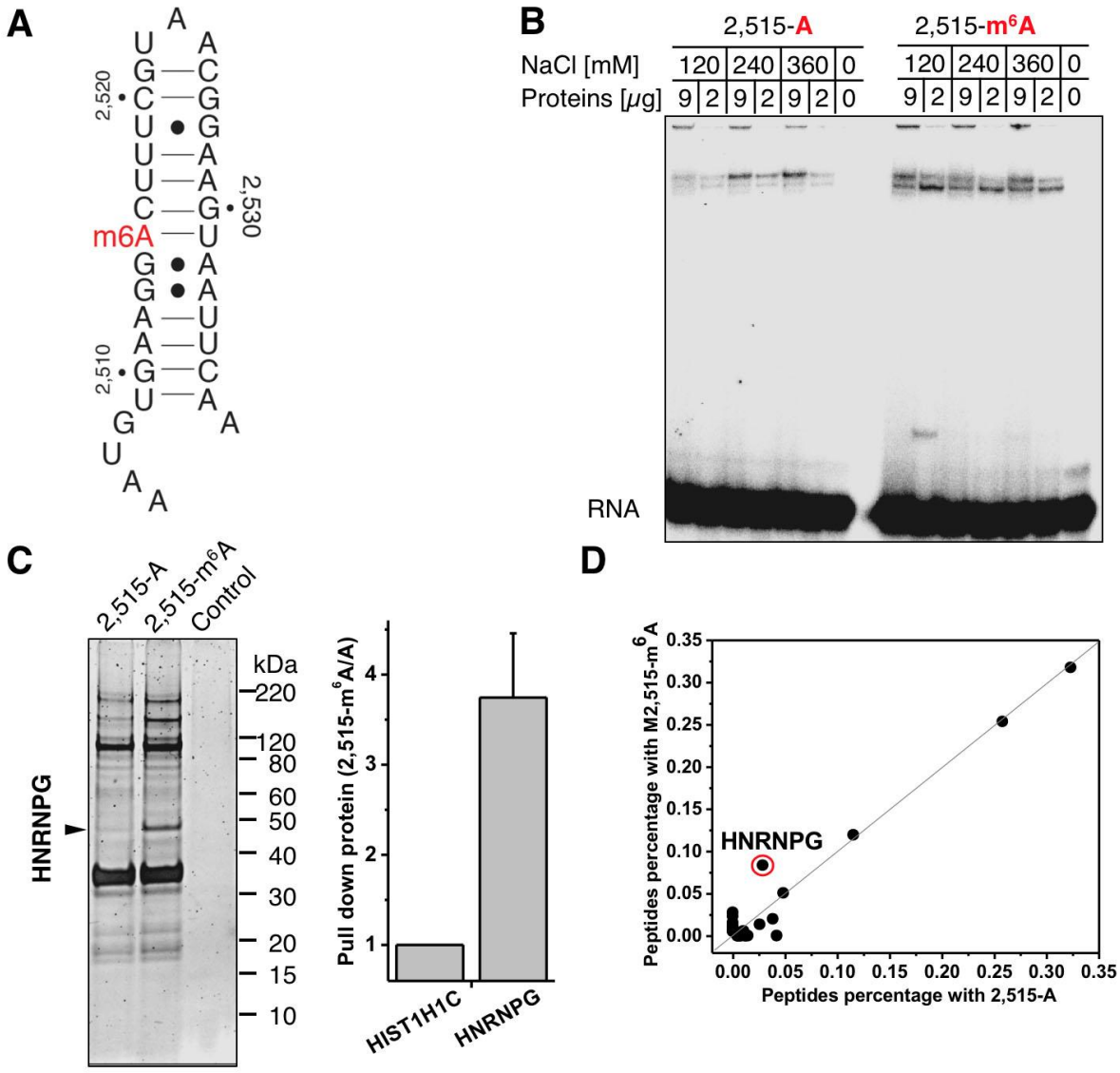


Figure 4.1. HNRNPG selectively binds m^{6}A modified RNAs. A) The m^{6}A residue at 2,515 site of the MALAT1 lncRNA is located in one hairpin-stem. **B**) Gel shift showing human nuclear proteins selectively recognize m^{6}A methylated RNAs. **C**) Denaturing gel electrophoresis showing pull down proteins by MALAT1 hairpins with and without m^{6}A at the 2,515 site. Control, pull down experiments without RNA. Quantitation data shown on the right. Data are mean ± standard deviation (s.d.); n=5, biological replicates. **D**) Quantitative mass spectrometry results showing pull down proteins by MALAT1 hairpins with and without m^{6}A at the 2,515 site.

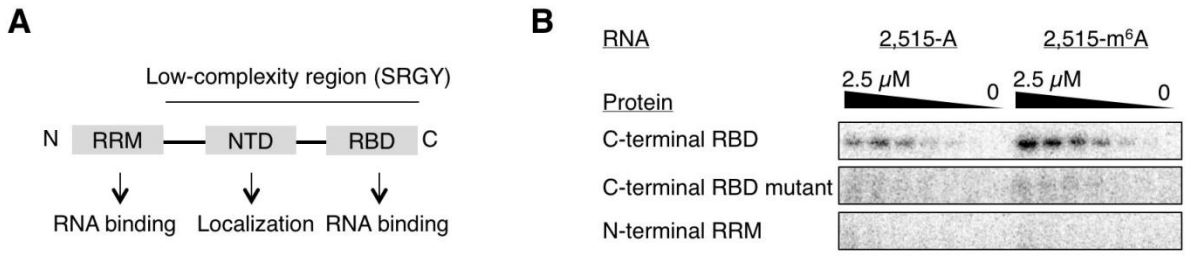


Figure 4.2. The C-terminal RNA binding domain of HNRNPG is m^{6}A responsive. A) Schematic representation of the full-length human HNRNPG protein. **B**) Ultraviolet crosslinking assay between HNRNPG RNA binding domains (RBDs) and MALAT1 hairpins with and without m^{6}A at the 2,515 site. C-terminal RBD mutant, three RGG repeats mutated into three FGG repeats. RRM, RNA Recognition Motif.

The RNA-binding properties of HNRNPG have been well characterized on the N-terminal RRM, which is a protein motif widely encountered in RNA binding proteins[41]. The N-terminal RRM binds A/C rich motifs in single stranded RNAs[41,42]. The rest region of HNRNPG are low-complexity (LC) squences, which is characterized by a biased amino acid composition[43]. The functions of the LC domains remain unknown until Kanhoush *et al.* show that this region contain two novel functional domains: a Nascent Targeting Domain (NTD) and a C-Terminal RNA Binding domain (RBD) (**Figure 4.2A**)[44]. Kanhoush et al. proposed that the NTD locates HNRNPG in the nucleus through protein-protein interactions, and the N-terminal RRM and C-terminal RBD are responsible for directly binding nascent transcripts. Kanhoush indicated that the C-ter RBD prefers to bind the A/G-rich motif located within a stable hairpin-stem loop. Consistent with this notion, the MALAT1 hairpin shows binding activities with the C-terminal RBD, but not the N-terminal RRM (**Figure 4.2B**). Remarkably, the m^{6}A methylation at 2,515 site greatly improved C-ter RBD binding (**Figure 4.2B**). We notice that there are three RGG repeats at the C-ter RBD, which are frequently included in the RNA binding domains of many RBPs and may contribute to bind the MALAT1 hairpin. Indeed, mutations at the RGG repeats

eliminate the binding activities of the C-ter RBD (**Figure 4.2B**). These findings indicate that HNRNPG selectively recognizes m^{6}A methylated RNAs through its C-ter RBD.

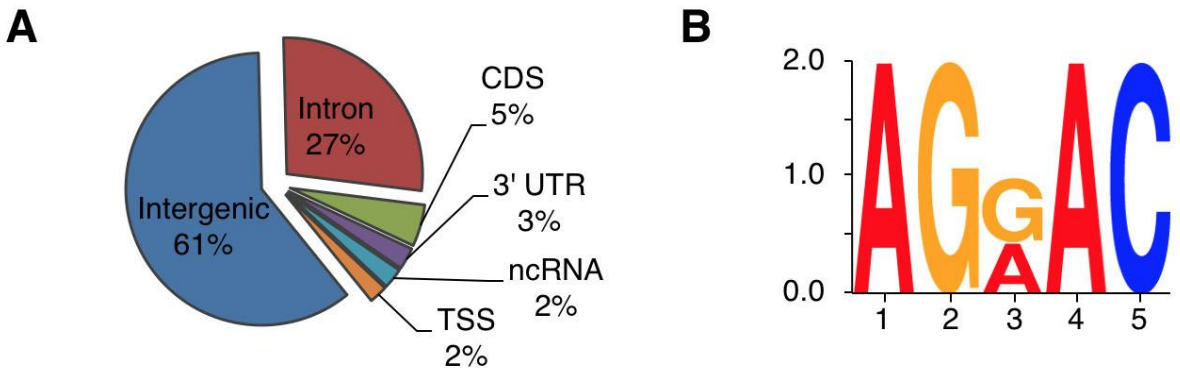


Figure 4.3. The RNA-binding properties of HNRNPG revealed by PAR-CLIP. A) Pie chart showing the regional distribution of HNRNPG binding sites. **B**) Weblogo showing the specific binding sequence of HNRNPG with PAR-CLIP peaks.

To characterize the large-scale binding of HNRNPG across the human transcriptome *in vivo*, we performed photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) experiments[45]. In total, we identified 354,057 HNRNPG binding sites in live HEK293T cells from two biological replicate experiments. The majority of HNRNPG binding sites are located in the intronic and intergenic regions (**Figure 4.3A**), consistent with its role in binding nascent RNA transcripts for pre-mRNA processing. In total, 106,300 AGRAC motif sites within HNRNPG binding sites were identified, and AGRAC (R = A or G) sequences emerge as the most enriched motif among HNRNPG binding sites (**Figure 4.3A**). These specific binding sequence of HNRNPG embeds the m^{6}A consensus motif 'GRAC'[18,19], and thus could be m^{6}A methylated. These results indicate that the m^{6}A role in recruiting HNRNPG at the MALAT1 hairpin-stem could be possibly generalized to many other HNRNPG target RNA transcripts.

