THE UNIVERSITY OF CHICAGO

MESSENGER RNA MODIFICATIONS AND READER PROTEINS IN THE REGULATION OF GENE EXPRESSION

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CELL AND MOLECULAR BIOLOGY

BY

ADAM WYLDER

CHICAGO, ILLINOIS

AUGUST 2023

Table of Contents

List of Figures

Acknowledgements

I am very grateful for the opportunities afforded by my advisor, Dr. Tao Pan, and I would also to thank my committee: Dr. David Pincus, Dr. Tobin Sosnick, and Dr. Heng-Chi Lee. I am grateful for the work of all past and present Pan Lab members, without which the work here could not have been completed or initiated. I thank Dr. Katherine Zhao for her initial work on hnRNPG and both she and Dr. Hailing Shi for providing additional information to assist in my chromatin fractionation protocols. I would like to thank Noah Peña for initial assistance in attempting a variety of E. coli YTHDC1 expression strategies and in addition for his assistance in the visualization of the MSR-Seq data presented in Chapter 4 of this thesis. I would also like to thank Dr. Jiangbo Wei for the FLAG-YTHDC1 plasmids central for Chapter 3. For the development of MSR-Seq and additional support and assistance, I would like to thank Dr. Chris Watkins and Dr. Chris Katanski. I would also like to thank Dr. Chris Katanski for providing learning resources on our bioinformatic pipelines. I would like to thank Sihao Huang for his development of Nano-SPA and collaboration on Chapter 5 of this thesis; his expertise in bioinformatics was instrumental in our work on nanopore sequencing. For teaching polysome profiling and assorted experimental advice throughout my stay, I would like to thank Dr. Wen Zhang. I would also like to thank the support of the University of Chicago Biological Science Division and the MGCB Training Grant for support without which this work would not be possible.

I would finally like to thank my family and personal friends. I am grateful for my parents Kris and Jim Wylder, my brother Matthew Wylder, and my grandma Ginny Baker; their love and support have gotten me where I am today. I would like to thank my long-time friends and Dungeons & Dragons group (Corr, C., Litzo., T., Vining, E., et. al). Finally, I

would like to thank my partner Rossy Natale who I have been fortunate enough to have by my side through graduate school.

This thesis includes text and figures derived from a published co-author manuscript in Molecular Cell by Cell Press, relies on data from a co-authored paper published in *Nature Communications* by SpringerNature, and a co-first author manuscript recently submitted to *Nature Biotechnology* by SpringerNature.

Abstract

Gene expression is regulated through a dynamic interconnected web of systems with its rich complexity only further appreciated following the Human Genome Project. The genome by itself not containing all information on a biological state has led to growing interest in the regulation of its expression, including through chemical modification deposited within mRNA transcripts. Here, I explore the impact of the two most abundant mRNA modifications, *N*6-methyladenosine (m^6A) and pseudouridine (Ψ) on gene expression by study of modification writer and reader enzymes, which respectively deposit and interact with transcript modifications. In Chapter 2, I show how m^6A reader protein hnRNPG can accommodate mutual, co-transcriptional interactions with RNA Polymerase II and m^6A-modified RNA through development of a spin-down assay for low-complexity protein interactions. In Chapter 3, I demonstrate that type I reader protein YTHDC1 interacts with RNA Polymerase II through its structured YTH-domain, which is outcompeted by its canonical ligand in m^6A. In Chapter 4, I perform a bioinformatic survey on the connection between m^6A writers and readers on tRNA. In Chapter 5, I detail our newly developed NanoSPA platform for direct, simultaneous long-read sequencing of m^6A and Ψ along the same mRNA transcript. Using this platform, we demonstrate crosstalk between m^6A and Ψ on modification concurrence and translational efficiency. Together, the work here expands the mechanistic understanding of mRNA modification writers and readers in regulation of gene expression.

Chapter 1

Introduction

Gene expression, the process by which information stored in genome imparts its downstream function, is dynamic and varied in all biological strata, from populations and species to specific cell types during stress (Frank, Matthew G. annis, Watkins 2019). It is the precise expression of a gene that ultimately shapes cellular function and fate and thereby refutes the genome as solely representative of the biological state. The genome is given more depth through transcriptional, post-transcriptional, translational, and post-translational regulation. These mechanisms connect, forming a dense regulatory network underpinning cellular physiology (Pope and Medzhitov 2018).

Herein, I detail interplay between chemical RNA modifications and the greater gene expression network.

1.1 Regulation of gene expression through chemical modification of messenger RNA

Chemically modified nucleotides were observed as early as the 1950s (COHN and VOLKIN 1951; WYATT and COHEN 1953), but the functional significance of this chemical diversity was not immediately apparent. Ten modifications were observed while sequencing the first biological RNA, an alanine transfer RNA (HOLLEY et al. 1965); the presence of a modified nucleotide in the third transfer RNA (tRNA) codon position led to the "wobble hypothesis," which postulates the importance of this modification in

expanding the triplet code through non-canonical interactions (Crick 1966). Initial characterizations found human ribosomal RNA (rRNA) and transfer RNA (tRNAs) highly enriched in modification relative to messenger RNA (mRNA), and as such became the focal point in their study. However, advances in sequencing technology have revealed the prominence of internal mRNA chemical modifications and allowed mechanistic inquiry into their significance. To date, over 100 unique chemical modifications have been identified in cellular RNA (Roundtree, Evans, et al. 2017).

Modification of eukaryotic mRNA was first studied in the context of the 5' m^7G cap (Ramanathan, Robb, and Chan 2016) and the 3' polyA tail (Fisher and Beal 2018). Modification of internal mRNA nucleotides primarily occurs at the ribose base, the most prevalent modifications being N1-methladenosine (m^1A), 5-methylcytosine (m^5C), N6-methyladenosine (m^6A), 5-hydroxymethylcytosine (hm^5C), N4-acetylcytidine (ac^4C), the uridine isomer pseudouridine (Ψ), and inosine (I) resulting from adenosine-to-inosine editing (Roundtree, Evans, et al. 2017; H. Shi et al. 2020; Arango et al. 2018). Research has shown that these modifications represent a regulatory layer beyond transcriptomics termed "epitranscriptomics," stemming from analogies to histone modification epigenetics (Helm and Motorin 2017). Broadly, these chemical modifications are introduced through enzymes referred to as "writers," removed in some cases by enzymes known as "erasers," and specifically recognized by "reader proteins," which impart the modification's downstream effect (**Figure 1.1**).

This thesis concerns the effect of the two most abundant of these modifications: m^6A and Ψ. Chapters 2 and 3 concern the co-transcriptional, RNA Polymerase II-mediated interactions of nuclear m^6A reader proteins, Chapter 4 is a bioinformatic survey

investigating the connection between m^{6}A-modified mRNA and tRNA processing, while Chapter 5 investigates the crosstalk between transcript m^{6}A and Ψ in addition to their effect on translation.

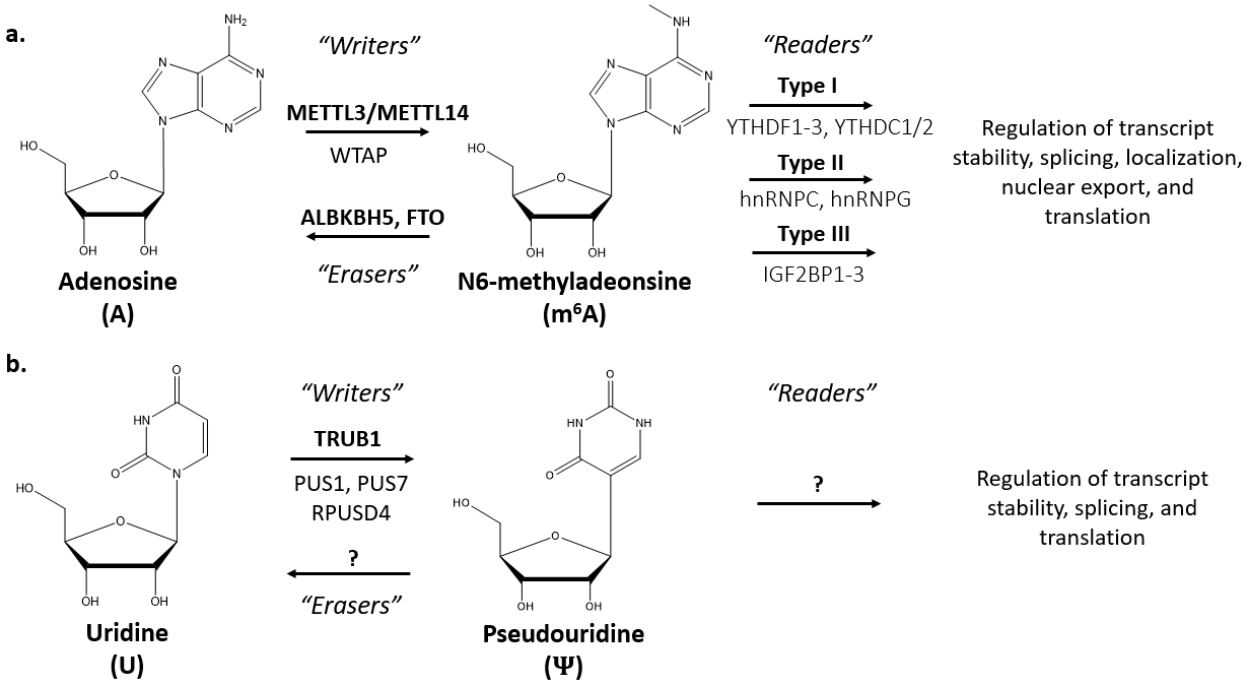


Figure 1.1 N6-methyladenosine(m6A) and pseudouridine (Ψ) are the two most abundant mRNA modifications in eukaryotes. (a). m^{6}A methyltransferases, or writer proteins, including core components METTL3 and METTL14 methylate the adenosine base at position 6 generating m^{6}A, whose abundance can be reversibly modified by demethylase, or eraser proteins, such as ALKBH5 and FTO. Once deposited, an m^{6}A moiety can be specifically recognized by one of three classes of "reader proteins." Once recognized, the modification can impact downstream gene regulation. (b). The uridine isomer Ψ is enzymatically generated by pseudouridine synthetases, or Ψ writers. The most predominant Ψ writer for mRNA in mammalian cell culture is TRUB1, while PUS1, PUS7, and RPUSD4 also act as writers for mRNA. No Ψ eraser or reader proteins have been identified in humans. Once deposited on a transcript, Ψ is known to impact stability, splicing, and translation.

1.1.1 N6-methyladenosine (m^6A)

*N*6-methyladenosine is the most abundant internal modification in eukaryotic mRNA, occurring on average at 1-3 Adenosines per transcript and thus making up 0.2-0.6% of transcript adenosines (Meyer et al. 2012). Along with its abundance, m^6A impacts nearly every stage of the mRNA lifecycle and is critical in development, cellular differentiation, and human disease (Roundtree, Evans, et al. 2017). Transcriptome-wide abundance and distribution of m^6A are regulated by the dynamic and opposing actions of two classes of enzymes: methyltransferases or "writers," including METTL3 and METTL14, and demethylases or "erasers," such as ALKBH5. Once deposited on the transcript, m^6A can impart downstream regulation following specific recognition by a class of proteins known as m^6A "reader proteins," (**Figure 1.1**).

Characterization of m^6A has been enabled and catalyzed both by advances in sequencing technology as well as the discovery that m^6A was reversible through the action of demethylases (Jia et al. 2011; Zheng et al. 2013). The development of m^6A-sequencing techniques enabled initial mapping to the transcriptome, and thus revealed over 12,000 m^6A peaks or modified regions in over 7,000 mRNA transcripts, confirming the prevalence of the conserved DRACH (D = A, G, or U; R = A or G, H = A, C, or U) consensus sequence, and demonstrated m^6A enrichment both near stop codons and within long, internal exons (Dominissini et al. 2012; Meyer et al. 2012). Firmly establishing the regulatory role and mechanism of m^6A and its mechanisms of action have required advances in mapping resolution and sequencing technology, which will be covered in greater depth in Chapter 1.3.

m^{6}A is a chemically stable modification with effects both on higher-order structures and energetics. The modification itself imparts a modest impact on the thermodynamic stability of secondary RNA structures (Kierzek and Kierzek 2003), but m^{6}A also induces changes to the base-pairing landscape. This is accomplished both by its destabilization of Hoogsteen base-pairing (Ashraf, Huang, and Lilley 2019) and Watson-Crick pairing with uridine in the *anti*-conformation; pairing which exhibits transient exchange to a mismatch-like conformation represents 1% of the population (B. Liu et al. 2021).

m^{6}A functions on mRNA stability, splicing, localization, nuclear export, and translation efficiency with a vital role in development and disease (Roundtree, Evans, et al. 2017). It is the dynamic interplay of m^{6}A levels on the transcriptome orchestrated by the opposing actions of writer and erasers ultimately followed by recognition through m^{6}A-reader proteins that imparts this breadth of regulatory function. Therefore, these three classes of proteins (writers, erasers, and readers) will be further detailed.

Writers

Deposition of m^{6}A occurs both co- and post-transcriptionally, catalyzed by the m^{6}A-methyltransferase complex (MTC); this complex's core is a heterodimer of proteins methyltransferase-like 3 (METTL3) and human methyltransferase-like 14 (METTL14) (Jianzhao Liu et al. 2014). The methyltransferase activity of the MTC is S-adenosyl methionine (SAM)-dependent, where SAM acts as a co-factor via donation of its methyl to the target adenosine. Both METTL3 and METTL14 contain an MT-A70 domain, which catalyzes adenosine methylation; however, METTL3 and not METTL14 were found to crosslink with SAM, which indicates METTL3's essential role in the core MTC heterodimer's

substrate accommodation (Bokar et al. 1997). It is within this heterodimer that METTL3 is active, as monomeric METTL3 is inert despite being stable in solution (Schöller et al. 2018).

In vivo, the METTL3/METTL14 heterodimer is formed in the cytoplasm, where it is thought the METTL3's nuclear localization signal enables the import of the dimer (Schöller et al. 2018). Once localized in the nucleus, the complex may associate with Wilms' tumor 1-associating protein (WTAP), which localizes the complex to nuclear speckles and assists in loading pre-mRNA processing factors (Ping et al. 2014). However, not every m^6A-modified site in mRNA is dependent on WTAP, as methylation sites have been classified by WTAP dependence (S. Schwartz, Mumbach, et al. 2014). The METTL3/METTL14 heterodimer itself exhibits a sequence specificity as it has a binding motif similar to the known RRACH m^6A consensus sequence (Dominissini et al. 2012); although, the MTC does not exhibit structural specificity (Jianzhao Liu et al. 2014). Knockdown of either METTL3, METTL14, or WTAP decreases *in vivo* m^6A abundance, but WTAP lacks methylase activity *in vitro*; WTAP's role in MTC localization resolves its lack of activity *in vitro* with its *in vivo* necessity.

Structural analysis has shown several cofactor binding sites in METTL3 are not conserved with METTL14, explaining SAM's preferential interaction with METTL3. METTL3 has the essential SAM binding residue D395 (Xiang Wang et al. 2016) in a catalytic DDPW motif within METTL3's greater catalytic loop not found in METTL14. Additionally, the essential DDPW motif W398 of METTL3 is thought to π-π stack with the m^6A-modified site, which increases the energetic favorability of methylation (Śledź and Jinek 2016). In contrast, METTL14 acts to stabilize the complex and increase affinity for SAM. METTL14 contains C-terminal Arginine-Glycine-Glycine (RGG) repeats, whose deletion abolishes the

catalytic activity of the METTL3/14 heterodimer (Schöller et al. 2018). While METTL14 does not contain the catalytic DDPW motif of METTL3, it is essential in the stabilization of METTL3, increasing complex affinity for SAM, and it energetically favors binding to the RNA substrate.

While METTL3, METTL14 and WTAP are the most well-studied m^{6}A writers, other proteins are also involved in the modification's installation in mRNA. KIAA1429, or vir-like m^{6}A methyltransferase-associated protein (VIRMA), is an additional methyltransferase complex component that interacts with WTAP and whose knockdown decreases transcript m^{6}A abundance (S. Schwartz, Mumbach, et al. 2014). METTL16 plays a comparatively minimal role, known only to act in SAM metabolism through modification of human SAM synthetase-encoding MAT2A transcripts; rather, METTL16's primary role is in methylation of the U6 spliceosomal small nuclear RNA (snRNA) (Pendleton et al. 2017). RNA-binding protein 15 (RBM15) interacts with RNA near m^{6}A-sites and is thought to recruit the METTL3/METTL14 heterodimer to its bound site, itself then integrated into the larger methyltransferase complex (Patil et al. 2016). Additionally, unique writers are found for noncoding RNA, including rRNA writers METTL5 and ZCCHC4 (Boulias and Greer 2023).

Erasers

Chemical modifications of RNA were initially believed to be irreversible, where the modification state was solely representative of the static action of a writer. While modification of rRNA by m^{6}A writers is still thought to be irreversible, the demonstration of m^{6}A's dynamic reversibility within mRNA transcripts revolutionized the collective understanding of the RNA modification landscape. The first m^{6}A demethylase or "eraser" discovered was fatty mass and obesity associated protein (FTO), which provided vital proof

of m^6A's reversibility and suggested possible generalized reversibility of mRNA modifications as a whole (Jia et al. 2011). FTO has been suggested to regulate alternative splicing and 3' processing (Bartosovic et al. 2017), and generates two potentially functional intermediates during its oxidation of m^6A in N6-hydroxylmethyladenosine (hm^6A) and N6-formyladenosine (f^6A) (Y. Fu et al. 2013). Additionally, FTO has been shown to oxidize m^6Am, expanding the diversity of the enzyme's substrates (Mauer et al. 2019). FTO is associated with body mass regulation in mice (Church et al. 2009, 2010), adipogenesis (Zhao et al. 2014), as is a known oncogene both in leukemia (Zejuan Li et al. 2017; Su et al. 2018) and glioblastoma (Cui et al. 2017).

Soon after the characterization of FTO, AlkB family member 5 (ALKBH5) was discovered -- a second Fe(II)- and α-ketoglutarate-dependent dioxygenase (Zheng et al. 2013). ALKBH5 prefers a consensus sequence, suggesting sequence specificity of its m^6A-demethylation activity compared to the broad activity of FTO. Like FTO, the importance of ALKBH5 has been demonstrated in cellular functions such as the nuclear export of mRNA. ALKBH5 is important developmentally; in mice, it is implicated both in mouse spermatogenesis and fertility (Zheng et al. 2013). Overexpression of ALKBH5 is required for glioblastoma stem cell proliferation and tumorigenesis and is a poor predictor of patient survival (Zhang et al. 2017).

Reader Proteins

While m^6A abundance is co-transcriptionally regulated through the opposing actions of writers and erasers, the modification imparts functionality via recognition by a specific class of RNA binding proteins known as m^6A-reader proteins. Three classes or

types of m^{6}A-reader proteins have currently been identified, each classified by a distinct m^{6}A-recognition mechanism described below.

Class I

YTH-domain family of proteins make up the first class of reader proteins, defined by direct recognition of m^{6}A-sites through their conserved, structured YT521-B homology (YTH)-domains. In humans, this family consists of proteins YTH domain family 1-3 (YTHDF1-3) and YTH domain-containing 1 and 2 (YTHDC1-2). YTH-family proteins are primarily localized to the cytosol, with YTHDC1 being the sole primarily nuclear-localized member (Roundtree, Evans, et al. 2017; Zhou and Pan 2018).

The YTH-domain itself displays specificity for m^{6}A-modified RNA than unmethylated RNA, with a binding dissociation constant to m^{6}A roughly 5- to 20-fold stronger than to the unmethylated nucleotide with variation due to sequence context (Xu et al. 2014; Patil et al. 2016). Multiple crystal structures of the YTH-domain of different YTH-family members both bound and unbound to m^{6}A-modified RNA have been published, revealing analogous mechanisms for the specific recognition of m^{6}A through two conserved tryptophan residues (Xu et al. 2014; Zhu et al. 2014). Studies have also suggested YTHDC1 and YTHDF1-3 bind directly to m^{1}A in RNA through their conserved YTH domain (X. Dai et al. 2018). In chapter 3, I present work that the YTH-domain of YTHDC1 is also capable of interacting with the C-terminal domain of RNA Polymerase II.

YTHDF1-3 are cytoplasmic localized m^{6}A reader proteins involved in m^{6}A-dependent gene regulation. YTHDF1 has been shown to upregulate translation, YTHDF2 to promote transcript decay, and YTHDF3 to enhance both translation and transcript decay.

(Xiao Wang et al. 2015, 2014; A. Li et al. 2017). By contrast, the nuclear m^{6}A-reader protein YTHDC1 functions accordingly upstream; YTHDC1 is involved in regulatory events including pre-mRNA processing and transcription in the nucleus (Kasowitz et al. 2018; Lee et al. 2021; N. Liu et al. 2017), pre-mRNA splicing (Xiao et al. 2016a), and stability of chromatin-associated RNAs (Jun Liu et al. 2020). Additionally, YTHDC1 is known to form membrane-less subnuclear complexes known as "YT-bodies," implicated in pre-mRNA processing (Rafalska et al. 2004).

The Class I YTH-family reader proteins are of significant interest in both human disease and development. In a variety of cancers, they have been implicated in nearly every step of disease progression and tumorigenesis (T. Liu et al. 2020; Bai et al. 2019; R. Shi et al. 2021). The ubiquity of their involvement in cancer leads to the proposal of YTHDF1 as a pan-cancer biomarker (J. Hu et al. 2021). YTHDF proteins are also part of an essential learning and memory pathway in *Drosophila* (Kan et al. 2021), while YTHDF2 is additionally involved in hematopoietic stem cell expansion (Zhenrui Li et al. 2018). In mice development, development, nuclear Class I reader protein YTHDC1 regulates alternative splicing and polyadenylation of developing oocytes (Kasowitz et al. 2018) and regulates LINE1 RNA scaffold function in ESCs and early embryos (C. Chen et al. 2021).

Class II

The second class of m^{6}A reader proteins recognizes their substrate through an "m^{6}A-switch" mechanism. Here, the accessibility of the reader protein's binding motif is subject to m^{6}A-dependent structural regulation, where the binding motif itself may or may not contain the m^{6}A moiety (N. Liu et al. 2017, 2015). Class II m^{6}A reader proteins include

heterogeneous nuclear ribonucleoproteins (hnRNPs) hnRNPC, hnRNPG, and possibly hnRNPA2b1.

Broadly, hnRNPs are critical in pre-mRNA regulation, including as splicing factors. hnRNPs can act as scaffolds, recruiting RNAs and assembling higher-ordered structures crucial for gene expression, demonstrated by their necessity in stem cell self-renewal and differentiation (Xie et al. 2021). hnRNPs are implicated in various stages of cancer development, and have been investigated as a potential pan-cancer diagnostic tool (Hao Li et al. 2020); for example, hnRNPC mediates m^{6}A-dependent alternative splicing in pancreatic ductal adenocarcinoma (X.-T. Huang et al. 2021) and also regulates alternative cleavage and polyadenylation in metastatic colon cancer (Fischl et al. 2019).

hnRNPG, also known as RBMX, recognizes m^{6}A-modified RNA through Arg-Gly-Gly (RGG) motifs embedded in its low-complexity body (N. Liu et al. 2017). Two regions within this low-complexity body contain RGG motifs, designated RGG1 and RGG2. The RGG2 region has been previously reported as responsible for interacting with splicing factor hTra2-β1 (Hofmann and Wirth 2002; Kanhoush et al. 2010). Chapter 2 of this thesis demonstrates, to our knowledge, the first interaction between RGG-regions and RNA Polymerase II in its study of hnRNPG. Additionally, chapter 2 addresses co-transcriptional m^{6}A-dependent regulation of alternative splicing by class II reader hnRNPG and acts to expand known interactions and functions of class II reader proteins.

Class III

Class III m^{6}A-reader proteins refer to proteins that use common RNA-binding domains and their flanking regions to specifically recognize m^{6}A-modified RNA (Zhou and Pan 2018). Insulin-like growth factor-2 mRNA-binding proteins 1, 2, and 3 (IGFBP1-3)

were identified as putative m^{6}A-reader proteins that bind m^{6}A-modified RNA at consensus GG(m^{6}A)C motifs (H. Huang et al. 2018). Once bound, the IGF2BPs are thought to stabilize a broad class of transcripts, thereby opposing the function of Class I m^{6}A reader protein YTHDF2 (H. Huang et al. 2018; Xiao Wang et al. 2014). Additionally, IGF2BPs are oncogenic and upregulated in most cancers; in the case of renal cell cancer tumorigenesis, their expression is driven by transcription factor EGR2 (Ying et al. 2021)

IGF2BP1-3 each contain two RNA Recognition Motifs (RRMs) and four hnRNPK-homology (KH) domains (H. Huang et al. 2018); KH-domains are among the most abundant RNA binding proteins in the human proteome, present in proteins such as hnRNPK (Valverde, Edwards, and Regan 2008). Accordingly, the KH-domains of the IGF2BPs impart binding specificity towards several transcripts (Bell et al. 2013). However, the role of the IGF2BPs' KH-domains expands beyond the canonical function of these domains in sequence specificity, as they have additionally been shown responsible for specificity towards the m^{6}A modification itself (H. Huang et al. 2018). It is this specificity in binding mediated through a common binding motif which alone does not bind m^{6}A, like the KH-domain (H. Huang et al. 2018), that differentiates Class III reader proteins from Class I, who have YTH-domains which alone are sufficient for direct m^{6}A-binding. While the exact mechanism of m^{6}A-binding by the IGF2BPs has yet to be definitively proven, hydrophobic residues in the IGF2BPs KH-domain and the surrounding regions may help offset the solvation penalty imposed by deposition of the hydrophobic m^{6}A moiety and thus drive-forward their interaction (Roundtree, Evans, et al. 2017).

1.1.2 Pseudouridine (Ψ)

Identified in 1951 (COHN and VOLKIN 1951), among the first ten modifications observed sequencing the first human biological RNA sample (HOLLEY et al. 1965), and the most abundant post-transcriptional modification, pseudouridine (Ψ) is perhaps now best known for its part of both Pfizer-BioNTech and Moderna Therapeutics COVID-19 vaccines. These vaccines contain Ψ derivative N1-methyl-pseudouridine modified mRNA, which seemingly increased efficacy relative to the unmodified CVnCoV mRNA vaccine (Morais, Adachi, and Yu 2021). This efficacy is thought partially explained by the enhanced translational capacity of Ψ as part of a non-immunogenic vector (Karikó et al. 2008). Outside of its use in vaccines, Ψ is found in such abundance that it has been described as "the fifth RNA nucleotide," (X. Li, Ma, and Yi 2016). This abundance is owed to Ψ's high enrichment both in tRNA and rRNA and to it being the second most abundant internal mRNA modification (S. Huang et al. 2021). Structurally, Ψ has an extra hydrogen bond when compared to uridine which can stabilize higher-order structures in transcripts (Kierzek et al. 2014).

Ψ is "written" through the actions of both RNA-dependent and RNA-independent pseudouridine synthases, which can install the modification in coding and non-coding RNA through isomerization of uridine (Hamma and Ferré-D'Amaré 2006). RNA-dependent writers are guided by H/ACA small nucleolar ribonucleoproteins (snoRNPs) to their target; in humans, RNA-dependent DKC1 relies on snoRNPs for localization to rRNA and snRNP targets (N. J. Watkins et al. 1998; Garus and Autexier 2021). Pseudouridinylated snRNPs then act in DKC1-dependent alternative splicing (Garus and Autexier 2021), demonstrating the importance of Ψ in splicing. Additionally, DKC1 has been studied as it relates to health

and disease; the protein is associated with nucleolar prominence and is a predictor of a poor breast cancer prognosis (Elsharawy et al. 2020).

RNA-independent or "stand-alone" Ψ writers are further divided by family, largely defined by homology with bacterial TruA, TruB, TruD, and RluA proteins (Hamma and Ferré-D'Amaré 2006). These Ψ writers include pseudouridine synthase (PUS) enzymes which can exhibit both sequence and structural specificity. Human PUS enzymes dynamically act on uridine not only in cytosolic and mitochondrial tRNA, but also within mRNA transcripts; such activities act in response to the environment (Carlile et al. 2014; Lovejoy, Riordan, and Brown 2014; S. Schwartz, Bernstein, et al. 2014). Human TruA family members include PUS1, PUSL1, and PUS3, human TruB family members include DKC1, TRUB1, and TRUB2, human PUS10 is a member of the Pus10 family, while human RPUSD1-4 are members of the RluA family (**Table 1.1.2**).

While the majority of Ψ writers act on rRNA and tRNA, four human Ψ writers (PUS1, TRUB1, PUS7, and RPUSD4) have been shown to act on mRNA substrates (Martinez et al. 2022); the functional significance of these Ψ mRNA writers has been demonstrated through observed regulation of alternative splicing by Ψ PUS1, PUS7 and RPUSD4 (Martinez et al. 2022). However, it has been shown that the predominant Ψ writer for mRNA in mammalian cell culture is TRUB1 with the second dominant writer being PUS7 (Safra et al. 2017). Chapter 5 of this thesis investigates TRUB1, demonstrating its importance in regulating translation and crosstalk with m^6A.

Unlike m^6A, no Ψ eraser has currently been identified. Ψ contains a stable C-C glycosidic bond, which thus reduces even the energetic favorability of non-enzymatic reversibility. There are no known Ψ reader proteins in humans, with the only suggested

reader protein being Prp5 RNA helicase in yeast (Wu et al. 2016). Interestingly, RNA containing known Ψ sites are enriched in binding to yeast methionine $tRNA^{Met}$ synthetase (MetRS); it has been suggested, as this enrichment is analogous to reader proteins and their substrate, MetRS acts as a sort of Ψ reader (Levi and Arava 2021). Regardless, no Ψ reader has been identified in mammals.

While predominately studied as it relates to immunology, Ψ has additional implications on human health, both with regard to its potential as a diagnostic tool, and given its involvement in several cancers and diseases (Cerneckis et al. 2022). Additionally, Ψ within mRNA has been suggested to affect translation *in vitro* (Eyler et al. 2019) and in yeast (Levi and Arava 2021). Chapter 5 of this thesis details Ψ's involvement in regulating translation in human HEK293T cell culture, in addition to its crosstalk with m^6A on translation efficiency and modification concurrence.

Gene	Family	Mechanism	Localization	Target	Target Sequence
PUS1	TruA	Stand-alone	Mitochondria, Nucleus	tRNA, mRNA	
PUSL1	TruA	Stand-alone	Mitochondria	tRNA	
PUS3	TruA	Stand-alone	Cytoplasm, Nucleus	tRNA	
DKC1	TruB	H/ACA Box sno-RNP	Nucleus	rRNA, snRNA	
TRUB1	TruB	Stand-alone	Nucleus, Cytoplasm, Mitochondria	tRNA, mRNA	GUΨYNANNC
TRUB2	TruB	Stand-alone	Mitochondria, Nucleus	tRNA	
PUS7	TruD	Stand-alone	Nucleus	rRNA, mRNA	UNΨAR
PUS7L	TruD	Stand-alone	Nucleus	rRNA	
PUS10	Pus10	Stand-alone	Nucleus, Mitochondria, Cytosol	tRNA	
RPUSD1	RluA	Stand-alone	Cytosol	tRNA, rRNA	
RPUSD2	RluA	Stand-alone	Nucleus, Mitochondria	tRNA	
RPUSD3	RluA	Stand-alone	Mitochondria, Nucleus	tRNA	
RPUSD4	RluA	Stand-alone	Mitochondria, Nucleus	tRNA, 16S rRNA, mRNA	

Table 1.1.0.2 The Human Genome Encodes 13 Ψ Writers. The 13 Ψ writers in the human genome (X. Li, Ma, and Yi 2016) are defined by their family and relation to known E. Coli Ψ writers. Five human Ψ writers, PUS1, TRUB1, PUS7, and RPUSD4 have been shown to act on mRNA substrates (Martinez et al. 2022) with TRUB1 being the dominant writer for mRNA in mammalian cell culture (Safra et al. 2017).

1.2 tRNA processing

Noncoding RNA (ncRNA) makes up 90% of cellular RNA, with the most abundant member of this class being transfer RNA (tRNA) (Phizicky and Hopper 2010). tRNA is an adapter connecting genetic information embedded in the mRNA triplet code to protein production; as such, tRNA represents a pivotal target for the regulation of translation. This regulation of tRNA is achieved at the level of expression, charging, and modification (C. P. Watkins et al. 2022). Before translation, the tRNA acceptor stem 3' CCA end of the tRNA is aminoacylated, or charged, with its cognate amino acid through the regulated action of aminoacyl-tRNA synthetases. The anticodon loop of tRNA accordingly contains the anticodon at positions 34, 35, and 36, which pairs with the mRNA codon in the ribosome A-site during translation. The charged tRNA is then able to provide an amino acid to the nascent polypeptide chain, thereby decoding the mRNA (Phizicky and Hopper 2010; Tsutomu Suzuki 2021).

Owing to their small size, structure, and heavily modified body, sequencing tRNA and simultaneously measuring its characteristic properties of abundance, charging, modification, and fragmentation has been challenging. My group established multiplex-small RNA sequencing (MSR-seq), a method for sequencing small RNAs, including tRNAs, which has dramatically increased datasets detailing tRNA properties we can simultaneously collect, thus motivating the study herein. Chapter 4 of this thesis details a preliminary bioinformatic survey of tRNA processing's dependence on the mRNA modification landscape.

1.3.1 tRNA biogenesis

Transcription of tRNA is performed by RNA Polymerase III in humans, which transcribes a precursor pre-tRNA. This precursor contains a 5'-leader processed by RNase P and a 3'-trailer removed by RNase Z. Additionally, the pre-tRNA is modified by the addition of a universal 3'-CCA by TRNT1. Some tRNAs contain introns in the anticodon loop which are processed by the tRNA splicing endonuclease (TSEN) complex, a complex that facilitates interactions with processing enzymes further downstream (Phizicky and Hopper 2010). Mature tRNA are classified by the amino acid present at the universal 3'-CCA, and then further by the sequence variation in either their codon (isoacceptors) or in their sequence outside of the codon (isodecoders); regulation of tRNA isoacceptor and isodecoder abundance ensures proper gene expression.

Human mitochondria encode their own set of 22 mitochondrial tRNAs (mt-tRNA) which are processed independently of human tRNA (Takeo Suzuki et al. 2020). Mitochondrial tRNA are generally processed similarly to bacterial tRNA, flanked as part of a polycistronic transcript. This polycistronic transcript is processed at the 5' end by RNase and at its 3'-end by mitochondrial RNase Z. Both mitochondrial and nuclear RNase Z is derived from the same gene (Rossmanith 2012), localization of human RNase Z^L being isoform-dependent (Rossmanith 2011). Interestingly, expression of RNase Z^L is decreased in cells lacking mitochondrial DNA, suggesting crosstalk between the mitochondrial genome and human RNase Z^L (Mineri et al. 2009). Similar to human cytosolic tRNA, mt-tRNA is heavily modified, including modifications unique to mitochondria (Rossmanith 2012).

1.3.2 Chemical modifications in tRNA

Chemical diversity imparted by modification critically expands tRNA's function. Human tRNA contains 13 chemical modifications on average, which fine-tune many tRNA properties including stability, localization, and folding (C. P. Watkins et al. 2022). Chemical modifications are often encountered in the third codon position of the tRNA anticodon, or the "wobble" position. Whereas the first two anticodon positions pair canonically, position three accommodates non-Watson-Crick interactions, or wobble pairing; wobble pairing expands the ability of tRNA to decode the triplet code via the tRNA anticodon. In the case of wobble inosine (I34), the nucleotide whose presence helped define the original wobble hypothesis (Crick 1966), decoding is expanded to read C, U, and A. For wobble uridine (U34), the site is often modified to various derivatives of 5-methyluridine (m^5U) (Tsutomu Suzuki 2021). Unmodified U34 decodes all four partner nucleotides through wobble pairing, whereas m^5U34 derivatives restrict decoding to codons containing A or G at the third position. Ribosome complex crystal structures have further supported the importance of m^5U34 derivates in cognate pairing (Kurata et al. 2008). Whereas m^5U derivatives restrict the decoding of U34 tRNA, chemical modification of wobble A34 and C34 tRNA expands decoding: for instance, in the cases of human A-to-I editing at A34 and C-to-f^5C conversion in mitochondrial tRNA (Tsutomu Suzuki 2021).

Modifications in tRNA are dynamic and have been shown to respond to the environment. For instance, abundance of methyl-3-cytosine (m^3C) at the wobble position (m^3C32) and in the variable loop (m^3C47D) is coordinated following either heat or arsenite stress (C. P. Watkins et al. 2022). Furthermore, the guanosine derivative queuosine (Q) at the wobble position (Q34) has demonstrated nutritional-dependence (Tuorto et al. 2018),

taurine modifications τm^{5}U34 and τm^{5}s^{2}U34 are accordingly taurine-dependent (Shu et al. 2020), and threonylcarbamoyladenosine (t^{6}A37) is sensitive to CO_2 (Lin et al. 2018). Chapter 4 of this thesis discusses the effect of m^{6}A writers and readers on tRNA modifications, specifically m^{3}C and m^{5}U derivatives.

