

**CHARACTERIZATIONS OF  
METHYLTRANSFERASE-LIKE PROTEIN 2  
IN RNA MODIFICATION**

**XU LUANG**

*(Bachelor's Degree, Major in Veterinary Medicine,  
Zhejiang University)*

**A THESIS SUBMITTED FOR THE DEGREE OF**

**DOCTOR OF PHILOSOPHY**

**CANCER SCIENCE INSTITUTE, SINGAPORE**

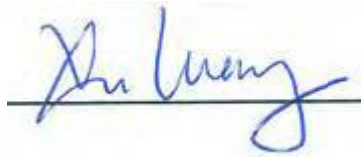
**NATIONAL UNIVERSITY OF SINGAPORE**

**2015**

## DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information that have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



Xu Luang

9 August 2015

## ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, Professor Fu Xin-Yuan. It was him that I got the chance to join his lab and got the chance to do research.

Prof. Fu is such a passionate scientist that I learn from him through daily life. Although his science career is delayed due to historical reasons, Prof. Fu still did excellent work on RNA splicing and STAT family proteins when he was doing graduate studies and post-doc training at U.S. His past research stories always inspire me on how to do research. Prof. Fu also invited many excellent scientists to give talks at Singapore and it give chance to communicate with top scientists in the field. He also managed to collaborate with top scientists for the purpose of moving forward the boundaries of science and encourage us to do so. I cannot say more thanks to his support to me when I failed to do a good job in course work during my first year to second year. His encouragement drive me through this journey.

I want to give thanks to Dr. LIU Xinyu, who taught me basic laboratory skills when I just join the group, including cloning, immunoblotting. He first set up and share the common online folders for all the reagents / data for the project. It really benefits me a lot at the beginning. I want to give thanks to all other lab members, to Dr. LIANG Junxin and Mr. Kyaw Soe Oo (John) for helping each other, sharing reagents, and having discussion on the project, to our lab manager QU Ge for her maintenance of the lab, to Dr. Lufei for his scientific advice.

I would like to give my sincere thanks to Professor Peter Dedon at SMART/MIT and his lab members: Dr. Yok Hian Chionh, Dr. Megan McBee,

and Mr. Fabian who kindly helped me on HPLC and Mass Spec analysis of 3-methylcytosine before I submitted thesis in August. I didn't want to quit or stop the project after submitting thesis, I want to thank Pete for this acceptance of me for joining the group (although not officially) to do more: getting advice, attending inspiring lab meetings, learning techniques from lab people. I learned really a lot. Special thanks gave to Dr. Chionh, Maggie Cai and Cui Liang for teaching me HPLC, Mass Spec, and many happy talks since last August.

I want to give thanks to Associate Professor GAO Yonggui for giving me the opportunity to work at his lab, to all your lab members who gave me help. I want to give special thanks to Mr. YE Fuzhou (now Dr. Ye), who patiently taught me FPLC purification of proteins, helped growing crystals for METTL2, 6, 8 proteins for so much endeavour and encouraged me through the journey.

I want to give thanks to all those people from various labs for the helping on this project, including but not limited to:

Dr. Leah Vardy at IMB, who gave me the chance to perform polysome profiling for several times under the help of Ms. Anisa, Ms. Vonny Leo;

Professor He Chuan at University of Chicago and his lab members on performing PAR-CLIP and discussions;

Dr. Guo Huili and Ms. Lan Jiemin for ribosome profiling.

My thanks go to Dr. Ernesto Guccione and Dr. Wu Qiang for your agreement for being my Thesis Advisory Committee members, your insightful suggestions and more importantly, your encouragements help me through this.

Of course, I want to give thanks to my parents for all your love to me all those years and letting me to pursue a career for my own will, a thousand words is not enough for my love to you both. I didn't contact with you that often, just some online chatting / telephone calls once a week, or even less frequently, it is my fault and I will fix this problem.

I also want to thank myself for my hard work, my passion on this project, my honesty towards data and my achievements from this journey. I have tried my best.

Last but not the least, I want to give special thanks to Cancer Science Institute, Singapore, which is the home to me in Singapore during the past years. Thanks to Prof. Daniel Tenen, Prof. Lorenz Poellinger, Prof. Hew Choy Leong for interviewing me at Hangzhou, China in Autumn 2010, to Ms. Selena GAN, Ms Yvette Soon, Ms. YAO Huijun, Mr. Rex NG Swee Siang, Ms. Dee Pham and all the CSI staff for providing laboratory and administrative support, to all the people who gave me help.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS.....</b>	<b>I</b>
<b>TABLE OF CONTENTS .....</b>	<b>IV</b>
<b>SUMMARY .....</b>	<b>VIII</b>
<b>LIST OF TABLES.....</b>	<b>IX</b>
<b>LIST OF FIGURES.....</b>	<b>XI</b>
<b>LIST OF SYMBOLS.....</b>	<b>XIV</b>
<b>1 CHAPTER 1. INTRODUCTION.....</b>	<b>1</b>
1.1 The importance of being modified: diversity and prevalence of RNA modifications .....	1
1.2 Detection and quantification of RNA modifications .....	11
1.3 Identification of enzymes responsible for specific RNA modifications .....	18
1.4 Methylation and methyltransferase (MTase).....	19
1.5 3-methylcytidine (m <sup>3</sup> C) and its corresponding methyltransferase...25	
1.6 m <sup>3</sup> C in tRNA has shown to be involved in codon-biased translation 28	
1.7 Phylogeny and potential functions of tRNA 32–38 pair .....	29
1.8 The study of translational control of gene expression by ribosome profiling.....	33
<b>2 CHAPTER 2. MATERIALS AND METHODS .....</b>	<b>37</b>

2.1	Plasmids constructions .....	37
2.2	Cell culture and treatment .....	37
2.3	Transfection .....	38
2.4	Real-time PCR .....	39
2.5	Western Blot .....	40
2.6	CRISPR/Cas9 gene targeting in both human and mice.....	44
2.7	Modifed quantitative PCR to measure m <sup>3</sup> C modification (qPCR- m <sup>3</sup> C) 45	
2.8	Primer extension analysis of modifications in RNA.....	47
2.9	Recombinant protein purification for MTase assay and growing crystals .....	49
2.10	<i>In Vitro</i> Transcription (IVT) of tRNA.....	50
2.11	Purification of specific tRNA species by solid-phase DNA probes	51
2.12	Reconstitution of 3-methylcytosine by recombinant Mettl2B.....	52
2.13	Methyltransferase assay using tritium labelled S- adenosylmethionine ( <sup>3</sup> H-SAM ) .....	53
2.14	Mass spectrometry in this project .....	54
2.15	Size exclusion HPLC and LC-MS/MS for quantification of modified ribonucleosides .....	54
2.16	RNA immunoprecipitation (RIP) assay .....	55
2.17	Polysome profiling-.....	56
2.18	Immunoprecipitation .....	60
2.19	Immunofluorescence staining for cultured cell lines (IF-IC) .....	61
2.20	Codon-run reporter constructs .....	61
2.21	Scripts for calculating codon numbers for mRNA .....	63

2.22	Clonogenic assay.....	66
<b>3</b>	<b>CHAPTER 3. RESULTS .....</b>	<b>68</b>
3.1	Sequence alignment of METTL2 with its paralogs and homologs..	68
3.2	Subcellular localizations of METTL2, 6 and 8.....	70
3.3	Generation of Mettl2A&2B double KO in HEK293T cell line .....	73
3.4	Recombinant METTL2 protein is prone to aggregate under normal purification conditions .....	77
3.5	tRNA <sup>Thr</sup> AGU and tRNA <sup>Thr</sup> UGU isoacceptors are substrates of METTL2 .....	81
3.6	Other substrates for METTL2, 6, 8 remain elusive .....	84
3.7	Ectopic expression of METTL2 in KO cells restore the modification which could block primer extension.....	86
3.8	Recombinant METTL2 could restore the modification on tRNA Thr <sup>AGU</sup> and Thr <sup>UGU</sup> .....	87
3.9	<i>in vitro</i> transcribed tRNAs are not methylated by recombinant METTL2 .....	88
3.10	A quantitative PCR based assay for m <sup>3</sup> C quantification.....	89
3.11	LC-MS/MS analysis of m <sup>3</sup> C contents in various RNA species.....	92
3.12	Identification of RNA species binding to METTL2, 8 by PAR-CLIP and validation by RIP, RT-PCR.....	98
3.13	METTL2 could methylate itself <i>in vitro</i> .....	101
3.14	Identification of METTL2, 6, 8 protein complex by Flag-IP mass spectrometry .....	104
3.15	METTL6 interact with Seryl-tRNA synthetase (SARS) in a RNA dependent manner.....	105



3.16	METTL2, 6, and 8 in global translation .....	108
3.17	Loss of METTL2 confers higher sensitivity to translation inhibitors and one alkylation agent .....	111
3.18	Codon-run to study effects of 3-methylcytosine on translation.....	113
3.19	Phylogeny and structural analysis tRNA 32-38 pair.....	116
<b>4</b>	<b>CHAPTER 4. DISCUSSIONS AND FUTURE WORK .....</b>	<b>122</b>
4.1	Limitation of Primer Extension assay .....	122
4.2	Limitation of RIP assay .....	123
4.3	Limitation of ribosome profiling .....	123
4.4	Self-methylation properties of METTL2, 6, 8.....	126
4.5	Methyl group on polar contacts .....	128
4.6	Enrichment of threonine tRNA <sup>AGU</sup> and tRNA <sup>UGU</sup> isoacceptors in eukaryotes .....	132
4.7	Future work for identification of all substrates and modification sites by METTL2 family proteins .....	134
4.8	Future structural study of METTL2 family proteins with co-factors and substrates. ....	135
4.9	Future work on m <sup>3</sup> C and codon biased translation .....	137
4.10	Models for m <sup>3</sup> C in translation and other cellular process .....	139
	<b>BIBLIOGRAPHY.....</b>	<b>141</b>
	<b>APPENDICES .....</b>	<b>158</b>

## SUMMARY

The biochemistry and biological functions of many RNA modifications remain elusive, especially in higher eukaryotes. 3-methylcytidine ( $m^3C$ ) is one kind of methylation modification located on certain tRNA species and hypothesized to affect translation.

This thesis reports that human METTL2 proteins catalyse  $m^3C$  formation at position 32 of threonine tRNA<sup>AGU</sup> and tRNA<sup>UGU</sup> isoacceptors *in vitro* and *in vivo*. LC-MS/MS analysis showed both human and mouse METTL2 contribute to 35% - 40% of total cellular  $m^3C$  formation, while reduction of methyltransferase-like 6 protein (METTL6), another paralog of METTL2, reduce  $m^3C$  contents by 10% - 15%. METTL6 is found to interact with seryl-tRNA synthetase (SARS) in an RNA dependent manner, suggesting serine tRNAs as potential METTL6 substrates. Reduction or deletion of *Mettl2*, *6* lead to significant decrease in cell proliferation rate and global translation changes in several cancer cell lines. The reduction also elicit higher sensitivity to one alkylating agent, MMS, consistent with the published data in yeast. Phylogenic and structural analysis of tRNA 32-38 pair indicates a potential role of  $m^3C$  in modulating the interaction between tRNA and the translation machinery. These studies support METTL2, *6* as RNA  $m^3C$  methyltransferase in both human and mouse and suggest a role of those proteins in fine-tuning translation.

## LIST OF TABLES

Table 1-1 Examples for methyltransferase-like proteins .....	24
Table 2-1 Resolving gel recipe.....	40
Table 2-2 Stacking gel recipe.....	41
Table 2-3 10X Running Buffer .....	42
Table 2-4 1X Running Buffer .....	42
Table 2-5 1X Transfer Buffer .....	42
Table 2-6 Gel sandwich in transfer .....	43
Table 2-7 5' end labelling of DNA probes for primer extension.....	48
Table 2-8 Oligos used for primer extension for human cytoplasmic and Mitochondria tRNAs .....	48
Table 2-9 In Vitro Transcription Reactions .....	52
Table 2-10 In vitro reconstitution of m3C .....	52
Table 2-11 Sucrose stock solution.....	57
Table 2-12 2X Resuspension Buffer (RSB) Recipe .....	57
Table 2-13 1x RSB with Cycloheximide .....	57
Table 2-14 1X Lysis Buffer .....	58
Table 2-15 Oligos for cloning Renilla luciferase and Firefly luciferase .....	62
Table 2-16 Oligos for Codon Runs .....	62
Table 3-1 mRNA transcripts identified as binding partners of METTL8 by PAR-CLIP.....	99
Table 3-2 METTL2, 6, 8 protein complex identified by Flag-IP Mass Spec	105
Table 3-3 Oligos for cloning Renilla luciferase and Firefly luciferase .....	113
Table 3-4 tRNA 32–38 base pairs frequency in three domains of life .....	117

Table 3-5 Threonine tRNA 32–38 base pairs frequency in three domains of life .....	118
Table 3-6 Summary of each 32-38 pair in the Figure 3-20.....	120
Table 4-1 Threonine codon percentage Ranking (content >16%) in human mRNAs .....	138
Table 4-2 Potential functions of 3-methylcytosine in various Cellular processes .....	140

## LIST OF FIGURES

Figure 1-1 A schematic of tRNA numbering and modifications found in cytoplasmic tRNA in <i>S. cerevisiae</i> .....	4
Figure 1-2 Principles of Primer extension assay and examples .....	13
Figure 1-3 Chemical structure of cytidine, 3-methylcytidine (m <sup>3</sup> C), and proposed mechanism for m <sup>3</sup> C formation .....	25
Figure 1-4 m <sup>3</sup> C occurs at position 32 of several tRNA species .....	27
Figure 3-1 Sequence alignment of the METTL2 with its paralogs and homologs.....	70
Figure 3-2 Subcellular localizations of METTL2, 6, 8 .....	72
Figure 3-3 CRISPR/Cas9 mediated Mettl2A & 2B gene knockout in 293T cells.....	76
Figure 3-4 Purification of METTL2 by Affinity and Size Exclusion Chromatography.....	81
Figure 3-5 Schematic diagram of human tRNAs and corresponding probes used in primer extension assay.....	82
Figure 3-6 Loss of METTL2A & 2B abolished the modification at position 32 in tRNA <sup>Thr</sup> AGU and tRNA <sup>Thr</sup> UGU isoacceptors .....	84
Figure 3-7 Reduction or loss of METTL2, 6 or 8 has no effects on the position 32 modification in other tRNA species tested.....	85
Figure 3-8 Ectopic expression METTL2 protein in KO cells restore the position 32 modification in tRNA Thr <sup>AGU</sup> and Thr <sup>UGU</sup> .....	86
Figure 3-9 <i>In Vitro</i> reconstitution of modification on tRNA Thr <sup>AGU</sup> and Thr <sup>UGU</sup> by recombinant METTL2 .....	88

Figure 3-10 METTL2B fail to restore position 32 modification on <i>in vitro</i> transcribed tRNA Thr <sup>AGU</sup> and Thr <sup>UGU</sup> .....	89
Figure 3-11 Quantitative PCR based assay for quantification of m <sup>3</sup> C (qPCR-m <sup>3</sup> C).....	92
Figure 3-12. Size exclusion HPLC and typical chromatography of LC-MS/MS for m <sup>3</sup> C.....	97
Figure 3-13. LC-MS/MS for quantitative m <sup>3</sup> C analysis in various RNA species .....	98
Figure 3-14 Validation of METTL2, 8 RNA binding species identified by PAR-CLIP.....	101
Figure 3-15 METTL2 could auto-methylate itself <i>in vitro</i> .....	103
Figure 3-16 METTL6 interact with seryl-tRNA synthetase (SARS) in a RNA dependent manner.....	107
Figure 3-17 METTL2, 6, 8 could affect global Translation .....	110
Figure 3-18 Loss of METTL2, 6, 8 confers higher sensitivity to translation inhibitors and alkylation damages .....	112
Figure 3-19 Codon-run to study effects of 3-methylcytosine on translation. ....	115
Figure 3-20 Interaction between tRNA Residue 32 and 38 extracted from solved structures.....	120
Figure 4-1 Methylation sites identified by mass spectrometry .....	127
Figure 4-2 Nitrogen methyl group on polar contacts .....	131
Figure 4-3 Enrichment of some codons and tRNA isoacceptors in Eukaryotics .....	133
Figure 4-4 Seryl-tRNA synthetase structures of <i>T. Thermophilus</i> and <i>homo sapiens</i> .....	137

Figure 4-5 Model for possible roles of METTL2 in tRNA modification and translation.....140

## LIST OF SYMBOLS

3mC or m <sup>3</sup> C	3-methylcytosine
aa	amino acid
aaRS	Amino Acyl-tRNA Synthetase
aa-tRNA	aminoacyl-tRNA
ABP140	Actin-binding protein, 140kDa
Anota	Analysis of translational activity
APS	Ammonium peroxodisulfate
Arg	Arginine
ATP	Adenosine 5'-triphosphate
bp	base pairs
BSA	bovine serum albumin
ChIP	chromatin immunoprecipitation
CHX	Cycloheximide
Cm	2'-O-methylcytidine
CRISPR	clustered regularly interspaced short palindromic repeats
DAPI	4', 6-diamidino-2-phenylindole
DMSO	dimethyl sulfoxide
DSB	double-strand break
dsDNA	double-stranded DNA
E.coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol tetraacetic acid
FBS	Fetal Bovine Serum
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
gDNA	genomic DNA
GFP	green fluorescent protein
Glu	glutamic acid
GST	glutathione-s-transferase



HA	Haemagglutinin
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid
hMETTL2A, 2B, 6, 8	Human Methyltransferase-like Protein 2A, 2B, 6 and 8
HRP	horseradish peroxidase
IB	immunoblotting
IgG	immunoglobulin g
IP	immunoprecipitation
IR	irradiation
kb	kilobases
KCl	potassium chloride
KD	Knockdown
kDa	kilodalton
KO	Knockout
LB	Luria-Bertani medium
m1A	1-methyladenosine
m2,2G	N2,N2-dimethylguanosine
M2/Mettl2	methyltransferase like protein 2
m7G	7-methylguanosine
Mass Spec	Mass Spectrometry
MEF	Mouse embryonic fibroblast
MMS	Methyl Methanesulfonate
mRNA	messenger RNA
MTase	Methyltransferase
Neo	Neomycin
Ni column	affinity column with nickel
NP-40	Tergitol-type NP-40
nt	nucleotide
OD	Optical density
PAGE	polyacrylamide gel electrophoresis
PAR-CLIP	Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation

PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
Puro	puromycin
qPCR	Quantitative Real-Time PCR
Riso-seq	Ribosome profiling
rRNA	ribosomal RNA
RSB	Resuspension Buffer
RT-PCR	Reverse transcription-PCR
SAM	S-adenosylmethionine
SARS	Seryl-tRNA synthetase
SDS	sodium dodecyl sulfate
Ser	Serine
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
TE	Tris-EDTA buffer
TEMED	Tetramethylethyl
Thr	Threonine
Tris	Tris(hydroxymethyl)aminomethane
TRM140	ABP140/tRNA Methyltransferase, 140kDa
tRNA	transfer RNA
UTR	untranslated region
WT	Wild type

## CHAPTER 1. INTRODUCTION

### 1.1 The importance of being modified: diversity and prevalence of RNA modifications

Naturally occurring DNA or RNA polymers were once thought to consist of only four canonical nucleosides A, T, C, G or A, U, C, G respectively. The presence of trace amount of non-canonical nucleoside in DNA was reported in 1948(1) and this nucleoside was identified as 5-methylcytosine in 1950 (2) and later became one of the focuses in epigenetic research together with 6-methyladenosine, histone modifications, etc. It was then found in 1951 that RNA also contains small but significant amount of non-canonical nucleosides (3) and it is determined to be 5-ribosyluracil, an isomer towards the canonical 1-ribosyluracil or uridine (4), also named the fifth base in RNA or pseudouridine ( $\Psi$ ) later. A great enthusiasm has been kept for discovering other thought minor or rare nucleosides in RNA or DNA polymers since then. For example, inosine was identified by Holley's lab at the first position of the anticodon of yeast tRNA Ala(5) and were quickly adapted by Crick's Wobble Hypothesis(6). Today the count of naturally occurring modified nucleosides is 114 according to Modomics (7). The number increases to 141 if residues that are doubly modified are included, like those containing a modified base and a methylation on the ribose, and there are approximately 50 modifications in mammals (7).

Those naturally occurring modifications are found in three major RNA species (tRNA, rRNA and mRNA) and some other minor RNA species, such as snRNA, miRNA and lncRNA, in Archaea, Bacteria and Eukarya. Although

different modifications show significant differences in abundance and site preference in various RNA species (7). These modified nucleosides are formed during maturation of nascent precursor RNA transcripts, or so called post-transcriptional modification or co-transcriptional modification, catalysed by modification enzymes in one step or several steps.

In general, the inherent ability of RNA is reinforced through the site selected installation of over 100 post-transcriptional modifications (7). However, their effects can be subtle and not easy to demonstrate either *in vivo* or *in vitro*. What's more, the investigation of modification enzymes was once limited by the presence of these enzymes in small cellular amounts, making them difficult to detect, let alone purify in sufficient amounts for enzymatic characterization before recombinant proteins are utilized for large scale purification. Additionally, the functional relevance of RNA modified was unclear because the genes coding for modification enzymes could be genetically deleted with only subtle or non-detectable effect on cell growth. Fortunately, interest in this research was sustained by the demonstration of a functional importance of several modified nucleotides in the anticodon loop of tRNA, namely the wobble nucleotide (position 34) and the nucleotide (position 37) that is next to the third anticodon nucleotide. Modifications at these sites correlate with clear-cut increases in the efficacy and accuracy of translation, a process that is termed translational fine-tuning. Moreover, RNA modifications have much more functions and will be discussed below. Nevertheless, remarkable progress has been achieved in understanding functions these modifications in RNA after years of investigation since the discovery of  $\Psi$  in RNA. In fact, modifications on RNA are one of the most

conserved properties. Here, RNA modification occurs in different RNA species will be briefly reviewed (notice one modification could occur in different RNA species), followed by characteristics of a few modified nucleosides and their corresponding enzymes, finalized by discussing one of the most simple, but important modification: methylation and their methyltransferases.

### **1.1.1 RNA modification in tRNA**

tRNA is the most extensively modified RNA species (8). As shown in Figure 1-1, the reason why tRNA is so heavily modified is still elusive. These modifications are hypothesized or proven to elicit various functions, including affecting translation rates, fidelity, tRNA stability, tRNA discrimination, etc. For example, It was shown the absence of inosine at wobble position 34 could generate coding errors, since inosine would extends anticodon-codon pairing capability through base pairing with U, A, and C, while A can only pairs with U. tRNA modification may increase tRNA half-life, which is relatively long, with estimates of 44 h in *Euglena gracilis*, 50 h in chicken muscle, and 3 d in avian liver (9-11), roughly comparable with the half-life of rRNA(10, 11). m1A58 on tRNA<sup>iMet</sup> is required for its stability(12), while Dnmt2 and Nsun2 are required for certain tRNA's stability(13). Longer half-life means more tRNA reuse and it would be regarded as a way of energy saving instead of de novo tRNA synthesis by transcription. In other words, modifications allow cells to respond to signals without the need for additional transcription. More evidence are showing modifications are temporally more responsive than the transcription apparatus, and result in conformational diversity, and population heterogeneity. The modified nucleosides are not recognized by native RNA

polymerases *in vivo*, offering a more flexible path for modification as a dynamic and regulatory mechanism. Moreover, initiator tRNA<sup>Met</sup> is distinguished from elongator tRNA<sup>Met</sup> through ribosylation at position A64 (12). The review below will focus on several tRNA modifications, and on how they offer advantage in its conformation and interaction with the translation apparatus.

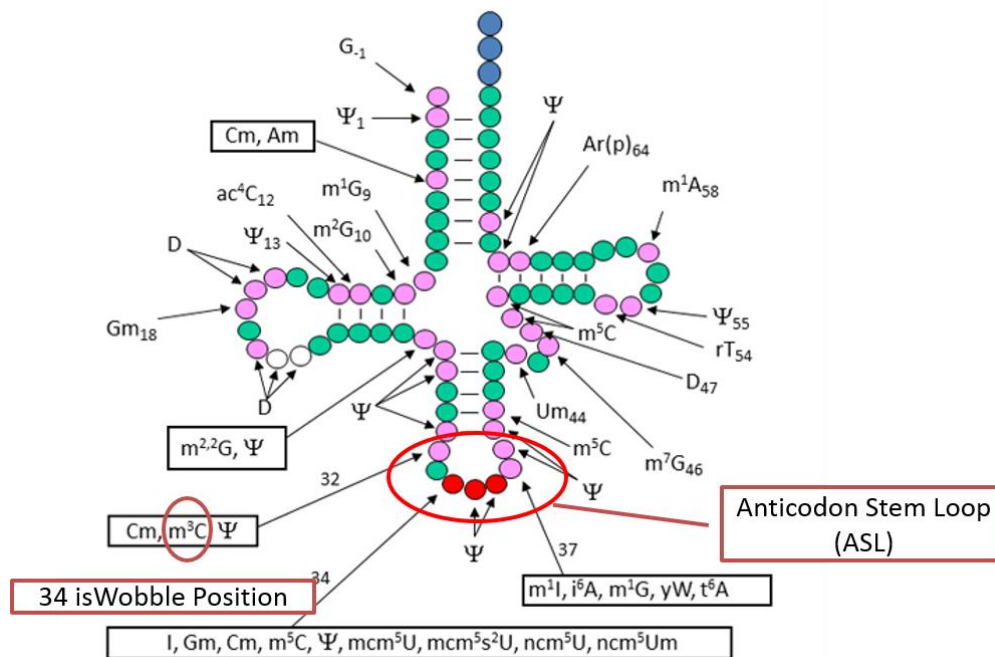


Figure 1-1 A schematic of tRNA numbering and modifications found in cytoplasmic tRNA in *S. cerevisiae*.

Adapted from figure 1 of (14). tRNA is shown in its usual secondary clover leaf structure form; Green circles are Residues that are unmodified in all yeast tRNA species; Pink circles are residues that are modified in some or all tRNA species; White circles are additional residues (20a and 20b) that are present in some, but not all, tRNAs and are sometimes modified; Red circles are anticodon residues, which are modified in some tRNAs; Blue circles are the CCA end. Conventional abbreviations are used; see the Modomics database

(<http://modomics.genesilico.pl>). For example: 2'-O-methylcytidine (Cm), N2,N2-dimethylguanosine (m<sup>2,2</sup>G), 3-methylcytidine (m<sup>3</sup>C). The pictured molecule starts at position -1 and is numbered consecutively from the next base (+1) to 76, with the insertion of two residues 20a and 20b. Several tRNA species, such as serine tRNA in cytoplasm have a longer variable arm starting after residue 44, and some tRNAs have different numbers of residues in the D-loop and the variable arm, but the anticodon is always numbered residues 34, 35, and 36, and the CCA end is always numbered residues 74, 75, and 76.

### **1.1.2 RNA modification in rRNA**

rRNA modifications are generally located in functional regions such as peptidyl transferase center (PTC), the A, P and E sites for tRNA and mRNA binding, the polypeptide exit tunnel, and sites of subunits interaction(15). These correlations suggest that nucleotide modifications are key to many rRNA functions. The long term hypothesis for the function of rRNA modifications is that they could modulate the ribosome's structures and functions and foster maturation. For instance, two most abundant modification in rRNA,  $\Psi$  and 2'-O-methyl ,could stabilize RNA base stacking (16) and prevents phosphate backbone hydrolysis by many ribonuclease (17), respectively. There are currently over 90 pseudouridines ( $\Psi$ ) and over 100 2' O-methylations on ribose and over 10 base methylations on human ribosomes (18). Although individual rRNA modifications is dispensable to some degree, they are indispensable when considered jointly. For instance, loss of certain rRNA modifications at peptidyl transfer centre will display decreased near-cognate tRNAs' discrimination capacity, reduced peptidyl transfer speed,

translation fidelity, and higher sensitivity to drugs targeting translation machinery. The exact abundance, and dynamics on ribosome rRNA modifications still require further investigation for possible usage in synthetic biology, such as synthetic ribosomes.

### **1.1.3 RNA modification in mRNA**

Although less abundant, there are several modifications found on messenger RNAs. In mammals, m<sup>7</sup>G is added towards the 5' end of mRNA, which is a critical step from precursor messenger RNA to mature messenger RNA. This cap is important in splicing, mRNA nuclear export, and it is crucial during translation initiation (19). Meanwhile, m<sup>6</sup>A, Ψ, m<sup>5</sup>C, I and 2'O methylation of ribose were found to be localized at the internal body of mRNA. Recently, m<sup>1</sup>A have been found as a new methyl mark on mRNA, mainly cluster at starting codons and hypothesized to affect translation (20, 21) .

mRNA modifications located in coding sequence need to be non-mutagenic, otherwise they will not preserve the genetic information from DNA. It is maybe the reason why most nucleoside modifications on mRNA are methylations (7), which minimize the disturbance towards the mRNA.

The N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) is the most frequent internal modification in eukaryotic mRNA. It was first discovered together with 5' G cap in 1970s (22, 23). It has been shown that m<sup>6</sup>A mainly localize within a G (m<sup>6</sup>A) C (Approx. 70%) or A (m<sup>6</sup>A) C (approx. 30%) consensus sequence. A large protein complex is found to install s this methylation and it turns out to contain Mettl3, Mettl14 and a splicing regulator, WTAP, (24, 25), However, due to large molecular weight of the complex (Approx.1000kDa), other proteins/co-factors which are partners to the enzymatic core remain to be identified. For



biological functions, m6A is critical for yeast meiosis(26), required for Notch signalling during oogenesis in fruit fly (27), plant development (28), and to be essential for self-renewal of mammalian embryonic stem cells (29).

Although C5-Methylcytosine is the mostly known for its role in eukaryotic genomic DNA, it is also of relatively high abundance in rRNAs and tRNAs. More than 10,000 possible m5C sites in mRNA and noncoding RNAs is revealed by a transcriptome-wide mapping (30). UTR and regions around Argonaute binding sites are enriched from m5C, and several sites were confirmed in using either anti-m5C immunoprecipitation coupled with next generation sequencing or RNA bisulfite sequencing. Some sites on mRNA are found to be bound by several m5C MTase targeting rRNA or tRNA, suggesting potential undiscovered mRNA substrates. There is a possible role of m5C metabolic processes since several genes harbouring m5C are found to be involved in those pathways (31). But the role of m5C in RNA largely remains to be elucidated.

#### **1.1.4 RNA modification in other RNA species**

Long non-coding RNA (lncRNA) are found to be involved in many cellular processes, such as chromosomal remodelling, transcription, X-inactivation and imprinting(32). It has been shown that some lncRNAs, such as NEAT1, MALAT1, contain multiple m6A sites (33-35) with poorly known functions. Another non-coding RNA species, Small nuclear RNAs ( snRNAs ), the RNA parts of the spliceosome, also harbours a couple of internal modifications such as Ψ, 2' O-methylation and m6A without knowing specific functions in snRNA besides the known general functions.

### 1.1.5 Some other common modifications

**Pseudouridine ( $\Psi$ )** : Pseudouridine is also named the 5<sup>th</sup> base, was the first found RNA modification identified in the 1950s (36). Isomerization allows the potential formation of an additional hydrogen bond which can affect the RNA 3D structure of RNA drastically. The presence of  $\Psi$  1 on mRNA could fine-tune translation by employing  $\Psi$  reader proteins or altering the secondary structures of mRNA. Changing the first uridine residue on the stop codon to  $\Psi$  can turn nonsense codon to sense codon (37). This finding expanded the genetic code, or in other words, it showed RNA modifications on mRNA could make transcripts recoded (37). Recently, hundreds of  $\Psi$  sites has been mapped in a transcriptome wide manner on yeast and human mRNA. These  $\Psi$  sites dynamic upon nutrition starvation or heat-shock (38, 39), suggesting introducing  $\Psi$  modification as a possible mechanism to adapt the translational speedily to various environmental stress.

**Inosine (I)** : Inosine was identified by Holley's group at the 1<sup>st</sup> position at anticodon of yeast tRNA<sup>Ala</sup> in 1965 and it is the second example of modified nucleotide in RNA after pseudouridine (40). It was quickly adapted by Crick's Wobble Hypothesis in 1966 (6). The A to I deamination at Wobble position at tRNAs' anticodon enables tRNA to read three synonymous codons: NNA, NNU and NNC (N for any nucleotides). Wobble at the third position of the anticodon together with other chemistry and structures caused by other modifications enable some 40 cytoplasmic tRNAs to translate the 61 Universal Genetic Codes and some 22 mitochondrial tRNAs to decode mitochondrial codes.

The non-canonical A - I base pair raised the question of how a purine- purine interaction is accommodated at the decoding site on the ribosome and it was answered with ribosome co-crystallized with tRNA. It was showed that the decoding site accommodates the distorted anticodon-codon double helix of A - I and the extensively modified anticodon -codon base pairs (41). Moreover, Paul Agris's lab demonstrated that 2-thiocytidine occurring in the anticodon loop of the isoaccepting tRNA<sup>Arg1</sup> species negates inosine's ability to read adenosine while allowing the recognition of U and C. Thus, sometimes multiple modifications that occur close in space work together to affect RNA dynamics conformations and function.

### **2'-O-Methylation**

2'-O-methylation occurs on the ribose and will protect RNA from cleavages by certain kinds of endonuclease. It has been synthesized into small interference RNA (siRNA) to improve the stability and avoid inflammatory signals during knocking down experiments (42) since 2'-O-methylation has a role in discrimination between self and non-self mRNA. Abolishment of 2'-O-methylation in human or mice coronavirus will trigger type I interferon response via the recognition by a RIG-I-like receptor, MDA5, which could binds dsRNA. 2' O-methylation catalysed by a MTase HEN1 is naturally present in Plant microRNAs. The function of such a modification is suggested to protect the 3' end of miRNA against poly-uridylation mediated degradation.

#### **1.1.6 Modifications resembles the side chains of amino acids**

RNA's modified nucleosides are more than simple modulators of structure and function. The genetic information, material consumption, and energy used in producing modified nucleosides by all organisms has evolved into a conserved

and ubiquitous investment. The chemistries of RNA modifications and their effects on structure seem similar to amino acid's side chains, either hydrophilic, polar and charged, and hydrophobic, aliphatic and aromatic. 10 physicochemical properties of modification chemistry, including spatial effects were attributed to their functions, as initiated by Paul Agris as early as 1996 (43).

Modification directs the RNA into the conformation most needed for the task at hand, whether that be tRNA, rRNA, snRNAs, mRNAs or miRNAs or lncRNAs. The sequences of the ribosome's peptidyl transferase site differ among organisms, but modifications are found conserved within the functional structure. The positions of pseudouridines and 2'-O-methylations in many types of RNA, rRNAs, snRNAs, from different organisms are conserved. These most ubiquitous of modifications are often found to facilitate RNA folding, structure, and conformational dynamics.

The modified wobble hypothesis presented in 1991 invoked the chemical and structural contributions of modified nucleosides to account for tRNA being either constrained in its codon recognition, or relaxed in its abilities to read multiple codons. Modified nucleoside prestructure RNA for function. The prestructuring lowers energy barriers to RNA/RNA and RNA/protein interactions. The fact that many modification enzymes are at the ready when RNA is transcribed makes one wonder whether the modification of any RNA can be an immediate temporal response to internal and external signals in a time frame shorter than that of gene expression. While RNA is inherently transient, we now know that some modifications are even more so with the discovery of demethylations.

### **1.1.7 Modifications represent an additional, operational code, a third code**

The Universal Genetic Code provided insight into the decoding of genes, while Nucleosides proximal to tRNA's 3'- aminoacyl-terminus, including the discriminator nucleotide (position 73), are recognition determinants for aminoacylation and as such are a second code discovered by Paul Schimmel. RNA modifications as determiners of functional structure, indicators of cellular function and health and a distinctive set of biomarkers for health and disease, are a third code. The field are gaining insights into this code with improvements in analytical sensitivity and high throughput methods that relate modified nucleoside chemistry and structure to cellular function and human health issues.

## **1.2 Detection and quantification of RNA modifications**

The difficulty in detecting the usual low abundance of RNA modifications is still a bottleneck in the field. Currently, enzymatic detection such as Reverse transcription-based methods, detection based on differences in physicochemical properties or chemical reactivity of different modifications are three major principles for the detection and quantification of RNA modifications(44), and will be discussed below.

### **1.2.1 Detection of RNA modifications by different enzymatic behaviors**

Arrest of primer extension during reverse transcription

It may result in different enzymatic reaction efficiency when cellular enzymes encounters a nucleosides with modification. Such different behaviours could be exploited for detection. One notable example is reverse transcriptase, which

may stall when meet certain modification in the RNA template. In general, modifications on the Watson-Crick edge of the nucleosides such as m<sup>3</sup>C, which is the focus of the thesis, and m1A, m3U, m22G, m1G are likely to stop the incorporation of a complementary nucleotide. 2'-OMe modification, which occurs at the ribose, could also present a hard stop for reverse transcriptase. Chemically small modifications, such as Ψ, m5U, m5C, m2G usually didn't stall the reverse transcription if located on the Hoogsteen edge. For bulky modifications at the Hoogsteen edge, the experimental conditions, such as the concentration of dNTPs or reaction temperatures affects the read-through (45). Generation a hard stop which could block/stall reverse transcription by certain chemicals, followed by comparing WT or KO samples by second generation sequencing present the most cutting edge methods for certain modification detection. For example, Ψ and m5C has been mapped at transcriptome level by treating with CMC, and bisulfides, respectively and followed by second generation sequencing (38, 39, 46).

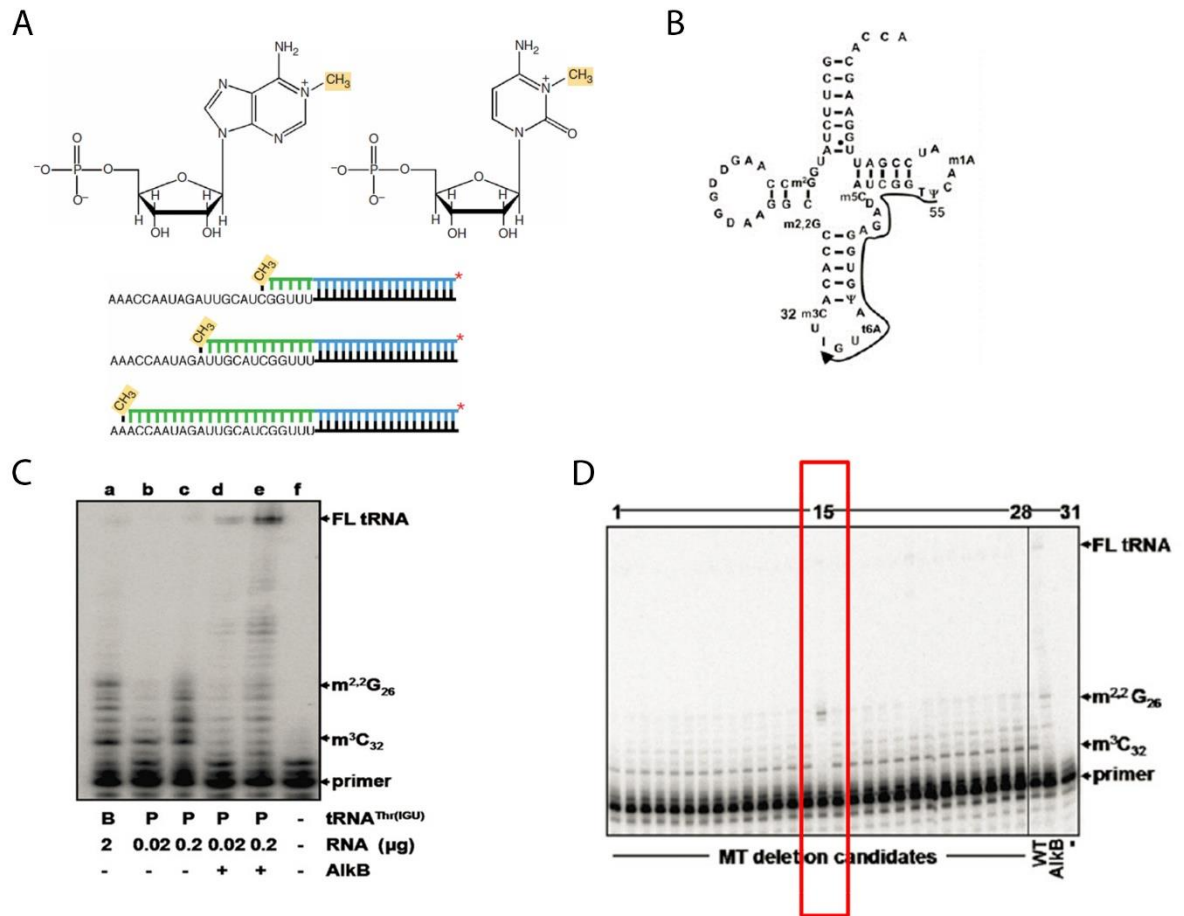


Figure 1-2 Principles of Primer extension assay and examples

Adapted from (47)

(A) Principles of primer extension assay, chemical structure of 1-methyladenosine and 3-methylcytidine is shown at the top panel; Reverse transcription would be blocked when reverse transcriptase meet certain modified nucleotide;

(B) Diagram of tRNA<sup>Thr</sup> IGU showing the localization of m<sup>3</sup>C

(C) tRNA<sup>Thr</sup> IGU was purified from WT cells and incubated with demethylase AlkB from *E. Coli*, and assayed by primer extension, showing the abolishment of extension block

(D) Primer extension block is abolished using tRNA from ABP140 KO strain

Ligation-based approach (restricted to Ψ, m<sup>6</sup>A and 2'-OMe)

Recently, an approach based on splint ligation has been developed(48).

However, this methods are restricted to Ψ, m<sup>6</sup> A and 2'-OMe so far.

Restriction enzyme-based sequencing (restricted to m6A and m5C)

Restriction enzyme-based sequencing, e.g. 6mA-REseq, relies on the identification of restriction enzymes with respective recognition sites and the sensitivity to the corresponding DNA modification (49). For 6mA-REseq, genomic DNA is digested with a 6mA-sensitive enzyme and randomly sheared, which results in an enrichment of unmethylated sequence motifs at the ends of the sequencing reads. In contrast, 6-methylated adenine prevents digestion and is enriched in inner body of the reads. Consequently, 6mA levels are readily inferred from the relative positions of the restriction enzyme consensus sequence (49, 50).

Single-Molecule Real-Time Sequencing (SMRT)

Genome wide 6-methyladenine locations at base-pair resolution were obtained in *E. coli* genomes using single-molecule real-time (SMRT) sequencing (51, 52). SMRT, a third-generation sequencing technique, is based on the processing of fluorescence-labeled nucleotides by DNA polymerases. The fluorescence label is not incorporated in the de novo synthesized strand but is cleaved away during the process. Meanwhile, the label emits light that is captured in the nanophotonic visualization chamber. High-fidelity polymerases are capable of synthesizing long continuous strands at a high speed, allowing a fast sequencing process and high read lengths. Importantly, the incorporation of a modified nucleotide, such as 6mA, presents different kinetics compared with unmodified adenine, allowing the direct inference of the modification status of each base.



### **1.2.2 Detection of RNA modifications by different physicochemical properties**

Most of the modifications are just minor decorations on the RNA, compared with the major sugar and base of RNA. A modified RNA cannot be easily distinguished from its unmodified form since a small modification such as adding a methyl group elicit minor changes to the RNA's properties.

Most of the HPLC/LC/MS/MS would digest the RNA polymers by nuclease and phosphatase. Modifications are identified by their retention characteristics in chromatography and capillary electrophoresis, or by their characteristic mass-to-charge ratio ( $m/z$ ), respectively.

Thin layer chromatography (TLC) is once the predominant way for the detection and identification of modifications. Although the  $R_f$  values of over 100 modified nucleotides or dinucleotides have been mapped in standard solvent systems, the technique is not commonly used after the advent of mass spec.

Liquid chromatography (LC) coupled with UV is still common in nucleoside detection. A comprehensive work on adapting different LC-UV methods to separate 65 nucleosides in a single run and identification via retention time and UV spectrum is given by(53).

Ultra-high performance liquid chromatography-triple quadrupole mass spectrometry coupled with multiple-reaction monitoring (UHPLC-MRM-MS/MS) allows the sensitive detection of nucleotide modifications, such as 5mC, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) at very low abundance (54). It has been modified to detect low abundance m6A in *C.elegants*(55), *Drosphila*(56) and *Chlamydomonas*(49) recently. Briefly, the

digested DNA is separated by reverse-phase ultra-high performance liquid chromatography (UHPLC) coupled with mass spectrometry detection using tandem mass spectrometers (MS/MS). Following detection of specific nucleotide modifications, quantification is achieved using a standard curve that is simultaneously analysed in the sample of interest. It is important to discard any potential contamination from Mycoplasma or bacterial DNA.