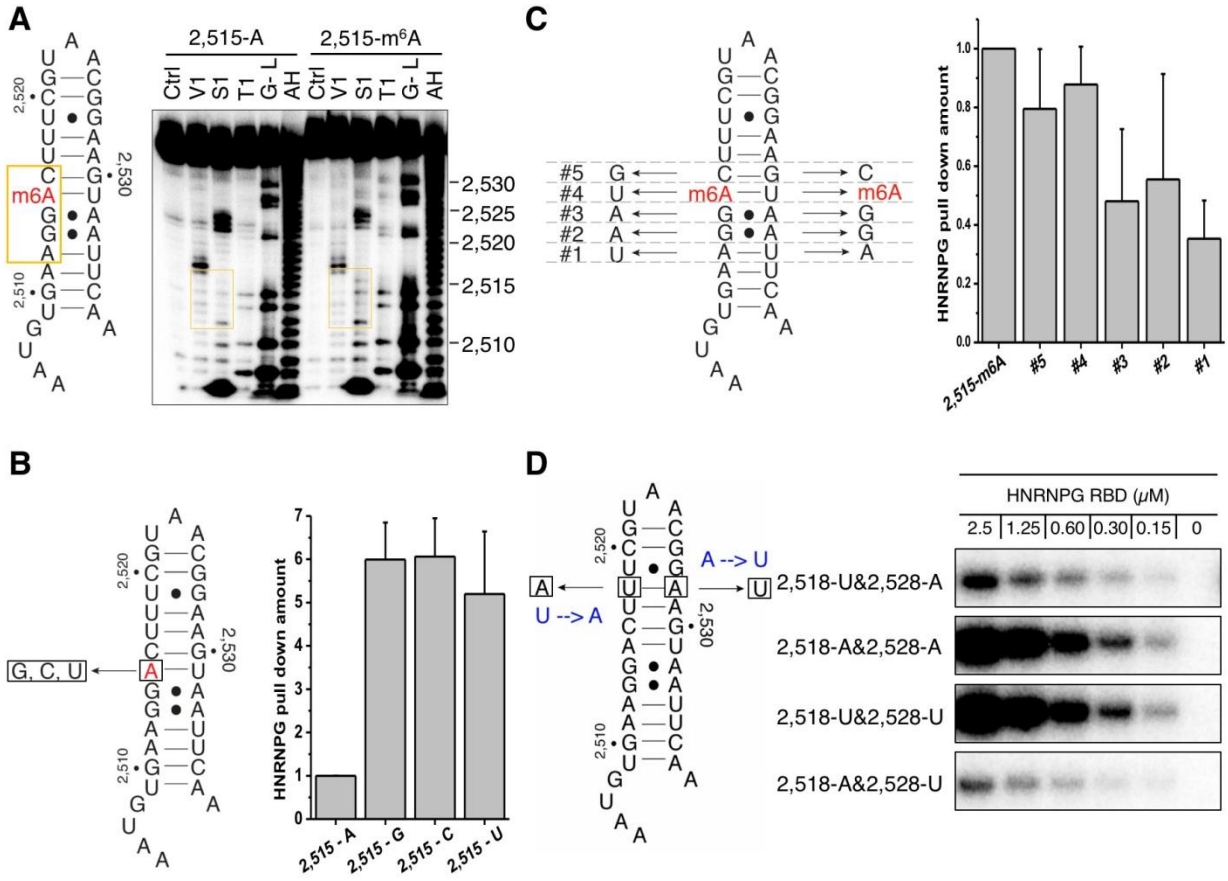


Figure 4.4. m^{6}A destabilizes the hairpin-stem structure to recruit HNRNPG. A) RNA structural probing showing that m^{6}A at 2,515 site destabilizes hairpin-stem and increases the accessibility of AGGAC motif, as highlighted in yellow. AH, alkaline hydrolysis; Ctrl, no nuclease added; G-L, G-ladder; S1, nuclease S1 digestion; T1, RNase T1 digestion; V1, RNase V1 digestion. **B**) RNA pull down results showing that the AGG motif within the AGGAC sequence is critical for binding HNRNPG. The positions of each base-pair (#1-5, respectively) were switched individually to mutate the AGGAC motif and retain the hairpin-stem stability. Their exact sequences were shown in the Supplemental Experimental Procedures. **C**) RNA pull down results showing that the destabilized hairpins, by substituting A at the 2,515 site with G/C/U, bind HNRNPG stronger. Data are mean ± standard deviation (s.d.); n=3, biological replicates. **D**) The C-terminal RNA binding domain of HNRNPG binds stronger with destabilized hairpins than the stable ones.

4.2.2 m^{6}A alters RNA structure to expose HNRNPG binding consensus motif

To identify how m^{6}A increases HNRNPG binding at the MALAT1 hairpin, we first performed RNA structural mapping assay for the MALAT1 hairpin to detect the m^{6}A effect on the hairpin structure change. We observed that the 2,515 m^{6}A residue destabilizes the hairpin

stem (**Figure 4.4A**), and the structural change is especially obvious at the RNA region adjacent to m^6A. In the presence of m^6A at the 2,515 site, there are significantly increased nuclease S1 digestion (single-strand specific) and markedly decreased RNase V1 digestion (double-strand/stacking specific) around the AGGAC motif, indicating that m^6A destabilizes the hairpin stem structure to increase the accessibility of the specific binding sequence of HNRNPG.

To validate it is the m^6A-induced hairpin-stem destabilization that increases HNRNPG binding, we synthesized new RNA oligos, substituting the A residue at the 2,515 site with the guanosine, cytosine and uridine respectively. We observed that the non-watson-crick base-pairing between uridine and the substituted nucleosides at the 2,515 site can destabilize the thermodynamic stability of the hairpin-stem, and significantly increase HNRNPG binding activities compared to the native non-methylated hairpin (**Figure 4.4B**). This result indicates it is the m^6A-induced hairpin-stem destabilization that increased HNRNPG binding, rather than the 2,515 m^6A itself. Besides, our RNA pull down experiments with hairpin mutants showed that mutation at individual nucleosides within the 'AGG' portion of HNRNPG specific binding sequence greatly decreased HNRNPG binding, while mutations at the 'AC' portion showed modest reduction in HNRNPG binding (**Figure 4.4C**). These results suggest that the 'AGG' motif is the HNRNPG direct recognition sites and the neighboring m^6A residue functions as RNA structure remodeler to increase its accessibility for HNRNPG recruitment.

This result validates the 'm^6A-switch' mechanism in regulating RNA-protein interactions. Therefore, HNRNPG is another RNA binding protein regulated by m^6A-switches. Besides, we show that the C-ter RBD of HNRNPG shows stronger binding activities with destabilized hairpin-stems than the stable hairpins (**Figure 4.4D**), validating that HNRNPG is regulated through m^6A-switches through the C-ter RBD.

4.2.3 Transcriptome-wide identification of m^{6}A-switches regulating HNRNPG Binding

To identify m^{6}A-switches regulating HNRNPG recognition across the transcriptome, we performed anti-HNRNPG PAR-CLIP–MeRIP experiments in HEK293T cells, which combined PAR-CLIP and Me-RIP experiments to identify methylated RNA binding sites in large scale[15]. We first performed PAR-CLIP experiments to isolate all HNRNPG-bound RNA fragments (input), then the anti-m^{6}A immunoprecipitation (MeRIP) was applied to enrich m^{6}A-containing HNRNPG bound RNA fragments (IP). Both the input and IP samples from two biological replicates were sent for RNA sequencing. In total, the PAR-CLIP–MeRIP approach identified 16,200 m^{6}A sites at HNRNPG binding sites ($\log_2$(IP/Input) > 0.5), including the m^{6}A at the MALAT1 2,515 site (**Figure 4.5A**).

To test whether the global m^{6}A reduction is able to decrease HNRNPG binding at their specific binding sequences, we knocked down *METTL3* and *METTL14*, the two m^{6}A writers known so far, and performed anti-HNRNPG PAR-CLIP experiments to measure the differential binding occupancies at their target AGRAC sequences. The global m^{6}A reduction has a dramatic effect on HNRNPG binding. In total, we identified 67,229 AGRAC sequences within HNRNPG PAR-CLIP peaks that show decreased HNRNPG binding strength upon *METTL3* and *METTL14* depletion. Among them, 37,750 sites were co-modulated by METTL3 and METTL14, which is consistent with previous reports that the two m^{6}A-writers work as a complex and have a significant amount of shared target sites. Remarkably, 13,161 methylated AGRAC motifs identified by PAR-CLIP–MeRIP experiments show decreased HNRNPG binding upon m^{6}A writers depletion (**Figure 4.5B**). These 13,161 methylated AGRAC motifs are termed high-confidence m^{6}A-switches for HNRNPG recruitment.

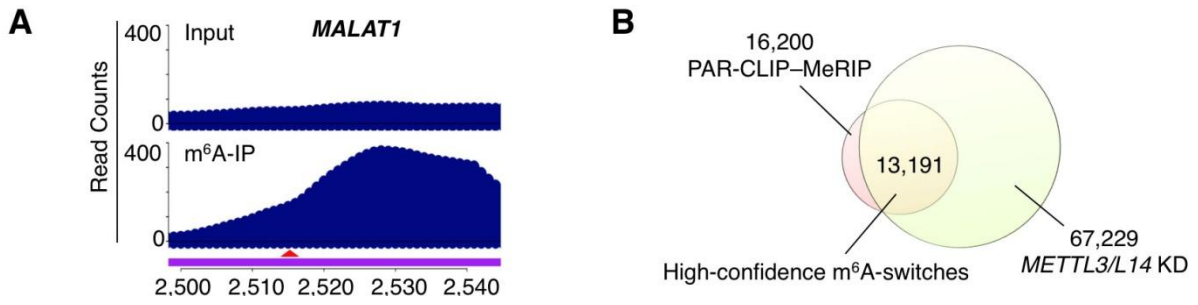


Figure 4.5. Transcriptome-wide identification of m^{6}A-switches recruiting HNRNPG. A) PAR-CLIP–MeRIP identified an m^{6}A residue around the MALAT1 2,515 site. **B**) Identification of high-confidence HNRNPG m^{6}A-switches through PAR-CLIP–MeRIP and PAR-CLIP experiments in m^{6}A-writers depleted HEK293T cells.

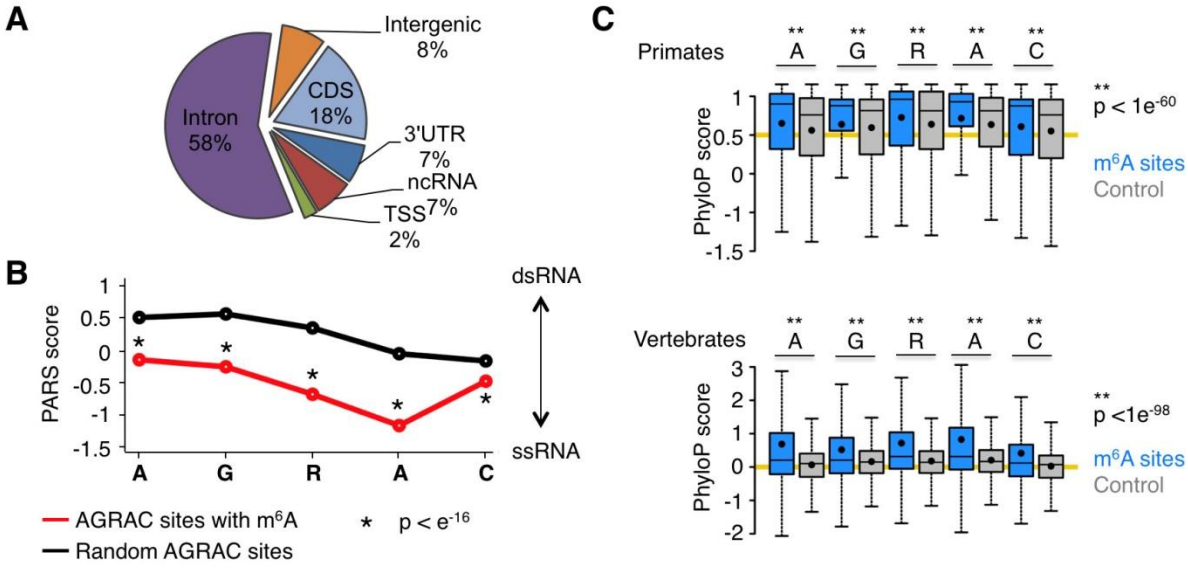
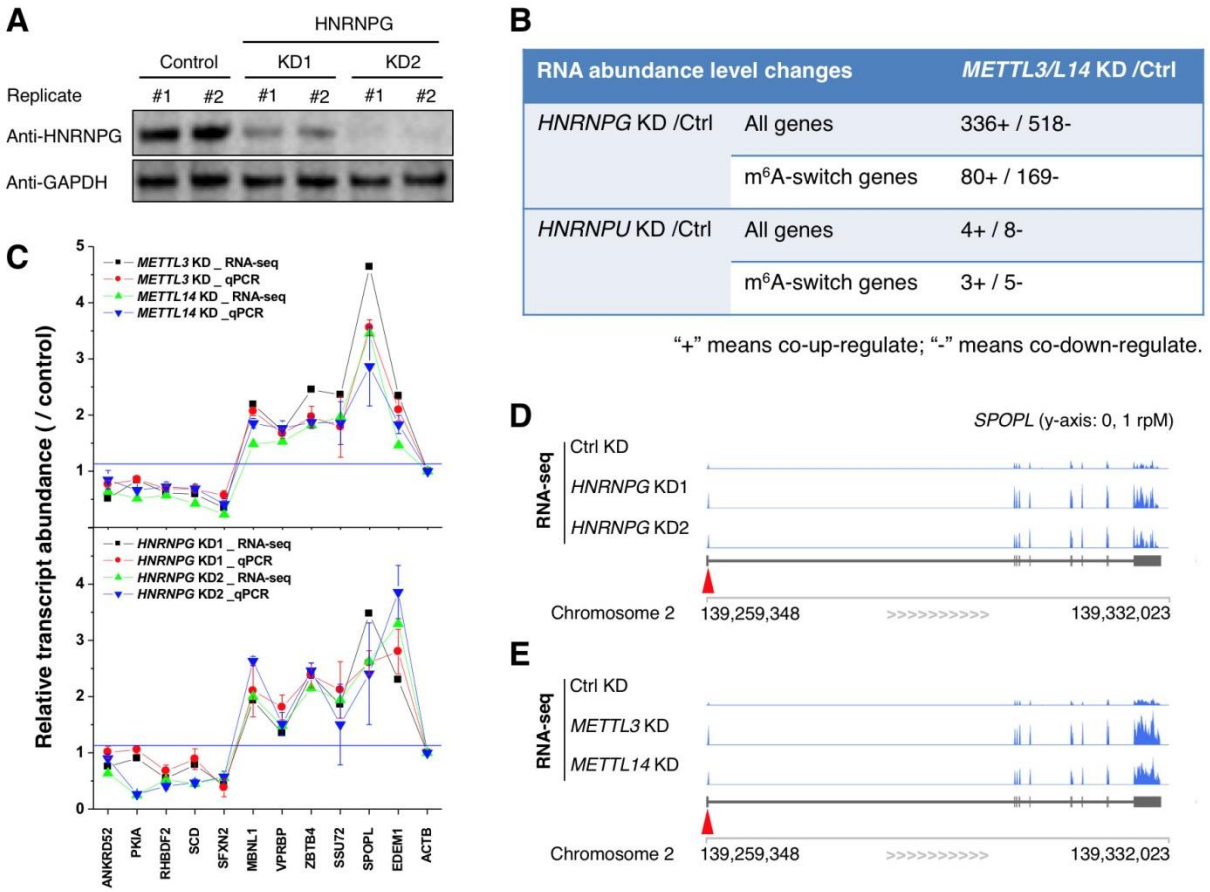


Figure 4.6. Properties of high-confidence HNRNPG m^{6}A-switches. A) Pie chart depicting the region distribution of high-confidence m^{6}A-swith sites. **B**) Cumulative distribution of high-confidence m^{6}A-switches (red) and control (black) regarding the S1/V1 cleavage preference. **C**) Phylogenetic conservation of high-confidence m^{6}A-switches among primates (upper panel) and vertebrates (lower panel).