1.3 Sequencing of RNA modifications

The discovery of RNA modifications predated the technology required for its functional characterization. Appreciation of mRNA modifications' regulatory function has only been afforded by continuous development and refinement of techniques which map these modifications to the transcriptome. Chapter 5 of this thesis describes new development in detection of mRNA modifications by direct sequencing, thereby enabling study of crosstalk between these modifications.

1.3.1. Antibody-based next-generation sequencing of m^6A

Immunoprecipitation-based techniques were the first breakthrough for high-throughput mapping of RNA modifications; the initial 100-nucleotide resolution mapping of m^6A sites (Dominissini et al. 2012; Meyer et al. 2012) was further refined by coupling to chemical or photo-crosslinking, which achieves near single-base resolution and thus allows specific identification of modified sites (K. Chen et al. 2015; Linder et al. 2015). These cross linking immunoprecipitation (CLIP)-based approaches were further expanded in techniques like m^6A-LAIC-seq, which aims to produce stoichiometric quantifications of modified-to-unmodified transcripts in a sample (Molinie et al. 2016).

While MeRIP-seq approaches have greatly expanded understanding of m^6A, coupling high-throughput sequencing to immunoprecipitations brings the inherit challenges of any antibody-based methodology to sequencing. The ideal antibody binds its antigen with high affinity and selectivity; however, the necessity of sterically accessible epitope for both affinity and selectivity precludes antibody-based techniques to reach this

ideal. In practice, antibodies have limited sensitivity and are prone to artifacts from non-specific interactions. In the case of application to m^{6}A-seq, sites mapped are often independent of core m^{6}A methyltransferase complex components (Helm and Motorin 2017). Regardless of these limitations, MeRIP-based approaches have been the most widespread method for sequencing m^{6}A.

1.3.2. Antibody-free next-generation sequencing of m^{6}A

Given the disadvantages of antibody-based methodology, several antibody-free methods for sequencing and mapping m^{6}A have been developed. MAZTER-seq and m^{6}A-ref-seq take advantage of endoRNases MazF and ChpBK to cut unmethylated RNA at ACA or UAC motifs with specificity over methylated motifs (Garcia-Campos et al. 2019; H.-X. Chen et al. 2022). MAZTER-seq is highly specific and requires comparatively little input, but its ACA-specificity limits application to 16% of mapped mammalian m^{6}A sites (Garcia-Campos et al. 2019).

Chemical labeling has also been used to increase m^{6}A mapping resolution. DART-seq takes advantage of the YTH-domain's ability to specifically interact with m^{6}A-modified RNA by fusing the domain to cytidine deaminase APOBEC1. This enables localization of APOBEC1 to YTH-bound m^{6}A-sites, thereby inducing a nearby C-to-U deamination signature near the bound site. Furthermore, DART-seq is compatible with long-read RNA-sequencing and thus distinguishes itself from methods solely compatible with short-read next-generation sequencing (Meyer 2019).

Metabolic labeling techniques have been developed for further specificity and quantification. m^{6}A-SEAL-seq uses oxidation by FTO for biotin-labeling with modified DTT.

This adduct enables streptavidin pulldown and downstream sequencing (Y. Wang et al. 2020). Another metabolic labeling approach is m^{6}A-label-seq, where cells are fed SAM analogue Se-allyl-L-selenohomocysteine, which leads to the generation of N^6-allyladenosine (a^{6}A) at m^{6}A-sites (Shu et al. 2020). Selective labeling techniques have also further enable quantification of m^{6}A stoichiometry in m^{6}A-SAC-seq (L. Hu et al. 2022).

1.3.3. Chemical sequencing of pseudouridine (Ψ)

Chemical-based methods have historically been used while sequencing Ψ, commonly employing N-cyclcohexyl N'-(2; morpholinoethyl) carbodiimide (CMC). CMC acts on Ψ, uridine, and guanosine, but the CMC-Ψ adduct, formed at the N3 position of Ψ, has alkaline hydrolysis resistance specific for this product. This can be taken advantage of during reverse transcription, as termination occurs one nucleotide adjacent to the CMC-Ψ adduct (S. Schwartz, Bernstein, et al. 2014; X. Li et al. 2015; Carlile et al. 2014). While CMC-seq has been critical in advancing study of Ψ, there has been discrepancies between independent Ψ mapping datasets; for instance, one comparison of CMC-seq Ψ mapping showed only 13 overlapping sites. Additionally, CMC-seq maps only a small amount of Ψ residues identified by LC-MS/MS. Detection and enrichment has been improved through methods like the azide-modified CMC-based CeU-seq, but resolution and stoichiometric information is still limited (X. Li et al. 2015).

Quantitative, transcriptome-wide mapping of Ψ was improved by the development of bisulfite-induced deletion sequencing (BID-seq), which accordingly reads deletion signature induced by bisulfite treatment. BID-seq provides stoichiometric information on the fraction of transcripts modified at a particular site (Q. Dai et al. 2023). This

stoichiometric information further supported TRUB1 as the dominant Ψ writer of the 13 identified Ψ mRNA writers (Q. Dai et al. 2023; Safra et al. 2017). Additionally, BID-Seq has revealed previously unidentified Ψ sites within mammalian mRNA stop codons (Q. Dai et al. 2023).

1.3.4 Long-read sequencing

Next-generation sequencing approaches historically sequence short read fragments, typically within the range of 25-300 bps. Short-read sequencing techniques have been instrumental in mapping sites of chemical modifications through the transcriptome and gaining insight into the modification landscape, but these sequencing techniques are not without limitation. Samples are digested and amplified during preparation, causing a loss of relative location and biasing reads for those preferentially amplified. A particular case-study of the difficulty of short-read sequencing is sequencing structural variation due to genomic repeats in DNA. Inaccuracies are abundant when aligning reads from elements of high structural variation resulting from this region of inaccessibility (Ho, Urban, and Mills 2020). Thus, mapping of centromeres, telomeres, and other arrays of tandem repeats has been limited when using short-read DNA sequencing approaches; likewise, higher-order structure in RNA limits accessibility during sequencing. When applied to RNA, the digestion step additionally limits mapping of splice isoforms, insertions, and multiple chemical modifications within the same transcript.

By contrast, long-read sequencing is the continuous sequencing of a single stretch of nucleotides, ranging from 1000 bp – 4 Mbp for DNA. In the example of genomic structural variation, long-read sequencing greatly enhanced their mapping (Ho, Urban, and Mills

2020; Chaisson et al. 2019), and even revealed preliminary population-scale variations in height, cholesterol, and anemia in these regions (De Coster, Weissensteiner, and Sedlazeck 2021). As of now, the two major long-read sequencing platforms are Oxford Nanopore Technology (ONT) PromethION and Pacific Biosciences (PacBio) high fidelity (HiFi), both methods do not require amplification of the nucleic acid inputs, in contrast to NGS approaches.

PacBio HiFi

PacBio HiFi, or single-molecule real-time (SMRT) sequencing uses a "SMRTbell," – a circular DNA molecule template containing a dsDNA insert flanked by hairpin adapters to sequence the DNA by fluorescent dNTP and measurement of emission following laser excitation of the fluorophore incorporation within a SMRT cell (Logsdon, Vollger, and Eichler 2020). Notably, PacBio long-read sequencing relies on DNA Polymerase and circularized DNA and is thus incompatible with direct RNA sequencing. However, full-length RNA can be sequenced indirectly by PacBio via "Iso-Seq," a cDNA synthesis and PCR amplification-based approach. (Au et al. 2013; Logsdon, Vollger, and Eichler 2020). In such an approach, the lack of compatibility with native RNA molecules loses direct modification information. PacBio DNA sequencing yields reads shorter than those from Nanopore platforms at 15-20 kbp, but its reads map with higher accuracy (Logsdon, Vollger, and Eichler 2020).

Oxford Nanopore Technology

Oxford Nanopore Technology immediately distinguishes itself from PacBio sequencing by direct sequencing compatible with both DNA and mRNA of both cDNA and

mRNA molecules. Nanopore sequencing platform PromethION can produce DNA reads of up to 4 Mbp (Payne et al. 2019). Nanopore sequencing works by attachment of the nucleic acid to an adapter loaded in motor proteins, allowing the proteins to unwind the nucleic acid to single-stranded form as electrical current drives it through the nanopores imbedded in the flow cell membrane. A given nucleotide and its neighbors generate characteristic change in current when being translocated through the pore due to their unique shape; the change in current through time can be analyzed in real-time and deconvoluted into sequencing information (Oikonomopoulos et al. 2020).

The ONT nanopore platform was used to directly sequence native RNA first in 2018, demonstrating reading of unmodified bases, m^6A, and m^5C in RNA and later for m^7G and Ψ (Chaisson et al. 2019; Smith et al. 2019; S. Huang et al. 2021). Direct mapping of modifications via nanopore is not limited by epitope binding as in antibody-based approaches, nor does it rely on chemical treatment which may lead to an abundance of false positives and negatives (S. Huang et al. 2021). The long reads afforded by nanopore sequencing further enables mapping of modifications within a single transcript, thus providing more detailed mapping in instances such as within splice isoforms. Learning-based approaches for detection of modifications as part of native nanopore sequencing have enabled increasingly precise mapping using assorted variables deciphered from training data (S. Huang et al. 2021). Chapter 5 of this thesis details NanoSPA, a nanopore based approach for the simultaneous, direct sequencing of m^6A and Ψ as part of a single transcript, demonstrating crosstalk between these two modifications on translation and their concurrence.

Chapter 2

Development of a biochemical assay to demonstrate low-complexity m6A-reader protein hnRNPG simultaneously binds phosphorylated CTD of RNAPII and RNA

Acknowledgement: This chapter is derived from an article published in Molecular Cell by Elsevier(Zhou et al. 2019). The authors in that article were: Katherine I. Zhou, Hailing Shi, Ruitu Lyu, Adam C. Wylder, Zaneta Matuszek, Jessica N. Pan, Chuan He, Marc Parisien, and Tao Pan. Author contributions: Conceptualization, K.I.Z. and T.P.; Methodology, K.I.Z., H.S., and T.P.; Software, M.P. and R.L.; Formal Analysis, K.I.Z., R.L., and M.P.; Investigation, K.I.Z., H.S., A.C.W., Z.M., and J.N.P.; Writing – Original Draft, K.I.Z.; Writing – Review & Editing, K.I.Z., H.S., C.H., M.P., and T.P.; Supervision, C.H. and T.P.

2.1 Introduction

The carboxy-terminal domain (CTD) of RNA polymerase II (RNAPII) plays a crucial role coordinating co-transcriptional processes with transcription (Bentley 2014). The CTD is composed of tandem heptad repeats that undergo various post-translational modifications, including phosphorylation of Ser5 by cyclin-dependent kinase 7 (CDK7) during promoter escape, and phosphorylation of Ser2 by CDK9 upon promoter-proximal pause release (Harlen and Churchman 2017). The various phosphorylation states of the CTD coordinate co-transcriptional processes with the stages of transcription:

Ser5 phosphorylated CTD (S5P-CTD) recruits 5′ capping factors at the 5′ end of the gene; Ser2 phosphorylated CTD (S2P-CTD) recruits constitutive splicing factors in the gene body and 3′ cleavage and polyadenylation factors at the 3′ end of the gene (Hsin and Manley 2012). While the CTD also contributes to alternative splicing, few alternative splicing factors are known to directly interact with the CTD (Muñoz, de la Mata, and Kornblihtt 2010; Kornblihtt et al. 2013). Although alternative splicing factors have been proposed to regulate splicing by modulating RNAPII pausing (Kornblihtt et al. 2013), previous studies have primarily examined this mode of splicing regulation at the transcriptome or interactome level (Ip et al. 2011; Shukla et al. 2011) and have not yet elucidated precise molecular mechanisms.

m^{6}A is deposited co-transcriptionally (Knuckles et al. 2017; Louloupi et al. 2018) and depletion of m^{6}A readers YTHDC1, hnRNPC, hnRNPG, and hnRNPA2B1 leads to changes in transcriptome-wide alternative splicing patterns(Kasowitz et al. 2018; N. Liu et al. 2015, 2017; Patil et al. 2016; Xiao et al. 2016b; Alarcón et al. 2015) with hnRNPG and YTHDC1 knockdown having the largest effects on alternative splicing. Recently, the extent to which m^{6}A functions in mRNA splicing has been called into question based on the low abundance of m^{6}A in introns within chromatin-associated mRNA (Ke et al. 2017). However, an independent study identified a much higher abundance of m^{6}A in nascent RNA (Louloupi et al. 2018). These conflicting results on the function of m^{6}A can be resolved by elucidating the concrete mechanisms behind m^{6}A-dependent regulation of alternative splicing.

Heterogeneous nuclear ribonucleoprotein G (hnRNPG) is an m^{6}A reader protein that binds RNA through RRM and Arg-Gly-Gly (RGG) motifs. Here, we show that hnRNPG

directly binds to the phosphorylated carboxy-terminal domain (CTD) of RNA polymerase II (RNAPII) using RGG motifs in its low-complexity region. Through interactions with the phosphorylated CTD and nascent RNA, hnRNPG associates co-transcriptionally with RNAPII and regulates alternative splicing transcriptome wide. m^{6}A near splice sites in nascent pre-mRNA modulates hnRNPG binding, which influences RNAPII occupancy patterns and promotes exon inclusion. Our results reveal an integrated mechanism of co-transcriptional m^{6}A-mediated splicing regulation, in which an m^{6}A reader protein uses RGG motifs to co-transcriptionally interact with both RNAPII and m^{6}A-modified nascent pre-mRNA to modulate RNAPII occupancy and alternative splicing.

The hnRNPG (also known as RBMX) protein is unique among known m^{6}A reader proteins in that it uses Arg-Gly-Gly (RGG) motifs in a low-complexity region to selectively bind m^{6}A-modified RNA (N. Liu et al. 2017). The low-complexity sequence of hnRNPG includes two regions each containing three RGG motifs (RGG1 and RGG2). RGG regions are among the most common RNA-binding domains (Gerstberger, Hafner, and Tuschl 2014) and tend to exhibit degenerate specificity for RNA (Ozdilek et al. 2017; Thandapani et al. 2013). In addition, RGG regions can function in protein-protein interactions (Thandapani et al. 2013). The RGG2 region of hnRNPG functions both in the selective binding of hnRNPG to m^{6}A-modified RNA (N. Liu et al. 2017) and the interaction of hnRNPG with the splicing factor hTra2-β1 (Hofmann and Wirth 2002; Kanhoush et al. 2010). The function of the RGG1 region of hnRNPG has not yet been elucidated. To our knowledge, no RGG region has previously been shown to interact directly with RNA polymerase II.

Here, we investigate how the CTD of RNAPII, the RGG regions of hnRNPG, and m^{6}A modification of nascent mRNA act together to regulate alternative splicing. We identify an

m^{6}A-mediated mechanism for alternative splicing regulation, in which the co-transcriptional interactions of hnRNPG simultaneously with RNAPII CTD and with nascent RNA influence RNAPII occupancy and affect exon inclusion. Through development of a biochemical assay for studying low-complexity protein interactions, we demonstrate a direct interaction between the RGG regions of hnRNPG and the phosphorylated CTD of RNAPII, showing that hnRNPG can interact with both nascent RNA and RNAPII CTD. We also demonstrate that three distinct regions of hnRNPG function to regulate alternative splicing, and hnRNPG-bound m^{6}A and RNAPII occupancy occur in specific patterns near hnRNPG-regulated exons. Our results lead to a model in which hnRNPG interacts co-transcriptionally with nascent RNA and RNAPII CTD, while m^{6}A in regulated exons modulates splicing through the interplay of hnRNPG binding and RNAPII occupancy.

2.2 Results

2.2.1 Development and validation of low-complexity spin assay

We found that RNAPII co-immunoprecipitated with hnRNPG from whole cell or chromatin extracts of human embryonic kidney (HEK) 293T cells; this co-immunoprecipitation decreased with hnRNPG knockdown (Zhou et al. 2019). This result is supported by a study of the RNAPII CTD interactome study, which showed enrichment of hnRNPG following immunoprecipitation of S2P or S5P RNAPII (Nojima et al. 2018). Since the RNAPII CTD is a docking site for many RNA processing factors and interacts with the low-complexity regions of multiple RNA-binding proteins (Burke et al. 2015; Harlen and Churchman 2017; Kwon et al. 2013; J. C. Schwartz et al. 2013), we examined whether

hnRNPG, which has an extensive low-complexity region (Dosztányi et al. 2005), can directly interact with the CTD. However, the tendency of hnRNPG to form large assemblies complicated the quantitative measurement of RNA binding using native gel shift or nitrocellulose filters (Zhou et al. 2019).

Therefore, I aimed to establish a biochemical binding assay compatible with low-complexity proteins. The "spin-down" assay (Figure 2.2.1a) I developed takes advantage of our observation that hnRNPG is soluble at 0.5 M NaCl but forms larger aggregates upon transfer to 0.15 M NaCl, as these aggregates can be pelleted in small volume mixtures by centrifugation, which allows separation of aggregated protein complexes from soluble components. The supernatant, containing the soluble fraction, can be separated from the pellet, which contains the aggregated protein complexes, allowing for comparison by gel electrophoresis.

This assay was applied to hnRNPG alone in a series of different salt buffers to understand the NaCl-dependence on aggregation, which showed hnRNPG was distributed between both soluble and pellet fractions in all NaCl concentrations excluding 0.5M (**Figure 2.2.1c**). Comparatively, GST-tagged RNAPII C-terminal domain (GST-CTD) was exclusive to the soluble fraction when added to the binding mixture (**Figure 2.2.1d**), suggesting it did not interact with hnRNPG aggregates. To test the dependence on RNAPII CTD phosphorylation state on interaction with hnRNPG, phosphorylated GST-CTD (GST-pCTD) was prepared by treatment with CDK7 or CDK9 prior to the spin-down assay, which phosphorylated Ser5 (S5P) or Ser2 (S2P), respectively. GST-pCTD remained soluble following phosphorylation by either kinase (**Figure 2.2.1e**). In conclusion, the data here

presented establishes utility of a small volume, low-complexity spin assay in visualization of soluble and aggregated protein distributions.

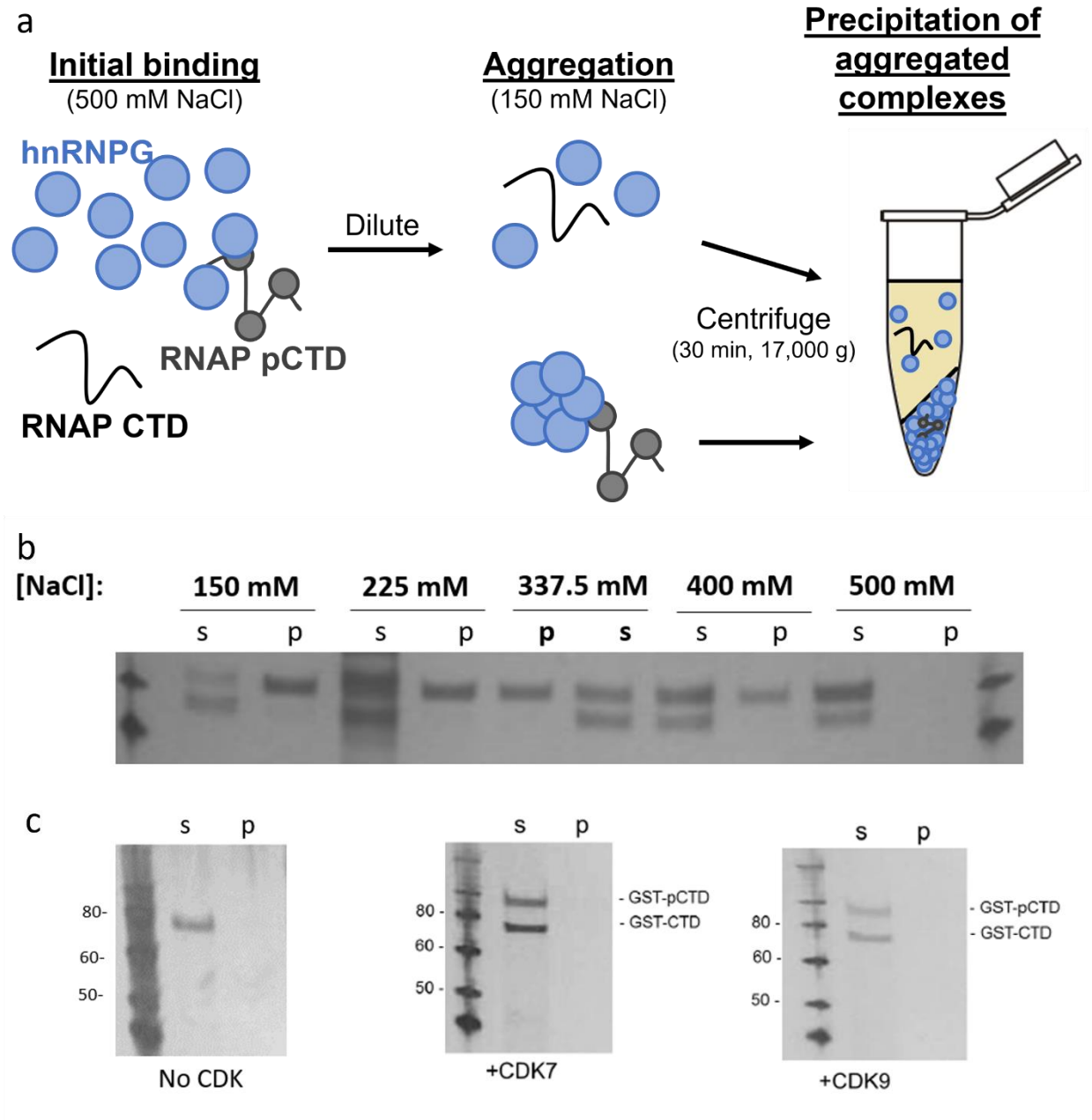


Fig 2.2.1 Development and validation of a spin-assay to measure interactions of low-complexity proteins. (a). Spin-assay Schematic. Initial binding of hnRNPG and binding partners (shown: RNAPII CTD, RNAPII pCTD) is done in 0.5 M NaCl buffer. Complex formation is induced by diluting the initial binding mixture into a lower NaCl concentration (shown: 0.15 M). Complexes are then centrifuged for 30 minutes at 17,000g, pelleting the aggregates. The supernatant containing the soluble (s) fraction can then be removed and

the pellet (p) containing hnRNPG-bound aggregates can be resuspended allowing downstream analysis. (b). Silver-stained denaturing protein gel showing an hnRNPG spin-assay salt titration. Initial binding performed at 0.5 M before diluting to the salt concentrations listed (c). Silver-stained denaturing protein gel of GST-CTD spin assay following no CDK treatment (left, S0P), CDK7 treatment (S5P, middle) and CDK9 treatment (S2P, right) showing GST-pCTD and GST-CTD remain exclusive to soluble fraction.

2.2.2 hnRNPG can accommodate simultaneous binding to RNAPII pCTD and RNA

Next, I used the spin assay to demonstrate direct interaction between hnRNPG and GST-pCTD *in vitro*. Following treatment with either CDK7 or CDK9, GST-pCTD co-precipitated with hnRNPG while unphosphorylated GST-CTD remained soluble (**Figure 2.2.2a, b**). Since hnRNPG is known to interact with RNA, including by acting as an m^6A reader protein, I sought to establish whether hnRNPG's binding to GST-pCTD exhibited RNA-dependence. Addition of a known RNA ligand of hnRNPG, an A- or m^6A-containing 34-nt hairpin from the long noncoding RNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) (Liu et al., 2017), did not affect co-precipitation of GST-pCTD in pelleted aggregate complexes. However, addition of either MALAT1-A or MALAT1-m^6A increased the fraction of hnRNPG found in the pellet (**Figure2.2.2c**), implying protein-RNA interactions help drive aggregation of hnRNPG. Taken together, these results demonstrate that hnRNPG has much higher affinity for phosphorylated than for unphosphorylated-CTD, and pCTD binding is unaffected by RNA binding.

While these protein stains demonstrate hnRNPG's binding to the phosphorylated RNAPII CTD is RNA-independent, they do not demonstrate whether hnRNPG accommodates simultaneous binding to both RNA and the pCTD. To detect the reciprocal relationship, the pCTD-dependence of hnRNPG-RNA interactions, the spin assay was

adapted to detect labeled RNA in soluble and pellet fractions through filter blotting.^{32}P-labeled MALAT1 hairpin containing A (2515-A) or m^{6}A (2515-m^{6}A) was tested along with varying concentrations of hnRNPG and GST-pCTD. This showed that RNA co-precipitated with hnRNPG aggregates in a pCTD-independent manner (**Figure 2.2.2d**). In conclusion, hnRNPG accommodates binding to phosphorylated RNAPII and RNA independent of one another.

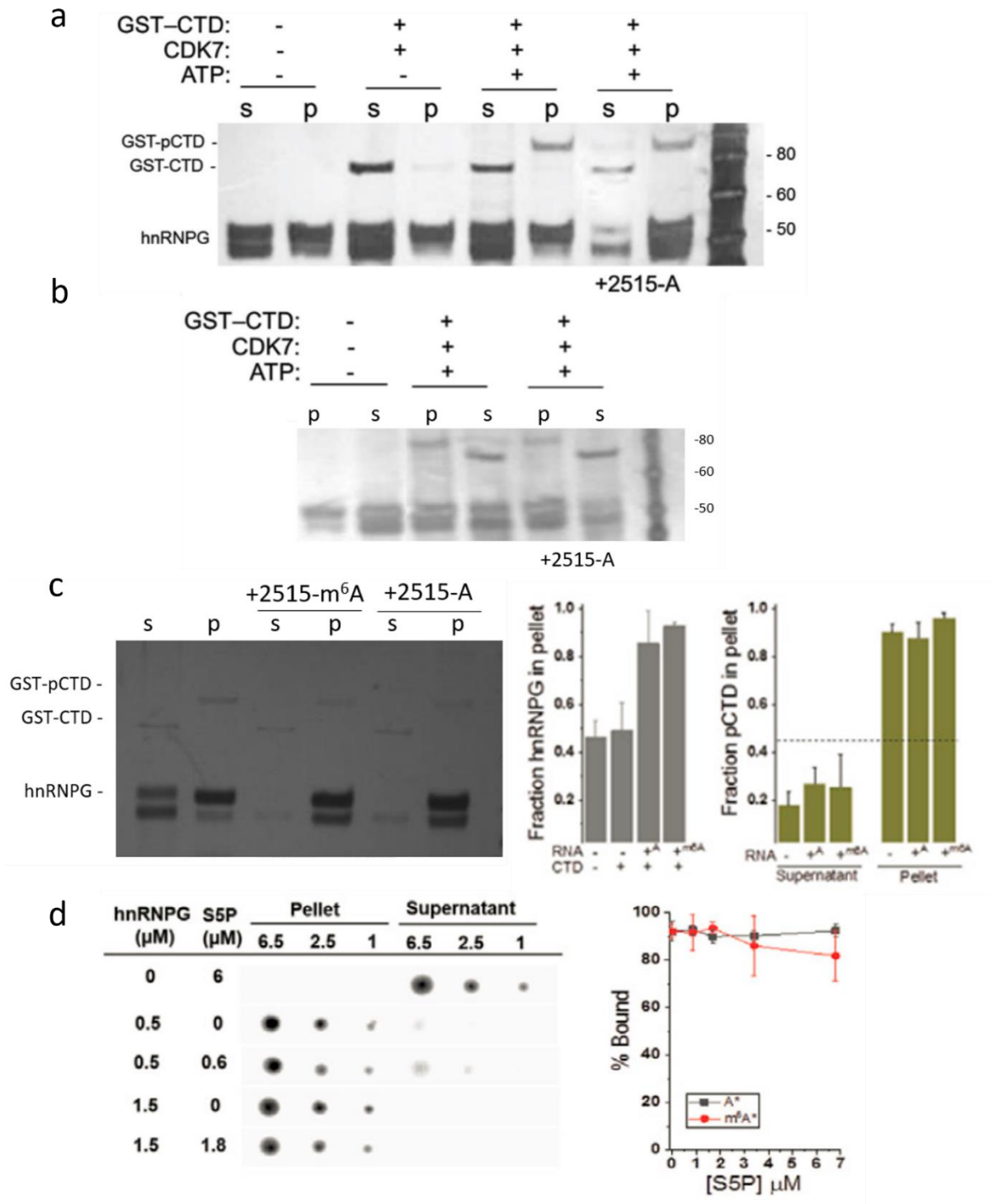


Fig 2.2.2 Spin-assay shows hnRNPG can accommodate simultaneous binding to RNAPII and RNA (a). Silver-stained denaturing protein gel showing GST-pCTD binding (S5P by CDK7) to recombinant hnRNPG by spin-down assay. s, supernatant; p, pellet; +2515-A, with equimolar MALAT1 RNA hairpin. and m^6A superscripts correspond to addition of MALAT1 RNA with A or m6A. hnRNPG runs as two bands due to variable N-

glycosylation. (b). Same as a, but showing S2P binding (CDK9). (c). Silver-stained denaturing protein gel showing increased hnRNPG pelleting following +2515-A, +2515-m^6A treatment. Fraction pCTD in the pellet refers to [GST-pCTD]/([GST-pCTD] + [GST-CTD]). Dashed line shows the fraction of hnRNPG in the pellet in the absence of GST-pCTD or RNA. Error bars: ±1 SD; n = 3 replicates. (d). Filter spotting showing RNA binding to recombinant hnRNPG by spin-down assay. Plot shows percent of bound MALAT1 RNA hairpin at varying concentrations of S5P-peptide. Error bars: ±1 SD; n = 3 independent experiments.

2.2.3 The RRM, RGG1, and RGG2 regions mediate hnRNPG binding to RNA and pCTD

hnRNPG contains three functional domains: an N-terminal structured RRM-domain and two highly disordered RGG (Arg-Gly-Gly) regions found in the protein's 309 amino-acid low-complexity region. To examine the effect of these three functional domains in hnRNPG on RNA and CTD binding, hnRNPG was purified with mutations in the RRM, RGG1, or RGG2 regions (Figure 2.2.3a). I used the spin-down assay to measure the impact of these domains on the proportion of hnRNPG aggregated and its propensity to bind and co-aggregate RNAPII pCTD. Following treatment with either CDK7 or CDK9 to phosphorylate GST-CTD, the RRM and RGG2 mutants, but not RGG1 mutant, bound GST-pCTD in the absence of RNA (**Figure 2.2.3b, c**). On addition of MALAT1 RNA hairpin, the RRM mutant remained capable of pCTD binding, but pCTD co-aggregation with RGG2 mutant hnRNPG was severely diminished. Mutations in RGG1 or RGG2, but not in RRM, reduced pelleting of hnRNPG compared to wildtype. These results were verified by a labeled filter blotting spin-down assay on mutant hnRNPG and ^{32}P-MALAT1 hairpin RNA, where hairpin RNA preferentially co-precipitated with wildtype hnRNPG compared to the three functional domain mutants (**Figure 2.2.3e**). In summary, the RRM and two RGG regions of hnRNPG all function in RNA

binding, while the RGG regions are required for the direct interaction with the RNAPII pCTD.

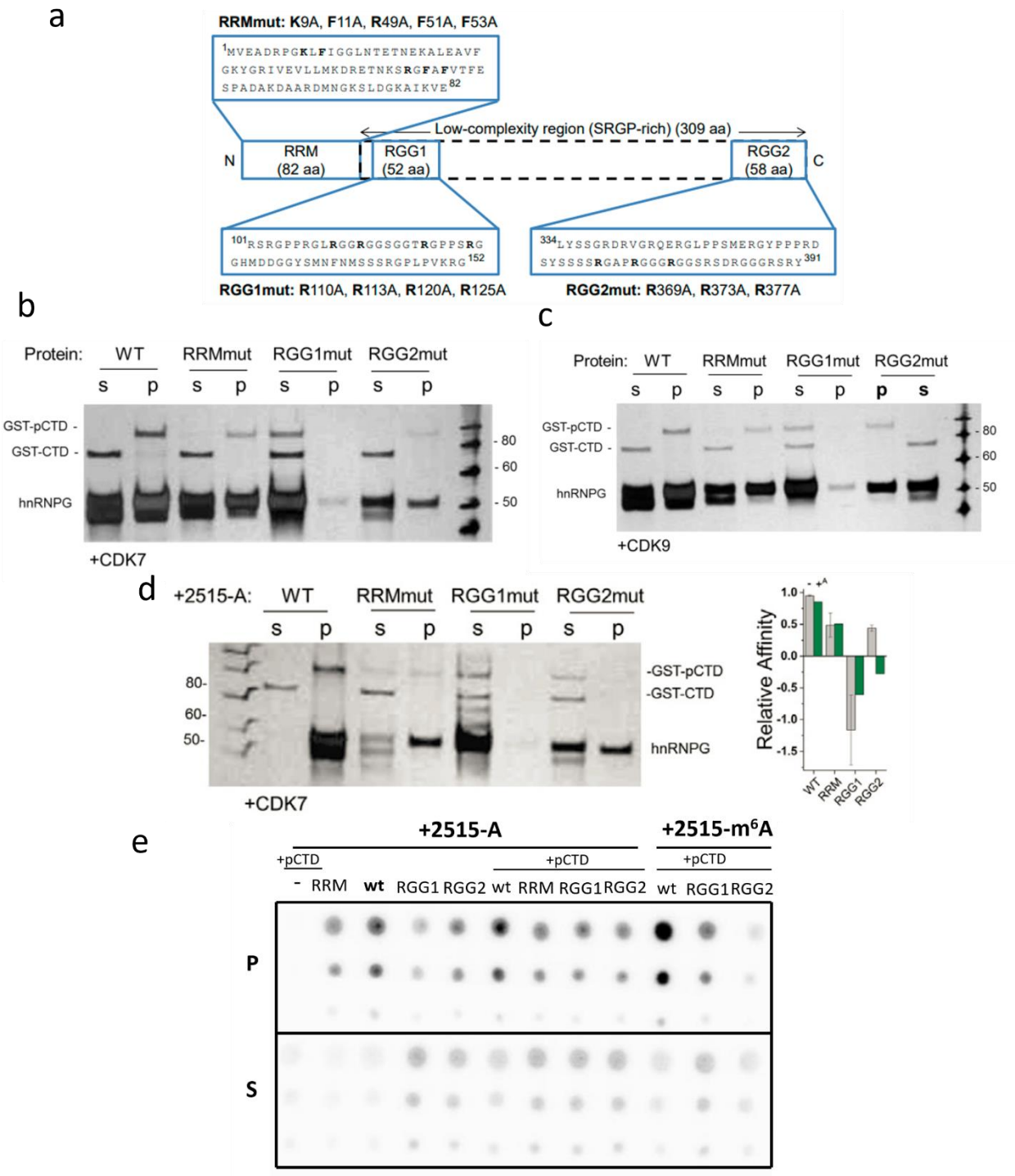


Fig 2.2.3 The RRM, RGG1 and RGG2 regions mediate hnRNPG binding to RNA and pCTD (a). Diagram showing the RRM, RGG1, RGG2, and low-complexity regions of full-

length hnRNPG, as well as the mutations introduced in the RRM, RGG1, and RGG2 regions to generate RRMmut, RGG1mut, and RGG2mut. (b). Silver-stained denaturing protein gel showing pCTD binding (S5P by CDK7) to wild-type and mutant hnRNPG proteins by the spin-down assay. Final NaCl concentration was 0.25 M; final hnRNPG concentration was 1.7 μM. s: supernatant fraction; p: pellet fraction. (c). Silver-stained denaturing protein gel showing pCTD binding (S2P by CDK9) to wild-type and mutant hnRNPG proteins by the spin-down assay. Final NaCl concentration was 0.25 M; final hnRNPG concentration was 1.7 μM. s: supernatant fraction; p: pellet fraction. (d). Silver-stained denaturing protein gel showing GST-pCTD binding (S5P by CDK7) to wild-type and mutant hnRNPG by spin-down assay. s, supernatant; p, pellet. Plot of relative affinity 1− %pCTD(pellet)/%pCTD(input) where y > 0 corresponds to pCTD binding. Error bars: data variation; n = 2 independent experiments. (e). Filter spotting showing RNA binding to recombinant hnRNPG by spin-down assay in the presence or absence of pCTD (S5P by CDK7). +2515-m6A refers to m6A-modified MALAT1 hairpin, whereas 2515-A is the unmodified equivalent; +pCTD

2.2.4 Investigating the phase separation of hnRNPG

The spin-down assay developed herein enables the investigation of protein aggregation's dependence on varying introduced factors, thus allowing investigation of the phase separation properties of low-complexity proteins. Because protein-protein and protein-RNA interactions which help drive aggregation and phase separation are dependent on salt concentration, the applicability of the spin-down assay was demonstrated in multiple salt concentrations. In the case of hnRNPG, its aggregation and binding properties were reproduced in both 0.25 and 0.37 M NaCl (**Figure 2.2.4a**).

The low-complexity region hnRNPG is enriched among the human proteome for repeats of the amino acid sequences "RDY," or "RDxY," (**Figure 2.2.4b**). We hypothesized that these motifs may be involved in driving phase separation of hnRNPG. Therefore, I used the spin-down assay on a 6-FAM-labeled 21-mer peptide consisting of part of the hnRNPG sequence enriched in RDxY motifs (6-FAM-SRDYPSSRDTRDYAPPPRDYT); this peptide alone was sufficient for small amounts of aggregation (**Figure 2.2.5c**). This provides

preliminary evidence that RDxY motifs may be driving aggregation of hnRNPG, but moreover demonstrates the versatility of the spin-down assay here developed.

