

# **Characterisation of RNA modifications in human cancer cells**

**Daniel Thomas Wing**

**Hughes Hall**

**August 2020**

**This thesis is submitted for the degree of Doctor of Philosophy**

## Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. The number of words in this thesis does not exceed the limit set by the Degree Committee. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other university or similar institution.

## Abstract

### Characterisation of RNA modifications in human cancer cells

Daniel Thomas Wing

Chemical modifications have been studied as key modulators of RNA biology for over 50 years. There are now 171 RNA modifications that have been discovered throughout the three kingdoms of life<sup>1</sup>. These chemical changes overlay the standard RNA information and have been linked to a wide range of biological processes and human diseases, such as cancer. However, the vast majority of these modifications are uncharacterised, defined on limited substrates and are subject to enzymatic pathways. This shortfall is partly due to lack of reagents and technologies to study these modifications.

In an effort to create better tools and reagents, I performed an in-depth characterisation of a group of new anti-RNA modification antibodies generated in collaboration with Abcam plc. Using this platform, I validated a new anti-m1G antibody and utilised this novel reagent to help uncover new roles for m1G and the enzymes that catalyse its deposition. Using pancreatic ductal adenocarcinoma cancer (PDAC) cells as a model system, I discovered that m1G methylation is present on mRNA where it is catalysed by TRMT5, an enzyme previously identified as a tRNA methyltransferase (TRMT). Utilising ribosome profiling, I found that loss of TRMT5 greatly affects global protein translation. Moreover, depletion of TRMT5 from the PDAC cell line, PANC-1, inhibited proliferation and migration during *in vitro* assays. In contrast, TRMT5 depletion from a non-cancerous pancreatic cell line, hTERT-HPNE cells, had no effect on proliferation. By carrying out immunohistochemistry on human PDAC tissue, I identified high TRMT5 expression in cancerous tissue, but not in surrounding normal fibroblasts. Importantly, inducible knockdown of TRMT5 in PANC-1 cells orthotopically injected into mice, significantly interfered with their ability to form metastatic tumours.

As a separate project I investigated the role of another TRMT RNA guanosine methyltransferase, TRMT1, in a human ovarian cancer cell line JHOC-5. TRMT1 is predicted to be an m2,2G and m2G methyltransferase in human cells. I found that its loss drastically diminished the level of m2,2G in small (<200nt) and large (>200nt) RNAs, but surprisingly did not affect m2G levels. Furthermore, silencing of TRMT1 by CRISPR knockout technology showed a significant negative impact on migration of JHOC-5 cells, implicating this enzyme in ovarian cancer biology.

In summary, I characterised a new set of RNA modification antibodies using a screening system that exceeded conventional assessment of modification versus non-modification binding. An antibody from this screen was instrumental in (i) characterising m1G as a new mRNA modification, (ii) showing that TRMTs exhibit target promiscuity beyond their canonical tRNA substrates and (iii) linking the m1G RNA modification pathway to the growth of pancreatic cancer cells. My work also identified human TRMT1 as a likely m2,2G RNA methyltransferase whose activity is required for the migration of ovarian cancer cells. Overall, my findings demonstrate the importance of these two RNA modification pathways for cancer cell growth and migration, thereby highlighting TRMT5 and TRMT1, and perhaps TRMTs in general, as potential therapeutic targets in cancer.

## Acknowledgements

I am hugely grateful to my supervisor, Tony Kouzarides, for giving me the fantastic opportunity to undertake my PhD work in his lab. With the culture of the lab being the perfect balance between hard work and fun, it is a truly enjoyable place to be. I am also extremely grateful to my day to day supervisor, Carlos Melo. Thank you for your guidance and for sharing your wisdom and experience over the years, I could not have gotten here without you. I would also like to thank Francisco and Luca for analysing all the sequencing data I have generated. Andy, thank you for all you have taught me, learning from you has been priceless in my development as a scientist. Thank you to everyone else in the lab for your help and support. In addition, I owe a great deal to my former manager at Abcam, Gary. Without you I would never have been granted such an incredible opportunity.

I would also like to thank Saeeda, who has supported me in every possible way through the PhD. Thank you for your patience and guidance. A big thank you to my friends Mark, Ben, Adam and Stuart for their continued support over the years. Finally, I would like to thank my family who have always believed in me and encouraged me.

I gratefully acknowledge the funding of this work by Cancer Research UK. I would also like to thank Abcam for funding my studies and allowing me to carry out the PhD during my employment.

# Table of contents

Abstract.....	2
Declaration.....	2
Acknowledgements.....	5
List of abbreviations.....	13
Chapter 1 – Introduction.....	16
1.1 RNA classes .....	16
1.1.1 Ribosomal RNA (rRNA).....	16
1.1.2 Transfer RNA (tRNA) .....	16
1.1.3 Messenger RNA (mRNA) .....	17
1.1.4 Non-coding RNAs – micro RNA (miRNA) and long non-coding RNA (lncRNA).....	17
1.2 RNA modifications.....	18
1.3 RNA modification nomenclature .....	18
1.4 Brief history of RNA modifications.....	20
1.5 Tools for studying RNA modifications.....	21
1.5.1 Antibody-based detection methods .....	21
1.5.2 Detection using natural or induced RT (reverse transcriptase) arrests and signatures .....	25
1.5.3 RNA mass spectrometry.....	29
1.5.4 Oxford Nanopore Direct RNA sequencing (DRS).....	30
1.5.5 Exploiting specific chemical reactivities of modified nucleosides .....	31
1.6 RNA modifying enzymes .....	32
1.6.1 RNA methyltransferases .....	32

1.6.2 N1-methylguanosine (m1G) enzymes.....	34
1.7 Internal mRNA modifications.....	36
1.7.1 m6A.....	36
1.7.2 m5C.....	37
1.7.3 Inosine (I).....	37
1.7.4 m7G.....	38
1.7.5 ac4C.....	38
1.7.6 2' O-methylation.....	38
1.7.7 m1A.....	39
1.8 Biological effects of RNA modifications.....	39
1.9 The epitranscriptome in human disease.....	40
1.10 The importance of studying novel RNA modifications.....	45
1.12 Aims of this PhD thesis.....	47
Chapter 2 – Materials and methods.....	48
2.1 General materials.....	48
2.1.1 Standard solutions.....	48
2.1.2 Vectors.....	49
2.2 Bacterial strains and transformation.....	49
2.2.1 <i>E. coli</i> strains.....	49
2.2.2 Growth of bacterial cultures.....	49
2.2.3 Transformation of <i>E. coli</i> by heat shock.....	50

2.3 Preparation of DNA.....	50
2.3.1 Large scale (maxi prep) plasmid DNA preparation from <i>E.coli</i> .....	50
2.3.2 Small scale (mini prep) plasmid DNA preparation from <i>E.coli</i> .....	50
2.4 DNA cloning.....	51
2.4.1 DNA restriction digests and dephosphorylation.....	51
2.4.2 Agarose gel electrophoresis.....	51
2.4.3. Purification of DNA from agarose gels.....	52
2.4.4 Ligations .....	52
2.5 Rabbit monoclonal antibody production (RabMAb®).....	52
2.5.1 Antigen preparation and immunization.....	52
2.6 Antibody characterisation.....	52
2.6.1 Modified RNA oligonucleotide design .....	52
2.6.2 RNA Oligonucleotide panel ELISA .....	54
2.6.3 Free nucleoside competition ELISA.....	54
2.6.4 RNA oligonucleotide panel HTRF® assay.....	55
2.6.5 meRIP-mass spectrometry (methylated RNA immunoprecipitation-mass spectrometry)..	55
2.6.6 Mass spectrometry (MS).....	57
2.7 Cell culture .....	57
2.7.1 Cell lines .....	57
2.7.2 HEK293-6e cells.....	58
2.7.3 siRNA gene knockdown.....	58

2.7.4 shRNA gene knockdown .....	60
2.7.5 CRISPR knockout .....	60
2.7.6 Overexpression of TRMT5.....	60
2.7.7 Cell proliferation assay.....	60
2.7.8 Wound healing assay .....	61
2.7.9 Transwell® migration assay.....	61
2.7.10 OP-puro global protein synthesis analysis.....	61
2.7.11 Global RNA synthesis analysis.....	62
2.8 RNA extraction, treatment, and analysis .....	63
2.8.1 Total RNA extraction by precipitation.....	63
2.8.2 DNase treatment of total RNA.....	63
2.8.3 Total RNA extraction by column .....	63
2.8.4 PolyA <sup>+</sup> RNA enrichment from total RNA .....	64
2.8.5 mRNA quantification.....	64
2.8.6 ELISA on total RNA .....	65
2.8.7 meRIP-seq .....	66
2.8.8 meRIP-qPCR for m1G validation .....	66
2.8.9 Library preparation from meRIP-seq .....	66
2.8.10 Oxford Nanopore Direct RNA sequencing (DRS).....	66
2.8.11 Ribosome profiling .....	66
2.9 Protein purification, detection and analysis .....	67

2.9.1 Lysate preparation for Western blot .....	67
2.9.2 SDS polyacrylamide gel electrophoresis (SDS-PAGE).....	67
2.9.3 Western blotting .....	67
2.9.4 Cellular fractionation .....	68
2.9.5 Antibodies .....	69
2.10 Bioinformatics .....	69
2.10.1 meRIP-seq .....	69
2.10.2 Ribo-seq .....	70
2.10.3 ChIP-seq .....	70
2.11 Chromatin immunoprecipitation (ChIP) .....	71
Chapter 3 – Generation of RNA modification specific antibodies .....	73
3.1 Introduction .....	73
3.2 Aims of this chapter .....	76
3.3 Results .....	76
3.3.1 Choice of modifications and antibody generation strategy.....	76
3.3.2 Anti-m1G characterisation.....	76
3.3.3 Anti-m2,2G characterisation.....	79
3.3.4 Anti-m6,6A characterisation .....	81
3.3.5 Anti-m4C characterisation .....	83
3.3.6 Anti-m2A characterisation .....	85
3.3.7 Anti-m3U characterisation.....	87

3.3.8 Anti-m1A characterisation .....	89
3.3.9 Anti-m1Y characterisation .....	91
3.4 Discussion.....	93
Chapter 4 – Characterization of N1-methylguanosine (m1G) in pancreatic cancer .....	96
4.1 Why study the m1G modification pathway in pancreatic cancer?.....	96
4.2 Introduction to the m1G enzymes.....	98
4.2.1 TRMT10A, B and C.....	98
4.2.2 TRMT5 .....	99
4.2.3 Pancreatic ductal adenocarcinoma (PDAC) .....	100
4.3 Aims of this chapter .....	101
4.4 Results.....	101
4.4.1 m1G methyltransferases in cancer .....	101
4.4.2 TRMT5 as a biomarker for pancreatic cancer .....	104
4.4.3 Is m1G present on mRNA?.....	106
4.4.4 Effect of TRMT5 loss on transcription and translation .....	108
4.4.5 Loss of TRMT5 reduces proliferation and migration of PANC-1 cells.....	114
4.4.6 PANC-1 mouse xenograft.....	118
4.4.7 Is TRMT5 present on chromatin?.....	120
4.5 Discussion.....	121
Chapter 5 - Characterization of N2, N2-methylguanosine (m2,2G) in ovarian cancer cells.....	125
5.1 Brief introduction.....	125

5.2 Aims of this chapter .....	126
5.3 Results .....	126
5.3.1 Expression level of TRMT1 in cancer cells lines .....	126
5.3.2 The JHOC-5 ovarian cancer cell model.....	128
5.3.2 Generation of a TRMT1 CRISPR knock-out cell line .....	128
5.3.3 <i>TRMT1</i> knockout RNA mass spectrometry .....	130
5.3.2 Biological effects of TRMT1 loss .....	132
5.3.5 Transcriptional and translational effect of TRMT1 loss from JHOC-5 cells.....	135
5.3.6 Mapping m <sup>2</sup> ,2G within the transcriptome .....	137
5.3.7 Chromatin association of TRMT1 in JHOC-5 ovarian cancer cell line .....	139
5.3.8 Genome-wide mapping of TRMT1 using chromatin immunoprecipitation (ChIP) .....	141
5.4 TRMT1 discussion.....	142
Chapter 6 - Final conclusions and overall discussion.....	145
6.1 In depth characterisation of novel RNA modification antibodies.....	145
6.2 TRMT5 and TRMT1 final discussion .....	146
Bibliography .....	148

## List of abbreviations

• 5-EU	Ethynyl uridine
• 5fc	5-formylcytosine
• 5hmC	5-hydroxymethylcytosine
• A	Adenosine
• ac4C	N4-acetylcytosine
• ADAR	Adenosine deaminases acting on RNA
• AKT	RAC-alpha serine/threonine-protein kinase
• ALKBH	AlkB homolog
• ALYREF	Aly/REF export factor
• Bp	Base pair
• BSA	Bovine serum albumin
• CCR4-NOT	Carbon catabolite repression negative on TATA-less
• CDKN2A	Cyclin-dependent kinase inhibitor 2A
• cDNA	Complementary DNA
• CDS	Coding sequence
• CEBPZ	CCAAT enhancer binding protein zeta
• ChIP-seq	Chromatin immunoprecipitation sequencing
• CLIP	Cross-linking immunoprecipitation
• cm <sup>5</sup> U	5-carboxymethyluridine
• CMCT	carbodiimide N-cyclohexyl-N0-(2-morpholinoethyl) carbodiimide metho-
• p-toluenesulfonate	
• CRISPR	Clustered regularly interspaced palindromic repeats
• CTU	Cytoplasmic tRNA 2-thiolation protein
• DCP2	Decapping mRNA 2
• DHU	Dihydrouridine
• DMEM	Dulbecco's modified eagle medium
• DNA	Deoxyribonucleic acid
• dNTP	Deoxyribonucleic acid triphosphate
• DOX	Doxycycline
• DRS	Direct RNA sequencing
• DTT	Dithiothreitol
• <i>E.coli</i>	<i>Escherichia coli</i>
• ECL	Enhanced chemiluminescence
• EDTA	Ethylenediaminetetraacetic acid
• ELISA	Enzyme-linked immunosorbent assay
• ELP3	Elongator complex protein 3
• ESI LC	Electrospray ionization liquid chromatography
• FACS	Fluorescence-activated cell sorting
• FBS	Fetal bovine serum
• FRET	Fluorescence resonance energy transfer
• FTO	Fat mass and obesity-associated protein
• G	Guanosine
• GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
• GFP	Green fluorescent protein
• GRO-seq	Genomic run-on sequencing
• HEK293	Human embryonic kidney 293 cells
• HGF	Hepatocyte growth factor
• HRP	Horseradish peroxidase
• HS	High sensitivity
• hTERT	Human telomerase reverse transcriptase
• HTRF <sup>®</sup>	Homogenous time resolved fluorescence
• I	Inosine
• IGF-1	Insulin-like growth factor 1
• IHC	Immunohistochemistry
• IP	Immunoprecipitation
• KD	Knockdown

• KO	Knockout
• KRAS	Gene for K-ras protein (part of RAS/MAPK pathway)
• LB	Lysogeny broth
• lncRNA	Long non-coding RNA
• m1A	N1-methyladenosine
• m1G	N1-methylguanosine
• m1I	N1-methylinosine
• m1Y	N1-methylpseudouridine
• m2,2G	N2,N2-dimethylguanosine
• m2A	N2-methyladenosine
• m2G	N2-methylguanosine
• m4C	N4-methylcytosine
• m5C	N5-methylcytosine
• m6A	N6-methyladenosine
• m7G	N7-methylguanosine
• m8A	N8-methyladenosine
• mcm <sup>5</sup> s2u	5-methoxycarbonylmethyl-2-thiouridine
• meRIP-MS	Methylated RNA immunoprecipitation-mass spectrometry
• meRIP-seq	Methylated RNA immunoprecipitation-sequencing
• METTL	Human methyltransferase like enzymes
• miCLIP	m6a individual nucleotide resolution cross-linking immunoprecipitation
• miRNA	Micro RNA
• MOPS	3-(N-morpholino)propanesulfonic acid
• mRNA	Messenger RNA
• MS	Mass spectrometry
• NAT10	RNA cytidine acetyltransferase
• NES	Nestin
• NSUN2	NOP2/Sun RNA methyltransferase 2
• Nt	Nucleotide
• OP-puro	O-Propargyl-puromycin
• PBS	Phosphate-buffered saline
• PBST	Phosphate-buffered saline Tween 20
• PDAC	Pancreatic ductal adenocarcinoma
• PolyA+	Poly(A) enriched RNA
• PUS	Pseudouridine synthases
• qPCR	Quantitative polymerase chain reaction
• RNA	Ribonucleic acid
• RabMAB <sup>®</sup>	Rabbit monoclonal antibody
• Ribo-seq	Ribosome profiling sequencing
• RIPA	Radioimmunoprecipitation assay buffer
• RPF	Ribosome protected fragments
• Rpm	Revolutions per minute
• rRNA	Ribosomal RNA
• RT	Reverse transcriptase
• <i>S.cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
• s2U	2-thiouridine
• SCARLET	Site-specific cleavage and radioactive-labelling followed by ligation assisted extraction and thin-layered chromatography
• sgRNA	Single guide RNA
• shRNA	Short hairpin RNA
• siRNA	Small interfering RNA
• SMAD4	SMAD family member 4
• snoRNA	Small nucleolar RNA
• snRNA	Small nuclear RNA
• SOC	Super optimal condition
• SRSF10	Serine/arginine-rich splicing factor 10
• SRSF3	Serine/arginine-rich splicing factor 3
• T	Thymidine
• TBE	Tris-borate-EDTA

- TBS Tris-buffered saline
- TBST Tris-buffered saline Tween 20
- TCGA The cancer genome atlas
- TE Tris-EDTA
- Tp53 Tumour protein p53
- Tris Tris(hydroxymethyl)aminomethane
- TRMT tRNA methyltransferase
- tRNA Transfer RNA
- TUT Terminal RNA uridyltransferase
- U Uridine
- UTR Untranslated region
- UV Ultraviolet
- YAP1 Yes1 associated transcriptional regulator
- YTHDF YTH domain-containing protein family
- $\Psi$  Pseudouridine

# Chapter 1 - Introduction

## 1.1 RNA classes

Ribonucleic acid (RNA) is a polymeric molecule essential for numerous biological processes. It is involved in the coding, decoding, regulation and expression of genes in all three domains of life. There are several classes of RNA molecules in mammalian cells, of which the most relevant ones are described below.

### 1.1.1 Ribosomal RNA (rRNA)

rRNA is the most abundant type of RNA molecule in most cells, accounting for ~80% of all cellular RNA. It forms a critical component of the ribosome and because it catalyses the translation of mRNA, is classified as an RNA enzyme or ribozyme. In eukaryotic cells there are four rRNA subunits; the 5S, 5.8S and 28S form part of the 60S large ribosomal subunit, and the 18S rRNA forms part of the 40S small ribosomal subunit. Eukaryotes possess many copies of the rRNA genes (300-400), which are organised into gene clusters, and account for ~0.5% of the whole genome. The ribosomal RNA genes are located in the nucleolus, which is a non-membrane bound structure within the nucleus.

### 1.1.2 Transfer RNA (tRNA)

tRNAs are highly structured adaptor molecules that form the physical link between mRNA, the ribosome and the forming polypeptide and make up ~15% of total RNA. While the mRNA carries the specific amino acid sequence of the encoded protein, the role of the tRNA is to decode that RNA sequence into amino acids. The most critical part of the tRNA is the three-nucleotide sequence that forms complementary base pairs with the mRNA codon and is termed the anticodon. At the 3' end of tRNA molecules, amino acids are covalently linked by aminoacyl tRNA synthetases. During the process of protein translation, tRNAs are transported to the ribosome by elongation factors. If the anticodon of the tRNA is complementary to the mRNA codon, the tRNA already present in the ribosome transfers the polypeptide chain to the 3' of the incoming tRNA. Importantly, tRNAs contain a high level of modified nucleotides. As discussed in detail later in this thesis, these modifications can be important in the tRNA interaction with mRNA during translation.

### 1.1.3 Messenger RNA (mRNA)

mRNA is a single-stranded RNA molecule that is produced from the corresponding gene in a process called transcription and accounts for only 1-5% of total cellular RNA. The mRNA is transcribed by RNA polymerase into a pre-mRNA, where introns are still present. Pre-mRNAs then undergo splicing to form the mature mRNA. These molecules are made up of four distinct regions; the 5' untranslated region (5' UTR), the coding sequence, the 3' untranslated region (3' UTR) and the polyA tail. The UTRs are located 5' of the start codon and 3' of the stop codon respectively and although not translated, are still considered exonic as they are present in the mature mRNA. These regions have been associated with many biological functions including mRNA stability, localisation and translation<sup>2,3,4</sup>. The coding regions contain the codons which are decoded and translated by the ribosome and tRNAs, into proteins. The polyA tail consists of an extended sequence of adenosine nucleotides, which protects the RNA from degradation, facilitates export from the nucleus and promotes translation.

### 1.1.4 Non-coding RNAs – micro RNA (miRNA) and long non-coding RNA (lncRNA)

miRNAs are small non-coding RNAs that function to regulate post-transcriptional gene expression through silencing of mRNAs. These molecules contain sequences complementary to their target mRNAs and after binding, they silence their targets through either enhancing degradation or inhibiting translation.

lncRNAs are another non-coding RNA type but are greater than 200nts in length. 78% of lncRNAs are described as being tissue-specific, compared to only ~19% of mRNAs<sup>5</sup>. There are currently 189,901 lncRNAs annotated on the database lncRNADB<sup>6</sup>. The majority of these have been identified through next generation sequencing methods and only a minority have been functionally characterised. The open source platform for lncRNA annotation lncRNAWiki shows that only 381 lncRNAs have associated functions, with the majority of these linked to transcriptional regulation<sup>7</sup>.

## 1.2 RNA modifications

RNA modifications occur in all living organisms and can be defined as changes to the chemical composition of ribonucleic acid that provide a mechanism for regulating the function and structure of the RNA molecule. They have been found in many different RNA species including tRNA, rRNA, lncRNA, miRNA, snRNA, snoRNA and mRNA (figure 1.1). All four RNA bases (adenine, guanine, cytosine and uracil) can be modified as well as the ribose sugar-phosphate backbone. Modifications can range from simple additions such as methylation, to complex hypermodifications that require multi-step synthesis. For example, 2'-O-methyluridine 5-oxyacetic acid methyl ester is a complex modification found at the anticodon wobble position of some bacterial tRNAs<sup>8</sup>.

## 1.3 RNA modification nomenclature

The focus of this thesis is methylation-based RNA modifications. RNA modifications are typically abbreviated to describe which nucleoside is modified, the type, position and number of modifications. For example, N6-methyladenosine is formed from the addition of a single methyl group at the sixth position to the adenine base and is therefore abbreviated to m6A. Other RNA modifications exist where multiple additions are made to the base. For example, N2,2 dimethylguanosine is produced from guanosine, when two methyl groups are added to the second position on the base and is abbreviated to m2,2G (see figure 1.2 for further details of nomenclature).



## 1.4 Brief history of RNA modifications

The first RNA modification was identified over 60 years ago in *Saccharomyces cerevisiae*<sup>10</sup>. The authors identified an unknown compound when analysing ribonucleic acid by 2D chromatography, which we now know to be pseudouridine ( $\psi$ ). Shortly after this discovery, a small number of methylated purines (including m1G, m2G and m7G) were found to exist in human urine samples<sup>11</sup>. These early experiments studied RNA modifications in total RNA extracts, but it was not until 1974 that m6A was shown to exist as an internal mRNA modification<sup>12-14</sup>. Indeed, after their discovery in the late 1950's, interest in RNA modifications waned and there was very little progress to elucidate their molecular function. This was largely due to two reasons: i) modifications were thought to be confined to a small number of transcripts and ii) the tools to study them were inadequate. This changed in 2011 when human fat mass and obesity-associated protein (FTO) was demonstrated to be an m6A RNA demethylase, meaning the modification could be removed from RNA<sup>15</sup>. This discovery strongly implied a dynamic existence of m6A in mRNA, opening the door to a multitude of potential effects on RNA biology. However, FTO was later found to have 100-fold higher catalytic activity towards m6Am than m6A<sup>16</sup>.

Two subsequent papers published in 2012, made use of m6A specific antibodies to map the epitranscriptomic mark across the transcriptome<sup>17,18</sup>. These studies marked the beginning of what is now termed "epitranscriptomics" and initiated a much broader investigation into the individual functions of each modification and their role in various RNA processes, as well as diseases. Novel modifications are now being rapidly discovered and characterised, and new functions of established modifications are showing how context dependent their role can be.

Analogous to histone modifying enzymes, mRNA modifying enzymes can be grouped into three categories based on their function: "writers", "erasers" and "readers". "Writers" are enzymes that catalyse (i) the addition of chemical groups (e.g. METTL3 and m6A), (ii) the addition of entire nucleotides (e.g. terminal uridylyltransferases (TUTs)) or (iii) the isomerisation of nucleotides (e.g. tRNA pseudouridine synthases (PUSs)). "Erasers" are enzymes that reverse the modification and "readers" are proteins or complexes that recognise and bind to a specific modification to exert effector functions.

To date, erasers of mRNA modifications have only been identified for m6A, m6Am and m1A. Two erasers of m1A have been identified: ALKBH3 and ALKBH1<sup>19,20</sup>. Similarly, two erasers have also been

identified that demethylate m6A: FTO and ALKBH5<sup>15,21</sup>. As mentioned previously, FTO demethylates m6Am with a higher efficiency than it does m6A<sup>16</sup>. Finally, reader proteins for mRNA modifications have been identified that specifically bind to m6A, m7G and inosine.

The research described in this thesis is focussed on identifying and characterising the role of specific RNA methylations and the corresponding methyltransferases (or writers). Therefore, the remainder of this introductory chapter will also be focussed towards RNA methylation.

## 1.5 Tools for studying RNA modifications

There are multiple techniques by which RNA modifications are detected, quantified and mapped. These techniques can be divided into several groups: antibody-based technologies, natural or induced reverse-transcriptase (RT) signatures, selective chemistry, mass spectrometry (MS) and Nanopore direct RNA sequencing. Each approach has its advantages and limitations but can be used in parallel to validate one another.

### 1.5.1 Antibody-based detection methods

There are now many known RNA modifications, but most lack any considerable characterisation. It is likely that the few modification pathways that are functionally defined represent the ‘tip of an iceberg’. This shortfall is partly due to a lack of reagents to detect modification. For some modifications, specific chemistry exists that allows mapping without antibodies, however the majority of work to date has heavily depended on specific antibodies. Table 1.1 lists all the current detection methods for internal mRNA modifications and indicates whether the method is antibody or chemistry based. It is clear from table 1.1 that antibodies have been crucial in mapping all known mRNA modifications to date, including m6A, m5C, m1A, ac4C and m7G. Even when selective chemical treatments are available, such as for m5C, antibodies provide an orthogonal approach to validate findings.

Raising antibodies that specifically recognise the modification of a single nucleotide is inherently challenging. Such tiny epitopes present huge scope for cross-reactivity with off-target modifications. For example, an antibody may need to discriminate between a single methyl group at the sixth

position of an adenosine base (m6A) and one at the first position (m1A). In addition to specificity issues, antibodies can have RNA sequence biases. What appears as a modification motif, could be an antibody bias. With many modifications that are structurally very similar, thorough antibody characterisation is critical before embarking on pull-down and sequencing type experiments.

Detection technique	Modification	Antibody or chemical	Antibody or chemical used	References
meRIP-seq	m6A	Antibody	Anti-m6A antibody	18
m6A-seq	m6A	Antibody	Anti-m6A polyclonal antibody	17
miCLIP	m6A, m6Am, m5C	Antibody	Anti-m6A polyclonal antibody, anti m5C monoclonal antibody	22,23,24
PA-m6A-seq	m6A	Antibody	Anti-m6A antibody	25
m6A-CLIP	m6A	Antibody	Anti-m6A antibody	23
SCARLET	m6A	Chemical	P <sup>32</sup> labelling	26
m6A-LAIC-seq	m6A	Antibody	Anti-m6A antibody	27
RNA-BisSeq	m5C	Chemical	Sodium bisulfite	28
Aza-IP	m5C	Antibody and chemical	Anti-V5 antibody, 5-azacytidine	29
m5C-RIP	m5C	Antibody	Anti-m5C monoclonal antibody	30
hMeRIP-seq	hm5C	Antibody	Anti-hm5C antibody	31
Pseudo-seq	Ψ	Chemical	CMC	32
PSI-seq	Ψ	Chemical	1-Cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho- <i>p</i> -toluenesulfonate	33
CeU-seq	Ψ	Chemical	N3-CMC	34
m1A-seq	m1A	Antibody	Anti-m1A antibody	35,36
m1A-ID-seq	m1A	Antibody	Anti-m1A antibody	37
m7G-meRIP	m7G	Antibody	Anti-m7G antibody	38
BoRed-seq	m7G	Chemical	Borohydride reduction	38
acRIP-seq	ac4C	Antibody	Anti-ac4C monoclonal antibody	39

**Table 1.1 - RNA modification transcriptome-wide mapping techniques.** Adapted from reference<sup>40</sup>. List of current RNA modification mapping techniques, and the related antibody or chemical used.

Research antibodies can be divided into two distinct categories, polyclonal and monoclonal. Polyclonal antibodies have historically been the workhorse of biological research. They are generally quick, simple and relatively cheap to produce compared to monoclonal antibodies. However, they have certain limitations. Once a batch of polyclonal antibody has been exhausted, it cannot be faithfully reproduced. The next production of the ‘same’ antibody (even if exactly the same immunogen is used), will be generated from different animals and the immune response and exact antibody repertoire will differ. This can make reproducibility a challenge, especially if the antibody was difficult to produce in the first instance. On the other hand, monoclonal antibodies have been around since the 1970’s, and largely circumvent the problems of batch-to-batch variability. However, they require more specialised equipment, reagents and skills. In addition, monoclonal antibodies are not completely free of

variability issues, as mouse or rabbit hybridoma cell lines are inherently unstable and liable to genetic drift. Modern recombinant technologies have further improved upon monoclonal antibodies by cloning the antibody variable domains into more reliable mammalian expression systems. The RNA modification antibodies utilised in the work presented here are all recombinantly cloned rabbit monoclonal antibodies (RabMAb®).

The first antibody-based enrichment of RNA modifications dates back to the 1980s<sup>41</sup>, but it was not until the advent of next generation sequencing that there was an explosion in the field. The mapping of human RNA methylomes using m6A antibodies and sequencing was, and still is, being followed up by a vast number of studies using similar technology. The basic principle of these techniques is the immunocapture of a specific modification present within fragmented RNA. The modification enriched RNA is then eluted from the antibody with either free modified nucleoside or digested off with ionic detergent and proteinase K. These techniques have been given various names but are generally referred to as methylated RNA immunoprecipitation (meRIP). The initial iteration of meRIP-seq was limited in its resolution to 100-200nt. Under these peaks of sequence enrichment, the exact position of the methylation remained essentially unidentified. To “call peaks” in the sequencing data sets, computational approaches are employed to look for read accumulation in regions of the immunoprecipitated RNA, when normalised to the input signal. This process was refined by the addition of a UV crosslinking step, where the antibody is covalently linked to the RNA, which greatly improved mapping resolution. This was an adapted version of cross-linking immunoprecipitation (CLIP), using anti-m6A antibodies, called “m6A individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP)”<sup>22</sup>. This technique was first employed using anti-m5C antibodies and identified the modification in non-coding RNA (ncRNA) that had not been detected previously<sup>42</sup>. Here the authors validated the technique by the detection of known m5C sites in tRNAs.

Despite being used heavily in mapping experiments, both meRIP and miCLIP have the inherent disadvantages of antibody-based techniques. False positives are commonplace in antibody-based mapping datasets. Sources of false positives have been scrutinised in m6A datasets and there are two potential causes of non-specific peaks: (i) non-specific binding of RNA to the antibody, which can be exacerbated by the non-random nature of RNA fragmentation<sup>43</sup>, (ii) specific, but modification independent, binding of specific RNA sequences to the antibody. This has been shown for one m6A antibody that can bind unmodified purine rich sequences in RNA<sup>44</sup>.

One crucial step that can be implemented to improve confidence in called peaks, is to deplete one or several enzymes known to deposit the modification being investigated. If a peak is called in a control sample but is decreased/lost upon disruption of a putative writer enzyme, this offers greater confidence that the peak contains a genuine modification site. Importantly though, this does not definitively demonstrate that the disrupted protein is the enzyme catalysing the modification. It could, for example, be a crucial co-factor or upstream regulator of the actual RNA modifying enzyme.

Underscoring the concern of antibody-based techniques, the validity of mapping results for some modifications, such as m1A has been questioned<sup>45</sup>. Moreover, the positions of 2'-O-methylated nucleotides have been recently mapped to locations that resemble a common sequencing artefact<sup>46</sup>. m5C mapping studies have also shown a high level of variability, where the number of sites has differed by almost 1000 fold between datasets<sup>47</sup>. Taken together, these reports highlight that discrepancies between antibody-based mapping experiments in the field of RNA cannot be ignored. Indeed, there are steps that can be taken to improve reproducibility. Batch-to-batch variation of polyclonal antibodies is always a potential source of disparity in biological research, especially when a high sensitivity technique such as meRIP-seq is being used. Monoclonal antibodies, and particularly genetically defined recombinant monoclonal antibodies, remove this risk and can be thoroughly characterised and remain so without fear of genetic drift or batch to batch changes.