### **1.2.3 Detection of RNA modifications by different chemical reactivity**

Knowledge on different chemical reactivity towards a variety of chemical between canonical or modified nucleotides is accumulated through years by biochemists and is a chest of treasure for detecting modifications. An important discovery from the 1970s along with the development of Maxam & Gilbert DNA/RNA sequencing method is the different sensitivity of certain modified nucleotides towards aniline induced chain scission via  $\beta$ -elimination(57). For example, m7G requires pre-treatment with NaBH<sub>4</sub>, while 3-methylcytosine (m<sup>3</sup>C) should be treated with hydrazine to allow chain scission. The latter reactions are actually one of the basis of direct DNA/RNA sequencing by the Maxam & Gilbert. As discussed earlier, the low abundant occurrence of modifications makes the application of physicochemical techniques such as HPLC/MS challenging. Chemical treatments towards different modifications offer advantages include tagging modifications for improved detection by other means, such as biotin, or adapting high throughput methods for the detection of modifications on a genome wide scale with reverse transcription based RNA seq.

Reagents which could selectively react with modified nucleotides only in a narrow window of well-defined experimental conditions have been identified

(58-60). Taking pseudouridines ( $\Psi$ ) as an example, there are at least 4 different chemicals that have been utilized to treat and detect  $\Psi$  (61). CMCT acylates G, U, inosine and  $\Psi$  residues under physiological conditions and alkaline treatment could leave only N3-acylated pseudouridines intact. The latter could arrest reverse transcription and can thus be detected by primer extension techniques similar and has been recently adapted towards identifying pseudouridines at whole transcriptome level (38, 39). Many other modifications could be chemically treated by different reagents and proceed to RT based RNA-seq profiling.

#### **1.2.4 Antibody based detection followed by second generation sequencing**

Producing antibodies targeting modified nucleosides have been performed as early as 1970s for inosine (62). Antibodies against m5C m6A for either DNA or RNA has been generated to be able to perform immunoprecipitation coupled with next-generation sequencing to identify of those modifications at genome wide(33, 34, 63). For example, 6mA-immunoprecipitation sequencing (6mA-IPseq) utilizes a specific antibody for 6-methyladenine to enrich modified fragments from the sequencing library. Following the alignment of sequencing reads to the reference genome, 6mA-modified regions present enriched mapping frequencies. Subsequent sequence enrichment analysis can reveal consensus recognition motifs for the adenine methyltransferases.

However, the immunoprecipitation based identification of modification genome wide or in whole transcriptome has only been developed for a limit number of modifications in general. A number of attempts on generating antibodies against modified ribonucleoside such as  $\Psi$ , m5C, m1A, m2,2G,

m1G, m7G, m1I, i6A, and ac4C seems to be ceased by the late 1990s(64-71), probably due to the low abundance of certain modifications and the improvements in Mass Spec sensitivity. Antibody against m<sup>3</sup>C in DNA has also been raised to use only in dot blot(72).

### **1.3 Identification of enzymes responsible for specific RNA modifications**

RNA modifications are carried out by specific enzymes. The identification and purification of RNA modifying enzymes for medium to low occurrence modifications were challenging tasks. The main obstacles includes the difficulty in obtaining pure individual RNA species as substrate, the presence of endogenous RNases, the typically low abundance of the enzymes involved in RNA modification, etc. Determining the presence or absence of the specific modified nucleotides, which is not an easy task before the advent of UHPLC and capable Mass Spectrometry as discussed in section 1.2. Nevertheless, a couple of enzymes are identified and characterized by using classic biochemical and genetic approaches such as using standard chromatographic techniques(73).

Fortunately, recent developments of the recombinant DNA and RNA technologies, together with techniques for chemical synthesis of DNA and RNA and detection of modified nucleosides in nucleic acids using Mass spectrometry(74) offers possibilities to obtain synthetic or semi-synthetic substrates. Subsequently, it is then possible to analyze the modified nucleoside content after incubation with cell extracts. Moreover “classical” biochemical and genetic methods for identification of new RNA MTases can be efficiently

supplemented by “reverse genetics” and large-scale “-omics” approaches, including biochemical genomics.

Comparative genomics has been shown useful to identify the responsible genes, as is the case for 5,6-dihydrouridine synthase which is one of the most common and abundant modifications on tRNAs (75), and is also present in some 23S RNA(76), but both the genes and enzymes in many organisms were unknown. By combining occurrence, chromosome localizations, and homologue searches, the Dus family of genes that contains orthologues in most sequenced species was identified(75). The last decade has witnessed great progress in the identification of the genes and corresponding enzymes that are required for modification, as well as in our understanding of their biochemistry and their biological roles. Thus, in *S. cerevisiae*, the vast majority of the genes that are responsible for modifications have now been identified, and a large number of modification genes and enzymes have been identified in other organisms (see the Modomics database, <http://modomics.genesilico.pl>).

#### **1.4 Methylation and methyltransferase (MTase)**

Methyltransferase is a subclass of transferase class enzymes that catalyze the transfer of a methyl group from one compound to another. There are currently over 300 types of Transmethylation reactions that are designated under Enzyme Commission 2.2.1: Methyltransferases. S-Adenosylmethionine (SAM or S-AdoMet) was used by the majority of methyltransferase as methyl group donors or co-factors in a transmethylation reaction, in which SAM donates its methyl-group towards a variety of substrates including macromolecules such as nucleic acids, proteins, lipids and secondary small metabolites such as

glycine, nicotinamide, indolethylamine, phenylethanolamine, etc, and the atomic targets could be C, N, O, S, halide ions, or As-cantered nucleophiles.

SAM was first discovered in 1952 as a conjugate of methionine and the adenosine moiety of ATP(77, 78). It is the second most widely used enzyme substrates following ATP, involving in a variety of metabolic pathways such as transmethylation, transsulfuration, and aminopropylation(79). Most transmethylation reactions catalyzed by SAM-dependent MTases are bimolecular nucleophilic substitution ( $SN_2$ ) reactions, in which one bond is broken and one bond is formed synchronously, generating S-adenosyl-homocysteine (SAH or AdoHcy) as the by-product. SAM-dependent MTase are inhibited by the product SAH, which can accumulate in the presence of homocysteine because the equilibrium of the SAH hydrolase reaction favors SAH formation (80).

#### **1.4.1 Structural features of MTase**

The structural and catalytic mechanism for methyltransfer reactions could be remarkably diverse since they could share little amino acid sequence similarity but adopted highly conserved structural folds (81). There are currently five designated families of SAM-dependent MTases ( from Classes I to V) based on their distinct structural features with Class I being the great majority(82). Another class of radical SAM-dependent MTases that methylate a range of substrates at unreactive, non-nucleophilic carbon and phosphorus centers was recently discovered (83).

Class I MTases with the Rossmann fold are the largest group of MTases, responsible for the majority of transmethylation. Structure of Class I MTases are strikingly similar, usually comprising a 7 stranded  $\beta$  sheets with a

structural turning point and a signature reversed  $\beta$  hairpin at the C terminal (represented as 6  $\uparrow$  7  $\downarrow$  5  $\uparrow$  4  $\uparrow$  1  $\uparrow$  2  $\uparrow$  3  $\uparrow$ ) The first  $\beta$  sheet usually ends up with GxGxG motif which is a signature for nucleotide binding. The only other strongly conserved position is an acidic residue at the end of  $\beta$ 2 that forms hydrogen bonds to both hydroxyls located on the sugar part of SAM(84).

Besides small molecule's methylation, the class I MTases also includes DNA 5-methylcytosine methyltransferases (Dnmts) (85). DNA methylation is essentially involved in a variety of cellular processes, including regulating gene expression, mutation repair, genetic imprinting, and maintenance of genome integrity, development and diseases etc.(86). The diversity of the Class I structure is further demonstrated by existence of a large family of RNA 5-methylcytosine MTases within this class (87), as well as RNA 6-methyladenosine MTase, etc. Although the largest group of protein methyltransferases (PMTs), which are the SET-domain protein methyltransferases, belong to the class V MTases, some of the PMTs belong to class I MTases. All known SET-domain MTases working on lysine residues on various nuclear proteins are involved in chromatin function and transcription regulation, including Rubisco and cytochrome C (88).

#### **1.4.2 A diverse set of structure features for each MTase class**

Almost all MTases catalyse the methyl transfer in an  $SN_2$ -like mechanism. However, distinct mechanisms have been employed to transfer the methyl-group to different atom in various substrates.

Methyl transfer to nitrogen atoms: [D/N/S]PP[Y/F] motif, or DPPY, is a common motif for substrate binding in N atom methylation and is present in DNA 6' N adenine and DNA 4' N cytosine MTases, and protein 5' N

glutamine MTase (89). The diversity of substrates indicates that the DPPY motif is not nucleotide specific, but is selective for nitrogen's conjugation to a planar system such as an amide moiety or a nucleotide base(90), (91). DPPY motif is absent in some small molecule's MTase, such as glycine-N MTase (GNMT)(92) and phenylethanolamine-N MTase (PNMT)(93) and both of them use acidic residue(s) to neutralize the positive charge on the substrate amino group (94), so is the case for METTL2, 6, 8 proteins, since they lack the DPPY motif, as shown in

Methyl transfer to carbon atoms: A significant number of crystal structures are solved for DNA C5-cytosine MTases with substrate bound, which transfer the methyl group on to carbon (95, 96). A Pro-Cys dipeptide is highly conserved within the active site of C5 cytosine MTases, and structurally resembles the PY part in DPPY motif. In M.HhaI, the 4' nitrogen of the cytosine is poised far away from SAM, in order to present the C5 atom as a more accessible methylation target. Methyl transfer towards C atoms is more difficult than to polarizable N atoms, the cytosine must first be activated by given a negative charge by a covalent bond formation, which will help the methyl transfer(97).

Methyl transfer to oxygen atoms: Several O-MTases' crystal structures have been solved, such as catechol O-MTase (COMT) (98), the Chalcone O-methyltransferase (ChOMT) and isoflavone O-methyltransferase (IOMT) (99) and the glutamate O-MTase CheR (100). An  $Mg^{2+}$  ion is required to orient the substrate-binding site, and a nearby lysine residue appears to deprotonate the substrate hydroxyl before targeting the methyl group on SAM. Neither ChOMT nor IOMT requires a metal ion, but they do need a histidine residue to deprotonate the hydroxyls in plant metabolites.



### 1.4.3 Methyltransferase like (METTL) proteins and their substrates

There are more than 30 genes assigned as methyltransferase like protein (Mettl) genes by the HUGO Gene Nomenclature Committee (HGNC), from Mettl1, Mettl2A&B, Mettl3 to Mettl25 etc. mainly due to the reason that they harbor a putative methyltransferase domain and have not been well characterized. Biochemistry and Biological functions of some of the Mettl proteins were revealed, as summarized in Table. METTL1 is identified as a 7-methylguanosine methyltransferase responsible for the formation of N(7)-methylguanine at position 46 (m7G46) in tRNA (101), METTL3 is identified as mRNA N6-adenosine Methyltransferase (102), and is recently found to form a stable heterodimer core complex with METTL14, which is also found to be able to catalyzes m6A mRNA methylation(103). METTL4's homolog in C.elegants is recently found to be the methyltransferase catalyzing m6A methylation in genomic DNA and raised the possibility that 6mA may be a carrier of heritable epigenetic information in eukaryotes (In press and personal Communication, Prof. Shi Yang, Harvard Medical School). Human METTL20 methylate electron transfer flavoprotein  $\beta$  (ETF $\beta$ ), thereby inhibiting its ability to mediate electron transfer from acyl-CoA dehydrogenases. METTL10 was found methylating lysine residues of eukaryotic elongation factor 1A in *Saccharomyces cerevisiae* (104), Mettl11A was found to be a N-terminal Xaa-Pro-Lys N-methyltransferase (105). As we could see, METTL1 to METTL25 was found to catalysing Methyl-Group transfer towards a diverse range of substrates including DNA, RNA, and Proteins. However, the majority of Mettl proteins in mammalian still remain poorly studied including Mettl2, 6, 8.

Table 1-1 Examples for methyltransferase-like proteins

	Aliases	Function
Mettl1	tRNA (Guanine(46))-N(7))-Methyltransferase	Catalyzes the formation of N(7)-methylguanine at position 46 (m7G46) in tRNA
Mettl3	mRNA M(6)A Methyltransferase subunit	This enzyme is involved in the posttranscriptional methylation of internal adenosine residues in eukaryotic mRNAs, forming N6-methyladenosine
Mettl4	m6A Mtase in DNA	N6A Mtase in Worm & Fly's genomic DNA
Mettl9	DORA Reverse Strand Protein	Probable methyltransferase
Mettl11	N-terminal Xaa-Pro-Lys N-methyltransferase 1	N-terminal Xaa-Pro-Lys N-methyltransferase 1
Mettl13	Antiapoptotic Protein FEAT	Probable methyltransferase
Mettl14	m6A Mtase in mRNA	N6-methyladenosine in mRNA
Mettl17	Protein RSM22 Homolog, Mitochondrial	May be a component of the mitochondrial small ribosomal subunit
Mettl18	Probable histidine methyltransferase	Probable histidine methyltransferase
Mettl19	tRNA Methyltransferase	Probable tRNA (Uracil-O(2))-

	44 Homolog (S. Cerevisiae)	Methyltransferase
Mettl21A-E	Probable histidine methyltransferase	Protein histidine methyltransferase

### 1.5 3-methylcytidine ( $m^3C$ ) and its corresponding methyltransferase

3-methylcytosine occurs at position 32 of certain tRNA species

3-methylcytidine ( $m^3C$ ) is one kind of methylation modification formed by adding a methyl group onto the 3' position in cytosine base attached to either a deoxyribose or ribose. 3-methylcytosine in DNA is an alkylation lesion. The N1 position on adenosine and N3-position on cytidine are nucleophilic centres most reactive to alkylating chemicals. The highly toxic m1A and both mutagenic and toxic  $m^3C$  are major products in single-stranded DNA or RNA only, since base pairing protects these positions from methylation in dsDNA.

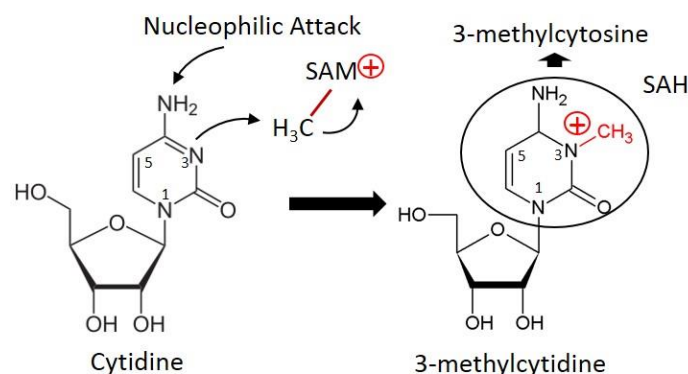


Figure 1-3 Chemical structure of cytidine, 3-methylcytidine ( $m^3C$ ), and proposed mechanism for  $m^3C$  formation

Cytosine is lined to ribose to form cytidine, in which the numbering of pyrimidine ring is indicated by numbers, notice the methyl group from S-adenosyl-methionine (SAM) is transferred to the position 3 nitrogen atom on

the pyrimidine ring by nucleophilic attack. The exact enzyme active site remain elusive.

3-methylcytidine was first discovered in 1963 from yeast RNA(106) and found in both eukaryotic rRNAs and tRNAs (107, 108). As shown in the Figure 1-4 , which is obtained from database(109), along with published papers, m<sup>3</sup>C is localized at position 32 of a number of threonine, serine and arginine's corresponding tRNA species from both cytoplasm and mitochondria (110) (111) . Interestingly, m<sup>3</sup>C is also found at e2 on the long variable loop of several cytoplasmic tRNA<sup>Ser</sup> species. In Bakers's yeast, m<sup>3</sup>C is located at position 32 of tRNA<sup>Thr1</sup> and tRNA<sup>Ser1</sup>(110, 112), which is at the junction of anticodon stem and anticodon loop, one residue away from anticodon (position 34-36). Residue cytidine at position 32, together with its modification, such as m<sup>3</sup>C, is proposed to regulate accurate codon recognition at the A-site, proper ribosomal conformation at the P-site, discrimination of synonymous codons, etc. Importantly, in the anticodon loop of tRNA, residue at 32 usually forms a non-canonical base pair with the residue at 38 (113). It has been proposed that the 32-38 base pair is important for tRNA-ribosomal interaction, which will be extensively reviewed in 1.7 and 3.19. It is proposed that if a methyl group is added on the N atom at position 3, one of the three hydrogen bonds between C and G or C and A could be impaired , but the biological significance of this change is a mystery. In short, the interactions among 32, 38 residues, ribosomal proteins and RNA could be vital towards tRNA's interaction with mRNA and ribosome, and reading frame maintenance and will be further discussed below 1.7 and 3.19.

Organism	Amino Acid	Acc-stem	D-stem	D-loop	D-stem	Ac-stem	Ac-loop	Ac-stem	V-region	T-stem	T-loop	T-stem	Acc-stem	CCA
		-1 1	8 10	14	22	26 27	32	39	44	49	53	61	66	73 74
<i>Bos murus, Liver</i>	-Arg	GGCUCCG	UK GCGC	AAD--GGAD-A	GCGC	A PPGGA	U3CP6A	UPCAA	AG-----7DU	TCGGG	TPCG <sup>m</sup> GU	CCCAG	CGGAGUC	G CCA
<i>Bos murus, Liver</i>	-Arg	GCCCCAG	UK LCCU	AAD--GGAD-A	AGGC	A PUGGC	UCU6A	GCCAG	GG-----ADU	GPGGG	TPCG <sup>m</sup> GU	CCCAG	CUGGGG?	C CCA
<i>Bos murus, Liver</i>	-Thr	GGCCUUG	UK LCUA	GCD--GGDC-A	AAGC	R CCUGU	UIGU6A	ACAGG	AG-----AD?	CUGGG	UPCG <sup>m</sup> AU	CCCAG	CGGGG?	U CCA
<i>Candida cylindracea</i>	-Ser	GGCAACU	UG GCMG	AGD--#GDD-A	AGGC	R UGACA	UCGA+A	PGUCA	3-GGGU---UU-----ACCCG-#	GCAGG	TPCG <sup>m</sup> AU	CCUGC	AGUUGUC	G CCA
<i>Candida cylindracea</i>	-Ser	GUJAAUG	UG GCGH	GAGD--GGDD-A	AGGC	R CUUCC	UCGU+A	GGAA	3-GGGC---UCU-----GCCCG-#	GCAGG	TPCG <sup>m</sup> AU	CCUUG	CAUUAAC	G CCA
<i>Drosophila melanogaster</i>	-Ser	GCA#UCG	UG GCMG	AGD--#GDD-A	AGGC	R UCUGA	UCGA+A	PCAGA	3-UCCC---UU-----GGGAG-?	GUAGG	TPCG <sup>m</sup> AU	CCUAC	CGGUCUG	G CCA
<i>Drosophila melanogaster</i>	-Ser	GACGAGG	UG GCMG	AGD--#GDD-A	AGGC	R PPGGA	UCGUEA	PCCAA	3-GUGC---UU-----GCACG-?	GUGGG	TPCG <sup>m</sup> AU	CCCAG	CCUUCG	G CCA
<i>Drosophila melanogaster</i>	-Ser	GCA#UCG	UG GCMG	AGC--#GDD-A	AGGC	R UCUGA	UIGA+A	PCAGA	3-UCCC---UU-----GGGAG-?	GUAGG	TPCG <sup>m</sup> AU	CCUAC	CGACUGC	G CCA
<i>Gallus gallus</i>	-Ser	AAGAAAG	A" LCAU	UA#GUGDUUG	AUGC	L PPUUG	UUGAHA	CCAAC	AU-----G-U	GAGGG	TPCG <sup>m</sup> UU	CCUUC	CUUUUUU	G CCA
<i>Lupinus luteus</i>	-Ser	GUCAGUA	UG GCMG	AGD--#GDD-A	AGGC	R ACAGA	UCGA`A	PCUGU	3-GAGAU---UU-----GCCUG-?	AGCGG	TPCG <sup>m</sup> AU	CCGCU	UACUGAC	G CCA
<i>Nicotiana rustica</i>	-Ser	GUCGUAU	UG UCCG	AGD--#GDD-A	AGGA	R ACAGA	UBGA`A	PCUGU	3-GGGC---UUC-----GCCCG-C	GCAGG	TPCG <sup>m</sup> AC	CCUGC	UGUCGAC	G CCA
<i>Nicotiana rustica</i>	-Ser	GUCGUAU	UG UCCG	AGD--#GDD-A	AGGA	R ACAGA	UCGA`A	PCUGU	3-GGGC---UUU-----GCCUG-C	GCAGG	TPCG <sup>m</sup> AU	CCUAC	UGUCGAC	G CCA
<i>Nicotiana rustica</i>	-Ser	GUGCCCU	UG GCCG	AGD--#GDD-A	AGGC	R PGUCG	UCGUEA	GPACA	3-GGGG---UUU-----CCCCG-C	GAGAG	TPCG <sup>m</sup> AU	CUCUC	AGGCGAC	G CCA
<i>Nicotiana rustica</i>	-Ser	GUGGACG	UG CCGG	AGD--#GDD-A	UCGG	R CAUGA	UIGA`A	PCAU	3-GGGC---UUU-----GCCCG-C	GCAGG	TPCG <sup>m</sup> AU	CCUGC	CGUCAC	G CCA
<i>Nicotiana rustica</i>	-Ser	GUGGACG	UG CCGG	AGD--#GDD-A	UCGG	R CAUGA	UIGA`A	PCAU	3-GGGC---UCU-----GCCCG-?	GCAGG	TPCG <sup>m</sup> AU	CCUGC	CGUCAC	G CCA
<i>Rattus norvegicus, Liver</i>	-Ser	GACGAGG	UG GCMG	AGD--#GDD-A	AGGC	R APGGA	UCGUEA	PCCAU	3-GPGC---UUU-----GCACG-?	GUGGG	TPCG <sup>m</sup> AU	CCCAG	CCUCGUC	G CCA
<i>Rattus norvegicus, Liver</i>	-Ser	GUAGUCG	UG GCMG	AGD--#GDD-A	AGGC	R APGGA	UIGA+A	ZCCAU	3-GGGG---UU-----CCCCG-?	GCAGG	TPCG <sup>m</sup> AU	CCUGC	CGACUAC	G CCA
<i>Rattus norvegicus, Morris hepatom</i>	-Ser	GUAGUCG	UG GCMG	AGD--#GDD-A	AGGC	R APGGA	UIGA+A	ZCCAU	3-GGGG---UU-----CCCCG-?	GCAGG	TPCG <sup>m</sup> AU	CCUGC	CGACUAC	G CCA
<i>Saccharomyces cerevisiae</i>	-Thr	GCUUCUA	UG LCCA	AGDD--GGD--A	AGGC	R CCACA	UIGU6A	PGUGG	AG-----AD?	AUCGG	TPCA <sup>m</sup> AU	CCGAG	UGGAAGC	A CCA
<i>Saccharomyces cerevisiae</i>	-Thr	GCUUCUA	UG LCCA	AGDD--GGD--A	AGGC	R CCACA	UIGU6A	PGUGG	AG-----AD?	AUCGG	TPCA <sup>m</sup> AU	CCGAG	UGGAAGC	A CCA
<i>Spinacia oleracea</i>	-Ser	GUGGACG	UG CCGG	AGD--#GDD-A	UCGG	R CAUGA	UIGA+A	PCAU	3-GGGC---UUU-----GCCCG-C	GCAGG	TPCG <sup>m</sup> AU	CCUGC	CGUCAC	G CCA

Figure 1-4 m<sup>3</sup>C occurs at position 32 of several tRNA species

The single quotation mark ‘ indicate 3-methylcytosine and were highlighted in yellow, figure are drawn from the tRNA database(114) by search against ‘ (3-methylcytosine) at position 32

ABP140 is the methyltransferase for 3-methylcytosine in *Saccharomyces cerevisiae*

In 2011, two groups published in the same issue the journal RNA, in which they found ABP140 (Actin-binding Protein 140 kDa) is the MTase responsible for installing m<sup>3</sup>C at position 32 on some tRNA<sup>Ser</sup> and tRNA<sup>Thr</sup> species in *S. cerevisiae* (47, 115), and they named it TRM140 (tRNA methyltransferase, 140 kDa). Both groups identified the enzyme via their in house genetic and biochemical screening, one used ribonucleome analysis and the other developed a primer extension assay screening. One group a knockdown of human Mettl2B and observe significant loss of 3-methylcytidine in HeLa cells using Mass spectrometry, further support Mettl2B as a human ortholog to Abp140 (115).

Disruption of the Abp140 gene did not lead to a lethal phenotype, only a double KO strain of ABP140 and TRMT1, which is responsible for m<sup>2,2</sup>G

modification, show a higher sensitivity towards cycloheximide, indicating a possible role of ABP140 in translation. Recently, a chemical genetic approach showed an ABP140 KO strain showed higher sensitivity to neomycin which is aminoglycoside antibiotic target the A-site decoding centre (116). This result implies that m<sup>3</sup>C at position 32 might modulate the A-site decoding when neomycin is added. Moreover, it is necessary to test phenotype of ABP140 KO under various stresses as mentioned earlier, including nutrient starvation, response certain chemicals, especially certain type of alkylation reagents, as discussed below in codon biased translation(117), to observe other phenotypes related to ABP140.

### **1.6 m<sup>3</sup>C in tRNA has shown to be involved in codon-biased translation**

As stated above, m<sup>3</sup>C is present at position 32 of tRNA Thr<sup>IGU</sup>, tRNA Ser<sup>UGA</sup>, and tRNA Ser<sup>CGA</sup> in *S. cerevisiae*, which reads tRNAs reading codons ACT, TCA and TCG, respectively(7). Position 32 is adjacent to the anticodon from 34 to 36 and has the potential to modulate codon- anticodon interactions. It was hypothesized that alkylation agents, such as MMS treatment would cause hypermodification of those tRNAs, thus leading to changes in translation rate and/or fidelity of mRNAs containing high levels of threonine and serine codons. One group observed SN<sub>2</sub> alkylating agents caused differential upregulation of m<sup>3</sup>C in tRNA, which is mechanically similar to H<sub>2</sub>O<sub>2</sub>-induced increases in m<sup>5</sup>C that led to selective translation of mRNAs enriched in Leu codon TTG(118). [javascript:void\(0\);](#) Moreover, the increase in mcm5U following MMS exposure has been linked to Trm9 and to the selective translation of mRNAs enriched with the Arg codon AGA(119). Following the line of logic for m<sup>5</sup>C and mcm5U, the group demonstrated a coordinated

translational stress response system involving stress-specific reprogramming of tRNA modifications that leads to selective translation of codon-biased mRNAs representing different classes of critical response proteins. tRNA modification subpatterns distinguished SN1 from SN2 alkylating agents, with SN2-induced increases in m<sup>3</sup>C in tRNA mechanistically linked to selective translation of threonine-rich membrane proteins from genes enriched with threonine's degenerate codons ACC and ACT (117).

### **1.7 Phylogeny and potential functions of tRNA 32–38 pair**

Located at the junction of anticodon loop and anticodon stem of tRNAs, residue at position 32 of tRNA are most frequently conserved as pyrimidines (Y) and usually forms a non-canonical base pair with the residue at 38(120), which is most frequently adenosine or uridine(120). Because of its critical location near the anticodon which is position 34, 35 and 36, 32-38 base pair has been involved in many tRNA functions including several post-transcriptional modifications(121), intron splicing(122), and aminoacylation (123). A comparison of over 5600 tRNA genomic sequences in *E. Coli* (114, 124) reveals an uneven base compositions at position 32–38. C-A, U-A and U-U are the most frequent and pairs containing G are least common. The majority of 32–38 pairs have been assigned to two different families of non-Watson-Crick base pairs. C32-A38, U32-C38, U32-A38 and C32-C38 pairs as well as corresponding pairs containing modified residues are included into family I. Cm32–A38 is one example for family I, in which a bifurcated hydrogen bond is formed between the 2' Oxygen atoms of cytidine and the 6' Nitrogen of adenosine. Family II contains the unmodified U32–U38 pair or U-U pairs containing Ψ. A single hydrogen bond is formed between the 3'

Nitrogen of Uridine 32 and the 4' Oxygen of Uridine 38. Of note, more than 90% of all 32–38 pairs belong to one of these two families, with several other base pair belongs to family III(125).

### **1.7.1 32-38 pairs in several tRNAs ensure uniform binding to the ribosomal A site and reading frame maintenance at P sites**

A series of studies showed that the 32-38 pair on tRNA<sup>Ala</sup> adjust tRNA affinity to the ribosome A site (126). When several different anticodons were cloned into tRNA<sup>Ala2</sup>, all mutated tRNAs interact with their corresponding codons at A site much weaker than WT tRNA<sup>Ala2</sup> (126). Further experiments revealed that tighter ribosomal binding was restored with WT tRNA sequence. In short, different 32-38 pairs will have different effects on changing the binding affinity of tRNA with the ribosome. It is hypothesized that tRNA<sup>Ala2</sup> has evolved to have 32-38 pairs that weaken its ribosomal binding to have similar binding affinity with other tRNAs. Interestingly, tRNA<sup>Pro</sup> in *E. Coli* which has a stable 5'-GGG-3' with 5'-CCC-3' pair, also harbours either a 32A-38U or the 32C-38G pair to decrease the affinity on tRNA binding. It also was observed that when either the modifications on tRNA or the esterified amino acid were removed, the binding of different tRNA towards A and P sites on ribosome were no longer uniform while nature occurring tRNAs showed similar affinity(127, 128), supporting the above hypothesis. More studies on ribosomal A sites' affinity to tRNA reveals that tRNA<sup>Gly</sup> of *E. Coli* binds tighter to both cognate and near-cognate GUC codon at the A-site of the ribosome when U to C mutation occurs at position 32 (129), while mutations at position 32 and 38 caused misincorporation of Ala at the near cognate GUC codon (113). Thus, the 32–38 pair has been used by multiple tRNAs to adjust



the affinity towards ribosome, and it is likely other tRNAs will adapt a similar strategy.

At the P-site of the ribosome, ribosomal protein S9 contacts residues 32, 33 and 34 on peptidyl-tRNAs(130, 131), which resulted in an increase in +1 frameshifting, indicating that the contact between S9 and residue 32 has a critical role in reading frame maintenance(132). Moreover, after the finding that tRNA suppressor efficiency is primarily defined by the ASL(133, 134), it has been shown that 32-38 pair modulate the functions of UAG, UAA and UGA tRNA suppressors (134).

### **1.7.2 The 32-38 pair and discrimination of synonymous codons**

Another study on pair 32-38 was showed in several *in vitro* translation assays using glycine tRNAs (129, 135, 136). For reading four glycine codons, GGA, GGT, GGC, GGG, *E. coli* use three tRNA<sup>Gly</sup> isoacceptors, whereas *Mycoplasma mycoides* only has tRNA<sup>Gly</sup>(UCC). The only different nucleotide between tRNA<sup>Gly</sup>UCC of *E. coli* and *M. mycoides* in the anticodon loop is position 32. It was shown that discrimination of the 3rd codon position was at least partially dependent on residue at position 32 since a U32 to C32 mutation lose the discrimination on glycine codons. Thus, it is position 32 that is able to fine tune the ability to recognize four glycine codons.

For mechanistic studies on the observation above, it was worthwhile to test *E. Coli* tRNA1<sup>Gly</sup> U32C mutant's binding to cognate codons and third position mismatches in the ribosomal A site, since previous studies demonstrated that 32-38 pair has a role in tRNA affinity to the ribosome (126). While wild type tRNA1<sup>Gly</sup> could form stable complexes with the near-cognate GGC, GGA and GGU glycine codons, the dissociation rate was too fast to measure and

indicates an over 50 fold faster dissociation from these mismatched codons than from the cognate GGG codon. In contrast, the mutated tRNA1<sup>Gly</sup> (U32C) dissociated from the three mismatched codons much more slower with a rate only about 12-fold faster than dissociating from its cognate GGG codon (128). An equilibrium binding assay also confirms the trends observed with the Koff assay, further confirms that tRNA1<sup>Gly</sup>U32C has a higher affinity to near-cognate codons than wild type tRNA1<sup>Gly</sup>. Therefore, the mutation in some way stabilized tRNA binding to ribosome and the nucleotides in WT tRNA weakens the binding(136).

It was suggested that 32-38 pair could adjust translational process by inducing conformation changes upon interaction with ribosomal residues(137, 138). However, the solved crystal structure of the A site shows no direct contacts between 32 or 38 nucleotides with ribosome(139) and biochemical experiments showed that only several 16S rRNA nucleotides has interaction with A site tRNA (140), all of which are located far from 32–38 pair (136, 141). It is possible that the 32–38 pair could indirectly modulate tRNA's affinity to the ribosome by modulating the anticodon loop conformation (120, 142).

### **1.7.3 Extended anticodon hypothesis**

The above evidence clearly suggest the importance of the 32-38 pair of aminoacyl-tRNAs in uniform and accurate codon recognition at both A-site and P-site. An extended anticodon hypothesis is proposed based on the evidence that tRNA anticodons correlated with nucleotides in the Anticodon Stem Loop (ASL) It is proposed residues 37 and 38, as well as several pairs in the anticodon loop-stem such as 32-38 pair, have evolved to suit the three

nucleotides in anticodon(143). In other words, each tRNA species may have evolved to match its anticodon to compensate for the different properties among aa-tRNAs , so that each aa-tRNA could be used relatively equivalently by the translational machinery(144).

One limitation of the above experiments is that binding or dissociation rates obtained from in vitro assay are far slower to those under normal translation. It is estimated that they are approximately  $10^3$ – $10^4$  slower than the rates of tRNAs' entrance and leave to the A site in cells. Nevertheless, X-ray solved structures support the above conclusion (120). On the other hand, it was shown that the accuracy of tRNA recognition *in vivo* mainly depends on the GTP hydrolysis kinetics and peptide bond formation, not on how tight tRNA bind to ribosome or how fast the disassociation rates (145). It is therefore needed to analyse the 32–38 pair in fine tuning translation using assays that can measure the tRNA kinetics in different steps in translation, such as ribosome profiling (146).

## **1.8 The study of translational control of gene expression by ribosome profiling**

As discussed above, the potential functions of 3-methylcytosine at position 32 in certain tRNAs in translation fine tuning calls for assay to observe the subtle effects. Here, a recent development of ribosome profiling by next generation sequencing is discussed for planned study of m<sup>3</sup>C.

### **1.8.1 mRNA abundance is far from a good proxy towards gene expression**

Measurements of messenger RNA (mRNA) abundance at global level using microarray or second generation sequencing have already greatly advanced the

study of regulation of gene expression(147) . However, there is a demanding need for direct identification and monitor proteins synthesized in cells for several reasons. First, mRNA levels cannot reflect real protein production since transcribed mRNA is subject to extensive post-transcriptional, translational and post-translational regulation (148-150) . Second, the effects from upstream ORFs (uORFs), internal ribosome entry sites (IRES), non-AUG start codons, and nonsense mediated read-through make it impossible to predict the exact protein product from the transcript sequence at the moment (151, 152). Third, ribosomal pausing during translation is revealed to help the co-translational folding and secretion of some proteins, thus prevent accurate predicting of protein production by mRNA abundance (153-155). In General, the correlation between mRNA abundance and protein remains poor frequently, thus levels of mRNA is far from a perfect proxy the gene expression.

### **1.8.2 Limitations of quantitative proteomics and polysome profiling followed by Microarray**

Advances in unbiased quantitative proteomics such as SILAC provides information about the identity and abundance of proteins at a time point (148, 149). However, there are still considerable limits on their ability to monitor protein translation on individual mRNA, regarding initiation, elongation or termination, as well as determine protein sequences independently or measure low-abundance proteins, thus cannot currently match the depth or the breadth of coverage that is obtainable from nucleic acid deep sequencing.

Polysome profiling followed by microarray, in which mRNAs are recovered from fractions of mRNA with different number of loaded ribosomes and sent

for microarray or RNA-seq, could provide a estimate of protein synthesis(156). However, this approach is hindered by the low resolution of fractions that have more than a few ribosomes per transcript, as well as that analyses of RNA from those fractions cannot indicates the position of ribosome on the transcript(157).

### **1.8.3 Global ribosome profiling as a promising tool to reveal functions of ‘minor’/’unimportant’ modifications**

Ribosome profiling (Riso-seq) is a method based on deep sequencing of ribosome-protected mRNA fragments. Purification and sequencing of these fragments provides a snapshot of all the ribosomes which are actively engaged in translation at a specific time point(158). This information can determine what proteins are being actively translated in a population of cells, providing information for investigating translational control; determining the rate of protein synthesis; studying frameshifting; predicting protein abundance, etc(157). The origins of the approach can be traced back to the finding that endoribonucleases treatment could convert polyribosome to monomeric ribosomes that is each bound to a short protected mRNA fragment. These assay named nuclease footprinting was first exploited to map the translation initiation sites of prokaryotic bacteriophage RNA(159) and eukaryotic reovirus RNAs(160). Because of the disability of adapting Sanger sequencing to genome-wide in that era, these experiments have to adopt cell-free translation systems tailed preferably with a single mRNA species (160). It was the advent of deep sequencing technologies, and the associated methods of cDNA libraries constructions that opened up the possibility of using the method to globally map the positions of ribosomes on mRNAs in intact cells.

One of the most important innovations of the method was to generate the protected mRNA fragments using *E. coli* RNase I, which has little sequence specificity, in contrast to the nucleases used previously (RNase A, RNase T1, or micrococcal nuclease). Such is the precision of RNase I cutting especially at the 5'-end, that the positions of ribosomes along the ORF can be determined with single nucleotide precision, and the footprints show a clear trinucleotide periodicity, which allows assignment of the translation reading frame and distinguishes footprints arising from translating ribosomes from RNA fragments that are protected for any other reasons. Ribosome profiling has been applied in mainly three aspects of translation study: the mechanism of protein synthesis, the identification of translated mRNA of the transcriptome, and translational control of gene expression (157, 161). The first of which is related with this Ph. D project and will be further discussed in Codon usage pattern part.

## CHAPTER 2. MATERIALS AND METHODS

### 2.1 Plasmids constructions

For recombinant DNA cloning and plasmid amplification, *E. coli* strain DH5 $\alpha$  / TOP10 (Invitrogen) was used as host. Bacteria cells were grown in LB liquid or on solid medium with ampicillin (100 $\mu$ g/ml) for selection of positive clones. Mouse Mettl2 was cloned from C57BL/6 E13 embryo liver cDNA pool and cloned into pPyCAGIP vector (gift from Dr. Wu Qiang, NUS) by XhoI/NotI. Human Mettl2 was cloned from human cervical cancer cell line HeLa S3 and inserted into pPyCAGIP vector by XhoI/NotI with Flag tag at N-terminus and HA tag at C-terminus. Point mutations at three Glycine residues and deletion mutation at SANT, SAM, motif were carried out following PCR based mutagenesis method (Quikchange). For bacterial expression, human Mettl2 WT and  $\Delta$ SAM cDNA were subcloned into pGEX6P1 vector by XhoI/EcoRI. pPyCAGIP-Flag-Mettl2-HA was further modified into pPyCAGIP-Flag-Mettl8-HA-T2A-EGFP by insertion of T2A peptide gagggcagaggaagtcttctaacaatgcggtgacgtggagg- agaatcctggccca and EGFP cDNA, which is cleaved in cells. All plasmids are verified by sequencing. pLKO.1 vectors harboring shRNA against Mettl2, Mettl8 and scramble control were either cloned or purchased from RNAi consortium through Sigma. A list of plasmids generated or purchased, and oligos used in generating plasmids are listed in the appendix.

### 2.2 Cell culture and treatment

All used cell lines, such as HEK293T, HeLa S3, A172, HCT116 (p53 wild type and isogenic p53 $^{-/-}$ ), U2OS and HepG2 were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal-bovine

serum (FBS) (Sigma, Gibco, thermoscientific) and antibiotics (100 units/ml penicillin and 100 µg/ml streptomycin) otherwise specified. H1229 cells were cultured in RPMI-1640 medium with the same supplements as above. The cells were incubated at 37 °C with a humidified atmosphere of 5% CO<sub>2</sub>. UV radiation is performed in Stratagene Stratalinker UV Crosslinker 2400. Gamma irradiation is performed in BioBeam 8000 irradiator (Gamma Service Medical GmbH, Leipzig, Germany) with <sup>137</sup>Cs as radiation source (dosage at 0.048Gy/sec).

## **2.3 Transfection**

### **2.3.1 Cationic liposomes based Transfection**

Lipofectamine 2000, Lipo RNAiMax or Mirus TransIT were used according to Manufacturer's instructions;

### **2.3.2 Calcium phosphate precipitation based transfection**

Calcium phosphate precipitation was done only in 293T/293 cells when confluency reaches 40%-60%. 20 µg of plasmid DNA in 450µl autoclaved sterile Milli-Q water was mixed with 50µl of 2.5M freshly-made calcium chloride. Then 500µl of 2×BES buffer (50 mM BES, 15 mM Na<sub>2</sub>HPO<sub>4</sub>, 280 mM NaCl. Adjust pH to 6.95 with 1 N NaOH) was added and the whole solution was vortexed immediately for 1 min followed by a 20 min incubation at room temperature to allow DNA precipitation. After incubation, the DNA solution was briefly spin down, pipetted up and down several times, and added evenly to dishes in a drop-wise manner. The transfected cells were incubated at 37 °C with 2.5% CO<sub>2</sub>. After about 16h, media was removed and replaced with 10ml fresh media. Then cells were moved back to 37 °C incubator with 5%



CO2 until cells reached full confluency for further passages, lysate/RNA collection or virus harvest.

### **2.3.3 Lenti-Virus mediated transduction**

Modified from Broad Institute, RNAi Consortium: To produce lentiviruses producing shRNA, shRNA-pLKO.1 plasmid or pLKO.1 control plasmid was transiently transfected into 293T packaging cells with 3rd generation lenti virus packaging plasmid and Envelop plasmid (VSVG & Gal-Pol) for producing viral particles by calcium phosphate precipitation. For the viral transduction in cells, Sigma protocol was followed. Briefly, cells to be transduced were seeded at appropriate density on Day 0. The next day lentiviruses were added to cells in growth media containing polybrene (6-8 µg/ml). Media was removed and replaced with fresh growth media on Day 2. After one or two days (cell type dependent) to wait for cells to robustly express puromycin resistant gene, media was removed and replaced with containing puromycin medium. After certain days of puromycin selection (concentration and cell type dependent), medium containing dead cells was removed and the remaining cell population was supposed to express shRNA of interest. The sequences of the shRNAs used are in Table 2-8. All the shRNA oligos were cloned or purchased from Sigma or IDT.

### **2.4 Real-time PCR**

Total RNA was harvested using Trizol reagent (Invitrogen) according to the manufacturer's instructions, purified by the RNeasy mini kit (Qiagen), and then reverse transcribed using the M-MLV Reverse Transcriptase system (Promega) or Superscript III (Invitrogen). The cDNA products were subjected to RT-PCR with KAPA SYBR® FAST Universal 2X qPCR Master Mix

(KK4600) using a 7300 or 7500 Real-Time PCR machine (Applied Biosystems). All mRNA expression values were normalized against the internal housekeeping gene GAPDH

## 2.5 Western Blot

Appropriate amount of lysate samples were electrophoresed on a polyacrylamide gel of a percentage dependent on the size of the target protein, transferred onto a PVDF membrane in cold room and probed with primary antibodies of interest at 4 °C overnight. The following experiment was processed using HRP conjugated secondary antibody (Biorad ) and enhanced chemiluminescence visualization method (various bands, e.g. Thermo Scientific).