2.3 Discussion

In this study, we showed that the RNA-binding protein hnRNPG uses RGG regions to directly interact with the phosphorylated CTD of RNAPII. hnRNPG binding to RNA and the RNAPII CTD can occur simultaneously, likely by the assembly of hnRNPG into large complexes. In cells, the RRM, RGG1, and RGG2 regions of hnRNPG functioned in the regulation of alternative splicing. We further found that hnRNPG-bound m^6A sites near the splice sites of regulated exons were associated with increased RNAPII occupancy and exon inclusion. Our results support a model in which hnRNPG assemblies interact co-transcriptionally with nascent RNA and the RNAPII CTD, while m^6A sites in nascent RNA promote hnRNPG binding to modulate RNAPII occupancy and alternative splicing.

One possible mechanism for hnRNPG- and m^6A-dependent alternative splicing regulation is through RNAPII pausing to provide an appropriate time window and spatial resolution to recruit splicing factors. RNAPII pausing during transcription elongation can be modulated by RNA sequence and structure, chromatin state, and RNA-protein interactions (Mayer, Landry, and Churchman 2017). Although RNAPII density by ChIP-seq could reflect RNAPII pausing, ChIP-seq has a limited resolution (100–200 nt) and cannot distinguish between RNAPII pausing and other etiologies of RNAPII occupancy. A future direction could utilize the native elongating transcript sequencing (NET-seq) method to

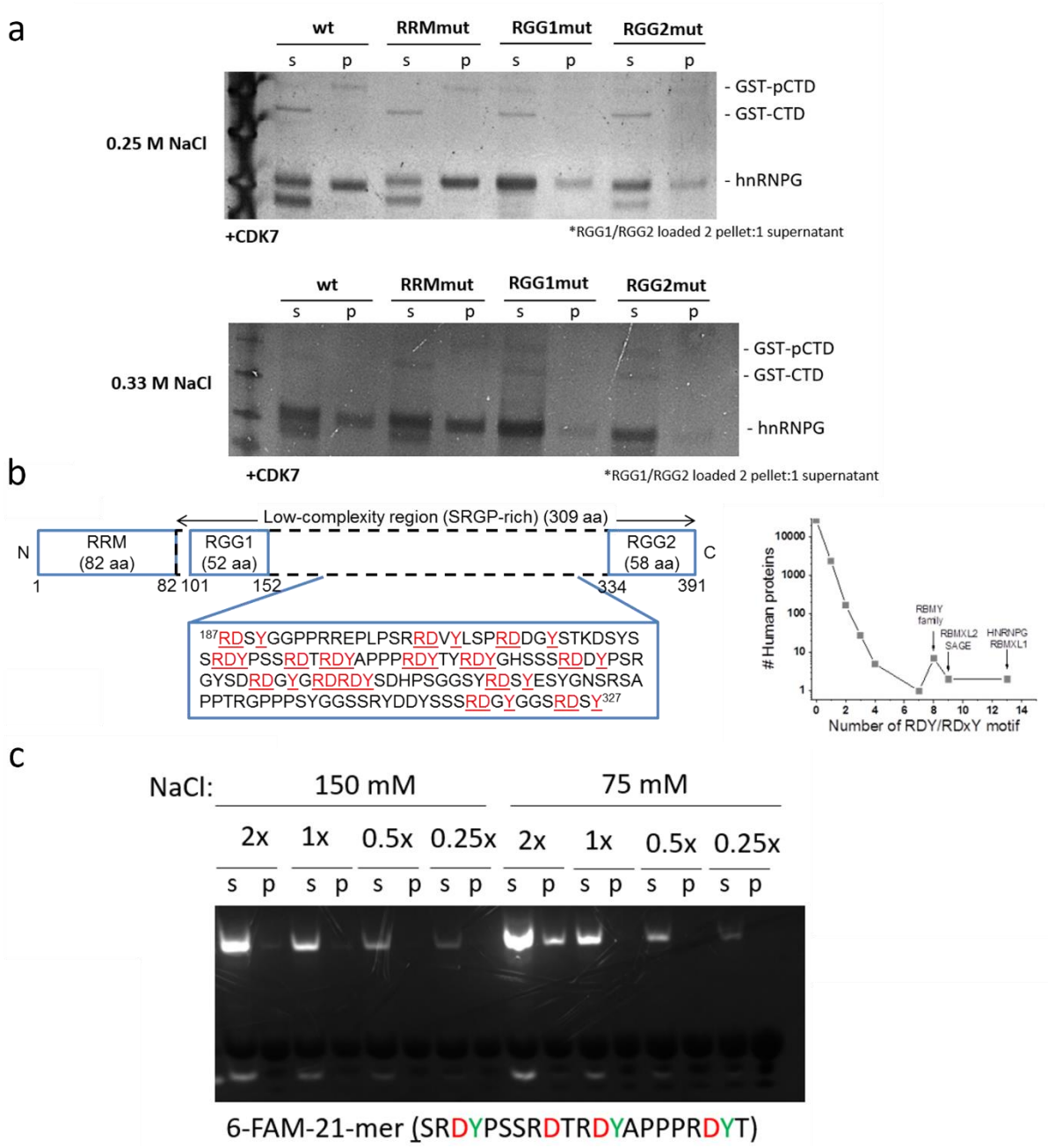


Fig 2.2.4 Spin-down assay as a generalized tool for studying phase separation (a). Silver-stained denaturing protein gel showing pCTD binding (S5P by CDK7) to wildtype and mutant hnRNPG proteins by the spin-down assay. Final NaCl concentration was 0.25 M (top), 0.33M (bottom); the final hnRNPG concentration was 1.7 μM. s: supernatant fraction; p: pellet fraction. (b). Left: hnRNPG domain structure, with the N-terminal 82 amino acid RRM-domain labeled, along with the RGG1 and RGG2 regions within the 309 amino acid low-complexity region. Enlarged is an embedded amino acid sequence enriched in the "RDY/RDxY" motif relative to the human proteome. Right: Quantification of the number of "RDY/RDxY" motifs throughout the human proteome. (c). Native protein gel showing 6-

FAM-21-mer (6-FAM-SRDYPSSRDTRDYAPPPRDYT) aggregation by the spin-down assay. Final concentrations 150 mM (left), 75 mM (right). 1x 6-FAM-21-mer refers to a final concentration 1.4 μM.

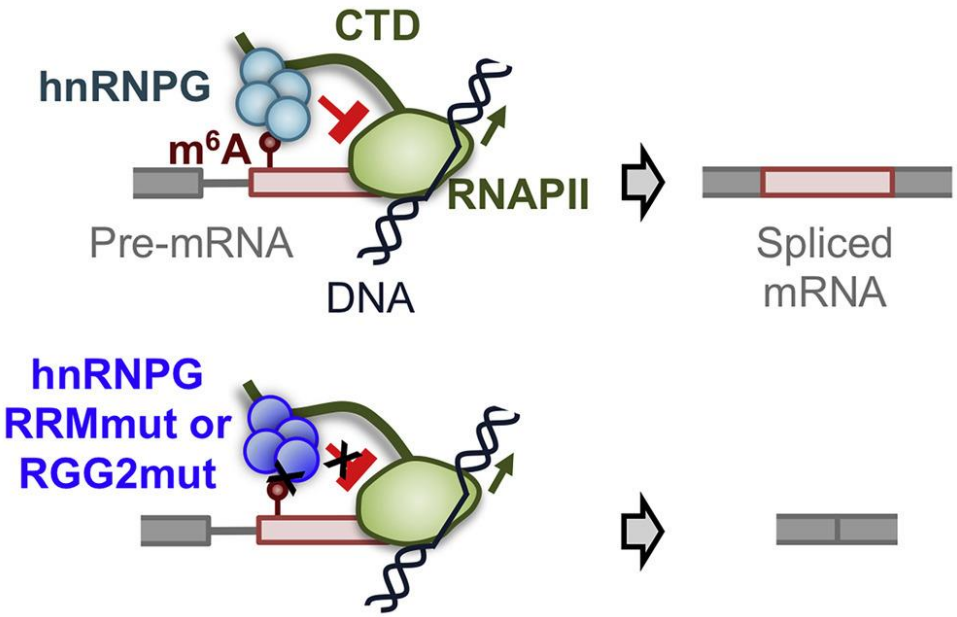


Fig 2.2.5 Model for m6A-dependent regulation of exon inclusion by hnRNPG. Red circle, m^6A site; red box, alternative exon; gray, constitutive exon; black lines, DNA; green, RNAPII with CTD as extended line; cyan, hnRNPG complex. RNAPII transcribes through the splice site, m^6A is installed, and hnRNPG interacts with m^6A, causing RNAPII to increase dwell time downstream of the m^6A site, which results in hnRNPG- and m^6A-dependent exon inclusion.

more accurately measure RNAPII pausing (Mayer, Landry, and Churchman 2017; Nojima et al. 2018) and m^{6}A-dependent splicing regulation. These published NET-seq results have shown that RNAPII density is higher in exons than in introns and is also high around splice sites. NET-seq upon the depletion of hnRNPG, METTL3, and/or METTL14 could provide the single-base resolution needed to firmly establish the potential role of RNAPII pausing in m^{6}A-dependent exon inclusion.

The RNAPII CTD coordinates co-transcriptional RNA processing with transcription by recruiting constitutive RNA processing factors through direct interactions. However, the role of the CTD in recruiting alternative splicing factors, particularly through direct interactions, is less clear (Bentley 2014). Our study demonstrates that an alternative splicing factor, hnRNPG, directly interacts with the RNAPII CTD. The selective interaction of hnRNPG with phosphorylated CTD and the effect of transcription inhibitors on the hnRNPG-RNAPII interaction support our conclusion that hnRNPG interacts with RNAPII co-transcriptionally. We also found that the direct interaction of hnRNPG with phosphorylated CTD depended on two RGG regions in its low-complexity region. To our knowledge, RGG regions have not previously been shown to directly bind the CTD. Because RGG motifs are the second most common RNA-binding motif among mammalian mRNA-binding proteins (Gerstberger, Hafner, and Tuschl 2014), our findings suggest that many other RNA-binding proteins might regulate pre-mRNA processing by interacting directly with the RNAPII CTD through their RGG regions. Moreover, RGG motifs are commonly found in low-complexity regions, so a direct interaction between RGG motifs and the CTD could also have implications for low-complexity region assembly. The CTD is known to interact with low-complexity regions of several RNA-binding proteins, and these interactions have

proposed functions in transcriptional regulation (Harlen and Churchman 2017). hnRNPG forms a distinct interaction with the RNAPII CTD for co-transcriptional splicing regulation.

The abundant m^{6}A modification in mRNA has been implicated in alternative splicing regulation. Most m^{6}A modifications are deposited co-transcriptionally and enriched in exonic regions near splice sites (Barbieri et al. 2018; Ke et al. 2017; Knuckles et al. 2017; Louloupi et al. 2018), but it is unknown how m^{6}A near splice sites can modulate co-transcriptional alternative splicing. Because alternative splicing regulation by RNA-binding proteins depends on their binding site positions (X.-D. Fu and Ares 2014), m^{6}A site positions, which determine m^{6}A reader protein binding sites, likely also influence alternative splicing regulation. We found that hnRNPG-regulated exons were associated with specific patterns of hnRNPG-bound m^{6}A sites and RNAPII occupancy, supporting a model in which hnRNPG binds m^{6}A sites near target exons and modulates RNAPII occupancy to promote exon inclusion. The regulation of alternative splicing involves a complex "splicing code" of *cis*- and *trans*-acting factors (X.-D. Fu and Ares 2014). We have shown that a *cis*-acting RNA modification, m^{6}A, modulates the regulation of alternative splicing by hnRNPG by recruiting hnRNPG to specific sites, while a *trans*-acting protein, hnRNPG, interacts with the transcription machinery and co-transcriptionally modulates RNAPII occupancy to regulate alternative splicing. The m^{6}A-dependent regulation of alternative splicing by hnRNPG demonstrates how the RNAPII CTD, RGG regions, and m^{6}A act together to modulate the co-transcriptional regulation of alternative splicing by a low-complexity m^{6}A reader protein.

2.4 Materials and methods

2.4.1. Cloning and purification of hnRNPG

Repeated attempts by us to overexpress and purify full-length hnRNPG protein in *E. coli* using multiple different plasmid constructs and expression strategies were unsuccessful, likely due to protease cleavage of the low complexity hnRNPG protein in cells. We therefore resorted to overexpressing the full-length protein from the baculovirus expression system where the full-length protein was secreted to avoid proteolysis.

The sequence encoding full-length human hnRNPG protein was amplified from human HeLa cDNA libraries (637203, Clontech) and subcloned into the pGEX-6p-1 vector using BamHI and XhoI restriction sites. Plasmids encoding the hnRNPG mutants RRMmut, RGG1mut, and RGG2mut were prepared by QuikChange mutagenesis (200524, Agilent) and Gibson assembly (E2611L, New England BioLabs), with the following mutations in the encoded proteins: K9A, F11A, R49A, F51A, and F53A in RRMmut; R110A, R113A, R120A, and R125A in RGG1mut; R369A, R373A, and R377A in RGG2mut. The mutant hnRNPG sequences were cloned into pCMV3-Flag-RBMX (HG16560-NF, Sino Biological) by Gibson assembly for expression in human cells. The wild-type and mutant hnRNPG sequences were cloned into the pAcGP67a vector using BamHI and NotI restriction sites for expression in insect cells. A His_8 tag was added to the N-terminus for affinity purification. A fast-folding variant of protein G, NuG2b (DTYKLVIVLNGTTFTYTTEAVDAATAEKVFKQYANDAGVDGEWTYDAATKTFTVTE (Lindorff-Larsen et al. 2011; Skinner et al. 2014), was added to the N-terminus to increase protein stability.

A baculovirus expression system was used for the expression of proteins in High Five insect cells as previously described (Langmead and Salzberg 2012). The secreted proteins were purified using nickel nitrilotriacetic agarose (Ni-NTA) resin (30230, Qiagen). The resin was washed with 10 mM Tris-Cl (pH 7.4), 1 M NaCl, 2.5 mM $MgCl_2$, 10% v/v glycerol buffer. The hnRNPG protein was released from the resin by cleavage C-terminal to the His_8–NuG2b tag with His-tagged Pre-Scission Protease (Z03092, GenScript) in storage buffer (10 mM Tris-Cl (pH 7.4), 500 mM NaCl, 2.5 mM $MgCl_2$, 10% v/v glycerol) at 4 °C overnight, and the protein was concentrated with a 30 kDa centrifugal filter (UFC803024) and either stored at 4 °C or flash-frozen in liquid nitrogen and stored at −80 °C. Immediately before use, hnRNPG protein stocks were spun at 21 K × *g* at 4 °C for 10 minutes, and the supernatant was used as the new hnRNPG stock. The concentrations of hnRNPG stocks were measured by Bradford assay (23236, Thermo).

2.4.2 hnRNPG spin-down assay for RNA binding

^{32}P-labeled RNA was pre-mixed ± hnRNPG protein at 0.5 M NaCl, then transferred to 0.15 M NaCl followed by centrifugation. Pellets were dissolved with 0.5 M NaCl buffer in the same volume as supernatant, spotted in µl indicated, and followed by phosphorimaging.

20 pmol of MALAT1 hairpin containing A (2515-A) or m^6A (2515-m^6A) were 5' ^{32}P-labeled using T4 PNK and purified by denaturing gel electrophoresis and stored in water. 8 µl of ^{32}P-RNA was mixed with 2 µl 100 mM TrisHCl (pH 7.4) and renatured by heating at 85°C for 2 minutes, then placed at room temperature for 5 minutes. 10 l 2X Buffer A (40 mM Tris-Cl (pH 7.4), 1 M NaCl, 5 mM $MgCl_2$, 0.4 mM EDTA) was then added, and this mixture was split into 5× 3.5 µl samples. 1.8 µl 9.2 µM hnRNPG protein in storage buffer

(10 mM TrisHCl (pH 7.4), 500 mM NaCl, 2.5 mM $MgCl_2$, 10% v/v glycerol) and 4.7 µl Buffer A (20 mM TrisHCl (pH 7.4), 500 mM NaCl, 2.5 mM $MgCl_2$, 0.2 mM EDTA) was then added. The final concentration for ^{32}P-RNA was less than 0.1 µM. For samples with unlabeled RNA (2515-A or 2515-m^{6}A), RNAPII CTD phosphor-S5 peptide (ab18488, abcam), or RNAPII CTD phosphor-S2 peptide (ab12793, abcam), they were diluted to the appropriate concentration in Buffer A and then added in lieu of Buffer A. Control reactions without hnRNPG substituted storage buffer (10 mM Tris-Cl (pH 7.4), 500 mM NaCl, 2.5 mM $MgCl_2$, 10% v/v glycerol) for hnRNPG. The binding mixtures were rotated at 4 °C overnight in a total volume of 10 µl.

hnRNPG protein precipitation was initiated by diluting the above binding mixture to a final NaCl concentration of150 mM by adding 23.3 µl Buffer B (20 mM TrisHCl (pH 7.4), 0 mM NaCl, 2.5 mM $MgCl_2$, 0.2 mM EDTA) and rotated at 4 °C for 1 hour. The mixtures were then centrifuged for 30 minutes at 17,000 g and 4 °C to pellet. Supernatant was separated from the pellet, and the pellet was resuspended in 33.3 µl Buffer A. 1, 2.5, and 6.25 µl of each sample was blotted onto filter paper. This blot was exposed on a phosphorimager plate and imaged using a BioRad Personal Molecular Imager. The resulting image was quantified using the Quantity One software.

2.4.3. hnRNPG spin-down assay for CTD binding

hnRNPG protein ± GST–CTD ± phosphorylation by CDK7 or CDK9 were pre-mixed at 0.5 M NaCl, then transferred to 0.15 M or 0.25 M NaCl followed by centrifugation. To phosphorylate the GST–CTD, 0.75 ng of recombinant GST-tagged human RNAPII CTD (SRP2120, Sigma) were combined with 0.36 ng of CDK7–Cyclin H–MNAT1 (PV3868,

Thermo) or CDK9–Cyclin T1 (14–685, Sigma) in 25 µL of 8 mM 3-(N-morpholino)propanesulfonic acid (MOPS) (pH 7), 0.2 mM EDTA, and 1 mM $MgCl_2$ supplemented with freshly added 0.1 mM ATP and 0.25 mM DTT, and then incubated at 30 °C for 1 hour. Control reactions without ATP were incubated under the same condition. Binding mixtures using 2x GST-CTD instead utilized 1.5 ng of recombinant GST-tagged human RNAPII CTD and 0.72 ng of CDK7 or CDK9.

Initial binding mixtures were generated by adding 2.45 µl of the above phosphorylation reaction to 1.1 µl of either 9.2 or 18.4 µM hnRNPG in storage buffer (10 mM TrisHCl (pH 7.4), 500 mM NaCl, 2.5 mM $MgCl_2$, 10% v/v glycerol) and 2.45 µl of 2x Buffer A (40 mM TrisHCl (pH 7.4), 1 M NaCl, 5 mM $MgCl_2$, 0.4 mM EDTA). For binding mixtures with RNA, 1.63 µl of 3X Buffer A (60 mM TrisHCl (pH 7.4), 1.5 M NaCl, 7.5 mM $MgCl_2$, 0.6 mM EDTA) and 0.82 µl of 25 or 50 µM 2515-A or 2515-m^6A was added instead. The binding mixtures were rotated overnight at 4 °C overnight in a total volume of 6 µl.

Protein precipitation was initiated by diluting binding mixtures to either final 150 or 250 mM NaCl by adding either 14 µL Buffer B (20 mM Tris-Cl (pH 7.4), 0 mM NaCl, 2.5 mM $MgCl_2$, 0.2 mM EDTA) or 2 µl 2X Buffer A and 12 µL Buffer B, respectively. These diluted binding mixtures were rotated for 1 hour at 4 °C in a total volume of 20 µl. The binding mixtures were then centrifuged for 1 hour at 17,000 g and 4 °C to pellet. Supernatant was separated from the pellet, and the pellet was resuspended in 20 µL Buffer B. 20 µL of 2x LDS (NP0008, Thermo), 50 mM EDTA was then added to each sample, and they were then incubated at 95 °C for 10 min. 20 µL of each sample was then loaded onto 12-well 4–12% polyacrylamide Bis-Tris gels (NP03322, Invitrogen) alongside a Novex™ Sharp Pre-stained Protein Standard (LC5800, Thermo) and ran at 150V for 1 hour. The gels

were then silver stained using the Pierce™ Silver Stain Kit (24612, Thermo). Gels were imaged using a UVP BioDoc-It™ Imaging System. The intensities of each band were quantified using Image Studio software.

Chapter 3

Expanding mechanisms by which nuclear reader proteins regulate m^{6}A-dependent gene expression

Acknowledgment: This chapter is derived from unpublished work. hnRNPG protein was courtesy of Dr. Katherine Zhou. YTHDC1 wild-type and mutant plasmids were courtesy of Professor Chaun He's lab. Size exclusion column during 6x-His-YTH purification done with the assistance of Tomasz Ślężak from the Professor Kossiakoff lab.

3.1 Introduction

Gene expression can be regulated co-transcriptionally through proteins that specifically recognize N^6-methyl-adenosine (m^{6}A) known collectively as m^{6}A-reader proteins. Two nuclear m^{6}A-reader proteins, YTH N6-Methyladenosine RNA Binding Protein C1 (YTHDC1) and heterogeneous nuclear ribonucleoprotein G (hnRNPG), are known to regulate m^{6}A-dependent alternative splicing (Alarcón et al. 2015; Kasowitz et al. 2018; N. Liu et al. 2015, 2017; Xiao et al. 2016a) and have been suggested to interact with one another (Hartmann et al. 1999; Huttlin et al. 2015, 2017); however, direct interaction between these two nuclear m^{6}A-reader proteins has yet to be demonstrated. The mechanisms behind the majority of m^{6}A-dependent alternative splicing events have yet to be identified, though Chapter 2 of this thesis describes one mechanism of m^{6}A-dependent alternative splicing which acts on a subset pre-mRNA with modification-dependent splicing (Zhou et al. 2019). Given the mechanism of m^{6}A in many m^{6}A-dependent splicing events remains to be defined, I investigated the properties of YTHDC1, the nuclear m^{6}A reader

with suggested hnRNPG interactions, with still uncharacterized m^{6}A-dependent activity in splicing.

YTHDC1 is the sole primarily nuclear-localized member of the YTH-family of m^{6}A-reader proteins, who directly recognize m^{6}A through conserved tryptophan residues embedded in their structured YTH-domains (Xiao et al. 2016a; Xu et al. 2014). The YTH-domain lies embedded within a largely unstructured whole protein (Xiao et al. 2016b); these low-complexity regions contribute to the formation of phase-separated processing complexes near nuclear speckles known as "YT-bodies" (Rafalska et al. 2004). Functionally, YTHDC1 is involved in transcription and m^{6}A-dependent pre-mRNA processing in the nucleus (Kasowitz et al. 2018; Lee et al. 2021; N. Liu et al. 2017) including alternative splicing and polyadenylation (Xiao et al. 2016a), stability of chromatin-associated RNAs (Jun Liu et al. 2020), and nuclear export (Roundtree, Luo, et al. 2017).

Here I detail the differing RNA and m^{6}A-modified RNA-dependence of YTHDC1 and hnRNPG on both interactions with one another and interactions with the C-terminal domain (CTD) of RNA Polymerase II (RNAPII). YTHDC1 can form direct protein-protein interactions with the C-terminal domain (CTD) of RNA Polymerase II (RNAPII) mediated through its structured YTH-domain, in contrast to hnRNPG's interactions with CTD which are predominately mediated through its unstructured and highly disordered RGG regions (Zhao et al. 2019). The YTH-domain's interactions with RNAPII are reduced following the addition of m^{6}A-RNA, suggesting the domain binds exclusively to m^{6}A or RNAP. This contrasts with hnRNPG's ability to accommodate simultaneous interactions and demonstrates divergent mechanisms by which m^{6}A reader proteins regulate m^{6}A-dependent, co-transcriptional gene expression.

3.2 Results

3.2.1 hnRNPG solubilizes YTHDC1 in vitro

Motivated by our studies on m^{6}A reader protein hnRNPG, I sought to investigate the mechanism of co-transcriptional, m^{6}A-dependent activity of nuclear reader protein YTHDC1. Purification of full-length YTHDC1 proved difficult for us and others with all tested *E. coli* expression procedures yielding no product (data not shown); indeed, *in vitro* structural and biochemical analysis of YTHDC1 has been limited to its more readily purifiable YTH-Domain (Woodcock et al. 2020). Purification of FLAG-YTHDC1 in transfected HEK293T cells was inconsistent, but one attempt yielded workable amounts of the eluted protein product of the correct molecular weight from anti-FLAG M2 beads (A2220 Sigma) after the addition of FLAG peptide (**Figure 3.2.1a**). The identity of this protein as YTHDC1 was further substantiated by anti-YTHDC1 Western Blot (**Figure 3.2.1c**).

The spin-down assay developed in Chapter 2 of this thesis enables the detection of low-complexity protein-protein interactions with low input and in small volumes; therefore, the spin-assay provided the unique opportunity to detect *in vitro* interactions and aggregative properties of YTHDC1 given low-yields of full-length low-complexity protein purification. I performed our spin-assay on YTHDC1 in the presence and absence of hnRNPG to determine whether the two proteins could interact and aggregate with one another *in vitro* (**Figure 3.2.1b**). For this assay, the proteins were initially pre-mixed in 500 mM NaCl, buffer conditions where the proteins remain soluble, before complex formation was induced by diluting to 150 mM NaCl. These complexes were then pelleted by

centrifugation, separated from the soluble supernatant, resuspended in buffer, run on SDS-PAGE, and visualized by silver-staining. When tested alone, YTHDC1 was solely in the aggregated pellet (P) fraction, whereas hnRNPG was distributed between both the pellet and the soluble supernatant (S) fraction. When YTHDC1 and hnRNPG were pre-mixed as part of the spin-assay, YTHDC1's distribution was disrupted, instead evenly distributed between the fractions. This observation implies the addition of hnRNPG solubilized a portion of YTHDC1 which would otherwise be aggregated. I confirmed the pelleting behavior of YTHDC1 when tested alone by Western Blot (**Figure 3.2.1c**). Taken together, we hypothesized a "fishing model," where a portion of hnRNPG bound to the transcriptionally active phosphorylation state of RNA Polymerase II's C-terminal domain (RNAPII pCTD) would interact with and thus fish YTHDC1 out from phase-separated granules. Populations of polymerase-bound hnRNPG which recruited YTHDC1 for assembly of downstream splicing machinery would thus account for YTHDC1- and hnRNPG-dependent pre-mRNA splicing.

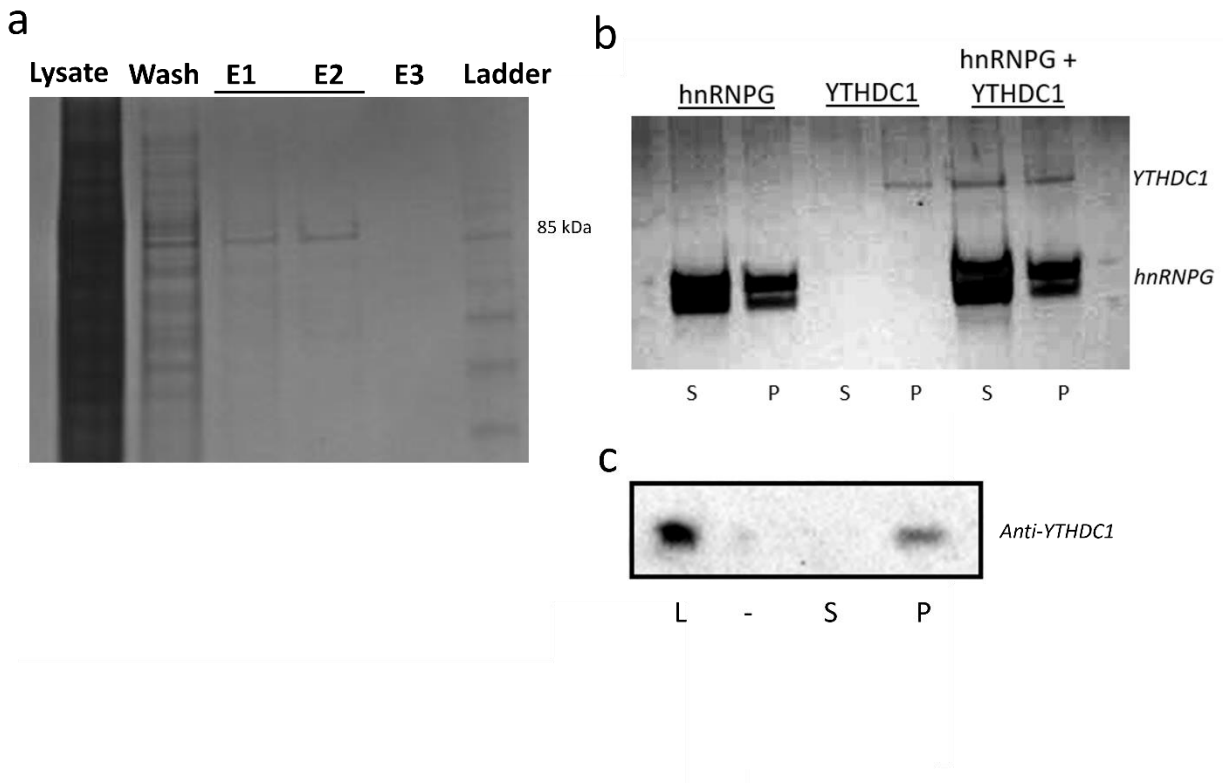


Figure 3.2. 1 YTHDC1 is solubilized by hnRNPG (a). Silver-stained protein gel showing purified FLAG-YTHDC1 from HEK293T cells transfected with FLAG-YTHDC1 mammalian expression plasmid by use of anti-FLAG M2 beads. Lanes from left-to-right: (1) Total cell lysate, input for anti-FLAG M2 beads (2) protein eluted following the addition of wash buffer; (3)-(5) elution fractions 1-3, respectively. No protein was observed by western blot in elution 3. (6) Molecular-weight protein standard ladder, the position corresponding to 85 kDa labeled. expected molecular weight of FLAG-YTHDC1 is ~86 kDa. (b). Silver-stained denaturing protein gel showing aggregation of hnRNPG and YTHDC1 alone and interactions with one another when pre-mixed by spin-assay. The final NaCl concentration was 0.25 M; the final hnRNPG concentration was 1.7 μM. s: supernatant fraction; p: pellet fraction. (c). Western blot against YTHDC1 of spin-assay performed on YTHDC1 alone, as in (b). l: ladder, -: empty lane, s: supernatant fraction, p: pellet fraction. Western Blot validation.

3.2.2 YTHDC1 interacts with the RNA Polymerase II CTD in an RNA-independent manner

Biochemical data generated by the spin-assay showing solubilization of YTHDC1 by hnRNPG corroborated literature evidence that YTHDC1 and hnRNPG can interact (Hartmann et al. 1999; Huttlin et al. 2015, 2017). We hypothesized a population of polymerase-bound hnRNPG would bind YTHDC1 directly as part of a co-transcriptional splicing complex for splice isoforms exhibiting YTHDC1 and hnRNPG co-dependence; therefore, we sought to demonstrate direct interaction of these two m^6A reader proteins in a co-transcriptional context by co-immunoprecipitation (CoIP) using human embryonic kidney (HEK) 293T chromatin extracts. First, extracts corresponding to cytosolic, nuclear, and chromatin fractions had their identities verified by Western blot against marker proteins for these subcellular fractions (**Figure 3.2.2a**). RNase treatment was employed to test the RNA-dependence of CoIP thus necessitating verification of chromatin extract RNA degradation on RNase treatment (**Figure 3.2.2b**), as direct, protein-protein interactions necessitate an RNase-independence. Indeed, the presence of RNA was abolished upon treating the chromatin fraction with RNase.

YTHDC1 co-immunoprecipitated with hnRNPG in chromatin extracts of HEK293T cells (**Figure 3.2.3**), corroborating previous evidence for their interaction. However, this CoIP was abolished following RNase treatment. The binary, RNase-dependence of YTHDC1 co-immunoprecipitation implies an indirect, RNA-mediated interaction with hnRNPG *in vivo*, as opposed to the protein-protein interactions we observed in our biochemical spin-down assay (**Figure 3.2.1b**). While this result may seemingly conflict with the prior

observations, the solubility of YTHDC1 induced by addition of hnRNPG does not necessarily demonstrate direct protein-protein interaction, but only a change in its environment that energetically disfavors aggregation in a particular biochemical context.

Given this result argues against direct protein-protein interactions for YTHDC1 and hnRNPG *in vivo*, other mechanisms by which YTHDC1 could be incorporated into co-transcriptional splicing assemblies were considered. In our previous report, we demonstrated that hnRNPG directly interacted with the transcriptionally active RNAPII pCTD (Zhou et al. 2019); therefore, I sought to investigate if both YTHDC1 and hnRNPG were both capable of interacting directly with the CTD of RNAPII. RNAPII pCTD indeed co-immunoprecipitated in an RNase-independent manner when tested, which replicates our previous result and acts as an internal control that further supports hnRNPG's RNase-dependent co-immunoprecipitation of YTHDC1 (**Figure3.2.2d**). RNAPII CTD co-immunoprecipitated with YTHDC1 in an RNase-independent manner, suggesting that both YTHDC1 and hnRNPG can make direct protein-protein interactions with RNAPII *in vivo* (**Figure 3.2.2d**). Taken together, these observations support a model where the CTD of RNAPII accommodates two populations of low-complexity, m^6A-reader protein-bound states: one bound to hnRNPG, and another to YTHDC1. These two proteins could act as platforms to recruit unique splicing condensates, explaining why sets of m^6A-dependent splicing events have a dependence on either hnRNPG or YTHDC1.

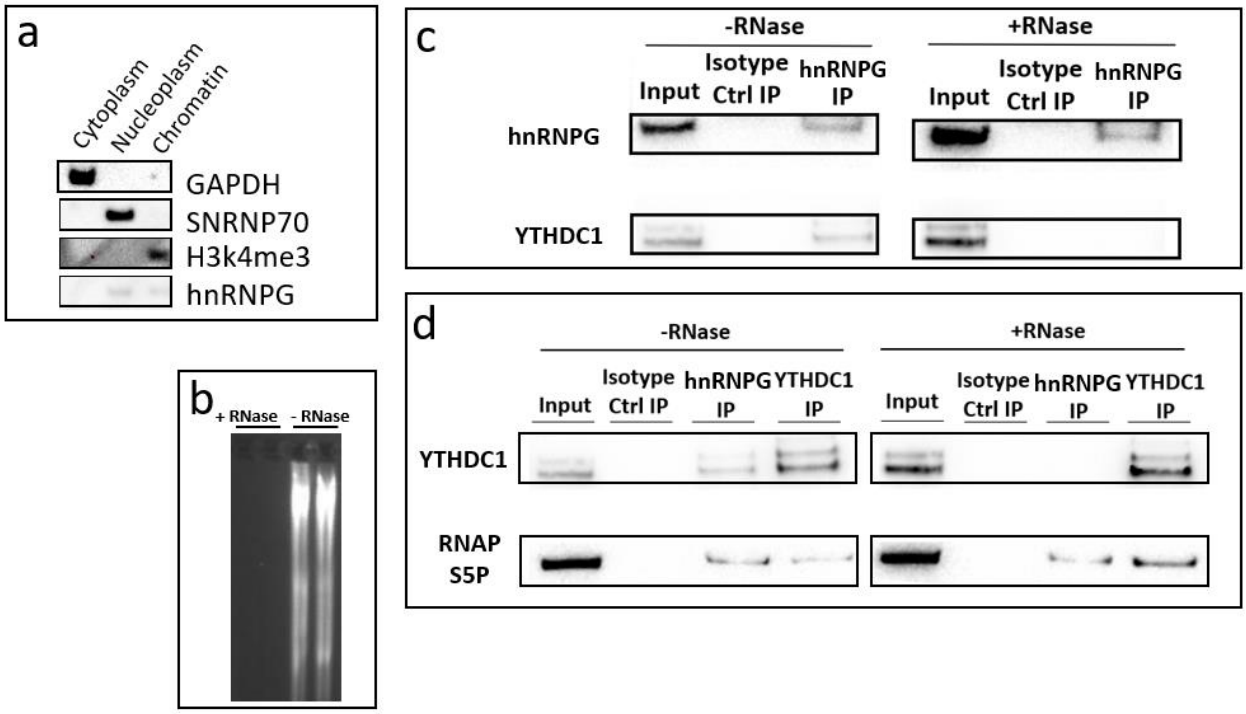


Figure 3.2. 2 YTHDC1 directly interacts with RNA Polymerase II (a). Western blot of representative cytoplasm, nucleoplasm, and chromatin extracts; these subcellular fractions were defined by the presence or absence of marker proteins for each subcellular compartment. Cytoplasm, nucleoplasm, and chromatin fractions are defined by the exclusive presence of GAPDH, SNRNP70, and H3k4me3, respectively. hnRNPG was found in both nucleoplasm and chromatin extracts. (b). Agarose gel of extract RNase A/T1 treatment. Both samples which received the RNase treatment, and which did not were incubated for 30 minutes at 37°C. (c). Western blot against hnRNPG (top) and YTHDC1 (bottom) of the chromatin extract input used for immunoprecipitation, isotype negative control immunoprecipitation, and YTHDC1 immunoprecipitation without RNase treatment (left) or with RNase treatment (right) showing RNase-dependent co-immunoprecipitation of YTHDC1 by hnRNPG. (d). Western blot against YTHDC1 (top) and RNA Polymerase II pCTD (RNAP S5P, bottom) of the chromatin extract input used for immunoprecipitation, isotype negative control immunoprecipitation, hnRNPG immunoprecipitation, and YTHDC1 immunoprecipitation without RNase treatment (left) or with RNase treatment (right) showing RNAP S5P co-immunoprecipitation with YTHDC1 in an RNase-independent manner.

