Gene Expression Regulation by the Decapping Enzyme NUDT12 and the RNA helicase YTHDC2

Li, Lingyun

UNIVERSITÉ DE GENÈVE FACULTÉ DES SCIENCES

Section de biologie

Département de biologie moléculaire et cellulaire Professeur Ramesh S. Pillai

Gene Expression Regulation by the Decapping Enzyme NUDT12 and the RNA Helicase YTHDC2

THÈSE

présentée aux Facultés de médecine et des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences en sciences de la vie,
mention Biosciences moléculaires

par

Lingyun LI

de

Dancheng, Chine

Thèse Nº nnnn

GENÈVE

2022

Table of Contents

Acknowledgements 1

Abstract (English) 3

Abstract (French) 5

Chapter 1. Regulation of mRNA 5′ cap by decapping enzyme NUDT12 7

Introduction 7

Canonical mRNA 5′ cap 7

Capping enzymes 8

Cap-adjacent methyltransferases 10

Biological functions of m7G cap 12

Emerging functions of cap 1 and cap 2 13

Noncanonical mRNA 5′ caps 15

NAD+ cap 16

mRNA decay 18

mRNA decapping enzymes 20

Bacterial decapping enzymes 21

Eukaryotic decapping enzymes 22

Eukaryotic deNADding enzymes 24

How are cytosolic NAD-RNAs degraded in mammalian cells? 25

Results 27

Discussion 71

Chapter 2. YTHDC2 binds 3′ UTR of RNA targets to regulate meiosis 74

Introduction 74

RNA modification 74

rRNA modification 74

tRNA modification 77
mRNA modification 79
N^1-Methyladenosine (m^1A) 80
5-Methylcytosine (m^5C) 81
N^6-Methyladenosine (m^6A) 82
m^6A writers 84
m^6A erasers 87
m^6A readers 89
YTHDF1-3 90
YTHDC1 91
YTHDC2 92
How YTHDC2 regulates spermatogenesis? 93
Results 94
Discussion 137
References 142

Acknowledgements

First, I would like to express my great gratitude to Ramesh who gave me the opportunity to pursue my PhD study at the University of Geneva and supervised me during the last four years. I still remember the first half year in the lab, I was really struggling to get used to work and live in a foreign city. My English was bad at that time so that I was not frequently chatting or discussing with my colleagues in the lab. Then Ramesh asked me to give presentation every week to try to improve my speaking and listening. I felt like I was in the elementary school, learning language from the teacher. Sometime it is painful but it deserves. Not only my English is improved, but also I learn the skills about how to make a nice presentation, how to give an impressive talk and so on. Beyond that, I am lucky to work on projects that I am interested and at the end of day, I can publish my papers. All of these things I really appreciate.

Second, I would like to thank all the people I work with in the lab. Magda taught me how to purify the protein and prepare grids. Hao and Kuan-ming were working on NUDT12 project with me. These two persons are the only ones that I can speak with Chinese when I feel it difficult to explain in English. They help me a lot in the lab and outside lab. I learn the mouse experiment from Raman who drove the YTHDC2 project that I was working on. I also learn the bioinformatics from David. He carried out the data analyses in my projects and each time when I was confused about the data, he tried to explain and make it clear to me. Mateusz, Elena, Michaela and Kyrylo are the PhD students. We stayed together every day in the lab and discussed almost everything. We went to pub for beer, home for party and mountain for hiking and skiing. I tried so many things with them that I never did before. I think such experiences will be a specific treasure for me. I also want to thank Graeme, Nancy, Joanna, Mark and Manor. Even though we didn't work together for long time, I still got lots of laugh and happiness from you guys. I also feel so grateful for what Fabienne and Pascal have done for me in the lab. They are just so kind and always willing to help when I ask.

Third, I would like to thank all the people from our department secretary, Cécile, Agnes. Fabrice and Veronica. I want to thank my TAC committee who allows me to study in UNIGE, especially Andreas who taught me a lot about cryo-EM and agreed to be my reference when I applied for postdoctoral position. I also want to thank people from Galej lab in Grenoble for their help with EM data collection and processing. Yashar and Christoph provided suggestion and assistance with

data collection and processing in UNIGE. Without the support from canton of Geneva and iGE3, I will not finish my PhD, so thank you for all the funding.

Fourth, I would like to thank my friends in Switzerland and in China. Sailan, Xiaoxu, Jing, Xiujiang, Guoqiang, Yannan, Shumin and Chunchen, they always motivate and inspire me when I talk to them, especially when I feel stressed and down. What a great fortune for me to have them as my friends.

Last but not the least, I would like to thank my family. Without any doubt, they are the engine that gives me power and drives me to where I am now. I am deeply influenced by my grandfather who is always telling me to be optimistic and be kind to people around. My parents, I do not talk too much with them, but they seem to know everything. They can easily notice the change of my mood when I have call with them, sometimes I am not in good attitude, sometimes I just complain, but they are never angry with me. They always encourage me and stand with me for my decision. They are the father and mother like anyone else, but the warmness and love they give to me is incomparable. My aunt and uncle, sister and brothers always make me happy each time I chat with them. I really hope my family could come to Geneva to attend my PhD defense, but I know it is quite difficult during the pandemic. At the moment, I miss them so much and I want to go back to see them as soon as possible. I love you all.我爱你们。

Abstract (English)

In the central dogma, mRNA transfers genetic information from DNA to protein. Such flow of information ensures execution of various cellular events, of which the accuracy and efficiency are conferred by collective regulation via different factors involved in almost every aspect of biological processes in the cell. Chemical modification of biological macromolecules is one of the critical aspects that affect gene expression regulation. For example, the epigenetics, including DNA methylation and histone modifications, impacts the genome accessibility and genome stability. Post-translational modifications of protein, like phosphorylation and ubiquitination, provide another regulatory layer of gene expression via the control of protein homeostasis. As for RNA, even if more than 170 modifications had been identified in different types of RNA, the majority of them remains mysterious in terms of their functions.

Chemical modifications of mRNA vary in regard to their structures and abundances. Mature mRNA has m^7G cap at its 5′ end and poly(A) tail at 3′ end. m^7G cap is a critical feature to protect mRNA from degradation and initiate translation via the recognition by eIF4E. However, in the last decade, the concept of alternative caps had emerged with the identification of novel caps present in mRNA 5′ end in both prokaryotic and eukaryotic organisms. Nicotinamide adenine dinucleotide (NAD) is a representative alternative cap, which can be incorporated into mRNA by RNA polymerase during transcription, generating NAD-capped RNA. But the biological consequence of NAD-RNAs is still poorly understood. During my PhD, we identified NUDT12 as a cytosolic NAD-RNA decapping enzyme. NUDT12 is active only as homodimers. Crystal structure of NUDT12 homodimer explains each monomer contributes to creation of the two functional catalytic pockets. NUDT12 acts on both m^7G-capped and NAD-capped RNA with preference on the later. Moreover, we identified an ~600 kDa dodecamer complex formed by bleomycin hydrolase (BLMH) and NUDT12, with BLMH being required for localization of NUDT12 to a few discrete cytoplasmic granules that are distinct from P-bodies. Both proteins downregulate gene expression when artificially tethered to a reporter RNA in vivo. Furthermore, loss of *Nudt12* results in a significant upregulation of circadian clock transcripts in mouse liver. Overall, our study points to a physiological role for NUDT12 in the cytosolic surveillance of NAD-RNAs.

Besides m^7G cap, mRNA bears internal modifications as well, such as m^1A, m^5C and m^6A. Among these modifications, m^6A is the most abundant one. m^6A influences mRNA splicing, export and stability. It is essential to maintain mouse embryonic development. The regulatory

functions of m^6A rely upon writers (RNA methyltransferase), erasers (RNA demethylase) and readers (m^6A recognition protein). YTHDC2 is a germ cell-specific m^6A reader in mammals. Previous studies revealed that YTHDC2 was required for mouse fertility because *Ythdc2* knockout mice were infertile. Such phenotype was a result of defective transition from mitosis to meiosis. Immunoprecipitation mass spectrometry indicated XRN1, an exoribonuclease, was associated with YTHDC2. Nevertheless, the function of this complex and the underlying molecular mechanism of YTHDC2 remain elusive. Here we identify U-rich motifs as binding sites of YTHDC2 on 3′ UTRs of testicular RNA targets. Although its YTH domain is an m^6A-binder in vitro, the YTH point mutant mice are fertile. Significantly, the 3′→5′ RNA helicase activity of YTHDC2 is essential for fertility. Biochemical studies decipher that the weak helicase activity of YTHDC2 is enhanced by its interaction with the XRN1. Single-cell transcriptomics indicate that Ythdc2 mutant mitotic germ cells transition into early stage of meiosis but accumulate a transcriptome with mixed mitotic/meiotic identity that leads to failure of progression into late stage of meiosis. Finally, our demonstration that Ythdc2 mutant zebrafish are infertile highlights its conserved role in animal germ cell development.

Abstract (French)

Selon le dogme central, l'ARN messager (ARNm) transfère l'information génétique de l'ADN aux protéines. Ce flux d'informations assure l'exécution de divers événements cellulaires, dont la précision et l'efficacité sont conférées, par régulation collective, via différents facteurs impliqués dans presque tous les aspects des processus biologiques de la cellule. La modification chimique des macromolécules biologiques est l'un des aspects critique qui affecte la régulation de l'expression génétique. Par exemple, l'épigénétique, qui comprend la méthylation de l'ADN et les modifications des histones, a un impact sur l'accessibilité et la stabilité du génome. Les modifications post-traductionnelles des protéines, comme la phosphorylation et l'ubiquitination, fournissent une autre couche de régulation de l'expression génétique via le contrôle de l'homéostasie des protéines. Quant à l'ARN, même si plus de 170 modifications ont été identifiées dans différents types d'ARN, la majorité d'entre elles restent mystérieuses quant à leurs fonctions.

Les modifications chimiques de l'ARNm varient en fonction de leurs structures et de leurs abondances. L'ARNm mature possède une coiffe m7G à son extrémité 5′ et une queue poly(A) à son extrémité 3′. La coiffe m7G est une caractéristique essentielle pour protéger l'ARNm de la dégradation et initier la traduction via la reconnaissance par eIF4E. Cependant, au cours de la dernière décennie, le concept de capsule alternative a émergé avec l'identification de nouvelles capsules présentes à l'extrémité 5′ de l'ARNm dans les organismes procaryotes et eucaryotes. Le nicotinamide adénine dinucléotide (NAD) est une coiffe alternative représentative, qui peut être incorporée dans l'ARNm par l'ARN polymérase pendant la transcription, générant un ARN coiffé de NAD. Mais la conséquence biologique des ARN à NAD est encore mal comprise. Au cours de mon doctorat, nous avons identifié NUDT12 comme une enzyme cytosolique de décapsulation des ARN à NAD. La protéine NUDT12 est active uniquement sous forme d'homodimères. La structure cristalline de l'homodimère de NUDT12 explique que chaque monomère contribue à la création de deux poches catalytiques fonctionnelles. NUDT12 agit à la fois sur l'ARN coiffé de m7G et sur celui coiffé de NAD, avec une préférence pour ce dernier. De plus, nous avons identifié un complexe dodécamère de ~600 kDa formé par la bléomycine hydrolase (BLMH) et NUDT12, la BLMH étant requise pour la localisation de NUDT12 chez des granules cytoplasmiques discrets distincts des P-bodies. Les deux protéines régulent à la baisse l'expression des gènes lorsqu'elles sont artificiellement attachées à un ARN rapporteur in vivo. En outre, la perte de Nudt12 entraîne une importante régulation à la hausse de la transcription des gènes de l'horloge circadienne dans

le foie de la souris. Globalement, notre étude indique un rôle physiologique pour NUDT12 dans la surveillance cytosolique des ARN à NAD.

Outre la coiffe m7G, l'ARNm porte également des modifications internes, telles que m1A, m5C et m6A. Parmi ces modifications, m6A est la plus abondante. Celle-ci influence l'épissage, l'exportation et la stabilité de l'ARNm. Elle est essentielle au maintien du développement embryonnaire de la souris. Les fonctions régulatrices de m6A reposent sur des enzymes qui écrivent (ARN méthyltransférase), qui effacent (ARN déméthylase) et qui lisent (protéine de reconnaissance m6A). YTHDC2 est une enzyme qui lit les m6A spécifiques des cellules germinales chez les mammifères. Des études antérieures ont révélé que YTHDC2 était nécessaire à la fertilité des souris car les souris knockout Ythdc2 étaient infertiles. Ce phénotype est le résultat d'une transition défectueuse de la mitose et de la méiose. La spectrométrie de masse par immunoprécipitation a indiqué que XRN1, une exoribonucléase, était associée à YTHDC2. Néanmoins, la fonction de ce complexe et le mécanisme moléculaire sous-jacent de YTHDC2 restent insaisissables. Nous avons identifié ici des motifs riches en U comme sites de liaison de YTHDC2 sur les 3′ UTR des cibles ARN testiculaires. Bien que son domaine YTH soit un liant m6A in vitro, les souris mutantes par mutation ponctuelle de YTH sont fertiles. De manière significative, l'activité hélicase ARN 3′→5′ de YTHDC2 est essentielle pour la fertilité. Des études biochimiques ont découvert que la faible activité hélicase de YTHDC2 est renforcée par son interaction avec XRN1. La transcriptomique unicellulaire indique que les cellules germinales mitotiques mutantes Ythdc2 passent au stade précoce de la méiose mais accumulent un transcriptome avec une identité mixte mitotique/méiotique qui conduit à l'échec de la progression vers le stade tardif de la méiose. Enfin, notre démonstration que le poisson zèbre mutant Ythdc2 est infertile souligne son rôle conservé dans le développement des cellules germinales d’autres espèces animales.

Chapter 1. Regulation of mRNA 5′ cap by decapping enzyme NUDT12

Introduction

In 1953, James Watson and Francis Crick published a paper in *Nature*, reporting the first structure of DNA (deoxyribonucleic acid) that two long strands of DNA form a double helix (Watson and Crick, 1953). From that moment, the door of molecular biology was opened and afterwards numerous research and discoveries established the foundation of it, namely, the central dogma. According to the central dogma, DNA which is composed of four nucleotides A, T, C, G encodes genetic information, and such information is transferred by messenger RNA (mRNA) which is transcribed by using DNA as the template from nucleus to cytoplasm. Finally, mRNA is translated (decoded) into protein to exert specific biological functions. This genetic informational flow is fast, complicated and prone to errors, which means in this process any error must be detected in time and eliminated immediately after to guard the integrity of genetics.

One of the critical controls of gene expression is occurring at the transcriptional layer, that is, the regulation of mRNA metabolism. mRNA metabolism covers the lifetime of an mRNA from birth to death: transcription of DNA to pre-mRNA, mRNA maturation, nuclear export, cellular delivery to specific locations, translation into proteins and decay. Each step is collectively regulated and seamlessly connected with other steps by various protein factors, including those enzymes which specifically act on mRNA 5′ terminus.

Canonical mRNA 5′ cap

In eukaryotic cells, when DNA is transcribed by RNA polymerase II to generate nascent mRNA in nucleus, the first event occurred to such nascent mRNA is the formation of 5′ cap. Eukaryotic mRNAs are capped by N^7-methyl guanosine (m^7G) that is linked to the first transcribed nucleotide of nascent transcripts through an inverted 5′ to 5′ triphosphate bridge (Wei et al., 1975b). Addition of this m^7G cap requires 3 steps of chemical reactions. First, RNA triphosphatase (TPase) removes the γ-phosphate from the 5′ triphosphate to generate 5′ diphosphorylated RNA as soon as 25 – 30 nucleotides are incorporated into the nascent transcripts. Second, RNA guanylyltransferase (GTase) transfers a GMP group from GTP to the 5′ end of that diphosphorylated RNA to generate G-ppp-RNA. Finally in the third step, guanine-N^7 methyltransferase (guanine-N7 MTase) deposits a methyl group at the N7 position of the guanine cap to form the m^7G cap structure (Ramanathan et al., 2016). This m^7GpppN (N is the first transcribed nucleotide) cap is also called cap 0. In most

cases, mRNA carrying cap 0 can be further methylated at the 2′-hydroxyl (2′-O) of ribose of the first transcribed nucleotide. This m^7GpppNm cap is called cap 1. In particular, if first transcribed nucleotide is adenosine, cap 1 can undergo further methylation at the N6 position to generate m^7Gpppm^{6}Am structure (Wei et al., 1975a). Moreover, certain mRNAs have the 2′-O-methylation at the second transcribed nucleotide, resulting in formation of cap 2 structure: m^7GpppNmpNm. Taken together, cap 0, cap 1 and cap 2 represent the canonical 5′ cap structure of mRNA in mammalian cells (Figure 1.1) (Reddy et al., 1992).

Capping enzymes

RNA triphosphatases catalyze the first reaction of m^7G cap biogenesis. These enzymes vary in different organisms and can be divided into two families: metal-dependent RNA triphosphatases of lower eukaryotes such as fungi and protozoa, and metal-independent RNA triphosphatases of higher eukaryotes, like metazoan and plants (Ghosh and Lima, 2010; Gu and Lima, 2005). In yeast,

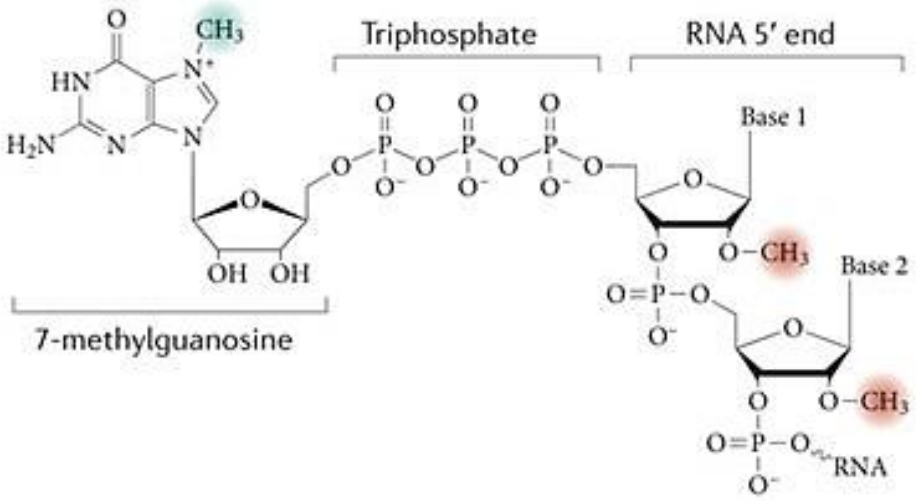


Fig 1.1 Chemical structure of eukaryotic mRNA 5′ cap (Decroly et al., 2011)

Cet1 is the RNA TPase that cleaves the γ-phosphate from pppRNA (Ho et al., 1998). Crystal structure of Cet1 indicates in the catalytic site several acidic residues that coordinate two metal ions interact with the pppRNA triphosphate moiety (Bisaillon and Shuman, 2001a, b). In contrast to Cet1, RNA TPase of metazoan hydrolyzes the phosphate independently of metal ion. These enzymes contain a conserved HCxxxxxR(S/T) motif in their P-loop from which the cysteine attacks the 5′ γ-phosphate to release ppRNA product. According to the structure of mouse RNA TPase Mce1, only pppRNA can be accessible to the positively charged catalytic pocket (Changela et al., 2001; Takagi et al., 1997).

RNA guanylyltransferases catalyze the second step that such enzymes transfer a GMP moiety from GTP to the 5′ diphosphate of RNA TPase-processed nascent transcripts, forming a capped 5′ end. This process consists of two rounds of reaction that a covalent lysyl-N-GMP intermediate is produced in the first step then the GMP is transferred to the 5′ diphosphate in the second step (Shuman and Hurwitz, 1981; Shuman et al., 1994). RNA GTases are conserved from virus to mammal and structurally they share the similar domain organization: the N-terminal nucleotidyltransferase (NTase) domain and the C-terminal Oligonucleotide/oligosaccharide binding (OB) domain. The NTase domain carries a nucleotide binding pocket where the enzymatic reaction is taken place and OB domain contributes to that catalytic reaction by adopting an "open-closed" conformational change with NTase domain (Fabrega et al., 2003; Hakansson et al., 1997).

RNA guanine-N^7 methyltransferases catalyze last step to generate m^7G cap by adding the methyl group to the N7 guanine of G(5′)ppp(5′)RNA. Mechanistically, RNA guanine-N7 MTase binds to the methyl donor S-adenosyl methionine (SAM) via one of its ligand binding sites then catalyzes the methyl transfer to guanine cap which is bound to the other ligand binding site. These two binding sites are located in a deep cleft in which it forms a classic Rossmann fold-like structure (Byszewska et al., 2014; Fabrega et al., 2004). Structural studies highlight that optimizing proximity and orientation of these two ligand binding sites is critical to the efficient methyl group transfer (Fabrega et al., 2004; Zheng et al., 2006).

Viral genome encodes its own enzymes required for cap genesis. These enzymes are versatile and some of them are multifunctional. For example, Vaccinia RNA capping enzyme forms a heterodimer which is composed of two subunits: the large subunit D1 and the small subunit D12. D1 possesses all the essential enzymatic activities for cap generation while as a regulatory module, D12 significantly stimulates D1 activities (Higman et al., 1994; Mao and Shuman, 1994). According to its structure, Vaccinia capping enzyme coordinates RNA TPase, GTase and guanine-N7 MTase activities within discrete modules in its protein architecture but links these three activities tightly to facilitates the m^7G cap formation (Figure 1.2) (De la Pena et al., 2007). Recently, a novel coronavirus, SARS-COV-2 which causes the COVID-19 pandemic, is shown to encode its capping enzyme Nsp14 to install m^7G cap to viral mRNAs (Chen et al., 2009a). Interestingly, SARS-COV-2 Nsp14 also harbors 3′ to 5′ exoribonuclease activity through its N-terminal ExoN domain. This ExoN domain is able to remove mismatched nucleotides to preserve the fidelity of coronavirus RNA (Liu et al., 2021).

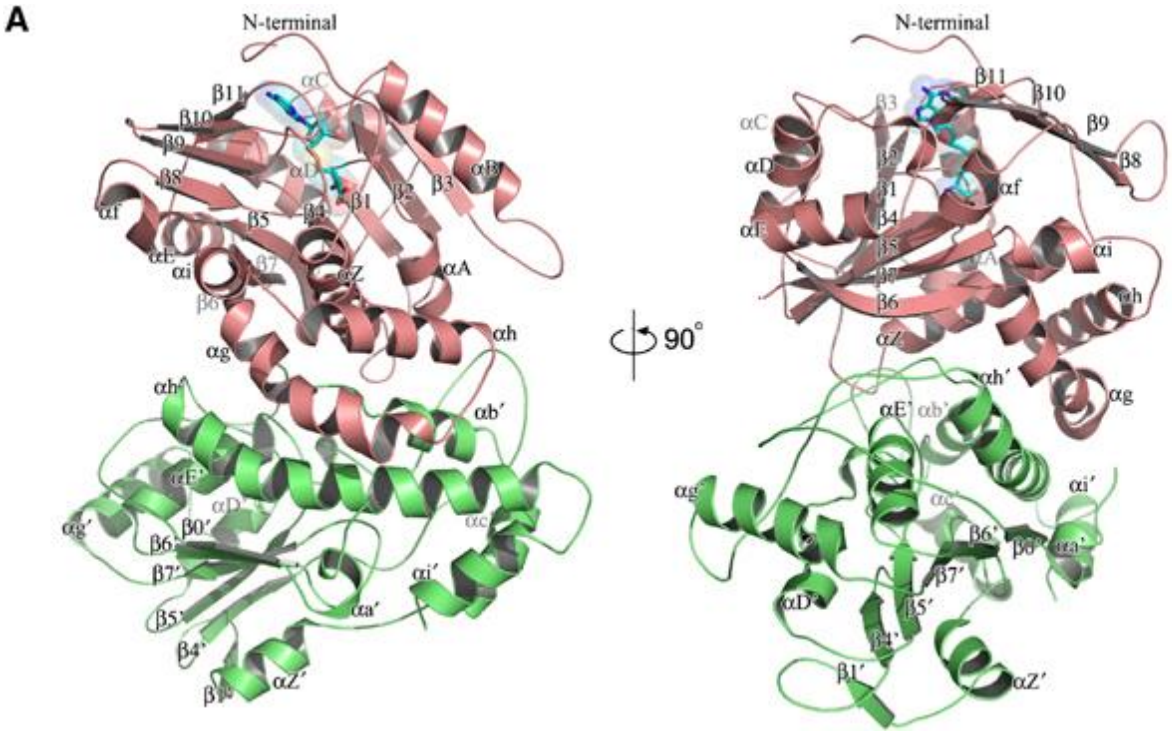


Fig 1.2. Structure of Vaccinia capping enzyme in complex with AdoHcy (PDB 2VDW). D1 subunit (pink) interacts with D12 subunit (green) to catalyze the m^7G cap generation.

Cap-adjacent methyltransferases

m^7G cap (cap 0) generation is the earliest event happening to 5′ end of nascent mRNA in eukaryotic cells. In mammals, almost all the mRNAs harbor 2′ methylation of ribose at the first transcribed nucleotide ($m^7GpppNm$, cap 1). This cap-specific ribose methylation is catalyzed by CMTR1. Initially named as ISG95 (interferon stimulated gene 95), this protein was reported to function as an RNA methyltransferase and its C-terminal WW domain interacted with C-terminal domain (CTD) of RNA polymerase II (Haline-Vaz et al., 2008). After that in 2010, the same protein was characterized as cap 1 2′-O-methyltransferase in human cells, hence, it was named as hMTR1 or CMTR1 (Belanger et al., 2010). Structural analysis of CMTR1 catalytic domain indicates that SAM and m^7G cap are located into the deep pocket where the first nucleotide adjacent to the cap is accommodated at the bend of the substrate, ready to be methylated (Smietanski et al., 2014). Functional studies uncover that CMTR1 can promote ribosomal protein and histone gene expression (Liang et al., 2022) and in Drosophila it directs mRNA to synapses (Haussmann et al., 2022).

Once the first transcribed nucleotide is adenosine, another RNA methyltransferase, called CAPAM or PCIF1, can further methylate N6 adenine to form m^7Gpppm^6Am 5′ end (Figure 1.3) (Akichika et al., 2019; Boulias et al., 2019; Sendinc et al., 2019). It is still not clear what is the

function of cap adjacent m^6Am since one group claimed that this modification has little effect on mRNA stability (Sendinc et al., 2019), whereas the other group reported loss of m^6Am reduces stability of certain m^6Am-containing mRNAs (Boulias et al., 2019). It seems that PCIF1 is not essential for mammals, at least for mouse because PCIF1 KO mice are viable and fertile, without any aberrant phenotype except the reduced body weight (Pandey et al., 2020). Transcripts are either upregulated or downregulated in different tissues and ribosome profiling analysis suggests removal of PCIF1 has no direct influence on mRNA translation (Pandey et al., 2020).

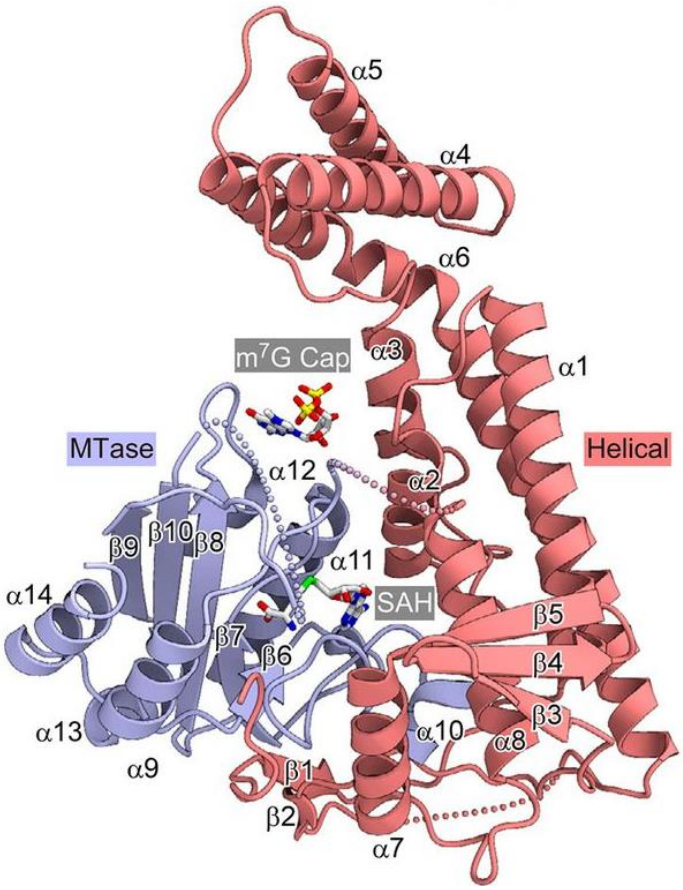


Fig 1.3. Structure of human CAPAM in complex with SAH and m^7G cap.

In some cases, a subset of mRNAs harbor 2′-O-methylation at both first and second transcribed nucleotide, forming cap 2. This methylation occurring at second transcribed nucleotide is catalyzed by CMTR2. Even if second 2′-O-methylation always appear with the cap 1 in mRNAs in vivo, in vitro methylation assay indicates that CMTR2 activity is independent of additional methylations (Werner et al., 2011). It is believed that CMTR1 functions in nucleus but regarding where the cap 2 methylation is taken place, it is still unclear. Physiological functions of CMTR2 remain enigmatic as well.

Biological functions of m^7G cap

As a prevalent 5′ end modification of mRNAs, m^7G cap is involved in versatile mRNA processes to regulate cellular biological functions. m^7G cap is co-transcriptionally added to nascent transcripts which need to be further processed to become mature mRNAs, including splicing and 3′ end modification. m^7G cap is a key marker to initiate pre-mRNA splicing (Inoue et al., 1989; Konarska et al., 1984; Ohno et al., 1987) and this is mediated by recruiting and binding to the nuclear cap-binding complex (CBC). CBC is composed of two proteins, CBC80 and CBC20. During transcription, once m^7G cap is installed at the 5′ end of mRNA precursor, CBC80 and CBC 20 forms a heterodimer to bind to m^7G cap. Subsequently, the cap-bound CBC interacts with U4/U6.U5 tri-snRNP to trigger splicing (Pabis et al., 2013). Not only is CBC involved in the assembly of spliceosome, but also it participates in pre-mRNA 3′ end processing, implicating a special role of CBC in coupling 5′ modification with 3′ processing to regulate pre-mRNA maturation (Flaherty et al., 1997). Except the relation with pre-mRNA processing via CBC, m^7G cap is also shown to have an influence on transcription termination and exosome-mediated degradation (Andersen et al., 2013). These findings reveal an essential function of m^7G cap structure in leading mRNA splicing and 3′ end formation through recruiting different protein complexes.

After splicing and polyadenylation, m^7G capped mature mRNA requires the transport from nucleus to cytoplasm to exert protein synthesis in ribosome. Many studies have reported that mRNA export relies on the interaction between CBC and nuclear export factors, for instance, REF (RNA export factor) (Nojima et al., 2007). REF is a component of transcription export complex TREX and it interacts with cap-bound CBC in a splicing-dependent manner (Cheng et al., 2006). These protein factors directly or indirectly communicate with nuclear pore complex to guide mRNA delivery to cytoplasm.

In cytoplasm, the mission of mRNA is to be translated by ribosome to synthesize nascent peptide which will further be folded and modified to form functional protein. The majority of mRNA translation is cap-dependent, which means the recognition of m^7G cap is a critical step to initiate translation. In eukaryotic cells, CBC stays attached to the mRNA 5′ cap in ribosome and recruits translation initiation factors such as eIF4G and eIF4A to start first round translation. This process needs the involvement of many other protein factors and ribosomal subunits as well (Chiu et al., 2004; Choe et al., 2012; Choe et al., 2014; Fortes et al., 2000). After the pioneer round of

translation, eIF4E will bind to the m^7G cap to replace CBC then eIF4F complex is recruited to the 5′ end of mRNA to initiate steady-state rounds of translation (Maquat et al., 2010). In eIF4F complex, eIF4G binds to the cap while interacts with poly(A) binding protein PABP1 and such interplays create a pseudo-circular structure of translated mRNA (Figure 1.4). This unique structure is implied that it provides accessibility of ribosome to the entire mRNA to ensure the intactness of translation (Preiss and Hentze, 1998, 1999; Smith et al., 2014). Taken together, as a specific anchor, m^7G cap is bound by various proteins complex in both nuclear and cytoplasm to drive mRNA splicing, export and translation. This cap also protects mRNA from 5′ to 3′ exoribonuclease-mediated decay.

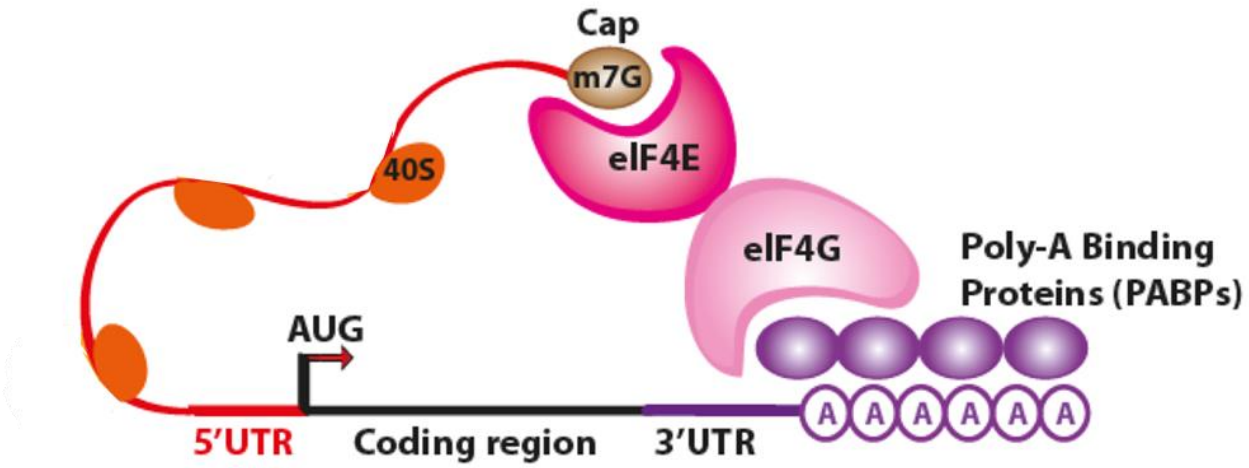


Fig 1.4. Schematic representation of cap-dependent translation. eIF4E, a cap binding protein interacts with eIF4G which further interacts with PABP1 to form a pseudo-circular structure of mRNA during translation, adapted from (Da Sacco and Masotti, 2012).

Emerging functions of cap 1 and cap 2

Although cap-adjacent 2′-O-methylations (here referred as cap 1 and cap 2) were identified long time ago, physiological functions of these modifications remain poorly described, particularly for cap 2. For small nuclear RNAs (snRNAs) such U1, U2, U4 and U5, 2′-O-methylation is present at the first transcribed nucleotide and this cap 1 structure is indispensable to maintain activities of snRNAs in splicing (Will and Luhrmann, 2011). Cap 1 is also verified to stimulate the ribosome binding and translation for viral mRNAs (Muthukrishnan et al., 1976; Muthukrishnan et al., 1978) as well as maternal mRNAs (Caldwell and Emerson, 1985; Kuge et al., 1998). Besides its regulation in mRNA splicing and translation, in recent years, cap 1 has emerged as a pivotal marker for distinguishing ‘self’ and ‘non-self’ RNA during innate immune response (Daffis et al., 2010). When pathogens like RNA virus invade cells, cytosolic protein factors like RIG-I (Retinoic Acid

Inducible Gene-I) and MAD5 are able to trigger the production of type I interferon to inhibit viral propagation via sensing viral RNA 2′-O-methylation (De la Pena et al., 2007; Wu et al., 2013). In other words, cap 1 functions as a negative regulator to prevent aberrant activation of RIG-I/MAD5 (Figure 1.5). This strategy can be utilized by virus that encodes its 2′-O-methyltransferase to evade the cellular immune surveillance, suggesting the role of cap 1 in viral evolution. Note that a recent case for such viral immune evasion is the SARS-COV-2 that it encodes its own 2′-O-methyltransferase NSP16 (Viswanathan et al., 2020).

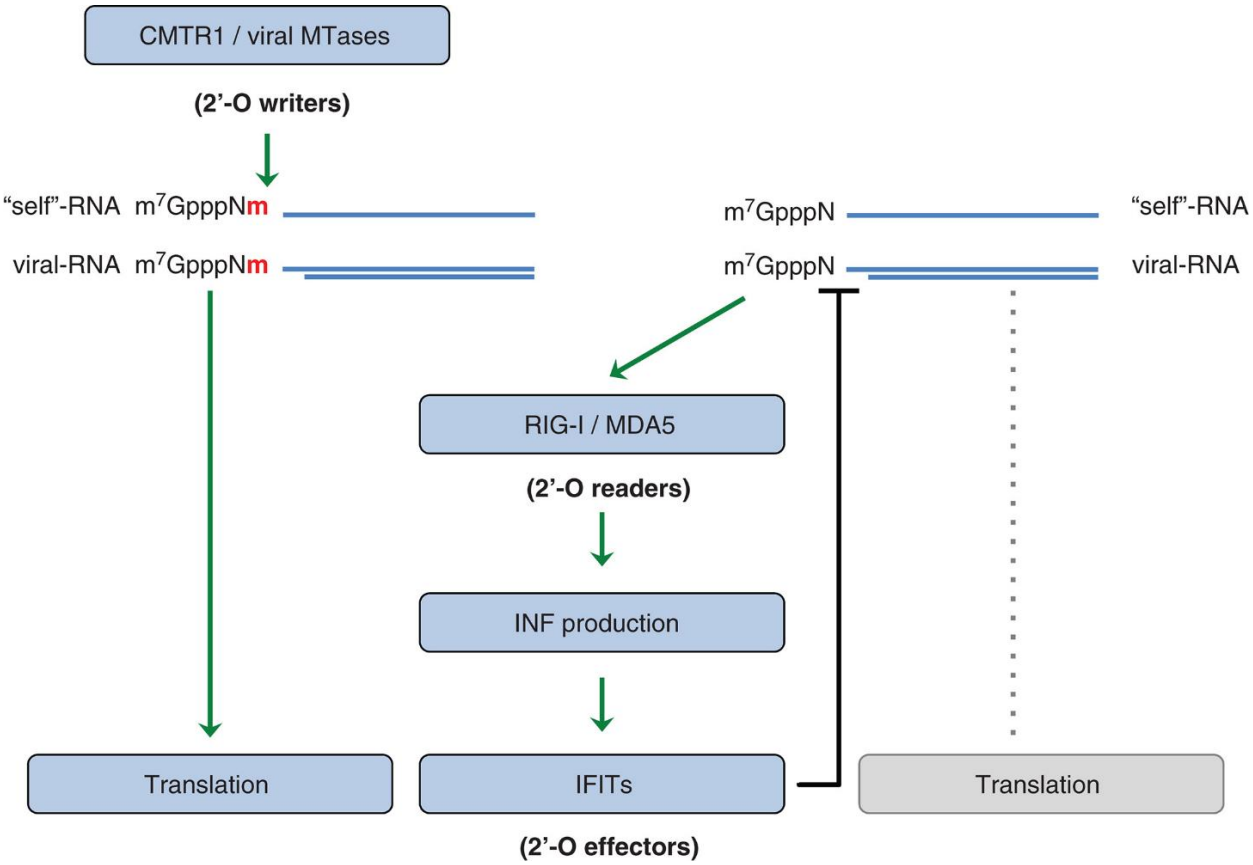


Fig 1.5. 2′-O-methylation (cap 1) in innate immune response. Cap 1 prevents the recognition by RIG-I/MAD5 which otherwise induce type I interferon (INF) production and expression of IFN-induced proteins (IFIT) (Inesta-Vaquera and Cowling, 2017).

Regarding to cap 2 modification, similarly to cap 1, it is also reported that the second transcribed nucleotide contributes to inhibition of IFIT1 action on cellular RNA, particularly on susceptible RNA sequences and at high IFIT1 concentrations (Abbas et al., 2017), indicating the potential regulation of cap 2 in innate immunity. Beyond this, whether cap 2 is related with other physiological functions is still unclear, especially considering its relatively low abundance compared with cap 1. More efforts should be made to precisely quantify cap 2 in mRNAs in future.

Noncanonical mRNA 5′ caps

m^7G cap is supposed to be the only cap modification existing in eukaryotic mRNA 5′ end for quite long time. However, in the last decades, discoveries of many other chemical structures present at both prokaryotic and eukaryotic mRNA 5′ ends diverse the concept of mRNA cap modification. These noncanonical or alternative caps represent a parallel regulation of mRNA 5′ end with classic m^7G cap.

So far, the identified noncanonical caps include metabolic cofactors, such as NAD^+/NADH (nicotinamide adenine dinucleotide), CoA (coenzyme A), FAD (flavin adenine dinucleotide), cell wall precursors UDP-GlcNAc (uridine diphosphate N-acetylglucosamine), dinucleotides polyphosphates, ADP-ribose and potentially other nucleoside derivatives (Bird et al., 2016; Cahova et al., 2015; Hudecek et al., 2020; Luciano and Belasco, 2020; Luciano et al., 2019; Wang et al., 2019). In contrast to m^7G cap which is deposited by dedicated enzymes, at least in part, installation of those noncanonical caps is proposed to depend on RNA polymerase II. During transcription, RNA polymerase II can incorporate abundant cellular metabolites into nascent transcripts as the first transcribed nucleotide to form the 5′ end (Figure 1.6). Nevertheless, RNA polymerase II is not the only enzyme which can add the noncanonical caps because in mammalian cells, intronic small nucleolar RNAs (snoRNAs) and small Cajal body RNAs (scaRNAs) are found to be NAD-capped and this capping process is independent of RNA polymerase II action (Jiao et al., 2017).

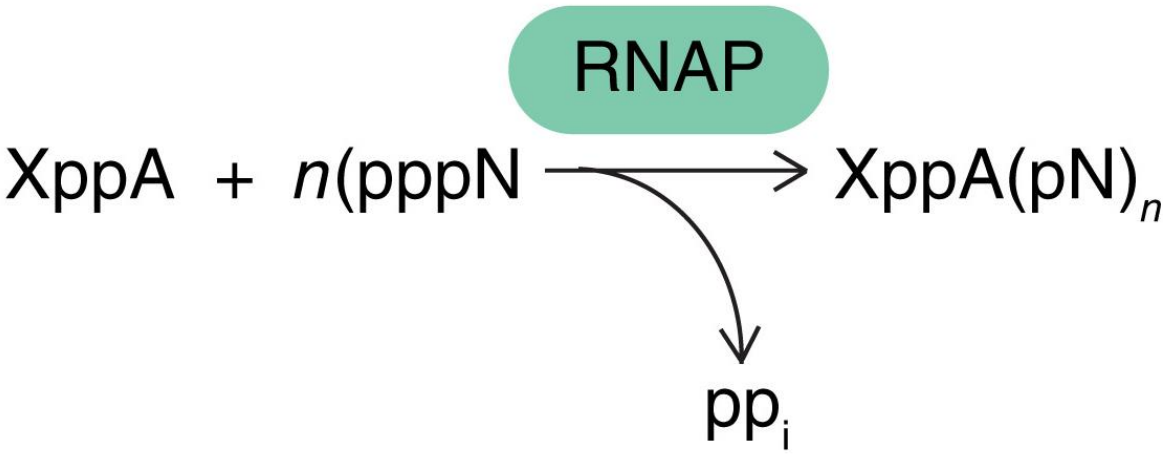


Fig 1.6. Schematic representation of noncanonical capping. RNA polymerase (RNAP) creates a capped transcript by using the small molecule as initiating substrate instead of canonical nucleoside triphosphate. This small molecule contains a nucleoside diphosphate (here adenosine diphosphate), and another moiety X bound to β-phosphate. RNAP uses incoming nucleoside triphosphate substrates (pppN) to elongate RNA chain, and moiety X is retained as noncanonical cap (Julius and Yuzenkova, 2019).

NAD^+ cap

Nicotinamide adenine dinucleotide (NAD) is a cofactor required for many cellular enzymes involved in various biological processes. In response to different environment, it changes between oxidized form (NAD^+) and reduced form (NADH). As a redox coenzyme, NAD^+ is reduced to NADH during beta oxidation, glycolysis, and the citric acid cycle. Such NADH is then transferred to mitochondria where it is oxidized in turn by electron transport chain through oxidative phosphorylation (Rich, 2003). Beyond this, NAD^+ is served as a co-substrate in ADP-ribosylation by various ADP-ribosyltransferases including poly (ADP-ribose) polymerase (PARPs), NAD-dependent DNA ligases and ADP-ribosyl cyclases (Lin, 2007). Recent excitement about NAD^+ is due to the discovery that this molecule can be added to the 5′ end of mRNA as an alternative cap (Figure 1.7).

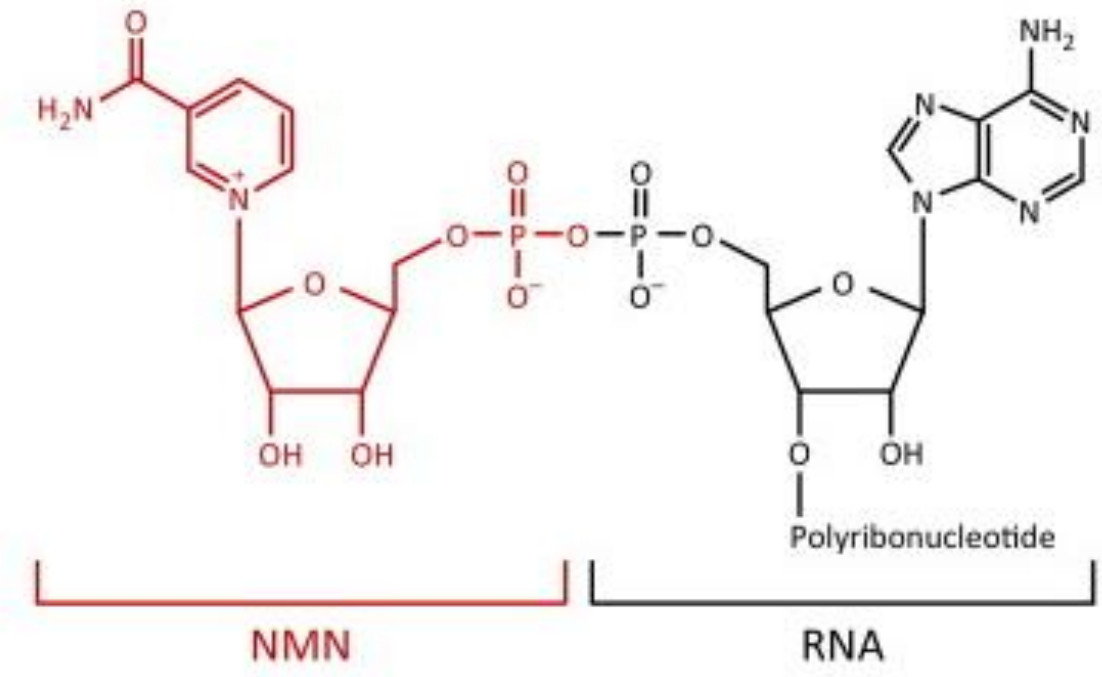


Fig 1.7. Chemical structure of NAD-RNA (Luciano and Belasco, 2015)

In 2009, David Liu's group screened small molecule-RNA conjugates and found that NAD-linked RNA is present in *Escherichia coli* and *Streptomyces venezuelae* (Chen et al., 2009b). However, due to the limit of liquid chromatography-mass spectrometry (LC-MS), they could not identify which RNA carried NAD^+ cap. This remaining question was not addressed until another group established a chemical approach to sequence NAD-capped RNAs in bacteria in 2015 (Cahova et al., 2015). In this paper, this group described a method that 5′ NAD-capped RNAs can

be biotin-labelled after Adenosine diphosphate-ribosylcyclase (ADPRC) treatment. The biotin-labelled RNAs will be captured by streptavidin beads and further subjected to next generation sequencing. Using this so-called NAD captureSeq (Figure 1.8), they discovered specific regulatory small RNAs (sRNAs) as well as 5′-terminal fragments of mRNAs coding for proteins involved in metabolic pathways, stress response control, and for poorly characterized proteins, are decorated with NAD^+ at the 5′ ends. For instance, sRNA RNAl is the most frequently NADylated RNA in *E. Coli*. Moreover, recently NAD^+ cap is identified in bacterial mRNAs as well. Those related genes encode proteins involved in DNA replication, sporulation, and oxidation/reduction processes (Frindert et al., 2018).

Not only in prokaryotic cells, but also in eukaryotic cells are NAD-RNAs discovered afterwards. Using NAD captureSeq, NAD^+ cap is found on subsets of nuclear and mitochondrial encoded mRNAs in *Saccharomyces cerevisiae* (Walters et al., 2017). In human cells, mainly small nuclear (snRNA), and small nucleolar (snoRNA) RNAs are found to bear 5′ NAD^+ cap (Jiao et al., 2017). Regarding the abundance, (0.6 ± 0.1) % mouse liver mRNA and (1.3 ± 0.3) % mouse kidney mRNA is 5′ NADylated, respectively (Wang et al., 2019).

Biological functions of NAD^+ cap are not well understood, but probably vary in different species. The most studied in vivo role of NAD^+ cap is its influence on RNA stability and this influence is opposite for bacteria and eukaryotes. NAD^+ cap is resistant to bacterial RppH-directed RNA decay in vitro (Cahova et al., 2015). While in eukaryotic cells, NAD-capped RNAs are rapidly degraded (Jiao et al., 2017). Another potential effect of NAD^+ cap is on mRNA translation. However, so far, no clear evidence has been found to support the hypothesis that NAD^+ cap could affect translation since neither the prokaryotic nor the eukaryotic translation apparatus can recognize the NAD^+ cap. Results from an in vivo experiment where luciferase reporter gene was modified by either 5′ triphosphate (uncapped) or 5′ NAD^+ indicated that NAD-capped reporter shared similar translation level with uncapped reporter (Jiao et al., 2017).

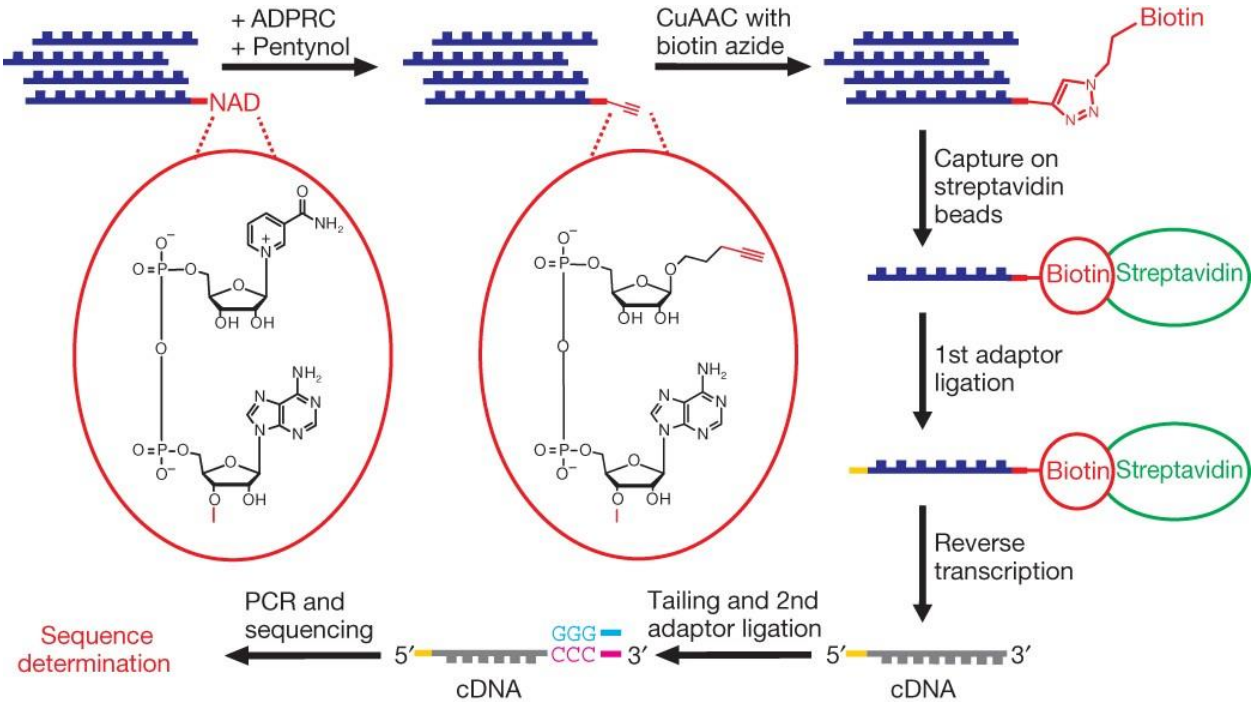


Fig 1.8. Schematic workflow of NAD captureSeq. ADPAC, adenosine diphosphate-ribosylcyclase; CuAAC, copper-catalysed azide-alkyne cycloaddition; pentynol, 4-pentyn-1-ol (Cahova et al., 2015).

mRNA decay

The dynamic balance of mRNA level in cytoplasm is important to regulate gene expression, which means on one hand, before protein translation starts, mRNA needs to be protected from cytosolic degradation. On the other hand, once translation process is completed, mRNA decay is executed to silence gene expression. In eukaryotic cells, there are two machineries acting on mRNA turnover: 5′-3′ RNA decay and 3′-5′ RNA decay. Numerous studies focused on 3′-5′ RNA decay pathway have provided a comprehensive understanding of how mRNA is degraded from its 3′ end: the RNA exosome, a multi-subunit, processive protein complex that recognizes mRNA 3′ terminus and degrades the targets. This review (Zinder and Lima, 2017) can be referred to explain how the catalytic activity of exosome is regulated and the structural basis of exosome-mediated RNA decay.

The 5′-3′ RNA decay pathway includes certain consecutive events that facilitate mRNA degradation in the 5′ to 3′ direction (Moore, 2005; Parker, 2012). First step of 5′-3′ RNA decay is to shorten poly(A) tail, or deadenylation, which leads to reduced occupation of PABP (poly(A) binding protein) and suppresses translation in the end. Two multisubunit protein complexes, PAN2-PAN3 and CCR4-NOT, carry out deadenylation in eukaryotes. The former complex initiates poly(A) tail trimming on PABP-bound mRNA and the later complex further accomplishes entire process of deadenylation (Collart, 2016; Wolf and Passmore, 2014). Shortly after

deadenylation, mRNA 5′ m^7G cap will be removed (decapping). This is the second step of 5′-3′ RNA decay, in which cytoplasmic decapping enzymes account for the removal of m^7G cap. Recently, it is proposed that the Pat-Lsm complex is a central regulator to bridge 3′ deadenylation to 5′ decapping. For example, in yeast, decapping coactivator complex Pat1-Lsm1-7 binds short oligo(A) tails and interacts with decapping enzyme Dcp2 simultaneously to promote mRNA degradation (Charenton et al., 2017; Chowdhury et al., 2007). Once 5′ m^7G cap is removed, such uncapped mRNA is subjected to 5′ to 3′ exoribonuclease-mediated decay (5′ exonucleolysis) to convert RNA body into nucleotides (Figure 1.9). XRN1 is the major exoribonuclease that degrades cellular 5′ monophosphorylated RNAs to finalize last step of 5′-3′ RNA decay (Jones et al., 2012).

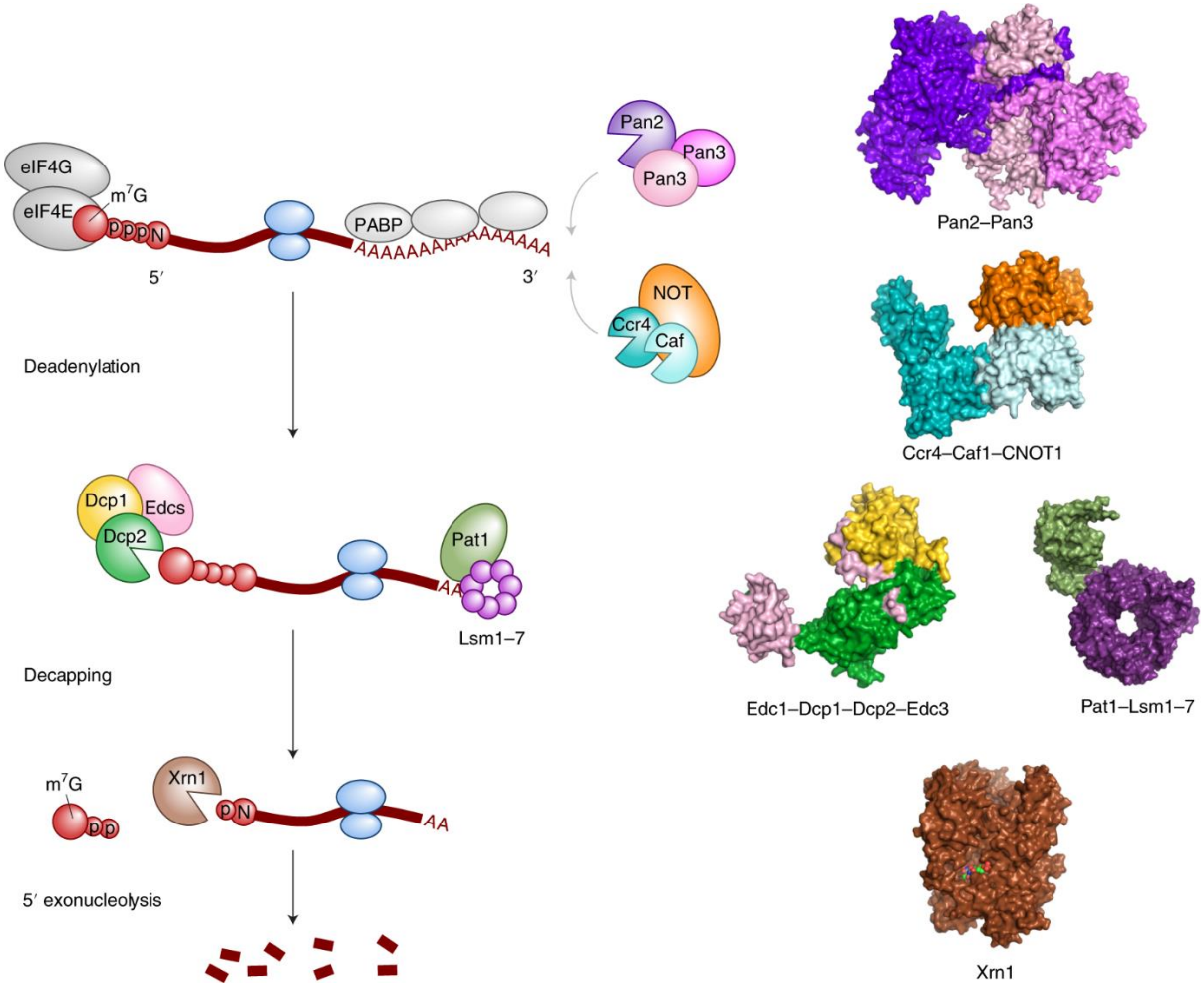


Fig 1.9. Schematic representation of major steps in the eukaryotic 5′-3′ RNA decay pathway. For each step, the corresponding structures of protein complexes are shown on the right. Pan2-Pan3, PDB: 4XR7; Ccr4-Caf1-CNOT1, PDB: 4B8C; Edc1-Dcp1-Dcp2-Edc3, PDB: 6AM0; Pat1-Lsm1-7, PDB: 4C8Q and Xrn1, PDB: 2Y35 (Mugridge et al., 2018).

In bacteria, two basic pathways are involved in mRNA decay: direct access pathway and 5′ end dependent pathway (Figure 1.10). In the direct access pathway, mRNA is cleaved by endonuclease to produce two RNA fragments: one with 5′ triphosphates and the other with 5′ monophosphate. Both two fragments can be further degraded by 3′-5′ and 5′-3′ exoribonucleases, respectively. Regarding the 5′ end dependent pathway, the nudix RNA pyrophosphohydrolase RppH converts 5′ triphosphorylated RNAs into 5′ monophosphorylated RNAs followed by 5′-3′ exoribonuclease or endonuclease digestion (Hui et al., 2014).

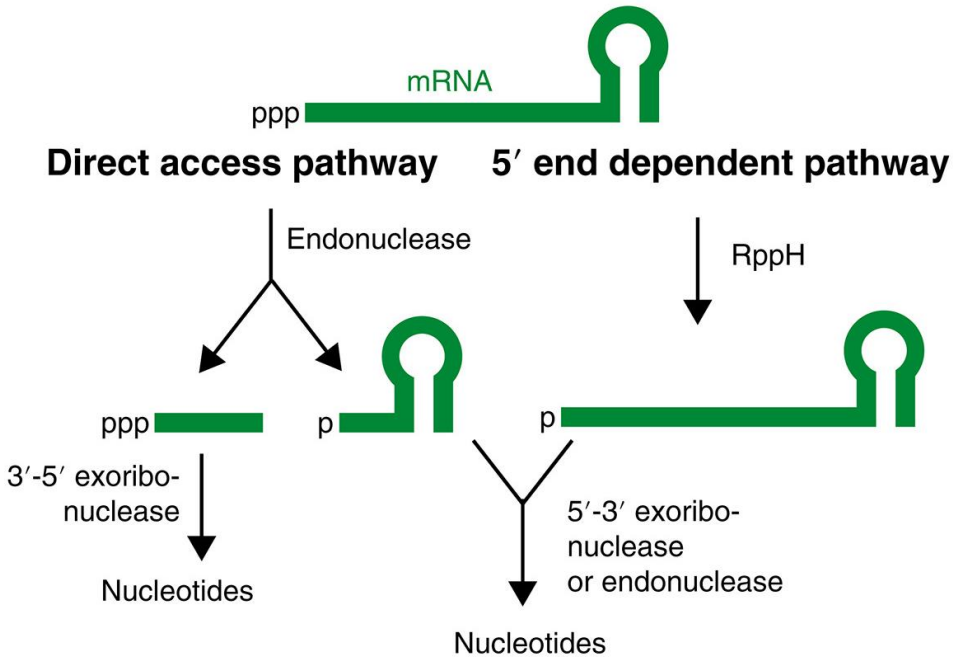


Fig 1.10. mRNA decay pathways in bacteria (Kramer and McLennan, 2019)

mRNA decapping enzymes

As mentioned above, in the 5′-3′ RNA decay pathway, a very critical step is the removal of 5′ m^7G cap. Across species, Nudix hydrolases are the principal enzymes carrying out this process. The Nudix (nucleoside diphosphate linked to another moiety X) superfamily contains diverse proteins possessing pyrophosphatase activity on a great variety of substrates, including nucleoside triphosphates, dinucleoside polyphosphates, dinucleotide coenzymes as well as the 5′ cap modification of mRNAs (Bessman et al., 1996; McLennan, 2006, 2013; Srouji et al., 2017). All the Nudix proteins include a conserved nudix box, a loop–helix–loop structural motif essential for catalysis. This motif binds to metal ions which is required for the hydrolysis of pyrophosphate bond present in all Nudix hydrolase substrates (Figure 1.11).

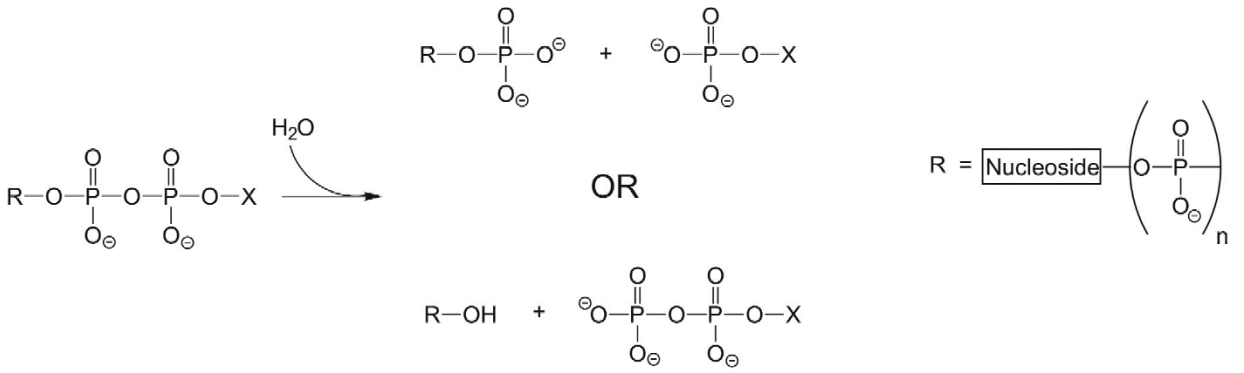


Fig 1.11. Chemical reaction catalyzed by Nudix enzymes. Canonical Nudix hydrolase substrates are nucleoside diphosphates linked to a variable moiety X. The nucleoside component (R) may include a variable number (n) of phosphate groups beyond the diphosphate, n = 0–4 (Srouji et al., 2017).

Bacterial decapping enzymes

In bacteria, two well studied Nudix hydrolases are RppH and NudC. Both enzymes act on pyrophosphate bond within different cap structures to generate 5′ monophosphorylated RNAs.

RppH was first identified as an Ap_5A/Ap_4A hydrolase because it is capable of cleaving the diadenosine polyphosphate at RNA 5′ end to produce monophosphorylated RNA (Bessman et al., 2001). In *E. coli*, RppH (*Ec*RppH) removes pyrophosphate from 5′ triphosphorylated RNAs. However, recent discoveries may implicate that *Ec*RppH proceeds to remove the β-phosphate after the γ-phosphate has first been cleaved via a yet unknown pathway (Luciano et al., 2017, 2018). In contrast to *Ec*RppH, *Bs*RppH (*B. Subtilis* RppH) removes γ- and β-phosphate consecutively, without the release of pyrophosphate (Richards et al., 2011). Both *Ec*RppH and *Bs*RppH require Both *Ec*RppH and *Bs*RppH require single-stranded nucleotides at the 5′ end of their substrates. While *Ec*RppH has a wide selectivity of nucleotides at first, second or third position of RNA 5′ end (Foley et al., 2015), *Bs*RppH has a stricter rule, with an absolute requirement for G at the second position, preference for a purine at the third position and a slight preference for A over G at the first position (Hsieh et al., 2013; Piton et al., 2013). After removal of diphosphates by RppH, RNAs with 5′ monophosphate are unstable and will be rapidly degraded.

NudC acts on NAD-capped RNAs in bacteria to release monophosphorylated RNA with the adenosine moiety of the NAD^+ cap at the 5′ end and nicotinamide mononucleotide (NMN) (Cahova et al., 2015). NudC functions as a homodimer (Figure 1.12), with each monomer contributing to stack the NAD^+ cap in the catalytic pocket where residues within and outside nudix box of both subunits collectively carry out deNADding activity (Hofer et al., 2016; Zhang et al., 2016b).

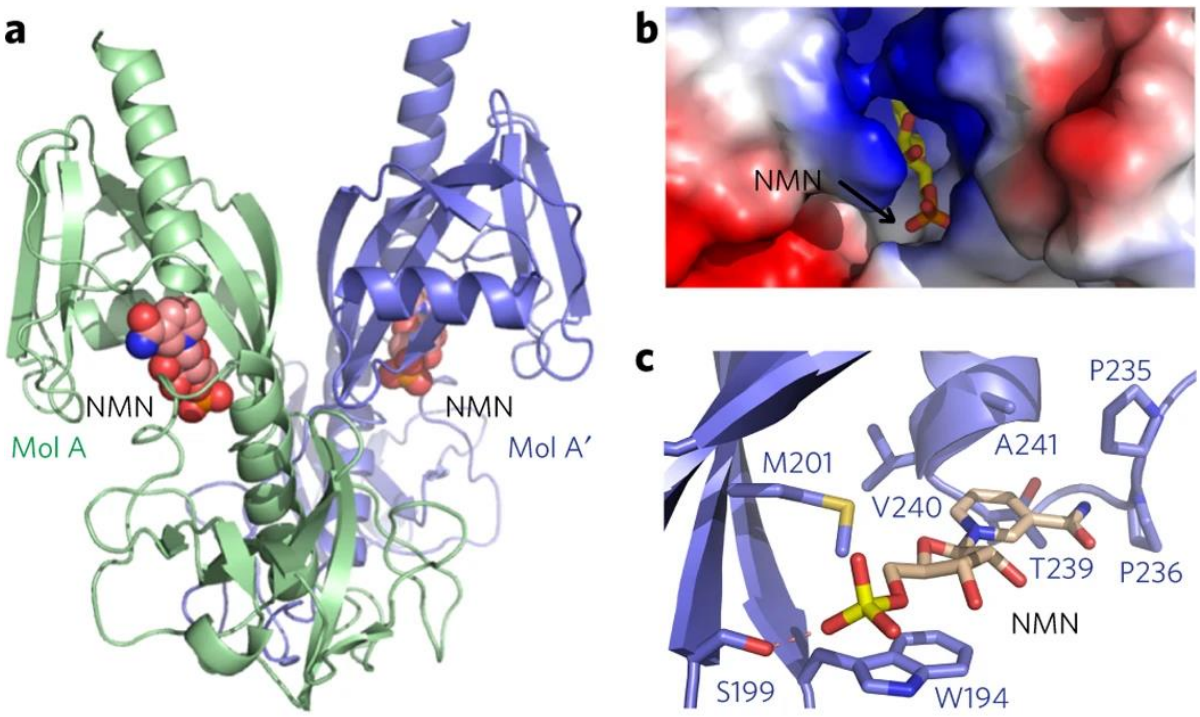


Fig 12. Crystal structure of *E. coli* NudC in complex with NMN (Hofer et al., 2016).

Eukaryotic decapping enzymes

Eukaryotic genomes encode a much larger number of Nudix proteins than the prokaryotes, for example, in human there are 22 NUDIX proteins acting on various substrates including 5′ capped RNAs. 8 out of 22 NUDIX proteins have been demonstrated to possess decapping activity in vitro, of which three enzymes, NUDT3, NUDT16 and NUDT20 (DCP2), are reported to decap RNA in vivo as well (Song et al., 2013).

Dcp2 was originally identified by genetic screen in yeast and its decapping activity was described later (Dunckley and Parker, 1999). Dcp2 is localized in cytoplasm and thought to be the main decapping enzyme that acts on 5′ ends of cytosolic mRNAs. Dcp2 cleaves 5 ′ capped mRNAs between the α- and β-phosphate, releasing a 7-methylguanosine diphosphate (m^7GDP) molecule and a 5′ monophosphorylated mRNA that can be subjected to 5′ exonucleolytic degradation. Enzymatic function of Dcp2 is conserved from yeast to human, suggesting the central role of this protein in governing 5′ RNA decapping in eukaryotes (van Dijk et al., 2002; Wang et al., 2002). Dcp2 forms a core decapping complex with Dcp1, of which Dcp1 is the regulatory module and Dcp2 is the catalytic module (Figure 1.13). Furthermore, catalytic activity of Dcp1-Dcp2 complex can be stimulated by decapping activators, including Pat1, Edc3/Edc4 and the Lsm1-7 complex (Grudzien-Nogalska and Kiledjian, 2017; Parker, 2012; Valkov et al., 2017). Alternatively, Dcp1-

Dcp2 complex can be recruited to mRNA targets by specific RNA binding proteins which recognize the *cis*-element of these targets to activate decapping (Li and Kiledjian, 2010). However, since Dcp2 is not ubiquitously expressed in mouse tissues, it may imply that in mammalian cells there are more Nudix proteins carrying out decapping on different subsets of mRNAs.

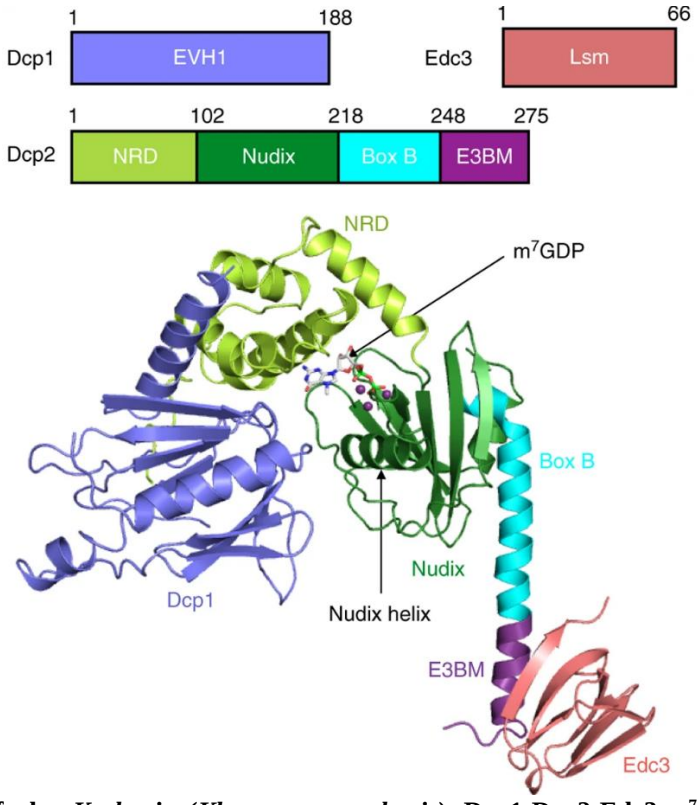


Fig 1.13. Structure of the *K. lactis* (*Kluyveromyces lactis)* Dcp1-Dcp2-Edc3-m^{7}GDP complex. Domain organization of each protein or fragment and their corresponding boundaries are indicated. m^{7}GDP is shown as grey sticks, and the divalent metal ions are shown as purple spheres. E3BM, Edc3-binding motif (Charenton et al., 2016).

Nudt3 and Nudt16 are another two Nudix hydrolases showing decapping activity in vivo (Grudzien-Nogalska et al., 2016; Lu et al., 2011; Song et al., 2010). In vitro decapping experiment showed that Nudt16 preferentially acts on methylated m^{7}G cap, producing m^{7}GDP and 5′ monophosphorylated mRNA, which is similar to Dcp2. Nudt16 can also cleave the unmethylated cap structure, in which both bonds between α- and β-, and β- and γ-phosphate will be hydrolysed. As for Nudt3, it cleaves its substrate either between the α and β phosphate, or between the β and γ phosphate with about equal preference (Song et al., 2013). Both Nudt16 and Nudt3 control a small subset of mRNA targets, with very little overlap between each protein and Dcp2. Nudt16 is reported to regulate small nucleolar RNA (snoRNA) decapping while Nudt3 appears to target those cell mobility-associated mRNAs (Grudzien-Nogalska et al., 2016; Lu et al., 2011).

Eukaryotic deNADding enzymes

With the discovery of noncanonical caps present at 5′ end of mRNAs in eukaryotes, it is speculated that certain enzymes within Nudix superfamily or outside can act on noncanonical caps as well. In bacteria, NudC removes the NAD^+ cap (deNADding). However, how NAD-RNA in eukaryotic cells is degraded remains unclear for quite long time.