### 2.5.1 Making resolving gel and stacking gel

Table 2-1 Resolving gel recipe

Resolving Gel* (15 ml )	8 %	10%	12%	15%
Water (ml)	7.0	5.9	5.0	3.5
30 % Acrylamide mixv(ml)	4.0	5.0	6.0	7.5
1.5 M Tris, pH 8.8 (ml)	3.8	3.8	3.8	3.8
10 % SDS (μl)	150	150	150	150
10 % Ammonium Persulfate (μl)	150	150	150	150
TEMED (μl)	36	24	24	24
Total (ml)	15	15	15	15

Table 2-2 Stacking gel recipe

Stacking Gel (4ml)	4%
Water (ml)	2.7
30 % Acrylamide mixv(ml)	0.67
1.5 M Tris, pH 8.8 (ml)	0.5
10 % SDS ( $\mu$ l)	40
10 % Ammonium Persulfate ( $\mu$ l)	40
TEMED ( $\mu$ l)	16
Total (ml)	4

Use ethanol to clean the glasses and green rack; Load resolving gel up to the green bottom line; Load 1mL of isopropanol at the top of the resolving gel; Wait for around 30 min; Remove water by inverting and placing the filter papers; Load stacking gel up to the top of the glass plate; Insert comb; Wait for 30-45 min; Remove comb by pulling it straight up slowly;

### 2.5.2 Sample preparation

Centrifuge samples at 13000 rpm for 2 min; Remove supernatant; Spin again at 13000 rpm for a few seconds; Use a smaller pipette to remove supernatant; Use 100 $\mu$ L of 50 mM Tris-HCl, 300 mM NaCl, including Protease inhibitors, and 0.1 Triton to Resuspend and lyse cells, boiled on 100°C hot plate; Add

30 $\mu$ L of loading dye to the two sets of tubes containing supernatant and pellet ; Denature at 100°C for 8-10 min ;

### 2.5.3 Sample loading and gel running

Table 2-3 10X Running Buffer

Tris	30g
Glycine	144g
MilliQ water	Up to 1L

Table 2-4 1X Running Buffer

10X running buffer	100 mL
20% SDS	5 mL
MilliQ water	900 mL

Insert the gel block into the tank, comb face inward; Add gel dam if only one gel is used; Fully fill the compartment between 2 gels with tank buffer; Load 30-35 $\mu$ L of sample into wells (40 $\mu$ L maximum) Amt of protein: 20 $\mu$ g; Load 6 $\mu$ L pre-stained marker in the first well; Half fill both side of the tank with tank buffer; Close the lid and connect the circuit ; Run gel at 100V~200 V, time depends ;

### 2.5.4 Transfer

Table 2-5 1X Transfer Buffer

10X running buffer	100mL
Methanol	200mL
MQ water	700mL

Prepare a container with methanol ; Use a forceps to remove the cover of the membrane and immerse it into methanol ; Prepare a container with transfer buffer ; Immerse the filter papers into the transfer buffer(4 filter papers per gel) ; Prepare transfer set-up accordingly. Avoid bubbles between gel and membrane;

Table 2-6 Gel sandwich in transfer

Black Node(-ve)
Black rack
Filter Pad
Filter Paper X2
Gel
Membrane
Filter Paper X2
Filter Pad
Transparent rack
Red/White Node(+ve)

Close cassette firmly and place it in module.(black face black) ; Place frozen ice pack into the tank( on the black side) ; Fill tank with 1X transfer buffer ; Run at 100V for 1.5-2hr until the marker bands reaches expected positions;

### **2.5.5 Blocking**

Dispense 10mL of 5% milk (blocking solution) into a container; Place the membrane in the blocking solution; Place on the shaker for 1hr, room temp or overnight at 4°C;

### **2.5.6 Primary antibody & secondary antibody**

Rinse membrane with 1XTBST 2 times (Optional) ; Wash membrane with 1XTBST for 15min, 2/3 times (Optional) ; Incubate membrane with Primary Antibody overnight at 4°C ; Antibody can be re-used for one more time ;

Rinse membrane with 1XTBST 2 times ; Wash membrane with 1XTBST for 15min, 3 times ; Incubate membrane with 2° antibody for 1hr at RT, or overnight at 4°C ; Discard the Secondary Antibody ;

### **2.5.7 Film development**

Rinse membrane with 1XTBST 2 times ; Wash membrane with 1XTBST for 15min, 2 times ; Wash membrane with 1XTBS for 15min ; Prepare a cling wrap ; Drain membrane and place it on the wrap ; Add 1mL of chemiluminescent HRP substrate to each membrane(ECL) ; (1:1-500mL white bottle +500mL black bottle) →975uL Big, 25uL Small bottle ; Wait 1 min for development and drain the extra HRP(ECL) ; Wrap the membrane, avoid bubbles ; Bring film, membrane, cassette, scissors, timer to dark room ; Cut film to appropriate size, cut the corner as a marker ; Place membrane and film into cassette for 30s ; Place film in the machine ; Adjust time accordingly ;

## **2.6 CRISPR/Cas9 gene targeting in both human and mice**

Vector pST1374-NLS-flag-linker-Cas9 containing human-codon-optimized Cas9 and human-gRNA-expression vector MLM3636 were from Addgene.

MLM3636 vectors with gRNAs targeting human Mettl2A, Mettl2B or both locus were self-made following instructions of Joung Lab. Cells were transiently transfected with Cas9 plasmid and gRNA expression vectors using lipofectamine/Mirus following manufacture's protocols. Cells were transfected in 24-well plates using 0.5 µg of the Cas9 expression plasmid and 0.3 µg of the RNA expression plasmid. Two days after transfection, cells were trypsinized and replated in 10 cm dishes with a highly diluted passage ratio for single clone selection. After about two weeks, single clones were picked and cultured in individual wells. The selection of positive clones was done first by western blot analysis of Mettl8 expression. Then genomic DNA of the Mettl8-null clones was extracted by phenol-chloroform method, and then subjected to PCR, TA cloning and sequencing to confirm the mutation of the target genes.

With the help from Model Animal Research Center of Nanjing University, our group successfully got METTL2, 6 and 8 constitutive KO mice. The mice are shipped to Singapore and maintained. Three pairs of sgRNA for D10A CAS9 has been assembled to the U6-promoter driven expression vector. The sgRNA efficiency has been verified. The sgRNA3 and sgRNA3 D10A has been transcribed and injected to the Mettl8<sup>+/-</sup> fertilized egg. Tails of the founder mice (45 founders). PCR has been performed to validate the mutation status of the loci. qPCR was performed to check mRNA of Mettl2, 6, or 8.

### **2.7 Modified quantitative PCR to measure m<sup>3</sup>C modification (qPCR-m<sup>3</sup>C)**

A quantitative PCR (qPCR) based method was modified to measure m<sup>3</sup>C modification (162). Each Serine or Threonine tRNA was reverse-transcribed in two reactions using two primers, in which reverse primer r1 was able to

anneal to the middle of this tRNA, overlap with the nucleotide at Cytosine 32, and reverse primer r2 was designed to target the region 3' to Cytosine 32. Subsequent qPCR was performed and generated two Ct Value Ct r1 and Ct r2, corresponding to data generated by reverse primer r1 and r2, respectively. The difference between Ct r1 and Ctr2 (hereafter dCtr2r1) was calculated by subtracting Ct r1 from Ct r2. The  $\Delta\text{Ct-r2r1}$  value was compared between RNA from WT or KD/KO cells

For isolation of total RNA from cells, Trizol reagents (Life Technologies) was used according to the manufacturer's protocol. Particular tRNA species, such as tRNA Thr (AGU), was further purified with biotin labelled oligos linked to Streptavidin Dyna beads, as described previously. The tRNA<sup>Ser</sup> & tRNA<sup>Thr</sup> isolated from wild-type (WT) and *Mettl2A/2B* KD/KO cells were subjected to qPCR-<sup>m</sup>3C. Threshold cycle (Ct) values calculated with the manufacturer's software (Life Technologies, 7300 System SDS Software) are indicated in this figure. For absolute quantification of <sup>m</sup>3C modification in crude total RNA, we subjected total RNA isolated from WT and *Mettl2A/2B* KD/KO cells to qPCR-<sup>m</sup>3C. Reverse Transcription and qPCR: The total amount of RNA isolated from cells and tissues was adjusted to 100 ng/ $\mu\text{L}$  in RNase-free water, unless otherwise indicated. To avoid genomic contamination, we digested 2  $\mu\text{L}$  (200 ng) total RNA with 5 U DNase I (Ambion) in a 20- $\mu\text{L}$  reaction at 37 °C for 20 min. We then heat-inactivated the DNase I at 75 °C for 10 min. After the DNase treatment, we mixed 2.5  $\mu\text{L}$  digested total RNA with 1  $\mu\text{L}$  20  $\mu\text{mol/L}$  reverse primer r1 or reverse primer r2, heat-denatured the mixture at 65 °C for 10 min, and cooled it rapidly on ice for at least 5 min. With the mixture still on ice, recombinant reverse transcriptase (M-MLV RT/ AMV RT



Promega) was added to a final concentration of 0.5 U/ $\mu$ L. Reverse transcription was performed in a total reaction volume of 10  $\mu$ L at 55 °C for 30 min and then heat-inactivated at 85 °C for 5 min. We subjected 2  $\mu$ L synthesized cDNA to qPCR with the SYBR Master Mix kit (KAPA) and the ABI PRISM 7300 Real-Time PCR System (Life Technologies) according to the manufacturers' instructions

## **2.8 Primer extension analysis of modifications in RNA**

Table 2 listed the oligos used in primer extension assay. To label oligos, 10-30 pmol primers were labeled with 2 times pmol (20-60pmol)  $\gamma$ -<sup>32</sup>P ATP (Perkin Elmer's BLU002Z500UC ) using T4 polynucleotide kinase (NEB), followed by removal of excess label using G25 chromatography column (GE healthcare). Primer extension assays were done as described in. In a 10  $\mu$ L reaction, 1-2 pmol of 5' end-labeled primers were annealed to 2  $\mu$ g total RNA (for testing cytoplasmic tRNA) or 10  $\mu$ g total RNA (for testing mitochondria tRNA ) in 0.5  $\mu$ L Superscript III buffer and by heating to 70°C for 5 min and slowing cooling (turning off heat block/water bath )done to 37 °C for 20-30 min. The annealed reaction was then extended using 1  $\mu$ L Superscript III (Invitrogen) and 0.11 mM of each dNTP (A, C, G, and T) for 30min at 50 °C - 55°C. The reaction was stopped by adding 1  $\times$  RNA load dye (formamide containing 0.1% bromophenol blue), heated for 5 min to 95 °C, cooled on ice, and loaded onto 15% polyacrylamide gel (20cm long. The longest gel in the institute) containing 6-8 M urea. The gel was dried on vacuum dryer (Bio-Rad) and exposed in a cassette with Phosphor Screen using different exposure time.

Table 2-7 5' end labelling of DNA probes for primer extension

Components	Amount
Primer	30 pmol
10X T4PNK Reaction	4 $\mu$ l
Buffer	
[ $\gamma$ -32P] ATP	50 pmol
6000Ci/mmol	
T4PNK	2 $\mu$ l (20 units)
Water	up to 40 $\mu$ l

#### Reverse Transcription

Table 2-8 Oligos used for primer extension for human cytoplasmic and Mitochondria tRNAs

Oligomer Name	Sequence	Oligo length	Predicted cDNA
m <sup>3</sup> C-G35-Thr(AGU)	GAACCCAGGATCTCTGTTTAC	22	25
m <sup>3</sup> C-G35-Thr(UGU)	GAACCTCGCGACCCCTGGTTTAC	22	24
m <sup>3</sup> C-G35-Thr(UGU2)	GAACCTCGCGACCCCTGGTTTAC	22	25
m <sup>3</sup> C-U39-Ser(AGA)	GAACCTGCGCGGGGAGACCCCAATGGA	27	29
m <sup>3</sup> C-U39-Ser(CGA)	GAACCCGCACACCCGAAACGCGGCTCG ACGGA	32	39
m <sup>3</sup> C-U39-Ser(UGA)	GAACCTGCGCGGGGAAACCCCAATGGA	27	34

m <sup>3</sup> C-G35- Thr(CGU)	GATCCATTGACCTCTGGGTAC	22	24
<sup>3</sup> C-mito- Ser(GCT)	TAGACATGGGGGCATGAGTTAG	22	24
m <sup>3</sup> C-mito- Ser(TGA)	TCGAACCCCCCAAAGCTGGTTTC	23	25

## 2.9 Recombinant protein purification for MTase assay and growing crystals

Either His-Tagged or GST-tagged Proteins was usually purified by affinity column first and by size exclusion column using FPLC.

### 2.9.1 Cell harvest

For large scale expression, usually 5 litres *E. coli* were grown and harvested.. Inoculate a single colony of BL21 *E. Coli* carrying desired plasmid into 20 ml LB medium (100 µg/ml Amp) and culture overnight at 37 °C. Dilute amount of overnight culture into 1 L LB medium to obtain an OD600 at less than 0.1 and culture at 37 °C for 3-4 hour when OD600 reaches approx. 0.5-0.8. Add IPTG to 0.1-0.5 mM and culture 12-16 hours at 15 °C. Pellet cells at in a polystyrene plastic bags in JA-10 rotors' bottles at 6,000 rpm for 20 mins and use liquid nitrogen to freeze the pellets to store at -80 °C or proceeded to affinity purification.

### 2.9.2 Disruption of cells by French press and column loading

Resuspend cell pellets in 50 - 100 ml lysis buffer (20 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, add 0.1 % (v/v) Triton X-100 and β-mecarpoethnal to 5mM just before resuspension. Incubate on ice for 20 mins with stirring and homogenize the cells using a homogenizer, otherwise it would clog the French

press system; Use French press to break the cell (PANDA system); Or use sonication 10 sec pulses with 10 sec intervals (30% amplitude) on ice until suspension becomes partially clarified, (around 30 mins), centrifuge at 15,000 rpm in JA-20 for 20 mins ; Load supernatant into column and collect the flow-through; For PANDA system: Open the machine half an hour before use, to cool down to 7 degree; Wash with milliQ water and later with the buffer you used to lysate the cells, add the cell supernatant and then adjust the pressure to 600-700 mPa; wait for three minutes between each run;

### **2.9.3 Wash and elution**

Wash column with wash buffer (20 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole, pH 8.0); Elute protein in an increasing concentration of imidazole in elution buffer (20 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, and 500 mM imidazole, pH 8.0). Analyse selected fractions by SDS-PAGE, collect desired fractions and dialyse or concentrate against buffer suitable for size exclusion chromatography. Perform gel filtration in a similar manner using Hiload, Superdex 16/60cm, one GEL filtration buffer receipt for METTL2, 6, 8 purification: 20mM Tris-HCl, 500mM NaCl, pH 7.5.

### **2.10 *In Vitro* Transcription (IVT) of tRNA**

tRNA was In vitro transcribed to serve as potential substrates for METTL2. T7 RNA Polymerase catalyses the synthesis of RNA in the 5'- 3' direction in the presence of a DNA template containing a T7 phage promoter. In this study, tRNA is In Vitro Transcribed to serve as substrates for 3-methylcytosine In Vitro reconstitution. PCR products containing T7 RNA Polymerase promoter in the correct orientation can be transcribed. Due to the complexity of multiple copies and secondary structure of tRNA gene, gBlocks® Gene

Fragments(IDT), which are double-stranded, sequence-verified genomic blocks are ordered (200ng) and further amplified by PCR and served as templates for IVT. PCR mixture is purified in a silica based PCR clean-up column (Qiagen/Thermo) to ensure better yields in the IVT reactions. PCR products is examined on agarose gels to estimate concentration and to confirm amplicon size prior to its use as a template in the IVT. Template size is important in optimizing optimal template input. The optimum template concentration ranges could be up to several micromolar for transcripts around 75nt, since shorter transcripts needs higher template concentrations. 1-2 µg of PCR fragments are used in a 50 µl in vitro transcription reaction. Reaction is assembled at room temperature in the following order in Table. Mix thoroughly, pulse-spin and incubate at 37°C for 2-4 hours. To remove template DNA, add 2 µl of DNase I (Ambion, RNase-free), mix and incubate at 37°C for 15 minutes. To remove potential RNase, add Protease K to 200 µg/ml with 0.5 % SDS. Transcribed tRNA was finally purified by Phenol/chloroform extraction and alcohol precipitation.

### **2.11 Purification of specific tRNA species by solid-phase DNA probes**

Modified from(*163*): Oligodeoxyribonucleotides (about 30 nt) with sequences complementary to the 3' side of target tRNAs were synthesized with biotinlyted 5' end, immobilized on streptavidin C1 dynabeads (Invitrogen) and used as solid-phase probes. A mixture of tRNAs was added to a suspension of the solid-phase probe in 2.4 M tetraethylammonium chloride and incubated for 10-30 min. Only a target tRNA hybridized with the immobilized probe at appropriate temperatures and was eluted out by heating. The sequences of the

3'-biotinylated DNA probes complementary to human tRNAThr AGU and tRNASer UGA were: respectively.

### 2.12 Reconstitution of 3-methylcytosine by recombinant Mettl2B

The reaction mixture (100  $\mu$ L) contained 1 times T7 RNA polymerase buffer, 200 ng of small RNAs (<200nt, by Mireasy kits), and 10-20  $\mu$ g recombinant Mettl2B, in the presence or absence of 0.5-1 mM Ado-Met. The reactions were incubated for 0.5 h at 30°C, 0.5 h at 37 °C. After incubation, the reactions was precipitated by 1.5 volume of 100% Ethanol and go through RNA microcolumns, eluted and assayed by Primer extension assay as mentioned above.

Table 2-9 In Vitro Transcription Reactions

Reagent	$\mu$ l
10X Reaction Buffer	5
ATP (100mM)	5
UTP (100mM)	5
CTP (100mM)	5
GTP (100mM)	5
Nuclease-free water	Xul
Template DNA from PCR	10ug
T7 RNA Polymerase Mix	2ul
Total reaction volume	50.0

Table 2-10 In vitro reconstitution of m3C

Reagent	$\mu$ l

Nuclease-free water	
10X RNA MTase Buffer	
tRNA	
Recombinant Protein	
SAM (32 mM)	
Total reaction volume	50.0

## **2.13 Methyltransferase assay using tritium labelled S-adenosylmethionine ( <sup>3</sup>H-SAM )**

### **2.13.1 In vitro <sup>3</sup>H-SAM binding assay**

10µg recombinant GST, GST-Mettl2 WT and ΔSAM protein were conjugated to 10 µl Glutathione Sepharose 4B at 4 for 30min before being incubated with 5µl <sup>3</sup>H-SAM (12-18 Ci/mmol (444-666 GBq/mmol)) at 30°C for 30min. After two washes with 0.5ml 100mM Tris-HCl at pH 7.5, the beads were transferred to scintillation tube and mixed with 2 ml scintillant cocktail (OptiPhase HiSafe 2, Perkin Elmer). The result was obtained from triplicate reactions on Wallac 1414 Liquid Scintillation counter controlled by the WinSpectral software system.

### **2.13.2 In vitro Methyltransferase assay using scintillation counter or fluorography**

*In vitro* Methyltransferase assay was either measured by a scintillation counter or by fluorography. For scintillation based assay, tRNA from either WT or METTL2, 6, 8 Knockdown or knockout cells were incubated with 2.5 µl <sup>3</sup>H-SAM and 10 µg recombinant proteins and went through G25 spin based size

exclusion columns to get rid of unincorporated  $^3\text{H-SAM}$ . Scintillation counting was compared between WT and KD/KO samples. For fluorography based methods, *In vitro* reactions were purified by G25 spin columns to get rid of unincorporated  $^3\text{H-SAM}$  and directly loaded on the 15% Urea SDS-PAGE, enhanced using 3H-Enhancer and precipitated using cold water and PEG8000, dried and exposed to X-ray film at  $-80\text{ }^\circ\text{C}$ . Double stranded DNA ladders were labelled by M.Sssl and  $^3\text{H-SAM}$

#### **2.14 Mass spectrometry in this project**

In this project, several experiments rely on the Mass spectrometry to obtain information, such as identification of automethylation sites on METTL2, 6, 8 proteins (performed at Zhao Yingmin's lab at Shanghai); resolving protein complex pulled down by Flag IP following manufacturers' protocols ; PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) followed by identification of crosslinked RNA species, performed by Prof. He Chuan's lab at University of Chicago; and quantification of 3-methylcytosine, which is described below

#### **2.15 Size exclusion HPLC and LC-MS/MS for quantification of modified ribonucleosides**

Total RNA quality and composition was determined using Bioanalyzer Nano 6000 Chips. tRNA were isolated from samples with RIN  $> 8$  were using size exclusion HPLC as previously described by(164). Briefly, total RNA was separated into its components using a Bio SEC-3 column (Agilent; ID: 4.6 mm, length: 300 mm, particle size:  $3\text{ }\mu\text{m}$ , pore size:  $100\text{ }\text{\AA}$ ) under isocratic elution of 100 mM ammonium acetate at flow rates of 1 mL/min at  $60\text{ }^\circ\text{C}$ . tRNA fractions were collected, concentrated with a 10 KDa MWCO spin filter, spin-



dialyzed against ultra-pure water and quantified by Ribogreen fluorescence (Invitrogen). 2 µg of tRNA from each sample were enzymatically hydrolyzed to nucleosides using benzonase (Sigma), nuclease P1 (Sigma) and bacterial alkaline phosphatase (Invitrogen) in the presence of antioxidants and deaminase inhibitors, cleaned up by spin-dialysis, resolved by HPLC using a Hypersil Gold aQ column (Thermo Scientific, ID: 3 mm, length: 50 mm, particle size: 1.9 µm) using a gradient elution between ultrapure water with 0.1% (v/v) formic acid and acetonitrile (LC-MS grade, Merck) with 0.1% (v/v) formic acid at 25 °C, and analyzed on a Agilent 6460 triple quadrupole mass spectrometer by multiple-reaction monitoring (MRM) in positive ion mode. Transitions for adenosine (A) 268.1 → 136.1 m/z, guanosine (G) 284.1 → 152.1 m/z, uridine (U) 245.1 → 113.1 m/z, cytidine (C) 244.1 → 112.1 m/z, m<sup>3</sup>C, m<sup>5</sup>C and 4-methylcytidine (m<sup>4</sup>C), 258.1 → 126.1 m/z, were used for relative quantification. Instrumentation for this process is described (165).

## **2.16 RNA immunoprecipitation (RIP) assay**

RIP assay were performed to validate METTL2, 8 binding targets identified by PAR-CLIP. RIP is analogous to IP, with different purpose. RIP immunoprecipitates a s RNA binding protein (RBP) and associated different RNA species that can be detected by RT-PCR, real-time PCR, microarrays or RNA-seq. Below is the RIP protocol adapted from (166, 167)

### **2.16.1 Collecting cytoplasmic and nuclear cell lysis and RIP**

Prepare fresh cell lysis buffer as in section 2.5.2 (plus adding RNase inhibitor) , Add lysis buffer to collected cell pellets keep on ice for 20 min

with frequent mixing. Centrifugation at 10,000 g for 10 min to pellet nucleus. It is of critical importance to avoid contamination by RNase or DNase.

### 2.16.2 RNA Immunoprecipitation

Add antibody (1 to 8 µg) to supernatant (5 mg-10 mg) and incubate for overnight at 4°C with rotation. Add protein A/G beads (40 µl) and incubate for 1 hr at 4°C with gentle rotation. Washing conditions need to be optimized. Pellet beads at 2,500 rpm for 30 sec or using magnetic rack if using dyna beads, remove supernatant, and resuspend beads in 500 µl RIP buffer. Repeat for a total of three RIP washes, followed by one wash in PBS. Purification of RNA by resuspending beads in TRIzol or other similar RNA extraction reagent. Elute RNA with nuclease-free water (e.g. 20 µl). Add approximately 15-25 µl (depending on yield) of either DEPC treated water or Nuclease free water to the RNA pellet.

### 2.16.3 Reverse transcription (RT)-PCR

Perform reverse transcription of DNase treated RNA according to manufacturer's instructions. If candidate target is obtained by PAR-CLIP or other assay, PCR or qPCR could be directly used towards suspected targets; if target is not known, cDNA libraries could be analysed by microarrays or 2<sup>nd</sup> generation sequencing. The control experiments should give no detectable products after PCR amplification.

## 2.17 Polysome profiling-

### 2.17.1 Preparation of 10% and 50% W/V sucrose:

Weigh 5g (10%) or 25g (50%) of sucrose into a 50ml Falcon tube; Add RNase free water to sucrose at 35 ml mark; Add the following:

Table 2-11 Sucrose stock solution

Components	Stock	Volume
75mM KCl	2M KCl	1.875ml
1.5mM MgCl	1M MgCl	75 $\mu$ l
10mM Tris-HCl pH7.4	1M Tris-Hcl pH 7.4	500 $\mu$ l

Place in boiling distilled water for 15 mins. Shake gently to mix sucrose every minute or so; Cool on ice; Add water to a final volume of 50ML; Filter through a 0.45um filter into a fresh 50ML falcon tube. Store at -20 °C until use.

### 2.17.2 Preparation of buffers

Table 2-12 2X Resuspension Buffer (RSB) Recipe

Components	Stock	Volume
20mM Tris-Hcl (pH7.4)	1M	200 $\mu$ l
300mM NaCl	5M	600 $\mu$ l
30mM MgCl	1M	300 $\mu$ l
ddWater		8.9 ml
Final Volume		10 ml

Table 2-13 1x RSB with Cycloheximide

(Resuspension Buffer, for 2 samples)

Components	Stock	Volume
2 x RSB	2 x	150 $\mu$ l
SuperasIn	20 U/ $\mu$ l	22.5 $\mu$ l
200 $\mu$ g/ml Cycloheximide	100mg/ml	0.6 $\mu$ l

Rnase free ddH <sub>2</sub> O		126.9 µl
Total		300 µl

Make fresh each day

Table 2-14 1X Lysis Buffer

(For 2 samples, make fresh each day)

Components	Stock	Volume
2 x RSB	2 x	500 µl
1% Triton X	10% stock	100 µl
2% Tween	10% stock	200 µl
1% deoxycholate	10% stock	100 µl
RNase free ddH <sub>2</sub> O		100 µl
Total		1000 µl

### 2.17.3 Cell harvest and sucrose gradient profiling

Cells are treated for 10 minutes with 100ug/ML Cyclohexamide at 37 °C incubator; Wash the cells with warm PBS (with 100ug/ML cycloheximide 1:1000); Cells are harvest by trypsinization (with 100ug/ml cycloheximide 1:1000) and neutralized with ice cold media with FBS (with 100ug/ml cycloheximide 1:1000) ; Spin the cells at 1500rpm for 5 mins 4°C. Add 1 ml of the cold PBS (with 100ug/ML cycloheximide 1:1000) Transfer into a new 1.5ml cold tube. Take out 50ul AND ADD 1ml TRIZOL. This is the total RNA; Spin and wash the cells with cold PBS for 2 mins, 2000rpm twice For SW41 gradient samples use the larger volumes. For SW60 gradient samples use the smaller volumes in brackets. Resuspend cell pellet in 300 (150) µl

Fresh RSB. Take out 320 (160)  $\mu$ l and put into a fresh cold 1.5ml tube. Add in the same volume of 320 (160)  $\mu$ l Fresh Lysis Buffer. Mix gently and leave it on ice for 10 mins. Shake gently every 2 minutes. Spin full speed for 3 mins to remove nuclei. Nuclear pellet will be smaller and whiter than the cell pellet. Transfer about 600 (300)  $\mu$ l extract into a new cold tube and spin full speed for 10 mins. Transfer into another new tube. Take out 10ul extract to measure the OD units (1/50 dilution). Take 50ul extract as unfractionated lysate. Add 1ml Trizol. SW41 only - Add 400ul of 10% sucrose to the top of the sucrose gradient (10%-50%) and weigh the tubes. Measure and load the same OD Units onto sucrose gradient. Total vol: 400 (250) ul. Top up with 2xRSB if necessary.

Making Sucrose gradients: Gradients should be made while the cells are being washed. Each spin cycle is 5 mins during which time gradient is prepared. Chill the centrifuge buckets on ice for 30 mins before spin. Insert kimiwipes into tube to collect condensation. Clean tubes, caps and syringe with RNase-away (Ambion) and wash with nuclease free water and leave to dry (Do this early in the morning); Add the room temp 10% filtered sucrose to the first mark ; Add the room temp 50% filtered sucrose to the second mark. Put in the long caps gently to avoid bubbles.

Use the biocomp gradient machine. Level the machine and choose the program. SW41-List- long caps 10%-50% w/v; After gradient formation, remove cap gently. Do not tilt the tube or disrupt the gradient.

Add 400ul 10% sucrose (200ul twice by p200) on top of gradient very gently to avoid mixing of sucrose with lower layers – No lid present. Weigh the tubes and adjust weight with 10% sucrose. Do not disturb the gradient. Add in the sample extract and if necessary, top up with 2xRSB; Label the side of each tube with sample name and note which bucket is used. Place gradients in the chilled buckets and load 400ul of sample. Seal bucket tightly with the lid.(O-rings must be recently greased); Spin at 8 °C, 36000rpm, 1.5hrs - 2hrs in beckman centrifuge

Collecting the fractions (after the 1.5-2hrs ultracentrifuge spin) Prepare and label RNase free 2ML eppendorfs on ice. Collect 1ml in each fractions. After fractionations, add in 110ul of 10% SDS and 12ul of proteinase k (20mg/ml). Incubate and shake at 42 °C, 30 mins. Samples are stored at -80 °C.

## **2.18 Immunoprecipitation**

Cell lysate was extracted in lysis buffer (50 mM Tris-HCl pH 7.5, 0.1 mM EGTA, 1% Triton X-100, 5 mM sodium pyrophosphate, 1 mM sodium orthovanadate, 50 mM sodium fluoride, 0.27 M sucrose, 0.1% (v/v) 2-ME, plus 1 tablet/50 ml of EDTA free complete protease inhibitor cocktail (Roche)). Protein lysates from cells were centrifuged at  $14,000 \times g$  for 5 min at 4 °C and the insoluble debris discarded. Protein concentrations were determined using Bio-Rad protein assay dye. 10 µl of Flag M2 beads (for Flag IP) (Sigma) or 2 µg of relevant antibody coupled to 10 µl of protein G-agarose beads (Thermo Scientific) were washed with 1ml lysis buffer before

incubation with 1mg total lysate for 1 h at 4 °C on a 1000rpm orbital shaker. Proteins bound to the beads were separated from the cell lysis by 10,000 g centrifugation for 1 min, washed twice with 1 ml of lysis buffer with 0.5 M NaCl and twice with 1 ml buffer B (50 mM Tris-HCl, 0.27 M sucrose and 0.1% (v/v) 2-ME, pH 7.5). After removing all the remaining supernatant, SDS-Laemmli buffer was added to the beads to denature the antibody and release the immunoprecipitated protein. The samples were boiled for 10 min before being subjected to SDS- PAGE electrophoresis and subsequent Western blotting analysis.

### **2.19 Immunofluorescence staining for cultured cell lines (IF-IC)**

Cells grown on glass cover slips were washed with PBS solution, fixed in 4% paraformaldehyde at room temperature for 10-15 min and permeabilized with PBS (0.2% Triton X-100) for 10 min, then blocked with 3% BSA in PBS (0.1% Triton X-100) for 1 hour at room temperature, followed by incubation with primary antibodies at 4 °C overnight. Cells were washed three times with 0.1% Triton X-100 in PBS, 10 min each time and incubated with fluorescence-labelled secondary antibodies in dark at room temperature for 1h. Nuclei counterstaining were performed using 0.005% DAPI staining. Coverslips with cells were mounted onto slides by VECTASHIELD Hard-set Mounting Medium (Vector Laboratories, H-1400). Microscopy Image acquisition and analysis was performed using Nikon A1R-A1 confocal microscopy system, which is owed by Cancer Science Institute, Singapore.

### **2.20 Codon-run reporter constructs**

pEF6 vector is inserted with codon optimized Renilla luciferase (Rluc) and Firefly luciferase (luc2) cloned from pMirGlo from Promega using oligos

below, with stop codon for Rluc excluded. Renilla luciferase was inserted into pEF6's via KpnI and BamHI RE sites located at Multiple cloning sites while firefly luciferase was inserted using EcoRI and NotI, thus leaving 5'-ACT AGT CCA GTG TGG TG-3' linker located between Renilla luciferase and Firefly luciferase.

Table 2-15 Oligos for cloning Renilla luciferase and Firefly luciferase

hRluc_F_KpnI	GTGAGGTACCATGGCTTCCAAGGTGTACGACC
hRluc_R_BamHI	ATGCGGATCCGAAGAAGCTCGTCAAGAAGGCGATAG
luc2_F_EcoRI	CTAGGAATTCATGGAAGATGCCAAAAACATTAAG
luc2_R_NotI	AGATAAGCGGCCGCTTACACGGCGATCTTGCCGCCCTTC

Runs of codons matching the anticodons of several human tRNAs which is approved or suspected to have m<sup>3</sup>C were generated by annealing DNA oligos containing BamHI/EcoRI overhangs at either ends and harboring five or ten successive, identical Threonine, Serine, or Glycine codons or Control codons, and inserted between the BamHI and EcoRI sites downstream of the Rluc2 and upstream of luc2. The Fluc/Rluc ratio on transfection of the constructs into cells from METTL2 KO or WT cells was determined using luciferase assay. Luciferases was done using passive lysis buffer according to manufacturer's instruction (Promega Dual Luciferase kit) and measured manually in triplicate.

Table 2-16 Oligos for Codon Runs

Sequence Name	Sequence 5'-3'
ThrACU10*F	<b>GATCC</b> ACTACTACTACTACTACTACTACTACTACTACTG
ThrACU10*R	<b>AATTC</b> AGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTG





```

for a in N:

    for b in N:

        for c in N:

            codons.append(a+b+c)

def doCount(seq):

    thisCount = [0 for i in range(len(codons))]

    if len(seq) % 3 != 0:

        print "Not divisible by 3"; return thisCount

    for i in range(0, len(seq), 3):

        thisCount[ codons.index(seq[i:i+3]) ] += 1

    return thisCount

ccdsToid = (113)

convert = False

f = gzip.open("refseq.UCSC.hg19.fa.gz", 'rb')

out = open("hg19.refseq.count.txt", 'w')

currSeq = None

currName = None

```

```

out.write("Name\t")

out.write("\t".join(codons) + "\n")

for r in f:

    if r[0] == '>':

        if currSeq != None:

            currCount = doCount( currSeq );

            if not convert: #or currName.split(' ')[0].split('_)[-1] in ccdsToid:

                out.write(currName + '\t')

                #out.write( ccdsToid[currName.split(' ')[0].split('_)[-1]] + '\t')

                out.write("\t".join(map(str,currCount)) + '\n')

            else:

                print "no conversion"

            currName = '_'.join(r.strip().split(' ')[0].split('_')[2:])

            currSeq = None

        else:

            if currSeq == None:

                currSeq = r.strip()

            else:

                currSeq += r.strip()

```

## 2.22 Clonogenic assay

Clonogenic assay of cells in vitro were adapted from(168). Cell counting and uniform seeding is of critical importance in this assay. Colonies at the bottom of the vessels were fixed in 3 - 4% paraformaldehyde (v/v) and stained using crystal violet 0.5% (wt /v) dissolved in water.

Treatment could be performed either before or after cells are plated into dishes. The first option is often used for a quick screening of the sensitivity of cells to different treatments. In the second option, cells are treated in dishes and subsequently re-plated in appropriate dilutions to assess clonogenic ability. The re-plating may be performed immediately after treatment (IP) or it may be delayed (DP) to allow recovery. It is often used in radiobiology to study lethal- or sub-lethal damage repair.

Stop cell growth when control dishes have obtained sufficiently large and round colonies. Wash gently with PBS and fix by covering plates with 3%-4% paraformaldehyde for 10 to 15 minutes, remove and rinse with PBS, add crystal violet solutions and stain for around 20 min and submerge the plates into tap water. Leave the vessels to dry at room temperature.

SeaPlaque<sup>TM</sup> Agarose with low melting temperature (Lonza, 50100) was used in Colony Formation. In a triplicate manner, 1.5ml culture medium with 0.6% agarose was first plated into each well of a 6-well plate. After base agarose solidified, another 1.5ml of 0.4% agarose was plated on top in culture medium containing 2500 cells per well. After 20 days colonies grown in soft agar were stained by 1 mg/ml Thiazolyl Blue Tetrazolium Bromide (Sigma, M5655) dissolve in water and then scanned in Bio-Rad Gel Doc EZ system. A

population with more than 50 cells was counted as one surviving colony.

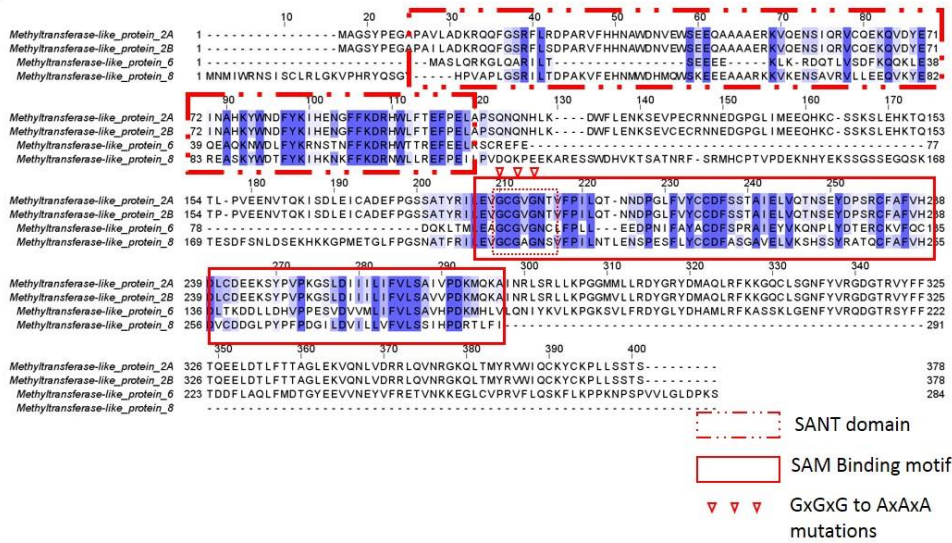
Colonies were counted using Quantity One software.

## CHAPTER 3. RESULTS

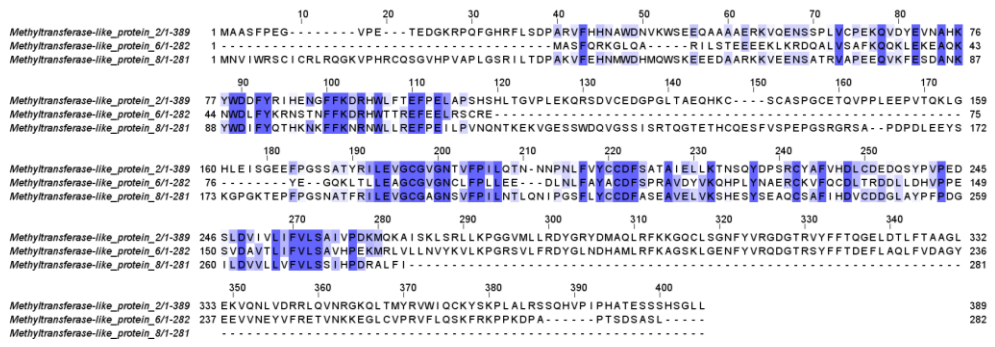
### 3.1 Sequence alignment of METTL2 with its paralogs and homologs

Mammalian homologs of ABP140 were searched against BLAST and found Methyltransferase-like protein 2A & 2B (Mettl2A, 2B) and Methyltransferase-like protein 2 (Mettl2) as human and mouse orthologues, respectively, with Mettl8 and Mettl6 ranking at second and third position in the BLAST list. As shown in Figure 3-1 (A), METTL2A, 2B, METTL6 and METTL8 are sharing high sequence homology and thus are considered as paralogs with each other. METTL2, 6, 8 possess the conserved SAM and nucleotide binding motif in TRM140, but lack the N-terminal region of TRM140. METTL6 is shorter than METTL2A, 2B, and 8 for lacking of another approximately 80 amino acids at N-terminal, the so-called SANT domain (169, 170). Searching for Mettl2A, 2B's homologs in mammals and sequence alignment by Clustal W indicates they are really conserved as shown Figure 3-1(B), and homologs are identified in zebra fish as METTL2A, 2B, 6, 8; *C. elegans* as Y53F4B.42, and even in *Trypanosoma Brucei* as XP\_827431 (GenBank: EAN77101.1). These phylogenetic studies showed TRM140 are quite conserved through evolution, and even creates three or four highly similar paralogs as Mettl2, 6, 8 or Mettl2A, 2B, 6, 8 in mammals, implying a potential indispensable role of these proteins.

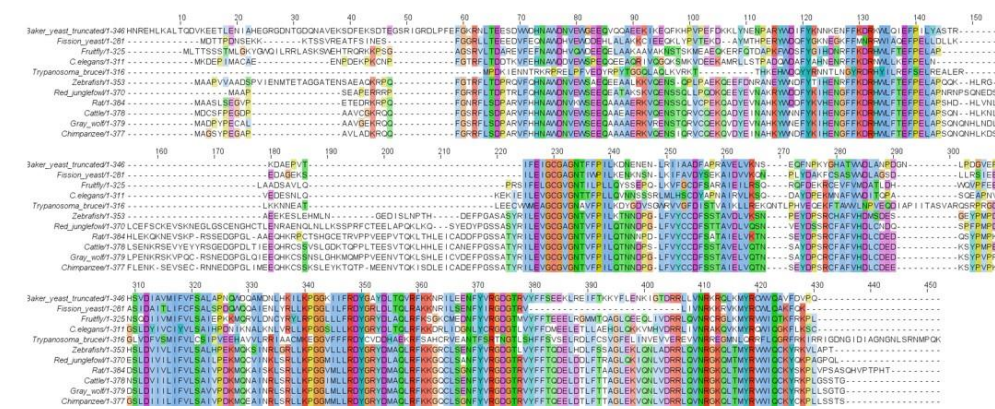
A



B



C



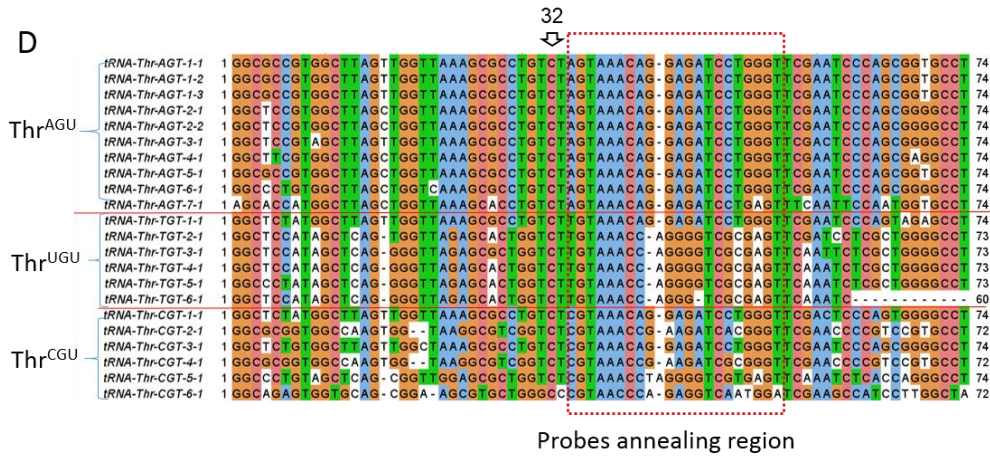


Figure 3-1 Sequence alignment of the METTL2 with its paralogs and homologs

(A) Clustal W sequence alignment of the human METTL2A, 2B, METTL6 and METTL8 proteins; with putative domain/motifs shown in red dashed boxes; Accession numbers are NP\_859076.3 (METTL2A), NP\_060866.2 (METTL2B), NP\_689609.2 (METTL6) and NP\_079046.2 (METTL8)

(B) Sequence alignment of full length mouse METTL2,6,8 proteins

(C) Sequence alignment showing the conservation of Mettl2, 6, 8 from baker's yeast to higher mammals, Accession numbers are as follows: *Drosophila melanogaster*: NP 647636.3; *Schizosaccharomyces pombe*: CAB76043.1; *Saccharomyces cerevisiae* YOR239W; *Pan troglodytes*: XP 001144324.1; *Gallus gallus*: NP 001006329.1; *Canis lupus*: XP 537604.3; *Bos Taurus*: NP 001068714.1; *Rattus norvegicus*: NP 001102309.1; *Caenorhabditis elegans*: NP 001040827.1; *Danio rerio*: NP 001017902.1; *Trypanosoma brucei*: XP 827431.1

(D) Alignment of all human threonine tRNA isodecoders. Sequences are obtained from GtRNAdb database (171)

### 3.2 Subcellular localizations of METTL2, 6 and 8

The study of localizations of METTL2, 6, 8 will provide insights into the subcellular localizations of the methyl transfer reactions. The antibody which successfully distinguished lysates in Western Blot between METTL2 WT and KO by Cas9/CRISPR were chosen for immunofluorescence staining. As shown in Figure 3-2 (A) and (B), METTL2 was found to be both in the nucleolus (as co-localized with the nucleolus marker: Fibrillarin) with a strong



signal and cytoplasm with a relatively weaker signals in and HeLaS3 cells. For another Glioma cell line A172, METTL2 is found in both nucleolus as well as cytoplasm, though signal intensity is close in both compartments. Consistently, the localization of METTL2 in the nucleolus was found in Snu387, HEK293T, HepG2, H1229, U251MG, HCT116, primary human fibroblasts, etc. (Confocal images available upon request). It is noteworthy to mention METTL2 signals are lost in the nucleolus in cells during Mitotic phase, indicating a possible cell cycle dependent localizations of METTL2.