We find that HNRNPG m^{6}A-switches are mostly located in introns (**Figure 4.6A**), suggesting their role in RNA splicing regulation. Besides, transcriptome-wide nuclease mapping



RNA abundance level changes		*METTL3/L14* KD /Ctrl
HNRNPG KD /Ctrl	All genes	336+ / 518-
	m^{6}A-switch genes	80+ / 169-
HNRNPU KD /Ctrl	All genes	4+ / 8-
	m^{6}A-switch genes	3+ / 5-

"+" means co-up-regulate; "-" means co-down-regulate.

Figure 4.7. HNRNPG and m^{6}A-writers co-regulate the expression of m^{6}A-switch-containing genes. A) HNRNPG was knockdown (KD) by two independent siRNAs (KD1 and KD2), confirmed by western blots. **B**) HNRNPG loss and METTL3/METTL14 depletion co-regulated the expression of many genes. Gene expression changes between control (Ctrl) and *HNRNPG*, *METTL3*/*METTL14*, *HNRNPU* knockdown HEK293T cells were analyzed by Cuffdiff2[46], and the absolute numbers of differentially expressed genes are shown. m^{6}A-switch genes refer to the 5,228 genes containing high-confidence m^{6}A-switches. The RNA-seq data from HNRNPU knockdown HEK293T cells (Gene Expression Omnibus accession GEO34995 data set[47]) were analyzed for comparison with a different mRNA-binding protein. HNRNPU did preferentially bind the 2,515-m^{6}A methylated MALAT1 hairpin. **C**) m^{6}A-writers depletion (upper panel) and HNRNPG loss (lower panel) cause similar changes in m^{6}A-switch containing transcripts, validated by RNA-seq and qPCR. **D**) The mRNA-seq peaks showing HNRNPG loss decreases the abundance of *SPOPL* transcripts. **E**) The mRNA-seq peaks showing *METTL3/L14* depletion decreases the abundance of *SPOPL* transcripts.

results indicate that AGRAC motifs regulated by m^{6}A-switches are more single-stranded/accessible than the AGRAC motifs in random genomic RNA regions (**Figure 4.6B**),

thus supporting our m^{6}A-switch model. Furthermore, HNRNPG m^{6}A-switches are significantly more conserved across species than random genomic regions (**Figure 4.6C**), indicating their physiological importance.

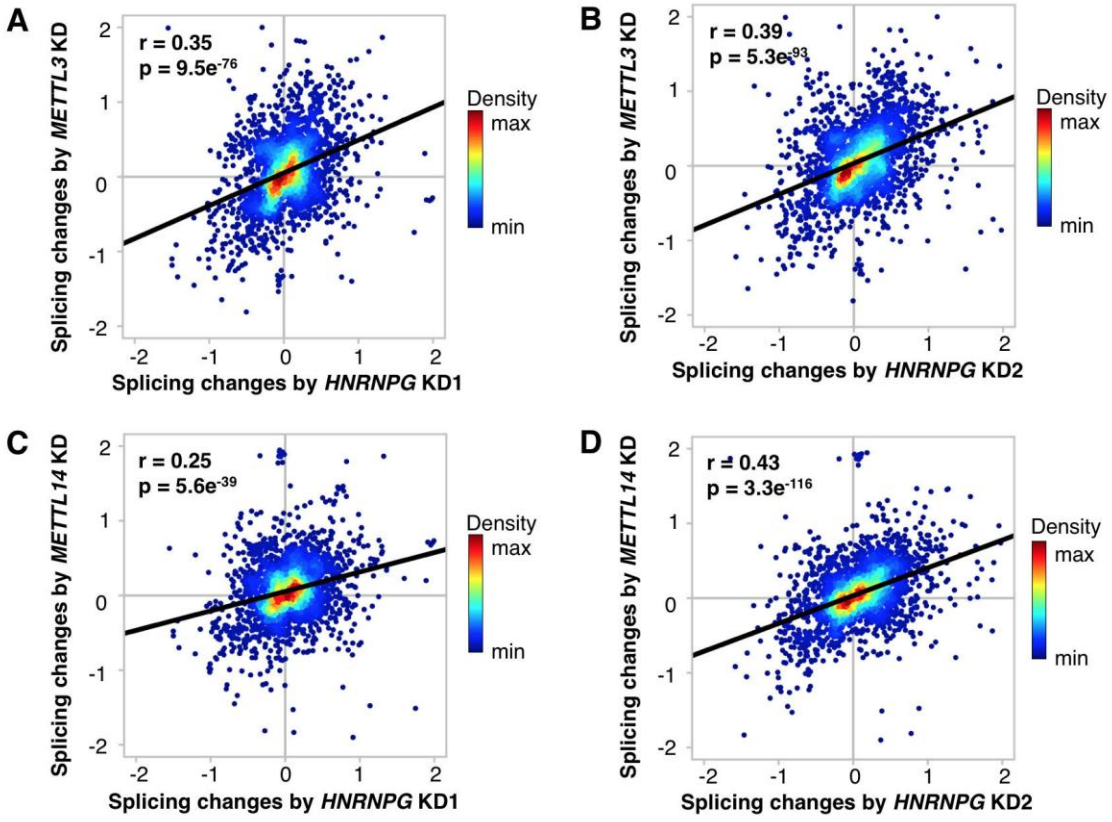


Figure 4.8. HNRNPG loss and m^{6}A writers depletion cause similar alternative splicing changes on m^{6}A-switch-containing transcripts. (**A-D**) Correlation between differentially-expressed exons with splicing changes following HNRNPG KD2 and m^{6}A-writers depletion. Annotated significantly differentially expressed exons were quantified in HNRNPG KD1 (**A, C**), HNRNPG KD2 (**B, D**) *METTL3* KD (**A, B**) and *METTL14* KD (**C, D**) HEK293T cells relative to control cells. Spearman correlations and p-values are shown within each panel.

4.2.4 m^{6}A-Selective HNRNPG Binding Events Regulate Alternative Splicing

HNRNPG has been known to affect transcription and alternative splicing on many genes. To figure out whether the m^{6}A-switch regulated HNRNPG binding events influence gene

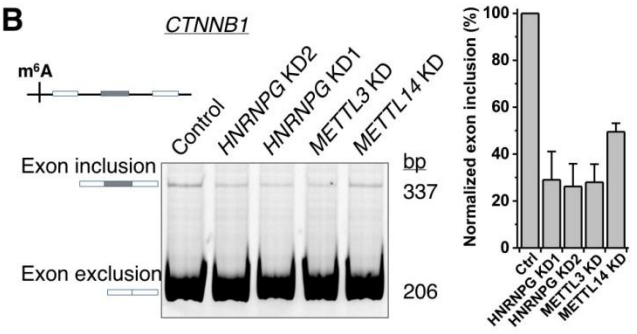
expression, we performed mRNA-seq experiments using HNRNPG knockdown and control HEK293T cells. Independent HNRNPG knockdowns were achieved with two different siRNAs (KD1 and KD2) at high efficiency (**Figure 4.7A**). We also analyzed our recently published mRNA-seq data using m^{6}A-writers (METTL3 and METTL14) depleted and control HEK293T cells[15]. Since the global m^{6}A reduction decreases HNRNPG binding at m^{6}A-switch sites, m^{6}A writers depletion should have similar effect on the m^{6}A-containing RNA transcripts as HNRNPG depletion.

A

No. of exons by HNRNPG m^{6}A-switch	*METTL3* KD /Ctrl	*METTL14* KD /Ctrl
HNRNPG KD1 / Ctrl	360 + / 495 -	244 + / 409 -
HNRNPG KD2 / Ctrl	412 + / 623 -	368 + / 608 -

"+" means co-up-regulate; "-" means co-down-regulate.

B



D

No. of co-regulated exons by HNRNPC	*METTL3* KD /Ctrl	*METTL14* KD /Ctrl
HNRNPG KD1 / Ctrl	7+ / 2-	6+ / 4-
HNRNPG KD2 / Ctrl	0+ / 0-	0+ / 0-

"+" means co-up-regulate; "-" means co-down-regulate.

C

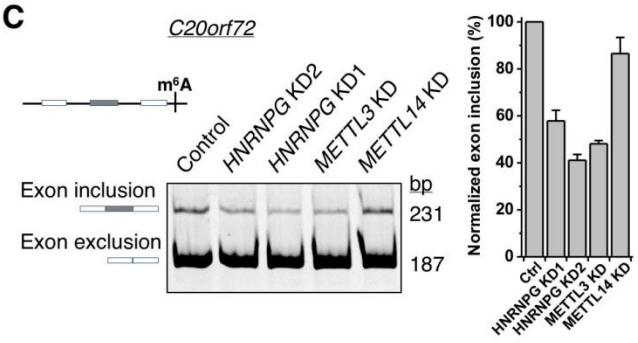


Figure 4.9. m^{6}A-switches recruit HNRNPG for alternative splicing regulation. A) The table listing the absolute number of co-regulated exons by HNRNP depletion and METTL3/L14 KD within m^{6}A-switch-containing genes. (**B-C**) PCR with reverse transcription (RT-PCR) results validating the m^{6}A-regulated exon inclusion in *CTNNB1* gene (**B**) and *C20orf72* gene (**C**). **D**) HNRNPG m^{6}A-switches and HNRNPC m^{6}A-switches co-regulate few splicing exons. The absolute numbers of differentially expressed exons are shown.

We first compared the RNA transcripts expression changes by Cuffdiff2. We find that m^{6}A writers depletion co-regulate the abundance of hundreds of m^{6}A-containing RNA transcripts with HNRNPG knockdown; in comparison, m^{6}A writers depletion co-regulate the expression of few genes with knockdown of another mRNA-binding protein, HNRNPU (**Figure 4.7B**), which did not show preferential interactions with the methylated MALAT1 hairpin (**Figures 4.1C-D**). The m^{6}A-dependent HNRNPG function in RNA abundance regulation, as validated by quantitative RT-PCR (**Figures 4.7C-E**), suggests the regulatory role of HNRNPG in gene expression.