One of the key discoveries in this field came out in 2017, in which human DXO (decapping exoribonuclease) protein was shown to remove NAD^+ cap from NAD-capped RNAs in nucleus (Jiao et al., 2017). Previous studies had identified mammalian Dxo has decapping activity on both methylated and unmethylated caps of RNA (Jiao et al., 2013). What's more, it also cleaves the pyrophosphate from triphosphate 5′ ends. In yeast, Dxo homologue Rai1 possesses 5′ pyrophosphohydrolase activity and forms a complex with 5′→3′ exoribonuclease Rat1 to degrade triphosphorylated RNAs in nucleus (Xiang et al., 2009). In mammals, Dxo combines these two activities, meaning it first removes the entire mRNA 5′ cap by cleaving after the first transcribed nucleotide, then processively degrades the mRNA body with its 5′→3′ exoribonuclease activity. Back to the paper (Jiao et al., 2017), human DXO was first time to identify as a nuclear NAD^+ decapping enzyme. Mechanistically, this protein removes the entire NAD^+ moiety, which is different with bacterial deNADding enzyme NudC, and degrades remaining RNA body after (Figure 1.14).

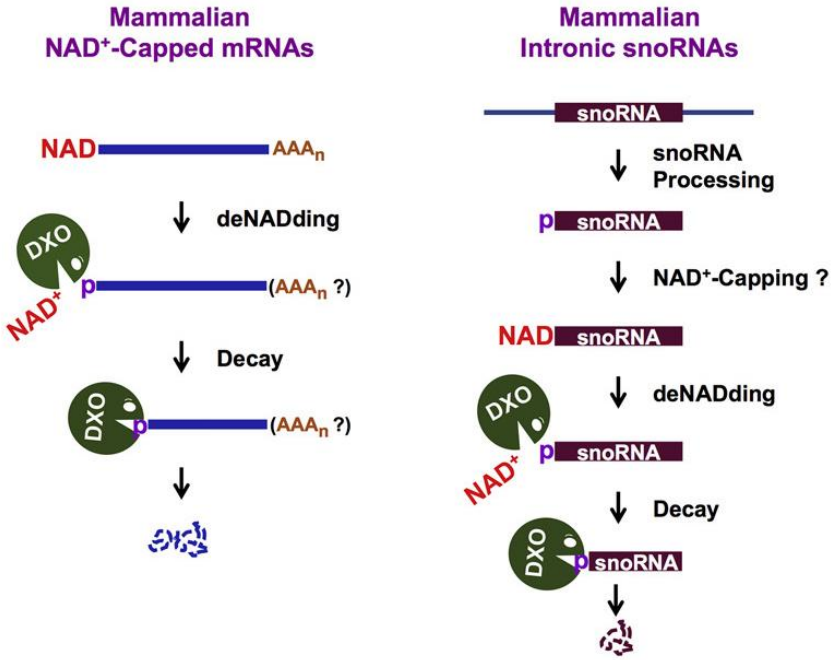


Fig 1.14. Human DXO is a deNADding enzyme (Jiao et al., 2017).

How are cytosolic NAD-RNAs degraded in mammalian cells?

So far, we know that mRNA 5′ end can be decorated with various modifications, including canonical m^7G cap and noncanonical NAD^+ cap. Installation and removal of 5′ caps depend on not only diverse enzymes which are directly catalysing reactions but also a broader protein factors that regulate these processes (Figure 1.15). In mammalian cells, Nudix hydrolases specifically remove mRNA 5′ cap m^7G, such as Dcp2 (Nudt20) and Nudt16, in cytoplasm and Dxo can remove the NAD^+ cap as well as m^7G in nucleus. Except these important findings, still one question is remaining here: what is the destiny of cytosolic NAD-RNAs.

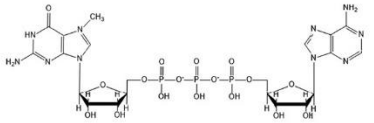
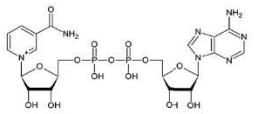
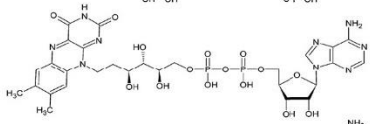
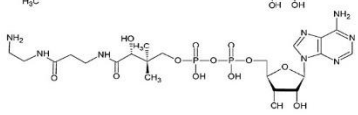
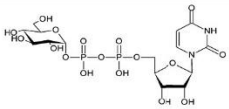
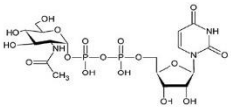
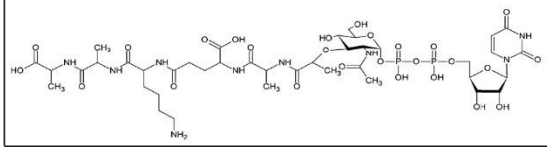
Capping molecules		Capping enzyme	Decapping enzyme
 m⁷G 7-methyl-guanosine, attached to 5' end of RNA	classic cap	phosphatase guanylyltransferase methyltransferase	Dcp2 (NUDIX) DXO
 NAD⁺ nicotinamide adenine dinucleotide	ADP cofactors	bacterial RNAP	NudC (possibly other unknown enzymes)
 FAD flavine adenine dinucleotide		eukaryotic RNApol II	NUDIX enzymes DXO
 DP-CoA dephospho coenzymeA		mitochondrial RNAP	unknown
 UDP-Glc UDP-glucose	Cell wall synthesis precursors	bacterial RNAP	unknown
 UDP-GlcNAc UDP-N-acetyl glucosamine			
 MurNAc-pentapeptide UDP-N-acetyl muramyle-pentapeptide			

Fig 1.15. RNA 5′ caps and their corresponding capping and decapping enzymes (Julius and Yuzenkova, 2019)

During my PhD, we searched Nudix superfamily and focused on those eight enzymes which were already demonstrated to have decapping activity towards m^7G-capped RNA in vitro as mentioned before (Song et al., 2013). Among those eight proteins, we are particularly interested

in Nudt12. Human NUDT12 is the largest protein in NUDIX family, including N-terminal ankyrin repeats (ANK) and C-terminal hydrolase domain. Strikingly, recombinant human NUDT12 protein was shown to hydrolyse NADH and NADPH in vitro (Abdelraheim et al., 2003). Considering the similar structure between NAD^+ and NADH, we were curious about whether NUDT12 could also act on NAD^+ and NAD-RNAs. The answer is yes. According to our data, human NUDT12 is a cytosolic NAD-RNA decapping enzyme. Like bacterial NudC, NUDT12 is active only as homodimer, with each monomer contributing to creation of the two functional catalytic pockets. We further identify BLMH (bleomycin hydrolase) as an interaction factor of NUDT12. The ANK repeats are required to maintain such interaction. In vitro, NUDT12 directly interacts with BLMH, forming an ~600-kDa dodecamer complex. In vivo, BLMH is required to localize NUDT12 into discrete granules which are distinct from P-bodies and both proteins downregulate gene expression when artificially tethered to a reporter RNA. RNA-seq analysis of mouse liver transcripts indicates that expression of circadian clock genes is upregulated in the absence of NUDT12, pointing to a potential physiological role of NUDT12 in cytoplasmic surveillance of NAD-RNAs.

Results

This section consists of a peer-reviewed paper entitled “Decapping Enzyme NUDT12 Partners with BLMH for Cytoplasmic Surveillance of NAD-Capped RNAs”, published in Cell Reports on 24th December 2019. We investigated human NUDT12 enzymatic activities and its functions in vivo. NUDT12 can remove both methylated m^7G cap and unmethylated G cap RNA in vitro, but generate different cleaved products. We further confirmed NUDT12 is a deNADding enzyme that cleaves bond within NAD, releasing NMN (nicotinamide mononucleotide) and 5′ AMP-RNA. We set up a competition assay to exam the decapping (m^7G cap) and deNADding (NAD^+ cap) and found that NUDT12 has a preference for NAD^+ cap over m^7G cap. We determined crystal structure of human NUDT12 in complex with m^7G cap, showing NUDT12 forms a homodimer and such dimerization is essential to maintain the enzymatic activity. Protein mutagenesis experiment elucidated that several residues from both subunits render NUDT12 selectivity towards different caps. To identify potential interaction factors of NUDT12, we performed immunoprecipitation-mass spectrometry (IP-MS) and showed that BLMH is associated with NUDT12 in vivo. We produced recombinant proteins and demonstrated NUDT12 forms a dodecamer with BLMH in vitro. With negative staining electron microscopy analysis, we proposed a model of this complex where hexameric BLMH forms a ring structure in the central and three subunits of NUDT12 homodimer are anchored to the periphery of the ring. BLMH has no detectable effect on decapping or deNADding activity of NUDT12 in vitro, but is required to localize NUDT12 to cytoplasmic discrete granules. To investigate physiological function of NUDT12, we made *Nudt12* KO mice. These mice are viable and fertile. We checked RNA expression in mouse liver and uncovered in the absence of Nudt12, expression of circadian transcripts are upregulated.

In this research, I produced all the recombinant proteins and performed in vitro biochemical experiments, including decapping assay and competition assay. To solve the human NUDT12 structure, I did limited proteolysis-coupled mass spectrometry to define the protein sequence boundaries for crystallization. I also performed negative staining electron microscopy analysis of NUDT12-BLMH complex and prepared sample for AUC and SEC-MALS experiments. Kuan-ming Chen created point mutants and solved the crystal structure together with Andrew A. McCarthy. Hao Wu created the mouse knockout mutant and performed all cell culture and mouse experiments. David Homolka analyzed sequencing data. Pasca Gos and Fabienne Fleury-Olela

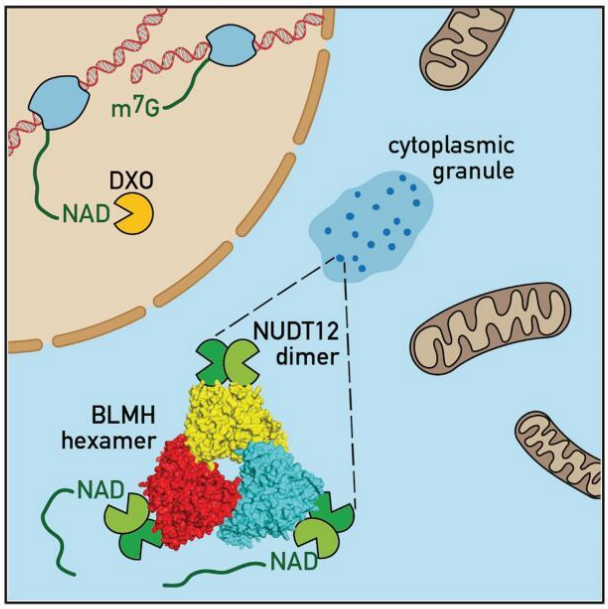
carried out circadian experiments with Hao Wu. Ramesh Pillai prepared and wrote the manuscript with input from everyone.

Article

Cell Reports

Decapping Enzyme NUDT12 Partners with BLMH for Cytoplasmic Surveillance of NAD-Capped RNAs

Graphical Abstract



Highlights

- Human NUDT12 is a decapping enzyme for NAD-capped RNAs
- Homodimerization of NUDT12 creates two interdependent catalytic pockets
- Interaction with BLMH facilitates localization of NUDT12 to cytoplasmic granules
- *Nudt12* knockout mice upregulate circadian clock transcripts

Authors

Hao Wu, Lingyun Li, Kuan-Ming Chen, ..., Fabienne Fleury-Olela, Andrew A. McCarthy, Ramesh S. Pillai

Correspondence

ramesh.pillai@unige.ch

In Brief

Wu et al. identify NUDT12 as a decapping enzyme for NAD-capped RNAs and demonstrate a role for the enzyme in regulation of circadian clock transcripts in mice. They reveal the existence of a BLMH-NUDT12 complex that facilitates localization of NUDT12 to distinct cytoplasmic granules.

Wu et al., 2019, Cell Reports *29*, 4422–4434
December 24, 2019
https://doi.org/10.1016/j.celrep.2019.11.108

CellPress

OPEN ACCESS
CellPress

Cell Reports
Article

Decapping Enzyme NUDT12 Partners with BLMH for Cytoplasmic Surveillance of NAD-Capped RNAs

Hao Wu,[1,3] Lingyun Li,[1,3] Kuan-Ming Chen,[1,3] David Homolka,[1] Pascal Gos,[1] Fabienne Fleury-Olela,[1] Andrew A. McCarthy,[2] and Ramesh S. Pillai[1,4,*]

[1]Department of Molecular Biology, Science III, University of Geneva, 30 Quai Ernest-Ansermet, 1211 Geneva 4, Switzerland
[2]European Molecular Biology Laboratory, Grenoble Outstation, 71 avenue des Martyrs, 38042 Grenoble, France
[3]These authors contributed equally
[4]Lead Contact
*Correspondence: ramesh.pillai@unige.ch
https://doi.org/10.1016/j.celrep.2019.11.108

SUMMARY

RNA polymerase II transcripts receive a protective 5′,5′-triphosphate-linked 7-methylguanosine (m^7G) cap, and its removal by decapping enzymes like DCP2 is critical for initiation of RNA decay. Alternative RNA caps can be acquired when transcription initiation uses metabolites like nicotinamide adenine dinucleotide (NAD), generating NAD-RNAs. Here, we identify human NUDT12 as a cytosolic NAD-RNA decapping enzyme. NUDT12 is active only as homodimers, with each monomer contributing to creation of the two functional catalytic pockets. We identify an ~600-kDa dodecamer complex between bleomycin hydrolase (BLMH) and NUDT12, with BLMH being required for localization of NUDT12 to a few discrete cytoplasmic granules that are distinct from P-bodies. Both proteins downregulate gene expression when artificially tethered to a reporter RNA *in vivo*. Furthermore, loss of *Nudt12* results in a significant upregulation of circadian clock transcripts in mouse liver. Overall, our study points to a physiological role for NUDT12 in the cytosolic surveillance of NAD-RNAs.

INTRODUCTION

Eukaryotic RNA polymerase II transcripts receive a 5′ cap structure as the first step of RNA processing (Furuichi and Shatkin, 2000; Shatkin, 1976; Wei et al., 1975). This 5′,5′-triphosphate-linked 7-methylguanosine (m^7G) cap plays a critical role in subsequent RNA processing events like splicing, export, stability, and translation (Sonenberg et al., 1979). RNAs with an alternative cap consisting of the metabolite nicotinamide adenine dinucleotide (NAD) were originally identified in *E. coli* (Chen et al., 2009) and later shown to be present on several bacterial small regulatory RNAs (Cahová et al., 2015). NAD caps are now detected in budding yeast (Walters et al., 2017), human cells (Jiao et al., 2017), and plants (Wang et al., 2019; Zhang et al., 2019).

Canonical RNA capping takes place after ~20 nt of the nascent RNA are transcribed. It involves the action of three enzymatic activities contained within two separate enzymes (Inesta-Vaquera and Cowling, 2017). In contrast, acquisition of an NAD cap is believed to be due to the direct use of NAD^+ or reduced NADH as the initiating nucleotide (instead of ATP) by bacterial and eukaryotic polymerases, including nuclear RNA polymerase II and the mitochondrial RNA polymerase (Bird et al., 2016, 2018). Efficiency of NAD incorporation is determined by the promoter DNA sequence at, and upstream, of the transcription start site (Bird et al., 2016, 2018). Development of a click-chemistry protocol called NADcaptureSeq has enabled identification of transcripts carrying the NAD cap structure (Cahová et al., 2015). It is now clear that short regulatory RNAs and 5′-terminal fragments are NAD-capped in *E. coli* (Cahová et al., 2015), while thousands of nuclear and mitochondrial mRNAs are enriched via this protocol when applied to yeast (Walters et al., 2017), human (Jiao et al., 2017), and plant (Wang et al., 2019; Zhang et al., 2019) total RNA. In all these systems, a large set of transcripts are NAD capped, with a certain proportion (up to 11%) of transcripts from any given gene carrying the alternative cap (Jiao et al., 2017). Finally, some intron-derived small nucleolar RNA (snoRNAs) are also enriched in such NADcaptureSeq libraries, pointing to the existence of a selective post-transcriptional NAD-capping mechanism (Jiao et al., 2017).

Presence of an m^7G cap offers stability to the transcript and influences further processing of the pre-mRNA. Based on the sequences recovered in NADcaptureSeq experiments, NAD capping does not negatively influence further RNA processing in the nucleus, as many of the NAD-RNAs are mature spliced and polyadenylated mRNAs (Jiao et al., 2017; Walters et al., 2017; Wang et al., 2019; Zhang et al., 2019). Furthermore, plant NAD-mRNAs are also found on polysomes, indicating that they may even be translated (Wang et al., 2019). Therefore, alternate capping with NAD may be used to regulate gene expression in both the nuclear and cytoplasmic compartments. A key aspect of this regulatory pathway is turnover of NAD-RNAs.

The eukaryotic decapping enzymes DCP2 and NUDT16 remove the m^7G cap from RNAs to promote decay. They belong to the highly conserved Nudix family of hydrolases (McLennan, 2006), members of which are defined by the presence of a so-called Nudix (nucleoside diphosphate linked to another moiety X) motif, which has glutamic acid residues responsible for their diphosphate hydrolysis activity (Bessman et al., 1996). The 5′ triphosphate end of bacterial RNAs is engaged by the bacterial Nudix protein RNA-pyrophosphohydrolase RppH that converts it into a 5′ monophosphate end, allowing RNA turnover by the endoribonuclease RNase E (Cahová et al., 2015). The NAD cap

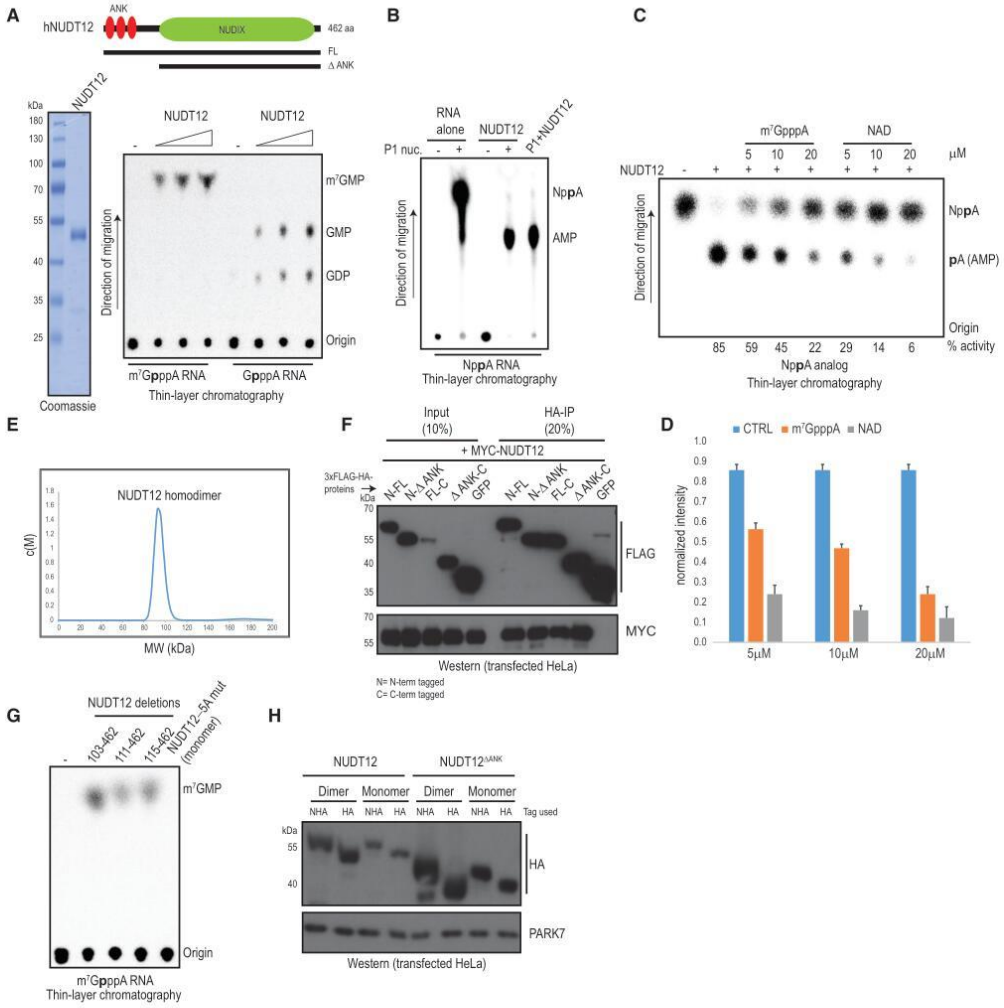


Figure 1. Human NUDT12 Is an NAD-RNA Decapping Enzyme

(A) Domain architecture of human NUDT12 with three ankyrin (ANK) repeats at the N terminus and a catalytic domain at the C terminus. Gel showing purity of the recombinant FL human NUDT12 protein. *In vitro* decapping assay with NUDT12 and ^{32}(P)-cap-labeled m^{7}G- or G-capped RNAs. Reactions were separated by thin-layer chromatography (TLC). See also Figures S1A and S1B.

(B) Decapping assay with NUDT12 and NAD-RNA. The reactions were treated with the endoribonuclease RNase P1 prior to TLC.

(C) Cleavage assay with NUDT12 and ^{32}(P)-NAD analog carried out in the presence of competing (concentrations used indicated in μM) cold m^{7}GpppA or NAD caps.

(D) Quantification from triplicate competition experiments similar to the one presented in (C). Error bars represent SD. See also Figure S1C.

(E) Analytical ultracentrifugation (AUC) analysis of NUDT12 shows it is a homodimer.

(F) Self-association of NUDT12 *in vivo* in human cell cultures is mediated via the C-terminal catalytic domain. The 3X-FLAG-HA-tagged NUDT12 proteins expressed have either an N-terminal tag or a C-terminal tag, as indicated. Tagged GFP protein was used as a negative control.

(legend continued on next page)

is considered to be protective in bacteria, as it inhibits RppH activity on RNA substrates *in vitro* (Cahová et al., 2015). Instead, another bacterial Nudix enzyme, NudC, is required to turnover NAD-RNA (Cahová et al., 2015). The NAD cap on human RNAs is seen as an aberration and is a signal for decay. The nuclear decapping exonuclease (DXO) that degrades incompletely capped RNAs is shown to remove the entire NAD moiety from such RNAs, leading to their degradation (Jiao et al., 2017). However, only a subset of NAD-RNAs are processed by DXO (Jiao et al., 2017), raising the possibility of existence of other activities. Here, we identify the mammalian Nudix protein NUDT12 as a cytoplasmic decapping enzyme for NAD-RNAs in human cell cultures and mice.

RESULTS

NUDT12 Is a Decapping Enzyme for NAD-RNAs

To identify mammalian enzymes capable of removing NAD caps from RNAs, we searched the Nudix family of hydrolases (McLennan, 2006), eight of which are already demonstrated to have decapping activity toward m^7G-RNA *in vitro* (Song et al., 2013). Among these eight proteins, we focused on NUDT12, as it is similar to the bacterial NudC (YjaD) enzyme that cleaves NAD-RNAs in *E. coli* (Cahová et al., 2015; Höfer et al., 2016; Zhang et al., 2016). Furthermore, recombinant human NUDT12 was demonstrated to hydrolyze NADH and NADPH *in vitro* (Abdelraheim et al., 2003). NUDT12 is one of the largest proteins in the family and is composed of three N-terminal ankyrin (ANK) repeats and a C-terminal catalytic domain containing the Nudix motif (Figure 1A).

To test its activity, purified full-length (FL) human NUDT12 protein was incubated with ^{32}P-cap-labeled RNAs, and reactions were resolved by thin-layer chromatography (STAR Methods). Consistent with previous findings (Song et al., 2013), NUDT12 efficiently liberated m^7GMP from the m^7G-capped RNA (m^7G**p**pppA-RNA; where **p** indicates ^{32}P label) and guanosine diphosphate (GDP) and guanosine monophosphate (GMP) from the unmethylated G-capped RNA (G**p**ppA-RNA) (Figures 1A and S1A). This cleavage reaction is different from that catalyzed by DCP2, which cleaves to release m^7GDP from the m^7G-capped RNA (Figure S1B; Wang et al., 2002). Importantly, when NUDT12 was incubated with a NAD-RNA (Np**p**A-RNA), it cleaves within the diphosphate bond of NAD to release the nicotinamide mononucleotide (NMN or Np). The radiolabeled nucleotide product (AMP or **p**A) that is left attached to the rest of the RNA substrate is visualized by treatment of the reactions with the ribonuclease RNase P1 that cleaves phosphodiester bonds within the RNA (Figure 1B). This cleavage mechanism of NUDT12 is different from that used by DXO, which removes the entire NAD moiety (Jiao et al., 2017). When incubated with free ^{32}P-labeled NAD (Np**p**A) as a substrate, cleavage activity of NUDT12 can be monitored by appearance of the labeled AMP moiety (Figure 1C). Demonstrating its specificity for NAD as a substrate, human NUDT12 activity is more efficiently competed by unlabeled NAD than the m^7GpppA cap analog at any of the concentrations tested (Figures 1C, 1D, and S1C). While this study was in preparation, similar findings were reported showing that mouse NUDT12 is a decapping enzyme for NAD-RNAs (Grudzien-Nogalska et al., 2019). These results demonstrate that NUDT12 is an NAD-decapping enzyme.

Homodimerization of Human NUDT12 Is Essential for Catalytic Activity and Stability *In Vivo*

Bacterial NudC exists as homodimers (Höfer et al., 2016; Zhang et al., 2016). Similarly, analytical ultracentrifugation (AUC) analysis reveals that ~50-kDa FL human NUDT12 exists as a homodimer in solution (Figure 1E). In line with this finding, immunoprecipitation experiments reveal that 3xFLAG-HA-tagged-NUDT12 and Myc-tagged-NUDT12 self-associate in transfected human HeLa cell cultures (Figure 1F). Deletion of the N-terminal ANK repeats (ΔANK) does not affect this self-association, pointing to a role for the catalytic domains in dimerization (Figure 1F). To examine the relevance of dimerization for activity, we sought to create a monomeric version of the protein. Based on the bacterial NudC structures (Höfer et al., 2016; Zhang et al., 2016), we created a human NUDT12 mutant that has five amino acid point mutations (5A-mut) that are predicted to disrupt intermolecular interactions (Figure S1D). Indeed, size-exclusion chromatography and AUC analysis (Figure S1E) indicates that the mutant (5A-mut) is monomeric. When tested for decapping activity with m^7G-RNA (Figure 1G) or with NAD-RNA (Figure S2C), the recombinant monomeric 5A-mut protein was completely inactive. Note that truncations at the N-term that remove the ANK repeats did not affect activity (Figure 1G). Interestingly, Western analysis of transfected HeLa cells shows that monomeric versions (FL or ΔANK) of NUDT12 accumulate to lower levels when compared to the wild-type dimeric protein (Figure 1H). Taken together, this shows that homodimerization of NUDT12 is essential for catalytic activity and protein stability *in vivo*.

Crystal Structure of Human NUDT12 in Complex with m^7GTP

To understand the structural basis for the requirement of homodimerization for catalytic activity, we crystallized the active (Figure 1G) catalytic core of human NUDT12 (111–462 aa) in complex with m^7GTP (PDB: 6SCX; Table 1) and determined the structure by X-ray crystallography (STAR Methods). The 2.9-Å structural model of human NUDT12 presents a homodimer. Each monomer is folded into three distinct regions, two tandem Nudix-fold domains (the N-terminal domain [NTD] and the C-terminal domain [CTD]), which are separated by a zinc-binding motif (Figure 2A). The overall structure is very similar (Figure 2B) to that recently described for the mouse NUDT12 complexed with the bound product AMP and three Mg^{2+} ions

(G) Decapping assay with homodimeric human NUDT12 carrying different N-terminal truncations, and a monomeric version (5A-mut). The monomeric version of the protein is inactive. See also Figures S1D and S1E.

(H) Western analysis of HA-tagged proteins expressed in HeLa cells. The FL NUDT12 protein or one lacking the ANK repeats are expressed, and either has homodimeric or monomeric (carrying 5A-mutations) versions.

See STAR Methods and Figures S1D and S1E. Note the lower abundance of the monomeric versions.

Table 1. Data Collection and Refinement Statistics

Crystal form	m^7GTP/Cd 6SCX	Apo
Wavelength (Å)	0.9801	0.9763
Resolution range (Å)	50-2.9 (3.2-2.9)	50-3.0 (3.3-2.8)
Space group	H32	$P6_222$
Unit cell (Å)	229.9, 229.9, 159.0, 90, 90, 120	234.9, 234.9, 459.9, 90, 90, 120
Unique reflections	21142 (1057)	72404 (3620)
Completeness (%)		
Spherical	60.1 (11.9)	41.3 (6.1)
Ellipsoidal	95.8 (88.1)	94.7 (76.9)
Mean < I/σI >	11.6 (1.8)	7.6 (2.0)
R_{pim} (%)	3.9 (50.6)	7.5 (45.0)
CC*	0.998 (0.633)	0.997 (0.723)
R_{work} (%)	23.2 (28.9)	
R_{free} (%)	26.0 (29.1)	
R_{msd} (bonds, Å)	0.007	
R_{msd} (angles, °)	0.90	

(PDB: 6O3P; root-mean-square deviation [RMSD] of 1.0 Å over 308 aa) (Grudzien-Nogalska et al., 2019) and the *E. coli* NudC homodimer (PDB: 5IW5) (Höfer et al., 2016). The only other major differences between the structures, apart from some alternate loop conformations, are an additional 8 aa observed at the N terminus of the human structure and replacement of Mg^{2+} and Zn^{2+} (of the zinc-binding motif) ions in the mouse structure (Figure 2B) with Cd^{2+} ions in the human structure (Figure 2A).

The CTD of each monomer contains the critical catalytic residues that constitute the Nudix motif. We clearly observed m^7GTP bound in each of the active sites, and it adopts at least two conformations (Figures 2C and 2D). The purine base is stacked between the aromatic residues F356 from the same monomer and Y318 from the other monomer, explaining the need for homodimerization. A similar stacking of the base of AMP is seen within the mouse NUDT12 structure (Figure 2E) or that of NAD in bacterial NudC (Figure 2F). Interestingly, the m^7GTP does not appear to be cleaved in the human structure, which is likely due to the addition of Cd^{2+}, where only two ions were observed.

To test the functional importance of key structural features in the human NUDT12, we designed point mutations (Figures S1D and S2A) and performed decapping assays. A core version of human NUDT12 (103–462 aa) lacking the ANK repeats is active in decapping m^7G- and NAD-RNAs (Figures 2G, S2B, and S2C). Mutation of the conserved Nudix motif residues E370A and E374A that coordinate the three catalytic magnesium ions in the mouse structure (Figure 2E) or the two cadmium ions in the human structure (Figures 2C and 2D) completely disables the enzyme (Figures 2G, S2B, and S2C). Similarly, mutation of F356A that stacks with the purine base of the nucleotide (Figures 2C–2E) is also detrimental for activity on both m^7G- and NAD-capped RNAs (Figures 2G, S2B, and S2C).

While the above mutations (labeled in red in Figure 2G) completely abolish activity, three mutations (marked in blue in Figure 2G) abolish activity only on m^7G-RNA, with no or only partial loss of activity on NAD-RNA (Figures 2G, S2B, and S2C). As previously described for the prokaryotic NudC (Höfer et al., 2016) and for mouse NUDT12 (Grudzien-Nogalska et al., 2019), the catalytic pocket in the CTD of each monomer is completed by contributions from the other protein chain. In particular, Y318 (from the zinc-binding motif) of the neighboring monomer stacks with the opposite face (relative to F356) of the purine base (Figures 2C and 2E). Mutation of Y318 results in specific loss of activity of NUDT12 for m^7G-capped RNA, with the variant retaining a reduced activity to NAD-RNA (Figures 2G, S2B, and S2C). Likewise, mutation of Y384 that resides at the back of the active site, in the interface between the two monomers (Figure S2D), makes human NUDT12 activity exclusive to NAD-RNAs (Figures 2G, S2B, and S2C). Y384 in NUDT12 makes a hydrogen bond contact with the other monomer via E364, which is part of the helix carrying the catalytic residues E370A and E374A (Figure S2E). Finally, mutation of W390 within the CTD abolishes activity on m^7G-RNA while only reducing activity on NAD-RNA (Figures 2G, S2B, and S2C). W390 in human NUDT12 is equivalent to W194 in NudC, where it is shown to stack with the nicotinamide moiety of the bound NAD (Höfer et al., 2016; Figure 2F). Thus, while homodimerization is essential for catalytic activity of human NUDT12, individual perturbation of the intermolecular contact points can modulate substrate specificity of the enzyme.

ANK Repeats of NUDT12 Associate with Bleomycin Hydrolase to Form an ~600-kDa Complex

To study how NUDT12 functions *in vivo*, we wanted to identify its associated factors. HA-tagged-NUDT12 was isolated from transfected human HeLa cell cultures and subjected to mass spectrometry (Figure S3A). We identified Bleomycin Hydrolase (BLMH) as a top candidate. Supporting this observation, endogenous NUDT12 was identified as the top interacting factor in tagged BLMH complexes (Figure S3A). BLMH is a cysteine peptidase that cleaves the anti-cancer peptide Bleomycin, reducing the intracellular levels of the drug (Akiyama et al., 1981; Lazo and Humphreys, 1983). A number of functions are attributed to BLMH, including hydrolysis of homocysteine lactone (Borowczyk et al., 2012; Zimny et al., 2006), and acting as a peptide trimmer in concert with the proteasome (Kim et al., 2009; Stoltze et al., 2000), but its primary biological function remains enigmatic (van der Linden et al., 2015).

We confirmed the association by demonstrating co-immunoprecipitation of Myc-BLMH with HA-NUDT12 from transfected HeLa cell lysates (Figure 3A). ANK repeats are established protein-protein interaction modules (Mosavi et al., 2004). Deletion (HA-NUDT12$^{\Delta ANK}$) of the three ANK repeats at the N terminus of NUDT12 abolishes this association with Myc-BLMH (Figure 3B). To examine if this interaction is direct, we co-expressed the proteins in Hi5 insect cells using a Baculovirus-mediated expression system (STAR Methods). Co-purification of untagged human BLMH with 6x-His-tagged human NUDT12 as a recombinant complex confirms a direct interaction between them (Figure 3C). Elution profile of the complex in a size-exclusion column suggests formation of a large entity (Figure 3D). Indeed, size-exclusion chromatography coupled to Multi-angle laser light scattering (SEC-MALLS) analysis estimates the molecular weight of the complex to be 561.6 kDa

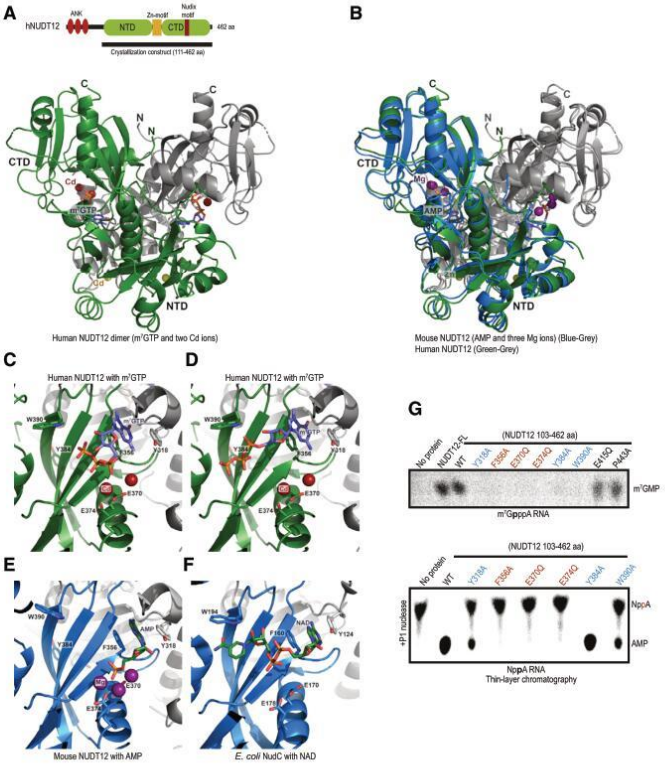


Figure 2. Crystal Structure of Human NUDT12 in Complex with m^{7}GTP

(A) Cartoon showing structural domains of human NUDT12 composed of N-terminal three ANK repeats followed by two tandem Nudix-fold domains: an N-terminal domain (NTD) and a C-terminal domain (CTD) that are separated by a zinc-binding motif. The CTD harbors the catalytic Nudix motif. Crystal structure of the homodimeric human NUDT12 (PDB: 6SCX) catalytic core bound with m^{7}GTP and Cd^{2+} ions. Note that the Cd^{2+} ions occupying the zinc-binding motif in human structure are indicated in yellow.

(B) An overlay of our human NUDT12 dimer crystal structure (PDB: 6SCX) with that of the mouse protein bound with AMP (PDB: 6O3P) (Grudzien-Nogalska et al., 2019) showing a high degree of overlap. The three magenta spheres represent the magnesium ions present in the mouse structure. The structural zinc ions are shown as yellow spheres.

(C and D) A zoomed-in view of the catalytic pocket of human NUDT12 showing the bound m^{7}GTP (two orientations are seen) and two Cd^{2+} ions. The bound m^{7}GTP remains uncleaved.

(E) Zoomed-in view of mouse NUDT12 with bound AMP and three catalytic Mg^{2+} ions.

(F) NAD bound in the *E. coli* NudC (PDB: 5IW5).

(G) Decapping assays with FL human NUDT12 or a core version lacking the ANK repeats (103–462 aa). Indicated point mutations were introduced into the core version of the protein. Mutations (red) abolishing activity on both m^{7}G-capped and NAD-RNAs or mutations (blue) that render the enzyme active only for NAD-RNAs are indicated.

See also Figures S2A–S2C.

(Figure 3E). Given that each protein is ~50 kDa, the complex is likely equivalent to a dodecamer (12-mer) composed of NUDT12 and BLMH.

Using negative-stain electron microscopy (EM), the complex is revealed to have a striking pattern consisting of a central doughnut with three radiating spokes (Figure 3F). Several particles have only one or two spokes, while others completely lack them. Human BLMH is reported to exist in solution as a hexamer, with the crystal structure of the complex appearing as a barrel composed of a double stack of interconnected trimers (O'Farrell et al., 1999). Based on our EM data and the fact that human NUDT12 alone forms homodimers (Figure 1E), we propose that the central doughnut is the BLMH hexamer, while the three spokes are formed by dimers of NUDT12 (Figure 3F), accounting for an ~600 kDa dodecamer.

We examined whether this complex formation alters decapping activity of NUDT12. When compared to NUDT12 dimers, we did not see any change in decapping activity of the NUDT12-BLMH complex on m^{7}G-capped RNA (Figure S3B). In addition, titration of BLMH into a reaction containing a fixed amount of either FL NUDT12 or NUDT12$^{\Delta ANK}$ did not alter decapping activity on m^{7}G-capped RNA (Figure 3G and S3C). Similarly, there was no effect of titrating-in BLMH into a reaction containing NUDT12 and labeled NAD (Figure 3H). Taken together, we find that the decapping enzyme NUDT12 forms a large complex with BLMH that does not affect the nuclease activity *in vitro*.

Localization of NUDT12 to Cytoplasmic Granules Depends on Interaction with BLMH

To examine where NUDT12 might be functioning *in vivo*, we examined subcellular localization of NUDT12 using rabbit polyclonal antibodies raised against human NUDT12. Immunofluorescence analysis indicates that endogenous NUDT12 in human HeLa cells is present in discrete cytosolic granules that number 2-3 per cell (Figure 4A). Tagged EGFP-NUDT12 and NUDT12-HA fusion proteins that were transiently expressed in transfected HeLa cells also mimic this localization pattern (Figures 4B and 4C). The EGFP-NUDT12 spots are distinct from other cytoplasmic granules like P-bodies, as they do not overlap with endogenous XRN1, a marker for P-body structures (Figure S4A). NUDT12 is reported to co-localize with peroxisomes (Abdelraheim et al., 2003), but we do not find such an association between EGFP-NUDT12 and peroxisomes stained with anti-catalase antibodies (Figure S4A).

Endogenous BLMH detected with rabbit polyclonal antibodies is mostly diffused in the cytoplasm, and so is a Myc-BLMH

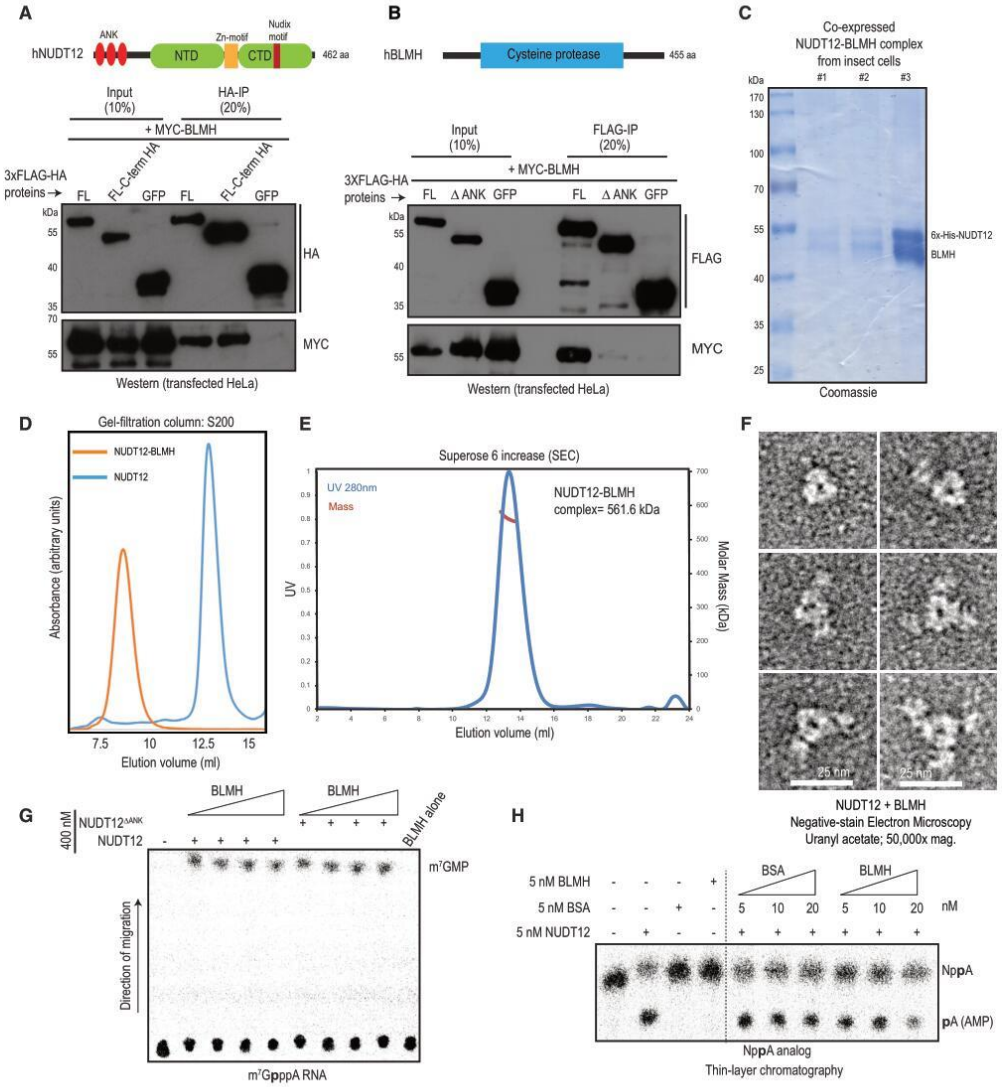


Figure 3. Human NUDT12 Associates with BLMH to Form an ~600-kDa Complex

(A) Cartoon showing domains of NUDT12. Western blot analysis showing co-immunoprecipitation of MYC-BLMH with HA-NUDT12 from transfected human HeLa cell cultures. Note that all proteins are N-terminal tagged with 3x-FLAG-HA or MYC. The use of a C-terminal HA-tagged NUDT12 version is indicated. The GFP protein is used as a negative control to evaluate nonspecific binding to beads. See also Figure S3A.

(B) Cartoon showing the functional domain in BLMH. MYC-BLMH uses the three ANK repeats of NUDT12 for interaction.

(C) Coomassie-stained gel showing co-purification of untagged BLMH with tagged NUDT12 as a pure recombinant complex.

(D) Gel-filtration elution profile of NUDT12 compared to that of the NUDT12-BLMH complex. The early elution of the complex indicates its large size.

(E) Size-exclusion-chromatography-coupled multi-angle laser light scattering (SEC-MALLS) of the complex estimates it to be ~561 kDa.

(legend continued on next page)

tagged version (Figure S4B). Nevertheless, when co-expressed, NUDT12-HA and Myc-BLMH co-localize in cytoplasmic granules in transfected HeLa cells (Figure 4D). Importantly, endogenous BLMH is also detected in EGFP-NUDT12 foci in HeLa cells (Figure S4C). To visualize the local cellular environment occupied by the EGFP-NUDT12 granules, we examined transfected HeLa cells by immunogold labeling and transmission electron microscopy. This reveals the protein to occupy electron-dense structures that are near various organelles including mitochondria, but no specific associations were noted (Figure 4E).

We reasoned that NUDT12 might interact with endogenous BLMH in HeLa cells to form large complexes, which may explain the formation of cytosolic granules. Indeed, deletion of the ANK repeats (HA-NUDT12$^{\Delta ANK}$) results in loss of the punctate localization pattern in transfected HeLa cells, with the protein appearing diffused (Figure 4F). Interestingly, when cells are treated with small interfering RNAs (siRNAs) targeting *Blmh* mRNA, endogenous NUDT12 granules are dispersed (Figure 4G). A similar dispersion of the EGFP-NUDT12 granules is seen when cells are treated with anti-*Blmh* siRNAs (Figure S4D). These experiments show that complex formation with BLMH helps to sequester NUDT12 into cytoplasmic granules.

To obtain additional evidence that BLMH and NUDT12 may function together *in vivo*, we made use of an artificial reporter tethering experiment in human cell cultures. Briefly, when the nuclease NUDT12 is tethered (via N-peptide fusions) to the BoxB hairpins on a luciferase reporter mRNA (Renilla BoxB), a strong inhibition of reporter gene expression is observed (Figures 4H, S5A, and S5B). Co-expression of a control HA-tagged version of the protein (that cannot be tethered) or N-peptide fusion of an unrelated fusion protein like LacZ does not elicit any repression of the BoxB reporter (Figures 4H, S5B, and S5C). Repression by NUDT12 is independent of complex formation with endogenous BLMH, as tethered NUDT12$^{\Delta ANK}$ displays a similar inhibitory effect (Figures 4H and S5B). We find that a tethered catalytic-dead (combined E370Q, E373Q, and E374Q mutations) version of NUDT12 also mediates repression, as it likely forms dimers with endogenous wild-type NUDT12. Consistently, a monomeric version of the catalytic-dead NUDT12 fails to repress gene expression (Figure S5B). Importantly, recombinant BLMH has no detectable nuclease activity *in vitro* (Figure 3G), but when BLMH is tethered to the reporter mRNA *in vivo*, it represses gene expression similar to that seen with NUDT12 (Figures 4H and S5B). We propose that this inhibitory effect of BLMH is mediated by recruitment of endogenous NUDT12 that associates with the tethered BLMH.

Overall, these studies reveal that NUDT12 associates with BLMH to form a complex that helps to compartmentalize the proteins in cytosolic granules. While our experiments do not indicate any impact of this complex on NUDT12 decapping activity *in vitro*, the cell culture tethering experiments reveal a potential for BLMH in recruiting NUDT12 nuclease activity to a reporter RNA.

NUDT12 Regulates Circadian Clock Transcript Levels in Mouse Liver

To examine the physiological role of NUDT12, we first determined its tissue expression pattern in mouse tissues. NUDT12 is widely expressed in multiple tissues of an adult mouse but abundantly present in liver and kidney (Figure 5A). We created a *Nudt12* knockout mouse mutant (STAR Methods; Figures 5B, S6A, and S6B). Western blot analysis shows a complete loss of the protein in the liver and kidney from knockout animals (Figure 5B).

Heterozygous *Nudt12*$^{+/-}$ and homozygous *Nudt12*$^{-/-}$ animals of both sexes are viable and fertile, with no other obvious phenotypes. Our efforts to quantify NAD-RNA levels by RNA mass spectrometry failed, probably due to their low abundance in cells (Jiao et al., 2017). However, metabolomic analyses of multiple tissues (liver, kidney, and testis) reveal an increase in NADH levels, but not NAD levels, in both the heterozygous and homozygous *Nudt12* mutant animals (Figure S6C). This is consistent with the observed preference of recombinant NUDT12 protein for hydrolysis of reduced NADH when compared to oxidized NAD (Abdelraheim et al., 2003). To examine any changes in the overall transcriptome, we carried out RNA sequencing (RNA-seq) analysis of total RNA from liver and kidney. Very few transcripts are differentially expressed in the knockout mutant (Figure 5C). When the altered (upregulated or downregulated) transcripts are compared between liver and kidney, very little overlap is seen, leading us to conclude that NUDT12 regulates transcripts that are largely specific to the different tissues (Figures 5D and S6D).

Examination of the upregulated transcripts in the liver reveals an enrichment of genes that are involved in circadian gene expression (Figure 5E). Mammalian circadian gene expression is set by the transcription factors CLOCK and BMAL1, which then activate transcription of *Cryptochrome* (*Cry* 1 and 2) and *Period* (*Per* 1, 2, and 3) (Albrecht, 2012; Schibler et al., 2015). Subsequent feedback inhibition of CLOCK and BMAL by CRY and PER proteins completes the loop. It is known that cellular NAD levels modulate circadian gene expression (Rutter et al., 2001). The NAD^+-dependent deacetylase SIRT1 regulates CLOCK acetylation targets like BMAL1 and histone H3 at circadian promoters (Nakahata et al., 2008). SIRT1 also promotes circadian gene expression via PER2 deacetylation (Asher et al., 2008). Expression of the NAD-biosynthesis enzyme nicotinamide phosphoribosyltransferase (NAMPT) and the NAD level also show circadian oscillations. Indeed, the *Nampt* gene itself is directly bound by CLOCK for upregulation of gene expression (Ramsey et al., 2009). In this context, we find that transcript levels of CLOCK downstream targets *Per* 1, 2, and 3 and *Cry* 2 are upregulated in the RNA-seq datasets

(F) Electron micrographs of negative-stained NUDT12-BLMH complexes. The magnification used and scale are indicated. The different views of the complex are shown, likely indicating a collapse of the complex during the preparation of the grid. The complex shows a central barrel of BLMH (double layer of trimers) with three spokes composed of NUDT12 dimers. Scale bar in nanometer (nm) is indicated.
(G) Decapping assay with m^7G-RNA and indicated NUDT12 proteins, carried out in presence of increasing amounts of BLMH.
(H) Cleavage assays with labeled NAD analog and FL NUDT12, carried out in presence of increasing amounts of BLMH or a control protein (BSA).
See also Figures S3B and S3C.

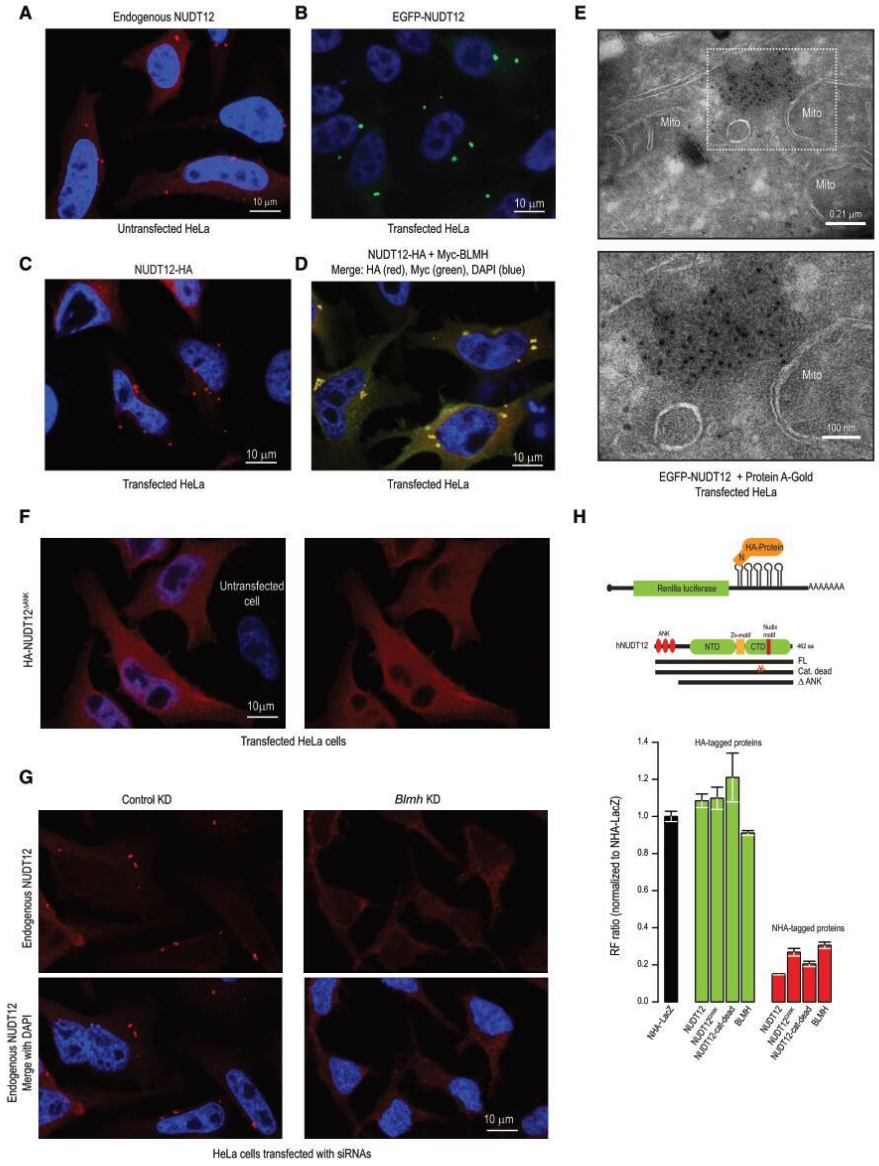


Figure 4. BLMH Facilitates Localization of NUDT12 to Cytoplasmic Granules and Downregulate Reporter Gene Expression

(A) Localization of endogenous NUDT12 to a few discrete cytoplasmic granules in human HeLa cell cultures. Scale bar in micrometer (μm) is indicated.

(B) Localization of EGFP-NUDT12 to discrete cytosolic granules and these do not overlap with P-bodies (see Figure S4A).

(C and D) Localization of HA-tagged NUDT12 (C) and its co-localization with tagged BLMH (D).

(E) Immunogold labeling and transmission electron microscopy of EGFP-NUDT12 (black spots) in transfected HeLa cells. The protein is present in electron-dense structures. Mitochondria (mito) are indicated.

(legend continued on next page)

from *Nudt12* knockout livers and to some extent in kidneys (Figure 5F). Since expression of these genes oscillate, we carried out RNA-seq analysis of mouse liver (n = 1) across a 24-h light-dark cycle, at 4-h intervals (Figure 5G; STAR Methods). This shows the expected peak expression of *Per1*, *Per2*, and *Cry2* at the beginning of the dark period (night; zeitgeber time [ZT] 12–16] in wild-type mouse liver (Figure 5G). In *Nudt12* knockout animals, the overall pattern is unchanged, but RNA levels are higher either during the daytime (ZT8) or at night (Figure 5G). This is further confirmed by qRT-PCR analysis using RNA isolated from triplicate biological replicates for each time point (Figure 5G). Our efforts at implementing the NADcaptureSeq protocol failed, so we are unable to determine the direct involvement of NAD-capping in regulation of circadian clock transcripts. Taken together, we propose that NUDT12 contributes to the tight regulation of circadian gene expression in mice.

DISCUSSION

In this study, we identify the Nudix family member NUDT12 as a cytosolic decapping enzyme for NAD-RNAs. NUDT12 is composed of three ANK repeats at the N terminus, followed by a catalytic domain at the C terminus. Here, we present the biochemical (Figure 1) and structural (Figure 2) analyses of the C-terminal catalytic domain of human NUDT12. While this study was in preparation, Grudzien-Nogalska et al. (2019) reported biochemical and structural analysis of mouse NUDT12, similar to what we report here. Our work goes beyond that study in identifying BLMH as an interaction partner of NUDT12 (Figure 3) in revealing the functional role of BLMH in localization and function of NUDT12 (Figure 4) and by investigating the physiological role of NUDT12 using a knockout mouse mutant (Figure 5).

Once transcribed, NAD-RNAs are subject to quality control by the nuclear DXO (Picard-Jean et al., 2018), which decaps a subset of such RNAs (Jiao et al., 2017). However, many transcripts do escape this checkpoint, as downstream processing steps like splicing and polyadenylation do not seem to be affected, and mature RNA is made. This remains a mystery, as a central role is proposed for the m^7G capping and the cap-binding complex (CBC) in these processing steps. How NAD-RNAs are exported to the cytoplasm is also not known, as the general export receptor Nxf1 accepts cargos that are assembled based on the presence of a 5′ m^7G cap and consequent splicing events (Köhler and Hurt, 2007). Once in the cytoplasm, the fate of NAD-RNA is not clearly understood. Previous studies showed that ribosomes do not engage reporter NAD-RNAs transfected into mammalian cell cultures (Jiao et al., 2017), while recent work identified plant NAD-RNAs in polysomal fractions (Wang et al., 2019). Until recently (Grudzien-Nogalska et al., 2019), and including this study, a sentinel responsible for cytoplasmic surveillance of NAD caps was not known.

General eukaryotic cytoplasmic decay of m^7G-RNA is initiated via deadenylation, followed by DCP2-mediated decapping. The 5′ monophosphate end generated by DCP2 is the correct substrate for the 5′→3′ exoribonuclease XRN1. NAD-RNAs are not decapped by DCP2 (Jiao et al., 2017), hence the importance of NUDT12. In the case of NUDT12-mediated NAD-RNA decapping, the 5′ monophosphate end generated is suitable for further degradation by XRN1. On the other hand, decapping of m^7G-RNA by NUDT12 is not desirable in the cytoplasm, as the generated diphosphate 5′ ends are not substrates for XRN1. In fact, the artificial tethering experiments show that when recruited to an m^7G-capped reporter RNA, NUDT12 has the ability to downregulate gene expression (Figure 4H), pointing to a need to regulate its accessibility to the general pool of mRNAs. This is probably done by sequestering NUDT12 in cytosolic granules that are distinct from the P-bodies (Figure S4A). Here, we show that the interaction partner BLMH is required for sequestering NUDT12 in cytoplasmic granules via formation of a large complex (Figure 3).

We show that BLMH forms a large complex with NUDT12 (Figure 3) to sequester NUDT12 into cytosolic granules (Figure 4). The downregulation of gene expression by BLMH tethering (Figure 4H) confirms the potential *in vivo* functional significance of this interaction. Beyond these, why BLMH, a cysteine peptidase, is required in this role is less clear. Loss of *Blmh* in mice results in low-penetrant neonatal death, with only 65% of the neonates surviving beyond that stage. Such *Blmh* mutant animals are fertile but display tail dermatitis (Schwartz et al., 1999). Curiously, yeast BLMH or Gal6 is shown to bind single-stranded DNA and RNA (Xu and Johnston, 1994; Zheng and Johnston, 1998). Yeast BLMH is cytosolic and forms a homohexameric barrel-like structure, with the active sites being located in the central cavity (Joshua-Tor et al., 1995). The sequence-independent, high-affinity binding of nucleic acids is mediated by charged residues at the entrance of the central cavity. Mutation of these residues does not affect peptidase activity *in vitro* but increases sensitivity to bleomycin toxicity *in vivo*, indicating a role for RNA binding in *in vivo* functions (Zheng and Johnston, 1998). Rat BLMH also shows RNA-binding activity (Takeda et al., 1996), but human BLMH does not bind nucleic acids (O'Farrell et al., 1999). Like NUDT12 (Figure 5), rat BLMH is also highly expressed in liver and kidney (Kamata et al., 2007). We believe that BLMH might have a life of its own as a peptidase but is moonlighting here as a sequestration factor for NUDT12.

Overall, the loss of NUDT12 and its potential *in vivo* activity in controlling cytosolic NAD-RNA levels is tolerated in mice (Figure 5). This is perhaps due to compensation from other enzymes like nuclear DXO or further enzymes currently unknown. We note

(F) Localization of NUDT12 lacking (ΔANK) the three ANK repeats. An untransfected cell is indicated for comparison.
(G) Endogenous NUDT12 loses its granular localization in cells treated with anti-*Blmh* siRNAs. This shows that localization of NUDT12 to the granules depends on presence of endogenous BLMH. See also Figure S4D.
(H) Artificial tethering of indicated proteins to the *Renilla* luciferase BoxB mRNA (via N-peptide-BoxB interaction). Luciferase activities were measured from cell lysates co-expressing indicated proteins and the *Renilla* BoxB reporter mRNA. The RF ratio (signal from *Renilla* divided with that from firefly reporter for normalization of transfection efficiency) luciferase activity measurements was normalized to that obtained from NHA-LacZ (set as 1). NUDT12 and BLMH are active in repressing the reporter. Error bars represent SD.
See also Figure S5.

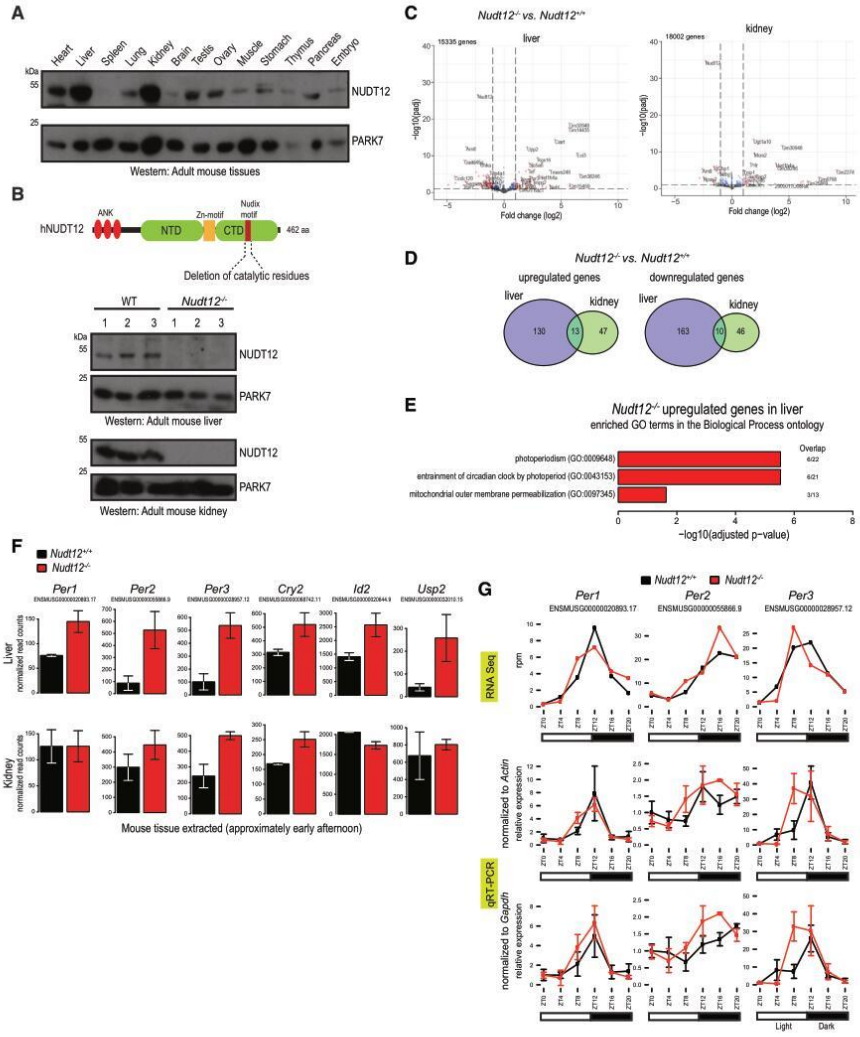


Figure 5. *Nudt12* Knockout Mouse Liver Displays an Elevated Level of Circadian Clock Transcripts

(A) Western analysis of NUDT12 with adult mouse tissues. PARK7 is used as a loading control.

(B) Cartoon showing creation of the *Nudt12* knockout mouse mutant. See also Figure S6A. Western analysis of tissue lysates from multiple biological replicates showing the absence of NUDT12 in liver and kidney of knockout animals.

(C) The volcano plots show the comparison of individual gene expression between *Nudt12*$^{-/-}$ and *Nudt12*$^{+/+}$. Genes with at least a 2-fold change and adjusted p value less than 0.1 are shown in red.

(D) The Venn diagrams compare the sets of differentially expressed genes in liver and kidney. Only a small proportion of genes are upregulated or downregulated in both tissues.

(E) Significantly enriched Gene Ontology (GO) terms in the Biological Process ontology are shown for the genes found to be upregulated in the liver of *Nudt12*$^{-/-}$.

(F) Expression of six *Nudt12*$^{-/-}$ upregulated genes associated with photoperiodism (GO:0009648) and entrainment of circadian clock by photoperiod (GO:0043153) is shown. The error bars represent SD between replicates. Note that for this experiment, animals were sacrificed for RNA extraction without any consideration of the circadian clock (but it was roughly in the afternoon).

(G) Gene expression analysis (at 4-h intervals) of three circadian clock genes using either RNA-seq (n = 1) or qRT-PCR (n = 3) in mouse livers over a 24-h controlled light-dark period. ZT, zeitgeber time. ZT0 corresponds to light turn ON at 7 a.m., while ZT12 corresponds to light turn OFF at 7 p.m. Error bars show SD.

that even in *E. coli* lacking NudC, no dramatic change was documented in the levels of small regulatory RNAs that are known to be NAD capped, again pointing to the existence of redundant mechanisms (Cahová et al., 2015). Indeed, although RppH was previously suspected to not decap NAD-RNAs (Cahová et al., 2015), the enzyme is able to remove NAD caps from certain RNAs *in vitro* (Grudzien-Nogalska et al., 2019). Although we document upregulation of circadian clock transcripts in the liver (Figures 5F and 5G), it is not known if this is due to NAD-capped versions of these RNAs being stabilized in the absence of NUDT12. We propose that NUDT12 participates in the tight regulation of circadian gene expression either directly by decapping NAD-RNAs from the clock genes or indirectly via regulation of other NAD-RNA substrates or by regulation of levels of the metabolites themselves.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animal Work
  - Nudt12 knockout mice
  - Genotyping
- METHOD DETAILS
  - Clones and constructs
  - Recombinant protein production
  - SEC-MALLS and AUC
  - Structure determination and refinement
  - Electron microscopy of NUDT12-BLMH complex
  - Antibodies
  - Cell culture
  - Transfection of cells for immunoprecipitation or immunofluorescence
  - Isolation of human NUDT12 and BLMH complexes from transfected cells for mass spectrometry
  - Mass spectrometry
  - Co-immunoprecipitations from transfected cells
  - Western Blot
  - Mouse multiple tissue western blot
  - Immunofluorescence of transfected cells
  - Immunogold labeling of GFP-NUDT12 in HeLa cells
  - Tethering experiments
  - NAD capped-RNA preparation by *in vitro* transcription
  - Preparation of RNA with labeled m^7G cap
  - *In vitro* decapping and deNADding assays
  - *In vitro* cleavage assays with labeled NAD
  - Metabolomics
  - Circadian experiments
  - Quantitative RT-PCR analysis of circadian gene expression
  - Preparation of RNA libraries
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - RNA-seq analysis
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.11.108.

ACKNOWLEDGMENTS

We thank Prof. Ueli Schibler for discussions and advice, Johanna Mattay for critical reading of the manuscript and kindly providing the DCP2 enzyme, and Nicolas Roggli for scientific illustration. We thank the following University of Geneva core facilities: iGE3 Genomics Platform for deep sequencing, the Transgenic Mouse Facility for mouse creation, the Bioimaging Centre at the Faculty of Sciences, and the Electron Microscopy Platform (PFMU). We thank the EMBL Genomics core facility for deep sequencing; the Functional Genomics Center, Zurich for proteomics analyses; Aline Le Roy of the Biophysical Platforme at IBS, Grenoble, France; and beamline staff at Diamond Light Source, Didcot, UK; EMBL-Hamburg, Germany, Synchrotron SOLEIL, Gif-sur-Yvette, France, especially M. Savko; and the Swiss Light Source, Villigen, Switzerland, especially V. Olieric. H.W. is grateful for a Long-Term Postdoctoral Fellowship from EMBO (ALTF 908-2017). This work was supported by grants to R.S.P. from the Swiss National Science Foundation (ERC Transfer Grant [GermMethylation], Project Grant [PIWIorigins], and funding from the NCCR RNA & Disease). Work in the Pillai lab is supported by the Republic and Canton of Geneva.

AUTHOR CONTRIBUTIONS

H.W. generated the mouse knockout mutant and performed all cell culture and mouse experiments; L.L. produced recombinant proteins and performed biochemical experiments; K.-M.C. created point mutants and solved the crystal structure of human NUDT12 with A.A.M.; D.H. did all computational analyses; F.F.-O. and P.G. carried out circadian experiments with H.W.; and R.S.P. prepared and wrote the manuscript with input from everyone.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 9, 2019
Revised: October 23, 2019
Accepted: November 26, 2019
Published: December 24, 2019

REFERENCES

Abdelraheim, S.R., Spiller, D.G., and McLennan, A.G. (2003). Mammalian NADH diphosphatases of the Nudix family: cloning and characterization of the human peroxisomal NUDT12 protein. Biochem. J. *374*, 329–335.

Akiyama, S., Ikezaki, K., Kuramochi, H., Takahashi, K., and Kuwano, M. (1981). Bleomycin-resistant cells contain increased bleomycin-hydrolase activities. Biochem. Biophys. Res. Commun. *101*, 55–60.

Albrecht, U. (2012). Timing to perfection: the biology of central and peripheral circadian clocks. Neuron *74*, 246–260.

Asher, G., Gatfield, D., Stratmann, M., Reinke, H., Dibner, C., Kreppel, F., Mostoslavsky, R., Alt, F.W., and Schibler, U. (2008). SIRT1 regulates circadian clock gene expression through PER2 deacetylation. Cell *134*, 317–328.

Bessman, M.J., Frick, D.N., and O'Handley, S.F. (1996). The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. J. Biol. Chem. *271*, 25059–25062.

Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. Trends Biochem. Sci. *37*, 49–57.

Bird, J.G., Zhang, Y., Tian, Y., Panova, N., Barvík, I., Greene, L., Liu, M., Buckley, B., Krásný, L., Lee, J.K., et al. (2016). The mechanism of RNA 5′ capping with NAD+, NADH and desphospho-CoA. Nature *535*, 444–447.

Bird, J.G., Basu, U., Kuster, D., Ramachandran, A., Grudzien-Nogalska, E., Towheed, A., Wallace, D.C., Kiledjian, M., Temiakov, D., Patel, S.S., et al. (2018). Highly efficient 5′ capping of mitochondrial RNA with NAD^+ and NADH by yeast and human mitochondrial RNA polymerase. eLife *7*, e42179.

Blighe, K. (2019). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling.

Borowczyk, K., Tisończyk, J., and Jakubowski, H. (2012). Metabolism and neurotoxicity of homocysteine thiolactone in mice: protective role of bleomycin hydrolase. Amino Acids *43*, 1339–1348.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., Vonrhein, C., et al. (2016). BUSTER (Global Phasing Ltd.).

Cahová, H., Winz, M.L., Höfer, K., Nübel, G., and Jäschke, A. (2015). NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. Nature *519*, 374–377.

Chen, H., and Boutros, P.C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics *12*, 35.

Chen, Y.G., Kowtoniuk, W.E., Agarwal, I., Shen, Y., and Liu, D.R. (2009). LC/MS analysis of cellular RNA reveals NAD-linked RNA. Nat. Chem. Biol. *5*, 879–881.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Cianci, M., Bourenkov, G., Pompidor, G., Karpics, I., Kallio, J., Bento, I., Roessle, M., Cipriani, F., Fiedler, S., and Schneider, T.R. (2017). P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. J. Synchrotron Radiat. *24*, 323–332.

Coleman, T.M., Wang, G., and Huang, F. (2004). Superior 5′ homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter. Nucleic Acids Res., e14.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Furuichi, Y., and Shatkin, A.J. (2000). Viral and cellular mRNA capping: past and prospects. Adv. Virus Res. *55*, 135–184.

Grudzien-Nogalska, E., Wu, Y., Jiao, X., Cui, H., Mateyak, M.K., Hart, R.P., Tong, L., and Kiledjian, M. (2019). Structural and mechanistic basis of mammalian Nudt12 RNA deNADding. Nat. Chem. Biol. *15*, 575–582.

Höfer, K., Li, S., Abele, F., Frindert, J., Schlotthauer, J., Grawenhoff, J., Du, J., Patel, D.J., and Jäschke, A. (2016). Structure and function of the bacterial decapping enzyme NudC. Nat. Chem. Biol. *12*, 730–734.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods *12*, 115–121.

Inesta-Vaquera, F., and Cowling, V.H. (2017). Regulation and function of CMTR1-dependent mRNA cap methylation. Wiley Interdiscip. Rev. RNA *8*, e1450.

Jiao, X., Doamekpor, S.K., Bird, J.G., Nickels, B.E., Tong, L., Hart, R.P., and Kiledjian, M. (2017). 5′ end nicotinamide adenine dinucleotide cap in human cells promotes RNA decay through DXO-mediated deNADding. Cell *168*, 1015–1027.e1010.

Joshua-Tor, L., Xu, H.E., Johnston, S.A., and Rees, D.C. (1995). Crystal structure of a conserved protease that binds DNA: the bleomycin hydrolase, Gal6. Science *269*, 945–950.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kamata, Y., Itoh, Y., Kajiya, A., Karasawa, S., Sakatani, C., Takekoshi, S., Osamura, R.Y., and Takeda, A. (2007). Quantification of neutral cysteine protease bleomycin hydrolase and its localization in rat tissues. J. Biochem. *141*, 69–76.

Kim, E., Kwak, H., and Ahn, K. (2009). Cytosolic aminopeptidases influence MHC class I-mediated antigen presentation in an allele-dependent manner. J. Immunol. *183*, 7379–7387.

Köhler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. *8*, 761–773.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44* (*W1*), W90–W97.

Lazo, J.S., and Humphreys, C.J. (1983). Lack of metabolism as the biochemical basis of bleomycin-induced pulmonary toxicity. Proc. Natl. Acad. Sci. USA *80*, 3064–3068.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Cryst. *40*, 658–674.

McLennan, A.G. (2006). The Nudix hydrolase superfamily. Cell. Mol. Life Sci. *63*, 123–143.

Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M., and Sliz, P. (2013). Collaboration gets the most out of software. eLife *2*, e01456.

Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., and Peng, Z.Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. Protein Sci. *13*, 1435–1448.

Nakahata, Y., Kaluzova, M., Grimaldi, B., Sahar, S., Hirayama, J., Chen, D., Guarente, L.P., and Sassone-Corsi, P. (2008). The NAD+-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control. Cell *134*, 329–340.

O'Farrell, P.A., Gonzalez, F., Zheng, W., Johnston, S.A., and Joshua-Tor, L. (1999). Crystal structure of human bleomycin hydrolase, a self-compartmentalizing cysteine protease. Structure *7*, 619–627.