Intriguingly, Ectopic overexpressed Flag-tagged, EGFP-tagged or YFP-tagged METTL2 was only found to be in the cytoplasm, perhaps due to disturbed nucleolus localization signals by tagging Flag, EGFP, or YFP. Intriguingly, both the nucleolus and cytoplasm localization signals of METTL2 was still maintained in T2-1 #3 & #7 METTL2 KO cell clones, indicating either the antibody recognize the smallest isoform of METTL2, which is not deleted in those KO clones, shown by Western Blot, or it is totally a non-specific signal.

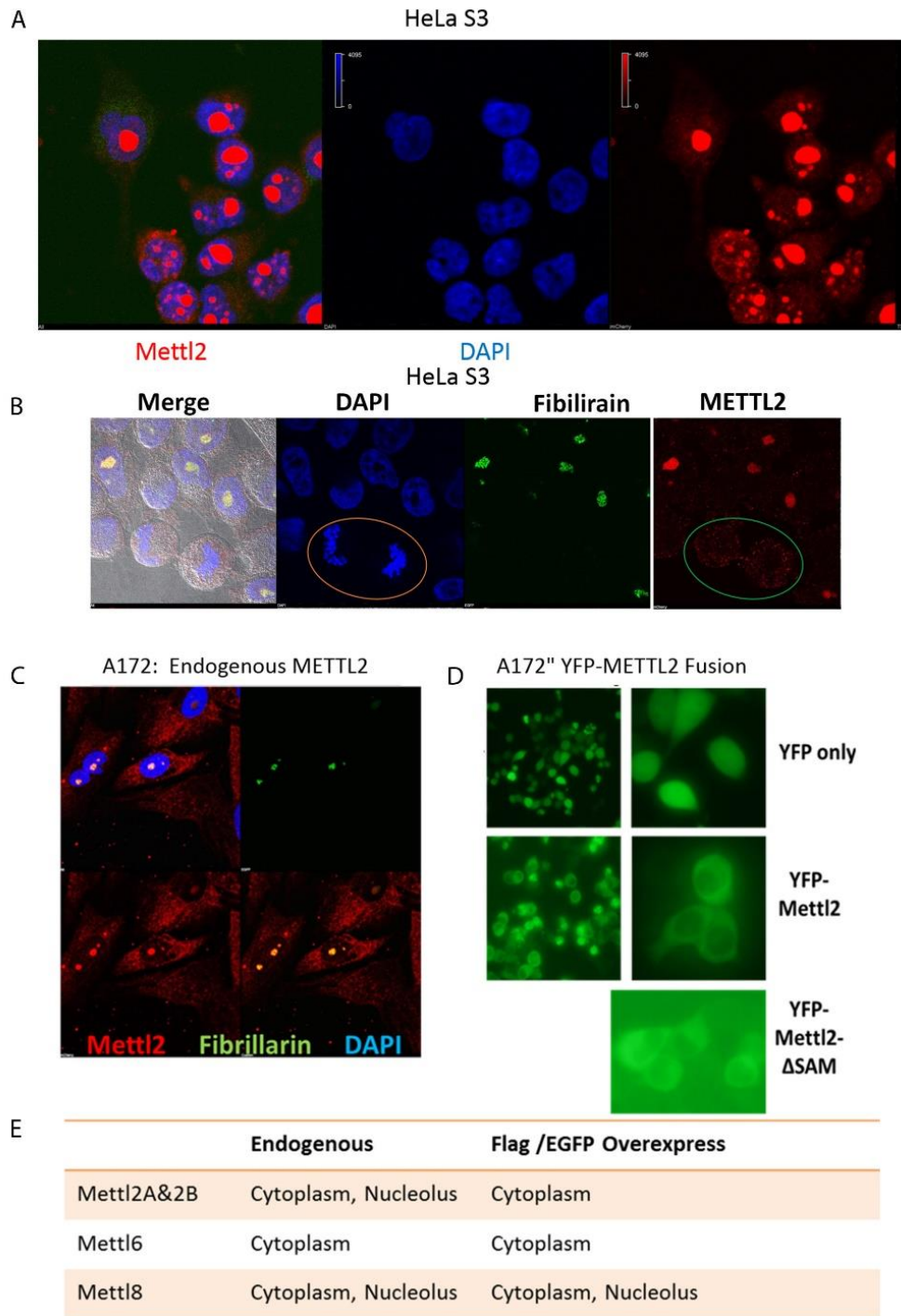


Figure 3-2 Subcellular localizations of METTL2, 6, 8

(A), (B) Endogenous METTL2 localizations in HeLa S3 cell line; cells during mitosis are cycled in (B).

(C) Endogenous METTL2 localizations in A172 human brain glioblastoma cell line.

(D) Ectopic overexpressed METTL2-YFP fusion proteins in HeLa S3 cell line.

(E) Summaries for METTL2, 6, 8 proteins' subcellular localizations.

Interestingly, METTL8 was found to be localized in the nucleus too, which indicates a possible interplay with ATM signalling pathway in the nucleus. Both endogenous and ectopic expressed METTL6 was found only in cytoplasm, supporting its interaction with Serine-tRNA synthase (SARS), though human seryl-tRNA synthetase was found to be in the nucleus to perform its non-canonical functions towards VEGFA (172, 173).

### **3.3 Generation of Mettl2A&2B double KO in HEK293T cell line**

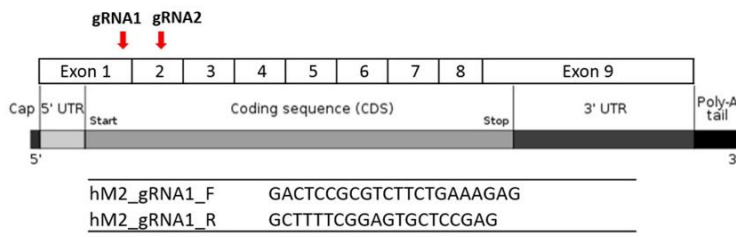
tRNA and rRNA have relatively longer half-life compared with other RNA species as stated in introduction. Moreover, shRNA or siRNA based knockdown usually cannot totally abolish mRNA targets. To ensure complete abolishment of potential modifications on tRNAs, we established CRISPR/Cas9 mediated knockout cell lines following the instruction of ZiFit to design targeting guiding RNA (gRNA) on both human Mettl2A and Mettl2B coding regions.

Three colonies using one gRNA with different deletion mutations which could cause pre-mature stop of translation were obtained and mutations were confirmed by TA cloning and Sanger sequencing as shown in Figure 3.3 (B) and western blot confirmed the knockout the METTL2A, 2B full-length protein at 43KDa Figure 3.3C using two monoclonal METTL2A&2B antibodies purchased from Abmart, designated as Abmart 1 and Abmart2, although extra bands were shown in WB on the right panel by using in house purified

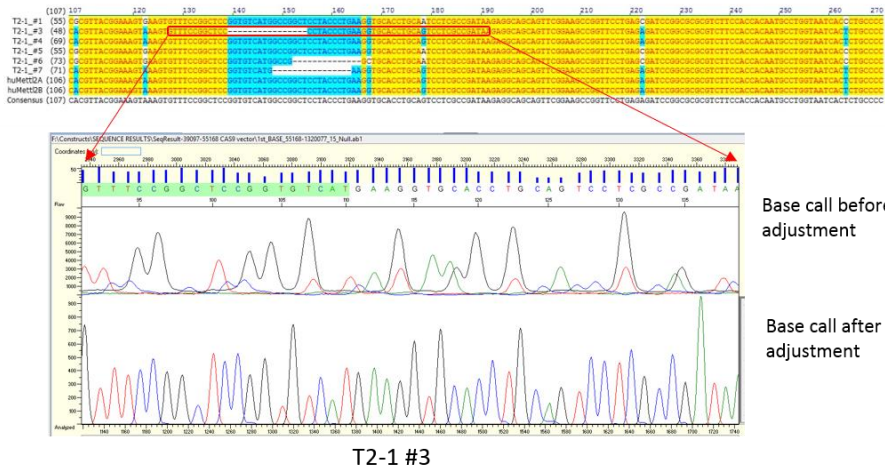
METTL2 antibody, probably due to non-specific binding. Real time PCR revealed over 20 fold reduction of METTL2

Off-target effect of the gRNA was evaluated according to bioinformatics tool developed by Zhang Feng group. There was no mutation on the top off-target candidates by sequencing. Nevertheless, whole genome sequencing or using double nickase Cas9 are needed to fully confirm no off-target effects in those clones. Those KO cells serves a better source of RNA than Knockdown cells as revealed by primer extension assay in Figure 3-6 and Figure 3-7, are being used in ribosome profiling.

**A Target both Mettl2A & 2B**



**B**



Clonal line#	Gene Zygosity	Mettl2A & 2B protein level	Mettl2A Allele 1			Mettl2A Allele 2			Mettl2B Allele 1			Mettl2B Allele 2		
			Insertion	Deletion	Mutation	Insertion	Deletion	Mutation	Insertion	Deletion	Mutation	Insertion	Deletion	Mutation
T2-1 #1	Homozygous	WT	-	-	-	-	-	-	-	-	-	-	-	-
T2-1 #3	Heterozygous	KO	-	16nt	-	-	16nt	-	-	-	3nt	-	-	3nt
T2-1 #6	Heterozygous	Partial KO	-	14 nt	-	-	14 nt	-	-	-	-	-	-	-
T2-1 #7	Heterozygous	KO	-	16 nt	-	-	16nt	-	-	16 nt	-	-	16 nt	-

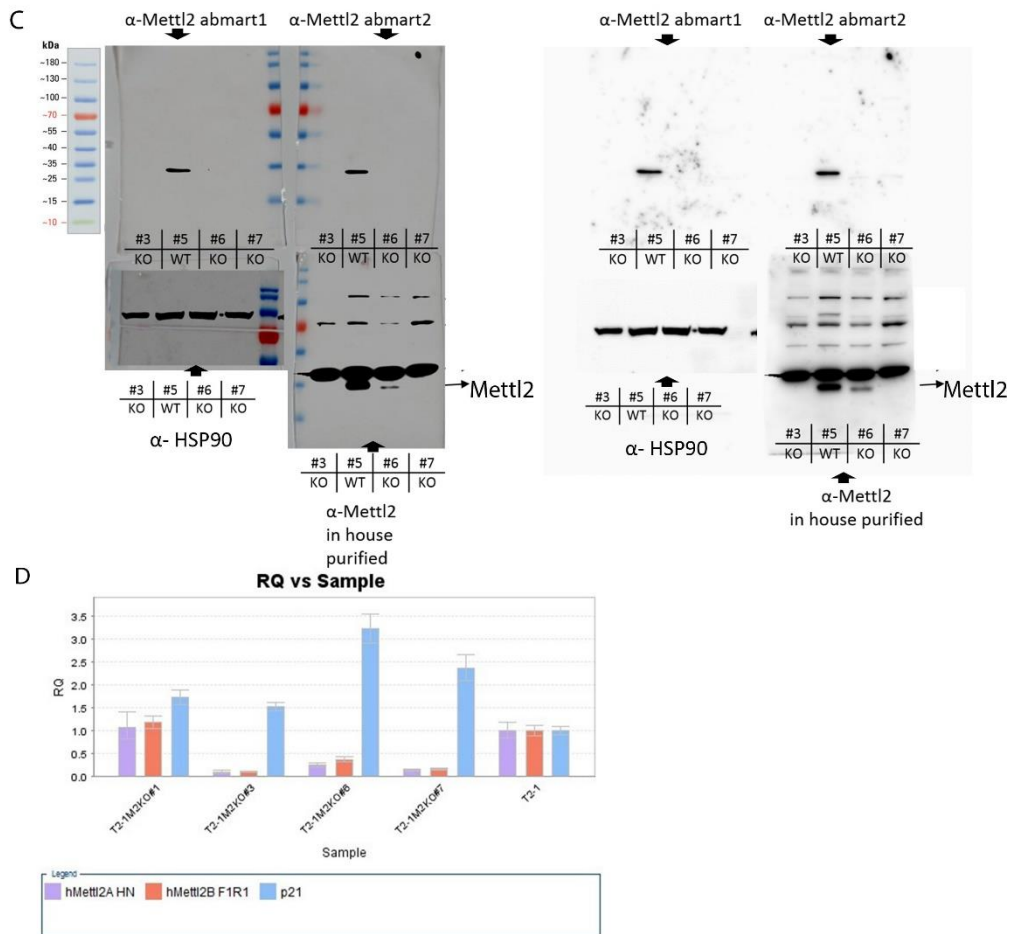


Figure 3-3 CRISPR/Cas9 mediated Mettl2A & 2B gene knockout in 293T cells

(A) Diagram of guide RNA targets on both human Mettl2A&2B mRNA and the sequence of gRNA1, which successfully generate three different M2 KO clones, is listed below.

(B) Genomic DNA Sequencing results of individual wildtype or knockout clones were aligned to the respective target region of gRNA1 used. Mutations or Deletions of various lengths were indicated in blue compared with human Mettl2A and Mettl2B ref\_seq sequence.

(C) Equal amount of lysates from different Mettl2 knockout clones and controls were resolved on SDS PAGE gel and probed with Hsp90 and three human Mettl2A&2B antibodies, designated as a-Mettl2 abmart1, a-Mettl2 abmart2, a-Mettl2 in house purified. (notice of some potential non-specific bands using in house purified Mettl2 antibody)

(D) Relative Quantitative Real time PCR analysis of WT HEK293T and a number of 293T Mettl2 KO clones (GAPDH mRNA as control).

### **3.4 Recombinant METTL2 protein is prone to aggregate under normal purification conditions**

To obtain sufficient amount of purified recombinant proteins for *in vitro* MTase assay and crystals growing for METTL2, 6, 8, large scale prokaryotic expressions and purifications of METTL2, 6, 8 proteins, together with certain mutants were performed. In all Class I MTase sequences, as reviewed in 1.4.1, GxGxG motif is a signature for co-factor SAM binding. For generation of methyl-transfer inactive METTL2B mutants, its GCGVG sequence, together with another extra 3 amino acid residues at 5 Prime and 1 amino acid residue at 3 Prime(169, 170), were deleted (27 nt deletion in total) and designated as M2-ΔSAM. For a more mild disruption of the protein structure, a GCGVGN to ACAVAN (designated M2-G3A, three Glycine to Alanine) mutants were obtained. METTL6 and METTL8 ΔSAM and G3A mutants were generated in a similar manner. Taken together, METTL2B and its abovementioned mutants, were subcloned in to either pGEX series or PET15b/21a series of prokaryotic expression vectors. Small scale tests were performed to obtain the optimal concentration of IPTG induction and temperature for *E. Coli* growth and test the efficiency of GST tag cleavage by Precision Protease (GE Healthcare). For most of the cases, 16 °C and 100 μM IPTG were able to induce protein expression quite well and GST tag were efficiently cut by the protease.

Figure 3.4 (A), right panels. Both GST-tagged and His-tagged METTL2 were expressed well in *E. Coli* BL21 strain in large scale, as sufficient amount of proteins were obtained after Nickle affinity columns Figure 3.4(A). However, there would be a substantial amount of proteins eluted within the void volume

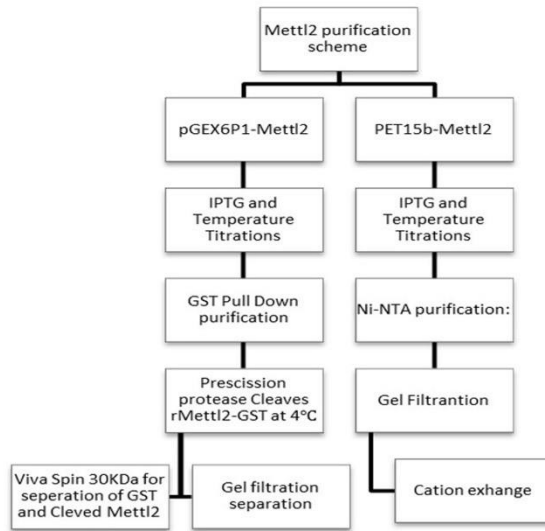
of the column as proteins went through size exclusion columns such as Superdex, Figure 3.4 B, indicating protein aggregations or precipitations.

METTL2, 8 has been predicted to contain a large portion of disordered region at N- terminal by XtalPred (174, 175), which may partially explain the occurrence of protein aggregations, and indicates it is hard to get crystals of diffraction quality. Indeed, a panel of crystal growth conditions (400 ) towards different truncated proteins were tested and failed to obtain high quality crystals for X-ray diffractions. (Screen conditions and photos of some crystals are available upon request)

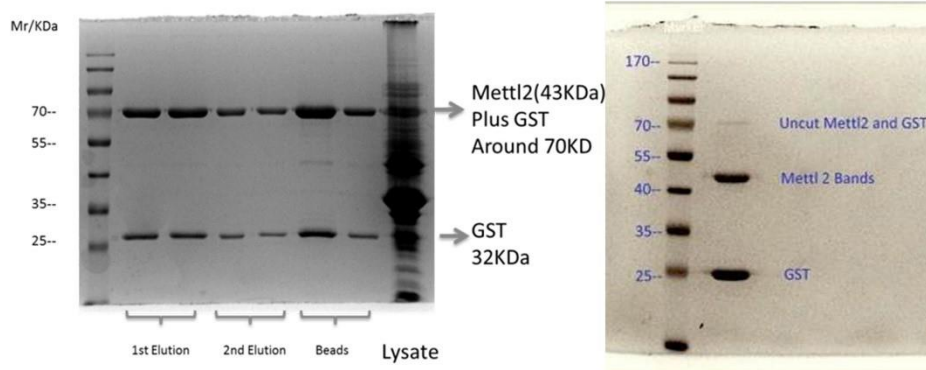
Besides the 43.5 kDa bands in the coomassie blue gels in later fractions in gel filtration step, which correspond to full length METTL2, there are higher molecular weight proteins which are eluted within the void volume of the column, indicating protein aggregates as discussed above. Curiously, those METTL2 protein aggregates could not be broken by reducing conditions with beta-mercaptoethanol ( $\beta$ -ME) in sample buffer with heating. The proteins migrates as approximately 66 kDa proteins in SDS-PAGE, indicating a dimer is formed through certain covalent bonds other than disulfide bond, which could be broken upon heating in reducing environment. Interestingly, this observation is also observed in ABP140, the yeast homologue of METTL2. In fact, the molecular weight of ABP140 was around 71.5 kDa based on amino acid sequence, but it is actually detected at around 140 kDa by SDS-PAGE and that's where the name ABP140 came from. Likewise, the recombinant ABP140 was detected at about 150 kDa due to the 19-kDa C-terminal tag (115). The property of METTL2 or ABP140 to form dimers may give hints to its substrates binding and its structure conformation.



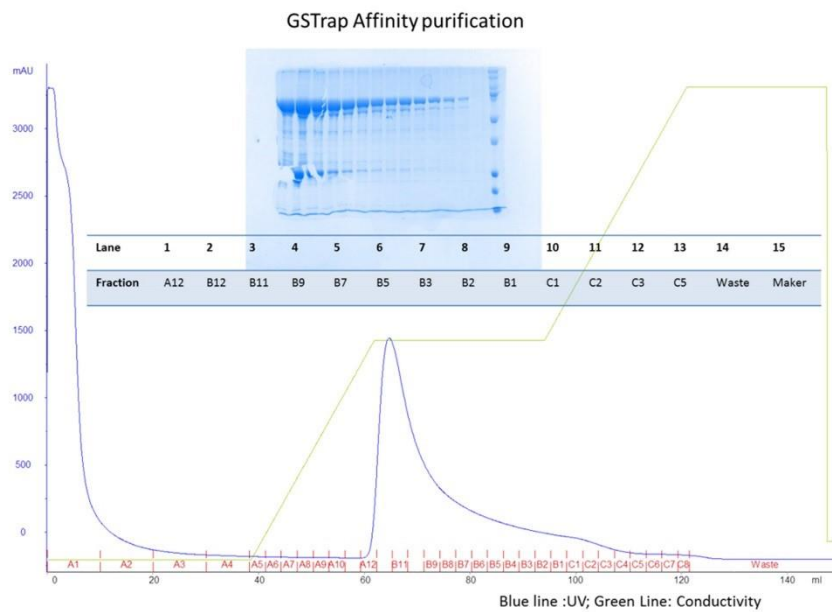
A



B

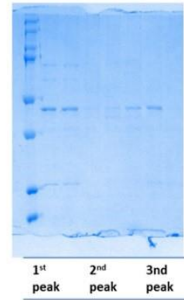
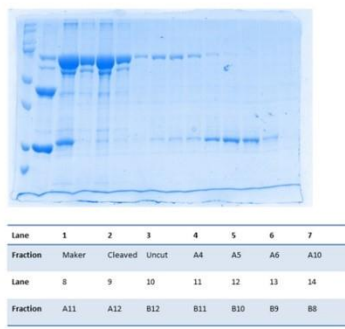


C

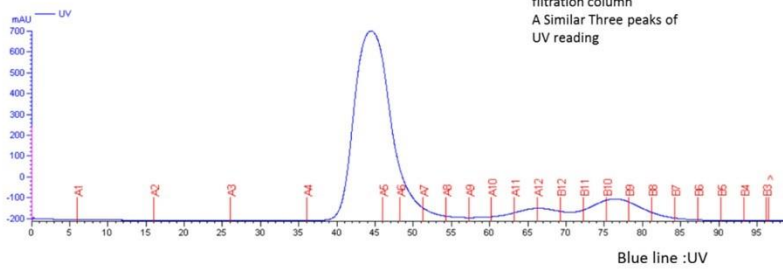


C. Contin.

Size Exclusion purification for GST-Mettl2 fusion protein

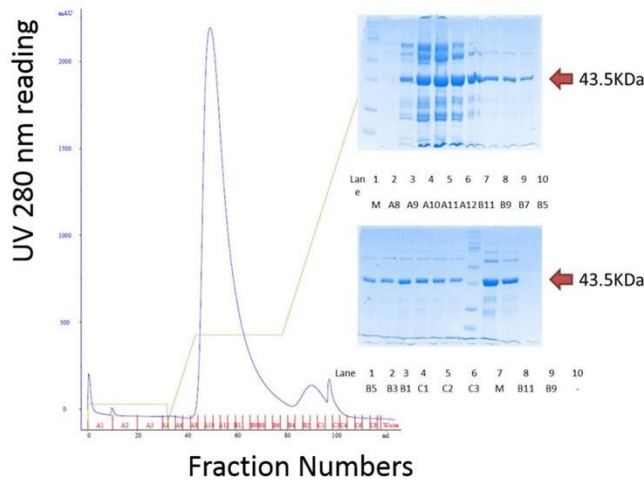


GST-Mettl2 cleaved with Precision protease running through gel filtration column  
A Similar Three peaks of UV reading

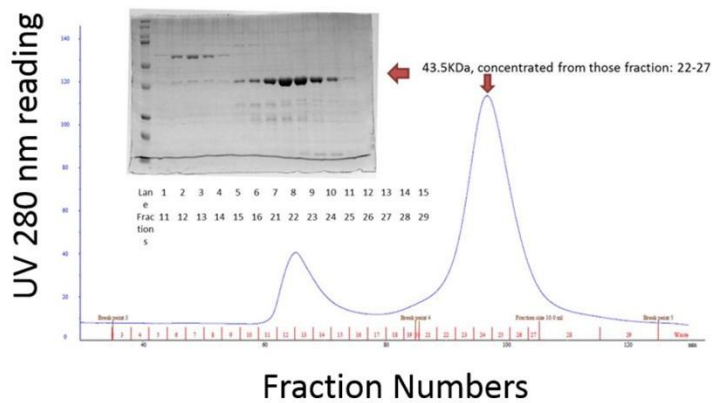


D

Affinity Nickel



Gel filtration



### Figure 3-4 Purification of METTL2 by Affinity and Size Exclusion Chromatography

(A) Schematic diagram for the FPLC purification of Mettl2 protein from *E. Coli*.

(B) Small scale test for the IPTG induction and Precision protease® cutting efficiency of GST-METTL2 fusion protein (right two panels).

(C) Large scale purification of GST tagged mouse METTL2, GST affinity purification at the upper panel and size exclusion purification at the bottom panel.

(D) Large scale purification of human His tagged METTL2B, Nickle affinity purification at the upper panel and size exclusion purification at the bottom panel.

### **3.5 tRNA<sup>Thr</sup> AGU and tRNA<sup>Thr</sup> UGU isoacceptors are substrates of METTL2**

Generally, all modifications on the Watson-Crick face of the nucleobase (including m<sup>1</sup>A, m<sup>3</sup>C, m<sup>3</sup>U, m<sup>2,2</sup>G, m<sup>1</sup>G, etc.) are prone to block the incorporation of a complementary nucleotide during reverse transcription by retrovirus derived reverse transcriptase, while small modifications at the Hoogsteen edge often allow unimpeded read-through (44). This difference in enzymatic behavior is exploited in the detection of modifications such as m<sup>3</sup>C (47, 115). To validate if human METTL2 could be responsible for m<sup>3</sup>C in certain tRNAs, primer extension assay were performed following published work (47, 115). In theory, if 3-methylcytosine is conserved in human during evolution, primer extension on threonine or serine tRNA species that contain m<sup>3</sup>C at position 32 will terminate at residue 33, while tRNAs lacking m<sup>3</sup>C will terminate at another downstream residue which could block reverse transcription, which is most likely m<sup>2,2</sup>G (N<sup>2</sup>, N<sup>2</sup>-dimethylguanosine) as from the yeast study(115). Schematic diagram of human tRNAs and corresponding probes used in primer extension assay are drawn in Figure 3-5.

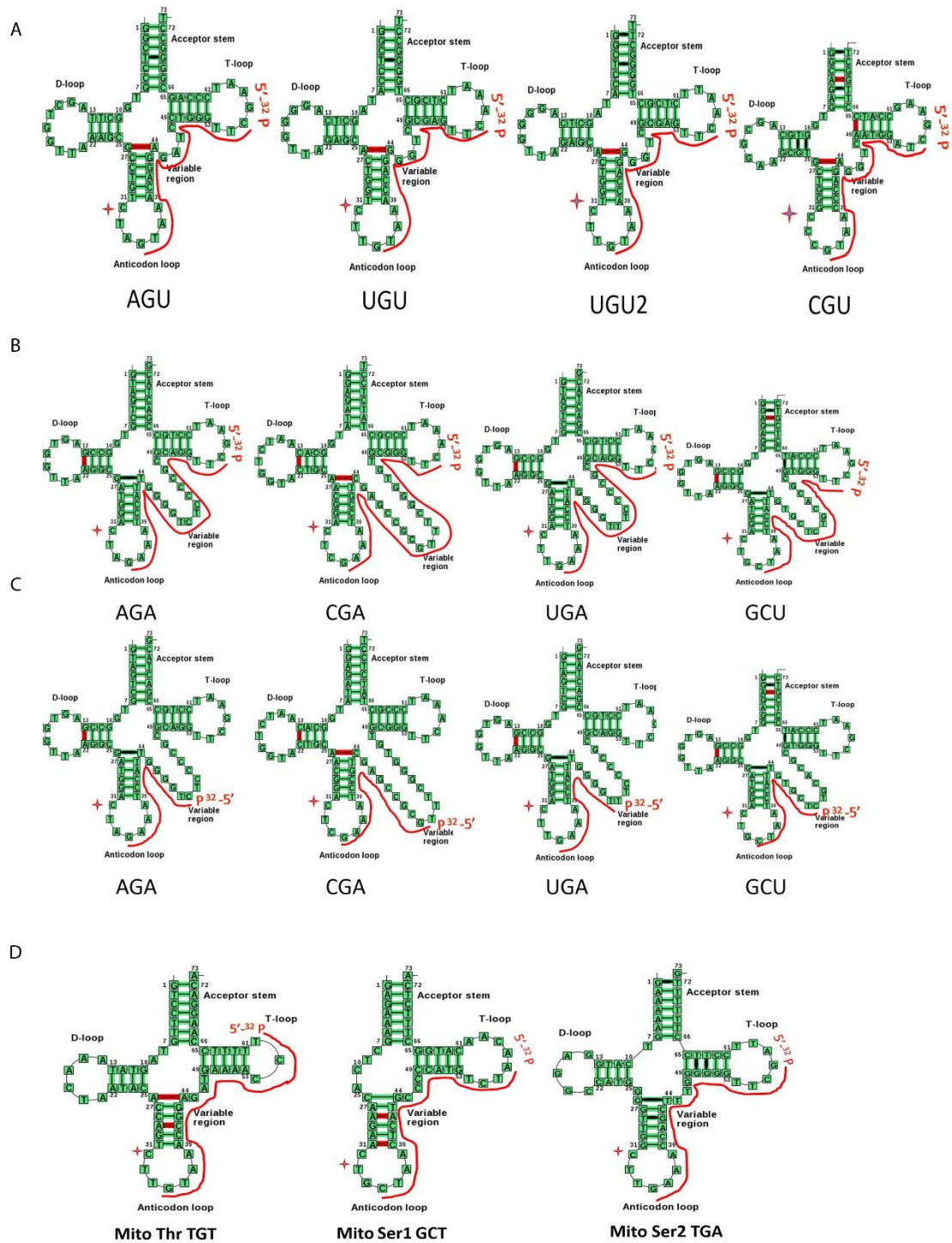


Figure 3-5 Schematic diagram of human tRNAs and corresponding probes used in primer extension assay.

All probes were labelled at 5' Prime by  $\gamma$ -<sup>32</sup>P-ATP and T4 PNK, the four point stars indicate cytosine residue at position 32 in tRNAs.

(A) Schematic of four human cytoplasmic Threonine tRNA Isoacceptors: AGU, UGU, UGU2, CGU, (Known modifications not shown, U is shown as

T), and probes used for primer extension assay for detecting potential m<sup>3</sup>C at position 32 (m<sup>3</sup>C<sub>32</sub>), designated as tThr<sup>AGU</sup>, tThr<sup>UGU</sup>, tThr<sup>CGU</sup>. They are used by reverse transcriptase to anneal to tRNA residues 3' 56 – 35 5';

(B) & (C) Schematic of cytoplasmic tRNA Serine Isoacceptors: AGA, CGA,UGA, GCU; Notice the redesign of probes in (C) to avoid long variable Arm in threonine tRNAs, since the probes in B showed a “stuttering” pattern in primer extension assay

(D) Mitochondria tRNAs Threonine UGU, Serine GCU and Serine UGA and probes. tRNA diagrams in green are adapted from tRNA database (114).

Of note, eukaryotes present not only more tRNA gene content than the other two kingdoms but also a higher variation in gene copy number among different isoacceptors and isodecoders (176). For example, there are ten Thr<sup>AGU</sup> decoder genes and six Thr<sup>UGU</sup> genes in human hg19 reference genome (171). Luckily, the high sequence similarity among tRNA isoacceptors (Figure 3-1C) made it possible to design only one or two <sup>32</sup>P labelled probes for primer extension assay. If m<sup>3</sup>C is conserved in human tRNAs during evolution, primer extension on threonine or serine tRNA species that contain m<sup>3</sup>C at position 32 will terminate at residue 33, while tRNAs lacking m<sup>3</sup>C will terminate at another downstream residue which could block reverse transcription, which is most likely m<sup>2,2</sup>G (N<sup>2</sup>, N<sup>2</sup>-dimethylguanosine) as inferred from the previous study (115). Indeed, primer extension for total RNA from wild type HEK293T cells generates a 24 nt size <sup>32</sup>P-ATP labelled band which is two nucleotides longer than the free probes, indicating a reverse transcription stopping event occurs at position 32 of both tRNA Thr<sup>AGU</sup> and tRNA Thr<sup>UGU</sup> (Figure 3-6A lane 1, lane 5 and lane 9). In contrast, primer extension using RNA from Mettl2A and 2B double KO cells generates a higher molecular weight bands, indicating loss of METTL2 would abolish the modification at position 32 (Figure 3-6A lane 2, 6, 10), which is 3-

methycytosine, as supported by the recent ARM-seq (177) and DM tRNA seq (176).

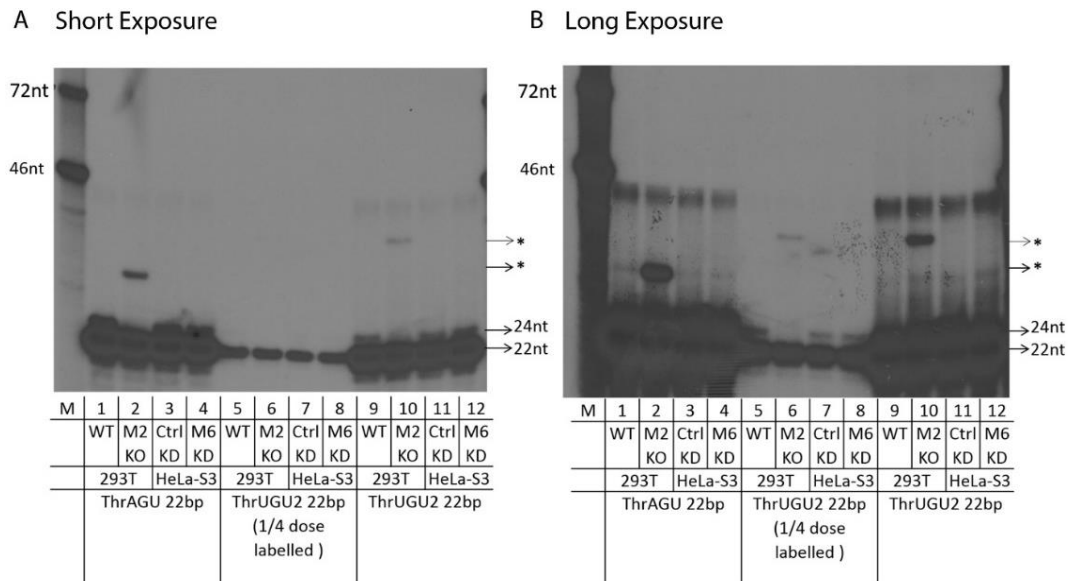


Figure 3-6 Loss of METTL2A & 2B abolished the modification at position 32 in tRNA<sup>Thr</sup>AGU and tRNA<sup>Thr</sup>UGU isoacceptors

(A). Autoradiography for various primer extension reactions resolved on 15% Urea PAGE. RNA used from different cells were indicated under lane number, for example, M2 KO under lane 2 indicated RNA used were extracted from METTL2 Knockout cells. Notice in lane 2 and 10, two higher size bands were pointed out as asterisks on the right which used RNA from METTL2 KO cells. Marker lane (M) contains ssDNA probes which is 56 nt and 72 nt in length, lane 1-4 used tThr<sup>AGU</sup> probes, lane 5 -12 used tThr<sup>UGU</sup>, lane 5-8 used 1/4 dose of  $\gamma$ -<sup>32</sup>P-ATP as compared to lane 1-4 and lane 9-12 (\* donates longer bands generated by using RNA from M2 KO cells).

(B) Longer exposure for (A), notice the same bands pattern in lane 9-12 in (A) appears in lane 5 - 8, in which less  $\gamma$ -<sup>32</sup>P-ATP is used in labelling probes.

### 3.6 Other substrates for METTL2, 6, 8 remain elusive

To check if METTL2, 6 and 8 could have other tRNA species as substrates, more probes were designed and labelled, including tRNA Thr<sup>CGU</sup>, (no functional tRNA Thr<sup>GGU</sup> isoacceptors in human) (114), and serine, arginine tRNA isoacceptors in both cytoplasm and mitochondria (). Primer extension targeting tRNA Thr<sup>CGU</sup> showed a stuttering pattern, indicates either a low

abundance of Thr<sup>CGU</sup> or poor specificity of the probes targeting Thr<sup>CGU</sup> (Figure 3-7). Other probes also failed to give a pattern as clear-cut as tRNA Thr<sup>AGU</sup> and Thr<sup>UGU</sup> probes, despite of trying various conditions of reactions, including different dNTP concentration, reaction time and incubation temperature, etc. Importantly, there is no clear difference between WT and KO, leaving other potential substrates of METTL2, 6, 8 elusive. It is not too surprising since reverse transcription profiles of modified RNAs are often prone to some kind of stuttering (doubling of the band on the gel) at the modified residue that blocks the reverse transcription (44).

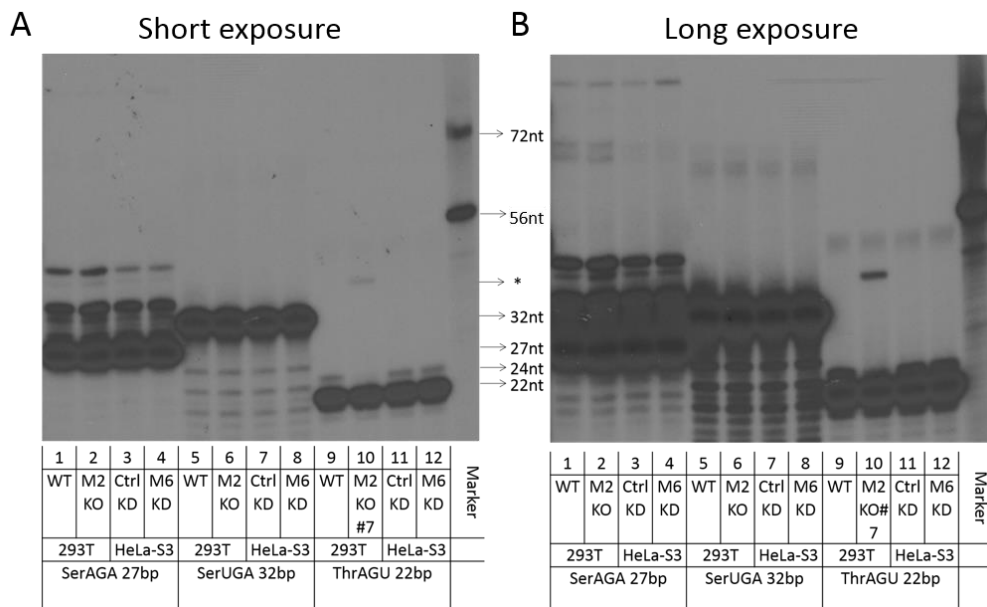


Figure 3-7 Reduction or loss of METTL2, 6 or 8 has no effects on the position 32 modification in other tRNA species tested

(A) Primer extension assay using other tRNA probes, lane 1-4 used Ser<sup>AGA</sup> probe, lane 5-8 used Ser<sup>UGA</sup> probe, lane 9 -12 used Thr<sup>AGU</sup> probe and KO clone #7 as repeat to the previous results. Marker lane contains ssDNA probes which are 46 nt and 72 nt in length.

(B) A longer exposure for film of (A), notice some larger size of fragments appears to be longer tRNA synthesized.

### 3.7 Ectopic expression of METTL2 in KO cells restore the modification which could block primer extension

Ectopic expression METTL2 Protein were performed to test if ectopic METTL2 could restore the modification *in vivo*. Indeed, the longer cDNA were generated in reactions which used RNA from METTL2 overexpressing cells. The longer cDNA were generated by both tRNA<sup>Thr</sup> AGU and tRNA<sup>Thr</sup> UGU probes, but not tRNA<sup>Thr</sup> CGU, which showed a stuttering pattern, as shown in (

Figure 3-8). LC-MS/MS analysis was also performed to check m<sup>3</sup>C level. Indeed, m<sup>3</sup>C is upregulated in cells overexpressing METTL2 (Figure 3-13C).

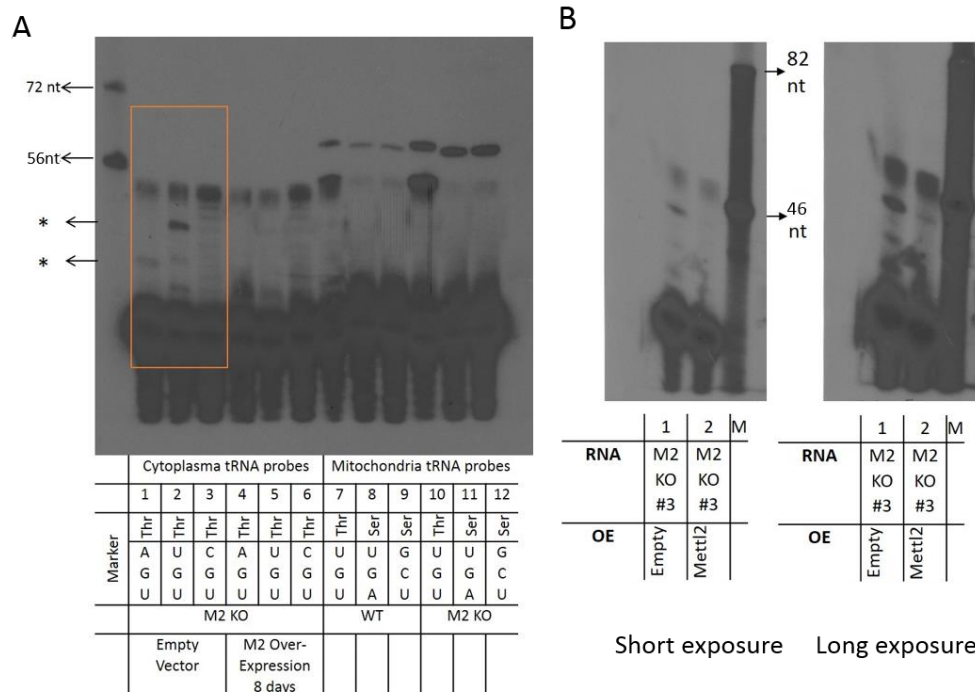


Figure 3-8 Ectopic expression METTL2 protein in KO cells restore the position 32 modification in tRNA Thr<sup>AGU</sup> and Thr<sup>UGU</sup>



(A) Primer extension assay for RNA isolated from *Mettl2* KO cells (lane 1 to 3), or KO cells overexpressing *Mettl2* mammalian expression vectors (lane 4 to 6). Lane 7 to lane 12 are reactions using mitochondria tRNA probes. The symbol \* indicates larger size bands generated by using RNA from M2 KO cells). Marker lane contains ssDNA probes which is 56 nt and 72 nt in length.

(B) One repeat for (A), in which cells overexpress ectopic METTL2 for 12 days are assayed for primer extension.

### 3.8 Recombinant METTL2 could restore the modification on tRNA

#### **Thr<sup>AGU</sup> and Thr<sup>UGU</sup>**

To test if Recombinant METTL2 could methylate tRNA<sup>Thr</sup> AGU and tRNA<sup>Thr</sup> UGU *in vitro*, Methyltransferase were carried out using FPLC purified WT METTL2 or two mutants M2- $\Delta$ SAM and M2-G3A, whose Methyltransferase activity are supposed to be lost, as described in 2.1. Indeed, as shown in (Figure 3-9 A), higher MW bands were gone in which WT METTL2 proteins were added, while G3A mutants failed to restore the modification as seen in (Figure 3-9 B), neither could the  $\Delta$ SAM mutants (now shown)

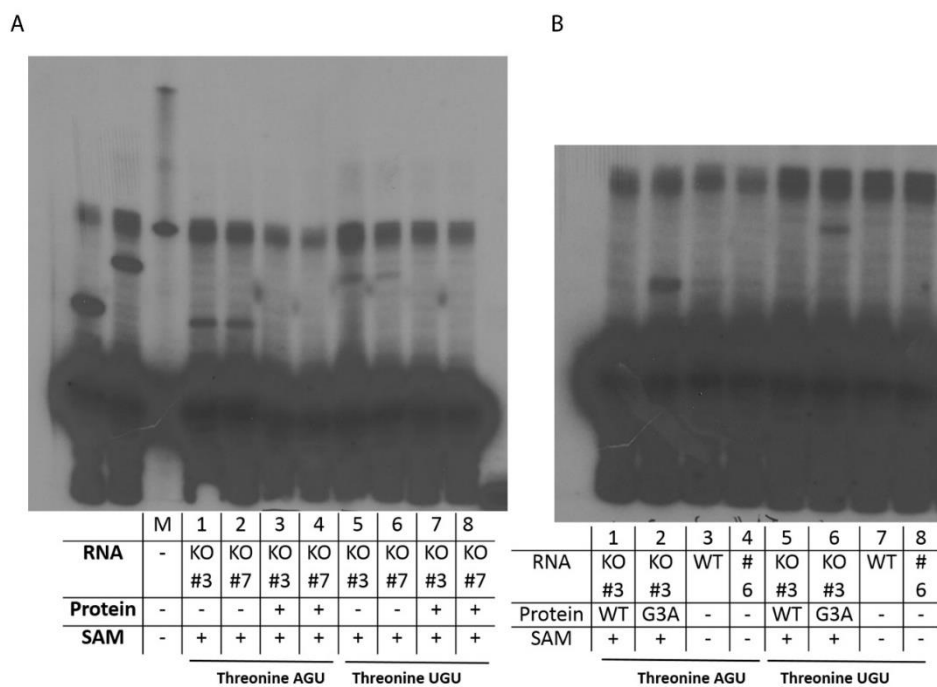


Figure 3-9 *In Vitro* reconstitution of modification on tRNA Thr<sup>AGU</sup> and Thr<sup>UGU</sup> by recombinant METTL2

(A) *In Vitro* methyltransferase assay. Protein + in lane 3,4,7,8 indicates N terminal 6×His tagged human Mettl2B protein purified from *E. coli* were added.