### 1.5.2 Detection using natural or induced RT (reverse transcriptase) arrests and signatures

Some modifications in the RNA template are known to either cause the RT enzyme to arrest or introduce non-Watson-Crick base pair (bp) compatible dNTPs into the nascent cDNA molecule during cDNA synthesis. This phenomenon may be due to either the inability to form a base pair with any canonical nucleotide or due to steric hindrance inhibiting the recognition of the modified residue by the RT. Such naturally existing RT signatures can also be used as a tactic to identify modification sites.

Table 1.2 shows a selection of RNA modifications and their effect on RTs. RT-based modification detection is attractive because there is no need for specific reagents for a given modification. However, despite this advantage, the techniques are not straightforward. Even an unmodified RNA template can produce RT stops due to nuclease cleavage or secondary structure of the RNA. Moreover, reverse transcription profiles of modified RNA are often prone to a stop-start error that can potentially mask other nearby modifications<sup>48</sup>.

Numerous software packages have been developed to identify unique misincorporation patterns to specific modifications. One of the earliest attempts to use these RT signatures to map RNA modifications was named HAMR (High-throughput Annotation of Modified Ribonucleotides)<sup>49</sup>. However, there are instances when multiple modifications are impossible to resolve as their mutational signatures are too similar, so modifications must be grouped. The success of this technique varies greatly depending on the modification. For example, m1A|m1I|ms2i6A could be predicted with a 98% accuracy, whereas m1G and m2G|m2,2G could only be predicted with 78% accuracy. Both m1G and m2,2G strongly favour the incorporation of T with varying amounts of C and A. Interestingly, for some modifications that were deemed RT signature silent, specific conditions were found which cause RT signatures to appear. For instance, 2'-O-methylations, which were previously thought to not produce an RT signature, were found to cause RT-stops at lower dNTP concentrations during primer extension. This technique was utilised to map 2'-O-modifications on rRNA<sup>50</sup>.

Chemical name	Abbreviation	RT block	RT Pause
2'-O-methyladenosine	Am	-	+ (low dNTP)
N1-methyladenosine	m1A	+	+
N2-methyladenosine	m2A	-	+
N6-methyladenosine	m6A	-	+
N6, N6 -dimethyladenosine	m6,6A	+	+
Inosine	I	-	-
2'-O-methylguanosine	Gm	-	+ (low dNTP)
N1-methylguanosine	m1G	+	+
N2-methylguanosine	m2G	-	+
N2, N2-dimethylguanosine	m2,2G	+	+
N7-methylguanosine	m7G	-	-
2'-O-methyluridine	Um	-	+ (low dNTP)
N3-methyluridine	m3U	+	+
5-methyluridine	m5U	-	-
5-carboxymethyluridine	Cm <sup>5</sup> U	-	-
Dihydrouridine	DHU	-	+
Pseudouridine	ψ	-	-
N1-methylpseudouridine	m1ψ	-	-
N3-methylpseudouridine	m3ψ	+	+
N1-methyl-3-(3-amino-carboxypropyl)-pseudouridine	m <sup>1</sup> acp <sup>3</sup> ψ	+	+
2'-O-methylcytidine	cM	-	+ (low dNTP)
N3-methylcytidine	m3C	+	+
N4-methylcytidine	m4C	-	+
N4-acetylcytidine	ac4C	-	+
5-methylcytidine	m5C	-	-
5-hydroxymethylcytidine	hm5C	-	-

**Table 1.2 - RNA modifications and their effects on avian myeloblastosis virus (AMV) reverse transcriptases.** Adapted from *Methods in Enzymology*, volume 425, RNA modification, page 37.

The concept of RT signatures to map modifications can be taken further by engineering enzymes to become more sensitive to modifications. In the first example of this approach, mutant Klen-Taq DNA polymerases were generated that exhibited sensitivity to 2'-O-methylation<sup>51</sup>. However, certain modifications, such as m6A, are generally removed during reverse transcription, so are not usually amenable to RT signature mapping. But recently, another mutant Klen-Taq DNA polymerase was identified that produces increased misincorporation in the strand opposite m6A residues<sup>52</sup>.

Moreover, manipulation of the RNA template itself can be used to identify the positions of modified nucleotides. For example, a technique termed AlkB-facilitated RNA methylation sequencing (ARM-seq) uses the bacterial demethylase AlkB to remove RT-blocking methylations from the RNA template<sup>53</sup>. Differential RT signatures are then analysed to determine modification positions. This

tactic has been used to map several modifications on tRNA including m1A, m3C and m1G. The *E.coli* demethylases can be genetically engineered to further increase their activity. For example, an AlkB D135S mutant has enhanced activity and has been used in an improved version of the technique termed DM-tRNA-seq<sup>54</sup>.

Like antibody techniques, RT-based methods also have several limitations. For example, it was found that without some kind of enrichment, only very few high confidence RT arrests can be identified<sup>55</sup>. Furthermore, the introduction of an RT signature can be heavily sequence dependent<sup>56</sup>. Notwithstanding, only a small subset of modifications do not affect Watson-Crick base pairing and are therefore RT silent during cDNA synthesis. These novel RT enzymes have been crucial in mapping tRNA and rRNA modifications but are yet to be widely adopted in transcriptome-wide studies. However, they offer a useful alternative to antibody-based detection methods.

### 1.5.3 RNA mass spectrometry

The overall abundance of modifications is an important metric for gauging how broadly a particular modification might influence the function of particular types of RNAs. Sequencing technologies have provided valuable insight into where modifications are located, but they cannot reliably provide information on abundance, because they rely on the biochemical properties of the antibody and/or chemical method used. Where antibodies have no application in discovering new modifications, mass spectrometry has been widely used in the discovery of novel RNA modifications. Currently, almost all of the modified nucleosides reported in the literature were either discovered by mass spectrometry or were structurally characterised by mass spectrometry<sup>57-58</sup>. The predominant strength of using mass spectrometry to study RNA modifications though is its ability to provide a complete and dynamic census of RNA modifications within a sample. The first example of this type of work was carried out by Dedon *et al*<sup>59</sup>. This study looked at changes in RNA modification in response to various stresses in yeast.

A single internal standard can be used to compare RNA modifications between two samples to give relative differences. This method, however, is not useful for accurate quantification and cannot be used to compare two different modifications. More precise measurements of individual modifications within a single sample can be obtained using isotope dilution methods<sup>60</sup>. This method requires isotopically labelled standards of every modification to be measured, so is not the most accessible method.

Electrospray ionization liquid chromatography-tandem mass spectrometry (ESI LC-MS/MS) has been the primary tool used in the identification and quantification of modified nucleosides at the monomer level. Although RNA can be analysed by MS at either the nucleotide or nucleoside state, the nucleoside approach is more popular. This technique was developed by the McCloskey lab in 1990 and remains popular today<sup>61</sup>. It is important that the RNA MS technique can separate isobaric nucleosides (i.e. molecules with the same normal mass, but different chemical formulas). In particular, isobaric species such as m1G and m2G must be differentiated through unique fragments and sufficiently high fragmentation energy to generate the unique fragments<sup>62-63</sup>.

Although primarily used for detection of nucleoside and nucleotide monomers, MS has also been applied to the mapping of modifications in oligonucleotides. Pitteri and colleagues used MS to analyse intact miRNAs<sup>64</sup>. However, this approach is yet to be applied to larger RNA species such as mRNA, tRNA and rRNA. Using MS to study RNA modifications in a sequence context is technically challenging

and is not usable at a transcriptome-wide scale. However, once modified candidates have been identified, it can offer a useful targeted approach to validate findings. This approach was recently used for detection of m7G modified miRNAs<sup>38</sup>.

#### 1.5.4 Oxford Nanopore Direct RNA sequencing (DRS)

A newly published method for detecting RNA modifications leverages Oxford Nanopore direct-RNA sequencing<sup>65</sup>. This technology has allowed for the first-time direct sequencing of full length, native RNA molecules, bypassing the need for reverse transcription or amplification. Nanopore technology uses a molecular motor to pull a single RNA or DNA molecule through a protein pore, which is embedded in a synthetic membrane. As the nucleobases pass through the pore, the flow of ions across the membranes is altered, depending on the composition of the bases, which is measured as a change in conductance through the pore.

Although it was designed as a sequencing method, numerous studies have found that the acquired data contain information about RNA modifications<sup>66,67,68</sup>. Critically, RNA modifications can induce shifts in the current intensity and the time the sequence spends in the pore (dwell time). Leger *et al* used m6A to validate an analytical pipeline<sup>65</sup>. Comparing RNA from METTL3 (an m6A writer) depleted cells to wild type cells they were able to detect m6A sites and cross validated these with antibody-based methods. One disadvantage of antibody-based pull-down sequencing methods is the lack of stoichiometric information obtained. Because Nanopore direct sequencing reads each RNA molecule separately, stoichiometric data are theoretically possible to obtain.

Although Nanopore direct RNA sequencing is not reliant on antibodies, it may be used to validate hypothetical sites discovered using antibodies. For instance, as with the m6A study described above, depletion of an RNA writer could allow modification detection by Nanopore direct RNA sequencing. However, this method would only show enzyme dependence, not direct catalytic activity. It is also feasible that downstream modification changes may arise due to depletion of one writer enzyme, producing false positives. However, catalytic activity could be shown directly if the RNA target were methylated *in vitro* with a recombinant enzyme.

### 1.5.5 Exploiting specific chemical reactivities of modified nucleosides

Various reagents have been found that specifically react with modified RNA nucleosides, or selectively react with non-modified RNA nucleosides. Once treated, the RNA modifications then produce either misincorporations or terminations during cDNA synthesis. These can then be detected during sequencing when compared to non-treated control samples. One of the most widely used of these techniques is bisulfite sequencing, which was adapted from 5mC detection in DNA<sup>28</sup>. In a non-modified cytosine, the bisulfite reagent acts as a nucleophile for addition to the base, which then leads to deamination. The presence of the methyl group on 5mC increases the electron density of the base, therefore repelling the nucleophile addition from the bisulfite. Despite being a widely established technique on abundant RNA species such as tRNA and rRNA<sup>69</sup>, results on mRNA are still contested<sup>47</sup>.

Chemical treatments exist for other modifications too. A technique to detect  $\psi$  through the addition of carbodiimide N-cyclohexyl-NO-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMCT) was developed in 1993 by Bakin *et al*<sup>70</sup>.  $\psi$  selectively reacts with carbodiimide moiety (CMC) to form a bulky adduct, which in turn produces an RT stop.

More recently, a selective chemical treatment has been developed to identify internal m7G modifications in miRNAs, termed BoRed-seq<sup>38</sup>. This technique uses borohydride reduction of m7G nucleosides to produce abasic sites. These sites are then labelled with biotin coupled aldehyde reactive probes. The tagged RNA can then be purified with streptavidin beads. This approach has been used alongside conventional antibody immunoprecipitations to increase confidence of called m7G sites<sup>38</sup>.

Like all the techniques discussed above, selective chemical treatments have their limitations. Critically, such treatments do not currently exist for the majority of modifications.

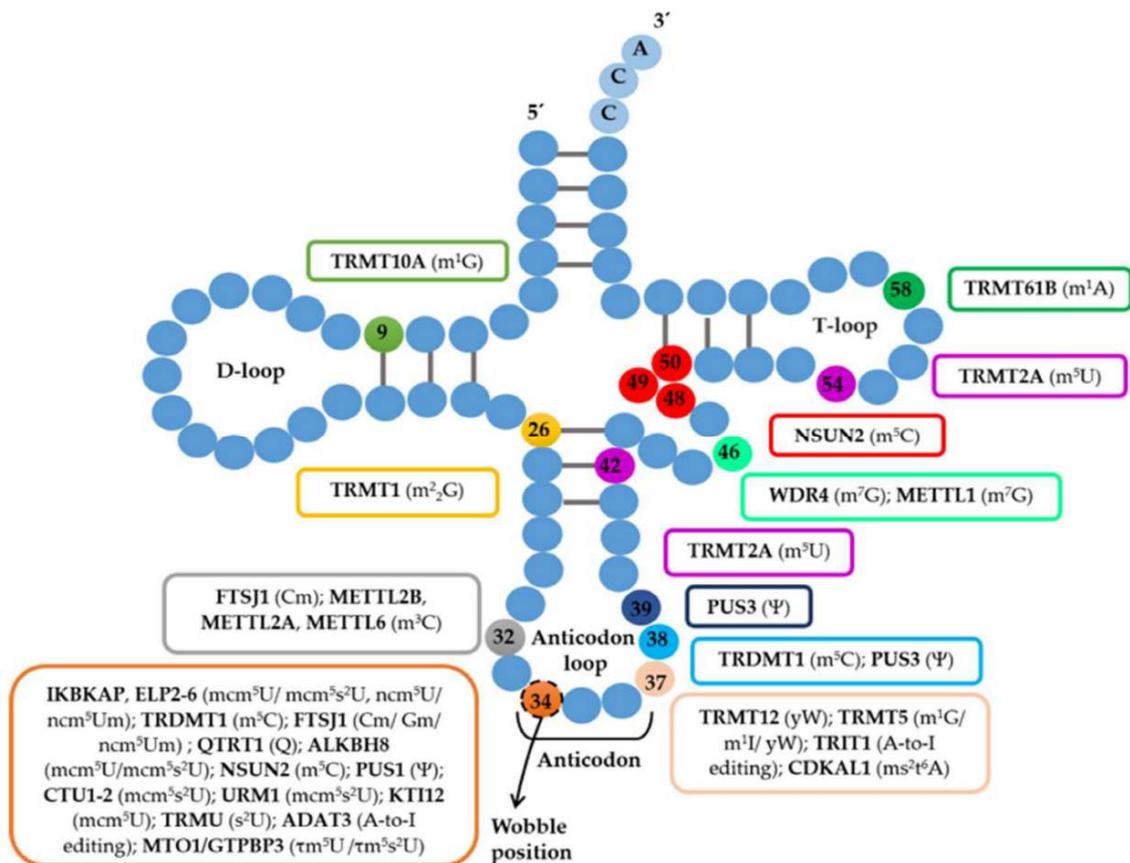
## 1.6 RNA modifying enzymes

### 1.6.1 RNA methyltransferases

Methyltransferases are a large family of enzymes that methylate their target substrate and are divided into several subgroups. The most common type and most important for this thesis is class I, which contain a Rossman fold. This tertiary structure is critical for binding the nucleotide *s*-adenosyl methionine (SAM), which acts as the methyl donor. Class II methyltransferases are defined by their SET domain (domains initially characterised in *Drosophilla* proteins, Su(var)3-9, Enhancer-of-zeste and Trithorax) and include well known histone methyltransferases. Class III methyltransferases are membrane associated enzymes.

The enzymes that catalyse methylation of RNA are termed RNA methyltransferases (RNMTs). They are a highly diverse family of proteins that share several common features including a methyltransferase domain and a SAM binding domain. There are several subgroups of RNMTs that are categorised based on their preferred or predicted substrate. For example, the methyltransferase like protein family (METTL) includes the well-known m6A writer proteins METTL3 and METTL14, which target mRNA<sup>17</sup>. However, many other members of the METTL family remain uncharacterised.

A subgroup of RNA methyltransferases called tRNA methyltransferases (TRMTs) bare structural similarities to the METTL family but are distinct because they are generally believed to exclusively target tRNAs. They are highly specific for their given target nucleotide and position. tRNAs are crucial players in the process of protein synthesis as they decode the information from mRNA. But in order to be fully active, tRNAs must undergo extensive post-transcriptional RNA modification, with approximately 17% of their nucleotides undergoing modification<sup>71</sup>. Inappropriate, or lack of modification can have a range of effects on translation, and subsequent protein expression, by increasing ribosome slippage or stalling. TRMT enzyme dysregulation plays a role in many diseases including neurodegenerative disorders and cancer<sup>72,73</sup>.



**Figure 1.3 - Known tRNA modifications and the enzymes that catalyse them.**

Illustration summarising the vast array of modifications identified on tRNA and the enzymes identified that catalyse them. Some enzymes can modify more than one site (e.g. TRMT2A) and some positions can be modified by different enzymes (position 34 and 37). Adapted from a paper by Pereira *et al*<sup>72</sup>.

### 1.6.2 N1-methylguanosine (m1G) enzymes

One of the most thoroughly studied modifications of tRNA is m1G at position 37 adjacent to the anticodon. This modification is conserved throughout all three domains of life, meaning it likely existed in an organism before the three kingdoms diverged<sup>74</sup>. This modification is performed in *E. coli* by TrmD, in *S. cerevisiae* by TRM5 and in humans, the orthologue is TRMT5. The alignment of these three enzymes is shown in figure 1.4. TRMT5 has been demonstrated to *in vitro* methylate tRNA<sup>leu</sup><sup>75</sup>. The loss of m1G from position 37 has global effects on translation and viability in both bacteria and yeast<sup>76</sup>. Loss of TrmD and TRM5 from bacteria and yeast respectively, severely impairs growth<sup>74,77</sup>. Although the loss of TRMT5 from human cells has not been widely investigated, TRMT5 mutations have been linked to multiple respiratory-chain deficiencies<sup>78</sup>.



**Figure 1.4 - Alignment of human TRM5, *S. cerevisiae* TRM5 and *E. coli* TrmD m1G methyltransferase protein sequences. Sequences were aligned using CLC bio (Qiagen) software.**

## 1.7 Internal mRNA modifications

For many decades' mRNA was viewed as a passive intermediate between DNA and protein, with little regulatory capacity. DNA modification was well established, and post translation protein modifications were widely studied, but only in 2011 did internal mRNA modifications become apparent as critical contributors to fine tune diverse regulatory processes in the cell<sup>15</sup>.

The relatively low abundance of mRNA (less than 5% of total RNA), combined with potentially low modification abundance, makes the study of mRNA modifications extremely challenging. To date, most of the known RNA modifications have been identified on tRNAs and rRNAs. However, a growing number of modifications (7 so far) are now being detected internally in mRNA.

### 1.7.1 m6A

m6A is the most abundant, and consequently, the most well characterised internal mRNA modification. The biological necessity of m6A has been known for decades. In the 1980s it was shown that m6A is crucial for proper mRNA processing and intracellular transport in mammals<sup>79</sup>. Mapping studies have shown m6A modifications occur approximately once every 700-800nt of mRNA<sup>80</sup>. The mark is primarily found in the consensus sequence G(m6A)C or A(m6A)C, but only a fraction of these sequences are modified<sup>81</sup>. Existence of a consensus sequence implies that the modification machinery has a certain selectivity.

Despite being the most abundant internal RNA modification, methods and the motivation to determine the localisation of modification sites were lacking until relatively recently. Two advances changed this. Firstly the discovery of FTO as an m6A eraser protein<sup>15</sup>, followed by the discovery of a second m6A eraser of the same family, ALKBH5<sup>21</sup>, implied a dynamic nature of the modification. Secondly, the rapid advancements of sequencing technologies, facilitated the first transcriptome-wide m6A mapping experiments<sup>17,18</sup>. These studies were the first to demonstrate that m6A primarily localises within the coding region and 3'UTR of mRNAs. Methods utilising cross-linking of antibodies to modification sites later improved resolution (miCLIP), allowing single base mapping of m6A marks<sup>23,25,82</sup>. However, these techniques cannot provide information about the proportion of modification at each target site. Consequently, a ligation-based primer extension method termed SCARLET was developed to quantitatively determine the fraction of mRNAs modified at a specific site (whilst also confirming putative modification sites). However, this technique is not a high throughput one<sup>83</sup>.

One of the predominant ways that m6A exerts its function is through associated reader proteins which specifically recognise m6A modified transcripts. Such reader proteins were identified via their binding to m6A methylated oligonucleotide affinity columns coupled to MS<sup>84</sup>. Even within a structurally similar family of reader proteins, each reader can exert very different effector functions. For example, the cytoplasmic m6A reader protein YTH domain-containing family protein 2 (YTHDF2), promotes degradation of its target RNA through recruitment of CCR4-NOT complex, but YTHDF1 and YTHDF3 promote translation<sup>85,86,87</sup>. Conversely, YTHDC1 is located in the nucleus and has been shown to regulate mRNA splicing, mRNA export and facilitate mRNA decay of specific transcripts<sup>88,89,90</sup>. YTHDC2 regulates mRNA stability, translation and has been shown to control spermatogenesis<sup>91</sup>. The YTH proteins represent just one family of m6A reader proteins and alone have a wide range of effects on RNA biology.

### 1.7.2 m5C

Like m6A, 5-methylcytosine (m5C) was discovered many years ago<sup>13</sup>. The first m5C mapping experiments were published even before m6A, leveraging bisulfite sequencing methods from DNA modification studies<sup>28</sup>. The distribution of this modification tends to be biased towards the untranslated regions, both 5' and 3'<sup>42</sup>. Of particular relevance to this thesis is the fact that the tRNA m5C methyltransferase NSUN2 was found to catalyse m5C on mRNA<sup>92</sup>. Indeed, it could be a common feature of tRNA modifying enzymes that they target additional RNA species, including mRNA. The primary function of m5C in mRNA appears to be nuclear export through the binding of a “reader” protein called ALYREF<sup>92</sup>. m5C in RNA can be further converted to 5-hydroxymethylcytosine 5hmC, which has also been found in lncRNA and mRNA<sup>93</sup>.

### 1.7.3 Inosine (I)

Genetically encoded adenosines can be deaminated in RNA to form inosine (I). This so-called RNA editing is performed by adenosine deaminases acting on RNA (ADAR) enzymes. Unlike other modifications, A to I editing events can be detected by sequencing alone. I nucleotides are reverse transcribed into Gs, so comparison of RNA sequencing data to the reference genome allows the identification of Is. Over 10<sup>8</sup> A to I editing sites have been so far identified by deep sequencing. Introduction of I is known to affect base pairing of RNAs, and although the biological function remains largely unknown, it has been suggested that it plays a role in RNA silencing<sup>94</sup>.

#### 1.7.4 m7G

m7G is well known for its presence in the cap structure of mRNA<sup>95</sup> but it is also found internally in various RNAs. In yeast, m7G is catalysed by Trm8p, facilitated by an essential co-factor, Trm82p. The best characterised enzyme performing internal m7G methylation is the TRMT8 yeast enzyme homologue METTL1 (methyl-transferase-like 1). Recently, the human homologue of Trm82p, WDR4 was also found to be a co-factor for METTL1<sup>96</sup>. WDR4 knockout was found to abolish m7G in tRNAs. Interestingly, m7G has recently been detected in both human mRNA and miRNA<sup>97,38</sup>. The latter study showed that a single m7G modification in the *let-7e-5p* miRNA alters the migratory capacity of lung cancer cells. The m7G is catalysed by METLL1 and disrupts a quadruplex structure in the miRNA, which in turn allows the pri-miRNA to be processed into the mature miRNA by the RNase, DICER. The mature miRNA then regulates its target mRNAs, silencing their expression. These downstream target RNAs are linked to the migratory phenotype and are therefore regulated by METTL1 catalytic activity.

#### 1.7.5 ac4C

Although most RNA modifications are methylation-based, there are numerous other chemical additions to RNA that have been identified. First identified in bacterial tRNAs, ac4C has been more recently detected in mammalian tRNAs and rRNAs where it is catalysed by N-acetyltransferase 10 (NAT10)<sup>98,1</sup>. Furthermore, acetylation of cytidine has been detected and characterised in mRNA using a new RabMAb<sup>®</sup> developed by Abcam. NAT10 dependent acetylation was mapped across the transcriptome and found to be enriched in coding sequences of mRNAs. Furthermore, NAT10 dependent acetylation increases the stability of transcripts and enhances substrate translation<sup>39</sup>.

#### 1.7.6 2'-O-methylation

Along with modifications to the nucleobase at various positions, the ribose sugar can also be methylated at the 2' hydroxyl group, which can occur on all four nucleotides. 2'-O-methylation often occurs in addition to base modification but can exist independently. All four variants (Am, Cm, Um, Gm) are found in tRNA and rRNA, but have not been identified in mRNA. However, the existence of m6A with additional 2'-O-methylation, termed m6Am, has been known since the 1970s, but only very recently has the writer enzyme, PCIF1, been discovered for this modification<sup>99</sup>. A novel technique termed m6Am-Exo-seq was used to map m6Am<sup>99</sup>. Using this technique, the authors found very little overlap between m6A and m6Am, indicating their functional distinction. The translational efficiency of transcripts containing m6Am modification is decreased, so the modification is likely an important regulator of gene expression. Despite initially being identified as an m6A demethylase<sup>15</sup>, FTO activity

on m6A containing RNA was later called into question. The unusually slow reaction time of FTO demethylation of m6A, together with the resistance of most m6A residues to FTO loss<sup>100</sup>, led to the belief that it may in fact be a non-specific reaction *in vitro*. It was subsequently shown that FTO has a 100-fold higher catalytic activity on m6Am than it does on m6A<sup>16</sup>.

### 1.7.7 m1A

N1-methyladenosine (m1A) is unlike m6A in that the methyl group is positioned at the interface of Watson-Crick base pairing, thereby disturbing base-pairing. m1A also creates a positive charge on the base, affecting RNA secondary structure and protein-RNA interactions. The number of m1A sites has been a point of controversy in the epitranscriptomics field<sup>101</sup>, but it is not controversial to say that m1A is far less abundant than m6A. Initially, m1A was being mapped to the translational start site and first splice site<sup>102</sup>. However, a recent paper from the Jaffrey laboratory asserts that the earlier m1A sites were in fact a result of antibody cross-reactivity<sup>45</sup>. The issue of antibody promiscuity in the field of epitranscriptomics will be addressed in more depth later in this thesis.

The rapidly growing list of internal mRNA modifications clearly shows that there could be others at low abundance. Notwithstanding, further work to understand the function of these modifications will be needed, in addition to the mechanisms that regulate their presence.

## 1.8 Biological effects of RNA modifications

Chemical modifications have the potential to influence mRNA structure and dynamics, splicing and maturation, RNA–protein interactions, translation, and stability. Changes to the base structure of RNA nucleotides can have profound effects on the stability of RNA molecules.  $\psi$ , m6A, 5fC, m1A and m1G have all been shown to affect the stability of RNAs and are therefore predicted to alter their secondary structure<sup>103,104,105,106,107,108</sup>. Modifications have also been shown to affect the half-life of mRNAs. m6Am modified RNAs demonstrate increased half-life in cells by repelling the decapping enzyme DCP2<sup>16</sup>. m6A can work through a similar mechanism to increase mRNA half-life, by interfering with the binding of the reader protein G3BP1<sup>84</sup>. In addition, RNA modifications can also influence the localization of transcripts, through reader protein interactions. The m6A reader protein YTHDF2 selectively binds to a subset of m6A modified RNAs resulting in their translocation from a translatable pool to cellular decay sites<sup>109</sup>. Another step in the mRNA life cycle that is influenced by m6A modification is splicing. The reader protein YTHDC1 recruits SRSF3 and blocks SRSF10, facilitating exon inclusion during

maturation<sup>88</sup>. Furthermore, m6A has been shown to influence translation by an array of different pathways. For example, the m6A methyltransferase METTL3 interacts with the translation initiation complex independently of its methyltransferase activity, enhancing translation<sup>110</sup>. The reader protein YTHDF1 also interacts with m6A modified transcripts to promote ribosome loading, thereby facilitating translation<sup>111</sup>. Interestingly, m6A on mRNA has been found to interfere with tRNA accommodation during translation elongation. Despite the fact that m6A still base pairs with uridine, the interaction between the m6A containing codon and the cognate tRNA anticodon is altered sufficiently to affect translation dynamics<sup>112</sup>. Unsurprisingly, the majority of mRNA modification dependent biological effects have been ascribed to m6A, but the mechanistic effects of other modifications are emerging. For instance, both m1A and ac4C have been found to alter translation when present on mRNA<sup>39,35</sup>.

## 1.9 The epitranscriptome in human disease

RNA modifications have been shown to modulate a variety of RNA processes (translation, splicing, localisation, degradation) and are therefore emerging as crucial regulators in cancer biology and other human diseases. Mutations in roughly half of the known RNA modifying enzymes have links to human diseases<sup>113</sup>. In this list, neurological diseases are highly represented, which correlates well with the high levels of many RNA modifications in neuronal tissues<sup>114,115</sup>.

A variety of cancers have been linked to mutations in RNA modifying enzymes including breast cancer, bladder cancer and leukaemia (see table 1.3). These enzymes can be overexpressed in cancer, as is the case with NSUN2 in breast cancer<sup>116</sup>, or conversely, they can be downregulated, as exemplified by disruption of TRM9L in human colon cancer cells that induces a G0/G1 cell cycle arrest and drastically reduced tumour growth *in vivo*<sup>117</sup>.

Like NSUN2, the METTL3/METTL14 complex components have been implicated as both oncogenes and tumour suppressors depending on the cancer type. In acute myeloid leukemia (AML), lung cancer and liver cancer METTL3 acts as an oncogene by either driving translation of oncogene mRNAs or degrading mRNAs of tumour suppressors<sup>118,110,119</sup>. However, the data for the role of METTL3 in lung cancer is conflicting. Barbieri *et al* found no effect of METTL3 knockout in lung cancer cells, but did describe a novel mechanism for METTL3 deposited m6A in cancer<sup>120</sup>. Here it was shown METTL3 is recruited to chromatin by CEBPZ, where it cotranscriptionally regulates several cancer-linked mRNAs, including those encoding transcription factors SP1 and SP2. This mechanism was demonstrated as

essential for maintaining proliferation in AML. Conversely, a tumour suppressor role of METTL3/METTL14 complex was identified in endometrial cancer. In this case down regulation of METTL3 leads to stabilization of AKT pathway mRNAs and therefore activation of the AKT pathway<sup>121</sup>. Interestingly, the m6A demethylase FTO can also act with pro-oncogenic functions in melanoma by removing m6A from crucial melanoma mRNAs, thereby increasing their stability<sup>122</sup>.

The gene encoding the m7G writer, METTL1, has also been described as both oncogene and tumour suppressor. METTL1 is significantly upregulated in glioblastoma and acts in an oncogenic manner, but in lung cancer has a tumour suppressor function<sup>123,38</sup>. This highlights how context dependent RNA modifications can be. The ultimate effect of an RNA modifying enzyme, be it oncogenic or tumour suppressive, depends on which downstream mRNAs are present in the particular cancer. This difference in effect of the METTL3/14 complex remains poorly understood, but it is likely that downstream cancer-specific targets are determining the outcome. If, for example METTL3 usually enhances expression of a downstream tumour suppressor in healthy cells, it will likely be downregulated in the cancer (like in the case of endometrial cancer). Conversely, if METTL3 is modifying an oncogene mRNA, it becomes upregulated and acts as an oncogene.

RNA modifying enzymes may present potential therapeutic targets. Recent work has shown that an inhibitor of the m6A demethylase FTO could suppress the growth of glioblastoma cells by decreasing the expression level of certain oncogenes<sup>124</sup>. It is increasingly obvious that RNA modifying enzymes can play a role in cancer biology, in either oncogenic or tumour suppressing capacities.