Paglia, G., Williams, J.P., Menikarachchi, L., Thompson, J.W., Tyldesley-Worster, R., Halldórsson, S., Rolfsson, O., Moseley, A., Grant, D., Langridge, J., et al. (2014). Ion mobility derived collision cross sections to support metabolomics applications. Anal. Chem. *86*, 3985–3993.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Picard-Jean, F., Brand, C., Tremblay-Létourneau, M., Allaire, A., Beaudoin, M.C., Boudreault, S., Duval, C., Rainville-Sirois, J., Robert, F., Pelletier, J., et al. (2018). 2′-O-methylation of the mRNA cap protects RNAs from decapping and degradation by DXO. PLoS ONE *13*, e0193804.

Pillai, R.S., Artus, C.G., and Filipowicz, W. (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. RNA *10*, 1518–1525.

R Core Team (2017). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Ramsey, K.M., Yoshino, J., Brace, C.S., Abrassart, D., Kobayashi, Y., Marcheva, B., Hong, H.K., Chong, J.L., Buhr, E.D., Lee, C., et al. (2009). Circadian clock feedback cycle through NAMPT-mediated NAD+ biosynthesis. Science *324*, 651–654.

Rutter, J., Reick, M., Wu, L.C., and McKnight, S.L. (2001). Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. Science *293*, 510–514.

Saini, C., Liani, A., Curie, T., Gos, P., Kreppel, F., Emmenegger, Y., Bonacina, L., Wolf, J.P., Poget, Y.A., Franken, P., and Schibler, U. (2013). Real-time

recording of circadian liver gene expression in freely moving mice reveals the phase-setting behavior of hepatocyte clocks. Genes Dev. *27*, 1526–1536.

Schibler, U., Gotic, I., Saini, C., Gos, P., Curie, T., Emmenegger, Y., Sinturel, F., Gosselin, P., Gerber, A., Fleury-Olela, F., et al. (2015). Clock-talk: interactions between central and peripheral circadian oscillators in mammals. Cold Spring Harb. Symp. Quant. Biol. *80*, 223–232.

Schmidt, E.E., and Schibler, U. (1995). Cell size regulation, a mechanism that controls cellular RNA accumulation: consequences on regulation of the ubiquitous transcription factors Oct1 and NF-Y and the liver-enriched transcription factor DBP. J. Cell Biol. *128*, 467–483.

Schwartz, D.R., Homanics, G.E., Hoyt, D.G., Klein, E., Abernethy, J., and Lazo, J.S. (1999). The neutral cysteine protease bleomycin hydrolase is essential for epidermal integrity and bleomycin resistance. Proc. Natl. Acad. Sci. USA *96*, 4680–4685.

Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. Cell *9*, 645–653.

Sonenberg, N., Rupprecht, K.M., Hecht, S.M., and Shatkin, A.J. (1979). Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. Proc. Natl. Acad. Sci. USA *76*, 4345–4349.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521.

Song, M.G., Bail, S., and Kiledjian, M. (2013). Multiple Nudix family proteins possess mRNA decapping activity. RNA *19*, 390–399.

Stoltze, L., Schirle, M., Schwarz, G., Schröter, C., Thompson, M.W., Hersh, L.B., Kalbacher, H., Stevanovic, S., Rammensee, H.G., and Schild, H. (2000). Two new proteases in the MHC class I processing pathway. Nat. Immunol. *1*, 413–418.

Takeda, A., Masuda, Y., Yamamoto, T., Hirabayashi, T., Nakamura, Y., and Nakaya, K. (1996). Cloning and analysis of cDNA encoding rat bleomycin hydrolase, a DNA-binding cysteine protease. J. Biochem. *120*, 353–359.

Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2017). STARANISO (Global Phasing Ltd.).

van der Linden, W.A., Segal, E., Child, M.A., Byzia, A., Drąg, M., and Bogyo, M. (2015). Design and synthesis of activity-based probes and inhibitors for bleomycin hydrolase. Chem. Biol. *22*, 995–1001.

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. Acta Crystallogr. D Biol. Crystallogr. *67*, 293–302.

Walters, R.W., Matheny, T., Mizoue, L.S., Rao, B.S., Muhlrad, D., and Parker, R. (2017). Identification of NAD+ capped mRNAs in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *114*, 480–485.

Wang, Z., Jiao, X., Carr-Schmid, A., and Kiledjian, M. (2002). The hDcp2 protein is a mammalian mRNA decapping enzyme. Proc. Natl. Acad. Sci. USA *99*, 12663–12668.

Wang, Y., Li, S., Zhao, Y., You, C., Le, B., Gong, Z., Mo, B., Xia, Y., and Chen, X. (2019). NAD^+-capped RNAs are widespread in the *Arabidopsis* transcriptome and can probably be translated. Proc. Natl. Acad. Sci. USA *116*, 12094–12102.

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

Xu, H.E., and Johnston, S.A. (1994). Yeast bleomycin hydrolase is a DNA-binding cysteine protease. Identification, purification, biochemical characterization. J. Biol. Chem. *269*, 21177–21183.

Zhang, D., Liu, Y., Wang, Q., Guan, Z., Wang, J., Liu, J., Zou, T., and Yin, P. (2016). Structural basis of prokaryotic NAD-RNA decapping by NudC. Cell Res. *26*, 1062–1066.

Zhang, H., Zhong, H., Zhang, S., Shao, X., Ni, M., Cai, Z., Chen, X., and Xia, Y. (2019). NAD tagSeq reveals that NAD^+-capped RNAs are mostly produced from a large number of protein-coding genes in *Arabidopsis*. Proc. Natl. Acad. Sci. USA *116*, 12072–12077.

Zheng, W., and Johnston, S.A. (1998). The nucleic acid binding activity of bleomycin hydrolase is involved in bleomycin detoxification. Mol. Cell. Biol. *18*, 3580–3585.

Zimny, J., Sikora, M., Guranowski, A., and Jakubowski, H. (2006). Protective mechanisms against homocysteine toxicity: the role of bleomycin hydrolase. J. Biol. Chem. *281*, 22485–22492.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-XRN1	Abcam	Cat. No. ab70259; RRID: AB_1271494
Rabbit anti-Catalase	Cell Signaling Technology	Cat. no. 12980S; RRID: AB_2798079
Rabbit anti-PARK7	Thermo Fisher Scientific	Cat. No. PA5-13404; RRID: AB_2160112
Anti-HA Affinity Matrix	Roche	Cat. no. 11815016001; RRID: AB_390914
Anti-FLAG M2 Magnetic Beads	Sigma	Cat. no. M8823
Rabbit anti-hNUDT12	This study	N/A
Rabbit anti-hBLMH	This study	N/A
Mouse anti-FLAG	Sigma	Cat. no. F3165
Rabbit anti-HA	Abcam	Cat. no. ab9110
Mouse anti-MYC	Cell Signaling Technology	Cat. no. 2276s
Rabbit polyclonal Anti-GFP	Thermo Fisher Scientific	A-6455
Protein-A-10nm Gold conjugate	Cytodiagnostics	SKU:AC-10-05
Pierce Anti-HA Agarose beads	Thermo Fisher	Cat. no. cat.no. #26181
EZview Red Anti-c-Myc Affinity Gel	Sigma	Cat. no. E6654
Anti-rabbit IgG HRP-linked	GE Healthcare	Cat. no. NA934
Anti-mouse IgG HRP-linked	GE Healthcare	Cat. no. NA931
Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	Cat. no. A11034
Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 594	Invitrogen	Cat. no. A11037
Goat anti-Mouse IgG (H+L), Superclonal Recombinant Secondary Antibody, Alexa Fluor 488	Invitrogen	Cat. no. A28175
Goat anti-Mouse IgG (H+L) Cross-Adsorbed ReadyProbes Secondary Antibody, Alexa Fluor 594	Invitrogen	Cat. no. R37121
Bacterial and Virus Strains		
DH10EMBacY bacterial strain	Bieniossek et al., 2012	N/A
Top10	This study	N/A
BL21(DE3)	This study	N/A
priHC bacterial strain	Bieniossek et al., 2012	N/A
Chemicals, Peptides, and Recombinant Proteins		
Sodium deoxycholate	Sigma	30968
Complete EDTA-free protease inhibitor	Roche	11 873 580 001
NAD,[adenylate-32P]	Perkin Elmer	BLU023X250UC
Guanosine 5′-triphosphate, [alpha-32P]	Perkin Elmer	BLU006H
m7G(5′)ppp(5′)A RNA Cap Structure Analog	NEB	S1405S
β-Nicotinamide adenine dinucleotide (NAD+)	NEB	B9007S
Ponceau S	Sigma	P3504
Dulbecco's modified Eagle Medium	Invitrogen	cat. No. 21969-035
fetal bovine serum	Thermo Fisher	cat. no. 10270106
Penicilline/Streptomycin	Thermo Fisher	cat. No. 15140122
Glutamine	Thermo Fisher	cat. no. 15140122
Trypsin-EDTA 0.05%	Thermo Fisher	cat. no. 25300-054
Lipofectamine 3000 Transfection Reagent	Invitrogen	L3000015
Opti-MEM Medium	Thermo Fisher	11058021

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
3xFLAG peptide trifluoroacetate salt	BACHEM	cat. no. H-7536
30% acrylamide (37.5:1)	National Diagnostic	cat. no. EC-890
N,N,N',N'-Tetramethylethylendiamin	Merck	cat. no. 1107320100
Tween20	SIGMA	cat. no. P7949
Amersham Prime Western Blotting Detection Reagen	GE Healthcare	RPN2232
SuperSignal West Femto Maximum Sensitivity Substrate	Thermo Fisher	cat. no. 34095
Pierce ECL 2 Substrate	Thermo Fisher	cat. no. 1896433A
Critical Commercial Assays/Kits		
MinElute Gel Extraction Kit	QIAGEN	28604
MEGAshortscript T7 Transcription Kit	Life technologies	Cat. no. AM1354
Pierce Detergent Compatible Bradford Assay Kit	Thermo Fisher	23246
Dual-Luciferase Reporter Assay System kit	Promega	E1910
SuperScript III Reverse Transcriptase kit	Thermo Fisher	18080085
LightCycler 480 Sybr Green I Master	Roche	04707516001
The TruSeq Stranded Total RNA kit	Illumina	RS-122-2301
Deposited Data		
Deep sequencing datasets	This study	GEO: GSE138019
All raw gel data will be deposited at Mendeley Data.	This study	https://doi.org/10.17632/ykzx9h339r.1
Structure: NUDT12-core in complex with m^7GTP and Cd^{2+}	This study	PDB: 6SCX
Diffraction images: NUDT12-core in complex with m^7GTP and Cd^{2+}	This study	https://doi.org/10.15785/SBGRID/716
Diffraction images: NUDT12-core	This study	https://doi.org/10.15785/SBGRID/719
Experimental Models: Cell Lines		
Sf9 insect cells for protein production	Eukaryotic Expression Facility, EMBL Grenoble, France	N/A
High Five insect cells for protein production	Eukaryotic Expression Facility, EMBL Grenoble, France	N/A
Experimental Models: Organisms/Strains		
Mouse: *Nudt12 knock-out*	This study	Available from Lead Contact
Oligonucleotides		
DNA and RNA oligos		See Table S1
Recombinant DNA		
pACEBac2	Bieniossek et al., 2012	N/A
Human *Nudt12* cDNA	This study	NP_113626
Human *Blmh* cDNA	This study	NP_000377
pIDK	Bieniossek et al., 2012	N/A
Software and Algorithms		
ENRICHR	Chen et al., 2013; Kuleshov et al., 2016	http://amp.pharm.mssm.edu/Enrichr/
R	R Core Team, 2017	https://www.r-project.org
DESeq2	Love et al., 2014	N/A
Bioconductor	Huber et al., 2015	https://www.bioconductor.org/
Salmon	Patro et al., 2017	N/A
tximport	Soneson et al., 2015	N/A
EnhancedVolcano	Blighe, 2019	https://github.com/kevinblighe/EnhancedVolcano
Phaser	McCoy et al., 2007	https://www.phaser.cimr.cam.ac.uk
XDS suite	Kabsch, 2010	http://xds.mpimf-heidelberg.mpg.de
autoPROC	Vonrhein et al., 2011	http://www.globalphasing.com/autoproc/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
STARANISO	Tickle et al., 2017	http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi
Coot	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot
BUSTER	Bricogne et al., 2016	http://www.globalphasing.com/buster
MOLPROBITY	Chen et al., 2010	http://molprobity.biochem.duke.edu
PyMOL	Molecular Graphics System, Version 1.8.6 Schrodinger, LLC	https://www.pymol.org/2/
SBgrid	Morin et al., 2013	https://sbgrid.org/
Other		
Chelating Sepharose Fast Flow beads	GE Healthcare	17-0575-01
StrepTrap HP	GE Healthcare	28-9075-46
Superdex S75 10/300 GL	GE Healthcare	17-5174-01
Superdex 200 10/300 GL	GE Healthcare	17-5175-01
MethaPhor agarose	Lonza	50180
Amersham Hyperfilm ECL	GE Healthcare	cat. no. 28906837
Amersham Protran 0.45 μm nitrocellulose membrane	GE Healthcare	cat. no. 10600002

LEAD CONTACT AND MATERIALS AVAILABILITY

All unique reagents generated in this study like polyclonal antibodies, the *Nudt12* knockout mouse mutant etc are available from the Lead Contact without any restriction. Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Ramesh S. Pillai (ramesh.pillai@unige.ch).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal Work

Mutant mice were generated at the Transgenic Mouse Facility of University of Geneva. The mice were bred in the Animal Facility of Sciences III, University of Geneva. The use of animals in research at the University of Geneva is regulated by the Animal Welfare Federal Law (LPA 2005), the Animal Welfare Ordinance (OPAn 2008) and the Animal Experimentation Ordinance (OEXA 2010). The Swiss legislation respects the Directive 2010/63/EU of the European Union. Any project involving animals has to be approved by the Direction Générale de la Santé and the official ethics committee of the Canton of Geneva, performing a harm-benefit analysis of the project. Animals are treated with respect based on the 3Rs principle in the animal care facility of the University of Geneva. We use the lowest number of animals needed to conduct our specific research project. Discomfort, distress, pain and injury is limited to what is indispensable and anesthesia and analgesia is provided when necessary. Daily care and maintenance are ensured by fully trained and certified staff. This particular work was approved by the Canton of Geneva (GE/18/19).

Nudt12 knockout mice

The *Nudt12* gene locus is located on mouse chromosome 17 and consists of seven exons (Figure S6A). We targeted the *Nudt12* locus in mouse embryos of the B6D2F1/J hybrid line (also called B6D2; The Jackson Laboratory, stock no. 100006). It is a cross between C57BL/6J (B6) and DBA/2J (D2), and heterozygous for all B6 and D2 alleles. Single-cell mouse embryos were injected with two different guide RNAs (gRNAs) that direct the DNA endonuclease Cas9 to delete a region within exon 6. Founder mice were identified by genotyping PCR and crossed with wild-type C57BL/6J (Janvier) partners to obtain germline transmission. We identified a mouse line with 58 bp deletion in the exon 6 that removes the critical catalytic residues that constitute the Nudix motif. It also creates a premature termination codon that leads to the removal of the transcripts via nonsense-mediated decay (NMD). Heterozygous *Nudt12*$^{+/-}$ and homozygous *Nudt12*$^{-/-}$ mice of both sexes were viable and fertile. Western blot analysis confirms the homozygous mutant mice to be clean knockouts as they lack the NUDT12 protein (Figure 5B). We did not find any obvious phenotypic alterations in the knockout mutant. However, sequence analysis of transcripts in the mutant mouse liver reveals a role for NUDT12 in tight regulation of circadian transcripts (Figures 5F and 5G).

Preparation of gRNAs

A cloning-free method was used to prepare DNA template for *in vitro* transcription of the chimeric crRNA-tracrRNA, termed single guide RNA (sgRNA or gRNA). In brief, a common reverse primer (CRISPR sgR primer) and a gene specific forward primer (CRISPR

F primer) with T7 promoter sequence was used to PCR amplify the single-stranded sgDNA template. Primer sequences are provided in Table S1.

Forward (F) primer design template:

5′-GAAATTAATACGACTCACTATAGGNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGC-3′
N represent the gene-specific sequence.

The following components were mixed to prepare the PCR reaction: 20 μl 5X Phusion HF buffer, 67 μl ddH2O, 2 μl 10 mM dNTPs, 5 μl of 10 μM CRISPR F primer, 5 μl of 10 μM CRISPR sgR primer, and 1 μl Phusion DNA polymerase. The PCR reaction was set as follows: 98°C for 30 s, 35 cycles of [98°C for 10 s, 60°C for 30 s and 72°C for 15 s], 72°C for 10 min, and finally at 4°C to hold the reaction. The PCR product (~110bp) was agarose gel-purified using mini-elute gel extraction kit (QIAGEN, cat. no. 28604). The purified DNA was used to produce gRNA by *in vitro* transcription via the T7 promoter. *In vitro* transcription was carried out with the MEGAshortscript T7 Transcription Kit (Life technologies; cat no. AM1354) for 4 hours at 37°C. Reactions were treated with DNase I to remove template DNA, phenol-chloroform extracted and precipitated with ethanol. Quality of the generated gRNA was verified by 1.5% agarose gel electrophoresis.

Denaturing formaldehyde-agarose gel electrophoresis

Quality of generated gRNAs were verified by 1.5% agarose-formaldehyde gel electrophoresis. Agarose gel was prepared by mixing 0.75 g agarose, 36.5 mL H_2O, 5 mL of 10x MOPS buffer (0.2 M MOPS, 80 mM sodium acetate, 10 mM EDTA) and 8.5 mL of 37% formaldehyde. Approximately, 4 μg of RNA was dissolved in the 4xRNA loading buffer (50% formamide, 6.5% formaldehyde, MOPS buffer 1x, bromophenol blue 0.2%, ethidium bromide 50 μg/ml) and heated to 65°C for 10 min. RNA was loaded into the gel and run at 70V for approximately 90 minutes. Gel was imaged in the E-Box VX5 (Vilber Lourmat, France) UV transilluminator.

Preparation of injection mix

We mixed 12.5 ng/μl of each gRNA with 25 ng/μl of Cas9 mRNA (ThermoFischer Scientifique; A29378), in injection buffer (10 mM Tris pH 7.5, 1 mM EDTA, pH 8.0). Prepare aliquots of 20 μL and store at −80°C.

Injection of mouse embryos of the hybrid background B6D2F1/J (black coat color) was carried out at the Transgenic Mouse Core Facility, University Medical Centre (CMU), University of Geneva. The B6D2F1/J hybrid line (also called B6D2; The Jackson Laboratory, stock no. 100006) is a cross between C57BL/6J (B6) and DBA/2J (D2), and heterozygous for all B6 and D2 alleles. The NMRI (Naval Medical Research Institute) mice, which have a white coat color were used as foster mothers.

Genotyping

Ear-punches of the weaned animals (21 days-old) were digested in 100 μl of buffer containing 10 mM NaOH, 0.1 mM EDTA for 120 min at 95°C. After centrifugation at 3000 rpm for 10 min, 50 μl of supernatant was transferred to a new tube containing 50 μl of TE buffer (20mM Tris-HCl, pH 8.0 and 0.1 mM EDTA). An aliquot of 2 μl of the digestion mix was used for PCR. Primers to detect bands corresponding to the *Nudt12*$^{+/+}$ (280 bp, WT), and *Nudt12*$^{-/-}$ (222 bp, knock out) alleles were WH-Oligo104 and WH-Oligo106 (Table S1). Identity of the bands was confirmed by Sanger sequencing.

Reaction mix for 25 μl PCR reactions: 1 × Taq buffer (without $MgCl_2$, ThermoFisher cat. no. B38), 2 mM $MgCl_2$, 0.5 μl dNTPs mix (stock 10 mM), 0.5 μl primer mix (stock 10 nM each), 2.0 μl tail DNA (100-200 ng), 0.5 μl Taq Pol (EMBL Protein Expression Facility, Heidelberg), water to make 25 μl final volume. Reactions were run using the following conditions: 94°C for 3 min, 35 cycles of [94°C for 30 s, 60°C for 30 s and 72°C for 30 s], 72°C for 5 min, and finally at 4°C to hold the reaction. Reactions were examined by 1.5% agarose gel electrophoresis (Figure S6B).

Only male mice were used for Western analysis (Figure 5B), RNaseq (Figure 5F) and metabolomics (Figure S6C) experiments. Both male and female mice were used for the circadian experiments (Figure 5G).

METHOD DETAILS

Clones and constructs

Constructs for mammalian cell expression

Coding sequences for full-length (FL) human NUDT12 (462 aa; Accession number NP_113626) and human BLMH (455 aa; Accession number NP_000377) were amplified from human HeLa cell total RNA by reverse transcription-PCR (RT-PCR). Mammalian expression vectors (pCI-neo vector backbone) allowing production of fusion proteins with an N-terminal 3xFLAG-HA or N-terminal N-HA-tag or C-terminal HA-tag from a cytomegalovirus (CMV) promoter were used. We also used a vector allowing production of proteins with an N-terminal 3xMYC tag (pcDNA3.1 vector backbone) from a cytomegalovirus (CMV) promoter.

Sequence of the N-terminal 3xFLAG-HA tag is as following:

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATGACGATGACAAGggcggcagcggcTACC
CATATGATGTTCCAGATTACGCT

The N-HA tag refers to the N peptide and an HA tag. The N peptide is useful for artificial tethering to Box B hairpins (Pillai et al., 2004). N-HA sequence is as following:

ATGGACGCACAAACACGACGTGAGCGTCGCGCTGAGAAACAAGCTCAATGGAAAGCTGCAAACccaccgctcgaggctgcccaagct tacaccatggccTACCCATATGATGTTCCAGATTACGCT

Sequence of C-terminal HA tag is as following:

ggcggcagcggcTACCCATATGATGTTCCAGATTACGCTTAA

Sequence of the N-terminal 3xMYC tag is as following:

ATGGAGCAAAAGCTCATCTCAGAGGAAGATCTGGAATTAGAGCAAAAGCTCATCTCAGAGGAAGATCTGGAATTAGAG CAAAAGCTCATCTCAGAGGAAGATCTG

The following deletions or point mutant versions were prepared (either N-term or C-term tagged):

$NUDT12^{\Delta ANK}$ [(1-4 fused to 99-462 (lacking 5-98 aa)]; FL NUDT12 catalytic dead mutant (combined mutation of E370Q, E373Q, E374Q).

Constructs expressing full-length NUDT12 that carries five mutations (5A mutant; Y384A, Y281A, F283A, C284A and C287A) designed to make it monomeric were also prepared. A deletion version of the monomeric protein lacking the ANK repeats and another monomeric version having additional mutations of the catalytically-important residues (combined mutation of E370Q, E373Q, E374Q) were also prepared.

Luciferase reporters for tethering assay:

Renilla luciferase reporter with 5BoxB hairpins at the 3′ end is expressed from the Thymidine Kinase promoter: phRL TK 5BoxB sp 36 (lab DNA bank ID: clone#122; Addgene ID: 115365) (Pillai et al., 2004)

Firefly luciferase plasmid is pGL2 driven by the SV40 promoter (lab DNA bank ID: clone#209)

Constructs for insect cell expression

For the production of full-length proteins, we used Baculovirus-mediated expression in insect cells. The full-length (1-462 aa) human NUDT12 (hNUDT12, UniProtKB - Q9BQG2) was cloned into the pACEBac2 acceptor vector (Bieniossek et al., 2012) for expression as an N-terminal 6xHis-TEV fusion in the insect cells. The full-length (1-455 aa) human BLMH (hBLMH, UniProtKB - Q13867) was cloned into pACEBac2 acceptor vector (Bieniossek et al., 2012) for expression as an N-terminal 6xHis-TEV fusion in the insect cells as well. For co-expression of human NUDT12 and BLMH, full-length coding sequence for hNUDT12 was cloned into the XhoI and KpnI restriction sites of the acceptor vector pACEBac2 to express the recombinant protein as an N-terminal 6xHis-TEV fusion. The full-length coding sequence for untagged hBLMH was cloned into the donor vector pIDK between KpnI/XhoI restriction sites. The proteins were co-expressed by taking advantage of the MultiBac system (Bieniossek et al., 2012), which allows the generation of multi-gene constructs via Cre-lox recombination. The acceptor and donor vectors were combined in Cre-mediated reaction in a total volume of 20 μl where 1 μg of each vector was mixed with 2 μl of 10x Cre buffer and 1 μl of Cre recombinase (NEB, cat no. M0298S). The reaction was incubated at 37°C for 1 h. After that, 10 μl of Cre reaction was transformed to 100 μl of competent TOP10 cells and plated on LB agar with appropriate antibiotics. The clones were verified by restriction digestion of the isolated plasmid, as well as by PCR.

Constructs for bacterial expression

Constructs lacking the N-terminal ankyrin repeats of hNUDT12 (103-462 aa or 115-462 aa or 111-462 aa) or its point mutants were cloned into the bacterial expression vector (pETM-11-SUMO vector; EMBL Protein Expression and Purification Core Facility) as 6xHis-Strep-SUMO-TEV fusions.

The following constructs were prepared: hNUDT12 (103-462 aa), hNUDT12 (103-462 aa) Y318A, hNUDT12 (103-462 aa) F356A, hNUDT12 (103-462 aa) E370Q, hNUDT12 (103-462 aa) E374Q, hNUDT12 (103-462 aa) Y384A, hNUDT12 (103-462 aa) W390A, hNUDT12 (103-462 aa) E415Q and hNUDT12 (103-462 aa) P443A.

For crystallization studies, we expressed the following human NUDT12 proteins: 103-462 aa, 115-462 aa, and 111-462 aa (this construct was used for structure determination).

All protein preparations used for biochemistry and structural work are without the tags unless specifically indicated. For NUDT12 (E370Q) and NUDT12 (E374Q) mutants, we were unable to remove the tag. Hence, tagged proteins were used for the experiments (Figure S2A). All the above protein preparations were homodimeric as determined by gel-filtration analyses.

To create the monomeric version of the NUDT12 (in the 103-462 aa backbone), we created a combined mutant with five point mutations (5A mutant; Y384A, Y281A, F283A, C284A and C287A) that is expected to disrupt intermolecular interactions, based on crystal structure of bacterial NudC homodimer (Höfer et al., 2016). The monomeric protein (Figure S1E) lacks *in vitro* biochemical activity

on m^7G-capped RNA (Figure 1G). However, note that NUDT12 protein (103-462 aa) carrying only the Y384A mutation (one of the five mutations needed for making the protein monomeric) lacks decapping activity on the m^7G-capped RNA, although it retains hydrolysis activity with ^{32}P-NAD (Figures 2G, S2B, and S2C).

Recombinant protein production

Production of full-length recombinant proteins was carried out in insect cell lines using the baculovirus expression system. The ovary-derived cell lines used are: High Five (Hi5) insect cell line originating from the cabbage looper (*Trichoplusia ni*) and the *Sf*9 cells derived from the fall army worm *Spodoptera frugiperda*. Briefly, recombinant full-length hNUDT12 coding sequence was cloned into pACEBac2 acceptor vector. Plasmids were transformed into DH10EMBacY competent cells for recombination with the baculovirus genomic DNA (bacmid). The bacmid DNA was extracted and transfected with FuGENE HD (Promega, cat. no. E231A) into the *Sf*9 insect cells for virus production. The supernatant (V_0) containing the recombinant baculovirus was collected after 72 to 96 hours post-transfection. To expand the virus pool, 3.0 mL of the V_0 virus stock was added into 25 mL of *Sf*9 (0.5×10^6/mL) cells. The resulting cell culture supernatant (V_1) was collected 24 h post-proliferation arrest. For large-scale expression of the protein, Hi5 cells were infected with virus (V_1) and cells were harvested 72 h after infection.

For bacterial expression, plasmids were transformed into the *E. coli* BL21(DE3) strain and expression was induced by addition of 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) when the culture density reached around 0.6 (OD_{600}). The proteins were then expressed overnight at 22°C following induction.

Purification of NUDT12 and BLMH

The insect cells were collected by centrifugation and lysed by sonication in buffer: 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5% Glycerol, 5 mM 2-mercaptoethanol, 40 mM Imidazole and protease inhibitor (Thermo Scientific, EDTA-free). After incubation for 1-2 hours with Ni^{2+} chelating Sepharose FF beads, the beads were washed by imidazole gradient washing buffer and finally bound protein was eluted with 250 mM imidazole in lysis buffer. Subsequently, His tag was cut by the TEV protease overnight in the dialysis buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM 2-mercaptoethanol). The cleaved tag was removed by a second purification on Nickel beads. The protein was further purified by gel-filtration chromatography (Superdex 200 increase for NUDT12 and Superose 6 increase for BLMH, GE Healthcare) in the buffer (25 mM Tris-HCl pH 8.0, 150 mM KCl, 5% glycerol and 1 mM DTT). The pure fractions were verified by SDS-PAGE electrophoresis and used for biochemical assays.

For NUDT12 mutants, bacterial cells were collected by centrifugation and suspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM Imidazole, 5% glycerol, 5 mM 2-mercaptoethanol and protease inhibitor). The subsequent procedures were the same as above and pure protein was flash frozen in liquid nitrogen till further use for biochemical assays.

Purification of NUDT12-BLMH complex

Insect cells co-expressing hNUDT12 and untagged hBLMH were resuspended in the lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM Imidazole, 5% glycerol and 5 mM 2-mercaptoethanol) supplemented with proteinase inhibitor (Thermo Scientific, EDTA-free), sonicated with MISONIX Sonicator S-4000 and the lysate was centrifuged at 18,000 rpm for 45 min at 4°C. The clarified supernatant was incubated at 4°C for 1h with the Ni^{2+} chelating Sepharose FF beads (GE Health; cat. no. 17057501). The beads were washed with an imidazole gradient in the wash buffer (with 40 mM, 50mM or 60 mM imidazole in lysis buffer). Finally, His-tag proteins bound to the beads were eluted with the elution buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 250 mM Imidazole, 5% glycerol and 5 mM 2-mercaptoethanol). Such a preparation of 6xHis-NUDT12 has the co-expressed untagged BLMH co-purifying with it (Figure 3C). For biochemical and biophysical experiments, the tag was removed. Briefly, the N-terminal 6xHis tag was cleaved overnight with TEV in the dialysis buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM 2-mercaptoethanol). After cleavage, second Ni-column purification was performed and flow-through containing the cleaved protein (complex) was collected. Proteins were further purified by gel filtration chromatography using Superose 6 increase 10/300GL (GE Healthcare) equilibrated with gel-filtration buffer containing: 25 mM Tris-HCl pH 8.0, 150 mM KCl, 5% glycerol and 1 mM DTT. The fractions were checked by SDS-PAGE electrophoresis and pure untagged hNUDT12-hBLMH protein complex was flash-frozen in liquid nitrogen after addition of 10% glycerol.

SEC-MALLS and AUC

Size-exclusion chromatography coupled to multi-angle laser light scattering (SEC-MALLS) and analytical ultra-centrifugation (AUC) experiments were carried out at the Grenoble Instruct-ERIC Center, within the Grenoble Partnership for Structural Biology (PSB), France.

For SEC-MALLS experiments, NUDT12-BLMH samples were stored at 4°C and centrifuged 15 min at 168,000 x g at 4°C before injection. An aliquot (100 μl) of the NUDT12-BLMH sample (3 mg/ml) was injected into a Superose 6 column (GE Healthcare), equilibrated at 4°C, with buffer: 25 mM Tris pH 8.0, 150 mM KCl, 5% Glycerol, 1mM DTT. An aliquot (90 μl) of BSA (2 mg/ml) was also injected into the same column as a control. Following experiments were conducted on a HPLC (Schimadzu) consisting of a degasser DGU-20AD, a LC-20AD pump, a autosampler SIL20-ACHT, a communication interface CBM-20A and a UV-Vis detector SPD-M20A, a column oven XL-Therm (WynSep) and a static light scattering detector miniDawn Treos, a dynamic light scattering detector DynaPro NANOSTAR, a refractive index detector Optilab rEX (Wyatt). The analysis was performed using the ASTRA software (version 5.4.3.20) and relevant parameters were: Solvent: water; Refractive index: 1.331; Temperature of LS: 35°C.

For AUC analysis, the NUDT12 (dimeric or monomeric 5A mutant version) samples (3.3 mg/ml) were diluted by 2-fold and 10-fold with buffer: 25 mM Tris pH 8.0, 150 mM KCl, 10% Glycerol, 1mM DTT, respectively. Then Sedimentation velocity experiments were

performed at 42,000 rpm and 4°C, on an analytical ultracentrifuge XLI, with a rotor Anti-50 (Beckman Coulter) and double-sector cells of optical path length 12, 3 and 1.5 mm equipped of Sapphire windows (Nanolytics). The reference is the buffer: 25 mM Tris pH 8.0, 150 mM KCl, 10% Glycerol, 1 mM DTT and acquisitions were made using absorbance at 280 nm and interference optics. For the analysis, all data were processed by Redate software (version 1.0.1) and analyses done with SEDFIT (version 16.1) and Gussi (version 1.3.2.). Relevant parameters were: for buffer, calculated with SEDNTERP software (version 20130813 Beta); for molecules, calculated with SEDFIT software (version 15.01b).

Crystallization and data collection

Optimal crystallization conditions for human NUDT12 (103-462 and 111-462 aa) were sought by robot screening at the High Throughput Crystallization Facility at EMBL Grenoble, France. Although both gave crystals that can be reproduced manually, only NUDT12 (111-462 aa) gave good diffraction. Crystals were manually produced: 2 μL protein solution at 10 mg/ml was manually mixed with 2 μL reservoir solution using the hanging drop method at room temperature. The reservoir condition used was 0.2 M tri-potassium citrate, 20% (w/v) PEG3350. The crystals were then flash-frozen at 100 K after transferring them to identical crystallization conditions containing 20% glycerol.

Diffraction data for this project were collected at Diamond Light Source (Didcot, UK) as proposal number MX21969, EMBL beamlines at the PETRA III storage ring (DESY, Hamburg, Germany) as proposal number MX-661, Synchrotron SOLEIL, (Gif-sur-Yvette, France) as proposal number 20181604, and the Swiss Light Source (Villigen, Switzerland) as proposal number e18077. Diffraction data reported here were collected on EMBL beamline P13 at the PETRA III storage ring (DESY, Hamburg, Germany) (Cianci et al., 2017) for native and PROXIMA 2 at SOLEIL (Gif-sur-Yvette, France) for crystals grown in the presence of 2mM 7-methyl-guanosine-5′-triphosphate (m^7GTP) and 2 mM $CdCl_2$. All data were integrated using the XDS suite (Kabsch, 2010). The diffraction data were highly anisotropic, with diffraction limits of ~3.0 Å and 2.9 Å along the best direction for native and m^7GTP/Cd crystal forms respectively, but only ~4.5 Å and 3.7 Å in the weakly diffracting directions. Therefore, data were processed using STARANISO (Tickle et al., 2017), as implemented in autoPROC (Vonrhein et al., 2011), which applies non-elliptical anisotropic limits based on a locally averaged mean $I/\sigma(I)$ cut-off, performs a Bayesian estimation of structure amplitudes, and applies an anisotropic correction to the data. Detailed crystallographic statistics are provided in Table 1.

Structure determination and refinement

While this study was in preparation, Grudzien-Nogalska et al. reported the crystal structure of mouse NUDT12 catalytic core in complex with AMP (PDB: 6O3P) (Grudzien-Nogalska et al., 2019). The human NUDT12 (111-462 aa) structure was therefore solved by molecular replacement using the mouse structure (PDB: 6O3P) as a search model with Phaser (McCoy et al., 2007). Several rounds of manual building with Coot (Emsley et al., 2010), and structure refinement with BUSTER (Bricogne et al., 2016) were carried out for all structures. As the m^7GTP/Cd co-crystal was clearly of better quality overall only this structure was refined to completion. MOLPROBITY (Chen et al., 2010) was used for model validation and all the crystallographic information is summarized in Table 1. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) for the m^7GTP/Cd form with the accession codes: 6SCX. Structural figures were prepared with PyMOL (Schrödinger, LLC). A comparison of the mouse NUDT12 (PBD: 6O3P) structure (Grudzien-Nogalska et al., 2019) with our human NUDT12 structure (PBD: 6SCX) reveals a very high degree of overlap (r.m.s deviation of 1.0 Å over 308 amino acids) (Figure 2B). Structural biology applications using in this project were compiled and configured by SBGrid (Morin et al., 2013).

Electron microscopy of NUDT12-BLMH complex

Carbon coated grid (CF300-CU-UL, Electron Microscopy Sciences) was glow-discharged and then used for negative-staining. Approximately, 5 ul of 0.01 mg/ml NUDT12-BLMH protein solution was placed onto the grid, incubated for 1 min and removed by filter paper. After that the grid was washed 3 times with buffer (25 mM Tris pH 8.0, 150 mM KCl, 1 mM DTT), then immediately stained with 2% uranyl acetate for 1 min. Staining reagent was removed by filter paper and the grids were allowed air-dry for 30 min. Electron microscopy was performed with Tecnai G2 Sphera microscope (FEI company) operated at 120 kV. All micrographs were acquired at 50,000 magnification corresponding to a pixel size of 0.2245 nm.

Antibodies

Commercial antibodies

Primary antibodies. mouse anti-FLAG (Sigma; F3165), rabbit anti-HA (abcam; ab9110), mouse anti-MYC (Cell Signaling Technology; 2276s), rabbit anti-XRN1 (Abcam; ab70259), rabbit anti-Catalase (Cell Signaling Technology; 12980S), rabbit anti-PARK7 (ThermoFischer Scientific; PA5-13404) were used for detecting proteins. Anti-HA affinity matrix (Roche; cat. no. 11815016001), anti-FLAG M2 Magnetic Beads (Sigma; cat. no. M8823), Pierce Anti-HA Agarose beads (ThermoFisher; cat.no. #26181), and EZview Red Anti-c-Myc Affinity Gel (Sigma; E6654) were used for immunoprecipitations.

Secondary antibodies. For western blot analyses, the following secondary antibodies conjugated to Horse Radish Peroxidase were used: anti-rabbit IgG HRP-linked (GE Healthcare; NA934), anti-mouse IgG HRP-linked (GE Healthcare; NA931); for immunofluorescence studies, the following secondary antibodies coupled to fluorescent dyes were used: anti-rabbit (Alexa Fluor 488; cat. no. A11034), anti-rabbit (Alexa Fluor 594; cat. no. A11037), anti-mouse (Alexa Fluor 488; cat. no. A28175), anti-mouse (Alexa Fluor 594; cat. no. R37121).

Antibodies generated for this study

We generated rabbit polyclonal antibodies to human NUDT12 and human BLMH. Two New Zealand White (NZW) rabbits were immunized with soluble antigens (Biotem, France). The antigen used was the purified recombinant human untagged NUDT12-BLMH complex produced in insect cells. For each injection, 1 mg/ml protein was used. After six injections (at day 0, 14, 28, 56, 70, 95), crude immune serum was collected (at day 39, 67, 82 and 104) and frozen. Since the crude sera contains a mixture of antibodies, we isolated specific antibodies to NUDT12 and BLMH via affinity-purification. Given that the human and mouse proteins are highly identical (NUDT12, 88% identical; BLMH, 93% identical), these antibodies allow us to detect proteins from both species. The antibodies were suitable for both western blot analysis and immunofluorescence studies.

Antibody production and purification

To purify antibodies against NUDT12, large amount of recombinant NUDT12 protein was resolved via SDS-PAGE and semi-dry transferred onto a nitrocellulose membrane. After reversible staining with Ponceau S (P3504, Sigma), part of the membrane containing the antigen was cut out and incubated overnight with 8 mL crude immune sera. After washes (1xPBS), bound antibodies were eluted with low pH solution (500 μL 0.1 M Glycine pH 2.5, 150 mM NaCl) and immediately neutralized with neutralization buffer (150 μL 0.5 M Tris-HCl pH 8.0, 150 mM NaCl). Antibodies were stored in 50% glycerol at −20°C. Rabbit polyclonal antibodies against BLMH were purified as above using the BLMH protein as the affinity matrix.

Cell culture

Human cervical carcinoma-derived HeLa cells were grown in Dulbecco's modified Eagle Medium (DMEM; Invitrogen, cat. No. 21969-035) supplemented with 10% fetal bovine serum (ThermoFisher; cat. no. 10270106), 1% Penicilline/Streptomycin (ThermoFisher; cat. No. 15140122), 1% 200 mM Glutamine (ThermoFisher; cat. no. 15140122), hereafter referred to as DMEM complete medium (DMEM CM), and maintained in an environment with 5% CO_2 at 37°C. When confluent, cells growing in 10cm^2 dish (Falcon, cat. no. 353003) were washed with warm (37°C) 1xPBS and incubated with 2 mL of Trypsin-EDTA 0.05% (ThermoFisher; cat. no. 25300-054) for 2 min to promote removal of cells from the growth surface. Subsequently, 10 mL warm DMEM media was added and cells were resuspended by pipetting. Cells were counted using a Bürker-Türk cell counter and appropriate cell numbers were seeded in cell culture dishes.

Transfection of cells for immunoprecipitation or immunofluorescence

For immunoprecipitation and mass spectrometry, cells were seeded into 10 cm^2 dishes and cultured as described above. After 12 hours, cells were transfected with plasmids by Lipofectamine 3000 Transfection Reagent (Invitrogen; L3000015). Briefly, 10 μg plasmids were diluted in 500 μL Opti-MEM Medium (ThermoFisher; 11058021) and 10 μL P3000 reagent. Simultaneously, 20 μL Lipofectamine 3000 were also diluted in 500 μL Opti-MEM Medium. After 5 min, they were mixed, vortexed vigorously for 15 s, and incubated for another 15 min. Next, the mixed reagents were transferred onto cells in the DMEM CM. After 6 hours, medium was exchanged with fresh DMEM CM and incubated for another 42 hours. For co-immunoprecipitation, cells were seeded into 6cm^2 dishes and cultured as described above, and 2 μg of each plasmid was transfected by Lipofectamine 3000 Transfection Reagent. The protocol is similar as above. For immunofluorescence, cells were seeded into Millicell EZ SLIDE 4-well glass (Sigma; PEZGS0416) and cultured as described above, and 500 ng of each plasmid was transfected by Lipofectamine 3000 Transfection Reagent. The protocol is same as above, but volume of transfection reagent was adjusted as per manufacturer recommendations.

Isolation of human NUDT12 and BLMH complexes from transfected cells for mass spectrometry

Transfected cells transiently expressing 3xFLAG-HA-NUDT12 and 3xFLAG-HA-BLMH were used for tandem-affinity purification. Cells were harvested 48h post-transfection. Briefly, the 10 cm^2 dishes were washed three times with ice cold PBS, and 1 mL of lysis buffer [20 mM Tris pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 0.1% sodium deoxycholate, 1 mM EDTA, 1 mM DTT, protease inhibitor (Complete Protease Inhibitor Cocktail Tablet; Roche, cat. no. 5056489001)] was added to the cells. Cells were removed from their growth surface using a cell scraper (Costar; cat. no. 3010) and transferred to 1.5 mL eppendorf tubes. After pipetting for 20 times, the suspended cells were incubated at 4°C for 1 hour for lysis. The total cell lysate was then spun at 13,000 x g for 15 min at 4°C. After centrifugation, the cleared lysate was transferred into a fresh tube. While 50 μl of lysate was transferred to a fresh tube and flash-frozen in liquid nitrogen to use as input, 950 μl was incubated with 20 μL of anti-FLAG M2 Magnetic Beads (Sigma; cat. no. M8823) for at least 3 hours at 4°C. After incubation, beads were collected and washed 5 times with wash buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.1% Triton X-100, 1 mM EDTA), and transferred into a fresh tube. To elute the complexes, the beads-bound complexes were incubated with 150 μL of FLAG elution buffer [150 μg/mL 3xFLAG peptide trifluoroacetate salt (BACHEM; cat. no. H-7536), 50 mM Tris pH 7.4, 150 mM NaCl], at 4°C for 20 min, with rocking (repeated cycles of 1200 rpm. 15 s., pause 30 s.). Incubation was repeated two more times (three times in total). All the eluted solutions were collected into a fresh tube. The eluted solution was incubated with 15 μL Anti-HA Affinity Matrix beads (Roche; cat. no. 11815016001) for 3 hours. Beads were collected by gentle centrifugation (500 x g for 1 min at 4°C) and the supernatant was discarded. Beads were washed 5 times with the same wash buffer and transferred into a new tube. 40 μl of 2x Laemmli buffer (4% SDS, 20% glycerol, 120 mM Tris-HCl pH 6.8, 10% β-mercaptoethanol, 0.02% bromophenol blue) was added. Beads were boiled at 95°C for 5 minutes and stored at −20°C until use for mass spectrometry.

Mass spectrometry

Mass spectrometry to confirm purified recombinant proteins was carried out at the Proteomics Core Facility, EMBL, Heidelberg. Identification of components within an immunoprecipitated complex was carried out at the Functional Genomics Center Zurich (ETH Zurich) using the shotgun mass spectrometry analysis. Database searches were performed using the Mascot (SwissProt, all species) search program. Applied settings, if not stated differently, are 1% protein false detection rate (FDR), min. 2 peptides per protein, 0.1% peptide FDR.

Co-immunoprecipitations from transfected cells

Harvest and lysis cells from 6 cm^2 dishes were as described above. Clear lysate of ~1ml was prepared and 50 μl removed as input, which was flash-frozen in liquid nitrogen and stored at −70°C. The remaining 950 μl was transferred into a new tube, 15 μl of the antibody/beads was added, and incubated for at least 3 hours at 4°C. Antibodies used are: Anti-FLAG M2 Magnetic Beads or Pierce Anti-HA Agarose beads or EZview Red Anti-c-Myc Affinity Gel (to immunoprecipitate Myc-BLMH to verify co-immunoprecipitation of 3xFLAG-HA-NUDT12). After incubation, beads were washed 5 times with wash buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.1% Triton X-100, 1 mM EDTA). Then beads were transferred into a fresh tube together with 40 μl of 2x Laemmli buffer (4% SDS, 20% glycerol, 120 mM Tris-HCl pH 6.8, 10% β-mercaptoethanol, 0.02% bromophenol blue). The beads and input [adding 10μl 6x Laemmli buffer (12% SDS, 60% glycerol, 360 mM Tris-HCL pH 6.8, 10% β-mercaptoethanol, 0.1% bromophenol blue)] were boiled at 95°C for 5 minutes and stored at −20°C until western blot analysis.

Western Blot

Whole cell lysates and suspensions of boiled beads were separated on SDS-PAGE gels in order to detect proteins of interest. SDS-PAGE gels were prepared using Ultra-Pure ProtoGel 30% acrylamide (37.5:1) (National Diagnostic; cat. no. EC-890) mixed with ultra-pure water and resolving gel buffer (0.375 M Tris, 0.1% SDS, pH 8.8) to obtain 8% resolving gel, and with stacking gel buffer (0.125 M Tris, 0.1% SDS, pH 6.8) to obtain 12% stacking gel. N,N,N',N'-Tetramethylethylendiamin (Merck, cat. no. 1107320100) and 10% ammonium persulfate were added to catalyze the polymerization reaction. Gel electrophoresis was performed at 90 V for 30 min. and then at 120 V for 90 min. After separation, proteins were blotted on the Amersham Protran 0.45 μm nitrocellulose membrane (GE Healthcare; cat. no. 10600002) overnight at 5 V at room temperature using Trans-Blot SD. Semi-Dry Transfer Cell system (Bio-Rad; cat. no. 1703940). After transfer, membranes were washed with Tris-buffered saline (TBS, 20 mM Tris, 150 mM NaCl, pH 7.6) and blocked for 1 hour at room temperature with 5% dry milk in TBS with 0.05% Tween20 (TTBS) (SIGMA; cat. no. P7949). After this, membranes were incubated with primary antibody (mouse anti-FLAG, 1:2000; mouse anti-MYC, 1:1000; rabbit anti-HA, 1:1000) for 3 hours at room temperature in 5% milk with TTBS. Then, membranes were washed 3 times for 10 minutes with TTBS and incubated with HRP-conjugated secondary antibody at 1:10 000 dilution, either with anti-rabbit IgG HRP-linked (GE Healthcare; NA934) or anti-mouse IgG HRP-linked (Invitrogen; a27025) for 1 hour at room temperature in TTBS. After 1 hour, membranes were washed 3 times for 10 minutes with TTBS and incubated with one of the following detection reagents: Amersham Prime Western Blotting Detection Reagent (GE Healthcare; RPN2232), SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher; cat. no. 34095) or Pierce ECL 2 Substrate (ThermoFisher; cat. no. 1896433A) for 5 min. at room temperature. Signal was detected using Amersham Hyperfilm ECL (GE Healthcare; cat. no. 28906837). The processed films were scanned using Perfection 3200 Photo scanner (Epson) with XSane image scanning software (ver. 0999).

Mouse multiple tissue western blot

Multiple tissues were isolated from an adult (> 60 days old) mouse. After flash-freezing in liquid nitrogen, a piece of different tissues were homogenized in 1ml lysis buffer [20 mM Tris pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 0.5% sodium deoxycholate, 1mM DTT, protease inhibitor (Complete Protease Inhibitor Cocktail Tablet, Roche; cat. no. 5056489001)]. After this, the lysate was transferred to a 1.5ml Eppendorf tube, centrifuged at 14000 g for 30min, and supernatant collected. An aliquot was taken to measure the concentration by Pierce Detergent Compatible Bradford Assay Kit (Thermo Fisher; 23246). Different tissue protein lysates were normalized to 1 mg/μl, SDS loading buffer added before boiling at 95°C for 10 min. Western blot analysis was performed as described above: for each well, 20 mg of protein was loaded. The blot was probed with affinity-purified anti-NUDT12 rabbit polyclonal antibodies. After stripping, the blot was re-probed with anti-PARK7 rabbit polyclonal antibodies, to serve as loading control.

Immunofluorescence of transfected cells

After transfection of cells for 48 hours, the media was removed and cells were washed three times with 1xPBS. Cells were fixed with 4% paraformaldehyde (PFA) by incubation at room temperature for 15 min. Cells were then washed three times with 1xPBS, followed by addition of 0.1M Glycine in 1xPBS, and incubated at room temperature for 4 min. Cells were then washed twice with 1xPBS. Cells were permeablized by incubation at room temperature for 4 min with 0.1% Triton X-100 in PBS. After this, cells were blocked for 15 min at room temperature with the blocking buffer (0.2% BSA, 1% goat serum, 0.1% Triton X-100 in PBS). During the block, primary antibodies were diluted in blocking buffer to proper concentration (each well need 100 μl) (rabbit anti-HA, 1:1000; mouse anti-FLAG, 1:2000; rabbit anti-NUDT12, 1:250; mouse anti-MYC, 1:1000; rabbit anti-XRN1, 1:250; rabbit anti-Catalase, 1:1000). The blocking buffer was removed from the well, and cells were incubated with primary antibody at room temperature for 3 hours. This was followed by removal of primary antibody, washed three times with 1xPBS. Proper dilutions of secondary antibodies in blocking buffer (all are

1:250) were prepared and cells were incubated with it for 1 hour at room temperature in dark. After this, cells were washed with 1xPBS three times. DAPI 0.5 μg/ml (diluted in PBS) was added to cells for 30 min at room temperature, and incubated in the dark. Wash 3 times with 1xPBS. Mount with VECTASHIELD Mounting Medium: Add one drop (10 ul) of Mount medium on slide, put the coverslip on it. Seal the coverslip to the slide with nail polish to prevent evaporation and the introduction of air bubbles. Store the slides in −20°C. Pictures were taken using Zeiss LSM710 confocal microscope (Bioimaging Center, University of Geneva).

Immunogold labeling of GFP-NUDT12 in HeLa cells

HeLa cells in a 10 cm^2 dish were transfected with a plasmid expressing N-terminal GFP-tagged NUDT12 (5 μg) following the protocol described above. After 2 days, the media was removed and the cells were washed three times with cold PBS. The cells were fixed at room temperature for 1 hour with 10 mL EM fixation buffer (2% Paraformaldehyde, 0.2% glutaraldehyde, in 0.1 M $NaP0_4$, pH 7,4 buffer). The cells were further washed three times with cold PBS before further sample preparation at the Electron Microscopy Platform (PFMU), CMU, University of Geneva. Briefly, cryo-sections were stained with rabbit polyclonal anti-GFP (ThermoFischer Scientifique, ref. A-6455), followed by detection with Protein-A-10nm Gold conjugate (Cytodiagnostics, SKU: AC-10-05). Electron microscopy was performed with Tecnai G2 Sphera microscope (FEI company) operated at 120 Kv (Bioimaging Facility of University of Geneva).

Tethering experiments

Human HeLa cells were seeded into 24-well plates. Once the cells reached 70%–80% confluency, they were transfected with required plasmids: NHA- or HA-fusion plasmid (320 ng), *Renilla* luciferase-BoxB plasmid (30 ng); firefly luciferase plasmid (150 ng). This makes a total of 500 ng plasmid in each well. The transfection reagent used was Lipofectamine 3000, with the same protocol as above. The N-peptide allows direct tethering of the fusion protein to the five BoxB hairpins present at the 3′UTR of the Renilla luciferase reporter (Pillai et al., 2004). The HA-tagged fusion (lacking the N-peptide) of the same protein is not expected to be tethered to the reporter, and serves as a negative control. The Firefly luciferase reporter plasmid serves as a normalization control for transfection efficiency in each well.

Luciferase reporter plasmids used:

Renilla luciferase reporter with 5BoxB hairpins at the 3′ end is expressed from the Thymidine Kinase promoter: phRL TK 5BoxB sp 36 (lab DNA bank ID: clone#122; Addgene ID: 115365) (Pillai et al., 2004)

Firefly luciferase plasmid is pGL2 driven by the SV40 promoter (lab DNA bank ID: clone#209)

The NHA- or HA-fusion plasmids used are as following:

pCI-neo-NHA-LacZ
pCI-neo-NHA-NUDT12
pCI-neo-HA-NUDT12
pCI-neo-NHA-NUDT12$^{\Delta ANK}$
pCI-neo-HA-NUDT12$^{\Delta ANK}$
pCI-neo-NHA-NUDT12 catalytic dead
pCI-neo-HA-NUDT12 catalytic dead
pCI-neo-NHA-BLMH
pCI-neo-HA-BLMH

Similar plasmids expressing the full-length monomeric version of NUDT12 and its mutants were also used.

After transfection for 48 hours, all the solutions were removed and luciferase activities of *Renilla* and Firefly luciferases measured in cell lysates using the Dual-Luciferase Reporter Assay System kit (Promega; E1910) and SPARK multimode microplate reader from TECAN. Briefly, 100 μl of 1xPLB (passive lysis buffer from kit) was added to each well, and rocked at room temperature for 15 min to lyse cells. After, the total lysate was transferred to a new Eppendorf tube, and spun down at 4°C for 10 min. 20 μl of clear lysate was added into each well of the 96 well plate for detection. 100 μl of LAR II was added to measure firefly luciferase activity; 100 μl of Stop & Glo reagent was added to measure *Renilla* luciferase activity. For each lysate/well, the Renilla activity (from the experimental *Renilla*-BoxB reporter RNA) was normalized with values obtained from firefly luciferase. Each plasmid transfection (eg NHA-LacZ) was done in triplicate wells (technical replicates) within an experiment that has a series of plasmids. Each set of experiments was repeated at least three times (biological replicates). Activity from tethering of various fusion proteins was normalized to that obtained from NHA-LacZ tethering, which was set as 1.

NAD capped-RNA preparation by *in vitro* transcription

A 40 nt NAD-capped RNA was generated by *in vitro* transcription as previously described (Jiao et al., 2017). The commonly used T7 class III promoter ϕ6.5 initiates using GTP, while the class II promoter ϕ2.5 initiates with ATP (Coleman et al., 2004). We prepared a single-stranded DNA template containing the T7 class II promoter ϕ2.5 at the 5′ end, with the transcription start site adenosine, followed by residues composed of T, C and G (there is no adenosine except at the TSS position). The template was amplified by PCR reaction to generate the dsDNA template for *in vitro* transcription (Table S1).

Sequence of DNA template used is indicated below. Only the top-strand (sense strand) is shown.
CAGTAATACGACTCACTATT**A**GGCCTCTCGCTCTGCTGGGTGTGCGCTTGCTTGGCTTGC.
The ϕ2.5 promoter is underlined. The transcription-start site adenosine is in bold.

To generate 5′ ^{32}P-cap-labeled NAD-RNA, rATP was replaced with ^{32}P-NAD$^+$ (Perkin Elmer, BLU023X250UC) in the reaction that was carried out at 37°C for 4 hours in 20 μL reaction containing 2 pmol DNA template, 2 μl 10x reaction buffer, 7.5 mM rNTP, 25 μCi ^{32}P-NAD$^+$ and 2 μl T7 enzyme mix (MEGAshortscript Kit, ThermoFisher Scientific). Unincorporated labeled NAD$^+$ was removed by Microspin G-25 Columns (GE Healthcare) and radioactive NAD-RNA was separated by resolving in a 15% urea gel. The gel band containing NAD-RNA was excised and the RNA eluted overnight in elution buffer (300 mM NaCl) at 25°C, with shaking at 750 rpm. The following day, labeled NAD-RNA was precipitated with ethanol and glycogen at −20°C for 30 min. After that, radioactive NAD-RNA was collected by centrifugation at 4°C for 15 min, and dissolved in water.

Preparation of RNA with labeled m^{7}G cap

^{32}P-labeled m^{7}G-capped RNA was generated by *in vitro* capping. We first prepared non-radioactive 5′ triphosphate RNAs by *in vitro* transcription using the same double-stranded DNA template described above for NAD-RNA synthesis.

In vitro transcription was carried out at 37°C for 4 hours in 20 μL reaction containing 2 pmol DNA template, 2 μl 10x reaction buffer, 7.5 mM rNTPs and 2 μl T7 enzyme mix (MEGAshortscript Kit, ThermoFisher Scientific). Then RNAs were extracted with phenol/chloroform and finally precipitated by ethanol. Purified 5′ triphosphorylated RNAs were directly used for capping reaction containing 10 pmol RNA, 2 μl 10x capping buffer (New England BioLabs), 1 μl Ribolock RNase inhibitor (Thermo Scientific), 60 μCi α -^{32}P-GTP (Perkin Elmer, BLU006H), 0.2 mM S-adenosylmethionine (New England BioLabs), 1 μl vaccinia virus capping enzyme (New England BioLabs) and 9.5 μl nuclease-free water. When G-capped RNAs without the N^7-methylation was required, the methyl donor S-adenosylmethionine was omitted. The reaction was carried out at 37°C for 4 hours, after which, capped RNAs were purified by passing through MicroSpin G-25 Columns (GE Healthcare) and further separated with 15% urea gel electrophoresis. Finally, radioactive m^{7}G-capped RNAs were extracted from the gel in 0.3 M NaCl, precipitated with ethanol, and dissolved in water, as described above.

In vitro decapping and deNADding assays

All decapping reactions contained the same amount of radioactively-labeled RNA. Briefly, ^{32}P-5′ end labeled NAD$^+$-capped, N^7-methylated or non-methylated G-capped RNAs were incubated with 200 nM to 600 nM recombinant proteins in 20 μl decapping reactions containing 100 mM KCl, 2 mM $MgCl_2$, 1 mM $MnCl_2$, 2 mM DTT, 10 mM Tris-HCl (pH 7.5) and incubated at 37°C for 30 min. Only for the NAD-RNA decapping reactions were, samples subsequently treated with 1 U of P1 nuclease (Sigma-Aldrich) for 30 min at 37°C (to digest the RNA into individual nucleotides to release the radioactive AMP that is left attached to the body of the RNA). To examine the influence of BLMH or BSA (as a control) on NUDT12 decapping activity, 5-20 nM of the proteins were added to the above reaction. Decapping products were resolved by PEI-cellulose TLC plates (MERCK) and developed in 0.45 M $(NH4)_2SO4$ in a TLC chamber at room temperature. When the resolving buffer front reaches ~2 cm from the top of the TLC plate, the plates were removed, allowed to air-dry, then wrapped in SaranWrap. Reaction products were visualized by exposure to Storage Phosphor screens (GE Health) and scanned with a Typhoon FLA 9500 scanner (GE Health). When required, signals were quantified by ImageQuantTL.

In vitro cleavage assays with labeled NAD

To evaluate hydrolysis of ^{32}P-NAD (Perkin Elmer, BLU023X250UC), reactions containing same amount of radioactive NAD were incubated with 10 nM of NUDT12 in reactions containing 100 mM KCl, 2 mM $MgCl_2$, 1 mM $MnCl_2$, 2 mM DTT, 10 mM Tris-HCl (pH 7.5) and incubated at 37°C for 30 min. For evaluating specificity of NUDT12, the above reactions were supplemented with 5-20 μM of either non-radioactive m^{7}GpppA (New England Biolabs) or non-radioactive NAD (New England Biolabs) cap analogs. For examining influence of BLMH or BSA (as a control) on NUDT12 activity, 5-20 nM of the proteins were added to the above reaction. Hydrolysis products were resolved by PEI-cellulose TLC plates (MERCK) and developed in 0.45 M $(NH4)_2SO4$ in a TLC chamber at room temperature. When the resolving buffer front reaches ~2 cm from the top of the TLC plate, the plates were removed, allowed to air-dry, then wrapped in SaranWrap. Reaction products were visualized by exposure to Storage Phosphor screens (GE Health) and scanned with a Typhoon FLA 9500 scanner (GE Health). Signals were quantified by ImageQuantTL. Total amount of ^{32}P-NAD (set as 100%) in each reaction is the sum of signals from uncleaved NAD (NppA) and hydrolysed product AMP. Activity of NUDT12 (amount of hydrolysed AMP) is represented as a percentage of this total (Figure 1D).

Metabolomics

Mass spectrometry analyses of metabolites were carried out with tissues (liver, kidney and testis) extracted from duplicate biological replicates of wild-type, duplicate heterozygous *Nudt12*$^{+/-}$ and quadruplicate homozygous *Nudt12*$^{-/-}$ knockout mice. The method followed is a protocol modified from a previous study (Paglia et al., 2014).

Metabolite extraction from tissue

Approximately 25 mg of fresh mouse tissue was placed in a homogenizer. Using 500 μL of methanol-water 4:1, the tissue was homogenized using a pistil. Transfer the homogenate to 2ml secure lock Eppendorf vial. Add 250 μL of fresh methanol-water 4:1 to the homogenator, move pistil a few times, transfer wash to the same 2ml secure lock Eppendorf vial. Repeat wash with additional 250 μL of fresh methanol-water 4:1, transfer wash to the same 2 mL secure lock Eppendorf vial, resulting in 1 mL tissue homogenate. Place

Eppendorf vial with tissue homogenate on dry ice until analysis. Homogenize other tissues in different homogenizers. After finishing all the homogenization, centrifuge all homogenates at 4°C at 13000 rpm for 15 min. Transfer 900 μL of clear supernatants to fresh 1.5 mL Eppendorf tubes. Samples were shipped (on dry ice) for analysis at the Functional Genomics Center Zurich (ETH Zurich) for LC-MS analysis.

Sample preparation for LC-MS analysis

50 μL methanol extract was dried under a nitrogen stream and reconstituted in 20 μL water (MS grade) and diluted with 80 μL injection buffer. The dilution was vortexed and centrifuged (16,000 x g, 4°C, 15 min). 50 μL of the supernatant was transferred to a glass vial with narrowed bottom (Total Recovery Vials, Waters) for LC-MS injection. In addition, method blanks, QC standards, and pooled samples were prepared in the same way to serve as quality control for the measurements. Injection buffer was composed of 90 parts of acetonitrile, 9 parts of methanol and 1 part of 5 M ammonium acetate.

LC-MS analysis

Metabolites were separated on a nanoAcquity UPLC (Waters) equipped with a BEH Amide capillary column (150 μm x130mm, 1.7 μm particle size, Waters), applying a gradient of 0.5 mM ammonium acetate in water (A) and 0.5 mM ammonium acetate in acetonitrile (B) from 10% A to 50% A over 10 min. The injection volume was 1 μL. The flow rate was adjusted over the gradient from 3 to 2 μl/min. The UPLC was coupled to Synapt G2Si mass spectrometer (Waters) by a nanoESI source. MS1 (molecular ion) and MS2 (fragment) data was acquired using negative polarization and MS^E over a mass range of 50 to 1200 m/z at MS1 and MS2 resolution of > 20'000.

Targeted Data analysis

Target metabolites (NAD^+ and NADH) were quantified by the area under the peak (AUP) of the MS1 extracted ion chromatogram (EIC) of the respective $[M-H]^-$ ion (m/z of 662.101 for NAD^+; m/z of 664.117 for NADH) extracted from raw data by using MassLynx v4.2 software (Waters). For all targets reference samples were used to locate and verify the correct peaks on the EIC of the samples by retention time and MS2 fragment information. AUP values reported (Figure S6C) are unit-less and are proportional to the concentration of the targets in the samples (relative quantification). AUP values can be used to calculate fold changes.

Untargeted Metabolomics Data analysis

Metabolomics datasets were evaluated in an untargeted fashion with Progenesis QI software (Nonlinear Dynamics), which aligns the ion intensity maps based on a reference dataset, followed by a peak picking on an aggregated ion intensity map. Detected ions were identified based on accurate mass, detected adduct patterns and isotope patterns by comparing with entries in the Human Metabolom Data Base (HMDB). A mass accuracy tolerance of 5 mDa was set for the searches. Fragmentation patterns were considered for the identifications of metabolites. All biological samples were analyzed at least in triplicate and quality controls were run on pooled samples and reference compound mixtures to determine technical accuracy and stability.

Circadian experiments

Circadian oscillations of gene expression use a timing system that is synchronized daily by light signals (*Zeitgebers*). Daily synchronization of the central pacemaker residing in the brain (SCN, suprachiasmatic nucleus) takes place by light received via the retina. On the other hand, circadian oscillators in the peripheral organs like liver, kidney etc receive signals from the feeding/fasting rhythms, which in turn is controlled by the SCN via rest/activity cycles. To study circadian oscillations in gene expression we placed mice (*Nudt12*$^{-/-}$ KO and wild-type littermates) in chambers where their exposure to light can be electronically regulated. *Zeitgeber* Time 0 corresponds to light ON (at 7 am in the morn), while ZT12 corresponds to light OFF (at 7 pm in the evening) which lasts until ZT24.

The locomotor activity (central clock oscillation) of mice were recorded as wheel-running activity (W.R.A.) and as spontaneous activity (S.A., through an infrared detection system of mice spontaneous movements) (Saini et al., 2013). Representative double-plot actograms obtained for two animals of each genotype. In each actogram the first seven days were recorded under 12 hr light/12 hr dark conditions (S.A. & W.R.A.), after which the light was turned off (W.R.A. only) and recording was continued in constant darkness for seven following days. The time spans during which the lights were completely switched off (continuous darkness) are marked by gray shading. The animals (W.R.A.) were then subjected to a 12 hr light-inversion regimen for 10 days (12hr dark period instead of the previous 12hr light period at the same given time-point, and similarly 12h light period instead of the previous 12hr dark period). No significant difference in locomotor activity was noted between the control and *Nudt12*$^{-/-}$ KO mice. Mice were then sacrificed at four hours intervals (ZT0, ZT4, ZT8, ZT12, ZT16, ZT20): one *Nudt12*$^{-/-}$ KO being sacrificed with one WT littermate at the same time-point. From ZT12 to ZT 24 (dark period), animals were sacrificed under red light. We repeated this experiment to get at least triplicate biological replicates for each time-point for each genotype. Livers were extracted from the sacrificed animals immediately and snap-frozen in liquid nitrogen, until RNA isolation.

For RNA extraction (Schmidt and Schibler, 1995), take approximately 0.5 g tissue, put it into a 50 mL conical tube, add 5 mL extraction buffer (composed of 250 g guanidium thiocyanate; 17.6 mL sodium citrate, 0.75 M, pH 7.0; 320 mL water, dissolve in 60°C. Before use, add 1/10 volume of sodium acetate, 2 M, pH 4.0; 1/100 volume β-mercaptoethanol), homogenize for 20 s. Then add 5 mL phenol-H_2O, mix well and stand on ice. Add 2 mL $CHCl_3$: isoamyl alcohol (49:1), mix well, stand on ice for 15 min, and transfer all the solution to a 15 mL falcon tube. Spin down for 20 min at 4000 rpm at 4°C, transfer the upper phase (approximately 5 ml) to a new 15 mL falcon tube, and 1 mL phenol: chloroform: isoamyl alcohol (24:24:2), mix well. Spin down for 15 min at 4000 rpm at 4°C, transfer upper phase (approximately 4 ml) to a new 15 mL falcon tube, add 4 mL isopropanol, and leave it at −20°C for 1 hour. After this, spin down at 4500 rpm for 30 min at 4°C, remove all the solution, and completely dry the tube. After, add 6 mL 4M lithium chloride to

suspend the precipitation, leave on ice for 5 min, and spin down at 4500 rpm for 15 min at 4°C, remove all the solution and dry the tube completely. Add 7 mL 75% ethanol to suspend the precipitation, leave at room temperature for 10 min, spin down at 4500 rpm for 15 min, remove all the solution and dry the tube completely. Repeat the 75% ethanol step once again. Dissolve the RNA precipitation with TE buffer to a proper concentration.

Quantitative RT-PCR analysis of circadian gene expression

Total RNA was extracted from mouse livers of WT and *Nudt12*$^{-/-}$ KO animals (triplicate biological replicates) sacrificed at time-points around the 24h period (ZT0, ZT4, ZT8, ZT12, ZT16, ZT20). Total RNA was isolated using the protocol described above.

Take 2.6 μg mouse liver RNA as template, do reverse transcription (random hexamer-primed) by using the SuperScript III Reverse Transcriptase kit (Thermo Fisher) to get the volume of 20 μl, dilute cDNA into 50 μl before use. Realtime-PCR solution was made with LightCycler 480 Sybr Green I Master (Roche) and run in Roche LightCycler 480 II machine. Normalization was done with values obtained with β-actin and GAPDH. Sequences of primers used are given in Table S1.

Preparation of RNA libraries

Total RNA from biological triplicates were used for the initial experiment with mouse liver and kidney samples. In this experiment, the time of animal sacrifice was approximately in the afternoon. RNA concentration was measured with a Qubit fluorimeter (Life Technologies) and RNA integrity assessed with a Bioanalyzer (Agilent Technologies). The TruSeq Stranded Total RNA kit with Ribo-Zero Gold was used for library preparation with 500 ng of total RNA as input. Library molarity and quality was assessed with the Qubit and Tapestation using a DNA High sensitivity chip (Agilent Technologies). Libraries were diluted at 2 nM and pooled before the clustering process on a HiSeq 4000 Single Read flow cell. Reads of 50 bases were generated using the TruSeq SBS reagents on the Illumina HiSeq 4000 sequencer (iGE3 Genomics Platform, University of Geneva).

For animals from the circadian experiment, we used total RNA from mouse liver of only one biological replicate [one animal each for wild-type (WT) and *Nudt12*$^{-/-}$ KO at ZT0, ZT4, ZT8 (*Nudt12*$^{+/-}$ HET was used instead of WT], ZT12, ZT16, ZT20; total of 6 WT and 6 KO animals) for preparation of RNaseq libraries, which was done as described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA-seq analysis

Reads were sorted into individual libraries based on the barcodes and mapped to the mouse genome (Ensembl release 95) using salmon v0.12.0 (salmon quant with options -l A -p 8 -o–gcBias–validateMappings) (Patro et al., 2017). Further analysis was performed using R version 3.6.0 (R Core Team, 2017) and Bioconductor (Huber et al., 2015). The DESeq function of DESeq2 1.24.0 bioconductor package (Love et al., 2014) was used to obtain log2 fold changes of gene expression between wild-type and mutant samples and the adjusted p values. Adjusted p value 0.1 was used as a threshold for statistical significance. The liver and kidney tissues were analyzed separately. The Volcano plots were plotted using EnhancedVolcano function from EnhancedVolcano 1.3.1 package (https://github.com/kevinblighe/EnhancedVolcano). The significantly differentially expressed genes with at least two times fold changes were highlighted in red (Figure 5C). The sets of genes upregulated and dowregulated in *Nudt12*$^{-/-}$ were compared between the tissues using Venn diagrams (VennDiagram 1.6.20) (Chen and Boutros, 2011) (Figure 5D). The group of genes found to be significantly up- or down-regulated in the mutant liver and spleen were searched for enriched Gene Ontology terms in the Biological Process ontology using ENRICHR (Chen et al., 2013; Kuleshov et al., 2016) and the enriched categories were shown (Figures 5E and S6D). Interestingly, six out of 22 genes associated with GO terms photoperiodism (GO:0009648) and entrainment of circadian clock by photoperiod (GO:0043153) were found to be upregulated in the liver of *Nudt12*$^{-/-}$. Their expression levels were plotted (Figure 5F). Expression of these genes during day-night cycle was analyzed by RNaseq and qRT-PCR. One replicate of *Nudt12*$^{-/-}$ and *Nudt12*$^{+/+}$ was analyzed by RNaseq (Figure 5G). The reads were mapped by salmon and the read counts were summarized to gene levels and normalized to library sizes. The expression of individual circadian clock genes was plotted as reads per million (rpm) (Figure 5G). Expression of selected genes was verified by qRT-PCR using three biological replicas. The expression was normalized to housekeeping genes *Actb* or *Gapdh* and to the ZT0 expression of the *Nudt12*$^{+/+}$ (Figure 5G).

DATA AND CODE AVAILABILITY

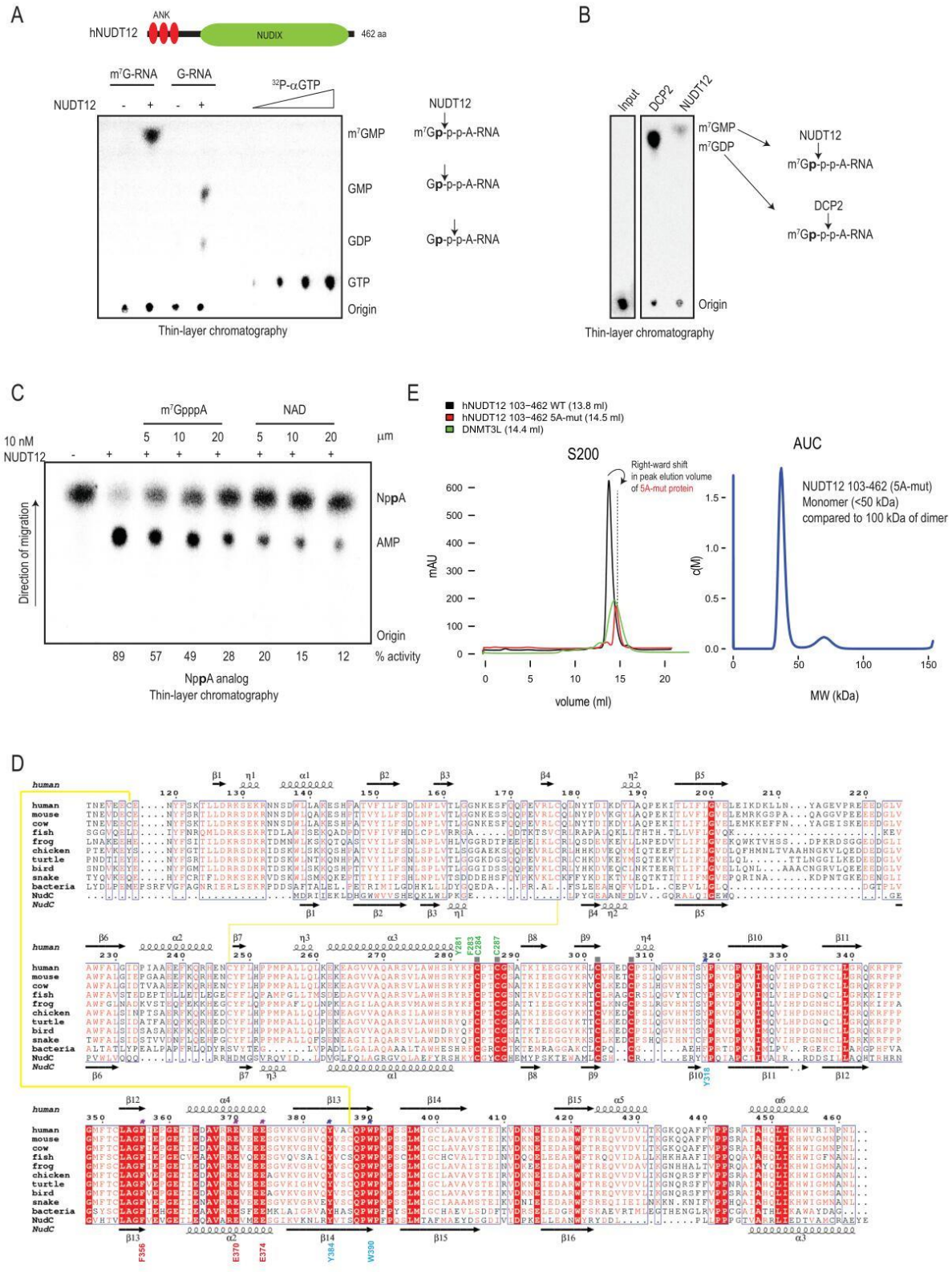
Deep sequencing data generated in this study are deposited with Gene Expression Omnibus (GEO) under the accession number GSE138019. Crystallographic data are deposited with Protein Data Bank under PDB accessions: 6SCX. Original/source data for figures in this paper are deposited with Mendeley Data under the accession https://doi.org/10.17632/ykzx9h339r.1.