3.2.3 YTHDC1's YTH-domain interacts directly with RNAPII in chromatin extracts

Having established the interaction between endogenous YTHDC1 and RNAPII in HEK293T cells, I next sought to investigate the sequence feature of YTHDC1 which mediated this interaction. YTHDC1's YTH-domain is the sole structured region, located in the center of the protein's amino acid sequence (**Figure 3.2.3**). The low-complexity portions outside of YTHDC1's YTH-domain contain regions enriched for several specific amino acid residues when compared to the human proteome. As in the case of hnRNPG, low-complexity regions can be responsible for mediating RNAPII interactions (Zhou et al. 2019), inspiring inquiry on whether these highly charged regions of YTHDC1 could mediate complex formation by polar interactions.

FLAG-YTHDC1 constructs were transfected into HEK293T cells to investigate the necessary YTHDC1 sequence component in anti-FLAG co-immunoprecipitation of RNAPII. Interestingly, the YTH-domain alone was sufficient to CoIP RNAPII (**Figure 3.2.3b**). As the YTH-domain is responsible for direct interaction with m^6A-modified RNA, whether this interaction was RNA-mediated was tested by RNase treatment. Surprisingly, the FLAG-YTH-domain alone was still sufficient to pulldown the actively transcribing polymerase following RNase treatment, supporting a direct protein-protein between the YTH-domain and RNAPII.

The YTH-domain's intermolecular interactions have solely been described in the context of RNA, making this FLAG-CoIP the first demonstration that the YTH-domain is capable of both protein-RNA and protein-protein interactions. To test the relative strength of the YTH-domain's interactions with m^6A-RNA and the RNAPII CTD, anti-FLAG immunoprecipitation of FLAG-YTH transfected cells was tested after adding methylated or

unmethylated RNA and compared to a control where RNA was not added (**Figure 3.2.3c**). The addition of A-RNA resulted in a modest decrease in co-immunoprecipitated CTD, while the addition of m^6A-RNA lead to a larger reduction. Taken together, this supports a competitive-binding model of YTHDC1, where the YTH domain can bind both RNA and RNAPII CTD but cannot accommodate their simultaneous binding. Furthermore, RNA acts as a competitor for YTH-binding, ejecting bound CTD when introduced.

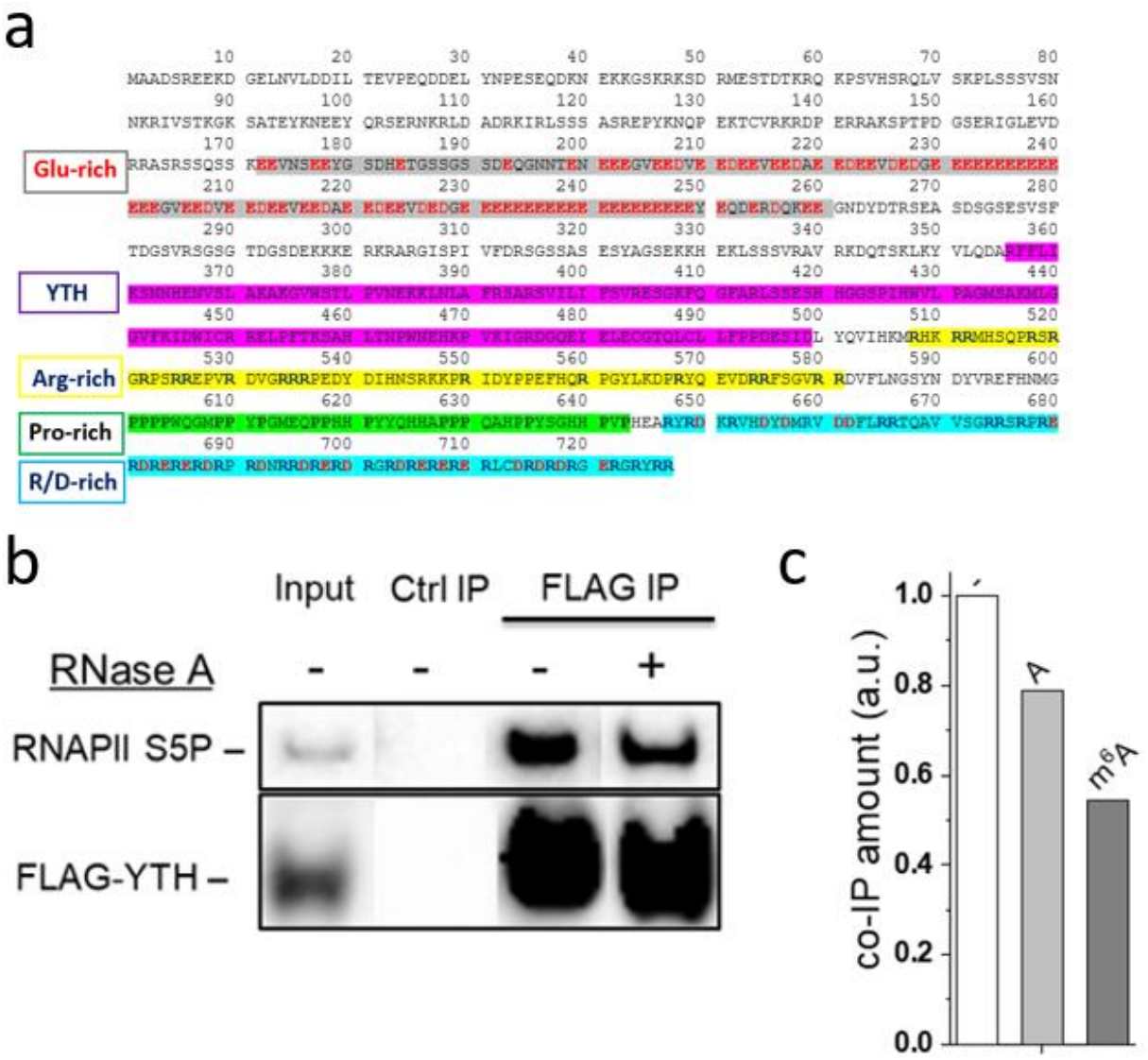


Figure 3.2. 3 The YTH-Domain of YTHDC1 interacts directly with RNAPII S5P (a). The amino acid sequence of full-length human YTHDC1. The structured YTH-domain is highlighted in purple, while other highlighted regions are enriched in the amino acid(s) denoted to the left, relative to the human proteome. (b). Western blot FLAG-IP shows FLAG-YTH interacts with the RNAPII S5P in an RNase-independent manner. Lanes from left to right (1) Input; chromatin extract used for IP (2) Ctrl IP; isotype control against FLAG IP (3, 4) FLAG IP. Lanes 1-3 are without RNase treatment, while lane 4 is following RNase A/T1 treatment. (b). Quantification of S5P co-immunoprecipitated with FLAG-YTH in the presence of A or m^6A containing RNA. A: 5'pAUAAAAACAGACUCUGUAGCGAUGUCAAA; m^6A:5'pAUAAAAm^6ACAGm^6ACUCUGUAGCGAUGUCAAA

3.2.4 His-YTH domain can pulldown S5P, S2P, S0P in vitro

Having established the ability of the YTH-domain to interact with RNAPII *in vivo*, I investigated whether the YTH-domain could also interact directly with the RNAPII CTD *in vitro*. Purification of YTH was performed using a 6x-His-tag scheme in *E. coli;* 6x-His-YTH expression plasmids were transformed into T7 Express competent *E. coli* cells and cultures induced with IPTG. The tagged peptide was then purified using sequential affinity and size-exclusion columns. Affinity purification was done using Ni-NTA resin, from which eluted impure fractions that contained a protein band on SDS-PAGE corresponding to the size of the tagged protein domain (**Figure 3.2.4a**). Ni-NTA column elution fractions containing this protein were concentrated using an Amnicom 3 kDa molecular weight cut-off filter. Following this, the concentrated sample was loaded onto a Size Exclusion Column, with the protein-containing fractions again pooled and then concentrated with a 3 kDa cut-off column. Coomassie staining confirmed that size-exclusion led to a loss of high molecular weight contaminants present in the Ni-NTA eluate (**Figure 3.2.4a**).

His-pulldown experiments were then performed on mixtures containing purified 6x-His-YTH and GST-RNAPII CTD to test the interaction between these tagged proteins. Phosphorylated GST-CTD (pCTD) was prepared by either CDK7 or CDK9 pre-treatment, generating phosphorylation at CTD Ser5 (S5P) or CTD Ser2 (S2P), respectively. 6x-His-YTH pulled down not only the two tested CTD phospho-isoforms (**Figure 3.2.4b**, left) but also the CTD in its unphosphorylated state (**Figure 3.2.4b**, right). This pulldown demonstrates direct interaction *in vitro* between the His-tagged YTHDC1 YTH-domain and GST-tagged RNAPII CTD, supporting direct protein-protein interactions between the YTH-domain and RNAPII. Interestingly, the YTH-domain's promiscuity in interacting with the RNAPII CTD

regardless of its phosphorylation state contrasts with hnRNPG, which interacted solely with the phosphorylated CTD indicative of a transcriptionally active polymerase (Zhou et al. 2019).

Taken collectively, this establishes a model where m^6A reader proteins exhibit unique interactions with RNAPII CTD where active transcription may modulate their activities (**Figure 3.2.4c**). In this model, the RNAPII CTD can bind directly to both hnRNPG and YTHDC1, creating at least two populations of reader-bound polymerases. While hnRNPG's effect on pre-mRNA exon inclusion is transcriptionally dependent (Zhou et al. 2019), the data shown here suggest m^6A-modified RNA, which may be nascent transcripts produced from the reader-bound polymerase, preferentially binds to YTHDC1's YTH-domain, ejecting the bound polymerase, instead leaving YTHDC1 tethered and spatially separated via the bound RNA.

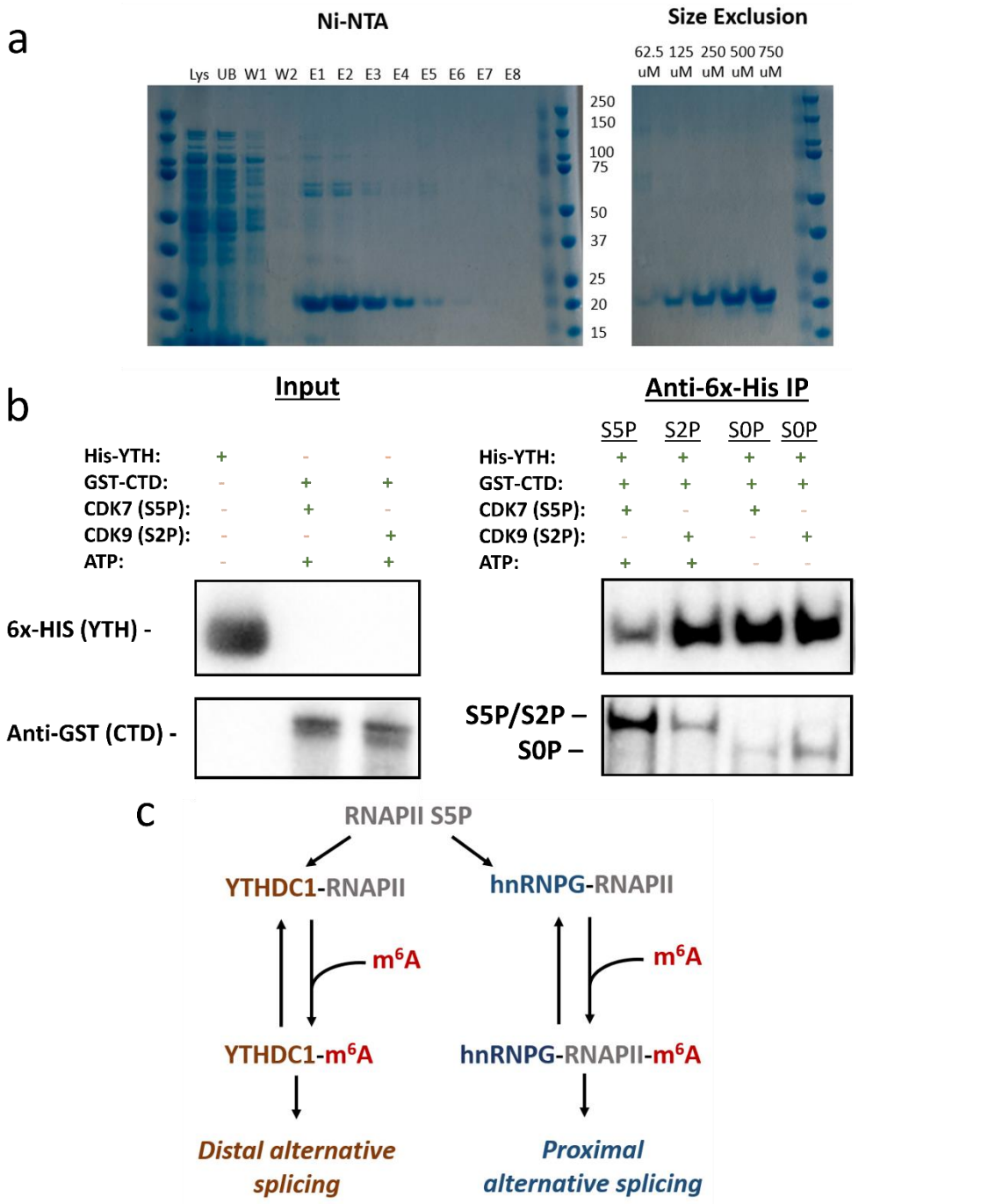


Figure 3.2. 4 6x-His-YTH-Domain Interacts with GST-CTD In Vitro (a). Sequential two-column purification of 6x-His-YTH domain visualized Coomassie-stained denaturing protein gels. Lysate from *E. Coli* culture was loaded into a Ni-NTA column (left) where elution fractions containing the tagged protein of interest were pooled and subsequently loaded into a size-exclusion column (right). Ni-NTA gel lanes numbered left to right: (1, 14) Protein standard ladder; (2) Input *E. Coli* lysate loaded in the column; (3) UB – unbound lysate eluate following resin binding; (4-5) W1-W2 – eluate following the first and second resin washes, respectively; (6-13) E1-E8 – fractions 1-8 following sequential additions of

elution buffer. The size-exclusion column gel was loaded with increasing concentrations of purified protein following size-exclusion to demonstrate the absence of high-molecular-weight contaminants present in the Ni-NTA column eluates. (b). Western blot of His-pulldown shows 6x-His-YTH domain pulls down RNAPII CTD. Blot done against 6X-His (top) and GST (bottom) to visualize 6x-His-YTH domain and GST-RNAPII CTD, respectively. Left: Input used in pull-down; Right: Anti-6x-His IP pull-down eluate. GST-CTD (S0P) was phosphorylated at Ser2 (S2P) or Ser5 by the addition of ATP and either CDK9 or CDK7, respectively. (c). Proposed model for m^{6}A-dependent co-transcriptional interactions, where at least two populations of reader-bound polymerases exist. In the YTHDC1-bound population, nascent m^{6}A-modified mRNA binds competitively to YTHDC1's YTH-domain, ejecting the bound polymerase and leading to distal m^{6}A-dependent effects on alternative splicing. For the hnRNPG-bound polymerase population, the reader protein accommodates binding to both nascent m^{6}A-modified transcripts and the polymerase, conferring proximal effects on alternative splicing.

3.3 Discussion

The data here presented demonstrate that the m^{6}A reader protein YTHDC1 directly interacts with the CTD of RNAPII. Pulldown of tagged proteins both *in vitro* and *in vivo* suggests that YTHDC1's m^{6}A-binding YTH-domain mediates direct protein-protein interactions with the RNAPII CTD; to date, there have been no reports that the YTH-domain of any YTH-family protein engages in intermolecular protein-protein interactions. Furthermore, data here shows that the introduction of m^{6}A-modified RNA untethers YTHDC1 from the CTD of RNAPII, suggesting competitive binding contrasting the accommodation to both interactions shown by another CTD-binding m^{6}A-reader protein, hnRNPG. These observations support a model where YTHDC1 interacts co-transcriptionally with the RNAPII CTD until the nascent RNA transcript is methylated, creating a substrate that outcompetes and thereby ejects the YTH-domain bound polymerase, resulting in YTHDC1 remaining tethered to active transcription solely by the nascent transcript itself.

YTHDC1 has been suggested to interact with hnRNPG (Hartman 1999, Huttlin, et al 2015, 2017), which could act as a way for the two proteins to coordinate their m^6A-dependent splicing activities (Zhou et al. 2019). However, testing this model of crosstalk has proven difficult owing to the low complexity and aggregation-prone nature of these full-length proteins. In my hands, repeated strategies and attempts at purification of YTHDC1 yielded a small amount of workable material, suited for application in my small-volume spin-assay. This assay demonstrated hnRNPG-dependent solubilization of YTHDC1, but this observation is not synonymous with direct protein-protein binding in this soluble fraction. Co-immunoprecipitation demonstrated that interactions between YTHDC1 and hnRNPG are RNase-dependent in HEK293T chromatin extract and thus that their cellular interactions are mediated by an RNA intermediate. Additionally, reports of the two protein's interactions could also be explained by our observation that YTHDC1 can bind directly to the CTD of RNAPII.

The YTH-domain directly recognizes m^6A-modified RNA through an internal hydrophobic cage also capable of accommodating adenosine at lower affinity (Y. Li et al. 2021). Reports have shown the YTH-domain of YTHDC1 and YTHDF1-3 also bind directly to m^1A in RNA (X. Dai et al. 2018), expanding the known versatility of protein-RNA interactions mediated by this domain. The data here presented suggests the YTH-domain interacts with a separate target, the CTD of RNAPII, independent of its phosphorylation state. This indiscriminate CTD-binding sets YTHDC1 apart from hnRNPG, an m^6A-reader protein that specifically interacts with the phosphorylated CTD of RNAPII, indicative of active transcription (Bentley 2014; Zhou et al. 2019). Thus, I propose that m^6A-reader

proteins' co-transcriptional interactions with the RNAPII CTD can be regulated by either CTD or RNA processing.

In this model, RNAPII CTD acts as a handle for YTHDC1 until the nascent transcript is methylated by the m^{6}A methyltransferase complex, generating a modified site that preferentially interacts with YTHDC1, thus ejecting the YTH-domain bound CTD. Because most m^{6}A modifications are deposited co-transcriptionally (Barbieri et al. 2018; Ke et al. 2017; Knuckles et al. 2017; Louloupi et al. 2018), this model accordingly allows for co-transcriptional regulation of CTD association. This contrasts hnRNPG, which binds exclusively to active over inactive RNAPII and can accommodate binding to m^{6}A simultaneously with the RNAPII pCTD. Both reader proteins could then act as a handle for the assembly of splicing complexes, forming proximal and distal to the polymerase for hnRNPG- and YTHDC1-bound m^{6}A sites, respectively.

3.4 Materials and methods

Experimental models and subjects

Human embryonic kidney (HEK) cell line HEK293T/17 (CRL11268) was obtained from the American Type Culture Collection (ATCC) and cultured in Dulbecco's Modified Eagle's Medium (DMEM) with high glucose and L-glutamine, without sodium pyruvate (HyClone, SH30022.01), with 10% FBS and 1% Pen–Strep (Penicillin–Streptomycin) in a 37°C incubator at 5% CO_2. Transfection of FLAG-tagged YTHDC1 constructs was done using

Lipofectamine 2000 (11668019, Thermo) following passage into culture media without 1% Pen–Strep as written below.

FLAG-YTH mutant plasmids were a gift from the Chaun He lab.

E. Coli expression plasmid was an N-terminal 6x-His tag with a TEV site before the construct within a pET28 backbone. Sequence (GGATCC) GGCACCAGCAAGCTGAAATACGT GCTGCAGGACGCGCGTTTCTTTCTGATCAAGAGCAACAACCACGAAAACGTTAGCCTGGCGAAG GCGAAAGGCGTGTGGAGCACCCTGCCGGTTAACGAGAAGAAACTGAACCTGGCGTTCCGTAGCGC GCGTAGCGTGATCCTGATTTTTAGCGTTCGTGAAAGCGGTAAATTCCAAGGCTTTGCGCGTCTGA GCAGCGAGAGCCACCACGGTGGCAGCCCGATTCATTGGGTGCTGCCGGCGGGTATGAGCGCGAAG ATGCTGGGTGGCGTTTTCAAAATCGACTGGATTTGCCGTCGTGAACTGCCGTTTACCAAGAGCGC GCACCTGACCAACCCGTGGAACGAGCACAAGCCGGTGAAAATCGGTCGTGATGGCCAGGAGATTG AACTGGAGTGCGGCACCCAACTGTGCCTGCTGTTCCCGCCGGACGAAAGCATCGATCTGTATCAG GTTATTCACAAAATGCGTCACTAA(CTCGAG)

3.4.1 FLAG-YTHDC1 purification

For purification of full-length FLAG-YTHDC1, 4 15 cm plates of HEK293T cells were transfected with 60 μg FLAG-tagged plasmid using Lipofectamine 2000 (11668019, Thermo) per manufacturer's protocol. Briefly, for each plate, 3 mL Opti-MEM I Reduced Serum Medium with 60 μg FLAG-tagged plasmid (YTHDC1) and 3 mL Opti-MEM I Reduced Serum Medium with 100 μL Lipofectamine 2000 were individually mixed and incubated at room temperature for 5 minutes, before their combination and an addition 20-minute room temperature incubation. The combined 6 mL DNA-Media-Lipofectamine mixture was

then added to the cell culture plate, which was then briefly mixed by gentle, manual rocking. Cells were incubated in this transfection media for 6 h at 37°C before being exchanged for fresh, antibiotic-free medium and incubated for an additional 24 hours.

Following transfection and incubation, the medium was removed from the plates and the cells were washed 3 times with 5 mL of 1x PBS. Each plate of cells was scraped and lysed in 2 mL Lysis buffer (Sigma lysis buffer supplemented with 10 μg/ml DNase I, 40 μg/ml RNase A, 100 U/ml RNase T1, 2 mM $MgCl_2$ and 1% v/v Nacalai protease inhibitor.) Cells were lysed and rotated for 2h at 4°C before centrifugation at 17,000 g for 30 minutes. To purify the FLAG-tagged protein from this lysate, 300 μL anti-FLAG M2 gel/beads (A2220 SIGMA) were first washed and prepared according to their manual. Briefly, for each sample 300 μL gently resuspended anti-FLAG M2 gel/beads were transferred to a fresh tube. This resin was then at 8,200 x g for 30 seconds, settled, and then the supernatant was removed. The packed gel was washed then twice with 2 mL TBS, then with 1 mL 0.1 M glycine HCl (pH 3.5), and then a final three washes of 2 mL TBS each. The supernatant was once again removed following the packing the resin by centrifugation. Cell lysate containing transfected FLAG-YTHDC1 was then added to the anti-FLAG gel tube and incubated overnight at 4 °C while rotating. The following day, the resin was centrifuged for 30 seconds at 8200 x g before washing three times with 1 mL wash buffer (50 mM Tris-HCl, pH 7.5, 300 mM KCl, 5% glycerol, 1 mM DTT). To this, 400 μl RNase T1/A mixture (40 μg/ml RNase A and 100 U/ml RNase T1 in 10 mM Tris-HCl pH 7.5, 300 mM NaCl, 2 mM $MgCl_2$, 1× protease inhibitor, 1 mM DTT, 0.1 U/μl DNase I) was added, then the sample was incubated at 800 rpm for 1 hour. The gel was then centrifuged for 30 seconds at 8200 x g, the supernatant removed, and subjected to an additional three washes in 1 mL wash buffer

each. To elute the protein, 1 mL wash buffer containing 300 µg FLAG peptide (Sigma, F3290) was added and then the resin was incubated at 4°C for 1 hour. Following another centrifugation, the supernatant was transferred to a fresh microcentrifuge tube before being concentrated by Amnicom 3 kDa cut-off filter. Protein identity was validated by SDS-PAGE and subsequent protease treatment, in addition to western blotting (see below).

3.4.2 hnRNPG-YTHDC1 spin assay

Full-length hnRNPG protein (see Chapter 2) ± full-length FLAG-YTHDC1 were pre-mixed at 0.5 M NaCl, then transferred to 0.15 M NaCl followed by centrifugation at 17000 x g for 30 minutes. Initial binding mixtures were generated by adding 2.45 µl of ~5 µM YTHDC1 in in storage buffer (10 mM TrisHCl (pH 7.4), 500 mM NaCl, 2.5 mM $MgCl_2$, 10% v/v glycerol) to 1.1 µl of either 9.2 or 18.4 µM hnRNPG in storage buffer (10 mM TrisHCl (pH 7.4), 500 mM NaCl, 2.5 mM $MgCl_2$, 10% v/v glycerol) and 2.45 µl of 2x Buffer A (40 mM TrisHCl (pH 7.4), 1 M NaCl, 5 mM $MgCl_2$, 0.4 mM EDTA). The binding mixtures were rotated overnight at 4 °C overnight in a total volume of 6 µl.

Protein precipitation was initiated by diluting binding mixtures to final 150 NaCl by adding 14 µL Buffer B (20 mM Tris-Cl (pH 7.4), 0 mM NaCl, 2.5 mM $MgCl_2$, 0.2 mM EDTA). These diluted binding mixtures were rotated for 1 hour at 4 °C in a total volume of 20 µL. The binding mixtures were then centrifuged for 1 hour at 17,000 g and 4 °C to pellet. Supernatant was separated from the pellet, and the pellet was resuspended in 20 µL Buffer B. 20 µL of 2x LDS (NP0008, Thermo), 50 mM EDTA was then added to each sample, and they were then incubated at 95 °C for 10 min. 20 µL of each sample was then loaded onto 12-well 4–12% polyacrylamide Bis-Tris gels (NP03322, Invitrogen) alongside a Novex™

Sharp Pre-stained Protein Standard (LC5800, Thermo) and ran at 150V for 1 hour. The gels were then silver stained using the Pierce™ Silver Stain Kit (24612, Thermo). Gels were imaged using a UVP BioDoc-It™ Imaging System. The intensities of each band were quantified using Image Studio software.

3.4.3 Chromatin CoIP

Preparation of HEK293T chromatin extraction was performed based on a published protocol (Brugiolo et al. 2017) following washing and collecting cells in ice-cold 1x PBS. Briefly, following collection in 1x PBS, cells were pelleted by centrifugation and then resuspended by flicking in 1.8 mL ice-cold NP-40 cytoplasmic lysis buffer (10 mM Tris-HCl (pH 7.5), 0.15% v/v NP40, 150 mM NaCl, 1% protease and phosphatase inhibitor (78440, Thermo)) prior to a 5-minute incubation on ice. The cell lysate was pipetted onto 2.5 volumes of sucrose cushion (NP-40 cytoplasmic lysis buffer in 24% w/v sucrose) pre-cooled to 4°C then centrifuged at 16,000 × *g* for 10 minutes at 4°C. The supernatant, containing the cytoplasmic fraction, was removed, and saved for downstream analysis. To wash the now pelleted nuclei, 1 mL of ice-cold wash buffer (0.1% v/v Triton X-100, 1 mM EDTA, 1x PB) was added and the pellet resuspended prior to 1-minute centrifugation at 1,200 x g.

Washed nuclei were resuspended in 200 μL glycerol buffer (20 mM Tris-HCl (pH 8.0), 75 mM NaCl, 0.5 mM EDTA, 50% v/v glycerol, 0.85 mM DTT) and then mixed with 200 mL nuclei lysis buffer (1% v/v NP-40, 20 mM HEPES (pH 7.5), 300mM NaCl, 1M Urea, 0.2mM EDTA, 1mM DTT) followed by thorough pulse-vortexing and a subsequent 15-

minute incubation on ice. The sample was then centrifuged at 12,000 x g for 10 minutes at 4°C and the supernatant saved as the nuclear extract fraction. The pelleted chromatin was then rinsed with 1 mL ice-cold 1x PBS supplemented with fresh 1 mM EDTA prior to digestion in 120 μL chromatin extraction buffer (20 mM Tris-Cl (pH 7.5), 100 mM KCl, 2 mM MgCl2, 2.5 mM $CaCl_2$, 0.3 M sucrose, and 0.1% v/v Triton X-100, 1% protease and phosphatase inhibitor) by 5 U/mL Micrococcal Nuclease (N3755, Sigma) at 4°C for 2 hours. 5 mM of EDTA was added to quench the digestion and the solution was centrifuged at 2000 x g for 10 minutes at 4°C to collect the supernatant containing the chromatin extract fraction. RNase treatment was performed by adding 2 μL RNase A/T1 Mix (2 mg/mL RNase A, 5000 U/mL RNase T1, ThermoFisher EN0551) and incubating for 30 minutes at 37°C.

For chromatin immunoprecipitation, 1/20 v/v anti-YTHDC1 (ab122340) or anti-hnRNPG (abcam 190352) was added to the chromatin extract to a final volume of 90-100 μl. Protein A beads (Thermo 10001D) were washed in 200 μL Wash Buffer (300 mM NaCl, 100 mM Tris (pH 8.0), 0.2 mM EDTA, 0.1% Triton X-100) and then resuspended in Wash Buffer (~0.05 mg/μl, such that 1.5 mg is contained in 30 μL) and rotated for 2 hours at 4°C. The beads were then collected, and the supernatant removed using a magnetic rack prior to being washed four times with 200 μL ice-cold Wash Buffer each wash. The sample was then eluted by boiling in 30 μL 4x LDS at 95°C for 5 minutes. 20 μl of each sample was then used for SDS-PAGE and western blot (see below).

3.4.4 His-YTH purification

Transformation was performed on high-efficiency T7 Express competent *E. Coli* cells (C2566H) with 100 ng plasmid per manufacturer's instructions. Briefly, cells and plasmid DNA were incubated on ice for 30 minutes, prior to 45 second heat shock at 42°C. The tubes then recovered from heat shock on ice for 6 minutes and were then grown in 1 mL LB at 37°C, 225 rpm for 60 minutes. Transformed cells were plated on ampicillin plates and incubated at 37°C overnight. One colony was picked and put into 5 mL LB medium (1/1000 v/v Ampicillin) then grown overnight at 37°C in a shaking incubator. Induced from this overnight culture, a total of 1L culture was grown to an OD600 of 1 prior to a 12-hour induction by 1 mM IPTG at 16°C.

The culture medium was then centrifuged, which resulted in a pellet resuspended in 30 mL Binding Buffer (1x Native Purification Buffer, pH 8, R901-01) by pipetting. This was then sonicated on ice in a pattern of 3 seconds on followed by 9 seconds off for 2 minutes. The lysate was then centrifuged at 3,000 x g for 15 minutes and the supernatant transferred to a new tube to remove debris. To prepare the affinity purification column (R90115, Thermo), Ni-NTA agarose was resuspended by inversion, then 1.5 mL of the gel was pipetted into a 10-mL column. The supernatant was removed after the resin settled by gravity. To the settled resin, 6 mL water was added, allowed to settle again, and again the supernatant was aspirated. The column was then washed twice in an identical manner using 6 mL Native Binding Buffer. 8 mL cell lysate was then added and incubated for 60 minutes. Following binding, the resin settled by gravity and the supernatant was removed. The resin was then washed four times with 8 mL Native Wash Buffer. After washing, protein was eluted by addition of 8 mL Native Elution Buffer (1x Native Purification Buffer,

4M Imidazole, pH 4) and collected in 1 mL fractions. Four columns were run total; fractions were then pooled and analyzed on SDS-PAGE. Elution fractions 3, 4, and 5 were pooled together before being concentrated with an Amicom 3 kDA MWCO column. Samples were then loaded onto a size exclusion column and 60 fractions were collected in 20 mM Tris (pH 7.5), 150 mM NaCl following a published protocol (Woodcock et al. 2020).

3.4.5 In vitro pulldown

500 ng GST–CTD (SRP2120, Sigma) was phosphorylated by 240 ng of CDK7–Cyclin H–MNAT1 (PV3868, Thermo) or CDK9–Cyclin T1 (14–685, Sigma) in a total of 20 μL 8 mM 3-(N-morpholino) propanesulfonic acid (MOPS) (pH 7), 0.2 mM EDTA, and 1 mM $MgCl_2$ supplemented with freshly added 0.1 mM ATP and 0.25 mM DTT. The reaction was incubated at 30 °C for 1 hour. Control reactions without GST–CTD, without kinase, or without ATP were incubated under the same conditions.

This 20-μL phosphorylation reaction was combined with 75 μL of wash buffer (300 mM NaCl, 100 mM Tris-Cl (pH 8), 0.2 mM EDTA, and 0.1% v/v Triton X-100), 240 pmol of YTH in 5 μL of storage buffer (10 mM Tris-Cl (pH 7.4), 500 mM NaCl, 2.5 mM $MgCl_2$, 10% v/v glycerol), and then rotated at 4°C overnight in a total volume of 100 μL. One tenth (10 μL) of the binding mixture was saved as the input.

To this mixture, 20 μg/mL anti-6x His antibody (ab137839) was added and then rotated overnight. Protein A/G Dynabeads were then added to a total of 20 μL per sample followed by 4-hour rotation; Dynabeads were washed four times in 200 μL Binding Buffer before use. The samples were eluted by boiling at 95°C for 5 minutes in a total of 40 μL

1xLDS supplemented 100 mM DTT and the supernatant containing eluted protein was removed separated from the protein A/G Dynabeads using a magnetic rack; 12 μL of each sample was then used for western blot (below).

3.4.6 Western blot

Samples were loaded onto 12-well 4–12% polyacrylamide Bis-Tris gels (NP03322, Invitrogen) and ran at 150V for 1 hour. The gels were then transferred to polyvinylidene fluoride membranes (IPVH00010, Millipore). Following transfer, membranes were blocked overnight in 10% w/v milk (1706404, Bio-Rad). The blots were probed with 1/1000 v/v anti-YTHDC1 (ab122340), 1/1000 v/v anti-hnRNPG (ab190352), 1/1000 v/v anti-FLAG (anti-DDDDK, ab1162), anti-6x-HIS (ab137839), anti-GST (Cell Signaling Technology, 2624), anti-SNRP70/U1-70k (ab83306), anti-Histone H3 (96C10) (Cell Signaling Technology, 96C10) followed by 1/10000 v/v sheep anti-mouse IgG (NA931V, Cytiva) or 1/10000 v/v donkey anti-rabbit IgG conjugated to horseradish peroxidase (NA934V, Cytiva). For GAPDH, the blot was visualized directly by HRP-conjugated anti-GAPDH (GA1R, Thermo). The blots were then visualized with ECL Prime Western Blotting Detection Reagents (RPN2232, Amersham) using a BioRad ChemiDoc MP

Chapter 4

m^6A-dependent coordination of mRNA and tRNA

Acknowledgments: This chapter is derived from unpublished data inspired by a co-authored publication in *Nature Communications* (C. P. Watkins et al. 2022). I would like to thank Dr. Chris Katanski for assistance with coding, Dr. Chris Watkins for his development of MSR-Seq, Mateusz Halucha for sequencing knockdown samples, and Noah Peña for data visualization in Figure 4.3.

4.1 Introduction

Transfer RNAs (tRNAs) are the most abundant family of small non-coding RNAs acting as adaptors during protein production, decoding the sequence information contained in the anticodon-paired mRNA by providing the cognate amino acid. Owing to their central role in translation, tRNAs are essential in cellular proliferation, fitness, and adaptation (Pan, T., 2018). tRNA's properties in translation can be regulated through their expression, including at the level of transcription by RNA Polymerase III, aminoacylation (charging), and chemical modification (C. P. Watkins et al. 2022). Thus, understanding the properties of tRNA is essential to our understanding of gene regulation.

Crucial functional diversity is imparted to tRNA through chemical modification. Human tRNA contains 13 chemical modifications on average (Tsutomu Suzuki 2021) which each fine-tunes many of the tRNA's properties, including stability, localization, and folding (C. P. Watkins et al. 2022). Mutations in tRNA modification enzymes have been linked to

several cancers (Tsutomu Suzuki 2021), underscoring the importance of regulating the tRNA modification landscape.

Accurate measurement of tRNA properties has been complicated by incompatibility with most common RNA-sequencing methods. The abundantly modified tRNA body combined with higher-order structures produces incomplete reverse transcription products during sequencing incompatible with downstream analysis (C. P. Watkins et al. 2022). Our group developed a multiplex small RNA-seq library preparation method (MSR-seq) to measure the properties of small RNAs, including tRNA. We demonstrated the use of MSR-seq to simultaneously measure tRNA abundance, charging, modification, and fragmentation in response to stress treatments (C. P. Watkins et al. 2022).