RNA modification		Human disease linked
Short Nomenclature	Full Name	
<b>m1Y</b>	1-methylpseudouridine	Bowen-Conradi Syndrome (EMG1)
<b>s2Um</b>	2-thio-2'-O-methyluridine	Acute infantile liver failure (TRMU) Cancer (TRMU) Deafness (TRMU) MERRF (tRNA/TRMU) MELAS syndrome (tRNA/TRMU) Mitochondrial associated deafness (TRMU) Mitochondrial Infantile Liver Disease (TRMU) Mitochondrial respiratory chain defects (TRMU) MLASA (TRMU)
<b>Um</b>	2'-O-methyluridine	Non-syndromic X-linked mental retardation (FTSJ1)
<b>m5U</b>	5-methyluridine	Breast cancer (TRMT2A)
<b>Y</b>	pseudouridine	Dyskeratosis congenital (DKC1) Pituitary tumorigenesis (DKC1) Prostate cancer (DKC1)  Lactic acidosis (PUS1) Mitochondrial myopathy (PUS1) MLASA (PUS1) Sideroblastic Anemia (PUS1)
<b>Cm</b>	2'-O-methylcytidine	Non-syndromic X-linked mental retardation (FTSJ1)
<b>m3C</b>	3-methylcytidine	Asthma (METTL2B) Breast cancer (METTL6) Lung Cancer (METTL6)
<b>m5C</b>	5-methylcytidine	Autosomal recessive Intellectual disability (NSUN2) Autistic features (NSUN2) Breast cancer (NSUN2) Cancer (NSUN2) Dubowitz syndrome (NSUN2) Intellectual disability syndromes (NSUN2) Noonan-like syndrome (NSUN2) Metabolism (NSUN2)  Hypotonia / Floppy baby syndrome (NSUN3) Lactic acidosis (NSUN3) Leber hereditary optic neuropathy (NSUN3)  Cri du chat syndrome (NOP2)  Breast cancer (TRDMT1) Spina bifida (TRDMT1) Metabolism (TRDMT1)  Williams-Beuren syndrome (WBSCR20/NSUN5)
<b>ac4C</b>	N4-acetylcytidine	Laminopathy (NAT10) Hutchinson-Gilford progeria syndrome (NAT10) Malignant disease (NAT10)
<b>m1G</b>	1-methylguanosine	Diabetes II (TRMT10A) Intellectual disability (TRMT10A)  X-linked intractable epilepsy (TRMT10C) Neurodevelopmental regression (TRMT10C) Multiple Respiratory Chain Deficiencies (TRMT10C) Mitochondrial respiratory chain defects (TRMT10C)  Head and Neck cancer (TRMT5) Multiple Respiratory Chain Deficiencies (TRMT5) Colorectal cancer (RG9MTD2)

<b>Gm</b>	2'-O-methylguanosine	Non-syndromic X-linked mental retardation (FTSJ1) Liver cancer (TARBP1) Cancer (RNMTL1)
<b>m7G</b>	7-methylguanosine	Intellectual disability (WDR4) Primordial dwarfism (WDR4)  Cancer (WBSCR22/TRMT112) Inflammation (WBSCR22/TRMT112) Neoplastic human lung pathology (WBSCR22/TRMT112) Williams-Beuren (WBSCR22/TRMT112)
<b>m2,2G</b>	N2,N2-dimethylguanosine	Intellectual disability (TRMT1)
<b>m2G</b>	N2-methylguanosine	Prostate cancer (TRMT11)
<b>yW</b>	wybutosine	Breast cancer (TRMT12) Leukemia (TRMT12)
<b>m1A</b>	1-methyladenosine	Mitochondrial respiratory chain defects (TRMT10C) Multiple Respiratory Chain Deficiencies (TRMT10C) Neurodevelopmental regression (TRMT10C) X-linked intractable epilepsy (TRMT10C)  Obesity (NML / Nucleomethyin)
<b>Am</b>	2'-O-methyladenosine	Non-syndromic X-linked mental retardation (FTSJ1)
<b>hm6A</b>	N6-hydroxymethyladenosine	Breast cancer (FTO) Cancer (FTO) Coronary heart disease (FTO) Diabetes II (FTO) Developmental delay (FTO) Intellectual disability/Mental retardation (FTO) Leukemia (FTO) Prostate cancer (FTO) Obesity (FTO) Zika virus (FTO)
<b>m6A</b>	N6-methyladenosine	Acute myelogenous leukemia (WTAP) Hypospadias (WTAP) Breast cancer (FTO) Cancer (FTO) Coronary heart disease (FTO) Diabetes II (FTO) Developmental delay (FTO) Intellectual disability/Mental retardation (FTO) Leukemia (FTO) Prostate cancer (FTO) Obesity (FTO) Zika virus (FTO)  Infertility (ALKBH5) Major depressive disorder (ALKBH5)

I	Inosine	<p>Intellectual disability (ADAT3) Strabismus (ADAT3)</p> <p>Aicardi-Goutieres syndrome (ADAR1) Chronic myeloid leukemia [ADAR1] Deliberate self-harm (ADAR1) Esophageal squamous cell carcinoma (ADAR1) Human hepatocellular carcinoma (ADAR1) Metastatic melanoma (ADAR1)</p> <p>ALS (ADAR2)[116-118] Alzheimer's disease (ADAR2) Glioblastoma multiforme (ADAR2)</p>
---	---------	---

**Table 1.3** - List of known RNA modifications with human disease links, their associated enzymes and RNA species where the modification has been found, focussing on methylation modifications and other primary modifications. Adapted from Jonkout *et al*, 2017<sup>125</sup>.

## 1.10 The importance of studying novel RNA modifications

There is mounting evidence that RNA modifying enzymes represent a new class of targets for cancer therapy. Based on a recent CRISPR screen, 174 of the 268 RNA modifying enzymes which were targeted, were essential in at least one cancer cell line<sup>126</sup>. Specific cancer types may have vulnerabilities to disruption of one or several RNA modifying enzymes. There are many factors that make RNA modifying enzymes attractive targets for drugging. For example, many RNA methyltransferases have solved structures, meaning there is potential for rational drug design<sup>127</sup>. Additionally, many RNA methyltransferases belong to the Rossmann fold family, which already contains enzymes which have been successfully drugged. For example, inhibitors of DOT1L, a histone lysine methyltransferase, are in clinical trials for the treatment of AML<sup>128</sup>. Furthermore, inhibitors have also been developed for METTL3<sup>129</sup> and FTO<sup>130</sup>, highlighting the potential for drugging these enzymes.

Most of the work on cancer-linked RNA modifications has focussed on mRNA, but tRNA modifications are gaining more attention. For example, the modification of U34 in the wobble position has been widely linked to cancer<sup>131</sup>. This modification occurs in a three-step process by a number of enzymes. Firstly, U34 is converted to cm<sup>5</sup>U by the elongator complex, followed by conversion into mcm<sup>5</sup>U by ALKBH8, before finally being converted into mcm<sup>5</sup>s<sup>2</sup>U by cytosolic thiouridylases subunits 1 and 2 (CTU1 and CTU2)<sup>131</sup>. This particular tRNA modification is present on three human tRNAs (tRNA<sup>Lys</sup><sub>(UUU)</sub>, tRNA<sup>Glu</sup><sub>(UUC)</sub> and tRNA<sup>Gln</sup><sub>(UUG)</sub>). A recent study found that some of the above RNA modifying proteins are upregulated in melanoma cells<sup>132</sup>. Elongator protein 1 (ELP1), elongator protein 3 (ELP3) and CTU1 and 2 are involved in the modification of U<sub>34</sub> tRNA. Upregulation of these four enzymes brought about specific upregulation of proteins involved in the metabolic switch to glycolysis, which is necessary in cancer cells. Interestingly, this important tRNA modification change did not alter translation at a global level, but rather only of certain mRNAs enriched in specific codons. Taken together this shows that tRNA modifying enzymes can have specific effects in cancer, where the cancer cells become addicted to the expression of RNAs containing sequences decoded by particular tRNAs. Depletion of such enzymes surprisingly does not seem to affect the normal cells in the same way. This raises the possibility of a therapeutic window in which to target these enzymes in cancer, despite their ubiquitous function.

Surveying TCGA data, one can identify many RNA modifying enzymes that are deregulated in cancer. Interestingly, RNA modifying enzymes are highly heterogenous across human tissue types and are often cancer specific. A comprehensive analysis of 13,358 paired tumour-normal human samples has

provided valuable information about which modifications and enzymes are potential regulators of specific cancer types<sup>133</sup>. 40 known RNA modifying proteins were significantly altered in expression level compared to their matched normal samples. Pancreatic cancer in particular appears to have a large proportion of RNA modifying enzymes dysregulated, which is discussed in more detail in section 4.1. Moreover, some RNA modifying enzymes have been shown to affect the chemotherapy sensitivity of some cancers by altering the methylation states of some tRNAs<sup>134</sup>. Therefore, if not direct therapeutic targets, some RNA modifying enzymes could be targeted as sensitizers.

These data show that there are many RNA modifying proteins involved in cancer, beyond the well-studied m6A pathway. CRISPR screens such as these also provide valuable tools for prioritising targets and their related modification pathways. Overall, RNA modifying enzymes act as fine tuners of a host of biological processes. Many of these processes are altered in disease states and can therefore be targeted as vulnerabilities. It is critical that we study these modifications and their pathways to understand the underlying mechanisms to eventually develop inhibitors.

Of particular relevance currently, is the role of RNA modifications in viral infections. m6A has been identified in the genomes and messenger RNAs of a wide range of RNA and DNA viruses including HIV, Influenza A and Herpes Simplex Virus<sup>135</sup>. The effect of m6A marks on splicing is often leveraged by viruses to produce mature RNAs, and m6A methylation inhibitors have been demonstrated to impair proper nuclear processing and export of viral RNA<sup>136</sup>. Not only are viral RNAs themselves often modified, but viral infections may alter the modification pattern in the host cells. For example, both HIV and Zika virus infection have been demonstrated to alter the m6A methylome in host cells<sup>137,138</sup>. Curiously, many of the altered transcripts are involved in viral replication and immune response, suggesting virally driven changes in modifications can promote infection. Furthermore, a recent study of the protein-protein interactions during SARS-CoV-2 Human infection identified the m2,2G methyltransferase TRMT1 as an interactor with the viral protein Nsp5<sup>139</sup>.

## 1.12 Aims of this PhD thesis

The overall purpose of this PhD thesis is to investigate novel substrates of tRNA modifying enzymes and their biological effect in cancer.

- i. Characterise new RNA modification antibodies, by validating their specificity and effectiveness in immunoprecipitation experiments for transcriptome-wide modification mapping.
- ii. Determine transcriptome-wide localisation of m1G RNA modification and the effects of TRMT5 loss in pancreatic cancer cells and mouse models.
- iii. Demonstrate TRMT1 as necessary for m2,2G modification in ovarian cancer cells and characterise the phenotype associated with its loss.

## Chapter 2 - Materials and methods

### 2.1 General materials

#### 2.1.1 Standard solutions

The composition of bacterial media and stock solutions used in this thesis are shown in Table 2.1. All restriction enzymes were purchased from New England Biolabs Ltd or Roche.

Name	Composition per litre
LB (lysogeny broth)	5g NaCl, 5g Yeast extract, 10g bacto-tryptone
SOC	20g bacto-tryptone, 5g bacto-yeast, 0.58g NaCl, 0.19g KCl, 10mM MgCl <sub>2</sub> and 10mM MgSO <sub>4</sub>
TE	10mM Tris, 1mM EDTA, pH8.0
TBE	90mM Tris-borate, 2mM EDTA, pH6.5
TBS	150mM NaCl, 20mM Tris-HCl, pH7.6
PBS	150mM NaCl, 2.5mM KCl, 10mM Na <sub>2</sub> HPO <sub>4</sub> , 2mM K <sub>2</sub> HPO <sub>4</sub>
PBST	PBS + 0.05% Tween 20
RIPA	10mM Tris pH8.0, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 140mM NaCl

**Table 2.1** - The composition of bacterial media and stock solutions.

## 2.1.2 Vectors

The plasmids used for various applications detailed below are summarised in Table 2.2.

Vector	Insert	Application
Lenti CRISPR V2 puro	<b>TRMT1 sg1</b> Forward-CACCGTCATGCAAGGATCGTCTCTG  Reverse-AAACCAGAGACGATCCTTGCATGAC	<i>TRMT1</i> knockout
Lenti CRISPR V2 puro	<b>TRMT1 sg2</b> Forward-CACCGAATACAGCAGCGATGGAGAA Reverse-AAACTTCTCCATCGCTGCTGTATTC	<i>TRMT1</i> knockout
pLKO-Tet-On	<b>TRMT5_sh</b> Forward- CCGGGCAGTTCAGTTCACCTGGTAACTCGAGTTAC CAGGTGAACTGAACTGCTTTTT  Reverse- AATTA AAAAGCAGTTCAGTTCACCTGGTAACTCGA GTTACCAGGTGAACTGAACTGC	TRMT5 inducible shRNA knockdown

**Table 2.2** - Plasmids used

## 2.2 Bacterial strains and transformation

### 2.2.1 *E. coli* strains

Chemically competent bacteria were produced in the Kouzarides lab.

**BL21(DE3):** F- ompT gal dcm lon hsdS<sub>B</sub>(r<sub>B</sub>-m<sub>B</sub>-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])

**DH5a:** F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r<sub>K</sub>-m<sub>K</sub>+), λ-

### 2.2.2 Growth of bacterial cultures

Plated cultures were grown at 37°C overnight on LB plates (LB with 20g bacto-agar per L). Bacterial suspension cultures were grown in LB media at 37°C overnight with shaking at 200rpm. To allow for the selection of bacteria transformed with plasmid DNA expressing the gene of interest, the media was supplemented with either ampicillin (100µg/ml) or kanamycin (100µg/ml) as appropriate.

### 2.2.3 Transformation of *E. coli* by heat shock

For the transformation of *E. coli*, 100ng plasmid DNA or 2µl ligation reaction was added to 50µL of chemically competent bacteria and incubated on ice for 20 minutes. Samples were heat shocked for 45 seconds at 42°C, returned to ice for 3 minutes following which 500µl of SOC media was added. This was incubated at 37°C for 1 hour. Bacteria were pelleted by centrifugation for 5 minutes at 500g, resuspended in 100µl SOC media and then plated.

## 2.3 Preparation of DNA

### 2.3.1 Large scale (maxi prep) plasmid DNA preparation from *E.coli*

The Qiagen Maxiprep kit (Qiagen, 12963) was used according to the manufacturer's instructions. 300ml LB media supplemented with antibiotic was inoculated with a single colony of the appropriate transformant and grown overnight at 37°C with shaking at 200rpm. Bacteria were harvested by centrifugation at 3500rpm for 10 minutes. The bacterial pellet was resuspended in 10ml P1 solution (50mM Tris-HCl, pH8.0, 10mM EDTA, 100µg/µl RNase A) and then lysed by the addition of 10ml P2 solution (200mM NaOH, 1% SDS). Following 5 minutes incubation at room temperature, 10ml ice-cold P3 solution (3.0M potassium acetate pH5.5) was added, the lysate was gently mixed by inversion and incubated on ice for 20 minutes. The lysate was cleared by centrifugation for 30 minutes at 4000rpm, 4°C. The cleared lysate was applied through muslin to a Qiagen 500 (anion-exchange) column, previously equilibrated with 10ml QBT buffer (750mM NaCl, 50mM MOPS, pH7.0, 15% isopropanol, 0.15% Triton-X100). The column was washed twice with 30ml QC buffer (1.0M NaCl, 50mM MOPS, pH7.0, 15% isopropanol). Plasmid DNA was eluted from the column with 15ml buffer QF (1.25M NaCl, 50mM Tris pH 8.5, 15% isopropanol) and precipitated by the addition of 10.5ml isopropanol. The DNA precipitate was centrifuged for 30 minutes at 4000rpm, 4°C. The DNA pellet was washed with 70% ethanol and air dried prior to being dissolved in TE or nuclease-free water. DNA concentration was determined by measuring the absorbance at 260nm using a NanoDrop® spectrophotometer.

### 2.3.2 Small scale (mini prep) plasmid DNA preparation from *E.coli*

The QIAprep Miniprep Kit (Qiagen, 27106) was used according to the manufacturer's instructions. 3ml LB media with the appropriate antibiotic was inoculated with a single bacterial colony and grown at 37°C with shaking at 200rpm overnight. Bacteria were pelleted by centrifugation at 3500rpm for 5 minutes. The bacterial pellet was resuspended in 250µl P1 buffer (50mM Tris-HCl pH8.0, 10mM EDTA 100µg/ml RNase A) and lysed by the addition of 250µl P2 buffer (200mM NaOH, 1% SDS). Following

incubation at room temperature for 5 minutes, 350µl ice-cold N3 buffer was added, the solution mixed by inversion and incubated on ice for 5 minutes. The debris was pelleted by centrifugation and the supernatant transferred to a Qiagen plasmid spin column and centrifuged again for 1 minute. The column was sequentially washed in 500µL PB buffer, 750µl PE buffer and centrifuged for 1 minute. After the final wash, the column was spun a second time to remove any residual buffer and then the DNA eluted in EB buffer (10mM Tris-HCl, pH8.5) by centrifugation for 1 minute.

## 2.4 DNA cloning

Several precautions were taken to maximise the generation of recombinant plasmids. (i) Both vector and insert fragments were gel-purified to exclude the presence of undigested plasmids in the ligation reaction. (ii) Recircularisation of vector fragments was minimised in two ways: (a) Where possible DNA fragments were inserted into vectors cut with two enzymes generating incompatible ends. (b) If compatible ends were generated, the 5' terminal phosphate was removed from vector DNA with alkaline phosphatase. Ligation by T4 DNA ligase requires presence of 5' phosphate groups on only one of the two DNA fragments.

### 2.4.1 DNA restriction digests and dephosphorylation

0.1-1µg DNA was digested with 5-10 units of restriction endonuclease for 1 hour at 37°C. The enzyme buffer and incubation temperature used were as recommended according to the manufacturer's instruction (New England Biolabs). 10 units of calf intestinal alkaline phosphatase (Roche) was added directly to the completed restriction digest of the vector DNA and incubated for a further 15 minutes 37°C. 6X agarose gel loading buffer (0.25% bromophenol blue, 30% glycerol in H<sub>2</sub>O) was added to the digestion before gel electrophoresis.

### 2.4.2 Agarose gel electrophoresis

DNA fragments were resolved using horizontal agarose gels. 0.8-2.0% agarose gels were made by dissolving agarose in 150ml Tris/Borate/EDTA (TBE) buffer, before adding ethidium bromide to a final concentration of 0.5µg/ml. The gel was poured into a tray and allowed to set. Once set, the gel was mounted in an electrophoresis tank and covered with TBE. DNA samples were loaded onto the gel in 1X agarose gel loading buffer, and electrophoresed at 140V. 1kb ladder (ThermoFisher, 10787026) was run alongside samples as a ladder.

### 2.4.3. Purification of DNA from agarose gels

DNA that was stained with ethidium bromide was visualised in an agarose gel by low wavelength UV light. The relevant fragments were excised from the gel with a scalpel. DNA was extracted from the agarose using the QIAquick Gel Extraction kit (Qiagen, 28704), according to the manufacturer's instructions. Three volumes of QG buffer were added to the agarose slice and this was incubated at 50°C for 10 minutes with shaking. 1 gel volume of isopropanol was added to the dissolved agarose, the mix was then applied to the Qiagen spin-column and centrifuged with the DNA being retained in the column. The column was then washed with 750µl of PE buffer and centrifuged for 1 minute. Any residual buffer was removed by a second spin. DNA was eluted in 30µL of EB buffer.

### 2.4.4 Ligations

Purified, linearized vector and insert DNA were ligated with T4 DNA ligase, using the Rapid DNA ligation Kit (ThermoFisher, K1423) according to the manufacturer's instructions. A molar ratio of 1:3 or 1:7 (vector: insert) DNA and 5 units of T4 DNA ligase was used as appropriate. The ligation reaction was carried out for 1 hour at room temperature or overnight at 4°C before transformation into chemically competent *E. coli*.

## 2.5 Rabbit monoclonal antibody production (RabMAb®)

### 2.5.1 Antigen preparation and immunization

Rabbit monoclonal RNA modification antibodies were provided by Abcam and were not commercially available at the time this PhD project was started. The exact details of the immunogen and immunisation protocol are commercially sensitive and not available for this thesis.

## 2.6 Antibody characterisation

### 2.6.1 Modified RNA oligonucleotide design

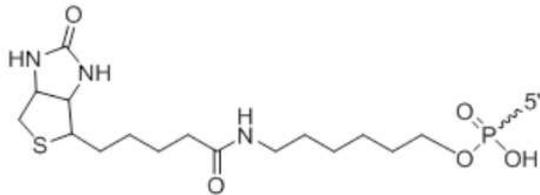
Two distinct RNA oligonucleotide designs were used during antibody screening and characterisation. Oligonucleotides labelled "TK" were designed at the Tony Kouzarides laboratory and those labelled "mN" were designed at Abcam Plc. The structural differences between the two designs are shown below. 2'-O-methylated nucleosides were used in the Abcam designs to improve RNase resistance

during long term storage. The TK oligos were designed with two adjacent modifications because this arrangement can occur naturally. For example, m6,6A is found in two adjacent positions in rRNA.

**TK:** 5' - rNrNrNrNrNrNrNrN (mod)(mod) rNrNrNrNrNrN (mod) rNrNrNrNrNrN (mod) rNrNrNrN (3'-2LCBi)

**mN:** 5'-Bi – mNmNmNmNmN (Mod\*) mNmNmNmNmN – 3'

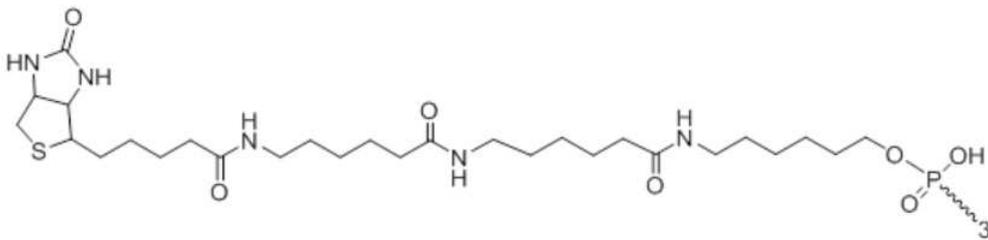
**5'-Bi:**



**Unit Molecular Weight:**

406.46

**2LCBi:**



**Unit Molecular Weight:**

632.78

**rN** - Mixed ribonucleoside bases in the same ratio (rA:rG:rC:rU = 25%:25%:25%:25%).

**mN** - Mixed 2'-O-methyl-nucleoside bases in the same ratio (mA:mC:mG:mU = 25%:25%:25%:25%).

**2LCBi** - Biotin is widely used throughout biological research to conjugate proteins and nucleic acids for biochemical assays. Because both streptavidin and avidin bind biotin with high affinity ( $K_d$  of 10<sup>-14</sup> mol/L to 10<sup>-15</sup> mol/L) and specifically, biotinylated proteins or nucleic acids of interest can be isolated from a sample by exploiting this highly stable interaction (<http://dharmacon.horizondiscovery.com/rnai-and-custom-rna-synthesis/custom-rna-synthesis/bases-and-modifications/>)

### 2.6.2 RNA Oligonucleotide panel ELISA

A panel of modified RNA oligonucleotides containing either a single modified nucleotide (mN) or multiple modified nucleotides (TK) (see above), were used to assess antibody specificity. Neutravidin was diluted in PBS to 10µg/ml and used to coat Nunc maxisorp plates (Thermo Fisher, 449824) at 100µl per well overnight at 4°C. Plates were then washed with PBST three times and blocked with 2% (w/v) bovine serum albumin in PBS for 1 hour at room temperature. RNA oligonucleotides were diluted to three concentrations (5µg/ml, 0.5µg/ml and 0.05µg/ml) in PBS and added to the appropriate wells in triplicate, and then incubated for 1 hour at room temperature. The plates were then washed again three times with PBST, before adding the primary antibody (1µg/ml in 2% (w/v) BSA, PBS) and then incubated for 1 hour at room temperature. Plates were then washed three times with PBST before adding secondary antibody (Abcam anti-rabbit ab6721/Dako anti-sheep P0163) at 1:1000 dilution in 2% BSA, PBS. The secondary antibody was then incubated for an hour at room temperature. Final washing of three PBST washes, followed by two PBS washes (to eliminate bubbles) were carried out and 50µl TMB substrate (VWR, 95059-290) was added for a colorimetric readout. Plates were developed for 1-15minutes until saturation before adding 50µl 0.25M sulphuric acid to stop the reaction. Plates were then measured for absorbance at 450nm on a Pherastar plate reader (BMG Labtech).

### 2.6.3 Free nucleoside competition ELISA

A streptavidin coated plate (ThermoFisher, 15218) was blocked for 1 hour at room temperature with 2% (w/v) BSA (blocking buffer). The plate was then washed three times with PBST. 100ng of m1G RNA oligonucleotide was diluted to 1µg/ml and 100µl was added to each well of a streptavidin coated 96-well plate, before incubating for 1 hour at room temperature. The plate was then washed again with PBST three times. m1G and m7G free nucleosides (Carbosynth, NM08574 and NM08037) were diluted to 0.2µM, 2µM, 4µM, 8µM, 16µM, 20µM and 40µM and 50µl was added to appropriate wells in triplicate. The anti-m1G RabMAb® was then diluted to 2µg/ml and 50µl was added to each well. This was then incubated for 1 hour at room temperature. The plate was then washed three times with PBST and anti-rabbit secondary antibody (Abcam anti-rabbit ab6721) was diluted 1:1000 in blocking buffer and 100µl was added to each well. The plate was then incubated for a further hour at room temperature. The plate was then washed a further three times with PBST and twice with PBS before adding 100µl Supersignal™ ELISA substrate (ThermoFisher, 37069). The plate was then analysed using a Clariostar plate reader (BMG Labtech).

#### 2.6.4 RNA oligonucleotide panel HTRF® assay

The same panel of modified RNA oligonucleotides as described in section 2.6.2 were used to confirm antibody specificity determined using the HTRF® (Homogenous Time Resolved Fluorescence) assay. A Lumi4®-terbium cryptate conjugated anti-rabbit secondary antibody (CisBio, 61PARTAF) was used as a FRET donor molecule and streptavidin D2 (CisBio, 610SADLF) was used as the FRET acceptor. The donor molecule was diluted to 1:2000 in PBS and the acceptor molecule was diluted to 1:250 in PBS. Modified RNA oligonucleotides were diluted to 5µg/ml, 0.5µg/ml and 0.05µg/ml in PBS. The antibodies to be tested were diluted to 1µg/ml in PBS. 5µl of each: FRET donor, FRET acceptor, antibody and RNA oligonucleotide were combined in wells of a 384 well plate in triplicate. After one-hour plates were read for emission at 665nm and 620nm on a PHERAstar plate reader (BMG Labtech). Emissions at 620nm (donor) are used as an internal reference, while emissions at 665nm are used as an indicator of the antibody binding. Final values for cross-reactivity were calculated using %  $\Delta F$  ([www.cisbio.com/drug-discovery/htrf-ratio-and-data-reduction](http://www.cisbio.com/drug-discovery/htrf-ratio-and-data-reduction)), which represents the signal to background of the assay.

#### 2.6.5 meRIP-mass spectrometry (methylated RNA immunoprecipitation-mass spectrometry)

100-150µg total RNA or 5-20µg polyA<sup>+</sup> RNA was fragmented to 100-200nt average size using RNA fragmentation reagents following the manufacturer's instructions (AM8740, ThermoFisher). 20-50µg of the fragmented RNA was diluted in a final volume of 18µl with RNase free water on ice. 2µl fragmentation reagent was quickly added to each tube whilst remaining on ice. The samples were mixed and briefly spun before incubation for 18 minutes at 70°C. The reaction was then stopped by removing all the tubes from the heat block simultaneously and placing on ice, followed by the addition of 2µl stop solution. The fragment sizes were confirmed using an Agilent Tapestation 2200 with RNA screentape (Agilent, 5067-5576).

Fragmented RNA was diluted with RNase free water to 755µl and heated at 75°C for 5 minutes to disrupt secondary structures, before being immediately placed on ice for >1 minute. 400U RNaseOUT™ (Life Technologies Ltd, 10777-019), 12.5µg RNA modification antibody and 200µl 5X IP buffer (50mM Tris-HCl pH7.4, 750mM NaCl, 0.5% Igepal CA-630) were then added to the RNA. The reaction was incubated overnight at 4°C on a rotating mixer to allow the antibody to bind to the modification of interest.

Protein A Dynabeads™ (Life Technologies Ltd, 100-04D) and Protein G Dynabeads™ (Life Technologies Ltd, 100-02D) were combined at a 1:1 ratio (allowing for 50µl per reaction) and washed with 1X IP buffer (10mM Tris-HCl pH7.4, 150mM, 0.1% IGEPAL® CA-630) and then resuspended in 1ml 1X IP buffer containing 0.5mg/ml bovine serum albumin (BSA) to block non-specific binding to the beads. The beads were then incubated for 2 hours at 4°C with rotation. The beads were then washed three times with 1X IP buffer and resuspended in the original volume. 50µl of blocked beads were used for each IP reaction.

The pre-blocked beads were then added directly to the fragmented RNA/antibody complex prepared previously. The reactions were then incubated for 2 hours at 4°C with rotation. The following wash steps were then used to reduce non-specific binding: low salt (10mM Tris pH7.4, 75mM NaCl, 0.1% Igepal CA-630), high salt (10mM Tris pH7.4, 200mM NaCl, 0.1% Igepal CA-630) and TEN buffer (10mM Tris-HCl pH7.4, 1mM EDTA, 0.05% Igepal CA-630). The RNA was then eluted using one of the two following methods:

- 1) For MS analysis of RNA modifications, the immunoprecipitated RNA was eluted from the beads with an SDS elution buffer + Proteinase K (5mM Tris pH7.5, 1mM EDTA, 0.05% SDS, 5µl Proteinase K) and incubated at 50°C for 2 hours with shaking.
- 2) For meRIP-seq, immunoprecipitated RNA was eluted by competition with free modified nucleoside. Each IP reaction was eluted with 450µl competition elution buffer (90µl 1X IP buffer, 6mM free nucleoside, 7µl RNaseOUT and RNase free water up to 450µl), then incubated for 2 hours at 4°C

Eluted RNA was recovered using Trizol LS reagent (ThermoFisher, 10296028), followed by isopropanol precipitation. 1ml of QIAzol was added to each sample before vortexing and then incubated for 5 minutes on ice. 200µl chloroform was then added to each sample and vortexed. The reactions were subsequently incubated for 2 minutes at room temperature. The samples were then centrifuged at 12 000g for 15 minutes at 4°C. The aqueous phase was transferred to a fresh tube and 500µl isopropanol followed by 1µl of GlycoBlue™ (AM9515, Thermo Fisher) was added to each sample. The reactions were then precipitated overnight at -80°C. The precipitated RNA pellet was washed twice with cold 75% ethanol before dissolving in RNase free water. The RNA was quantified using a Qubit HS RNA assay kit (ThermoFisher, Q32852) and analysed for size distribution using Agilent high sensitivity screentape (Agilent, 5067-5579). If any RNA was successfully pulled down by the antibody being used, the RNA was sent for MS analysis by collaborators at Storm Therapeutics.

## 2.6.6 Mass spectrometry (MS)

Mass spectrometry analysis of RNA modifications was carried out in collaboration with Storm Therapeutics.

Nucleosides were prepared from RNA by enzymatic digestion, using a cocktail of Benzonase (Merck), Phosphodiesterase 1 (Merck), and Antarctic Phosphatase (New England Biolabs) as described previously<sup>140</sup>. The reactions were filtered using an Amicon 30kDa MWCO spin-column (Merck) to remove protein and the filtrate was mixed with a 2x loading buffer containing 0.1% formic acid and an internal standard (<sup>13</sup>C-labeled uridine generated from 645672-1MG Merck KGaA, previously treated with Antarctic Phosphatase). The samples were loaded onto an ACQUITY UPLC HSS T3 Column, 100 Å, 1.8µm, 1 mm X 100 mm (Waters Corp., Milford, MA, USA) and resolved using a gradient of 2%–10% acetonitrile in 0.1% formic acid over 10 minutes. Mass spectrometric analysis was performed in positive ion mode on an Orbitrap Q Exactive HF (Thermo Fisher, Waltham, MA, USA) mass spectrometer. Standard dilutions of all experimental nucleosides were prepared and analysed in parallel. There were three technical replicates of each sample and the analytical processing was performed using XCalibur Software (Thermo Fisher).

## 2.7 Cell culture

### 2.7.1 Cell lines

Various cancer cell lines were maintained in the Kouzarides' laboratory, Gurdon Institute, Cambridge. The cells were cultured in the media outlined below, in 10cm petri dishes, until approximately 80% confluence before passaging. To passage, the media was aspirated, and the cells were washed with 10ml PBS, before adding 1ml 1X trypsin-EDTA. The plates were then incubated at 37°C, with frequent agitation until the cells had detached. 3ml media was then added to inactivate the trypsin, and the cells were divided as necessary. Media was then added to each plate up to a volume of 10-12ml.

Cell line	Media
hTERT-HPNE	RPMI 20% FBS with 1% penicillin/streptomycin
hTERT-HPNE E6/E7/K-RasG12D	RPMI 20% FBS with 1% penicillin/streptomycin
PANC-1	DMEM 10% FBS with 1% penicillin/streptomycin
JHOC-5	DMEM 10% FBS with 1% penicillin/streptomycin

**Table 2.3** - Cell lines and the growth media used.

### 2.7.2 HEK293-6e cells

HEK293EBNA1-6E cells were maintained at Abcam Plc, Cambridge. The cells were passaged every two days, seeding at  $3 \times 10^5$  cells/ml. They were grown in Freestyle™ 293 Expression medium at 37°C, 5% CO<sub>2</sub>, with shaking at 120rpm. Cell pellets were used for bulk RNA extraction for meRIP-MS experiments.

### 2.7.3 siRNA gene knockdown

The ON-TARGETplus siRNA (Dharmacon) reverse transfection system was used to knockdown potential RNA methyltransferase enzymes. siRNAs were dissolved in RNase free water to 20µM. siRNA reverse transfections were performed in either 10cm dishes, 6-well plates or 96-well plates depending on the experiment. Protocols for each size transfection are outlined below.

#### 10cm dishes:

Cells were passaged the day before transfection into 10cm dishes. 16µl Dharmafect I was added to 784µl Opti-MEM for each transfection and incubated at room temperature for 5 minutes. Meanwhile, 40µl siRNA was added to 240µl Opti-MEM media (ThermoFisher, 31985070). 800µl of Dharmfect I-Opti-MEM mixture was added to each siRNA reaction and incubated at room temperature for 15 minutes. During the incubation time, cells were trypsinised and resuspended in DMEM + 10% FBS without penicillin/streptomycin. 7ml of cell suspension was added to a 10cm dish for each transfection and the transfection reaction was added to the cells. The following day, media was changed to DMEM + 10% FBS + penicillin/streptomycin. Transfected cells were either harvested at 48 hours, 72 hours or transfected for a second time to improve knockdown efficiency.

**6-well dishes:**

Cells were passaged on the day of transfection, and 120 000-200,000 cells were added to a single well in 2ml DMEM 10% FBS without penicillin/streptomycin. For each transfection 4µl Dharmafect I and 196µl Opti-MEM media were combined and incubated for 5 minutes at room temperature. During the incubation, 10µl of siRNA was added to 90µl Opti-MEM. The 100µl siRNA+ Opti-MEM was then added to the Dharmafect I- Opti-MEM and incubated for 15 minutes at room temperature. The entire 300µl transfection mixture was then added dropwise to the well containing cells. The following day the media was refreshed with DMEM 10% FBS + penicillin/streptomycin.