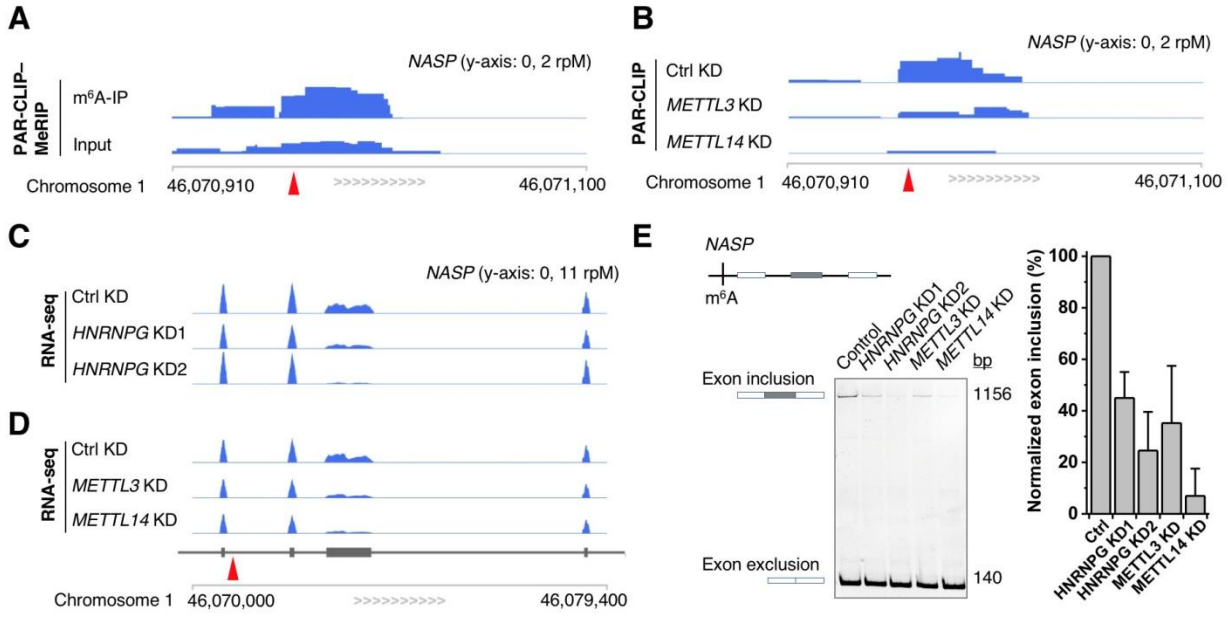


Figure 4.10. The m^{6}A-selective HNRNPG binding regulates alternative splicing. A) HNRNPG-based PAR-CLIP–MeRIP peak indicates the m^{6}A presence. **B**) HNRNPG-based PAR-CLIP peaks showing the m^{6}A depletion decreases HNRNPG binding. **C**) The mRNA-seq peaks showing HNRNPG loss decrease one exon inclusion in the *NASP* gene. **D**) The mRNA-seq peaks showing m^{6}A-writers depletion decrease one exon inclusion in the *NASP* gene. **E**) Semi-quantitative PCR with reverse transcription (RT-PCR) validating the HNRNPG m^{6}A-switch regulated exon inclusion.

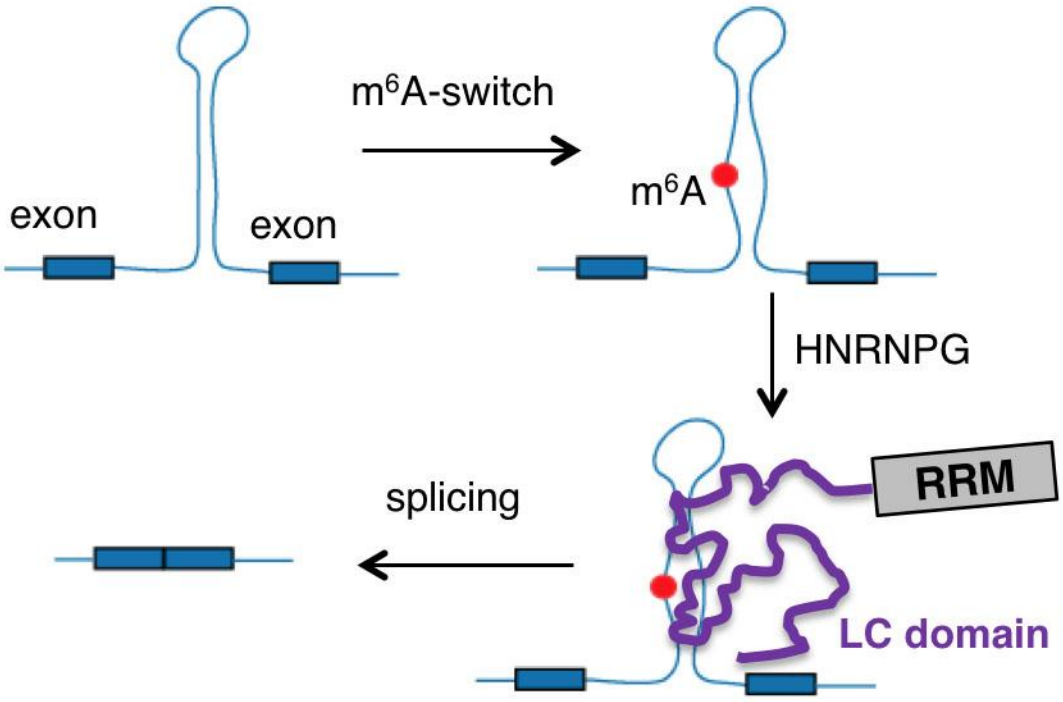


Figure 4.11. Proposed model: m^{6}A-switches recruit HNRNPG via its C-terminal low-complexity domain to perform alternative splicing steps.

We then identified changes in splicing patterns using the DEXSeq software[48], and we observed a good correlation in splicing changes between HNRNPG depletion and m^{6}A writers depletion (**Figures 4.8A-D**), indicating that m^{6}A writers and HNRNPG co-regulate alternative splicing on m^{6}A-containing RNA transcripts. In total, *HNRNPG* KD1 co-up/down-regulate the differential usage of 360/495 exons with *METTL3* KD and co-up/down-regulate 244/409 exons with *METTL14* KD; *HNRNPG* KD2 co-up/down-regulate 412/623 exons with *METTL3* KD and co-up/down-regulate 368/608 exons with *METTL14* KD (**Figure 4.9A**). We confirmed the alternative inclusion of several exons by RT-PCR experiments (**Figures 4.9B-C and 4.10**). Take the *NASP* gene for example, the HNRNPG PAR-CLIP–MeRIP data indicates the m^{6}A presence

at the AGGAC sequence (**Figure 4.10A**). And *METTL3* and *METTL14* KD both decreased HNRNPG binding at the AGGAC sequence (**Figure 4.10B**). The mRNA-seq data shows that the reduced HNRNPG binding due to m^6A writers depletion decreased the inclusion level of one nearby exon, to similar extent as *HNRNPG* knockdown, as validated by RT-PCR (**Figures 4.10C-E**). Taken together, these results indicate that the m^6A-switch-dependent HNRNPG binding events influence the alternative splicing events (**Figure 4.11**).

Comparing with the exons regulated by HNRNPC m^6A-switches as previously reported[15], the number of exons regulated by HNRNPG m^6A-switches is about four times more than the exons regulated by HNRNPC m^6A-switches. Thus, HNRNPG m^6A-switches affect more splicing events than HNRNPC m^6A-switches. But we found few exons that are co-regulated by both m^6A-switches (**Figure 4.9D**), suggesting that HNRNPC and HNRNPG m^6A-switches regulate separate subsets of exons. These findings depict the exclusive alternative splicing regulation networks by HNRNPG and HNRNPC m^6A-switches.

4.2.5 Conclusion and discussion

Our findings reveal that m^6A-switches regulate HNRNPG binding for pre-mRNA processing events. The m^6A consensus motif is embedded within HNRNPG specific binding sequences, and m^6A methylation destabilizes structured RNA regions to increase their accessibility for HNRNPG recruitment. Quite recently, Alarcon *et al.* reported HNRNPA2B1 is also m^6A reader in the nucleus responsible for m^6A-dependent RNA processing. Alarcon indicated that HNRNPA2B1 binds m^6A methylated sequences, but it remains unclear how HNRNPA2B1 recognizes m^6A methylation. Considering that HNRNPA2B1 does not contain the YTH-domain, it is very likely that HNRNPA2B1 binding is also mediated by m^6A-switches.

Together with our previously reported HNRNPC m^6A-switches, these results suggest that m^6A-switches may emerge as generalized mechanisms for regulating RNA-protein interactions in the nucleus, and many other nuclear pre-mRNA processing enzymes could be regulated by m^6A-switches. Figuring out the detailed mechanisms regarding the m^6A-modualted RNA-protein interactions would help decipher the intricate complicated cellular RNA-protein networks for gene regulation.

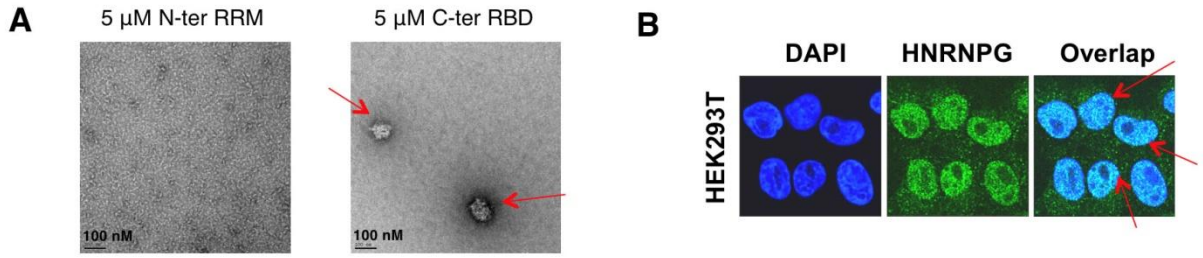


Figure 4.12. HNRNPG exist in the form of granules *in vitro* and *in vivo*. A) Electron microscopy images of the C-ter RBD and N-ter RRM of HNRNPG at 5 μM concentration. The C-terminal RBD forms aggregates, as marked by the red arrows. **B**) Immunostaining results showing the cellular location of HNRNPG proteins in HEK293T cells. The cellular puncta containing HNRNPG were labeled with red arrows.

Our findings also help elucidate the obscure RNA recognition modes for HNRNPG. Compared to the well-characterized N-terminal RRM, the C-terminal RBD has been studied less. The C-ter RBD is located within low-complexity domain, which is composed by biased amino acids. The C-ter RBD has been shown to bind the A/G-rich motif in one hairpin structure[44]. Here, we show that the RNA recognition mode of the C-ter RBD is regulated by m^6A-switches. However, we do not exclude the possibility that other regions of HNRNPG cooperate the selective recognition for the m^6A mark. The multiple RNA recognition modes of HNRNPG indicate that there are many other HNRNPG binding events and corresponding downstream

functions that are independent of m^{6}A modification. It remains an interesting question how these distinct RNA binding domains act in concert to regulate specific splicing events.

Our findings also reveal the association between m^{6}A readers and low-complexity domains. Low-complexity domains are known to promote the formation of RNA granules, which are membrane-free structures composed of RNA molecules and RNA-binding proteins (RBPs). Indeed, both previous data[42,44] and our observations (**Figures 4.12A-B**) support that HNRNPG forms RNA granules *in vitro* and *in vivo*. However, it remains unclear what other protein factors reside in the same RNA granules with HNRNPG. Actually, two previously identified direct m^{6}A 'readers' (YTHDF1 and YTHDF2) also contain the low-complexity domain (QN-rich sequences) at the their C-terminus, and the QN-rich sequences recruit YTHDF1/2 to RNA granules (stress granules and P-bodies respectively) for gene regulation[10,11]. Interestingly, other m^{6}A 'readers' known so far (HNRNPC and HNRNPA2B1) are also identified in RNA granules formed *in vitro*[43]. The prevalence of LC domains in m^{6}A readers suggest that the cellular locations of m^{6}A-readers and their bound mRNAs are likely controlled by some RNA granules. It remains unclear how LC domains contribute to the granule formation, but the phosphorylation status of LC domains has been shown to control the assembly/disassembly of RNA granules, thus adding another layer of control to the m^{6}A readers' cellular location.

In summary, we illustrate how m^{6}A alters RNA structure to recruit HNRNPG binding for alternative splicing regulation. These findings improve our understanding of the intricate RNA-protein interactions for pre-mRNA processing steps, and have great implications for the m^{6}A-dependent cellular biology.

4.3 Experimental section

4.3.1 Mammalian cell culture, siRNA knockdown and Western blot

Human embryonic kidney cell line HEK293T (CRL-11268) were obtained from American Type Culture Collection (ATCC) and were cultured under standard conditions. Control siRNA (1027281, Qiagen), METTL3 siRNA (SI04317096, Qiagen), METTL14 siRNA (SI04317096, Qiagen) or HNRNPG siRNA (SI00700084 and SI00700077, Qiagen) were transfected into HEK293T cells at a concentration of 40 nM using lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's instructions. Cells were collected 48 hours after the transfection, shock-frozen in liquid nitrogen, and stored at -80 °C for further studies. Western blot analysis using METTL3- (HPA038002, Sigma), METTL14- (HPA038002, Sigma), HNRNPG- (sc-14581, Santa Cruz), GAPDH- (A00192-40, Genescript) specific antibodies was performed under standard procedures. All synthetic oligos were synthesized by Q.D.

4.3.2 Gel shift and RNA pull down assays

HeLa nuclear extracts were isolated using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (78833, Thermo Scientific) according to the manufacturer's instructions. The purified radioactively-labeled RNA oligos were refolded by heating at 90 °C for 1 min, then 30 °C for 5 min. 3 µl HeLa nuclear extract and 6 µl refolded RNA were incubated at room temperature (RT) for 30 min and then at 4 °C for 2 hrs. Each sample was mixed with 1 µl 50% glycerol, separated on the 8% native 1x TBE gel, and visualized by phosphorimaging using the Personal Molecular Imager (Bio-Rad).