Motivated by our advancements in small RNA sequencing, I investigated whether tRNA properties were subject to regulation by N6-methyl-adenosine (m^6A) in RNA (m^6A-RNA). Cellular m^6A levels are dynamically and co-transcriptionally regulated by the opposing actions of methyltransferase ('writer') and demethylase ('eraser') proteins (Roundtree, Evans, et al. 2017). Following deposition by the m^6A methyltransferase complex (MTC), a heterodimer of METTL3 and METTL14 (Xiang Wang et al. 2016), the modification can be recognized by an m^6A-reader protein, allowing m^6A to impart its downstream effect on gene regulation (Roundtree, Evans, et al. 2017). Two reader proteins include YTH N6-Methyladenosine RNA Binding Protein C1 (YTHDC1) and heterogeneous nuclear ribonucleoprotein G (hnRNPG), they primarily localize in the nucleus and are involved in phase separation (Zhou et al. 2019). The involvement of m^6A in mRNA on regulation of tRNA is not understood and represents a potential connection between mRNA modification and tRNA modification. I, therefore, aimed to perform a broad initial

bioinformatic survey on the involvement of m^{6}A writers and reader proteins on tRNA regulation to take advantage of our MSR-Seq platform.

4.2 Results

4.2.1. tRNA modification enzyme transcript abundance is disrupted on m^{6}A writer and reader knockdown

Chemical modifications of tRNA fine-tune crucial properties and function (C. P. Watkins et al. 2022). Our MSR-Seq platform allows for simultaneous measurement of tRNA abundance, charging, fragmentation, and modification, motivating our studies on the mRNA transcripts of the tRNA modification enzymes. I aimed to analyze the expression of tRNA modification enzymes in existing datasets following knockdown of m^{6}A writer and reader proteins in existing data sets to examine the m^{6}A-dependence of these transcript's regulation (GSE 56010).

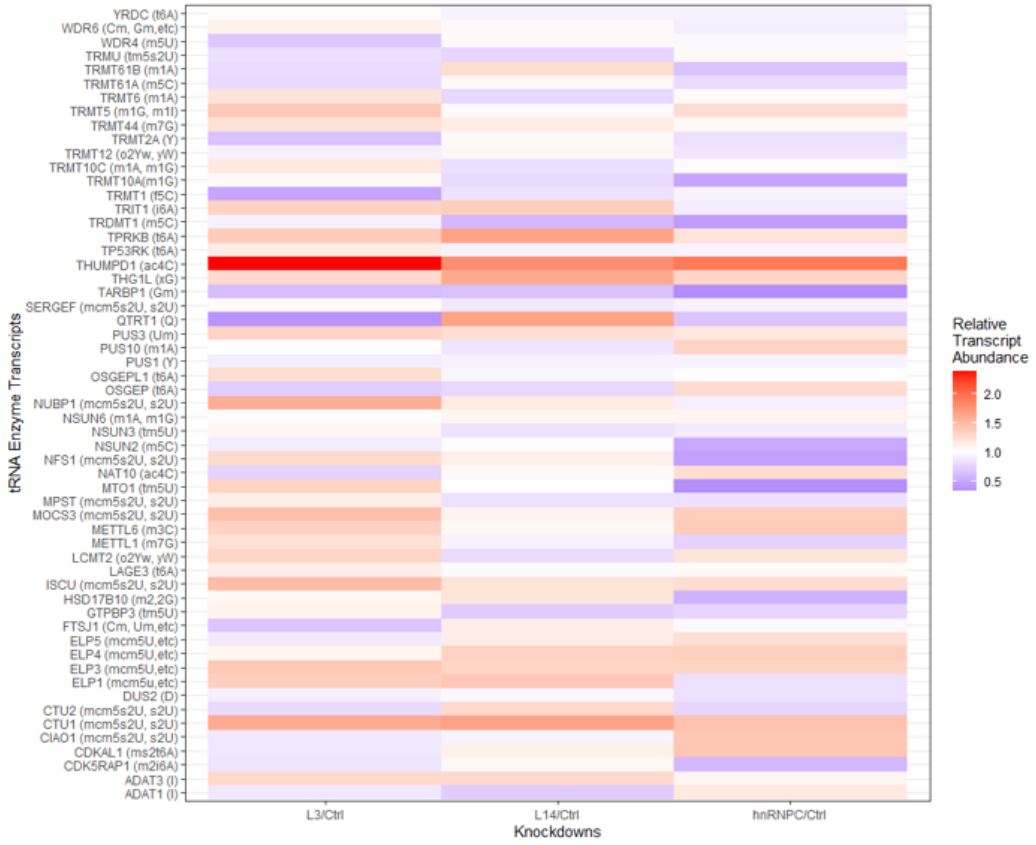
Knockdowns were performed on both components of the core MTC heterodimer, METTL3, and METTL14, in addition to the m^{6}A reader heterogeneous nuclear ribonucleoproteins C1/C2 (HNRNPC) in HEK293T cells. First, reads were mapped to the human genome from control and knockdown mRNA-seq data, generating a transcript read count for each gene in each knockdown condition. The data was then filtered for reads mapped to tRNA enzyme transcripts; transcript abundance for each knockdown condition was calculated relative to the negative control and visualized by heatmap (**Figure 4.2.1a**).

This analysis shows that the mRNA expression of tRNA modification enzymes is disrupted following the knockdown of either m^{6}A writer or reader proteins. To account for

potential false positives in the data, I applied additional filters (**Figure 4.2.2b**). The average transcript enrichment following METTL3 and METTL14 knockdown was taken ('MTC Average') and used to classify the dataset; for the set of transcripts with 'MTC Average' > 1, transcripts where METTL3 or METTL14 KD relative abundance < 1 were omitted. For MTC average < 1, data sets where METTL3 or METTL14 KD abundance >1 were omitted. This filter removes transcript hits whose abundance is differentially disrupted depending on which component of the MTC core heterodimer is knocked down. Additionally, transcripts with an 'MTC Average' between 0.95-1.10 were removed.

The number of tRNA enzymes for each modification with altered abundance was then counted following filtering (**Figure 4.2.1c**). A majority of affected enzymes target the methoxycarbonylmethyl-2-thiouridine (mcm^5s^2U) pathway, including components of the Elongator Complex composed of ELP1-6 proteins. Thus, the analysis here supports m^6A-dependent regulation of tRNA enzyme transcripts involved in installing mcm^5s^2U and its derivatives at the tRNA wobble U34 position. Additionally, THUMPD1 (ac^4C) and TARBP1 (Gm) have the highest enrichment and depletion following MTC knockdown. Sorting by enrichment following knockdown of m^6A reader protein hnRNPC shows that THUMPD1 (ac^4C) is the most enriched mRNA transcript, with TARBP1 (Gm) being the most depleted, matching with the MTC knockdown data. This concurrence may suggest that regulation of these tRNA modification enzyme transcript's abundance is both m^6A- and hnRNPC-dependent. Taken together, our analysis supports m^6A-dependent regulation of tRNA modification enzyme transcripts.

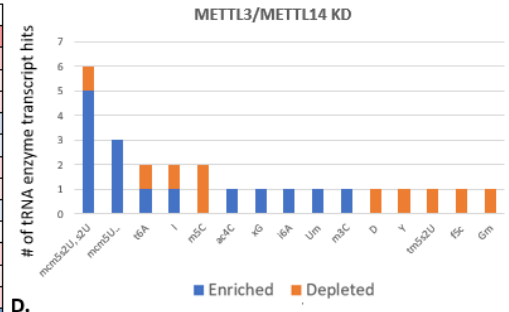
A.



B.

	MTC Average	METTL3	METTL14	hnRNPC
THUMPD1 (ac4C)				
CTU1 (mcm5s2U, s2U)				
TPRKB (t6A)				
THG1L (xG)				
ELP1 (mcm5u, ...)				
NUBP1 (mcm5s2U, s2U)				
ELP3 (mcm5U, ...)				
ISCU (mcm5s2U, s2U)				
TRIT1 (i6A)				
ADAT3 (I)				
MOCS3 (mcm5s2U, s2U)				
PUS3 (Um)				
ELP4 (mcm5U, ...)				
NFS1 (mcm5s2U, s2U)				
METTL6 (m3C)				
TRMT44 (m7G)				
HSD17B10 (m2,2G)				
NSUN2 (m5C)				
DUS2 (D)				
PUS1 (Y)				
CIAO1 (mcm5s2U, s2U)				
TRMU (tm5s2U)				
ADAT1 (I)				
TRDMT1 (m5C)				
OSGEP (t6A)				
TRMT1 (f5C)				
TARBP1 (Gm)				

C.



D.

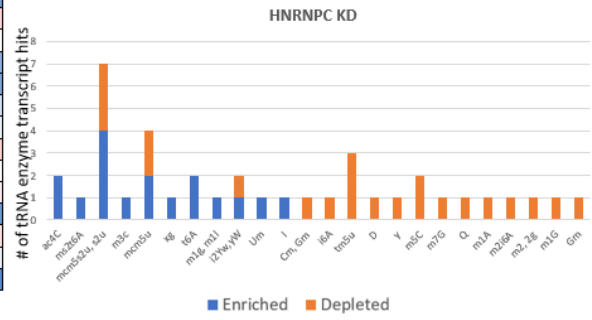


Figure 4.2. 1 tRNA modification enzyme transcript abundance is disrupted on m6A writer and reader knockdown. (a). RNA-seq data filtered to reads mapping to tRNA enzyme transcripts following knockdown of METTL3, METTL14, hnRNPC, or with a negative control siRNA was visualized by heatmap. The proportion of total reads mapped to each enzyme transcript in experimental condition knockdowns was normalized to the control knockdown condition to calculate "relative transcript abundance," or abundance of a transcript in the experimental relative to control conditions. Enzyme transcript names are listed along with the modification their act on. (b) Heatmap generated in a further filtered to remove transcripts where a change in relative abundance shows differential dependence on METTL3 and METTL14. (c). Quantification of modification targets of enriched (relative transcript abundance > 1.1) and depleted (relative transcript abundance < 0.9) transcripts following knockdown of METTL3 or METTL14. (d). As in c but following hnRNPC knockdown.

4.2.2 hnRNPG-PAR-CLIP data analysis shows enrichment and depletion of certain tRNAs

MSR-seq has enabled accurate measurement of tRNA abundance, which motivated exploratory bioinformatic analyses of existing datasets. Heterogeneous nuclear ribonucleoprotein G (hnRNPG) is an RNA-binding protein involved in the formation of subnuclear RNA processing bodies. We previously established that hnRNPG acts in co-transcriptional regulation of m^6A-dependent pre-mRNA splicing, during which we generated hnRNPG-PAR-CLIP datasets (Zhou et al. 2019; Brugiolo et al. 2017). I analyzed this previously published PAR-CLIP data to gain insight into potential interactions between hnRNPG and tRNA (GSE74085, GSE114311).

The hnRNPG-PAR-CLIP reads were mapped to the library of human tRNAs, which allowed for the calculation of the abundance of mapped reads by normalizing to the total read count. hnRNPG-PAR-CLIP tRNA abundances were then compared to tRNA abundances determined by from MSR-Seq (C. P. Watkins et al. 2022). We found that hnRNPG-PAR CLIP

enriches (log_2(PAR-CLIP abundance/MSR-Seq Abundance > 1) specific isoacceptors relative to MSR-Seq (eg. CysGCA, GlnTTG, HisGTG, SerCGA) and depletes (log_2(PAR-CLIP abundance/MSR-Seq Abundance < -1) many others (eg ArgCCG, ArgTCG, LeuCAA, IleTAT, ThrAGR) (**Figure 4.2.2a**). Interestingly, hnRNPG-PAR-CLIP enriches the abundance of many mitochondrial tRNAs (mt-tRNAs), depleting only a select few (eg. mtArgTCG) (**Figure 4.2.2b**). Taken together, our data suggest that hnRNPG interacts with a specific subset of human tRNAs as well as many mt-tRNAs.

Given the previously established ability of both hnRNPG (Chapter 2) and YTHDC1 (Chapter 3) to directly interact with RNA Polymerase II, these reader proteins may represent a platform for crosstalk and coordination of activities between multiple polymerases. Therefore, I asked whether YTHDC1 was able to interact with components of RNA Polymerase III. To test this hypothesis, anti-FLAG immunoprecipitations were performed following transfection with either empty FLAG or FLAG-YTHDC1 constructs. The data show that RNA Polymerase III Subunit A (POLR3A) co-immunoprecipitates with FLAG-YTHDC1 in HEK293T cells following lipofectamine transfection. This offers support for the hypothesis that m^6A-reader proteins can interact with RNAPIII where they can potentially mediate crosstalk to RNAPII, and taken together with our support for hnRNPG-tRNA interactions implicates m^6A reader proteins in tRNA processing.

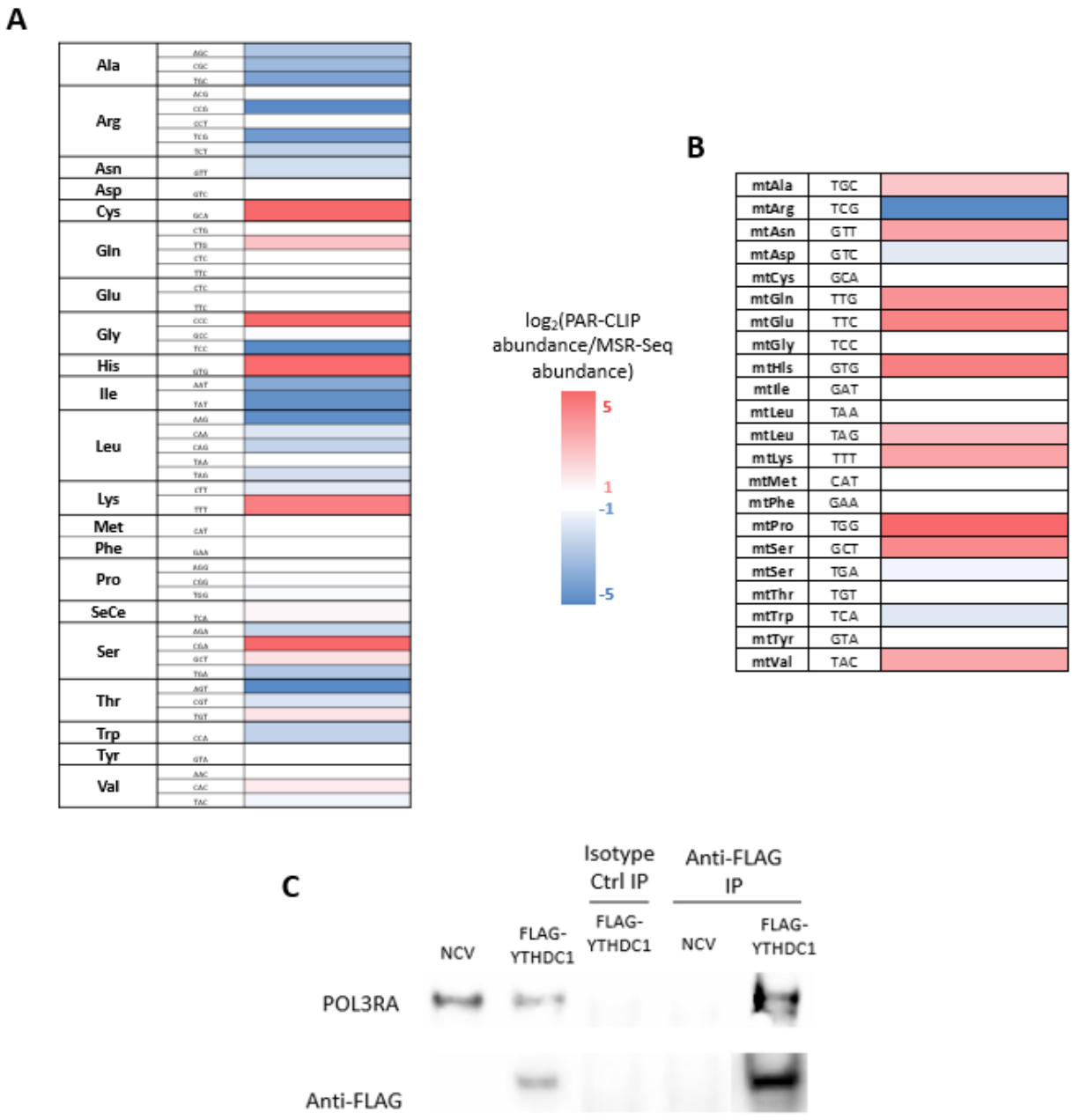


Figure 4.2. 2 hnRNPG-PAR-CLIP data analysis shows the enrichment and depletion of certain tRNAs (a). Heatmap showing tRNAs enriched and depleted in hnRNPG-PAR-CLIP relative to MSR-Seq data. Abundance for both hnRNPG-PAR-CLIP and MSR-Seq reads were calculated relative to the sum of total reads, and enrichment was calculated by taking the $\log_2$ ratio of the two abundances, where > 1 (red) corresponds to enrichment and < -1 (blue) corresponds to depletion. Isoacceptors and their cognate amino acids are labeled. (b). As in a, but visualizing mt-tRNA enrichment rather than human tRNA enrichment. c. POLR3A pulled down with YTHDC1. (c). Western blot shows RNAPIII component POLR3A co-precipitates with FLAG-YTHDC1 during anti-FLAG immunoprecipitation. Membranes

were blotted against POL3RA (top) and FLAG (for FLAG-YTHDC1, bottom). Lanes from left to right: (1) Input, cells transfected with empty FLAG non-coding vector (NCV); (2) Input, cells transfected with the full-length FLAG-YTHDC1 construct; (3) Immunoprecipitation performed with isotype negative control antibody to control for isotype non-specific interactions of the anti-FLAG antibody; (4-5) Anti-FLAG immunoprecipitation following transfection with either the empty FLAG vector or full-length FLAG-YTHDC1 construct, respectively.

4.2.3 MSR-Seq supports possible role of m^6A in tRNA m^3C modification

The preliminary bioinformatic survey here performed on m^6A writers and readers broadly supports a connection between these proteins and tRNA regulation. We performed MSR-Seq following knockdown (KD) of METTL3, METTL14, hnRNPG, YTHDC1, or a scrambled control RNA to better understand the importance of m^6A writers and readers on the tRNA landscape. We found that writer and reader knockdown showed little global change in tRNA abundance, but that specific tRNAs were enriched (**Fig 4.2.3a**). One such isoacceptor is LysTTT, whose abundance was increased following METTL3 KD. Interestingly, our PAR-CLIP analysis supports hnRNPG interaction with this isoacceptor (**Figure 4.2.2a**).

Methyl-3-cytosine (m^3C) can occur at position 32 in the anticodon loop of $tRNA^{Ser}$, $tRNA^{Thr}$, $tRNA^{Arg}$(CCT), $tRNA^{Arg}$(TCT) written by METTL2A, METTL2B, METTL6, or at position 47d(e2) in the variable of the loop of $tRNA^{Ser}$ and $tRNA^{Leu}$(CAG) with an unknown writer (Bohnsack et al. 2022). Knockdown of hnRNPG, METTL3, and METTL14 had a modest impact on the mutation rate at position 32, but the effect was more dramatic at position 47d, implying a global change in modification levels at this site (**Figure 4.2.3b**). Knockdown of hnRNPG and METTL3 increased the abundance of $tRNA^{Leu}$(CAG)/$tRNA^{Ser}$ isodecoders with C47d at position 47d relative to $tRNA^{Ser}$ with T47d (**Fig 4.2.3c**). This data

suggests hnRNPG may be involved in regulating the abundance of m^{3}C47d in an m^{6}A-dependent manner, and further supports the importance of m^{6}A writer and reader proteins in tRNA processing.

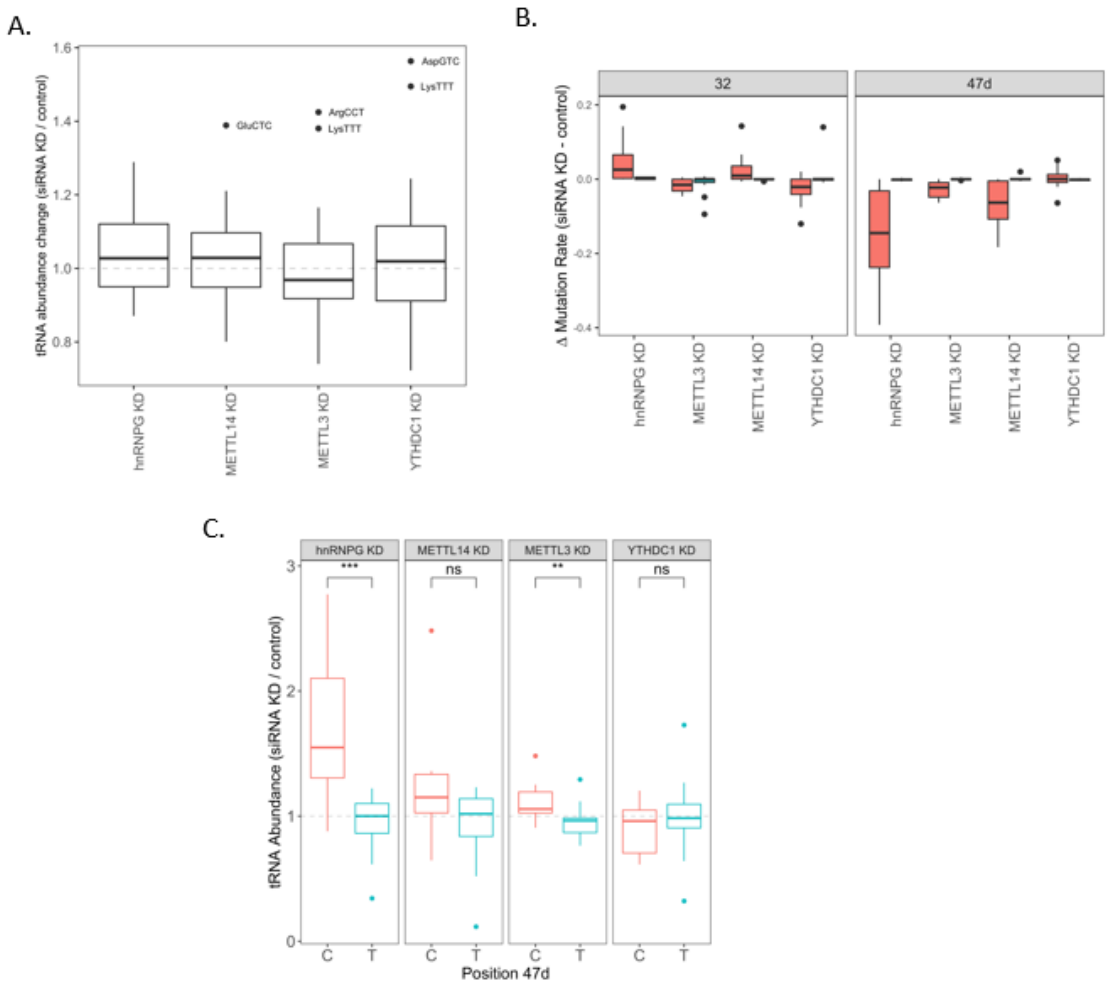


Figure 4.2. 3 MSR-Seq supports the possible role of m^{6}A in tRNA m^{3}C modification (a). tRNA abundance change following si-hnRNPG, si-METTL14, si-METTL3 or si-YTHDC1 transfection in YTHDC1 relative to s as determined by MSR-Seq. Abundance ratios normalized to si-Ctrl. Isoacceptors whose abundance is significantly changed on writer or reader knockdown are listed as points in the boxplot. (b). m^{3}C modification determined by the change in mutation rate (writer/reader siRNA KD – siCTRL). Red: m^{3}C47d(e2) modified tRNA$^{Ser/Leu}$ isodecoders. Blue: tRNA^{Ser} isodecoders with U47d(e2) sequence. (c). The abundance of tRNA's containing C (site of m^{3}C47d modification) or T at position 47d following m^{6}A writer or reader knockdown normalized to the siCTRL. tRNA with cytosine at position 47d have their abundance significantly increased following the knockdown of hnRNPG or METTL3.

4.3 Conclusions

Data collection has historically been a rate-limiting step in research, but advancements in high-throughput sequencing have generated an abundance of datasets that have yet to be fully analyzed. Our development of a multiplex small RNA-seq library preparation method (MSR-seq) (C. P. Watkins et al. 2022) motivated testing the connection of general tRNA processing to the system of m^6A writers and readers explored through this thesis. The aim of this broad bioinformatic survey on the properties of tRNA and their processing factors on the m^6A machinery was to provide initial evidence for or against the m^6A-dependence of tRNA expression and modification, which can be used to inform future mechanistic inquiries. Indeed, these "fishing experiments" provide preliminary corroboration that the m^6A modification connects properties of mRNA to tRNA.

Bioinformatic analysis of m^6A writer and reader knockdown datasets showed transcript abundance of THUMPD1 was the most enriched of all tRNA modification enzymes following knockdown of METTL3, METTL14, or HNRNPC. THUMPD1 is a conserved adaptor that regulates the production of tRNA N4-acetylcytidine (ac^4C) by the catalytic RNA cytidine acetyltransferase, NAT10 (Sharma et al. 2015). The THUMPD1 promoter has transcription factor binding sites for heterogeneous nuclear ribonucleoproteins L and H1 (hnRNPL and hnRNPH1, respectively) (Fishilevich et al. 2017), which supports the involvement of hnRNPs, a group of proteins associated with forming membrane-less processing bodies, in THUMPD1 regulation. In relation to health and disease, mutation of THUMPD1 results in syndromic intellectual disability (Broly et al. 2022). In contrast to the enrichment of ac^4C adapter THUMPD1, TARBP1 transcripts,

which act on 2′O-methylation of G34 (de Crécy-Lagard et al. 2019), were the most depleted of all tRNA modification enzymes following METTL3, METTL14 or HNRNPC knockdown.

The most consistent targeted modification by affected enzymes following writer or reader knockdown (**Figure 4.2.1c, d**) was methoxycarbonylmethyl-2-thiouridine (mcm^5s^2U) and its derivates, a modification at the tRNA U34 wobble position in $tRNA^{Glu/Gln/Lys}$ isoacceptors. Analysis of hnRNPG-PAR-CLIP data suggests that hnRNPG interacts with $tRNA^{Gln}$(TTG) and $tRNA^{Lys}$(TTT) which can contain mcm^5s^2U. The generation of mcm^5s^2U occurs through a series of biological intermediates (Lentini, Ramos, and Fu 2018). While the exact atomic mechanism is unknown, it is known the elongator complex generates initial 5-carbonylmethyluridine(cm^5U) and 5-carbamoylmethyluridine (ncm^5U) at the site of interest (Karlsborn et al. 2016; Kolaj-Robin and Séraphin 2017; Lentini, Ramos, and Fu 2018). The bioinformatic analysis here presented shows that the knockdown of core m^6A writers and reader hnRNPC lead to an increase in transcript abundance for elongator complex subunits ELP1, 3, and 4 (**Figure 4.2.1b**). This analysis suggests the involvement of m^6A in regulating elongator complex formation as part of mcm^5s^2U production at the wobble position, and by extension suggests an indirect involvement of m^6A in mRNA decoding.

Interestingly, our analyses of hnRNPG-PAR-CLIP demonstrated enrichment of both specific human tRNAs and many mt-tRNAs (**Figure 4.2.2a, b**), which suggests these tRNAs interact directly with hnRNPG. While hnRNPG is not localized to mitochondria, the mechanism by which the nucleus and the mitochondria communicate is not fully understood (Shaukat et al. 2021). It is thought mt-tRNA may be further processed in the cytoplasm to enable nuclear crosstalk (Bruni, Lightowlers, and Chrzanowska-Lightowlers

2017; Jády, Ketele, and Kiss 2018), and that part of this crosstalk is enabled by the generation of mitochondrial tRNA fragments (mt-tRFs) (Shaukat et al. 2021). It is possible that hnRNPG is part of the crosstalk between the mitochondria and the processing in the nucleus, acting as a platform to interact with mt-tRNAs.

An outstanding question in the field is how eukaryotic polymerases coordinate their transcriptional activities. Our observations that m^6A-reader proteins can interact directly with RNAPII (Chapters 2, 3) lead to the hypothesis that these reader proteins may also interact with RNA polymerase III, thus enabling crosstalk to coordinate transcription between the two polymerases. The anti-FLAG immunoprecipitation here shows that RNA Polymerase III Subunit A (POLR3A) co-immunoprecipitates with FLAG-YTHDC1 in transfected HEK293T cells, similarly to its co-precipitation of the RNA Polymerase II CTD (Chapter 3). While the RNase dependence or the impact of the FLAG-tag on POLR3A pulldown was not tested, this IP result offers preliminary evidence for RNAPII-RNAPIII cross-talk mediated by YTHDC1, although whether this interaction with RNAPIII is a direct protein-protein interaction remains to be answered.

Our MSR-Seq data demonstrated a potential connection between m^6A and tRNA methyl-3-cytosine (m^3C), particularly at position 47d(e2). m^3C47d occurs in the variable of loop of $tRNA^{Ser}$; therefore, our data is further supported by observed enrichment of SerCGA reads in the hnRNPG-PAR-CLIP data. However, there is currently no known writer of m^3C47d (Bohnsack et al. 2022), which makes it challenging to study the mechanism by which m^6A regulates m^3C47d levels. Still, this data provides further support for regulation of tRNA modification by m^6A.

In summary, I performed a bioinformatic survey to generate potential connections between tRNA and m^6A-writers and readers and identified several possible candidates. These analyses suggest interactions between m^6A reader protein YTHDC1 and RNAPIII, reader protein hnRNPG interactions with both human and mitochondrial tRNA, involvement of m^6A writers METTL3 and METTL14 as well as reader hnRNPC on regulating tRNA modification enzyme transcript abundance, particularly those relating to mcm^5s^2U and its derivatives, and between m^6A writers and readers on the regulation of tRNA m^3C47d abundance.

4.4 Materials and Methods

4.4.1. tRNA modification enzyme abundance data analysis

Previous METTL3, METTL14, hnRNPC, and control knockdown RNA-seq data was analyzed (GSE56010)**.** Total reads mapped to the human genome were first calculated for later normalization, and then the data was filtered for reads which mapped to known or predicted tRNA enzymes. For each gene, the abundance was determined by normalization to total read count. Relative abundance following knockdown relative to the control was taken by dividing normalized experimental knockdown read abundance to normalized control knockdown read abundance. A three color heatmap (red-white-blue) was generated using bbplot showing relative reads. Data was further filtered using 'MTC Average,' the average transcript enrichment following METTL3 and METTL14 knockdown to remove genes whose expression was differentially disturbed in each MTC knockdown condition. For the set of transcripts with 'MTC Average' > 1, transcripts where METTL3 or METTL14 KD relative

abundance < 1 were omitted. For MTC average < 1, data sets where METTL3 or METTL14 KD abundance >1 were omitted. Transcripts with 'MTC Average' values between 0.95-1.10 were additionally removed to account for noise.

4.4.2 hnRNPG-PAR-CLIP data analysis

PAR-CLIP-hnRNPG datasets (GSE74085, GSE114311) were uploaded to UChicago's midway2 cluster for analysis. bowtie2 (2.3.3.1) was used to align filtered reads to both human and mitochondrial tRNA (Langmead and Salzberg 2012). The total read counts for all reads mapped to tRNA was calculated and used to normalize reads of each tRNA gene. hnRNPG-PAR-CLIP data was compared to MSR-seq data (C. P. Watkins et al. 2022) for which reads for each tRNA gene were likewise normalized to the sum of total tRNA reads. Following this, enrichment was calculated by $\log_2$(hnRNPG-PAR-CLIP/MSR-Seq) where a value > 1 corresponds to enriched and a value < -1 corresponds to depleted. Data was separated into mt-tRNA and human tRNA and a three-color (red-white-blue) heatmap was generated by Microsoft Excel.

4.4.3 FLAG-YTHDC1 Immunoprecipitation

For FLAG-YTHDC1 transfection, 15 cm plates of HEK293T cells were transfected with 60 μg FLAG-tagged plasmid using Lipofectamine 2000 (11668019, Thermo) per manufacturer's protocol. Briefly, both 3 mL Opti-MEM I Reduced Serum Medium with 60 μg FLAG-tagged plasmid (YTHDC1) and 3 mL Opti-MEM I Reduced Serum Medium with 100 μl Lipofectamine 2000 were individually mixed and incubated at RT for 5 minutes. The two mixtures were then combined and incubated for an additional 20 minutes at RT, before the combined 6 mL DNA-Lipofectamine-2000-Medium mix was added to the cell culture plate

and mixed gently by rocking back and forth. Cells were initially incubated for 6 h at 37°C before having medium changed to a fresh, antibiotic-free medium and cultured for an additional 24 hours.

Anti-FLAG (ab122340) was added to the lysate to a total volume of 90-100 μL /100 mM plate of cells. The mixture was then rotated overnight at 4°C. Protein A (Thermo 10001D) were washed in 200 μL Wash Buffer (300 mM NaCl, 100 mM Tris (pH 8.0), 0.2 mM EDTA, 0.1% Triton X-100) and then resuspended in Wash Buffer (~0.05 mg/μL, such that 1.5 mg is in 30 μL) and rotated for 2 hours at 4°C. The beads were then collected, and the supernatant removed using a magnetic rack before being washed four times with 200 μL ice-cold Wash Buffer. The sample was then eluted by boiling in 30 μL 4x LDS at 95°C for 5 minutes. 20 μL of each sample was then used for SDS-PAGE and western blot.

Samples loaded onto 12-well 4–12% polyacrylamide Bis-Tris gels (NP03322, Invitrogen) and ran at 150V for 1 hour. The gels were then transferred to polyvinylidene fluoride membranes (IPVH00010, Millipore). The membranes were blocked in 10% w/v milk (1706404, Bio-Rad). The blots were probed with 1/1000 v/v anti-YTHDC1 (ab122340), 1/1000 v/v anti-hnRNPG (ab190352), The blots were probed with 1/1000 v/v anti-YTHDC1 (ab122340), 1/1000 v/v anti-hnRNPG (ab190352), 1/1000 v/v anti-FLAG (anti-DDDDK, ab1162), anti-6x-HIS (ab137839), anti-GST (Cell Signaling Technology, 2624), anti-SNRP70/U1-70k (ab83306), anti-Histone H3 (96C10) (Cell Signaling Technology, 96C10) followed by 1/10000 v/v sheep anti-mouse IgG (NA931V, Cytiva) or 1/10000 v/v donkey anti-rabbit IgG conjugated to horseradish peroxidase (NA934V, Cytiva). For GAPDH, the blot was visualized directly by HRP-conjugated anti-GAPDH (GA1R,

Thermo). The blots were then visualized with ECL Prime Western Blotting Detection Reagents (RPN2232, Amersham) using a BioRad ChemiDoc MP

4.4.4 MSR-seq sample preparation

Human embryonic kidney (HEK) cell line HEK293T/17 (CRL11268) was obtained from the American Type Culture Collection (ATCC) and cultured in Dulbecco's Modified Eagle's Medium (DMEM) with high glucose and L-glutamine, without sodium pyruvate (HyClone, SH30022.01), with 10% FBS and 1% Pen–Strep (Penicillin–Streptomycin) in a 37°C incubator at 5% CO_2. Transfection of siRNA was done using Lipofectamine 2000 (11668019, ThermoFisher) following passage into culture media without 1% Pen–Strep. For each knockdown condition, two 10 cm plates were each transfected using 30 uL Lipofectamine 2000 and 600 pmol siRNA (METTL3: Qiagen Cat. No. SI04137096 FlexiTube siRNA; METTL14: Qiagen Cat. No. SI00459942 FlexiTube siRNA; hnRNPG Qiagen Cat No. SI00700077 FlexiTube siRNA) after diluting into 3.0 mL Gibco™ Opti-MEM™ I Reduced Serum Medium (Cat. No. 31985-062). Transfections were incubated for 37°C incubator at 5% CO_2 for 48 hours. RNA purified by Trizol Reagent was sequenced with by Mateusz Halucha per (C. P. Watkins et al. 2022) and data analysis performed with Noah Peña, who generated the graphs used in Figure 4.2.3.

Chapter 5

Simultaneous m^{6}A and Pseudouridine Nanopore Profiling Reveals Coordination in Translation

Acknowledgements: This chapter is taken from a Brief Communication submitted to Nature Biotechnology on May 2nd, 2023. The authors of that manuscript are Sihao Huang, Adam C. Wylder, and Tao Pan, where S.H. and A.C.W. are co-first authors. S.H. performed all computational work including NanoSPA pipeline development and sequencing data analysis. A.C.W. performed all experimental work including si-knockdowns, polysome profiling, and nanopore sequencing. S.H. and T.P. conceived the project, S.H., A.C.W. and T.P. designed the experiments and wrote the paper.

We thank Dr. Lisheng Zhang for providing analyzed m^{6}A-SAC-seq data prior to publication and David Pan for contribution to coding. This work was supported by NIH (RM1 HG008935 to T.P.). The nanopore seq data of this study are submitted to NCBI GEO (accession number: GSE230936). The published nanopore WT samples (replicate 1 and 2) used to validate the m^{6}A model are GSM5467024 and GSM5467025. NanoSPA codes and pipeline are available on GitHub (https://github.com/sihaohuanguc/NanoSPA).