(B) G3A mutated human METTL2B proteins are added in lane 2 and lane 6 reactions and incubated with RNA from KO cells and SAM, in contrast with lane 1 and lane 5, in which WT huMettl2B proteins were added. #6 indicates another cell clone, which failed to show the abolishment of modification at position 32.

### 3.9 *in vitro* transcribed tRNAs are not methylated by recombinant METTL2

Unmodified tRNAs were *in vitro* transcribed using T7 RNA polymerase and were served as a potential substrate for METTL2. No m<sup>3</sup>C<sub>32</sub> formation was observed in this transcript as shown in (Figure 3-10 ), again, tested by primer extension. Some modifications in tRNAs might act as positive determinant for m<sup>3</sup>C formation at position 32. This result suggested that tertiary structure of native mature tRNAs, instead of unmodified tRNAs, are required for efficient

m<sup>3</sup>C32 formation in tRNA<sup>Thr</sup> AGU and tRNA<sup>Thr</sup> UGU and both *in vitro* transcribed tRNAs are not good substrates for METTL2 protein. (unlabelled lanes on the right side of the films are control lanes from previous reactions as a marker)

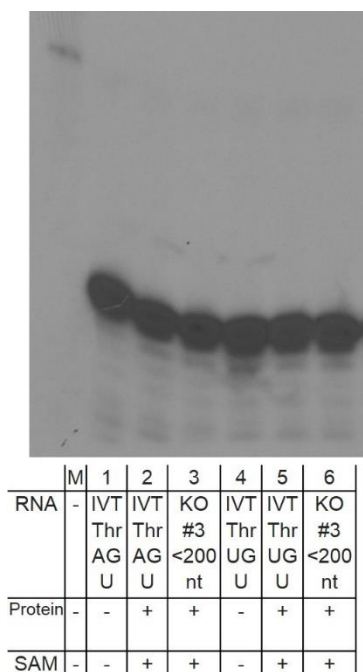


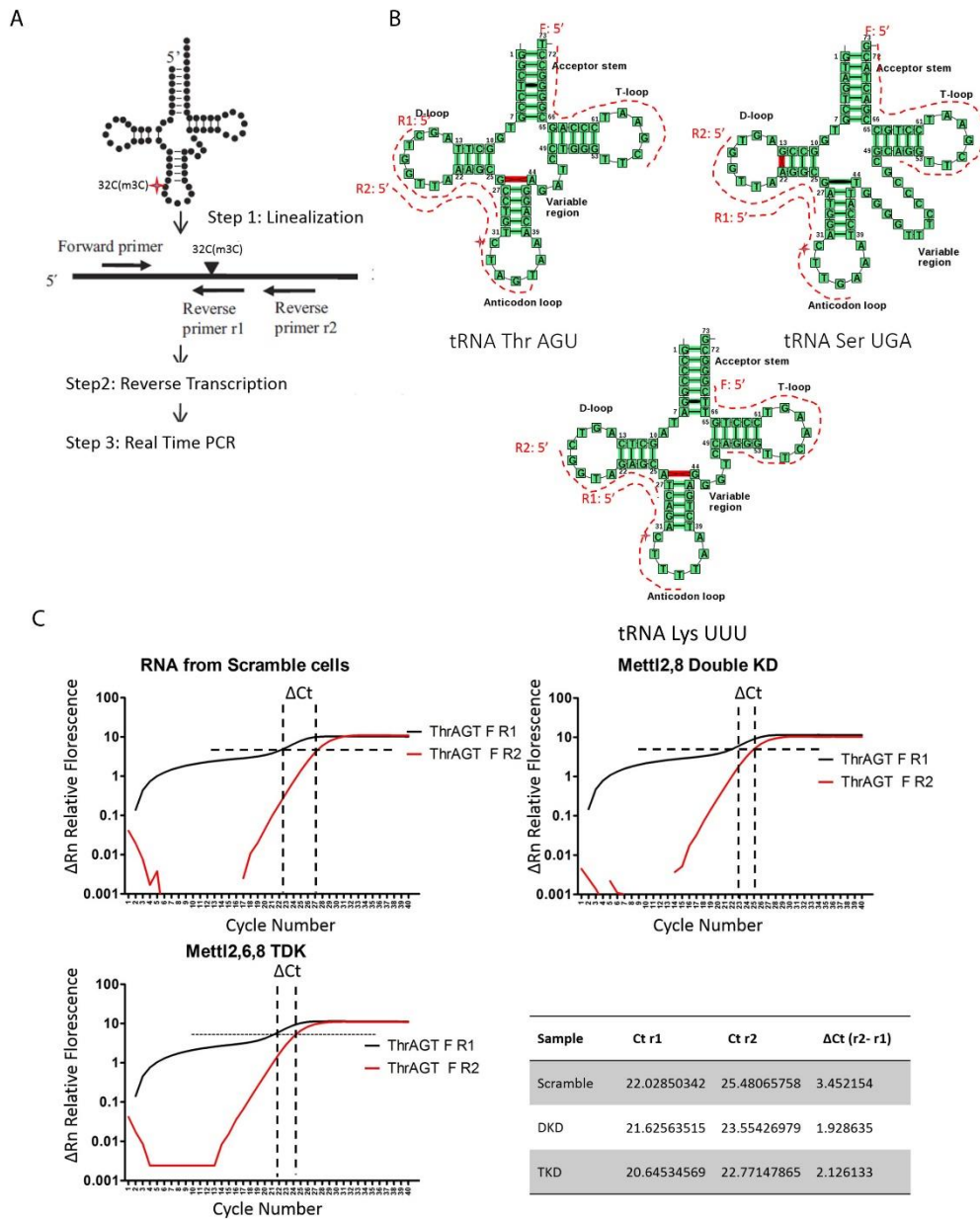
Figure 3-10 METTL2B fail to restore position 32 modification on *in vitro* transcribed tRNA Thr<sup>AGU</sup> and Thr<sup>UGU</sup>.

Primer extension using *in vitro* transcribed tRNA Thr<sup>AGU</sup> and Thr<sup>UGU</sup>, as described section 2.10. Notice lane 3, 6 are repeats previous results

### 3.10 A quantitative PCR based assay for m<sup>3</sup>C quantification

Due to the tedious and low throughput nature of primer extension assay, a quantitative PCR (qPCR) based assay were developed for Measurement of 3-methylcytosine modification in tRNA<sup>Thr</sup> AGU and tRNA<sup>Thr</sup> UGU (qPCR-m<sup>3</sup>C), inspired and modified from measuring 2-methylthio modification(162). tRNA<sup>Thr</sup> AGU, tRNA<sup>Thr</sup> UGU and some other tRNAs, such as tRNA<sup>Ser</sup>UGA was reverse transcribed with 2 unique primers, with reverse primer r1

designed to target a region which includes the Cytosine at 32, and reverse primer r2 was designed to anneal to the region downstream of C32. Subsequent qPCR was performed to detect the corresponding cDNAs and determined the Ct Value for both r1 and r2, the gap between Ct r1 and Ct r2 (designated  $\Delta\text{Ct r2-r1}$ ) showed a difference between RNA from WT and RNA from METTL2 knockdown, or METTL2, 6, 8 triple Knockdown, as show in Figure 3-11B. Notably, the  $\Delta\text{Ct}$  value is not increased in Triple Knockdown, indicating METTL6 plays a minor role in threonine tRNA modification (Figure 3-11D). However, a similar design to tRNA<sup>ser</sup>UGA failed to get reasonable Ct r1 and Ct r2, since the Ct r1 is supposed to be smaller than Ct r2, as shown in (C). These results may reflect a similar situation to stuttering pattern in traditional primer extension assay, probably due to the long variable arm in cytoplasm serine tRNAs. qPCR-m3C were actually developed before the primer extension assay in the above figures, and it offered a potential faster and semi-quantitative way for assessing 3-methylcytosine modification levels in certain tRNAs. Further titrations needed to be performed if a quantitative measurements are desired.



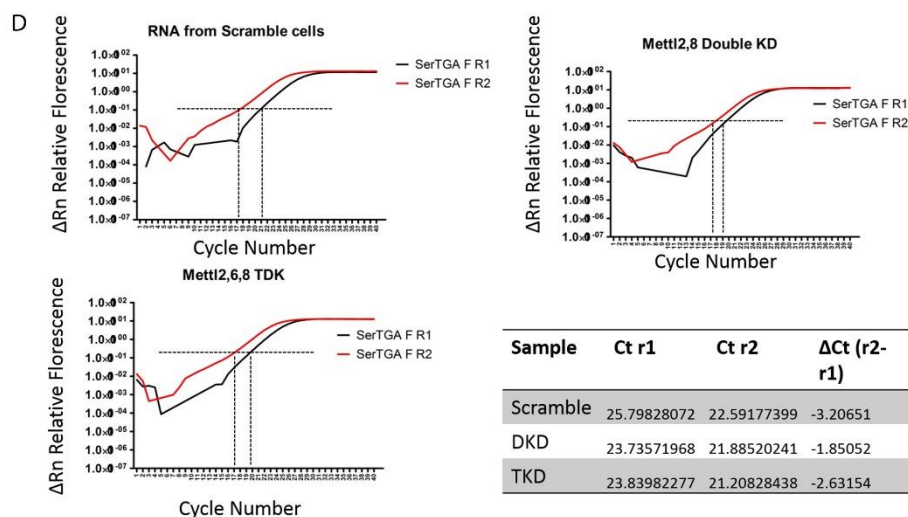


Figure 3-11 Quantitative PCR based assay for quantification of  $m^3C$  (qPCR- $m^3C$ )

(A) Cartoon showing the work flow of the assay

(B) Schematic diagram of human tRNAs and corresponding probes used in primer Extension

(C) and (D)  $m^3C$  Real time PCR results for tRNAThr<sup>AGU</sup> and Ser<sup>UGA</sup>, respectively

### 3.11 LC-MS/MS analysis of $m^3C$ contents in various RNA species

To support the above finding by primer extension assay and study the contribution of  $m^3C$  by METTL2 family proteins, quantitative analysis of ribonucleotide modifications is performed by size exclusion HPLC and triple quadrupole mass spectrometry, following the methods of our group(178, 179).

tRNA were purified from total RNA preparations extracted from transformed HEK293T and HeLa cell lines by HPLC (Figure 3-12A). tRNA were isolated from each sample and quantified. Less than 2  $\mu$ g was obtained from samples T2-1#3b and T2-1#6. Nonetheless, all samples were processed and analysed by LC-MS for ribonucleoside composition. Except for T2-1#6, isolated tRNA proved to be over 90% pure by Bioanalyzer fluorescent electrophoresis, thus,

we expect little contamination from other cellular RNAs. LC-MS/MS analysis for 7 ribonucleosides (adenosine (A), guanosine (G), uridine (U), cytidine (C), m<sup>3</sup>C, m<sup>5</sup>C and 4-methylcytidine (m<sup>4</sup>C) ) were conducted. Each ribonucleoside is positively identified by aligning commercial standards by their retention times, accurate masses and collision induced disassociation fragmentation patterns.

Instrument sensitivity and dynamic range was determined through external calibration curves for C, m<sup>3</sup>C and m<sup>5</sup>C. Injection to injection variance across sample runs was determined to be 18.9% though CV of spiked <sup>15</sup>N<sub>5</sub>dA internal standards. Relative quantification for the m<sup>3</sup>C content in tRNA was determined by the normalized ratios of m<sup>3</sup>C: total AUCG or m<sup>3</sup>C: m<sup>5</sup>C.

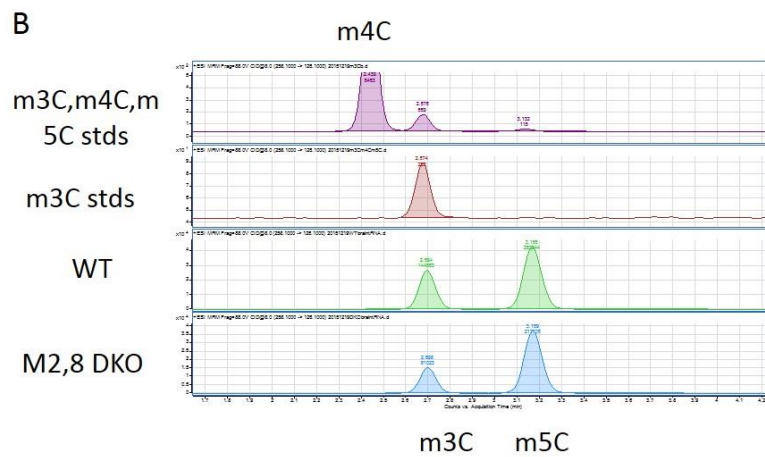
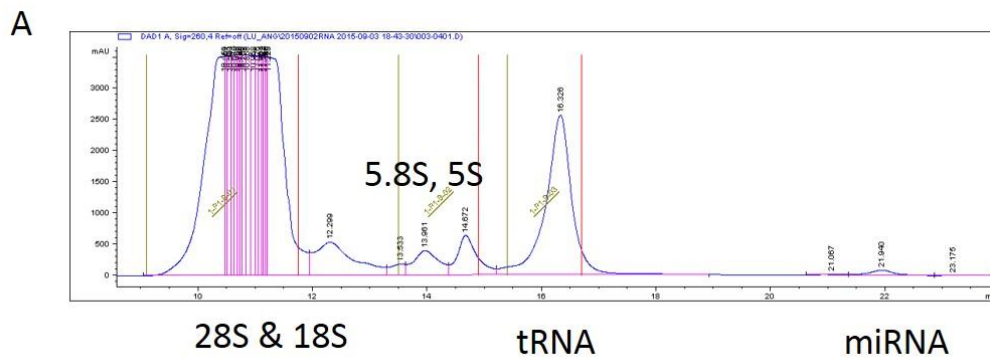
In the first series of experiments, low levels of m<sup>3</sup>C (82-297 fmol/μg tRNA; limit of quantitation 5.62 fmol) was detected for in all samples, save for sample T2-#3+DKD, wherein no m<sup>3</sup>C signal was detected. In all, relative to T2-1#1, T2-1#1b, T2-1#1c controls, A modest 22% decrease in m<sup>3</sup>C levels was observed in T2-1 #3 and T2-1 #3. A further 5% decrease was detected in T2-1#3+772. A complete loss of m<sup>3</sup>C signal in T2-#3+DKD were observed. In HeLa cells, relative to 830, which is a scramble shRNA control, a 26% reduction in m<sup>3</sup>C signal was observed in 772, which is Mettl6 knockdown and a further 8% loss was detected in triple knockdown. However, it was found one enzyme phosphodiesterase II was added at a much lower concentration compared with the previous protocol of our group (179), so a modified recipe was adapted.

After several rounds of experiments, METTL2A and 2B are found to contribute to about 40 % ( $\pm 5\%$ ) of the  $m^3C$  level in tRNAs (Figure 3-13A), while knocking down METTL6 contributes to approximately 12% ( $\pm 5\%$ ). However, we haven't find statistically significant contribution towards  $m^3C$  formation from METTL8 (Figure 3-13B). To check those  $m^3C$  contributions in mice, knockout mice for *Mettl2*, 6, 8 genes were created by Cas9/CRISPR and the sequence of KO loci was confirmed by genotyping and mRNA levels were checked by real-time PCR. Similar to human *Mettl2A* and 2B KO HEK293T cell line, *Mettl2* KO showed around 40% reduction of  $m^3C$  contents in tRNA/total RNA compared with WT, while *Mettl6* KO showed around 20%  $m^3C$  reduction (Figure 3-13D, E).  $m^3C$  contents remains similar level between WT and *Mettl8* KO, *Mettl2* KO and *Mettl2*, 8 Double KO, *Mettl6* KO and *Mettl6*, 8 Double KO, implying no contribution/little contribution of  $m^3C$  modification from mouse *Mettl8* (Figure 3-14D, E). Importantly, mRNA level of *Mettl2*, 6, 8 were checked in all kinds of KO and WT, and it indicates no upregulation of any *Mettls* in another one's KO.

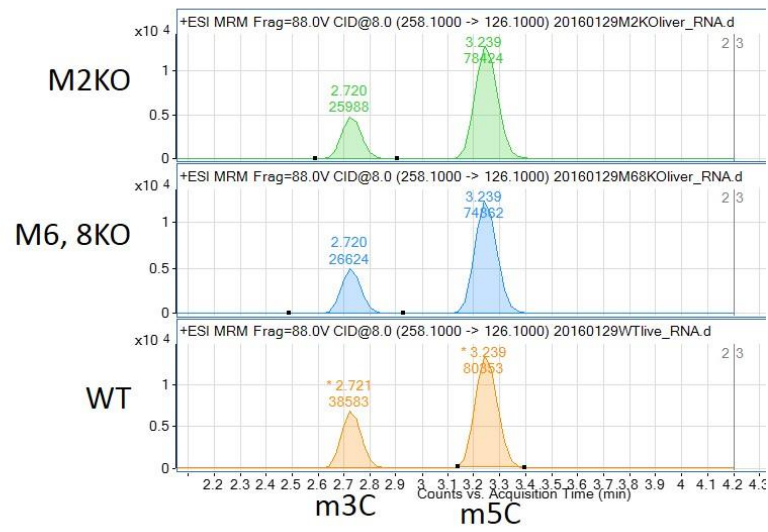
To check if  $m^3C$  is located within other RNA species, the fractions of 28S and 18S rRNA, the two largest rRNA components were collected (Figure 3-12A), concentrated, digested and analyzed. We observe there is  $m^3C$  signals, albeit much less abundant than in tRNA, as compared either to total AUCG area under curve, or the adjacent  $m^5C$  peak's area under curve (Figure 3-12C). We calculated the  $m^3C$ / total AUCG and  $m^3C/m^5C$  ratios across 20 samples, containing either WT or certain KO, and 9 samples are tRNA and 11 samples are total RNA (Figure 3-12E). 28S and 18S rRNA contributes a high percentage ( $>80\%$  by UV signal Figure 3-12A) of total RNA, and 28S and



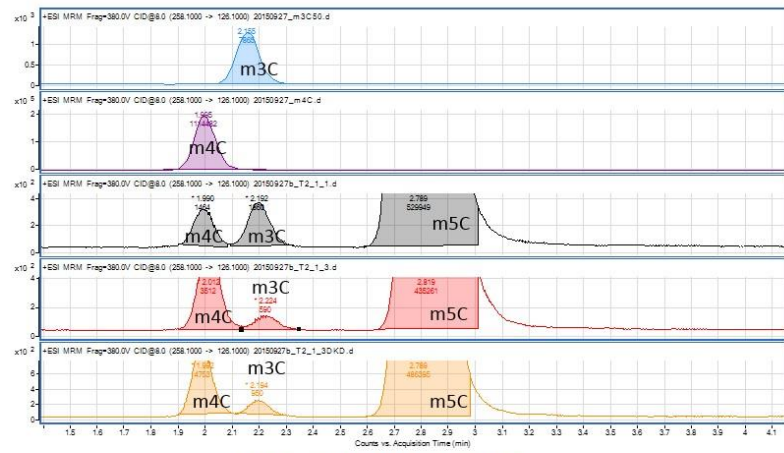
18S rRNA contains high levels of  $m^3C$ . If there's no  $m^3C$  located in other RNA species besides tRNA, we would expect a much lower  $m^3C/m^5C$  ratio in those 11 total RNA samples. It implies  $m^3C$  could reside within other RNA species such as mRNA, 5S rRNA, 5.8S rRNA, etc. The  $m^3C$  signal seen in 28S and 18S could come from long mRNAs (co-purified with 18S or 28S rRNA, or tRNA bound to rRNA).



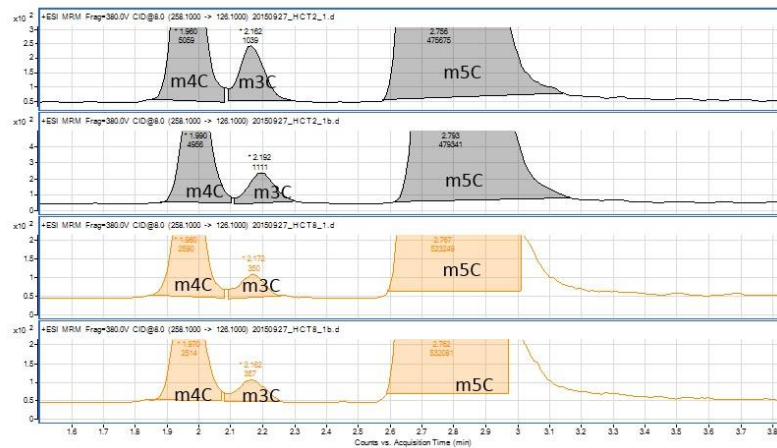
C



D



rRNA 293T WT Vs M2KO



rRNA HCT116 WT Vs M8KO

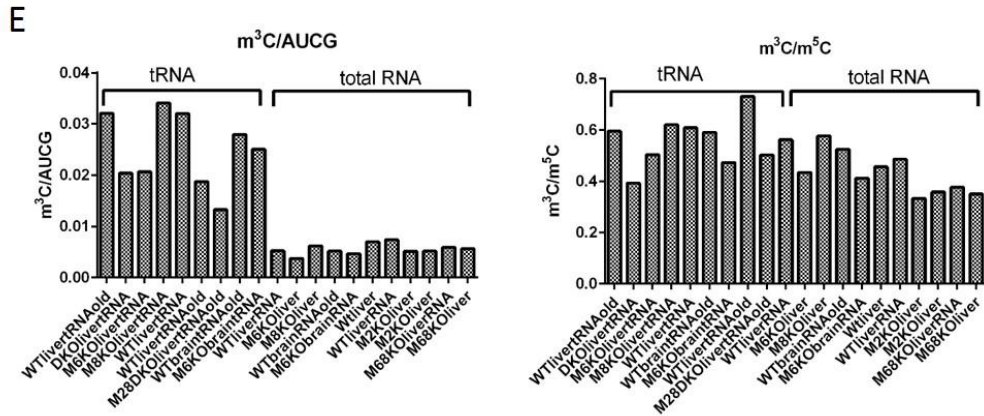


Figure 3-12. Size exclusion HPLC and typical chromatography of LC-MS/MS for m<sup>3</sup>C

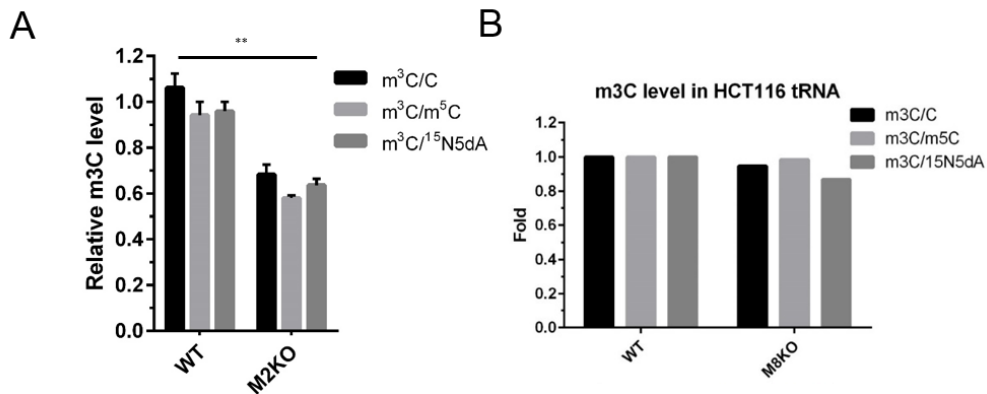
(A) Typical size exclusion HPLC chromatography, peaks for various RNA species are labelled.

(B) Typical LC-MS/MS for m<sup>3</sup>C, m<sup>4</sup>C and m<sup>5</sup>C (258.1-126.1 m/z) chemical standards and two tRNA samples analyzed.

(C) Typical LC-MS/MS for 258.1-126.1 m/z chemical standards and three total RNA samples analyzed.

(D) Typical LC-MS/MS for 258.1-126.1 m/z in 28S and 18S rRNA from WT and METTL2 or 8 KO.

(E) m<sup>3</sup>C versus total AUCG and m<sup>3</sup>C versus m<sup>5</sup>C ratios across 20 samples, 9 of which are tRNA samples and 11 of which are total RNA.



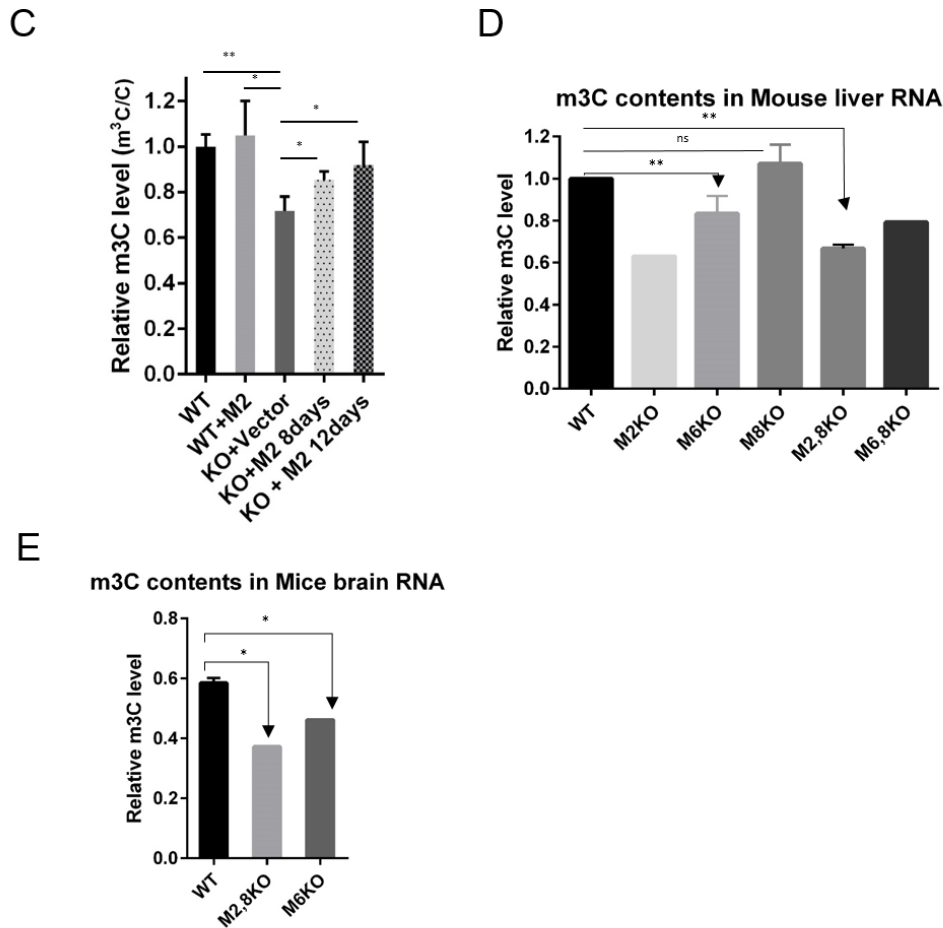


Figure 3-13. LC-MS/MS for quantitative m<sup>3</sup>C analysis in various RNA species

(A) LC-MS/MS measurements of m<sup>3</sup>C RNA samples used in the assay, by normalizing against total C, m<sup>5</sup>C or <sup>15</sup>NdA

(B) m<sup>3</sup>C contents in tRNAs from HCT116 WT and Mettl8 KO cell line.

(C) m<sup>3</sup>C level in METTL2 overexpressed cells

(D) m<sup>3</sup>C contents in mice liver tissue.

(E) m<sup>3</sup>C contents in mice whole brain RNA

Error bars are for Biological triplicate, Student t-test; \* p < 0.05; \*\* p < 0.01

### 3.12 Identification of RNA species binding to METTL2, 8 by PAR-CLIP and validation by RIP, RT-PCR

Photoactivatable Ribonucleoside Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) is a used for discover RNA binding proteins (RBPs)'binding sites on various RNA species. It uses photoreactive

ribonucleoside analogs, including 4-thiouridine (4-SU) and 6-thioguanosine (6-SG) as cell culture supplements. 365 nm UV light induces efficient crosslinking of RNA binding protein with photoreactive nucleoside-labelled RNA species. After immunoprecipitation of the RBP, cross-linked RNA was isolated, reversed transcribed into cDNAs, and deep sequenced using second generation sequencing. Flag-METTL8 is ectopically overexpressed in HeLa S3 cells and PAR-CLIP is performed at University of Chicago using Flag antibody. As shown in (A), 1095 peaks were identified for METTL8, and binding motif were calculated. However, the p-Value for the peaks are not as significant as a typical RBP, such as YTH domain family members, which are m6A binding proteins(180). Nevertheless, Gene Ontology Enrichment is performed to show transcripts which encodes nucleus structures, RNA binding proteins, etc. as top enriched clusters by using DAVID(181) , as shown in Figure 3-14 (B).

Table 3-1 mRNA transcripts identified as binding partners of METTL8 by PAR-CLIP

Gene ID	Symbol	Description	binding sites
90632	LINC00473	long intergenic non-protein coding RNA 473	4
100302692	FTX	FTX transcript, XIST regulator (non-protein coding)	4
164045	HFM1	HFM1, ATP-dependent DNA helicase homolog (S. cerevisiae)	3
29123	ANKRD11	ankyrin repeat domain 11	3
3376	IARS	isoleucyl-tRNA synthetase	3
7919	DDX39B	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39B	3
4643	MYO1E	myosin IE	3
29123	ANKRD11	ankyrin repeat domain 11	3
3376	IARS	isoleucyl-tRNA synthetase	3
7919	DDX39B	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39B	3

90632	LINC00473	long intergenic non-protein coding RNA 473	3
554226	ANKRD30 BL	ankyrin repeat domain 30B-like	3

With at least three bindings sites and ranking from highest binding sites and confidence

Due to high homologue of METTL8 with METTL2, top candidates for binding with METTL8 listed in Table 3-1 were tested by RIP using METTL2 antibody Abmart1/2. As shown in Figure 3-14 (C) and (D), NEAT1, which is a long non-coding RNA, forming paraspeckles with NONO and SFPQ, is confirmed, so do IARS and threonine tRNA AGU. It is noteworthy that NONO and SFPQ is identified as protein binding partners of METTL2 by Flag-IP Mass Spec as well as IP western confirmation.

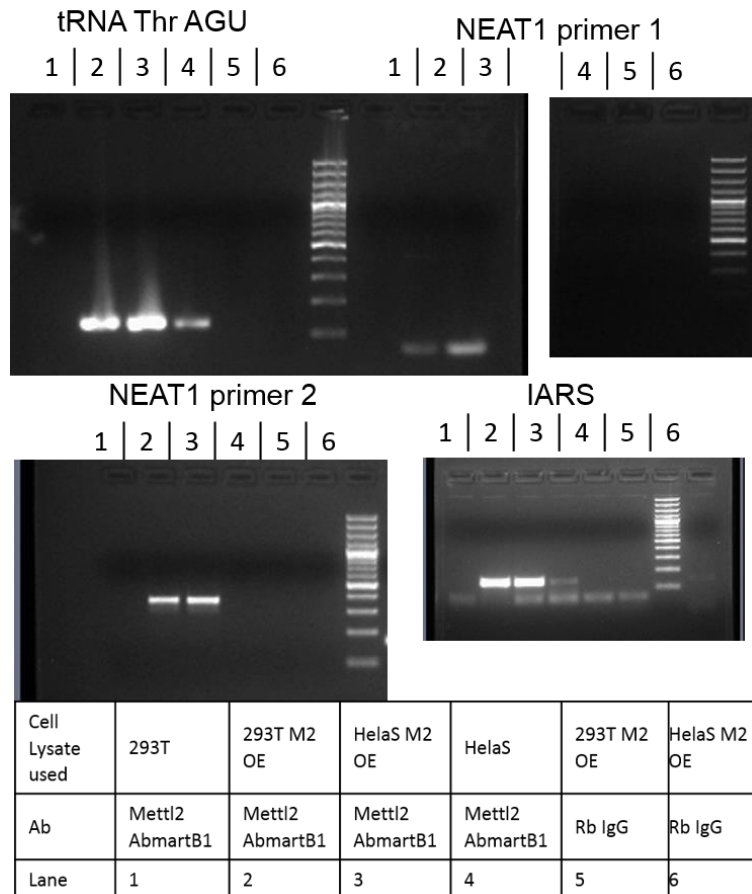


Figure 3-14 Validation of METTL2, 8 RNA binding species identified by PAR-CLIP

The antibody (Ab) used in the RIP are summarized in the table at the bottom.

### 3.13 METTL2 could methylate itself *in vitro*

Besides primer extension assay and Mass Spectrometry in finding substrates for RNA/DNA MTase, direct *in vitro* MTase assay could test if the suspecting protein is a MTase towards certain RNA substrates. The reaction is set up by incubating recombinant proteins, with its potential substrates, and isotope labelled S-adenosylmethionine, which is usually C14 or H3 labelled S-adenosylmethionine ( $^{14}\text{C-SAM}$ , or  $^3\text{H-SAM}$ ), resolved on denaturing Urea gel and exposed using autoradiography. Although rarely used, a good example is finding DNMT2 as a tRNA MTase, instead of being a DNA MTase like its

close paralogs, DNMT1 and DNMT3(182). DNMT2 is shown be able to add methyl group from S-adenosylmethionine towards tRNA from Dnmt2 KO cells. To test if METTL2 could methylate tRNAs from Mettl2 KO cells, a similar assay was done following the methods for DNMT2(182). However, In Figure 3-15 (A) & (B) The bands in the lower part of the gel turns out to be METTL2 itself, since the bands are still there without add RNA from either WT or KO cells. Similar results were obtained using GST tagged proteins as shown in (C) or using <sup>14</sup>C labelled S-adenosylmethionine as shown in (D). Indicating recombinant METTL2 could add <sup>3</sup>H-labeled methylgroup towards protein itself, instead of the tRNAs.

It is not surprising a potential RNA/DNA MTase could methylate itself in vitro, since in vitro reactions creates a highly concentrated environment for proteins and many recombinant Methyltransferase could show automethylation activity *in vitro*. FPLC purified proteins as in 3.4 were used in those assay.



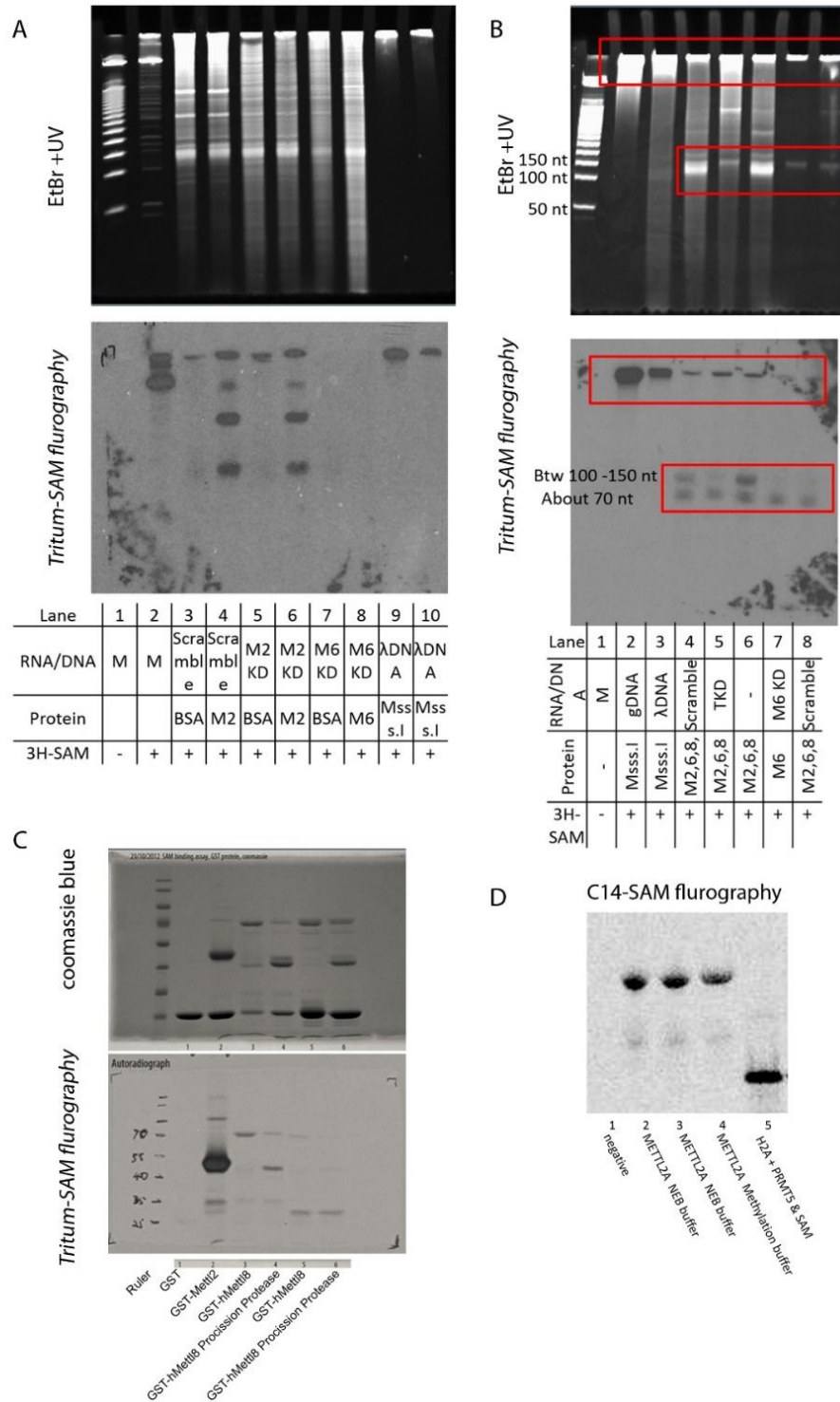


Figure 3-15 METTL2 could auto-methylate itself *in vitro*

(A) *In Vitro* MTase assay using FPLC purified recombinant METTL2B proteins, with RNA from either scramble shRNA transfected cells or Mettl2, 6, 8 stable knocking down cells. EtBr staining of RNA is at upper panel, autoradiography is at lower panel.

(B) Repeat of another batch of His tagged METTL2 proteins in *in Vitro* MTase assay, no RNA was added in lane 6.

(C) Repeat of (A) and (B) by using purified GST tagged METTL2 proteins.

(D) Repeat of *in vitro* MTase assay using C<sup>14</sup>-labelled SAM, using histone H2A methylated by PRMT5 and C<sup>14</sup>-SAM as control.

### **3.14 Identification of METTL2, 6, 8 protein complex by Flag-IP mass spectrometry**

To elucidate the protein complex formed *in vivo*, FLAG tagged Mettl2, 6, 8 vectors were constructed and transfected into HeLaS3 cells and after puromycin selection, several stable cell lines were generated. Immunoprecipitation using FLAG antibody were performed against various cell lysate, and the IP mixture were separated on SDS-PAGE and sent for Mass spectrometry.

Proteins which are of high confidence or verified by reverse IP or flag IP again were shown in Table 3-2. As shown in Figure 3-16 (A), Seryl-tRNA synthetase was identified the only high confidence partners with METTL6, while NONO-SFPQ, some eEF3 were found in complex with METTL2 and METTL8. Seryl-tRNA synthetase raised great interest since it is one of the twenty amino-acryl syntheses which specifically interact with seryl-tRNA and aminoacylate serine tRNAs. Moreover, METTL2A and 2B were just found to target threonine tRNAs, and the other only known tRNA species which bear 3-methylcytosine is tRNA serine and arginine in cytoplasm, and tRNA serine and threonine in mitochondria. This interaction is confirmed by Flag-IP, followed by Western Blot. The finding of METTL6 interact with SARS may imply tRNA serine could be the substrates of METTL6, although not confirmed by primer extension assay, as shown in Figure 3-7, no difference between scramble control and METTL6 KD are observed.

Table 3-2 METTL2, 6, 8 protein complex identified by Flag-IP Mass Spec

Flag tagged Bait	Binding Partners
Mettl2A,2B	<b>SFPQ, NONO, Nucleolin</b> , EF1A1, EF1A2, EF1A3, HNRPK
Mettl6	<b>SARS</b> , MEP50, ICLN(Methylosome subunit pICln)
Mettl8	<b>SFPQ, NONO</b>

Proteins in bold are verified by Flag IP, Western Blot

### 3.15 METTL6 interact with Seryl-tRNA synthetase (SARS) in a RNA dependent manner

To validate the Mass Spec finding, FLAG IP were performed against cell lysate which has overexpressed FLAG-METTL6 and FLAG-METTL6- $\Delta$ SAM. FLAG-METTL6 interacts with endogenous SARS, as consistent with Mass Spec finding, while FLAG-METTL6- $\Delta$ SAM failed to show the interaction. To further narrow down the region responsible for the interaction between SARS and METTL6, a series of truncations were made for GST tagged SARS protein, which has 1-172 amino acids, 1-152 amino acids, 153-460 amino acids, 153-514 amino acids of SARS, all counted from the N terminal of SARS. Intriguingly, only endogenous SARS were immunoprecipitated by FLAG-METTL6, while all the truncated SARS failed to bind full length METTL6, which may indicated either there are multiple domains required on SARS or the tertiary structures is of critical importance in recognizing METTL6, see discussion.

Given the fact that SARS interact with tRNA serine, a hypothesis that SARS and METTL6 may interact through tRNA serine. IP mixture were digested with RNase A or DNase I before the washing steps and resolved on SDS-

PAGE. Interestingly, the binding between SARS and METTL6 are lost upon RNase incubation, but not DNase treatment. These results indicates SARS interacts with METTL6 in a RNA dependent manner, although the identity of the RNA species has not been confirmed as serine tRNA. The failure of using primer extension to test 3mC at position 32 in serine tRNAs also render the substrates of METTL6 elusive.

The interaction between METTL6 and SARS's inspired the hypothesis that METTL2A, 2B may interact with Threonyl-tRNA synthetase. However, the FLAG IP towards METTL2 Overexpressed cells failed to discover TARS as binding partners with METTL2A, 2B.