Cell Reports, Volume *29*

Supplemental Information

Decapping Enzyme NUDT12 Partners with BLMH for Cytoplasmic Surveillance of NAD-Capped RNAs

Hao Wu, Lingyun Li, Kuan-Ming Chen, David Homolka, Pascal Gos, Fabienne Fleury-Olela, Andrew A. McCarthy, and Ramesh S. Pillai



Mutations made to human NUDT12

Mutations that abolish activity on both m7G-capped RNA and NAD: F356, E370, E374

Mutations that abolish activity only on m7G-capped RNA: Y318, Y384, W390

Mutations that render the protein monomeric (5A-mut): Y281, F283, C284, C287 and Y384

Figure-S1

Figure S1. NUDT12 hydrolase has a preference for NAD. Related to Figure 1.
(A) Decapping assays with m^7G-capped or G-capped RNA (unmethylated) and full-length NUDT12. Reactions were resolved by thin-layer chromatography (TLC). The ^{32}P-α-GTP radionucleotide is used as a positional marker. The scheme on the right shows the ^{32}P-labelled phosphate (in bold) in the RNA substrate, and the cleavages introduced by NUDT12 on the different substrates. See also Figure 1. (B) Comparison of decapping activity of human DCP2 and NUDT12 with m^7G-capped RNA. Note the different products generated. (C) Hydrolase activity of NUDT12 on labelled NAD (Np**p**A) tested in the presence of indicated competing cold cap analogs. This is a replicate of the experiments done for data presented in Figure 1C-D. (D) Sequence alignment of NUDT12 orthologues from prokaryotes to human. The numbering on the top refers to amino acids from the human NUDT12 protein. The secondary structure features as identified in our human NUDT12 catalytic core structure (PDB: 6SCX) is shown on the top, while that of *E. coli* NudC (PDB: 5IW5) is shown below. Different amino acids mutated in this study are indicated in different colours. Two cysteines involved in disulfide bridge formation are connected with a yellow line. There is a second potential disulfide bridge between Cys-117 and Cys-387 in the human structure. (E) Elution profiles of 40.9 kDa NUDT12 (103-462 aa) proteins in a gel-filtration column (S200). A mutant version carrying five point mutations (5A-mut) in the dimerization interface elutes later than the wildtype protein. Its elution profile resembles that of 48.0 kDa DNMT3L, while the dimeric wildtype NUDT12 version elutes earlier. We consider this 5A-mut as monomeric. Analytical Ultra-centrifugation (AUC) estimates the molecular weight of the 5A-mut protein as 41 kDa, confirming it to be monomeric. See Figure 1E for AUC data on wildtype homodimeric human NUDT12 protein.

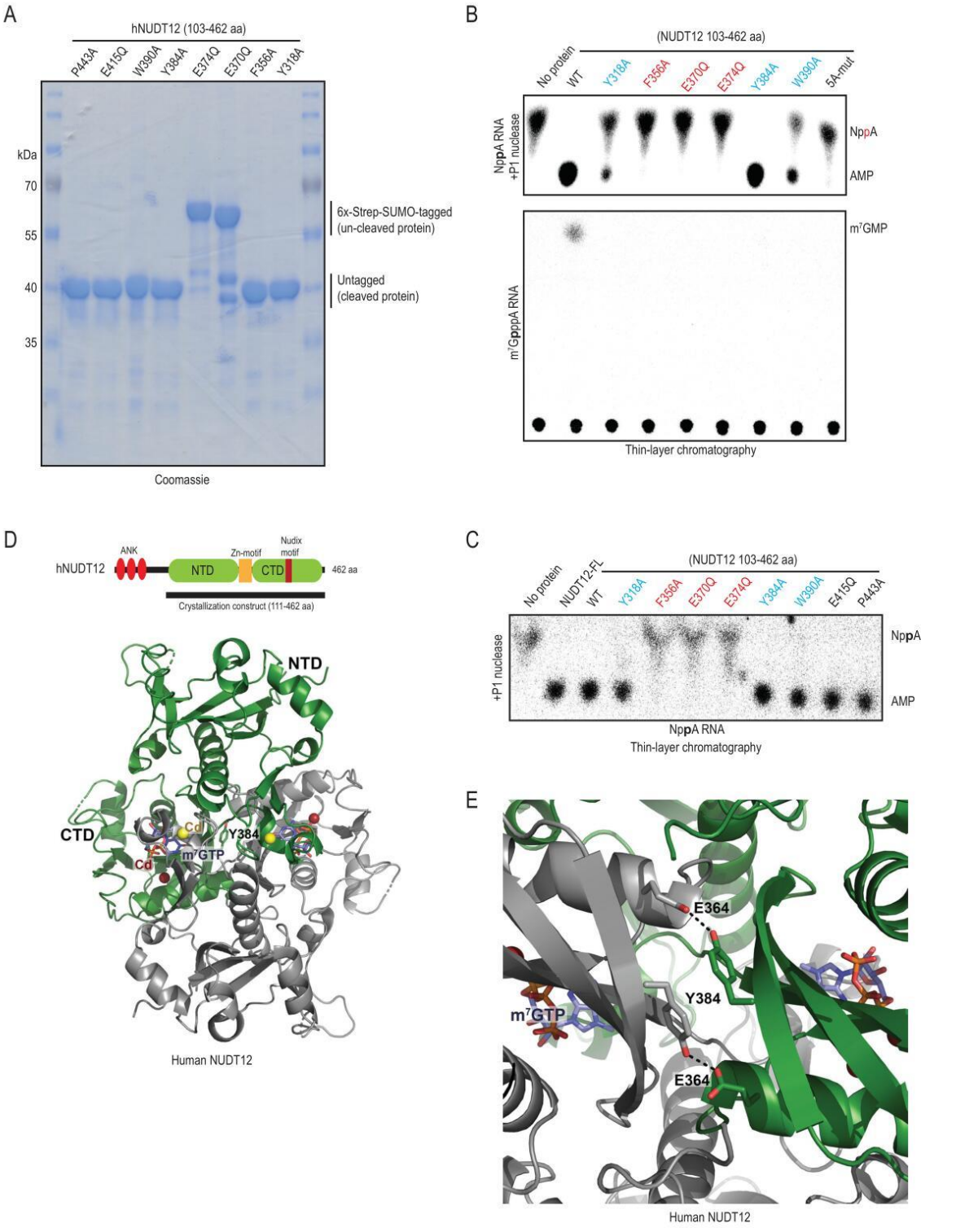
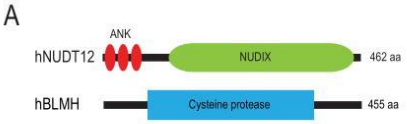

A
hNUDT12 (103-462 aa)
P443A
E415Q
W390A
Y384A
E374Q
E370Q
F356A
Y318A
kDa
70
55
40
35
6x-Strep-SUMO-tagged
(un-cleaved protein)
Untagged
(cleaved protein)
Coomassie
B
(NUDT12 103-462 aa)
No protein
WT
Y318A
F356A
E370Q
E374Q
Y384A
W390A
5A-mut
NppA RNA
+P1 nuclease
NppA
AMP
m7GMP
m7GpppA RNA
Thin-layer chromatography
D
hNUDT12
ANK
NTD
Zn-motif
CTD
Nudix
motif
462 aa
Crystallization construct (111-462 aa)
NTD
CTD
Cd
Y384
m7GTP
Human NUDT12
C
(NUDT12 103-462 aa)
No protein
NUDT12-FL
WT
Y318A
F356A
E370Q
E374Q
Y384A
W390A
E415Q
P443A
+P1 nuclease
NppA
AMP
NppA RNA
Thin-layer chromatography
E
E364
Y384
m7GTP
E364
Human NUDT12

Figure-S2

Figure S2. Mutations in the homodimerization interface of human NUDT12 modulates its decapping activity. Related to Figures 2.

(A) Purified recombinant mutant versions of the catalytic core (103-462 aa) of human NUDT12 used for cleavage assays. Note that tags on some proteins were not cleavable, thus used in their tagged form. (B) Decapping assay with labelled NAD-RNA (Np**p**A-RNA) or m^7G-RNA (m7G-**p**pp-RNA) and wildtype NUDT12 core protein (103-462 aa) or its mutated versions. See also Figure 2G. (C) Decapping assay with NAD-RNA (Np**p**A-RNA) and wildtype NUDT12 core protein (103-462 aa) or its mutated versions. Mutations (red) abolishing activity on both m^7G-capped and NAD-RNAs or mutations (blue) that render the enzyme active only for NAD-RNAs are indicated. See also Figure 2G. (D) A top-view showing the location of the Y384 in the interface of the human NUDT12 homodimer and close to the two catalytic pockets. Cadmium (Cd) ions in the catalytic pocket are shown in red, while those occupying the Zn-binding motif are shown in yellow. (E) Hydrogen-bonding mediated by Y384 with a glutamic acid residue from the other monomer, which is part of the helix carrying the catalytic residues E370A and E374A.

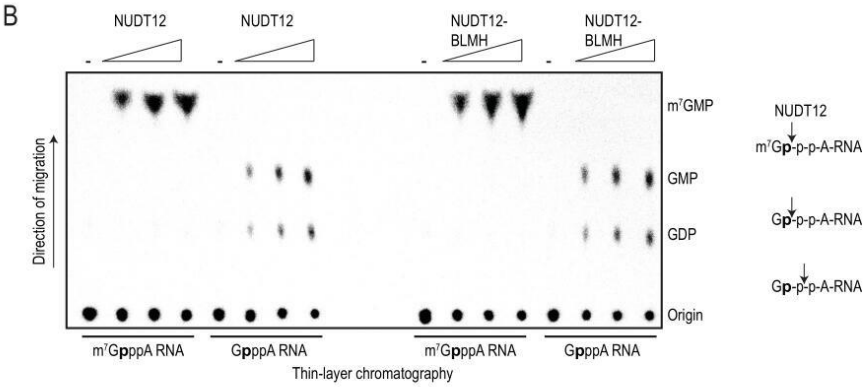
A



Name	MW	NUDT12-1	NUDT12-2	GFP-1	GFP-2
NUDT12	52 kDa	23	19	0	0
BLMH	53 kDa	8	3	0	0
KPYM	58 kDa	4	19	0	0
HS71A	70 kDa	3	6	0	0
TBB5	50 kDa	3	2	0	0
DCD	11 kDa	2	3	0	0
EF2	95 kDa	2	12	0	0
HBA	15 kDa	2	2	0	0
PGAM1	29 kDa	2	2	0	0

Name	MW	BLMH-1	BLMH-2	BLMH-3	GFP-1	GFP-2
BLMH	53 kDa	30	30	32	0	0
NUDT12	52 kDa	7	8	8	0	0
JUP	82 kDa	6	2	6	0	0
HSP90B1	92 kDa	5	3	3	0	0
ATP5F1A	60 kDa	4	4	5	0	0
CCT3	61 kDa	4	2	3	0	0
EEF1G	50 kDa	4	4	4	0	0
WDR77	37 kDa	3	4	3	0	0
RPS3	27 kDa	3	3	3	0	0
RPS25	14 kDa	2	3	2	0	0
SERPINB12	46 kDa	2	2	2	0	0

B



C

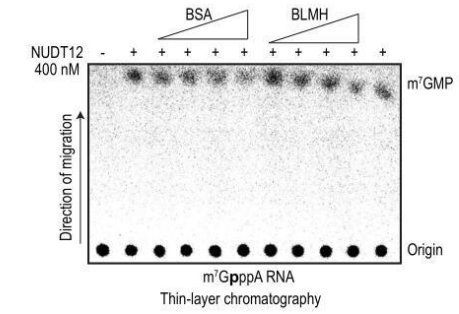


Figure-S3

Figure S3. BLMH does not alter decapping activity of NUDT12 in vitro. Related to Figure 3.
(A) Mass spectrometry analysis of tagged-NUDT12 and BLMH complexes isolated from transfected HeLa cell cultures. (B) Decapping assays with m^7G- and G-capped RNAs using NUDT12 or the NUDT12-BLMH complex. The complex does not alter activity of NUDT12. (C) Titration of BLMH or a control protein BSA into a reaction containing a constant amount of NUDT12, followed by addition of the ^{32}P-cap-labelled m^7G-RNA substrate. There is no alteration of NUDT12 decapping activity. See also Figure 3G-H.

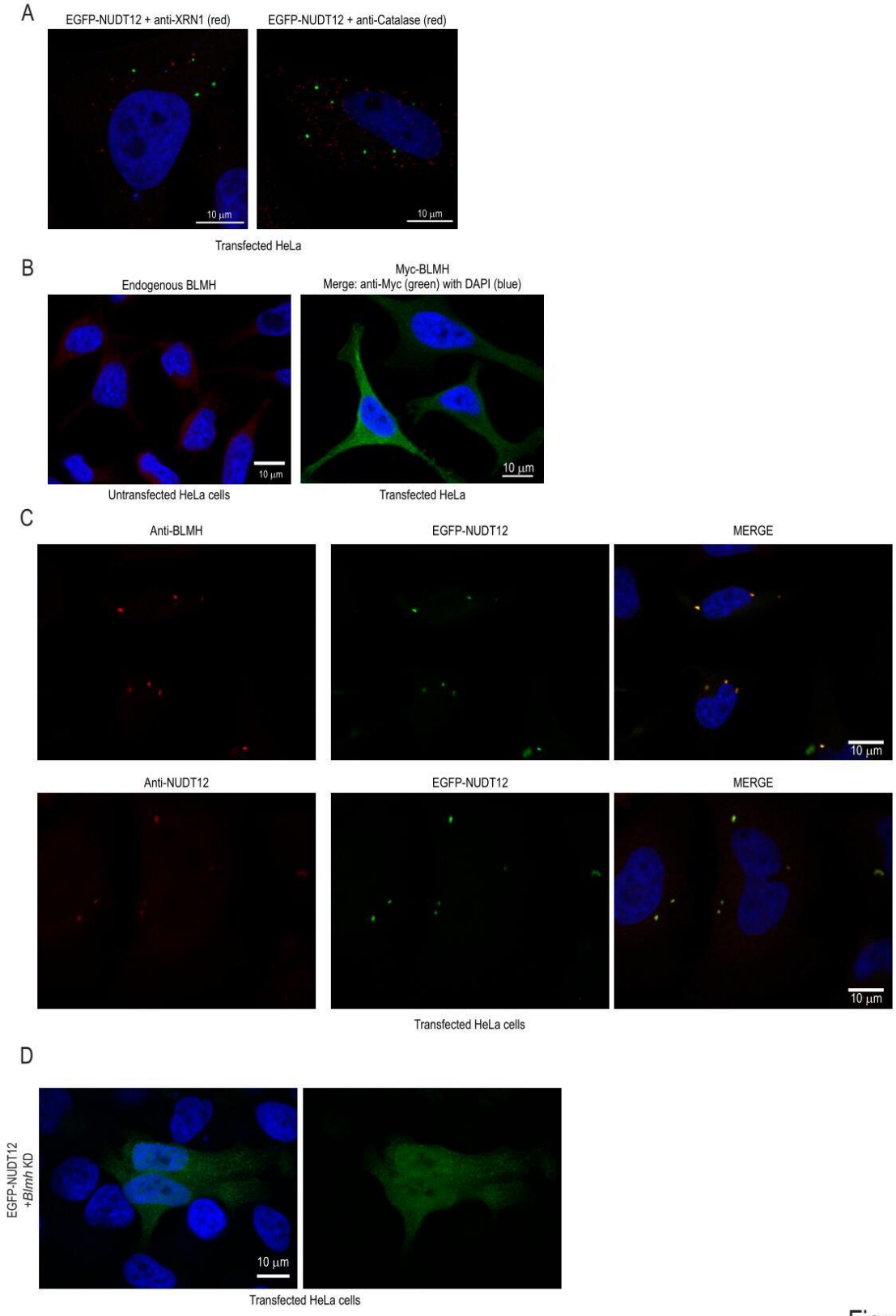

A
EGFP-NUDT12 + anti-XRN1 (red)
EGFP-NUDT12 + anti-Catalase (red)
10 μm
10 μm
Transfected HeLa
B
Endogenous BLMH
Myc-BLMH
Merge: anti-Myc (green) with DAPI (blue)
10 μm
10 μm
Untransfected HeLa cells
Transfected HeLa
C
Anti-BLMH
EGFP-NUDT12
MERGE
10 μm
Anti-NUDT12
EGFP-NUDT12
MERGE
10 μm
Transfected HeLa cells
D
EGFP-NUDT12
+Blmh KD
10 μm
Transfected HeLa cells

Figure-S4

Figure S4. Localization of NUDT12 to cytoplasmic granules depends on BLMH. Related to Figure 4.
(A) Immunostaining of endogenous XRN1 in transfected HeLa cells expressing EGFP-NUDT12. XRN1, a marker for P-bodies is present in dotty structures in the cytoplasm, and these do not overlap with that occupied by NUDT12. Immunostaining of peroxisomes (as marked by anti-catalase antibodies) reveals no overlap with granules occupied by EGFP-NUDT12. (B) Endogenous BLMH is diffusely distributed in the cytoplasm of HeLa cells. A similar distribution is seen for Myc-BLMH. (C) Endogenous BLMH is detected in EGFP-NUDT12 granules. (D) Knockdown of *Blmh* mRNA in HeLa cells with siRNAs, result in disruption of EGFP-NUDT12 granules, with the protein being diffused in the cell. See Figure 4G.

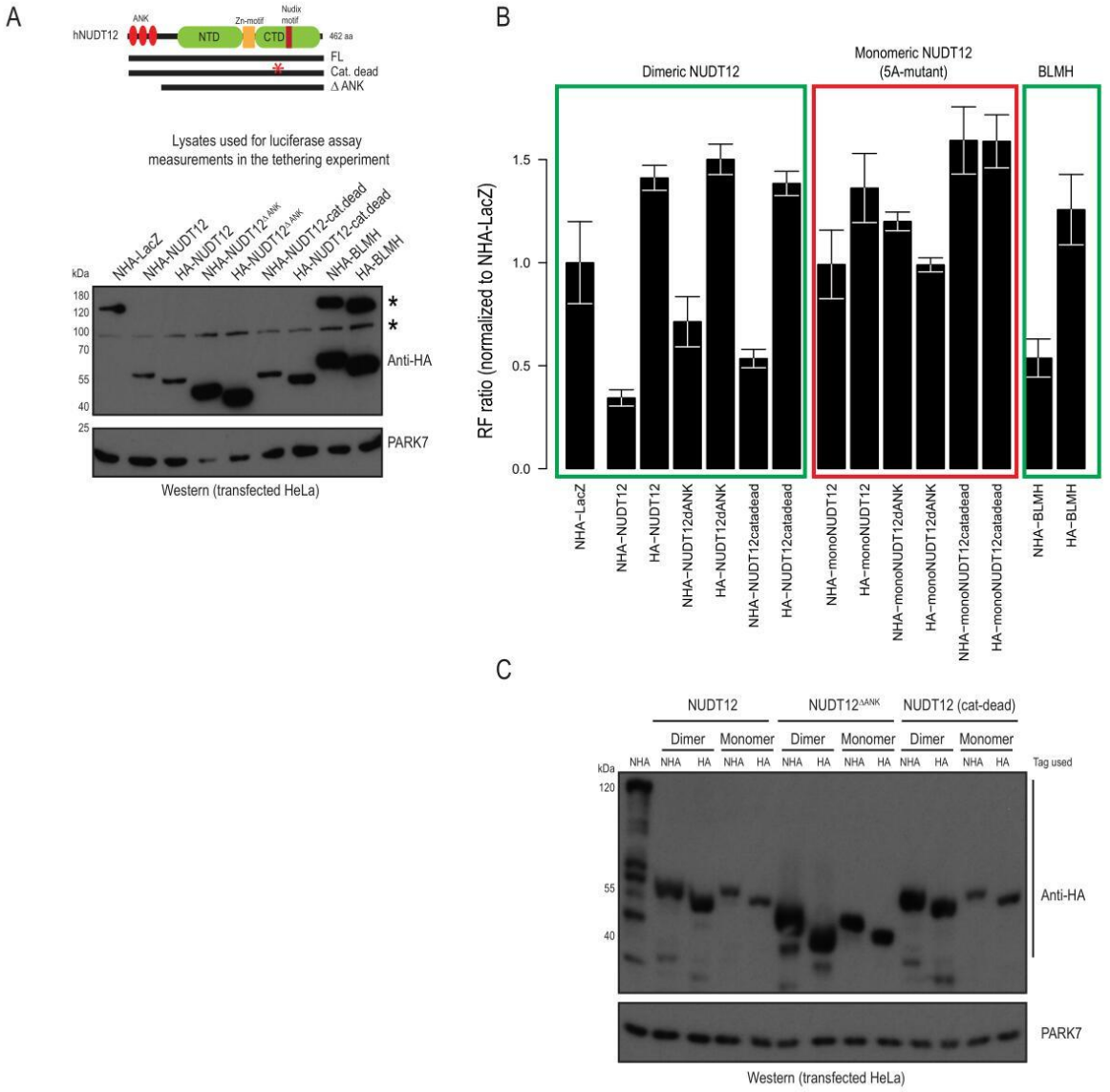
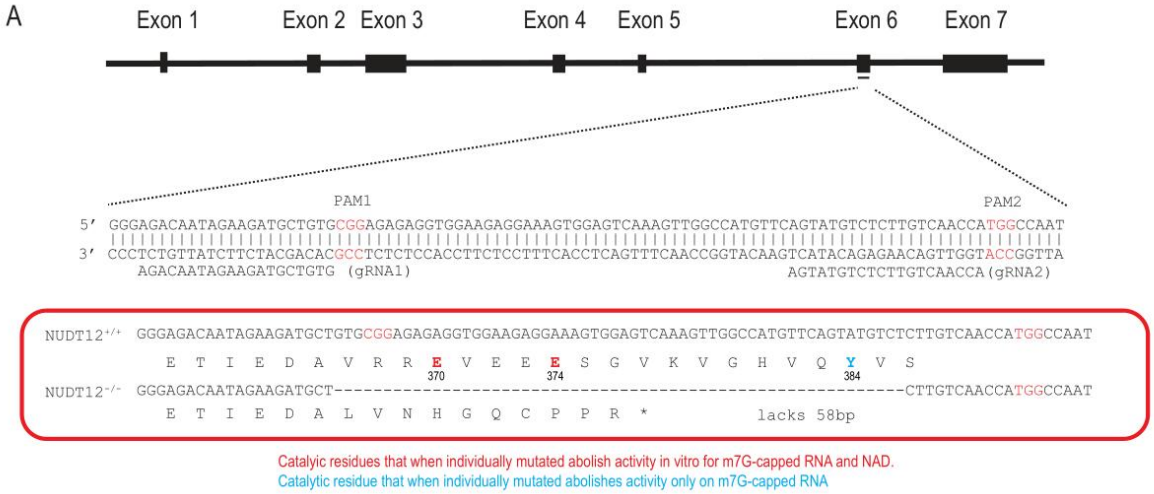
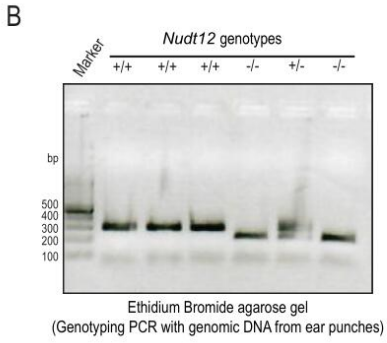

A
hNUDT12
ANK
Zn-motif
Nudix motif
NTD
CTD
462 aa
FL
Cat. dead
Δ ANK
Lysates used for luciferase assay measurements in the tethering experiment
NHA-LacZ
NHA-NUDT12
HA-NUDT12
NHA-NUDT12ΔANK
HA-NUDT12ΔANK
NHA-NUDT12-cat.dead
HA-NUDT12-cat.dead
NHA-BLMH
HA-BLMH
kDa
180
120
100
70
55
40
25
Anti-HA
PARK7
Western (transfected HeLa)
B
Dimeric NUDT12
Monomeric NUDT12 (5A-mutant)
BLMH
RF ratio (normalized to NHA-LacZ)
0.0
0.5
1.0
1.5
NHA−LacZ
NHA−NUDT12
HA−NUDT12
NHA−NUDT12dANK
HA−NUDT12dANK
NHA−NUDT12catadead
HA−NUDT12catadead
NHA−monoNUDT12
HA−monoNUDT12
NHA−monoNUDT12dANK
HA−monoNUDT12dANK
NHA−monoNUDT12catadead
HA−monoNUDT12catadead
NHA−BLMH
HA−BLMH
C
NUDT12
NUDT12ΔANK
NUDT12 (cat-dead)
Dimer
Monomer
NHA
HA
Tag used
kDa
120
55
40
Anti-HA
PARK7
Western (transfected HeLa)

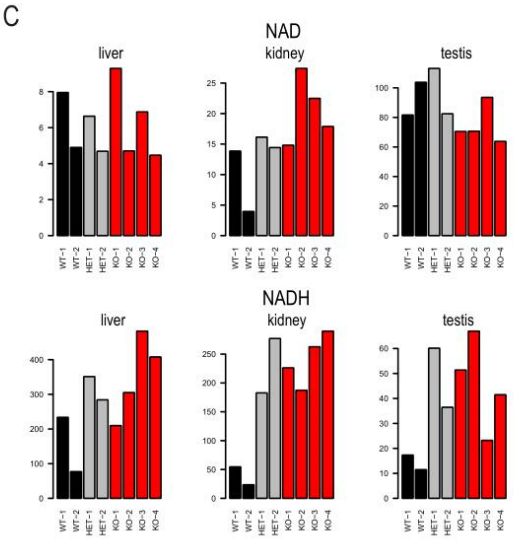
Figure-S5

Figure S5. Atificial tethering of NUDT12 and BLMH to a reporter RNA downregulates gene expression. Related to Figure 4.

(A) Western analysis showing expression of tagged proteins co-expressed with the *Renilla* BoxB reporter in transfected HeLa cells. These are the lysates used for measurements reported in Figure 4H. PARK7 is used loading control. (B) Artificial tethering of monomeric (5A-mutant; see STAR Methods) NUDT12 to the *Renilla* BoxB reporter does not downregulate gene expression, confirming the fact that homodimerization is essential for catalytic activity. (C) Western blot to show expression of the various tagged proteins in lysates from experiment presented in panel B of this figure. PARK7 is used loading control.


A
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6
Exon 7
PAM1
PAM2
5′ GGGAGACAATAGAAGATGCTGTGCGGAGAGAGGTGGAAGAGGAAAGTGGAGTCAAAGTTGGCCATGTTCAGTATGTCTCTTGTCAACCATGGCCAAT
3′ CCCTCTGTTATCTTCTACGACACGCCTCTCTCCACCTTCTCCTTTCACCTCAGTTTCAACCGGTACAAGTCATACAGAGAACAGTTGGTACCGGTTA
AGACAATAGAAGATGCTGTG (gRNA1)
AGTATGTCTCTTGTCAACCA (gRNA2)
NUDT12+/+ GGGAGACAATAGAAGATGCTGTGCGGAGAGAGGTGGAAGAGGAAAGTGGAGTCAAAGTTGGCCATGTTCAGTATGTCTCTTGTCAACCATGGCCAAT
E T I E D A V R R E V E E E S G V K V G H V Q Y V S
370
374
384
NUDT12-/- GGGAGACAATAGAAGATGCT----------CTTGTCAACCATGGCCAAT
E T I E D A L V N H G Q C P P R * lacks 58bp
Catalyic residues that when individually mutated abolish activity in vitro for m7G-capped RNA and NAD.
Catalytic residue that when individually mutated abolishes activity only on m7G-capped RNA


B
Nudt12 genotypes
Marker
+/+
+/+
+/+
-/-
+/-
-/-
bp
500
400
300
200
100
Ethidium Bromide agarose gel
(Genotyping PCR with genomic DNA from ear punches)


C
NAD
liver
kidney
testis
NADH
liver
kidney
testis
WT-1
WT-2
HET-1
HET-2
KO-1
KO-2
KO-3
KO-4

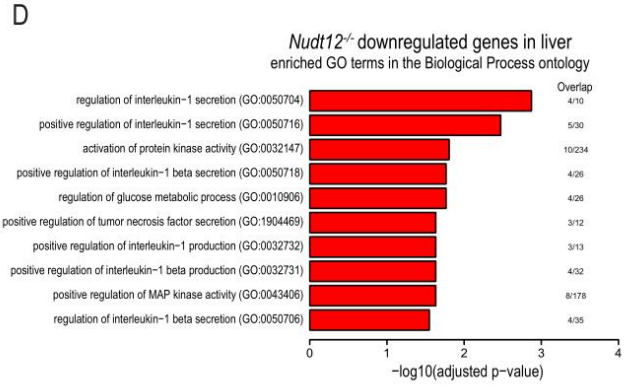

D
Nudt12-/- downregulated genes in liver
enriched GO terms in the Biological Process ontology
Overlap
regulation of interleukin-1 secretion (GO:0050704) 4/10
positive regulation of interleukin-1 secretion (GO:0050716) 5/30
activation of protein kinase activity (GO:0032147) 10/234
positive regulation of interleukin-1 beta secretion (GO:0050718) 4/26
regulation of glucose metabolic process (GO:0010906) 4/26
positive regulation of tumor necrosis factor secretion (GO:1904469) 3/12
positive regulation of interleukin-1 production (GO:0032732) 3/13
positive regulation of interleukin-1 beta production (GO:0032731) 4/32
positive regulation of MAP kinase activity (GO:0043406) 8/178
regulation of interleukin-1 beta secretion (GO:0050706) 4/35
0
1
2
3
4
−log10(adjusted p−value)

Figure-S6

Figure S6. Creation of the *Nudt12* knockout mouse. Related to Figure 5.
(A) Strategy for gene targeting of the *Nudt12* locus in mouse single-cell embryos. Two guide RNAs direct endonuclease activity of Cas9 within the exon 6, deleting a region encoding the critical catalytic residues of the Nudix motif. This generates a premature stop codon, likely leading to nonsense-mediated decay of the mRNA. Western analysis shows complete lack of the protein in the homozygous knockout mutant. See Figure 5B. Key residues that are catalytically important and tested in vitro are marked. (B) Genotyping PCR of genomic DNA from ear-punches. The PCR products are separated in an agarose gel and stained with ethidium bromide. (C) Metabolomic analysis of tissue lysates from liver, kidney and testis of wildtype (WT), heterozygous (HET) and homozygous (KO) *Nudt12* animals. The values for NAD^+ and NADH for individual biological replicates are represented in the histogram. A modest increase in amount of NADH in tissue lysates from the knockout animals is noted. See Figure 5. (D) Significantly enriched GO terms in the Biological Process ontology are shown for the genes found to be downregulated in the liver of knockout *Nudt12*$^{-/-}$ animals.

Table S1. DNA oligonucleotides used in this study.
Related to STAR Methods and Figure 1-5.

Name	Sequence	Comments
WH-Oligo36	GAAATTAATACGACTCACTATAGGAGACAATAGAAGATGCTGTGGTTTTAGAGCTAGAAATAGC	CRISPR F primer-1(gene specific sequence in red)
WH-Oligo37	GAAATTAATACGACTCACTATAGGAGTATGTCTCTTGTCAACCAGTTTTAGAGCTAGAAATAGC	CRISPR F primer-2(gene specific sequence in red)
WH-Oligo60	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC	Common reverse primer for preparing template for guide RNA
WH-Oligo104	ATTGAATGGGGGCTCTTTC	Forward primer for NUDT12 KO mouse genotyping
WH-Oligo106	CTATTTCATTCTTGTCAACTTTAATTTCTGTA	Reverse primer for NUDT12 KO mouse genotyping
WH-Oligo-344	AATGCCGACCAGCGTGTCAT	Mouse Per1 RT-PCR forward primer
WH-Oligo-345	GGCTGCTGTTTCTGCATGGC	Mouse Per1 RT-PCR reverse primer
WH-Oligo-346	TGCTGCTGGCAGAGAGGGTA	Mouse Per2 RT-PCR forward primer
WH-Oligo-347	GGGACCGCCCTTTCATCCAC	Mouse Per2 RT-PCR reverse primer
WH-Oligo-348	AGCAGCCAGAGAGACACGGA	Mouse Per3 RT-PCR forward primer
WH-Oligo-349	TGCACTGGCTGATGCCAGAC	Mouse Per3 RT-PCR reverse primer
WH-Oligo-364	GTTGTCTCCTGCGACTTCA	Mouse GAPDH RT-PCR forward primer
WH-Oligo-365	GGTGGTCCAGGGTTTCTTA	Mouse GAPDH RT-PCR reverse primer
WH-Oligo-366	CTGTCCCTGTATGCCTCTG	Mouse Beta-actin RT-PCR forward primer
WH-Oligo-367	TGTCACGCACGATTTCC	Mouse Beta-actin RT-PCR reverse primer
LY-Oligo1	CAGTAATACGACTCACTATTAGGCCTCTCGCTCTGCTGGGTGTGCGCTTGCTTGGCTTGC	ssDNA template (modified phi 2.5-CA2) to make dsDNA template for in vitro transcription of RNA used for decapping assay

Supplemental Table S2. List of all deep-sequencing libraries created in this study.

experiment	sample	reads
RNAseq of liver Nudt12 WT: ZT0	WHR138	24709554
RNAseq of liver Nudt12 WT: ZT4	WHR139	20616702
RNAseq of liver Nudt12 WT: ZT8	WHR140	16575483
RNAseq of liver Nudt12 WT: ZT12	WHR141	30855826
RNAseq of liver Nudt12 WT: ZT16	WHR142	27778343
RNAseq of liver Nudt12 WT: ZT20	WHR143	28329457
RNAseq of liver Nudt12 KO: ZT0	WHR144	27085202
RNAseq of liver Nudt12 KO: ZT4	WHR145	20855363
RNAseq of liver Nudt12 KO: ZT8	WHR146	19551353
RNAseq of liver Nudt12 KO: ZT12	WHR147	20549638
RNAseq of liver Nudt12 KO: ZT16	WHR148	20313356
RNAseq of liver Nudt12 KO: ZT20	WHR149	22523142
RNAseq of liver Nudt12 WT	WHR54	31532545
RNAseq of liver Nudt12 WT	WHR55	30589525
RNAseq of liver Nudt12 WT	WHR56	30440609
RNAseq of liver Nudt12 KO	WHR57	29397498
RNAseq of liver Nudt12 KO	WHR58	30723446
RNAseq of liver Nudt12 KO	WHR59	29883954
RNAseq of kidney Nudt12 WT	WHR60	33074431
RNAseq of kidney Nudt12 WT	WHR61	29696795
RNAseq of kidney Nudt12 WT	WHR62	15583722
RNAseq of kidney Nudt12 KO	WHR63	31631183
RNAseq of kidney Nudt12 KO	WHR64	37047861
RNAseq of kidney Nudt12 KO	WHR65	31734097

Discussion

In the first chapter, we described a Nudix hydrolase, human NUDT12 that cleaves both canonical m^7G cap and noncanonical NAD^+ cap of mRNAs. NUDT12 preferentially acts on NAD^+ cap/NAD-RNA in vitro, which can be partially explained by its structure that in the catalytic pocket, except three essential residues (F356, E370 and E374) required for both decapping and deNADding activity, mutations of remaining residues only impair decapping activity while have very mild or no effect on deNADding activity. We failed to crystalize NUDT12 in complex with NAD^+ cap, hence, we could not provide the details about how NUDT12 accommodates NAD^+ cap within its binding pocket. However, since the catalytic domain is highly conserved between bacterial deNADding enzyme NudC (Hofer et al., 2016; Zhang et al., 2016b) and human NUDT12, we could still compare each other and speculate how NUDT12 recognizes different caps and how the preference towards NAD^+ cap is rendered.

NUDT12 directly interacts with BLMH, forming a dodecamer in vitro. BLMH is a cytoplasmic cysteine peptidase that forms a hexamer in vitro (O'Farrell et al., 1999). The only known function of this protein is that it can hydrolyse the anticancer drug, BLM to protect cell from of toxicity of chemotherapy. Mice lacking *Blmh* have low-penetrant neonatal death. Such mice are fertile but display tail dermatitis (Schwartz et al., 1999). Beyond this, the normal physiological function of this protease is unclear. BLMH has no significant influence on enzymatic activity of NUDT12, according to our in vitro decapping assay, however, we found that BLMH is required to sequester NUDT12 into cytosolic discrete granules in vivo. There granules are not P-bodies, nor stress granules. Inhibition of mRNA biogenesis does not interfere with the formation of such granules (data not shown here), suggesting RNAs are not co-localized with NUDT12-BLMH complex. Furthermore, when BLMH was artificially tethered to a luciferase reporter, it led to downregulation of the expression of this reporter, resembling that effect of NUDT12. Notably, human BLMH has no nucleic acid binding ability (O'Farrell et al., 1999), even though it is evolutionarily conserved and its homologues in yeast and rat bind either single-stranded DNA or RNA (Takeda et al., 1996; Xu and Johnston, 1994; Zheng and Johnston, 1998). More efforts should be made to better understand what the biological consequence of NUDT12-BLMH complex is.

We made *Nudt12* KO mice and such mutant mice have no aberrant phenotype, suggesting Nudt12 might be redundant, to some extent. As mentioned above, human nuclear DXO is a

deNADding enzyme as well. Function of Nudt12 in mammalians could be compensated by Dxo in nucleus and/or by other unknown enzymes in cytoplasm. Here we point out that loss of NudC in *E. coli* causes no dramatic change in regards to the level of small regulatory RNAs which are known to be NAD^+ capped (Cahova et al., 2015), again implicating the existence of a parallel mechanism beside Nudt12.

Although Nudt12 is not essential for mouse viability and fertility, we notice that in the absence of Nudt12, a subset of circadian gene is upregulated in mouse liver. We don't know if this is due to increased stability of these circadian transcripts which are NAD-capped after removal of Nudt12. We tried to identify what kinds of RNA carry 5′ NAD^+ cap in mouse liver or other tissues by using NAD captureSeq, but we failed. Nevertheless, it is still important to make it clear what are the NAD-capped RNAs and how abundant they are in mammalian, especially human cells, considering recently a few papers describing alternative approaches to quantify NAD-RNAs just came out (Hu et al., 2021; Yu et al., 2021).

Since the discovery of NAD-RNA in bacteria in 2015 and later in yeast and mammals, one frequently asked question is why cells produce NAD-RNAs. In other words, what are the biological functions of NAD-RNA? So far, what we know is that NAD-RNAs are fragile, of which the half-life is very short compared with m^7G-capped RNAs. Is that meaning NAD-RNAs are toxic such that once they are present inside cells, they must be eliminated rapidly? One hypothesis about the generation of NAD-RNA is that when cells are under stressful conditions, the cellular NAD^+ level will be elevated, which means that the chance of generating NAD-RNA is increased. Therefore, NAD-RNAs can be a signal when cells are exposed to stress. The biogenesis of NAD-RNA is thought to be facilitated by RNA polymerase II in nucleus, even if no clear evidence can validate this up to now. Whereas in cytoplasm, how the NAD-RNAs are generated is still a mystery. One possibility is that nuclear NAD-RNAs can be exported to cytoplasm. But as is known that decapping enzyme Dxo governs NAD-RNA decay in nucleus (Jiao et al., 2017), why those NAD-RNAs that exported to cytoplasm can escape such surveillance. On the other hand, RNA export relies upon the communication between nuclear export factors and CBC which binds m^7G cap (Cheng et al., 2006; Nojima et al., 2007). No data is shown that CBC recognizes NAD^+ as well. Hence, NAD-RNA transport from nucleus to cytoplasm seems impossible. Then the other possibility of biogenesis of cytoplasmic NAD-RNAs is that these NAD-RNAs are produced

locally via a mechanism yet to be discovered. At least in yeast, certain NAD-RNAs are generated in mitochondria (Walters et al., 2017).

Human genome encodes 22 NUDIX hydrolases, of which only eight enzymes are active against m^7G cap in vitro (Song et al., 2013). Apart from DCP2 (NUDT20), NUDT16 and NUDT12 which is described in this thesis, our knowledge about other NUDIX proteins is quite limited. Recently, NUDT2 is another protein of interest in this filed because of the identification of Ap_4A-capped RNAs (Hudecek et al., 2020; Luciano and Belasco, 2020; Luciano et al., 2019). Human NUDT2 is the Ap_4A hydrolase that catalyzes the asymmetric hydrolysis of diadenosine 5',5'''-P1,P4-tetraphosphate (Ap_4A) to yield AMP and ATP (Ge et al., 2013; Thorne et al., 1995). It also acts on m^7G cap. Physiological function of NUDT2 is not well known but at least this protein seems to be essential for neurodevelopment because mutation of NUDT2 causes intellectual disability in human (Anazi et al., 2017; Diaz et al., 2020; Yavuz et al., 2018). Whether this phenotype is linked to Ap_4A-RNA is unclear. Depletion of NUDT2 results in dramatic increase of intracellular level of Ap_4A molecule which in turn activates the immune response (Shu et al., 2019). Again, the underlying mechanism needs to be illustrated.

The diverse 5′ cap modifications and their corresponding capping/decapping enzymes make the regulation of RNA 5′ ends complex. The regulatory network of classic m^7G cap is more or less established owing to numerous studies on it in the past decades. However, with the discoveries of more and more noncanonical caps, such as NAD^+, Ap_4A and FAD, our cognition about RNA 5′ ends needs to be expanded. One direction would be systematically and quantitatively measure RNA 5′ ends. This requires a much more precise approach to detect RNA 5′ modifications, considering their low abundance in the cells. Moreover, once a novel modification is discovered, efforts should be made to identify those related enzymes and more importantly, to provide insights into mechanism of how this modification functions in vivo. Since RNA 5′ modifications are also linked to innate immune response, inflammation and tumorigenesis, development of new diagnostic methods and potential therapies that target relevant enzymes and/or pathways, deserves more attention.

Chapter 2. YTHDC2 binds 3′ UTR of RNA targets to regulate meiosis

Introduction

Since the proposal of central dogma by Crick in 1957, hundreds of thounds papers have been published in this field, with a vast majority being focused on DNA and protein. Among those studies, one of the most popular research interest is to investigate modifications present in DNA and protein. Up to now, it is known that modifications of DNA/chromatin, including methylation, histone modification, comprise the basic concept of epigenetics that control gene expression. Modifications of protein, such as phosphorylation, acetylation and ubiquitination which are called post-translational modifications, make another layer of gene expression regulation. However, in the last decade, with more and more literatures related with quantification and functiona of RNA modifications coming out, a novel layer that controls gene expression in the RNA level starts to emerge.

RNA modification

It is believed that RNAs from all kingdoms of life can be decorated with chemical modifications in great diversity. The earliest discovery of RNA modification can date back to 1960, when pseudouridine was identified from highly abundant samples (Cohn, 1960). To date, more than 170 chemical modifications have been identified in various types of RNA (Boccaletto et al., 2022). Most modifications are present in transfer RNA (tRNA) and ribosomal RNA (rRNA), of which the enzymatic reactions and fucntions are relatively well characterized. However, in the past decade, owing to the improvement of high-resolution mapping approachs and next generation sequencing, many other modifications are found in mammalian messanger RNA (mRNA) and small nuclear RNA (snRNA), particularly for mRNA methylation. These modifications are dynamic, reversible and critical in diverse cellular processes involved in RNA metabolsim. Dysfunction of RNA modifications is linked to abnormal cell development and human diseases. RNA modifications or in other words, the epitranscriptomics, represent an emerging crucial way to regulate gene expression.

rRNA modification

Ribosomal RNA (rRNA) is the most abundant RNA in a cell and it is extensively modified during transcription and following maturation in the nucleolus, nucleus and cytosol. Recently, more than 130 individual chemical modifications in rRNA are identified based on the high resolution

structure of human ribosome (Natchiar et al., 2017). Among these modifications, 2′-O-methylation of the ribose and converstion of U to pseudouridine (Ψ) are the two major types, making up 95% of rRNA modification. To date, 55 2′-O-methylation sites and 45 Ψ sites have been identified in yeast rRNAs, while in human rRNAs 60 2′-O methylations and 25 Ψ pseudouridylations are experimentally validated (Birkedal et al., 2015; Natchiar et al., 2017; Taoka et al., 2016). As an essential subunit of ribosome, rRNA serves to maintain protein synthesis. Therefore, rRNA modifications that frequently occur in or close to functionally crucial sites of ribosome are thought to make contribution to the protein quality control during translation (Sloan et al., 2017).

In general, there are two mechanisms that introduce chemical modifications into rRNAs at different stages during ribosome biogenesis. One is guided by small nucleolar RNAs (snoRNAs) and the other is facilitated by stand-alone enzymes. Modifications guided by snoRNAs are 2′-O-methylation and pseudouridylation, which are catalyzed by two classes of snoRNPs, termed box C/D and box H/ACA snoRNPs respectively (Ganot et al., 1997; Kiss-Laszlo et al., 1996; Ni et al., 1997; Watkins and Bohnsack, 2012). Eukaryotic snoRNPs, which are functionally conserved, are ribonucleoprotein complexes that comprise a snoRNA and its associated catalytic protein. Such snoRNA binds its pre-rRNA targets with base-pairing to guide the catalytic component to modify specific site. Box C/D snoRNAs form a single hairpin structure through their conserved sequence motifs (C box, 5′- RUGAUGA-3′ and D box, 5′-CUGA-3′). Structural analysis of a box C/D snoRNA from *Sulfolobus solfataricus* (Figure 2.1) indicates that extensive pairing between snoRNA and pre-rRNA positions the catalytic site of methyltransferase fibrillarin to modify the ribose of target, 5 nucleotides upstream of the D/D′ box (Lin et al., 2011; Tollervey et al., 1993). Particularly, additional complementarity of snoRNA-pre-rRNA occurring close to the methylated site could stimulate methylation (van Nues et al., 2011). In contrast to box C/D snoRNAs, H/ACA box snoRNAs containing a conserved H box (5′-ANANNA-3′) and an ACA sequence form a double hairpin structure. Through base-pairing with snoRNA, pre-rRNA is accommodated in the pseudouridylation pocket localized in one hairpin, in which non-based-paired uridine is exposed to pseudouridine synthetase to complete isomerisation, as is shown by the structure of an H/ACA box sRNP from *Pyrococcus furiosus* (Figure 2.1) (Lafontaine et al., 1998). Notably, there are several exceptions that 2′-O-methylation and pseudouridylation are independent of snoRNAs. For example, in yeast pseudouridine synthetase Pus7 directly catalyzes pseudouridylation in a

consensus motif in 5S rRNA and 2′-O-methyltransferase Sbp1 can methylate G2922 of the 25S rRNA (Decatur and Schnare, 2008; Lapeyre and Purushothaman, 2004).

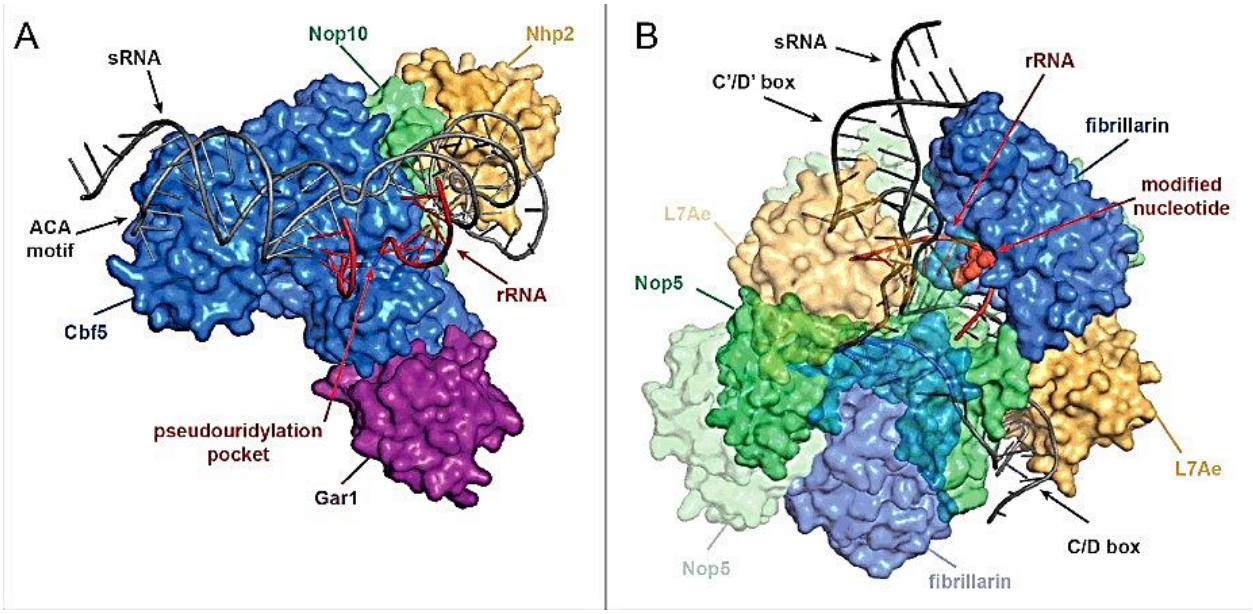


Fig 2.1. Structures of box C/D and box H/ACA snoRNP complexes. A, crystal structure of an H/ACA box sRNP from *Pyrococcus furiosus*. Cbf5 is the pseudouridine synthetase that deposits Ψ at rRNA target. Gar1, Nop10 and Nhp2 are the protein components that stablize the tertiary fold of the RNA and ensure correct positioning of the target nucleotide in the Cbf5 active site. B, structure of a box C/D sRNP from *Sulfolobus solfataricus*. Fibrillarin catalyzes 2′-O-methylation in rRNA with assistance from other protein subunits including L7Ae and Nop5 (Duan et al., 2009; Lin et al., 2011; Sloan et al., 2017).

Besides 2′-O-methylations and pseudouridylations, eukaryotic rRNAs are also modified at the base position. Those base modifications, such as N^7-methylguanosine (m^7G) and acetylated cytosine (ac^4C) present either in small or large ribosomal subunits, are installed by various enzymes including RNA methyltransferases, acetyltransferases and pseudouridine synthetases. rRNA methyltransferases, like the cap-adjacent methyltransferase, belong to the Rossman-like fold family and use S-adenosylmethionine (SAM) as a methyl group donor. As for rRNA acetylation, NAT10 (Kre33 in yeast) is demonstrated to carry out acetylation of cytosines of the 18S rRNA (Ito et al., 2014; Sharma et al., 2015). What's more, unlike other base-modifying enzymes which appear to exclusively act on rRNAs, NAT10/Kre33 acetylates $tRNA^{Ser}$ and $tRNA^{Leu}$ as well, but these processes require additional cofactor THUMPD1/Tan1 (Sharma et al., 2015).

rRNA modifications often appear in a cluster in ribosome where modified residues are residing in or nearby functionally important regions including decoding and tRNA binding sites (the A-, P- and E-sites), the peptidyltransferase center and the intersubunit interface (Sloan et al., 2017). Such

spatial and structural arrangement suggests that chemical modifications of rRNA play an important role in regulating ribosomal function to ensure efficient and accurate protein synthesis. On one hand, rRNA modifications stabilize the overall structure of ribosome. On the other hand, the heterogeneity of these modifications at certain sites or across species reflect the complex of ribosome functions during evolution.

tRNA modification

Transfer RNAs (tRNA) deliver amino acids to ribosome during mRNA translation. They are the second abundant RNA after rRNA but harbor the largest number and the greatest diversity of chemical modifcations in cells. So far, more than 100 different modifications have been found in both cytoplasmic and mitochondral tRNAs. The classic and most common tRNA consists of 76 nucleotides and is highly conserved across species (Rich and RajBhandary, 1976). In human cell, each tRNA is decorated with, on average, 11-13 different modifications which are installed during tRNA maturation (Schimmel, 2018). These modificatioins which are deposited by so many various enyzmes (Figure 2.2), range from simple methylation or isomerization, including N^1-methyladenosine (m^1A), 5-methylcytosine (m^5C), pseudouridine (Ψ) and inosine (I), to complex multistep chemical modifications, such as N^6-threonylcarbamoyladenosine (t^6A) and 5-methoxycarbonylmethyl-2-thiouridine ($mcm5s^2U$) (Delaunay and Frye, 2019). Function of each modification depends on both its location in tRNA and its chemical structure. For instance, anticodon loop in tRNA displays a great variety of modifications, with the wobble position (position 34) bearing the highest diversity (Figure 2.2). Modifications at wobble position expand and optimize codon usage, hence, they render tRNAs more efficiency of decoding message from mRNAs during gene-specific translation (Chan et al., 2012; Hanson and Coller, 2018; Schaffrath and Leidel, 2017).

Under certain conditions, tRNA modifications can also impact function through direct involvement in folding process. In *S. cerevisiae*, m^1A58 is present in 23 out of 34 tRNAs and is installed by the Trm6p/Trm61p methyltransferase complex (Anderson et al., 1998). Mutations of this complex results in a slow growth phenotype, attributable to the trunover of unmethylated $tRNA_i^{Met}$. m^1A58 is identified in human tRNA as well $tRNA_3^{Lys}$, where human homologs TRMT6 and TRMT61A catalyze the methylation. Overexpression of human TRMT6/TRMT61A complex in yeast restores the formation of m^1A58 in $tRNA_i^{Met}$, suggesting m^1A58 is evolutionarily conserved (Ozanick et al., 2005). Another case is m^7G46 in $tRNA^{Val(AAC)}$, which is methylated by

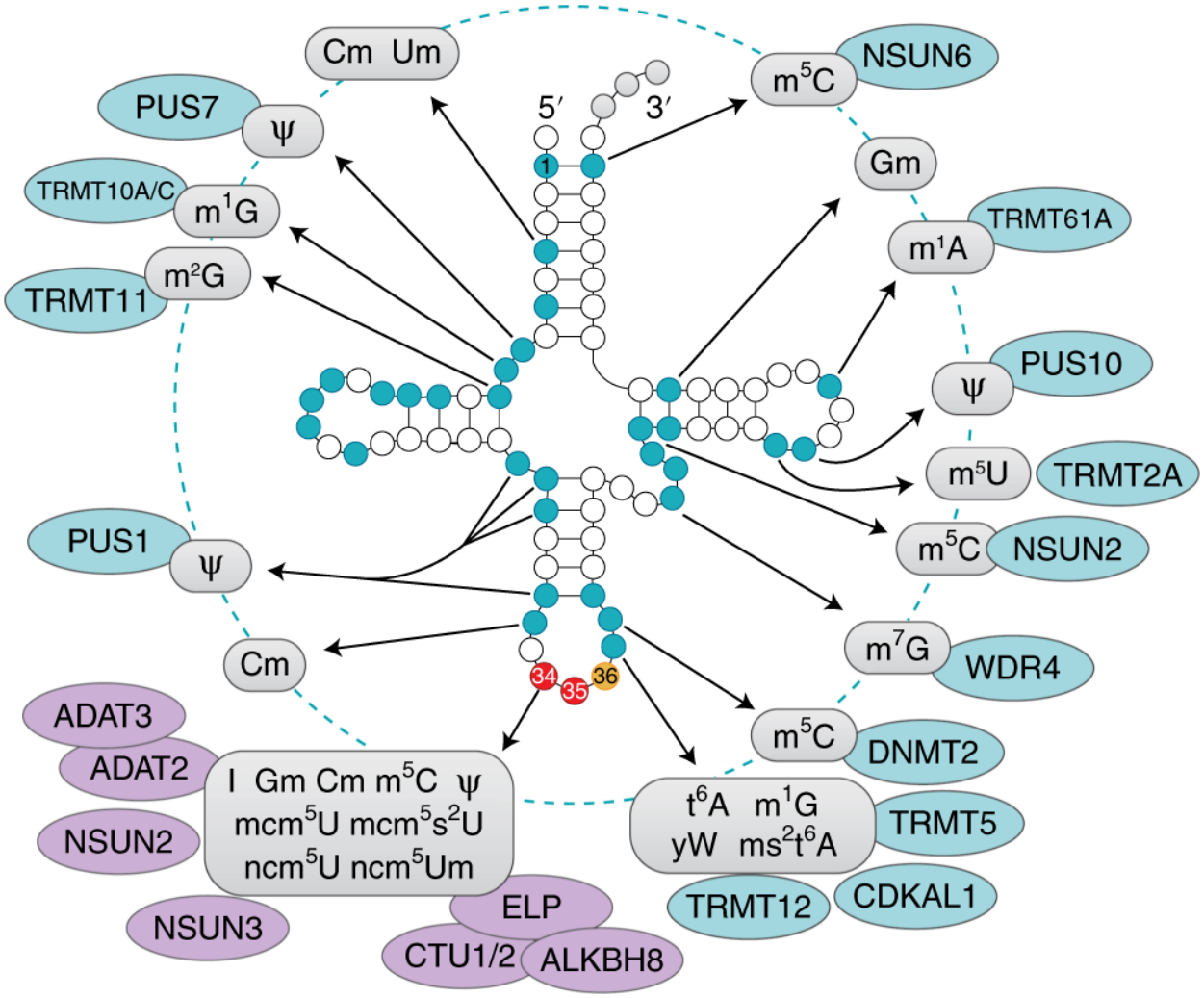


Fig 2.2. Schematic representation of tRNA chemical modifications. Modified positions and types along with their corresponding depositing enzymes are indicated. Enzymes involved in modifications at wobble positions are highlighted in purple. Positions 34–36 indicate anticodons (Delaunay and Frye, 2019).

the methyltransferase complex Trm8p/Trm82p in yeast. Double deletions of related genes (*trm8*-Δ *trm4*-Δ) cause loss of m^7G46 modification and dramatic decrease of cellular $tRNA^{Val(AAC)}$ accompanied by loss of aminoacylation. Such mutants have a temperature-sensitive growth defect (Alexandrov et al., 2006). Together, modifications outside of the anticodon region can play extremely important role in stablizing structure of endogeous tRNAs.

In addition to influence tRNA stability and tRNA-mRNA pairing, modifications can also affect the cleavage of tRNA. One of the instance is that *Drosophila* Dnmt2 can introduce m^5C38 in three types of tRNA, $tRNA^{Asp(GTC)}$, $tRNA^{Val(AAC)}$ and $tRNA^{Gly(GCC)}$. Its enzymatic activity is essential for *Drosophila* viability. Interestingly, Dnmt2-mediated methylation protects tRNAs against ribonuclease cleavage under stress conditions, indicating a novel function of tRNA modification

during the biogenesis of tRNA-derived small RNAs (Schaefer et al., 2010). Similarly, in human cells, NSUN2 is another m^5C methyltransferase that catalyzes the reaction at multiple sites in tRNAs. Loss of m^5C in tRNA increases its endonucleolytic cleavage, leading to an accumulation of 5′ tRNA-derived small RNA fragments. Accumulation of those fragments in the absence of NSUN2 is linked to reduced protein translation, activation of stress pathways and disordered neuro-development. Whereas NSUN2-mediated tRNA methylation protects from endonucleolytic cleavage into small RNA fragments (Blanco et al., 2014). Based on these studies, it is possible that tRNA modifications can have an impact on cell signaling processes where dysregulation of them may lead to disease and cancer.

In summary, in response to different microenvironments, tRNAs adopt at least two mechanisms that control protein synthesis via chemical modifications. First, modifications outside the anticodon loop ensure a stable structure of tRNA so that they modulate the global de novo protein synthesis. Second, modifications within the anticodon loop determine the translation of codon-specific genes (Delaunay and Frye, 2019).

mRNA modification

Recent excitement in the field of RNA modification was due to the quantification and mechanistic studies on reversible modifications in mRNA, of which mRNA methylation in particular, attracted the most interest. In addition to 5′ m^7G cap and 3′ polyadenylation tail, mRNA also bears various internal chemical decorations (Figure 2.3), including base isomerization (U to Ψ), methylation of bases (m^1A, m^5C and m^6A) as well as 2′ postion of riboses (N_m and m^6A_m), deamination (A to I) and oxidation of m^5C to 5-hydroxymethylcytosine (hm^5C) (Li et al., 2016b). Identification and quantification of these modifications have been benefited tremendously from recent advances in developing novel methods that couple analytic chemistry and high-throughput sequencing to allow us map each modification at or near single-base resolution (Moshitch-Moshkovitz et al., 2022). Together with extensive characterization of the modification "effectors", including enzymes involved in reversible chemical reactions that alter the modification level and protein factors which can recognize such marker, and their functional studies, our understanding towards mRNA modification is expanded dramatically. We know modifications in DNA and histone represent the epigenetic control of gene expression. Likewise, RNA modifications comprise the novel layer of gene expression, which is termed as epitranscriptomics.

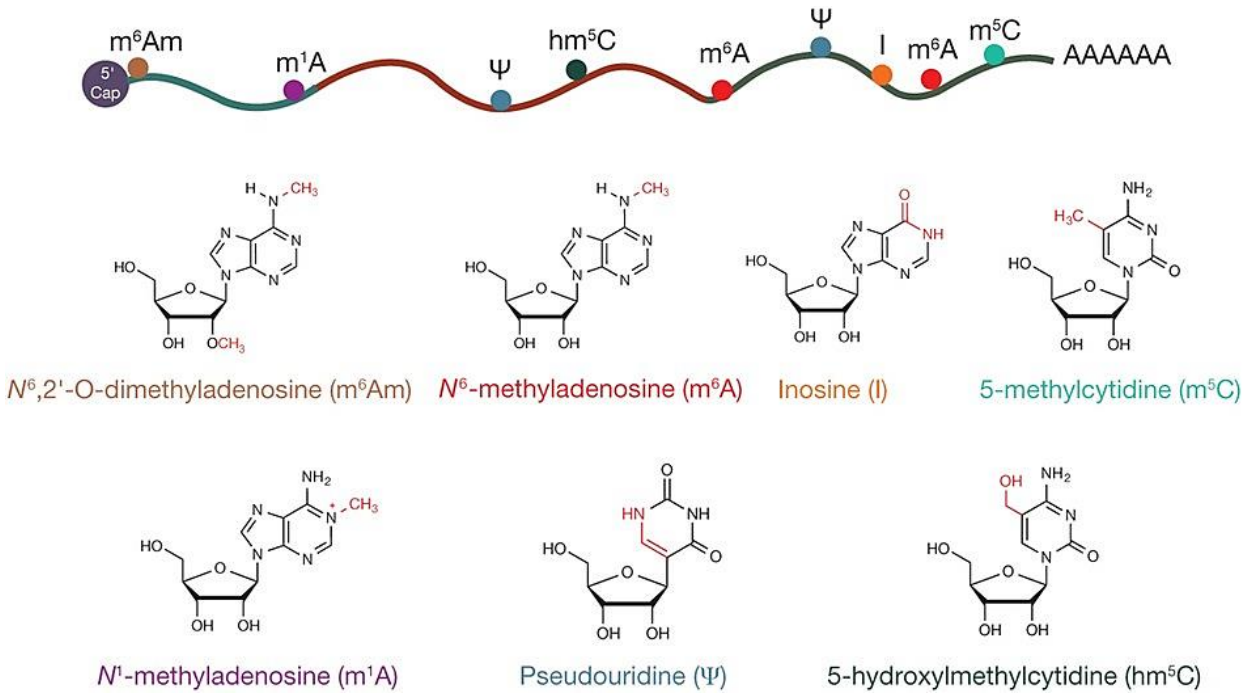


Fig 2.3. Chemical modifications in mRNA (Li et al., 2016b).

N^1-Methyladenosine (m^{1}A)

m^{1}A was identified in nucleic acids almost sixty years ago (Dunn, 1961). After decades, its distribution and function in tRNA and rRNA are well documented that m^{1}A is a prevalent modification in tRNA with a major impact on structure and function. This can be explained by its occurrence at the Watson-Crick base-pairing interface where this positively charged base can significantly influence protein-RNA interactions and RNA secondary structures via electrostatic effects (Agris, 1996; Roundtree et al., 2017a). However, m^{1}A in mRNA is less studied because of its low abundance in mRNA. This situation was changed in 2016 when two groups independently reported that m^{1}A is present in thousands of different transcripts in eukaryotic cells, at an estimated average transcript stoichiometry of 20% in humans (Dominissini et al., 2016; Li et al., 2016a). One year later, m^{1}A was also found in mitochondrial transcripts (Li et al., 2017b; Safra et al., 2017). By exploiting m^{1}A-seq or m^{1}A-ID-seq, m^{1}A is mapped uniquely to the 5′ untranslated regions (UTRs) with enrichment around start codon upstream of the first splice site (Dominissini et al., 2016; Li et al., 2016a). Enzymes that catalyze the deposition of m^{1}A in cytosolic mRNAs is still unclear, even though some m^{1}A sites are found in tRNA T-loop-like structures, where TRMT6/TRMT61A complex is thought to introduce this modification (Safra et al., 2017). Regarding mitochondrial m^{1}A, at least one defined m^{1}A site in ND5 mRNA is installed by methyltransferase TRMT10C. This methylation pattern is tissue-specific and tighty controled

during development (Safra et al., 2017). This modification can be removed by RNA demethylase ALKHB3 and this reversible process is quite dynamic in response to various types of cellular stress (Dominissini et al., 2016; Li et al., 2016a). Role of m^1A in regulaing mRNA translation is debatable. One group claimed that m^1A on mRNA is a repressive marker for translation bacasue it blocks base pairing (Safra et al., 2017). Whereas, this conclusion is conteracted by another finding that m^1A correlates with elevated translation in human Hela and mouse embryonic cells (Dominissini et al., 2016). A better approach of evaluating the global effect of m^1A on translation is crucial to understand its physiological function.

5-Methylcytosine (m^5C)

The earliest discovery of m^5C in protein coding RNA dates back to 1970s. At that time, it was found that m^5C is much less abundant than other types of mRNA modifications, such 2′-O-methylation and m^6A (Desrosiers et al., 1974; Dubin and Taylor, 1975). Taking advantage of bisulfite sequencing that was first applied to map m^5C in DNA (Schaefer et al., 2009), transcriptome-wide sequencing of m^5C in mRNA and lncRNA (long noncoding RNA) in human cells was achieved (Squires et al., 2012) and the distribution of this methylated marker in transcripts appears to favour untranslated regions, where several m^5C sites are localized to the binding sites for Argonaute proteins in particular (Squires et al., 2012). Consistent with previous studies, even though the abundance of m^5C varies among organisms (Amort et al., 2017; David et al., 2017; Huang et al., 2019), overall number of m^5C in mRNAs is rather low and it can even reach as low as single digits in mammals (Legrand et al., 2017). The tRNA methyltransferase NSUN2 was reported to methylate cytosine in two mRNAs: CINP and NAPRT1 (Squires et al., 2012) and recently, another protein NSUN6 was demonstrated to specifically methylate cytosines at the consensus motif CTCCA which is located in loops of hairpin in 3′ UTR (Selmi et al., 2021). However, it is not clear whether mRNA m^5C can be directly removed by other enzymes. Interestingly, m^5C is recognized by several proteins, including ALYREF and YBX1 (Chen et al., 2019; Yang et al., 2017; Zou et al., 2020). ALYREF is the mRNA export adaptor protein. It binds m^5C-containing mRNAs with its K171 residue to guide mRNA export (Yang et al., 2017). YBX1 is the Y box binding protein and its W65 residue in the conserved cold-shock domain is crucial to recognize m^5C modification. In bladder cancer cells, YBX1 promotes cancer pathogenesis and progression via targeting m^5C-containing oncogene-related transcripts (Chen et al., 2019).

In *Drosophila melanogaster*, m^5C in mRNA can be further oxidized by Tet-family enzymes to

form 5-hydroxymethylcytosine (hm^5C) (Fu et al., 2014). Using hMeRIP-seq (methylated RNA immunoprecipitation followed by sequencing), hm^5C is detected in over 1500 mRNAs with enrichment in the exonic and intronic regions (Delatte et al., 2016). The Drosophila Tet ortholog which carries out hydroxymethylation and hm^5C-containing RNAs are found to be abundant in brain. Mutation of Tet in fruit fly leads to impaired brain development, accompanied by decreased RNA hydroxymethylation, suggesting hm^5C might be important for Drosophila neurodevelopment (Delatte et al., 2016).

N^6-Methyladenosine (m^6A)

Among all the internal chemical modifications, m^6A is the most abundant one in mRNA. Although discovery of this modification is rather early (Desrosiers et al., 1974), systematic quantification of m^6A is only available after 2012 when two groups described an antibody-based m^6A mapping method, whereby they quantified m^6A in mouse and human cells and revealed the m^6A methylome (Dominissini et al., 2012; Meyer et al., 2012). It turns out that mRNA m^6A methylation is unexpectedly prevalent in mammalian transcriptomes and this modification is highly conserved from mouse to human, implicating m^6A may have very essential functions in regulating mRNA metabolism. These two papers undoubtably ignite the field of RNA modification, since then, numerous studies have been focused on m^6A and its biological functions and regulations are unravelled.

Statistical analyses of m^6A reveal that in human cells, approximately 0.15 – 0.16 percentage of adenosine in polyadenylated RNAs is converted to m^6A and m^6A to A ratios are significantly affected by highly expressed genes in the transcriptome (He and He, 2021; Liu et al., 2020; Liu et al., 2014; Wei et al., 2018). What's more, percentage of transcripts containing m^6A in the whole transcriptome reaches 25% in average and this number can go up to 60% if those relatively lowly expressed genes are taken into consideration (He and He, 2021). In regards to the average number of m^6A per transcript, ~ 3 m^6A residues are detected in one transcript based on LC-MS/MS and ~ 1.7 m^6A peaks per transcript are quantified according to m^6A-seq respectively (Dominissini et al., 2012; He and He, 2021; Perry et al., 1975). Notably, these numbers can be either overestimated or underestimated because many transcripts are unmethylated while many have more m^6A sites than the average.

When mapping m^6A position in the whole transcriptome, it is uncovered that m^6A is preferentially enriched in the vicinity of stop codons and in 3′ untranslated regions (3′ UTRs)

(Figure 2.4). Furthermore, m^6A commonly cocurs at a consensus motif DRACH, where D equals A, G or T; R equals A or G; H can be either A, C or U; A is the methylated site (Dominissini et al., 2012; Meyer et al., 2012). Multiple DRACH motifs can be next to each other but not all of them carry m^6A. In some cases, m^6A is identified outside DRACH sequence as well (Doxtader et al., 2018; Mendel et al., 2018; Pendleton et al., 2017). However, one unsolved question is that how this unique pattern of m^6A distribution is defined. In addition, m^6A in protein coding sequences appears to more frequently present in longer internal exons (> 200 nt) and this is not because these long exons occupy the majority of all internal exons (only 15% internal exons are > 200 nt) (Ke et al., 2017). Together, deposition of m^6A in mRNAs is specific and also context-dependent, which means this process is probably tightly regulated in mammalian cells.

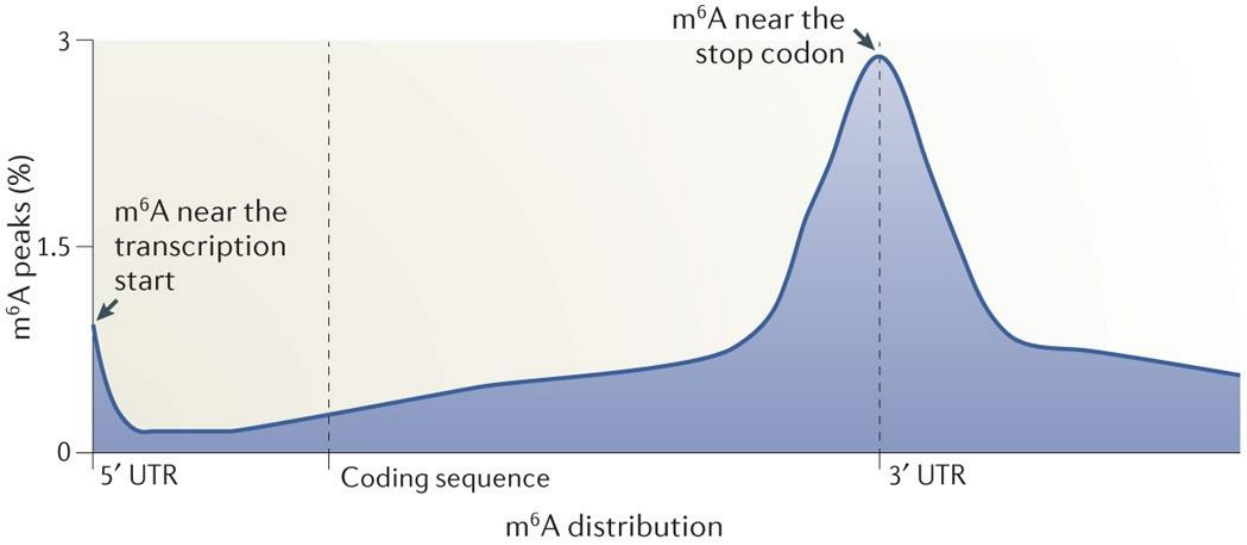


Fig 2.4. Metagene profile of m^6A distribution along the entire transcript. One m^6A peak at the transcription start site is attributed to N^6,2′-O-dimethyladenosine (m^6Am). m^6A is also found in coding regions and is particulary enriched near stop codon in 3′ UTR (Meyer and Jaffrey, 2014).

m^6A is involved in many cellular processes to regulate gene expression. m^6A alters transcription, pre-mRNA splicing, mRNA exprot and translation (Wang et al., 2014a; Wang et al., 2015; Xiao et al., 2016; Zhou et al., 2019). It also affects protein-RNA interactions by changing RNA structures (Liu et al., 2015). Given its diverse roles in the regulation of gene expression, m^6A is indeed a critical regulatory target in various physioligical processes and aberrant abundance or dysregulation of m^6A can cause severe consequences. Such multifaced nature of m^6A in gene expression regulation is relied upon its protein "effectors", including "writers" which install m^6A marker in transcripts; "erasers" which remove this marker and "readers" which specifically recognize this marker to modulate different processes (Figure 2.5).