**96-well plates:**

10 000 cells were added per well in 85µl DMEM 10% FBS without penicillin/streptomycin. For each well transfection 0.2µl Dharmafect I and 9.8µl Opti-MEM were combined and incubated for 5 minutes at room temperature. During the incubation, 0.5µl of siRNA was added to 4.5µl Opti-MEM. The 10µl Dharmafect I–Opti-MEM was then added to the siRNA and incubated for 15 minutes at room temperature. The 15µl transfection mixture was then added to the cells. The following day the media was refreshed with DMEM 10% FBS + penicillin/streptomycin.

**Protocol for double transfection:**

In most cases a double transfection was used. An initial transfection was carried out as above and then 48 hours later the cells were trypsinised, counted and re-used in a second transfection.

#### 2.7.4 shRNA gene knockdown

The pLKO-Tet-On inducible shRNA system (Addgene) was used to knockdown TRMT5 in PANC-1 cells. AgeI and EcoRI were used to clone the oligo into the pLKO-Tet-on vector. Oligos were reconstituted in ddH<sub>2</sub>O to 0.1 nmol/ $\mu$ l before vortexing for a few seconds to mix and dissolve. 11.25 $\mu$ l of each oligo and 2.5 $\mu$ l 10X annealing buffer (1M NaCl, 100 mM Tris-HCl, pH7.4) were mixed. They were then placed in a heat block at 95°C, before switching off the heat block to allow the reaction to cool to room temperature. 1 $\mu$ l of the oligo mixture was diluted 1:400 in 0.5X annealing buffer (1 $\mu$ l + 399 $\mu$ l 0.5X buffer). pLKO-Tet-On vector was digested with AgeI and EcoRI and gel-purified. Ligation reactions were set up as follows: oligo dilution (or 0.5X buffer as a negative control) 20ng gel-purified digested pLKO-Tet-On, 1 $\mu$ l 10X ligase buffer 1 $\mu$ l T4 DNA ligase 1 $\mu$ l Water 6 $\mu$ l Total 10 $\mu$ l. Ligations were incubated at room temperature for 1-3 hours.

#### 2.7.5 CRISPR knockout

The LentiCRISPR V2<sup>141</sup> system was used to generate a knockout of RNA methyltransferases TRMT5 and TRMT1. sgRNAs were designed to target the first exon. 293T cells were transfected with the lentivCRISPRv2 TRMT5 sgRNAs together with psPAX2 (Addgene#12260) and VsVg (Addgene#8454). Virus particles were harvested 48 hours post-transfection. Target cells were then transduced with the virus with polybrene to increase the efficiency of transduction. 48 hours after infection transduced cells were selected for by treatment with puromycin (2 $\mu$ g/ml).

#### 2.7.6 Overexpression of TRMT5

pCMV6 plasmid containing human TRMT5 gene was transfected into PANC-1 cells using FuGENE<sup>®</sup>HD transfection reagent (E2311, Promega). 200,000 PANC-1 cells were diluted in 2ml DMEM +10% FBS media and seeded into 6-well dishes. 6 $\mu$ g plasmid DNA was added to 188 $\mu$ l opti-MEM. 12 $\mu$ l FuGENE<sup>®</sup> transfection reagent was then added and incubated for 15 minutes at room temperature. This was then subsequently added to the cells dropwise.

#### 2.7.7 Cell proliferation assay

120,000 PANC-1 cells or 150,000 cells were plated in 6-well dishes (Thermo Fisher, 140675). At each time point the media was aspirated and the cells were washed in 1ml sterile PBS. Cells were then trypsinised with 0.5ml 1X trypsin-EDTA and neutralised with 0.5ml of media. The contents of the wells were transferred to 15ml centrifuge tubes and centrifuged at 1300rpm for 3 minutes, before being resuspended in 1ml of media. 10 $\mu$ l of cell suspension was then used for trypan blue counting.

### 2.7.8 Wound healing assay

Cells were grown to confluence in 6-well dishes. A single scratch was produced in each well using a 1ml pipette tip. The monolayer was then washed to remove debris and fresh media was added. To ensure the same field was imaged at each time point, a line was drawn on the underside of the well perpendicular to the scratch. Each image was then taken with the line on the edge of the field of view as a point of reference. Images were analysed using ImageJ. The size of the wound was measured using the ImageJ plugin MRI wound healing tool ([http://dev.mri.cnrs.fr/projects/imagej-macros/wiki/Wound\\_Healing\\_Tool](http://dev.mri.cnrs.fr/projects/imagej-macros/wiki/Wound_Healing_Tool)).

### 2.7.9 Transwell® migration assay

The day before the Transwell® assay cells were passaged in the morning. The cells were then serum starved overnight in DMEM + penicillin/streptomycin media. The following day, 0.44ml DMEM 10% FBS + penicillin/streptomycin was added to the bottom of a 24-well plate. Transwell® chambers were then added to the wells with sterile tweezers, ensuring no bubbles were introduced. The cells were then washed, trypsinised and neutralised in DMEM 2% FBS + penicillin/streptomycin. The cells were then washed a second time before being resuspended in DMEM + penicillin/streptomycin serum free at 300,000 cells per ml. 0.25ml of cell suspension (75,000 cells) was then added to each Transwell® chamber. The plates were then incubated overnight at 37°C. The following day the Transwells® were removed and washed by submerging in PBS. The migrated cells were then fixed and stained using crystal violet and imaged on Zeiss stemi 305 microscope. Migration was quantified using ImageJ software and the colony counting function or colony area plugin<sup>142</sup>.

### 2.7.10 OP-puro global protein synthesis analysis

For analysis of global protein synthesis, a commercial kit was used from (ab235624, Abcam). Transfected cells or CRISPR knockout cell lines were grown to ~80% confluency in 6-well plates before fixation, permeabilization and staining as outlined in the kit protocol. 350µl fixative solution was added to each well containing cells and incubated for 15 minutes at room temperature. The fixative solution was then removed, and the wells were washed with 500µl 1X wash buffer. 350µl permeabilization buffer was then added to each well and incubated for 10 minutes at room temperature. During the incubation a 1X protein labelling cocktail was prepared with the following ratios: 93µl PBS, 1µl copper reagent (100X), 1µl fluorescent azide (100X) and 5µl reducing agent (20X). The permeabilization buffer was removed and 350µl 1X protein labelling cocktail was added to each well. The plates were then incubated in the dark for 30 minutes at room temperature. The stain was

removed, and the wells were then washed with 500µl 1X wash buffer. 350µl of 1X DNA stain was then added to each well and incubated for 20 minutes at room temperature in the dark. Cells were then harvested using 1X trypsin and scraping into FACS tubes. The cells were then centrifuged for 5 minutes at 1500g and washed once with 1X PBS. Staining was then measured by flow cytometry using a BD Aria IIu cell sorter and subsequently analysed using FlowJo.

#### 2.7.11 Global RNA synthesis analysis

For analysis of global RNA synthesis, a commercial kit was used (ab228561, Abcam). Transfected cells or CRISPR knockout cell lines were grown to 80% confluency before following the kit protocol. 350µl fixative solution was added to each well containing cells and incubated for 15 minutes at room temperature, protected from light. The fixative solution was then removed, and the wells were washed twice with 500µl 1X wash buffer. 350µl permeabilization buffer was then added to each well and incubated for 10 minutes at room temperature. During the incubation an RNA reaction and DNA staining cocktail was prepared with the following ratios: 93µl PBS, 1µl copper reagent (100X), 1µl fluorescent azide (100X) and 5µl reducing agent (20X). The permeabilization buffer was removed and 350µl 1X protein labelling cocktail was added to each well. The plates were then incubated in the dark for 30 minutes at room temperature. The stain was removed, and the wells were then washed three times with 500µl 1X wash buffer. 350µl of 1X DNA stain was then added to each well and incubated for 20 minutes at room temperature in the dark. Cells were then harvested using 1X trypsin and scraping into FACS tubes. The cells were then centrifuged for 5 minutes at 1500g and washed once with 1X PBS. Staining was then measured by flow cytometry using a BD Aria IIu cell sorter and subsequently analysed using FlowJo software.

## 2.8 RNA extraction, treatment, and analysis

### 2.8.1 Total RNA extraction by precipitation

RNA was extracted using QIAzol lysis reagent (Qiagen Ltd, 79306). Cell pellets were washed with PBS and 1ml QIAzol was used to lyse  $\sim 10^6$  cells. Cells were mixed by pipetting until homogenous, followed by a 5-minute incubation at room temperature to allow complete lysis. 200 $\mu$ l chloroform was added per 1ml of QIAzol reagent and the reaction was mixed vigorously for >30 seconds by inversion, followed by a 3-minute incubation at room temperature. RNA was separated by centrifugation at 10,000rpm, at 4°C for 20 minutes. The RNA-containing aqueous phase was transferred to a fresh tube. 500 $\mu$ l isopropanol was added per 1ml of QIAzol reagent to precipitate the RNA and the reaction was incubated for either 10 minutes at room temperature (for high RNA concentrations) or >1 hour at -80°C when expected RNA concentration was low. The precipitation reactions were then centrifuged at 10,000rpm, at 4°C for 20 minutes to pellet the RNA. The RNA pellet was then washed with ice cold 75% ethanol (v/v) to remove residual salts and retain RNA. The pellet was then air dried for 10 minutes, before being dissolved in RNase free water and quantified using a Nanodrop (ThermoFisher) or Qubit HS RNA.

### 2.8.2 DNase treatment of total RNA

Turbo DNA free™ kit (ThermoFisher, AM1907) was used to remove potential DNA contaminants from RNA samples. A maximum of 15 $\mu$ g RNA was incubated with 1 $\mu$ l Turbo DNase at 37°C for 30 minutes. Inactivation reagent included in the kit was used to stop the reaction, followed by treatment at 65°C for 15 minutes. The inactivation reagent was removed by centrifugation at 10,000g for 2 minutes. The treated RNA was then transferred to a fresh tube and purified using Zymo clean and concentrator™ 25 columns (Cambridge Bioscience Ltd, R1018) to remove DNase and buffer.

### 2.8.3 Total RNA extraction by column

RNA was extracted using QIAzol lysis reagent as in section 2.9.1. Once aqueous phase was transferred to a fresh tube, 1.5 volumes 100% ethanol was added. The mixture was then transferred to RNAeasy spin columns and centrifuged for 30 seconds at 8000g. 350 $\mu$ l buffer RW1 was then added to wash the column and centrifuged for 30 seconds at 8000g. 10 $\mu$ l DNase I stock solution was added to 70 $\mu$ l RDD buffer and mixed. 80 $\mu$ l of the solution was then added to each RNAeasy column and incubated for 15 minutes at room temperature to remove DNA contamination. Another 350 $\mu$ l RW1 buffer was added to the column followed by centrifugation for 30 seconds at 8000g. This was followed by two washes

with 500µl RPE buffer, the first was performed for 30 seconds at 8000g, followed by a second for 2 minutes at 8000g. RNA was then eluted in RNase free water.

#### 2.8.4 PolyA<sup>+</sup> RNA enrichment from total RNA

Dynabeads™ mRNA Purification Kit (ThermoFisher, 61006) was used to enrich for polyA<sup>+</sup> RNA from total RNA, following manufacturer's instructions.

#### 2.8.5 mRNA quantification

##### **First strand cDNA synthesis:**

The Tetro cDNA synthesis kit™ (Bioline) was used according to manufacturer's instructions to synthesise first strand cDNA. 1µg purified RNA was mixed with 4µl 5XRT Buffer, 1µl reverse transcriptase (200U/µl), 1µl RNase inhibitor (10U/µl), 1µl dNTP mix (10mM total) and 1µl random hexamer primer mix. The reaction volume was made up to 20µl with RNase free water. The reactions were incubated for 10 minutes at 25°C, followed by 30 minutes at 45°C. The reactions were then terminated at 85°C for 5 minutes before being held at 4°C. cDNA was then diluted 1:4 for use in qPCR.

##### **Quantitative real-time PCR (qPCR):**

cDNA synthesised as above was used to quantify mRNA levels. Quantitative real-time PCR was performed with the ABI Prism 7300 (Applied Biosystems) sequence detection system, using SYBR Green as the detector. A 10µl reaction was used on 96-well plates (5µl SYBR Green master-mix, 0.4µl forward primer (10µM), 0.4µl reverse primer (10µM), 2µl cDNA (diluted 1:4 in RNase free water), 2.2µl RNase free water). PCR amplification was performed with an initial step of 10 minutes of 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 55-60°C. Primers are summarised below in table 2.4.

Target gene	Sequence
TRMT5	AGAAATCTTGAGAGCTGTGCTT
	TGGAAATTTCCGGTACATATTGTCA
TRMT10A	ATTCAACGATGTTACGCAGAAAAC
	TTTGTCAATTTTCATCCATGTTT
TRMT10B	TACATCCTCGGTGGGCTTGT
	ATGATAGTTTTTCCATTCTGG
GAPDH	TGCACCACCAACTGCTTAGC
	GGCATGGACTGTGGTCATGAG
Nestin	CAGCCTGGAGGTGGCCACGTACA
	TAAGAAAGGCTGGCACAGGT
Nestin m1G region	CTGGAGCAGTCTGAGGAAGTG
	GGACTCTCTATCTCCTCCCTCT
Nestin negative region	AGAGGAGGCGTTCCCTGCTGAGA
	ATCTTCCAGGATCGGGGTGTACGTT
BRD4	TGGAGCTTGCGGGCCATGGCCA
	CTAAACTGGAGGCCCGTGAGTAC

**Table 2.4** - Primers used in qPCR

### 2.8.6 ELISA on total RNA

DNA coating solution (ThermoFisher, 17250) was used to capture total RNA or polyA+ RNA directly onto white 96-well maxisorp plates (Fisher scientific, 10394751). Total RNA was added to wells containing 100µl coating solution and incubated at room temperature for 2 hours. After RNA immobilisation, wells were washed three times with PBST. Primary antibodies were diluted in PBST (0.05% tween 20) at 1µg/ml and 100µl was added to the immobilised RNA, before incubation at room temperature for 1 hour. Wells were then washed again three times with PBST. HRP conjugated anti-rabbit secondary (Abcam, ab205718) was then diluted in PBST at 1:5000 and 100µl was added to each well, before incubation at room temperature for 1 hour. Wells were washed again three times with PBST and 100µl ECL substrate (ThermoFisher, 37070) was added. Plates were then analysed immediately using a Clariostar or PHERAstar plate reader (BMG Labtech).

### 2.8.7 meRIP-seq

Carried out as described in section 2.6.5, followed by library preparation.

### 2.8.8 meRIP-qPCR for m1G validation

Large RNA (>200nt) was prepared from total RNA using Zymo RNA clean and concentrator 25 (Zymo, R1017) as outlined in the manufacturer's instructions. This was then fragmented to ~200nt as previously described in section 2.6.5. 5µg large RNA was used per immunoprecipitation reaction with 10µg anti-m1G antibody using the same protocol as section 2.6.5. All eluted RNA was used to produce cDNA and for subsequent qPCR.

### 2.8.9 Library preparation from meRIP-seq

Immunoprecipitated RNA was converted into cDNA libraries for sequencing using NEXTflex Rapid Directional RNA-seq kit (Newmarket Scientific, NOVA-5138-08). Libraries were multiplexed at equal molarities and submitted for 50bp paired-end sequencing on an Illumina HiSeq 1500®.

### 2.8.10 Oxford Nanopore Direct RNA sequencing (DRS)

PolyA<sup>+</sup> RNA was purified as described in section 2.8.4 and analysed by Nanopore direct RNA sequencing. The nanopore direct sequencing was carried out by Dr Paulo Amaral, Kouzarides' laboratory as described in a recent publication<sup>65</sup>.

### 2.8.11 Ribosome profiling

Ribosome profiling was carried out at the Loayza-Puch laboratory, German Cancer Research Center (Deutsches Krebsforschungszentrum) as follows. PANC-1 cells underwent knockdown using either scramble, TRMT5 or TRMT10A siRNA as described in 2.7.3. Cells were treated with cycloheximide (100 µg/ml) for 8 to 10 minutes, washed with ice-cold phosphate-buffered saline containing cycloheximide (100 µg/ml), pelleted, and lysed in buffer A (20 mM Tris-HCl, pH 7.8, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 1% Triton X-100, 2 mM DTT, 100 µg/ml cycloheximide, 1X complete protease inhibitor). Lysates were centrifuged at 5,000 rpm and the supernatant was treated with 2U/µl of RNase I (Invitrogen, Grand Island, NY, USA) for 40 minutes at room temperature. Lysates were fractionated on a linear sucrose gradient (7% to 47%) using the SW-41Ti rotor at 36,000 rpm for 2 h. Fractions enriched in monosomes were pooled and treated with proteinase K (Roche, Mannheim, Germany) in a 1% SDS solution. Released RNA fragments were purified using Trizol reagent and precipitated in the presence of

glycogen. For library preparation, RNA was gel-purified on a denaturing 10% polyacrylamide urea (7 M) gel. A section corresponding to 30 to 33 nucleotides, the region where most of the ribosome-protected fragments are comprised, was excised, eluted and ethanol precipitated. The resulting fragments were 3'-dephosphorylated using T4 polynucleotide kinase (New England Biolabs Inc. Beverly, MA, USA) for 6 h at 37°C in 2-(N-morpholino)ethanesulfonic acid (MES) buffer (100 mM MES-NaOH, pH 5.5, 10 mM MgCl<sub>2</sub>, 10 mM β-mercaptoethanol, 300 mM NaCl). 3' adaptor was added with T4 RNA ligase 1 (New England Biolabs Inc. Beverly, MA, USA) for 2.5 h at 37°C. Ligation products were 5'-phosphorylated with T4 polynucleotide kinase for 30 minutes at 37°C. 5' adaptor was added with T4 RNA ligase 1 for 18 h at 22°C.

## 2.9 Protein purification, detection and analysis

### 2.9.1 Lysate preparation for Western blot

Urea lysis buffer (50mM Tris pH6.8, 6M urea, 2% SDS) + protease inhibitor cocktail was added to cell pellets to lyse the cells at a ratio of 5 volumes lysis buffer to 1 volume of pellet. The lysates were immediately placed on ice and incubated for 15 minutes. During the incubation the cells were pipetted up and down to aid lysis. Lysates were then sonicated for 3 cycles of 30 seconds ON, followed by 30 seconds OFF. Lysates were then centrifuged for 5 minutes, 8000g at 4°C. The supernatant was then transferred to a fresh tube. Protein concentration was then determined using Qubit Protein Assay kit (Thermo Fisher, Q33211).

### 2.9.2 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

10-50µg of cell lysate was resolved using NuPAGE® 4-12% Bis Tris Gels (Life Technologies Ltd, NP0336BOX). When detecting proteins of 10-30kDa, MES SDS running buffer was used (Life Technologies Ltd, NP0002) and gels were run for 30 minutes at 200V. When detecting proteins of >30kDa, MOPS SDS running buffer was used (Life Technologies Ltd, NP0001) and gels were run for 55 minutes at 200V.

### 2.9.3 Western blotting

Proteins were transferred to nitrocellulose membranes (Life Technologies Ltd, LC2000) using the mini-trans blot cell (BioRad). Following electrophoresis, gels were transferred into Tris glycine transfer buffer (384mM glycine, 25mM Tris). The gel was sandwiched next to nitrocellulose (Amersham

Protram), between 2 pieces of Whatman™ 3MM paper and between 2 sponge pads. The sandwich was inserted into the transfer chamber filled with transfer buffer ( such that the nitrocellulose was facing the positive anode. Proteins were transferred to the nitrocellulose by applying a constant current (0.5A) for 1.5 hours, using an ice block to chill the buffer. After the transfer, the nitrocellulose membrane was stained with Ponceau S to confirm equal transfer. Membranes were then blocked for at least 1 hour in 2% milk or 5% BSA in TBST (TBS with 0.1% Tween-20 (v/v)), at room temperature with gentle agitation. The membranes were then probed with primary antibody, diluted to an assay dependant concentration in blocking buffer, overnight at 4°C with gentle agitation. Blots were then washed four times for 10 minutes each with TBST. Secondary antibodies (Abcam, anti-mouse ab6789 and anti-rabbit ab6721) were diluted to 1:5000 and 1:40000 respectively in blocking buffer. The membranes were then incubated with secondary antibody for 1 hour at room temperature, with gentle agitation. The membranes were then washed four times for 10 minutes each with TBST. Membranes were then developed using ECL following manufacturer's instructions (Promega, W1015). After incubation, excess fluid was drained off and the membrane was exposed to X-ray film (Fuji) or analysed using a Chemidoc™ imaging system (BioRad).

#### 2.9.4 Cellular fractionation

Subcellular fractionation was carried out as described in a recent paper by Gillotin *et al*<sup>143</sup>.

## 2.9.5 Antibodies

The following antibodies were used in the listed concentrations for Western blotting, immunoprecipitation and flow cytometry assays.

Antibody	Description	WB	ChIP
Anti-TRMT5	Rabbit polyclonal, Sigma HPA000843-100µl	1/500	NA
Anti-TRMT5	Rabbit polyclonal, PTG	NA	4µg
Anti-TRMT10A	Rabbit polyclonal, Sigma HPA047601-100µl	1/500	NA
Anti-TRMT10B	Rabbit polyclonal, Sigma SAB2702016	1/500	NA
Anti-TRMT1	Rabbit polyclonal, Abcam ab186019	1µg/ml	4µg
Anti-GAPDH	Rabbit monoclonal, Abcam ab181602	1/10000	NA
Anti-β-Tubulin	Rabbit polyclonal, Abcam ab6046	1/500	NA
Anti-Vinculin	Mouse monoclonal, Abcam ab130007	1µg/ml	NA
Anti-Nestin	Rabbit monoclonal, Abcam ab105389	1/100	NA

**Table 2.5** - Antibodies and the associated assay dependent concentrations.

## 2.10 Bioinformatics

### 2.10.1 meRIP-seq

Reads were mapped against the reference genome sequence (GRCh38, v98) using *STAR* aligner (v2.7.1). Single mapper reads were kept using *samtools* tools and PCR duplicates were removed by *Picard Markduplicates* tool. *BAMscale scale* tool generated scaled bigWig tracks of all samples for visualization purposes. Using all biological replicates per condition, *MACS2* (v2.2.5) was used to roughly estimate peak size (300bp) before peak calling. Called peaks between WT-IP and WT-input samples (TRMT5 and TRMT10) by *MACS2* (--nomodel --extsize 300bp). In parallel, the *MetDiff R*

package was used to call peaks following the same experimental design (WT-IP and WT-input). Significant peaks from both *MACS2* and *MetDiff* were merged in a single bed file for reference and counts under peaks per replicate were obtained by the *featureCounts* function (subread suite). The *edgeR* and *DESeq2* R package was used to filter (min. 20 reads along all samples) and to normalize read counts per peak. Wald test was used to detect differential binding between WT vs KO conditions considering input and IP samples (FDR = 0.10), as well as variances between two different RIP-seq experiments.

### 2.10.2 Ribo-seq

Reads were mapped against the reference genome sequence (GRCh38, v98 - longest CDS reference) using *bowtie2*, where rDNA and mtRNA were previously filtered from FASTA files. Single mapper reads were kept using *samtools* tools and duplicates were removed by *umi\_tools dedup* tool. After alignment, P-site coverage of CDS reads (codon occupancy analysis) and whole transcript coverage (transcriptional efficiency analysis) were produced using own R scripts. Transcriptional efficiency analysis was performed between wild type and knockout conditions. After filtering transcripts with fewer than 20 reads, RPF and input reads were normalized (median of ratios normalization) and compared using the *Xtail* R package. Only the first *Xtail* pipeline was considered to select up and down TE genes (p-value = 0.05 and absolute TE log2FC > 0.5). Codon occupancy analysis was performed by comparing observed and expected reads on P-site per codon from every transcript. After summarizing observed count reads per codon, linear regression against expected counts (reference sequence) was performed. Residuals to the linear trend were calculated and compared between wild type and knockout replicates by *t-test*. For every codon, t-statistics from all genes were considered to assess significant enrichment/depletion (p-value < 0.001) in a particular subset of genes (i.e. up or down TE).

### 2.10.3 ChIP-seq

Reads were mapped against the reference genome sequence (hg38) using ENCODE ChIP-seq pipeline phase 3 (<https://github.com/ENCODE-DCC/chip-seq-pipeline2>). Duplicated, low quality and multi-mapped reads were removed, and scaled track files (.bw files) were also generated from these final alignment files (.bam files). No binding differences were found between input and IP samples, using *MACS2* for peak calling (--extsize 50) on separate replicates.

## 2.11 Chromatin immunoprecipitation (ChIP)

~10-20 million cells were used per ChIP reaction. The medium was poured off each plate of cells to be used and 15ml 1% formaldehyde solution diluted in FA buffer (50mM HEPES-KOH, 100mM NaCl, 1mM EDTA, 5mM EGTA) was added to each plate. The cross-linking reaction was left for 10 minutes at room temperature. 1/20<sup>th</sup> volume 2.5M glycine was added to each plate to quench the reaction, followed by two washes with cold PBS. The cells were then removed by scraping and centrifuged at 1500g for 5 minutes at 4°C. The supernatant was discarded, and the pellet was snap frozen in liquid nitrogen for later use.

50µl Protein A and 50µl Protein G Dynabeads were combined for each ChIP reaction and washed twice with 1ml PBS + 0.5% BSA before being resuspended in the original volume. Antibody was added to each aliquot of washed beads and incubated for 4 hours, at 4°C, with agitation. The cross-linked cell pellet was resuspended in 10ml LB1 buffer (50mM HEPES-KOH pH7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X100) and then incubated for 10 minutes at 4°C on a roller. The cells were centrifuged at 2000g for 4 minutes at 4°C, and the supernatant was discarded. The pellet was then resuspended in 10ml LB2 buffer (10mM Tris-HCl pH8, 200mM NaCl, 1mM EDTA, 0.5mM EGTA) and incubated for 5 minutes at 4°C on a roller. The cells were centrifuged at 2000g for 5 minutes at 4°C and the supernatant was discarded. The pellet was resuspended in 3ml LB3 buffer (10mM Tris-HCl pH8, 100mM NaCl, 1mM EDTA 0.5mM EGTA, 0.1% Na deoxycholate, 0.5% N-laurylsarcosine) and 20µl was taken as an unsonicated control.

15ml sonication tubes were used with ~200µl sonication beads (Diagenode). The beads were washed with PBS before being used. 500µl of lysate was added to each sonication tube and vortexed. 20 cycles for PANC-1 cells and 17 cycles for JHOC-5 cells were used (30 seconds ON followed by 30 seconds OFF). The tubes were combined to give 6ml of nuclear extract and stored on ice. 20µl from each reaction was removed to check the fragment size on a gel: 180µl decrosslinking buffer (200mM NaCl, 0.5% SDS) + 4µl Proteinase K (20mg/ml) + 10µl RNase (33791500, Roche) was added to each 20µl and incubated for 1hour at 65°C. The DNA was then purified on a ChIP clean and concentrator column (Zymo). Once the size was checked, the IP reactions were set up as follows.

The bead/antibody mix was added to the nuclear extract and incubated overnight at 4°C with rotation. The following day the beads were washed four times with RIPA buffer and once with TBS. The elution was then carried out in three steps. Firstly, 200µl elution buffer was added to the beads (50mM Tris-

HCl pH8, 10mM EDTA, 1% SDS) and incubated for four hours at 65°C. Secondly 200µl TE buffer and 15µl RNase was added and incubated at 37°C for 30 minutes. Thirdly, 4µl proteinase K was added and incubated at 55°C for 1 hour. 2% input sample was treated in the same way as CHIP samples, by taking 120µl of input and adding 80µl elution buffer. After elution, all samples were purified on CHIP clean and concentrator columns (Zymo) and eluted in 42µl.

## Chapter 3 - Generation of RNA modification specific antibodies

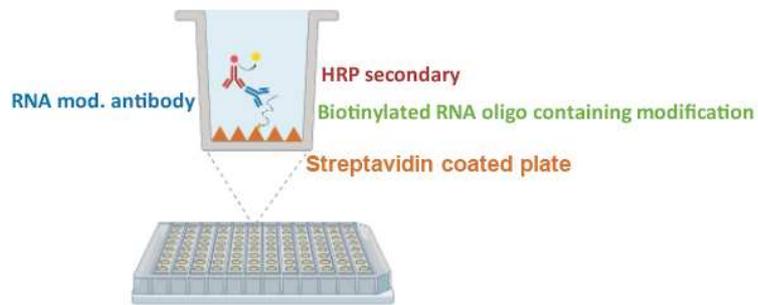
### 3.1 Introduction

The use of antibodies, and particularly antibodies coupled to next generation sequencing, has been crucial to the explosion of research on RNA modifications. However, these techniques are not without their pitfalls. It has become clear that antibody cross-reactivity and batch variability can account for false positives. It is therefore critical to apply thorough characterisation to each new antibody.

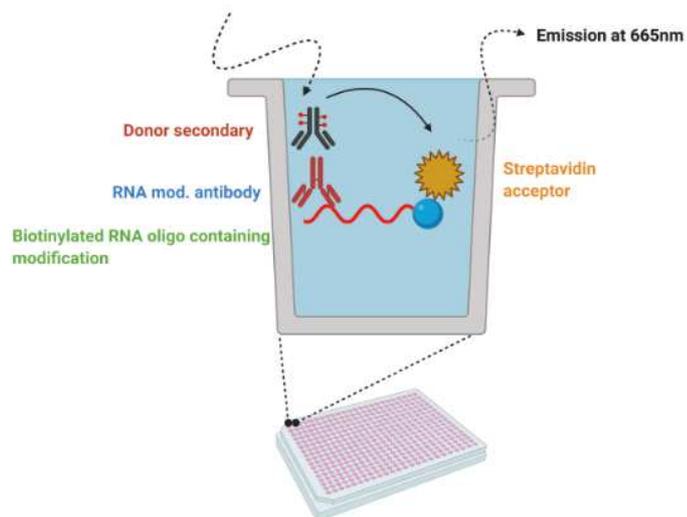
In order to study novel modifications, or to search for novel substrates, a collaboration between Abcam Plc and the Kouzarides lab was established. Rabbit monoclonal antibodies (RabMAbs<sup>®</sup>) were generated by Abcam Plc and the subsequent characterisation was carried out as part of this thesis to determine which antibodies would be suitable for use in biological experiments. This chapter covers the characterisation of these antibodies and feeds into the subsequent biological and mechanistic investigation. Antibody performance in the assays here contributed to the choice of modifications to pursue in later work.

Modified RNA oligonucleotides are commercially available and can be used in simple binding assays, such as dot blot and ELISA, to determine the specificity of new antibodies. However, as mentioned previously, there is an ever-expanding list of RNA modifications and the modified oligonucleotides available are not exhaustive. Moreover, some of these RNA modification oligonucleotides are prohibitively expensive for research purposes. Standard characterisation protocols in the RNA modification field, often involve dot blots. Dot blot assays can be informative tests for antibody validation, but require large amounts of oligonucleotide to work, typically in the range of several micrograms. In this work, a panel of methylation modified RNA oligonucleotides was chosen, to interrogate the antibodies. An ELISA type assay where the RNA was captured onto a 96-well plate was used (figure 3.1A), and a homogenous FRET (fluorescence resonance energy transfer) type assay termed HTRF<sup>®</sup> was also used (figure 3.1B). Both of these assays require significantly less RNA oligonucleotide material than dot blot assays and therefore allow testing of a broader range of modifications.

A



B



**Figure 3.1 - Specificity characterisation assays. A)** Diagram of anti-RNA modification antibody ELISA characterisation. **B)** Diagram of anti-RNA modification antibody HTRF<sup>®</sup> characterisation.

HTRF<sup>®</sup> (homogenous time resolved fluorescence) is an immunoassay that is becoming popular as a replacement for high throughput ELISA screening. The technology combines FRET and time resolved measurement (TR). A fluorescent signal from the donor is produced after excitation by an energy source. Depending on proximity, this then triggers an energy transfer to the acceptor, which in turn produces fluorescence at a specific wavelength. Traditional FRET assays are plagued by background emission, but this background fluorescence is short lived and can therefore be eliminated by introducing a time delay (50-150µseconds) between excitation and measurement. Because this is carried out in a homogenous format, there is no need for separation of unbound components, eliminating the need for washing or centrifugation steps. Additionally, some antibodies and antigens do not interact efficiently in ELISA assays due to the surface interference. In this work, surface interference could mask the binding of antibodies to particular modifications leading to false negatives. In addition, the HTRF<sup>®</sup> assay is routinely carried out in 384 well format, therefore less reagent is used. The modified RNA oligonucleotides are in some instances very costly, so limiting the amount used has potential economic advantages for future antibody validation. Furthermore, the HTRF<sup>®</sup> assay is much quicker to carry out than ELISA.

The principle applications of RNA modification antibodies are based on immunoprecipitation, but not all antibodies can perform in these types of assays. It is therefore vital that any potential antibodies are tested for their ability to immunoprecipitate the modification of interest. Here we used MS to measure potential enrichment of the modification of interest. However, this assay also has limitations; for example, if an antibody designed to be specific to m1G is cross reactive to m2G, it may appear to be enriching m1G from a mixed population. However, this apparent enrichment could be due to co-precipitation of m1G to nearby m2G nucleotides. It therefore cannot be ruled out that the modification of interest is not co-precipitating with another cross-reactive modification. Nonetheless, in conjunction with specificity screening, we can be reasonably confident in our antibodies. As demonstrated in later chapters, confidence in the antibodies can be further increased with miCLIP experiments when there are known substrates and modification positions. Furthermore, using a non-antibody-based technique to validate potential modification sites is critical to be confident of new modification sites. Such techniques could be based on MS, reverse transcriptase signatures, nanopore or specific chemistry.