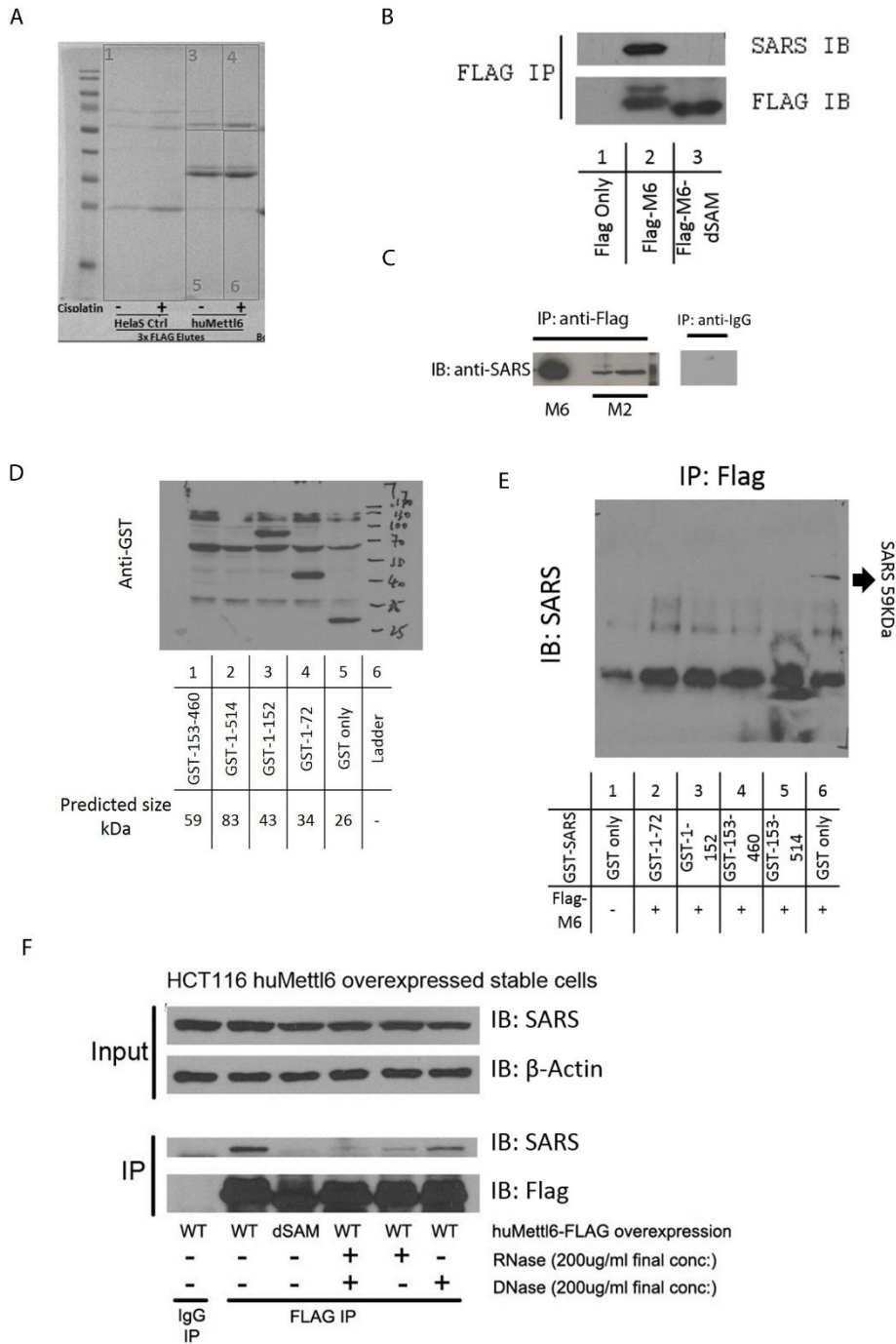


Figure 3-16 METTL6 interact with seryl-tRNA synthetase (SARS) in a RNA dependent manner

(A) Commassie blue staining for FLAG immunoprecipitated mixtures from FLAG only and FLAG-METTL6 ectopic expressed cell lysates, the boxed regions are sent for Mass Spec.

(B) FLAG-IP from cells which overexpress FLAG only vector, FLAG\_METTL6 and FLAG-METTL6- $\Delta$ SAM and blot against SARS

(C) Flag IP in either METTL2 or METTL6 overexpressed cells

- (D) Expression check of several GST tagged-truncated SARS proteins
- (E) FLAG-IP in cells overexpressed of various GST tagged-truncated SARS proteins
- (F) Treatment of FLAG-IPed mixture with or without RNase, DNase and blot against FLAG (for IP quality check,) and SARS, check lysate for ectopic protein expression levels by blotting against SARS and  $\beta$ -Actin

### **3.16 METTL2, 6, and 8 in global translation**

Due to the potential link between certain tRNA modifications and translation. Polysome profiling was performed to check if Knocking down or overexpression mutant METTL2, 6, 8 would affect global translation. Of note, each plate of cells should not be too confluent, otherwise it would lead to slowing down of translation machinery, thus caused biased results. Each sample presented in the Figure 3-17 were harvest at 70 % ~ 90 % confluence.

Polysome profiling was performed to check if the knocking down or ectopic overexpression of mutant METTL2 family proteins would affect global translation. Cells in every profile were harvested at 70% to 90% confluence. All cells are grown in standard normal growth conditions (For most cell lines, it is DMEM high sucrose and 10 % FBS). There is no statistically significant difference between wild type and METTL2 KO clones (clone 3 and clone 7) for the polysome versus monosome area under curve (AUC) ratio (P/M ratio) (Figure 3-17A). In another study, knocking down of METTL2A and 2B induced a higher monosome peak and lower polysome peaks than control scramble shRNA groups. Moreover, knocking down or knock out METTL8 also failed to show drastic difference with wild type control. However, overexpressing METTL8  $\Delta$ SAM mutant caused global down regulation of translation as shown by the higher monosome peak and lower polysome peaks,

indicating a role of METTL8  $\Delta$ SAM mutants in affecting global translation (Figure 3-17B). It could be possible that overexpression of METTL8  $\Delta$ SAM mutants would cause some cellular stress such as protein misfolding, or it could serve as dominant negative towards endogenous METTL8. Future work need to be done to clarify the difference between different cell lines. Moreover, Mettl6 or SARS would cause a global change in translation, as demonstrated by lower P/M ratio (Figure 3-17C,D). However, METTL6 single KO failed to show a significant downregulation of P/M ratio. In all, in all the cell lines we tested, we failed to observe a significant change in P/M ratio, indicating the function of METTL2, 6 and 8 remains elusive.

As shown in Figure 3-17 (A), knocking down METTL2 (both 2A & 2B) affected a global translation by inducing a higher monosome peak and lower polysome peaks. Although the total area under curve (starting from 40S monosome peak and ending at the end of fractionation) is not identical between two samples in Figure 3-17 (A), there's still a significant decrease in polysome fractions as compared in KD cells by normalizing each OD of METTL2 KD cells to WT.

As mentioned in introduction, overexpressing mutant METTL8, or loss of METTL8 would cause up-regulation of ATM, at two fold up-regulation at mRNA level, but 5-10 fold up-regulation at protein level accessed by Western blot. To address mechanism for up-regulation of ATM, METTL8 WT and  $\Delta$ SAM mutant proteins were transfected into HCT116 cells. After puromycin selection and colony picking up, stable cells overexpressing METTL8 WT and  $\Delta$ SAM mutant were generated and proceeded to polysome profiling as shown in Figure 3-17 (B). Overexpressing METTL8  $\Delta$ SAM mutant caused global

down regulation of translation as shown by the higher monosome peak and lower polysome peaks, indicating a role of METTL8  $\Delta$ SAM mutants in affecting global translation apparatus. Moreover, Mettl6 or SARS knockdown would cause a global down regulation of translation machinery, as demonstrated by polysome: monosome ratio. Either Mettl6 or SARS KD will caused a lower polysome peaks and higher monosome peaks.

However, a METTL8 KD or KO failed to show a similar pattern as METTL8  $\Delta$ SAM mutants, neither in some other METTL2/6 KO, indicating future work need to be done to clarify the inconsistency between different cell lines, gene Knockdown/Knockout methods.

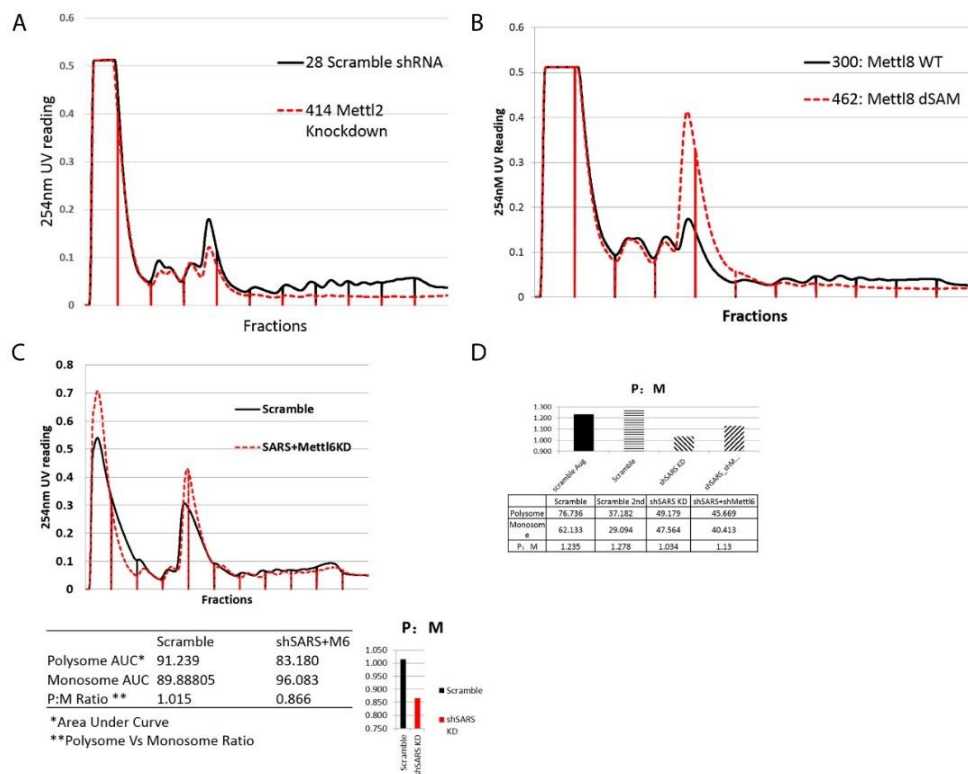


Figure 3-17 METTL2, 6, 8 could affect global Translation



- (A) Polysome profiling comparing between cells stably expressing scramble shRNA and Mettl2 shRNA;
- (B) Polysome profiling comparing between cells stably expressing wild type METTL8 and METTL8  $\Delta$ SAM mutant;
- (C) Polysome profiling comparing between cells stably expressing scramble shRNA and Mettl6, Seryl-tRNA synthetase double shRNA;
- (D) Another experiments repeating (C)

### **3.17 Loss of METTL2 confers higher sensitivity to translation inhibitors and one alkylation agent**

Measuring cell line's viability, growth rates and sensitivity to chemicals that targets various cellular pathways are common themes for gene function studies. Cell growth rates were accessed by counting cell numbers using haemocytometer at different time points, or use MTS assay. In several cell lines tested ( 293T, HeLaS3, HCT116, and A172 cell lines), KD / KO METTL2 alone could confer reduced cell growth rates. Of note, A172 cell lines showed a much profound reduced growth rates upon Mettl2 KD, as shown in Figure 3-18 (A), so did HCT116, as accessed by colony formation assay as shown in Figure 3-18 (B). Double KD of Mettl2 & 8 or a triple KD reduced the cell growth rates and viability more pronounced in HCT116 and HeLaS3 cell lines.

Due to the sensitivity of ABP140 KO strain to neomycin and ABP140&TRMT1 Double KO to CHX(115), sensitivity of METTL2, 6, 8 KD/KO to both drugs were performed by MTS assay; A172 cells showed a higher sensitivity to Geneticin and CHX, as shown by MTS assay, as in Figure 3-18 (C). However, the effects are not statistically significant in some other cell lines, which need more thorough investigation.

ABP140 KO yeast showed a significant higher sensitivity to MMS alkylation damage(117). To test if this is true in mammalian cell lines, MMS treatment plus MTS assay were performed and indeed, Mettl2, 8 double KD and Mettl2, 6, 8 triple KD showed a higher sensitivity to MMS at various concentrations ranged from 1 mM to 5 mM, comparable to similar window in yeast.

It is of note that only repeatable results were shown in Figure 3-18. Further studies in more cell lines or Mettl2, 6 KO mice are needed support those phenotypes observed

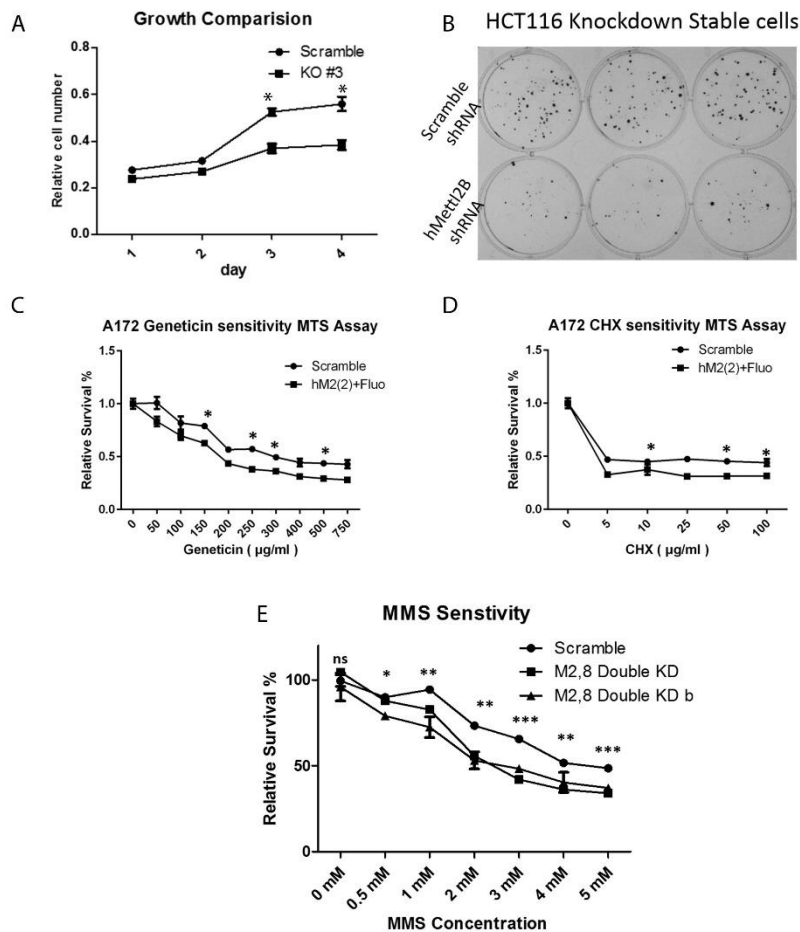


Figure 3-18 Loss of METTL2, 6, 8 confers higher sensitivity to translation inhibitors and alkylation damages

(A) Representative growth curve shows loss of METTL2 confers lower growth rate in 293T cell line;

(B) Representative colony formation assay comparing control knockdown and METTL2 stable knockdown

(C), (D) and (E): MTS assay to assess sensitivity to Geneticin (Gibco), which is an analog to neomycin, Cycloheximide and MMS;

(\*\*\* indicate p value < 0.005, \*\* indicates p value < 0.01, \* indicates p value < 0.05, student T tests)

### 3.18 Codon-run to study effects of 3-methylcytosine on translation

Inactivation of certain tRNA modification enzymes can impair translational decoding process, as measured by decreased translation of relevant 'codon-runs' from several publications (183, 184). To investigate the possible effects of the loss of 3-methylcytosine observed in the METTL2 KO cells, reporter plasmids were constructed, containing both Renilla luciferase (Rluc; internal reference) and firefly luciferase (luc2) genes, in which luc2 translation, but not Rluc translation, depends on in-frame translation with a codon-run of five or ten identical codons(183, 184).

Making Codon-run reporter constructs: pEF6 vector is inserted with codon optimized Renilla luciferase (Rluc) and Firefly luciferase (luc2) cloned from pMirGlo of Promega using oligos listed in the Table below, with stop codon for Renilla luciferase excluded. Renilla luciferase was inserted into pEF6's via KpnI and BamHI RE sites located at Multiple cloning sites while firefly luciferase was inserted using EcoRI and NotI, thus leaving 5'-ACT AGT CCA GTG TGG TG-3' linker located between Renilla luciferase and Firefly luciferase.

Table 3-3 Oligos for cloning Renilla luciferase and Firefly luciferase

hRluc_F_KpnI	GTGAGGTACCATGGCTTCCAAGGTGTACGACC
--------------	----------------------------------

hRluc_R_BamHI	ATGCGGATCCGAAGAAGAACTCGTCAAGAAGGCGATAG
luc2_F_EcoRI	CTAGGAATTCATGGAAGATGCCAAAAACATTAAG
luc2_R_NotI	AGATAAGCGGCCGCTTACACGGCGATCTTGCCGCCCTTC

Runs of codons matching the anticodons of several human tRNAs which is approved or suspected to have m<sup>3</sup>C were generated by annealing DNA oligos containing BamHI/EcoRI overhangs at either ends and harboring five or ten successive, identical Threonine, Serine, or Glycine/Serine codons or Control codons, and inserted between the BamHI and EcoRI sites downstream of the Rluc and upstream of luc2. The luc2/Rluc ratios obtained from control cells or Knockdown/Knockout cells were determined using luciferase assay, which used passive lysis buffer according to manufacturer's instruction (Promega's Dual Luciferase kit) and measured manually.

(ACN)10 and (UCN/AGY)10 indicate runs of 10 identical threonine and serine codons, respectively, and are translated in frame with firefly luciferase (luc2). Reporter constructs containing the indicated codon runs, were transfected into WT or METTL2 KO cells. The ratio between luc2 and Rluc was obtained and used to calculate the efficiency of translation of the codon run.

However, the differences between METTL2 KO and WT were generally negligible, showing that the altered m<sup>3</sup>C modification pattern of METTL2 KO tRNAs have little effect on translation efficiency in this experimental system. Nevertheless, it is noteworthy that the absolute luminance difference between WT and Mettl2 KO regarding threonine codons and Scramble shRNA, SARS

KD, Mettl6 KD is obvious, as shown in the Figure (B) and (C) below. Since Renilla luciferase harbours several threonine AGU/UGU and serine codons, it is not surprising that there is a difference in absolute readings for both luciferase between WT and KO.

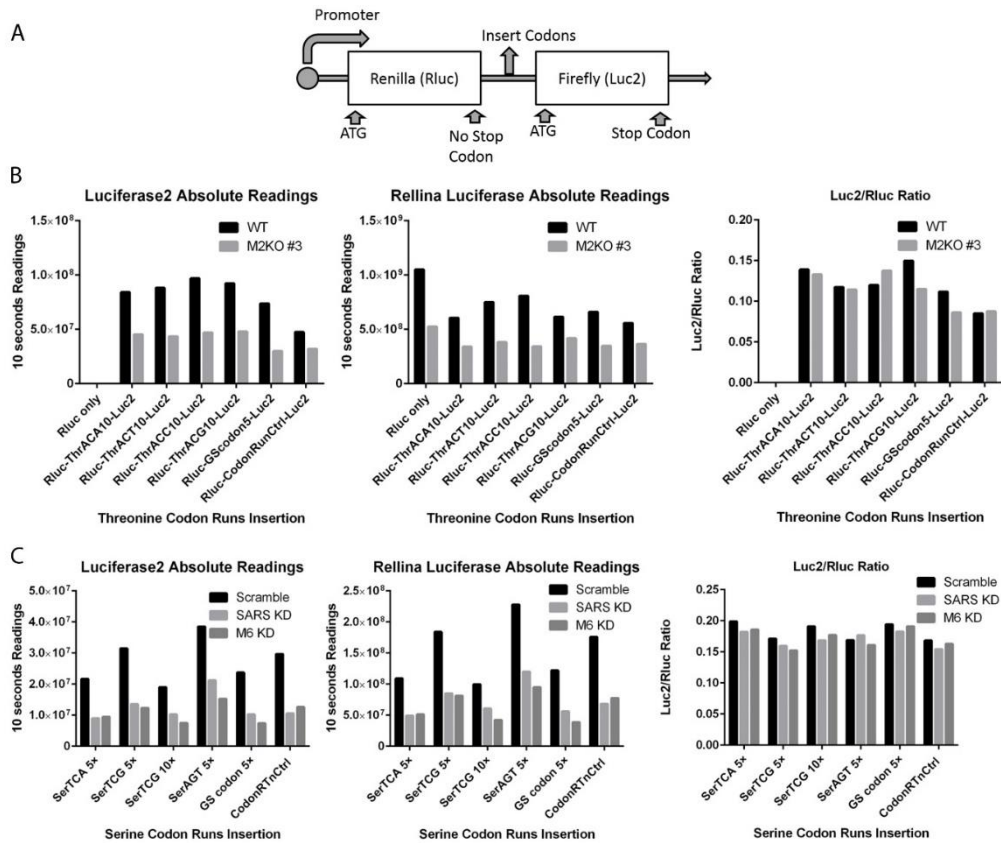


Figure 3-19 Codon-run to study effects of 3-methylcytosine on translation

(A) Diagram of Codon-run reporter constructs, which is cloned from pEF6/V5-His A vector, driven by the pEF1 $\alpha$  promoter

(B) Threonine codon-runs, 10 seconds' absolute readings of firefly luc2 and Renilla luciferase are shown in the left and middle panel, with ratios of luc2/Rluc on the right panel, comparing 293T Mettl2 WT and KO, cells are seeded at same density and transfected with vectors for 24 hours

(C) Serine codon-runs, 10 seconds' absolute readings of firefly luc2 and renilla luciferase are shown in the left and middle panel, with ratios of luc2/Rluc on the right panel, comparing HCT116 Control KD, Seryl-tRNA synthetase KD, and Mettl6 KD, cells are seeded at same density and transfected with vectors for 24 hours

### 3.19 Phylogeny and structural analysis tRNA 32-38 pair

The anticodon loop of tRNA is universally conserved to contain seven nucleotides and is important in interacting with other proteins or RNA, such as ribosome or amino-acyl tRNA synthetase. As introduced in section 1.7, the 32-38 base pairs located at the junction of anticodon stem and loop, it is shown in some tRNAs, the 32-38 pair has contact with ribosomal RNA to ensure uniform tRNA binding to the A site and play a role in discrimination of synonymous codons and reading frame maintenance at P sites. Thus, the pair extends the sequence properties of the anticodon loop in addition to the usually modified purine at 37 and the U-turn at position 33.

Phylogenetic analysis was performed using a tRNA database (11), its design allows user to get numbers of tRNA with specific nucleotide at position 32 and position 38 by typing in, for example, C\*\*\*\*\*A (\* indicates any nucleotide) query into the database, thus a specific number of this kind of 32-38 pair out of all tRNAs for different domain of life, or certain species, or certain tRNA isoacceptors are obtained. As shown in the Table 3-4 below, 32(Cytidine or modified cytidine)-38(Adenosine) base pair outnumber other base pairs in all three domains of life. However, by taking a closer look at Threonine tRNAs, it is noteworthy that Cytidine (32) - Adenosine (38) pair on threonine tRNAs has a much higher percentage in Eukaryotes than in either Bacteria or Archaea. In fact, all human threonine tRNAs has cytosine and adenosine at position 32 and 38 respectively, based on the tRNA database(11), while threonine tRNA in *E. Coli* all have U-A pair at 32-38 positions, indicating a difference in tRNA anticodon stem loop during evolution and natural selection. Moreover, all human threonine tRNAs has uridine and

adenosine at position 33 and 37, thus human threonine tRNAs all have CU\*\*\*AA at anticodon loop (\*\*\*) indicates anticodon).

Based on the Table 3-4 and Table 3-5, and from previous studies(125), approximately 90% of the 32-38 oppositions can be classified to two families, in which each pair are isosteric towards each other, one with a larger (approx. 86 %, including 32C-38A pair) and the other with a smaller (approx. 7 %) population. The remaining (approx. 7 %) of the pairs have been grouped to a 3rd family, since they cannot be designated into the first two. Of note, the Y(pyrimidine)32-R38 base-pairs are not isosteric upon base pair reversal, may provide a structural explanation for the conservation of pyrimidine at position 32.

Table 3-4 tRNA 32–38 base pairs frequency in three domains of life

	Fa mi ly	32–38 pair	Eukaryota		Bacteria		Archaea	
			Number	%	Number	%	Number	%
1	I	CA	1300	58.0	3312	52.0	760	69.9
2	I	UA	78	3.5	1137	17.9	119	10.9
3	II	UU	335	14.9	661	10.4	36	3.3
4	I	CC	256	11.4	539	8.5	102	9.4
5	I	UC	188	8.4	304	4.8	13	1.2
6	III	AU	1	0.0	119	1.9	1	0.1
7	III	AA	2	0.1	93	1.5	2	0.2
8	III	UG	4	0.2	85	1.3	44	4.0
9	III	CG	11	0.5	74	1.2	7	0.6
10		AC	5	0.2	16	0.3	3	0.3
11	III	CU	57	2.5	12	0.2	0	0.0
12	III	GA	2	0.1	9	0.1	0	0.0
13	III	GU	3	0.1	3	0.0	0	0.0
14		AG	0	0.0	2	0.0	0	0.0
15		GG	1	0.0	1	0.0	0	0.0
16		GC	0	0.0	0	0	0	0.0
		Total	2243	100	6367	100	1087	100

\*calculated from tRNA database(109)

Table 3-5 Threonine tRNA 32–38 base pairs frequency in three domains of life

Family	32–38 pair	Eukaryota		Bacteria		Archaea	
		Number	%	Number	%	Number	%
I	CA	117	88.6	152	35.8	40	58.0
I	UA	14	10.6	209	49.3	29	42.0
II	UU	1	0.8	0	0.0	0	0.0
I	CC	0	0.0	1	0.2	0	0.0
I	UC	0	0.0	0	0.0	0	0.0
III	AU	0	0.0	0	0.0	0	0.0
III	AA	0	0.0	62	14.6	0	0.0
III	UG	0	0.0	0	0.0	0	0.0
III	CG	0	0.0	0	0.0	0	0.0
	AC	0	0.0	0	0.0	0	0.0
III	CU	0	0.0	0	0.0	0	0.0
III	GA	0	0.0	0	0.0	0	0.0
III	GU	0	0.0	0	0.0	0	0.0
	AG	0	0.0	0	0.0	0	0.0
	GG	0	0.0	0	0.0	0	0.0
	GC	0	0.0	0	0.0	0	0.0
	Total	132	100	424	100	69	100

\*calculated from tRNA database(109)

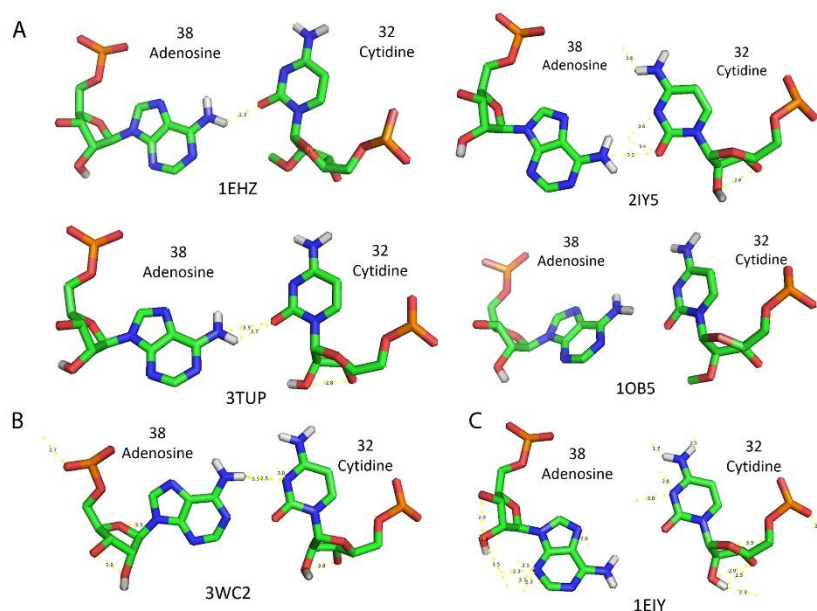
For structural analysis tRNA 32-38 pair, based on the available sequences and crystal structures, it is demonstrated the relatively conservation of a bifurcated hydrogen bond interaction between residues 32 and 38 located at the stem/loop junction in most of the tRNA<sup>Phe</sup> (12). This interaction leads to the formation of a non-canonical base-pair Figure 3-20 (A) .

However, a few exceptions are discovered for this bifurcated hydrogen bonds by searching available PDB structures. For example, one tRNA/synthetase complex between threonine tRNA synthetase (TARS) and tRNA threonine from *E. Coli* (Protein Data Bank: 1QF6) showed the U-A pair is impaired upon aminoacyl-synthetase binding. As shown in the Figure 3-20 (D), the uridine residue is interacting with one asparagine residue from TARS, indicating a potential tRNA conformation change upon aaRS interaction.



Another example is from PDB entry 1E1Y, in which Cytosine and Adenosine do not have polar interaction between each other. Moreover, cytosine is bonding with N atom at 3' position instead of the oxygen atom on the pyrimidine ring of cytosine in PDB entry 3WC2. Strikingly, 32 Cytosine interact with 37 adenosine instead of 38 adenosine in several arginine tRNAs ( PDB 2ZUE and 2ZUF )

These examples searched from PDB and presented Figure 3-20 (B-F) further demonstrated the pairing diversity of 32C-38A pair in various structures. In all, the extended sequence signature by 32-38 nucleotides at the junction of the anticodon loop and stem underscores the importance of both nucleotides. Lacking of tRNA structures with 3mC calls for further functional and structural studies to address the function of modification.



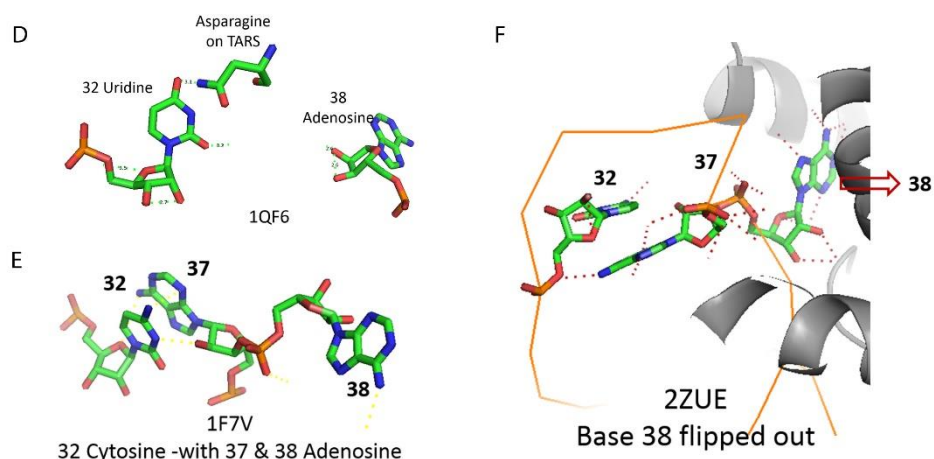


Figure 3-20 Interaction between tRNA Residue 32 and 38 extracted from solved structures

Each stick diagram is extracted from one PDB accession file indicated below each diagram. Carbon: Green; Nitrogen: Blue; Hydrogen: White (only some of the polar hydrogen shown); Oxygen: Red; Phosphor: Orange.

(A). C-A pair at 32-38, notice the pair occurs between O atom on cytosine and N atom on adenosine;

(B). C-A pair at 32-38, notice the pair occurs between 3' N atom on cytosine and N atom on adenosine;

(C). No polar interaction between C-A pair at 32-38;

(D). 32 Uridine may interact with protein in complex, such as TARS

(E) & (F). 32 Cytosine interact with 37 adenosine instead of 38 adenosine, phosphate backbone is shown in orange ribbons in (F)

Table 3-6 Summary of each 32-38 pair in the Figure 3-20

PDB Accession	tRNA	32	38	32 atom	38 atom	Specis
1EHZ	Phe	Cm	A	O	N	Baker's Yeast
2IY5	Phe	C	A	O	N	Th.Thermophilus
3TUP	Phe	Cm	A	O	N	Human mito
1OB5	Phe	Cm	A	O	N	Bacteria
3WC2	Phe	C	A	N	N	C. albicans
1E1Y	Phe	C	A		-	Th.Thermophilus
1QF6	Thr	U	A		-	<i>E. Coli</i>
1F7V*	Arg	C	A		-	Baker's Yeast
2ZUE*	Arg	C	A		-	Pyrococcus horikoshii
2ZUF*	Arg	C	A		-	Pyrococcus

					horikoshii
- no polar contacts between 32 and 38 bases					
* 32 Cytosine interacts with 37 adenosine instead of 38 adenosine					

## CHAPTER 4. DISCUSSIONS AND FUTURE WORK

The chapter will present the limitations of some of the methods used and some of the results obtained, besides discussing the scientific achievements from this thesis project. Ongoing or future work is also to be described in the chapter.

### 4.1 Limitation of Primer Extension assay

Despite many advantages as shown in the introduction 1.2.1 and results 3.5, the use of reverse transcription for detection of modified nucleotides in RNAs is not very straightforward. In fact, the reverse transcription profile, even for an unmodified RNA, depends on several parameters, especially the presence of Pyrimidine - A (C-A and U-A) bonds, which are very sensitive to nuclease cleavage (some commercial reverse transcriptase is a nuclease), and also on some RNA secondary structure which could also generate RT pauses / blocks. Therefore, the simple presence of a pause in a primer extension profile does not necessarily indicate the presence of a modified residue. In addition, reverse transcription profiles of modified RNAs are often prone to some kind of stuttering (doubling of the bands on the gel) at the modified residue that blocks the reverse transcription (Figure 1-2) (61, 185). The reasons for such stuttering remains unclear, but this possibility complicates the detection of nearby modified residues and should be considered during analysis of reverse transcription patterns.

It is indeed the case for primer extension assay for tRNAs other than threonine tRNA<sup>AGU</sup> or tRNA<sup>UGU</sup>. Indeed, the appearance of pauses in an RT profile may only serve as an indication of the possible occurrence of modified residues at these positions. The use of other complementary techniques of RNA analysis,

such as HPLC and Mass Spec will serve as ultimate ways to confirm the RT results. For primer extension shown in the results 3.5, 20 centimetre gels were used, which is typically used for EMSA experiments. Ideally, a 40 centimetre sequencing length gel should be used to generate a lane pattern at single nucleotide resolution.

#### **4.2 Limitation of RIP assay**

PAR-CLIP has been used by one group to identify binding partners for NSUN6 (186), in which Flag IP of Flag-tagged NSUN6 showed a significant higher percentage binding towards tRNAs than Flag only vectors, thus leading to identifications of certain tRNA species as substrates for NSUN6.

RIP assay is commonly used to validate the RNA species which is found to bind to RNA binding proteins by PAR-CLIP. However, good negative control and optimization of washing steps should be performed to exclude non-specific binding. Ideally, antibody which can target endogenous protein of interest should be used in RIP, instead of using anti-tag antibody to immunoprecipitate ectopic overexpressed proteins. Indeed, generation of bands using primers targeting GAPDH mRNA, which indicates a potential non-specific binding in RIP assay. Future PAR CLIP or more validation need to be performed to address this issue.

#### **4.3 Limitation of ribosome profiling**

Thought ribosome profiling has not been finished before the submission of this thesis. Limitations and cautions are discussed below for future reference.

#### **4.3.1 Issues in analyzing ribo-seq together with parallel RNA-seq**

Transcriptional or post-transcriptional processes will modulate the abundance or dynamics of the polysome-associated RNA by controlling mRNA. It is therefore essential to include mRNA abundance data when analysing ribosome profiling data. RNA-seq for cytoplasmic or whole-cell mRNA are usually obtained in parallel to allow for such corrections. In the past, translational-efficiency scores calculated as the log<sub>2</sub> ratio of the amount of polysome-associated RNA or RPFs to the amount of cytosolic RNA were used to determine translational activity of an mRNA. However, analysis based on translational-efficiency scores showed poor correlation. The analysis of translational activity (ANOTA) was developed and its results correlate well with changes in protein amounts (187, 188). A third method called 'RPF-based' method has been developed for the analysis of ribosomal profiling (189), (190), However, the performance of this method in monitoring changes in the proteome has not been studied. In all, better analytic methods should be developed and curated for studying polysome profiling data (191).

#### **4.3.2 Sample harvest issues**

One highly concerned issue for ribosome profiling is the extent to which the cell harvest and lysis procedure may distort the ribosome distribution. Unless the material can be cooled to near 0°C extremely rapidly, polysome run-off will occur, depleting ribosomes from all initiation sites, all short ORFs and the 5'-proximal part of long ORFs (192).

To avoid this, many protocols would incubate the cells with CHX for 5 to 10 min before cooling down. This prevents ribosome run-off, but it will give much more footprints at initiation sites than at steady state, because CHX

doesn't inhibit scanning or initiation, and so an additional 80S initiation complex can be added to every mRNA which has a vacant initiation site at the time when the inhibitor starts to “bite.”. For highly efficient sites that capture virtually all scanning ribosomes, it is estimated that 100% occupancy of vacant initiation sites will occur within <10 seconds of the onset of inhibition by CHX. It follows that if an initiation site is mutated to be so inefficient that it captures only 10% of scanning ribosomes, 100% occupancy of all vacant sites will still be achieved in < 2 min after the onset of inhibition. So a 3 to 5 min incubation with CHX could result in footprints of equal magnitude at initiation sites that really differ in intrinsic efficiency by a factor of 10 or more. Added to this, there is a polarity effect in that an 80S ribosome stalled at a 5'-proximal site will prevent any subsequent scanning ribosome from reaching downstream sites, and this could result in the observed footprint signal at a uORF initiation site with mediocre context actually being greater than the signal at a highly efficient downstream start site (192).

It would be a similar situation arise if cells are incubated for 3-5 min with harringtonine, which inhibits neither initiation nor on-going elongation (or polysome run-off), but block the transition from initiation to elongation, so that it leaves clear footprints of 80S ribosomes stalled at every initiation site. This has proved to be an excellent way of identifying initiation sites, but, like CHX, it can completely obscure their relative utilization frequency under steady state.

### **4.3.3 Other cautions in analyzing ribosome profiling**

The precise positions at which ribosomes are associated with the mRNA could be located after the alignment of RPFs to the reference genome. This unique

property has been employed for detailed mechanistic studies on nearly every in translation process, including the initiation, elongation and termination.

Moreover, ribosome profiling in embryonic stem cells has shown noncoding RNAs binds to ribosome, which suggests that non-coding RNAs could be translated into short peptides. However, this finding was re-analysed (193) and found that RPF sequencing obtained reads alone are not sufficient to judge if an RNA is translated or not. More recently, data obtained with cells treated with harringtonin, were insufficient for the prediction of whether an RNA was translated or not.

Of note, many RNA binding proteins could protect, or partially protect its binding RNA species from nuclease digestion. Highly versatile RNA structures may offer another line of defence against nuclease digestion. Both issues should be evaluated when ribosome profiling data are used to study translation process.

#### **4.4 Self-methylation properties of METTL2, 6, 8**

As shown in Figure 3-15, the high self-methylation activity by recombinant METTL2, 6, 8 indicates certain residues of those proteins could be methylated and it is determined by mass spectrometry under the help from Prof. Zhao Yinming's lab at Shanghai, China. Protein methyltransferase usually exhibit methyltransferase activity towards only one kind of amino acid residue. However, both arginine and lysine residues are methylated in the case of mice METTL2 and METTL6, as shown in Figure 4-1A,B,C. Interestingly, R129 residue in mice METTL6 is conserved to R225 in METTL2, but absent in



METTL8. As mentioned in 3.13 METTL2 could methylate itself *in vitro*, it is possible for a RNA/DNA MTase to methylate itself *in vitro*, since *in vitro* reactions creates a highly concentrated environment for proteins, co-factors and substrates and many recombinant Methyltransferase could show automethylation activity *in vitro*, such as PRMT5, even its canonical substrates are histone 3.

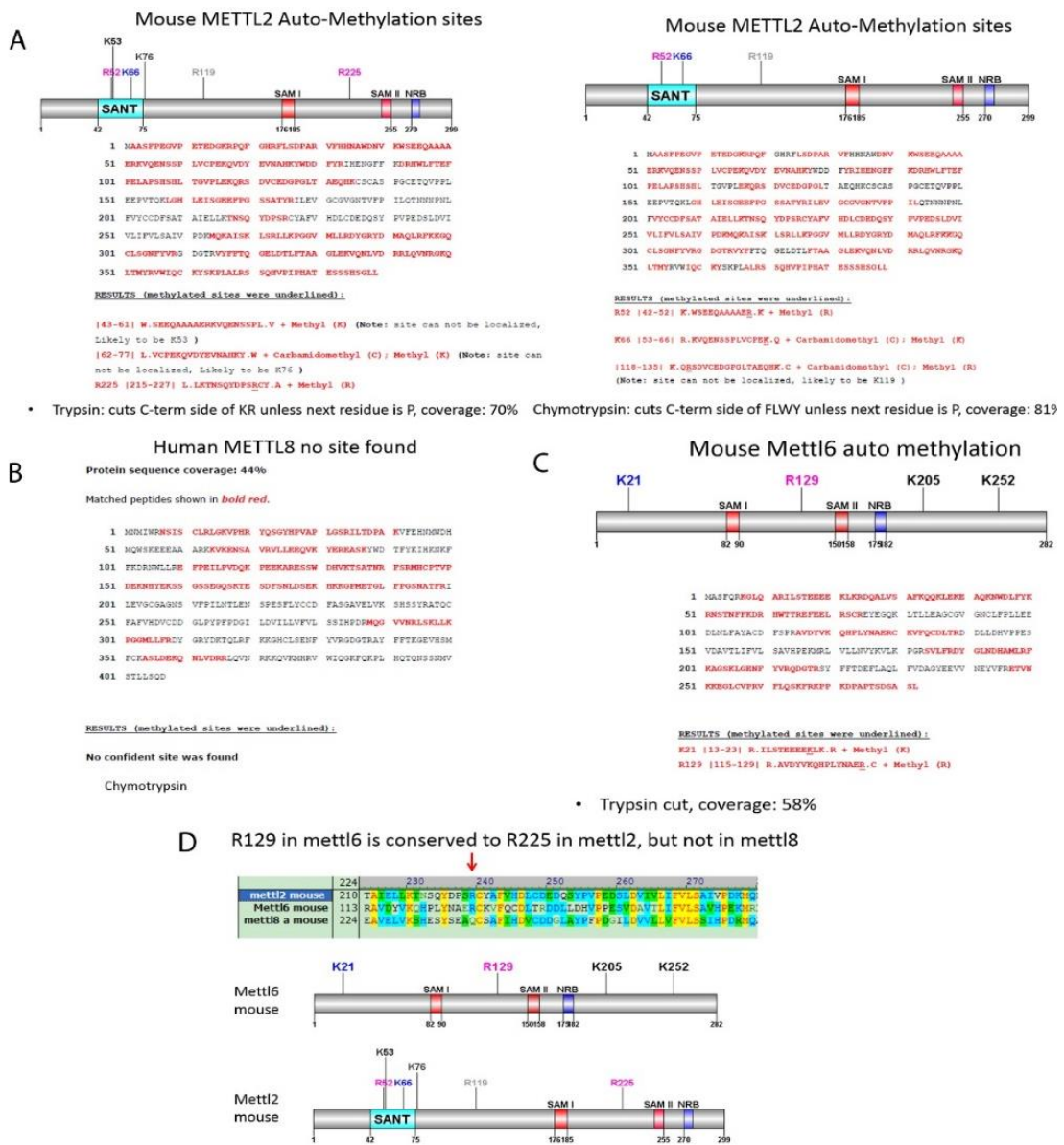


Figure 4-1 Methylation sites identified by mass spectrometry (In collaboration with Prof. Zhao Yinming, Shanghai)

(A) (B) (C). Mass spectrometry identified several lysine (K) or arginine (R) residues being methylated in mouse METTL2 and METTL6. The methylated residues are indicated above the diagram for METTL2 or METTL6. (D) Residue R129 in METTL6 is conserved to R225 in METTL2, but it is absent in METTL8

#### **4.5 Methyl group on polar contacts**

In a methyl group, carbon is linked to three hydrogen atoms. Carbon is more electronegative than hydrogen. So carbon pulls bonded electrons nearer to carbon and carbon acquires a slight extra negative charge and those correspondingly hydrogens lose hold over bonded electrons and acquires a slight positive charge. Thus carbon get a little more electronic charge and it will push electrons towards the group to which it is linked.

Thus methyl group will be electron releasing or electron donating compared to hydrogens, thus they act to prevent the carbon from absorbing some of the excessive negative charge of the conjugate. In an ethyl group ( $\text{CH}_3\text{-CH}_2\text{-}$ ), this effect will be more pronounced than methyl group. The net electron releasing effect of ethyl group is more than that of a methyl group. For example, a tertiary carbon in a tertiary alcohol can absorb less negative charge since it has more methyl groups than a primary alcohol.

##### **4.5.1 Methyl group on N3 cytosine may introduce a transient positive charges upon protonation**

There are three naturally occurring modified nucleotide that could be introduced a transient positive charges(43): 1-methyladenosine, 3-methylcytidine and 7-methylguanosine in cells, although this study has not been thorough(43). Simple methylation of one of the heterocyclic ring nitrogen of either the purine or pyrimidine base can result in a positively

charged quaternary nitrogen. Approximately, 25% of tRNAs harbours m1A<sup>+</sup> at position 58 at TΨC stem-and-loop domain, and approximately 40% have m7G<sup>+</sup> at position 46 of the variable arm.