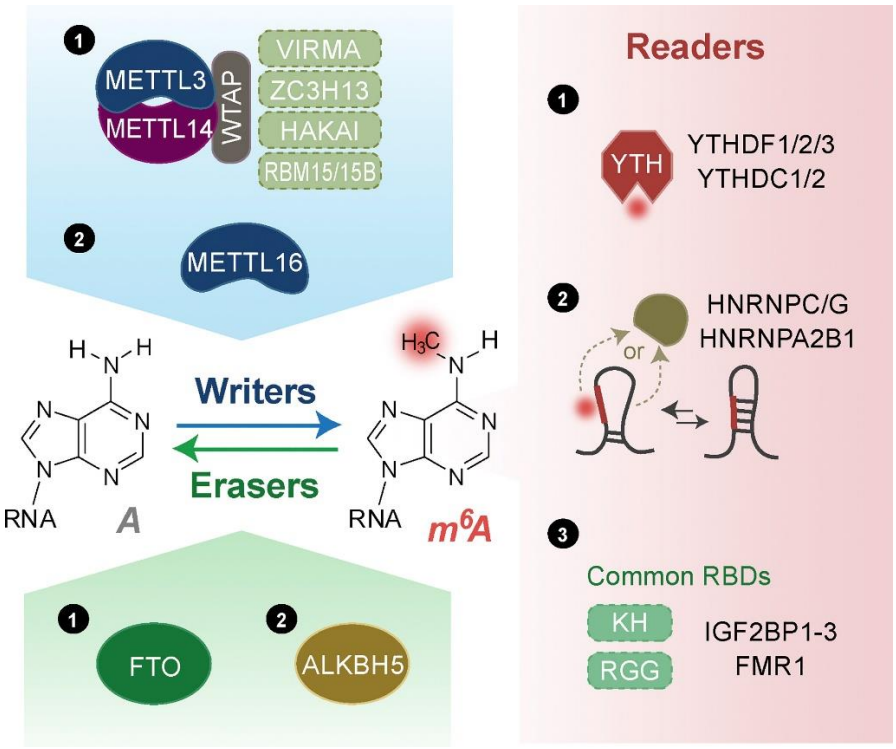


Fig 2.5. m^{6}A writers, erasers and readers (Shi et al., 2019). METTL3-METTL14-WTAP methyltransferase complex accounts for the majority of m^{6}A in mammalian transcriptomes and METTL16 methylates adenosine in a structured region. m^{6}A can be removed by FTO and ALKBH5 which belong to RNA demethylase. YTH proteins including YTHDF1/2/3 and YTHDC1/2, HNRNP proteins that can recognize m^{6}A in different contexts.

m^{6}A writers

Almost twenty years after the discovery of m^{6}A in protein coding RNA, corresponding methyltransferase complex was purified and the AdoMet-binding subunit was cloned subsequently, namely METTL3 (Bokar et al., 1994; Bokar et al., 1997). Followed by identification of METTL14 being a direct interaction factor of METTL3, this heterodimer is characterized as the methyltransferase responsible for the vast majority of m^{6}A in mouse and human transcriptomes (Liu et al., 2014; Wang et al., 2014b). Biochemical and structural works demonstrate that METTL3 is the catalytic module while METTL14 as an allosteric activator is essential for RNA binding (Figure 2.6) (Sledz and Jinek, 2016; Wang et al., 2016b; Wang et al., 2016c). Further studies reported a larger complex consisting of accessory subunits, including Wilms tumor 1-associating protein (WTAP), Vir like m^{6}A methyltransferase associated protein (VIRMA), zinc finger CCCH-type containing 13 (ZC3H13), RNA binding motif protein 15/15B (RBM15/15B) and an E3 ubiquitin ligase HAKAI that are important for the activity and specificity of this m^{6}A writer (Patil et al., 2016; Ping et al., 2014; Ruzicka et al., 2017; Wen et al., 2018; Yue et al., 2018). For example, WTAP interacts with METTL3/METTL14 to recruit them to mRNA targets and is required for

their nuclear localization as well as the methyltransferase activity (Ping et al., 2014). ZC3H13 contributes to m^{6}A methylation by anchoring WTAP-VIRMA-HAKAI-ZC3H13 complex in nucleus (Wen et al., 2018). RBM15/RBM15B interacts and directs m^{6}A methyltransferase complex to methylate adenosine sites proximal to m^{6}A consensus motifs (Patil et al., 2016). These regulatory components ensure co-transcriptional installation of m^{6}A by METTL3/METTL14 core subunits in nucleus, however, other studies also give hint that METTL3 may form an assembly in cytoplasm (Alarcon et al., 2015; Choe et al., 2018). It is not clear whether such cytosolic METTL3-containing complex resembles the nuclear one.

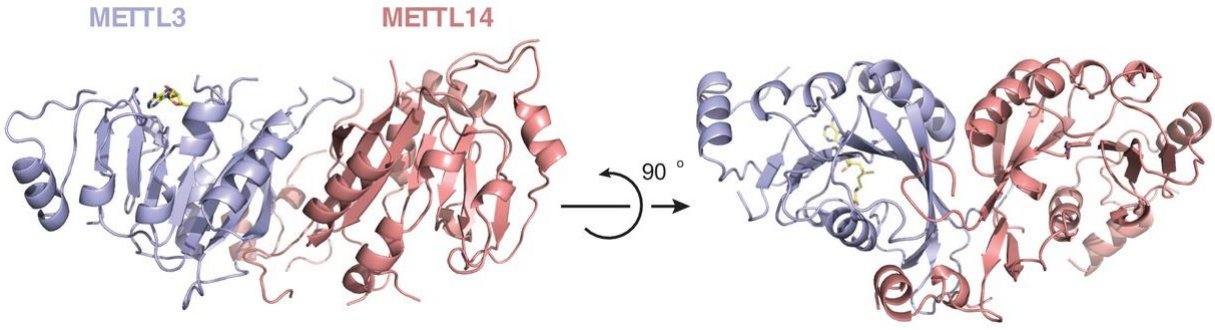


Fig 2.6. Crystal structure of METTL3-METTL14 heterodimer. The SAM donor bound to METTL3 is shown as stick format and colored yellow (Sledz and Jinek, 2016).

As the principal m^{6}A methyltransferase, METTL3/METTL14 and their homologs exert essential functions in multiple physiological processes. Deletion Ime4 (METTL3 in yeast) or mutation its catalytic domain impedes the progression of meiosis and induces foraging pseudohyphal (PH) development in yeast (Agarwala et al., 2012). Drosophila Ime4 is essential for viability and targets Notch signaling pathway to maintain oogenesis (Hongay and Orr-Weaver, 2011). In Arbidopsis, removal of METTL3 ortholog Mta results in a 90% reduction of m^{6}A, accompanied by defects in growth, apical dominance and embryonic development (Bodi et al., 2012; Zhong et al., 2008). *Mettl3* or *Mettl14* KO mice are lethal because embryos in these mutant mice fail to thrive at E5.5 (Batista et al., 2014; Geula et al., 2015). Conditional knockout *Mettl3*/*Mettl14* in different mouse tissues further indicate a central regulatory role of this complex in spermatogenesis (Lin et al., 2017), brain development and function (Xu et al., 2020; Yoon et al., 2017), innate immune response (Rubio et al., 2018; Winkler et al., 2019) and cardiac homeostasis (Dorn et al., 2019). Apart from its physiological functions, METTL3-METTL14 complex is also linked to a broad range of cancer types and type 2 diabetes (De Jesus et al., 2019;

Huang et al., 2020; Ianniello et al., 2019; Yang et al., 2019). All these independent works point out the importance of regulating METTL3/METTL14 in normal physiology as well as in disease.

Though most m^{6}A sites in mammalian transcriptomes are deposited by METTL3-METTL14 in linear regions, still a small proportion is catalyzed by additional methyltransferases. Recently, METTL16 was shown to install m^{6}A in certain RNA targets (Aoyama et al., 2020; Nance et al., 2020; Pendleton et al., 2017; Warda et al., 2017). Structure of METTL16 reveals that like METTL3, its catalytic domain adopts the classic Rossmann-like fold and uses SAM as the methyl group donor (Figure 2.7) ((Doxtader et al., 2018; Mendel et al., 2018). Until now, two targets of METTL16 are confirmed: U6 snRNA which is involved in pre-mRNA splicing and MAT2A mRNA whose gene encodes the SAM synthetase MAT2A. Instead of methylating adenosine in DRACH motif, METTL16 catalyzes m^{6}A methylation in a consensus sequence ((UAC**A**GARAA) that located in a structural region (Doxtader et al., 2018; Mendel et al., 2018). Additionally, METTL16 was also shown to bind a long noncoding RNA (Brown et al., 2016).

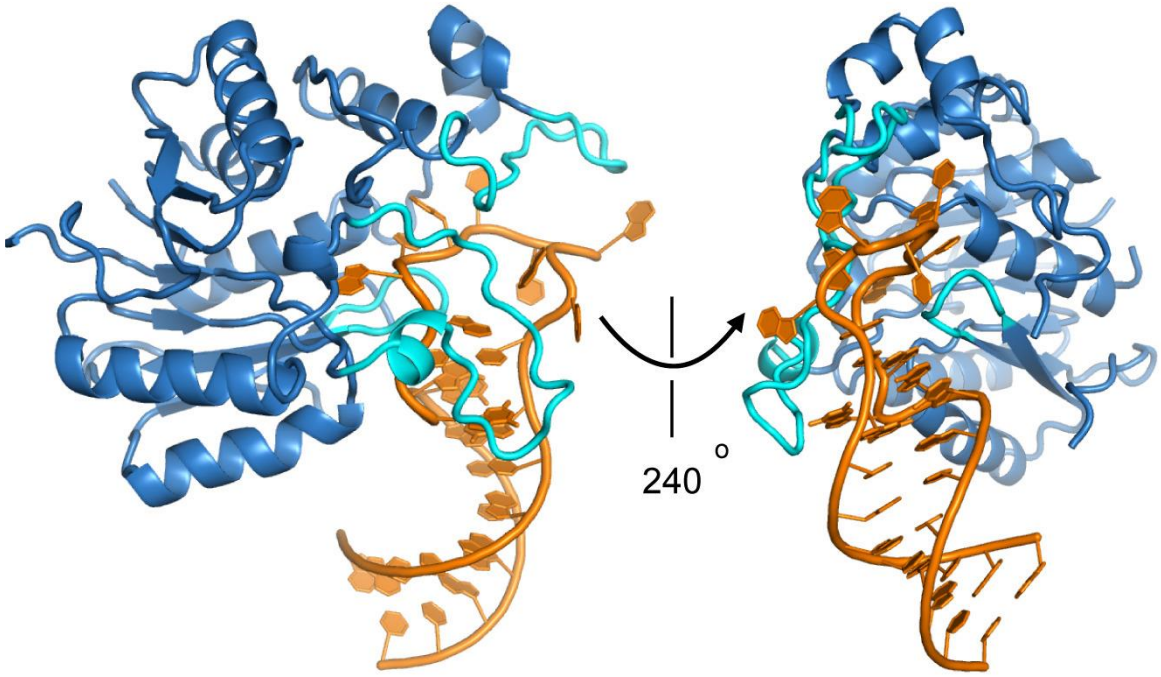


Figure 2.7. Crystal structure of METTL16 methyltransferase domain bound to MAT2A Hairpin 1. Protein is shown in blue, RNA in orange, and "clamp" loops interacting with the RNA in cyan (Doxtader et al., 2018).

In human cells, METTL16 regulates MAT2A mRNA splicing in response to cellular SAM levels (Pendleton et al., 2017). In mouse, knockout Mettl16 leads to embryonic lethality. Conditional knockout of Mettl16 in adult mouse testis also causes infertility due to lack of mature germ cells (Mendel et al., 2018). C. elegans genome encode a methyltransferase METT-10, the

ortholog of METTL16, and it deposits an m^6A marker on the the 3′ splice site (AG) of the SAM synthetase pre-mRNA, indicating the conserved function of METTL16. Presence of m^6A at the 3′ splice site inhibits proper splicing of worm synthetase RNA and this inhibitory effect is conserved in mouse. Mechanistically, 3′ splice site m^6A blocks the binding of splicing factor U2AF35 to the 3′ splice site, implicating that site-specific m^6A may be critical for splicing regulation (Mendel et al., 2021).

Besides METTL3/METTL14 and METTL16, another methyltransferases are also demonstrated to modify m^6A in other types of RNA. METTL5-TRMT112 and ZCCHC4 catalyze m^6A methylation in human 18S rRNA and 28S rRNA, respectively (Ma et al., 2019; van Tran et al., 2019). The diversity of m^6A methyltransferase reflects that as an important post-transcriptional modification, m^6A can be added either in a type-specific or context-specific manner and such addition requires tight and dynamic regulation.

m^6A erasers

One remarkable aspect of m^6A modification is that apart from its deposition by RNA methyltransferases, the methyl group can also be removed by RNA demethylases. Until now, two enzymes are verified to be able to convert m^6A to A on mRNAs, including fat mass and obesity-associated protein (FTO) and ALKBH5 (Jia et al., 2011; Zheng et al., 2013). These erasers belong to the Alkb family of dioxygenases, of which catalytic reaction requires involvement of α-ketoglutarate, ferrous and oxygen (Falnes et al., 2002; Gerken et al., 2007). Substrates of Alkb enzymes are composed of DNA and RNA, and each protein appears to have specificity towards different substrates.

According to its name, FTO was initially identified as a protein linked to obesity (Church et al., 2010; Dina et al., 2007; Fischer et al., 2009). Later, its demethylase activity was reported that FTO could remove 3-methylthymidine (m^3T) and 3-methyluridine (m^3U) in single-stranded DNA and RNA, respectively (Gerken et al., 2007; Jia et al., 2008). Based on these two studies, the relatively higher activity towards RNA nucleotides was observed. Followed by another work published in 2011, FTO was shown to efficiently demethylate m^6A in mRNA in vitro and knockdown FTO leads to the increased m^6A level in HeLa cells , suggesting that m^6A-containing mRNAs could be the main target of FTO (Jia et al., 2011). However, several studies pointed out that m^6A abundance was not mediated by FTO in other tissues. For instance, analysis of mouse *Fto*-knockout brain transcriptome did not show a significant increase in m^6A levels (Hess et al., 2013). Neither the

increase of m^6A sites in mRNA was observed from mouse embryos and cells lacking Fto (Mauer et al., 2017). Using antibody-independent approach to map m^6A also revealed that overall abundance of m^6A was not changed by FTO depletion (Garcia-Campos et al., 2019). Taken together, FTO-mediated m^6A seems to be complex and to some extent, is irrelevant to physiological processes.

In addition to m^6A, a big advance in defining FTO substrate is that m^6A_m can be demethylated by FTO to form A_m as well. In vitro biochemical experiments showed a much higher demethylation rate towards m^6A_m compared with m^6A. As discussed above, m^6A_m can be found in mRNA as well as in snRNA at the first transcribed nucleotide. Quantification of cellular m^6A_m level in poly(A) RNA in *FTO*-knockout cells indicated m^6A_m appears to tolerate the loss of FTO (Mauer et al., 2017). Whereas for snRNA, *FTO*-knockout cells showed >50% $m6A_m$ stoichiometries, confirming snRNAs are FTO targets (Mauer et al., 2019). Considering the fact that snRNAs mediate pre-mRNA splicing and *FTO*-knockout cells display splicing defects (Bartosovic et al., 2017), it is postulated that FTO-mediated $m6A_m$ is a key regulatory information during splicing.

ALKBH5 is the second identified m^6A demethylase in mammalian cells (Zheng et al., 2013). Unlike FTO which is localized in both nucleus and cytoplasm, ALKBH5 is exclusively enriched in nucleus, supporting that ALKBH5 probably demethylates nuclear RNA targets to regulate those events occurring in nucleus, such as mRNA transcription, splicing and export. Knockdown ALKBH5 increased cellular m^6A levels, while overexpression ALKBH5 decreased m^6A more significantly (Zheng et al., 2013). ALKBH5 is not essential for mouse viability but is required for fertility, as *ALKBH5*-knockout mice testes showed defective spermatogenesis. RNA-seq analysis indicated that in mouse testis, around 1500 genes were affected by ALKBH5 in terms of their expressions, including those genes involved in p53 functional interaction network (Zheng et al., 2013).

Physiological functions of ALKBH5 remain unclear except its essence for mouse germ cells, but certain cancer cells are correlated with elevated expression of ALKBH5, such as glioblastoma stem cells (Dixit et al., 2017; Zhang et al., 2017). ALKBH5 can demethylate the transcription factor FOXM1 pre-mRNAs, resulting in enhanced RNA stability and expression. The other example is breast cancer stem cell. When these cells are exposed to hypoxic environment, upregulated expression of ALKBH5 is also detected (Thalhammer et al., 2011; Zhang et al., 2016a). Similar to FOXM1, NANOG mRNA which encodes a pluripotency factor, is demethylated

by ALKBH5 to promote breast cancer stem cell phenotype. Overall, not only ALKBH5 is crucial for sperm development, but also its upregulation is correlated with tumorigenesis under certain conditions.

To sum up, in contrast to writers which install m^6A modification, FTO and ALKBH5 represent an additional mechanism that regulates distribution and abundance of m^6A in mammalian transcriptomes by removing such marker. This dynamic balance further ensures the readers to regulate cellular processes via recognizing this modification.

m^6A readers

The multifaceted nature of m^6A is relying upon its reader proteins that can specifically recognize this chemical marker. In vertebrates, five proteins are found to sense m^6A methylation through a highly conserved YT521-B homology (YTH) domain. These YTH domain-containing proteins, based on their amino acid sequences, are divided into three groups: YTH domain family 1-3 (YTHDF1-3), YTH domain-containing 1 (YTHDC1) and YTH domain-containing 2 (YTHDC2). The YTH domain was previously predicted to a putative RNA-binding domain before it was demonstrated to recognize m^6A (Stoilov et al., 2002). The first clue implicating YTH domain is an m^6A-binding domain came from an in vitro m^6A pull-down assay where two YTHDF proteins (YTHDF2-3) were detected (Dominissini et al., 2012). Later on, structure of the human YTHDF2 YTH domain bound to m^6A was determined, indicating that YTH domain recognizes m^6A and binds to RNAs in an m^6A-dependent manner (Li et al., 2014; Zhu et al., 2014).

In human, these five m^6A readers differ in protein sequence and length, except the YTH domain (Figure 2.8). Several structural studies have revealed that YTH domain, containing approximate 150 amino acids, consists of four α helices and six β sheets. The m^6A is accommodated within an aromatic cage in the YTH domain where two or three tryptophan residues engage with the methyl moiety (Luo and Tong, 2014; Theler et al., 2014; Xu et al., 2015; Xu et al., 2014). All the readers appear to recognize m^6A in the DRACH motif without obvious sequence preference, except YTHDC1. YTHDC1 shows the strongest binding affinity when bound to RNA target having upstream G next to m^6A (GGm^6ACU), which could result from the formation of hydrogen bond between carbonyl group of G and nitrogen group from main chain (Xu et al., 2014).

YTH proteins regulate almost every aspect of RNA metabolism, including pre-mRNA splicing, mRNA export, mRNA decay and translation. It is believed that these m^6A readers function in a context-dependent way to influence specific physiological processes.

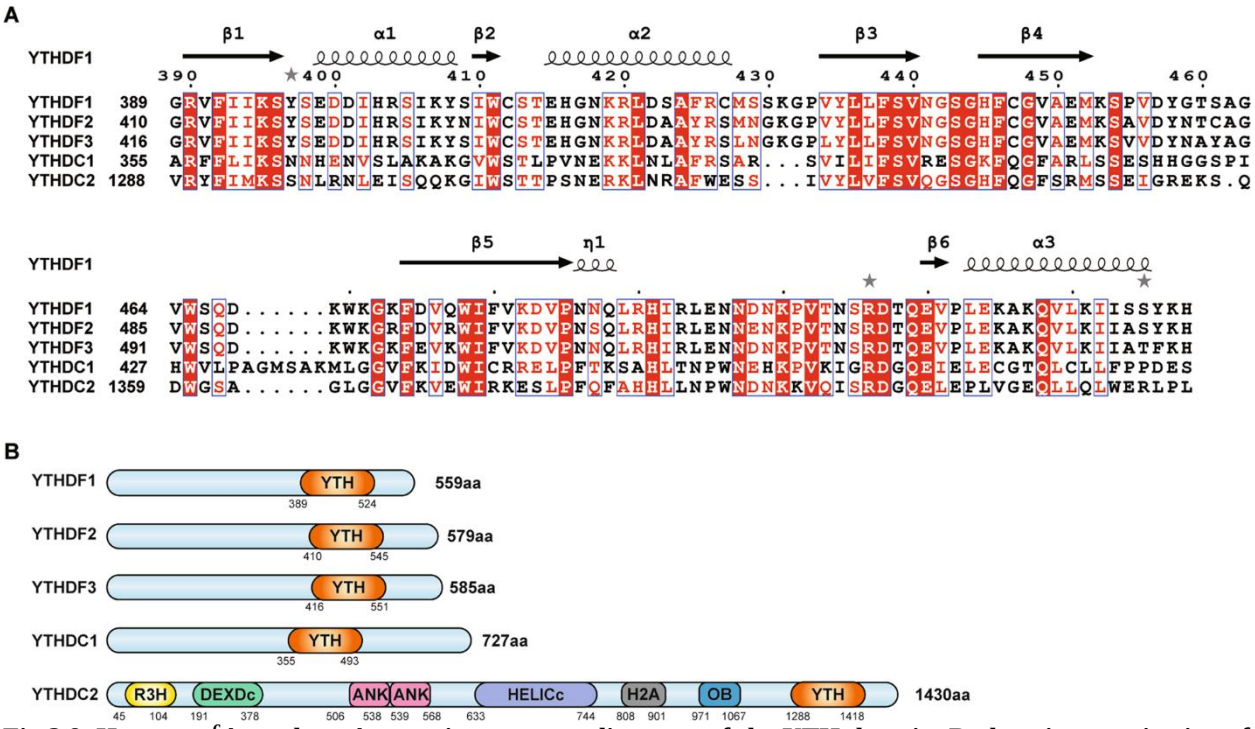


Fig 2.8. Human m^{6}A readers. A, protein sequence alignment of the YTH domain. B, domain organization of each human m^{6}A reader. All the five proteins include the C-terminal conserved YTH domain (Liao et al., 2022).

YTHDF1-3

YTHDF readers are localized in cytoplasm, hence, these proteins mainly modulate m^{6}A-associated events occurring in cytosol. Compared with YTHDC proteins, besides the YTH domain, YTHDF proteins also share high sequence identity over the rest region which contains low-complexity sequences and prion-like P/Q/N-rich domains (Patil et al., 2018). Recent work indicated that these unstructured regions are required for YTHDF proteins that undergo phase separation and m^{6}A can enhance this process (Ries et al., 2019).

Early studies proposed that YTHDF proteins specifically mediate m^{6}A RNA processing. YTHDF1 was shown to enhance the translation of m^{6}A-containing mRNAs via recognizing m^{6}A sites in 3′ UTR and interacting with translation initiation factor eIF3 (Wang et al., 2015). Moreover, YTHDF1 could promote mRNA translation elongation by recognizing m^{6}A located in 5′ UTR and binding to eEF2 (Li et al., 2020). Similar effect was also observed for Snail mRNA where YTHDF1 targets m^{6}A sites in its CDS region to induce translation elongation (Lin et al., 2019). According, the influence of YTHDF1 on mRNA translation appears to be context-specific. YTHDF1 is unlikely to be involved in mRNA decay because knockdown YTHDF1 has little effect on the stability of m^{6}A-containig mRNAs (Wang et al., 2015).

Among YTHDF proteins, function of YTHDF2 was first characterized, with a major impact on mRNA decay. Human YTHDF2 predominantly resides in 3′ UTR in more than 3000 m^6A-containing transcripts. Depletion of YTHDF2 leads to the accumulation of these transcripts in active polysome as well as their increased cellular level, suggesting YTHDF2 guides m^6A mRNAs for degradation (Wang et al., 2014a). This negative effect on mRNA stability requires both the C-terminal YTH domain and N-terminal region which directly interacts with the superfamily homology (SH) domain of CNOT1, a component of CCR4-NOT deadenylation complex. YTHDF2 binds to the m^6A marker of mRNAs and recruits CCR4-NOT complex to degrade transcripts (Du et al., 2016). Furthermore, YTHDF2 can trigger Endoribonucleolytic cleavage of m^6A-containing RNAs through interacting with HRSP12 which will further recruit endoribonucleases RNase P and RNase MRP (Park et al., 2019). This role of YTHDF2 in directing mRNA degradation is consistent with its localization in P-bodies which is known to be involved in mRNA decay (Wang et al., 2014a). Overall, YTHDF2 regulates m^6A-containing mRNA degradation in cytoplasm.

At present, function of YTHDF3 is not clear. One hypothesis is that when YTHDF3 is close to YTHDF1, it promotes mRNA translation. Likely, YTHDF3 promotes the degradation of m^6A-modified mRNA in synergy with YTHDF2 (Li et al., 2017a; Shi et al., 2017). However, in 2020, two groups proposed a unified model for YTHDF proteins in regards to their regulatory function on m^6A. YTHDF proteins share the same mRNA targets rather than different mRNAs. Instead regulate m^6A-containing mRNA translation, all YTHDF proteins indeed act redundantly to promote mRNA degradation. They clearly showed that knockdown individual YTHDF protein almost has no impact on mRNA stability, but only when all three YTHDF proteins were depleted simultaneously, half-life of m^6A-containing mRNAs is significantly decreased, indicating in the cytoplasm, YTHDF proteins collectively mediate m^6A mRNA decay (Lasman et al., 2020; Zaccara and Jaffrey, 2020).

YTHDC1

As the founding member of YTH family proteins, YTHDC1 is the only known m^6A reader that localizes in nucleus, particularly in the YT bodies that are adjacent to nuclear speckles (Nayler et al., 2000). In agreement with its nuclear localization, YTHDC1 regulates pre-mRNA splicing by connecting trans- and cis-regulatory elements and affects alternative splicing patterns. YTHDC1 may bind pre-mRNA targets as soon as they are transcribed and methylated, then splicing factors,

such as SRSF3 and SRSF10 (serine/arginine-rich splicing factor), are recruited by YTHDC1 to the targets so as to facilitate splicing. Remarkably, these splicing factors interact with YTHDC1 in a competitive manner. On one hand, binding of SRSF3 at the N-terminal domain of YTHDC1 prevents the binding of SRSF10, leading to exon inclusion of pre-mRNA during splicing. One the other hand, once the N-terminal region is bound to SRSF10, YTHDC1 promote exon skipping (Xiao et al., 2016). Notably, the exon inclusion is probably predominant in the YTHDC1-mediated splicing events because YTHDC1 shows a much higher affinity with SRSF3 over SRSF10. Not only involved in alternative splicing, but also YTHDC1-SRSF3 complex modulates 3′ end alternative polyadenylation of pre-mRNA via interacting with processing factor CPSF6 (Kasowitz et al., 2018).

After splicing, YTHDC1 was shown to direct m^6A-containing mRNAs to export to cytoplasm. In this process, YTHDC1 and SRSF3 collaborate with the nuclear RNA export factor (NXF1) to facilitate the exportation of methylated mRNAs. NXF1 binds SRSF3 and this interaction can be enhanced by YTHDC1 (Roundtree et al., 2017b). Alternatively, m^6A-containing mRNA export can be driven by the transcription and export complex (TREX) via recruiting YTHDC1 to the mRNA target (Lesbirel et al., 2018).

Beyond its regulation of m^6A in mRNA, YTHDC1 was also reported to act on noncoding RNA. For example, YTHDC1 preferentially recognizes m^6A sites on XIST and is essential for XIST function. XIST is a long noncoding RNA that is heavily modified by m^6A. This noncoding RNA contributes to the inactivation of X chromosome and transcriptional repression of genes on X chromosome (Patil et al., 2016). It is postulated that binding of YTHDC1 to XIST may cause additional recruitment of proteins that involved in epigenetic silencing.

YTHDC2

YTHDC2 is the largest protein among YTH family proteins. Compared with other YTH proteins which only contain C-terminal YTH domain and remaining low-complexity regions, YTHDC2 includes multiple domains along its entire length. There is a N-terminal R3H domain, a central DEAH-box helicase domain where two RecA domains are interrupted by two ankyrin repeats, a helicase-associated 2 domain (HA2) followed by OB-fold (oligonucleotide/oligosaccharide-binding fold) and the C-terminal YTH domain (Figure 2.8). In line with its domain organization, YTHDC2 was demonstrated to possess ATP-dependent RNA helicase activity. It resolves the duplex region of RNA targets in a 3′ to 5′ manner (Jain et al., 2018; Wojtas et al., 2017).

Unlike other YTH proteins which are ubiquitously expressed in different tissues, YTHDC2 is highly enriched in germ cells (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018; Wojtas et al., 2017). *Ythdc2* knockout mice are infertile. In the knockout testes, mitotic germ cells enter early stage of meiosis but undergo premature spermatogenesis and apoptosis in the end. By preforming Immunoprecipitation coupled to mass spectrometry (IP-MS), YTHDC2 is associated with another germ cell-specific protein MEIOC as well as the 5′ to 3′ exoribonuclease XRN1 in mouse testis (Abby et al., 2016; Jain et al., 2018; Wojtas et al., 2017). Knockout *Meioc* phenocopies that of *Ythdc2* in mouse, implicating that YTHDC2 may form a complex with MEIOC to regulate meiosis-associated transcripts. Interaction between YTHDC2 and XRN1 is specific because all the other YTH proteins do not bind to XRN1 (Kretschmer et al., 2018). However, it is unclear whether YTHDC2 interacts with XRN1 and MEIOC simultaneously.

How YTHDC2 regulates spermatogenesis?

As mentioned above, YTHDC2 is essential for mouse fertility but the underlying mechanism remains elusive. Although its C-terminal YTH domain binds to m^6A in vitro, the binding affinity is the weakest among all YTH proteins (Wojtas et al., 2017; Xu et al., 2015). Analysis of YTHDC2-binding sites in transcripts by CLIP experiment have shown a limited overlap with m^6A sites (Patil et al., 2016). All these data point to a question: is m^6A recognition ability of YTHDC2 relevant to its in vivo function, particular in the germ cells. YTHDC2 is associated with XRN1 and MEIOC, however, we don't know whether the interactions are direct or not, and if so, what are the physiological functions of this complex?

During my PhD, I collaborate with my colleagues to try to address these questions. We decipher that YTHDC2 preferentially binds to U-rich motifs in the 3′ UTR of testicular transcripts. Mice carrying m^6A reading-deficient Ythdc2 are fertile, indicating the YTH domain is not required for mouse fertility. However, the helicase activity of YTHDC2 is essential for mammalian germ cells. YTHDC2 directly interacts with XRN1 and the coiled coil domain of MEIOC. In vitro biochemical assay shows that XRN1 stimulates the ATPase and helicase activity of YTHDC2. Using single-cell RNA sequencing, we uncover a unique role of YTHDC2 in ensuring proper separation of mitotic transcripts from meiotic transcripts when spermatogonial cells enter meiosis, because loss of *Ythdc2* leads to a mixed identity of transcriptome in meiotic germ cells that fail to complete the whole process of meiosis. In the end, we observe that Ythdc2 knockout zebrafish is infertile as well, suggesting YTHDC2 is functionally conserved in regulating animal fertility.

Results

This section includes a peer-reviewed paper entitled “The XRN1-regulated RNA helicase activity of YTHDC2 ensures mouse fertility independently of m^{6}A recognition”, published in Molecular Cell on 18th March 2022. In this paper, we described that by performing iCLIP experiment, we discovered that YTHDC2 specifically locates to the 3′ UTR of testicular transcripts by using U-rich motif as the loading site. We further did in vitro RNA binding assay and found that YTHDC2 preferentially binds to poly(U) oligonucleotides rather than poly(A) oligonucleotides. Strikingly, such U-rich sequence is not overlapped with the DRACH motif. We introduced a point mutation (W1375A) in mouse genome and found that heterozygous and homozygous Ythdc2^{YTH} knockin mice with the mutant YTH domain are viable and fertile, indicating m^{6}A reading capacity of Ythdc2 is not required for mouse fertility. Since YTHDC2 was already demonstrated as an RNA helicase, we wondered whether its helicase activity is relevant to the phenotype. We created a mouse knockin mutant that has a point mutation in the ATP hydrolysis motif DEVH (E332Q). Mice carrying catalytic-dead Ythdc2 are viable but infertile. Even crossed with wildtype mice, such Ythdc2$^{+/cat\text{-}dead}$ mice can’t produce litter, indicating this E332Q mutation is dominant negative. RNA-seq analyses showed dramatic gene expression changes in catalytic-dead mutant testis compared with wildtype testis. Those genes that are significantly downregulated are normally highly expressed in meiotic spermatocytes. Hence, the helicase activity of YTHDC2 is indispensable for mouse fertility. With IP-MS, we confirmed that YTHDC2 is associated with XRN1 as well as MEIOC in mouse testis. Moreover, we identified RBM46 as a novel interaction factor of YTHDC2. We further produced their recombinant proteins and found that YTHDC2 directly interacts with XRN1 via its ankyrin repeats. YTHDC2 also binds to the coiled coil domain of MEIOC in vitro. To investigate the biological consequence of these complexes, we established a fluorescence-based helicase assay and observed that YTHDC2 itself is a weak helicase, probably due to its unique domain organization where two RecA domains are inserted by two ankyrin repeats. However, XRN1, which binds to the ankyrin repeats, stimulates the helicase activity of YTHDC2. We speculated that XRN1 can allosterically regulate the conformational change of YTHDC2 during unwinding process. In the end, we did single-cell sequencing of transcriptomes from wildtype and *Ythdc2* knockout testis. Our data indicated that YTHDC2 governs a fine separation of the transcriptomes when germ cells transition from mitosis to meiosis. In the absence

of YTHDC2, cells do enter into meiosis but still express mitotic genes, and such a mixed transcriptome fails to support further progression into late stage of meiosis.

In this project, I produced all the recombinant proteins required for experiments. I also carried out all the in vitro biochemical assays. Radha Raman Pandey created mouse mutant, performed iCLIP experiment and prepared sequencing libraries. Pascal Gos prepared FACS-purified germ cells. Kyrylo Krasnykov did the single-cell sequencing analysis. David Homolka analyzed mass spectrometry data and made all the figures. Mateusz Mendel and Richard J. Fish produced zebrafish mutant and analyzed phenotype. Radha Raman Pandey and Ramesh S. Pillai designed experiments and Ramesh S. Pillai wrote the manuscript with input from everyone.

Apart from the protein production and biochemical experiments, in order to understand how YTHDC2 interacts with XRN1, I also did cryo electron microscopy (cryo-EM) analysis of YTHDC2-XRN1 complex. I made a great effort to try to solve the structure of this complex, but I failed in the end. I could only resolve a model of XRN1 but not YTHDC2, nor the complex. Therefore, we did not put the cryo-EM data in our paper. Here I just show some preliminary results. First, I did negative staining to check the quality of sample (Figure 2.9). The sample appeared to be good for cryo-EM. Then I prepared cryogenic grids and we collected several datasets at EMBL

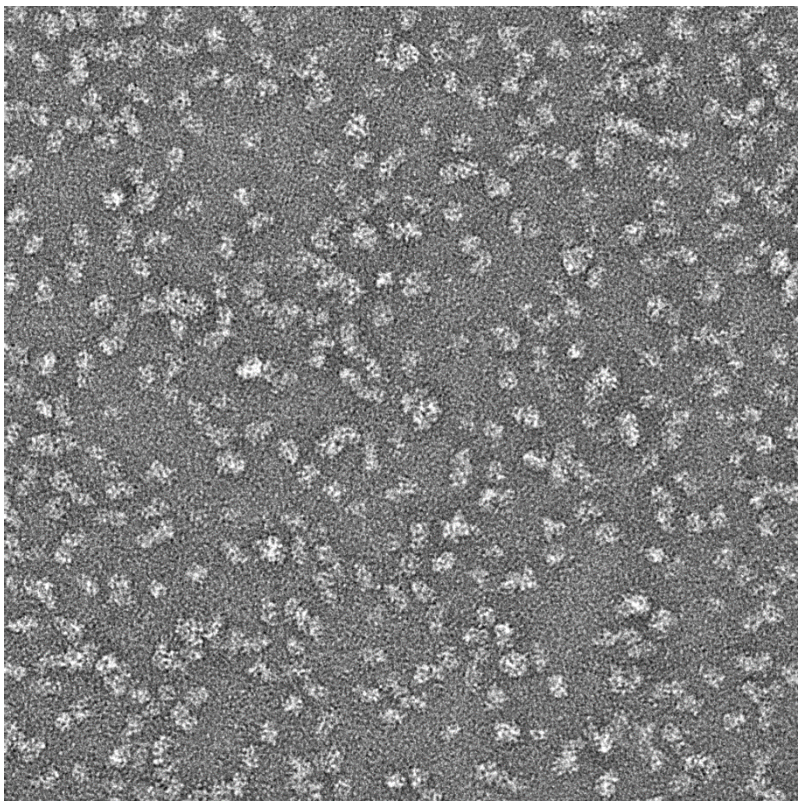
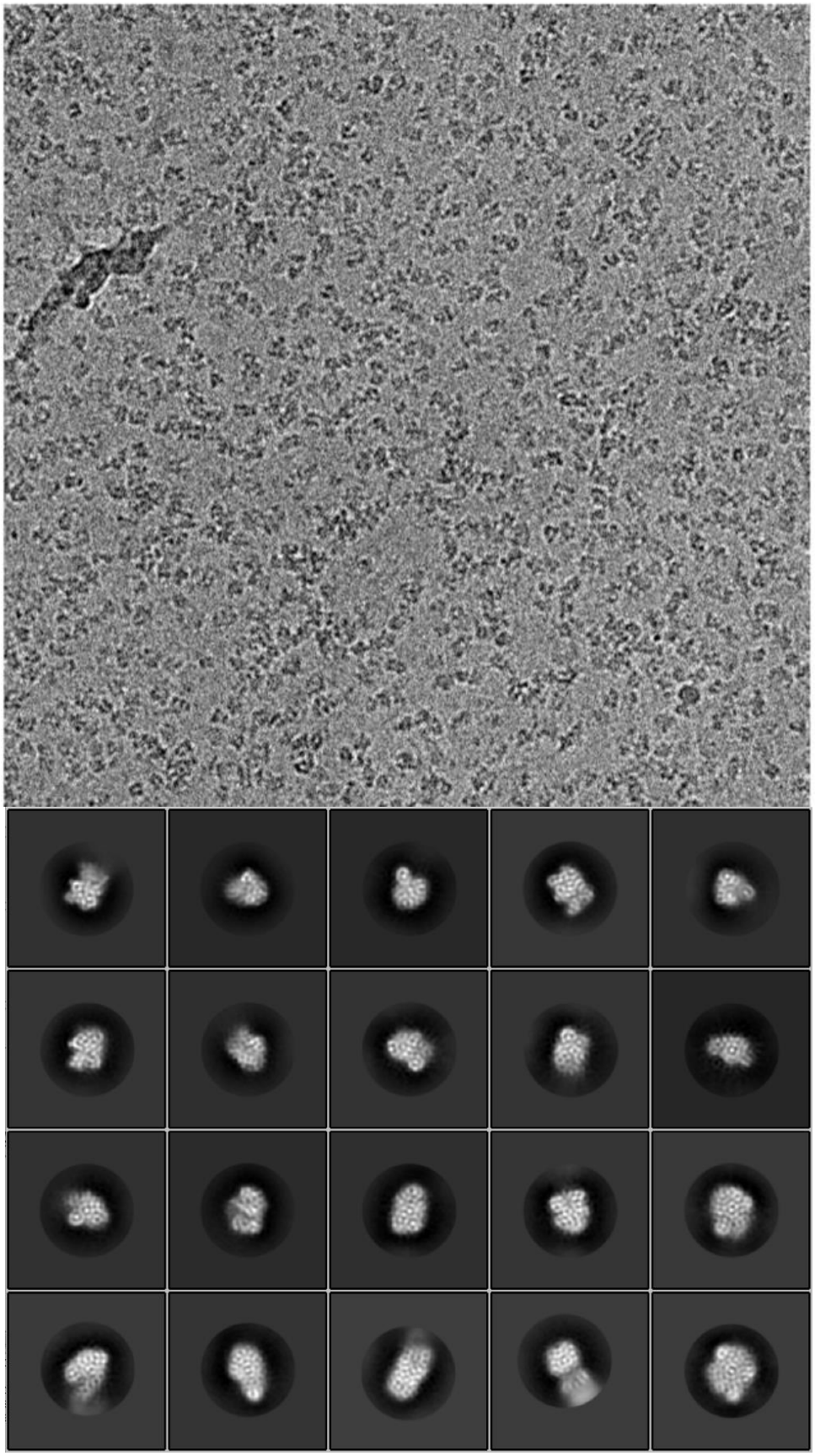


Fig 2.9. Negative staining of YTHDC2-XRN1 (1-1279)

Grenoble and CryoGEnic facility in University of Geneva. One of the dataset allowed us to resolve a 3D volume of XRN1 at 4 angstrom (Figure 2.10). I also obtained a model supposed to be YTHDC2 (Figure 2.11), but I could not validate it because the resolution was quite bad. The main problem is that this complex is fragile and easy to fall apart in ice, even when I use crosslinking reagent to stabilize the complex.

There is a cover for this paper, which is designed by Anne-Claire Godet (THREONINE).



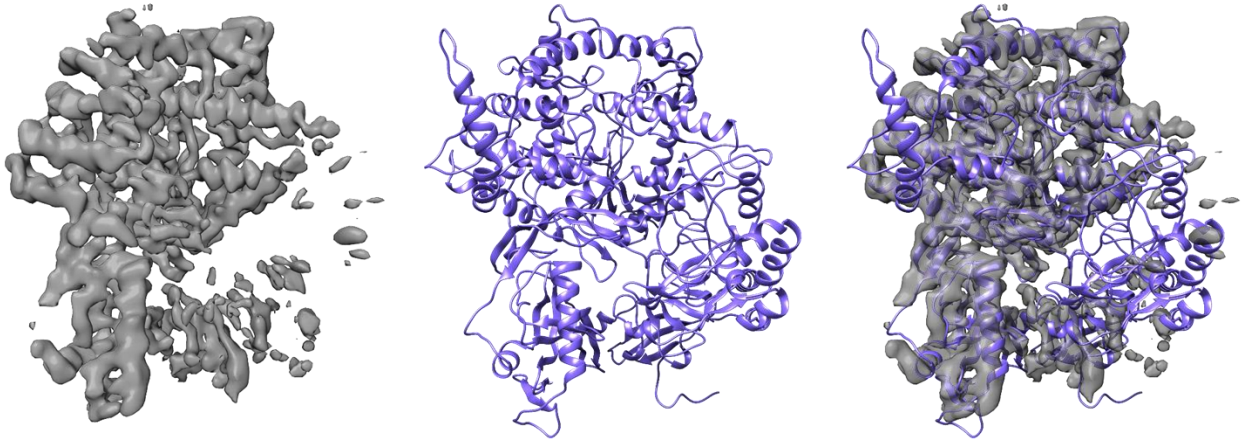
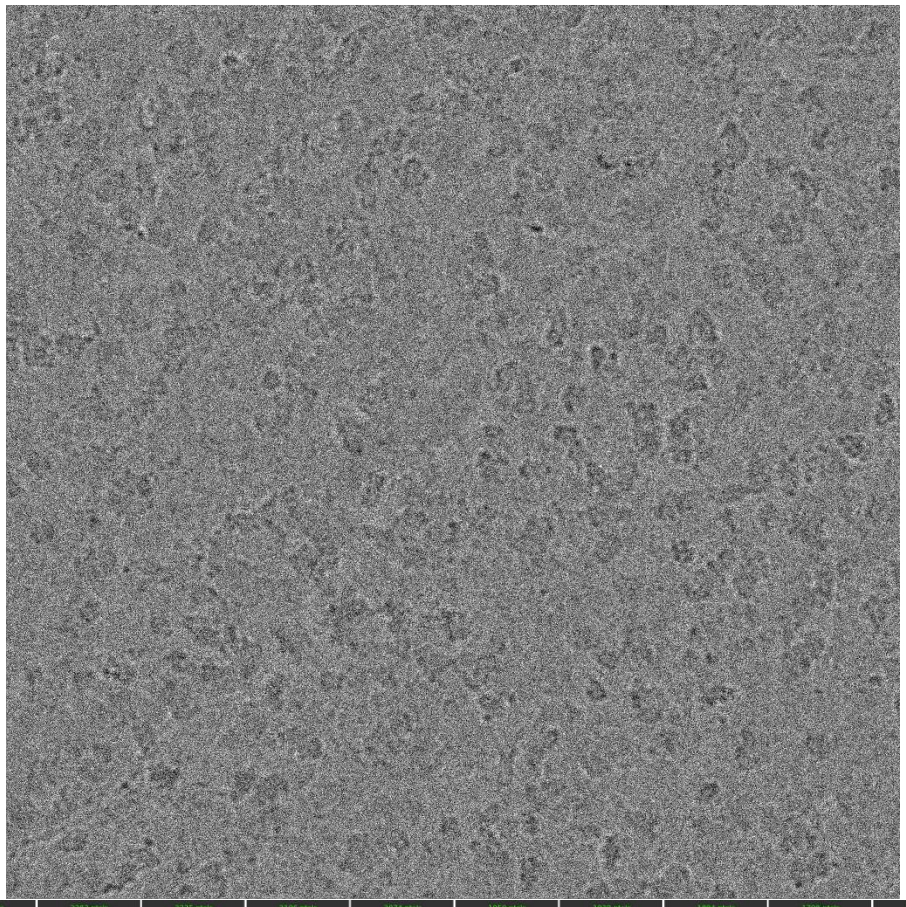


Figure 2.10. Micrograph, 2D classification and 3D reconstruction (XRN1 dataset). The resolved model of XRN1 is fitted with Drosophila XRN1 structure (PDB: 2Y35).




2747 ptcls
7.6 A 2 ess
2283 ptcls
7.9 A 1 ess
2225 ptcls
9.1 A 2 ess
2106 ptcls
8.6 A 1 ess
2074 ptcls
10.4 A 2 ess
1950 ptcls
8.3 A 2 ess
1939 ptcls
8.4 A 2 ess
1894 ptcls
8.5 A 2 ess
1708 ptcls
10.2 A 2 ess
1677 ptcls
8.8 A 2 ess
1668 ptcls
9.7 A 2 ess
1619 ptcls
9.4 A 2 ess
1584 ptcls
9.7 A 2 ess
1557 ptcls
10.1 A 2 ess
1533 ptcls
9.5 A 2 ess
1523 ptcls
10.5 A 2 ess
1516 ptcls
10.9 A 2 ess
1515 ptcls
10.0 A 2 ess
1511 ptcls
10.7 A 2 ess
1508 ptcls
9.9 A 2 ess
1477 ptcls
8.8 A 2 ess
1467 ptcls
9.1 A 2 ess
1462 ptcls
9.4 A 2 ess
1458 ptcls
9.4 A 2 ess
1449 ptcls
11.1 A 2 ess
1429 ptcls
9.8 A 2 ess
1420 ptcls
9.2 A 2 ess
1417 ptcls
9.5 A 2 ess
1417 ptcls
8.8 A 2 ess
1410 ptcls
9.8 A 2 ess
1394 ptcls
9.9 A 2 ess
1392 ptcls
10.8 A 2 ess
1383 ptcls
9.5 A 2 ess
1380 ptcls
9.6 A 2 ess
1378 ptcls
10.3 A 2 ess
1374 ptcls
10.6 A 2 ess
1359 ptcls
9.2 A 2 ess
1353 ptcls
11.1 A 2 ess
1346 ptcls
10.1 A 2 ess
1340 ptcls
10.6 A 2 ess
1329 ptcls
9.2 A 2 ess
1328 ptcls
10.4 A 2 ess
1314 ptcls
9.1 A 2 ess
1313 ptcls
10.1 A 2 ess
1304 ptcls
10.5 A 2 ess
1301 ptcls
8.8 A 2 ess
1296 ptcls
11.4 A 2 ess
1293 ptcls
10.8 A 2 ess
1293 ptcls
9.4 A 2 ess
1288 ptcls
10.3 A 2 ess
1280 ptcls
11.4 A 2 ess
1277 ptcls
8.6 A 2 ess
1263 ptcls
9.8 A 2 ess
1239 ptcls
9.9 A 2 ess
1217 ptcls
10.3 A 2 ess
1208 ptcls
9.4 A 2 ess
1205 ptcls
11.3 A 2 ess
1204 ptcls
8.9 A 2 ess
1188 ptcls
9.5 A 2 ess
1184 ptcls
9.4 A 2 ess
1180 ptcls
10.4 A 2 ess
1174 ptcls
9.3 A 2 ess
1166 ptcls
10.2 A 2 ess
1165 ptcls
9.0 A 2 ess
1157 ptcls
10.6 A 2 ess
1155 ptcls
10.6 A 2 ess
1153 ptcls
8.4 A 3 ess
1150 ptcls
10.3 A 2 ess
1149 ptcls
9.2 A 2 ess
1147 ptcls
9.6 A 2 ess
1142 ptcls
10.1 A 2 ess
1134 ptcls
9.8 A 2 ess
1128 ptcls
10.3 A 2 ess
1117 ptcls
11.2 A 2 ess
1110 ptcls
10.2 A 2 ess
1105 ptcls
10.3 A 2 ess
1058 ptcls
10.6 A 2 ess
1043 ptcls
11.6 A 2 ess
1022 ptcls
10.7 A 2 ess
907 ptcls
9.7 A 2 ess

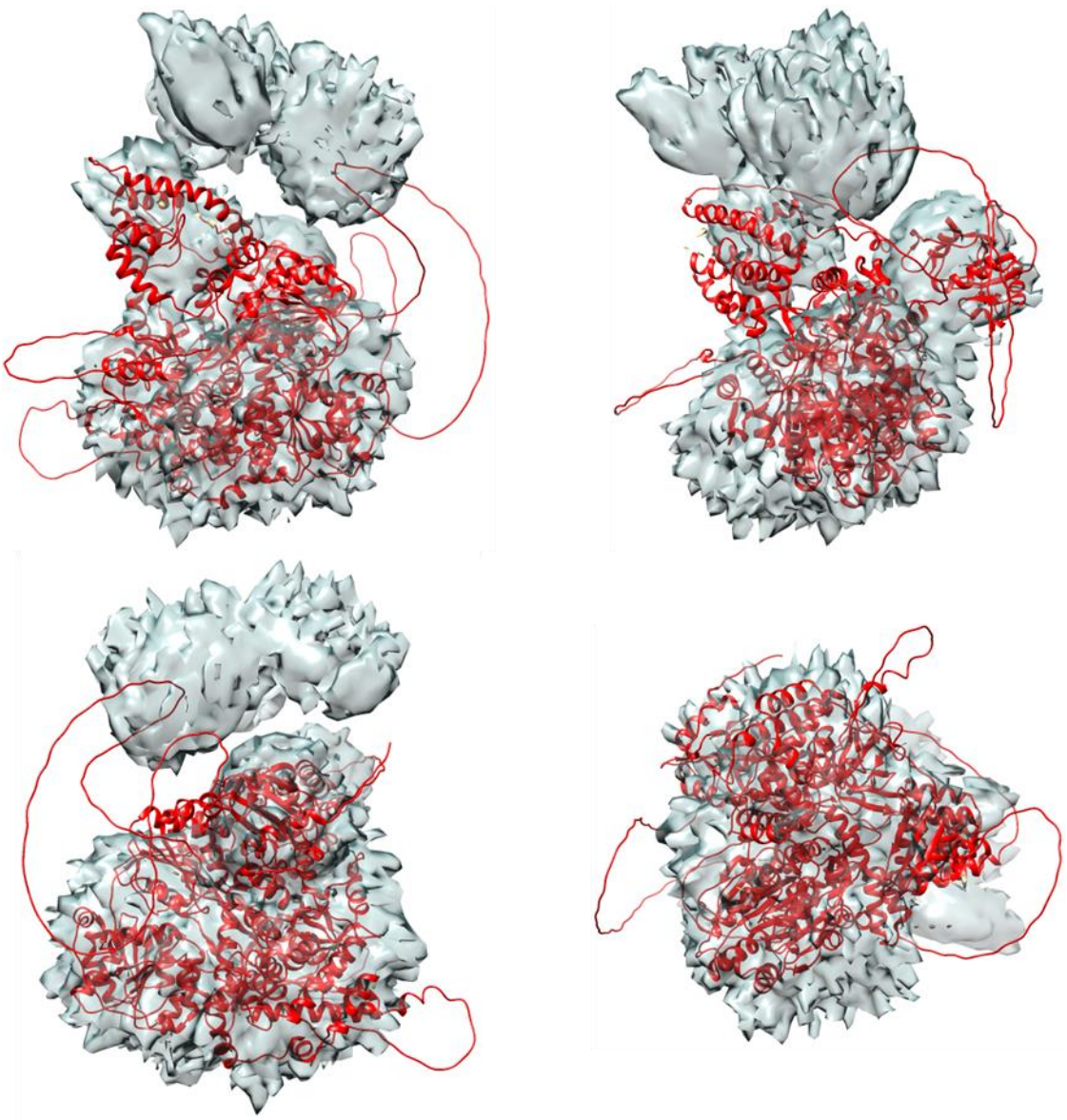


Fig 2.11. Micrograph, 2D classification and 3D reconstruction of another dataset which a potential volume of YTHDC2 is resolved. The 3D model is fitted with predicted structure of YTHDC2 by AlphaFold (Jumper et al., 2021; Varadi et al., 2022).

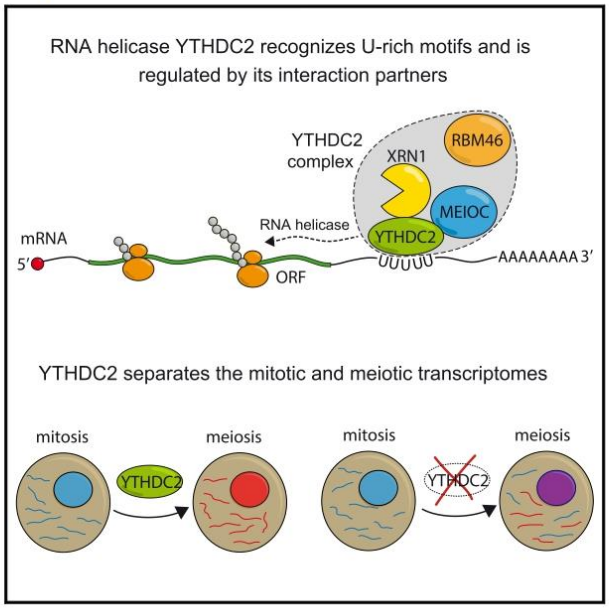

Molecular Cell
Volume 82
Number 9
May 5, 2022
CellPress
THREONINE

Molecular Cell

Article

The XRN1-regulated RNA helicase activity of YTHDC2 ensures mouse fertility independently of m^{6}A recognition

Graphical abstract



Highlights

- m^{6}A-reading capacity of YTHDC2 is not essential for mouse fertility
- YTHDC2 binds U-rich motifs in 3′ UTR of mRNAs
- The essential RNA helicase activity of YTHDC2 is enhanced by XRN1
- *Ythdc2* mutant germ cells have a mixed mitotic-meiotic identity

Authors

Lingyun Li, Kyrylo Krasnykov, David Homolka, ..., Richard J. Fish, Radha Raman Pandey, Ramesh S. Pillai

Correspondence

raman.pandey@unige.ch (R.R.P.), ramesh.pillai@unige.ch (R.S.P.)

In brief

Li et al. show that the m^{6}A-reading capacity of YTHDC2 is not essential for mouse fertility, but instead it binds U-rich motifs on RNAs and depends on its essential RNA helicase activity. YTHDC2 allows fine separation of the mitotic and meiotic transcriptomes as germ cells transition into meiosis.

Li et al., 2022, Molecular Cell *82*, 1678–1690
May 5, 2022
https://doi.org/10.1016/j.molcel.2022.02.034

CellPress

CellPress
OPEN ACCESS

Molecular Cell

Article

The XRN1-regulated RNA helicase activity of YTHDC2 ensures mouse fertility independently of m^{6}A recognition

Lingyun Li,[1,3] Kyrylo Krasnykov,[1,3] David Homolka,[1] Pascal Gos,[1] Mateusz Mendel,[1] Richard J. Fish,[2] Radha Raman Pandey,[1,*] and Ramesh S. Pillai[1,4,*]

[1]Department of Molecular Biology, Science III, University of Geneva, 30 Quai Ernest-Ansermet, CH-1211 Geneva 4, Switzerland
[2]Department of Genetic Medicine and Development, Faculty of Medicine, University of Geneva, 1 rue Michel-Servet, CH-1211 Geneva 4, Switzerland
[3]These authors contributed equally
[4]Lead contact
*Correspondence: raman.pandey@unige.ch (R.R.P.), ramesh.pillai@unige.ch (R.S.P.)
https://doi.org/10.1016/j.molcel.2022.02.034

SUMMARY

The functional consequence of N^6-methyladenosine (m^{6}A) RNA modification is mediated by "reader" proteins of the YTH family. YTH domain-containing 2 (YTHDC2) is essential for mammalian fertility, but its molecular function is poorly understood. Here, we identify U-rich motifs as binding sites of YTHDC2 on 3′ UTRs of mouse testicular RNA targets. Although its YTH domain is an m^{6}A-binder *in vitro*, the YTH point mutant mice are fertile. Significantly, the loss of its 3′→5′ RNA helicase activity causes mouse infertility, with the catalytic-dead mutation being dominant negative. Biochemical studies reveal that the weak helicase activity of YTHDC2 is enhanced by its interaction with the 5′→3′ exoribonuclease XRN1. Single-cell transcriptomics indicate that *Ythdc2* mutant mitotic germ cells transition into meiosis but accumulate a transcriptome with mixed mitotic/meiotic identity that fail to progress further into meiosis. Finally, our demonstration that *ythdc2* mutant zebrafish are infertile highlights its conserved role in animal germ cell development.

INTRODUCTION

RNAs are decorated with several chemical modifications, with N^6-methyladenosine (m^{6}A) being abundant on mRNAs (Fu et al., 2014; Roignant and Soller, 2017; Schwartz, 2016). The RNA methyltransferase "writer" complex composed of the METTL3-METTL14 heterodimer (Liu et al., 2014; Śledź and Jinek, 2016; Wang et al., 2016b) installs the modification on thousands of transcripts within the sequence context DR**A**CH (D = A, G or U; R = G or A; H = A, C or U; **A** = m^{6}A) (Dominissini et al., 2012; Meyer et al., 2012; Schwartz et al., 2013). Another writer, METTL16, catalyzes the mark on extended motifs (UAC**A**GAGAA or UAC**A**GAAAC) that are in the single-stranded region of a secondary structure on select RNAs (Doxtader et al., 2018; Mendel et al., 2018; Mendel et al., 2021; Pendleton et al., 2017; Ruszkowska et al., 2018; Warda et al., 2017). METTL3 (Batista et al., 2014; Geula et al., 2015) and METTL16 (Mendel et al., 2018) are essential for mouse embryonic development.

Presence of the m^{6}A methylation can have two major consequences: they can either repel proteins that would otherwise normally bind the unmethylated sequence or enable specific binding by "reader" proteins that can recognize the mark (Edupuganti et al., 2017). As an example of the former situation, 3′ splice site m^{6}A is shown to repel binding of the splicing factor U2AF35, inhibiting RNA splicing (Mendel et al., 2021; Watabe et al., 2021; Yoshida et al., 2020). However, much of the biology of m^{6}A is believed to be mediated by "reader" proteins, with the conserved YTH family of proteins being the best studied (Patil et al., 2018; Stoilov et al., 2002; Zhang et al., 2010). The signature YTH domain of these family members specifically recognizes m^{6}A within an aromatic pocket to bind RNAs (Li et al., 2014; Luo and Tong, 2014; Theler et al., 2014; Xu et al., 2014; Zhu et al., 2014). Subsequently, they recruit additional factors to alter splicing, RNA stability or translation. There are five mammalian members of the YTH family, with nuclear YTHDC1 being involved in regulating splicing, RNA stability, and alternative polyadenylation site choice (Kasowitz et al., 2018; Xiao et al., 2016). The YTHDF1, YTHDF2, and YTHDF3 are cytosolic and have redundant roles in regulating RNA stability or translation (Lasman et al., 2020; Wang et al., 2014; Wang et al., 2015; Zaccara and Jaffrey, 2020). The YTHDF proteins recruit the CCR4-NOT deadenylation complex to destabilize RNAs (Du et al., 2016; Liu et al., 2020; Wang et al., 2014). Consistent with their important roles, loss of mouse YTHDC1 (Kasowitz et al., 2018) and YTHDF2 (Ivanova et al., 2017; Lasman et al., 2020) causes embryonic lethality.

YTHDC2 is the largest member of the YTH family, is cytosolic, and expressed exclusively in the animal germline. Loss of *Ythdc2* results in male and female infertility in mice (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018; Liu et al., 2021; Wojtas et al., 2017). Destabilization of RNA and modulation of translation are

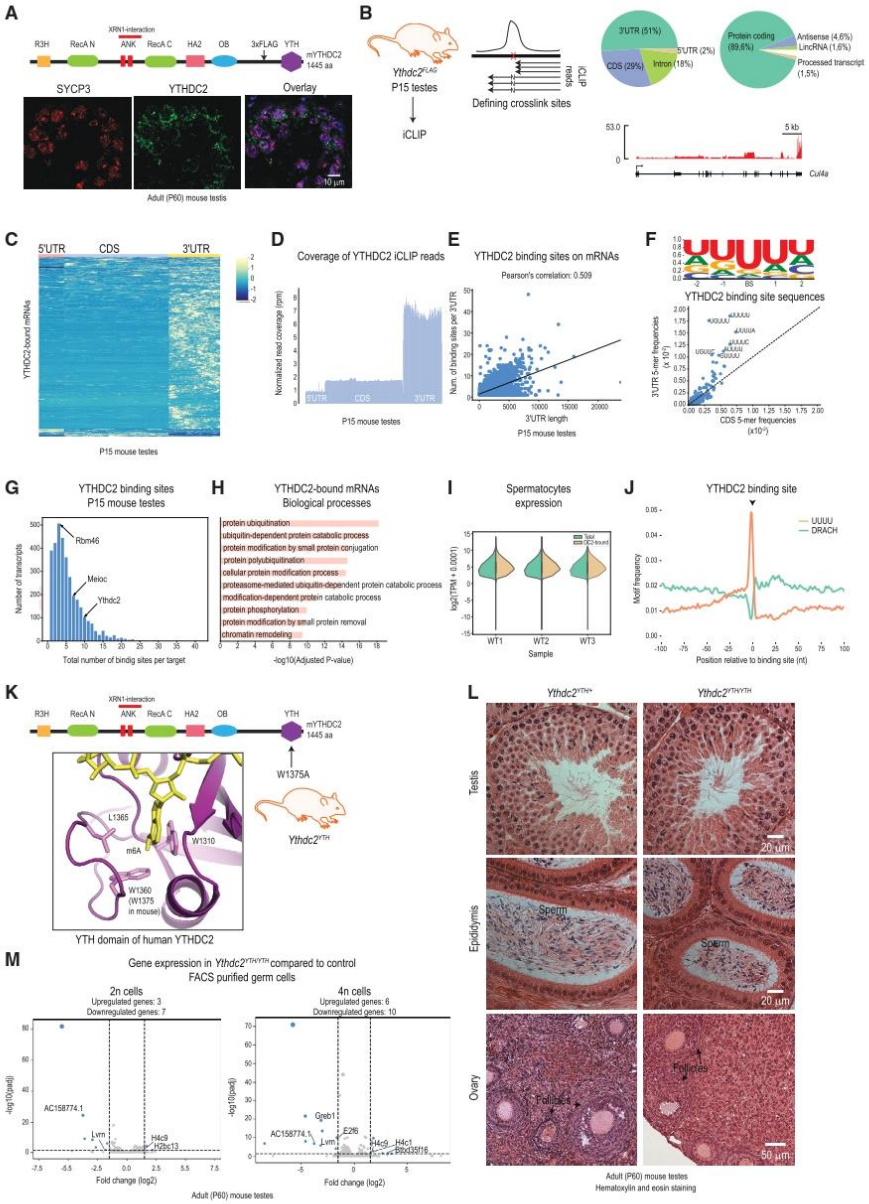


Figure 1. YTHDC2 binds to the 3′ UTR of protein-coding genes through a U-rich motif

(A) Cartoon showing domain organization of mouse YTHDC2 protein. The 3xFLAG tag sequence was inserted in the mouse *Ythdc2* gene locus between the OB-fold and the YTH domain. Immunofluorescence detection with anti-FLAG (green) and anti-SYCP3 (red) antibodies in adult (P60) testes sections of the *Ythdc2*^{FLAG} knockin mice. Scale bar in µm is indicated.

(legend continued on next page)

reported as consequences in the *Ythdc2* mutant germ cells, but precise mechanisms involved are unclear. YTHDC2 is a member of the DExH family of RNA helicases that is endowed with accessory RNA-binding domains like the R3H domain, oligo-nucleotide binding (OB) fold, and the YTH domain (Figure 1A). The YTH domain of YTHDC2 has the weakest RNA binding of all the YTH family members (Wojtas et al., 2017; Xu et al., 2015). For the same m^{6}A RNAs tested, binding affinities are at least 15- to 24-fold weaker for YTHDC2 when compared to YTHDC1 (Xu et al., 2015). This raises the question whether the specific m^{6}A-reading capacity of its YTH domain is relevant *in vivo*, especially given the presence of other RNA-binding domains. YTHDC2 possesses a 3′→5′ RNA helicase activity (Jain et al., 2018; Wojtas et al., 2017), but its physiological relevance is not known, since the fly ortholog Bgcn, which is also important for fertility, lacks residues consistent with such an activity (Bailey et al., 2017; Jain et al., 2018; Li et al., 2009; Ohlstein et al., 2000). Immunoprecipitated YTHDC2 complexes contain the 5′→3′ exoribonuclease XRN1 (Kretschmer et al., 2018; Wojtas et al., 2017) and a largely disordered protein called MEIOC (Abby et al., 2016; Soh et al., 2017; Wojtas et al., 2017). Whether these associations are direct, and how they impact the function of YTHDC2 is not known. Finally, it was shown that *Ythdc2* mutant germ cells transition from mitotic spermatogonia to enter meiosis, but do not stay on course to complete meiosis (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018; Wojtas et al., 2017). It is unclear why germ cells fail to make a successful transition into meiosis in the absence of *Ythdc2*.

Here, we identify binding sites for YTHDC2 as U-rich motifs in the 3′ UTR of testicular RNAs and reveal that its m^{6}A-binding property is dispensable *in vivo* in mice. However, its RNA helicase activity is essential for mammalian fertility. Biochemical studies reveal that its helicase activity is restrained by structural inserts placed within the RNA helicase domain, but this design allows positive regulation by its interaction partner XRN1. Using single-cell transcriptomics, we identify a role for YTHDC2 in ensuring proper separation of transcriptomes when the germ cells transition from mitosis to meiosis, because mutant cells display a mixed transcriptome identity that is not conducive for proper meiotic progression. Finally, using a zebrafish knockout model, we demonstrate a conserved role for YTHDC2 in ensuring animal fertility.

RESULTS

YTHDC2 binds RNA targets with U-rich motifs

To understand the role of YTHDC2 in mouse testes, we wished to identify its RNA targets. Our previous attempts with a rabbit polyclonal antibody to purify YTHDC2-bound RNAs from mouse testes were unsuccessful. Therefore, we created a knockin mouse line expressing the 3x-FLAG-tagged fusion protein (STAR Methods). We inserted the tag at an internal location in a surface-accessible loop (defined by limited proteolysis of the recombinant YTHDC2 protein) between the OB fold and the YTH domain (Figures 1A and S1A). Such homozygous knockin *Ythdc2*^{FLAG} mice are viable and fertile.

The maximal expression of YTHDC2-FLAG protein was observed in the pachytene stage spermatocytes labeled with the meiotic marker SYCP3 (Figure 1A). To identify RNA targets of YHTDC2, we used testes from *Ythdc2*^{FLAG} animals at postnatal day 15 (P15), the developmental time-point when pachytene spermatocytes dominate the cell types present in the seminiferous tubules. Testicular cell suspensions were exposed to UV light, and the crosslinked YTHDC2-RNA complexes were isolated by anti-FLAG affinity beads (STAR Methods). After deep-sequencing of associated RNAs using the iCLIP protocol (Huppertz et al., 2014), the reads were mapped to the mouse genome (Figure S1B).

We defined the crosslink sites based on previously established criteria (Krakau et al., 2017). Briefly, a pile up of iCLIP

(B) Testes from *Ythdc2*^{FLAG} mice (P15) were used to perform iCLIP experiment to identify the YTHDC2-bound RNAs. Binding sites were identified as loci having a pile-up of reads with 3′ ends terminating at the potential crosslinking site or by reads crossing it but having an increased rate of indels. Pie charts showing proportion of YTHDC2 binding sites in different RNA biotypes, and the location of the YTHDC2 binding sites over different RNA features. The coverage of iCLIP reads over a representative target is shown. Notice the enrichment on the 3′ UTR.

(C) Heatmap showing z-scores of the iCLIP read coverage per binned YTHDC2-bound transcripts. See also Table S2.

(D) Metaplot showing an average iCLIP read coverage over YTHDC2 target transcripts.

(E) Scatterplot and the correlation analysis of the number of YTHDC2 binding sites against 3′ UTR length.

(F) Sequence logo showing a U-rich motif centered on the YTHDC2 binding site (BS) in a 5-mer window. Scatterplot comparing the 3′ UTRs and CDS frequency of all the pentamers centered on the YTHDC2 crosslink site (mostly a uridine). Top seven 5-mers found at the crosslink site are highlighted. See also Figure S1G.

(G) Histogram showing a distribution of the total number of binding sites per transcript.

(H) Gene ontology analysis of YTHDC2 targets. Only terms with the adjusted p value < 0.5 were plotted.

(I) Violin plots showing the distribution of expression counts (TPM + 0.0001) of all genes (green) in spermatocytes from P15 wild-type mouse testes, next to the distribution of expression counts (TPM + 0.0001) of YTHDC2 targets (orange). Three biological replicates are shown. Although transcripts bound by YTHDC2 have a broad expression profile, they show a slightly higher expression.

(J) Distribution of putative m^{6}A methylation sites (DRACH) and a U-rich sequence (UUUU) in a 200-nt window flanking the YTHDC2 binding site (position 0). See also Figures S1E and S1F.

(K) A modeled m^{6}A nucleotide in the hydrophobic pocket of the YTH domain of human YTHDC2 from (Wojtas et al., 2017). Key residues required for m^{6}A recognition are indicated. Mutation of W1360 is shown to abolish m^{6}A binding. The corresponding W1375 in the *Ythdc2*^{YTH} was mutated to assess physiological relevance of the YTH domain *in vivo*. See also Figure S1H.

(L) Eosin and hematoxylin-stained adult testis and ovary sections from heterozygous and homozygous *Ythdc2*^{YTH} knockin mutant mice. No difference was observed. Scale bar in μm is indicated.

(M) Ribo-depleted RNA-seq (in biological triplicates) analysis of FACS-purified 2n (spermatogonia) and 4n (spermatocytes) cells from mouse testes of heterozygous and homozygous *Ythdc2*^{YTH} knockin mutant mice. Volcano plots show the differential gene expression between the genotypes. Genes showing significant differential expression (padj < 0.05 and log2FC > 1.5 or log2FC < −1.5) are marked in blue. Significantly dysregulated pseudogenes are masked.

reads at a particular locus is considered to be a strong indicator of a binding site. The nucleotide that is crosslinked to the protein acts as a roadblock during reverse transcription, leading to termination of reads. In those cases where the reverse transcriptase traverses the crosslinked nucleotide, it introduces indels at the site (Figure 1B). These features were used to identify over ~34K YTHDC2 binding sites on thousands of testicular transcripts (Table S2). Most (90%) of these transcripts are protein-coding mRNAs, with majority (51%) of the binding sites being in the 3′ UTR of the RNAs (Figures 1B and S1D). When mapped to specific transcripts, a clear enrichment of the iCLIP reads on the 3′ UTR is visible (Figure 1B). A similar pattern emerges when reads are mapped to all YTHDC2-bound mRNAs (Figure 1C). A meta-plot of reads mapping to all the YTHDC2-bound protein-coding transcripts confirms this general trend for the protein to be mostly bound to the 3′ UTR, followed by some enrichment in the coding domain sequence (CDS) (Figure 1D). There is no strong correlation between the 3′ UTR length and the number of binding sites (Figure 1E). This enrichment is consistent with the known 3′→5′ RNA helicase activity of YTHDC2 (Jain et al., 2018; Wojtas et al., 2017), requiring a 3′ end loading on the RNA substrate.

Examination of the sequences at the crosslink site shows that a uridine is the preferred nucleotide, while the flanking nucleotides are also mostly uridines. Thus, as defined by this UV crosslinking approach, a U-rich motif defines the binding site of YTHDC2 (Figure 1F). We collected all 5-nt motifs centered on the crosslink site and examined their frequency of occurrence in the 3′ UTR and the coding domain sequence (CDS) (Figures 1F, S1E, and S1F). This analysis reveals two aspects of the YTHDC2 binding sites: first, several U-rich motifs (UUUUU, UGUUU, UGUUC, or AUUUU, etc.) dominate the crosslink site, with the allowance of additional nucleotides within it; second, these motifs are mostly present in the 3′ UTRs. Electrophoretic mobility shift assays (EMSA) with a radiolabelled RNA and recombinant human YTHDC2 confirms this preference for U-rich sequence to be directly determined by the protein itself (Figure S1G). Taken together, this suggests that YTHDC2 is loaded at the 3′ end of mRNAs using U-rich motifs as landing sites.

Transcripts from over 5000 genes are identified as bound by YTHDC2 in the P15 mouse testes (Table S2), with most transcripts having anywhere from 1 to 5 binding sites (Figure 1G). Interestingly, *Ythdc2* and its interaction partner *Meioc* are represented in this list (Figure 1G). A Gene Ontology (GO) analysis reveals a predominant presence of mRNAs associated with the ubiquitination pathway (Figure 1H), with ~500 transcripts belonging to this category (Table S2). The presence in the iCLIP dataset is not due to any aberrant high expression of these genes, as we find the YTHDC2-bound transcripts to have a broad expression pattern in the germ cells (Figure 1I). We propose that collective regulation of the few thousand transcripts that are bound by YTHDC2 in P15 testes is important for germ cell development.

m^{6}A-binding activity of YTHDC2 is dispensable for its function *in vivo*

YTHDC2 belongs to the YTH family of m^{6}A readers (Patil et al., 2018). The isolated YTH domain of YTHDC2 binds m^{6}A RNAs *in vitro*, albeit with binding affinities (K_d = 6-24 μM for YTHDC2) (Wojtas et al., 2017; Xu et al., 2015) that are at least 15-24 fold weaker than that for other YTH family members (K_d = 0.4-1 μM for YTHDC1) (Xu et al., 2015). This raises the question whether YTHDC2, which has several RNA-binding modules, depends on the YTH domain for its *in vivo* functions.