The *in vitro* pull down assay was performed as described [6]. The eluted protein samples were separated on 4-12% polyacrylamide Bis-Tris gels (NP0321BOX, Invitrogen) and stained with SYPRO-Ruby (S12000, Invitrogen) according to the manufacturer's instructions. Protein in

gel slices or the entire pulled down protein samples were digested with trypsin and identified using LC-MS/MS by the Donald Danforth Plant Science Center (Washington University, St. Louis. MO).

The RNA oligos used in **Figures 4.1 B-D**:

2,515-A: 5'- AAUGUGAAGGACUUUCGUAACGGAAGUAAUUCAA-Biotin;

2,515-m^{6}A: 5'- AAUGUGAAGGm^{6}ACUUUCGUAACGGAAGUAAUUCAA-Biotin.

The RNA oligos used in **Figure 4.4B**:

2,515-G: 5'- AAUGUGAAGGGCUUUCGUAACGGAAGUAAUUCAA-Biotin;

2,515-C: 5'- AAUGUGAAGGCCUUUCGUAACGGAAGUAAUUCAA-Biotin;

2,515-U: 5'- AAUGUGAAGGUCUUUCGUAACGGAAGUAAUUCAA-Biotin.

The RNA oligos used in **Figure 4.4C**:

#1: 5'-AAUGUGAUGGm^{6}ACUUUCGUAACGGAAGUAAAUCAA-Biotin;

#2: 5'-AAUGUGAAAGm^{6}ACUUUCGUAACGGAAGUAGUUCAA-Biotin;

#3: 5'-AAUGUGAAGAm^{6}ACUUUCGUAACGGAAGUGAUUCAA-Biotin;

#4: 5'-AAUGUGAAGGUCUUUCGUAACGGAAGm^{6}AAAUUCAA-Biotin;

#5: 5'-AAUGUGAAGGm^{6}AGUUUCGUAACGGAACUAAUUCAA-Biotin.

4.3.3 Protein expression and *in vitro* UV cross-linking assays

For expression of the N-terminal RNA Recognition Motif (RRM, residues 1-83) and C-terminal RNA binding domain (RBD, residues 294-391) of human HNRNPG proteins, fragments encoding HNRNPG domains were amplified using PCR from human HeLa cDNA libraries (637203, Clontech) and subcloned into pGEX-6p-1 expression vectors using BamHI and XhoI restriction sites. Plasmid DNA was transformed into E. coli BL21-CodonPlus(DE3)-RP cells

(Stratagene), and the GST-fusion proteins were expressed and purified as described [49]. The *in vitro* UV crosslinking assay were performed as previously described [49].

The RNA oligos used in **Figure 4.4D**:

2,518-U&2,528-A: 5'-AAUGUGAAGGACUUUCGUAACGGAAGUAAUUCAA;

2,518-A&2,528-A: 5'-AAUGUGAAGGACUAUCGUAACGGAAGUAAUUCAA;

2,518-U&2,528-U: 5'-AAUGUGAAGGACUUUCGUAACGGUAGUAAUUCAA;

2,518-A&2,528-U: 5'-AAUGUGAAGGACUAUCGUAACGGUAGUAAUUCAA.

4.3.4 RNA structural probing

The synthetic RNA oligos were 5′ end-labeled with γ-^{32}P-ATP by T4 PNK (70031, Affymetrix), gel purified, and re-folded. Structural probing assay with RNase T1, nuclease S1 and RNase V1 was performed as previously described [50].

4.3.5 PARCLIP and PARCLIP-MeRIP

PAR-CLIP procedures were performed as previously reported [45] with the following modification. HEK293T cells in 15 cm plates treated following normal PAR-CLIP procedures were lysed and digested with a combination of RNase I (Ambion, AM2295, 12 μl 1/50 diluted with H_2O) and Turbo DNases (2 μl) for 3 mins at 37 °C, shaking at 1,100 rpm. The lysate was then immediately cleared by spinning at 14,000 rpm, 4 °C for 30 min, and placed on ice for further use. HNRNPG binding sites were identified by PARalyzer v1.1[51] with default settings.

PAR-CLIP–MeRIP experiment applied m^{6}A-antibody immunoprecipitation[52,53,54] to the HNRNPG PAR-CLIP RNA samples (from HEK293T cells in eight 15-cm plates). The HNRNPG PAR-CLIP RNA sample was incubated with m^{6}A-specific antibody (202003, SYSY),

RNase inhibitor (80 units, Sigma-Aldrich), human placental RNase inhibitor (NEB) in 200 μl 1x IP buffer (50 mM Tris-HCl pH 7.4, 750 mM NaCl and 0.5% (vol/vol) Igepal CA-630) at 4 °C for 2 hours under gentle shaking conditions. For each PAR-CLIP–MeRIP experiment, 20 μl protein-A beads (Invitrogen) were washed twice with 1 ml 1x IP buffer, blocked with 2 hours incubation with 100 μl 1× IP buffer supplemented with BSA (0.5 mg/ml), RNasin and Human placental RNase inhibitor, and then washed twice with 100 μl 1x IP buffer. The pre-blocked protein-A beads were then combined with the prepared immuno-reaction mixture and incubated at 4 °C for 2 hours, followed by three washes with 100 μl 1× IP buffer. After that, the RNA was eluted by 1 hour incubation with 20 μl elution buffer (1× IP buffer and 6.7 mM m^6A, Sigma-Aldrich) under gentle shaking conditions, and purified by ethanol precipitation. The purified RNA sample (IP) as well as the input PAR-CLIP RNA sample (Input control) were used for library construction by Truseq small RNA sample preparation kit (Illumina).

Libraries for both PAR-CLIP and PAR-CLIP–MeRIP experiments were prepared using TruSeq Small RNA Sample Preparation Kit (RS-200-0012, Illumina) according to the manufacturer's instructions, and then sequenced by Illumina Hiseq2000 with single end 50-bp read length. The control and IP samples from PAR-CLIP–MeRIP experiments (same case for the control and knockdown (KD) samples from *METTL3/L14* KD experiments) were sequenced together in one flowcell on two lanes, and the reads from two lanes of each sample were combined for remaining analysis. The raw seq data was trimmed using the Trimmomatic computer program version 0.30 [55] to remove adaptor sequences, and mapped to the Human genome version hg19 by Bowtie 1.0.0[56] without any gaps and allowed for at most two mismatches.

4.3.6 Detection and distribution analysis of m^6A sites within HNRNPG binding sites

Detection of m^6A sites by the PAR-CLIP–MeRIP sample involves comparing the read counts of the IP sample with that of the control (Ctrl) sample as follows: (1) we identified all AGRAC motifs within HNRNPG PAR-CLIP peaks; (2) we performed transcriptome-wide scanning to compare read counts of each AGRAC motif in (1) from both Control and IP samples to calculate the fold change score, score = $\log_2(Counts_{IP}/Counts_{Control})$. m^6A sites were considered as AGRAC motifs with $\log_2(Counts_{IP}/Counts_{Control})$ larger than 0.5.

Detection of m^6A sites by the PAR-CLIP sample involves comparing the read counts of the IP sample with that of the control (Ctrl) sample as follows: (1) we identified all AGRAC motifs within HNRNPG PAR-CLIP peaks; (2) we performed transcriptome-wide scanning to compare read counts of each AGRAC motif in (1) from both Control and METTL3/L14 KD samples to calculate the fold change score, score = $\log_2(Counts_{KD}/Counts_{Control})$. m^6A sites were considered as AGRAC motifs with $\log_2(Counts_{IP}/Counts_{Control})$ less than 0.5.

The high-confidence m^6A sites (HC m^6A) within HNRNPG binding sites fulfill the following two requirements: (1) $\log_2(Counts_{IP}/Counts_{Control}) > 0.5$; (2) $\log_2(Counts_{IP}/Counts_{Control}) < 0.5$. Pie charts illustrating distribution of HC m^6A sites within each segment were made using the following hierarchy: intron > ncRNA > 3′UTR > 5′UTR > CDS > intergenic.

4.3.7 RNA-sequencing

RNA-seq experiments were performed on two replicate RNA samples from *HNRNPG* KD (KD1 and KD2) as well as control HEK293T cells (48 hours after transfection). Total RNA samples were extracted according to RNeasy plus kit (Catalog # 74104, Qiagen). Libraries were prepared according to the TruSeq Stranded mRNA LT Sample Prep Kit (Catalog # RS-122-9005DOC). KD and control samples were sequenced together in one flowcell on four lanes,

respectively. All samples were sequenced by illumina Hiseq 2000 with pair end 100-bp read length. The reads from the four lanes of each sample were combined for all analysis. The RNA-seq data was mapped using the splice-aware alignment algorithm TopHat version 1.1.4 [57] based on the following parameters: tophat –num-threads 8 –mate-inner-dist 200 –solexa-quals –min-isoform-fraction 0 –coverage-search-segment-mismatches 1. Gene expression level changes were analyzed using cuffdiff [46]. Differential splicing was determined using DEXSeq [48] based on Cufflinks-predicted, nonoverlapping exons.

4.3.8 Evolutionary conservation, graphic and statistical analysis

Phylogenetic conservation analysis was performed by comparing PhyloP scores at the m^6A-located AGRAC motifs to those of randomly selected sequences. The PhyloP scores were accessed from the precompiled phyloP scores[58] (ftp://hgdownload.soe.ucsc.edu/goldenPath/hg19/phyloP46way/) under both primates and vertebrates categories. P-values were evaluated using the Mann-Whitney-Wilcoxon test. The random selection was done separately for primates and for vertebrates.

Sequence logos were generated using the WebLogo package. R statistical package was used for all statistical analysis (unless stated otherwise).

4.3.9 RT-PCR quantitation

Total RNA samples were extracted from HEK293T cells and reverse transcribed using SuperScript® III First-Strand Synthesis System (Life Technologies, #18080-051). In order to validate the splicing changes identified from our RNA-seq data, we performed RT-PCR measurements using Thermo Scientific™ Taq™ DNA Polymerase under the following conditions: 95 °C for 3 mins, 30 cycles of [95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min] and

then finally 72 °C for 10 min. For the target alternate exon, we designed and used primers annealing to both neighboring constitutive exons. The PCR products were separated on 10% 1X TBE gel and SYBR-gold stained. In order to validate the gene expression level changes identified from our RNA-seq data, we performed qRT-PCR measurements using Power SYBR® Green PCR Master Mix (Life Technology, # 4367659) under the following conditions: 50 °C for 3 mins followed by 95 °C for 10 mins, 40 cycles of [95 °C for 15 s, 60 °C for 1 min] and then 40 °C for 1 min and 95 °C for 15 s and finally 60 °C for 30 s.

The primer sequences are listed as below (Gene name: forward primer; reverse primer):

C20orf72: ACAGCGGATGATTCTGGAAC; TTCCTGGGGTGAAAGTATGC;
NASP: TGTGCATGTGGAAGAGGAAG; GAAGGTGTGCATGTGGAAGA;
CTNNB1: GAAAATCCAGCGTGGACAAT; CAGGACTTGGGAGGTATCCA;
PKIA (primers_1): CCTGGTTTCCCCAAAGAAGT; TGATTGGAAACCTTCTTGTCTTT;
PKIA (primers_2): TGGTAGCAATGACTGATGTGG; ACTTGCAGAGGAAACCAGGA
RHBDF2 (primers_1): AGAGCCAGAGACCCAAGACA; CCCAAGACTCAGAGAGGCA;
RHBDF2 (primers_2): GAGTACCCAGGAAGCTGCAC; TACAGATGCTCCGGTGTCAA;
SCD (primers_1): TGTTCGTTGCCACTTTCTTG; GGGGGCTAATGTTCTTGTCA;
SCD (primers_2): CTCCACTGCTGGACATGAGA; AATGAGTGAAGGGGCACAAC;
SFXN2 (primers_1): GCCAGACTGGTCTCGAACTC; ACGGTCCCCTTTTTAGCACT;
SFXN2 (primers_2): CCTGGGATTGGTCGAAAAG; AAATGCCACCAGTTACAGCC;
ANKRD52 (primers_1): CTGTGCCGAGACTTTAAGGG; GCGAGTATCCGCTGTAATCC;
ANKRD52 (primers_2): AGACGCTGGTGAATCTGGAC; GCTGTAAGCACCTCCACACA;
MBNL1 (primers_1): AATATCTTCATCCACCCCCA; TTGGCTAGTTGCATTTGCTG;
MBNL1 (primers_2): GCTGCATCTGTCTATGCCAA; CGAATTTCCAAGCTGCTTTC;
VPRBP (primers_1): GCTGACAAAAGAGGCTGACC; GCTGAGGATGAGCAGTAGGG;
VPRBP (primers_2): TGATAGAATATGGCCCAGCG; CCAATTGCAGGCAATAGAAA;

ZBTB4 (primers_1): TTCCATGCCTTTGGATCTTC; ATTTGGGGGTCAAGATAGGG;

ZBTB4 (primers_2): GCTCACTTCAGCCCCACTAC; AGACGAGGAAGAGGAGGAGG;

SSU72 (primers_1): GCACTTCCCGACATACCTGT; GCACAATGACAGCAGCATCT;

SSU72 (primers_2): AAATAAGAGAATCAAGCCCCG; TTCCACCACCTGGTCATACA;

SPOPL (primers_1): GCTGGAGTCGTAACTCGGAA; CGCTCCTAAACTTCTTCCCC;

SPOPL (primers_2): GGAGGTTTGTTCTGGTGCAT; GCCCTTAAGAAGCACACTGG;

EDEM1 (primers_1): AGCCTCCTTTCTGCTCACAG; GGTGTTTTCAAAAGCAGGGA;

EDEM1 (primers_2): ATGAGCATCTTCGGGAATTG; AACTCATGAGGTTTCGGCCT.