However, neither of both mononucleosides carries a net charge under physiological conditions. The m1A tautomer prevalent at pH 7 has no net charge and m7G is a zwitterion. It is tertiary structure of the molecules around the nucleosides that protonated modified nucleosides since NMR analyses of native, but <sup>13</sup>CH<sub>3</sub>-enriched yeast and *E. coli* tRNA<sup>Phe</sup> demonstrate that the two methylated nucleosides are positively charged in the tRNA(194).

Protonation occurs through tertiary structure hydrogen bonding, m1A58 to T54 and m7G46<sup>+</sup>, to G22, and C13. Thus, the purpose of methylations at A58 and G46 in tRNAs is thought to produce site-specific electrostatic charges within the tertiary structure of the tRNA and not to add methyl groups to the molecular structure. The positive charges are transient because of their dependence on the weak tertiary interactions that can easily be disrupted by interaction with a protein. The study of other charged nucleosides has not been as thorough. With H-bonded tertiary interactions similar to those of m1A<sup>+</sup>, the nucleoside m<sup>3</sup>C<sup>+</sup> would be positively charged. Thus, methylation of the 3' position of C has the analogous property of introducing a transient Positive charge(43).

#### **4.5.2 3' Methyl group on cytosine will increase the pKa of pyrimidine ring**

It is noteworthy that a roughly 4 % of Cytosine at the position 32 is methylated, and thus protonated(125). Adding a methyl group at 3' position

on cytosine would increase the pKa value of cytosine, it is proposed that a closer pKa value between electron donor and acceptor would increase the strength of hydrogen bonds(195). Thus, adding a methyl group at 3' position will decrease the hydrogen bond strength formed by its adjacent O atom. Meanwhile, it is the adding of a methyl group at 3' Nitrogen, which increase the pKa of the pyrimidine ring that render 3-methylcytosine protonated under physiological conditions. The protonated 3-methylcytosine could serve as an additional hydrogen bond donor at 3' position on the pyrimidine ring.

3-methylcytosine may affect some C-A pairing by a steric clash, based on the solved structures at tRNA 32-38 pairs as shown in the Figure 3 19 (A). Those Cytidines have polar contact with adenosine through 2' Oxygen on C to Nitrogen on A, indicated as. O...H-N. However, there are some exceptions, i.e. 3WC2, in which cytosine bonds adenosine through 3' N atom. It is easy to anticipate that a methyl group on the 3' N atom would create a steric hindrance which could disrupt the hydrogen bonds in certain C(32)-(A)38 pair. Cytosine modifications shows an example of modifying nucleobase to optimize RNA triplex forming. The natural RNA modification has inspired chemists to create man-made modified nucleobase to optimize RNA triplex forming in many biological and biotech applications. Regarding C•G-C (Figure 4-2 (A), (B)), because of the relatively low pKa (about 4.5) of monomer Cytosine, a low pH is required to fully protonate cytosine in a triplex-forming Peptide nucleic acid (PNA) to form a C<sup>+</sup>•G-C base triple. A neutral PNA Nucleobase pseudoisocytosine (J), Figure 4-2 (C), has been synthesized to form a J•G-C base triple(196). The pKa of the J monomer is

about 9.4 and incorporation of J residues in a PNA significantly alleviates the pH dependence for PNA•RNA2 triplex formation(197).

A PNA nucleobase 2-aminopyridine (M) in Figure 4-2 (B) with a monomer pKa of 6.7 is promising for the recognition of a Watson–Crick G-C pair to form an M<sup>+</sup>•G-C base triple(198). The replacement of cytosine residues by M residues allows the formation of stable and sequence-specific PNA•RNA2 triplexes at near-physiological pH conditions(197).

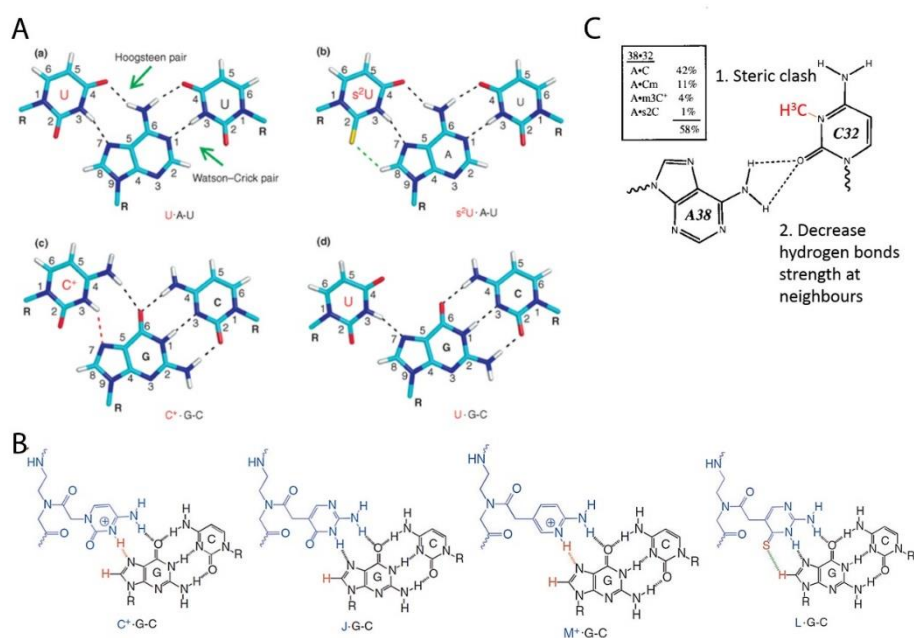


Figure 4-2 Nitrogen methyl group on polar contacts

Adapted from(197)

(A) Major-groove base triples of U·A-U (a), s<sup>2</sup>U·A-U (b), C<sup>+</sup>·G-C (c), and U·G-C (d). Notice in C<sup>+</sup>·G-C, the hydrogen bond formed between protonated N3 in cytosine and N7 in G is indicated by a red dashed line. The H, C, N, O, and S atoms are shown in white, cyan, blue, red, and yellow, respectively.

(B) Major-groove base triples formed between peptide nucleic acid (PNA) bases (blue) and RNA Watson-Crick base pairs (black). Red dashed lines indicates hydrogen bonds formed between protonated 3' Nitrogen in C<sup>+</sup> and N7 in G, and protonated N1 in M<sup>+</sup> and N7 in G. J is for pseudoisocytosine. M<sup>+</sup> is for the protonated form of 2-aminopyridine. L is for thio-pseudoisocytosine. In the base triple L·G-C, the van der Waals interaction between the sulfur

atom of L and H8 atom of G is indicated by a green dashed line. Other hydrogen bonds between the bases are indicated by grey dashed lines.

(C) Proposed effects by the methyl group at N3 position of the pyrimidine ring on cytosine.

#### **4.6 Enrichment of threonine tRNA<sup>AGU</sup> and tRNA<sup>UGU</sup> isoacceptors in eukaryotes**

The concentration of each tRNA is determined by its number of gene copies in the genome(199). Thus, tRNA gene copies correlates well with tRNA isoacceptor abundances and thus able to play a key role in affecting codon translation efficiency. Therefore, the study of tRNA gene content bias may help explaining codon usage biases in extant genomes.

For any actively dividing cell, the translation efficiency of a given codon is determined by the amount of tRNA in the cell(200). The variability in tRNA gene number is extreme in some cases: certain tRNA species are absent in entire branches of the phylogenetic tree, whereas others are clearly predominant, e.g., there are no GGU threonine isoacceptors in the eukaryotes genome, and less than 5 copies of CGU isoacceptors, meanwhile, there are no GGA, ACU, serine tRNA isoacceptors, and less than 5 copies of CGA isoacceptors.

Using principal component analysis, tRNA isoacceptors that became positively selected (increased in number) in Bacteria and Eukarya were identified(201). The results indicate that the appearance of UMs and hetADATs contributed to the divergence of eukaryotic and bacterial genomes from their archaeal counterparts. The effect of the Inosine 34 in eukaryotes and xo5U34 in bacteria caused by these enzymes increased the decoding

capacity of modified tRNAs which, therefore, were positively selected during evolution(201). The diverse codon usage biases displayed by Bacteria and Eukarya are, at least partly, due to the different modification strategies used to improve translation efficiency. However, other modifications have not been taken into calculation, and certain modification and the extended anticodon hypothesis may support other modification's role in Unequal Enrichment of tRNA Isoacceptors in different Kingdom.

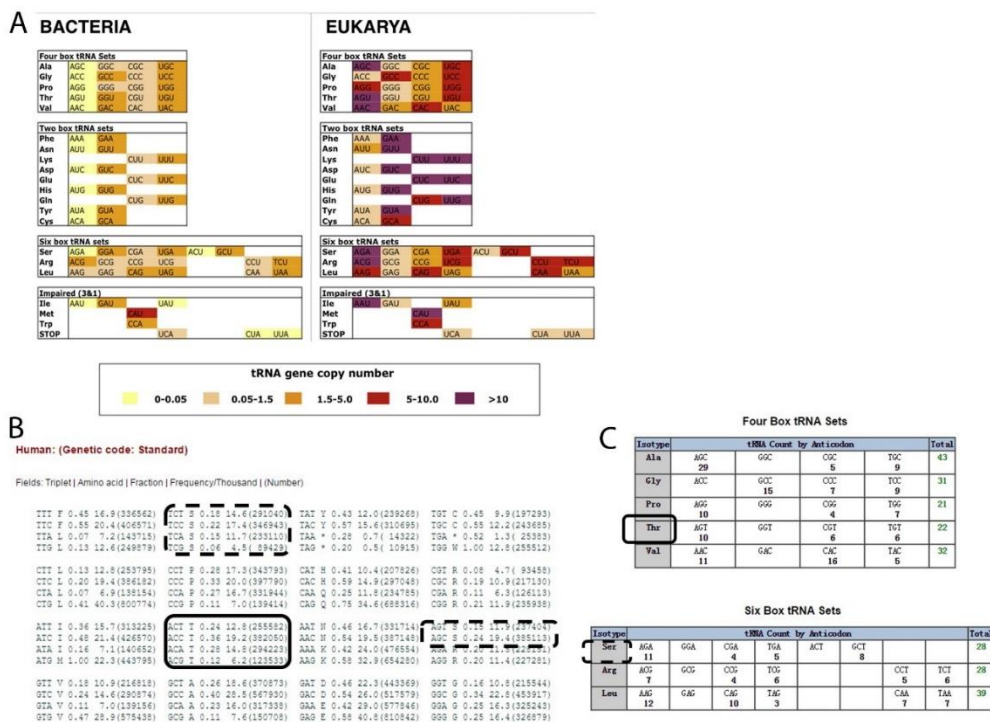


Figure 4-3 Enrichment of some codons and tRNA isoacceptors in Eukaryotics  
 (A) tRNA gene copy numbers in Bacteria and Eukarya, adapted from (202), different colors around codons different copy number of tRNAs  
 (B) Human codon percentage, adapted from(171)  
 (C) tRNA gene copy numbers in human, adapted from(171)

#### **4.7 Future work for identification of all substrates and modification sites by METTL2 family proteins**

As presented in the results, only threonine tRNA<sup>AGU</sup> and tRNA<sup>UGU</sup> isoacceptors have been confirmed as substrates for METTL2. Other modification sites have been elusive, as indicated from the mass spectrometry data and Figure 3-12, Figure 3-13.

It is difficult to use primer extension to find every modification site for METTL2,6,8, since many of the probes for primer extension shows a stutter pattern and it requires a tremendous amount of work. Alternatively, m3C-seq has been planned to be able detect all 3-methylcytosine sites by comparing WT or METTL2, 6, 8 KO cells, which could be easily developed from the methods used from global profiling pseudouridine ( $\Psi$ ) (38, 39). 1-methyladenosine seq to profile all possible m1A sites in tRNA, rRNA (Poster at 2015 RNA editing Gordon conference and personal communication). A similar approach has been applied towards m<sup>3</sup>C.

Single molecule real time sequencing (SMRT) offers another way to profile 3-methylcytosine in whole genome or transcriptome. SMRT is a single molecule DNA sequencing method, in which a single molecular of DNA polymerase enzyme is fixed at the bottom of a zero-mode waveguide (ZMW) with a single molecule of DNA as a template(203). Four different fluorescent dyes is attached to one of four DNA bases. The fluorescent tag is cleaved off and diffuses out of the observation area of the ZMW, when a nucleotide is incorporated by the DNA polymerase, with a detector monitoring the fluorescent signal of the nucleotide incorporation(203).



This third generation sequencing technology has been utilized to discover a broad spectrum of DNA base modifications, such as 5-methylcytosine ( $m^5C$ ), 6-methyladenine ( $m^6A$ ), N4-methylcytosine ( $m^4C$ ), etc (as shown by Pacific Biosciences webpage). in which variant enzyme kinetics identify modified DNA bases. Of note, six human Pols belonging to the three families: B (Pol  $\delta$ ), X (Pols  $\beta$  and  $\lambda$ ) and Y (Pols  $\kappa$ ,  $\iota$  and  $\eta$ ) have been compared for the properties to read through 3-methylcytidine. It was shown that  $m^3C$  would impair B-family, block X-family, but not Polymerases in Y-family. Although both Pols  $\eta$  and  $\iota$  could install nucleotides opposite  $m^3C$ , they could introduce erroneous dATP and dTTP, respectively. Meanwhile, Pols  $\kappa$  and  $\eta$  are most efficient in extending  $m^3C$  base-paired primers. In conclusion, the presence of functional Pol  $\eta$  is mandatory to efficiently overcome 3-methylcytosine ( $m^3C$ ) by mediating complete bypass or extension. In future, protein engineering could use the natural properties of Pols such as Pols  $\eta$ , or alter the properties of natural DNA polymerase to map  $m^3C$  transcriptome wide using SMRT.

#### **4.8 Future structural study of METTL2 family proteins with co-factors and substrates.**

Extensive crystallization trials failed to yield diffracting crystals of METTL2, 6, or 8 proteins, with only some low quality crystals/ precipitates obtained, as shown in Figure 4-4 (D). This growth conditions generate low quality crystals no good diffraction data is obtained. Current NMR is unable to study the proteins at the size range of METTL2, 6, 8, while cryo-EM could only solve higher molecular weight proteins at this moment.

METTL2, 6, 8 has been predicted to contain a large portion of disordered region at N- terminal by XtalPred (174). Without any new strategy, it would

be quite difficult to crystallize human METTL2 alone (or other METTL proteins). Growing co-crystal complex between METTL2, co-factor SAM or SAH, and threonine tRNAs may have the possibility to stabilize the disorder issue in METTL2 proteins. There are several studies in which a series of mutation are introduced to reduce the entropy of surface exposed residues. Those mutations have successfully helped in protein crystallizations (or called surface entropy reduction, SER strategy) (204). In particular, a mutation of 3 residues, K100A, E101A, E103A helped obtained the crystal structure of human nicotinamide N-methyltransferase (hNNMT) (205), which is relatively close to METTL2, 6, 8 proteins by sequence alignment.

As shown in Figure 4-4(A) and (B), Seryl-tRNA synthetase of *T. Thermophilus* in complex with a part of tRNA<sup>Ser</sup> has been solved, without residue at position 32. While only apo structure of human SARS has been solved, without tRNA substrates. In sum, no crystals structures for tRNAs with m<sup>3</sup>C at position 32 has been solved. Once solved, it would provide insights into potential functions of this modifications at this position.

Alternatively, a co-crystal structure between METTL2, with tRNA, or together with corresponding aminoacyl-tRNA synthetase would be another possibility to grow out crystals and to study at functions of 3-methylcytosine. As shown in a cartoon in Figure 4-4(C).

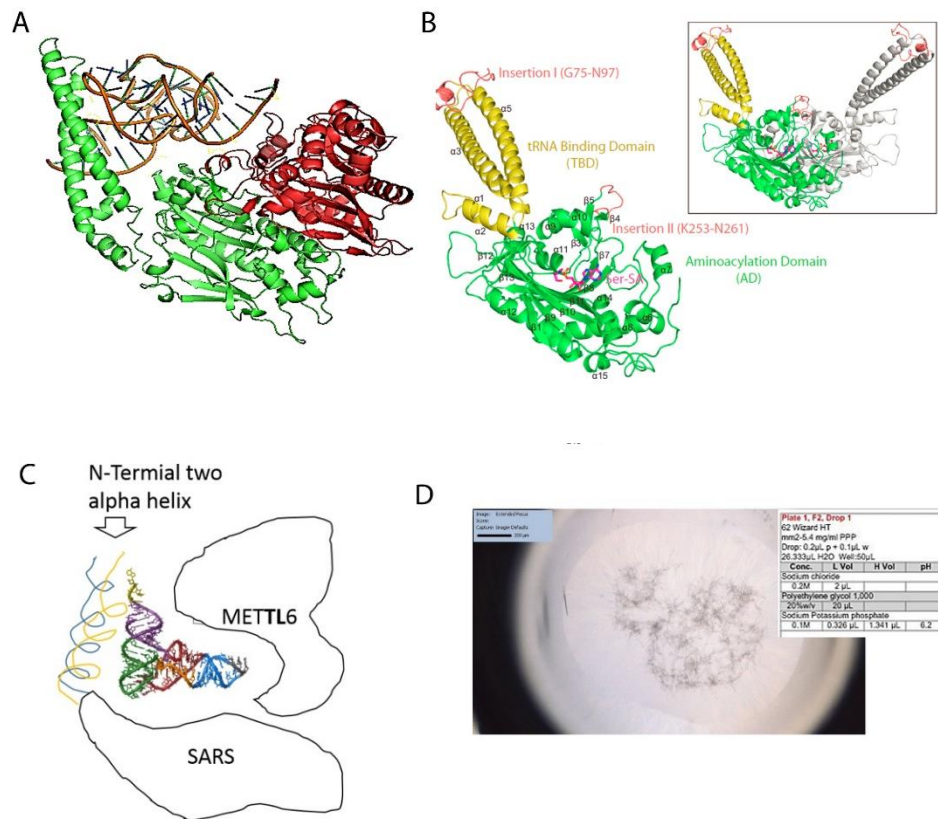


Figure 4-4 Seryl-tRNA synthetase structures of *T. Thermophilus* and *homo sapiens*

(A) Seryl-tRNA synthetase of *T. Thermophilus* in complex with tRNA<sup>Ser</sup> (PDB: 1SER), With tRNA phosphate sugar backbone in orange and bases in blue; two subunits of SARS are shown in Green and Red, respectively, Figures are made by Pymol;(notice the incomplete tRNA molecules)

(B) Human SARS structure without tRNA, (PDB: 4L87), Adapted from (172).

(C) Proposed interaction model for METTL6 and SARS. The two alpha helix are relocated at the N-terminal of SARS.

(D) Some small crystals and many precipitates have been observed in one buffer condition screened.

#### 4.9 Future work on m<sup>3</sup>C and codon biased translation

3-methylcytosine has been proposed to have a role in codon biased translation for certain stress related transcripts (117)

In order to take a glimpse of threonine codon percentage in human transcriptome, and prepare for further ribosome profiling analysis. Codon percentage across human transcriptome is calculated. In short, ref\_seq mRNAs were downloaded from USCS, excluding 5' UTR and 3' UTR. Then codon counting for each mRNA for each 61 codons (excluding stop codons) were calculated using Python scripts as written below and analysed using Excel. Each ref\_seq entry was ranked by percentage of one of the codons, thus a codon percentage ranking for each 61 codons were generated. The Python scripts are presented in section 2.21, and Threonine codon percentage Ranking is presented Table 4-1.

In general, the membrane proteins, such as mucin family proteins has the highest threonine percentage. It is similar to the data from yeast(117). Of note,

Table 4-1 Threonine codon percentage Ranking (content >16%) in human mRNAs

	Threonine codons %	Ref_seq Accession	
1	33.65	NM_001198815	mucin 22 (MUC22)
2	24.05	NM_001040105	mucin 17, cell surface associated
3	23.94	NM_005960	mucin 3A
4	22.36	NM_001164462	mucin 12, cell surface associated; similar to mucin 11
5	21.64	NM_001127710	proteoglycan 4
6	21.11	NM_001127709	proteoglycan 4
7	20.98	NM_002458	mucin 5B, oligomeric mucus/gel-forming
8	20.78	NM_001303232	proteoglycan 4 (PRG4), transcript variant E
9	20.67	NM_001127708	proteoglycan 4
10	20.21	NM_005807	proteoglycan 4
11	19.4	NM_018406	mucin 4, cell surface associated
12	19.22	NM_001010909	mucin 21, cell surface associated
13	18.67	NM_080870	diffuse panbronchiolitis critical region 1
14	18.66	NM_001010909	mucin 21, cell surface associated
15	18.6	NM_001129895	similar to HGC6.3
16	18.08	NM_001099414	hepatitis A virus cellular receptor 1

17	18.08	NM_012206	hepatitis A virus cellular receptor 1
18	18.08	NM_001173393	hepatitis A virus cellular receptor 1, transcript variant 3
19	17.72	NM_024690	mucin 16, cell surface associated
20	17.66	NM_001308156	hepatitis A virus cellular receptor 1
21	17.58	NM_058173	mucin-like 1
22	17.22	NM_182832	placenta-specific 4
23	17.11	NM_022338	chromosome 11 open reading frame 24
24	16.93	NM_005961	mucin 6, oligomeric mucus/gel-forming
25	16.84	NM_001142403	CD164 molecule, sialomucin
26	16.75	NM_052932	transmembrane protein 123
27	16.57	NM_001080400	KIAA1881
28	16.46	NM_001142404	CD164 molecule, sialomucin
29	16.4	NM_001145006	mucin 7, secreted
30	16.22	NM_003006	selectin P ligand
31	16.16	NM_006016	CD164 molecule, sialomucin
32	16.03	NM_016242	endomucin

Importantly, three threonine codons: ACA, ACT, ACC's rankings follows a similar pattern to total threonine codons as in Table 4-1. However, ACG, codon specific ranking did not show a similar enrichment, the highest mRNA which harbors ACG codons only has about 5.5 % percentage and they are no longer mucin family proteins and the ranking is shown in Appendices 3.

#### **4.10 Models for m<sup>3</sup>C in translation and other cellular process**

According to the literature review and results obtained so far, Figure 4-5 is drawn to show proposed important future work to perform, Table 4-2 Table 4-2 Potential functions of 3-methylcytosine in various Cellular processes listed several potential functions and proposed assays for studying 3-methylcytosine in various cellular processes.

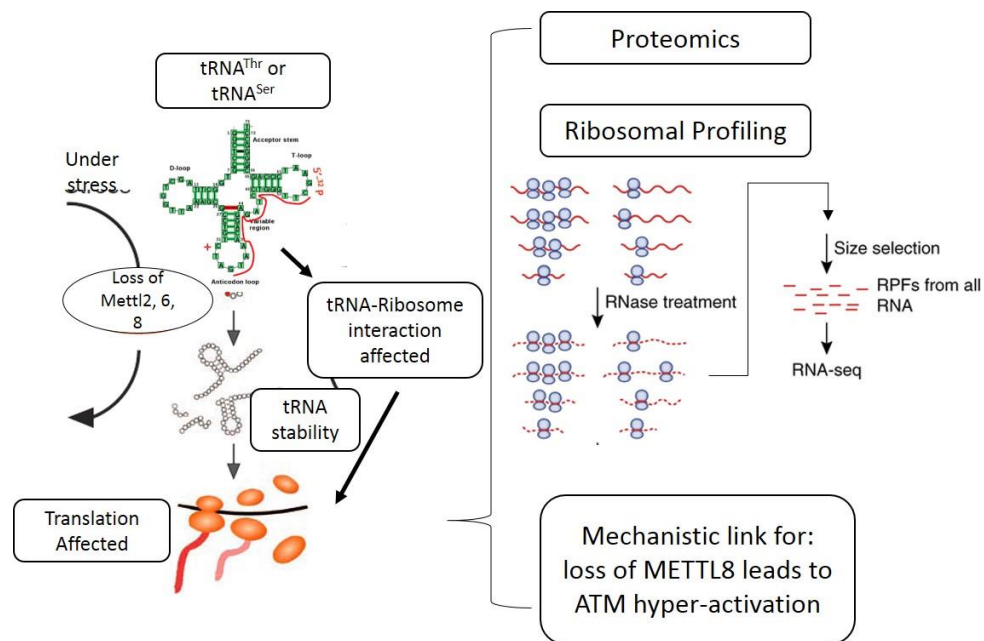


Figure 4-5 Model for possible roles of METTL2 in tRNA modification and translation

Loss of METTL2, 6, 8 may also affect ribosomal interaction with certain tRNA contains  $m^3C$  (or  $m^2, ^2G$ ), which is being further characterized by ribosomal profiling and proteomics. Loss of  $m^3C$  may also affect tRNA stability, which should be studied by Northern blot.

Table 4-2 Potential functions of 3-methylcytosine in various Cellular processes

Cellular process	Aspects	Assay
Translation	Rates	Polyribosome profiling; $^{35}S$ incorporation
	Fidelity	Ribo-seq
	Frameshifting	Reporter assay; Ribo-seq;
tRNA/rRNA itself	Maturation/ half-life/ cleavage	North blot (cleavage)
Immune system	w/wo modification affecting Retrovirus RT process(206-208)	Primer extension;
		Retrovirus infection and titer determination
Under various stress	amino acid starvation; overdosing(with aaRS); oxidation; elevated temperature	Usually followed by MTS assay

## Bibliography

1. R. D. Hotchkiss, The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* **175**, 315-332 (1948).
2. G. R. Wyatt, Occurrence of 5-methylcytosine in nucleic acids. *Nature* **166**, 237-238 (1950); published online EpubAug 5 (
3. W. E. Cohn, E. Volkin, Nucleoside-5'-Phosphates from Ribonucleic Acid. *Nature* **167**, 483-484 (1951).
4. F. F. Davis, F. W. Allen, Ribonucleic Acids from Yeast Which Contain a 5th Nucleotide. *J Biol Chem* **227**, 907-915 (1957).
5. R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, Marquise.M, S. H. Merrill, J. R. Penswick, A. Zamir, Structure of a Ribonucleic Acid. *Science* **147**, 1462-& (1965).
6. F. H. C. Crick, Codon-Anticodon Pairing - Wobble Hypothesis. *J Mol Biol* **19**, 548-& (1966).
7. M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic acids research* **41**, D262-267 (2013); published online EpubJan (10.1093/nar/gks1007).
8. B. El Yacoubi, M. Bailly, V. De Crécy-Lagard, in *Annual Review of Genetics*. (2012), vol. 46, pp. 69-95.
9. P. A. Kanerva, P. H. Maenpaa, Codon-specific serine transfer ribonucleic acid degradation in avian liver during vitellogenin induction. *ACTA CHEM. SCAND.* **35**, 379-385 (1981).
10. M. Nwagwu, M. Nana, Ribonucleic acid synthesis in embryonic chick muscle, rates of synthesis and half-lives of transfer and ribosomal RNA species. *Journal of Embryology and Experimental Morphology* **Vol. 56**, 253-267 (1980).
11. U. Karnahl, C. Wasternack, Half-life of cytoplasmic rRNA and tRNA, of plastid rRNA and of uridine nucleotides in heterotrophically and photoorganotrophically grown cells of *Euglena gracilis* and its apoplast mutant W3BUL. *International Journal of Biochemistry* **24**, 493-497 (1992)10.1016/0020-711X(92)90044-2).
12. J. Anderson, L. Phan, R. Cuesta, B. A. Carlson, M. Pak, K. Asano, G. R. Björk, M. Tamame, A. G. Hinnebusch, The essential Gcd10p-Gcd14p nuclear complex 15 required for 1- methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes and Development* **12**, 3650-3662 (1998).
13. F. Tuorto, R. Liebers, T. Musch, M. Schaefer, S. Hofmann, S. Kellner, M. Frye, M. Helm, G. Stoecklin, F. Lyko, RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nature structural & molecular biology* **19**, 900-905 (2012); published online EpubSep (10.1038/nsmb.2357).

14. E. M. Phizicky, A. K. Hopper, tRNA biology charges to the front. *Genes and Development* **24**, 1832-1860 (2010)10.1101/gad.1956510).
15. W. A. Decatur, M. J. Fournier, rRNA modifications and ribosome function. *Trends in Biochemical Sciences* **27**, 344-351 (2002)10.1016/S0968-0004(02)02109-6).
16. D. R. Davis, Stabilization of RNA stacking by pseudouridine. *Nucleic acids research* **23**, 5020-5026 (1995).
17. G. Kawai, Y. Yamamoto, T. Kamimura, T. Masegi, M. Sekine, T. Hata, T. Iimori, T. Watanabe, T. Miyazawa, S. Yokoyama, Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry®* **31**, 1040-1046 (1992).
18. D. D. Piekna-Przybylska, W. A. Decatur, M. J. Fournier, The 3D rRNA modification maps database: With interactive tools for ribosome analysis. *Nucleic acids research* **36**, D178-D183 (2008)10.1093/nar/gkm855).
19. S. F. Mitchell, S. E. Walker, M. A. Algire, E. H. Park, A. G. Hinnebusch, J. R. Lorsch, The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. *Molecular Cell* **39**, 950-962 (2010)10.1016/j.molcel.2010.08.021).
20. D. Dominissini, S. Nachtergaele, S. Moshitch-Moshkovitz, E. Peer, N. Kol, M. S. Ben-Haim, Q. Dai, A. Di Segni, M. Salmon-Divon, W. C. Clark, G. Zheng, T. Pan, O. Solomon, E. Eyal, V. Hershkovitz, D. Han, L. C. Dore, N. Amariglio, G. Rechavi, C. He, The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441-446 (2016); published online EpubFeb 25 (10.1038/nature16998).
21. X. Li, X. Xiong, K. Wang, L. Wang, X. Shu, S. Ma, C. Yi, Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nature chemical biology* **12**, 311-316 (2016); published online EpubMay (10.1038/nchembio.2040).
22. R. Desrosiers, K. Friderici, F. Rottman, Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 3971-3975 (1974).
23. R. C. Desrosiers, K. H. Friderici, F. M. Rottman, Characterization of novikoff hepatoma mRNA methylation and heterogeneity in the methylated 5' terminus. *Biochemistry* **14**, 4367-4374 (1975).
24. X. L. Ping, B. F. Sun, L. Wang, W. Xiao, X. Yang, W. J. Wang, S. Adhikari, Y. Shi, Y. Lv, Y. S. Chen, X. Zhao, A. Li, Y. Yang, U. Dahal, X. M. Lou, X. Liu, J. Huang, W. P. Yuan, X. F. Zhu, T. Cheng, Y. L. Zhao, X. Wang, J. M. R. Danielsen, F. Liu, Y. G. Yang, Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Research* **24**, 177-189 (2014)10.1038/cr.2014.3).
25. J. Liu, Y. Yue, D. Han, X. Wang, Y. Fu, L. Zhang, G. Jia, M. Yu, Z. Lu, X. Deng, Q. Dai, W. Chen, C. He, A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine



- methylation. *Nature chemical biology* **10**, 93-95 (2014)10.1038/nchembio.1432).
26. J. C. Shah, M. J. Clancy, IME4, a gene that mediates MAT and nutritional control of meiosis in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **12**, 1078-1086 (1992).
  27. C. F. Hongay, T. L. Orr-Weaver, *Drosophila* inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 14855-14860 (2011)10.1073/pnas.1111577108).
  28. S. Zhong, H. Li, Z. Bodi, J. Button, L. Vespa, M. Herzog, R. G. Fray, MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *The Plant cell* **20**, 1278-1288 (2008); published online EpubMay (10.1105/tpc.108.058883).
  29. Y. Wang, Y. Li, J. I. Toth, M. D. Petroski, Z. Zhang, J. C. Zhao, N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nature cell biology* **16**, 191-198 (2014); published online EpubFeb (10.1038/ncb2902).
  30. J. E. Squires, H. R. Patel, M. Nousch, T. Sibbritt, D. T. Humphreys, B. J. Parker, C. M. Suter, T. Preiss, Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic acids research* **40**, 5023-5033 (2012)10.1093/nar/gks144).
  31. V. Khoddami, B. R. Cairns, Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nature Biotechnology* **31**, 458-464 (2013)10.1038/nbt.2566).
  32. J. L. Rinn, H. Y. Chang, in *Annual Review of Biochemistry*. (2012), vol. 81, pp. 145-166.
  33. D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, G. Rechavi, Topology of the human and mouse m<sup>6</sup>A RNA methylomes revealed by m<sup>6</sup>A-seq. *Nature* **484**, 201-206 (2012)10.1038/nature11112).
  34. K. D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason, S. R. Jaffrey, Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635-1646 (2012); published online EpubJun 22 (10.1016/j.cell.2012.05.003).
  35. N. Liu, M. Parisien, Q. Dai, G. Zheng, C. He, T. Pan, Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19**, 1848-1856 (2013)10.1261/rna.041178.113).
  36. F. F. Davis, F. W. Allen, Ribonucleic acids from yeast which contain a fifth nucleotide. *The Journal of biological chemistry* **227**, 907-915 (1957).
  37. J. Karjolich, Y. T. Yu, Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **474**, 395-399 (2011)10.1038/nature10165).
  38. S. Schwartz, D. A. Bernstein, M. R. Mumbach, M. Jovanovic, R. H. Herbst, B. X. Leon-Ricardo, J. M. Engreitz, M. Guttman, R. Satija, E. S. Lander, G. Fink, A. Regev, Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and

- mRNA. *Cell* **159**, 148-162 (2014); published online EpubSep 25 (10.1016/j.cell.2014.08.028).
39. T. M. Carlile, M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoli, W. V. Gilbert, Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143-146 (2014)10.1038/nature13802).
  40. M. B. Shapiro, C. R. Merrill, D. F. Bradley, J. E. Mosimann, R. W. Holley, Reconstruction of Protein and Nucleic Acid Sequences - Alanine Transfer Ribonucleic Acid. *Science* **150**, 918-& (1965).
  41. A. Weixlbaumer, F. V. t. Murphy, A. Dziergowska, A. Malkiewicz, F. A. Vendeix, P. F. Agris, V. Ramakrishnan, Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nature structural & molecular biology* **14**, 498-502 (2007); published online EpubJun (10.1038/nsmb1242).
  42. A. D. Judge, G. Bola, A. C. H. Lee, I. MacLachlan, Design of noninflammatory synthetic siRNA mediating potent gene silencing in vivo. *Molecular Therapy* **13**, 494-505 (2006)10.1016/j.ymthe.2005.11.002).
  43. P. F. Agris, The importance of being modified: roles of modified nucleosides and Mg<sup>2+</sup> in RNA structure and function. *Progress in nucleic acid research and molecular biology* **53**, 79-129 (1996).
  44. S. Kellner, J. Burhenne, M. Helm, Detection of RNA modifications. *RNA Biology* **7**, 237-247 (2010)10.4161/rna.7.2.11468).
  45. Y. Kirino, Y. I. Goto, Y. Campos, J. Arenas, T. Suzuki, Specific correlation between the wobble modification deficiency in mutant tRNAs and the clinical features of a human mitochondrial disease. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7127-7132 (2005)10.1073/pnas.0500563102).
  46. S. Edelheit, S. Schwartz, M. R. Mumbach, O. Wurtzel, R. Sorek, Transcriptome-Wide Mapping of 5-methylcytidine RNA Modifications in Bacteria, Archaea, and Yeast Reveals m<sup>5</sup>C within Archaeal mRNAs. *PLoS Genetics* **9**, (2013)10.1371/journal.pgen.1003602).
  47. S. D'Silva, S. J. Haider, E. M. Phizicky, A domain of the actin binding protein Abp140 is the yeast methyltransferase responsible for 3-methylcytidine modification in the tRNA anti-codon loop. *RNA* **17**, 1100-1110 (2011); published online EpubJun (10.1261/rna.2652611).
  48. X. Zhao, Y. T. Yu, Detection and quantitation of RNA base modifications. *RNA* **10**, 996-1002 (2004)10.1261/rna.7110804).
  49. Y. Fu, G. Z. Luo, K. Chen, X. Deng, M. Yu, D. Han, Z. Hao, J. Liu, X. Lu, L. C. Dore, X. Weng, Q. Ji, L. Mets, C. He, N(6)-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. *Cell* **161**, 879-892 (2015); published online EpubMay 7 (10.1016/j.cell.2015.04.010).
  50. H. Heyn, M. Esteller, An Adenine Code for DNA: A Second Life for N<sup>6</sup>-Methyladenine. *Cell*, (2015)10.1016/j.cell.2015.04.021).
  51. G. Fang, D. Munera, D. I. Friedman, A. Mandlik, M. C. Chao, O. Banerjee, Z. X. Feng, B. Losic, M. C. Mahajan, O. J. Jabado, G. Deikus, T. A. Clark, K. Luong, I. A. Murray, B. M. Davis, A. Keren-Paz, A. Chess, R. J. Roberts, J. Korlach, S. W. Turner, V. Kumar, M. K. Waldor, E. E. Schadt, Genome-wide mapping of methylated

- adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nature Biotechnology* **30**, 1232-+ (2012); published online EpubDec (Doi 10.1038/Nbt.2432).
52. T. A. Clark, I. A. Murray, R. D. Morgan, A. O. Kislyuk, K. E. Spittle, M. Boitano, A. Fomenkov, R. J. Roberts, J. Korlach, Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic acids research* **40**, (2012); published online EpubFeb (ARTN e29 DOI 10.1093/nar/gkr1146).
  53. C. W. Gehrke, K. C. Kuo, Ribonucleoside analysis by reversed-phase high-performance liquid chromatography. *Journal of Chromatography A* **471**, 3-36 (1989)10.1016/S0021-9673(00)94152-9).
  54. S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-1303 (2011)10.1126/science.1210597).
  55. E. L. Greer, M. A. Blanco, L. Gu, E. Sendinc, J. Liu, D. Aristizabal-Corrales, C. H. Hsu, L. Aravind, C. He, Y. Shi, DNA Methylation on N(6)-Adenine in *C. elegans*. *Cell* **161**, 868-878 (2015); published online EpubMay 7 (10.1016/j.cell.2015.04.005).
  56. G. Zhang, H. Huang, D. Liu, Y. Cheng, X. Liu, W. Zhang, R. Yin, D. Zhang, P. Zhang, J. Liu, C. Li, B. Liu, Y. Luo, Y. Zhu, N. Zhang, S. He, C. He, H. Wang, D. Chen, N(6)-methyladenine DNA modification in *Drosophila*. *Cell* **161**, 893-906 (2015); published online EpubMay 7 (10.1016/j.cell.2015.04.018).
  57. A. M. Maxam, W. Gilbert, A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 560-564 (1977).
  58. B. S. Watson, T. L. Hazlett, J. F. Eccleston, C. Davis, D. M. Jameson, A. E. Johnson, Macromolecular arrangement in the aminoacyl-tRNA·elongation factor Tu·GTP ternary complex. A fluorescence energy transfer study. *Biochemistry* **34**, 7904-7912 (1995).
  59. J. A. Plumbridge, H. G. Baumert, M. Ehrenberg, R. Rigler, Characterisation of a new, fully active fluorescent derivative of *E. coli* tRNA(Phe). *Nucleic acids research* **8**, 827-843 (1980).
  60. C. H. Yang, D. Soll, Covalent attachment of a fluorescent group to 4 thiouridine in transfer RNA. *Journal of Biochemistry* **73**, 1243-1247 (1973).
  61. A. Bakin, J. Ofengand, Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* **32**, 9754-9762 (1993); published online EpubSep 21 (
  62. H. Inouye, S. Fuchs, M. Sela, U. Z. Littauer, Anti-inosine antibodies. *BBA Section Nucleic Acids And Protein Synthesis* **240**, 594-603 (1971)10.1016/0005-2787(71)90717-9).
  63. M. Weber, J. J. Davies, D. Wittig, E. J. Oakeley, M. Haase, W. L. Lam, D. Schubeler, Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature genetics* **37**, 853-862 (2005); published online EpubAug (10.1038/ng1598).

64. K. Itoh, S. Ishiwata, N. Ishida, M. Mizugaki, Diagnostic use of anti-modified nucleoside monoclonal antibody. *Tohoku Journal of Experimental Medicine* **168**, 329-331 (1992).
65. K. Itoh, K. Suzuki, S. Ishiwata, T. Tezuka, M. Mizugaki, T. Suzuki, Application of a recombinant Fab fragment from a phage display library for sensitive detection of a target antigen by an inhibition ELISA system. *Journal of Immunological Methods* **223**, 107-114 (1999)10.1016/S0022-1759(98)00209-9).
66. C. Jayabaskaran, Antibodies specific for 1-methylguanosine as a probe of tRNA conformation. *Biochemistry International* **8**, 561-571 (1984).
67. L. Levine, T. P. Waa'kes, L. Stolbach, Serum levels of N2,N2 dimethylguanosine and pseudouridine as determined by radioimmunoassay for patients with malignancy. *Journal of the National Cancer Institute* **54**, 341-343 (1975).
68. M. Masuda, T. Nishihira, K. Itoh, M. Mizugaki, N. Ishida, S. Mori, An immunohistochemical analysis for cancer of the esophagus using monoclonal antibodies specific for modified nucleosides. *Cancer* **72**, 3571-3578 (1993)10.1002/1097-0142(19931215)72:12<3571::AID-CNCR2820721205>3.0.CO;2-9).
69. L. Montesano, D. G. Glitz, Wheat germ cytoplasmic ribosomes. Localization of 7-methylguanosine and 6-methyladenosine by electron microscopy of immune complexes. *J Biol Chem* **263**, 4939-4944 (1988).
70. T. W. Munns, M. K. Liszewski, H. F. Sims, Characterization of antibodies specific for N6-methyladenosine and for 7-methylguanosine. *Biochemistry* **16**, 2163-2168 (1977).
71. C. Reynaud, C. Bruno, P. Boullanger, J. Grange, S. Barbesti, A. Niveleau, Monitoring of urinary excretion of modified nucleosides in cancer patients using a set of six monoclonal antibodies. *Cancer Letters* **61**, 255-262 (1992)10.1016/0304-3835(92)90296-8).
72. S. Dango, N. Mosammaparast, M. Sowa, L.-J. Xiong, F. Wu, K. Park, M. Rubin, S. Gygi, J. Harper, Y. Shi, DNA unwinding by ASCC3 helicase is coupled to ALKBH3-dependent DNA alkylation repair and cancer cell proliferation. *Mol Cell* **44**, 373-384 (2011)10.1016/j.molcel.2011.08.039).
73. H. Grosjean, R. Benne, *Modification and editing of RNA*. (ASM Press, Washington, DC, 1998).
74. H. B. R. Grosjean, *Modification and editing of RNA*. (ASM Press, Washington, DC, 1998).
75. A. C. Bishop, J. Xu, R. C. Johnson, P. Schimmel, V. de Crecy-Lagard, Identification of the tRNA-dihydrouridine synthase family. *J Biol Chem* **277**, 25090-25095 (2002); published online EpubJul 12 (10.1074/jbc.M203208200).
76. M. Sprinzl, C. Horn, M. Brown, A. Ioudovitch, S. Steinberg, Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic acids research* **26**, 148-153 (1998); published online EpubJan 1 (
77. G. L. Cantoni, The Nature of the Active Methyl Donor Formed Enzymatically from L-Methionine and Adenosinetriphosphate. *J Am Chem Soc* **74**, 2942-2943 (1952).