We first asked whether the U-rich motifs (YTHDC2 crosslink sites) identified on the testicular transcripts are near potential m^{6}A methylation sites (as defined by DRACH motifs). When examined in a 200-nt window flanking the crosslink site, we do not find any preference for having the DRACH motifs (Figure 1J). To directly test the *in vivo* relevance of the YTH domain, we edited the mouse genome to introduce a point mutation that was previously shown to abolish m^{6}A-specific RNA binding *in vitro* (Wojtas et al., 2017). This converts an aromatic tryptophan (W1375A) in the hydrophobic pocket of the YTH domain, to an alanine (Figures 1K and S1H). Heterozygous and homozygous *Ythdc2*^{YTH} knockin mice with the mutant YTH domain were found to be viable and fertile. These animals were maintained in our facility for over 2 years without any long-term defects in fertility. Histological examination of testes and ovaries from the homozygous *Ythdc2*^{YTH} knockin mice did not show any differences (Figure 1L). Mature sperm was identified in the epididymis of the homozygous mutant, while ovarian follicles at different stages of development were also observed (Figure 1L). Testicular transcriptome analysis of sorted 2n and 4n germ cells from adult (P60) mice also did not reveal any dramatic differences between the control heterozygous and homozygous *Ythdc2*^{YTH} knockin animals (Figure 1M). Taken together, we find that the ability to read m^{6}A modification is not essential for the *in vivo* roles of YTHDC2 in the mouse germline.

YTHDC2 directly associates with XRN1 and MEIOC but likely in separate complexes *in vivo*

Next, we focused on the YTHDC2-associated factors previously identified by immunoprecipitation experiments. This includes the 5′→3′ exoribonuclease XRN1 (Kretschmer et al., 2018; Wojtas et al., 2017) and a largely disordered protein with a coiled-coil domain called MEIOC (Figure 2A) (Abby et al., 2016; Soh et al., 2017). To determine if these associations are due to direct interactions, we expressed recombinant versions of the human proteins. YTHDC2 was produced as a full-length recombinant protein in the insect cell expression system, while the two others were produced as truncations: XRN1 (1-1279 aa) with a large deletion in the C-terminal disordered region, and the isolated conserved C-terminal coiled-coil (CC) domain of MEIOC (STAR Methods).

When co-expressed together, YTHDC2 and XRN1 (1-1279 aa) co-purify, and elute as a complex during size-exclusion chromatography (Figures 2A and S2A). Likewise, co-expressed YTHDC2 and the MEIOC-CC domain also co-purify together (Figure 2B). The two ankyrin (ANK) repeats of YTHDC2 are essential for interaction with XRN1 (Kretschmer et al., 2018; Wojtas et al., 2017). Deleting these repeats in the YTHDC2$^{\Delta ANK}$ protein did not affect its co-purification with MEIOC-CC domain (Figure 2B), indicating that YTHDC2 associates with its two partner proteins via distinct domains. Taken together, we demonstrate direct interactions between YTHDC2 and XRN1 (1-1279

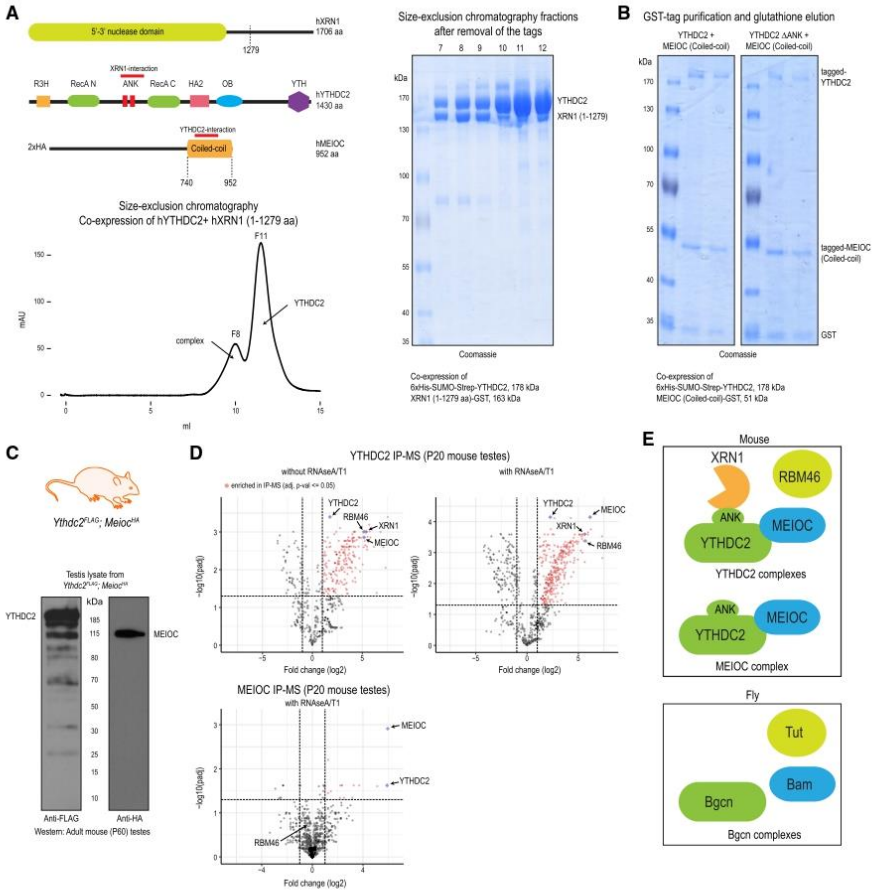


Figure 2. YTHDC2 directly interacts with XRN1 and MEIOC, and associates with RBM46

(A) Cartoon showing domain organization of human YTHDC2, XRN1 and MEIOC. The interaction regions in the proteins are marked. Size-exclusion chromatography protein elution trace showing the presence of YTHDC2-XRN1(1-1279 aa) complex and free YTHDC2 protein. Indicated fractions were analyzed by SDS-PAGE to reveal presence of YTHDC2- XRN1(1-1279 aa) complex (fractions 7-9) and free YTHDC2 (fractions 10–12). See also Figure S2A.

(B) Co-expression YTHDC2 and MEIOC-coiled-coil (CC) domain in insect cells as tagged proteins, followed by purification over glutathione-Sepharose beads to isolate MEIOC-CC. SDS-PAGE analysis shows the co-purification of YTHDC2 with MEIOC-CC (left panel). Deletion of the ANK repeats of YTHDC2 does not affect this association (right panel).

(C) Western blot of testicular lysate showing expression of YTHDC2-3xFLAG and 2xHA-MEIOC from double-knockin *Ythdc2*^{FLAG};*Meioc*^{HA} mice.

(D) Purification of protein complexes using anti-FLAG or anti-HA affinity beads. Complexes were treated with RNases (RNase A/T1), when indicated. Mass spectrometry identification of complex components is shown. The volcano plot shows the enrichment of proteins in purifications from knockin mouse testes lysates compared to that from wildtype mouse testes. See also Table S3.

(E) Cartoon showing the different protein associations of mouse YTHDC2 as established in this study in mouse testes. Complex components previously reported to be associating with YTHDC2 fly ortholog Bgcn is also shown.

aa), and between YTHDC2 and the coiled-coil (CC) domain of MEIOC.

To determine if the three proteins exist together *in vivo*, we generated *Meioc*^{HA} knockin mice expressing the HA-MEIOC tagged protein (Figure S2B) and crossed it with the *Ythdc2*^{FLAG} knockin mice. Such double knockin *Ythdc2*^{FLAG};*Meioc*^{HA} mice homozygous for both tagged alleles (Figure 2C) are viable and fertile, confirming that the tagged proteins were fully functional. We isolated endogenous complexes from P20 mouse testes using anti-FLAG or anti-HA affinity beads and subjected them to

mass spectrometry. YTHDC2-FLAG complexes enrich for both XRN1 and MEIOC (Figure 2D). Consistent with their direct interactions with YTHDC2 (Figures 2A and 2B), their presence in the endogenous complexes is resistant to treatment with RNases (Figure 2D). Interestingly, while HA-MEIOC complexes contained YTHDC2, we did not find XRN1 (Figures 2D, S2C, and S2D). This leads us to conclude that YTHDC2 exists in separate complexes with XRN1 and MEIOC *in vivo* (Figure 2E). While there are several other factors present, we report the presence of RNA binding motif protein 46 (RBM46) in the YTHDC2 purifications. Pointing to a role for RBM46 in animal fertility, zebrafish *rbm46* mutants reveal arrested meiosis and are infertile (Dai et al., 2021). Furthermore, the fly ortholog of RBM46 is the RNA recognition motif containing protein called Tumorous testis (Tut), also shown to be associated with Bgcn (Chen et al., 2014), the fly ortholog of YTHDC2. Taken together, we conclude that YTHDC2 forms specific sub-complexes *in vivo* in mouse testes and identify RBM46 as an additional component of the mouse YTHDC2 complex.

Ythdc2 RNA helicase catalytic-dead mutation is dominant-negative and blocks spermatogenesis

Having identified the RNA targets and protein complexes of YTHDC2, we next examined the relevance of its RNA helicase activity. RNA helicases utilize energy from ATP hydrolysis for unwinding RNA or displacing proteins or for mediating conformational changes of RNA-protein complexes (Bleichert and Baserga, 2007; Linder and Jankowsky, 2011). We created a mouse knockin mutant that has a point mutation in the ATP hydrolysis motif DEVH (DEVH → DQVH; E332Q) (Figure S3A). This mutation was shown to abolish the *in vitro* ATP hydrolysis and RNA unwinding activity of human YTHDC2 (Wojtas et al., 2017). Single-mouse embryo injections with the CRISPR construct (STAR Methods) resulted in viable heterozygous catalytic-dead (*Ythdc2*$^{+/cat\text{-}dead}$) founder mice of both sexes (Figure S3A). When crossed with wildtype mating partners, such heterozygous *Ythdc2*$^{+/cat\text{-}dead}$ catalytic-dead mutant mice did not produce any litters, indicating that the E332Q mutation is dominant negative. Some homozygous *Ythdc2*$^{cat\text{-}dead/cat\text{-}dead}$ founder mice were obtained and were also found to be infertile.

To examine the defect in spermatogenesis, we sequenced the testis transcriptomes from P14 catalytic-dead mutant and control wildtype animals. Analysis indicates that most of the gene expression changes are downregulation of transcripts in the catalytic-dead mutant (Figures 3A and S3B). Comparison with transcriptomes of purified germ cell populations present in wildtype mouse testes indicates that most of the downregulated genes are those normally highly expressed in meiotic spermatocytes (Figure 3B). This points to arrested spermatogenesis in the catalytic-dead mutant testes, with germ cells not making a proper progression through meiosis. Thus, the RNA helicase activity of YTHDC2 is essential for meiosis in the mouse male germline.

Ankyrin repeats constrain helicase activity of YTHDC2 but interaction with XRN1 enhances it

YTHDC2 is a member of the DExH family of processive RNA helicases that show RNA-dependent ATPase activity (Jankowsky and Fairman, 2007; Pyle, 2008). This drives its 3′ →5′ RNA unwinding activity (Jain et al., 2018; Wojtas et al., 2017). The helicase domain of YTHDC2 is unusual in being interrupted by two ankyrin (ANK) repeats. The predicted (AlphaFold) (Jumper et al., 2021) core helicase domain of human YTHDC2 reveals how the repeats are hitched to the central β sheet of the C-terminal RecA module (Figure 3C). ANK repeats are protein-protein interaction domains (Mosavi et al., 2004), which in the case of YTHDC2 are required for interaction with XRN1 (Kretschmer et al., 2018; Wojtas et al., 2017).

To examine the relevance of the ANK repeats, we purified the YTHDC2$^{\Delta ANK}$ protein lacking the repeats (Figure 3D) and used it in a fluorescence-based RNA unwinding assay to compare its activity to that of the wildtype protein. The assay uses an RNA duplex with a single-stranded 3′ overhang as the substrate (Figure 3E), where a longer single-stranded RNA is annealed to two shorter complementary single-stranded RNAs: one fluorescently labeled with Cy3 and another with the quencher BHQ. Unwinding of the duplexes by YTHDC2 will separate the Cy3 away from the quencher, generating the fluorescence signal. When the reactions were primed by addition of ATP, we observed a low level of unwinding activity with the full-length YTHDC2 (Figures 3F and S3C). In comparison, the YTHDC2$^{\Delta ANK}$ protein presented strikingly robust activity. This suggests that presence of the two ANK repeats between the RecA modules constrains the RNA unwinding activity of YTHDC2 (Figure 3C).

Interaction partners are shown to modulate the ATPase/unwinding activities of RNA helicases (Bleichert and Baserga, 2007; Bono et al., 2006; Mathys et al., 2014). Therefore, we asked whether XRN1 or MEIOC can modulate the helicase activity of YTHDC2. We prepared a catalytic-dead (E178Q) version of XRN1 (1-1279) and used it in RNA unwinding reactions. Strikingly, in comparison to the YTHDC2 alone, presence of XRN1^{E178Q}(1-1279) enhanced the activity of YTHDC2 (Figures 3G, S3D, and S3E), and this was a dose-dependent effect (Figure S3F). As expected, XRN1 alone did not show any unwinding activity (Figures S3E and S3F). Consistent with the requirement of the ANK repeats for interaction, addition of XRN1^{E178Q}(1-1279) to the YTHDC2$^{\Delta ANK}$ protein did not alter its helicase activity (Figure 3H and S3H). This suggests that interaction of XRN1 to the ANK repeats may alter the conformation of the RecA domains (Figure 3C), such that they take up more readily a closed state that favors ATP hydrolysis and RNA unwinding (Linder and Jankowsky, 2011). Indeed, ATP hydrolysis assays showed a ~2-fold increase in activity of YTHDC2 in the presence of XRN1 (Figures 3I and S3I). The enhancement of the helicase activity by XRN1 is specific, as addition of the MEIOC-CC domain, whose interaction with YTHDC2 is not via the ANK repeats (Figure 2B), did not alter the RNA unwinding activity (Figures 3G, S3D, and S3G). We conclude that YTHDC2 has a weak helicase activity per se, but its structural design allows enhancement of this activity by its interaction partner XRN1.

Single-cell transcriptomics identifies a mixed transcriptome in the *Ythdc2* mutant germ cells

To examine the gene regulation consequences due to loss of *Ythdc2*, we conducted single-cell transcriptome analysis. Histological analyses have shown that mitotic spermatogonia (SG) in the *Ythdc2* KO testes differentiate by initiating the meiotic transcription program (Bailey et al., 2017; Hsu et al., 2017; Jain

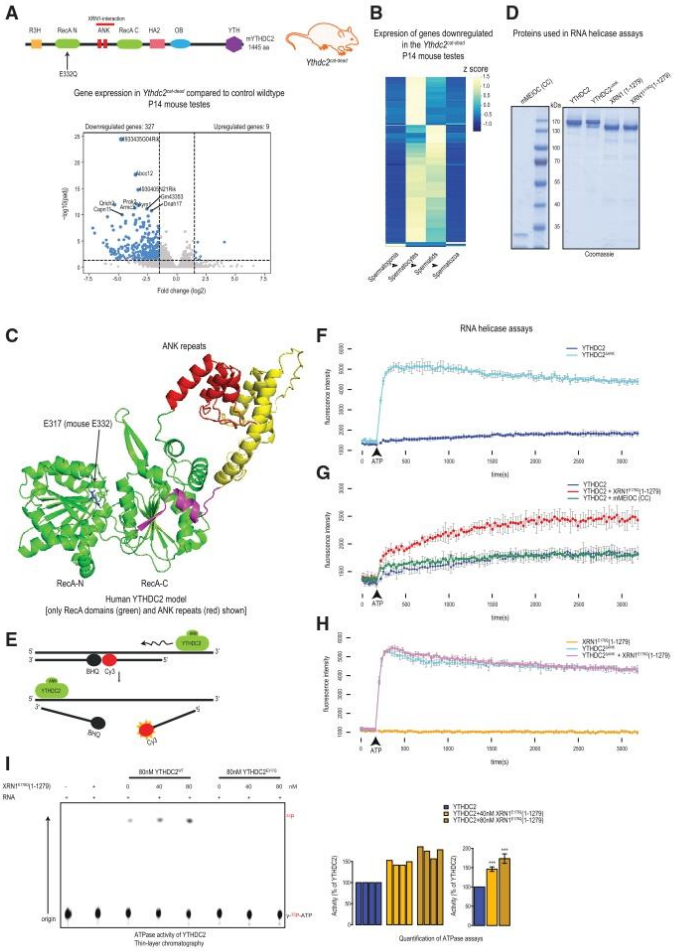


Figure 3. The essential 3′→5′ RNA helicase activity of YTHDC2 is modulated by XRN1

(A) Creation of the RNA helicase catalytic-dead *Ythdc2*$^{cat\text{-}dead}$ mouse mutant with the E332Q mutation in the N-terminal RecA domain. See also Figure S3A. Volcano plot showing differential testicular gene expression analysis between wildtype and *Ythdc2*$^{cat\text{-}dead}$ helicase mutant P14 mice. Significant genes (padj < 0.05 and log2FC > 1.5 or log2FC < −1.5) are shown in blue.

(B) Heatmap showing the expression pattern of significantly dysregulated genes (from the *Ythdc2*$^{cat\text{-}dead}$ mutant) in wildtype spermatogenic cell types. Z-scores from TPM counts are plotted. The genes downregulated in the mutant are those normally highly expressed in the spermatocytes and spermatids.

(C) Predicted (AlphaFold) structure of human YTHDC2, showing only the two RecA domains and the intervening two ankyrin (ANK) repeats. The helicase catalytic site glutamic acid (E) mutated in the *Ythdc2*$^{cat\text{-}dead}$ mouse mutant is indicated.

(D) Coomassie-stained gel (SDS-PAGE) of the recombinant human or mouse (m) proteins used for RNA unwinding assays.

(E) Schematic showing fluorescence-based RNA helicase assay. The Cy3 fluorescence is quenched in the duplex state. See also Table S1.

(F–I) (F) RNA helicase activity of full-length YTHDC2 and the ΔANK mutant. Deletion of ANK repeats enhances the helicase activity. Error bars refer to SD (G) RNA helicase activity of YTHDC2 in the presence of XRN1^{E178Q} (1-1279) or mouse MEIOC-CC domain. The addition of XRN1^{E178Q} (1-1279) enhances YTHDC2 helicase activity (red line). Error bars refer to SD (H) RNA helicase activity of the YTHDC2 ΔANK mutant is not affected by addition of XRN1^{E178Q} (1-1279). See also Figures S3E and S3F. Error bars refer to SD (I) thin-layer chromatography (TLC) analysis of ATP hydrolysis assays with full-length YTHDC2 or its catalytic-dead version. XRN1^{E178Q} (1-1279) was added in some of the reactions at the indicated concentrations. RNA is needed to stimulate the activity of YTHDC2. See also Figure S3I. Bar plots show the quantified YTHDC2 ATPase activity from four independent experiments. Addition of XRN1^{E178Q} (1-1279) enhances the ATPase activity of YTHDC2. Error bars refer to SD. The one tailed t test was used to test whether activity is higher than 100% (***p value < 0.001).

et al., 2018; Wojtas et al., 2017), but the mutant meiotic cells do not maintain this trajectory and undergo apoptosis. To examine more closely the cellular gene expression during this transition phase, we purified testicular germ cells from P10 animals that have replicated their DNA (4n) prior to entry into meiosis. Single-cell RNaseq libraries were prepared using the 10X Genomics technology (Figure 4A). We constructed single-cell maps based on the UMAP algorithm using expression data for 3000 highly variable genes (Figure 4B). Identity of the cell populations was established based on previous single-cell RNaseq of testicular cell types (Ernst et al., 2019).

The heterozygous *Ythdc2*$^{+/-}$ testes from P10 animals contained cells expected at this stage of spermatogenesis: undifferentiated spermatogonia (SG), differentiating SG, and meiotic leptotene/zygotene spermatocytes (L/Z SCs). While the knockout *Ythdc2*$^{-/-}$ testes had a similar set of cell types, it lacked the meiotic cells. Instead, two additional cell populations were present in the knockout testes (Figure 4B). Based on previous gene expression and histological analyses using the meiotic marker protein SYCP3 (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018; Wojtas et al., 2017), *Ythdc2* mutant germ cells do make it into meiosis. Marker analyses reveal that the KO-specific cell populations have a mixed identity by having the transcriptomes present in both the differentiating SG and the meiotic cells (Figure 4C). They express several meiotic markers, including SYCP3 (Figure 4C), and markers normally expressed only in

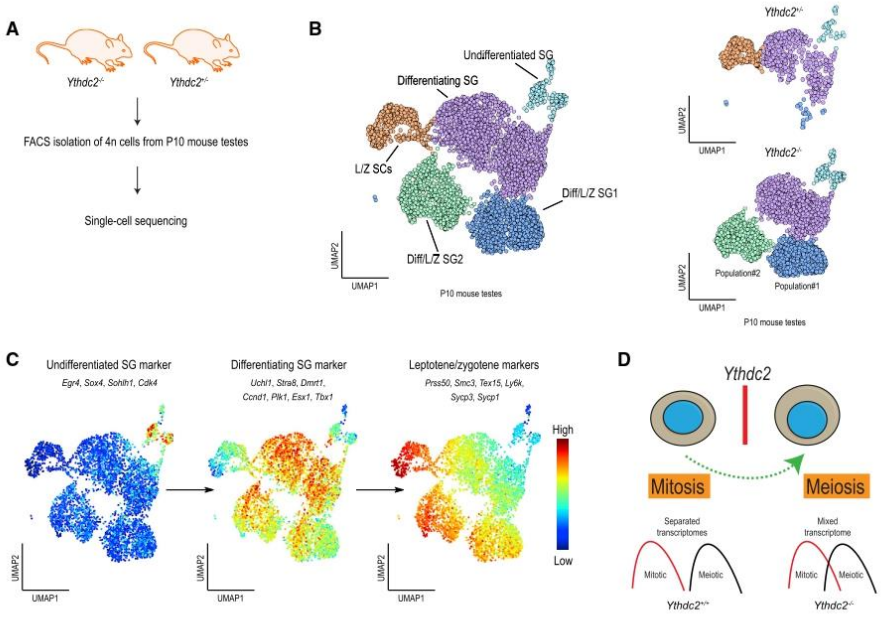


Figure 4. Loss of mouse *Ythdc2* results in germ cells with a mixed mitotic/meiotic identity

(A) Single-cell purification of 4n germ cells from mouse testes of *Ythdc2*$^{+/-}$ and *Ythdc2*$^{-/-}$ (KO) P10 animals.

(B) UMAP-based integrated single-cell map of *Ythdc2*$^{+/-}$ and *Ythdc2*$^{-/-}$ (KO) testicular 4n germ cells (left). Annotated cell types are assigned a specific color. Genotype-separated single-cell maps are shown (right). Notice the presence of two unique cell populations present only in the KO testes.

(C) Analysis of the indicated mitotic and meiotic marker genes expression. UMAP-based single-cell maps are ordered based on the developmental trajectory: mitotic spermatogonia (SG) marker expression (left), differentiating SG cell marker expression (middle), and leptotene/zygotene meiotic markers expression (right). Note that differentiating SG and meiotic markers are simultaneously expressed in the two populations of cells found exclusively in the *Ythdc2* KO. See also Figures S4B and S4C.

(D) Model showing how YTHDC2 acts during the mitotic-meiotic transition. In the absence of *Ythdc2*, germ cells display a mixed identity with expression of markers present in both differentiating SG and meiotic cells.

differentiating SG (Figure 4C). Interestingly, immunofluorescence analysis of *Meioc* KO germ cells reveals a simultaneous expression of the meiotic marker SYCP3 and the mitotic marker cyclin A2 (CCNA2) (Soh et al., 2017). Similarly, we find the expression of *Ccna2* in the two mixed-identity cell populations in the *Ythdc2* KO testes (Figure S4C). Taken together, our data show that YTHDC2 is required for a fine separation of the transcriptomes when germ cells make the transition from mitosis to meiosis (Figure 4D). In the absence of YTHDC2, cells transition into meiosis but still express mitotic genes, and such a mixed transcriptome fails to support further progression into meiosis.

ythdc2 has a conserved role in ensuring fertility in the zebrafish germline

To explore the conserved role of YTHDC2, we generated *ythdc2* mutant zebrafish (STAR Methods). The mutant allele has a 46-base-pair deletion (*ythdc2*$^{\Delta 46}$) in the region encoding the N-terminal RecA module of the RNA helicase domain (Figure S5A). Heterozygous (*ythdc2*$^{\Delta 46/+}$) adult fish are viable and fertile. When in-crossed, the heterozygous (*ythdc2*$^{\Delta 46/+}$) animals give rise to embryos (examined at 1 and 5 days post-fertilization; dpf) in the expected Mendelian ratios (Figure 5A). Interestingly, all the homozygous adult *ythdc2*$^{\Delta 46/\Delta 46}$ individuals we identified are phenotypically males, and none were phenotypically female. This distinction is made on the basis of a flatter body plan and reddish color of males compared to females. We note that masculinization of the homozygous individuals is also reported for the *rbm46* zebrafish mutants (Dai et al., 2021). The homozygous *ythdc2*$^{\Delta 46/\Delta 46}$ males are infertile, as crosses with wildtype females failed to produce any viable, fertilized embryos (Figure 5B). To study the spermatogenic defect, we examined the gonads in 6-month-old wild-type, heterozygous, and homozygous mutant adult males using standard histological staining. The wildtype and heterozygous gonads revealed multiple seminiferous tubules full of spermatozoa, surrounded by spermatogonia and spermatocytes at different stages and associated cell types, as expected. Somatic Leydig and Sertoli cells are hardly visible (Figures 5C, S5B, and S5C). Strikingly, the homozygous *ythdc2*$^{\Delta 46/\Delta 46}$ gonads are hollowed out with empty tubules lacking any trace of meiotic germ cells. An interstitium

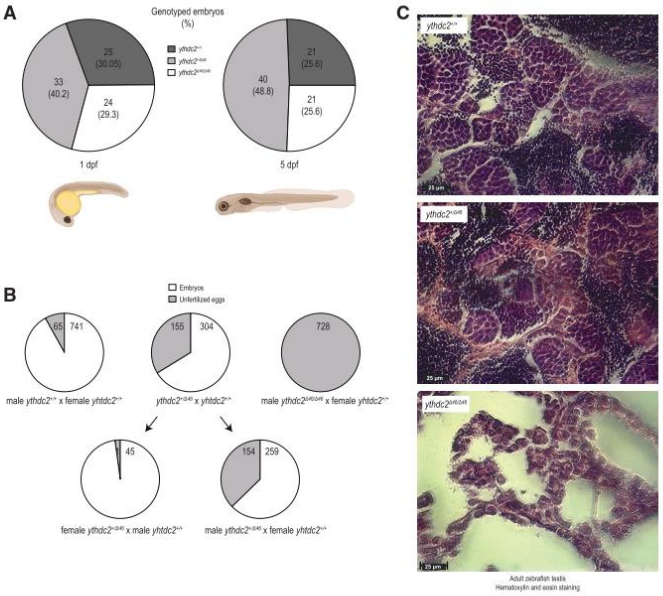


Figure 5. Conserved role of zebrafish *ythdc2* in male fertility

(A) Generation of *ythdc2* knockout (Δ46) zebrafish. Heterozygous crosses give rise to animals in the expected Mendelian ratios.

(B) Crosses of heterozygous and homozygous *ythdc2*$^{-\Delta 46/-\Delta 46}$ knockout males or females with wild-type partners. Crosses with homozygous knockout males do not produce any embryos, indicating that they are infertile. See also Figure S5B.

(C) Hematoxylin and Eosin staining of adult zebrafish testes showing that the infertile homozygous *ythdc2*$^{-\Delta 46/-\Delta 46}$ fish lack sperm. Scale bar in μm is indicated.

remains, but there are no cells corresponding to the germline. Taken together, this analysis confirms a conserved role for YTHDC2 in germline development from insects (Chen et al., 2014; Li et al., 2009) to zebrafish (Figure 5) and mice (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018; Liu et al., 2021; Wojtas et al., 2017).

DISCUSSION

We (Wojtas et al., 2017) and others (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018) previously demonstrated that YTHDC2 regulates m^{6}A transcripts in the mammalian germline. This hinged on the following observations. First, the isolated YTH domain of YTHDC2 has m^{6}A-reading activity *in vitro* and has a structural similarity to the m^{6}A-binding pocket of other YTH proteins (Hsu et al., 2017; Jain et al., 2018; Wojtas et al., 2017; Xu et al., 2015), although the binding is weaker compared to the domain found in other YTH proteins (Xu et al., 2015). Second, m^{6}A-enriched transcripts are enriched in YTHDC2 complexes (Bailey et al., 2017; Hsu et al., 2017). Third, the levels of these YTHDC2-bound RNA targets are increased in the absence of *Ythdc2* (Hsu et al., 2017) or levels of m^{6}A-enriched RNAs are increased in the *Ythdc2* mutant (Wojtas et al., 2017). Together, these suggested a role for YTHDC2 in post-transcriptional regulation by regulating RNA abundance or translation (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018; Wojtas et al., 2017).

In this study, we examined the *in vivo* relevance of its m^{6}A-reading activity. This question is especially relevant for a protein like YTHDC2, as it is decorated with multiple RNA-binding domains (R3H domain, RNA helicase module, and the OB fold). We created mice with a single-amino-acid mutation in the YTH domain that is shown to abolish the m^{6}A-binding *in vitro* (Wojtas et al., 2017). Such mutant mice did not reveal any impact on fertility (Figure 1L). Therefore, we conclude that despite the isolated YTH domain having an affinity for m^{6}A RNAs *in vitro* (Hsu et al., 2017; Wojtas et al., 2017; Xu et al., 2015), YTHDC2 may not depend on its m^{6}A-reading activity for RNA binding *in vivo* (Figure 1). Instead, our iCLIP experiments identify U-rich motifs in the 3′ end of the RNAs as the binding sites for YTHDC2 (Figure 1F), with no enrichment of m^{6}A methylation motifs in its vicinity (Figure 1J). The previously made correlations between m^{6}A methylation and YTHDC2-association and RNA regulation, needed a sound genetic test of *in vivo* relevance, which we now provide. In the light of this result, we propose that m^{6}A recognition via the YTH domain is not the dominant force for RNA-binding by YTHDC2 *in vivo*. Instead, multivalent interactions via the other RNA binding domains (R3H, helicase domain and OB fold) provide this RNA-recognition ability. We find that recombinant human YTHDC2 on its own prefers to bind an oligo-U RNA sequence over an oligo-A RNA (Figure S1G). Structural studies of the DExH RNA helicase MLE shows how base-specific interaction with the OB fold can specify the binding to U-rich sequences (Prabu et al., 2015). Furthermore, the R3H domain is a single-stranded nucleic acid binding domain (Grishin, 1998), which in the polyA-specific ribonuclease (PARN) contributes to RNA binding but does not specify the target (Wu et al., 2005). In conclusion, we propose that in identifying its RNA targets, YTHDC2 prioritizes sequence-specific RNA binding over recognition of RNA modifications. This is reminiscent of the extreme case of the fission yeast YTH protein Mmi1 (Harigaya et al., 2006), where the YTH domain lacks the ability to recognize m^{6}A (Wang et al., 2016a), but instead uses a distinct structural feature in the domain for high-affinity binding to hexanucleotide motifs present in its RNA targets (Hiriart et al., 2012; Yamashita et al., 2012).

Once on the 3′ UTR (Figures 1C and 1D), YTHDC2 depends on its RNA helicase activity for its *in vivo* functions in the mouse

germline (Figure 3). Precisely what it achieves is currently not known. It probably translocates along the RNA in a 3′→5′ direction to unwind RNA structures or evict bound protein factors. It is also possible that it remodels RNA-protein complexes bound at the 3′ end of the RNA. Given that the ATPase (Figure 3I) and helicase activities (Figures 3F–3H) of YTHDC2 are regulated by the 5′→3′ exoribonuclease XRN1, this interaction likely provides a mechanism to tie the enhancement of RNA helicase activity to consequences like RNA degradation. Interestingly, the C-terminal low-complexity region of XRN1 is shown to directly associate with decapping and deadenylation factors, bridging enzymatic activities that act at both ends of the mRNA to promote mRNA decay (Chang et al., 2019). While we did not precisely map the domain on XRN1 for association with YTHDC2, this C-terminal low-complexity region is a good candidate to mediate it (Figure 2A). Furthermore, the precise contributions of MEIOC and RBM46 to the function of YTHDC2 is not clear. MEIOC interacts with its conserved coiled-coil (CC) domain to a region of YTHDC2 that is distinct from that used by XRN1, and this does not modulate the RNA helicase activity (Figure 3G).

YTHDC2 likely regulates thousands of RNA targets in the mouse germline. It directly binds ~5,000 transcripts (Table S2) via U-rich motifs (Figure 1F) in the P15 mouse testes, where pachytene spermatocytes are the dominant cell type in the seminiferous tubules. Transcripts acting in the ubiquitination pathway form a large class of the YTHDC2 targets (Table S2 and Figure 1H). While it is known that the ubiquitination pathway plays an essential role in removal of specific transcription factors or histones at defined steps during mouse spermatogenesis (Nakagawa et al., 2017), how regulation of this pathway by YTHDC2 contributes to progression through meiosis remains to be understood. Pointing to its relevance beyond the early mitosis-meiosis transition stage, conditional deletion of *Ythdc2* in the pachytene spermatocytes causes abrogation of germ cell development, leading to infertility (Liu et al., 2021).

Examination of single-cell transcriptomes of germ cells in the P10 mouse testes shows that YTHDC2 plays a critical role in enabling a fine separation of the transcriptomes as mitotic differentiating spermatogonia enter meiosis (Figure 4). In the absence of *Ythdc2*, mutant germ cells initiate meiosis, but accumulate a mixed transcriptome with mitotic/meiotic identity (Figures 4B and 4D). Such mutant germ cells fail to progress further in meiosis and undergo apoptosis (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018; Wojtas et al., 2017). We favor a model where YTHDC2 regulates the decay of mitotic RNA targets during this mitosis-meiosis transition stage. We see an enrichment of U-rich motifs in the transcriptomes represented in the undifferentiated and differentiating spermatogonia, and in one of the mixed-identity *Ythdc2* KO-specific cell populations (Figure S4D). There is also an enrichment of ubiquitination pathway genes in these cell types when compared to the meiotic cells (Figure S4E). In contrast, the spermatocyte transcriptomes in the control P10 mouse testes are depleted of such U-rich motifs (Figure S4D). While the U-rich motifs may be one of the specificity elements in the YTHDC2 targets, other YTHDC2 complex components (like RBM46) may contribute to RNA target identification. YTHDC2 is now shown to be essential for germ cell development in the insect (Li et al., 2009), fish (Figure 5) and mouse (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018; Liu et al., 2021; Wojtas et al., 2017) systems. Beyond this conservation of the role in fertility, at the molecular level the protein seems to use distinct mechanisms. The insect ortholog Bgcn lacks the YTH domain and amino acid residues consistent with ATPase and helicase activities (Jain et al., 2018; Ohlstein et al., 2000), but forms a complex with Bam and Tut (Chen et al., 2014), orthologs of mouse MEIOC and RBM46. Understanding this catalytic and non-catalytic contributions of YTHDC2 proteins to gene regulation will be an exciting question for the future.

Limitations of study

Although we show that the YTH domain of YTHDC2 is dispensable in the mouse germline, maintenance of this m^6A-reading capacity might indicate that it may be used in other contexts that are not examined in this study. YTHDC2 facilitates translation of *Hif1a* mRNA in mouse adipocytes under conditions of increased m^6A levels on the RNA (when the demethylase FTO is deleted), promoting white-to-beige fat transition (Wu et al., 2021). Another study reported that YTHDC2 is recruited by the m^6A-modified internal ribosome entry sites (IRES) of the hepatitis C virus (HCV) to promote cap-independent translation (Kim and Siddiqui, 2021). Furthermore, YTHDC2 is shown to facilitate translation of mRNAs with coding-region m^6A (Mao et al., 2019). Therefore, ruling out any role for the m^6A-reading activity of YTHDC2, especially outside the germline, maybe premature. We identified RNAs encoded by ~5,000 genes as YTHDC2 targets (Table S2), but how their collective regulation contributes to a successful mitotic-meiotic transition remains to be identified. It is also not clear if further specificity elements beyond U-rich motifs exist to define such targets. Contributions of YTHDC2-associated factors like MEIOC and RBM46 to its function need to be elucidated. One challenge we faced is in obtaining full-length MEIOC protein, even when co-expressed together with YTHDC2. This may indicate that additional factors are missing in the complex nucleated by YTHDC2, and our mass spectrometry data (Table S3) will facilitate further studies in this direction. Finally, the precise molecular mechanism by which YTHDC2 regulates its RNA targets remains to be discovered.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animal work
  - Genotyping PCR for $Ythdc2^{FLAG}$ mouse line
- METHOD DETAILS
  - Clones and constructs
  - Baculovirus-mediated insect cell protein expression
  - Purification of recombinant proteins and complexes
  - YTHDC2 RNA helicase assay

- ATPase assay
- Electrophoretic Mobility Shift Assay (EMSA)
- Antibodies
- Isolation of endogenous YTHDC2 and MEIOC complexes from mouse testes
- Tissue lysates
- Western Blot
- Histology and Immunofluorescence
- FACS purification of mouse germ cells for RNaseq and single-cell RNA sequencing
- Preparation of RNA libraries

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - iCLIP
  - Analysis of RNA-seq data
  - Analysis of single cell sequencing data

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2022.02.034.

ACKNOWLEDGMENTS

We thank Prof. Bart Deplancke for advice on single-cell transcriptome analysis, and Nicolas Roggli for scientific illustrations. Help from the following core facilities at University of Geneva is acknowledged: Transgenic mouse facility, Zebrafish Core Facility, iGE3 Genomics Platform, Histology, and Bioimaging Center. We also thank the Genomics Core at EMBL Heidelberg, the Functional Genomics Center, Zurich for proteomics analyses. L.L. is supported by an iGE3 PhD Fellowship. This work was supported by grants to R.S.P. from the Swiss National Science Foundation, Switzerland (ERC Transfer Grant CRETP3_166923, Project Grant 310030B_185386, and Sinergia Grant CRSII5_183524, and funding from the NCCR RNA & Disease 51NF40_182880). Work in the Pillai lab is supported by the Republic and Canton of Geneva.

AUTHOR CONTRIBUTIONS

L.L. purified recombinant proteins and performed biochemical assays; K.K. performed analysis of deep sequencing datasets; P.G. prepared FACS-purified germ cells; D.H. performed analysis of mass spectrometry data and assembled all the figures; M.M. and R.J.F. prepared and analyzed the zebrafish mutant; R.R.P. created mouse mutants, generated deep sequencing libraries from mouse samples, performed iCLIP and designed the study together with R.S.P.; manuscript preparation and writing by R.S.P. with input from everyone.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 21, 2021
Revised: February 22, 2022
Accepted: February 23, 2022
Published: March 18, 2022

REFERENCES

Abby, E., Tourpin, S., Ribeiro, J., Daniel, K., Messiaen, S., Moison, D., Guerquin, J., Gaillard, J.C., Armengaud, J., Langa, F., et al. (2016). Implementation of meiosis prophase I programme requires a conserved retinoid-independent stabilizer of meiotic transcripts. Nat. Commun. *7*, 10324.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Bailey, A.S., Batista, P.J., Gold, R.S., Chen, Y.G., de Rooij, D.G., Chang, H.Y., and Fuller, M.T. (2017). The conserved RNA helicase YTHDC2 regulates the transition from proliferation to differentiation in the germline. eLife *6*, e26116.

Bassett, A., and Liu, J.L. (2014). CRISPR/Cas9 mediated genome engineering in Drosophila. Methods *69*, 128–136.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., et al. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell *15*, 707–719.

Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. Trends Biochem. Sci. *37*, 49–57.

Bleichert, F., and Baserga, S.J. (2007). The long unwinding road of RNA helicases. Mol. Cell *27*, 339–352.

Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell *126*, 713–725.

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

Busch, A., Brüggemann, M., Ebersberger, S., and Zarnack, K. (2020). iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites. Methods *178*, 49–62.

Chang, C.-T., Muthukumar, S., Weber, R., Levdansky, Y., Chen, Y., Bhandari, D., Igreja, C., Wohlbold, L., Valkov, E., and Izaurralde, E. (2019). A low-complexity region in human XRN1 directly recruits deadenylation and decapping factors in 5′-3′ messenger RNA decay. Nucleic Acids Res. *47*, 9282–9295.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Chen, D., Wu, C., Zhao, S., Geng, Q., Gao, Y., Li, X., Zhang, Y., and Wang, Z. (2014). Three RNA binding proteins form a complex to promote differentiation of germline stem cell lineage in Drosophila. PLoS Genet. *10*, e1004797.

Dai, X., Cheng, X., Huang, J., Gao, Y., Wang, D., Feng, Z., Zhai, G., Lou, Q., He, J., Wang, Z., and Yin, Z. (2021). Rbm46, a novel germ cell-specific factor, modulates meiotic progression and spermatogenesis. Biol. Reprod. *104*, 1139–1153.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Doxtader, K.A., Wang, P., Scarborough, A.M., Seo, D., Conrad, N.K., and Nam, Y. (2018). Structural Basis for Regulation of METTL16, an S-Adenosylmethionine Homeostasis Factor. Mol. Cell *71*, 1001–1011.e4.

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., and Wu, L. (2016). YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. Nat. Commun. *7*, 12626.

Edupuganti, R.R., Geiger, S., Lindeboom, R.G.H., Shi, H., Hsu, P.J., Lu, Z., Wang, S.-Y., Baltissen, M.P.A., Jansen, P.W.T.C., Rossa, M., et al. (2017). N^6-methyladenosine (m^6A) recruits and repels proteins to regulate mRNA homeostasis. Nat. Struct. Mol. Biol. *24*, 870–878.

Ernst, C., Eling, N., Martinez-Jimenez, C.P., Marioni, J.C., and Odom, D.T. (2019). Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. Nat. Commun. *10*, 1251.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biology *16*, 278.

Fish, R.J., Freire, C., Di Sanza, C., and Neerman-Arbez, M. (2021). Venous Thrombosis and Thrombocyte Activity in Zebrafish Models of Quantitative and Qualitative Fibrinogen Disorders. Int. J. Mol. Sci. *22*, 655.

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m^{6}A RNA methylation. Nat. Rev. Genet. *15*, 293–306.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S., et al. (2015). Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science *347*, 1002–1006.

Grishin, N.V. (1998). The R3H motif: a domain that binds single-stranded nucleic acids. Trends Biochem. Sci. *23*, 329–330.

Harigaya, Y., Tanaka, H., Yamanaka, S., Tanaka, K., Watanabe, Y., Tsutsumi, C., Chikashige, Y., Hiraoka, Y., Yamashita, A., and Yamamoto, M. (2006). Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature *442*, 45–50.

Hiriart, E., Vavasseur, A., Touat-Todeschini, L., Yamashita, A., Gilquin, B., Lambert, E., Perot, J., Shichino, Y., Nazaret, N., Boyault, C., et al. (2012). Mmi1 RNA surveillance machinery directs RNAi complex RITS to specific meiotic genes in fission yeast. EMBO J. *31*, 2296–2308.

Hsu, P.J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., Qi, M., Lu, Z., Shi, H., Wang, J., et al. (2017). Ythdc2 is an N^6-methyladenosine binding protein that regulates mammalian spermatogenesis. Cell Res. *27*, 1115–1127.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods *12*, 115–121.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. Methods *65*, 274–287.

Ivanova, I., Much, C., Di Giacomo, M., Azzi, C., Morgan, M., Moreira, P.N., Monahan, J., Carrieri, C., Enright, A.J., and O'Carroll, D. (2017). The RNA m(6)A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence. Mol. Cell *67*, 1059–1067.e4.

Jain, D., Puno, M.R., Meydan, C., Lailler, N., Mason, C.E., Lima, C.D., Anderson, K.V., and Keeney, S. (2018). *ketu* mutant mice uncover an essential meiotic function for the ancient RNA helicase YTHDC2. eLife *7*, e30919.

Jankowsky, E., and Fairman, M.E. (2007). RNA helicases–one fold for many functions. Curr. Opin. Struct. Biol. *17*, 316–324.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

Kasowitz, S.D., Ma, J., Anderson, S.J., Leu, N.A., Xu, Y., Gregory, B.D., Schultz, R.M., and Wang, P.J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. PLoS Genet. *14*, e1007412.

Kim, G.-W., and Siddiqui, A. (2021). N6-methyladenosine modification of HCV RNA genome regulates cap-independent IRES-mediated translation via YTHDC2 recognition. Proc Natl Acad Sci USA *118*, e2022024118.

Krakau, S., Richard, H., and Marsico, A. (2017). PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. Genome Biol. *18*, 240.

Kretschmer, J., Rao, H., Hackert, P., Sloan, K.E., Höbartner, C., and Bohnsack, M.T. (2018). The m^{6}A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5′-3′ exoribonuclease XRN1. RNA *24*, 1339–1350.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44* (W1), W90–W97.

Kurtenbach, S., and Harbour, J.W. (2019). SparK: a publication-quality NGS visualization tool. Preprint at bioRxiv. https://doi.org/10.1101/845529.

Lasman, L., Krupalnik, V., Viukov, S., Mor, N., Aguilera-Castrejon, A., Schneir, D., Bayerl, J., Mizrahi, O., Peles, S., Tawil, S., et al. (2020). Context-dependent functional compensation between Ythdf m^{6}A reader proteins. Genes Dev. *34*, 1373–1391.

Li, Y., Minor, N.T., Park, J.K., McKearin, D.M., and Maines, J.Z. (2009). Bam and Bgcn antagonize Nanos-dependent germ-line stem cell maintenance. Proc. Natl. Acad. Sci. USA *106*, 9304–9309.

Li, F., Zhao, D., Wu, J., and Shi, Y. (2014). Structure of the YTH domain of human YTHDF2 in complex with an m(6)A mononucleotide reveals an aromatic cage for m(6)A recognition. Cell Res. *24*, 1490–1492.

Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. Nat. Rev. Mol. Cell Biol. *12*, 505–516.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95.

Liu, J., Gao, M., Xu, S., Chen, Y., Wu, K., Liu, H., Wang, J., Yang, X., Wang, J., Liu, W., et al. (2020). YTHDF2/3 are required for somatic reprogramming through different RNA deadenylation pathways. Cell Rep. *32*, 108120.

Liu, R., Kasowitz, S.D., Homolka, D., Leu, N.A., Shaked, J.T., Ruthel, G., Jain, D., Keeney, S., Luo, M., and Pillai, R.S. (2021). YTHDC2 Is Essential for Pachytene Progression and Prevents Aberrant Microtubule-Driven Telomere Clustering in Male Meiosis. Cell Reports *37*, 110110.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular Systems Biology *15*, e8746.

Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. *5*, 2122.

Luo, S., and Tong, L. (2014). Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. Proc. Natl. Acad. Sci. USA *111*, 13834–13839.

Mao, Y., Dong, L., Liu, X.-M., Guo, J., Ma, H., Shen, B., and Qian, S.-B. (2019). m^{6}A in mRNA coding regions promotes translation via the RNA helicase-containing YTHDC2. Nat. Commun. *10*, 5332.

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. Mol. Cell *54*, 751–765.

Mendel, M., Chen, K.M., Homolka, D., Gos, P., Pandey, R.R., McCarthy, A.A., and Pillai, R.S. (2018). Methylation of Structured RNA by the m(6)A Writer METTL16 Is Essential for Mouse Embryonic Development. Mol. Cell *71*, 986–1000.e11.

Mendel, M., Delaney, K., Pandey, R.R., Chen, K.-M., Wenda, J.M., Vågbø, C.B., Steiner, F.A., Homolka, D., and Pillai, R.S. (2021). Splice site m^{6}A methylation prevents binding of U2AF35 to inhibit RNA splicing. Cell *184*, 3125–3142.e25.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell *149*, 1635–1646.

Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., and Peng, Z.Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. Protein Sci. *13*, 1435–1448.

Nakagawa, T., Zhang, T., Kushi, R., Nakano, S., Endo, T., Nakagawa, M., Yanagihara, N., Zarkower, D., and Nakayama, K. (2017). Regulation of mitosis-meiosis transition by the ubiquitin ligase β-TrCP in male germ cells. Development *144*, 4137–4147.

Ohlstein, B., Lavoie, C.A., Vef, O., Gateff, E., and McKearin, D.M. (2000). The Drosophila cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles. Genetics *155*, 1809–1819.

Özeş, A.R., Feoktistova, K., Avanzino, B.C., Baldwin, E.P., and Fraser, C.S. (2014). Real-time fluorescence assays to monitor duplex unwinding and ATPase activities of helicases. Nat. Protoc. *9*, 1645–1661.

Patil, D.P., Pickering, B.F., and Jaffrey, S.R. (2018). Reading m^6A in the Transcriptome: m^6A-Binding Proteins. Trends Cell Biol. *28*, 113–127.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Pendleton, K.E., Chen, B., Liu, K., Hunter, O.V., Xie, Y., Tu, B.P., and Conrad, N.K. (2017). The U6 snRNA m(6)A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell *169*, 824–835.e14.

Prabu, J.R., Müller, M., Thomae, A.W., Schüssler, S., Bonneau, F., Becker, P.B., and Conti, E. (2015). Structure of the RNA Helicase MLE Reveals the Molecular Mechanisms for Uridine Specificity and RNA-ATP Coupling. Mol. Cell *60*, 487–499.

Pyle, A.M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. Annu. Rev. Biophys. *37*, 317–336.

R Core Team. (2018). R Foundation for Statistical Computing; Vienna, Austria: 2018. R: A language and environment for statistical computing. https://www.R-project.org.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Roehr, J.T., Dieterich, C., and Reinert, K. (2017). Flexbar 3.0 - SIMD and multicore parallelization. Bioinformatics *33*, 2941–2942.

Roignant, J.Y., and Soller, M. (2017). m^6A in mRNA: An Ancient Mechanism for Fine-Tuning Gene Expression. Trends Genet. *33*, 380–390.

Ruszkowska, A., Ruszkowski, M., Dauter, Z., and Brown, J.A. (2018). Structural insights into the RNA methyltransferase domain of METTL16. Sci. Rep. *8*, 5311.

Schwartz, S. (2016). Cracking the epitranscriptome. RNA *22*, 169–174.

Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G., et al. (2013). High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell *155*, 1409–1421.

Śledź, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. eLife *5*, e18434.

Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. *27*, 491–499.

Soh, Y.Q.S., Mikedis, M.M., Kojima, M., Godfrey, A.K., de Rooij, D.G., and Page, D.C. (2017). Meioc maintains an extended meiotic prophase I in mice. PLoS Genet. *13*, e1006704.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., et al. (2013). Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. Cell Rep *3*, 2179–2190.

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

Theler, D., Dominguez, C., Blatter, M., Boudet, J., and Allain, F.H. (2014). Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. Nucleic Acids Res. *42*, 13911–13919.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388–1399.

Wang, C., Zhu, Y., Bao, H., Jiang, Y., Xu, C., Wu, J., and Shi, Y. (2016a). A novel RNA-binding mode of the YTH domain reveals the mechanism for recognition of determinant of selective removal by Mmi1. Nucleic Acids Res. *44*, 969–982.

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

Warda, A.S., Kretschmer, J., Hackert, P., Lenz, C., Urlaub, H., Höbartner, C., Sloan, K.E., and Bohnsack, M.T. (2017). Human METTL16 is a N^6-methyladenosine (m^6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. EMBO Rep. *18*, 2004–2014.

Watabe, E., Togo-Ohno, M., Ishigami, Y., Wani, S., Hirota, K., Kimura-Asami, M., Hasan, S., Takei, S., Fukamizu, A., Suzuki, Y., et al. (2021). m^6A-mediated alternative splicing coupled with nonsense-mediated mRNA decay regulates SAM synthetase homeostasis. EMBO J. *40*, e106434.

Wojtas, M.N., Pandey, R.R., Mendel, M., Homolka, D., Sachidanandam, R., and Pillai, R.S. (2017). Regulation of m6A Transcripts by the 3′→5′ RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. Mol. Cell *68*, 374–387.e12.

Wu, M., Reuter, M., Lilie, H., Liu, Y., Wahle, E., and Song, H. (2005). Structural insight into poly(A) binding and catalytic mechanism of human PARN. EMBO J. *24*, 4082–4093.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biology *19*, 15.

Wu, R., Chen, Y., Liu, Y., Zhuang, L., Chen, W., Zeng, B., Liao, X., Guo, G., Wang, Y., and Wang, X. (2021). m6A methylation promotes white-to-beige fat transition by facilitating Hif1a translation. EMBO Rep. *22*, e52348.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.S., Hao, Y.J., Sun, B.F., Sun, H.Y., Li, A., Ping, X.L., Lai, W.Y., et al. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Mol. Cell *61*, 507–519.

Xu, C., Wang, X., Liu, K., Roundtree, I.A., Tempel, W., Li, Y., Lu, Z., He, C., and Min, J. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. Nat. Chem. Biol. *10*, 927–929.

Xu, C., Liu, K., Ahmed, H., Loppnau, P., Schapira, M., and Min, J. (2015). Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. J. Biol. Chem. *290*, 24902–24913.

Yamashita, A., Shichino, Y., Tanaka, H., Hiriart, E., Touat-Todeschini, L., Vavasseur, A., Ding, D.-Q., Hiraoka, Y., Verdel, A., and Yamamoto, M. (2012). Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. Open Biol. *2*, 120014.

Yoshida, H., Park, S.-Y., Sakashita, G., Nariai, Y., Kuwasako, K., Muto, Y., Urano, T., and Obayashi, E. (2020). Elucidation of the aberrant 3′ splice site selection by cancer-associated mutations on the U2AF1. Nat. Commun. *11*, 4744.

Zaccara, S., and Jaffrey, S.R. (2020). A unified model for the function of YTHDF proteins in regulating m6A-modified mRNA. Cell *181*, 1582–1595.e18.

Zhang, Z., Theler, D., Kaminska, K.H., Hiller, M., de la Grange, P., Pudimat, R., Rafalska, I., Heinrich, B., Bujnicki, J.M., Allain, F.H., and Stamm, S. (2010). The YTH domain is a novel RNA binding domain. J. Biol. Chem. *285*, 14701–14710.

Zhu, T., Roundtree, I.A., Wang, P., Wang, X., Wang, L., Sun, C., Tian, Y., Li, J., He, C., and Xu, Y. (2014). Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. Cell Res. *24*, 1493–1496.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Polyclonal rabbit anti-SYCP3	Novus Biologicals	Cat. No. NB300-231, RRID:AB_1522841
Polyclonal rabbit anti-HA	Santa Cruz	Cat. No. sc-805, RRID:AB_631618
Monoclonal mouse anti-FLAG	Sigma	Cat. No. F3165, RRID: AB_259529
Anti-HA affinity matrix	Roche	Cat. No. 11815016001, RRID:AB_390914
Anti-FLAG M2 Magnetic Beads	Sigma	Cat. No. M8823, RRID:AB_2637089
Goat anti-mouse Alexa Fluor 488	ThermoFischer Scientific	Cat. No. A28175, RRID: AB_2536161
Goat anti-rabbit Alexa Fluor 594	ThermoFischer Scientific	Cat. No. A11037, RRID: AB_2534095
HRP-conjugated anti-rabbit IgG	GE Healthcare	Cat. No. NA934, RRID:AB_772206
HRP-conjugated anti-mouse IgG	Thermo Fisher	Cat. No. A27025, RRID:AB_2536089
Bacterial and virus strains		
BL21(DE3) bacterial strain	NEB	Cat. No. C2527H
DH10EMBacY bacterial strain	(Bieniossek et al., 2012)	N/A
Chemicals, peptides, and recombinant proteins		
γ-^{32}P-Adenosine triphosphate	Perkin Elmer	Cat. No. NEG002A001MC
γ-^{32}P- Adenosine triphosphate	Hartmann Analytic	Cat. No. SRP-501
40% Acrylamide/Bis Solution 19:1	Bio-Rad	Cat. No. 1610144
N,N,N',N'-Tetramethylethylendiamin	Merck	Cat. No. 1107320100
S. pyrogenes Cas9 mRNA	ThermoFischer Scientific	Cat. No. A29378
S. pyrogenes Cas9 protein	IDT	Cat. No. 1081058
S. pyrogenes Cas9 protein	PNAbio	Cat. No. CP01-20
FuGENE HD transfection reagent	Promega	Cat. No. E231A
Cre Recombinase	NEB	Cat. No. M0298
D-desthiobiotin	IBA	Cat. No. 2-1000-001
Ammonium persulphate	AppliChem	Cat. No. A2941
ATP	Thermo Scientific	Cat. No. R0441
Glycogen, RNA grade	Thermo Scientific	Cat. No. R0551
Lithium Chloride	Merck	Cat. No. 1056790250
Formic acid	Fluka	Cat. No. 06440
T4 Polynucleotide Kinase	Thermo Scientific	Cat. No. EK0031
Protease Inhibitor Tablets, EDTA-free	Thermo Scientific	Cat. No. A32965
L-Glutathione reduced	Merck	Cat. No. G4251-25G
Critical commercial assays/kits		
NEBNext® Multiplex Small RNA Library Prep Set for Illumina®	NEB	Cat. No. E7300
MinElute Gel Extraction Kit	QIAGEN	Cat. No. 28604
MEGAshortscript™ T7 Transcription Kit	Life technologies	Cat. no. AM1354
Mini elute kit	QIAGEN	Cat. No. 28004
Platinum™ II Hot-Start PCR Master Mix	ThermoFisherScientific	Cat. No. 14000012
Illustra MicroSpin G-25 micro columns,	GE Healthcare	Cat. No. 27-5325-01
Phire Green Hot Start II PCR Master Mix	Thermo Scientific	Cat. No. F126L
Deposited data		
Deep sequencing datasets	This study.	GEO: GSE196427
All raw gel data are deposited at Mendeley Data.	This study.	https://doi.org/10.17632/wrnxpfmxnd.1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Cell lines		
Sf21 insect cells for protein production	Eukaryotic Expression Facility, EMBL Grenoble, France	N/A
High Five (Hi5) insect cells for protein production	Eukaryotic Expression Facility, EMBL Grenoble, France	N/A
Experimental models: Organisms/strains		
Mouse: *Ythdc2 knock-out*	(Wojtas et al., 2017)	EMMA: EM11456
Mouse: *Ythdc2-cat-dead*	This study.	Dominant-negative mutation, not maintained.
Mouse: *Ythdc2-YTH*	This study.	Available from Lead Contact.
Mouse: *Ythdc2-3xFLAG*	This study.	Available from Lead Contact.
Mouse: *2xHA-Meioc*	This study.	Available from Lead Contact.
Zebrafish: *ythdc2*^{D46}	This study.	Available from Lead Contact.
Oligonucleotides		
DNA and RNA oligos	N/A	See Table S1
Recombinant DNA		
pACEBac2	(Bieniossek et al., 2012)	N/A
pACEBac2S	(Wojtas et al., 2017)	Modified pACEBac2 expressing N-terminal 6xHis-Strep-SUMO-TEV fusion protein.
pACEBac2-GST	This study	Modified pACEBac2 expressing C-terminal TEV-GST fusion protein.
pIDK	(Bieniossek et al., 2012)	N/A
Software and algorithms		
FastQC v0.11.5	N/A	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Flexbar v3.5.0	N/A	https://github.com/seqan/flexbar
SeqAn v2.4.0	N/A	https://github.com/seqan/seqan
STAR v2.7.9a	(Dobin et al., 2013)	N/A
umi_tools dedup v1.1.1	(Smith et al., 2017)	N/A
samtools merge v1.7	(Li et al., 2009)	N/A
PureCLIP v1.3.1	(Krakau et al., 2017)	N/A
SparK v2.6.2	(Kurtenbach and Harbour, 2019)	N/A
Kallisto v0.46.2	(Bray et al., 2016)	N/A
Salmon	(Patro et al., 2017)	https://combine-lab.github.io/salmon/
DESeq2	(Love et al., 2014)	N/A
seaborn package v0.10.0	N/A	https://seaborn.pydata.org/
matplotlib v3.1.3	N/A	https://matplotlib.org/
ENRICHR	(Chen et al., 2013; Kuleshov et al., 2016)	http://amp.pharm.mssm.edu/Enrichr/
Cellranger pipeline v3.0.2	N/A	https://github.com/10XGenomics/cellranger
Python v3.6.9	N/A	https://www.python.org/
Scanpy v1.4.5	(Wolf et al., 2018)	N/A
Scran bioconductor package	(Lun et al., 2016)	N/A
mnnpy package	N/A	https://pypi.org/project/mnnpy/
MAST algorithm	(Finak et al., 2015)	https://github.com/RGLab/MAST
Bioconductor	(Huber et al., 2015)	https://www.bioconductor.org/
R	(R Core Team, 2018)	https://www.r-project.org

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BLAST	(Altschul et al., 1997)	https://blast.ncbi.nlm.nih.gov/
edgeR	(Robinson et al., 2010)	https://bioconductor.org/packages/release/bioc/html/edgeR.html
Other		
Chelating Sepharose Fast Flow beads	GE Healthcare	Cat. No. 17-0575-01
StrepTrap HP	GE Healthcare	Cat. No. 28-9075-46
Superdex S75 10/300 GL	GE Healthcare	Cat. No. 17-5174-01
Superdex 200 10/300 GL	GE Healthcare	Cat. No. 17-5175-01
MethaPhor agarose	Lonza	Cat. No. 50180
Protino Glutathione Agarose 4B	Macherey-Nagel	Cat. No. 745500.10
TLC PEI Cellulose F	Merck	Cat. No. 1055790001
SF 900 III SFM	GIBCO	Cat. No. 12658027
Express Five SFM	GIBCO	Cat. No. 10486025
CellCarrier-96 Black, Optically Clear Bottom, Tissue Culture Treated, Sterile, 96-Well with Lid	Perkin Elmer	Cat. No. 6005550
Strep-Tactin Sepharose resin	IBA	Cat. No. 2-1201-010

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ramesh S. Pillai (ramesh.pillai@unige.ch).

Materials availability

All unique reagents including plasmids and animal models generated in this study are available from the Lead Contact without any restrictions for academic research purposes. The host institution of the Lead Contact requires the signing of a Materials Transfer Agreement (MTA) for both academic and non-academic users.

Data and code availability

Single-cell RNA-seq data and other deep sequencing data generated in the study are deposited with the Gene Expression Omnibus (GEO: GSE196427). This accession number is also provided in the key resources table. Original data including western blots, thin-layer chromatography plate scans, the raw immunofluorescence and histological microscopy images are deposited at Mendeley and are publicly available as of the date of publication. The DOI is listed in the key resources table.

This paper does not report original code. A detailed description of the use of publicly available programs is mentioned in the methods, and also listed in key resources table. Any additional information required to reanalyze the data reported in this paper is available from the Lead Contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal work

Mutant mice were generated at the Transgenic Mouse Facility, University of Geneva. The mice were bred in the Animal Facility of Sciences III, University of Geneva. Mutant fish were generated and maintained in the University of Geneva zebrafish core facility.

The use of animals in research at the University of Geneva is regulated by the Animal Welfare Federal Law (LPA 2005), the Animal Welfare Ordinance (OPAn 2008) and the Animal Experimentation Ordinance (OEXA 2010). The Swiss legislation respects the Directive 2010/63/EU of the European Union. Any project involving animal work has to be approved by the Direction Générale de la Santé and the official ethic committee of the State of Geneva performing a harm-benefit analysis of the project. Animals are treated with respect based on the 3Rs principle in the animal care facility of the University of Geneva. We use the lowest number of animals needed to conduct our specific research project. Discomfort, distress, pain and injury is limited to what is indispensable and anesthesia and analgesia is provided when necessary. Daily care and maintenance are ensured by fully trained and certified staff. This research with mouse and zebrafish is licensed by the Geneva Cantonal authorities under the authorization number Pillai GE-162-19B.

Ythdc2 mutant mice

The *Ythdc2* gene is located on mouse chromosome 18 and consists of 30 exons (Figure S1A). The generation of *Ythdc2* knockout mouse line is described previously (Wojtas et al., 2017). The exon 7 encodes for parts of the N-terminal RecA domain of the RNA helicase module in YTHDC2, and specifically the ATPase motif DEVH. YTH domain is encoded in exon 27-29. We targeted the endogenous *Ythdc2* locus in mouse embryos of the hybrid background B6D2F1 (50% C57BL/6 and 50% DBA/2) with specific guide RNAs and homology repair templates to create *Ythdc2* mutant lines inactivating the RNA helicase domain and the m^6A-binding capacity of the YTH domain.

Preparation of sgRNAs. A cloning-free method (Bassett and Liu, 2014) was used to prepare DNA template for *in-vitro* synthetic single guide RNA (sgRNA) transcription. Briefly, a common reverse primer (CRISPR sgR primer) and a gene specific CRISPR forward (F) primer with T7 promoter sequence was used to PCR amplify the sgDNA template (Table S1). The following components were mixed to prepare the PCR reaction: 20 μl 5X Phusion HF buffer, 67 μl ddH2O, 2 μl 10 mM dNTPs, 5 μl, 10 μM CRISPR F primer, 5 μl 10 μM CRISPR sgR primer and 1 μl Phusion DNA polymerase. The PCR reaction was set as follows: 98°C 30 s, 35 cycles of [98°C, 10 s; 60°C, 30 s and 72°C, 15 s], 72°C 10 min, 4°C hold. PCR product (~110 bp) was agarose gel-purified using mini-elute gel extraction kit (QIAGEN; Cat. No. 28604). The purified DNA was used to produce sgRNA by *in vitro* transcription reaction using T7 promoter. *In vitro* transcription was carried out with the MEGAshortscript T7 Transcription Kit (Life technologies; Cat. No. AM1354) for 4 h at 37°C. Reactions were treated with TURBO DNase I (37°C for 15 min) to remove template DNA, phenol-chloroform extracted and precipitated with ethanol. Quality of the generated gRNAs were verified by 1% agarose gel electrophoresis.

For single-cell sequencing, we used purified 4n germ cells from P10 males of heterozygous and homozygous *Ythdc2* KO aninmals (Figure 5A). Biological duplicates were used for preparing each single-cell RNaseq library.

sgRNA generation for Ythdc2-3xFLAG-tagged mouse line

Our previous attempts to tag (3xFLAG tag with a linker) the N- and C-termini of mouse YTHDC2 resulted in homozygous knockin mice that were infertile. This suggested that tagging at these positions resulted in destabilization of the endogenous protein in mouse testes. Therefore, we chose an internal location to place the tag. Purification of the recombinant human YTHDC2 protein (Wojtas et al., 2017) revealed the presence of a minor fraction of the protein that lacked the YTH domain. Peptide mass fingerprinting revealed this to be due to proteolysis within a putative loop just before the YTH domain. Given that this internal loop is likely to be surface-accessible, we placed an in-frame 3xFLAG tag (with a linker) to create the $Ythdc2^{FLAG}$ knockin mouse. Such homozygous knockin mice are fertile.

Forward (F) primer for sgRNA generation for C-terminal 3xFLAG tag in mouse YTHDC2: RRoligo868
GAAATTAATACGACTCACTATAGGCTTGGAGATCCAATAATAGTGTTTTAGAGCTAGAAATAGC
ssDNA homology oligo: RRoligo867 for insertion of 3xFLAG in exon 25 of *Ythdc2* (Strand +)
5′CCCATGTCTTCAGAAGAACTTCCTTTGGCTTCATCTTGGAGATCCAATAATAGT**ATG**GACTACAAAGACCATGACGGTGATTAT
AAAGATCATGACATCGATTACAAGGATGACGATGACAAGGGTGGAGGCGGGTCTATGAGGAAAAGTACAGCTGACACTGAGTTTG
CGGATGGATCTACTACTGGAGAAAGG 3′

We used mouse testes from P15 $Ythdc2^{FLAG}$ animals for iCLIP experiments (Figure 1) and mouse testes from P20 $Ythdc2^{FLAG}$; $Meioc^{HA}$ animals for anti-FLAG protein complex purifications (Figure 2).

sgRNA generation for $Ythdc2^{YTH}$ mutant mouse line

Forward (F) primer for sgRNA generation for YTH domain mutation W1375A in mouse YTHDC2: RRoligo1064 (Table S1).
5′GAAATTAATACGACTCACTATAGGGTCCAGCTGACCCCCAGTCCGTTTTAGAGCTAGAAATAGC3′
Homology oligo for YTH domain mutation W1375A in mouse *Ythdc2* (strand -): RRoligo1060
5′AAATGGTGTGCAAACTGAAAGGGAAGGCTTTCTTTGCGTATCCACTCCACTTTAAATACTCCACCTAGTCCAGCTGACCC**TGCA**
TCCTGGCTCTTCTCCCTTCCTATCTCAGAAGACATCCTAGAAAATCCCTGTATAAAAATGAGGGAG 3′

Preparation of injection mix. We mixed 12.5 ng/μL of each of the gRNA with 12.5 ng/μL of a 181 nt ssDNA homology repair template (IDT), and 25 ng/μL of Cas9 mRNA (ThermoFischer Scientifique; Cat. No. A29378) in injection buffer (10 mM Tris pH 7.5, 1 mM EDTA pH 8.0). Injection mixture can be frozen at −80°C in aliquots (20 μL) prior to use.

Injection of mouse embryos of the hybrid background B6D2F1 (50% C57BL/6 and 50% DBA/2, black coat color) was carried out at the Transgenic Mouse Core Facility, University Medical Centre (CMU), University of Geneva. The NMRI (Naval Medical Research Institute) mice, which have a white coat color were used as foster mothers.

The 2n and 4n testicular germ cells from adult (P60) $Ythdc2^{YTH}$ homozygous and heterozygous mice were used for RNasequencing (Figure 1M). Biological triplicates were used.

Ythdc2 E332Q mutant and $Meioc^{HA}$ knockin mouse line

For these mouse knockin lines, the gRNAs were commercially synthesized [Integrated DNA Technologies (IDT)] in two parts.

Purchase and annealing of gRNA. We purchased a tracrRNA (67 bases long, IDT; Cat. No. 1072533) which is the common part of the gRNA that binds to the Cas9 protein, and a crRNA (IDT) that has complementarity of 20 bases with the genomic DNA sequence of interest plus common additional bases (GUUUUAGAGCUAUGCU) for pairing with the 3′ end of tracrRNA. The complete gRNA was obtained by annealing 200 pmol each of tracrRNA and crRNA in 1xTE buffer (1X TE buffer, pH 7.5, IDT; Cat. No. 11-05-01-05) to a final volume of 10 uL. The gRNAs mix was heated at 95°C for 5 min and then allowed to cool at room temperature for 30 min to get annealed functional gRNA. The annealed gRNA was stored at −20°C until used for injection into mouse embryos.

crRNA: /Alt1/NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAUGCU /Alt2/

N represent the gene-specific sequence. The sequence of crRNAs used are provided below.

crRNA and ssDNA repair oligo for Meioc^{HA} mouse line

SgRNA for 2xHA tag mouse at MEIOC N-term

sgRNA1: TCCGCCGCTCACCTCCATGG

CrRNA synthesized from IDT for *Meioc* N-term targeting sgRNA1

RRoligo1142: 5′/AlTR1/rUrCrCrGrCrCrGrCrUrCrArCrCrUrCrCrArUrGrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU/AlTR2/3′

Homology ssDNA oligo for 2xHA at MEIOC N-term (reverse complement)

5′TCCCTCAGGCCTTGGGGGTGACGGGGCCGGCAGGTGTCTCCGCCGCTCACCTCCATAGACCCGCCTCCACCAGCGTAATC
TGGAACATCATATGGGTAGCCTCCACCAGCGTAATCTGGAACATCATATGGGTACATGGGGGCCCCCCAGCTCCTCCTGAGCTA
GTTAACAGCTCCCGGGGTACGGGGGTGAGGGAG 3′

crRNA and ssDNA oligo for Ythdc2$^{cat\text{-}dead}$ mouse mutant line

We made a point mutation (E332Q) within the ATP hydrolysis motif (DEVH → DQVH) present in the N-terminal RecA module of the RNA helicase domain to create the *Ythdc2$^{cat\text{-}dead}$* mice. A similar mutation (E317Q) in the recombinant human YTHDC2 inactivates the ATPase and RNA helicase activities (Wojtas et al., 2017).

sgRNA: TTCATGCACTTCATCCTGCA

CrRNA synthesized from IDT for *Ythdc2* E332Q locus targeting sgRNA

RR_crRNA_1196

5′/AlTR1/rUrUrCrArUrGrCrArCrUrUrCrArUrCrCrUrGrCrArGrUrUrUrUrArGrArGrCrUrArUrGrCrU/AlTR2/ 3′

The sequence of the ssDNA repair template for E332Q mutation is as follows:

TATTTAATTTAAACACTATAAATCCAGCAGTGTTTCCTGTTAAAGGTGAACCGTTCACAGTATTTTACCTGATCCATCATTTTGTTTC
TGTTTCTCTTCCTTGCAG**GACCAA**GTGCATGAAAG**A**GATCGATTCAGTGATTTTTTGCTTACTAAGTTAAGAGATTTGTTGCAAAAG
CACCCAACTTTG

We used mouse testes from P20 *Ythdc2^{FLAG};Meioc^{HA}* animals for anti-HA protein complex purifications (Figure 2). For *Ythdc2$^{cat\text{-}dead}$* mutant quadruplicate P14 testes were used (Figure 3).

Preparation of injection mix. Just before injection into mouse embryos, the annealed gRNAs (final concentration 0.6 pmol/μL) were mixed with the Cas9 protein at 30 ng/μL final concentration (IDT; Cat. No. 1081058) in a final volume of 9 μL. The mix was incubated at room temperature for 10 min for complex formation. For knockin mouse generation a ssDNA homology repair template (IDT; 20 ng/μL final concentration) was also added to the injection mix and volume was adjusted with 1x TE buffer, pH 7.5, to 100 μL. The injection mix was centrifuged at 13,000 rpm for five minutes at 4°C, and 50 μL of supernatant was collected and stored on ice.

Injection of mouse embryos of the hybrid background B6D2F1/J (black coat color) was carried out at the Transgenic Mouse Core Facility, University Medical Centre (CMU), University of Geneva. The NMRI (Naval Medical Research Institute) mice, which have a white coat color were used as foster mothers.

Founder mice of hybrid background B6D2F1 were identified by genotyping and crossed with wildtype C57Bl6/J (Jackson Labs) partners to obtain germline transmission. Intercrosses between heterozygous (*Ythdc2$^{KI/+}$*) knockin (KI) males and females were used to obtain homozygous (*Ythdc2$^{-KI/KI}$*) knockin males and females. Fertility was functionally assessed by crossing them with fertile heterozygous or wildtype partners. The same mutant animal was mated with a fertile partner for 1 week each, using a total of at least three different fertile partners successively. At least 3 homozygous males and females were tested. Further analyses of fertility were carried out using histological analysis of dissected ovaries and testes for comparison to the fertile control animals (Figure 1L). The heterozygous and homozygous *Ythdc2$^{cat\text{-}dead}$* founder mice of both sexes were found to be infertile when crossed with control fertile partners (Figure 3A), which prevented us from establishing the colony. All analyses with *Ythdc2$^{cat\text{-}dead}$* mutant were conducted with founder mice (Figure 3A).

Genotyping

Routine genotyping was done using ear punches of the weaned animals (which is also used for identification purposes). Ear punches were digested in 100 μl of buffer containing 10 mM NaOH, 0.1 mM EDTA for 90 min at 95°C. After centrifugation at 3000 rpm for 10 min, 50 μL of supernatant was transferred to a new tube containing 50 μl of TE buffer (20mM Tris-HCl, pH 8.0 and 0.1 mM EDTA).