4.4 References

1. Nilsen, T.W. & Graveley, B.R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457-63 (2010).
2. Beyer, A.L., Christensen, M.E., Walker, B.W. & LeStourgeon, W.M. Identification and characterization of the packaging proteins of core 40S hnRNP particles. *Cell* **11**, 127-38 (1977).
3. Krecic, A.M. & Swanson, M.S. hnRNP complexes: composition, structure, and function. in *Curr Opin Cell Biol*, **11**, 363-71 (1999).
4. Bokar, J.A. The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. *Fine-tuning of RNA functions by modification and editing* (ed. Grosjean, H.) 141-178 (Springer-Verlag, Berlin, Heidelberg, New York, 2005).
5. Desrosiers, R., Friderici, K. & Rottman, F. Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells. *Proceedings of the National Academy of Sciences* **71**, 3971-3975 (1974).
6. Dominissini, D. et al. Topology of the human and mouse m^6A RNA methylomes revealed by m^6A-seq. *Nature* **485**, 201-206 (2012).
7. Meyer, K. et al. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. *Cell* **149**, 1635-1646 (2012).
8. 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).
9. Schwartz, S. et al. High-Resolution Mapping Reveals a Conserved, Widespread, Dynamic mRNA Methylation Program in Yeast Meiosis. *Cell* **155**, 1409-1421 (2013).
10. Wang, X. et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-120 (2014).
11. Wang, X. et al. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388-99 (2015).
12. Wang, Y. et al. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**, 191-198 (2014).

13. Fustin, J.-M. et al. RNA-Methylation-Dependent RNA Processing Controls the Speed of the Circadian Clock. *Cell* **155**, 793-806 (2013).
14. Batista, P.J. et al. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707-19 (2014).
15. Liu, N. et al. N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560-564 (2015).
16. Alarcon, C.R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S.F. N6-methyladenosine marks primary microRNAs for processing. *Nature* **519**, 482-485 (2015).
17. Alarcon, C.R. et al. HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell* **162**, 1299-308 (2015).
18. Wei, C.-M., Gershowitz, A. & Moss, B. Methylated nucleotides block 5′ terminus of HeLa cell messenger RNA. *Cell* **4**, 379-386 (1975).
19. Wei, C.M. & Moss, B. Nucleotide sequences at the N6-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry* **16**, 1672-6 (1977).
20. Bokar, J.A., Rath-Shambaugh, M.E., Ludwiczak, R., Narayan, P. & Rottman, F. Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. *Journal of Biological Chemistry* **269**, 17697-17704 (1994).
21. Liu, J. et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol* **10**, 93-95 (2014).
22. Ping, X.-L. et al. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res* **24**, 177-189 (2014).
23. Wang, Y. et al. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**, 191-8 (2014).
24. Jia, G. et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **7**, 885-7 (2011).
25. Zheng, G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* **49**, 18-29 (2012).
26. Xu, C. et al. Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. in *Nat Chem Biol* **10**, 927-9 (2014).
27. Zhou, K.I. et al. N-Methyladenosine Modification in a Long Noncoding RNA Hairpin Predisposes Its Conformation to Protein Binding. *J Mol Biol*. (2015)
28. Roost, C. et al. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *J Am Chem Soc* **137**, 2107-15 (2015)
29. Kierzek, E. & Kierzek, R. The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic Acids Res* **31**, 4472-80 (2003).
30. Spitale, R.C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486-490 (2015).
31. Delbridge, M.L., Lingenfelter, P.A., Disteche, C.M. & Graves, J.A. The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. *Nat Genet* **22**, 223-4 (1999).
32. Mazeyrat, S., Saut, N., Mattei, M.G. & Mitchell, M.J. RBMY evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene. *Nat Genet* **22**, 224-6 (1999).

33. Martinez-Contreras, R. et al. hnRNP proteins and splicing control. *Adv Exp Med Biol* **623**, 123-47 (2007).
34. Nasim, M.T., Chernova, T.K., Chowdhury, H.M., Yue, B.G. & Eperon, I.C. HnRNP G and Tra2beta: opposite effects on splicing matched by antagonism in RNA binding. *Hum Mol Genet* **12**, 1337-48 (2003).
35. Takemoto, T. et al. RBMX is a novel hepatic transcriptional regulator of SREBP-1c gene response to high-fructose diet. in *FEBS Lett* **581**, 218-22 (2007).
36. Zhao, S., Korzan, W.J., Chen, C.C. & Fernald, R.D. Heterogeneous nuclear ribonucleoprotein A/B and G inhibits the transcription of gonadotropin-releasing-hormone 1. *Mol Cell Neurosci* **37**, 69-84 (2008).
37. Matsunaga, S. et al. RBMX: a regulator for maintenance and centromeric protection of sister chromatid cohesion. *Cell Rep* **1**, 299-308 (2012).
38. Adamson, B., Smogorzewska, A., Sigoillot, F.D., King, R.W. & Elledge, S.J. A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat Cell Biol* **14**, 318-28 (2012).
39. Tsend-Ayush, E. et al. RBMX gene is essential for brain development in zebrafish. *Dev Dyn* **234**, 682-8 (2005).
40. Dichmann, D.S., Fletcher, R.B. & Harland, R.M. Expression cloning in Xenopus identifies RNA-binding proteins as regulators of embryogenesis and Rbmx as necessary for neural and muscle development. *Dev Dyn* **237**, 1755-66 (2008).
41. Moursy, A., Allain, F.H. & Clery, A. Characterization of the RNA recognition mode of hnRNP G extends its role in SMN2 splicing regulation. *Nucleic Acids Res* **42**, 6659-72 (2014).
42. Heinrich, B. et al. Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in pre-mRNA. *J Biol Chem* **284**, 14303-15 (2009).
43. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149, 753-67 (2012).
44. Kanhoush, R. et al. Novel domains in the hnRNP G/RBMX protein with distinct roles in RNA binding and targeting nascent transcripts. *Nucleus* **1**, 109-22 (2010).
45. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-41 (2010).
46. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-78 (2012).
47. Huelga, S.C. et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* **1**, 167-78 (2012).
48. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, 2008-17 (2010).
49. Zarnack, K. et al. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152**, 453-66 (2013).
50. Liu, N. et al. Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19**, 1848-56 (2013).
51. Corcoran, D.L. et al. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* **12**, R79 (2011).
52. Dominissini, D. et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-6 (2012).

53. Meyer, K.D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635-46 (2012).
54. Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N. & Rechavi, G. Transcriptome-wide mapping of N(6)-methyladenosine by m(6)A-seq based on immunocapturing and massively parallel sequencing. *Nat Protoc* **8**, 176-89 (2013).
55. Lohse, M. et al. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res* **40**, W622-7 (2012).
56. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
57. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
58. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110-21 (2010).

Chapter 5. Summary and Perspective

5.1 Introduction

N^6-methyladenosine (m^{6}A) is the most abundant internal modification in eukaryotic messenger RNA (mRNA). Recent discoveries on the locations, functions and mechanisms of m^{6}A have shed light onto a new layer of gene regulation at the RNA level, opening the field of m^{6}A-coded RNA epigenetics. Here, this summary and perspective aims to reveal the mechanisms of m^{6}A functions.

5.2 Detection of m^{6}A status

Initially, m^{6}A functional studies were hindered by the lack of methods for its detection, which is challenging because m^{6}A is indistinguishable from A during reverse transcription, and there are no known chemical reagents which can specifically label m^{6}A. In 2012, two groups independently combined m^{6}A antibody-immunoprecipitation and deep sequencing methods to identify tens of thousands of m^{6}A sites in mammalian mRNA at ~100-200 base resolution[1,2]. This method, which we refer to here as m^{6}A/MeRIP-seq, identified peaks in ~25% of all transcripts in human cells. Surprisingly, these m^{6}A peaks are not evenly distributed along the transcript, but are enriched near stop codons and in the 3' untranslated region (UTR), although the modification also occurs in the coding region and 5' UTR. The physiological relevance of this asymmetric distribution remains unclear. Subsequent improvements of the technology have further refined m^{6}A detection, for example Schwartz *et al*. optimized the anti-m^{6}A pull down protocol and eliminated many false positives by using methylation-deficient yeast strains[3]. In addition, ultraviolet light-induced antibody-RNA cross-linking followed by reverse transcription has enabled single-nucleotide-resolution mapping of m^{6}A of a human transcriptome[4,5]. These

studies have confirmed the RRACH (R=A,G; H=A,C,U) motif as the m^{6}A consensus motif, as characterized by early studies[6],. However, the majority of RRACH motifs in cellular RNA do not harbor the modification. Instead, m^{6}A levels in mRNA/lncRNA are fractional or incomplete, as shown by Site-specific Cleavage And Radioactive-labeling followed by Ligation-assisted Extraction and TLC (SCARLET)[7]. This method revealed modification fractions ranging from 7-88% among dozens of m^{6}A sites in human cells. Together with results from m^{6}A site measurements before the deep sequencing era, this finding indicates that not all copies of a transcript that contains an m^{6}A peak are methylated.

Table 5.1. The phenotypes of m^{6}A perturbation in cells and organisms.

Species / cell lines	**m^{6}A perturbation**	**Phenotype**	**Ref**
Human&Mice / ESC	*METTL3* KO	Impaired ESC exit from self-renewal toward differentiation	[8,9]
Human / HEK293T	*METTL3*, *METTL14* KD	Slowed cell proliferation rate	[10]
Human / U2OS; Mice / embryonic fibroblasts	*METTL3* KD	Circadian period elongation	[11]
Human / HeLa	*METTL3* KD	Cell death	[12]
Human&Mice	*FTO* KO	Obesity	[13,14]
Mice/ESC	*METTL3*, *METTL14* KD	Lost self-renewal capability	[15]
Mice	*ALKBH5* KO	Spermatogenesis	[16]
Drosophila	*IME4* KO	Viability	[17]
Budding yeast	*IME4* KO	Sporulation stop	[18]
Arabidopsis thaliana	*MTA* KD, KO	Embryonic development, plant growth patterning	[19,20]

ESC: embryonic stem cells; KD: knockdown; KO: knockout; IME4, MTA: homolog of METTL3.

5.3 The selective m^{6}A methylation by m^{6}A 'writers', 'erasers' and other machineries

Analogous to DNA methylation and histone modifications, the m^{6}A RNA modification can be added by 'writers' and removed by 'erasers' (**Figure 5.1**). m^{6}A 'writers' were initially identified from HeLa nuclear extracts as components of multi-protein complexes, which

contained two enzymes - METTL3 and METTL14 - that catalyze m^6A methylation[15,21]. Further studies identified Wilms' tumour 1-associating protein (WTAP) as one important protein factor in the RNA methylation complex. WTAP translocates the METTL3-METTL14 complex to mRNA targets in nuclear speckles, thus increasing the RNA-binding capabilities of the m^6A methyltransferases and improving the methylation efficiency[20,21,22]. Knockdown of *METTL3* or *METTL14* causes ~40% global m^6A decrease in human mRNAs, while *WTAP* depletion leads to even larger drops in m^6A levels[21,23]. A large fraction (around 30%) of human m^6A-writers' target RNA sites are located in introns, suggesting that m^6A methylation might occur co-transcriptionally either before or during splicing[15,21]. Since various pre-mRNA splicing factors colocalize with the m^6A-methyltransferases in the nuclear speckles, the intronic m^6A residues may regulate splicing factor binding to influence alternative splicing patterns.