5.1 Introduction

*N*6-methyladenosine (m^6A) and pseudouridine (Ψ) are the top two most abundant internal mammalian mRNA modifications measured by quantitative mass spectrometry (Roundtree, Evans, et al. 2017). M^6A is the most extensively studied mRNA modification; it participates in many cellular processes including mRNA stability, splicing, export, localization, and translation(Roundtree, Evans, et al. 2017; Frye et al. 2018). The best known Ψ function is innate immune avoidance when in delivered mRNAs, as shown in the successful COVID-19 mRNA vaccines (Jackson et al. 2020); Ψ has also been shown to affect splicing and translation (Karikó et al. 2008; Eyler et al. 2019; Anderson et al. 2010; Martinez et al. 2022).

A major knowledge gap in the biological studies of m^6A and Ψ is how they enhance or antagonize each other in the same mRNA transcript. The investigation of coordinated m^6A and Ψ function requires mapping methods that can simultaneously report m^6A and Ψ in the same sequencing library. So far, Illumina sequencing of m^6A and Ψ has always been performed separately and independently. Nanopore sequencing has also been employed to map either m^6A or Ψ in numerous studies and pipelines (L. Hu et al. 2022; S. Huang et al. 2021) but these studies also considered m^6A or Ψ separately; therefore, no prior studies could reveal the potential crosstalk between m^6A and Ψ in the mRNA transcriptome. Here we develop a nanopore sequencing pipeline named **Nano**pore **S**imultaneous investigation for **P**seudouridine and m^6**A** (NanoSPA) that analyzes m^6A and Ψ modifications in the same data. We apply NanoSPA to both the human transcriptome with or without knocking down the m^6A writer METTL3 or one of the thirteen Ψ writers to reveal their co-dependence, and to polysome profiling to reveal their effects and co-dependence on translation.

5.2 Results

5.2.1 NanoSPA method for simultaneous analysis of m^{6}A and Ψ and their relationship in the transcriptome

To investigate m^{6}A and Ψ at the same time, we designed a workflow (**fig. 5.2.1a**) containing two machine learning modules which predicted these two modifications. For nanopore sequencing of Ψ, we previously developed and published the NanoPsu pipeline (S. Huang et al. 2021). The challenge here was to make a new model for nanopore sequencing of m^{6}A compatible with NanoPsu to enable simultaneous analysis of both modifications. For this purpose, new m^{6}A training data would be crucial. We took advantage of the recently published m^{6}A-SAC-seq data by Illumina sequencing, which mapped m^{6}A at single-base resolution and with modification stoichiometry transcriptome-wide (L. Hu et al. 2022). Using the HeLa m^{6}A-SAC-seq data together with our nanopore direct mRNA sequencing data, we selected high confidence 100% modified m^{6}A sites for model training (see methods). To further raise accuracy, we limited our model to A and m^{6}A sites within the GGACU motif, the most prevalent m^{6}A motif in the transcriptome (L. Hu et al. 2022). By fixing the motif content, we were able to extract features not only from the A site itself (**fig. 5.2.1b**), but also from the flanking G, G, C and U sites (**fig. S5.2.1a**), which improved the model's performance. We trained multiple feedforward neural networks (FNN) and selected the most optimal one (**fig. S5.2.1b**) as the final model, which had an AUC of 0.9879 on testing set and predicted most m^{6}A and A sites correctly (**fig. 5.2.1c, d**).

To further validate the performance of m^{6}A prediction by NanoSPA, we predicted the m^{6}A sites in a published nanopore human transcriptome sample (S. Huang et al. 2021) and compared it with the corresponding m^{6}A sites revealed by m^{6}A-SAC-seq in the same cell line. We successfully recovered 88.79% (2717/3060) m^{6}A sites called by m^{6}A-SAC-seq. We revealed an additional 1433 GGACU m^{6}A sites using nanopore sequencing. Higher m^{6}A fraction revealed in m^{6}A-SAC-seq resulted accordingly in increased probability for m^{6}A calling by NanoSPA (**fig. 5.2.1e**). Furthermore, high-confidence m^{6}A sites called by m^{6}A-SAC-seq were also called with high confidence by NanoSPA (**fig. S5.2.1c**). Our predicted GGACU m^{6}A sites had a metagene profile consistent with known m^{6}A enrichment around the stop codon (**fig. 5.2.1f**). These results confirmed that our model can predict m^{6}A in the GGACU motif with high accuracy.

To investigate the relationship between m^{6}A and Ψ, our experimental design (**fig. S5.2.2a, S5.2.2b**) used siRNA knockdown by negative control siRNA (siCTRL), against core m^{6}A writer METTL3, and against the most utilized of the 13 Ψ writers in cultured cell lines, TRUB1 (Safra et al. 2017). We performed polyA-selection and ran direct nanopore RNA sequencing of biological replicates, which yielded good mapping coverages (**fig. S5.2.2c**). Applying NanoSPA on the siCTRL samples, we found that transcript groups with more Ψ had fewer m^{6}A sites (**fig. 5.2.1g**). Conversely, transcript groups with more m^{6}A had less Ψ modification (**fig. 5.2.1h**). These results suggest m^{6}A and Ψ are globally antagonistic to each other and that both modifications are less likely to co-occur on the same transcripts.

To further evaluate this m^{6}A and Ψ relationship, we analyzed the changes of m^{6}A and Ψ upon their writer knockdowns. As expected, METTL3 knockdown reduced m^{6}A in all transcript groups regardless of their Ψ status (**fig. S5.2.2d**), whereas TRUB1 knockdown

reduced Ψ modification globally (**fig. S5.2.2e**). We compared the RNA expression of siCTRL and writer knockdown samples (**fig. S5.2.2f**) and found that m^6A or Ψ writer level reduction primarily affected genes involved in metabolic processes (**fig. S5.2.2g, S5.2.2h**). The antagonistic nature of m^6A and Ψ from the siCTRL data suggested that TRUB1 knockdown should increase m^6A modification, which we indeed observed (**fig. 5.2.1i**). We also observed an appreciable increase in Ψ upon METTL3 knockdown in transcripts that contained modified GGACU m^6A sites, but not in transcripts where GGACU motifs were not modified (**fig. 5.2.1j**), consistent with the antagonistic relationship between m^6A and Ψ.

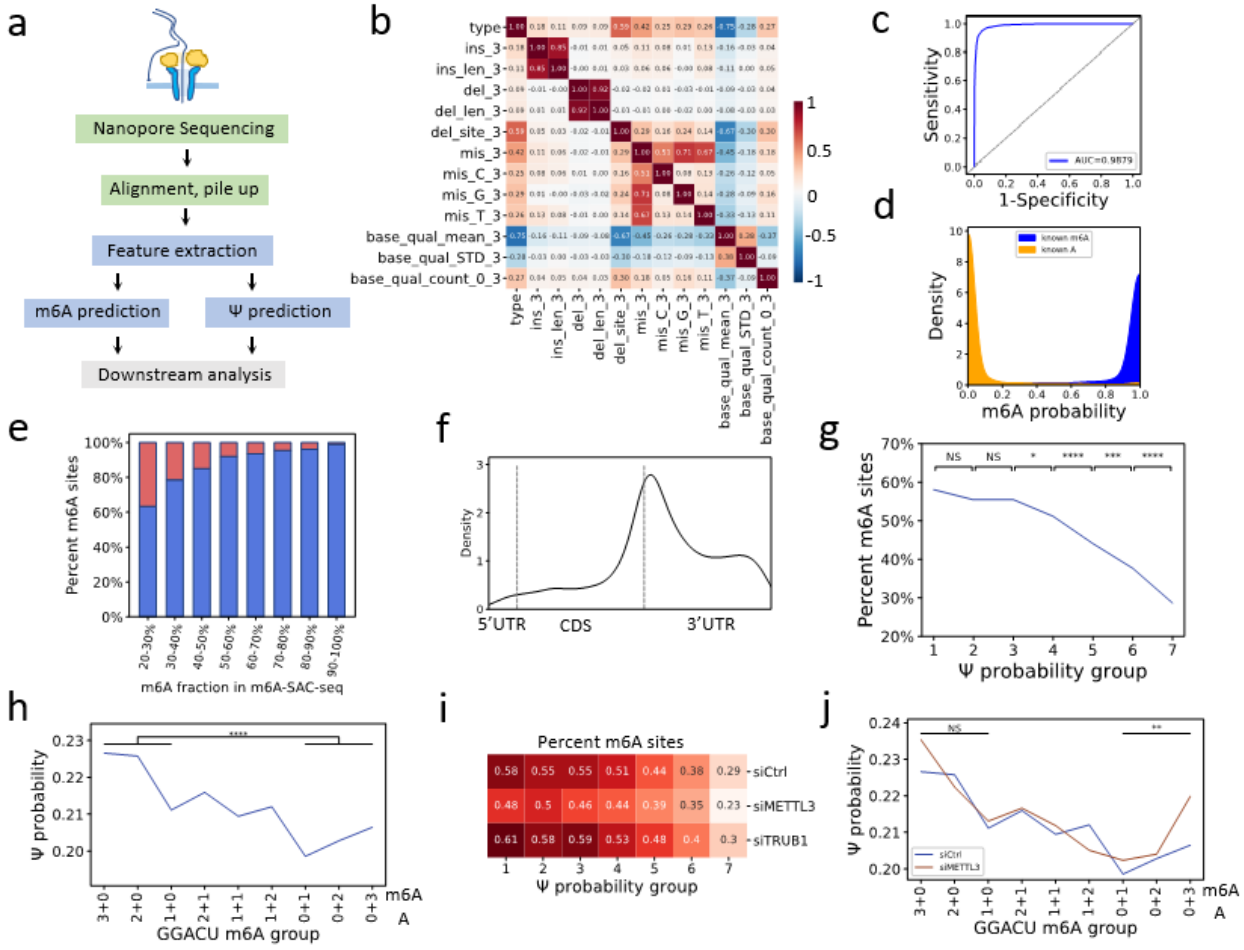


Figure 5.2. 1 NanoSPA method for simultaneous analysis of m6A and Ψ and their relationship in the transcriptome. ns: not significant, *: p<0.05, **: p<0.01, ***: p<10^{-3}, ****: p<10^{-4}. a. Workflow of the NanoSPA pipeline. b. Correlation between features of A sites in GGACU motif and modification state (m^{6}A=1, A=0). Ins, insertion rate after the base. Ins_len, insertion length mean. Del, deletion rate after the base. Del_len, deletion length mean. Del_site, deleted site ratio (the site is in a deletion). Mis, overall mismatching ratio. Mis_C, mutation to C ratio. Mis_G, mutation to G ratio. Mis_T, mutation to T ratio. Base_qual_mean, average base quality score. Base_qual_STD, base quality score standard deviation. Base_qual_count_0, ratio of bases with a quality score 0 at a site. c. ROC curve of the final FNN model on testing set. AUC is 0.9879. d. Prediction result of the annotated m^{6}A and A sites in the testing set. e. Comparing m^{6}A sites grouped by m^{6}A fraction in m^{6}A-SAC-seq data and the percentage of the m^{6}A sites revealed by NanoSPA. Blue, m^{6}A sites covered by NanoSPA; red, m^{6}A sites not covered by NanoSPA.f. Meta gene distribution of the GGACU m^{6}A sites by NanoSPA. g. Percentage of m^{6}A sites based on transcripts grouped by mean Ψ probability. The transcripts are distributed into 7 groups evenly based on mean Ψ probability, with group 1 having the lowest mean Ψ probability and group 7 the highest. h. Mean Ψ probability of transcripts based on the number of A and m^{6}A sites in GGACU motifs in the transcript. Group 3+0: transcripts with 3 or more A and 0 m^{6}A.... 2+1: 2 A and 1 m^{6}A

sites…. 0+2: 0 A and 2 m^{6}A, and so on. i. Mean m^{6}A percentage based on Ψ grouped transcripts in siCTRL, siMETTL3 and siTRUB1. j. Comparing mean Ψ probability of transcripts based on number of A and m^{6}A sites in GGACU motifs in the transcript.

5.2.2. Role of m^{6}A and Ψ in translation

To investigate the effect of m^{6}A and Ψ on translation, we performed polysome profiling without and with knockdown of the m^{6}A or Ψ writers as above (**fig. 5.2.2a, S5.2.2a, S5.2.2b**). For the siCTRL sample (**fig. S5.2.3a**), gene ontology analysis showed that transcripts with the greatest reduction of levels in the polysome belonged to genes involved in protein synthesis, such as ribosomal proteins (**fig. 5.2.2b, S5.2.3b**) and transcripts with the greatest increase of levels in the polysome belonged to cellular organelles (**fig. S5.2.3c**). The overall m^{6}A level in the polysome was higher than the input, and this increase was even higher for the transcript groups with high Ψ levels (**fig. 5.2.2c**). The overall Ψ level in the polysome was lower than the input; however, this decrease was more pronounced for transcripts in which GGACU motifs were not modified (**fig. 5.2.2d**). These results suggest that m^{6}A and Ψ were synergistic on the polysome transcripts, contrasting their antagonistic nature in the input.

The opposing m^{6}A and Ψ effects on input and polysome transcripts made the writer knockdown results difficult to predict. Even though METTL3 knockdown clearly showed a decrease in the input m^{6}A levels (**fig. S5.2.2d**), m^{6}A levels in polysome transcripts persisted at similar levels as in the siCTRL (**fig. S5.2.3d**). At the same time, Ψ levels increased in the polysome transcripts in siMETTL3 relative to siCTRL (**fig. S5.2.3e**), which could be partially derived from increased Ψ in the input mRNA. As for TRUB1 knockdown, which reduced Ψ levels in the input (**fig. S5.2.2e**), Ψ levels of the polysome transcripts

were down regardless of the m^{6}A group (**fig. 5.2.2e**), but m^{6}A level remained similar in the input and polysome transcripts (**fig. S5.2.3f**). These results suggest that polysome accumulation of transcript m^{6}A and Ψ has a hierarchical relationship, where Ψ exerts a larger effect over m^{6}A.

We further examined the relationship between m^{6}A, Ψ, and translation efficiency (TE), the ratio of polysome over input mRNA for each transcript. The transcript group more enriched in polysome GGACU m^{6}A sites also had a higher TE value (**fig. 5.2.2f**). Transcripts in all TE groups showed higher m^{6}A levels on the polysome (**fig. 5.2.2g**), indicating that m^{6}A enrichment on the polysome was indeed beneficial for translation. Transcripts with more Ψ on polysome over input had higher TE values (**fig. 5.2.2h**), indicating that Ψ on the polysome also promoted translation. Although the opposing effect of m^{6}A and Ψ in the input and polysome made the writer knockdown results not readily predictable, we expected the results to reflect the hierarchical relationship of m^{6}A and Ψ in polysome accumulation described above. TE changes upon METTL3 knockdown were not significant (**fig. S5.2.3g**) and transcripts with m^{6}A on polysome still had higher TE (**fig. 5.2.2i**), which mirrored the m^{6}A level changes on the polysome (**fig. S5.2.3d**). TE changes according to transcript Ψ probability groups upon METTL3 knockdown was like siCTRL (**fig. S5.2.3h**). Finally, TE value decreases upon TRUB1 knockdown were very significant in magnitude (**fig. 5.2.2j**) regardless of m^{6}A modification state on input or polysome (**fig. 5.2.2k**), consistent with the large Ψ level decrease (**fig. 5.2.2e**). TRUB1 knockdown also affected polysome m^{6}A based on TE, this effect switched over from negative to positive in low and high TE groups, respectively (**fig. 5.2.2l**).

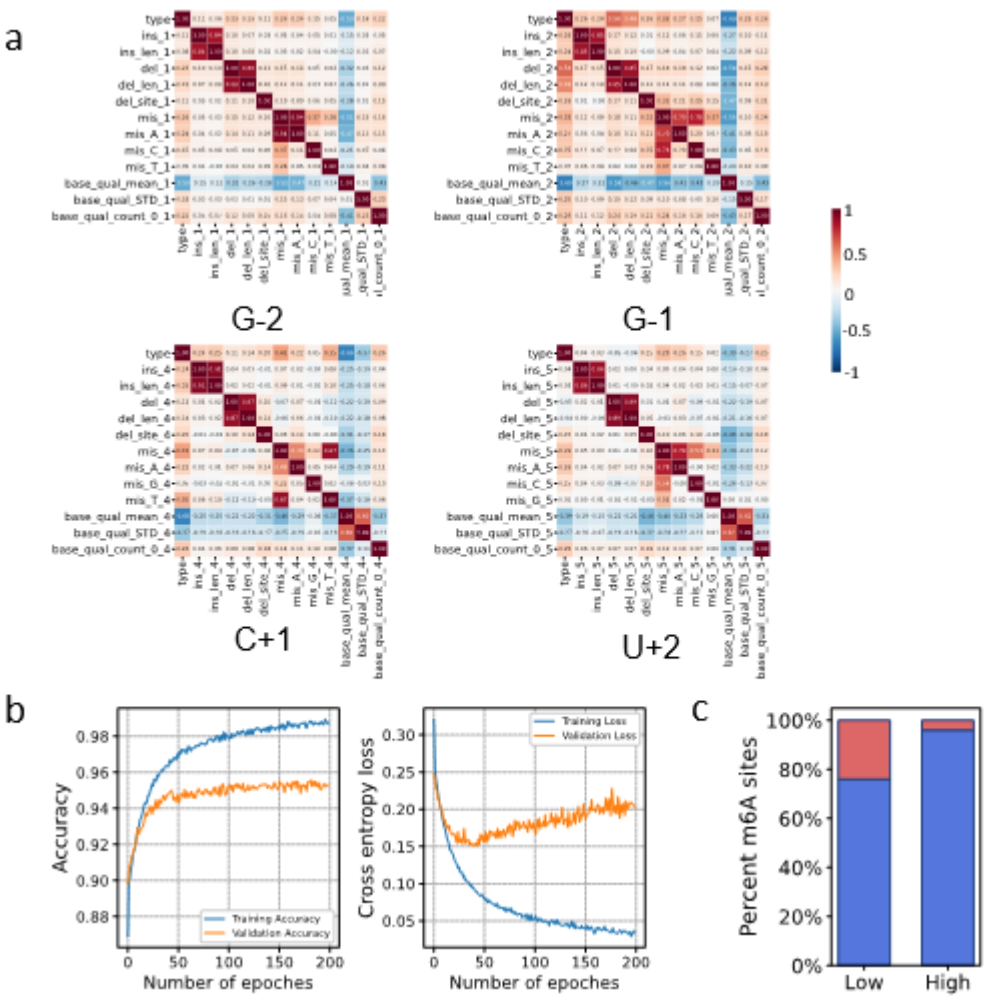


Figure S5.2. 1 Additional NanoSPA model building and m6A comparison to NGS m6A-SAC-seq. (a) Feature correlation of flanking sites in $G^{-2}G^{-1}AC^{+1}U^{+2}$ motif with modification state (1=m^{6}A, 0=A). Ins: insertion, del: deletion, mis: mismatch, base_qual: base quality. (b) Training process of the final model. Left panel, training, and validation accuracy. Right panel, training, and validation loss. The model after epoch 42 is the final model. (c) m^{6}A sites from m^{6}A-SAC-seq data are grouped by confidence (high=over 20% mutation in both replicates, low=others) and the percentage of the m^{6}A sites identified by NanoSPA. Blue, m^{6}A sites covered by NanoSPA; red, m^{6}A sites not covered by NanoSPA.

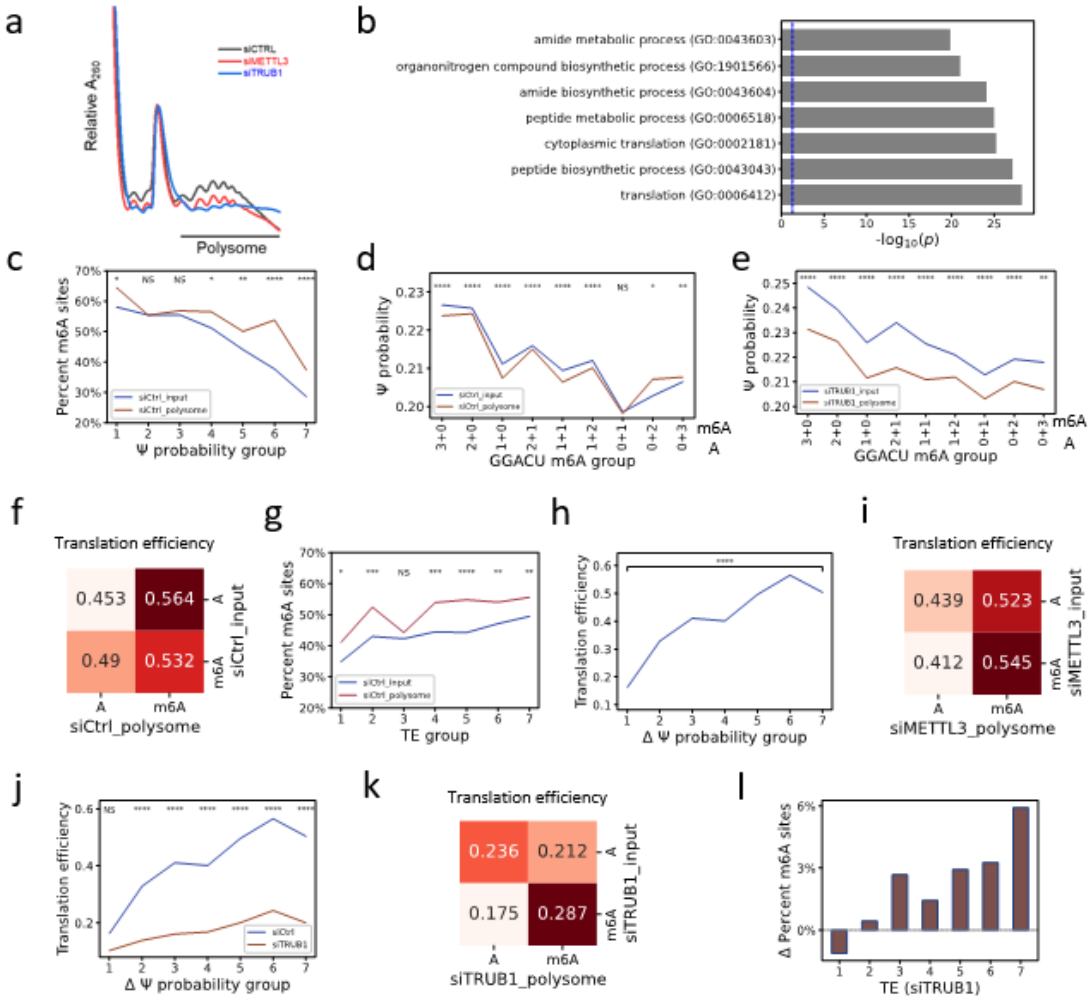


Figure 5.2. 2 Role of m^{6}A and Ψ in translation. ns: not significant, *: p<0.05, **: p<0.01, ***: p<10^{-3}, ****: p<10^{-4}. (a). Polysome profiles of siCTRL, siMETTL3, and siTRUB1 samples. Knockdown of the m^{6}A writer or a major Ψ writer resulted in global decrease in translation. The "polysome" line shows the combined fractions used for RNA extraction and nanopore sequencing. (b). Biological process gene ontology (GO) analysis of top 200 highest expressed transcripts in siCTRL input over polysome. (c). Percentage of m^{6}A sites of siCTRL input and polysome transcripts based on mean Ψ probability of transcripts. The transcripts are distributed into 7 groups evenly based on mean Ψ probability, with group 1 having the lowest mean Ψ probability and group 7 the highest. (d). Mean Ψ probability of siCTRL input and polysome transcripts based on number of GGACU A and m^{6}A sites in the transcript groups. Group 3+0: transcripts with 3 or more A and 0 m^{6}A…. 2+1: 2 A and 1 m^{6}A sites…. 0+2: 0 A and 2 m^{6}A, and so on. (e). Mean Ψ probability of siTRUB1 input and polysome transcripts based on number of GGACU A and m^{6}A sites in the transcript groups. (f). Translation efficiency (TE) of siCTRL transcripts grouped by the modification state of the GGACU motif in input or polysome. A: unmodified, m^{6}A: modified. (g). Percentage of m^{6}A sites of siCtrl input and polysome transcripts grouped by TE. The transcripts are distributed into 7 groups evenly based on TE, with group 1 having the lowest TE and group

7 the highest. (h). TE comparison of siCTRL transcripts grouped by differential mean Ψ probability. The transcripts are distributed into 7 groups evenly based on Ψ probability difference between polysome and input, with group 1 having the lowest values and group 7 the highest. (i). Translation efficiency (TE) of siMETTL3 transcripts grouped by the modification state of the GGACU motif in input or polysome. A: unmodified, m^6A: modified. (j). TE comparison of siCTRL and siTRUB1 transcripts grouped by differential mean Ψ probability. The transcripts are distributed into 7 groups evenly based on Ψ probability difference between polysome and input, with group 1 having the lowest values and group 7 the highest. (k). Translation efficiency (TE) of siTRUB1 transcripts grouped by the modification state of the GGACU motif in input or polysome. A: unmodified, m^6A: modified. (l). Differential percentage of m^6A sites on siTRUB1 input and polysome transcripts grouped by TE. The transcripts are distributed into 7 groups evenly based on TE, with group 1 having the lowest TE and group 7 the highest.

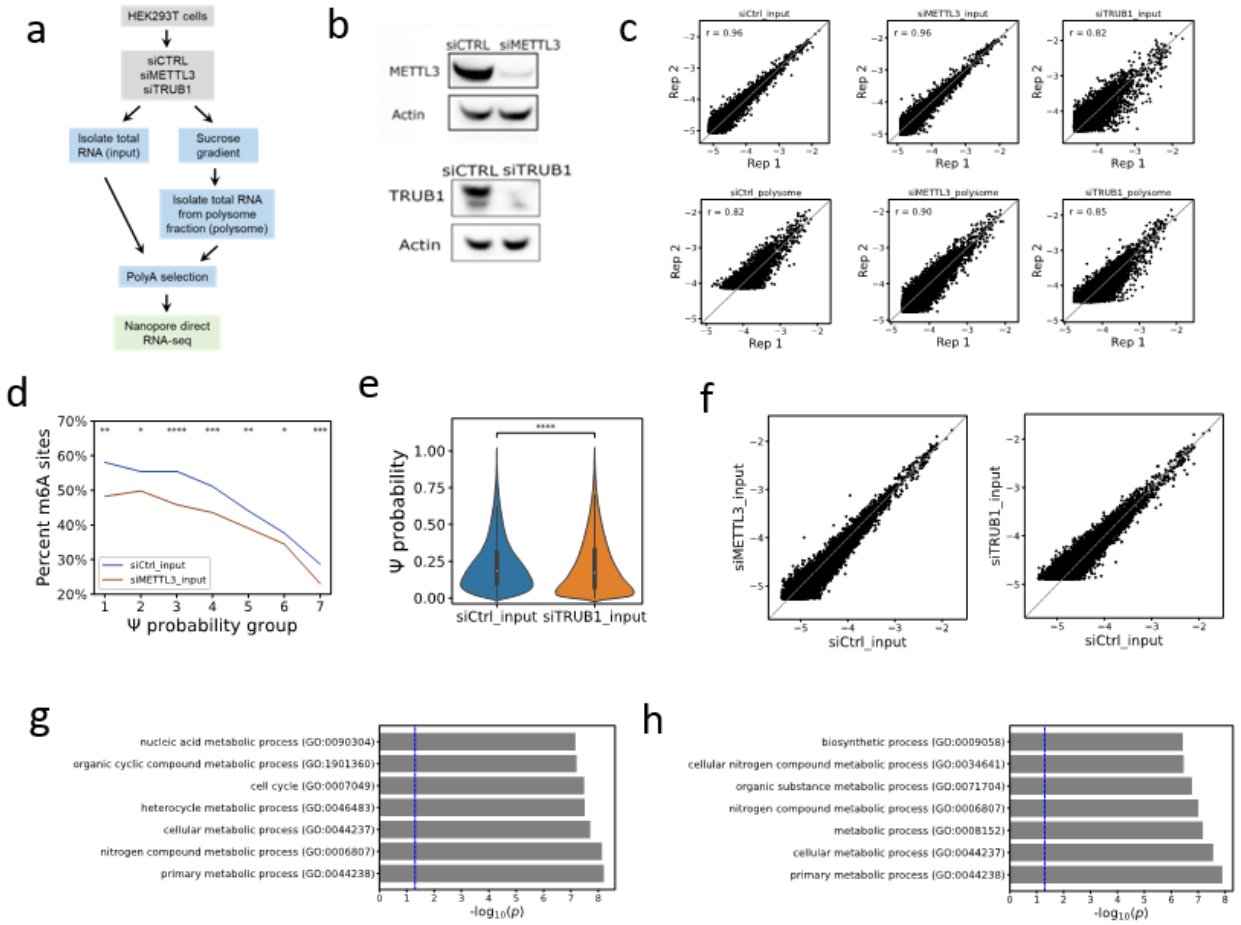


Figure S5.2. 2 Experimental results of samples and additional NanoSPA results. ns: not significant, *: p<0.05, **: p<0.01, ***: p<10^{-3}, ****: p<10^{-4}. (a) Experimental design. A total of 12 nanopore data sets were generated: siCTRL, siMETTL3, siTRUB1; input and polysome, two biological replicates. (b) Western blots showing knockdown of the METTL3 or the TRUB1 protein. Actin is the loading control. (c) RNA expression of biological replicates of each condition. (d) Comparison of GGACU m^6A site percentage of siCTRL and siMETTL3 based on transcripts grouped by mean Ψ probability as in Fig. 1g. (e) Comparison of Ψ probability distribution of siCTRL and siTRUB1. (f) Comparison of RNA expression of siCTRL versus siMETTL3 or siTRUB1. (g) Biological process gene ontology (GO) of top 200 highest expressed transcripts in siCTRL over siMETTL3. (h) Biological process GO of 200 highest expressed transcripts in siCTRL over siTRUB1.

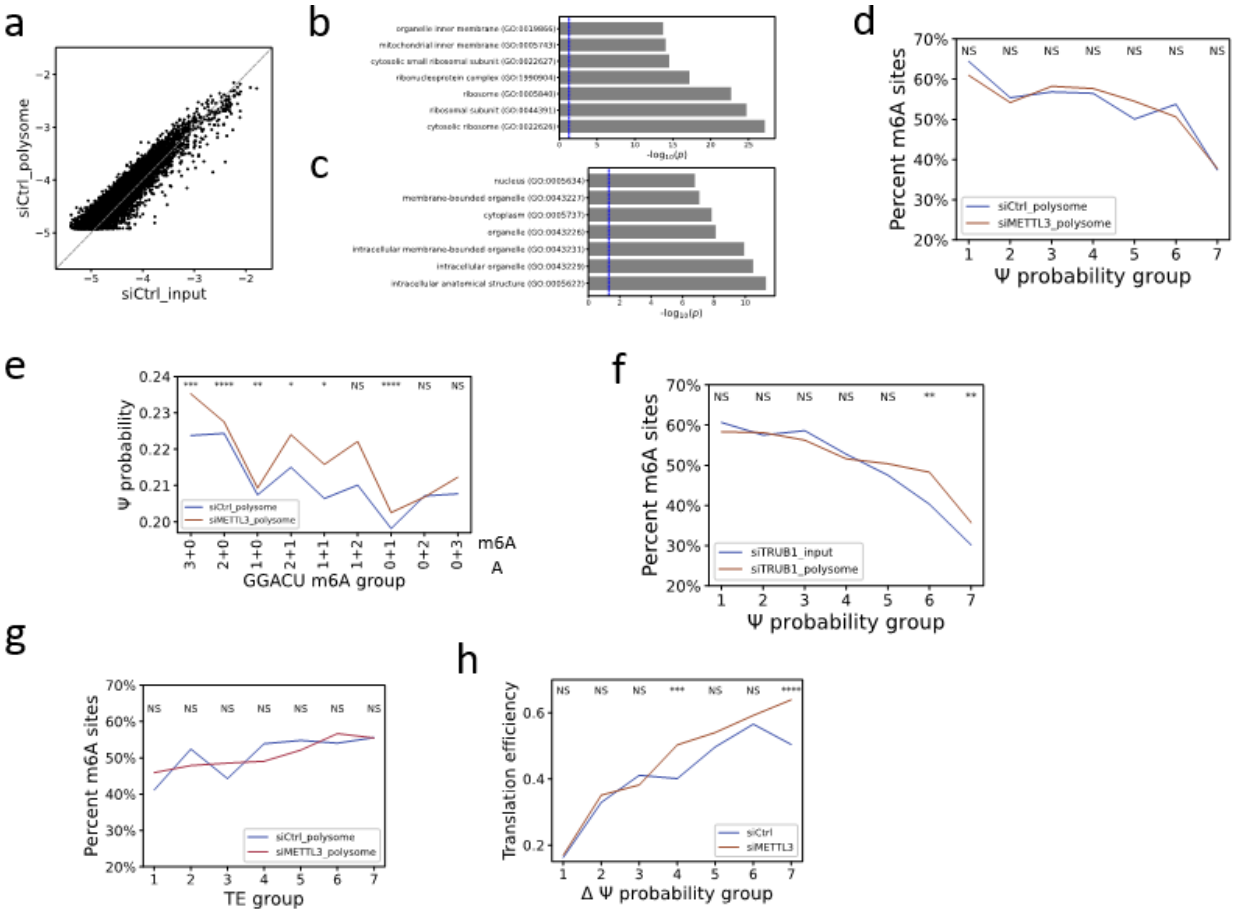


Figure S5.2. 3 Additional NanoSPA results of polysome profiling. ns: not significant, *: p<0.05, **: p<0.01, ***: p<10^{-3}, ****: p<10^{-4}. (a) Comparison of siCTRL RNA expression between input and polysome. (b) Cellular Component GO of top 200 highest expressed transcripts in siCTRL input over polysome. (c) Cellular Component GO of top 200 highest expressed transcripts in siCTRL polysome over input. (d) Percentage of m^{6}A sites of siMETTL3 input and polysome based on transcripts grouped by mean Ψ probability as in Fig. 1g. (e) Mean Ψ probability of siMETTL3 input and polysome transcripts based on the number of A and m^{6}A site groups as in Fig. 1h. (f) Percentage of m^{6}A sites of siTRUB1 input and polysome transcripts based on mean Ψ probability groups. (g) Percentage of m^{6}A sites on siCTRL and siMETTL3 polysome transcripts grouped by TE as in Fig. 2g. (h) TE comparison of siCTRL and siMETTL3 transcripts grouped by differential mean Ψ probability.

5.3 Conclusions

In summary, we developed a machine learning pipeline for nanopore sequencing to analyze m^6A and Ψ simultaneously. We identified an antagonistic effect of m^6A and Ψ in total mRNA input, but a synergistic effect of m^6A and Ψ for polysome-associated mRNA. Furthermore, m^6A and Ψ have a hierarchical effect on promoting translation efficiency in which Ψ takes precedent over m^6A.

5.4 Methods

5.4.1 Cell culture and siRNA knockdown

Human embryonic kidney (HEK) HEK293T/17 cells (CRL11268) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) with high glucose and L-glutamine, without sodium pyruvate (HyClone, SH30022.01) and with 10% FBS in a 37°C incubator at 5% CO_2 to seed for reverse transfection by RNAiMax (Sigma 13778150). For each knockdown condition, 75 µL Lipofectamine RNAiMax was added to three separate 150 mm plates containing 300 pmol siRNA (siCTRL, MISSION® siRNA Universal Negative Control #1; siMETTL3, SASI_HS_00044317; siTRUB1, SASI_Hs02_0036419) in 5 mL Opti-MEM™ I Reduced Serum Medium (31985070) and incubated at 37°C with 5% CO_2 for 20 minutes. HEK293T cells grown to 80% confluency were washed and detached in 1x Phosphate Buffer Saline (PBS) before pelleting by centrifugation at 500 x g for 3 minutes. The cell pellet was resuspended in media and cells were counted using an Invitrogen™ Countess™ 3 FL Automated Cell Counter, for which a 10 µL aliquot of cells was mixed with 10 µL trypan blue and loaded into a chamber slide. For control and METTL3 knockdowns, ~3.5 x 10^6

cells were seeded into three 150 mm plates containing the appropriate siRNA-lipofectamine mixture. As these conditions were not viable for TRUB1 knockdown cells, ~5 x 10^6 cells were seeded instead. The plates were mixed by gently rocking and incubated in a 37°C with 5% CO_2 for 72 hours.

5.4.2 Polysome profiling

Polysome profiling procedures were adapted from a previous publication (Xiao Wang et al. 2014). For each knockdown condition, three 150 mm plates of ~80% confluent HEK293T cells were treated with 100 µg/mL cycloheximide (CHX AC3574200500, Fisher Scientific) for 7 minutes at 37°C. Media was removed from the plates and the cells were washed twice with 10 mL of ice-cold 1x PBS containing 100 µg/mL CHX prior to being scraped and collected in 5 mL 1x ice-cold PBS. Cells were pelleted by centrifugation at 500 x g for 5 minutes, then the three plates for each knockdown conditions were combined in 0.8 mL Lysis Buffer (20 mM HEPES, pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, 1% Triton X-100, 100 µg/mL CHX supplemented with fresh 1x Roche protease inhibitor and 1% Superase inhibitor) and lysed by rotating at 4°C for 20 minutes. Cell debris was pelleted by 15-minute centrifugation at 16,000 x g. To this lysate, 4 µL T4 Turbo DNAse (Invitrogen, AM2238) was added, and the mixture was incubated at room temperature for 15 minutes. 180 µL of this lysate was saved as "Input" for RNA downstream nanopore sequencing, and 20 µL was saved for western blot (see below).