## 3.2 Aims of this chapter

- i. Characterise new anti-RNA modification antibodies by ELISA and HTRF<sup>®</sup> to assess specificity.
- ii. Determine modification enrichment levels by RNA immunoprecipitation followed by mass spectrometry (meRIP-MS).

## 3.3 Results

### 3.3.1 Choice of modifications and antibody generation strategy

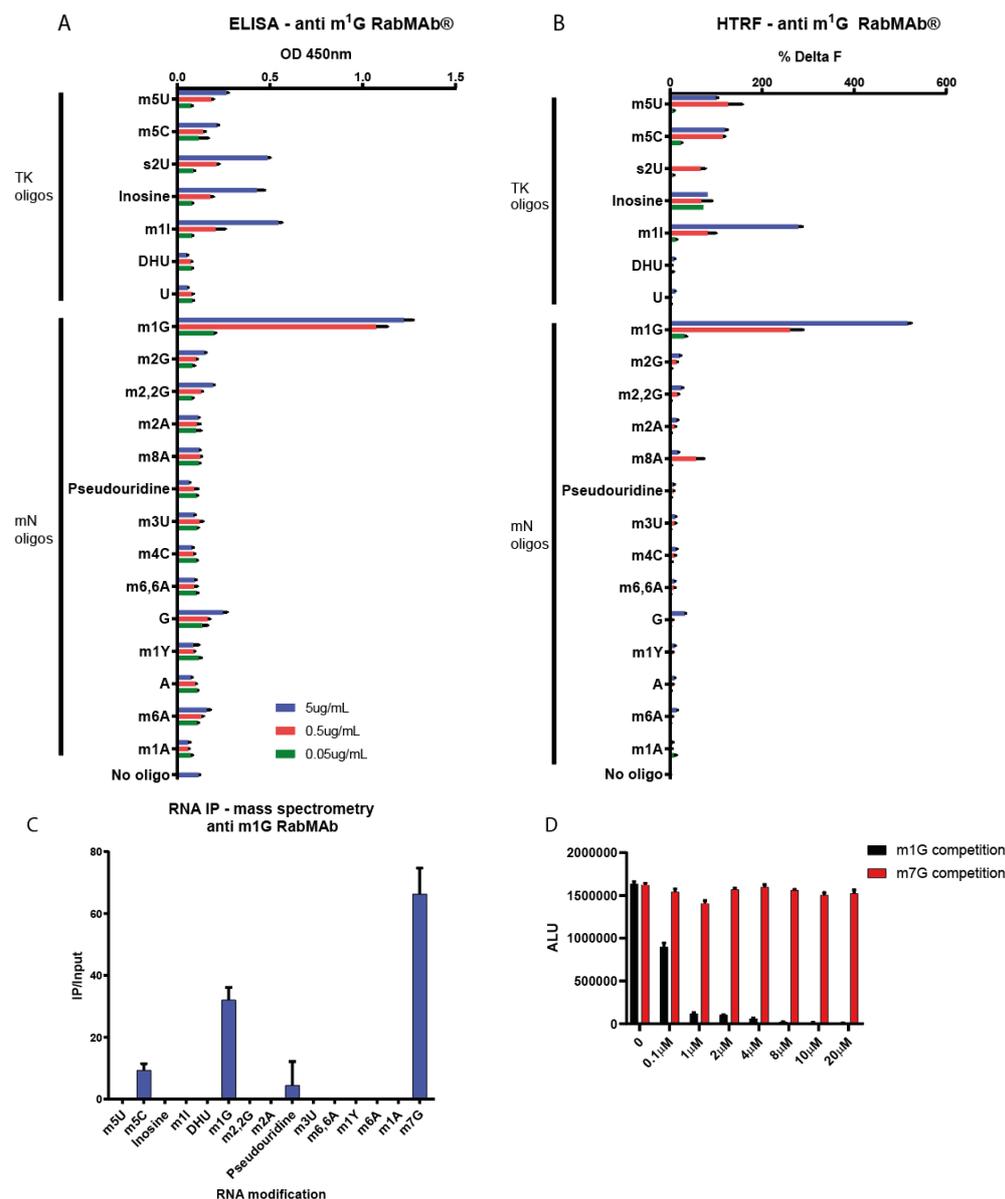
Target modifications for these new antibodies were chosen based on several criteria. Firstly, primary modifications were prioritised to avoid the confounding factors of multi-step pathways. Secondly, given methylation is the most common type of RNA modification, we decided to focus on these. In addition, the Kouzarides lab has decades of experience working with methyltransferases, so methylation modifications were a natural choice. Finally, targets were prioritized where there was no existing commercially available antibody.

### 3.3.2 Anti-m1G characterisation

The anti-m1G antibody was first validated against a panel of modified RNA oligonucleotides in both ELISA and HTRF<sup>®</sup> (figure 3.2A and B respectively). Figure 3.2A shows that by ELISA, at the highest concentration of oligonucleotide, there was a small degree of cross-reactivity to inosine, m1I and s2U. By HTRF<sup>®</sup> assay, only minor cross-reactivity to m1I at the highest concentration of oligonucleotide was observed (figure 3.2B). The HTRF<sup>®</sup> data did not show the same cross-reactivity to inosine and s2U as ELISA, rather it appears to be background noise. For these three potential cross-reactive modifications, the ELISA data are more reliable because the signal increases with oligonucleotide concentration in a dose-dependent manner. Therefore, the anti-m1G antibody is largely specific for m1G modification in the context of modifications tested by both ELISA and HTRF<sup>®</sup>, but there may be low level cross-reactivity to a number of other modifications.

The anti-m1G antibody was then used for immunoprecipitation assays followed by meRIP-MS of the eluted RNA. In figure 3.2C, we show that the modification of interest is enriched compared to input RNA by ~35 fold. This indicates the antibody is useful for immunoprecipitation and potentially transcriptome-wide mapping experiments. Other modifications besides m1G were enriched above the

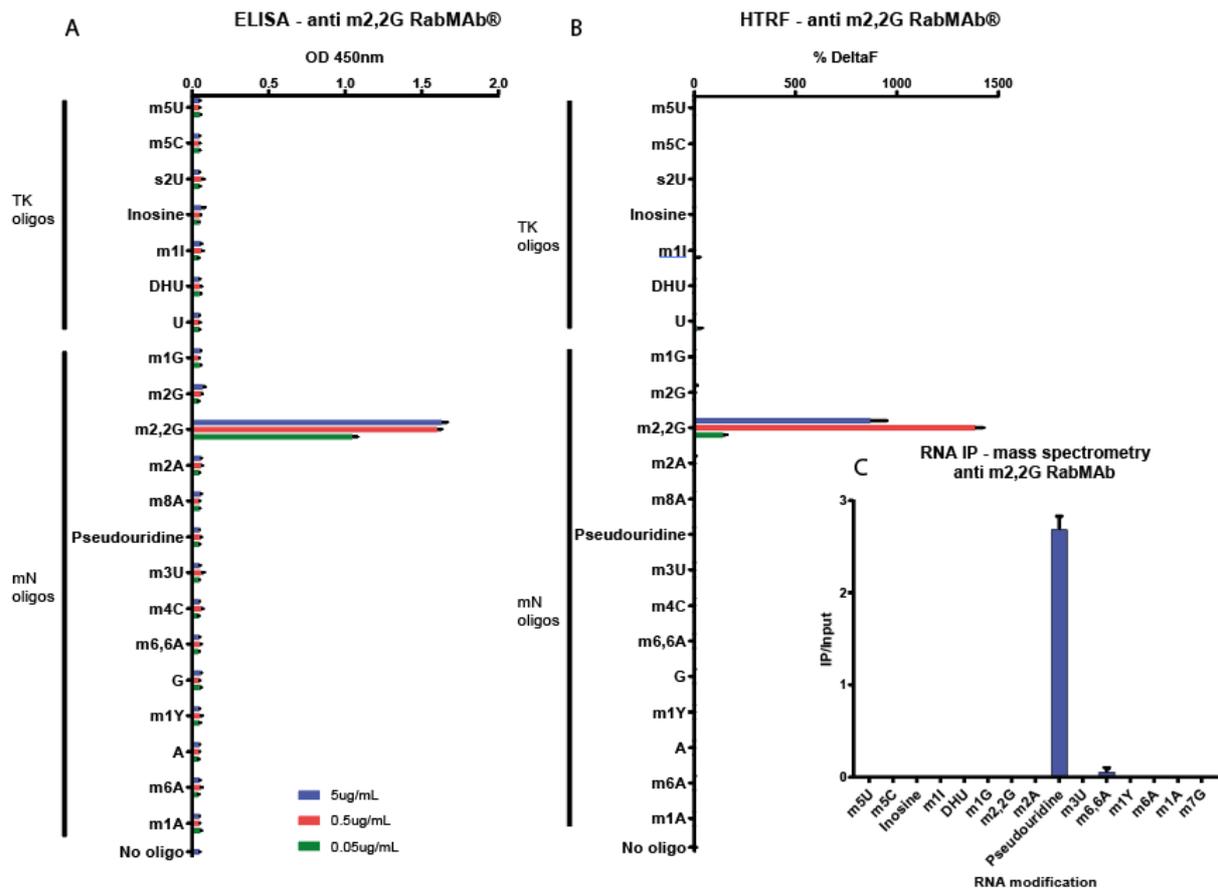
level in input RNA, notably m7G, pseudouridine and m5C. It is possible that this is antibody cross-reactivity but could be attributed to co-precipitation of these modifications because of their proximal association with m1G and therefore present on the same fragments. For example, if m1G modifications are colocalising with these pseudouridine and m5C modifications, I would expect to see an enrichment in the meRIP-MS assay despite lack of antibody cross-reactivity. Unfortunately, m7G containing RNA oligonucleotides were not available from commercial suppliers. I therefore used free m7G and m1G nucleosides in a competition ELISA to determine which were able to interfere with the antibody binding the m1G oligonucleotide. Using this assay m1G free nucleoside can compete with the m1G oligonucleotide therefore reducing signal. However, even at 20 $\mu$ M, free m7G nucleoside did not compete the antibody away from binding the m1G oligonucleotide.



**Figure 3.2 - Anti-m<sup>1</sup>G RabMAb® characterisation.** A panel of biotinylated RNA oligonucleotides were used in both ELISA and HTRF® assays. TK and mN describe two distinct oligonucleotide designs which are described in section 2.7.1. **A)** RNA oligonucleotides were captured onto streptavidin coated plates in triplicate, at the three concentrations stated. These were probed with an anti-m<sup>1</sup>G primary antibody, followed by detection with HRP conjugated secondary antibody. Binding was quantified by colorimetric reaction and absorbance at 450nm on a BioTek plate reader. **B)** RNA oligonucleotides were used in an HTRF® assay to assess antibody specificity. The assay was designed with goat anti-rabbit Tb (Terbium cryptate) as the FRET donor and streptavidin d2 as the acceptor molecule. Terbium cryptate emission is measured at 620nm and streptavidin d2 emission is measured at 665nm. Binding is shown as %DeltaF (signal-background/background). The emission was measured on a PHERAstar plate reader (BMG). **C)** methylated RNA immunoprecipitation was carried out on total RNA derived from HEK293-6e cells. After bound RNA was eluted from the antibody, it was hydrolysed to single nucleotides and analysed by RNA MS. Each sample was analysed in three technical replicates. Each modification is represented as normalised to the relevant non modified nucleoside, and then IP/Input. **D)** m<sup>1</sup>G RNA oligonucleotide was used in ELISA as described in A), but with the addition of varying concentrations of m<sup>1</sup>G or m<sup>7</sup>G free nucleoside with the primary antibody step.

### 3.3.3 Anti-m<sup>2,2</sup>G characterisation

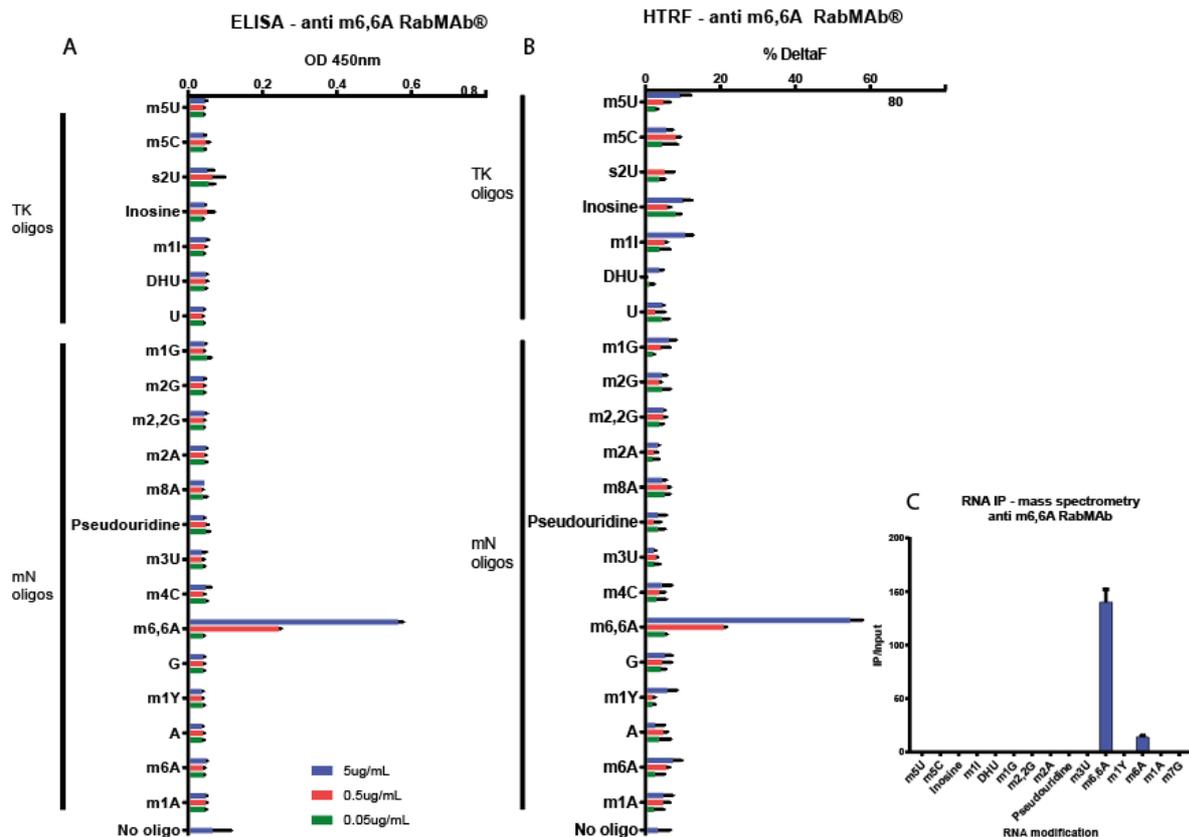
The anti-m<sup>2,2</sup>G antibody was first validated by ELISA and HTRF<sup>®</sup> against the panel of modified RNA oligonucleotides. By both ELISA (figure 3.3A) and HTRF<sup>®</sup> (figure 3.3B) the antibody showed exceptional specificity for the target modification. Furthermore, this specificity was observed at all three oligonucleotide concentrations, showing the antibody can be highly sensitive too. However, when used in meRIP-MS experiments (figure 3.3C), the antibody did not enrich the modification of interest above input and there was only a very small enrichment of a non-target modification, pseudouridine.



**Figure 3.3 - Anti-m2,2G RabMAb® characterisation.** A panel of biotinylated RNA oligonucleotides were used in both ELISA and HTRF® assays. TK and mN describe two distinct oligonucleotide designs which are described in section 2.7.1. **A)** RNA oligonucleotides were captured onto streptavidin coated plates in triplicate, at three concentrations. These were probed with an anti-m2,2G primary antibody, followed by detection with HRP conjugated secondary antibody. Binding was quantified by colorimetric reaction and absorbance at 450nm on a BioTek plate reader. **B)** RNA oligonucleotides were used in an HTRF® assay to assess antibody specificity. The assay was designed with goat anti-rabbit Tb (Terbium cryptate) as the FRET donor and streptavidin d2 as the acceptor molecule. Terbium cryptate emission is measured at 620nm and streptavidin d2 emission is measured at 665nm. Binding is shown as %DeltaF (signal-background/background). The emission was measured on a PHERAstar plate reader (BMG). **C)** methylated RNA immunoprecipitation was carried out on total RNA derived from HEK293-6e cells. After bound RNA was eluted from the antibody, it was hydrolysed to single nucleotides and analysed by RNA MS. Each sample was analysed in three technical replicates. Each modification is represented as normalised to the relevant non modified nucleoside, and then IP/Input.

### 3.3.4 Anti-m6,6A characterisation

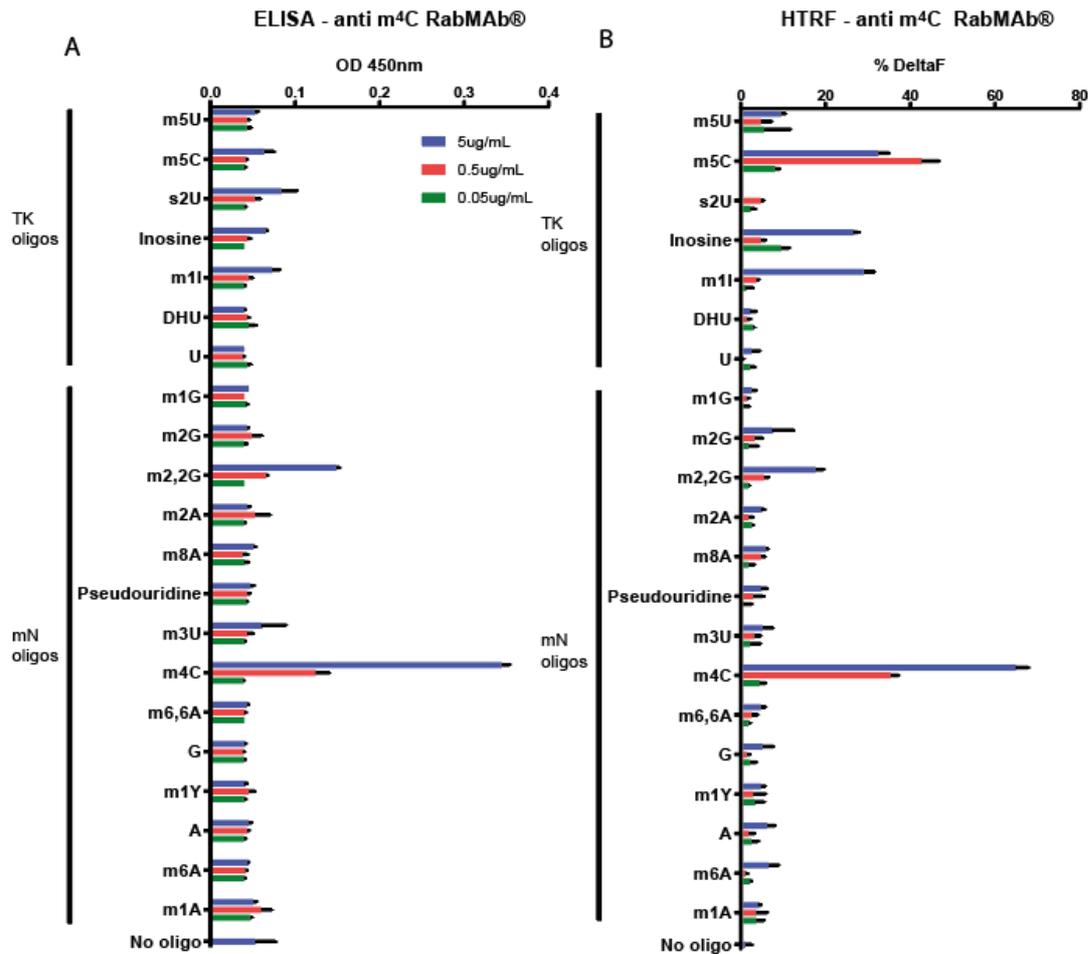
The anti-m6,6A antibody was first validated by ELISA and HTRF® against the panel of modified RNA oligonucleotides. This antibody has good specificity towards m6,6A compared to the other modifications in the panel, as seen by both ELISA (figure 3.4A) and HTRF® (figure 3.4B). Moreover, the sensitivity of this antibody is comparatively high, giving signal at 0.5µg/ml oligonucleotide. The antibody was then tested in meRIP-MS experiments (figure 3.4C). This assay indicates that this antibody is capable of enriching m6,6A modification from total RNA above input (~150 fold).



**Figure 3.4 - Anti-m6,6A RabMAb® characterisation.** A panel of biotinylated RNA oligonucleotides were used in both ELISA and HTRF® assays. TK and mN describe two distinct oligonucleotide designs which are described in section 2.7.1. **A)** RNA oligonucleotides were captured onto streptavidin coated plates in triplicate, at three concentrations. These were probed with an anti-m6,6A primary antibody, followed by detection with HRP conjugated secondary antibody. Binding was quantified by colorimetric reaction and absorbance at 450nm on a BioTek plate reader. **B)** RNA oligonucleotides were used in an HTRF® assay to assess antibody specificity. The assay was designed with goat anti-rabbit Tb (Terbium cryptate) as the FRET donor and streptavidin d2 as the acceptor molecule. Terbium cryptate emission is measured at 620nm and streptavidin d2 emission is measured at 665nm. Binding is shown as %DeltaF (signal-background/background). The emission was measured on a PHERAstar plate reader (BMG). **C)** methylated RNA immunoprecipitation was carried out on total RNA derived from HEK293-6e cells. After bound RNA was eluted from the antibody, it was hydrolysed to single nucleotides and analysed by RNA MS. Each sample was analysed in three technical replicates. Each modification is represented as normalised to the relevant non modified nucleoside, and then IP/Input.

### 3.3.5 Anti-m4C characterisation

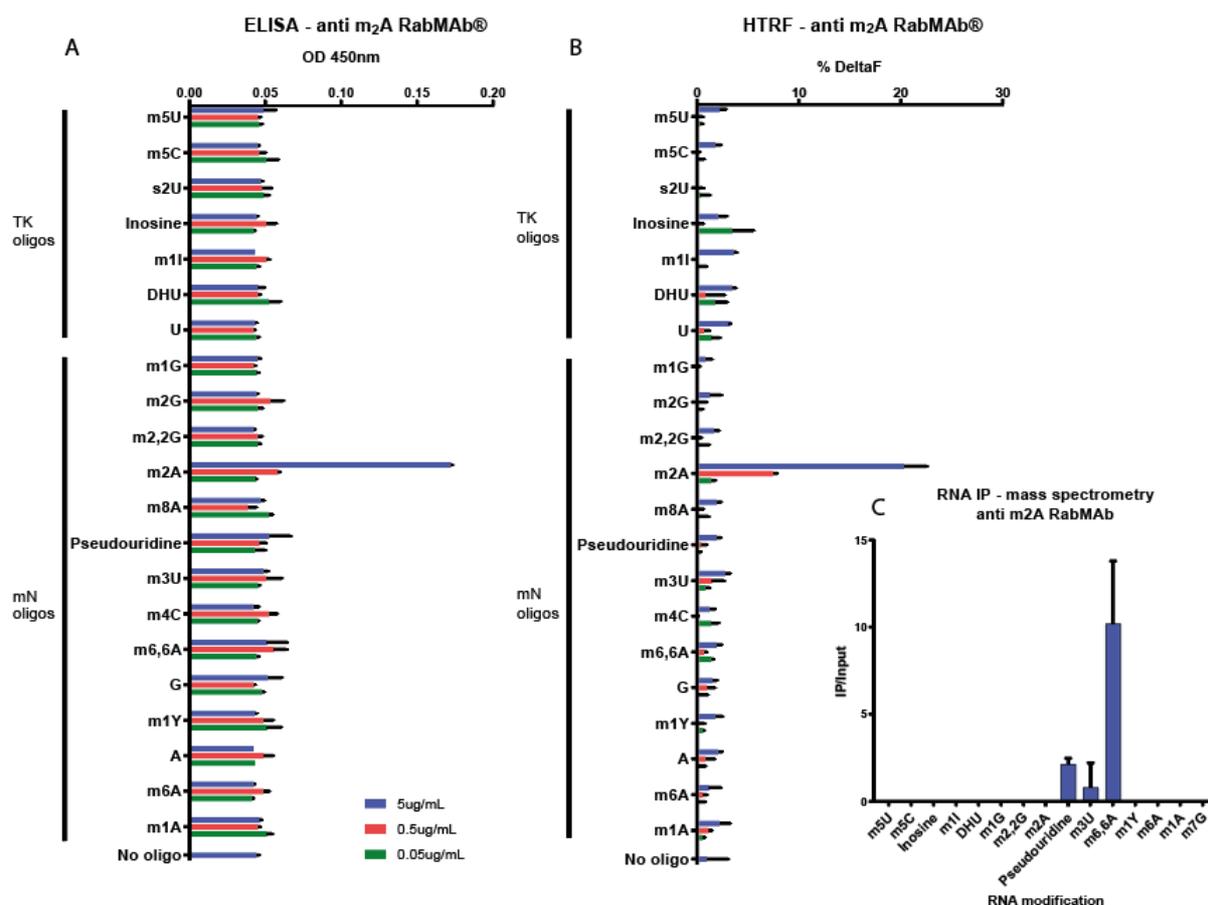
A recent paper showed that METTL15 catalyses m4C on mitochondrial rRNA and is crucial for mitoribosome biogenesis<sup>144</sup>. The anti-m4C antibody was tested for specificity against the panel of RNA oligonucleotides in both ELISA and HTRF<sup>®</sup> assays. This antibody showed good specificity toward the target modification in ELISA, with a small amount of cross-reactivity to m2,2G (figure 3.5A). By HTRF, significant cross-reactivity to m5C was observed, and a small amount of cross-reactivity to inosine, m1I, m2G and m2,2G (figure 3.5B). However, the signal in both the ELISA and HTRF<sup>®</sup> assay was lower than several of the other antibodies, indicating the sensitivity may be low. At the time of testing, there was no m4C standard available for RNA MS, so no meRIP-MS experiments were performed for this antibody.



**Figure 3.5 - Anti-m<sup>4</sup>C RabMAb<sup>®</sup> characterisation.** A panel of biotinylated RNA oligonucleotides were used in both ELISA and HTRF<sup>®</sup> assays. TK and mN describe two distinct oligonucleotide designs which are described in section 2.7.1. **A)** RNA oligonucleotides were captured onto streptavidin coated plates in triplicate, at three concentrations. These were probed with an anti-m<sup>4</sup>C primary antibody, followed by detection with HRP conjugated secondary antibody. Binding was quantified by colorimetric reaction and absorbance at 450nm on a BioTek plate reader. **B)** RNA oligonucleotides were used in an HTRF<sup>®</sup> assay to assess antibody specificity. The assay was designed with goat anti-rabbit Tb (Terbium cryptate) as the FRET donor and streptavidin d2 as the acceptor molecule. Terbium cryptate emission is measured at 620nm and streptavidin d2 emission is measured at 665nm. Binding is shown as %DeltaF (signal-background/background). The emission was measured on a PHERAstar plate reader (BMG).

### 3.3.6 Anti-m2A characterisation

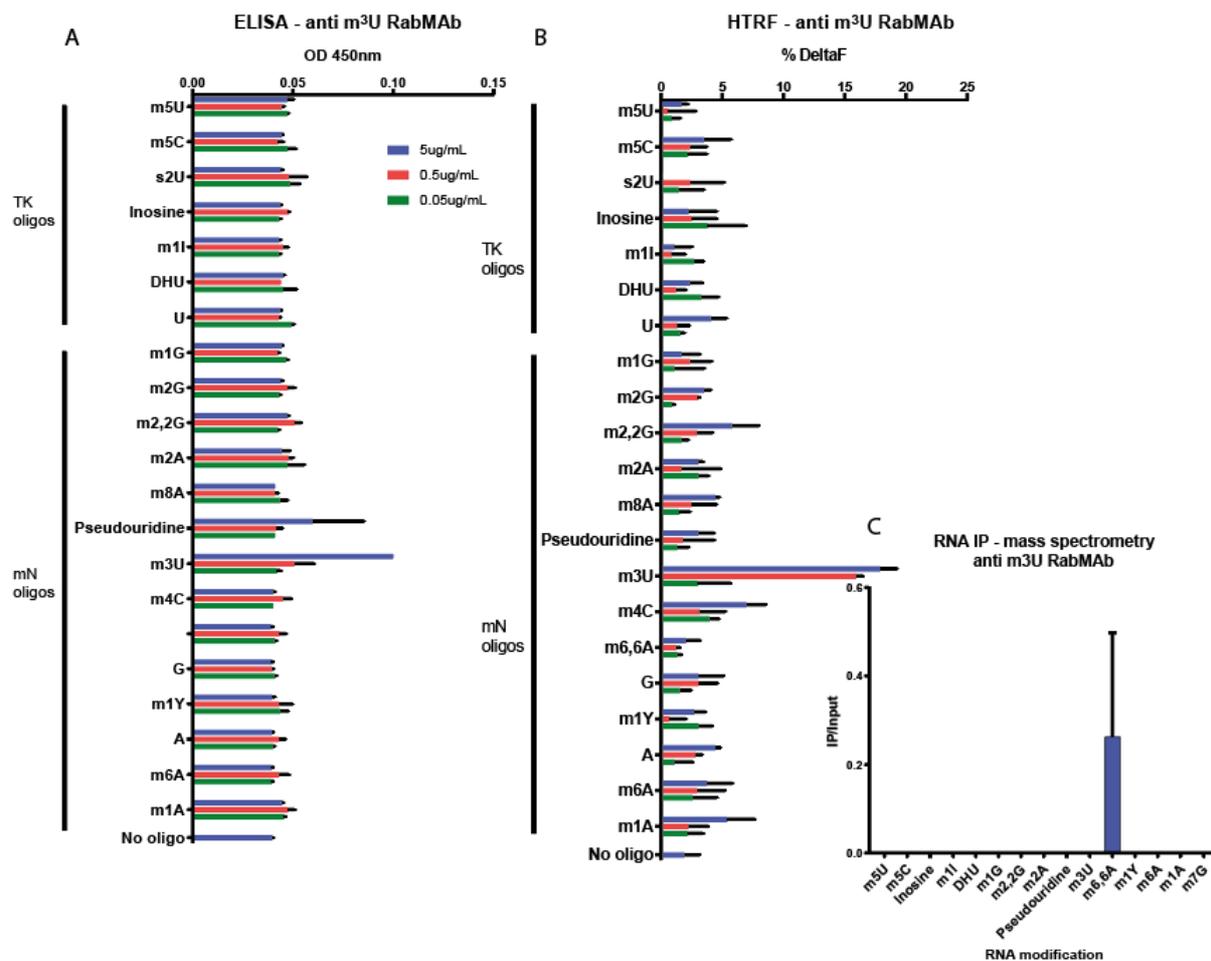
The anti-m2A antibody was tested for specificity against the panel of RNA oligonucleotides in both ELISA and HTRF<sup>®</sup> assays (figure 3.6A and figure 3.6B respectively). This antibody is specific for the target modification as observed in both assays. In the ELISA assay, the signal is only detected at the highest concentration of RNA oligonucleotide (5µg/ml). However, in the HTRF<sup>®</sup> assay the signal was present in both 5µg/ml and 0.5µg/ml oligonucleotide concentrations. I then used RNA immunoprecipitation coupled with mass spectrometry (meRIP-MS) to ascertain whether the antibody can enrich the target modification above the level in input RNA. There was no evidence of any enrichment of m2A above the level of input with this antibody (figure 3.6C). A small degree of enrichment of other non-target modifications (m6,6A, m3U and pseudouridine) was evident, but given this was not reflected in the specificity screening, it was likely background noise.



**Figure 3.6 - Anti-m<sub>2</sub>A RabMAb® characterisation.** A panel of biotinylated RNA oligonucleotides were used in both ELISA and HTRF® assays. TK and mN describe two distinct oligonucleotide designs which are described in section 2.7.1. **A)** RNA oligonucleotides were captured onto streptavidin coated plates in triplicate, at three concentrations. These were probed with an anti-m<sub>2</sub>A primary antibody, followed by detection with HRP conjugated secondary antibody. Binding was quantified by colorimetric reaction and absorbance at 450nm on a BioTek plate reader. **B)** RNA oligonucleotides were used in an HTRF® assay to assess antibody specificity. The assay was designed with goat anti-rabbit Tb (Terbium cryptate) as the FRET donor and streptavidin d2 as the acceptor molecule. Terbium cryptate emission is measured at 620nm and streptavidin d2 emission is measured at 665nm. Binding is shown as %DeltaF (signal-background/background). The emission was measured on a PHERAstar plate reader (BMG). **C)** methylated RNA immunoprecipitation was carried out on total RNA derived from HEK293-6e cells. After bound RNA was eluted from the antibody, it was hydrolysed to single nucleotides and analysed by RNA MS. Each sample was analysed in three technical replicates. Each modification is represented as normalised to the relevant non modified nucleoside, and then IP/Input.