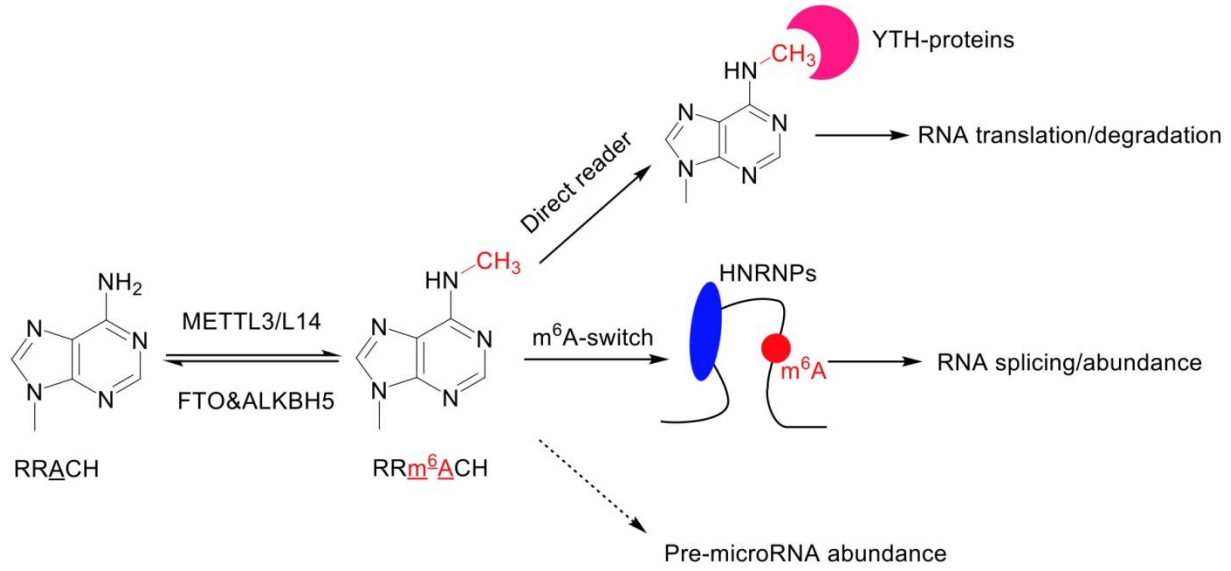


Figure 5.1. m^6A-coded RNA epigenetics: writers, erasers and readers. m^6A can be added by 'writers' METTL3/L14 within RRACH consensus motif (R=A,G; H=A,C,U), and removed by 'erasers' FTO/ALKBH5. m^6A can directly recruit the 'readers' YTH-domain proteins to regulate the stability and translation efficiency of cognate RNAs. Through the m^6A-switch mechanism, m^6A can alter its local RNA structure and affect HNRNPs binding to influence the RNA splicing and abundance. m^6A can also affect the pre-microRNA abundance.

The discovery of m^6A erasers in 2011 and 2012 indicated that the m^6A modification is reversible[16,24]. The first identified m^6A eraser, FTO (fat mass and obesity-associated protein), is strongly associated with human body weight[25,26,27,28,14], though the disease mutations in the locus encoding FTO are now known to influence the transcription of adjacent genes rather than FTO itself [29]. Despite the controversial role of FTO in diseases, the second identified m^6A eraser ALKBH5 is associated with the mouse fertility[16]. Both FTO and ALKBH5 belong to the non-heme Fe(II)- and α-ketoglutarate(KG)-dependent dioxygenase AlkB family of proteins. FTO oxidizes m^6A to generate N^6-hydroxymethyladenosine as an intermediate modification, and N^6-formyladenosine as a further oxidized product[30]. These demethylation intermediates, present in isolated mRNAs from human cells as well as mouse tissues, could potentially also possess regulatory functions, as has been described for the oxidized derivatives of 5-methylcytocine in DNA. Besides the already identified m^6A writers and erasers, there may be other proteins yet to be discovered that are responsible for regulating the levels of m^6A regulations in the cell.

One essential issue regarding m^6A methylation is how the methylation machinery selectively targets a subset of consensus motifs in the transcriptome. It remains unclear why RRACH motifs adjacent to stop codons are more likely to be methylated. Cis-elements such as neighboring RNA sequences or structures likely play a role in the methylation specificity. Schwartz *et al.* showed that m^6A has a static topology across multiple dynamic systems, although the m^6A/MeRIP-seq experiments they used cannot directly quantify m^6A levels and thus may be blind to quantitative differences in methylation levels between samples[23]. Meanwhile, several other studies indicate the m^6A amounts are various among different mouse tissues[1,2]. The tissue-specific m^6A patterns argue for trans-regulatory elements. One crucial way to better understand

the selective m^6A-deposition is to identify additional factors interacting with the methylation/demethylation machineries.

Recently, Chen *et al.* suggested that microRNAs (miRNAs) can act as a trans-regulatory factor for the methylation specificity[31]. miRNAs are important post-transcriptional regulators in eukaryotes. The authors showed that miRNAs modulate METTL3 binding to mRNAs via a sequence pairing mechanism to perform m^6A RNA modification. They showed that the m^6A on individual mRNAs and overall cellular m^6A abundance can be affected by the miRNA expression or sequences. Although Argonaute (AGO) proteins are known to mediate the binding of miRNAs to their target mRNAs, AGO proteins do not participate in the regulation of m^6A formation. Thus, it is likely that miRNAs associate with other yet unknown RNA-binding proteins to regulate m^6A deposition. These results reveal the functions of miRNAs in guiding m^6A writers to specific RNA sites for methylation, and partially explain the methylation specificity mechanism. Another essential issue of the methylation control is how the m^6A 'writers' cooperate with 'erasers' or other possible protein factors to tightly regulate the m^6A dynamics in response to various stress stimuli. Quite recently, Zhou *et. al* found that, in response to heat shock stress, certain adenosines at the 5' untranslated region (5'UTR) of newly synthesized mRNAs are preferentially m^6A methylated[32]. They showed that this stress-increased 5' UTR methylation is the result of competitive binding between the m^6A 'eraser' FTO and m^6A 'reader' YTHDF2, which is one well characterized m^6A-specific RNA-binding protein. Upon heat shock stress, YTHDF2 translocates from the cytoplasm to the nucleus and preserves the 5'UTR methylation of stress-induced transcripts by blocking m^6A 'eraser' FTO from demethylation. Furthermore, this stress-responsive 5'UTR m^6A methylation promotes cap-independent translation initiation, but how exactly m^6A recruits the translation machinery needs

further investigation. These findings present the first model about the control of m^6A dynamics in response to stress stimuli.

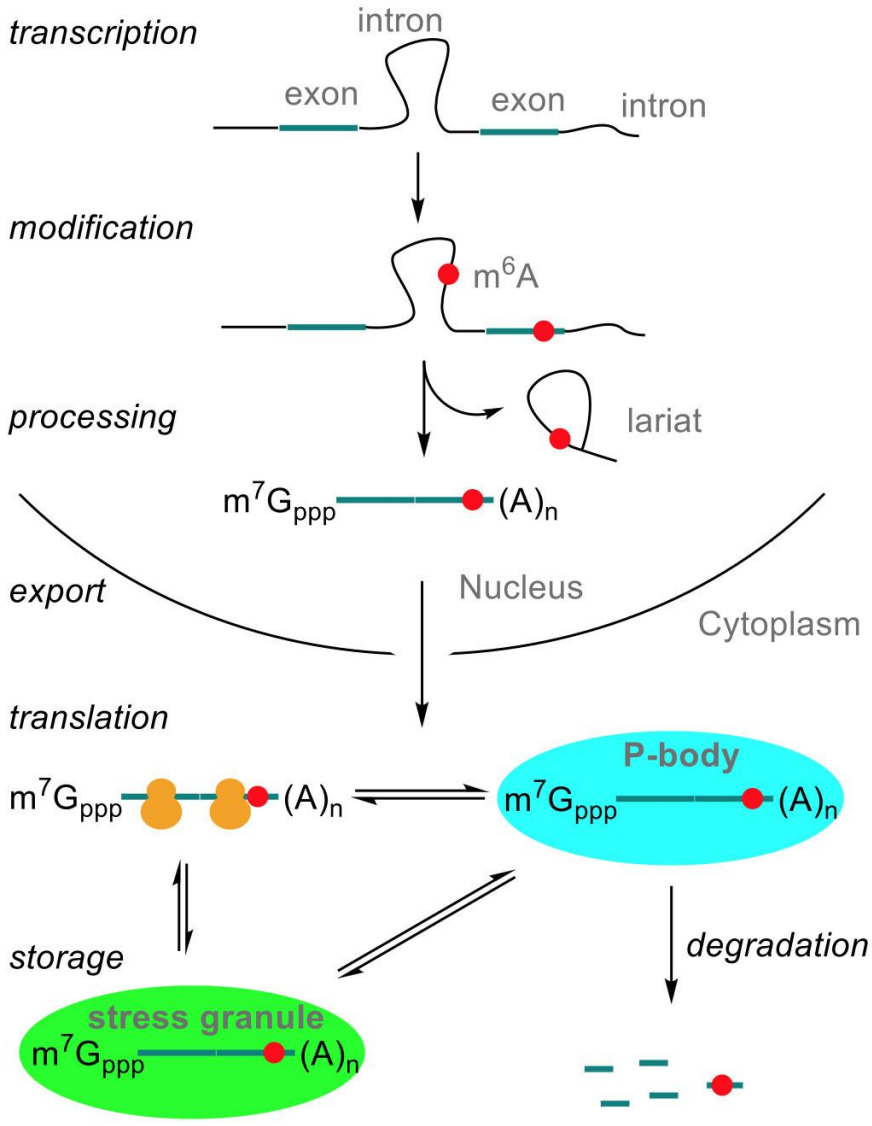


Figure 5.2. m^6A-regulated RNA metabolism in mammalian cells. Both introns and exons of the nascent RNA transcripts can be m^6A modified. m^6A affects the post-transcriptional regulators binding to influence RNA processing steps, such as alternative splicing. Afterwards, methylated transcripts are exported into the cytoplasm, and transferred among polysomes (translation), stress granules (storage) and P-bodies (degradation) to regulate gene expression.

5.4 Mechanisms for m^6A function

m^6A can exert its function through the recruitment of specific binding proteins, the so-called m^6A 'readers', but the modification also affects RNA structure, with its own consequences on RNA-mediated processes. In this section, we discuss the currently known m^6A readers and their mechanisms, as well as other functional consequences of m^6A.

5.4.1 Direct m^6A readers: YTH-domain proteins

m^6A can be directly recognized by the YTH-domain proteins. In 2012, Dominissini *et al.* used the m^6A methylated RNA bait to pull down RNA-binding proteins from the human lysate, and found that the YTH-domain proteins selectively bind the m^6A methylated RNA bait compared with the non-methylated counterparts[1]. The mammalian YTH-domain proteins have 5 members: YTHDF1-3, mostly localized in the cytoplasm and YTHDC1-2, localized in the nucleus[33-35,36]. A co-crystal structure shows that a hydrophobic pocket within the YTH domain directly accommodates the methyl-group of m^6A[35,37]. The functions of YTHDF1 and YTHDF2 have been characterized in detail[33,34]. While YTHDF2 recognizes many m^6A-containing mRNAs that are not being translated and recruits them to cytoplasmic P-bodies for degradation, YTHDF1 selectively binds actively m^6A-modified mRNAs and promotes their protein synthesis by interacting with translation machinery[34],[33]. Wang *et al.* showed that YTHDF1can promote ribosome loading of m^6A-modified mRNAs, and interact with initiation factors to faciliate translation initiation. The YTHDF1 promoted translation efficiency is particularly obvious during the stress response. Wang *et al.* suggested that, in response to arsenite treatment, the YTHDF1 is driven to stress granules and mediate/stabilize the formation of stalled initiation complex within the stress granules. Once stress is released, the YTHDF1-bound mRNAs can rapidly resume translation initiation from its stalled position. Curiously, YTHDF1 and YTHDF2 share a large set of common target mRNAs. Wang *et al.* suggested that YTHDF1 binds the

shared mRNA targets earlier to promote their translation before they are recognized by YTHDF2 to facilitate their decay [33]. Thus, YTHDF1 and YTHDF2 cooperate to improve the translation efficiency of transcripts that have short life times.