5-50% sucrose gradient (20 mM HEPES, pH 7.6, 100mM KCl, 5 mM $MgCl_2$, 100 µg/mL CHX supplemented with fresh 1x Roche protease inhibitor and 1% SUPERase Inhibitor) was prepared in SETON 7042 tubes using a Biocomp Gradient Station. 600 µL

sucrose gradient was removed from the top of each balanced sucrose gradient tube and then replaced by gently pipetting 600 µL of the respective knockdown cell lysate on top the gradient. Sucrose gradients were centrifuged for 3 hours at 28,000 RPM in an Optima L-100XP centrifuge using a Beckman SW28 rotor. Sucrose gradient fractions were collected and absorbances continuously measured using a Biocomp gradient station.

For each knockdown replicate, the 30 generated fractions were split in half and 0.9 mL TRIzol™ Reagent was added to each tube. Samples were incubated for 5 minutes at room temperature prior to addition of 0.18 mL chloroform and an additional 3-minute incubation. The aqueous layer was separated by centrifugation for 15 minutes at 12,000 × g and 4°C and then added to a new tube. The sample was frozen overnight at -80°C following addition of 0.45 mL isopropanol. RNA was precipitated by centrifugation for 15-minute at 12,000 × g and 4°C. The supernatant was removed, and the RNA pellet resuspended and washed in 0.9 mL 75% ethanol before being centrifuged for 5 minutes at 7,500 x g and 4°C. The supernatant was removed, and the pellet air-dried for ten minutes prior to resuspension in 10 µL dH_2O. The two tubes for each fraction were combined prior to combining all fractions disome and after.

Input and polysome RNA samples were made to 150 µL in dH_2O and cleaned by adding 280 µL Beckman RNAClean XP beads. Samples were mixed by pipetting up and down, incubated at room temperature for 5 minutes, pelleted on a magnetic rack for 5 minutes, washed with 1000 µL 70% EtOH three times, and air-dried for ten minutes before eluting into 150 µL dH_2O. Poly(A)-selection of cleaned RNA samples was then done using Promega PolyATract mRNA Isolation System IV per the manufacturer's protocol. Briefly, RNA samples were made to 500 µL and incubated at 65°C for 10 minutes before addition of

3 µL Biotin-dT and 1x SSC. Once cooled, this sample was added to 100 µL of washed poly(A) magnetic beads in 0.5x SSC. Following a 10-minute incubation at room temperature, RNA-bound beads were captured using a magnetic rack, washed in 0.1X SSC, and eluted in a total of 250 µL dH_2O.

PolyA+ RNA was concentrated using Zymo Oligo Clean & Concentrator columns per manufacturer's protocol prior to Nanopore sequencing. Briefly, 500 µL Oligo Binding Buffer and 2 mL absolute ethanol were added to 250 µL PolyA+ RNA and mixed by pipetting. The sample was transferred to a Zymo-Spin™ IC Column in a collection tube and centrifuged. The sample was washed using 750 µL DNA Wash buffer prior to elution in 11 µL nuclease-free dH_2O. For quality control, polyA+ RNA was submitted to the University of Chicago Genomics facility and analyzed using an Agilent 2100 Bioanalyzer system. 5 µL of RNA containing ~1 ng/µL was analyzed using the RNA Pico/High Sensitivity Assay (input sensitivity of 0.05-5 ng/µL) to confirm RNA integrity.

5.4.3 WT cell sample culture

HeLa cells (ATCC) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) with high glucose and L-glutamine, without sodium pyruvate (HyClone, SH30022.01) with 10% FBS and 1% Pen–Strep (Penicillin–Streptomycin). ~3.5 x 10^6 cells were seeded into three 150 mm plates with the same media without 1% Pen–Strep (Penicillin–Streptomycin). The following RNA extraction and purification steps are the same as above.

5.4.4 Western blot

Samples were prepared by adding 1x LDS and 100 mM DTT before boiling at 95°C for 5 minutes. Samples were loaded onto 12-well 4–12% polyacrylamide Bis-Tris gels (NP03322, Invitrogen) and ran at 150V for 1 hour. The gels were then transferred to polyvinylidene fluoride membranes (IPVH00010, Millipore). The membranes were blocked overnight in 10% w/v milk (1706404, Bio-Rad). The blots were probed with 1/1000 v/v anti-actin (clone C4 MAB1501), 1/1000 v/v anti-METTL3 (ab195352), or 1/500 v/v anti-TRUB1 (1250-1-AP) in 5% w/v milk (1706404, Bio-Rad) followed by 1/10000 v/v sheep anti-mouse IgG (NA931V, Cytiva) or 1/10000 v/v donkey anti-rabbit IgG conjugated to horseradish peroxidase (NA934V, Cytiva) in 5% w/v milk (1706404, Bio-Rad). The blots were then visualized with ECL Prime Western Blotting Detection Reagents (RPN2232, Amersham) using a BioRad ChemiDoc MP.

5.4.5 Nanopore direct RNA sequencing

The library preparation of direct RNA seq samples followed instructions from Oxford Nanopore Technology for Direct RNA Sequencing Kit (SQK-RNA002). Concisely, 500 ng of Poly(A)$^+$ RNA sample was used to perform a run. The RT Adaptor (RTA) was ligated to the 3' end of Poly(A)$^+$ RNA using T4 DNA ligase (NEB M0202S), followed by reverse transcription by SuperScript III Reverse Transcriptase (ThermoFisher 12574018). The RT product was then purified by 1.8x RNAClean XP beads (72 µL) (Beckman Coulter A63987). A second RNA Adaptor (RMX) was then attached to the 3' end of Poly(A)$^+$ RNA by T4 DNA ligase (NEB M0202S). The RNA product was purified with 1x RNAClean XP beads (40 µL)

and eluted with 21 µl Elution Buffer. The sample was loaded onto a R9.4.1 flow cell (FLO-MIN106D) and then run on the MinION sequencer for 72 hours.

5.4.6 Nanopore data pre-processing

Raw sequencing data files were uploaded to UChicago midway2 cluster for pre-processing. Base calling was performed by guppy base caller (version 3.2.2+9fe0a78). Then, the reads were aligned to human genome (GRCh38.p13) by minimap2(Heng Li 2018) (version 2.18-r1015) with parameters -ax splice -uf -k14. The mapped reads were piled up by samtools(Heng Li et al. 2009)(v1.11) and features for modifications prediction were extracted by customized python scripts (https://github.com/sihaohuanguc/NanoSPA).

5.4.7 Model training for m^{6}A prediction

The wild-type HeLa cell nanopore seq sample was used to train the model for m^{6}A prediction. The data was pre-processed as described above. The annotation of 599 100% modified high confidence m^{6}A and 636 random unmodified A sites in GGACU motif was achieved from the m^{6}A-SAC-seq study(L. Hu et al. 2022). High confidence was defined as >20% mutation rate in both replicates. Data augmentation was performed by shuffling the reads of a site and sampling 16 random reads as a group with 20 groups for each site. After data augmentation, we achieved 11980 m^{6}A and 12720 A sites for model training. Then, 55 features (see fig. 1b, S1a) for the centered, -2, -1, +1, and +2 sites were collected for all generated sites. The dataset was then split into 60% training, 20% validation and 20% testing sets. Feedforward neural network (FNN) and Recurrent neural network (RNN)

models were trained on training set and evaluated by validation loss (cross entropy). The final model with the best performance was an FNN with two hidden layers (128 and 64 nodes, drop rate 0.1 and 0.2 after each layer respectively), with learning rate 0.001. The model with the lowest validation loss was stored (epoch 42) when the validation loss was 0.1503 and the validation accuracy was 0.9493. The final model performance was evaluated on the testing set by AUC (area under curve) of ROC (Receiver Operating Characteristic).

5.4.8 HEK293T cell data processing

The HEK293T cell samples were sequenced and pre-processed as described above. The mapped reads of two biological replicates were combined to increase the data analysis throughput. We obtained 1.1-3.6 million mapped reads for the input, and 1.2-1.6 million mapped reads for the polysome samples. For Ψ prediction, features of all U sites with >20 coverage were collected and performed by NanoPSU. For m^6A prediction, features of all A sites within GGACU motif and with >20 coverage in all 5 sites were collected and then performed prediction by the model generated above. At the same time, the expression counts of genes were calculated as the maximum peak height of the reads piled at the gene regions. Relative expression level of a gene was calculated as the expression count of a gene divided by the sum of expression counts of the sample. Genes with less than 15 coverage were filtered. Translation efficiency (TE) of a gene was calculated as the expression level in polysome sample divided by expression level in the input sample. The gene information was provided by the comprehensive gene annotation file (gencode. v41.annotation. gff3) in the GENCODE database (https://www.gencodegenes.org). The gene ontology (GO) analysis

was performed using the Gene Ontology Resource (http://geneontology.org) (Ashburner et al. 2000; “The Gene Ontology Resource: Enriching a GOld Mine.” 2021)

Chapter 6

Future directions and perspectives

6.1 Coordination of m^{6}A-RNA and RNAPII CTD by the YTH-Domain

Chapter 3 of this thesis presents data showing the YTH-domain of YTHDC1 interacting directly with the CTD of RNA Polymerase II in addition to its previously described binding to m^{6}A-modified RNA. This observation was supported by CTD co-immunoprecipitation both with endogenous full-length YTHDC1 (**Figure 2.2.1**), and transfected FLAG-tagged YTH-domain (**Figure 2.2.2**), in addition to co-precipitation during *in vitro* His-pulldown using purified 6x-His-YTH-domain (**Figure 2.2.3**). This data supports a competitive model for YTH-CTD interaction, where the addition of m^{6}A-RNA causes dissociation between the YTH-domain and the RNAPII CTD.

To characterize the mechanism and function of the YTH-domain's RNAPII binding interactions, its polymerase-binding activity needs to be abolished by mutagenesis of the YTH-domain without significantly changing its m^{6}A-binding ability. Crystal structures of both m^{6}A-bound and unbound YTH domain from YTHDC1 have revealed its mechanism of m^{6}A binding (Xu et al. 2014) through two conserved tryptophan residues, W377 and W428, along with L439 that create an aromatic cage that recognizes the m^{6}A moiety, analogous to the recognition of histone modifications by aromatic cages in chromodomains. Alanine substitution at either conserved tryptophan abolishes binding to m^{6}A (Jacobs and Khorasanizadeh 2002). The aromatic cage can also recognize adenosine, but the methyl group of m^{6}A is thought to contribute van der Waal interactions worth ~1.7 kcal/mol,

representing a ~16-fold stronger interaction than the unmodified base. Molecular dynamics have shown that this energetic contribution is driven both through stabilizing the aromatic cage through preferential interactions and through stabilizing a bound-like conformation of the GG(m^6A)CU oligo itself (Y. Li et al. 2021). It is possible that competition between the CTD of RNA Polymerase II and m^6A-RNA may be due to energetically favorable interactions with m^6A relative to the RNAPII CTD, analogous to the preferential energetics of YTH's interaction with m^6A over A. In this case, W377A and W428A mutants should also abolish CTD binding, although such binding mutants would not be able to disentangle the functional significance of m^6A binding from the significance of RNAPII binding.

Further crystal structures of unbound YTH have revealed conformational changes within the YTH-domain binding pocket. Side-chain flipping was observed at Met438; in the flipped conformation, Met438 goes from surface-facing to facing the binding pocket. In the canonical outwards-facing structure, molecular dynamic simulations showed rotation of Trp428, which then obstructs the aromatic cage. In the structure where Met438 faces the binding pocket, this obstruction is recovered by the entry of Leu439 to the binding site (Y. Li et al. 2021). Sidechain flipping of Met438 impacting the stability of rotating Trp428 implies mutation to Alanine at this site could lock the conformation, which may change binding affinity for the CTD of RNA Polymerase II. Furthermore, restoration of the aromatic cage by Leu439 may be important in favoring m^6A as a binding entity relative to the CTD.

Structural changes occur outside of the binding pocket when comparing YTH domain holo (with m^6A RNA bound) and apo (no RNA bound) structures. Notably, a connecting loop adopts a different conformation between these structures which may be

responsible for mediating interaction with the CTD, and thus mutating this loop to a stretch of alanines may abolish CTD binding, but not m^{6}A binding (**Figure 6.1**).

a.

	Structural Significance	Effect on M6A Binding
Trp377Ala	Aromatic cage	Decreases
Trp428Ala	Aromatic cage	Decreases
Met438Ala	Flips into binding pocket	entropy-enthalpy compensation
Ser378Ala	Stability of aromatic cage	entropy-enthalpy compensation
Thr379Val	Water-mediated m6A interaction	Decreases
424-446Ala	Loop; conformation changes on m6A binding	Unknown

b.

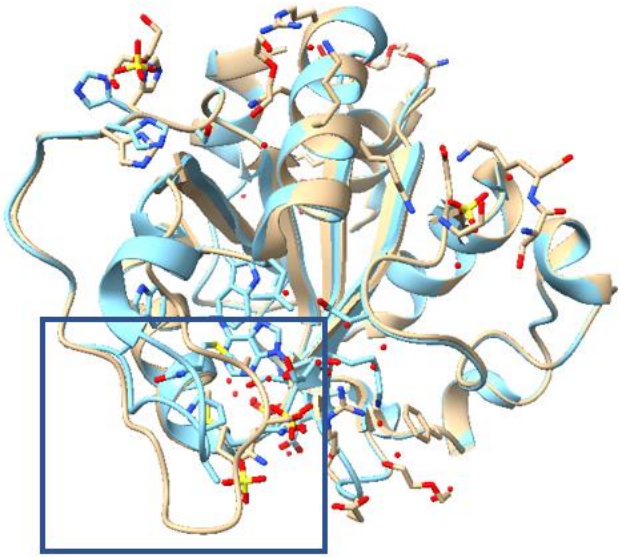


Figure 6. 1 Mutagenesis scheme to show coordination of m6A-RNA and RNAPII CTD by the YTH-Domain. (a). Table showing proposed mutants, their structural significance, and reported effect on m^{6}A-binding energetics on mutation (Y. Li et al. 2021). (b). Overlay of m6A-bound (blue) and unbound (brown) YTHDC1 YTH-domain structures (PDB: 6zcn, 6zd9). Conformational change induced on ligand binding is boxed.

6.2 Coordination of m^{6}A-RNA and tRNA properties

The bioinformatic survey employing MSR-Seq data presented in Chapter 4 of this thesis supports a connection between m^{6}A in mRNA and tRNA expression and modification, in particular the methyl-3-cytosine (m^{3}C) at position 47d(e2). Further study

of this connection is challenged by the currently unknown writer of m^3C47d (Bohnsack 2022). If such a writer were to be discovered, analyzing the knockdown of this m^3C47d writer could provide essential data connecting m^6A in mRNA to m^3C47d in tRNA. Analysis of hnRNPG-PAR-CLIP data suggested that this m^6A reader protein interacted with tRNAs. PAR-CLIP of other m^6A reader proteins like YTHDC1 could expand on potential connections between mRNA and tRNA. Coupling MSR-seq to reader-protein cross-linking could also generate more specific information on the nature of these interactions, including enrichment of modifications and the effect of tRNA charging.

Additionally, Chapter 4 demonstrates the co-immunoprecipitation of RNA Polymerase III subunit POL3RA with m^6A-reader protein YTHDC1. While these preliminary results offer an exciting potential mechanism for eukaryotic polymerase crosstalk, the specifics of this interaction first needs to be investigated further. Immunoprecipitation of multiple endogenous reader proteins should be performed to see if this result broadly applies to m^6A readers. Given the complex assembly and multi-subunit nature of RNA Polymerase III, CoIP eluate should be blotted against a variety of RNAPIII subunits in the presence and absence of RNase treatment. As in Chapter 6.1, a YTHDC1 mutagenic screen that aims to identify binding mutants would allow further investigation into the importance of these interactions and crosstalk to RNAPII.

6.3. Coordination of m^6A and pseudouridine

In Chapter 5 of this thesis, I presented data on the crosstalk between m^6A and Ψ in total RNA and polysome transcripts which demonstrates an antagonistic effect of m^6A and Ψ co-occurrence and a synergistic, hierarchical effect of m^6A and Ψ on translation. Our

NanoSPA platform enables simultaneous detection of m^{6}A and Ψ within a single transcript, and thus allows investigation of crosstalk between these two modifications. Our biological investigation focused primarily on writer knockdowns: core m^{6}A MTC component METTL3, and the primary Ψ writer on mRNA in cultured cell lines, TRUB1 (Safra et al. 2017). Whereas METTL3 is a necessary component of the m^{6}A MTC, TRUB1 is one of 13 Ψ writers (X. Li, Ma, and Yi 2016). In addition to TRUB1, PUS1, PUS7, and RPUSD4 have been shown to act on mRNA transcripts, with PUS1, PUS7, and RPUSD4 having demonstrated involvement in alternative pre-mRNA splicing and polyadenylation (Martinez et al. 2022). Given proper gene perturbation, our platform enables testing the crosstalk of Ψ writers on m^{6}A abundance and concurrence. Furthermore, our study's biological significance focused on Ψ's connection to translation; to our knowledge, this is the first time nanopore RNA sequencing has been used for polysome profiling. By expanding data analysis to include alternative splicing, specifically the impact of the writers already implicated, NanoSPA can be adapted to demonstrate potential crosstalk of Ψ and m^{6}A modifications in splicing. Given this thesis covers two m^{6}A reader proteins with known involvement in pre-mRNA splicing, YTHDC1, and hnRNPG, it may also be of interest to alter their expression to study their potential involvement in mediating modification crosstalk.

References

Alarcón, Claudio R, Hani Goodarzi, Hyeseung Lee, Xuhang Liu, Saeed Tavazoie, and Sohail F Tavazoie. 2015. "HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events." *Cell* 162 (6): 1299–1308. https://doi.org/10.1016/j.cell.2015.08.011.

Anderson, Bart R, Hiromi Muramatsu, Subba R Nallagatla, Philip C Bevilacqua, Lauren H Sansing, Drew Weissman, and Katalin Karikó. 2010. "Incorporation of Pseudouridine into MRNA Enhances Translation by Diminishing PKR Activation." *Nucleic Acids Research* 38 (17): 5884–92. https://doi.org/10.1093/nar/gkq347.

Arango, Daniel, David Sturgill, Najwa Alhusaini, Allissa A Dillman, Thomas J Sweet, Gavin Hanson, Masaki Hosogane, et al. 2018. "Acetylation of Cytidine in MRNA Promotes Translation Efficiency." *Cell* 175 (7): 1872-1886.e24. https://doi.org/10.1016/j.cell.2018.10.030.

Ashburner, Michael, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25 (1): 25–29. https://doi.org/10.1038/75556.

Ashraf, Saira, Lin Huang, and David M J Lilley. 2019. "Effect of Methylation of Adenine N6 on Kink Turn Structure Depends on Location." *RNA Biology* 16 (10): 1377–85. https://doi.org/10.1080/15476286.2019.1630797.

Au, Kin Fai, Vittorio Sebastiano, Pegah Tootoonchi Afshar, Jens Durruthy Durruthy, Lawrence Lee, Brian A Williams, Harm van Bakel, et al. 2013. "Characterization of the Human ESC Transcriptome by Hybrid Sequencing." *Proceedings of the National Academy of Sciences of the United States of America* 110 (50): E4821-30. https://doi.org/10.1073/pnas.1320101110.

Bai, Yang, Chunxing Yang, Runliu Wu, Lihua Huang, Shenlei Song, Wanwan Li, Peichen Yan, Changwei Lin, Daojiang Li, and Yi Zhang. 2019. "YTHDF1 Regulates Tumorigenicity and Cancer Stem cell-like Activity in Human Colorectal Carcinoma." *Frontiers in Oncology* 9: 332. https://doi.org/10.3389/fonc.2019.00332.

Barbieri, Isaia, Konstantinos Tzelepis, Luca Pandolfini, Junwei Shi, Gonzalo Millán-, Samuel C Robson, Demetrios Aspris, et al. 2018. "Europe PMC Funders Group Promoter-Bound METTL3 Maintains Myeloid Leukaemia via M6A- Dependent Translation Control" 552 (7683): 126–31. https://doi.org/10.1038/nature24678.Promoter-bound.

Bartosovic, Marek, Helena Covelo Molares, Pavlina Gregorova, Dominika Hrossova, Grzegorz Kudla, and Stepanka Vanacova. 2017. "N6-Methyladenosine Demethylase FTO Targets Pre-MRNAs and Regulates Alternative Splicing and 3'-End Processing." *Nucleic Acids Research* 45 (19): 11356–70. https://doi.org/10.1093/nar/gkx778.

Bell, Jessica L, Kristin Wächter, Britta Mühleck, Nikolaos Pazaitis, Marcel Köhn, Marcell Lederer, and Stefan Hüttelmaier. 2013. "Insulin-like Growth Factor 2 MRNA-Binding Proteins (IGF2BPs): Post-Transcriptional Drivers of Cancer Progression?" *Cellular and Molecular Life Sciences : CMLS* 70 (15): 2657–75. https://doi.org/10.1007/s00018-012-1186-z.

Bentley, David L. 2014. "Coupling MRNA Processing with Transcription in Time and Space." *Nature Reviews Genetics* 15 (3): 163–75. https://doi.org/10.1038/nrg3662.

Bohnsack, Katherine E, Nicole Kleiber, Nicolas Lemus-Diaz, and Markus T Bohnsack. 2022. "Roles and Dynamics of 3-Methylcytidine in Cellular RNAs." *Trends in Biochemical Sciences* 47 (7):

596–608. https://doi.org/10.1016/j.tibs.2022.03.004.

Bokar, J A, M E Shambaugh, D Polayes, A G Matera, and F M Rottman. 1997. "Purification and CDNA Cloning of the AdoMet-Binding Subunit of the Human MRNA (N6-Adenosine)-Methyltransferase." *RNA (New York, N.Y.)* 3 (11): 1233–47.

Boulias, Konstantinos, and Eric Lieberman Greer. 2023. "Biological Roles of Adenine Methylation in RNA." *Nature Reviews Genetics* 24 (3): 143–60. https://doi.org/10.1038/s41576-022-00534-0.

Broly, Martin, Bogdan V Polevoda, Kamel M Awayda, Ning Tong, Jenna Lentini, Thomas Besnard, Wallid Deb, et al. 2022. "THUMPD1 Bi-Allelic Variants Cause Loss of TRNA Acetylation and a Syndromic Neurodevelopmental Disorder." *The American Journal of Human Genetics* 109 (4): 587–600. https://doi.org/https://doi.org/10.1016/j.ajhg.2022.02.001.

Brugiolo, Mattia, Valentina Botti, Na Liu, Michaela Müller-McNicoll, and Karla M Neugebauer. 2017. "Fractionation ICLIP Detects Persistent SR Protein Binding to Conserved, Retained Introns in Chromatin, Nucleoplasm and Cytoplasm." *Nucleic Acids Research* 45 (18): 10452–65. https://doi.org/10.1093/nar/gkx671.

Bruni, Francesco, Robert N Lightowlers, and Zofia M Chrzanowska-Lightowlers. 2017. "Human Mitochondrial Nucleases." *The FEBS Journal* 284 (12): 1767–77. https://doi.org/10.1111/febs.13981.

Burke, Kathleen A., Abigail M. Janke, Christy L. Rhine, and Nicolas L. Fawzi. 2015. "Residue-by-Residue View of In Vitro FUS Granules That Bind the C-Terminal Domain of RNA Polymerase II." *Molecular Cell* 60 (2): 231–41. https://doi.org/10.1016/j.molcel.2015.09.006.

Carlile, Thomas M, Maria F Rojas-Duran, Boris Zinshteyn, Hakyung Shin, Kristen M Bartoli, and Wendy V Gilbert. 2014. "Pseudouridine Profiling Reveals Regulated MRNA Pseudouridylation in Yeast and Human Cells." *Nature* 515 (7525): 143–46. https://doi.org/10.1038/nature13802.

Cerneckis, Jonas, Qi Cui, Chuan He, Chengqi Yi, and Yanhong Shi. 2022. "Decoding Pseudouridine: An Emerging Target for Therapeutic Development." *Trends in Pharmacological Sciences* 43 (6): 522–35. https://doi.org/10.1016/j.tips.2022.03.008.

Chaisson, Mark J P, Ashley D Sanders, Xuefang Zhao, Ankit Malhotra, David Porubsky, Tobias Rausch, Eugene J Gardner, et al. 2019. "Multi-Platform Discovery of Haplotype-Resolved Structural Variation in Human Genomes." *Nature Communications* 10 (1): 1784. https://doi.org/10.1038/s41467-018-08148-z.

Chen, Chuan, Wenqiang Liu, Jiayin Guo, Yuanyuan Liu, Xuelian Liu, Jun Liu, Xiaoyang Dou, et al. 2021. "Nuclear m(6)A Reader YTHDC1 Regulates the Scaffold Function of LINE1 RNA in Mouse ESCs and Early Embryos." *Protein & Cell* 12 (6): 455–74. https://doi.org/10.1007/s13238-021-00837-8.

Chen, Hong-Xuan, Zhang Zhang, Dong-Zhao Ma, Li-Qian Chen, and Guan-Zheng Luo. 2022. "Mapping Single-Nucleotide m(6)A by m(6)A-REF-Seq." *Methods (San Diego, Calif.)* 203 (July): 392–98. https://doi.org/10.1016/j.ymeth.2021.06.013.

Chen, Kai, Zhike Lu, Xiao Wang, Ye Fu, Guan-Zheng Luo, Nian Liu, Dali Han, et al. 2015. "High-Resolution N(6) -Methyladenosine (m(6) A) Map Using Photo-Crosslinking-Assisted m(6) A Sequencing." *Angewandte Chemie (International Ed. in English)* 54 (5): 1587–90. https://doi.org/10.1002/anie.201410647.

Church, Chris, Sheena Lee, Eleanor A L Bagg, James S McTaggart, Robert Deacon, Thomas Gerken, Angela Lee, et al. 2009. "A Mouse Model for the Metabolic Effects of the Human Fat Mass and Obesity Associated FTO Gene." *PLoS Genetics* 5 (8): e1000599. https://doi.org/10.1371/journal.pgen.1000599.

Church, Chris, Lee Moir, Fiona McMurray, Christophe Girard, Gareth T Banks, Lydia Teboul, Sara Wells, et al. 2010. "Overexpression of Fto Leads to Increased Food Intake and Results in Obesity." *Nature Genetics* 42 (12): 1086–92. https://doi.org/10.1038/ng.713.

COHN, WALDO E, and ELLIOT VOLKIN. 1951. "Nucleoside-5′-Phosphates from Ribonucleic Acid." *Nature* 167 (4247): 483–84. https://doi.org/10.1038/167483a0.

Coster, Wouter De, Matthias H Weissensteiner, and Fritz J Sedlazeck. 2021. "Towards Population-Scale Long-Read Sequencing." *Nature Reviews Genetics* 22 (9): 572–87. https://doi.org/10.1038/s41576-021-00367-3.

Crécy-Lagard, Valérie de, Pietro Boccaletto, Carl G Mangleburg, Puneet Sharma, Todd M Lowe, Sebastian A Leidel, and Janusz M Bujnicki. 2019. "Matching TRNA Modifications in Humans to Their Known and Predicted Enzymes." *Nucleic Acids Research* 47 (5): 2143–59. https://doi.org/10.1093/nar/gkz011.

Crick, F. H. C. 1966. "Codon--Anticodon Pairing: The Wobble Hypothesis." *Journal of Molecular Biology* 19: 548–55.

Cui, Qi, Hailing Shi, Peng Ye, Li Li, Qiuhao Qu, Guoqiang Sun, Guihua Sun, et al. 2017. "M(6)A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells." *Cell Reports* 18 (11): 2622–34. https://doi.org/10.1016/j.celrep.2017.02.059.

Dai, Qing, Li-Sheng Zhang, Hui-Lung Sun, Kinga Pajdzik, Lei Yang, Chang Ye, Cheng-Wei Ju, et al. 2023. "Quantitative Sequencing Using BID-Seq Uncovers Abundant Pseudouridines in Mammalian MRNA at Base Resolution." *Nature Biotechnology* 41 (3): 344–54. https://doi.org/10.1038/s41587-022-01505-w.

Dai, Xiaoxia, Tianlu Wang, Gwendolyn Gonzalez, and Yinsheng Wang. 2018. "Identification of YTH Domain-Containing Proteins as the Readers for N1-Methyladenosine in RNA." *Analytical Chemistry*. United States. https://doi.org/10.1021/acs.analchem.8b01703.

Dominissini, Dan, Sharon Moshitch-Moshkovitz, Schraga Schwartz, Mali Salmon-Divon, Lior Ungar, Sivan Osenberg, Karen Cesarkas, et al. 2012. "Topology of the Human and Mouse M6A RNA Methylomes Revealed by M6A-Seq." *Nature* 485 (7397): 201–6. https://doi.org/10.1038/nature11112.

Dosztányi, Zsuzsanna, Veronika Csizmok, Peter Tompa, and István Simon. 2005. "IUPred: Web Server for the Prediction of Intrinsically Unstructured Regions of Proteins Based on Estimated Energy Content." *Bioinformatics* 21 (16): 3433–34. https://doi.org/10.1093/bioinformatics/bti541.

Elsharawy, Khloud A, Omar J Mohammed, Mohammed A Aleskandarany, Ayman Hyder, Hekmat L El-Gammal, Mohamed I Abou-Dobara, Andrew R Green, Leslie W Dalton, and Emad A Rakha. 2020. "The Nucleolar-Related Protein Dyskerin Pseudouridine Synthase 1 (DKC1) Predicts Poor Prognosis in Breast Cancer." *British Journal of Cancer* 123 (10): 1543–52. https://doi.org/10.1038/s41416-020-01045-7.

Eyler, Daniel E, Monika K Franco, Zahra Batool, Monica Z Wu, Michelle L Dubuke, Malgorzata

Dobosz-Bartoszek, Joshua D Jones, Yury S Polikanov, Bijoyita Roy, and Kristin S Koutmou. 2019. “Pseudouridinylation of MRNA Coding Sequences Alters Translation.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (46): 23068—23074. https://doi.org/10.1073/pnas.1821754116.

Fischl, Harry, Jonathan Neve, Zhiqiao Wang, Radhika Patel, Alastair Louey, Bin Tian, and Andre Furger. 2019. “HnRNPC Regulates Cancer-Specific Alternative Cleavage and Polyadenylation Profiles.” *Nucleic Acids Research* 47 (14): 7580–91. https://doi.org/10.1093/nar/gkz461.

Fisher, Andrew J, and Peter A Beal. 2018. “Structural Basis for Eukaryotic MRNA Modification.” *Current Opinion in Structural Biology* 53 (December): 59–68. https://doi.org/10.1016/j.sbi.2018.05.003.

Fishilevich, Simon, Ron Nudel, Noa Rappaport, Rotem Hadar, Inbar Plaschkes, Tsippi Iny Stein, Naomi Rosen, et al. 2017. “GeneHancer: Genome-Wide Integration of Enhancers and Target Genes in GeneCards.” *Database* 2017 (January): bax028. https://doi.org/10.1093/database/bax028.

Frank, Matthew G. annis, Watkins, Maier. 2019. “乳鼠心肌提取 HHS Public Access.” *Physiology & Behavior* 34 (80): 678–87. https://doi.org/10.1016/j.tig.2018.03.007.The.

Frye, Michaela, Bryan T Harada, Mikaela Behm, and Chuan He. 2018. “RNA Modifications Modulate Gene Expression during Development.” *Science (New York, N.Y.)* 361 (6409): 1346–49. https://doi.org/10.1126/science.aau1646.

Fu, Xiang-Dong, and Manuel Jr Ares. 2014. “Context-Dependent Control of Alternative Splicing by RNA-Binding Proteins.” *Nature Reviews. Genetics* 15 (10): 689–701. https://doi.org/10.1038/nrg3778.

Fu, Ye, Guifang Jia, Xueqin Pang, Richard N Wang, Xiao Wang, Charles J Li, Scott Smemo, et al. 2013. “FTO-Mediated Formation of N6-Hydroxymethyladenosine and N6-Formyladenosine in Mammalian RNA.” *Nature Communications* 4 (1): 1798. https://doi.org/10.1038/ncomms2822.

Garcia-Campos, Miguel Angel, Sarit Edelheit, Ursula Toth, Modi Safra, Ran Shachar, Sergey Viukov, Roni Winkler, et al. 2019. “Deciphering the ‘m(6)A Code’ via Antibody-Independent Quantitative Profiling.” *Cell* 178 (3): 731-747.e16. https://doi.org/10.1016/j.cell.2019.06.013.

Garus, Alexandre, and Chantal Autexier. 2021. “Dyskerin: An Essential Pseudouridine Synthase with Multifaceted Roles in Ribosome Biogenesis, Splicing, and Telomere Maintenance.” *RNA (New York, N.Y.)* 27 (12): 1441–58. https://doi.org/10.1261/rna.078953.121.

Gerstberger, Stefanie, Markus Hafner, and Thomas Tuschl. 2014. “A Census of Human RNA-Binding Proteins.” *Nature Reviews. Genetics*. England. https://doi.org/10.1038/nrg3813.

Hamma, Tomoko, and Adrian R Ferré-D’Amaré. 2006. “Pseudouridine Synthases.” *Chemistry & Biology* 13 (11): 1125–35. https://doi.org/10.1016/j.chembiol.2006.09.009.

Harlen, Kevin M., and L. Stirling Churchman. 2017. “The Code and beyond: Transcription Regulation by the RNA Polymerase II Carboxy-Terminal Domain.” *Nature Reviews Molecular Cell Biology* 18 (4): 263–73. https://doi.org/10.1038/nrm.2017.10.

Hartmann, Annette M, Oliver Nayler, Franz Werner Schwaiger, Axel Obermeier, and Stefan Stamm. 1999. “The Interaction and Colocalization of Sam68 with the Splicing-Associated Factor YT521-B in Nuclear Dots Is Regulated by the Src Family Kinase P59 Fyn.” *Molecular Biology of*

the Cell. Vol. 10.

Helm, Mark, and Yuri Motorin. 2017. "Detecting RNA Modifications in the Epitranscriptome: Predict and Validate." *Nature Reviews. Genetics* 18 (5): 275–91. https://doi.org/10.1038/nrg.2016.169.

Ho, Steve S, Alexander E Urban, and Ryan E Mills. 2020. "Structural Variation in the Sequencing Era." *Nature Reviews Genetics* 21 (3): 171–89. https://doi.org/10.1038/s41576-019-0180-9.

Hofmann, Yvonne, and Brunhilde Wirth. 2002. "HnRNP-G Promotes Exon 7 Inclusion of Survival Motor Neuron (SMN) via Direct Interaction with Htra2-B1." *Human Molecular Genetics* 11 (17): 2037–49. https://doi.org/10.1093/hmg/11.17.2037.

HOLLEY, R W, J APGAR, G A EVERETT, J T MADISON, M MARQUISEE, S H MERRILL, J R PENSWICK, and A ZAMIR. 1965. "STRUCTURE OF A RIBONUCLEIC ACID." *Science (New York, N.Y.)* 147 (3664): 1462–65. https://doi.org/10.1126/science.147.3664.1462.

Hsin, Jing Ping, and James L. Manley. 2012. "The RNA Polymerase II CTD Coordinates Transcription and RNA Processing." *Genes and Development* 26 (19): 2119–37. https://doi.org/10.1101/gad.200303.112.

Hu, Jian, Dongxu Qiu, Anze Yu, Jiao Hu, Hao Deng, Huihuang Li, Zhenglin Yi, Jinbo Chen, and Xiongbing Zu. 2021. "YTHDF1 Is a Potential Pan-Cancer Biomarker for Prognosis and Immunotherapy." *Frontiers in Oncology* 11: 607224. https://doi.org/10.3389/fonc.2021.607224.

Hu, Lulu, Shun Liu, Yong Peng, Ruiqi Ge, Rui Su, Chamara Senevirathne, Bryan T Harada, et al. 2022. "M6A RNA Modifications Are Measured at Single-Base Resolution across the Mammalian Transcriptome." *Nature Biotechnology* 40 (8): 1210–19. https://doi.org/10.1038/s41587-022-01243-z.

Huang, Huilin, Hengyou Weng, Wenju Sun, Xi Qin, Hailing Shi, Huizhe Wu, Boxuan Simen Zhao, et al. 2018. "Recognition of RNA N(6)-Methyladenosine by IGF2BP Proteins Enhances MRNA Stability and Translation." *Nature Cell Biology* 20 (3): 285–95. https://doi.org/10.1038/s41556-018-0045-z.

Huang, Sihao, Wen Zhang, Christopher D Katanski, Devin Dersh, Qing Dai, Karen Lolans, Jonathan Yewdell, A Murat Eren, and Tao Pan. 2021. "Interferon Inducible Pseudouridine Modification in Human MRNA by Quantitative Nanopore Profiling." *Genome Biology* 22 (1): 330. https://doi.org/10.1186/s13059-021-02557-y.

Huang, Xi-Tai, Jian-Hui Li, Xiao-Xu Zhu, Chen-Song Huang, Zhuo-Xing Gao, Qiong-Cong Xu, Wei Zhao, and Xiao-Yu Yin. 2021. "HNRNPC Impedes m(6)A-Dependent Anti-Metastatic Alternative Splicing Events in Pancreatic Ductal Adenocarcinoma." *Cancer Letters* 518 (October): 196–206. https://doi.org/10.1016/j.canlet.2021.07.016.