### 3.3.7 Anti-m3U characterisation

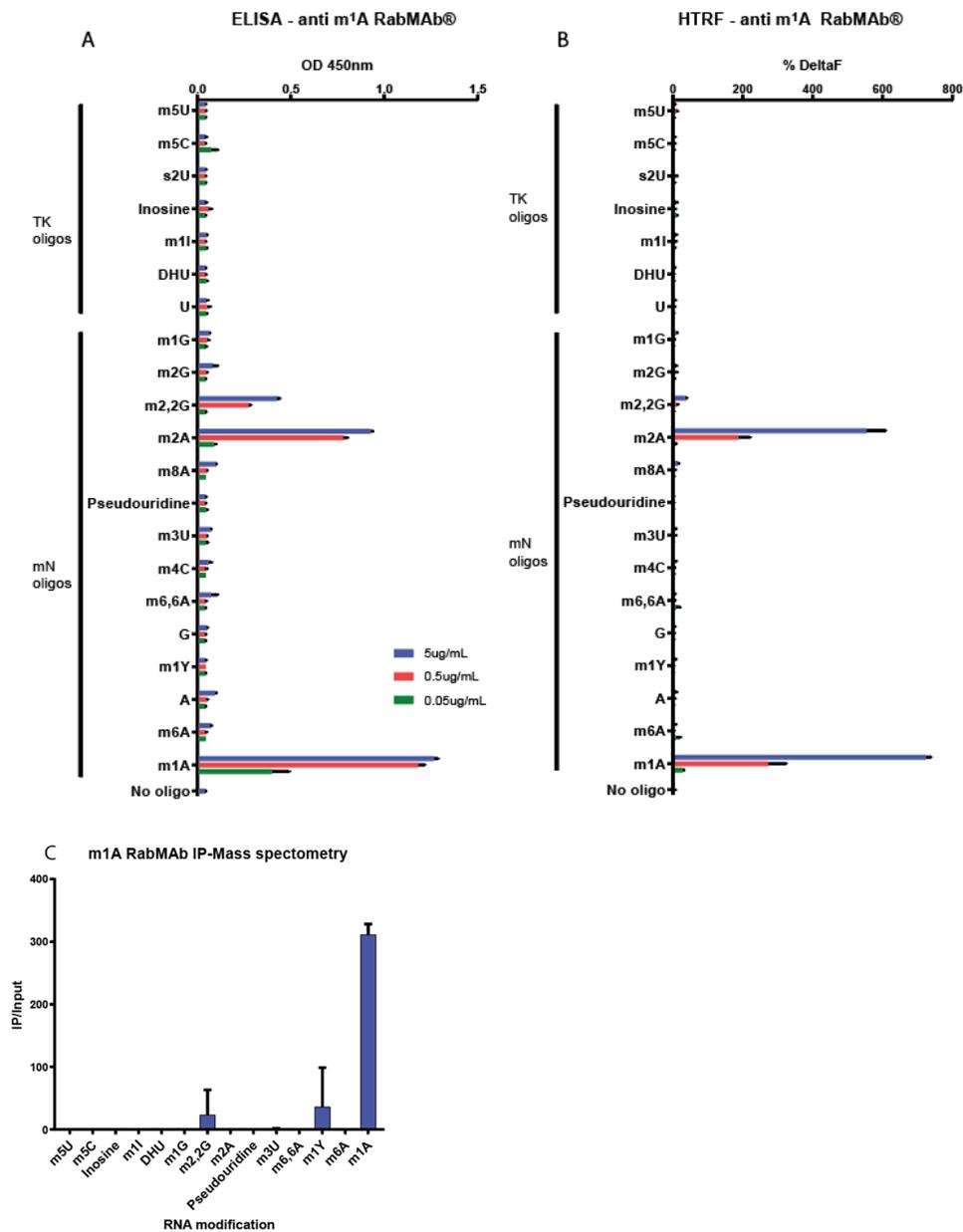
The anti-m3U antibody was tested for specificity against the panel of RNA oligonucleotides in both ELISA and HTRF<sup>®</sup> assays. By ELISA (figure 3.7A), this antibody shows good specificity to the target at the highest target concentration, but the signal was very low. In the HTRF<sup>®</sup> assay (figure 3.7B), the signal was higher and was observed at both 5µg/ml and 0.5µg/ml concentrations of oligonucleotide. This antibody was then tested for its ability to enrich the target modification above input levels by meRIP-MS. However, no enrichment of the target modification was observed with this assay. There was a small level of enrichment of a non-target modification m6,6A.



**Figure 3.7 - Anti-m<sup>3</sup>U RabMAb<sup>®</sup> characterisation.** A panel of biotinylated RNA oligonucleotides were used in both ELISA and HTRF<sup>®</sup> assays. TK and mN describe two distinct oligonucleotide designs which are described in section 2.7.1. **A)** RNA oligonucleotides were captured onto streptavidin coated plates in triplicate, at three concentrations. These were probed with an anti-m<sup>3</sup>U primary antibody, followed by detection with HRP conjugated secondary antibody. Binding was quantified by colorimetric reaction and absorbance at 450nm on a BioTek plate reader. **B)** RNA oligonucleotides were used in an HTRF<sup>®</sup> assay to assess antibody specificity. The assay was designed with goat anti-rabbit Tb (Terbium cryptate) as the FRET donor and streptavidin d2 as the acceptor molecule. Terbium cryptate emission is measured at 620nm and streptavidin d2 emission is measured at 665nm. Binding is shown as %DeltaF (signal-background/background). The emission was measured on a PHERAstar plate reader (BMG). **C)** methylated RNA immunoprecipitation was carried out on total RNA derived from HEK293-6e cells. After bound RNA was eluted from the antibody, it was hydrolysed to single nucleotides and analysed by RNA MS. Each sample was analysed in three technical replicates. Each modification is represented as normalised to the relevant non modified nucleoside, and then IP/Input.

### 3.3.8 Anti-m1A characterisation

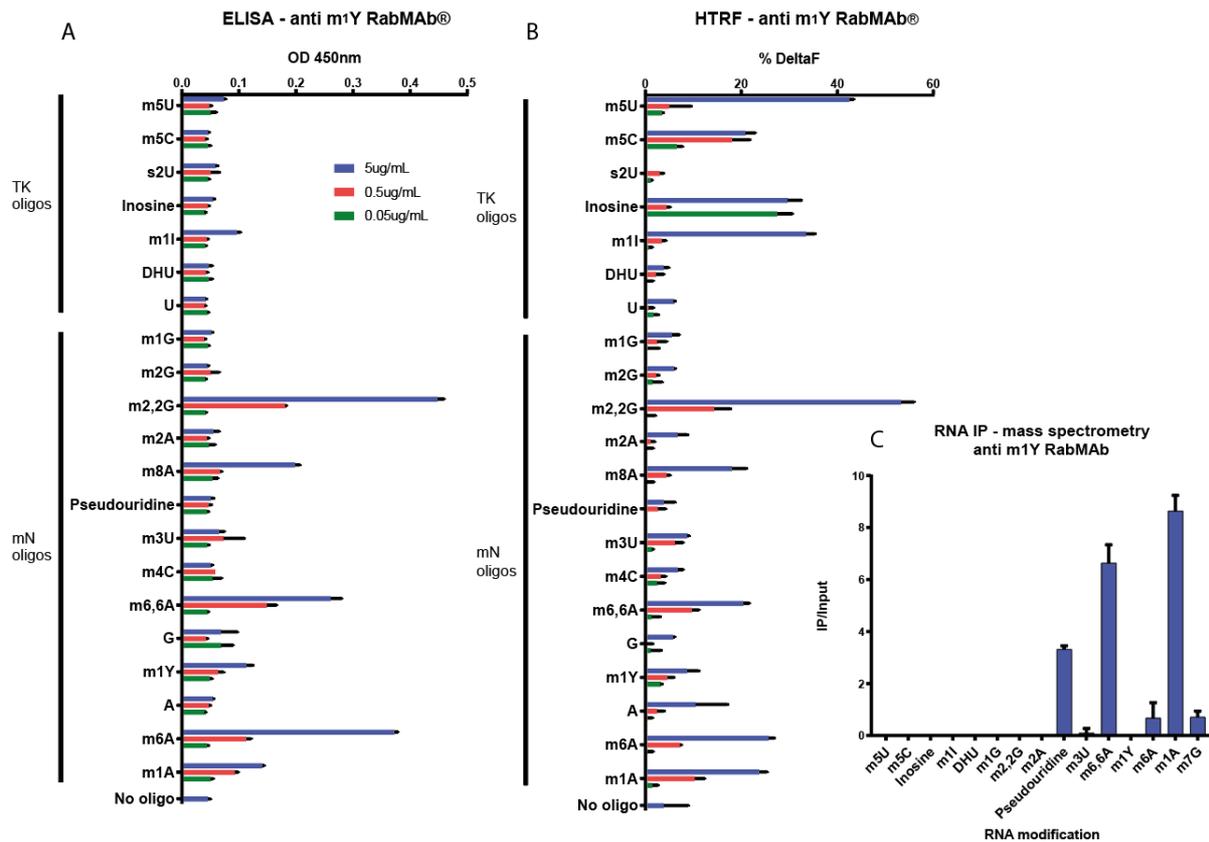
The anti-m1A antibody was tested for specificity against the panel of RNA oligonucleotides in both ELISA and HTRF<sup>®</sup> assays. Using the ELISA assay (figure 3.8A), m1A was detected at all three concentrations so the sensitivity of this antibody appears very high, but I observed significant cross-reactivity to m2A and m2,2G at the higher oligonucleotide concentrations. In the HTRF<sup>®</sup> assay (figure 3.8B) I also observed strong reactivity to the target modification, although it was not detectable at the lowest oligonucleotide concentration of 0.05µg/ml. Additionally, the cross-reactivity to m2A was present to an even greater degree than seen in the ELISA assay. In this case, the signal from the m2A oligonucleotide was almost at the same level as that of the target m1A modification oligonucleotide. However, the m2,2G cross-reactivity was not observable in the HTRF<sup>®</sup> data. The anti-m1A antibody was then tested in meRIP-MS to see if it was suitable for future meRIP-seq experiments. Despite the apparent cross-reactivity in the binding assays, the anti-m1A antibody enriched the target modification over three hundred times above the level in the input RNA (figure 3.8C). Notably, m2A was not enriched by immunoprecipitation.



**Figure 3.8 - Anti-m1A RabMAB® characterisation.** A panel of biotinylated RNA oligonucleotides were used in both ELISA and HTRF® assays. TK and mN describe two distinct oligonucleotide designs which are described in section 2.7.1. **A)** RNA oligonucleotides were captured onto streptavidin coated plates in triplicate, at three concentrations. These were probed with an anti-m1A primary antibody, followed by detection with HRP conjugated secondary antibody. Binding was quantified by colorimetric reaction and absorbance at 450nm on a BioTek plate reader. **B)** RNA oligonucleotides were used in an HTRF® assay to assess antibody specificity. The assay was designed with goat anti-rabbit Tb (Terbium cryptate) as the FRET donor and streptavidin d2 as the acceptor molecule. Terbium cryptate emission is measured at 620nm and streptavidin d2 emission is measured at 665nm. Binding is shown as %DeltaF (signal-background/background). The emission was measured on a PHERAstar plate reader (BMG). **C)** methylated RNA immunoprecipitation was carried out on total RNA derived from HEK293-6e cells. After bound RNA was eluted from the antibody, it was hydrolysed to single nucleotides and analysed by RNA MS. Each sample was analysed in three technical replicates. Each modification is represented as normalised to the relevant non modified nucleoside, and then IP/Input.

### 3.3.9 Anti-m1Y characterisation

The anti-methylpseudouridine (m1Y) antibody was tested against the panel of modified RNA oligonucleotides in ELISA and HTRF<sup>®</sup> assay (figure 3.9A and figure 3.9B respectively). The anti-m1Y antibody was found to be highly nonspecific and reacted with many modifications to a greater degree than the target. The majority of the cross-reactivity occurs concurrently in both ELISA and HTRF<sup>®</sup> assays. However, there are some modifications which produce signal in HTRF<sup>®</sup>, but not in ELISA (m5U, m5C, inosine, and m1I). The anti-m1Y antibody was then used in meRIP-MS to assess its suitability for meRIP experiments (figure 3.9C). However, it was not capable of enriching the modification of interest in immunoprecipitation assays.



**Figure 3.9 - Anti-m1Y RabMAb® characterisation.** A panel of biotinylated RNA oligonucleotides were used in both ELISA and HTRF® assays. TK and mN describe two distinct oligonucleotide designs which are described in section 2.7.1. **A)** RNA oligonucleotides were captured onto streptavidin coated plates in triplicate, at three concentrations. These were probed with an anti-m1Y primary antibody, followed by detection with HRP conjugated secondary antibody. Binding was quantified by colorimetric reaction and absorbance at 450nm on a BioTek plate reader. **B)** RNA oligonucleotides were used in an HTRF® assay to assess antibody specificity. The assay was designed with goat anti-rabbit Tb (Terbium cryptate) as the FRET donor and streptavidin d2 as the acceptor molecule. Terbium cryptate emission is measured at 620nm and streptavidin d2 emission is measured at 665nm. Binding is shown as %DeltaF (signal-background/background). The emission was measured on a PHERAstar plate reader (BMG). **C)** methylated RNA immunoprecipitation was carried out on total RNA derived from HEK293-6e cells. After bound RNA was eluted from the antibody, it was hydrolysed to single nucleotides and analysed by RNA MS. Each sample was analysed in three technical replicates. Each modification is represented as normalised to the relevant non modified nucleoside, and then IP/Input.

### 3.4 Discussion

When designing a characterisation strategy for these antibodies there were two primary questions that were considered: 1) is the antibody specific for the target modification and 2) can the antibody enrich the modification from total RNA? Specificity is crucial to determine which modification is present under peaks of sequencing type experiments, and the ability to immunoprecipitate the modification is prerequisite to carry out transcriptome-wide mapping assays.

Many of the antibodies tested showed high specificity, but the signal was weak (anti-m2A, anti-m3U and anti-m4C). These specific but weak antibodies did not work in the meRIP-MS assays where tested, suggesting a signal above a certain threshold in the ELISA/HTRF® assays may be necessary for an antibody to be useful in pull-down techniques. Conversely, the anti-m2,2G antibody demonstrated high specificity and signal strength against the oligonucleotides but was unsuccessful in immunoprecipitation. This cannot be explained by low affinity to the modification because of the strong signal seen in other assays. We also know that m2,2G exists on several tRNAs, so the IP failure cannot be due to lack of availability of the target modification.

It is possible that the wash conditions used in the immunoprecipitation were too stringent for this antibody. The wash buffers in the IP assay contain higher concentrations of sodium chloride and tween than in the specificity assays. Furthermore, the oligonucleotides used are only 10-20 nucleotides in length, but the native RNA used in the immunoprecipitation assays were fragmented to approximately 100 nucleotides. It is plausible that this could play a role in the differences seen between the assays. Furthermore, the m2,2G in native fragments of RNA could adopt local secondary structures that render the modification inaccessible to the antibody. Although heat treatment at 70°C was carried out to remove secondary structure, there could be some structures produced by m2,2G that are not removed or reformed after heating.

The anti-m1G, anti-m1A and anti-m6,6A antibodies were the only ones to enrich the modification above the level of input and demonstrate acceptable specificity by ELISA/HTRF®. Regarding the enrichment of non-target modifications with the anti-m1G antibody, this could be something that could be removed by fragment size optimisation. Assuming that these enrichments (m7G, m5C and pseudouridine) are not due to antibody promiscuity, but rather by proximity of the modifications to the target m1G; a smaller fragment size used in this assay may eradicate the apparent cross-reactivity by removing the co-precipitation effect. Using the competition ELISA with free m1G and m7G

nucleosides increases confidence that the enrichment seen of m7G is not due to antibody cross-reactivity. Because free m1G nucleoside was seen to successfully block anti-m1G RabMAb<sup>®</sup> binding to the oligonucleotide, this assay demonstrated that the same approach could be used to elute m1G containing RNA from the antibody in subsequent meRIP-seq experiments.

At the time of writing this thesis, the anti-m1A and anti-m6,6A antibodies were commercially available through Abcam Plc (ab208196 and ab208198), but the anti-m1G antibody was not. Having access to a unique tool made m1G an attractive target modification to pursue.

Although there is often significant overlap between the two specificity assays (ELISA and HTRF<sup>®</sup>), I observed examples where cross-reactivity appeared in one assay but not the other. This usually occurs with the HTRF<sup>®</sup> assay uncovering cross-reactivity that the ELISA did not show. It is likely that the HTRF<sup>®</sup> assay is slightly more sensitive than the ELISA. Furthermore, the ELISA protocol uses multiple wash steps that could be reducing low affinity cross-reactivity. Because HTRF<sup>®</sup> is a homogenous assay, there are no wash steps and therefore weaker interactions may be detected. Furthermore, these characterisation data show how challenging it can be to raise good quality RNA modification antibodies. RNA modifications as antibody epitopes are extremely small compared to a standard peptide antigen, which makes raising a high affinity antibody rarer. Furthermore, the difference between a modified nucleotide and a non-modified one is clearly very small, meaning an antibody has to be exceptionally specific, especially given the overwhelming proportion of nucleotides in a cell are unmodified. It is possible that cross-reactivity exists outside the scope of this testing, meaning we must be cautious when relying too heavily on antibodies in the field. There are very few commercially available RNA modification antibodies, likely due to the difficulty in raising affinity reagents to such small and challenging targets. The standard level of characterisation may not be sufficient to avoid artefacts in mapping studies. For example, the anti-m1A antibody has a significant level of cross-reactivity to m2A, as seen by both the ELISA and HTRF<sup>®</sup> assays. This antibody was recently used to highlight the apparent overestimation of m1A sites in previous datasets<sup>36</sup>. But even using this more stringent characterisation, the m2A cross-reactivity I observed was not tested for.

The common validation in much published research is to simply screen against the modified nucleoside and the non-modified version, which in this case could miss crucial information. The field in general may need to move towards more wide-ranging testing of cross-reactivity. However, cross-reactivity does not necessarily render an antibody completely useless (for example the anti-m1A RabMAb<sup>®</sup>). It may be that the cross-reactive modification is not present in a particular system or RNA

type. Alternatively, a cross reactive antibody could be used to identify candidate modification sites. Used in conjunction with knockdowns/outs and orthogonal methods, useful information may be obtained.

In summary, these results show several of these antibodies are suitable to take forward for use in transcriptome-wide mapping studies (m1G, m6,6A and m1A). The characterisation undertaken set a very high bar for the antibodies, and expectedly, the majority were not up to the standard to use further. The anti-m1G antibody particularly, is specific and significantly enriches the modification of interest.

## Chapter 4 - Characterization of N1-methylguanosine (m1G) in pancreatic cancer

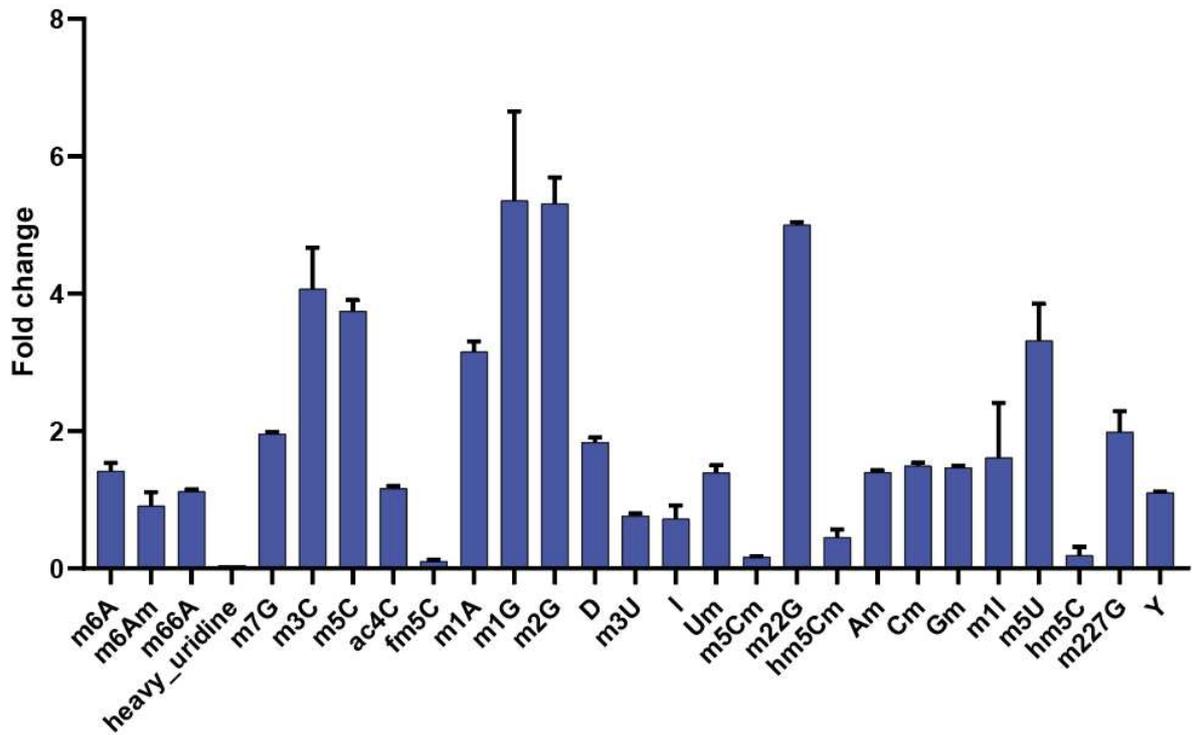
### 4.1 Why study the m1G modification pathway in pancreatic cancer?

Pancreatic cancer is one of the deadliest human cancers, with very poor survival rates. According to Cancer Research UK, only 3% of patients survive for 5 years or more. One of the reasons it is so fatal is because by the time it has been diagnosed, the cancer has invariably metastasised to other sites in the body. The ability of pancreatic cancers to metastasise is linked to poor prognosis, so if this process could be interrupted, disease outcomes could be improved.

Furthermore, the Kouzarides' laboratory has an established collaboration with Dr Sonia Melo (Ipatimup institute), an expert in the field of pancreatic cancer. This collaboration makes available numerous mouse models in which to investigate how RNA modifications may be linked to a transformed phenotype. It was decided to leverage this collaboration to the advantage of this project. For these reasons, I decided to explore RNA modification pathways in pancreatic cancer.

In collaboration with Storm Therapeutics, RNA MS was performed to ascertain the modification status of various pancreatic cell lines compared to non-transformed cell lines. This identified N1-methylguanosine (m1G) as one of the most elevated RNA modifications in the pancreatic cancer cell line compared to normal (figure 4.1). This was also a modification to which we had successfully raised an antibody against (Chapter 3). Therefore, I decided to investigate the relevance of the m1G modification pathway in pancreatic cancer.

### hTERT-HPNE vs PANC-1 RNA modifications fold changes



**Figure 4.1 - Changes in RNA modification levels between transformed (PANC-1) and non-transformed (hTERT-HPNE) pancreatic cell lines.** Cellular total RNA was purified and analysed by RNA MS in triplicate. Each modification was normalised to the average of the four canonical nucleosides. The fold change between the non-transformed (hTERT-HPNE) and transformed and cell line (PANC-1) for each modification was then calculated.

## 4.2 Introduction to the m1G enzymes

In *Saccharomyces cerevisiae*, m1G is deposited on a subset of tRNAs at position 37 by Trm5, and at position 9 by Trm10<sup>77,145</sup>. In humans, there are four putative m1G methyltransferases predicted from homology to yeast Trm5 and Trm10. These are TRMT5, TRMT10A, TRMT10B and TRMT10C. Very recently there has been increased interest in these enzymes, but their role in cancer has not yet been experimentally defined.

### 4.2.1 TRMT10A, B and C

TRMT10A is predicted to be a cytoplasmic variant and has already been linked to diabetes and intellectual disabilities<sup>146,147</sup>. A case study of a family with diabetes mellitus found that the patients had a G206R mutation in the *TRMT10A* gene, which did not affect tRNA binding, but methyltransferase activity was <0.1% that of the wild type TRMT10A<sup>147</sup>. The authors concluded that the mutation in the *TRMT10A* gene abolished its ability to bind the methyl donor SAM, which affected glucose homeostasis through apoptosis of  $\beta$  cells and eventually led to diabetes in the patients. A more recent study shed further light on the mechanism behind *TRMT10A* mutations and early onset diabetes<sup>148</sup>. This study confirmed TRMT10A as a tRNA m1G methyltransferase using *in vitro* assays with recombinant TRMT10A. Moreover, small RNAs were purified from patients with TRMT10A mutations and analysed for m1G content. Both tRNA<sup>Glu</sup><sub>(CUG)</sub> and tRNA<sup>iniMeth</sup><sub>(CAU)</sub> were found to be targets of TRMT10A at position 9. The authors then used an iPSC derived  $\beta$ -like cell model to determine the effect of TRMT10A disruption in the pancreas. TRMT10A silencing in this system increased oxidative stress and apoptosis. To assess whether these two observations were linked, an antioxidant compound was used to reduce the oxidative stress, and indeed the apoptosis was decreased as well. This study highlighted how a tRNA methyltransferase mutation and the subsequent hypomethylation can cause a disease state.

Vilardo *et al* recently demonstrated that TRMT10B is an m1A methyltransferase, despite the homology to yeast m1G methyltransferase Trm10<sup>149</sup>. In this work, HAP1 haploid knockout cell lines were produced for both TRM10A and TRMT10B. RT signatures were analysed to determine which modifications were affected by the loss of each enzyme. Knockout of TRMT10A caused a loss of signature in 17 of the 18 nuclear encoded tRNAs, but the TRMT10B knockout only affected the signature of position 9 on tRNA<sup>Asp</sup><sub>(GTC)</sub>. Interestingly, this tRNA has an adenosine at this position, which was recently identified as the only case of m1A9 in human nuclear encoded tRNAs<sup>150</sup>. As expected, this study demonstrated that loss of these enzymes had no effect on mitochondrial tRNA methylation.

Moreover, it was also shown that neither TRMT10A nor TRMT10B loss had any significant effect on growth rates of the HAP1 cell lines. This study concluded that TRMT10A is strictly guanosine specific and TRMT10B seems to only m1A methylate tRNA<sup>Asp</sup><sub>(GTC)</sub>.

TRMT10C has recently also been demonstrated to catalyse m1A on tRNA and so has not been studied in this work<sup>151</sup>.

#### 4.2.2 TRMT5

Methylation of the N1 atom of guanosine (m1G) at position 37 of tRNA is one of the most ancient RNA modifications known (figure 4.2A)<sup>74</sup>. It is present in all organisms, as well as mitochondria and chloroplasts. This modification is deposited by a tRNA methyltransferase encoded by *trmD* in bacteria, *Trm5* in *Saccharomyces cerevisiae* and TRMT5 in mammals<sup>152</sup>.

The human tRNA methyltransferase, TRMT5 has recently been identified as a cancer vulnerability in a large CRISPR screen<sup>126</sup>. In this study, TRMT5 deletion produced loss of fitness in 50% of the 165 cell lines tested and loss of fitness in a subsection of cell lines in all 13 cancer types tested. Interestingly, from the same screen TRMT10A was only found to be essential in 3% of cell lines and TRMT10B was not found to be essential in any cell line used. This screen suggests of the three putative m1G methyltransferases investigated in this chapter, TRMT5 is likely to be the most important in cancer pathologies.

Defective expression of mitochondrial encoded genes has been linked to TRMT5 in several independent studies. Several patients exhibiting lactic acidosis and multiple mitochondrial respiratory-chain-complex deficiencies in skeletal muscle, were found to have *TRMT5* mutations which resulted in hypomethylation at position 37 of mitochondrial tRNAs<sup>78</sup>. Additionally, using a FLAG-tagged construct, the authors showed TRMT5 localizes within mitochondria in HeLa cells. The link between TRMT5 and mitochondrial disorders is not limited to loss of methylation, gain of TRMT5 methylation has also been found to cause disease too. A mutation in a tRNA sequence, that introduces a novel TRMT5 modification site, has been shown to have clinical implications in inherited hypertension<sup>153</sup>. The aberrant 4435A-G mutation in mitochondrial tRNA<sup>Met</sup> causes changes in structure and function of this tRNA. Significant decreases in aminoacylation and steady state tRNA<sup>Met</sup> levels resulted in a decrease in expression of mitochondrial DNA-encoded proteins. This in turn caused

respiratory deficiency, markedly diminished mitochondrial ATP levels and membrane potential, and increased the production of reactive oxygen species in mutant cell lines.

In other eukaryotic systems, TRMT5 has been reported as a nuclear protein. *Arabidopsis thaliana* AtTRM5 has been identified as a Trm5 orthologue. Using a GFP tagged construct, AtTRM5 was found to localize to the nucleus and not the mitochondria. Loss of this enzyme results in significantly reduced levels of m1G and m1I. Unlike loss of TRMT10A and TRMT10B, loss of AtTRM5 drastically reduces growth rate.

#### 4.2.3 Pancreatic ductal adenocarcinoma (PDAC)

Pancreatic ductal adenocarcinoma (PDAC) is one of the deadliest forms of cancer and at present highly untreatable. This is owing to late presentation of symptoms and resistance to therapeutic intervention. One of the defining features of pancreatic cancer is the large volume of desmoplastic stroma that can constitute up to 90% of the tumour volume<sup>154</sup>. Using transcriptomic approaches, two tumour associated stroma subtypes have been identified: a 'normal' subtype and an 'activated' subtype<sup>155</sup>. These subtypes greatly affect the prognosis of the patient, with the highly 'activated' subtype being more severe. The communication between the surrounding associated stroma and the PDAC itself strongly influences the growth and metastasis of the tumour. Interestingly, factors expressed in the activated version of the stroma have been shown to induce a basal-like subtype of PDAC, through upregulation of the oncogene *YAP1*<sup>156</sup>. Moreover, stromal signalling using HGF and IGF-1 have been shown to result in the phosphorylation of the prometastatic molecule Annexin A2<sup>157</sup>. This hallmark feature of crosstalk between PDAC and surrounding stroma offers a potential intervention point for therapeutics.

One of the most common mutations in PDAC is oncogenic mutation of *KRAS*, which is present in over 90% of cases. Additional signature mutations include the inactivation of tumour suppressors *TP53*, *SMAD4* and *CDKN2A*<sup>158</sup>. Despite being highly untreatable, there are some promising therapeutics being produced. For example, a small molecule inhibitor to *KRAS*<sup>G12C</sup> is showing potential in clinical trials, but this mutation is present in only ~1.5% of PDAC patients<sup>159</sup>. Extensive tumour heterogeneity presents further difficulty in producing effective treatments. It is therefore crucial to identify new common vulnerabilities that can be targeted.

## 4.3 Aims of this chapter

- i. Determine which enzyme(s) is/are responsible for m1G deposition.
- ii. Characterise proliferation and migration phenotypes of TRMT5 depleted cells.
- iii. Construct TRMT5 shRNA inducible knockdown cell line for mouse experiments.
- iv. Establish biological outcomes of TRMT5 depletion using an *in vivo* mouse model.
- v. Map m1G modification across transcriptome in PANC-1 cell line.
- vi. Demonstrate translational effect of TRMT5 depletion by Ribo-seq.