For experimental mice that were euthanized before the weaning age, tail biopsies were digested in 500 μL tail buffer (50 mM Tris-HCl, pH 8.0, 100 mM EDTA, 100 mM NaCl, 1% SDS) with 2.5 μg of Proteinase K at 55°C overnight. After spinning at 16000 × g for 10 min to remove hairs, supernatants were transferred into a new tube and the DNA was precipitated by adding 500 μL of isopropanol. Samples were spun in the centrifuge at 16000 × g for 10 min, and the resulting pellet was washed with 1 mL 70% Ethanol. The pellet was dried and resuspended for at least 1 h at 37°C in 100 to 150 μL of 10 mM Tris-HCl, pH 8.0. Approximately, 1-1.5 μL of genomic DNA were used for PCR with Phire Green Hot Start II PCR Master Mix (Thermo Scientific, Cat. No. F126L).

Genotyping PCR for Ythdc2^{FLAG} mouse line

Primers RRoligo869 and RRoligo870 (Table S1) were used to amplify the edited *Ythdc2* gene locus followed by Sanger sequencing with RRoligo869 to identify the 3xFLAG-tag insertion.

Product size: 301 (KI), 214bp (WT)

PCR condition: 98°C-30 s, 35 cycles of (98°C −5 s, 60°C- 5 s, 72°C- 10 s), and 72°C −60 s.

Genotyping PCR for Ythdc2 W1375A mutant mouse line. Primers RRoligo1092 and RRoligo1093 (Table S1) were used to amplify the edited *Ythdc2* gene locus followed by sanger sequencing with RRoligo1093 to identify the *Ythdc2* W1375A mutation.

Product size: 282bp.

PCR condition: 98°C-30 s, 35 cycles of (98°C −5 s, 60°C- 5 s, 72°C- 10 s), and 72°C −60 s.

Genotyping PCR for Meioc^{HA} mouse line. Primers RRoligo1151 and RRoligo1152 (Table S1) were used to amplify the edited *Meioc* gene locus followed by sanger sequencing with RRoligo1151 to identify the *Meioc^{HA}* allele.

Product size: 282bp. 323bp (wt), 400 bp (2xHA-tagged KI allele)

PCR condition: 98°C-30 s, 35 cycles of (98°C −5 s, 63°C- 5 s, 72°C- 10 s), and 72°C −60 s. PCR reaction was supplemented with 3% DMSO.

Genotyping primer for Ythdc2$^{cat\text{-}dead}$ mutation. Primers RRoligo1358 and RRoligo1359 (Table S1) were used to amplify the edited *Ythdc2* gene locus followed by sanger sequencing with RRoligo1360 to identify the *E332Q* catalytic-dead mutation.

Product size: 600bp

PCR condition: 98°C-30 s, 35 cycles of (98°C −5 s, 60°C- 5 s, 72°C- 10 s), and 72°C −60 s.

Zebrafish Ythdc2 mutant

The *Danio rerio* (zebrafish) *ythdc2* gene is located on chromosome 5 and predicted to comprise 30 exons. To create a knockout allele of *ythdc2*, we engineered a small deletion in exon 5 of early zebrafish embryos using the CRISPR-Cas9 genome editing tool. Exon 5 encodes for part of the RNA helicase domain. Two guide RNAs were mixed with recombinant *S. pyrogenes* Cas9 protein (PNAbio; Cat. No. CP01-20) and injected into embryos of the Tubingen background. After screening multiple founders, we retained one line that germline-transmitted a 46-nucleotide deletion in *ythdc2*, resulting in removal of 123 to 456 aa from the N-terminal RecA module of the RNA helicase domain (*ythdc2$^{\Delta 46}$*). The genomic sequences for guide RNA designs, with PAM sequences in brackets, were ACGCAGCAGCGCCCCTGAGC(TGG) and TGCTCCTGGAGCTCGTAGAC(CGG).

The founder was crossed with wildtype partners to establish a colony of heterozygous *ythdc2$^{+/\Delta 46}$* animals. In-crosses between heterozygotes gave rise to *ythdc2$^{+/+}$*, *ythdc2$^{+/\Delta 46}$* and *ythdc2$^{\Delta 46/\Delta 46}$* embryos in expected Mendelian ratios. This was determined by genotyping 1 dpf embryos and 5 dpf larvae (Figure 5A). Adult homozygous *ythdc2$^{\Delta 46/\Delta 46}$* males were crossed with wildtype or heterozygous partners but did not result in any viable embryos (Figure 5B), indicating male infertility. All the homozygous mutant animals were males, and no females (based on observable phenotypes) were found.

Early embryos of the Tubingen (TU) laboratory zebrafish background were injected with approximately 1 nL per embryo of reagents for genome editing, essentially as described previously (Fish et al., 2021).

Genotyping

Zebrafish embryos, larvae or fin sections from adults were genotyped using PCR primers (Table S1) to amplify genomic DNA flanking the *ythdc2$^{\Delta 46}$* deletion site. The deletion was sufficiently large to discriminate wildtype from Δ46 alleles and assign *ythdc2$^{+/+}$*, *ythdc2$^{+/\Delta 46}$* and *ythdc2$^{\Delta 46/\Delta 46}$* genotypes based on visual examination of the products resolved in an agarose gel (Figure S5A).

METHOD DETAILS

Clones and constructs

Constructs for protein expression in insect cells

The coding sequence for full-length human YTHDC2 (1430 aa; UniProtKB - Q9H6S0) was cloned into pACEBac2S vector (modified from pACEBac2) (Bieniossek et al., 2012) to express as N-terminal 6xHis-Strep-SUMO-TEV fusion protein. The same vector was used to clone a catalytic-dead version of human YTHDC2 (hYTHDC2^{E317Q}) and YTHDC2 with ANK repeats deletion (hYTHDC2$^{\Delta ANK}$). Note we put a linker (GSAGSAAGSGEFLE) to join the two RecA domains after removal of the ANK repeats, as previously done for a mammalian expression construct (Wojtas et al., 2017). Human XRN1 truncation (1-1279 aa; UniProtKB - Q8IZH2) coding sequence was cloned into pACEBac2-GST vector (modified from pACEBac2) to express recombinant protein with C-terminal TEV-GST fusions. A catalytic-dead (E178Q) version of human XRN1(1-1279)^{E178Q} was cloned into pACEBac2-GST vector as well for use in RNA helicase assays.

To co-express full-length YTHDC2 with XRN1 (1-1279 aa), the required coding sequence for XRN1 was cloned into pIDK vector and recombined with pACEBac2S-YTHDC2 via MultiBac (Figure 2A). The coding sequence for human MEIOC coiled-coil domain (hMEIOC_CC, 740-952 aa, UniProtKB - A2RUB1) was cloned into pACEBac2-GST vector to co-express with full-length YTHDC2 protein (Figure 2B).

Since the human MEIOC_CC domain construct was not soluble when expressed alone, we expressed the mouse MEIOC coiled-coil domain (mMEIOC_CC, 721-950 aa, UniProtKB - A2AG06) as a soluble protein from the pACEBac2S vector for use in RNA helicase assays (Figure 3G, S3D, and S3G). This expressed stretch of mouse MEIOC has 91% identity with the human protein. All the clones were verified by restriction digestion of the isolated plasmids, as well as by PCR.

Baculovirus-mediated insect cell protein expression

The expression and purification of human YTHDC2 or its variants were as previously described (Wojtas et al., 2017). Similarly, for purification of individual proteins, the pACEBac2-hXRN1(1-1279)-GST or its catalytic-dead version, the pACEBac2-hMEIOC_CC-GST and pACEBac2S-mMEIOC_CC plasmids were separately transformed into DH10EMBacY competent cells for preparation of

the bacmids. The bacmid DNA was extracted and transfected with FuGENE HD (Promega; Cat. No. E231A) into the Sf9 insect cells for virus production. The supernatant (V0) containing the recombinant baculovirus was collected after 48 to 72 h post-transfection. To expand the virus pool, 2.0 mL of the V0 virus stock was added into 25 mL of Sf9 (0.5 – 0.8 × 10^6/mL) cells. The resulting cell culture supernatant (V1) was collected 24 h post-proliferation arrest. For large-scale expression of the protein, Hi5 cells were infected with virus (V1) and cells were harvested 72 h after infection.

For production of YTHDC2-XRN1 complex, the pIDK-XRN1 (1-1279 aa) plasmid was recombined with pACEBac2S-YTHDC2 via Cre-recombination (NEB; Cat No. M0298S). This reaction product was transformed into DH10EMBacY competent cells for preparation of the MultiBac bacmid. The bacmid DNA was then extracted and transfected with FuGENE HD into the Sf9 insect cells for virus production. For large-scale expression of the protein, Hi5 cells were infected with virus (V1) and cells were harvested 72 h after infection.

For production of YTHDC2-hMEIOC_CC complex, Hi5 cells were co-infected with baculoviruses expressing either 6x-His-Strep-Sumo-YTHDC2 or hMEIOC_CC-GST. Cells were harvested 72 h after infection.

Purification of recombinant proteins and complexes

Purification of recombinant full-length human YTHDC2 or its variants is described (Wojtas et al., 2017).

Hi5 cells expressing XRN1 (1-1279 aa) or its catalytic mutant (E178Q) versions were resuspended by lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5% glycerol and 5 mM 2-mercaptoethanol) supplemented with proteinase inhibitor (Thermo Scientific, EDTA-free). Cells were disrupted by sonication and then cell lysate was centrifuged at 18,000 rpm for 50 min at 4°C. The clarified supernatant was incubated with Protino Glutathione Agarose 4B (Macherey-Nagel; Cat. No. 745500.10) at 4°C for 2 h. After incubation, beads were washed with lysis buffer and finally bound protein was eluted with 15 mM L-Glutathione reduced in lysis buffer. The C-terminal GST tag was removed by TEV cleavage overnight at 4°C during dialysis against the buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol). After tag cleavage, protein was further purified through Superdex 200 increase 10/300 column (GE Healthcare) equilibrated with buffer (25 mM HEPES-KOH, pH 7.5, 150 mM KCl, 5% glycerol, 1 mM DTT) S2. Fractions of pure protein were confirmed by SDS-PAGE electrophoresis, flash-frozen in liquid nitrogen in 10% glycerol and stored at −80°C.

Hi5 cells expressing mouse MEIOC coiled-coil domain (mMEIOC_CC) were resuspended by lysis buffer (40 mM imidazole, 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5% glycerol and 5 mM 2-mercaptoethanol) supplemented with proteinase inhibitor (Thermo Scientific, EDTA-free). Cells were disrupted by sonication and then cell lysate was centrifuged at 18,000 rpm for 50 min at 4°C. The clarified supernatant was incubated with Ni^{2+} chelating Sepharose FF beads (GE Health; Cat. No. 17057501) at 4°C for 2 h. After incubation, beads were washed by 50/60 mM imidazole in lysis buffer and the bound protein was eluted by 250 mM imidazole in buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5% glycerol and 5 mM 2-mercaptoethanol). The N-terminal 6XHis-Strep-SUMO tag was removed by TEV cleavage overnight at 4°C during dialysis against buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol). After dialysis, the cleaved tag was removed by a second nickel column purification, and the flow-through containing untagged protein was concentrated and subjected to size-exclusion chromatography (Superdex 75 increase 10/300 column, GE Healthcare) equilibrated with buffer (25 mM Tris-HCl, pH 8.0, 150 mM KCl, 5% glycerol, 1 mM DTT). Fractions of pure protein were confirmed by SDS-PAGE electrophoresis, flash-frozen in liquid nitrogen in 10% glycerol and stored at −80°C.

Hi5 cells co-expressing YTHDC2 and XRN1 (1-1279 aa) from a MultiBac construct were resuspended in the lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 40 mM Imidazole, 5% glycerol and 5 mM 2-mercaptoethanol) supplemented with proteinase inhibitor (Thermo Scientific, EDTA-free). Cells were disrupted by sonication and clarified by centrifugation at 18,000 rpm for 50 min at 4°C. The supernatant was incubated at 4°C for 2 h with the Ni^{2+} chelating Sepharose FF beads (GE Health; Cat. No. 17057501). After incubation, beads were washed with an imidazole gradient in the wash buffer (40 mM or 50mM imidazole, 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5% glycerol and 5 mM 2-mercaptoethanol). Proteins bound to the beads were eluted with elution buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 250 mM Imidazole, 5% glycerol and 5 mM 2-mercaptoethanol). Such eluate was then subjected to a second Strep column (IBA; Cat. No. 2-1201-010) purification, and finally bound protein was eluted by 5 mM desthiobiotin in buffer: 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5% glycerol and 5 mM 2-mercaptoethanol. The N-terminal 6xHis-Strep-SUMO-TEV tag was cleaved overnight with TEV in the dialysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol). After dialysis, cleaved tag was removed via second nickel column and flow-through containing the untagged protein was pooled and concentrated. Concentration was done using the Pierce Protein Concentrator (Thermofisher; Cat. No. 88522). Briefly, the columns were filled up with the buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol), spun at 4000xg, then the protein was loaded into the column and spun till the desired volume (500-600 μL) was reached. Proteins were further purified by gel filtration chromatography using Superdex 200 increase 10/300 GL (GE Healthcare) equilibrated with the buffer (25 mM HEPES-KOH pH 7.6, 150 mM KCl, 5% glycerol, 1 mM DTT). The fractions were checked by SDS-PAGE electrophoresis and the pure protein complex (Figure 2A) was flash-frozen in liquid nitrogen after addition of 10% glycerol and stored at −80°C.

To validate the direct interaction between YTHDC2 and hMEIOC_CC, Hi5 cells co-expressing 6xHis-Strep-Sumo-YTHDC2 and hMEIOC_CC-GST were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5% glycerol and 5 mM 2-mercaptoethanol) supplemented with proteinase inhibitor (Thermo Scientific, EDTA-free). Cells were disrupted by sonication and the cell lysate was centrifuged at 18,000 rpm for 50 min at 4°C. The clarified supernatant was incubated with Protino Glutathione Agarose 4B (Macherey-Nagel; Cat. No. 745500.10) at 4°C for 2 h. After incubation, beads were washed with lysis buffer and finally bound protein was eluted

by 15mM L-Glutathione reduced in lysis buffer. The glutathione eluate was checked by SDS-PAGE electrophoresis. Purification of GST-tagged hMEIOC_CC domain co-purified the 6xHis-Strep-Sumo-YTHDC2 (Figure 2B), confirming direct interactions.

YTHDC2 RNA helicase assay

The fluorescence-based helicase assay was modified from (Özeş et al., 2014).

Preparation of RNA loading strand for annealing by in vitro transcription

A 61 nt RNA was generated by *in vitro* transcription using T7 promoter. The reaction was carried out at 37°C for 4 h in a 20 μL reaction containing 6 pmol DNA template (Table S1), 2 μL 10x reaction buffer, 2 μL of 75 mM CTP/UTP/GTP/ATP and 2 μL T7 enzyme mix (MEGAshortscript T7 Transcription Kit, Cat. No. AM1354). Generated RNAs were purified by G-25 column (illustra MicroSpin G-25 micro columns, Cat. No. 27-5325-01) to remove free nucleotides, followed by phenol-chloroform extraction.

Preparation of double-stranded RNA for unwinding experiment

To prepare double-stranded RNA for the unwinding experiment, two RNAs were synthesized (Integrated DNA Technologies): a probe strand carrying 3′ Cy3 fluorophore and a quench strand carrying 5′ BHQ quencher (Figure 3E and Table S1). The annealing reaction containing 1 μL of 10 μM probe strand, 3 μL of 10 μM quench strand, 1 μL of 10 μM loading strand (prepared above), and 45 μL 1x annealing buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl) was carried out at 95°C for 5 min and then slowly cooled down to 25°C in thermomixer (Eppendorf). After that, 1 μL of 100 μM DNA competitor (Table S1) was added into the reaction and incubated on ice until further use.

Fluorescence-based helicase assay

For all experiments, 160 nM protein was mixed with 20 nM annealed RNA in unwinding buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 2 mM DTT) in total 100 μL reaction. The reaction mixture was loaded into 96-well plate (Perkin Elmer, Cat. No. 6005550) and the background fluorescence was measured at 37°C in Spark 10M plate reader (Tecan). The excitation wavelength was set at 540 nm and emission wavelength at 585 nm. The bandwidth for both excitation and emission was 20 nm. The background signal was measured every 15 s for 12 cycles then 100 μL of 4 mM ATP was injected into the well to induce the unwinding reaction. Change of the fluorescence intensity was recorded over time for in total 100 cycles with 30 s interval. The final concentration of protein in the reaction was 80 nM protein, 10 nM dsRNA and 2 mM ATP in unwinding buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 2 mM DTT). All the measurements were set at gain 160 with Z-position at 33327 μm in SPARKCONTROL V2.1. Three individual traces for each reaction were recorded at the same time and the average of fluorescence, together with standard deviation were calculated to make the curves (Figures 3F–3H and S3C–S3H).

ATPase assay

ATP hydrolysis reactions (20 μL) contained the following components: 20 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 2 mM DTT and 10 μCi/mL γ-^{32}P-ATP (Hartmann Analytic; Cat. No. SRP-501). The same dsRNA (Table S1) as in the RNA helicase assay was used at final 10 nM concentration. Wildtype (WT) or catalytic-dead (E317Q) human YTHDC2 (Wojtas et al., 2017) was added at indicated concentrations (Figure 3I and S3I). When needed XRN1^{E178Q}(1-1279) was also included. Reactions were incubated at 37°C for 30 min. The reaction products were spotted on a thin layer chromatography (TLC) plate (Merck; Cat. No.1055790001), air-dried for 3-4 min, and migrated with 0.5 M LiCl, 0.5 M formic acid, in a migration chamber for 40 min. Free phosphate can be distinguished from ATP because it migrates faster during TLC. The plate was then dried at room temperature and exposed to Storage Phosphor screens (GE Health; Cat. No. 28-9564-75) and scanned with a Typhoon FLA 9500 scanner (GE Health; Cat. No. 29-1885-90). Four independent experiments were performed and quantified by ImageQuant TL. The values were normalized to YTHDC2 activity and plotted (Figure 3I). One tailed t.test was used to test whether adding XRN1^{E178Q} increases the activity above 100%.

Electrophoretic Mobility Shift Assay (EMSA)

To examine the RNA-binding property of hYTHDC2 we carried out electrophoretic mobility shift assay (EMSA) (Figure S1G). We first prepared 5′-end labeled single-stranded RNA (Table S1) and purified it by gel-elution after urea-PAGE. Synthetic RNA (20 pmol; Microsynth, CH) was 5′-end labeled with 20 μCi of γ-^{32}P-ATP (Hartmann Analytic; Cat. No. SRP-501) and 10 U of polynucleotide kinase (Thermo Scientific; Cat. No. EK0031) at 37°C for 1 h. The 20 μL reactions were cleared of unincorporated radioactive nucleotides by passing through a G25 microspin column (illustra MicroSpin G-25 micro columns, Cat. No. 27-5325-01), mixed with gel loading dye and separated by 18% urea-PAGE (1xTBE) at constant power of 18 W. The gel was briefly exposed for 10 min to Storage Phosphor screens (GE Health; Cat. No. 28-9564-75) and scanned with a Typhoon FLA 9500 scanner (GE Health; Cat. No. 29-1885-90). The band corresponding to the major radioactive species was cut out and eluted in 300 μL of 300 mM NaCl overnight by mixing (Thermomixer at 750 rpm) at 25°C. Labeled RNA was directly precipitated with 100% cold ethanol and glycogen (1 μL of 20 mg/mL stock) at −20°C for 30 min. After centrifugation, the RNA was resuspended in 20 μL of nuclease-free water.

For forming RNA-protein complexes, we incubated recombinant full-length wild-type human YTHDC2 with radioactive-labeled RNAs in 20 μL reactions containing: 20 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$ and 2 mM DTT. Proteins were used at three different concentrations: 37, 74 and 148 nM. Incubations were carried out on ice for 1 h. Reactions were mixed with 1 μL of loading dye (1x TBE, 10% glycerol, bromophenol blue and xylene cyanol) and separated out by 6% native PAGE (0.25 xTBE). Electrophoresis (running buffer was 0.25 xTBE) was carried out at 3W, 4°C for 1 h till the bromophenol blue dye is 6 cm above the end of the gel. Note that since the theoretical isoelectric point (pI) of full-length human YTHDC2 is 8.68, we adjusted pH of the gel and running buffer to

10.5. This should ensure that the proteins carry an overall negative charge under native conditions. The gel was exposed to Storage Phosphor screens and scanned with a Typhoon FLA 9500 scanner (Figure S1G).

Antibodies

Commercial antibodies

Anti-SYCP3 (Novus Biologicals; NB300-231) to detect the mouse protein. Mouse anti-HA monoclonal was a kind gift of Marc Buhler, FMI, Basel, Switzerland, while rabbit anti-HA (Santa Cruz; sc-805) and monoclonal mouse anti-FLAG (Sigma; Cat. no. F3165) were purchased. Anti-HA affinity matrix (Roche; Cat. No. 11815016001), anti-FLAG M2 Magnetic Beads (Sigma; Cat. No. M8823) were used for immunoprecipitations.

For immunofluorescence studies the following secondary antibodies coupled to fluorescent dyes were used: anti-rabbit (Alexa Fluor 594; Cat. No. A11037), anti-mouse (Alexa Fluor 488; Cat. No. A28175). For western blot analyses the following secondary antibodies conjugated to Horse Radish Peroxidase were used: anti-rabbit IgG HRP-linked (GE Healthcare; NA934), anti-mouse IgG HRP-conjugated (Thermo Fisher; Cat. No. A27025).

iCLIP

Testes from five P15 mice from *Ythdc2*^{FLAG}*; Meioc*^{HA} double homozygous knockin line (Figure 2C) were used for one experiment. The five pairs of P15 testes were digested with collagenase IV (1mg/mL in DMEM, 2ml/pair testis) at 33°C with gentle intermittent mixing in a Thermomixer (Eppendorf). After 15 min, digested testes were removed from thermomixer and gently mixed with a 1ml pipette for several times to get a single-cell suspension. The testes suspension was spun at 300xg for 5 min at 4°C. The medium was removed and the cells were collected in 10 mL cold DMEM. Cells were transferred into a 10 cm^2 Petri dish (cell culture grade) and UV cross-linked with 200mj/cm^2 (254 nm wavelength UV lamps) for 3 times on ice. After UV cross-linking, cells were collected in 15 mL falcon tubes and spun at 500xg for five minutes in at 4°C. The cell pellet was washed twice with ice-cold 1X PBS and snap-frozen in liquid nitrogen and stored at −80°C until use. The remaining procedure for iCLIP was performed as described by (Huppertz et al., 2014). RNase I digestion was done at 1:200 dilution for 3 min at 37°C with 1100 rpm mixing in a Thermomixer (Eppendorf). Anti-FLAG M2 Magnetic Beads (Sigma; Cat. No. M8823) (100μl/reaction) were used for immunoprecipitation. The final PCR amplification was performed with Platinum II Hot-Start PCR Master Mix (ThermoFisherScientific; Cat. No. 14000012) for 30 cycles. The PCR product was precipitated with sodium acetate and ethanol, separated on 3% MetaPhor agarose (Lonza; Cat. No. 50180) in 1xTAE buffer. The DNA band corresponding to 100-200 bp was cut from the gel and purified using MiniElute kit (QIAGEN; Cat. No. 28004). Purified iCLIP libraries were sequenced on NextSeq 500 illumina sequencer for 75 cycles single end. The data for the iCLIP analysis is from two biological replicates (Figure S1B).

Isolation of endogenous YTHDC2 and MEIOC complexes from mouse testes

Testes from P20 *Ythdc2*^{FLAG}*; Meioc*^{HA} double homozygous knockin line (Figure 2C) were used. We used eight testes for each experiment. Testes were lysed in 3.0ml lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% glycerol, 5 mM EDTA pH 8.0, 0.5 mM DTT, 0.1% NP-40 and EDTA-free protease inhibitor from ThermoScientific) using a glass homogenizer, on ice. The lysate was cleared by centrifugation at 20000xg for first 10 min at 4°C. The supernatant was collected in a fresh tube and centrifuged again for 20 min at 20000xg 4°C. A portion (5%) of supernatant was saved as input, with the remaining supernatant being divided into two parts. One part was supplemented with 200 μl of 250 μg/mL 3xFLAG peptide (Sigma) and serves as the control. We used 100 μl per portion of anti-FLAG magnetic beads (Sigma; Cat. No. M8823), which were washed three times with 1ml lysis buffer. The lysate was then added to the washed beads and incubated at 4°C for 3 h with continuous rotation (10 rpm). After the incubation, beads were washed five times with lysis buffer (5 min for each wash at 4°C with gentle mixing). During last wash the beads were transferred to a fresh tube. The protein complexes were eluted twice with 150 μl 3xFLAG peptide (Sigma) at 37°C, 1000 rpm for 15 min in a Thermomixer (Eppendorf). The eluted complexes from two elutions were pooled. An aliquot of 20 μL from the pooled elutions was tested by PAGE and Silver staining. The experiments were performed in biological triplicates. In a separate experiment, the lysate was supplemented with RNaseA/T1 to degrade all the RNA before purifying the protein complexes (Figure 2D).

Purification of MEIOC-associated proteins was performed similarly using anti-HA affinity matrix (Roche; Cat. No. 11815016001) from testes lysate of either P20 wild type (control IP) or from P20 *Ythdc2*^{FLAG}*; Meioc*^{HA} double homozygous knockin line (Figure 2C). The lysate was supplemented with RNaseA/T1 to degrade RNA before purifying the protein complexes (Figure 2D). The MEIOC-associated proteins were eluted by boiling the beads with SDS sample loading buffer. The experiments were performed in biological triplicates.

The eluted complexes were processed at the Functional Genomics Center Zurich (ETH Zurich) using shotgun mass spectrometry analysis. Database searches were performed using the Mascot (SwissProt, mouse) search program. Applied settings: 1% protein false detection rate (FDR), min. 1 peptides per protein, 0.1% peptide FDR. To search for the proteins associated with YTHDC2-3xFLAG and 2xHA-MEIOC in mouse P20 testes, the peptide counts obtained by immunoprecipitation and mass spectrometry (IP-MS) were assigned to genes and processed using edgeR (Robinson et al., 2010). The counts were compared between the positive and control samples. For YTHDC2-3XFLAG complex purifications the control is the P20 knockin testes lysate pre-incubated with the FLAG peptide, while for 2xHA-MEIOC complex purification, the control is P20 wild-type mouse testes. The obtained log2FC values and associated adjusted p values were used to construct the volcano plots (Figure 2D). The genes with logFC > = 1 and adj. p value <= 0.05 were considered to be significantly enriched. Venn diagrams compare the enriched genes between different

immunoprecipitations (Figure S2C). The enrichment of GO Biological Process terms associated with the enriched genes was analyzed (Figure S2D) using ENRICHR.

Tissue lysates

Tissues were extracted from euthanized mice, flash-frozen in liquid nitrogen, and kept at −80°C until extract preparation. Tissues were lysed using a cell disruptor (Janke & Kunkel; IKA-WERK) in 1 mL of lysis buffer [10% glycerol, 50 mM Tris-HCl, pH8.0, 150 mM NaCl, 5 mM MgCl2, 0.5% sodium deoxycholate, 1% Triton X-100, complete protease inhibitor cocktail (Roche; Cat. No. 5056489001)]. After disruption, SDS was added to a final of 2% and gently sonicated for 10 s to breakdown genomic DNA. Aliquots were frozen in −80°C and separated by 8% SDS-PAGE, transferred to a nitrocellulose membrane and probed with antibodies.

Western Blot

Whole cell lysates were separated on SDS-PAGE gels in order to detect proteins of interest. In short, SDS-PAGE gels were prepared using Ultra-Pure ProtoGel 30% acrylamide (37.5:1) (National Diagnostic; Cat. No. EC-890) mixed with ultra-pure water and resolving gel buffer to obtain 8% resolving gel (0.375 M Tris, 0.1% SDS, pH 8.8), and with stacking gel buffer to obtain 8% stacking gel (0.125 M Tris, 0.1% SDS, pH 6.8). N,N,N',N'-Tetramethylethylendiamin (Merck; Cat. No. 1107320100) and 10% ammonium persulfate were added to catalyze the polymerization reaction. Gel electrophoresis was performed at 90 V for 30 min. and then at 120 V for 90 min. After separation, proteins were blotted on the Amersham Protran 0.45 μm nitrocellulose membrane (GE Healthcare; Cat. No. 10600002) overnight at 5 V at room temperature using Trans-Blot SD. Semi-Dry Transfer Cell system (Bio-Rad; Cat. No. 1703940). After transfer, membranes were washed with Tris-buffered saline (TBS, 20 mM Tris, 150 mM NaCl, pH 7.6) and blocked for 1 h at room temperature with 5% dry milk in TBS with 0.05% Tween20 (TTBS) (Sigma; Cat. No. P7949). After 1 h, membranes were incubated with primary antibody for 1 h at RT in 5% milk with TTBS. Then, membranes were washed 5 times for 5 min with TTBS and incubated with HRP-conjugated secondary antibody at 1:10 000 dilution, with either anti-rabbit IgG HRP-linked (GE Healthcare; Cat. No. NA934) or HRP-conjugated anti-mouse IgG (ThermoFischer; Cat. No. A27025) for 1 h at RT in 5% milk in TTBS. After 1 h, membranes were washed 5 times for 5 min with TTBS followed by 3 washes for 5 min with TBS and incubated with one of the detection reagents: Amersham Prime Western Blotting Detection Reagent (GE Healthcare; Cat. No. RPN2232), SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher; Cat. No. 34095) or Pierce ECL 2 Substrate (ThermoFisher; Cat. No. 1896433A) for 5 min. at room temperature. Signal was detected using Amersham Hyperfilm ECL (GE Healthcare; Cat. No. 28906837). The processed films were scanned using Perfection 3200 Photo scanner (Epson) with XSane image scanning software (ver. 0999).

Histology and Immunofluorescence

Histology of mouse tissue sections

To prepare the paraffin sections, isolated testes or ovaries were fixed in Bouin's solution (Sigma; Cat. No. HT10132) overnight at room temperature. Samples were transferred into the embedding cassettes (Simport; Cat. No. M508-3) and subsequently washed for 1 h in tap water and then transferred to 1x PBS. Specimens were sent to the histology platform (CMU, University of Geneva) where they were dehydrated in 70% (2 × 2 h), 90% (1 h), 95% (1 h) and 100% ethanol (3 × 30 min) followed by incubation in HistoSAV-xylolersatz (3 times 30 min; Biosystems). The solution was removed and replaced with paraffin, and incubated at 56-58°C. Testes were then transferred into plastic molds (Polysciences mold S-22; Cat. No. NC0397999) filled with paraffin, and paraffin was allowed to solidify at room temperature. The sections (~5 μm thickness) were cut using microtome (Leica RM2135) and mounted on the Superfrost Plus microscope slides (Thermo Fisher; Cat. No. 4951PLUS4). The sections were allowed to stretch for 24 h at 42°C and then were stored at room temperature.

For histological analysis, the slides containing the paraffin sections were placed in a glass slide holder filled with HistoSAV (3 × 5 min) to remove the paraffin. For rehydration, the slides were incubated in 3x 100% ethanol, 96% ethanol, 70% ethanol, 50% ethanol and water (3 min for each step). Sections were stained with Hematoxylin solution (Merck) for 3-5 min and rinsed in running tap water. Then, sections were stained with Eosin Y solution (Sigma Aldrich; Cat. No. E4382) for 3 to 5 min and washed with water. For dehydration, the sections were incubated in 50% (30 s), 70% (30 s), 96% (30 s), 100% ethanol (2 min) and HistoSAV (3 × 3 min). Few drops of Neo-Mount (Merck) were deposited on the sections and immediately covered with coverslips. Pictures were taken using microscope AXIO Imager M2 (Zeiss).

Immunofluorescence of mouse tissue sections

Collected testes were washed in PBS and immediately fixed in 10 mL of 2% paraformaldehyde at 4°C for 3 h on a rotating wheel. Tissues were washed twice in 1x PBS and dehydrated in 15% sucrose in 1x PBS for almost 3 h (till the testes/ovaries sink to the bottom of the falcon tube). After a further dehydration step in 30% sucrose overnight, tissues were embedded in a cryomold (Tissue-Tek) filled with Andwin Scientific Tissue-Tek CRYO-OCT Compound (Fisher Scientific; Cat. No. 14-373-65) and frozen on dry ice. Embedded tissues were cut at histology platform (CMU, University of Geneva) using cryostat (Leica CM3050). The 7 μm tissue sections were mounted on Superfrost Plus glass slides and stored at −80°C. For immunofluorescence experiments, sections were allowed to dry at room temperature for 30 min and fixed in cold 4% paraformaldehyde in PBS (on ice) for 10 min. Slides were then washed in PBS at RT (2 × 5 min), and once in distilled water (5 min). Next, antigen retrieval was performed with Heat-Induced Epitope Retrieval (HIER). Briefly, slides were submerged in 600 mL of 10 mM Citrate Buffer pH 6.0 (2.1 g citric acid, 1.0 g NaOH pellet dissolved in 1000 mL MiliQ water) and heated in a microwave at full power (600 W) for 20 min. Tissues were allowed to cool down at room

temperature for at least 45 min, washed in PBS and permeabilized in 0.3% Triton X-100 in PBS at RT for 10 min. Slides were washed twice in TBS-T (TBS; 20 mM Tris, 150 mM NaCL, pH 7.6, and 0.1% Tween20) and blocked for minimum 30 min at room temperature in a humidified chamber in blocking buffer (5% normal goat serum in TBS-T). Primary antibodies were diluted in blocking buffer at different concentrations (see below) and incubated overnight at 4°C. Next day, slides were washed twice in TBS-T, incubated with secondary antibody diluted in blocking buffer (anti-mouse or anti-rabbit conjugated to Alexa 488, or 594 fluorophore) in a humidified chamber for 45-60 min (dilution 1:1000). Slides were washed twice in TBS-T and incubated with DAPI (0.5 μg/mL, Bio-Rad; Cat. No. 10043282) for 5-15 min to counterstain the nuclei. Sections were finally washed twice in TBS-T, once in ddH_2O and mounted with Slowfade Gold Antifade Reagent (Life technologies; Cat. No. S36942). Pictures were taken using Zeiss LSM780 confocal microsocope (Bioimaging Center, University of Geneva).

Primary antibodies concentrations: anti-SCP3 1:200 (Novus Biologicals; NB300-231).

Histology of zebrafish tissue sections

6-month-old adult male zebrafish were euthanized and fixed whole in 4% formal/PBS for 3 days at 4°C. Samples were mounted into histology cassettes and rinsed with water prior to dehydration, paraffin embedding and sectioning. Sections were mounted onto microscope slides and stained with hematoxylin/eosin. The latter steps were performed by the University of Geneva, faculty of medicine histology platform. Samples were imaged using a Zeiss AxioCam instrument.

FACS purification of mouse germ cells for RNaseq and single-cell RNA sequencing

Male pups (P10) or adult (P60) mice were euthanized for testes collection. Testes were removed and placed in a Petri dish (3 cm) with PBS and transferred in another Petri dish containing 1.5 mL DMEM supplemented with 1mg/mL type IV collagenase (Sigma; Cat. No. C5138). All extraneous tissue and the tunica are removed, and the seminiferous tubules teased apart. The samples are transferred into a 2ml Eppendorf and put at 35°C in a water bath for 15 min. The tubes were regularly inverted for mixing. After centrifugation for 5 min at 278x g (1610 rpm in Mikro22R) testes are resuspended in pre-warmed 500 μL 0.125% Trypsin and incubated for 6 min at 35°C. The Tryptic digest was stopped by addition of DMEM with 10% FBS. The solution is mixed gently into a suspension by pipetting at least 10 times. Centrifuge for 5 min. at 200xg (1360 rpm in Mikro22R). Resuspend the cells with 1ml or more of FACS buffer (1x PBS with 1% BSA). Filter with a 100 μm cup Filcon (BD; Cat. No. 340639). Check the viability with trypan blue (stains only dead cells), count and dilute in FACS buffer at about 4x106 cell/mL. Put on ice and move to FACS instrument.

We used Hoechst or Drak5 to stain DNA just before sorting (9 μL of Drak5 for 3.9x106 cell/mL) to isolate 1n, 2n and 4n cells. The 2n and 4n testicular germ cells from adult (P60) *Ythdc2*^{YTH} homozygous and heterozygous mice were used for RNasequencing (Figure 1M). Biological triplicates were used.

For single-cell sequencing, we used purified 4n germ cells from P10 males of heterozygous and homozygous *Ythdc2* KO aninmals (Figure 5A). Biological duplicates were used.

For *Ythdc2*$^{cat\text{-}dead}$ mutant quadruplicate P14 testes were used.

Preparation of RNA libraries

Sequencing of testicular RNA

Total RNA was isolated using the Trizol reagent (ThermoFisher Scientific; Cat. No. 15596026) from the required mouse testes samples (whole testes or FACS purified germ cells). RNA concentration was measured with a Qubit fluorimeter (Life Technologies) and RNA integrity assessed with a Bioanalyzer (Agilent Technologies). The TruSeq Stranded Total RNA kit with Ribo-Zero Gold was used for library preparation with 500 ng of total RNA as input. Library molarity and quality was assessed with the Qubit and Tapestation using a DNA High sensitivity chip (Agilent Technologies). Libraries were diluted at 2 nM and pooled before the clustering process on a HiSeq 4000 Single Read flow cell. Reads of 50 bases were generated using the TruSeq SBS reagents on the Illumina HiSeq 4000 sequencer (iGE3 Genomics Platform, University of Geneva).

Single cell sequencing

Mouse testes from *Ythdc2* heterozygous (HET) and homozygous *Ythdc2* KO P10 mice (Wojtas et al., 2017) were collected and single-cell suspensions prepared for FACS purification of 4n germ cells, as described above. We pooled cells from two individual mice of the same genotype together to obtain sufficient number of cells. The cells were captured on the 10x Genomics Chromium system. The Chromium Single Cell 3′ Reagent Kit version 2 kit was used for library preparation according to manufacturer's specifications. Library molarity and quality was assessed with the Qubit and Tapestation and loaded on a Paired-end 100 flow cell. Thus, each dataset consists of cells from two P10 mice (four pairs of testes). For *Ythdc2* HET we used one dataset, while for *Ythdc2* KO two datasets were used.

QUANTIFICATION AND STATISTICAL ANALYSIS

iCLIP

Two replicate YTHDC2 iCLIP datasets from P15 mice were used in this analysis (Figure S1B). The analysis strategy was mainly based on the previously published approach (Busch et al., 2020). Quality of the library was checked with FastQC v0.11.5 and adapters were removed with Flexbar v3.5.0 and SeqAn v2.4.0 (Roehr et al., 2017). Trimmed reads were aligned to the mouse genome reference (GRCm38) by STAR v2.7.9a (Dobin et al., 2013) with the following iCLIP-specific parameters:–outFilterMismatchNoverReadLmax

0.04–outFilterMismatchNmax 999–outFilterMultimapNmax 1–alignEndsType Extend5pOfRead1. Generated BAM files were deduplicated with umi_tools dedup v1.1.1 (Smith et al., 2017). The BAM files were merged together with samtools merge v1.7 (Li et al., 2009) and PureCLIP v1.3.1 (Krakau et al., 2017) was used to identify binding sites from the merged BAM file. The binding sites BED file generated by the PureCLIP algorithm was further annotated by intersecting binding site coordinates with the gene coordinates obtained from the GRCm38 genome annotation file (Ensembl release 102) by using bedtools intersect. Pie charts (Figure 1B) were produced with matplotlib pie function by intersecting binding sites genomic coordinates with either 5′UTR/CDS/3′UTR/intron genomic coordinates, or with gene coordinates with the regard to gene biotype. *Cul4a* gene coverage track (Figure 1B) was plotted using SparK v2.6.2 (Kurtenbach and Harbour, 2019). Heatmap of z-scores (Figure 1C) and metaplot (Figure 1D) of the average binned coverage across all genes encoding YTHDC2-bound transcripts (transcript with the highest number of exons was considered as the canonical isoform) were produced based on the average coverage per bin of the binned transcript areas (5′ UTR = 50 bins, CDS = 200 bins, 3′ UTR = 100 bins) with either matplotlib bar or seaborn clustermap functions, respectively. Total number of annotated binding sites was taken to produce scatterplots (Figures 1E and 1F) with the matplotlib scatter function. Binding site coordinates were extended to 5-mers and their sequences were taken from the GRCm38 genome reference file (Ensembl release 102) both for 3′ UTR- and CDS-specific sites, and their frequencies were compared (Figure 1F). Total number of 5-mer sequences was overlapped, and position-specific nucleotide frequencies were computed to make a sequence logo with logomaker python package. Highly confident binding sites were retained based on the binding site coverage criteria ($0 < \log2(\text{BS coverage}) < 4.5$) to plot the histogram of the total number of binding sites per YTHDC2 binding target (Figure 1G). GO term analysis of the highly confident YTHDC2 binding targets was performed with ENRICHR (Chen et al., 2013; Kuleshov et al., 2016) (Figure 1H). Total number of binding sites was taken to perform the comparative analysis of the UUUU-stretch frequencies versus DRACH motif frequencies (Figure 1J). Motif frequencies (Figures 1J, S1E, and S1F) were computed by using sliding window approach with the window size of 5 and step = 1 in case of non-match, otherwise step = 5 (motif length). Frequencies were normalized per total number of motifs identified at a certain position.

Analysis of RNA-seq data

Quality control of the demultiplexed libraries was performed with FastQC. Sequencing adapters were removed with Flexbar software. For the analysis of the catalytic dead mutant (Figures 3A and S3B), reads were aligned to the mouse transcriptome (GRCm38 assembly) with kallisto quant RNA-seq quantification tool v0.46.2 (Bray et al., 2016). All the other RNA-seq datasets were aligned to the transcriptome reference using salmon quant v1.3.0 (Patro et al., 2017). Transcript counts were collapsed into gene counts and DESeq function of the DESeq2 bioconductor package (Huber et al., 2015; Love et al., 2014) was used to obtain shrunken log2 fold changes of gene expression between control and mutant samples, and the adjusted p values. Adjusted p value 0.05, as well as the absolute value of log2 fold change 1.5 were defined as thresholds of statistical significance. All the visualizations were done in Python 3.6.9. Volcano plot of differential gene expression in the *Ythdc2*^{YTH} mutant was plotted using scatterplot function of the seaborn package (v0.10.0) with additional matplotlib (v3.1.3) customizations (Figure 3A). Heatmap of the differentially expressed genes in different stages of spermatogenesis was plotted with clustermap function with z_score and weighted hierarchical clustering options of the seaborn package (Figure 3B). Genes found to be significantly up- or downregulated in the mutant were searched for enriched Gene Ontology terms in the Biological Process ontology using ENRICHR (Chen et al., 2013; Kuleshov et al., 2016) and plotted (Figure S3B). RNA sequencing data from different purified spermatogenic populations- Spermatogonia, spermatocytes, spermatids, and spermatozoa (Figure 3B)- were obtained from (Soumillon et al., 2013), and processed according to the aforementioned procedure.

Analysis of single cell sequencing data

The whole single-cell RNaseq (scRNA-seq) analysis strategy was based on a previously published approach Luecken and Theis, 2019. Demultiplexed raw sequencing data were processed by cellranger pipeline (v3.0.2) developed by 10x Genomics. Reads were aligned to the custom mm10 mouse genome reference by cellranger count. Cell count matrix, gene and barcode tables were taken as final output for further processing. Scanpy scRNA-seq toolkit v1.4.5 was mainly used to integrate, normalize, and process the data. Once integrated, scRNA-seq dataset was filtered based on the following criteria: 1) transcript counts per cell (cells containing less than 1500 or more than 40000 transcripts were discarded) using filter_cells function; 2) genes expressed per cell (cells with less than 2000 genes expressed were discarded) using filter_cells function; 3) mitochondrial gene fraction (cells with more that 20% of mt genes per total number of genes expressed were discarded); 4) gene occurrence (genes which were found in less than 20 cells were discarded) with filter_genes function. Finally, 4528 cells were retained for further analysis (1151 cells of *Ythdc2*$^{+/-}$ and 3377of *Ythdc2*$^{-/-}$ genotypes). Scran bioconductor package (Lun et al., 2016) was used for the normalization with default parameters on count matrices before constructing single-cell maps. To reduce the dimensionality of the data and construct single-cell maps UMAP algorithm was applied to the normalized count matrices based on 3000 highly variable genes identified (Figure 4). Batch correction was done by MNN algorithm with mnnpy package. Leiden clustering algorithm (leiden function of the scanpy package) was used to identify cell populations with further cluster annotation based on scRNA-seq atlas of the spermatogenesis (Ernst et al., 2019). MAST algorithm was used to perform differential gene expression analysis (diffxpy.api python package). We gratefully acknowledge the extensive use of R (R Core Team, 2018) and BLAST (Altschul et al., 1997) in this study.

Molecular Cell, Volume *82*

Supplemental information

The XRN1-regulated RNA helicase activity of YTHDC2 ensures mouse fertility independently of m^6A recognition

Lingyun Li, Kyrylo Krasnykov, David Homolka, Pascal Gos, Mateusz Mendel, Richard J. Fish, Radha Raman Pandey, and Ramesh S. Pillai

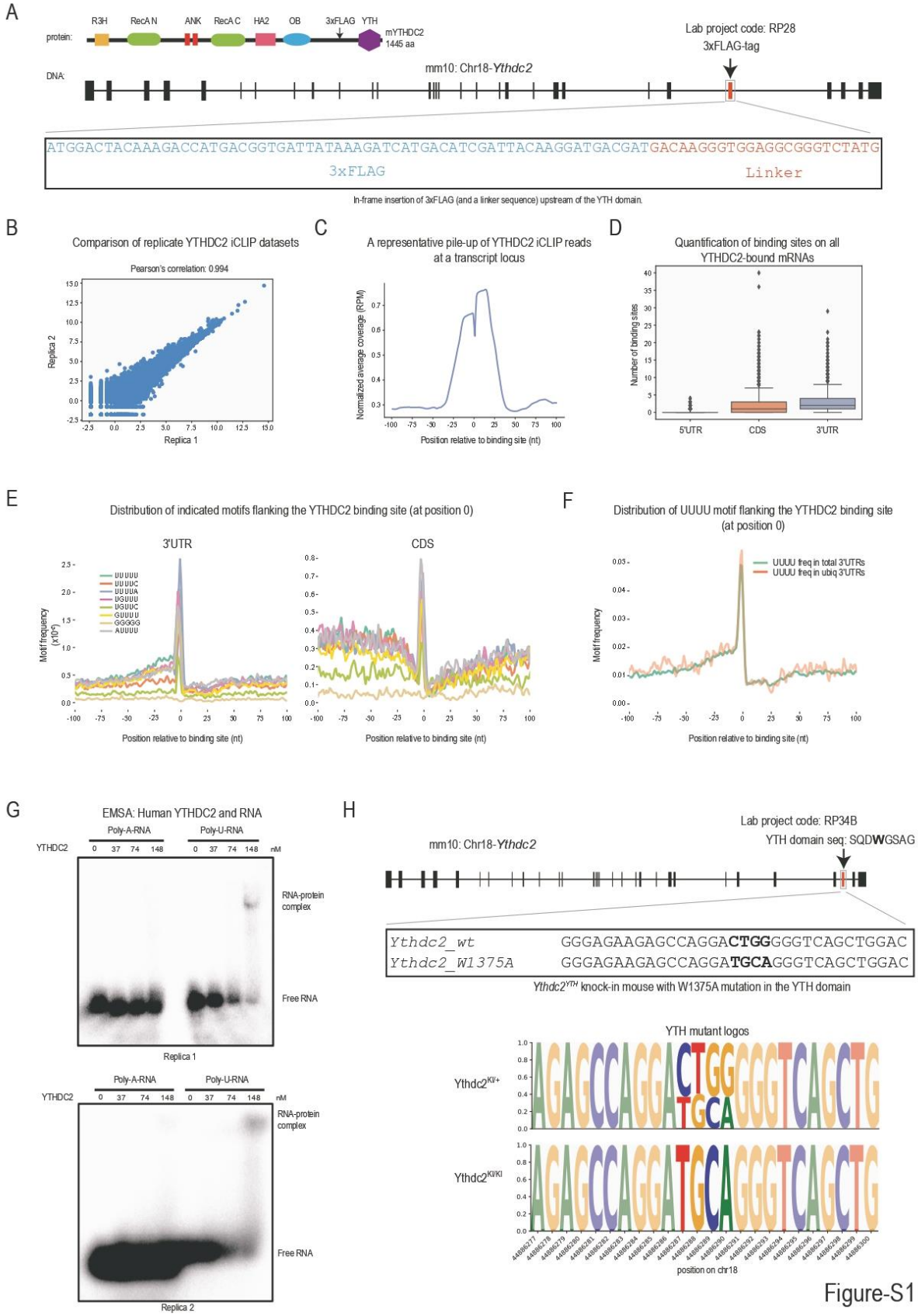

A
protein:
R3H
RecA N
ANK
RecA C
HA2
OB
3xFLAG
YTH
mYTHDC2
1445 aa
DNA:
mm10: Chr18-Ythdc2
Lab project code: RP28
3xFLAG-tag
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGGTGGAGGCGGGTCTATG
3xFLAG
Linker
In-frame insertion of 3xFLAG (and a linker sequence) upstream of the YTH domain.
B
Comparison of replicate YTHDC2 iCLIP datasets
Pearson's correlation: 0.994
Replica 2
Replica 1
C
A representative pile-up of YTHDC2 iCLIP reads at a transcript locus
Normalized average coverage (RPM)
Position relative to binding site (nt)
D
Quantification of binding sites on all YTHDC2-bound mRNAs
Number of binding sites
5'UTR
CDS
3'UTR
E
Distribution of indicated motifs flanking the YTHDC2 binding site (at position 0)
3'UTR
CDS
Motif frequency (x10^3)
UUUUU
UUUUC
UUUUA
UGUUU
UGUUC
GUUUU
GGGGG
AUUUU
Position relative to binding site (nt)
F
Distribution of UUUU motif flanking the YTHDC2 binding site (at position 0)
UUUU freq in total 3'UTRs
UUUU freq in ubiq 3'UTRs
Motif frequency
Position relative to binding site (nt)
G
EMSA: Human YTHDC2 and RNA
Poly-A-RNA
Poly-U-RNA
YTHDC2
0 37 74 148 0 37 74 148 nM
RNA-protein complex
Free RNA
Replica 1
Replica 2
H
mm10: Chr18-Ythdc2
Lab project code: RP34B
YTH domain seq: SQDWGSAG
Ythdc2_wt GGGAGAAGAGCCAGGACTGGGGGTCAGCTGGAC
Ythdc2_W1375A GGGAGAAGAGCCAGGATGCAGGGTCAGCTGGAC
Ythdc2^YTH knock-in mouse with W1375A mutation in the YTH domain
YTH mutant logos
Ythdc2^KI/+
Ythdc2^KI/KI
position on chr18

Figure-S1

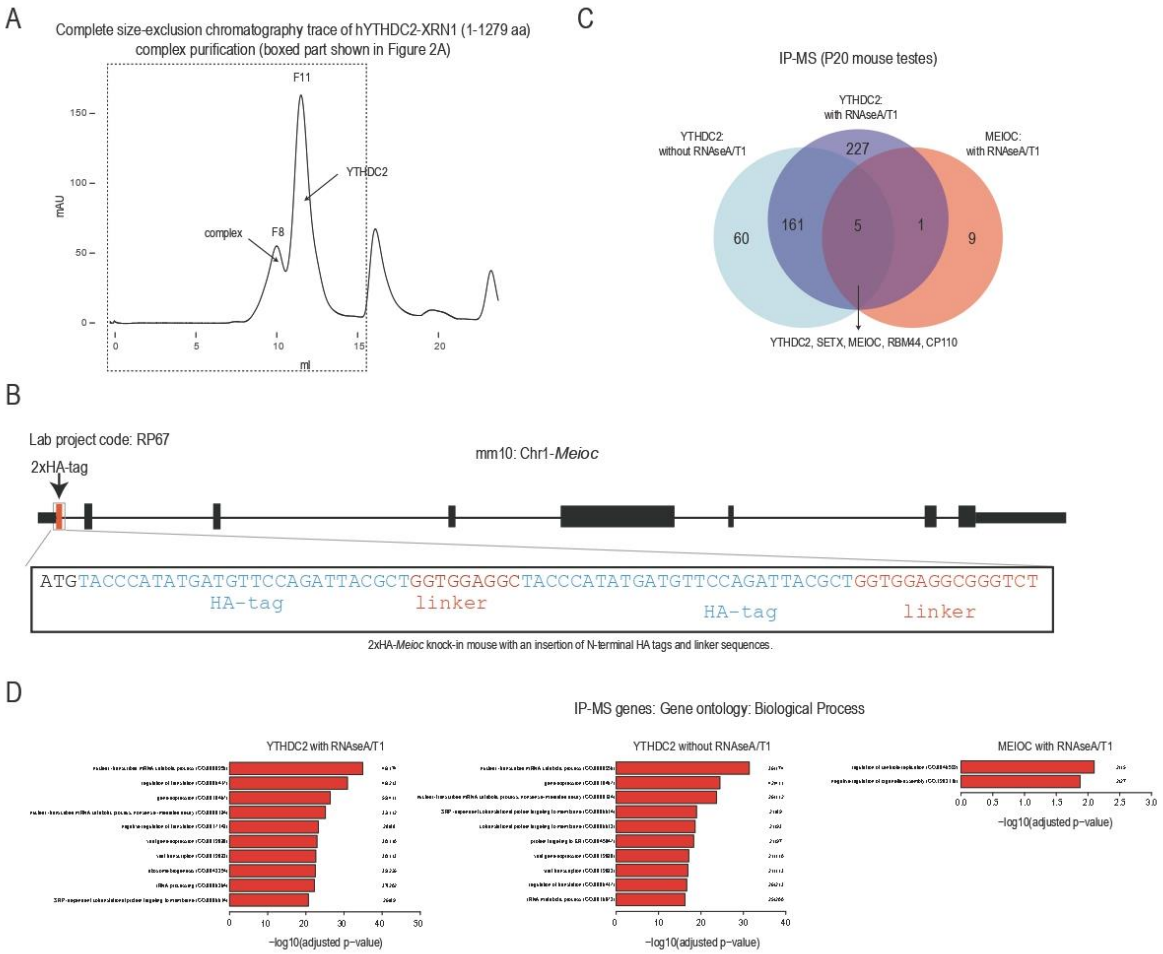

A
Complete size-exclusion chromatography trace of hYTHDC2-XRN1 (1-1279 aa)
complex purification (boxed part shown in Figure 2A)
F11
YTHDC2
complex
F8
mAU
ml
C
IP-MS (P20 mouse testes)
YTHDC2:
with RNAseA/T1
YTHDC2:
without RNAseA/T1
MEIOC:
with RNAseA/T1
227
60
161
5
1
9
YTHDC2, SETX, MEIOC, RBM44, CP110
B
Lab project code: RP67
2xHA-tag
mm10: Chr1-Meioc
ATGTACCCATATGATGTTCCAGATTACGCTGGTGGAGGCTACCCATATGATGTTCCAGATTACGCTGGTGGAGGCGGGTCT
HA-tag
linker
HA-tag
linker
2xHA-Meioc knock-in mouse with an insertion of N-terminal HA tags and linker sequences.
D
IP-MS genes: Gene ontology: Biological Process
YTHDC2 with RNAseA/T1
YTHDC2 without RNAseA/T1
MEIOC with RNAseA/T1
−log10(adjusted p−value)

Figure-S2

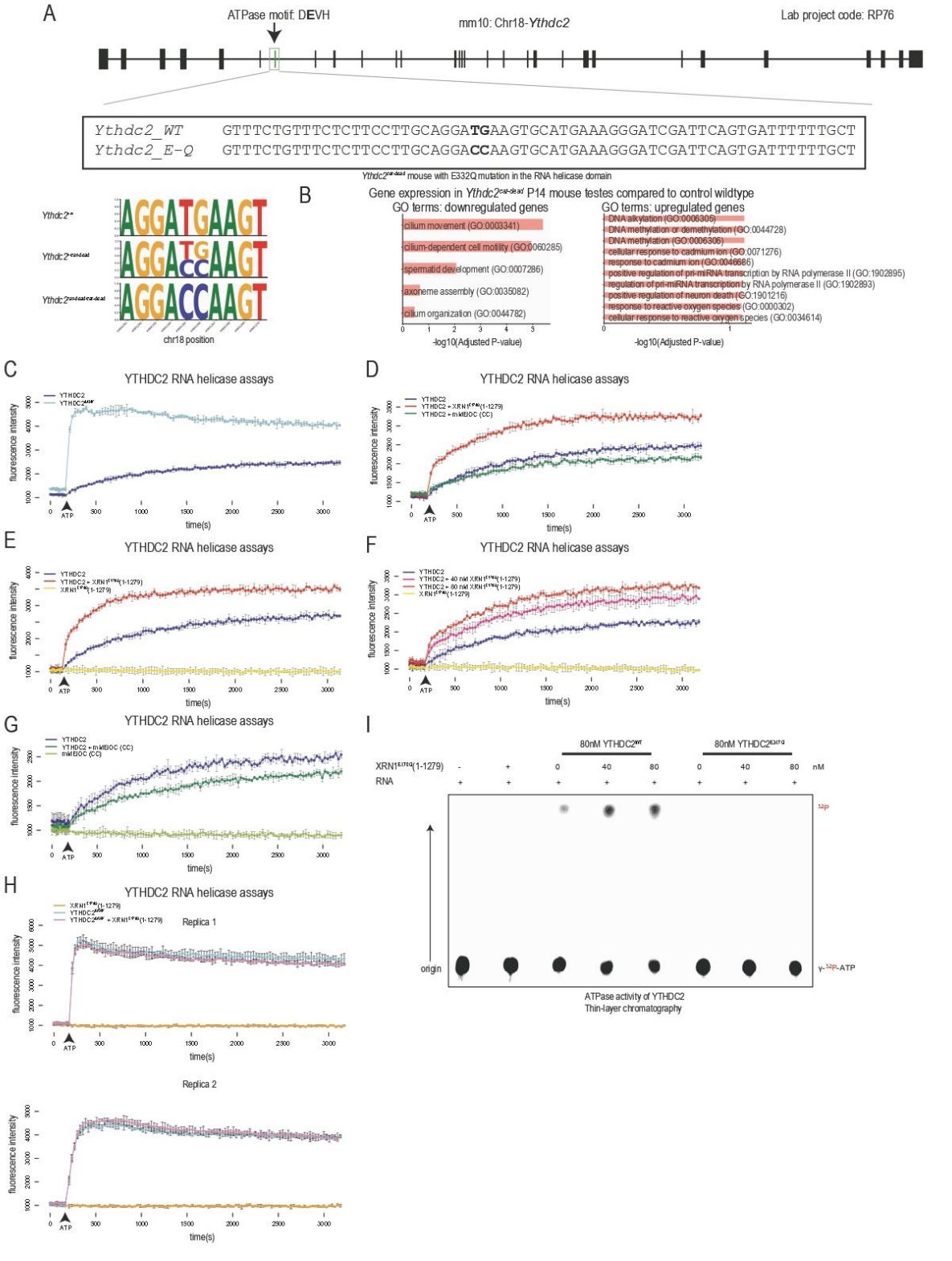

A
ATPase motif: DEVH
mm10: Chr18-Ythdc2
Lab project code: RP76
Ythdc2_WT GTTTCTGTTTCTCTTCCTTGCAGGATGAAGTGCATGAAAGGGATCGATTCAGTGATTTTTTGCT
Ythdc2_E-Q GTTTCTGTTTCTCTTCCTTGCAGGACCAAGTGCATGAAAGGGATCGATTCAGTGATTTTTTGCT
Ythdc2 mouse with E332Q mutation in the RNA helicase domain
B
Gene expression in Ythdc2 P14 mouse testes compared to control wildtype
GO terms: downregulated genes
cilium movement (GO:0003341)
cilium-dependent cell motility (GO:0060285)
spermatid development (GO:0007286)
axoneme assembly (GO:0035082)
cilium organization (GO:0044782)
-log10(Adjusted P-value)
GO terms: upregulated genes
DNA alkylation (GO:0006305)
DNA methylation or demethylation (GO:0044728)
DNA methylation (GO:0006306)
cellular response to cadmium ion (GO:0071276)
response to cadmium ion (GO:0046686)
positive regulation of pri-miRNA transcription by RNA polymerase II (GO:1902895)
regulation of pri-miRNA transcription by RNA polymerase II (GO:1902893)
positive regulation of neuron death (GO:1901216)
response to reactive oxygen species (GO:0000302)
cellular response to reactive oxygen species (GO:0034614)
-log10(Adjusted P-value)
chr18 position
C
YTHDC2 RNA helicase assays
D
YTHDC2 RNA helicase assays
E
YTHDC2 RNA helicase assays
F
YTHDC2 RNA helicase assays
G
YTHDC2 RNA helicase assays
H
YTHDC2 RNA helicase assays
Replica 1
Replica 2
I
80nM YTHDC2WT
80nM YTHDC2E378Q
XRN1E378Q(1-1279)
RNA
nM
origin
ATPase activity of YTHDC2
Thin-layer chromatography

Figure-S3

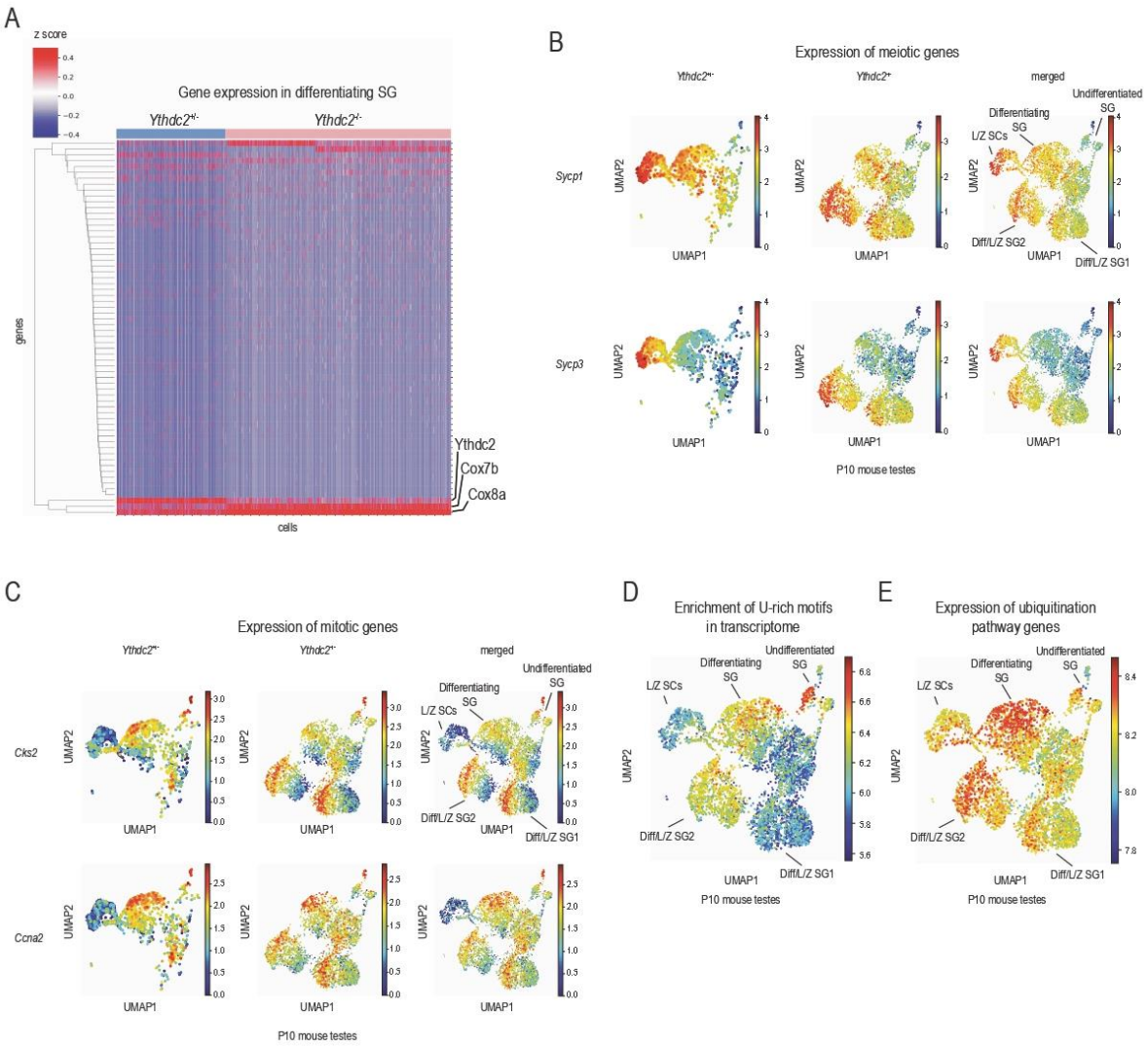
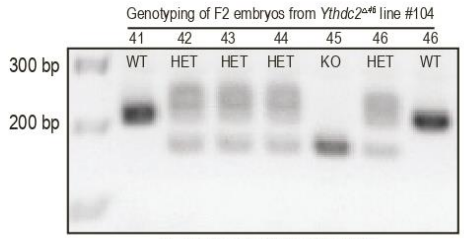

A
z score
Gene expression in differentiating SG
Ythdc2+/-
Ythdc2-/-
genes
cells
Ythdc2
Cox7b
Cox8a
B
Expression of meiotic genes
Ythdc2+/+
Ythdc2-/-
merged
Sycp1
Sycp3
UMAP1
UMAP2
Undifferentiated SG
Differentiating SG
L/Z SCs
Diff/L/Z SG2
Diff/L/Z SG1
P10 mouse testes
C
Expression of mitotic genes
Cks2
Ccna2
D
Enrichment of U-rich motifs in transcriptome
E
Expression of ubiquitination pathway genes

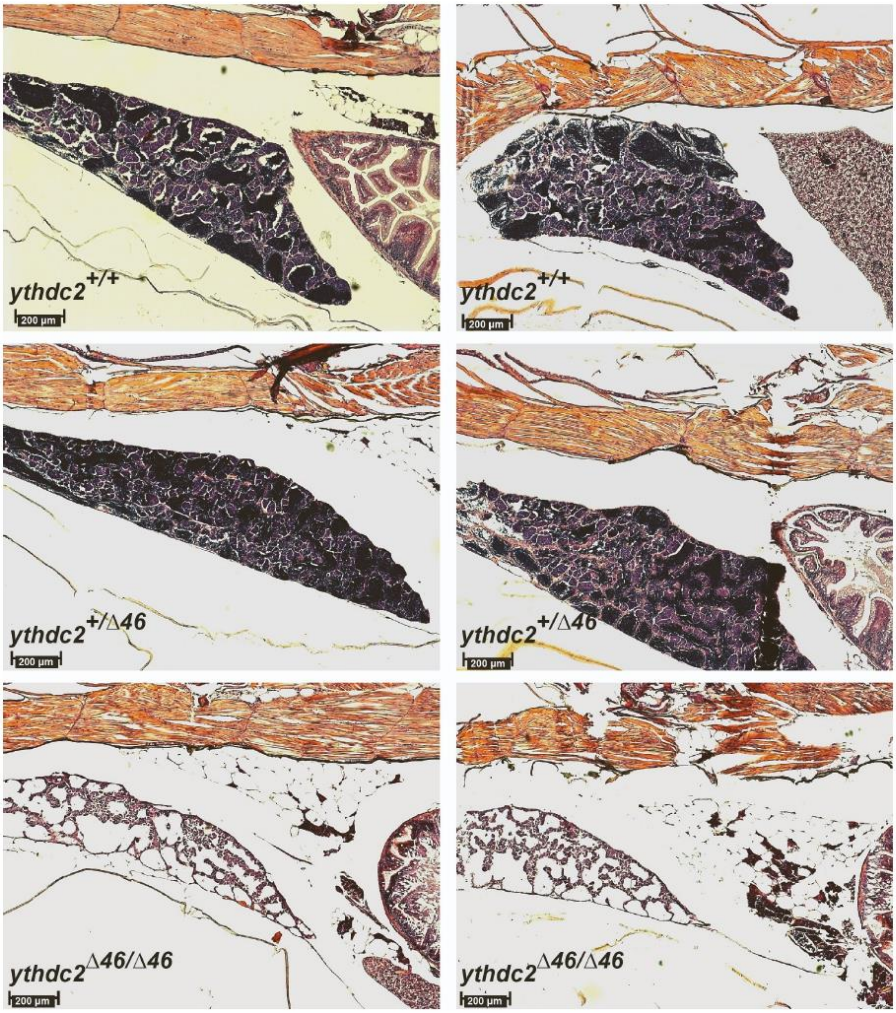
Figure-S4

A



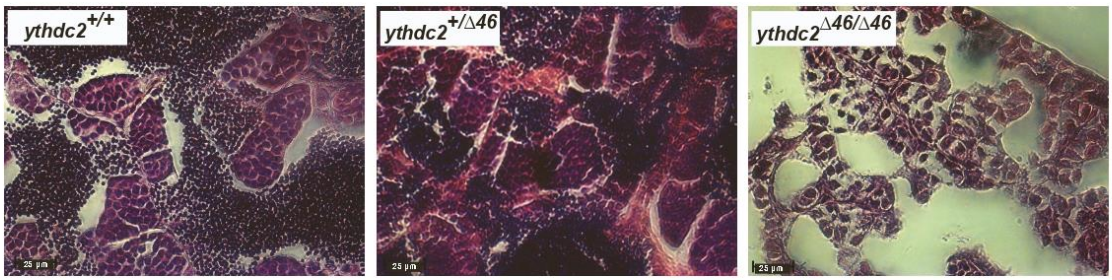
2% agarose gel, 1xTBE, 90V, 55 min.

B



Magnification: 5x Adult zebrafish testis Hematoxylin and eosin staining

C



Magnification: 40X Adult zebrafish testis Hematoxylin and eosin staining

Figure-S5

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Recognition of m^{6}A is not critical for in vivo function of YTHDC2. Related to Figure 1.

(A) Creation of the *Ythdc2*^{FLAG} knock-in mouse model. The 3xFLAG tag sequence is inserted into a surface-accessible loop (determined using limited proteolysis data, also confirmed by AlphaFold) between the OB-fold and the YTH domain. Sequence of the inserted tag/linker is shown. (B) Scatterplot of the log2-transformed counts from the two iCLIP libraries. Pearson's correlation analysis of the replicate similarity is shown. (C) Testes from *Ythdc2*^{FLAG} mice (P15) were used to perform iCLIP experiment. Normalized iCLIP read metaprofile coverage over YTHDC2 binding sites (position 0) shown (D) Box plots showing the distributions of the number of binding sites in the indicated transcript regions. (E) Distribution frequency of the top seven U-rich pentamers centred on the YTHDC2 crosslink site (a uridine at position 0) in a 200 nt widow. The distribution within the 3′ UTR and CDS of YTHDC2-bound RNA targets is shown. A negative control pentamer (GGGGG) is not enriched at the crosslinked site. (F) Distribution frequency of a U-rich motif in the 3′ UTRs of all YTHDC2 targets vs all ubiquitin pathway-related genes (~500 genes) in the vicinity of the YTHDC2 crosslink site. (G) Electrophoretic mobility shift assay (EMSA) to detect RNA binding of full-length YTHDC2. The 5′-end labelled poly-A-RNA or poly-U-RNA oligos were used for the binding reaction, and subsequently separated in a native gel. Only poly-U-RNA binds to YTHDC2. (H) Creation of the *Ythdc2*^{YTH} knock-in mutant mouse model. A point mutation of the critical tryptophan (W1375A) in the YTH domain of mouse YTHDC2. The tryptophan (in bold) and a few flanking amino acids are shown. The designed mutation is shown in the box below. Sequence logos prepared from RNAseq data of testicular RNAs from individual *Ythdc2*^{YTH} mice showing their heterozygous and homozygous knock-in status.

Figure S2. YTHDC2 forms a complex with XRN1 and MEIOC in vivo. Related to Figure 2.

(A) YTHDC2 and XRN1(1-1279 aa) were co-expressed in insect cells, tags removed and fractionated by size-exclusion chromatography. Protein elution trace showing the presence of YTHDC2-XRN1(1-1279 aa) complex and free YTHDC2 protein. Indicated fractions were analysed by SDS-PAGE in Figure 2A. (B) Generation of 2xHA-tagged MEIOC knock-in mice. The inserted tag/linker sequence is shown below. (C) Comparison of proteins commonly identified in the purifications of YTHDC2-3xFLAG and 2xHA-MEIOC complexes from *Ythdc2*^{FLAG};*Meioc*^{HA} mouse testes. Some purifications were treated with RNases, as indicated. (D) Gene ontology analysis of proteins enriched in the purifications.

Figure S3. Creation of the *Ythdc2* RNA helicase mutant and in vitro RNA helicase assays with purified proteins. Related to Figure 3.

(A) Creation of the RNA helicase catalytic-dead (cat-dead) mouse mutant with the E322Q mutation in the N-terminal RecA domain. This converts the catalytic motif from DEVH to DQVH. The designed mutation is shown in the box below. Sequence logo prepared from RNAseq data of testicular RNAs from wildtype, heterozygous and homozygous *Ythdc2*$^{cat\text{-}dead}$ knock-in mice. (B) Gene ontology analysis

of genes up- and downregulated in the *Ythdc2*$^{cat\text{-}dead}$ mutant. Only terms with the adjusted p-value < 0.5 were plotted. (C-G) Fluorescence-based RNA helicase assay was used to study the unwinding activity of full-length YTHDC2. The Cy3 fluorescence is quenched in the duplex state. Unwinding activity is initiated after addition of ATP and measured over time (s). (C) Comparison of activity with YTHDC2$^{\Delta ANK}$ version. (D-G) Activity of YTHDC2 in the presence of its interaction partners XRN1 [XRN1^{E178Q} (1-279)] or MEIOC-coiled-coil (CC) domain. (H) Replicates of RNA helicase assay with the YTHDC2$^{\Delta ANK}$ version in the presence or absence of XRN1^{E178Q} (1-279). (I) Thin-layer chromatography (TLC) analysis of ATP hydrolysis assays with full-length YTHDC2 or its catalytic-dead version. XRN1^{E178Q} (1-279) was added in some of the reactions at the indicated concentrations. RNA is needed to stimulate the activity of YTHDC2. Addition of XRN1^{E178Q} (1-279) enhances the ATPase activity of YTHDC2.

Figure S4. Single-cell analyses of 4n germ cells from the *Ythdc2* knockout mutant. Related to Figure 4.

(A) Heat map showing differential gene expression in the single-cell transcriptome datasets of differentiating spermatogonia (SG) that are present in P10 mouse testes from the *Ythdc2*$^{+/-}$ and *Ythdc2*$^{-/-}$ (KO) genotypes. Shown are 64 genes with Wald test p-val < 0.05. (B) Expression analysis of the indicated meiotic markers (Sycp1, Sycp3). (C) Expression analysis of the indicated mitotic markers (*Ccna2*, *Cks2*). Note that the indicated mitotic marker genes are expressed in the undifferentiated and differentiated spermatogonia (SG) of both control and *Ythdc2* KO testes. Interestingly, these markers are also co-expressed in the two unique populations of cells found only in the *Ythdc2* KO together with meiotic markers *Sycp1* and *Sycp3*. (D) UMAP-based single-cell map showing the sum of log2-transformed frequencies of top ten U-rich motifs identified with iCLIP at the 3′ UTR in all the expressed genes per cell. (E) UMAP-based single-cell map illustrating the total expression for ubiquitination pathway genes per cell.