5.4.2 m^6A affects RNA structure

m^6A can also alter the stability and conformation of RNA molecules. Several studies have measured the m^6A effect on the base pairing of RNA duplexes[38],[39]. The experiments by Kierzek *et al.* with various adenine N-substitutions discovered m^6A destabilizes base-paired regions of RNA. By NMR analysis, Roost *et al.* explained that the m^6A-U pair in a stable RNA duplex is paired in canonical Watson-Crick geometry, which rotates the N^6-methyl group into the high energy anti- conformation. One single m^6A residue within this energetically unfavorable orientation can destabilize the RNA duplex by 0.5-1.7 kcal/mol. Besides, Roost *et al.* also observed that m^6A in unpaired positions base stacks stronger than the unmodified base, thus contributing to its stabilization in single-stranded locations. Roost *et al.* explained that m^6A in unpaired RNA regions adopt the relaxed (syn) conformation[40], thus placing the methyl group in contact with adjacent base pair and likely adding favorable hydrophobic stabilization.

Consistent with these thermodynamic measurement studies, transcriptome-wide RNA structural mapping analysis from human cells reveals that the secondary RNA structures adjacent to the m^6A sites are more single-stranded than the non-methylated RNA region. Thus, it is likely that m^6A functions as the RNA secondary structure remodeler within eukaryotic mRNA/lncRNAs. These m^6A-modulated structures can serve as another layer of information, affecting RNA-protein, but also RNA-RNA and RNA-DNA interactions transcriptome-wide. Potentially, co-transcriptional m^6A methylation could also affect the folding of the newly synthesized transcript, andcause the formation of different secondary/tertiary RNA structures of

the transcripts. RNA transcripts with different structures probably interact with different downstream enzymatic factors, go through various biological pathways and end up with distinct RNA fates.

5.4.3 Indirect m^{6}A readers: m^{6}A-switches

The concept that m^{6}A modulates RNA-protein interactions via effects on RNA secondary structure is exemplified by a recent report demonstrating that m^{6}A can influence the accessibility of RNA-binding motifs for their RNA-binding proteins (RBPs)[10]. The mechanism regulating RNA-protein interactions based on m^{6}A-dependent RNA structural remodeling is termed 'm^{6}A-switch'. Liu *et al.* found that m^{6}A can alter the local RNA structure to increase the accessibility of RNA-binding motifs for HNRNPC binding. HNRNPC, belonging to the large family of heterogeneous nuclear ribonucleoproteins (HNRNPs), is one abundant nucleus-localized protein to bind and process nascent RNA transcripts into mature mRNAs. HNRNPC binds its U-tract RNA-binding motifs in nascent mRNA exposed by m^{6}A, rather than the m^{6}A itself. The U-tract motifs are frequently buried within their local RNA structures (such as hairpin-stems), thus inhibiting their interactions with HNRNPC. Liu *et al.* find that the U-tract motifs located in hairpin-stems regularly base pair with the m^{6}A-consensus motif (RRACH), which can m^{6}A methylated by m^{6}A writers. Once methylated, the m^{6}A residues will destabilize the hairpin-stem structure and make the U-tract motif more single-stranded or accessible to facilitate HNRNPC binding. It is likely that the RNA binding sites of HNRNPC outnumber HNRNPC protein copies in the cell. Through the m^{6}A-switch mechanism, the m^{6}A methylation gives direction to HNRNPC regarding which specific RNA binding sites to bind. HNRNPC has been associated with RNA transcription, stability and alternative splicing regulation. And changes in HNRNPC binding patterns directly impact the downstream functions of HNRNPC at their target substrates.

In total, thousands of m^6A-switches were identified that regulate HNRNPC binding, and influence RNA abundance and alternative splicing in human cells.

Besides HNRNPC, m^6A-switches could possibly control the accessibility of the RNA-binding motifs for many other RBPs, thus influencing wide-ranging biological activities. Further investigation of the m^6A-altered RNA structure and its functions requires highly-sensitive methods to detect m^6A and secondary/tertiary RNA structures *in vivo*. Recently, Spitale *et al.* developed the *in vivo* click selective 2′-hydroxyl acylation and profiling experiment (icSHAPE) to obtain the global view of *in vivo* RNA structure profiles at base-resolution[41]. They observed that RNA structures surrounding and including the m^6A sites are more single stranded *in vivo*. Besides, the global m^6A depletion by METTL3 knockout decreases the single-stranded tendency around methylated bases, validating the role of m^6A in distabilizing base pairing in RNA helices.

5.4.4 The association between low-complexity domains and m^6A readers

Interestingly, all m^6A readers (both direct and indirect) identified so far (YTH-domain proteins, HNRNPC, HNRNPA2B1, HNRNPG) contain a low-complexity (LC) domain, which is characterized by a biased amino acid composition. In fact, many RNA-binding proteins contain such LC domains, which can promote the formation of RNA granules[42]. RNA granules, including stress granules and P-bodies of mammalian cells, are membrane-free cellular structures composed of RNA molecules and RNA-binding proteins (RBPs). Although RNA granules were first observed more than 100 years ago, the mechanism underlying their formation has been elusive. McKnight and colleagues show that RNA granules consist of mainly RNA binding proteins containing LC domains, and the LC domains are indispensable for RNA granules formation both *in vitro* and *in vivo*. For example, the YTHDF2 protein contains a QN-rich domain, which plays an essential role in recruiting YTHDF2-RNA complex to P-bodies for RNA

decay. In addition, the METTL14 has a Gly-rich domain at the C-terminal region which may also regulate its subnuclear localization[36]. It remains unclear how LC domains contribute to the granule formation, but evidences show at phosphorylation on the LC domains can destroy the granule formation, thus adding a new layer of complexity of granule formation regulation. The prevalence of LC domains in m^6A-dependent proteins suggest that m^6A-readers and their bound mRNAs are located at specific cellular locations, interacting with other RNA binding proteins within the same RNA granules. Aside from YTH-domain proteins, other proteins may exist that directly recognize the m^6A residue and their identification would be of great importance to understand the biological pathways regulated by the m^6A modification.

5.4.5 m^6A-dependent microRNA maturation

m^6A has also been associated with miRNA biogenesis. Tavazoie and coworkers reported the role of m^6A in promoting the initiation of miRNA biogenesis[43]. They found that m^6A methylation in primary microRNAs (pre-miRNAs) is required for their recognition and processing by DGCR8, an RNA-binding protein in the microprocessor complex. Global m^6A reduction by depletion of METTL3 decreased the interactions between DGCR8 and pri-miRNAs, thus causing a global reduction of mature miRNAs. Quite recently, Tavazoie and coworkers found that HNRNPA2B1 is one mediator of the m^6A-dependent nuclear RNA processing events. Similar as HNRNPC, HNRNPA2B1 belongs to the nuclear HNRNP protein families, and binds nascent RNA transcripts for pre-mRNA processing. Tavazoie found that HNRNPA2B1 recognize m^6A marks in the pre-mRNA and recruit DGCR8 protein complexes to faciliate miRNA maturation[44]. Besides, HNRNPA2B1 also mediates the m^6A-dependent alternative splicing regulation. Tavazoie pointed out that HNRNPA2B1 cannot explain all m^6A-dependent nuclear RNA processing events, there may exist other m^6A readers yet to be discovered. The

detailed mechanism how HNRNPA2B1 recognizes m^6A residues is not clear. Since HNRNPA2B1 does not contain the YTH-domain, we speculate that HNRNPA2B1 is regulated by the m^6A-switch mechanism. Collectively, these studies reveal the potential regulatory network between the m^6A-dependent biology and the miRNA-based pathway for co-operative gene regulation.

5.5 Conclusions

m^6A-coded RNA epigenetics is entering an exponential growth phase. In the last years, we have witnessed significant advances in m^6A profiling techniques and gained understanding in m^6A mechanisms and functions. The identification of m^6A-responsive RNA-binding proteins has revealed that m^6A regulates its cognate RNAs from birth to decay. Besides, m^6A can also modulate RNA structures, microRNA biology and protein localization. Thus, the effects of m^6A are wide-spread. However, it remains unclear how these multiple m^6A functions are coordinated within the cell, how they are coupled with m^6A biogenesis and removal, or how and if m^6A functions are regulated in a cell type/state-dependent manner.

Future studies should improve the resolution and sensitivity of m^6A profiling methods, explore additional m^6A regulatory mechanisms, and further investigate the cellular signaling pathways involving m^6A and their reader proteins. These studies will be of great importance to better understand the complex regulatory networks underlying m^6A-encoded RNA biology and enable applications to human health and disease.

5.6 References:

1. Dominissini, D. et al. Topology of the human and mouse m^6A RNA methylomes revealed by m^6A-seq. *Nature* **485**, 201-206 (2012).
2. Meyer, K. et al. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. *Cell* **149**, 1635-1646 (2012).

3. Schwartz, S. et al. High-Resolution Mapping Reveals a Conserved, Widespread, Dynamic mRNA Methylation Program in Yeast Meiosis. *Cell* **155**, 1409-1421 (2013).
4. Chen, K. et al. High-resolution N(6) -methyladenosine (m(6) A) map using photo-crosslinking-assisted m(6) A sequencing. *Angew Chem Int Ed Engl* **54**, 1587-90 (2014).
5. Linder, B. et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12**, 767-72 (2015).
6. Wei, C.-M., Gershowitz, A. & Moss, B. Methylated nucleotides block 5′ terminus of HeLa cell messenger RNA. *Cell* **4**, 379-386 (1975).
7. 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).
8. Batista, P.J. et al. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707-19 (2014).
9. Geula, S. et al. Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* 347, 1002-6 (2015).
10. Liu, N. et al. N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560-564 (2015).
11. Fustin, J.-M. et al. RNA-Methylation-Dependent RNA Processing Controls the Speed of the Circadian Clock. *Cell* **155**, 793-806 (2013).
12. Bokar, J.A. The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. in *Fine-tuning of RNA functions by modification and editing* 141-177 (Springer, 2005).
13. Gerken, T. et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469-72 (2007).
14. Fischer, J. et al. Inactivation of the Fto gene protects from obesity. *Nature* **458**, 894-8 (2009).
15. Wang, Y. et al. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**, 191-8 (2014).
16. Zheng, G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* **49**, 18-29 (2012).
17. Hongay, C.F. & Orr-Weaver, T.L. *Drosophila* Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proceedings of the National Academy of Sciences* **108**, 14855-14860 (2011).
18. Agarwala, S.D., Blitzblau, H.G., Hochwagen, A. & Fink, G.R. RNA methylation by the MIS complex regulates a cell fate decision in yeast. *PLoS Genet* **8**, e1002732 (2012).
19. Bodi, Z. et al. Adenosine methylation in Arabidopsis mRNA is associated with the 3' end and reduced levels cause developmental defects. *Frontiers in Plant Science* **3** (2012).
20. Zhong, S. et al. MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *The Plant Cell Online* **20**, 1278-1288 (2008).
21. Liu, J. et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol* **10**, 93-95 (2014).
22. Ping, X.-L. et al. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res* **24**, 177-189 (2014).
23. Schwartz, S. et al. Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep* **8**, 284-96.

24. Jia, G. et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **7**, 885-7 (2011).
25. Boissel, S. et al. Loss-of-Function Mutation in the Dioxygenase-Encoding FTO Gene Causes Severe Growth Retardation and Multiple Malformations. *American Journal of Human Genetics* **85**, 106-111 (2009).
26. Church, C. et al. Overexpression of Fto leads to increased food intake and results in obesity. *Nat Genet* **42**, 1086-1092 (2010).
27. Dina, C. et al. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* **39**, 724-726 (2007).
28. Frayling, T.M. et al. A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science* **316**, 889-894 (2007).
29. Smemo, S. et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371-5 (2014).
30. Fu, Y. et al. FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat Commun* **4**, 1798 (2013).
31. Chen, T. et al. m(6)A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. *Cell Stem Cell* **16**, 289-301 (2015).
32. Zhou, J. et al. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* 526, 591-4 (**2015**).
33. Wang, X. et al. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388-99 (2015).
34. Wang, X. et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-120 (2014).
35. Xu, C. et al. Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat Chem Biol* **10**, 927-9 (2014).
36. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m6A RNA methylation. *Nat Rev Genet* **15**, 293-306 (2014).
37. Theler, D., Dominguez, C., Blatter, M., Boudet, J. & Allain, F.H. Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. *Nucleic Acids Res* **42**, 13911-9 (2014).
38. Kierzek, E. & Kierzek, R. The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic Acids Res* **31**, 4472-80 (2003).
39. Roost, C. et al. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *J Am Chem Soc* **137**, 2107-15 (2015).
40. Sternglanz, H. & Bugg, C.E. Conformation of N6-methyladenine, a base involved in DNA modification: restriction processes. *Science* **182**, 833-4 (1973).
41. Spitale, R.C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486-490 (2015).
42. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-67 (2012).
43. Alarcon, C.R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S.F. N6-methyladenosine marks primary microRNAs for processing. *Nature* **519**, 482-485 (2015).
44. Alarcon, C.R. et al. HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell* **162**, 1299-308 (2015).