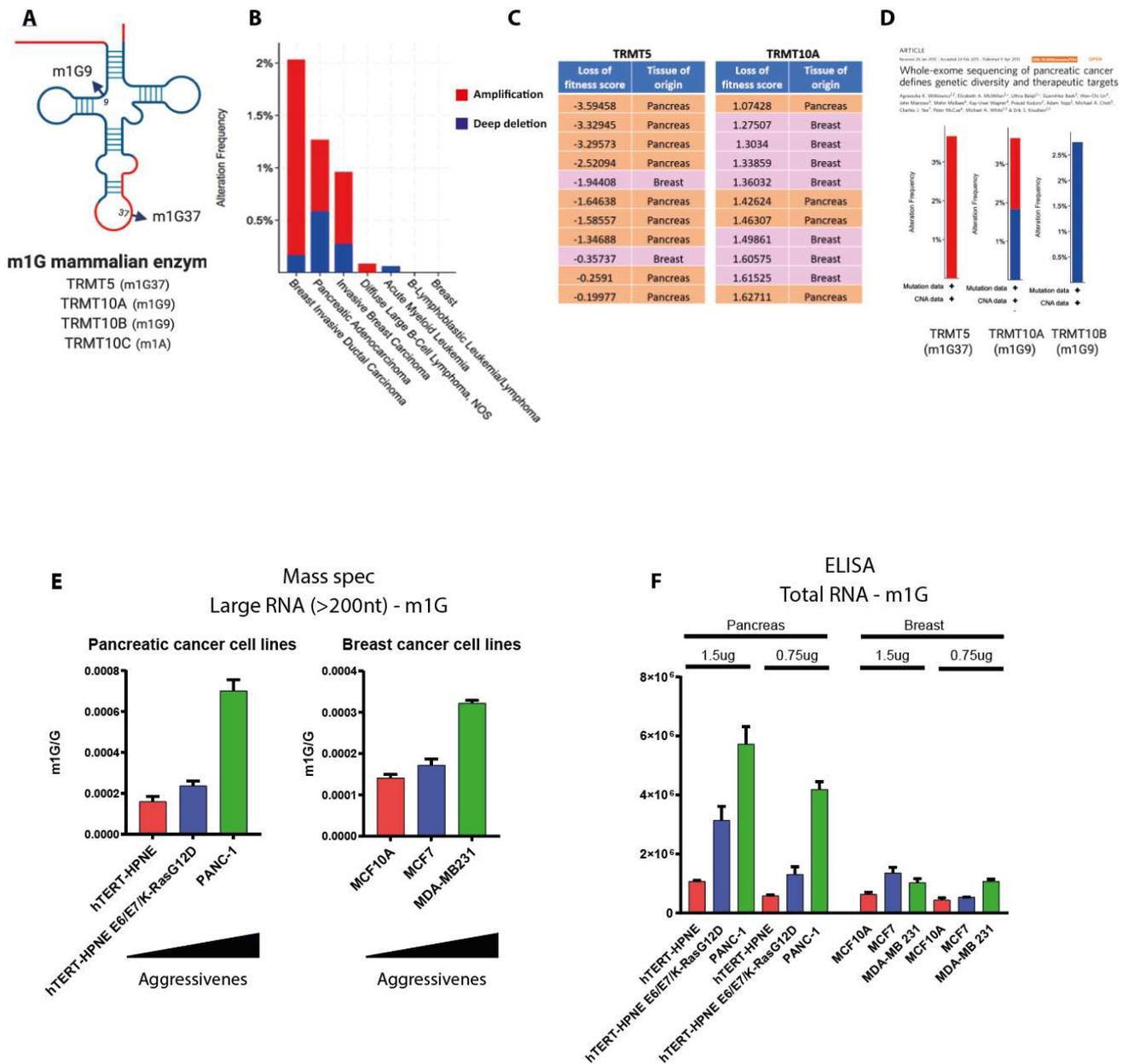
## 4.4 Results

### 4.4.1 m1G methyltransferases in cancer

For the reasons previously stated, the model system chosen for the analysis of the m1G modification pathway was human pancreatic cancer. To further validate this choice, I surveyed The Cancer Genome Atlas (TCGA) for dysregulation of TRMT5, TRMT10A and TRMT10B combined. Each of the enzymes is dysregulated in different cancer types, but when merged (i.e. any of the three methyltransferases is dysregulated) the highest levels were identified in breast invasive ductal carcinoma, invasive breast carcinoma and pancreatic adenocarcinoma (figure 4.2B). These dysregulations constitute both gene amplifications and deletions. Next, I analysed the loss of fitness scores from a recent CRISPR screen<sup>126</sup> for both TRMT5 and TRMT10A, focussing on pancreatic and breast cancer cell lines (figure 4.2C). A lower loss of fitness score indicates a greater effect on cell proliferation when the gene in question is disrupted. The scores are significantly lower for TRMT5 loss, than for TRMT10A. Furthermore, pancreatic cell lines generally have lower scores compared to breast cell lines. Finally, I interrogated TRMT5 and TRMT10A TCGA datasets individually (figure 4.2D). Here, it is clear that when TRMT5 is dysregulated, it is primarily due to its amplification. Conversely, when TRMT10A is dysregulated it is equally divided between amplification and deletion events<sup>160</sup>. Because TRMT5 tends to become amplified when it is dysregulated in cancer patients, this implies that it could be acting as an oncogenic driver. In the long term this makes TRMT5 a more attractive therapeutic target because it is easier to inhibit than it is to enhance enzyme activity. Historically there has been more success developing inhibitors to oncogenes than activators of tumour suppressors<sup>161</sup>. Taken together, these data reinforced my decision to focus on TRMT5 in pancreatic cancer cells.

To determine whether m1G levels vary between different cell lines from the same tissue or between cell lines from different tissues, large RNA (>200nt) was extracted from panels of pancreatic and breast cancer cell lines of varying degrees of aggressiveness. As a comparison, breast cancers were also investigated here based on TCGA data showing dysregulation of m1G enzymes in these cancers. The RNA was analysed for m1G content by both RNA MS (figure 4.2E) and using the newly characterised anti-m1G antibody in a direct ELISA (figure 4.2F). For the ELISA, total RNA was captured onto a 96-well plate using a commercially available coating reagent and probed with the anti-m1G antibody. In both pancreatic and breast cancer cell lines, there is a striking correlation between the m1G level in RNA and the aggressiveness of the cancer cell line (figure 4.2E and F). For example, hTERT-HPNE pancreatic cells, derived from early stage metaplastic cells, immortalised by human telomerase, have very low m1G levels. When transformed with KRAS (hTERT-HPNE E6/E7/K-RasG12D), these cells represent an intermediate between healthy cells and highly aggressive cancer cells and interestingly, have intermediate levels of m1G. Conversely, the highly aggressive PANC-1 cell line is from fully developed carcinoma and has at least 3-fold higher m1G levels than hTERT-HPNE.

Notably, the pancreatic cell lines exhibit higher levels of m1G compared to breast cell lines. This is true when comparing both the less aggressive hTERT-HPNE and MCF10A cell lines, and the more aggressive PANC-1 and MDA-MB-231 cell lines. By RNA MS analysis, m1G is on average twice as high in the pancreatic cell lines compared to breast cell lines (figure 4.2E). By m1G ELISA the difference appears even more pronounced (figure 4.2F). The observation that m1G levels are generally higher in the pancreatic cell lines than the breast cell lines tested, provides further evidence for the importance of m1G and TRMT5 in pancreatic cancer.

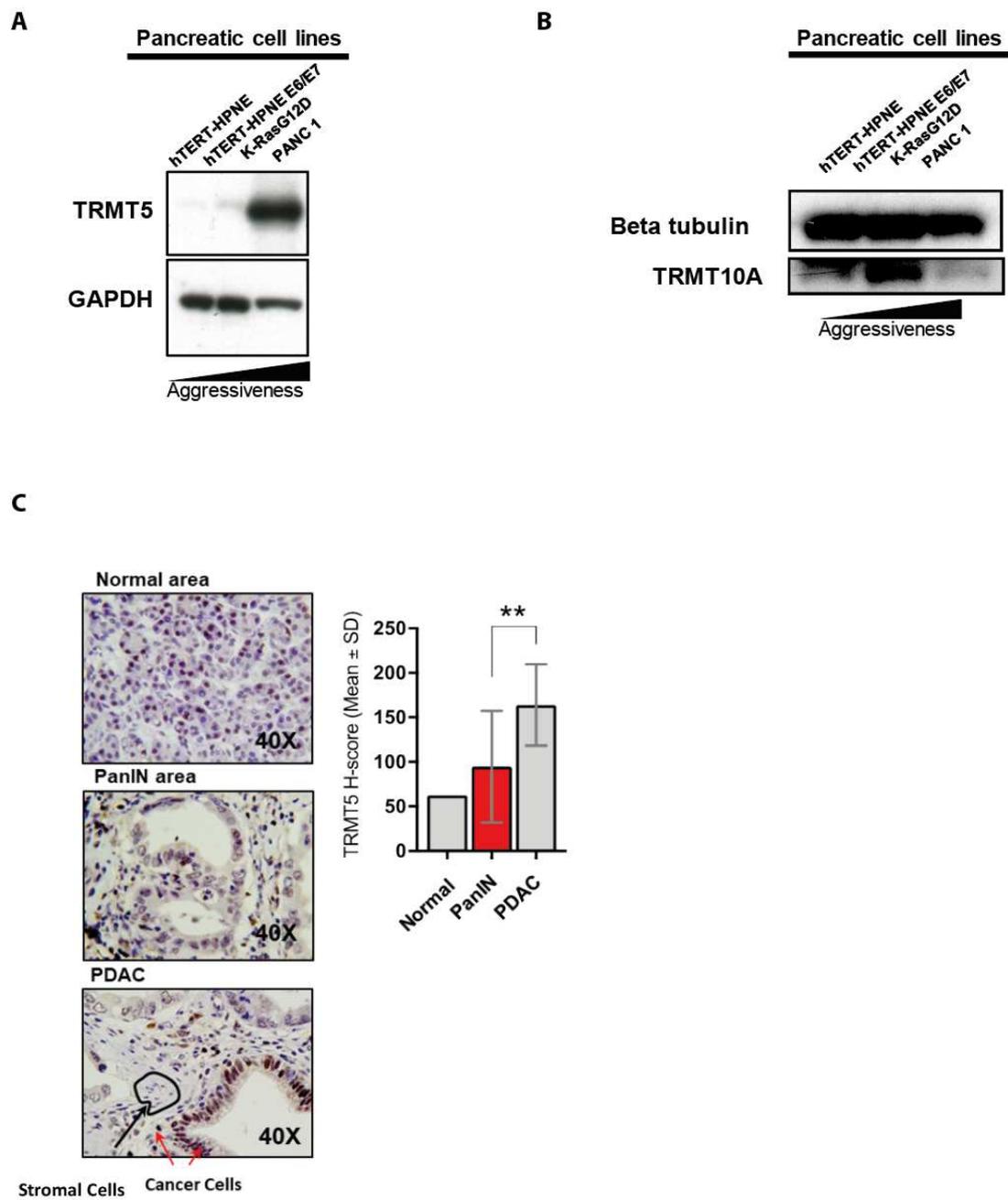


**Figure 4.2 - m1G levels and cancer cell line aggressiveness. A)** Schematic showing known human m1G modification sites on tRNA and the putative methyltransferases based on *S.cerevisiae* homology. **B)** TCGA data showing cancer types where either TRMT5, TRMT10A or TRMT10B are most highly deregulated. **C)** Loss of fitness scores for the top pancreatic and breast cancer cell lines from a CRISPR screen by Behan *et al*<sup>126</sup>. **D)** TRMT5, TRMT10A and TRMT10B alteration frequencies are shown from whole exome sequencing of pancreatic cancer patients<sup>160</sup>. **E)** RNA MS was carried out on large RNA (>200nt) from a panel of pancreatic and breast cancer cell lines. m1G/G content was calculated. **F)** Direct ELISA was carried out on RNA at two concentrations from a panel of pancreatic and breast cancer cell lines. The RNA was coated onto maxisorp 96-well plates and probed using the anti-m1G RabMab®.

#### 4.4.2 TRMT5 as a biomarker for pancreatic cancer

The level of m1G in pancreatic and breast cancer cell lines appear to directly correlate with the degree of aggressiveness. Therefore, I next tested whether the levels of the enzymes, TRMT5 and/or TRMT10A follow a similar pattern. Using a Western blotting approach, I determined the expression levels of TRMT5 and TRMT10A in the pancreatic cancer cell lines used previously (figure 4.3A and B). In a similar pattern to m1G levels, the expression levels of TRMT5 correlate with the aggressiveness of the cancer cell line (figure 4.3A). For example, the non-transformed, immortalised pancreatic cell line hTERT-HPNE has minimal expression of TRMT5 protein and concurrently low levels of m1G. hTERT-HPNE transformed with KRAS has increased TRMT5 expression and m1G levels. Finally, in the aggressive PDAC cell line PANC-1, the TRMT5 expression level is strikingly high, in line with the observed m1G level. On the other hand, the protein expression level of TRMT10A does not scale with aggressiveness in the same way as TRMT5 (figure 4.3B). In fact, the TRMT10A level in PANC-1 is lower than that of the two less aggressive cell lines.

To supplement these findings in cell lines, immunohistochemistry was carried out by the Sonia Melo laboratory on human PDAC samples (figure 4.3C). Here I observed highly elevated levels of TRMT5 in the pancreatic duct of human pancreatic cancer samples. Interestingly however, the surrounding cancer associated stromal cells show comparatively low levels of TRMT5 expression. This lends further evidence to the hypothesis that TRMT5 and m1G modification are linked to pancreatic cancer and are potentially required for pancreatic cancer cells to thrive. Together, these data led me to focus on TRMT5 as the most important m1G methyltransferase in pancreatic cancer.

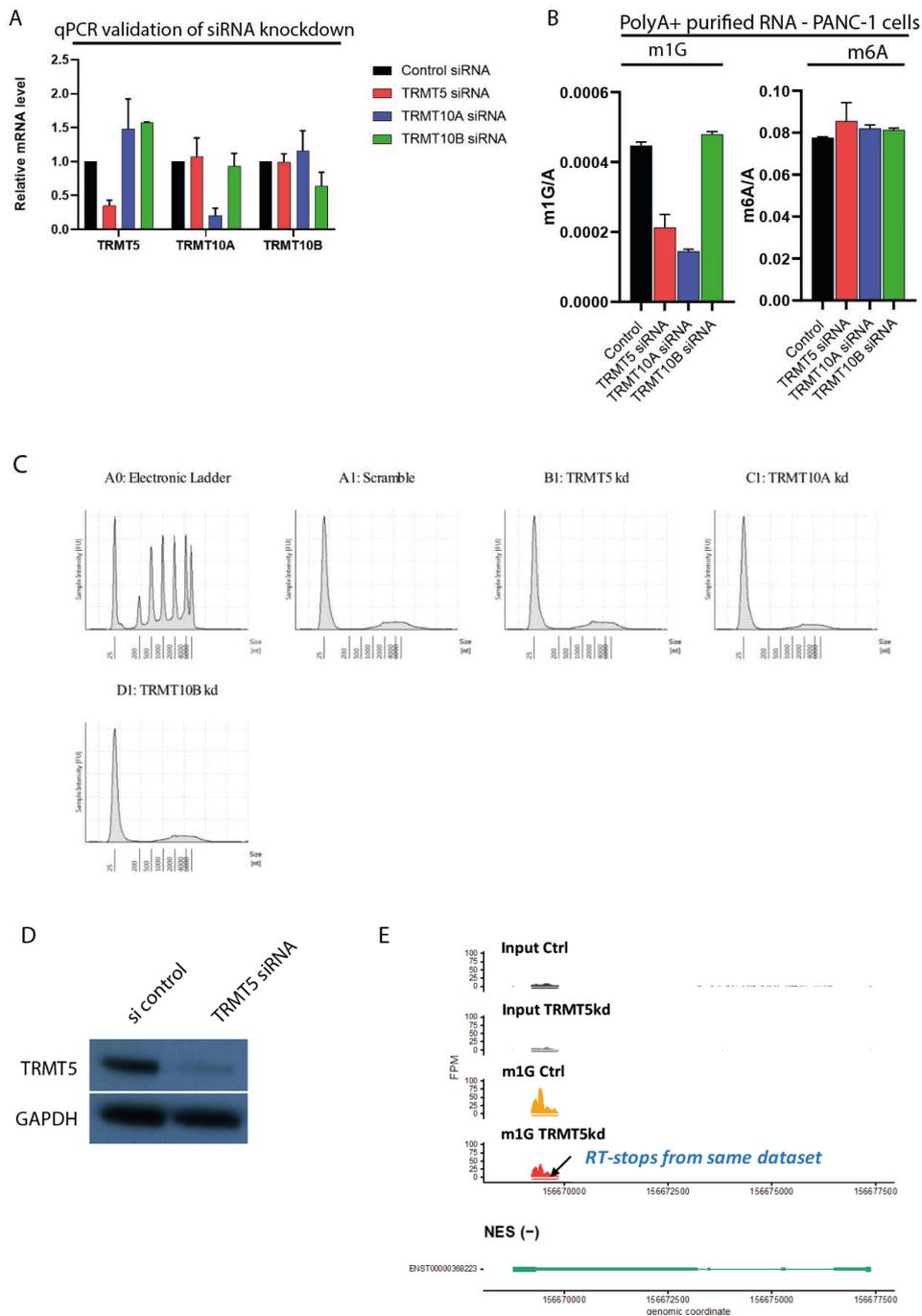


**Figure 4.3 - TRMT5 m1G methyltransferase as a pancreatic cancer target. A, B)** Western blotting of TRMT5 and TRMT10A was carried out on a panel of pancreatic cell lines of increasing aggressiveness. GAPDH and beta tubulin were probed as loading controls. **C)** Immunohistochemistry was carried out on human PDAC tissue sections using anti-TRMT5 antibody. Graph and statistical test were carried out using mean values of each PDAC patient (overall staining in zones of normal tissue, PanIN tissue and PDAC tissue). Statistical test used: Unpaired t-test, P-value = 0.0079. Errors bars show standard deviation of the mean.

#### 4.4.3 Is m1G present on mRNA?

With an increasing number of RNA modifications being identified on mRNA, I hypothesised that in addition to tRNA, m1G is also being installed on mRNA. Each of the putative m1G methyltransferases TRMT5, TRMT10A and TRMT10B were depleted in PANC-1 cells using siRNA and knockdown was confirmed by qPCR (figure 4.4A). Because m1G is known to exist on tRNA, care was taken to limit tRNA contamination. Large RNA (>200nt) was isolated from total RNA to subtract the tRNA population. From this fraction, highly purified polyA+ RNA was then enriched and analysed using the Agilent high sensitivity tapestation, where no tRNA or rRNA peaks were detectable (figure 4.4C). RNA modifications were then quantified in the resulting polyA+ RNA using RNA MS. m1G and m6A content were calculated by normalising to unmodified adenosine. Here, only depletion of TRMT5 and TRMT10A decreased m1G levels in this RNA fraction (figure 4.4B). Furthermore, m6A levels were not affected by knockdown of any of the enzymes. However, the percentage of m1G in the polyA+ RNA fraction is very low, suggesting that if m1G is indeed present on mRNA, it is only on a small number of RNAs or m1G modified targets are rarely modified (i.e. low stoichiometry).

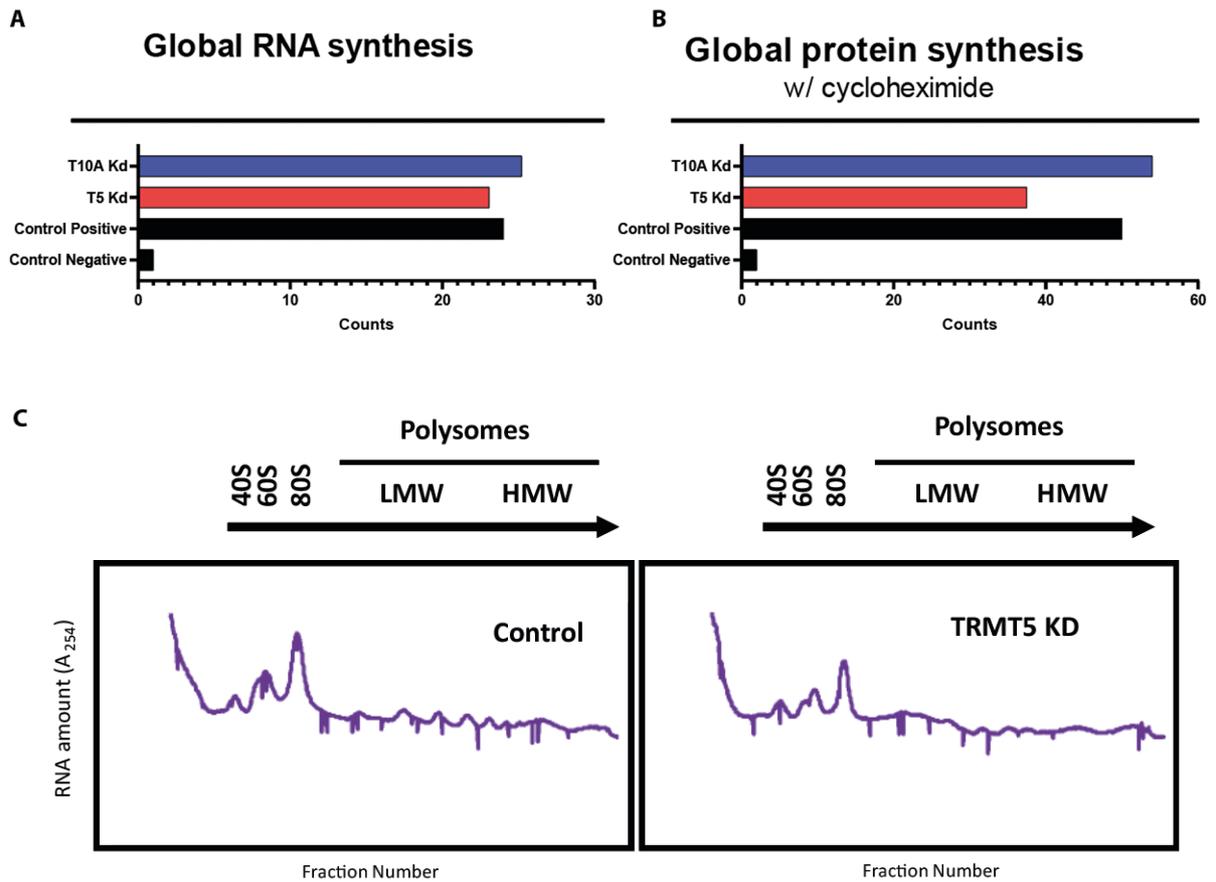
Using the new anti-m1G antibody characterised in chapter 3.3.2, m1G meRIP-seq was then carried out to map the modification across the polyA+ enriched RNA fraction of the transcriptome. Due to the potential artefacts arising from antibody-based mapping studies, the meRIP-seq was used to search for TRMT5 dependent m1G peaks. Firstly, TRMT5 was depleted from PANC-1 cells using a double 48hour siRNA knockdown. Knockdown efficiency was confirmed by Western blotting (figure 4.4D). Total RNA was then harvested after four days using QIAzol, followed by polyA+ RNA purification. The meRIP-seq identified 83 statistically significant TRMT5 dependent peaks on mRNA. These preliminary findings suggest that m1G modification may be present on mRNA. A peak was identified on the terminal exon of *Nestin* mRNA. Nestin has been previously linked to pancreatic cancer, which made it a relevant target to follow up<sup>162</sup>. Furthermore, the same meRIP-seq dataset was analysed using HAMR (high-throughput annotation for modified ribonucleotides) to locate RT arrests<sup>49</sup>. Using this method an RT stop was found to exist within the m1G-enriched nestin peak (figure 4.4E).



**Figure 4.4 - m1G as a novel mRNA modification. A)** siRNA knockdown efficiency was validated using qPCR. cDNA was produced from total RNA prior to polyA+ purification and used to measure mRNA levels of each of the target enzymes relative to *GAPDH* mRNA levels. Three technical replicates of the qPCR were carried out to produce error bars. **B)** Each of the m1G methyltransferases was knocked down in PANC-1 cells using siRNA. Total RNA was harvested 48 hours post transfection, from which polyA+ RNA was purified using NEXTFLEX® Poly(A) beads 2.0. This RNA was then analysed by RNA MS and m1G/A and m6A/A ratios were calculated. **C)** Purified polyA+ RNA was analysed for purity on HS RNA tapesturation to check for contaminating rRNA. **D)** Western blotting of TRMT5 and GAPDH was carried out to confirm the loss of TRMT5 after double siRNA knockdown for meRIP-seq. **E)** Genome browser snapshot from *nestin* mRNA (FPM = fragments per million mapped reads). RT stop detected using HAMR is shown.

#### 4.4.4 Effect of TRMT5 loss on transcription and translation

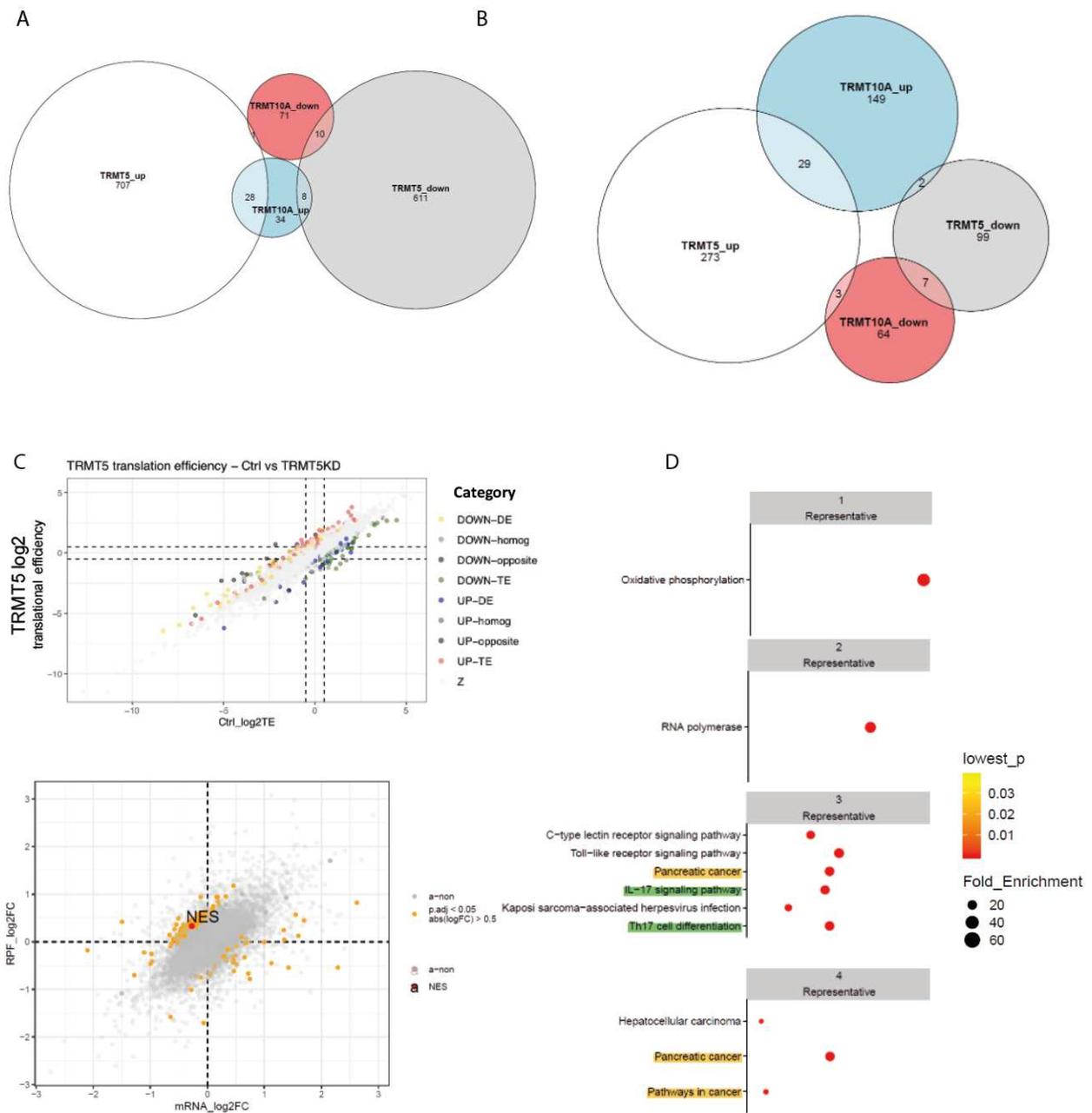
m<sup>1</sup>G is known to exist in two positions on a subset of tRNAs<sup>145,75</sup>. TRMT5 is responsible for deposition of m<sup>1</sup>G at position 37 of some tRNAs. This modification has been shown to be important in the maintenance of reading frame fidelity<sup>163,164,165</sup>. It is therefore expected that disrupting TRMT5 in PANC-1 cells will generate a substantial effect on translation. There may also be feedback mechanisms affecting transcription. If translation is greatly impeded by reduced reading frame fidelity, a negative feedback loop may downregulate expression at the transcriptional level. Here, I used two commercial kits (ab235634 and ab228561, Abcam) followed by flow cytometry to quantify the global transcription and translation rates in PANC-1 upon siRNA knockdown of TRMT5 and TRMT10A (figure 4.5A and 4.5B respectively). The protein synthesis kit uses OP-puro, which forms covalent bonds with nascent polypeptides. These truncated polypeptides are then detected by a click chemistry reaction with fluorescent azide. The RNA synthesis kit utilises 5-EU (ethynyl uridine), which is incorporated into nascent RNA but not DNA. The 5-EU modified RNA is then detected by a click chemistry reaction with fluorescent azide. Using these kits, I observed RNA transcription levels that are marginally lower in TRMT5 depleted cells compared to siRNA control, although this is not a large effect. There is a slight increase in transcription levels when TRMT10A is depleted, but similarly this is not a large change. Interestingly, global protein synthesis levels are significantly reduced when TRMT5 is lost, however a small increase in protein synthesis upon TRMT10A depletion was observed. Therefore, polysome profiling was carried out on PANC-1 cells after knock down of TRMT5 to further investigate the translational effect.



**Figure 4.5 - Loss of TRMT5 and TRMT10A effect on transcription and translation in PANC-1 cells. A, B) TRMT5 and TRMT10A were knocked down in PANC-1 cells by siRNA before using global protein synthesis and RNA synthesis kits to measure the rate of translation and transcription respectively. Signal was measured using flow cytometry and the number of counts is shown (n=1). C) PANC-1 cells were transfected with either control siRNA or TRMT5 siRNA and polysome profiles were compared.**

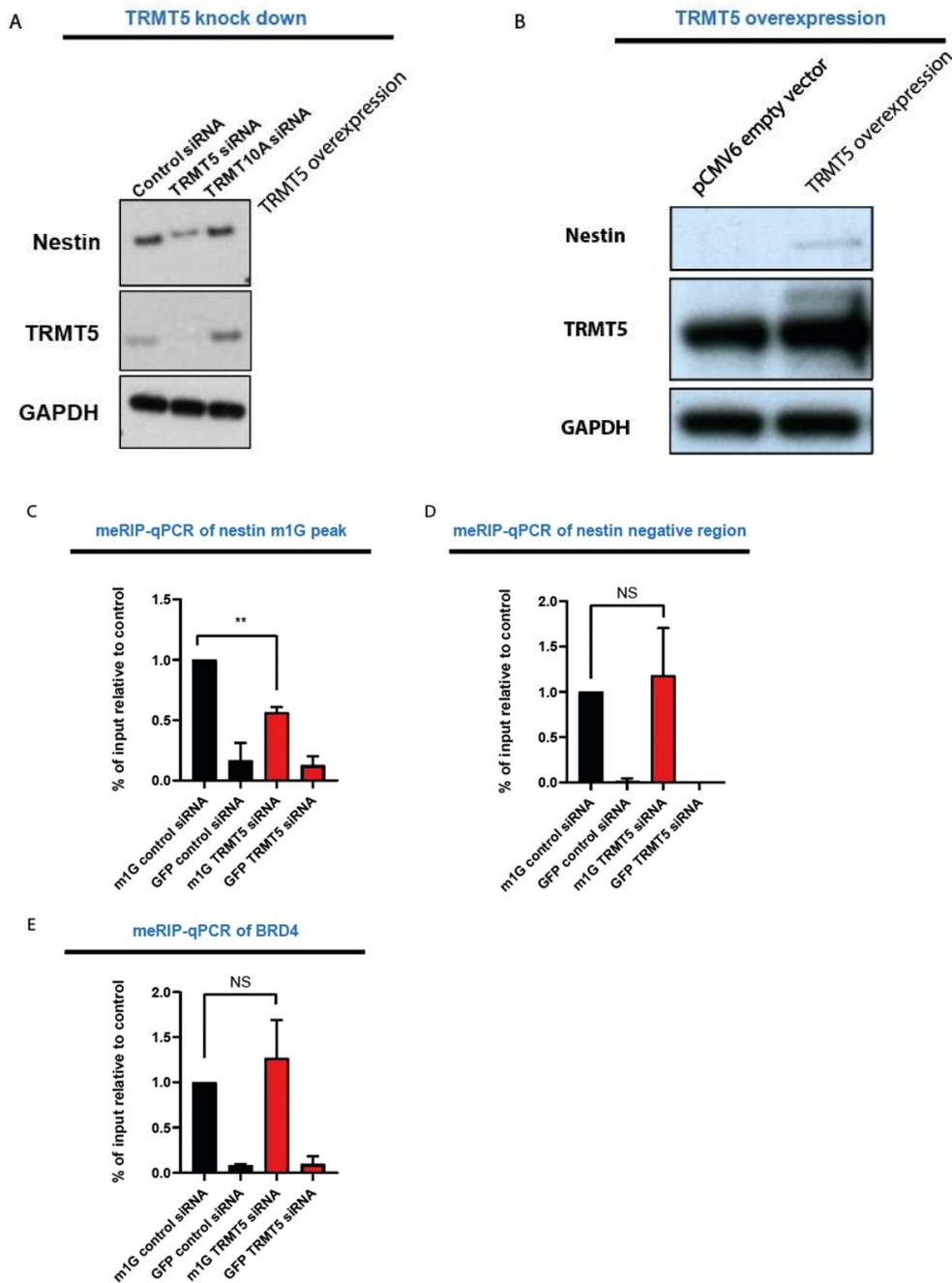
Figure 4.5 demonstrates that TRMT5 depletion from PANC-1 cells causes a marked effect on global protein translation but is unlikely to have a transcriptional effect. The polysome profiling shows a clear disruption in the low molecular weight polysome fraction, with less distinct peaks in the TRMT5 knockdown sample suggesting a disruption to translation (figure 4.5C). These effects may be due to tRNA modification changes, mRNA modification changes or both. Therefore, after establishing a global effect on translation when TRMT5 is depleted from PANC-1 cells, I wanted to understand if this effect was happening equally across the transcriptome or at a transcript specific level. Subsequently, using Ribo-seq, I investigated the translational effects of TRMT5 depletion at a more granular level.

Ribo-seq was carried out by collaborators at Loayza-Puch laboratory, Deutsches Krebsforschungszentrum. PANC-1 cells underwent double 48 hours siRNA knockdown before Ribo-seq and concurrent RNA-seq. Standard RNA-seq was carried out alongside the Ribo-seq so that RNA level changes could be accounted for during the analysis. Ribosome protected fragments (RPFs) were defined for the transcriptome to measure the relative translation efficiency. If a region of mRNA is highly protected by ribosomes, it is overrepresented in the sequencing data and is assigned a higher translational efficiency value. Interestingly, the knockdowns of TRMT5 and TRMT10A appear to have opposing effects with very little overlap (figure 4.6A). mRNA levels were cross compared with translational efficiency scores to cluster them into four groups (figure 4.6B). The X axis represents the change in mRNA level and the Y axis represents the change in translation efficiency. I was particularly interested in transcripts whose level was unaffected by TRMT5 knockdown, but whose translational efficiency was altered; these transcripts are highlighted in blue. Using the anti-m1G antibody I showed that m1G may be present on a small number of mRNA molecules and using Ribo-seq we found that TRMT5 knockdown was having specific effects on translation at distinct transcriptomic locations. I hypothesised that m1G on mRNA would exacerbate translational aberrations caused by loss of tRNA m1G modification. Therefore, the Ribo-seq data set was used in conjunction with the meRIP-seq data set to deconvolute mRNA m1G effects from tRNA effects. Here, only *nestin* mRNA gave both a TRMT5 dependent peak by meRIP-seq and reduced translational efficiency upon TRMT5 knock down as measured by Ribo-seq (figure 4.6C). RPF were found to increase 0.603-fold upon TRMT5 knock down. Furthermore, KEGG pathway analysis was carried out on the Ribo-seq dataset to identify pathways which were most affected by disruption of TRMT5 (figure 4.6D). Interestingly, pancreatic cancer was one of the significant pathways altered.