Huttlin, Edward L, Raphael J Bruckner, Joao A Paulo, Joe R Cannon, Lily Ting, Kurt Baltier, Greg Colby, et al. 2017. "Architecture of the Human Interactome Defines Protein Communities and Disease Networks." *Nature* 545 (7655): 505–9. https://doi.org/10.1038/nature22366.

Huttlin, Edward L, Lily Ting, Raphael J Bruckner, Fana Gebreab, Melanie P Gygi, John Szpyt, Stanley Tam, et al. 2015. "The BioPlex Network: A Systematic Exploration of the Human Interactome." *Cell* 162 (2): 425–40. https://doi.org/10.1016/j.cell.2015.06.043.

Ip, Joanna Y., Dominic Schmidt, Qun Pan, Arun K. Ramani, Andrew G. Fraser, Duncan T. Odom, and

Benjamin J. Blencowe. 2011. “Global Impact of RNA Polymerase II Elongation Inhibition on Alternative Splicing Regulation.” *Genome Research* 21 (3): 390–401. https://doi.org/10.1101/gr.111070.110.

Jackson, Lisa A, Evan J Anderson, Nadine G Rouphael, Paul C Roberts, Mamodikoe Makhene, Rhea N Coler, Michele P McCullough, et al. 2020. “An MRNA Vaccine against SARS-CoV-2 - Preliminary Report.” *The New England Journal of Medicine* 383 (20): 1920–31. https://doi.org/10.1056/NEJMoa2022483.

Jacobs, Steven A, and Sepideh Khorasanizadeh. 2002. “Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail.” *Science* 295 (5562): 2080–83. https://doi.org/10.1126/science.1069473.

Jády, Beáta E, Amandine Ketele, and Tamás Kiss. 2018. “Dynamic Association of Human MRNP Proteins with Mitochondrial TRNAs in the Cytosol.” *RNA (New York, N.Y.)* 24 (12): 1706–20. https://doi.org/10.1261/rna.066738.118.

Jia, Guifang, Ye Fu, Xu Zhao, Qing Dai, Guanqun Zheng, Ying Yang, Chengqi Yi, et al. 2011. “N6-Methyladenosine in Nuclear RNA Is a Major Substrate of the Obesity-Associated FTO.” *Nature Chemical Biology* 7 (12): 885–87. https://doi.org/10.1038/nchembio.687.

Kan, Lijuan, Stanislav Ott, Brian Joseph, Eun Sil Park, Wei Dai, Ralph E Kleiner, Adam Claridge-Chang, and Eric C Lai. 2021. “A Neural m(6)A/Ythdf Pathway Is Required for Learning and Memory in Drosophila.” *Nature Communications* 12 (1): 1458. https://doi.org/10.1038/s41467-021-21537-1.

Kanhoush, Rasha, Brent Beenders, Caroline Perrin, Jacques Moreau, Michel Bellini, and May Penrad-Mobayed. 2010. “Novel Domains in the HnRNP G/RBMX Protein with Distinct Roles in RNA Binding and Targeting Nascent Transcripts.” *Nucleus* 1 (1): 109–22. https://doi.org/10.4161/nucl.1.1.10857.

Karikó, Katalin, Hiromi Muramatsu, Frank A Welsh, János Ludwig, Hiroki Kato, Shizuo Akira, and Drew Weissman. 2008. “Incorporation of Pseudouridine into MRNA Yields Superior Nonimmunogenic Vector with Increased Translational Capacity and Biological Stability.” *Molecular Therapy : The Journal of the American Society of Gene Therapy* 16 (11): 1833–40. https://doi.org/10.1038/mt.2008.200.

Karlsborn, Tony, A K M Firoj Mahmud, Hasan Tükenmez, and Anders S Byström. 2016. “Loss of Ncm(5) and Mcm(5) Wobble Uridine Side Chains Results in an Altered Metabolic Profile.” *Metabolomics : Official Journal of the Metabolomic Society* 12 (12): 177. https://doi.org/10.1007/s11306-016-1120-8.

Kasowitz, Seth D., Jun Ma, Stephen J. Anderson, N. Adrian Leu, Yang Xu, Brian D. Gregory, Richard M. Schultz, and P. Jeremy Wang. 2018. “Nuclear M6A Reader YTHDC1 Regulates Alternative Polyadenylation and Splicing during Mouse Oocyte Development.” *PLoS Genetics* 14 (5): 1–28. https://doi.org/10.1371/journal.pgen.1007412.

Ke, Shengdong, Amy Pandya-Jones, Yuhki Saito, John J. Fak, Cathrine Broberg Vågbø, Shay Geula, Jacob H. Hanna, Douglas L. Black, James E. Darnell, and Robert B. Darnell. 2017. “M6A MRNA Modifications Are Deposited in Nascent Pre-MRNA and Are Not Required for Splicing but Do Specify Cytoplasmic Turnover.” *Genes and Development* 31 (10): 990–1006. https://doi.org/10.1101/gad.301036.117.

Kierzek, Elzbieta, and Ryszard Kierzek. 2003. “The Thermodynamic Stability of RNA Duplexes and

Hairpins Containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines." *Nucleic Acids Research* 31 (15): 4472–80. https://doi.org/10.1093/nar/gkg633.

Kierzek, Elzbieta, Magdalena Malgowska, Jolanta Lisowiec, Douglas H Turner, Zofia Gdaniec, and Ryszard Kierzek. 2014. "The Contribution of Pseudouridine to Stabilities and Structure of RNAs." *Nucleic Acids Research* 42 (5): 3492–3501. https://doi.org/10.1093/nar/gkt1330.

Knuckles, Philip, Sarah H. Carl, Michael Musheev, Christof Niehrs, Alice Wenger, and Marc Bühler. 2017. "RNA Fate Determination through Cotranscriptional Adenosine Methylation and Microprocessor Binding." *Nature Structural and Molecular Biology* 24 (7): 561–69. https://doi.org/10.1038/nsmb.3419.

Kolaj-Robin, Olga, and Bertrand Séraphin. 2017. "Structures and Activities of the Elongator Complex and Its Cofactors." *The Enzymes* 41: 117–49. https://doi.org/10.1016/bs.enz.2017.03.001.

Kornblihtt, Alberto R., Ignacio E. Schor, Mariano Alló, Gwendal Dujardin, Ezequiel Petrillo, and Manuel J. Muñoz. 2013. "Alternative Splicing: A Pivotal Step between Eukaryotic Transcription and Translation." *Nature Reviews Molecular Cell Biology* 14 (3): 153–65. https://doi.org/10.1038/nrm3525.

Kurata, Shinya, Albert Weixlbaumer, Takashi Ohtsuki, Tomomi Shimazaki, Takeshi Wada, Yohei Kirino, Kazuyuki Takai, Kimitsuna Watanabe, V Ramakrishnan, and Tsutomu Suzuki. 2008. "Modified Uridines with C5-Methylene Substituents at the First Position of the TRNA Anticodon Stabilize U.G Wobble Pairing during Decoding." *The Journal of Biological Chemistry* 283 (27): 18801–11. https://doi.org/10.1074/jbc.M800233200.

Kwon, Ilmin, Masato Kato, Siheng Xiang, Leeju Wu, Pano Theodoropoulos, Hamid Mirzaei, Tina Han, Shanhai Xie, Jeffry L. Corden, and Steven L. McKnight. 2013. "XPhosphorylation-Regulated Binding of RNA Polymerase II to Fibrous Polymers of Low-Complexity Domains." *Cell* 155 (5): 1049. https://doi.org/10.1016/j.cell.2013.10.033.

Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Lee, Joo-Hyung, Ruoyu Wang, Feng Xiong, Joanna Krakowiak, Zian Liao, Phuoc T Nguyen, Elena V Moroz-Omori, et al. 2021. "Enhancer RNA M6A Methylation Facilitates Transcriptional Condensate Formation and Gene Activation." *Molecular Cell* 81 (16): 3368-3385.e9. https://doi.org/10.1016/j.molcel.2021.07.024.

Lentini, Jenna M, Jillian Ramos, and Dragony Fu. 2018. "Monitoring the 5-Methoxycarbonylmethyl-2-Thiouridine (Mcm5s2U) Modification in Eukaryotic TRNAs via the γ-Toxin Endonuclease." *RNA (New York, N.Y.)* 24 (5): 749–58. https://doi.org/10.1261/rna.065581.118.

Levi, Ofri, and Yoav S Arava. 2021. "Pseudouridine-Mediated Translation Control of MRNA by Methionine Aminoacyl TRNA Synthetase." *Nucleic Acids Research* 49 (1): 432–43. https://doi.org/10.1093/nar/gkaa1178.

Li, Ang, Yu-Sheng Chen, Xiao-Li Ping, Xin Yang, Wen Xiao, Ying Yang, Hui-Ying Sun, et al. 2017. "Cytoplasmic m(6)A Reader YTHDF3 Promotes MRNA Translation." *Cell Research*. England. https://doi.org/10.1038/cr.2017.10.

Li, Hao, Jingwei Liu, Shixuan Shen, Di Dai, Shitong Cheng, Xiaolong Dong, Liping Sun, and Xiaolin Guo. 2020. "Pan-Cancer Analysis of Alternative Splicing Regulator Heterogeneous Nuclear Ribonucleoproteins (HnRNPs) Family and Their Prognostic Potential." *Journal of Cellular and*

Molecular Medicine 24 (19): 11111–19. https://doi.org/10.1111/jcmm.15558.

Li, Heng. 2018. "Minimap2: Pairwise Alignment for Nucleotide Sequences." *Bioinformatics (Oxford, England)* 34 (18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics (Oxford, England)* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Li, Xiaoyu, Shiqing Ma, and Chengqi Yi. 2016. "Pseudouridine: The Fifth RNA Nucleotide with Renewed Interests." *Current Opinion in Chemical Biology* 33: 108–16. https://doi.org/https://doi.org/10.1016/j.cbpa.2016.06.014.

Li, Xiaoyu, Ping Zhu, Shiqing Ma, Jinghui Song, Jinyi Bai, Fangfang Sun, and Chengqi Yi. 2015. "Chemical Pulldown Reveals Dynamic Pseudouridylation of the Mammalian Transcriptome." *Nature Chemical Biology* 11 (8): 592–97. https://doi.org/10.1038/nchembio.1836.

Li, Yaozong, Rajiv Kumar Bedi, Lars Wiedmer, Xianqiang Sun, Danzhi Huang, and Amedeo Caflisch. 2021. "Atomistic and Thermodynamic Analysis of N6-Methyladenosine (M6A) Recognition by the Reader Domain of YTHDC1." *Journal of Chemical Theory and Computation* 17 (2): 1240–49. https://doi.org/10.1021/acs.jctc.0c01136.

Li, Zejuan, Hengyou Weng, Rui Su, Xiaocheng Weng, Zhixiang Zuo, Chenying Li, Huilin Huang, et al. 2017. "FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N(6)-Methyladenosine RNA Demethylase." *Cancer Cell* 31 (1): 127–41. https://doi.org/10.1016/j.ccell.2016.11.017.

Li, Zhenrui, Pengxu Qian, Wanqing Shao, Hailing Shi, Xi C He, Madelaine Gogol, Zulin Yu, et al. 2018. "Suppression of m(6)A Reader Ythdf2 Promotes Hematopoietic Stem Cell Expansion." *Cell Research* 28 (9): 904–17. https://doi.org/10.1038/s41422-018-0072-0.

Lin, Huan, Kenjyo Miyauchi, Tai Harada, Ryo Okita, Eri Takeshita, Hirofumi Komaki, Kaoru Fujioka, et al. 2018. "CO(2)-Sensitive TRNA Modification Associated with Human Mitochondrial Disease." *Nature Communications* 9 (1): 1875. https://doi.org/10.1038/s41467-018-04250-4.

Linder, Bastian, Anya V Grozhik, Anthony O Olarerin-George, Cem Meydan, Christopher E Mason, and Samie R Jaffrey. 2015. "Single-Nucleotide-Resolution Mapping of M6A and M6Am throughout the Transcriptome." *Nature Methods* 12 (8): 767–72. https://doi.org/10.1038/nmeth.3453.

Lindorff-Larsen, Kresten, Stefano Piana, Ron O Dror, and David E Shaw. 2011. "How Fast-Folding Proteins Fold." *Science (New York, N.Y.)* 334 (6055): 517–20. https://doi.org/10.1126/science.1208351.

Liu, Bei, Honglue Shi, Atul Rangadurai, Felix Nussbaumer, Chia-Chieh Chu, Kevin Andreas Erharter, David A Case, Christoph Kreutz, and Hashim M Al-Hashimi. 2021. "A Quantitative Model Predicts How M6A Reshapes the Kinetic Landscape of Nucleic Acid Hybridization and Conformational Transitions." *Nature Communications* 12 (1): 5201. https://doi.org/10.1038/s41467-021-25253-8.

Liu, Jianzhao, Yanan Yue, Dali Han, Xiao Wang, Ye Fu, Liang Zhang, Guifang Jia, et al. 2014. "A METTL3–METTL14 Complex Mediates Mammalian Nuclear RNA N6-Adenosine Methylation." *Nature Chemical Biology* 10 (2): 93–95. https://doi.org/10.1038/nchembio.1432.

Liu, Jun, Xiaoyang Dou, Chuanyuan Chen, Chuan Chen, Chang Liu, Meng Michelle Xu, Siqi Zhao, et al.

2020. “N (6)-Methyladenosine of Chromosome-Associated Regulatory RNA Regulates Chromatin State and Transcription.” *Science (New York, N.Y.)* 367 (6477): 580–86. https://doi.org/10.1126/science.aay6018.

Liu, Nian, Qing Dai, Guanqun Zheng, Chuan He, Marc Parisien, and Tao Pan. 2015. “N(6)-Methyladenosine-Dependent RNA Structural Switches Regulate RNA-Protein Interactions.” *Nature* 518 (7540): 560–64. https://doi.org/10.1038/nature14234.

Liu, Nian, Katherine I. Zhou, Marc Parisien, Qing Dai, Luda Diatchenko, and Tao Pan. 2017. “N6-Methyladenosine Alters RNA Structure to Regulate Binding of a Low-Complexity Protein.” *Nucleic Acids Research* 45 (10): 6051–63. https://doi.org/10.1093/nar/gkx141.

Liu, Tao, Qinglv Wei, Jing Jin, Qingya Luo, Yi Liu, Yu Yang, Chunming Cheng, et al. 2020. “The M6A Reader YTHDF1 Promotes Ovarian Cancer Progression via Augmenting EIF3C Translation.” *Nucleic Acids Research* 48 (7): 3816–31. https://doi.org/10.1093/nar/gkaa048.

Logsdon, Glennis A, Mitchell R Vollger, and Evan E Eichler. 2020. “Long-Read Human Genome Sequencing and Its Applications.” *Nature Reviews Genetics* 21 (10): 597–614. https://doi.org/10.1038/s41576-020-0236-x.

Louloupi, Annita, Evgenia Ntini, Thomas Conrad, and Ulf Andersson Vang Ørom. 2018. “Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of M6A in Splicing Efficiency.” *Cell Reports* 23 (12): 3429–37. https://doi.org/10.1016/j.celrep.2018.05.077.

Lovejoy, Alexander F, Daniel P Riordan, and Patrick O Brown. 2014. “Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine Synthases Modify Specific MRNAs in S. Cerevisiae.” *PLOS ONE* 9 (10): 1–15. https://doi.org/10.1371/journal.pone.0110799.

Martinez, Nicole M, Amanda Su, Margaret C Burns, Julia K Nussbacher, Cassandra Schaening, Shashank Sathe, Gene W Yeo, and Wendy V Gilbert. 2022. “Pseudouridine Synthases Modify Human Pre-MRNA Co-Transcriptionally and Affect Pre-MRNA Processing.” *Molecular Cell* 82 (3): 645-659.e9. https://doi.org/10.1016/j.molcel.2021.12.023.

Mauer, Jan, Miriam Sindelar, Vladimir Despic, Théo Guez, Ben R Hawley, Jean-Jacques Vasseur, Andrea Rentmeister, et al. 2019. “FTO Controls Reversible m(6)Am RNA Methylation during SnRNA Biogenesis.” *Nature Chemical Biology* 15 (4): 340–47. https://doi.org/10.1038/s41589-019-0231-8.

Mayer, Andreas, Heather M Landry, and L Stirling Churchman. 2017. “Pause & Go: From the Discovery of RNA Polymerase Pausing to Its Functional Implications.” *Current Opinion in Cell Biology* 46 (June): 72–80. https://doi.org/10.1016/j.ceb.2017.03.002.

Meyer, Kate D., Yogesh Saletore, Paul Zumbo, Olivier Elemento, Christopher E. Mason, and Samie R. Jaffrey. 2012. “Comprehensive Analysis of MRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons.” *Cell* 149 (7): 1635–46. https://doi.org/10.1016/j.cell.2012.05.003.

Meyer, Kate D. 2019. “DART-Seq: An Antibody-Free Method for Global M6A Detection.” *Nature Methods* 16 (12): 1275–80. https://doi.org/10.1038/s41592-019-0570-0.

Mineri, Rossana, Norman Pavelka, Erika Fernandez-Vizarra, Paola Ricciardi-Castagnoli, Massimo Zeviani, and Valeria Tiranti. 2009. “How Do Human Cells React to the Absence of Mitochondrial DNA?” *PloS One* 4 (5): e5713. https://doi.org/10.1371/journal.pone.0005713.

Molinie, Benoit, Jinkai Wang, Kok Seong Lim, Roman Hillebrand, Zhi-xiang Lu, Nicholas Van Wittenberghe, Benjamin D Howard, et al. 2016. “M6A-LAIC-Seq Reveals the Census and

Complexity of the M6A Epitranscriptome." *Nature Methods* 13 (8): 692–98. https://doi.org/10.1038/nmeth.3898.

Morais, Pedro, Hironori Adachi, and Yi Tao Yu. 2021. "The Critical Contribution of Pseudouridine to MRNA COVID-19 Vaccines." *Frontiers in Cell and Developmental Biology* 9 (November): 1–9. https://doi.org/10.3389/fcell.2021.789427.

Muñoz, Manuel J., Manuel de la Mata, and Alberto R. Kornblihtt. 2010. "The Carboxy Terminal Domain of RNA Polymerase II and Alternative Splicing." *Trends in Biochemical Sciences* 35 (9): 497–504. https://doi.org/10.1016/j.tibs.2010.03.010.

Nojima, Takayuki, Kenny Rebelo, Tomás Gomes, Ana Rita Grosso, Nicholas J. Proudfoot, and Maria Carmo-Fonseca. 2018. "RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-Transcriptional Splicing." *Molecular Cell* 72 (2): 369-379.e4. https://doi.org/10.1016/j.molcel.2018.09.004.

Oikonomopoulos, Spyros, Anthony Bayega, Somayyeh Fahiminiya, Haig Djambazian, Pierre Berube, and Jiannis Ragoussis. 2020. "Methodologies for Transcript Profiling Using Long-Read Technologies." *Frontiers in Genetics* 11: 606. https://doi.org/10.3389/fgene.2020.00606.

Ozdilek, Bagdeser A., Valery F. Thompson, Nasiha S. Ahmed, Connor I. White, Robert T. Batey, and Jacob C. Schwartz. 2017. "Intrinsically Disordered RGG/RG Domains Mediate Degenerate Specificity in RNA Binding." *Nucleic Acids Research* 45 (13): 7984–96. https://doi.org/10.1093/nar/gkx460.

Patil, Deepak P, Chun-Kan Chen, Brian F Pickering, Amy Chow, Constanza Jackson, Mitchell Guttman, and Samie R Jaffrey. 2016. "M(6)A RNA Methylation Promotes XIST-Mediated Transcriptional Repression." *Nature* 537 (7620): 369–73. https://doi.org/10.1038/nature19342.

Payne, Alexander, Nadine Holmes, Vardhman Rakyan, and Matthew Loose. 2019. "BulkVis: A Graphical Viewer for Oxford Nanopore Bulk FAST5 Files." *Bioinformatics* 35 (13): 2193–98. https://doi.org/10.1093/bioinformatics/bty841.

Pendleton, Kathryn E, Beibei Chen, Kuanqing Liu, Olga V Hunter, Yang Xie, Benjamin P Tu, and Nicholas K Conrad. 2017. "The U6 SnRNA m(6)A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention." *Cell* 169 (5): 824-835.e14. https://doi.org/10.1016/j.cell.2017.05.003.

Phizicky, Eric M, and Anita K Hopper. 2010. "TRNA Biology Charges to the Front." *Genes & Development* 24 (17): 1832–60. https://doi.org/10.1101/gad.1956510.

Ping, Xiao-Li, Bao-Fa Sun, Lu Wang, Wen Xiao, Xin Yang, Wen-Jia Wang, Samir Adhikari, et al. 2014. "Mammalian WTAP Is a Regulatory Subunit of the RNA N6-Methyladenosine Methyltransferase." *Cell Research* 24 (2): 177–89. https://doi.org/10.1038/cr.2014.3.

Pope, Scott D, and Ruslan Medzhitov. 2018. "Emerging Principles of Gene Expression Programs and Their Regulation." *Molecular Cell* 71 (3): 389–97. https://doi.org/10.1016/j.molcel.2018.07.017.

Rafalska, Ilona, Zhaiyi Zhang, Natalya Benderska, Horst Wolff, Annette M. Hartmann, Ruth Brack-Werner, and Stefan Stamm. 2004. "The Intranuclear Localization and Function of YT521-B Is Regulated by Tyrosine Phosphorylation." *Human Molecular Genetics* 13 (15): 1535–49. https://doi.org/10.1093/hmg/ddh167.

Ramanathan, Anand, G Brett Robb, and Siu-Hong Chan. 2016. “MRNA Capping: Biological Functions and Applications.” *Nucleic Acids Research* 44 (16): 7511–26. https://doi.org/10.1093/nar/gkw551.

Rossmanith, Walter. 2011. “Localization of Human RNase Z Isoforms: Dual Nuclear/Mitochondrial Targeting of the ELAC2 Gene Product by Alternative Translation Initiation.” *PloS One* 6 (4): e19152. https://doi.org/10.1371/journal.pone.0019152.

———. 2012. “Of P and Z: Mitochondrial TRNA Processing Enzymes.” *Biochimica et Biophysica Acta* 1819 (9–10): 1017–26. https://doi.org/10.1016/j.bbagrm.2011.11.003.

Roundtree, Ian A, Molly E Evans, Tao Pan, and Chuan He. 2017. “Dynamic RNA Modifications in Gene Expression Regulation.” *Cell* 169 (7): 1187–1200. https://doi.org/10.1016/j.cell.2017.05.045.

Roundtree, Ian A, Guan-Zheng Luo, Zijie Zhang, Xiao Wang, Tao Zhou, Yiquang Cui, Jiahao Sha, et al. 2017. “YTHDC1 Mediates Nuclear Export of N6-Methyladenosine Methylated MRNAs.” Edited by Nick J Proudfoot. *ELife* 6: e31311. https://doi.org/10.7554/eLife.31311.

Safra, Modi, Ronit Nir, Daneyal Farouq, Ilya Vainberg Slutskin, and Schraga Schwartz. 2017. “TRUB1 Is the Predominant Pseudouridine Synthase Acting on Mammalian MRNA via a Predictable and Conserved Code.” *Genome Research* 27 (3): 393–406. https://doi.org/10.1101/gr.207613.116.

Schöller, Eva, Franziska Weichmann, Thomas Treiber, Sam Ringle, Nora Treiber, Andrew Flatley, Regina Feederle, Astrid Bruckmann, and Gunter Meister. 2018. “Interactions, Localization, and Phosphorylation of the m(6)A Generating METTL3-METTL14-WTAP Complex.” *RNA (New York, N.Y.)* 24 (4): 499–512. https://doi.org/10.1261/rna.064063.117.

Schwartz, Jacob C., Xueyin Wang, Elaine R. Podell, and Thomas R. Cech. 2013. “RNA Seeds Higher-Order Assembly of FUS Protein.” *Cell Reports* 5 (4): 918–25. https://doi.org/10.1016/j.celrep.2013.11.017.

Schwartz, Schraga, Douglas A Bernstein, Maxwell R Mumbach, Marko Jovanovic, Rebecca H Herbst, Brian X León-Ricardo, Jesse M Engreitz, et al. 2014. “Transcriptome-Wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of NcRNA and MRNA.” *Cell* 159 (1): 148–62. https://doi.org/10.1016/j.cell.2014.08.028.

Schwartz, Schraga, Maxwell R Mumbach, Marko Jovanovic, Tim Wang, Karolina Maciag, G Guy Bushkin, Philipp Mertins, et al. 2014. “Perturbation of M6A Writers Reveals Two Distinct Classes of MRNA Methylation at Internal and 5’ Sites.” *Cell Reports* 8 (1): 284–96. https://doi.org/10.1016/j.celrep.2014.05.048.

Sharma, Sunny, Jean-Louis Langhendries, Peter Watzinger, Peter Kötter, Karl-Dieter Entian, and Denis L J Lafontaine. 2015. “Yeast Kre33 and Human NAT10 Are Conserved 18S RRNA Cytosine Acetyltransferases That Modify TRNAs Assisted by the Adaptor Tan1/THUMPD1.” *Nucleic Acids Research* 43 (4): 2242–58. https://doi.org/10.1093/nar/gkv075.

Shaukat, Athanasios-Nasir, Eleni G Kaliatsi, Vassiliki Stamatopoulou, and Constantinos Stathopoulos. 2021. “Mitochondrial TRNA-Derived Fragments and Their Contribution to Gene Expression Regulation.” *Frontiers in Physiology* 12: 729452. https://doi.org/10.3389/fphys.2021.729452.

Shi, Hanhan, Peiwei Chai, Renbing Jia, and Xianqun Fan. 2020. “Novel Insight into the Regulatory

Roles of Diverse RNA Modifications: Re-Defining the Bridge between Transcription and Translation." *Molecular Cancer* 19 (1): 78. https://doi.org/10.1186/s12943-020-01194-6.

Shi, Rongkai, Shilong Ying, Yadan Li, Liyuan Zhu, Xian Wang, and Hongchuan Jin. 2021. "Linking the YTH Domain to Cancer: The Importance of YTH Family Proteins in Epigenetics." *Cell Death & Disease* 12 (4): 346. https://doi.org/10.1038/s41419-021-03625-8.

Shu, Xiao, Jie Cao, Mohan Cheng, Siying Xiang, Minsong Gao, Ting Li, Xiner Ying, et al. 2020. "A Metabolic Labeling Method Detects m(6)A Transcriptome-Wide at Single Base Resolution." *Nature Chemical Biology* 16 (8): 887–95. https://doi.org/10.1038/s41589-020-0526-9.

Shukla, Sanjeev, Ersen Kavak, Melissa Gregory, Masahiko Imashimizu, Bojan Shutinoski, Mikhail Kashlev, Philipp Oberdoerffer, Rickard Sandberg, and Shalini Oberdoerffer. 2011. "CTCF-Promoted RNA Polymerase II Pausing Links DNA Methylation to Splicing." *Nature* 479 (7371): 74–79. https://doi.org/10.1038/nature10442.

Skinner, John J, Wookyung Yu, Elizabeth K Gichana, Michael C Baxa, James R Hinshaw, Karl F Freed, and Tobin R Sosnick. 2014. "Benchmarking All-Atom Simulations Using Hydrogen Exchange." *Proceedings of the National Academy of Sciences of the United States of America* 111 (45): 15975–80. https://doi.org/10.1073/pnas.1404213111.

Śledź, Paweł, and Martin Jinek. 2016. "Structural Insights into the Molecular Mechanism of the m(6)A Writer Complex." *ELife* 5 (September). https://doi.org/10.7554/eLife.18434.

Smith, Andrew M, Miten Jain, Logan Mulroney, Daniel R Garalde, and Mark Akeson. 2019. "Reading Canonical and Modified Nucleobases in 16S Ribosomal RNA Using Nanopore Native RNA Sequencing." *PloS One* 14 (5): e0216709. https://doi.org/10.1371/journal.pone.0216709.

Su, Rui, Lei Dong, Chenying Li, Sigrid Nachtergaele, Mark Wunderlich, Ying Qing, Xiaolan Deng, et al. 2018. "R-2HG Exhibits Anti-Tumor Activity by Targeting FTO/m(6)A/MYC/CEBPA Signaling." *Cell* 172 (1–2): 90-105.e23. https://doi.org/10.1016/j.cell.2017.11.031.

Suzuki, Takeo, Yuka Yashiro, Ittoku Kikuchi, Yuma Ishigami, Hironori Saito, Ikuya Matsuzawa, Shunpei Okada, et al. 2020. "Complete Chemical Structures of Human Mitochondrial TRNAs." *Nature Communications* 11 (1): 4269. https://doi.org/10.1038/s41467-020-18068-6.

Suzuki, Tsutomu. 2021. "The Expanding World of TRNA Modifications and Their Disease Relevance." *Nature Reviews Molecular Cell Biology* 22 (6): 375–92. https://doi.org/10.1038/s41580-021-00342-0.

Thandapani, Palaniraja, Timothy R. O'Connor, Timothy L. Bailey, and Stéphane Richard. 2013. "Defining the RGG/RG Motif." *Molecular Cell* 50 (5): 613–23. https://doi.org/10.1016/j.molcel.2013.05.021.

"The Gene Ontology Resource: Enriching a GOld Mine." 2021. *Nucleic Acids Research* 49 (D1): D325–34. https://doi.org/10.1093/nar/gkaa1113.

Tuorto, Francesca, Carine Legrand, Cansu Cirzi, Giuseppina Federico, Reinhard Liebers, Martin Müller, Ann E Ehrenhofer-Murray, Gunnar Dittmar, Hermann-Josef Gröne, and Frank Lyko. 2018. "Queuosine-Modified TRNAs Confer Nutritional Control of Protein Translation." *The EMBO Journal* 37 (18). https://doi.org/10.15252/embj.201899777.

Valverde, Roberto, Laura Edwards, and Lynne Regan. 2008. "Structure and Function of KH Domains." *The FEBS Journal* 275 (11): 2712–26. https://doi.org/https://doi.org/10.1111/j.1742-4658.2008.06411.x.

Wang, Xiang, Jing Feng, Yuan Xue, Zeyuan Guan, Delin Zhang, Zhu Liu, Zhou Gong, et al. 2016. "Structural Basis of N(6)-Adenosine Methylation by the METTL3-METTL14 Complex." *Nature* 534 (7608): 575–78. https://doi.org/10.1038/nature18298.

Wang, Xiao, Zhike Lu, Adrian Gomez, Gary C Hon, Yanan Yue, Dali Han, Ye Fu, et al. 2014. "N6-Methyladenosine-Dependent Regulation of Messenger RNA Stability." *Nature* 505 (7481): 117–20. https://doi.org/10.1038/nature12730.

Wang, Xiao, Boxuan Simen Zhao, Ian A Roundtree, Zhike Lu, Dali Han, Honghui Ma, Xiaocheng Weng, Kai Chen, Hailing Shi, and Chuan He. 2015. "N(6)-Methyladenosine Modulates Messenger RNA Translation Efficiency." *Cell* 161 (6): 1388–99. https://doi.org/10.1016/j.cell.2015.05.014.

Wang, Ye, Yu Xiao, Shunqing Dong, Qiong Yu, and Guifang Jia. 2020. "Antibody-Free Enzyme-Assisted Chemical Approach for Detection of N6-Methyladenosine." *Nature Chemical Biology* 16 (8): 896–903. https://doi.org/10.1038/s41589-020-0525-x.

Watkins, Christopher P, Wen Zhang, Adam C Wylder, Christopher D Katanski, and Tao Pan. 2022. "A Multiplex Platform for Small RNA Sequencing Elucidates Multifaceted TRNA Stress Response and Translational Regulation." *Nature Communications* 13 (1): 2491. https://doi.org/10.1038/s41467-022-30261-3.

Watkins, N J, A Gottschalk, G Neubauer, B Kastner, P Fabrizio, M Mann, and R Lührmann. 1998. "Cbf5p, a Potential Pseudouridine Synthase, and Nhp2p, a Putative RNA-Binding Protein, Are Present Together with Gar1p in All H BOX/ACA-Motif SnoRNPs and Constitute a Common Bipartite Structure." *RNA (New York, N.Y.)* 4 (12): 1549–68. https://doi.org/10.1017/s1355838298980761.

Woodcock, Clayton B, John R Horton, Jujun Zhou, Mark T Bedford, Robert M Blumenthal, Xing Zhang, and Xiaodong Cheng. 2020. "Biochemical and Structural Basis for YTH Domain of Human YTHDC1 Binding to Methylated Adenine in DNA." *Nucleic Acids Research* 48 (18): 10329–41. https://doi.org/10.1093/nar/gkaa604.

Wu, Guowei, Hironori Adachi, Junhui Ge, David Stephenson, Charles C Query, and Yi-Tao Yu. 2016. "Pseudouridines in U2 SnRNA Stimulate the ATPase Activity of Prp5 during Spliceosome Assembly." *The EMBO Journal* 35 (6): 654–67. https://doi.org/10.15252/embj.201593113.

WYATT, G R, and S S COHEN. 1953. "The Bases of the Nucleic Acids of Some Bacterial and Animal Viruses: The Occurrence of 5-Hydroxymethylcytosine." *The Biochemical Journal* 55 (5): 774–82. https://doi.org/10.1042/bj0550774.

Xiao, Wen, Samir Adhikari, Ujwal Dahal, Yu-Sheng Chen, Ya-Juan Hao, Bao-Fa Sun, Hui-Ying Sun, et al. 2016a. "Nuclear m(6)A Reader YTHDC1 Regulates MRNA Splicing." *Molecular Cell* 61 (4): 507–19. https://doi.org/10.1016/j.molcel.2016.01.012.

Xiao, Wen, Samir Adhikari, Ujwal Dahal, Yu Sheng Chen, Ya Juan Hao, Bao Fa Sun, Hui Ying Sun, et al. 2016b. "Nuclear M6A Reader YTHDC1 Regulates MRNA Splicing." *Molecular Cell* 61 (4): 507–19. https://doi.org/10.1016/j.molcel.2016.01.012.

Xie, Wen, Hecheng Zhu, Ming Zhao, Lei Wang, Shasha Li, Cong Zhao, Yao Zhou, et al. 2021. "Crucial Roles of Different RNA-Binding HnRNP Proteins in Stem Cells." *International Journal of Biological Sciences* 17 (3): 807–17. https://doi.org/10.7150/ijbs.55120.

Xu, Chao, Xiao Wang, Ke Liu, Ian A Roundtree, Wolfram Tempel, Yanjun Li, Zhike Lu, Chuan He, and Jinrong Min. 2014. "Structural Basis for Selective Binding of M6A RNA by the YTHDC1 YTH

Domain." *Nature Chemical Biology* 10 (11): 927–29. https://doi.org/10.1038/nchembio.1654.

Ying, Yufan, Xueyou Ma, Jiajie Fang, Shiming Chen, Weiyu Wang, Jiangfeng Li, Haiyun Xie, et al. 2021. "EGR2-Mediated Regulation of m(6)A Reader IGF2BP Proteins Drive RCC Tumorigenesis and Metastasis via Enhancing S1PR3 MRNA Stabilization." *Cell Death & Disease* 12 (8): 750. https://doi.org/10.1038/s41419-021-04038-3.

Zhang, Sicong, Boxuan Simen Zhao, Aidong Zhou, Kangyu Lin, Shaoping Zheng, Zhike Lu, Yaohui Chen, et al. 2017. "M(6)A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program." *Cancer Cell* 31 (4): 591-606.e6. https://doi.org/10.1016/j.ccell.2017.02.013.

Zhao, Xu, Ying Yang, Bao-Fa Sun, Yue Shi, Xin Yang, Wen Xiao, Ya-Juan Hao, et al. 2014. "FTO-Dependent Demethylation of N6-Methyladenosine Regulates MRNA Splicing and Is Required for Adipogenesis." *Cell Research* 24 (12): 1403–19. https://doi.org/10.1038/cr.2014.151.

Zheng, Guanqun, John Arne Dahl, Yamei Niu, Peter Fedorcsak, Chun-Min Huang, Charles J Li, Cathrine B Vågbø, et al. 2013. "ALKBH5 Is a Mammalian RNA Demethylase That Impacts RNA Metabolism and Mouse Fertility." *Molecular Cell* 49 (1): 18–29. https://doi.org/10.1016/j.molcel.2012.10.015.

Zhou, Katherine I., Hailing Shi, Ruitu Lyu, Adam C. Wylder, Żaneta Matuszek, Jessica N. Pan, Chuan He, Marc Parisien, and Tao Pan. 2019. "Regulation of Co-Transcriptional Pre-MRNA Splicing by M6A through the Low-Complexity Protein HnRNPG." *Molecular Cell* 76 (1): 70-81.e9. https://doi.org/10.1016/j.molcel.2019.07.005.

Zhou, Katherine I, and Tao Pan. 2018. "An Additional Class of m(6)A Readers." *Nature Cell Biology* 20 (3): 230–32. https://doi.org/10.1038/s41556-018-0046-y.

Zhu, Tingting, Ian A Roundtree, Ping Wang, Xiao Wang, Li Wang, Chang Sun, Yuan Tian, Jie Li, Chuan He, and Yanhui Xu. 2014. "Crystal Structure of the YTH Domain of YTHDF2 Reveals Mechanism for Recognition of N6-Methyladenosine." *Cell Research*. England. https://doi.org/10.1038/cr.2014.152.