Figure S5. Spermatogenic defect in the homozygous *ythdc2* knockout male zebrafish. Related to Figure 5.

(A) Agarose gel showing genotyping PCR of zebrafish embryos to detect germline transmission of the *ythdc2*$^{\Delta 46}$ allele. (B) Hematoxylin and Eosin staining of adult zebrafish testes showing that the infertile homozygous *ythdc2*$^{\Delta 46/\Delta 46}$ fish lack spermatozoa. A lower magnification image compared to figure 5 is shown, highlighting the general testes histology for two fish of each *ythdc2* genotype. (C) Images at higher magnification. Scale bar in μm is indicated.

Table S1: List of RNA and DNA oligos used. Related to STAR Methods.

Helicase assay oligonucleotides		
CAGTAATACGACTCACTATAGGAGGGGCCGG TGGGGCCTGCGTCTTTACGGTGCTTAAAACA AAACAAAACAAAACAAAC	DNA template sequence for T7 in vitro transcription of loading strand RNA (only sense strand shown)	
AGCACCGUAAAGACGC	RNA with last cytidine will be labeled with Cy3	
GGCCCCACCGGCCCCUCC	RNA with first guanosine will be labeled with BHQ	
GCGTCTTTACGGTGCT	DNA competitor	
EMSA		
AAAAAAAAAAAAAAAA	polyA-RNA	17 nt
UUUUUUUUUUUUUUUU	polyU-RNA	17 nt
Synthetic CRISPR guide RNA creation for deletion in mice		
AAAGCACCGACTCGGTGCCACTTTTTCAAGT TGATAACGGACTAGCCTTATTTTAACTTGCT ATTTCTAGCTCTAAAAC	Common reverse DNA oligo for sgRNA generation	CRISPR sgR primer, RRoligo694
GAAATTAATACGACTCACTATAGGGTCCAGC TGACCCCCAGTCCGTTTTAGAGCTAGAAATA GC	Forward primer for sgRNA for YTH mutant (gene-specific sequence)	CRISPR F primer, RRoligo1064
GAAATTAATACGACTCACTATAGGCTTGGAG ATCCAATAATAGTGTTTTAGAGCTAGAAATA GC	Forward primer for sgRNA for 3xFLAG Ythdc2 line (gene-specific sequence)	CRISPR F primer, RRoligo868
Mouse genotyping primers		
GAGCGCCATCTAAACCTTGG	Genotyping PCR for Ythdc2-3xFLAG mouse line	RRoligo869
CCATCCGCAAACTCAGTGTC	Genotyping PCR for Ythdc2-3xFLAG mouse line	RRoligo870
TGCTTTCATCCTCTCCCTCA	*Genotyping PCR for Ythdc2 W1375A*	RRoligo1092
TGAGCAGTAACAACAGCAACA	*Genotyping PCR for Ythdc2 W1375A*	RRoligo1093
GCAATGTAACGGCCTCTTCG	*Genotyping PCR for 2xHA-Meioc mouse line*	RRoligo1151
CTTCTCCTCAGCGTCTCTCC	*Genotyping PCR for 2xHA-Meioc mouse line*	RRoligo1152
CCCTTGCCTAACCAAACAGC	*Genotyping primer for Ythdc2 E332Q mouse mutation*	RRoligo1358
GTTCACATCCAAGGCAGCAC	*Genotyping primer for Ythdc2 E332Q mouse mutation*	RRoligo1359
AAGGATGTGCTACCATGCCT	*To identify E332Q mutation*	RRoligo1360
Zebrafish genotyping primers		
AGACAGTAAAAGCAGGACCAGTG	Zebrafish Ythdc2 mutant genotypin	ythdc2 genF
TGTTGACCTGTGTGGTTTTCCC	Zebrafish Ythdc2 mutant genotypin	ythdc2 genR

Discussion

In this chapter, I described our following studies after the first YTHDC2 paper from the lab was published in 2007 (Wojtas et al., 2017). We would like to address three main questions: 1. What are the targets of YTHDC2 in mouse testis? 2. What makes YTHDC2 essential for mouse fertility, its m^6A reading ability or helicase activity? 3. What is the mechanism in regulating meiosis by YTHDC2? To seek the answer for the first question, my colleague Raman optimized the protocol for iCLIP experiment and prepared RNA samples and libraries for next generation sequencing. After analyzing the sequencing data, we found that YTHDC2 preferentially binds to 3′ UTR in testicular transcripts. This is not because these transcripts have longer 3′ UTR. Moreover, the binding sites of YTHDC2 are enriched in uridines (UUUUU), implicating YTHDC2 uses such U-rich motif as loading site to locate at the 3′ UTR. We speculated that this feature is conferred by the OB fold because this domain is thought to preferentially bind to U-rich sequence, for example, RNA helicase MLE displays uridine specificity via its OB fold (Prabu et al., 2015). I then performed electrophoretic mobility shift assay (EMSA) and confirmed that YTHDC2 selectively binds to U-rich oligonucleotide instead of A-rich oligonucleotide in vitro. However, it is still unclear why YTHDC2 specifically binds to 3′ UTR. On one hand, the OB fold may contribute to this process through recognizing U-rich sequence present at the 3′ UTR. On the other hand, it is possible that other proteins recruit YTHDC2 to 3′ UTR in vivo. Using immunoprecipitation coupled to mass spectrometry, we identified a putative RNA-binding protein RBM46 that is associated with YTHDC2 in mouse germ cells. Interestingly, a recent work reported that depletion of Rbm46 in zebrafish caused infertility (Dai et al., 2021). In the mutants, although spermatogonial cells were almost normal, the spermatogenesis was impaired, which was very similar to what we observed in YTHDC2 mutant zebrafish. The function of RBM46 in mouse and human remains elusive, but at least in germ cells, it is possible that RBM46 binds to targeted transcripts and recruits YTHDC2 to the 3′ UTR. Therefore, creation of *Rbm46* knockout mouse, checking its phenotype and investigating the interaction between YTHDC2 and RBM46 would benefit the understanding of how RBM46 and YTHDC2 collectively regulate germ cell development to ensure the proper meiosis.

Another finding from the iCLIP sequencing results is that among those thousands of transcripts regulated by YTHDC2 in mouse testis, a large class belongs to the ubiquitination pathway. As is known β-TrCP, a substrate recognition component of an E3 ubiquitin ligase complex, can

ubiquitinate transcription factor DMTR1 for degradation and thereby facilitate the proper transition from mitosis to meiosis in mouse testis (Nakagawa et al., 2017), but whether β-TrCP is directly targeted by YTHDC2 and how the ubiquitination pathway is mediated by YTHDC2 to maintain normal spermatogenesis remain to be understood.

Even though the YTH domain of YTHDC2 harbors m^6A recognition capacity in vitro, its binding affinity is the weakest among all YTH proteins in mammalian cells (Wojtas et al., 2017). We also compared the YTHDC2 binding sites (UUUUU) with m^6A methylation sites (DRACH) in a 200-nt window and found that U-rich motif is not in the vicinity of DRACH motif. These discoveries let us doubt if the m^6A reading ability is relevant to the mouse phenotype. Concerning with this, we introduced a point mutation (W1375A) in the YTH domain to abrogate m^6A recognition by mouse Ythdc2. After a long-term observation, we confirmed YTH mutant mice were viable and fertile, indicating in fact the YTH domain is dispensable for mouse fertility. Probably YTHDC2 does not use its YTH domain to bind to the targets and this is reminiscent of the extreme case of the yeast YTH protein Mmi1, where Mmi1 YTH domain specifically recognizes hexanucleotide U(U/C)AAAC that are enriched in the determinant of selective removal (DSR) region in meiotic-specific mRNAs. The YTH domain cannot bind m^6A-containig RNAs (Hiriart et al., 2012; Wang et al., 2016a; Yamashita et al., 2012). Accordingly, it would make sense to rethink the YTH proteins in different species, especially those possessing weak binding affinity towards m^6A, which may act in an m^6A-independent way to mediate specific biological processes.

YTHDC2 includes two RecA domains and is already demonstrated to be a 3′ to 5′ RNA helicase. We created catalytic-dead YTHDC2 mutant mice and these mice were infertile, indicating the helicase activity is essential for mouse fertility. However, what is the role of helicase domain in regulating germ cell meiosis is currently unknown. YTHDC2 probably translocates along the target from 3′ to 5′ to resolve the RNA structures or preclude bound protein factors. It is also possible that YTHDC2 remodels RNA-protein complexes bound at the 3′ UTR. What makes YTHDC2 unique as an RNA helicase is that two ankyrin repeats insert into the helicase domain. Such design allows the ankyrin repeats to serve as a module where other protein factors can target it to regulate YTHDC2 helicase activity, here, in our case, we showed XRN1 enhances the helicase activity of YTHDC2 by directly binding to this module. To study the allosteric effect of XRN1 and decipher the interface, I tried to determine the structure of YTHDC2-XRN1 complex by crosslinking mass spectrometry and cryo-EM. At the same time, I also plan to use hydrogen-

deuterium exchange mass spectrometry (HDX-MS) to study the molecular dynamics of YTHDC2. The aim is that with structural information, we can design mutations to dissect YTHDC2 and XRN1 without affect their major functions in the cell. For example, our data indicated a short region in the C-terminal domain of XRN1 likely binds to the ankyrin repeats. It is known that C-terminus of XRN1 is the low-complex region and interacts with many other protein factors, such as decapping activator EDC4 (Braun et al., 2012). By specifically mutating residues in XRN1 that required for interaction with YTHDC2 without touching either its N-terminal nuclease domain or other regions in the C-terminus that interplay with other protein factors will be the best model to study the function of YTHDC2-XRN1 complex in vivo. Furthermore, we can also mutate residues in the ankyrin repeats to block the formation of YTHDC2-XRN1 complex and observe what is the consequence. However, we do not know if altering ankyrin repeats will impact its helicase activity because we noticed that removal of ankyrin repeats makes YTHDC2 surprisingly active in vitro. Considering the fact that helicase activity of YTHDC2 is required for mouse fertility, one experiment would be interesting to exam delta ankyrin repeats mouse. What will happen if mouse genome encodes a super active helicase? Can we detect structural changes in YTHDC2-bound RNA targets in wildtype and mutant germ cells, particularly in the 3′ UTR? All these experiments will help us to understand how the helicase activity of YTHDC2 that can be modulated by XRN1 contributes to facilitate meiosis in mammalian germ cells.

Though MEIOC directly binds to YTHDC2 via its C-terminal coiled coil domain, it is unclear where it binds in YTHDC2. Ankyrin repeats are probably not the region, based on our data (unpublished). Except the coiled coil domain, the remaining region is intrinsically disordered, which makes it difficult to express MEIOC recombinant protein. I also tried to co-express YTHDC2 and MEIOC, but failed either. MEIOC knockout mouse has the same phenotype as YTHDC2 knockout (Abby et al., 2016; Soh et al., 2017), implying that MEIOC probably regulates the same targets of YTHDC2 to promote meiosis in germ cells. In insect, MEIOC ortholog Bam forms a complex with YTHDC2 ortholog Bgcn, but Bgcn lacks the YTH domain and probably has no ATPase and helicase activities (Jain et al., 2018; Ohlstein et al., 2000). YTHDC2, together with MEIOC, appears to adapt different mechanisms to regulate germ cell development during evolution.

Single-cell sequencing of testicular transcriptome reveals that YTHDC2 mediates the fine separation of the transcriptomes when mitotic spermatogonia enter meiosis. In the YTHDC2-

deficient testis, mitotic germ cells can replicate and transition into meiosis. However, such meiotic cells still express genes that are related with mitosis, meaning both mitotic transcripts and meiotic transcripts are present in cells which are already in the early stage of meiosis. This mixed identity of transcriptomes leads to defective differentiation for meiotic cells so that no sperm is produced and these cells probably undergo apoptosis in the end. The reason causing mixed transcriptomes in meiotic cells is still poorly understood. Nevertheless, our data favor a model that YTHDC2 promotes the decay of mitotic RNA targets during the mitosis-meiosis transition stage via recruiting exoribonuclease XRN1. In the future, it would be interesting to test this model by mutating the U-rich sequence in one or two YTHDC2 targets and check if that affects its regulation by YTHDC2 or we can artificially tether YTHDC2 to a reporter mRNA and detect the consequence for that reporter.

So far, what we have discussed is about the m^6A-independent function in regulating mammalian germ cell development. Whereas, we can not exclude the possibility that YTHDC2 functions replying upon its m^6A reading ability in other scenarios. Several studies point to its role in enhancing translation. For example, when FTO is depleted in mouse adipocytes, as a result, YTHDC2 recognizes increased m^6A markers in *Hif1a* mRNA to induce mRNA translation and promote white-to-beige fat transition (Wu et al., 2021). Another paper shows that YTHDC2 recognizes methylated adenosine within the internal ribosome entry site (IRES) in HCV (Hepatitis C virus) RNA genome to support its cap-independent translation (Kim and Siddiqui, 2021). Furthermore, YTHDC2 also targets m^6A in mRNA coding sequences (CDS), probably to resolve mRNA secondary structures, and stimulates translation (Mao et al., 2019). Therefore, it would be risky to rule out any m^6A-dependent function of YTHDC2, especially outside the germline. Future studies should be focused on investigating m^6A-dependent and m^6A-independent roles of YTHDC2 in different contexts.

In recent years, the importance of chemical modifications that are represented by m^6A methylation in gene expression regulation becomes more and more appreciated. As discussed in this thesis, to date, more than 170 chemical modifications have been identified in RNA. Some of them are quite abundant in the cells but some are subtle. Undoubtably, many chemical modifications play essential functions in almost every aspect of RNA metabolism. At the same time, many puzzles remain to be solved in this field. Is the modification reversible? What are the writers and erasers for this modification? More importantly, what are the readers? Addressing

these questions will expand our knowledge of how RNA modifications contribute to elegant fine-tuning of gene expression and eventually, benefit us for developing potential RNA-based therapies.

References

Abbas, Y.M., Laudenbach, B.T., Martinez-Montero, S., Cencic, R., Habjan, M., Pichlmair, A., Damha, M.J., Pelletier, J., and Nagar, B. (2017). Structure of human IFIT1 with capped RNA reveals adaptable mRNA binding and mechanisms for sensing N1 and N2 ribose 2'-O methylations. Proc Natl Acad Sci U S A *114*, E2106-E2115.

Abby, E., Tourpin, S., Ribeiro, J., Daniel, K., Messiaen, S., Moison, D., Guerquin, J., Gaillard, J.C., Armengaud, J., Langa, F., *et al.* (2016). Implementation of meiosis prophase I programme requires a conserved retinoid-independent stabilizer of meiotic transcripts. Nat Commun *7*, 10324.

Abdelraheim, S.R., Spiller, D.G., and McLennan, A.G. (2003). Mammalian NADH diphosphatases of the Nudix family: cloning and characterization of the human peroxisomal NUDT12 protein. Biochem J *374*, 329-335.

Agarwala, S.D., Blitzblau, H.G., Hochwagen, A., and Fink, G.R. (2012). RNA methylation by the MIS complex regulates a cell fate decision in yeast. PLoS Genet *8*, e1002732.

Agris, P.F. (1996). The importance of being modified: roles of modified nucleosides and Mg^{2+} in RNA structure and function. Prog Nucleic Acid Res Mol Biol *53*, 79-129.

Akichika, S., Hirano, S., Shichino, Y., Suzuki, T., Nishimasu, H., Ishitani, R., Sugita, A., Hirose, Y., Iwasaki, S., Nureki, O., *et al.* (2019). Cap-specific terminal N-6-methylation of RNA by an RNA polymerase II-associated methyltransferase. Science *363*, 141-141.

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

Alexandrov, A., Chernyakov, I., Gu, W., Hiley, S.L., Hughes, T.R., Grayhack, E.J., and Phizicky, E.M. (2006). Rapid tRNA decay can result from lack of nonessential modifications. Mol Cell *21*, 87-96.

Amort, T., Rieder, D., Wille, A., Khokhlova-Cubberley, D., Riml, C., Trixl, L., Jia, X.Y., Micura, R., and Lusser, A. (2017). Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. Genome Biol *18*, 1.

Anazi, S., Maddirevula, S., Faqeih, E., Alsedairy, H., Alzahrani, F., Shamseldin, H.E., Patel, N., Hashem, M., Ibrahim, N., Abdulwahab, F., *et al.* (2017). Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. Mol Psychiatry *22*, 615-624.

Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., *et al.* (2013). The human cap-binding complex is functionally connected to the nuclear RNA exosome. Nat Struct Mol Biol *20*, 1367-1376.

Anderson, J., Phan, L., Cuesta, R., Carlson, B.A., Pak, M., Asano, K., Bjork, G.R., Tamame, M., and Hinnebusch, A.G. (1998). The essential Gcd10p-Gcd14p nuclear complex is required for 1-

methyladenosine modification and maturation of initiator methionyl-tRNA. Genes Dev *12*, 3650-3662.

Aoyama, T., Yamashita, S., and Tomita, K. (2020). Mechanistic insights into m6A modification of U6 snRNA by human METTL16. Nucleic Acids Res *48*, 5157-5168.

Bailey, A.S., Batista, P.J., Gold, R.S., Chen, Y.G., de Rooij, D.G., Chang, H.Y., and Fuller, M.T. (2017). The conserved RNA helicase YTHDC2 regulates the transition from proliferation to differentiation in the germline. Elife *6*.

Bartosovic, M., Molares, H.C., Gregorova, P., Hrossova, D., Kudla, G., and Vanacova, S. (2017). N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3'-end processing. Nucleic Acids Res *45*, 11356-11370.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., *et al.* (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell *15*, 707-719.

Belanger, F., Stepinski, J., Darzynkiewicz, E., and Pelletier, J. (2010). Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. J Biol Chem *285*, 33037-33044.

Bessman, M.J., Frick, D.N., and O'Handley, S.F. (1996). The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. J Biol Chem *271*, 25059-25062.

Bessman, M.J., Walsh, J.D., Dunn, C.A., Swaminathan, J., Weldon, J.E., and Shen, J. (2001). The gene ygdP, associated with the invasiveness of Escherichia coli K1, designates a Nudix hydrolase, Orf176, active on adenosine (5')-pentaphospho-(5')-adenosine (Ap5A). J Biol Chem *276*, 37834-37838.

Bird, J.G., Zhang, Y., Tian, Y., Panova, N., Barvik, I., Greene, L., Liu, M., Buckley, B., Krasny, L., Lee, J.K., *et al.* (2016). The mechanism of RNA 5' capping with NAD+, NADH and desphospho-CoA. Nature *535*, 444-447.

Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J., and Nielsen, H. (2015). Profiling of ribose methylations in RNA by high-throughput sequencing. Angew Chem Int Ed Engl *54*, 451-455.

Bisaillon, M., and Shuman, S. (2001a). Functional groups required for the stability of yeast RNA triphosphatase in vitro and in vivo. J Biol Chem *276*, 30514-30520.

Bisaillon, M., and Shuman, S. (2001b). Structure-function analysis of the active site tunnel of yeast RNA triphosphatase. J Biol Chem *276*, 17261-17266.

Blanco, S., Dietmann, S., Flores, J.V., Hussain, S., Kutter, C., Humphreys, P., Lukk, M., Lombard, P., Treps, L., Popis, M., *et al.* (2014). Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. Embo J *33*, 2020-2039.

Boccaletto, P., Stefaniak, F., Ray, A., Cappannini, A., Mukherjee, S., Purta, E., Kurkowska, M., Shirvanizadeh, N., Destefanis, E., Groza, P., *et al.* (2022). MODOMICS: a database of RNA modification pathways. 2021 update. Nucleic Acids Res *50*, D231-D235.

Bodi, Z., Zhong, S., Mehra, S., Song, J., Graham, N., Li, H., May, S., and Fray, R.G. (2012). Adenosine Methylation in Arabidopsis mRNA is Associated with the 3' End and Reduced Levels Cause Developmental Defects. Front Plant Sci *3*, 48.

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

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

Boulias, K., Toczydlowska-Socha, D., Hawley, B., Liberman, N., Takashima, K., Zaccara, S., Guez, T., Vasseur, J.J., Debart, F., Aravind, L., *et al.* (2019). Identification of the m(6)Am Methyltransferase PCIF1 Reveals the Location and Functions of m(6)Am in the Transcriptome. Mol Cell *75*, 631-+.

Braun, J.E., Truffault, V., Boland, A., Huntzinger, E., Chang, C.T., Haas, G., Weichenrieder, O., Coles, M., and Izaurralde, E. (2012). A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. Nat Struct Mol Biol *19*, 1324-1331.

Brown, J.A., Kinzig, C.G., DeGregorio, S.J., and Steitz, J.A. (2016). Methyltransferase-like protein 16 binds the 3'-terminal triple helix of MALAT1 long noncoding RNA. Proc Natl Acad Sci U S A *113*, 14013-14018.

Byszewska, M., Smietanski, M., Purta, E., and Bujnicki, J.M. (2014). RNA methyltransferases involved in 5 ' cap biosynthesis. Rna Biol *11*, 1597-1607.

Cahova, H., Winz, M.L., Hofer, K., Nubel, G., and Jaschke, A. (2015). NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. Nature *519*, 374-377.

Caldwell, D.C., and Emerson, C.P., Jr. (1985). The role of cap methylation in the translational activation of stored maternal histone mRNA in sea urchin embryos. Cell *42*, 691-700.

Chan, C.T., Pang, Y.L., Deng, W., Babu, I.R., Dyavaiah, M., Begley, T.J., and Dedon, P.C. (2012). Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat Commun *3*, 937.

Changela, A., Ho, C.K., Martins, A., Shuman, S., and Mondragon, A. (2001). Structure and mechanism of the RNA triphosphatase component of mammalian mRNA capping enzyme. Embo J *20*, 2575-2586.

Charenton, C., Gaudon-Plesse, C., Fourati, Z., Taverniti, V., Back, R., Kolesnikova, O., Seraphin, B., and Graille, M. (2017). A unique surface on Pat1 C-terminal domain directly interacts with Dcp2 decapping enzyme and Xrn1 5'-3' mRNA exonuclease in yeast. Proc Natl Acad Sci U S A *114*, E9493-E9501.

Charenton, C., Taverniti, V., Gaudon-Plesse, C., Back, R., Seraphin, B., and Graille, M. (2016). Structure of the active form of Dcp1-Dcp2 decapping enzyme bound to m(7)GDP and its Edc3 activator. Nat Struct Mol Biol *23*, 982-986.

Chen, X., Li, A., Sun, B.F., Yang, Y., Han, Y.N., Yuan, X., Chen, R.X., Wei, W.S., Liu, Y., Gao, C.C., *et al.* (2019). 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. Nat Cell Biol *21*, 978-990.

Chen, Y., Cai, H., Pan, J., Xiang, N., Tien, P., Ahola, T., and Guo, D.Y. (2009a). Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. P Natl Acad Sci USA *106*, 3484-3489.

Chen, Y.G., Kowtoniuk, W.E., Agarwal, I., Shen, Y., and Liu, D.R. (2009b). LC/MS analysis of cellular RNA reveals NAD-linked RNA. Nat Chem Biol *5*, 879-881.

Cheng, H., Dufu, K., Lee, C.S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. Cell *127*, 1389-1400.

Chiu, S.Y., Lejeune, F., Ranganathan, A.C., and Maquat, L.E. (2004). The pioneer translation initiation complex is functionally distinct from but structurally overlaps with the steady-state translation initiation complex. Genes Dev *18*, 745-754.

Choe, J., Lin, S., Zhang, W., Liu, Q., Wang, L., Ramirez-Moya, J., Du, P., Kim, W., Tang, S., Sliz, P., *et al.* (2018). mRNA circularization by METTL3-eIF3h enhances translation and promotes oncogenesis. Nature *561*, 556-560.

Choe, J., Oh, N., Park, S., Lee, Y.K., Song, O.K., Locker, N., Chi, S.G., and Kim, Y.K. (2012). Translation initiation on mRNAs bound by nuclear cap-binding protein complex CBP80/20 requires interaction between CBP80/20-dependent translation initiation factor and eukaryotic translation initiation factor 3g. J Biol Chem *287*, 18500-18509.

Choe, J., Ryu, I., Park, O.H., Park, J., Cho, H., Yoo, J.S., Chi, S.W., Kim, M.K., Song, H.K., and Kim, Y.K. (2014). eIF4AIII enhances translation of nuclear cap-binding complex-bound mRNAs by promoting disruption of secondary structures in 5'UTR. Proc Natl Acad Sci U S A *111*, E4577-4586.

Chowdhury, A., Mukhopadhyay, J., and Tharun, S. (2007). The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. RNA *13*, 998-1016.

Church, C., Moir, L., McMurray, F., Girard, C., Banks, G.T., Teboul, L., Wells, S., Bruning, J.C., Nolan, P.M., Ashcroft, F.M., *et al.* (2010). Overexpression of Fto leads to increased food intake and results in obesity. Nat Genet *42*, 1086-1092.

Cohn, W.E. (1960). Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. J Biol Chem *235*, 1488-1498.

Collart, M.A. (2016). The Ccr4-Not complex is a key regulator of eukaryotic gene expression. Wiley Interdiscip Rev RNA *7*, 438-454.

Da Sacco, L., and Masotti, A. (2012). Recent insights and novel bioinformatics tools to understand the role of microRNAs binding to 5' untranslated region. Int J Mol Sci *14*, 480-495.

Daffis, S., Szretter, K.J., Schriewer, J., Li, J., Youn, S., Errett, J., Lin, T.Y., Schneller, S., Zust, R., Dong, H., *et al.* (2010). 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. Nature *468*, 452-456.

Dai, X., Cheng, X., Huang, J., Gao, Y., Wang, D., Feng, Z., Zhai, G., Lou, Q., He, J., Wang, Z., *et al.* (2021). Rbm46, a novel germ cell-specific factor, modulates meiotic progression and spermatogenesis. Biol Reprod *104*, 1139-1153.

David, R., Burgess, A., Parker, B., Li, J., Pulsford, K., Sibbritt, T., Preiss, T., and Searle, I.R. (2017). Transcriptome-Wide Mapping of RNA 5-Methylcytosine in Arabidopsis mRNAs and Noncoding RNAs. Plant Cell *29*, 445-460.

De Jesus, D.F., Zhang, Z., Kahraman, S., Brown, N.K., Chen, M., Hu, J., Gupta, M.K., He, C., and Kulkarni, R.N. (2019). m(6)A mRNA Methylation Regulates Human beta-Cell Biology in Physiological States and in Type 2 Diabetes. Nat Metab *1*, 765-774.

De la Pena, M., Kyrieleis, O.J.P., and Cusack, S. (2007). Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap N7 methyl-transferase. Embo J *26*, 4913-4925.

Decatur, W.A., and Schnare, M.N. (2008). Different mechanisms for pseudouridine formation in yeast 5S and 5.8S rRNAs. Mol Cell Biol *28*, 3089-3100.

Decroly, E., Ferron, F., Lescar, J., and Canard, B. (2011). Conventional and unconventional mechanisms for capping viral mRNA. Nat Rev Microbiol *10*, 51-65.

Delatte, B., Wang, F., Ngoc, L.V., Collignon, E., Bonvin, E., Deplus, R., Calonne, E., Hassabi, B., Putmans, P., Awe, S., *et al.* (2016). RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science *351*, 282-285.

Delaunay, S., and Frye, M. (2019). RNA modifications regulating cell fate in cancer. Nat Cell Biol *21*, 552-559.

Desrosiers, R., Friderici, K., and Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc Natl Acad Sci U S A *71*, 3971-3975.

Diaz, F., Khosa, S., Niyazov, D., Lee, H., Person, R., Morrow, M.M., Signer, R., Dorrani, N., Zheng, A., Herzog, M., *et al.* (2020). Novel NUDT2 variant causes intellectual disability and polyneuropathy. Ann Clin Transl Neurol *7*, 2320-2325.

Dina, C., Meyre, D., Gallina, S., Durand, E., Korner, A., Jacobson, P., Carlsson, L.M., Kiess, W., Vatin, V., Lecoeur, C., *et al.* (2007). Variation in FTO contributes to childhood obesity and severe adult obesity. Nat Genet *39*, 724-726.

Dixit, D., Xie, Q., Rich, J.N., and Zhao, J.C. (2017). Messenger RNA Methylation Regulates Glioblastoma Tumorigenesis. Cancer Cell *31*, 474-475.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., *et al.* (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201-206.

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M.S., Dai, Q., Di Segni, A., Salmon-Divon, M., Clark, W.C., *et al.* (2016). The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. Nature *530*, 441-446.

Dorn, L.E., Lasman, L., Chen, J., Xu, X., Hund, T.J., Medvedovic, M., Hanna, J.H., van Berlo, J.H., and Accornero, F. (2019). The N(6)-Methyladenosine mRNA Methylase METTL3 Controls Cardiac Homeostasis and Hypertrophy. Circulation *139*, 533-545.

Doxtader, K.A., Wang, P., Scarborough, A.M., Seo, D., Conrad, N.K., and Nam, Y. (2018). Structural Basis for Regulation of METTL16, an S-Adenosylmethionine Homeostasis Factor. Mol Cell *71*, 1001-1011 e1004.

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., and Wu, L. (2016). YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. Nat Commun *7*, 12626.

Duan, J., Li, L., Lu, J., Wang, W., and Ye, K. (2009). Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. Mol Cell *34*, 427-439.

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

Dunckley, T., and Parker, R. (1999). The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. Embo J *18*, 5411-5422.

Dunn, D.B. (1961). The occurrence of 1-methyladenine in ribonucleic acid. Biochim Biophys Acta *46*, 198-200.

Fabrega, C., Hausmann, S., Shen, V., Shuman, S., and Lima, C.D. (2004). Structure and mechanism of mRNA cap (guanine-N7) methyltransferase. Mol Cell *13*, 77-89.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol Cell *11*, 1549-1561.

Falnes, P.O., Johansen, R.F., and Seeberg, E. (2002). AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. Nature *419*, 178-182.

Fischer, J., Koch, L., Emmerling, C., Vierkotten, J., Peters, T., Bruning, J.C., and Ruther, U. (2009). Inactivation of the Fto gene protects from obesity. Nature *458*, 894-898.

Flaherty, S.M., Fortes, P., Izaurralde, E., Mattaj, I.W., and Gilmartin, G.M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3' processing. Proc Natl Acad Sci U S A *94*, 11893-11898.

Foley, P.L., Hsieh, P.K., Luciano, D.J., and Belasco, J.G. (2015). Specificity and evolutionary conservation of the Escherichia coli RNA pyrophosphohydrolase RppH. J Biol Chem *290*, 9478-9486.

Fortes, P., Inada, T., Preiss, T., Hentze, M.W., Mattaj, I.W., and Sachs, A.B. (2000). The yeast nuclear cap binding complex can interact with translation factor eIF4G and mediate translation initiation. Mol Cell *6*, 191-196.

Frindert, J., Zhang, Y., Nubel, G., Kahloon, M., Kolmar, L., Hotz-Wagenblatt, A., Burhenne, J., Haefeli, W.E., and Jaschke, A. (2018). Identification, Biosynthesis, and Decapping of NAD-Capped RNAs in B. subtilis. Cell Rep *24*, 1890-1901 e1898.

Fu, L., Guerrero, C.R., Zhong, N., Amato, N.J., Liu, Y., Liu, S., Cai, Q., Ji, D., Jin, S.G., Niedernhofer, L.J., *et al.* (2014). Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc *136*, 11582-11585.

Ganot, P., Bortolin, M.L., and Kiss, T. (1997). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell *89*, 799-809.

Garcia-Campos, M.A., Edelheit, S., Toth, U., Safra, M., Shachar, R., Viukov, S., Winkler, R., Nir, R., Lasman, L., Brandis, A., *et al.* (2019). Deciphering the "m(6)A Code" via Antibody-Independent Quantitative Profiling. Cell *178*, 731-747 e716.

Ge, H., Chen, X., Yang, W., Niu, L., and Teng, M. (2013). Crystal structure of wild-type and mutant human Ap4A hydrolase. Biochem Biophys Res Commun *432*, 16-21.

Gerken, T., Girard, C.A., Tung, Y.C., Webby, C.J., Saudek, V., Hewitson, K.S., Yeo, G.S., McDonough, M.A., Cunliffe, S., McNeill, L.A., *et al.* (2007). The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. Science *318*, 1469-1472.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S., *et al.* (2015). Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. Science *347*, 1002-1006.

Ghosh, A., and Lima, C.D. (2010). Enzymology of RNA cap synthesis. Wires Rna *1*, 152-172.

Grudzien-Nogalska, E., Jiao, X., Song, M.G., Hart, R.P., and Kiledjian, M. (2016). Nudt3 is an mRNA decapping enzyme that modulates cell migration. RNA *22*, 773-781.

Grudzien-Nogalska, E., and Kiledjian, M. (2017). New insights into decapping enzymes and selective mRNA decay. Wiley Interdiscip Rev RNA *8*.

Gu, M.G., and Lima, C.D. (2005). Processing the message: structural insights into capping and decapping mRNA. Curr Opin Struc Biol *15*, 99-106.

Hakansson, K., Doherty, A.J., Shuman, S., and Wigley, D.B. (1997). X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. Cell *89*, 545-553.

Haline-Vaz, T., Silva, T.C.L., and Zanchin, N.I.T. (2008). The human interferon-regulated ISG95 protein interacts with RNA polymerase II and shows methyltransferase activity. Biochem Bioph Res Co *372*, 719-724.

Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. Nat Rev Mol Cell Biol *19*, 20-30.

Haussmann, I.U., Wu, Y.Y., Nallasivan, M.P., Archer, N., Bodi, Z., Hebenstreit, D., Waddell, S., Fray, R., and Soller, M. (2022). CMTr cap-adjacent 2 '-O-ribose mRNA methyltransferases are required for reward learning and mRNA localization to synapses. Nature Communications *13*.

He, P.C., and He, C. (2021). m(6) A RNA methylation: from mechanisms to therapeutic potential. Embo J *40*, e105977.

Hess, M.E., Hess, S., Meyer, K.D., Verhagen, L.A., Koch, L., Bronneke, H.S., Dietrich, M.O., Jordan, S.D., Saletore, Y., Elemento, O., *et al.* (2013). The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry. Nat Neurosci *16*, 1042-1048.

Higman, M.A., Christen, L.A., and Niles, E.G. (1994). The mRNA (guanine-7-)methyltransferase domain of the vaccinia virus mRNA capping enzyme. Expression in Escherichia coli and structural and kinetic comparison to the intact capping enzyme. J Biol Chem *269*, 14974-14981.

Hiriart, E., Vavasseur, A., Touat-Todeschini, L., Yamashita, A., Gilquin, B., Lambert, E., Perot, J., Shichino, Y., Nazaret, N., Boyault, C., *et al.* (2012). Mmi1 RNA surveillance machinery directs RNAi complex RITS to specific meiotic genes in fission yeast. Embo J *31*, 2296-2308.

Ho, C.K., Pei, Y., and Shuman, S. (1998). Yeast and viral RNA 5 ' triphosphatases comprise a new nucleoside triphosphatase family. J Biol Chem *273*, 34151-34156.

Hofer, K., Li, S., Abele, F., Frindert, J., Schlotthauer, J., Grawenhoff, J., Du, J., Patel, D.J., and Jaschke, A. (2016). Structure and function of the bacterial decapping enzyme NudC. Nat Chem Biol *12*, 730-734.

Hongay, C.F., and Orr-Weaver, T.L. (2011). Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. Proc Natl Acad Sci U S A *108*, 14855-14860.

Hsieh, P.K., Richards, J., Liu, Q., and Belasco, J.G. (2013). Specificity of RppH-dependent RNA degradation in Bacillus subtilis. Proc Natl Acad Sci U S A *110*, 8864-8869.

Hsu, P.J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., Qi, M., Lu, Z., Shi, H., Wang, J., *et al.* (2017). Ythdc2 is an N(6)-methyladenosine binding protein that regulates mammalian spermatogenesis. Cell Res *27*, 1115-1127.

Hu, H., Flynn, N., Zhang, H., You, C., Hang, R., Wang, X., Zhong, H., Chan, Z., Xia, Y., and Chen, X. (2021). SPAAC-NAD-seq, a sensitive and accurate method to profile NAD(+)-capped transcripts. Proc Natl Acad Sci U S A *118*.

Huang, H.L., Weng, H.Y., Deng, X.L., and Chen, J.J. (2020). RNA Modifications in Cancer: Functions, Mechanisms, and Therapeutic Implications. Annu Rev Canc Biol *4*, 221-240.

Huang, T., Chen, W., Liu, J., Gu, N., and Zhang, R. (2019). Genome-wide identification of mRNA 5-methylcytosine in mammals. Nat Struct Mol Biol *26*, 380-388.

Hudecek, O., Benoni, R., Reyes-Gutierrez, P.E., Culka, M., Sanderova, H., Hubalek, M., Rulisek, L., Cvacka, J., Krasny, L., and Cahova, H. (2020). Dinucleoside polyphosphates act as 5'-RNA caps in bacteria. Nat Commun *11*, 1052.

Hui, M.P., Foley, P.L., and Belasco, J.G. (2014). Messenger RNA degradation in bacterial cells. Annu Rev Genet *48*, 537-559.

Ianniello, Z., Paiardini, A., and Fatica, A. (2019). N(6)-Methyladenosine (m(6)A): A Promising New Molecular Target in Acute Myeloid Leukemia. Front Oncol *9*, 251.

Inesta-Vaquera, F., and Cowling, V.H. (2017). Regulation and function of CMTR1-dependent mRNA cap methylation. Wiley Interdiscip Rev RNA *8*.

Inoue, K., Ohno, M., Sakamoto, H., and Shimura, Y. (1989). Effect of the cap structure on pre-mRNA splicing in Xenopus oocyte nuclei. Genes Dev *3*, 1472-1479.

Ito, S., Horikawa, S., Suzuki, T., Kawauchi, H., Tanaka, Y., Suzuki, T., and Suzuki, T. (2014). Human NAT10 is an ATP-dependent RNA acetyltransferase responsible for N4-acetylcytidine formation in 18 S ribosomal RNA (rRNA). J Biol Chem *289*, 35724-35730.

Jain, D., Puno, M.R., Meydan, C., Lailler, N., Mason, C.E., Lima, C.D., Anderson, K.V., and Keeney, S. (2018). ketu mutant mice uncover an essential meiotic function for the ancient RNA helicase YTHDC2. Elife *7*.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G., *et al.* (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol *7*, 885-887.

Jia, G., Yang, C.G., Yang, S., Jian, X., Yi, C., Zhou, Z., and He, C. (2008). Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. FEBS Lett *582*, 3313-3319.

Jiao, X., Chang, J.H., Kilic, T., Tong, L., and Kiledjian, M. (2013). A mammalian pre-mRNA 5' end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing. Mol Cell *50*, 104-115.

Jiao, X., Doamekpor, S.K., Bird, J.G., Nickels, B.E., Tong, L., Hart, R.P., and Kiledjian, M. (2017). 5' End Nicotinamide Adenine Dinucleotide Cap in Human Cells Promotes RNA Decay through DXO-Mediated deNADding. Cell *168*, 1015-1027 e1010.

Jones, C.I., Zabolotskaya, M.V., and Newbury, S.F. (2012). The 5' --> 3' exoribonuclease XRN1/Pacman and its functions in cellular processes and development. Wiley Interdiscip Rev RNA *3*, 455-468.

Julius, C., and Yuzenkova, Y. (2019). Noncanonical RNA-capping: Discovery, mechanism, and physiological role debate. Wiley Interdiscip Rev RNA *10*, e1512.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., *et al.* (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589.

Kasowitz, S.D., Ma, J., Anderson, S.J., Leu, N.A., Xu, Y., Gregory, B.D., Schultz, R.M., and Wang, P.J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. PLoS Genet *14*, e1007412.

Ke, S., Pandya-Jones, A., Saito, Y., Fak, J.J., Vagbo, C.B., Geula, S., Hanna, J.H., Black, D.L., Darnell, J.E., Jr., and Darnell, R.B. (2017). m(6)A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes Dev *31*, 990-1006.

Kim, G.W., and Siddiqui, A. (2021). N6-methyladenosine modification of HCV RNA genome regulates cap-independent IRES-mediated translation via YTHDC2 recognition. Proc Natl Acad Sci U S A *118*.

Kiss-Laszlo, Z., Henry, Y., Bachellerie, J.P., Caizergues-Ferrer, M., and Kiss, T. (1996). Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell *85*, 1077-1088.

Konarska, M.M., Padgett, R.A., and Sharp, P.A. (1984). Recognition of cap structure in splicing in vitro of mRNA precursors. Cell *38*, 731-736.

Kramer, S., and McLennan, A.G. (2019). The complex enzymology of mRNA decapping: Enzymes of four classes cleave pyrophosphate bonds. Wiley Interdiscip Rev RNA *10*, e1511.

Kretschmer, J., Rao, H., Hackert, P., Sloan, K.E., Hobartner, C., and Bohnsack, M.T. (2018). The m(6)A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5'-3' exoribonuclease XRN1. RNA *24*, 1339-1350.

Kuge, H., Brownlee, G.G., Gershon, P.D., and Richter, J.D. (1998). Cap ribose methylation of c-mos mRNA stimulates translation and oocyte maturation in Xenopus laevis. Nucleic Acids Res *26*, 3208-3214.

Lafontaine, D.L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev *12*, 527-537.

Lapeyre, B., and Purushothaman, S.K. (2004). Spb1p-directed formation of Gm2922 in the ribosome catalytic center occurs at a late processing stage. Mol Cell *16*, 663-669.

Lasman, L., Krupalnik, V., Viukov, S., Mor, N., Aguilera-Castrejon, A., Schneir, D., Bayerl, J., Mizrahi, O., Peles, S., Tawil, S., *et al.* (2020). Context-dependent functional compensation between Ythdf m(6)A reader proteins. Genes Dev *34*, 1373-1391.

Legrand, C., Tuorto, F., Hartmann, M., Liebers, R., Jacob, D., Helm, M., and Lyko, F. (2017). Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. Genome Res *27*, 1589-1596.

Lesbirel, S., Viphakone, N., Parker, M., Parker, J., Heath, C., Sudbery, I., and Wilson, S.A. (2018). The m(6)A-methylase complex recruits TREX and regulates mRNA export. Sci Rep *8*, 13827.

Li, A., Chen, Y.S., Ping, X.L., Yang, X., Xiao, W., Yang, Y., Sun, H.Y., Zhu, Q., Baidya, P., Wang, X., *et al.* (2017a). Cytoplasmic m(6)A reader YTHDF3 promotes mRNA translation. Cell Res *27*, 444-447.

Li, F., Zhao, D., Wu, J., and Shi, Y. (2014). Structure of the YTH domain of human YTHDF2 in complex with an m(6)A mononucleotide reveals an aromatic cage for m(6)A recognition. Cell Res *24*, 1490-1492.

Li, X., Xiong, X., Wang, K., Wang, L., Shu, X., Ma, S., and Yi, C. (2016a). Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. Nat Chem Biol *12*, 311-316.

Li, X., Xiong, X., and Yi, C. (2016b). Epitranscriptome sequencing technologies: decoding RNA modifications. Nat Methods *14*, 23-31.

Li, X., Xiong, X., Zhang, M., Wang, K., Chen, Y., Zhou, J., Mao, Y., Lv, J., Yi, D., Chen, X.W., *et al.* (2017b). Base-Resolution Mapping Reveals Distinct m(1)A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. Mol Cell *68*, 993-1005 e1009.

Li, Y., and Kiledjian, M. (2010). Regulation of mRNA decapping. Wiley Interdiscip Rev RNA *1*, 253-265.

Li, Z., Peng, Y., Li, J., Chen, Z., Chen, F., Tu, J., Lin, S., and Wang, H. (2020). N(6)-methyladenosine regulates glycolysis of cancer cells through PDK4. Nat Commun *11*, 2578.

Liang, S., Silva, J.C., Suska, O., Lukoszek, R., Almohammed, R., and Cowling, V.H. (2022). CMTR1 is recruited to transcription start sites and promotes ribosomal protein and histone gene expression in embryonic stem cells. Nucleic Acids Research *50*, 2905-2922.

Liao, J., Wei, Y., Liang, J., Wen, J., Chen, X., Zhang, B., and Chu, L. (2022). Insight into the structure, physiological function, and role in cancer of m6A readers-YTH domain-containing proteins. Cell Death Discov *8*, 137.

Lin, H. (2007). Nicotinamide adenine dinucleotide: beyond a redox coenzyme. Org Biomol Chem *5*, 2541-2554.

Lin, J., Lai, S., Jia, R., Xu, A., Zhang, L., Lu, J., and Ye, K. (2011). Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. Nature *469*, 559-563.

Lin, X., Chai, G., Wu, Y., Li, J., Chen, F., Liu, J., Luo, G., Tauler, J., Du, J., Lin, S., *et al.* (2019). RNA m(6)A methylation regulates the epithelial mesenchymal transition of cancer cells and translation of Snail. Nat Commun *10*, 2065.

Lin, Z., Hsu, P.J., Xing, X., Fang, J., Lu, Z., Zou, Q., Zhang, K.J., Zhang, X., Zhou, Y., Zhang, T., *et al.* (2017). Mettl3-/Mettl14-mediated mRNA N(6)-methyladenosine modulates murine spermatogenesis. Cell Res *27*, 1216-1230.

Liu, C., Shi, W., Becker, S.T., Schatz, D.G., Liu, B., and Yang, Y. (2021). Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. Science *373*, 1142-+.

Liu, J., Li, K., Cai, J., Zhang, M., Zhang, X., Xiong, X., Meng, H., Xu, X., Huang, Z., Peng, J., *et al.* (2020). Landscape and Regulation of m(6)A and m(6)Am Methylome across Human and Mouse Tissues. Mol Cell *77*, 426-440 e426.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., *et al.* (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol *10*, 93-95.

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

Lu, G., Zhang, J., Li, Y., Li, Z., Zhang, N., Xu, X., Wang, T., Guan, Z., Gao, G.F., and Yan, J. (2011). hNUDT16: a universal decapping enzyme for small nucleolar RNA and cytoplasmic mRNA. Protein Cell *2*, 64-73.

Luciano, D.J., and Belasco, J.G. (2015). NAD in RNA: unconventional headgear. Trends Biochem Sci *40*, 245-247.

Luciano, D.J., and Belasco, J.G. (2020). Np4A alarmones function in bacteria as precursors to RNA caps. Proc Natl Acad Sci U S A *117*, 3560-3567.

Luciano, D.J., Levenson-Palmer, R., and Belasco, J.G. (2019). Stresses that Raise Np4A Levels Induce Protective Nucleoside Tetraphosphate Capping of Bacterial RNA. Mol Cell *75*, 957-966 e958.

Luciano, D.J., Vasilyev, N., Richards, J., Serganov, A., and Belasco, J.G. (2017). A Novel RNA Phosphorylation State Enables 5' End-Dependent Degradation in Escherichia coli. Mol Cell *67*, 44-54 e46.

Luciano, D.J., Vasilyev, N., Richards, J., Serganov, A., and Belasco, J.G. (2018). Importance of a diphosphorylated intermediate for RppH-dependent RNA degradation. Rna Biol *15*, 703-706.

Luo, S., and Tong, L. (2014). Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. Proc Natl Acad Sci U S A *111*, 13834-13839.

Ma, H., Wang, X., Cai, J., Dai, Q., Natchiar, S.K., Lv, R., Chen, K., Lu, Z., Chen, H., Shi, Y.G., *et al.* (2019). N(6-)Methyladenosine methyltransferase ZCCHC4 mediates ribosomal RNA methylation. Nat Chem Biol *15*, 88-94.

Mao, X.D., and Shuman, S. (1994). Intrinsic Rna (Guanine-7) Methyltransferase Activity of the Vaccinia Virus Capping Enzyme D1 Subunit Is Stimulated by the D12 Subunit - Identification of Amino-Acid-Residues in the D1 Protein Required for Subunit Association and Methyl-Group Transfer. J Biol Chem *269*, 24472-24479.

Mao, Y., Dong, L., Liu, X.M., Guo, J., Ma, H., Shen, B., and Qian, S.B. (2019). m(6)A in mRNA coding regions promotes translation via the RNA helicase-containing YTHDC2. Nat Commun *10*, 5332.

Maquat, L.E., Tarn, W.Y., and Isken, O. (2010). The pioneer round of translation: features and functions. Cell *142*, 368-374.

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A.V., Patil, D.P., Linder, B., Pickering, B.F., Vasseur, J.J., Chen, Q., *et al.* (2017). Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. Nature *541*, 371-375.

Mauer, J., Sindelar, M., Despic, V., Guez, T., Hawley, B.R., Vasseur, J.J., Rentmeister, A., Gross, S.S., Pellizzoni, L., Debart, F., *et al.* (2019). FTO controls reversible m(6)Am RNA methylation during snRNA biogenesis. Nat Chem Biol *15*, 340-347.

McLennan, A.G. (2006). The Nudix hydrolase superfamily. Cell Mol Life Sci *63*, 123-143.

McLennan, A.G. (2013). Substrate ambiguity among the nudix hydrolases: biologically significant, evolutionary remnant, or both? Cell Mol Life Sci *70*, 373-385.

Mendel, M., Chen, K.M., Homolka, D., Gos, P., Pandey, R.R., McCarthy, A.A., and Pillai, R.S. (2018). Methylation of Structured RNA by the m(6)A Writer METTL16 Is Essential for Mouse Embryonic Development. Mol Cell *71*, 986-1000 e1011.

Mendel, M., Delaney, K., Pandey, R.R., Chen, K.M., Wenda, J.M., Vagbo, C.B., Steiner, F.A., Homolka, D., and Pillai, R.S. (2021). Splice site m(6)A methylation prevents binding of U2AF35 to inhibit RNA splicing. Cell *184*, 3125-3142 e3125.

Meyer, K.D., and Jaffrey, S.R. (2014). The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat Rev Mol Cell Biol *15*, 313-326.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell *149*, 1635-1646.

Moore, M.J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. Science *309*, 1514-1518.

Moshitch-Moshkovitz, S., Dominissini, D., and Rechavi, G. (2022). The epitranscriptome toolbox. Cell *185*, 764-776.

Mugridge, J.S., Coller, J., and Gross, J.D. (2018). Structural and molecular mechanisms for the control of eukaryotic 5'-3' mRNA decay. Nat Struct Mol Biol *25*, 1077-1085.

Muthukrishnan, S., Morgan, M., Banerjee, A.K., and Shatkin, A.J. (1976). Influence of 5'-terminal m7G and 2'--O-methylated residues on messenger ribonucleic acid binding to ribosomes. Biochemistry *15*, 5761-5768.

Muthukrishnan, S., Moss, B., Cooper, J.A., and Maxwell, E.S. (1978). Influence of 5'-terminal cap structure on the initiation of translation of vaccinia virus mRNA. J Biol Chem *253*, 1710-1715.

Nakagawa, T., Zhang, T., Kushi, R., Nakano, S., Endo, T., Nakagawa, M., Yanagihara, N., Zarkower, D., and Nakayama, K. (2017). Regulation of mitosis-meiosis transition by the ubiquitin ligase beta-TrCP in male germ cells. Development *144*, 4137-4147.

Nance, D.J., Satterwhite, E.R., Bhaskar, B., Misra, S., Carraway, K.R., and Mansfield, K.D. (2020). Characterization of METTL16 as a cytoplasmic RNA binding protein. PLoS One *15*, e0227647.

Natchiar, S.K., Myasnikov, A.G., Kratzat, H., Hazemann, I., and Klaholz, B.P. (2017). Visualization of chemical modifications in the human 80S ribosome structure. Nature *551*, 472-477.

Nayler, O., Hartmann, A.M., and Stamm, S. (2000). The ER repeat protein YT521-B localizes to a novel subnuclear compartment. J Cell Biol *150*, 949-962.

Ni, J., Tien, A.L., and Fournier, M.J. (1997). Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell *89*, 565-573.

Nojima, T., Hirose, T., Kimura, H., and Hagiwara, M. (2007). The interaction between cap-binding complex and RNA export factor is required for intronless mRNA export. J Biol Chem *282*, 15645-15651.

O'Farrell, P.A., Gonzalez, F., Zheng, W., Johnston, S.A., and Joshua-Tor, L. (1999). Crystal structure of human bleomycin hydrolase, a self-compartmentalizing cysteine protease. Structure *7*, 619-627.

Ohlstein, B., Lavoie, C.A., Vef, O., Gateff, E., and McKearin, D.M. (2000). The Drosophila cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles. Genetics *155*, 1809-1819.

Ohno, M., Sakamoto, H., and Shimura, Y. (1987). Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure. Proc Natl Acad Sci U S A *84*, 5187-5191.

Ozanick, S., Krecic, A., Andersland, J., and Anderson, J.T. (2005). The bipartite structure of the tRNA m1A58 methyltransferase from S. cerevisiae is conserved in humans. RNA *11*, 1281-1290.
Pabis, M., Neufeld, N., Steiner, M.C., Bojic, T., Shav-Tal, Y., and Neugebauer, K.M. (2013). The nuclear cap-binding complex interacts with the U4/U6.U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. RNA *19*, 1054-1063.

Pandey, R.R., Delfino, E., Homolka, D., Roithova, A., Chen, K.M., Li, L.Y., Franco, G., Vagbo, C.B., Taillebourg, E., Fauvarque, M.O., *et al.* (2020). The Mammalian Cap-Specific m(6)Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues. Cell Rep *32*.

Park, O.H., Ha, H., Lee, Y., Boo, S.H., Kwon, D.H., Song, H.K., and Kim, Y.K. (2019). Endoribonucleolytic Cleavage of m(6)A-Containing RNAs by RNase P/MRP Complex. Mol Cell *74*, 494-507 e498.

Parker, R. (2012). RNA degradation in Saccharomyces cerevisae. Genetics *191*, 671-702.

Patil, D.P., Chen, C.K., Pickering, B.F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S.R. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature *537*, 369-373.

Patil, D.P., Pickering, B.F., and Jaffrey, S.R. (2018). Reading m(6)A in the Transcriptome: m(6)A-Binding Proteins. Trends Cell Biol *28*, 113-127.

Pendleton, K.E., Chen, B., Liu, K., Hunter, O.V., Xie, Y., Tu, B.P., and Conrad, N.K. (2017). The U6 snRNA m(6)A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell *169*, 824-835 e814.

Perry, R.P., Kelley, D.E., Friderici, K., and Rottman, F. (1975). The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5' terminus. Cell *4*, 387-394.

Ping, X.L., Sun, B.F., Wang, L., Xiao, W., Yang, X., Wang, W.J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.S., *et al.* (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell Res *24*, 177-189.

Piton, J., Larue, V., Thillier, Y., Dorleans, A., Pellegrini, O., Li de la Sierra-Gallay, I., Vasseur, J.J., Debart, F., Tisne, C., and Condon, C. (2013). Bacillus subtilis RNA deprotection enzyme RppH recognizes guanosine in the second position of its substrates. Proc Natl Acad Sci U S A *110*, 8858-8863.

Prabu, J.R., Muller, M., Thomae, A.W., Schussler, S., Bonneau, F., Becker, P.B., and Conti, E. (2015). Structure of the RNA Helicase MLE Reveals the Molecular Mechanisms for Uridine Specificity and RNA-ATP Coupling. Mol Cell *60*, 487-499.

Preiss, T., and Hentze, M.W. (1998). Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. Nature *392*, 516-520.

Preiss, T., and Hentze, M.W. (1999). From factors to mechanisms: translation and translational control in eukaryotes. Curr Opin Genet Dev *9*, 515-521.

Ramanathan, A., Robb, G.B., and Chan, S.H. (2016). mRNA capping: biological functions and applications. Nucleic Acids Res *44*, 7511-7526.

Reddy, R., Singh, R., and Shimba, S. (1992). Methylated cap structures in eukaryotic RNAs: structure, synthesis and functions. Pharmacol Ther *54*, 249-267.

Rich, A., and RajBhandary, U.L. (1976). Transfer RNA: molecular structure, sequence, and properties. Annu Rev Biochem *45*, 805-860.

Rich, P.R. (2003). The molecular machinery of Keilin's respiratory chain. Biochem Soc Trans *31*, 1095-1105.

Richards, J., Liu, Q., Pellegrini, O., Celesnik, H., Yao, S., Bechhofer, D.H., Condon, C., and Belasco, J.G. (2011). An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in Bacillus subtilis. Mol Cell *43*, 940-949.

Ries, R.J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B.F., Patil, D.P., Kwak, H., Lee, J.H., and Jaffrey, S.R. (2019). m(6)A enhances the phase separation potential of mRNA. Nature *571*, 424-428.

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

Roundtree, I.A., Luo, G.Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., *et al.* (2017b). YTHDC1 mediates nuclear export of N(6)-methyladenosine methylated mRNAs. Elife *6*.

Rubio, R.M., Depledge, D.P., Bianco, C., Thompson, L., and Mohr, I. (2018). RNA m(6) A modification enzymes shape innate responses to DNA by regulating interferon beta. Genes Dev *32*, 1472-1484.

Ruzicka, K., Zhang, M., Campilho, A., Bodi, Z., Kashif, M., Saleh, M., Eeckhout, D., El-Showk, S., Li, H., Zhong, S., *et al.* (2017). Identification of factors required for m(6) A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. New Phytol *215*, 157-172.

Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., Erlacher, M., Rossmanith, W., Stern-Ginossar, N., and Schwartz, S. (2017). The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. Nature *551*, 251-255.

Schaefer, M., Pollex, T., Hanna, K., and Lyko, F. (2009). RNA cytosine methylation analysis by bisulfite sequencing. Nucleic Acids Res *37*, e12.

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., and Lyko, F. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes Dev *24*, 1590-1595.

Schaffrath, R., and Leidel, S.A. (2017). Wobble uridine modifications-a reason to live, a reason to die?! Rna Biol *14*, 1209-1222.

Schimmel, P. (2018). The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. Nat Rev Mol Cell Biol *19*, 45-58.

Schwartz, D.R., Homanics, G.E., Hoyt, D.G., Klein, E., Abernethy, J., and Lazo, J.S. (1999). The neutral cysteine protease bleomycin hydrolase is essential for epidermal integrity and bleomycin resistance. Proc Natl Acad Sci U S A *96*, 4680-4685.

Selmi, T., Hussain, S., Dietmann, S., Heiss, M., Borland, K., Flad, S., Carter, J.M., Dennison, R., Huang, Y.L., Kellner, S., *et al.* (2021). Sequence- and structure-specific cytosine-5 mRNA methylation by NSUN6. Nucleic Acids Res *49*, 1006-1022.

Sendinc, E., Valle-Garcia, D., Dhall, A., Chen, H., Henriques, T., Navarrete-Perea, J., Sheng, W.Q., Gygi, S.P., Adelman, K., and Shi, Y. (2019). PCIF1 Catalyzes m6Am mRNA Methylation to Regulate Gene Expression. Mol Cell *75*, 620-+.

Sharma, S., Langhendries, J.L., Watzinger, P., Kotter, P., Entian, K.D., and Lafontaine, D.L. (2015). Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMPD1. Nucleic Acids Res *43*, 2242-2258.

Shi, H., Wang, X., Lu, Z., Zhao, B.S., Ma, H., Hsu, P.J., Liu, C., and He, C. (2017). YTHDF3 facilitates translation and decay of N(6)-methyladenosine-modified RNA. Cell Res *27*, 315-328.

Shi, H., Wei, J., and He, C. (2019). Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. Mol Cell *74*, 640-650.

Shu, S., Paruchuru, L.B., Tay, N.Q., Chua, Y.L., Foo, A.S.Y., Yang, C.M., Liong, K.H., Koh, E.G.L., Lee, A., Nechushtan, H., *et al.* (2019). Ap4A Regulates Directional Mobility and Antigen Presentation in Dendritic Cells. iScience *16*, 524-534.

Shuman, S., and Hurwitz, J. (1981). Mechanism of mRNA capping by vaccinia virus guanylyltransferase: characterization of an enzyme--guanylate intermediate. Proc Natl Acad Sci U S A *78*, 187-191.

Shuman, S., Liu, Y.Z., and Schwer, B. (1994). Covalent Catalysis in Nucleotidyl Transfer-Reactions - Essential Motifs in Saccharomyces-Cerevisiae Rna Capping Enzyme Are Conserved in Schizosaccharomyces-Pombe and Viral Capping Enzymes and among Polynucleotide Ligases. P Natl Acad Sci USA *91*, 12046-12050.

Sledz, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. Elife *5*.

Sloan, K.E., Warda, A.S., Sharma, S., Entian, K.D., Lafontaine, D.L.J., and Bohnsack, M.T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. Rna Biol *14*, 1138-1152.

Smietanski, M., Werner, M., Purta, E., Kaminska, K.H., Stepinski, J., Darzynkiewicz, E., Nowotny, M., and Bujnicki, J.M. (2014). Structural analysis of human 2'-O-ribose methyltransferases involved in mRNA cap structure formation. Nat Commun *5*, 3004.

Smith, R.W., Blee, T.K., and Gray, N.K. (2014). Poly(A)-binding proteins are required for diverse biological processes in metazoans. Biochem Soc Trans *42*, 1229-1237.

Soh, Y.Q.S., Mikedis, M.M., Kojima, M., Godfrey, A.K., de Rooij, D.G., and Page, D.C. (2017). Meioc maintains an extended meiotic prophase I in mice. PLoS Genet *13*, e1006704.

Song, M.G., Bail, S., and Kiledjian, M. (2013). Multiple Nudix family proteins possess mRNA decapping activity. RNA *19*, 390-399.

Song, M.G., Li, Y., and Kiledjian, M. (2010). Multiple mRNA decapping enzymes in mammalian cells. Mol Cell *40*, 423-432.

Squires, J.E., Patel, H.R., Nousch, M., Sibbritt, T., Humphreys, D.T., Parker, B.J., Suter, C.M., and Preiss, T. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res *40*, 5023-5033.

Srouji, J.R., Xu, A., Park, A., Kirsch, J.F., and Brenner, S.E. (2017). The evolution of function within the Nudix homology clan. Proteins *85*, 775-811.

Stoilov, P., Rafalska, I., and Stamm, S. (2002). YTH: a new domain in nuclear proteins. Trends Biochem Sci *27*, 495-497.

Takagi, T., Moore, C.R., Diehn, F., and Buratowski, S. (1997). An RNA 5'-triphosphatase related to the protein tyrosine phosphatases. Cell *89*, 867-873.

Takeda, A., Masuda, Y., Yamamoto, T., Hirabayashi, T., Nakamura, Y., and Nakaya, K. (1996). Cloning and analysis of cDNA encoding rat bleomycin hydrolase, a DNA-binding cysteine protease. J Biochem *120*, 353-359.

Taoka, M., Nobe, Y., Yamaki, Y., Yamauchi, Y., Ishikawa, H., Takahashi, N., Nakayama, H., and Isobe, T. (2016). The complete chemical structure of Saccharomyces cerevisiae rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. Nucleic Acids Res *44*, 8951-8961.

Thalhammer, A., Bencokova, Z., Poole, R., Loenarz, C., Adam, J., O'Flaherty, L., Schodel, J., Mole, D., Giaslakiotis, K., Schofield, C.J., *et al.* (2011). Human AlkB homologue 5 is a nuclear 2-oxoglutarate dependent oxygenase and a direct target of hypoxia-inducible factor 1alpha (HIF-1alpha). PLoS One *6*, e16210.

Theler, D., Dominguez, C., Blatter, M., Boudet, J., and Allain, F.H. (2014). Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. Nucleic Acids Res *42*, 13911-13919.

Thorne, N.M., Hankin, S., Wilkinson, M.C., Nunez, C., Barraclough, R., and McLennan, A.G. (1995). Human diadenosine 5',5"'-P1,P4-tetraphosphate pyrophosphohydrolase is a member of the MutT family of nucleotide pyrophosphatases. Biochem J *311 (Pt 3)*, 717-721.

Tollervey, D., Lehtonen, H., Jansen, R., Kern, H., and Hurt, E.C. (1993). Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. Cell *72*, 443-457.

Valkov, E., Jonas, S., and Weichenrieder, O. (2017). Mille viae in eukaryotic mRNA decapping. Curr Opin Struct Biol *47*, 40-51.

van Dijk, E., Cougot, N., Meyer, S., Babajko, S., Wahle, E., and Seraphin, B. (2002). Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. Embo J *21*, 6915-6924.

van Nues, R.W., Granneman, S., Kudla, G., Sloan, K.E., Chicken, M., Tollervey, D., and Watkins, N.J. (2011). Box C/D snoRNP catalysed methylation is aided by additional pre-rRNA base-pairing. Embo J *30*, 2420-2430.

van Tran, N., Ernst, F.G.M., Hawley, B.R., Zorbas, C., Ulryck, N., Hackert, P., Bohnsack, K.E., Bohnsack, M.T., Jaffrey, S.R., Graille, M., *et al.* (2019). The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. Nucleic Acids Res *47*, 7719-7733.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., *et al.* (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res *50*, D439-D444.

Viswanathan, T., Arya, S., Chan, S.H., Qi, S., Dai, N., Misra, A., Park, J.G., Oladunni, F., Kovalskyy, D., Hromas, R.A., *et al.* (2020). Structural basis of RNA cap modification by SARS-CoV-2. Nat Commun *11*, 3718.

Walters, R.W., Matheny, T., Mizoue, L.S., Rao, B.S., Muhlrad, D., and Parker, R. (2017). Identification of NAD+ capped mRNAs in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *114*, 480-485.

Wang, C., Zhu, Y., Bao, H., Jiang, Y., Xu, C., Wu, J., and Shi, Y. (2016a). A novel RNA-binding mode of the YTH domain reveals the mechanism for recognition of determinant of selective removal by Mmi1. Nucleic Acids Res *44*, 969-982.

Wang, J., Alvin Chew, B.L., Lai, Y., Dong, H., Xu, L., Balamkundu, S., Cai, W.M., Cui, L., Liu, C.F., Fu, X.Y., *et al.* (2019). Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. Nucleic Acids Res *47*, e130.

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

Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., Gong, Z., Wang, Q., Huang, J., Tang, C., *et al.* (2016c). Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex. Nature *534*, 575-578.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., *et al.* (2014a). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117-120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388-1399.

Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014b). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat Cell Biol *16*, 191-198.

Wang, Z., Jiao, X., Carr-Schmid, A., and Kiledjian, M. (2002). The hDcp2 protein is a mammalian mRNA decapping enzyme. Proc Natl Acad Sci U S A *99*, 12663-12668.

Warda, A.S., Kretschmer, J., Hackert, P., Lenz, C., Urlaub, H., Hobartner, C., Sloan, K.E., and Bohnsack, M.T. (2017). Human METTL16 is a N(6)-methyladenosine (m(6)A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. EMBO Rep *18*, 2004-2014.

Watkins, N.J., and Bohnsack, M.T. (2012). The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. Wiley Interdiscip Rev RNA *3*, 397-414.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737-738.

Wei, C., Gershowitz, A., and Moss, B. (1975a). N6, O2'-dimethyladenosine a novel methylated ribonucleoside next to the 5' terminal of animal cell and virus mRNAs. Nature *257*, 251-253.

Wei, C.M., Gershowitz, A., and Moss, B. (1975b). Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. Cell *4*, 379-386.

Wei, J., Liu, F., Lu, Z., Fei, Q., Ai, Y., He, P.C., Shi, H., Cui, X., Su, R., Klungland, A., *et al.* (2018). Differential m(6)A, m(6)Am, and m(1)A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. Mol Cell *71*, 973-985 e975.

Wen, J., Lv, R., Ma, H., Shen, H., He, C., Wang, J., Jiao, F., Liu, H., Yang, P., Tan, L., *et al.* (2018). Zc3h13 Regulates Nuclear RNA m(6)A Methylation and Mouse Embryonic Stem Cell Self-Renewal. Mol Cell *69*, 1028-1038 e1026.

Werner, M., Purta, E., Kaminska, K.H., Cymerman, I.A., Campbell, D.A., Mittra, B., Zamudio, J.R., Sturm, N.R., Jaworski, J., and Bujnicki, J.M. (2011). 2'-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. Nucleic Acids Res *39*, 4756-4768.
Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*.

Winkler, R., Gillis, E., Lasman, L., Safra, M., Geula, S., Soyris, C., Nachshon, A., Tai-Schmiedel, J., Friedman, N., Le-Trilling, V.T.K., *et al.* (2019). m(6)A modification controls the innate immune response to infection by targeting type I interferons. Nat Immunol *20*, 173-182.

Wojtas, M.N., Pandey, R.R., Mendel, M., Homolka, D., Sachidanandam, R., and Pillai, R.S. (2017). Regulation of m(6)A Transcripts by the 3'-->5' RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. Mol Cell *68*, 374-387 e312.

Wolf, J., and Passmore, L.A. (2014). mRNA deadenylation by Pan2-Pan3. Biochem Soc Trans *42*, 184-187.

Wu, B., Peisley, A., Richards, C., Yao, H., Zeng, X., Lin, C., Chu, F., Walz, T., and Hur, S. (2013). Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5. Cell *152*, 276-289.

Wu, R., Chen, Y., Liu, Y., Zhuang, L., Chen, W., Zeng, B., Liao, X., Guo, G., Wang, Y., and Wang, X. (2021). m6A methylation promotes white-to-beige fat transition by facilitating Hif1a translation. EMBO Rep *22*, e52348.

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., and Tong, L. (2009). Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. Nature *458*, 784-788.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.S., Hao, Y.J., Sun, B.F., Sun, H.Y., Li, A., Ping, X.L., Lai, W.Y., *et al.* (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Mol Cell *61*, 507-519.

Xu, C., Liu, K., Ahmed, H., Loppnau, P., Schapira, M., and Min, J. (2015). Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. J Biol Chem *290*, 24902-24913.

Xu, C., Wang, X., Liu, K., Roundtree, I.A., Tempel, W., Li, Y., Lu, Z., He, C., and Min, J. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. Nat Chem Biol *10*, 927-929.

Xu, H., Dzhashiashvili, Y., Shah, A., Kunjamma, R.B., Weng, Y.L., Elbaz, B., Fei, Q., Jones, J.S., Li, Y.I., Zhuang, X., *et al.* (2020). m(6)A mRNA Methylation Is Essential for Oligodendrocyte Maturation and CNS Myelination. Neuron *105*, 293-309 e295.

Xu, H.E., and Johnston, S.A. (1994). Yeast bleomycin hydrolase is a DNA-binding cysteine protease. Identification, purification, biochemical characterization. J Biol Chem *269*, 21177-21183.

Yamashita, A., Shichino, Y., Tanaka, H., Hiriart, E., Touat-Todeschini, L., Vavasseur, A., Ding, D.Q., Hiraoka, Y., Verdel, A., and Yamamoto, M. (2012). Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. Open Biol *2*, 120014.

Yang, X., Yang, Y., Sun, B.F., Chen, Y.S., Xu, J.W., Lai, W.Y., Li, A., Wang, X., Bhattarai, D.P., Xiao, W., *et al.* (2017). 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. Cell Res *27*, 606-625.

Yang, Y., Shen, F., Huang, W., Qin, S., Huang, J.T., Sergi, C., Yuan, B.F., and Liu, S.M. (2019). Glucose Is Involved in the Dynamic Regulation of m6A in Patients With Type 2 Diabetes. J Clin Endocrinol Metab *104*, 665-673.

Yavuz, H., Bertoli-Avella, A.M., Alfadhel, M., Al-Sannaa, N., Kandaswamy, K.K., Al-Tuwaijri, W., Rolfs, A., Brandau, O., and Bauer, P. (2018). A founder nonsense variant in NUDT2 causes a recessive neurodevelopmental disorder in Saudi Arab children. Clin Genet *94*, 393-395.

Yoon, K.J., Ringeling, F.R., Vissers, C., Jacob, F., Pokrass, M., Jimenez-Cyrus, D., Su, Y., Kim, N.S., Zhu, Y., Zheng, L., *et al.* (2017). Temporal Control of Mammalian Cortical Neurogenesis by m(6)A Methylation. Cell *171*, 877-889 e817.

Yu, X., Vandivier, L.E., and Gregory, B.D. (2021). NAD-seq for profiling the NAD(+) capped transcriptome of Arabidopsis thaliana. STAR Protoc *2*, 100901.

Yue, Y., Liu, J., Cui, X., Cao, J., Luo, G., Zhang, Z., Cheng, T., Gao, M., Shu, X., Ma, H., *et al.* (2018). VIRMA mediates preferential m(6)A mRNA methylation in 3'UTR and near stop codon and associates with alternative polyadenylation. Cell Discov *4*, 10.

Zaccara, S., and Jaffrey, S.R. (2020). A Unified Model for the Function of YTHDF Proteins in Regulating m(6)A-Modified mRNA. Cell *181*, 1582-1595 e1518.

Zhang, C., Samanta, D., Lu, H., Bullen, J.W., Zhang, H., Chen, I., He, X., and Semenza, G.L. (2016a). Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m(6)A-demethylation of NANOG mRNA. Proc Natl Acad Sci U S A *113*, E2047-2056.
Zhang, D., Liu, Y., Wang, Q., Guan, Z., Wang, J., Liu, J., Zou, T., and Yin, P. (2016b). Structural basis of prokaryotic NAD-RNA decapping by NudC. Cell Res *26*, 1062-1066.

Zhang, S., Zhao, B.S., Zhou, A., Lin, K., Zheng, S., Lu, Z., Chen, Y., Sulman, E.P., Xie, K., Bogler, O., *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*, 591-606 e596.

Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.M., Li, C.J., Vagbo, C.B., Shi, Y., Wang, W.L., Song, S.H., *et al.* (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell *49*, 18-29.

Zheng, S.S., Hausmann, S., Hausmann, P., Liu, Q.S., Ghosh, A., Schwer, B., Lima, C.D., and Shuman, S. (2006). Mutational analysis of Encephalitozoon cuniculi mRNA cap (guanine-N7) methyltransferase, structure of the enzyme bound to sinefungin, and evidence that cap methyltransferase is the target of sinefungin's antifungal activity. J Biol Chem *281*, 35904-35913.
Zheng, W., and Johnston, S.A. (1998). The nucleic acid binding activity of bleomycin hydrolase is involved in bleomycin detoxification. Mol Cell Biol *18*, 3580-3585.

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., and Fray, R.G. (2008). MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. Plant Cell *20*, 1278-1288.

Zhou, K.I., Shi, H., Lyu, R., Wylder, A.C., Matuszek, Z., Pan, J.N., He, C., Parisien, M., and Pan, T. (2019). Regulation of Co-transcriptional Pre-mRNA Splicing by m(6)A through the Low-Complexity Protein hnRNPG. Mol Cell *76*, 70-81 e79.

Zhu, T., Roundtree, I.A., Wang, P., Wang, X., Wang, L., Sun, C., Tian, Y., Li, J., He, C., and Xu, Y. (2014). Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. Cell Res *24*, 1493-1496.

Zinder, J.C., and Lima, C.D. (2017). Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. Genes Dev *31*, 88-100.

Zou, F., Tu, R., Duan, B., Yang, Z., Ping, Z., Song, X., Chen, S., Price, A., Li, H., Scott, A., *et al.* (2020). Drosophila YBX1 homolog YPS promotes ovarian germ line stem cell development by preferentially recognizing 5-methylcytosine RNAs. Proc Natl Acad Sci U S A *117*, 3603-3609.