78. G. L. Cantoni, Methyl Donor Formed Enzymatically from L-Methionine and Adenosinetriphosphate. *Fed Proc* **11**, 330-331 (1952).
79. W. A. Loenen, S-adenosylmethionine: jack of all trades and master of everything? *Biochemical Society transactions* **34**, 330-333 (2006); published online EpubApr (10.1042/bst20060330).
80. S. J. Moat, I. F. W. McDowell, HOMOCYSTEINE IN HEALTH AND DISEASE. *Brain* **125**, 682-683 (2002); published online EpubMarch 1, 2002 (10.1093/brain/awf065).
81. J. R. Wilson, C. Jing, P. A. Walker, S. R. Martin, S. A. Howell, G. M. Blackburn, S. J. Gamblin, B. Xiao, Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell* **111**, 105-115 (2002)10.1016/S0092-8674(02)00964-9).
82. H. L. Schubert, R. M. Blumenthal, X. Cheng, Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci* **28**, 329-335 (2003); published online EpubJun (10.1016/S0968-0004(03)00090-2).
83. X. Zhang, The tRNA methyltransferase NSun2 stabilizes p16(INK4) mRNA by methylating the 3'-untranslated region of p16. *Nat Commun.* **3**, 712 (2012).
84. E. B. Fauman, R. M. Blumenthal, X. Cheng, Structure and evolution of AdoMet-dependent methyltransferases. *S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions*, 1-38 (1999).
85. A. W. Struck, M. L. Thompson, L. S. Wong, J. Micklefield, S-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *Chembiochem : a European journal of chemical biology* **13**, 2642-2655 (2012); published online EpubDec 21 (10.1002/cbic.201200556).
86. A. Bird, DNA methylation patterns and epigenetic memory. *Genes and Development* **16**, 6-21 (2002); published online Epub// (10.1101/gad.947102).
87. R. Reid, P. J. Greene, D. V. Santi, Exposition of a family of RNA m(5)C methyltransferases from searching genomic and proteomic sequences. *Nucleic acids research* **27**, 3138-3145 (1999); published online EpubAug 1 (
88. P. Z. Kozbial, A. R. Mushegian, Natural history of S-adenosylmethionine-binding proteins. *BMC Structural Biology* **5**, (2005)10.1186/1472-6807-5-19).
89. K. Goedecke, M. Pignot, R. S. Goody, A. J. Scheidig, E. Weinhold, Structure of tile N6-adenine DNA methyltransferase M·TaqI in complex with DNA and a cofactor analog. *Nature Structural Biology* **8**, 121-125 (2001)10.1038/84104).
90. T. Malone, R. M. Blumenthal, X. Cheng, Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J Mol Biol* **253**, 618-632 (1995)10.1006/jmbi.1995.0577).
91. X. Zhang, L. Zhou, X. Cheng, Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *EMBO Journal* **19**, 3509-3519 (2000).

92. Y. Huang, J. Komoto, K. Konishi, Y. Takata, H. Ogawa, T. Gomi, M. Fujioka, F. Takusagawa, Mechanisms for auto-inhibition and forced product release in glycine N-methyltransferase: Crystal structures of wild-type, mutant R175K and S-adenosylhomocysteine-bound R175K enzymes. *J Mol Biol* **298**, 149-162 (2000)10.1006/jmbi.2000.3637).
93. J. L. Martin, J. Begun, M. J. McLeish, J. M. Caine, G. L. Grunewald, Getting the adrenaline going: Crystal structure of the adrenaline-synthesizing enzyme PNMT. *Structure* **9**, 977-985 (2001)10.1016/S0969-2126(01)00662-1).
94. H. L. Schubert, R. M. Blumenthal, X. Cheng, Many paths to methyltransfer: A chronicle of convergence. *Trends in Biochemical Sciences* **28**, 329-335 (2003)10.1016/S0968-0004(03)00090-2).
95. S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, Hhal methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369 (1994)10.1016/0092-8674(94)90342-5).
96. X. Cheng, R. J. Roberts, AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic acids research* **29**, 3784-3795 (2001).
97. R. Reid, P. J. Greene, D. V. Santi, Exposition of a family of RNA m5C methyltransferases from searching genomic and proteomic sequences. *Nucleic Acids Research* **27**, 3138-3145 (1999)10.1093/nar/27.15.3138).
98. J. Vidgren, L. A. Svensson, A. Liljas, Crystal structure of catechol O-methyltransferase. *Nature* **368**, 354-358 (1994)10.1038/368354a0).
99. C. Zubieta, X. Z. He, R. A. Dixon, J. P. Noel, Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant o-methyltransferases. *Nature Structural Biology* **8**, 271-279 (2001)10.1038/85029).
100. S. Djordjevic, A. M. Stock, Crystal structure of the chemotaxis receptor methyltransferase CheR suggests a conserved structural motif for binding S-adenosylmethionine. *Structure* **5**, 545-558 (1997).
101. A. Alexandrov, M. R. Martzen, E. M. Phizicky, Two proteins that form a complex are required for 7-methylguanosine modification of yeast tRNA. *RNA* **8**, 1253-1266 (2002)10.1017/S1355838202024019).
102. J. A. Bokar, M. E. Shambaugh, D. Polayes, A. G. Matera, F. M. Rottman, Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA* **3**, 1233-1247 (1997).
103. J. Liu, Y. Yue, D. Han, X. Wang, Y. Fu, L. Zhang, G. Jia, M. Yu, Z. Lu, X. Deng, Q. Dai, W. Chen, C. He, A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nature chemical biology* **10**, 93-95 (2014); published online EpubFeb (10.1038/nchembio.1432).
104. R. S. Lipson, K. J. Webb, S. G. Clarke, Two novel methyltransferases acting upon eukaryotic elongation factor 1A in *Saccharomyces cerevisiae*. *Archives of Biochemistry and Biophysics* **500**, 137-143 (2010); published online Epub8/15/ (10.1016/j.abb.2010.05.023).
105. V. M. Richon, D. Johnston, C. J. Sneeringer, L. Jin, C. R. Majer, K. Elliston, L. F. Jerva, M. P. Scott, R. A. Copeland, Chemogenetic analysis of human protein methyltransferases. *Chemical biology &*

- drug design* **78**, 199-210 (2011); published online EpubAug (10.1111/j.1747-0285.2011.01135.x).
106. R. H. Hall, Isolation of 3-methyluridine and 3-methylcytidine from soluble ribonucleic acid. *Biochemical and Biophysical Research Communications* **12**, 361-364 (1963).
  107. Y. Iwanami, G. Brown, Methylated bases of ribosomal ribonucleic acid from HeLa cells. *Arch Biochem Biophys* **126**, 8-15 (1968)10.1016/0003-9861(68)90553-5).
  108. B. Maden, M. Salim, The methylated nucleotide sequences in HELA cell ribosomal RNA and its precursors. *J Mol Biol* **88**, 133-152 (1974)10.1016/0022-2836(74)90299-x).
  109. F. Jühling, M. Mörl, R. K. Hartmann, M. Sprinzl, P. F. Stadler, J. Pütz, tRNADB 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic acids research* **37**, D159-D162 (2009)10.1093/nar/gkn772).
  110. J. Weissenbach, I. Kiraly, G. Dirheimer, Structure primaire des tRNAThr la et b de levure de bière. *Biochimie* **59**, 381-391 (1977)10.1016/S0300-9084(77)80314-3).
  111. M. Sprinzl, K. S. Vassilenko, Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic acids research* **33**, D139-D140 (2005)10.1093/nar/gki012).
  112. M. V. Olson, G. S. Page, A. Sentenac, P. W. Piper, M. Worthington, R. B. Weiss, B. D. Hall, Only one of two closely related yeast suppressor tRNA genes contains an intervening sequence. *Nature* **291**, 464-469 (1981)10.1038/291464a0).
  113. M. Olejniczak, O. Uhlenbeck, tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* **88**, 943-950 (2006)10.1016/j.biochi.2006.06.005).
  114. F. Juhling, M. Morl, R. K. Hartmann, M. Sprinzl, P. F. Stadler, J. Putz, tRNADB 2009: compilation of tRNA sequences and tRNA genes. *Nucleic acids research* **37**, D159-162 (2009); published online EpubJan (10.1093/nar/gkn772).
  115. A. Noma, S. Yi, T. Katoh, Y. Takai, T. Suzuki, Actin-binding protein ABP140 is a methyltransferase for 3-methylcytidine at position 32 of tRNAs in *Saccharomyces cerevisiae*. *RNA* **17**, 1111-1119 (2011); published online EpubJun (10.1261/rna.2653411).
  116. M. Alamgir, V. Erukova, M. Jessulat, A. Azizi, A. Golshani, Chemical-genetic profile analysis of five inhibitory compounds in yeast. *BMC Chemical Biology* **10**, (2010)10.1186/1472-6769-10-6).
  117. C. T. Chan, W. Deng, F. Li, M. S. DeMott, I. R. Babu, T. J. Begley, P. C. Dedon, Highly Predictive Reprogramming of tRNA Modifications Is Linked to Selective Expression of Codon-Biased Genes. *Chemical research in toxicology*, (2015); published online EpubApr 13 (10.1021/acs.chemrestox.5b00004).
  118. C. T. Chan, W. Pang Yi Fau - Deng, I. R. Deng W Fau - Babu, M. Babu Ir Fau - Dyavaiah, T. J. Dyavaiah M Fau - Begley, P. C. Begley Tj Fau - Dedon, P. C. Dedon, Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. published online Epub20120704 DCOM- 20121102 (D - nlm: nihms425607)

- D - NLM: PMC3535174 EDAT- 2012/07/05 06:00 MHDA- 2012/11/03 06:00 CRDT- 2012/07/05 06:00 PHST- 2011/12/14 [received] PHST- 2012/05/31 [accepted] AID - ncomms1938 [pii] AID - 10.1038/ncomms1938 [doi] PST - epublish).
119. U. Begley, M. Dyavaiah, A. Patil, J. P. Rooney, D. DiRenzo, C. M. Young, D. S. Conklin, R. S. Zitomer, T. J. Begley, Trm9-catalyzed tRNA modifications link translation to the DNA damage response. *Mol Cell* **28**, 860-870 (2007)10.1016/j.molcel.2007.09.021).
  120. R. P. Fahlman, M. Olejniczak, O. C. Uhlenbeck, Quantitative analysis of deoxynucleotide substitutions in the codon-anticodon helix. *J Mol Biol* **355**, 887-892 (2006)10.1016/j.jmb.2005.11.011).
  121. P. F. Agris, Decoding the genome: A modified view. *Nucleic acids research* **32**, 223-238 (2004)10.1093/nar/gkh185).
  122. M. I. Baldi, E. Mattoccia, E. Bufardecì, S. Fabbri, G. P. Tocchini-Valentini, Participation of the intron in the reaction catalyzed by the *Xenopus* tRNA splicing endonuclease. *Science* **255**, 1404-1408 (1992).
  123. R. Giege, M. Sissler, C. Florentz, Universal rules and idiosyncratic features in tRNA identity. *Nucleic acids research* **26**, 5017-5035 (1998); published online EpubNov 15 (
  124. M. Sprinzl, C. Steegborn, F. Hubel, S. Steinberg, Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic acids research* **24**, 68-72 (1996); published online EpubJan 1 (
  125. P. Auffinger, E. Westhof, Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *Journal of Molecular Biology* **292**, 467-483 (1999)10.1006/jmbi.1999.3080).
  126. M. Olejniczak, T. Dale, R. P. Fahlman, O. C. Uhlenbeck, Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding. *Nature Structural and Molecular Biology* **12**, 788-793 (2005)10.1038/nsmb978).
  127. R. P. Fahlman, T. Dale, O. C. Uhlenbeck, Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Molecular Cell* **16**, 799-805 (2004)10.1016/j.molcel.2004.10.030).
  128. R. P. Fahlman, O. C. Uhlenbeck, Contribution of the esterified amino acid to the binding of aminoacylated tRNAs to the ribosomal P- and A-sites. *Biochemistry* **43**, 7575-7583 (2004)10.1021/bi0495836).
  129. F. Lustig, T. Borén, C. Claesson, C. Simonsson, M. Barciszewska, U. Lagerkvist, The nucleotide in position 32 of the tRNA anticodon loop determines ability of anticodon UCC to discriminate among glycine codons. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 3343-3347 (1993).
  130. A. Korostelev, H. Noller, The ribosome in focus: new structures bring new insights. *Trends Biochem Sci* **32**, 434-441 (2007)10.1016/j.tibs.2007.08.002).
  131. S. Näsval, K. Nilsson, G. Björk, The ribosomal grip of the peptidyl-tRNA is critical for reading frame maintenance. *J Mol Biol* **385**, 350-367 (2009)10.1016/j.jmb.2008.10.069).
  132. A. Korostelev, H. F. Noller, The ribosome in focus: new structures bring new insights. *Trends in Biochemical Sciences* **32**, 434-441 (2007)10.1016/j.tibs.2007.08.002).



133. L. A. Raftery, M. Yarus, Systematic alterations in the anticodon arm make tRNA(Glu)-Suoc a more efficient suppressor. *The EMBO journal* **6**, 1499-1506 (1987); published online EpubMay (
134. M. Yarus, S. Cline, L. Raftery, P. Wier, D. Bradley, The translational efficiency of tRNA is a property of the anticodon arm. *J Biol Chem* **261**, 10496-10505 (1986); published online EpubAug 15 (
135. F. Lustig, T. Boren, Y. S. Guindy, P. Elias, T. Samuelsson, C. W. Gehrke, K. C. Kuo, U. Lagerkvist, Codon discrimination and anticodon structural context. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 6873-6877 (1989).
136. C. Claesson, F. Lustig, T. Boren, C. Simonsson, M. Barciszewska, U. Lagerkvist, Glycine codon discrimination and the nucleotide in position 32 of the anticodon loop. *J Mol Biol* **247**, 191-196 (1995)10.1006/jmbi.1994.0132).
137. L. A. Raftery, M. Yarus, Systematic alterations in the anticodon arm make tRNA(Glu)-Suoc a more efficient suppressor. *EMBO Journal* **6**, 1499-1506 (1987).
138. D. Smith, L. Breeden, E. Farrell, M. Yarus, The bases of the tRNA anticodon loop are independent by genetic criteria. *Nucleic acids research* **15**, 4669-4686 (1987)10.1093/nar/15.11.4669).
139. J. M. Ogle, D. E. Brodersen, Clemons W.M, Jr., M. J. Tarry, A. P. Carter, V. Ramakrishnan, Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897-902 (2001)10.1126/science.1060612).
140. D. Moazed, H. F. Noller, Binding of tRNA to the ribosomal A and P sites protects two distinct sets of nucleotides in 16 S rRNA. *J Mol Biol* **211**, 135-145 (1990)10.1016/0022-2836(90)90016-F).
141. J. M. Ogle, F. V. Murphy Iv, M. J. Tarry, V. Ramakrishnan, Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111**, 721-732 (2002)10.1016/S0092-8674(02)01086-3).
142. S. S. Phelps, S. Joseph, D. E. Draper, Non-bridging phosphate oxygen atoms within the tRNA anticodon stem-loop are essential for ribosomal a site binding and translocation. *J Mol Biol* **349**, 288-301 (2005)10.1016/j.jmb.2005.03.079).
143. M. Yarus, Translational efficiency of transfer RNA's: Uses of an extended anticodon. *Science* **218**, 646-652 (1982).
144. T. Dale, O. C. Uhlenbeck, Amino acid specificity in translation. *Trends in Biochemical Sciences* **30**, 659-665 (2005)10.1016/j.tibs.2005.10.006).
145. K. B. Gromadski, M. V. Rodnina, Kinetic Determinants of High-Fidelity tRNA Discrimination on the Ribosome. *Molecular Cell* **13**, 191-200 (2004)10.1016/S1097-2765(04)00005-X).
146. T. Pape, W. Wintermeyer, M. V. Rodnina, Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E.coli ribosome. *EMBO Journal* **17**, 7490-7497 (1998)10.1093/emboj/17.24.7490).
147. P. O. Brown, D. Botstein, Exploring the new world of the genome with DNA microarrays. *Nature genetics* **21**, 33-37 (1999)10.1038/4462).

148. D. Baek, J. Villén, C. Shin, F. D. Camargo, S. P. Gygi, D. P. Bartel, The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008)10.1038/nature07242).
149. M. Selbach, B. Schwanhäusser, N. Thierfelder, Z. Fang, R. Khanin, N. Rajewsky, Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58-63 (2008)10.1038/nature07228).
150. N. Sonenberg, A. G. Hinnebusch, New Modes of Translational Control in Development, Behavior, and Disease. *Molecular Cell* **28**, 721-729 (2007)10.1016/j.molcel.2007.11.018).
151. O. Namy, J. P. Rousset, S. Naphthine, I. Brierley, Reprogrammed Genetic Decoding in Cellular Gene Expression. *Molecular Cell* **13**, 157-168 (2004)10.1016/S1097-2765(04)00031-0).
152. C. Touriol, S. Bornes, S. Bonnal, S. Audigier, H. Prats, A. C. Prats, S. Vagner, Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biology of the Cell* **95**, 169-178 (2003)10.1016/S0248-4900(03)00033-9).
153. A. A. Komar, T. Lesnik, C. Reiss, Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Letters* **462**, 387-391 (1999)10.1016/S0014-5793(99)01566-5).
154. A. K. K. Lakkaraju, C. Mary, A. Scherrer, A. E. Johnson, K. Strub, SRP Keeps Polypeptides Translocation-Competent by Slowing Translation to Match Limiting ER-Targeting Sites. *Cell* **133**, 440-451 (2008)10.1016/j.cell.2008.02.049).
155. S. L. Wolin, P. Walter, Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO Journal* **7**, 3559-3569 (1988).
156. Y. Arava, Y. Wang, J. D. Storey, C. L. Liu, P. O. Brown, D. Herschlag, Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3889-3894 (2003)10.1073/pnas.0635171100).
157. N. T. Ingolia, Ribosome profiling: New views of translation, from single codons to genome scale. *Nature Reviews Genetics* **15**, 205-213 (2014)10.1038/nrg3645).
158. N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-223 (2009)10.1126/science.1168978).
159. J. A. Steitz, Polypeptide chain initiation: Nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature* **224**, 957-964 (1969)10.1038/224957a0).
160. M. Kozak, How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109-1123 (1978); published online EpubDec (
161. R. Jackson, N. Standart, The awesome power of ribosome profiling. *RNA* **21**, 652-654 (2015); published online EpubApr (10.1261/rna.049908.115).
162. P. Xie, F. Y. Wei, S. Hirata, T. Kaitsuka, T. Suzuki, T. Suzuki, K. Tomizawa, Quantitative PCR measurement of tRNA 2-methylthio modification for assessing type 2 diabetes risk. *Clinical chemistry* **59**,

- 1604-1612 (2013); published online EpubNov (10.1373/clinchem.2013.210401).
163. H. Tsurui, Y. Kumazawa, R. Sanokawa, Y. Watanabe, T. Kuroda, A. Wada, K. Watanabe, T. Shirai, Batchwise purification of specific tRNAs by a solid-phase DNA probe. *Analytical biochemistry* **221**, 166-172 (1994); published online EpubAug 15 (10.1006/abio.1994.1393).
  164. Y. H. Chionh, C. H. Ho, D. Pruksakorn, I. Ramesh Babu, C. S. Ng, F. Hia, M. E. McBee, D. Su, Y. L. Pang, C. Gu, H. Dong, E. G. Prestwich, P. Y. Shi, P. R. Preiser, S. Alonso, P. C. Dedon, A multidimensional platform for the purification of non-coding RNA species. *Nucleic acids research* **41**, e168 (2013); published online EpubSep (10.1093/nar/gkt668).
  165. The general mode of translation inhibition by macrolide antibiotics. *Proceedings of the National Academy of Sciences* **111**, 15958-15963 (2014)10.1073/pnas.1417334111).
  166. A. M. Khalil, M. Guttman, M. Huarte, M. Garber, A. Raj, D. Rivea Morales, K. Thomas, A. Presser, B. E. Bernstein, A. van Oudenaarden, A. Regev, E. S. Lander, J. L. Rinn, Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* **106**, 11667-11672 (2009); published online EpubJul 14 (10.1073/pnas.0904715106).
  167. J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, H. Y. Chang, Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323 (2007); published online EpubJun 29 (10.1016/j.cell.2007.05.022).
  168. N. A. Franken, H. M. Rodermond, J. Stap, J. Haveman, C. van Bree, Clonogenic assay of cells in vitro. *Nature protocols* **1**, 2315-2319 (2006)10.1038/nprot.2006.339).
  169. K. R. Badri, Y. Zhou, U. Dhru, S. Aramgam, L. Schuger, Effects of the SANT domain of tension-induced/inhibited proteins (TIPs), novel partners of the histone acetyltransferase p300, on p300 activity and TIP-6-induced adipogenesis. *Molecular and cellular biology* **28**, 6358-6372 (2008); published online EpubOct (10.1128/MCB.00333-08).
  170. S. Jakkaraju, X. Zhe, D. Pan, R. Choudhury, L. Schuger, TIPs are tension-responsive proteins involved in myogenic versus adipogenic differentiation. *Developmental cell* **9**, 39-49 (2005); published online EpubJul (10.1016/j.devcel.2005.04.015).
  171. P. P. Chan, T. M. Lowe, GtRNADB: a database of transfer RNA genes detected in genomic sequence. *Nucleic acids research* **37**, D93-D97 (2009); published online EpubJan (
  172. X. Xu, Y. Shi, X. L. Yang, Crystal structure of human Seryl-tRNA synthetase and Ser-SA complex reveals a molecular lever specific to higher eukaryotes. *Structure* **21**, 2078-2086 (2013); published online EpubNov 5 (10.1016/j.str.2013.08.021).
  173. X. Xu, Y. Shi, H. M. Zhang, E. C. Swindell, A. G. Marshall, M. Guo, S. Kishi, X. L. Yang, Unique domain appended to vertebrate tRNA synthetase is essential for vascular development. *Nature communications* **3**, 681 (2012)10.1038/ncomms1686).

174. L. Slabinski, L. Jaroszewski, L. Rychlewski, I. A. Wilson, S. A. Lesley, A. Godzik, XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403-3405 (2007); published online EpubDec 15 (10.1093/bioinformatics/btm477).
175. M. Sickmeier, J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic, A. K. Dunker, DisProt: the Database of Disordered Proteins. *Nucleic acids research* **35**, D786-793 (2007); published online EpubJan (10.1093/nar/gkl893).
176. G. Zheng, Y. Qin, W. C. Clark, Q. Dai, C. Yi, C. He, A. M. Lambowitz, T. Pan, Efficient and quantitative high-throughput tRNA sequencing. *Nature methods* **12**, 835-837 (2015); published online EpubSep (10.1038/nmeth.3478).
177. A. E. Cozen, E. Quartley, A. D. Holmes, E. Hrabeta-Robinson, E. M. Phizicky, T. M. Lowe, ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nature methods* **12**, 879-884 (2015); published online EpubSep (10.1038/nmeth.3508).
178. D. Su, C. T. Chan, C. Gu, K. S. Lim, Y. H. Chionh, M. E. McBee, B. S. Russell, I. R. Babu, T. J. Begley, P. C. Dedon, Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nature protocols* **9**, 828-841 (2014); published online EpubApr (10.1038/nprot.2014.047).
179. C. T. Chan, M. Dyavaiah, M. S. DeMott, K. Taghizadeh, P. C. Dedon, T. J. Begley, A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS genetics* **6**, e1001247 (2010); published online EpubDec (10.1371/journal.pgen.1001247).
180. X. Wang, C. He, Reading RNA methylation codes through methyl-specific binding proteins. *RNA Biology* **11**, (2014)10.4161/rna.28829).
181. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57 (2009).
182. M. G. Goll, F. Kirpekar, K. A. Maggert, J. A. Yoder, C. L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, T. H. Bestor, Methylation of tRNA<sup>Asp</sup> by the DNA methyltransferase homolog Dnmt2. *Science* **311**, 395-398 (2006)10.1126/science.1120976).
183. D. Fu, J. A. N. Brophy, C. T. Y. Chan, K. A. Atmore, U. Begley, R. S. Paules, P. C. Dedon, T. J. Begley, L. D. Samson, Human AlkB homolog ABH8 is a tRNA methyltransferase required for wobble uridine modification and DNA damage survival. *Molecular and cellular biology* **30**, 2449-2459 (2010)10.1128/MCB.01604-09).
184. U. Begley, M. Dyavaiah, A. Patil, J. P. Rooney, D. DiRenzo, C. M. Young, D. S. Conklin, R. S. Zitomer, T. J. Begley, Trm9-catalyzed tRNA modifications link translation to the DNA damage response. *Mol Cell* **28**, 860-870 (2007); published online EpubDec 14 (10.1016/j.molcel.2007.09.021).
185. R. Denman, J. Colgan, K. Nurse, J. Ofengand, Crosslinking of the anticodon of P site bound tRNA to C-1400 of E.coli 16S RNA does

- not require the participation of the 50S subunit. *Nucleic acids research* **16**, 165-178 (1988); published online EpubJan 11 (
186. S. Haag, A. S. Warda, J. Kretschmer, M. A. Gunnigmann, C. Hobartner, M. T. Bohnsack, NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. *RNA*, (2015); published online EpubJul 9 (10.1261/rna.051524.115).
  187. O. Larsson, N. Sonenberg, R. Nadon, anota: Analysis of differential translation in genome-wide studies. *Bioinformatics* **27**, 1440-1441 (2011); published online EpubMay 15 (10.1093/bioinformatics/btr146).
  188. O. Larsson, N. Sonenberg, R. Nadon, Identification of differential translation in genome wide studies. *Proc Natl Acad Sci U S A* **107**, 21487-21492 (2010); published online EpubDec 14 (10.1073/pnas.1006821107).
  189. A. B. Olshen, A. C. Hsieh, C. R. Stumpf, R. A. Olshen, D. Ruggero, B. S. Taylor, Assessing gene-level translational control from ribosome profiling. *Bioinformatics* **29**, 2995-3002 (2013); published online EpubDec 1 (10.1093/bioinformatics/btt533).
  190. C. R. Stumpf, M. V. Moreno, A. B. Olshen, B. S. Taylor, D. Ruggero, The translational landscape of the mammalian cell cycle. *Mol Cell* **52**, 574-582 (2013); published online EpubNov 21 (10.1016/j.molcel.2013.09.018).
  191. T. Tebaldi, E. Dassi, G. Kostoska, G. Viero, A. Quattrone, tRanslatome: an R/Bioconductor package to portray translational control. *Bioinformatics* **30**, 289-291 (2014); published online EpubJan 15 (10.1093/bioinformatics/btt634).
  192. C. A. Piccirillo, E. Bjur, I. Topisirovic, N. Sonenberg, O. Larsson, Translational control of immune responses: from transcripts to translomes. *Nature immunology* **15**, 503-511 (2014); published online EpubJun (10.1038/ni.2891).
  193. M. Guttman, P. Russell, N. T. Ingolia, J. S. Weissman, E. S. Lander, Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240-251 (2013)10.1016/j.cell.2013.06.009).
  194. W. S. Smith, B. Nawrot, A. Malkiewicz, P. F. Agris, RNA modified uridines VI: Conformations of 3-(3-(s)-amino-3-carboxypropyl]uridine (acp3U) from tRNA and 1-methyl-3-[3-(s)-amino-3-carboxypropyl]pseudouridine (m1acp3Ψ) from rRNA. *Nucleosides and Nucleotides* **11**, 1683-1694 (1992).
  195. P. Acharya, P. Cheruku, S. Chatterjee, S. Acharya, J. Chattopadhyaya, Measurement of nucleobase pKa values in model mononucleotides shows RNA-RNA duplexes to be more stable than DNA-DNA duplexes. *Journal of the American Chemical Society* **126**, 2862-2869 (2004); published online EpubMar 10 (10.1021/ja0386546).
  196. M. Egholm, L. Christensen, K. L. Dueholm, O. Buchardt, J. Coull, P. E. Nielsen, Efficient pH-independent sequence-specific DNA binding by pseudoisocytosine-containing bis-PNA. *Nucleic acids research* **23**, 217-222 (1995); published online EpubJan 25 (

197. G. Devi, Y. Zhou, Z. Zhong, D. F. Toh, G. Chen, RNA triplexes: from structural principles to biological and biotech applications. *Wiley interdisciplinary reviews. RNA* **6**, 111-128 (2015); published online EpubJan-Feb (10.1002/wrna.1261).
198. S. A. Cassidy, P. Slickers, J. O. Trent, D. C. Capaldi, P. D. Roselt, C. B. Reese, S. Neidle, K. R. Fox, Recognition of GC base pairs by triplex forming oligonucleotides containing nucleosides derived from 2-aminopyridine. *Nucleic acids research* **25**, 4891-4898 (1997); published online EpubDec 15 (
199. T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, Y. Pilpel, An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344-354 (2010)10.1016/j.cell.2010.03.031).
200. T. Ikemura, Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J Mol Biol* **151**, 389-409 (1981)10.1016/0022-2836(81)90003-6).
201. E. M. Novoa, M. Pavon-Eternod, T. Pan, L. Ribas De Pouplana, A role for tRNA modifications in genome structure and codon usage. *Cell* **149**, 202-213 (2012)10.1016/j.cell.2012.01.050).
202. E. M. Novoa, M. Pavon-Eternod, T. Pan, L. R. de Pouplana, A Role for tRNA Modifications in Genome Structure and Codon Usage. *Cell* **149**, 202-213 (2012); published online EpubMar 30 (
203. W. contributors. (Wikipedia, The Free Encyclopedia, 2015).
204. D. R. Cooper, T. Boczek, K. Grelewska, M. Pinkowska, M. Sikorska, M. Zawadzki, Z. Derewenda, Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta crystallographica. Section D, Biological crystallography* **63**, 636-645 (2007); published online EpubMay (10.1107/S0907444907010931).
205. Y. Peng, D. Sartini, V. Pozzi, D. Wilk, M. Emanuelli, V. C. Yee, Structural basis of substrate recognition in human nicotinamide N-methyltransferase. *Biochemistry* **50**, 7800-7808 (2011); published online EpubSep 13 (10.1021/bi2007614).
206. A. N. Suhasini, R. Sirdeshmukh, Onconase action on tRNA(Lys3), the primer for HIV-1 reverse transcription. *Biochem Biophys Res Commun* **363**, 304-309 (2007); published online EpubNov 16 (10.1016/j.bbrc.2007.08.157).
207. M. Rigourd, V. Goldschmidt, F. Brule, C. D. Morrow, B. Ehresmann, C. Ehresmann, R. Marquet, Structure-function relationships of the initiation complex of HIV-1 reverse transcription: the case of mutant viruses using tRNA(His) as primer. *Nucleic acids research* **31**, 5764-5775 (2003); published online EpubOct 1 (
208. C. Isel, R. Marquet, G. Keith, C. Ehresmann, B. Ehresmann, Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. *J Biol Chem* **268**, 25269-25272 (1993); published online EpubDec 5 (



## Appendices

### Appendix 1. Plasmids generated /used in this project

	Stock	Name	Brief comments
1	376	pPyCAGIP-Flag-Mettl2-Myc	Overexpression in mammalian cell lines
2	377	pPyCAGIP-Flag-Mettl2-T2A-EGFP	Same
3	378	pPyCAGIP-Flag-hMettl2-T2A-EGFP	Same
4	497	pEF6-Flag-hMettl2-myc	Expression good in western,
5	528	pPyCAGIP-Flag-Mettl2	no myc tag at C terminal
6	683	pEYFPC1-hMettl2A	mettl2A
7	686	pPyCAGIP-Flag-hMettl2A-T2A-EGFP	mettl2A
8	688	pPyCAGIP-Flag-hMettl2A-T2A-mCherry	mettl2A
9	756	pPYCAGIP-Flag-hMettl2A-ΔSAM-T2A-	27nt deletion in SAM binding
10	519	pPyCAGIP-Flag-Mettl2-ΔSAM-Myc	27nt deletion in SAM binding
11	529	pPyCAGIP-Flag-Mettl2ΔSAM	no myc tag at C terminal
12	691	pEYFPC1-hMettl2-ΔSAM	
13	520	pPyCAGIP-Flag-Mettl2-G3A-Myc	Three glycine to Alanine
14	530	pPyCAGIP-Flag-Mettl2G3A	no myc tag at C terminal
15	534	pPyCAGIP-GST-Mettl2-G3A-Myc	GST tag
16	995	pTRE3G-Flag-hM2B-IRES-EGFP	BamHI/NdeI
17	996	pTRE3G-Flag-hM2-IRES-EGFP-BSD	
18	997	pTRE3G-Flag-hM2-ΔSAM-IRES-EGFP-	27nt deletion in frame
19	998	pTRE3G-Flag-hM2-short-IRES-EGFP-	
20	999	pTRE3G-Flag-hM2B-G3A-IRES-EGFP-	
21	1025	pPyCAGIP-GST-hSARS-1-72	SARS truncations
22	1026	pPyCAGIP-GST-hSARS-1-152	
23	1027	pPyCAGIP-GST-hSARS-1-514	
24	1028	pPyCAGIP-GST-hSARS-153-460	
25	1029	pPyCAGIP-GST-hSARS-153-514	
26	338	pGEX6P1-mMettl2	GST tag purification of mouse



27	375	pET15b-mMettl2	His tag purification mouse
28	498	pGEX6P1-Mettl2 mouse $\Delta$ SAM	SAM motif 27ny deleted
29	499	pGEX6P1-Mettl2 mouse G3A	extra mutation just before
30	856	pFASTB-Hta-huMettl2A	human
31	857	pFASTB-Hta-huMettl2A- $\Delta$ SAM	human
32	858	pET15b-huMettl2A	human
33	859	pET15b-huMettl2A- $\Delta$ SAM	human
34	1002	pet15b-hMettl2-G3A-Mutation	Very good 3 mutation
35	1003	pet15b-hMettl2-G3A-Mutation backup	Very good 3 mutation
36	1004	pet15b-hMettl2-insertion	
37	1050	pET21b-huSARS	His tag at C-Terminal
38	413	pLVx-puro-mettl2 mouse shRNA1	
39	414	pLVx-puro-mettl2 human shRNA1	1054nt on CDS
40	561	pLVx-GFP-hMettl2 shRNA1	GFP version
41	562	pLVx-GFP-hMettl2 shRNA2	
42	563	pLVx-puro-Mettl2 shRNA3	target elsewhere
43	564	pLVx-GFP-Mettl2 shRNA3	target elsewhere
44	636	pLVx-GFP-hMettl2 shRNA1	missing, LA
45	637	pLVx-GFP-hMettl2 shRNA1	
46	638	pLVx-puro-Mettl2-3'UTR-shRNA1	
47	680	pLKO puro scramble	
48	786	pLKO-puro-mettl2-shRNA1	3UTR
49	787	pLKO-puro-mettl2-shRNA2	CDS
50	788	pLKO-puro-mettl2-shRNA3	CDS
51	789	pLKO-puro-mettl2-shRNA4	CDS
52	790	pLKO-puro-mettl2-shRNA5	3UTR
53	791	pLKO-puro-hMettl2A-shRNA1	3UTR
54	792	pLKO-puro-hMettl2A-shRNA2	CDS, unique on 2A, Same as
55	793	pLKO-puro-hMettl2B-shRNA3	CDS
56	794	pLKO-puro-hMettl2B-shRNA4	CDS
57	795	pLKO-puro-hMettl2B-shRNA5	3UTR unique on 2B
58	807	pST1374-NLS-flag-linker-Cas9	Xingxu Huang lab

59	854	MLM3636-huMettl2-1	
60	855	MLM3636-huMettl2-2	
61	860	MLM3636-Mettl6-Cas9-F1+R1	Mouse M6 gRNA
62	861	MLM3636-Mettl6-Cas9-F2+R2	
63	932	MLM3636-TRM1 gRNA1	work with Mettl2 together to
64	933	MLM3636-TRM1L gRNA1	work with Mettl2 together to
65	936	pSpCas9(BB)-2A-GFP (PX458)	from DAI, GIS
66	937	pLKO-TRMT1 shRNA1	
67	938	pLKO-TRMT1 shRNA2	
68	939	pLKO-TRMT1-like shRNA1	
69	940	pLKO-TRMT1-like shRNA2	
70	946	pRGEN_Cas9-HA_CMV expression vector	Toolgen Korean
71	947	pRGEN_Mm-Mettl8_U6_SG_1	mouse M8 gRNA1
72	948	pRGEN_Mm-Mettl8_U6_SG_2	mouse M8 gRNA2
73	1055	pST1374-NLS-flag-linker-Cas9 D10A	Cas9 Nickase
74	1016	pEF6-Rluc-SerUCA5-Luc2	
75	1017	pEF6-Rluc-SerUCG5-Luc2	
76	1019	pEF6-Rluc-Luc2	For Codon Reporter
77	1020	pcDNA3.1-Rluc-Luc2	For Codon Reporter
78	1034	pEF6-Rluc-SerAGU5-Luc2	
79	1035	pEF6-Rluc-SerUCG10-Luc2	
80	1036	pEF6-Rluc-GScodon5-Luc2	
81	1037	pEF6-Rluc-CodonRunCtrl-Luc2	
82	1042	pEF6-Rluc	No Luc2
83	1101	pEF6-Rluc-ThrACA10-Luc2	
84	1102	pEF6-Rluc-ThrACT10-Luc2	
85	1103	pEF6-Rluc-ThrACC10-Luc2	
86	1104	pEF6-Rluc-ThrACG10-Luc2	

## Appendix 2. Primers excluding those listed in Materials and Methods

### Appendix 2.1 Primers for cloning;

hM2_cDNA_F1	GTGAAGATCTGGCCGGCT CCTACCCTGAAGGTG	map to human Mettl2B cDNA start with BglII overhang
hM2_cDNA_F2	CTTCGGTACCGCCACCAT GGACTACAAGGACGACGA	map to Flag tag start with KpnI, Kozak, ATG
hM2_cDNA_F3	CTTCGGTACCGCCACCAT GGCCGGCTCCTACCCTGA AGGTG	map to hMettl2A&B cDNA start with kozak, ATG, KpnI in overhang
hM2_cDNA_F4	acgcGTCGACtcGCCGGCTC CTACCCTGAAGGTG	map to hMettl2A&B cDNA start without ATG, Sall in overhang, tc added to avoid frame shift
hM2_cDNA_R1	GTCAGCTAGCACGCGTGC TGGTGCTGGACAGAAGGG GCT	map to human mettl2B cDNA end without stop codon, NheI in overhang
hM2_cDNA_R2	AGATAAGCGGCCGcGCT GGTGCTGGACAGAAGGG GCT	map to human mettl2B cDNA end without stop codon, c added to avoid shift, NotI in overhang
hM2_cDNA_R4	AGATAAGCGGCCGCTCAG CTGGTGCTGGACAGAAGG GGCT	map to human mettl2A cDNA end with stop codon, (mettl2b has TAA as stop codon)NotI in overhang
hM2_cDNA_R5	CTAGTCTAGATCAGCTGGT GCTGGACAGAAGGGGCT	map to human mettl2A cDNA end with stop codon, (mettl2b has TAA as stop codon)XbaI in overhang

## Appendix 2.2 Primers for introducing mutations

hMettl2_dSAM_F1	CTGGATCCTCAGCCACCTACCGAaacacagctcttccaattttacaacg
hMettl2_dSAM_R1	tcgtttgtaaaattggaagactgtgtTCGGTAGGTGGCTGAGGATCCA
hM2_G3A_F	aacacagctcttccaattttacGcCTGTGcTGTGGcAcacctaccgaataactggag
hM2_G3A_R	AACCTCCAGTATTCGGTAGGTGTGCCACAGCACAGGCGT
M2_dSAM_F	cctggctcctctgccacctaccgaAACACAGTCTTTCCAATTTACAA
M2_dSAM_R	gttattgttagtttgtaaaattggaagactgtgtTCGGTAGGTGGCAGAGGA

M2_dSANT_F	cgtcttccatcacaatgcctggCATGAAAATGGGTTTTTCAAGGATAG
M2_dSANT_R	TCCTTGAAAAACCCATTTTCATGCCAGGCATTGTGATGG
hM8_dSAM_F2	accctttccagatgggatcGTCTTTGTGCTCTCTTCTATTCATCCTG
hM8_dSAM_R2	gaatagaagagagcacaagacGATCCCATCTGGAAAAGGGTAAGG
hMettl2 U6 F1	ACACCCTGCAATCCTCGCCGATAAAG
hMettl2 U6 R1	AAAACCTTATCGGCGAGGATTGCAGG
hMettl2 U6 F2	ACACCTGCACCTTCAGGGTAGGAGCG
hMettl2 U6 R2	AAAACGCTCCTACCCTGAAGGTGCAG
hM2_gRNA1_F_seq	GACTCCGCGTCTTCTGAAAGAG
hM2_gRNA1_R_seq	GCTTTTTCGGAGTGCTCCGAG
hTRMT1 U6 F1	ACACCGCTGCCGAATACAGCAGCGAG
hTRMT1 U6 R1	AAAACCTCGCTGCTGTATTCGGCAGCG
hTRMT1L U6 F1	ACACCGGAGCTGCTGCCCTGGAGAG
hTRMT1L U6 R1	AAAACCTCTCCAGGGGCAGCAGCTCCG

### Appendix 2.3 Primers for real time PCR

mettl2_qPCR_F1	TCAGCAAGCTAAGCCGACTC
mettl2_qPCR_R1	TCCAGCTCACCTTGTGTGAA
mettl2_qPCR_F2	ATGCCTGGGACAACGTGAAG
mettl2_qPCR_R2	TCAGGTGACTGTGGCTAGGT
hMettl2A_qPCR_F1	ACACAAGAGGAACTGGACACGCT
hMettl2A_qPCR_R1	GGCAGCAGGTGCCTCTCAGC
hMettl2B_qPCR_F1	TGAAGTGTTTCCGGCTCCGGTGT
hMettl2B_qPCR_R1	ACCACTCCACATTGTCCCAGGCAT
hMettl6_qPCR_F1	AGCTAAGATCATGTAGAGAGTTTGA
hMettl6_qPCR_R1	CACAGACTCTGGCGGTACAT
hMettl6_qPCR_F2	GCCAGAGTCTGTGGATGTTGTT
hMettl6_qPCR_R2	GAGCTGAGCCAGGAAGTCATCAG
GAPDH (human)	ACCGTCAAGGCTGAGAACGGGA
GAPDH (human)	CCCTGCAAATGAGCCCCAGCC
GAPDH (mouse)	GGTTGTCTCCTGCGACTTCAACAGC

GAPDH (mouse)	CGAGTTGGGATAGGGCCTCTCTTGC
mettl8_hu_qPCR_F5	GTTCGAGGAGATGGTACCAGAGCA
mettl8_hu_qPCR_R5	TGCCTTGAATCCACACTCGGTGC
mettl8_hu_qPCR_F6	AGGGGAAGTCCACAGTATGTTCTGC
mettl8_hu_qPCR_R6	TTGCCTTGAATCCACACTCGGTGC
mettl8_hu_qPCR_F7	GTGGCCCCTCTGGGATCAAGGA
mettl8_hu_qPCR_R7	TTCCCTCAACAGCCAATTACGATCC
p21_qPCR_F1	TGTCCTTGGGCTGCCTGTTTTCAGC
p21_qPCR_R1	TGCCGCATGGGTTCTGACGGA
p21_qPCR_F2	CGATGGAACTTCGACTTTGTCA
p21_qPCR_R2	GCACAAGGGTACAAGACAGTG
huSARS_qPCR_F1	TCGGGCAGACAACCTGAACA
huSARS_qPCR_R1	CAAACCGCTCTGCTTCCAAC
huSARS_qPCR_F2	CCACTACCCGTACCATCTGC
huSARS_qPCR_R2	CAAGCATCGGTGACCTCCAT
huATM_qPCR_F1	AGTGGCAGAAACACTCCCAG
huATM_qPCR_R1	CCCTTTCAGGGAGCTGAGTG
huATM_qPCR_F2	ATTCCAGCAGACCAGCCAAT
huATM_qPCR_R2	GGTCATCACGGCCCTTAACA
NEAT1_qPCR_F1	TTGTTCCAGAGCCCATGAT
NEAT1_qPCR_R1	TGAAAACCTTTACCCCAGGA
NEAT1_qPCR_F2	GATCTTTTCCACCCCAAGAGTACATAA
NEAT1_qPCR_R2	CTCACACAAACACAGATTCCACAAC

#### Appendix 2.4 Oligos for making Codon Runs

Sequence Name	Sequence 5'-3'
ThrACU10*F	GATCCACTACTACTACTACTACTACTACTACTG
ThrACU10*R	AATTCAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTG
ThrACC10*F	GATCCACCACCACCACCACCACCACCACCACCACCG
ThrACC10*R	AATTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTG
ThrACA10*F	GATCCACAACAACAACAACAACAACAACAACAACAG



12	4.39	NM_001078173	family with sequence similarity 127, member
13	4.39	NM_001078172	family with sequence similarity 127, member
14	4.29	NM_153200	endothelial differentiation-related factor 1
15	4.26	NM_005663	Wolf-Hirschhorn syndrome candidate 2
16	4.23	NM_001281297	endothelial differentiation-related factor 1
17	4.23	NM_033258	guanine nucleotide binding protein (G
18	4.10	NM_001267558	programmed cell death 6
19	4.09	NM_016643	zinc finger protein 771
20	4.09	NM_001142305	zinc finger protein 771
21	4.08	NM_000997	ribosomal protein L37
22	4.03	NM_003792	endothelial differentiation-related factor 1
23	4.00	NM_020412	chromatin modifying protein 1B
24	3.97	NM_015381	family with sequence similarity 19
25	3.92	NM_198708	hydroxysteroid (11-beta) dehydrogenase 1-
26	3.92	NM_001145250	solute carrier family 29 (nucleoside