**Figure 4.6 - Ribo-seq and RNA-seq of TRMT5 depleted PANC-1 cells. A)** Transcripts were clustered into groups based on RNA-seq reads upon TRMT5 and TRMT10A knockdown. **B)** Transcripts were clustered into groups based on the translational efficiency being increased or decreased upon TRMT5 or TRMT10A knockdown. **C)** The translational efficiency for each mRNA was plotted against mRNA fold change. Transcripts were clustered into four distinct groups. Transcripts with a translational efficiency change (up/down), but no mRNA fold change are highlighted in blue. These targets were cross referenced with meRIP-seq data to identify *nestin* as a target. **D)** KEGG pathway analysis was applied to the Ribo-seq dataset.

Western blotting was carried out on TRMT5 depleted PANC-1 cells and interestingly the level of nestin is diminished (figure 4.7A). Conversely, when TRMT5 was overexpressed in PANC-1 cells, nestin levels are increased (figure 4.7B). Furthermore, to validate the finding that *nestin* mRNA is m1G methylated a meRIP-qPCR approach was carried out. qPCR primers were designed that amplified the region where the m1G peak was identified, in addition to a primer pair spanning a negative region of the *nestin* mRNA. meRIP was carried out in the same manner as for meRIP-seq, but the resulting RNA was converted to cDNA and used in qPCR reactions. Here I observed that m1G signal in the peak region is diminished upon TRMT5 knock down (figure 4.7C), but there is no difference in signal in the negative region (figure 4.7D). Moreover, *BRD4* was identified from the previous meRIP-seq as an mRNA where m1G methylation is not likely to be present. qPCR primers were designed for this mRNA as a negative control (figure 4.7E). Both the *nestin* negative region and *BRD4* show lower levels of enrichment than the *nestin* m1G region and neither are affected by loss of TRMT5. The negative region of *nestin* only showed ~9% the level of enrichment of the m1G region and the *BRD4* RNA only showed ~3% of the enrichment of the m1G *nestin* region (data not shown). This provides additional validation that there is in fact m1G methylation on *nestin* mRNA.

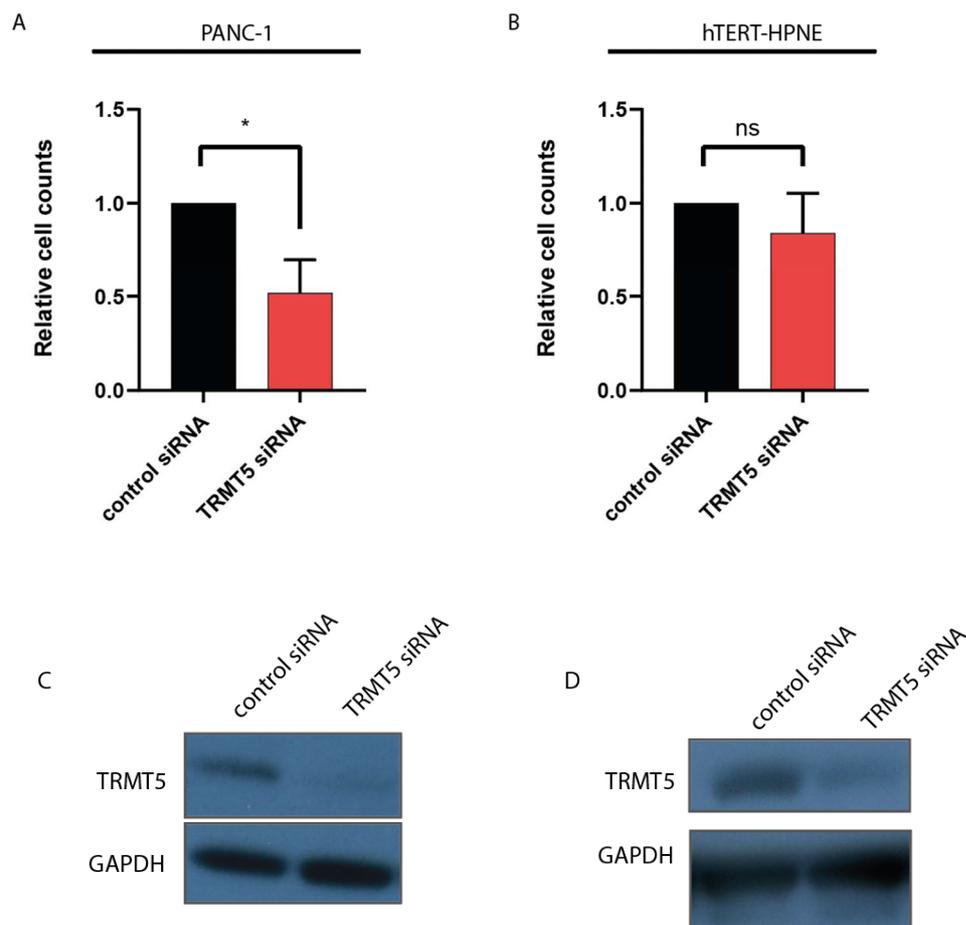


**Figure 4.7 - Nestin m1G validation. A)** PANC-1 cells were treated with either control siRNA, TRMT5 siRNA or TRMT10A siRNA. Western blotting was used to measure the levels of nestin, TRMT5 and GAPDH. **B)** TRMT5 containing pCMV6 plasmid was used to transfected PANC-1 cells, which were harvested after 48 hours. Nestin, TRMT5 and GAPDH levels were determined by Western blotting. **C-E)** meRIP-qPCR was carried out on large RNA (>200nt) from control siRNA and TRMT5 depleted PANC-1 cells to validate the loss of m1G signal. Primers were designed to amplify the m1G region of *nestin*, a negative m1G region and a negative RNA (BRD4), where m1G was not identified from meRIP-seq. % of input was normalised against m1G meRIP-qPCR in control siRNA sample. P values were calculated using a paired t test (n=3) and are 0.0034, 0.6069 and 0.3813 respectively.

#### 4.4.5 Loss of TRMT5 reduces cell count and migration of PANC-1 cells

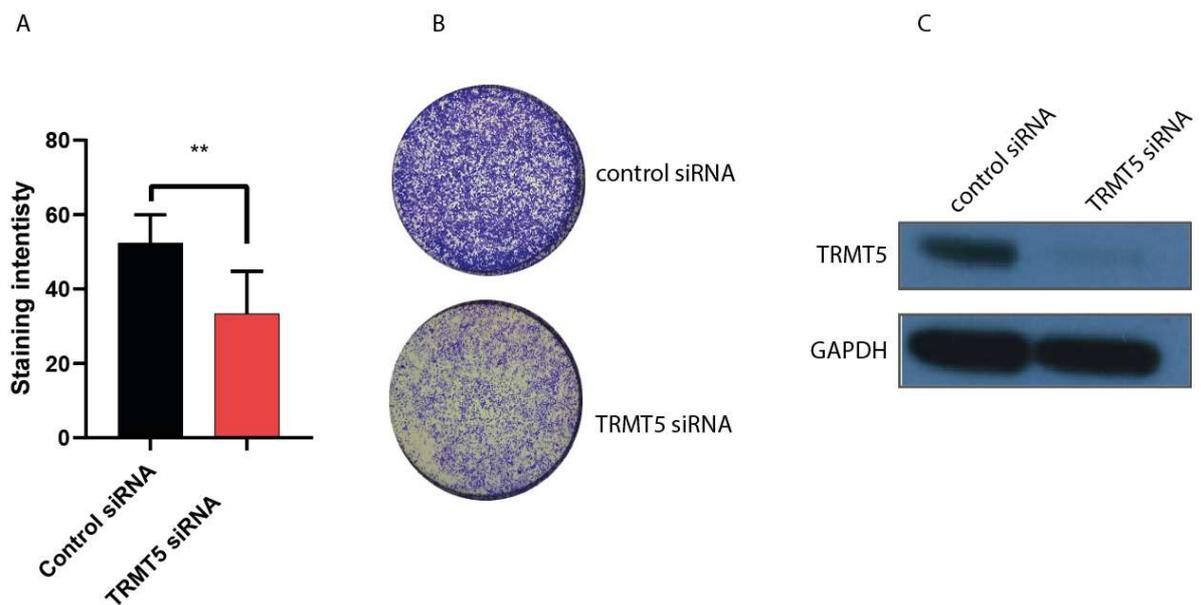
One of the hallmarks of cancer is the ability of the cells to sustain proliferative signalling<sup>166</sup>. Cell lines manifest this characteristic as an unnaturally fast growth rate. Oncogenic proteins are characterised by their contribution to the maintenance of cancer. The data here show an oncogenic effect of TRMT5 in pancreatic cancer cell line PANC-1 (figure 4.8).

TRMT5 was depleted by siRNA and cell counts were carried out to assess any effect on cell proliferation. I observed a statistically significant reduction in the number of PANC-1 cells when TRMT5 was disrupted (figure 4.8A). There was an approximately 50% reduction in cell count when TRMT5 was knocked down after four days. Interestingly, when the same knockdowns and counts were carried out on non-cancerous hTERT-HPNE, there was no significant effect on cell number (figure 4.8B). This cancer-specific effect highlights TRMT5 as a potential vulnerability. The observed phenotype could be due to decreased proliferation or increased death in the cells lacking TRMT5.



**Figure 4.8 - The effect of TRMT5 disruption on cell counts in PANC-1 and hTERT-HPNE cells. A, B)** TRMT5 was knocked down in PANC-1 and hTERT-HPNE by siRNA and cells were counted 72 hours post transfection (n=4).  $P = 0.0082$  and  $0.0131$  respectively using paired t test. **C, D)** Western blotting of TRMT5 and GAPDH was carried out to validate the knockdown of TRMT5.

Another hallmark of cancer is the ability to activate invasion and metastasis. To study the involvement of TRMT5 in the migratory response of PANC-1 cells towards a serum gradient, a standard membrane-based Transwell® assay was used (principle of the Transwell® assay is explained in chapter 2.7.9). Double transfection with two combined siRNAs was used to deplete TRMT5. 24 hours after the second transfection, serum starved PANC-1 cells were seeded into Transwell® chambers and incubated for a further 24h. TRMT5 knockdown was confirmed by Western blotting (Figure 4.9C). Migrated cells were stained with crystal violet and imaged (figure 4.9B). On average the number of migrated cells was 36% lower in the TRMT5 depleted cells, relative to the control cells (figure 4.9A). To conclude, these results show there is a significant reduction in the migratory capacity of PANC-1 cells upon TRMT5 depletion.



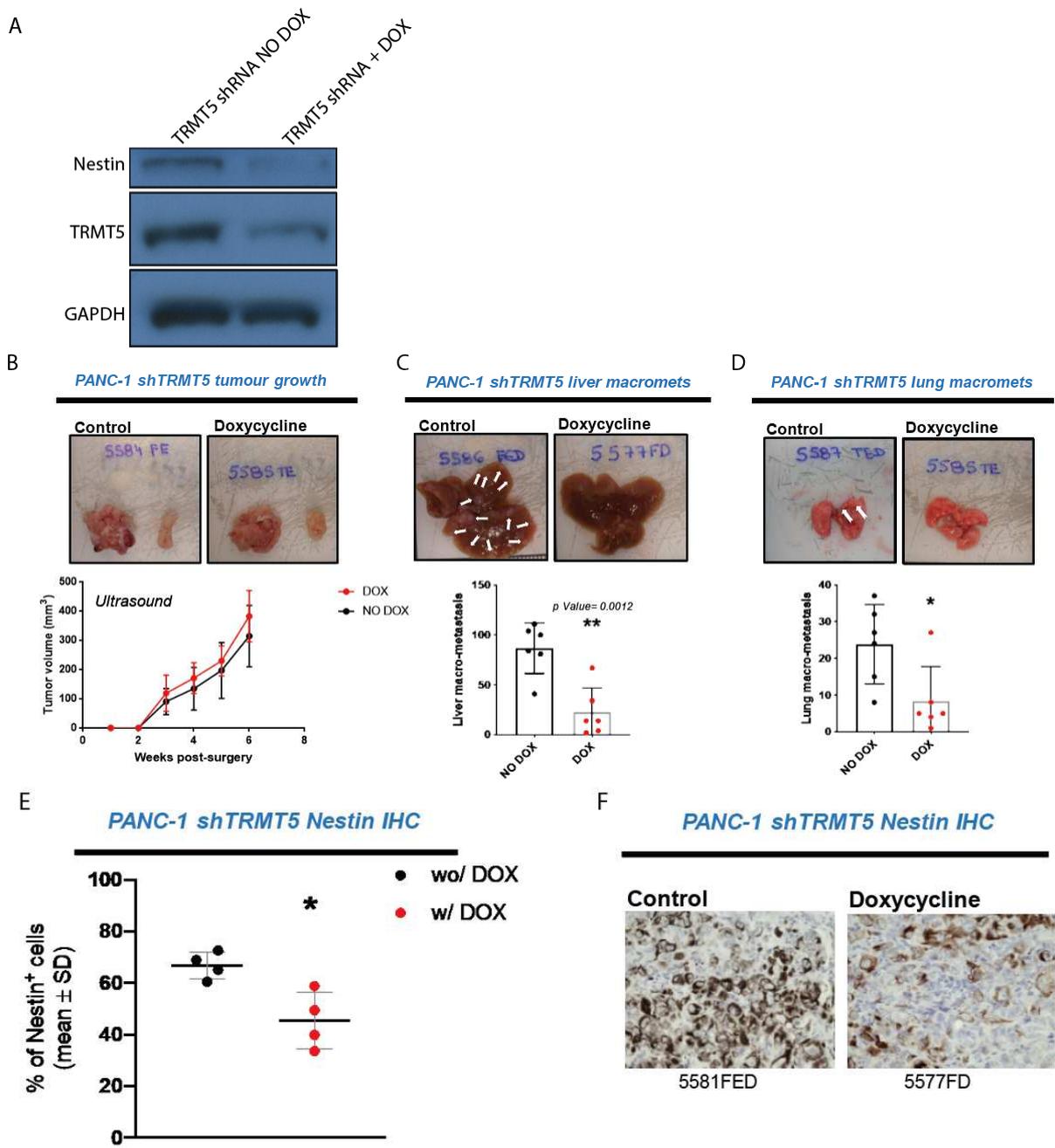
**Figure 4.9 - Loss of TRMT5 in PANC-1 cells causes migratory defect.** PANC-1 cells were double transfected with either TRMT5 siRNA or control siRNA as described in the materials and methods. 24 hours after the second transfection, the cells were serum starved overnight. The following day 75,000 cells were seeded into a Transwell® migration assay. The next day the Transwells® were washed and stained with crystal violet. **A)** Migration was quantified using ImageJ plugin “ColonyArea”. Significance was assessed using a paired T-test ( $P = 0.0322$ ,  $n=4$ ). **B)** Example of crystal violet stained Transwells®. **C)** Cells were taken during assay set up for westerns blotting of TRMT5 GAPDH.

TRMT5 loss from PANC-1 cells *in vitro* induces a migratory defect, suggesting TRMT5 could be essential to drive metastasis in pancreatic cancer *in vivo*. To investigate whether the metastasis of PANC-1 cells *in vivo* was influenced by loss of TRMT5, I decided to use a mouse orthotopic xenograft system to explore this phenotype further.

#### 4.4.6 PANC-1 mouse xenograft

Collaborators in the Sonia Melo laboratory carried out pancreas orthotopic xenograft experiments using an inducible shRNA TRMT5 knockdown PANC-1 cell line that was constructed as part of this thesis work. Following transduction of PANC-1 cells with the TRMT5 shRNA, a single clone was isolated using a FACS cell sorter. This single clone was used in the following experiments.

PANC-1 cells with induced shRNA show decreased levels of TRMT5 and nestin by Western blotting (figure 4.10A). The same inducible shRNA TRMT5 cell line without doxycycline (non-induced) was used as a negative control. Immunocompromised mice were used so that an immune reaction was not generated against the injected PANC-1 cells. The TRMT5 shRNA PANC-1 cells were injected directly into the pancreas and pancreatic tumour formation was monitored by ultrasound. The mice were administered doxycycline via food pellets to induce the knockdown and control mice were fed pellets which did not contain doxycycline. Three endpoints were measured to assess the impact of TRMT5 depletion on pancreatic tumour growth kinetics: (i) tumour weight, (ii) tumour volume and (iii) the number of macromets at the lung and liver. As seen by the comparable size of the pancreatic tumours in DOX+ and DOX- mice, TRMT5 depletion did not impair the tumour growth of pancreatic cells in mice (figure 4.10B). However, TRMT5 depletion significantly impaired the ability of pancreatic tumours to metastasise (figure 4.10C, D). Pancreatic tumours usually metastasise to sites on the lungs and/or liver, therefore macromets at both locations were counted 6-weeks post-surgery to assess the relative metastasis between TRMT5 depleted and negative control treated mice. At both the lungs and liver, there is a striking decrease in the number of macromets when injected PANC-1 cells were depleted of TRMT5. Furthermore, using IHC, nestin expression levels were measured in both groups of mice, shRNA induced and non-induced (figure 4.10E+F). As with the PANC-1 *in vitro* experiments, this also demonstrated that nestin expression is reduced upon TRMT5 knockdown.



**Figure 4.10 - Metastatic deficiency of TRMT5 depleted PANC-1 cells in mice.** **A)** Western blotting of Nestin, TRMT5 and GAPDH was used to confirm the knockdown of TRMT5 after induction for 4 days with doxycycline in PANC-1 cells. **B)** Tumour volume was measured for 6 weeks post-surgery using ultrasound in mice administered with and without doxycycline containing food pellets. **C)** Macromets were counted on the livers of euthanized mice with and without DOX ( $P = 0.0012$  using unpaired t test). **D)** Lung macromets were counted ( $P = 0.0241$  using unpaired t test). Example tumours and macromets are shown for illustration. **E)** Nestin IHC quantification carried out on mouse xenograft tumour, showing induced (W/ DOX) and non-induced mice (wo/DOX),  $n=4$ . **F)** Example nestin IHC images in mouse xenograft tumour.

#### 4.4.7 Is TRMT5 present on chromatin?

Owing to the recent discovery that METTL3 interacts with chromatin and deposits m6A cotranscriptionally on specific mRNAs<sup>120</sup>, I hypothesised that other RNA modifying enzymes could be using a similar mechanism. To explore the possibility that TRMT5 binds chromatin, I used ChIP-seq to search for TRMT5 binding peaks. PANC-1 cells were cross linked with formaldehyde, nuclear extracts were purified, and chromatin was extracted as described in chapter 2. Following the chromatin immunoprecipitation there were detectable levels of DNA eluted, as seen by both Qubit and high sensitivity DNA tapestation analysis. However, after producing ChIP-seq libraries and next generation sequencing there were no statistically significant peaks detected.

## 4.5 Discussion

The role of RNA modifications in cancer is a rapidly growing field and the concept is introducing new possibilities for therapeutic intervention. Many of the enzymes that catalyse these modifications are dysregulated in cancers and require functional characterisation to understand how they could be driving cancer progression. Functional genomic approaches to target identification, such as that by Behan *et al*, show just how many RNA modifying enzymes are potential targets for drugging<sup>126</sup>.

In collaboration with Storm Therapeutics I analysed RNA MS data in a PDAC cell line versus an untransformed pancreatic cell line (figure 4.1). It is clear that several RNA modifications are overrepresented in the PDAC cell line. Highest among these are m1G, m2G and m2,2G. With the availability of the newly characterised anti-m1G RabMAb®, I decided to pursue the role of m1G in PDAC. In pancreatic cancer, m1G levels increase with the aggressiveness of the cancer cell line, but notably the m1G levels are higher in PANC-1, than in the breast cancer cell line MDA-MB231 (figure 4.2E), which suggests m1G could be more important in pancreatic cancer cells than in breast cancer cells. The same pattern was observed using the anti-m1G antibody by ELISA (figure 4.2F). Pancreatic cancer cells are highly proliferative and metabolically active, which could explain the high levels of m1G observed. It is reasonable to assume that increased metabolic activity would require increased translation machinery and therefore m1G modified tRNAs. Furthermore, specific tRNA pools could be in higher demand in pancreatic cancer, where m1G modified ones could be overrepresented. However, we do not know if increased m1G levels are a cause or effect of cancer. It would be interesting to see in future work if TRMT5 overexpression could push non-cancerous cell lines into a transformed state. This would provide evidence of a causal link.

Using the CRISPR screen by Barnett *et al*<sup>126</sup> alongside TCGA data I identified the putative m1G methyltransferases TRMT5 and TRMT10A as strong candidates to study in pancreatic cancer cells. The four TRMT enzymes that were predicted to deposit m1G in humans from yeast homology have recently been reduced to just two<sup>167,168</sup>. Both TRMT10B and TRMT10C have been found to catalyse the formation of N1 adenosine modification, despite their homology to *saccharomyces cerevisiae* m1G methyltransferase trm10. Based on the literature, only TRMT5 and TRM10A appear to be responsible for m1G deposition in humans<sup>75</sup>. These findings correspond with the RNA MS data I have presented, where only TRMT5 and TRMT10A loss reduced m1G levels, not TRMT10B (figure 4.4B). However, this does not rule out the possibility that TRMT10B can also modify guanosine, but to a lesser degree. It

could act in a similar way to the m6A demethylase FTO, which has much stronger activity towards m6Am, but can still demethylate m6A<sup>169</sup>.

Interrogating the two m1G methyltransferases in CRISPR screen databases and TCGA databases, I identified TRMT5 as the more promising cancer target (Figure 4.2C+D). Furthermore, in a similar pattern to the m1G levels, TRMT5 expression correlates with cell line aggressiveness (figure 4.3A). However, this is not the case for TRMT10A expression (figure 4.3B). In the CRISPR screen mentioned previously, loss of TRMT5 produces very strong deleterious effects on the fitness of cancer cells. This is especially promising considering loss of fitness scores in this study are corrected for against matched healthy cells. Therefore, it appears TRMT5 is of particular importance for cancer cells compared to normal cells. Why do cancer cells appear to be more dependent on TRMT5 than TRMT10A? The m1G deposited by TRMT5 lies at position 37 of tRNA which is more important for reading frame fidelity and translation in general<sup>170</sup>. Moreover, loss of m1G at position 37 has been widely reported to affect growth rates in numerous systems including bacteria, yeast and plants<sup>152,74,171</sup>. Healthy cells which are proliferating at a much slower rate are therefore likely to be less impacted by loss of TRMT5. TRMT10A deposits m1G at position 9, which is likely more important for tRNA structure. Alternatively, it is possible that TRMT10A modified tRNAs are held in excess by the cell, such that they are not a rate limiting resource. Even upon knockdown of TRMT10A by siRNA, there may be enough residual fully modified tRNAs to compensate for the loss. It may be useful to create a CRISPR knockout of TRMT10A to investigate a more chronic effect of its loss. Nevertheless, from the data presented here, it appears TRMT5 is the only one of the two that we can show is modifying mRNA. The disparity in effect on cancer cells could be, at least in part, due to this novel function and distinct from translation altogether.

Analysis of stringently purified polyA+ RNA by RNA MS indicated m1G may be present on mRNA (figure 4.4B). However, this data is not conclusive evidence of m1G presence on mRNA but was enough of an indication to proceed with further investigation. Therefore, using the anti-m1G antibody, meRIP-seq was carried out. meRIP-seq identified only a small number of statistically significant TRMT5 dependent peaks. There are several potential reasons for this. m1G may be a very rare modification on mRNA, and the number of targets I detected represents an accurate assessment of m1G modified mRNAs. Or alternatively, our tools may not be sensitive enough to detect other m1G modified targets. Compared to m6A, m1G is indeed very rare (as seen by RNA MS). However, m1G may be more similar to m1A, where only a few RNAs are genuinely modified and at low frequency. Jaffrey *et al* suggest that modifications that produce RT stops (such as m1G) are likely transient and only act to target RNAs for

degradation<sup>36</sup>, which would make detection by meRIP-seq challenging. In future experiments, it would be interesting to assess the half-life of m1G target RNAs upon TRMT5 loss, to test this hypothesis.

Probing the effect of TRMT5 loss on global transcription and translation I observed a significant reduction in translation when TRMT5 was knocked down. Further investigation by Ribo-seq demonstrated that TRMT5 and TRMT10A loss were having opposing effects on a number of mRNAs. An effect on translation was expected given the presence of m1G on several tRNAs, however, I wanted to understand if these effects were coming purely from loss of tRNA modification, or if m1G present on mRNA was having an additive effect.

The meRIP-seq dataset was therefore cross referenced with the Ribo-seq dataset. This was in order to produce a list of mRNAs where firstly TRMT5 dependent m1G peaks were identified, and secondly an effect on translation was observed. The only statistically significant candidate mRNA that was produced from this analysis was *nestin*. Nestin, is a class VI intermediate filament expressed in 30% of PDAC cases. Moreover, its expression in PDAC correlates positively with invasion<sup>172</sup>. Knockdown of *nestin* by shRNA has previously been shown to attenuate invasion and migration measured by scratch and chamber assays<sup>162</sup>. Therefore, I then sought to find if *TRMT5* knockdown would affect nestin protein expression by Western blotting. Indeed, disruption of TRMT5 markedly decreases nestin expression levels (figure 4.7). Furthermore, overexpression of TRMT5 has the converse effect, providing further evidence that TRMT5 and nestin expression levels are linked. This is not direct mechanistic evidence that TRMT5 mediated m1G deposition is solely responsible for the change in nestin expression, but the correlation is as expected. To investigate the mechanistic effect of m1G on *nestin* mRNA, it is crucial to determine the location of the modification more precisely. Based on the data presented here, only the approximate region can be determined under the peak. miCLIP could be carried out with the new anti-m1G antibody to provide single base resolution. Once an exact site is identified, a silent mutation could be introduced to change the guanosine, without altering the amino acid sequence. This would eliminate the mRNA m1G modification but allow the tRNA m1G modification at position 37 to remain. Admittedly, even an mRNA mutation that did not alter amino acid sequence could have an effect on mRNA biology. Essentially, deconvoluting the mRNA effect from the tRNA effect will be highly challenging.

In future work I would be interested to use *nestin* mRNA as a substrate for *in vitro* methyltransferase activity assays. The RNA could either be transcribed *in vitro* (IVT) or purified from PANC-1 cells using probes. TRMT5 could either be produced recombinantly (as well as catalytic dead mutants) or

immunoprecipitated from PANC-1 cells for use in such assays. If successful, this would directly demonstrate that TRMT5 can catalyse m1G on *nestin* mRNA. Alternatively, a targeted nanopore DRS approach could be used on the same IVT/purified *nestin* mRNA. In an attempt to use an antibody independent method, I used HAMR<sup>49</sup> to search for RT stops, and interestingly a site was identified within the *nestin* m1G peak. This provides further confidence of the validity of the antibody-based data. The Kouzarides lab is currently pursuing a collaboration to develop an m1G-specific RT enzyme to validate antibody derived m1G targets. Overall, it is important to validate *nestin* m1G methylation using an antibody independent method.

The link between *nestin* and migration of pancreatic cancer is well established, so I wanted to determine if TRMT5 loss would influence migratory capacity of PANC-1 cells. Figure 4.9 shows that indeed TRMT5 loss does attenuate the migration of PANC-1 cells. It may be that this is, at least in part, driven by the decrease in *nestin* expression. The m1G modification in *nestin* mRNA appears to be necessary for proper *nestin* expression and this is TRMT5 dependent. In future work I aim to try and rescue this phenotype with either TRMT5 or *nestin* exogenous expression. If *nestin* expression can rescue the migratory defect in TRMT5 depleted cells, that would strongly imply that the mechanism is *nestin*-based. Interestingly, Matsuda *et al* found that when *nestin* was knocked down by shRNA in PANC-1 cells, there was an increase expression of filamentous F-actin<sup>162</sup>. It would be interesting to determine if knocking down TRMT5 phenocopied this.

Collaborators in the Sonia Melo laboratory allowed me to investigate the effects of TRMT5 loss using an *in vivo* mouse xenograft system. Although knockdown of TRMT5 did not produce any difference in tumour size in the mice, interestingly the number of metastasised tumours was significantly lower. This observation is concordant with the *in vitro* PANC-1 migration phenotype I observed. *In vitro* there was an obvious proliferation defect, but this was not apparent in the mice. This could be a direct consequence of the migratory defect. For instance, the tumours in the control mice (W/O DOX) were releasing more cells, due to the higher metastasis, which could have been reducing their average size. This would have artificially made the proliferation rates look equivalent in the DOX vs W/O DOX samples. To circumvent this potentially confounding factor, ki67 staining could be carried out in future to get a more accurate picture of proliferative activity of the tumours. Furthermore, IHC shows that TRMT5 knockdown reduces expression of *nestin*, which reinforces the proposed mechanism of TRMT5 dependent *nestin* expression by m1G modification. This highlights TRMT5 as a potential target for therapeutic intervention. Although it does not appear to be a candidate for treating the PDAC itself, it could be targeted to reduce the risk of metastasis.

## Chapter 5 - Characterization of N2, N2-methylguanosine (m2,2G) in ovarian cancer cells

### 5.1 Brief introduction

Along with m1G, m2,2G is also highly overrepresented in PANC-1 RNA compared to the non-cancerous cell line hTERT-HPNE (figure 4.1). Furthermore, the concentration of m2,2G has been shown to be significantly increased in urine samples from cancer patients compared to normal control samples<sup>173</sup>. Taken together, these data highlight m2,2G as a modification that may have a particular relevance in cancer. Although the anti-m2,2G antibody characterised in Chapter 3 did not show enrichment of the modification by meRIP-mass spec, it did show exceptional specificity and signal to noise in ELISA and HTRF<sup>®</sup> assays. Therefore, based on its potential links to cancer and the potential usefulness of the anti-m2,2G antibody, I decided to further explore the role of m2,2G in cancer.

In order to facilitate the study of RNA modifications, identification of the relevant modification enzymes is of great importance. Fortunately, the m2,2G methyltransferase in *Saccharomyces cerevisiae*, tRNA methyltransferase 1 (Trm1p), was one of the earliest tRNA methyltransferases discovered<sup>174</sup>. Trm1p is transported into the nucleus and mitochondria, where it catalysed m2,2G at position 26 of many tRNAs<sup>175</sup>. In addition, Trm1p deletion mutants exhibit changes in tRNA stability that are compounded when combined with deletion of Trm4p or the La binding protein<sup>176</sup>. Furthermore, X-ray crystallography studies have demonstrated that the presence of m2,2G can help prevent alternative tRNA conformations<sup>177</sup>.

In humans two potential orthologues of the yeast Trm1 have been identified: TRMT1L C1ORF25 (Trm1-like protein) and TRMT1<sup>178</sup>. However, although both are predicted to catalyse m2,2G formation, only TRMT1 has been functionally proven to catalyse the modification *in vitro*<sup>179</sup>. This modification is found at position 26 on 9 different tRNAs in humans. Additionally, in humans, TRMT1 has been identified as a novel cause of autosomal recessive intellectual disability (ARID) through genome wide SNP analysis<sup>180</sup>. Individuals with mutations in both copies of *TRMT1* exhibit cognitive impairment, neurodevelopmental delays and facial dysmorphism. TRMT1L has also been linked to similar neurological development problems in mice<sup>181</sup>.

## 5.2 Aims of this chapter

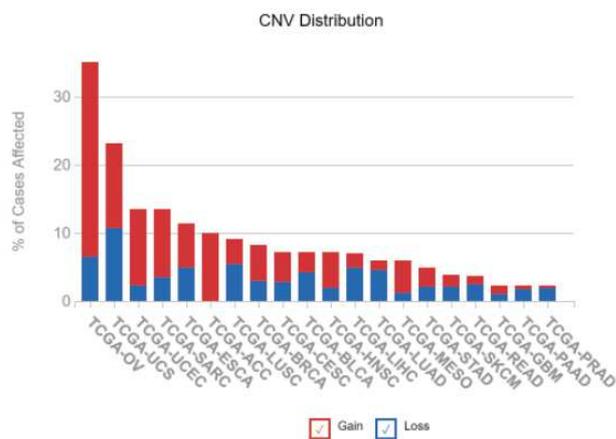
The overarching objective of this chapter is to evaluate TRMT1's role in human cancer. To assess this, the following aims will be pursued:

- i. Determine a system in which to study TRMT1's role in cancer.
- ii. Determine the phenotypic effects of cancer cells in which TRMT1 has been silenced.

## 5.3 Results

### 5.3.1 Expression level of TRMT1 in cancer cells lines

The TCGA database indicates that ovarian cancer has the highest gain of *TRMT1* compared to all other cancer types. Out of 585 cases analysed, 167 cases have gain of *TRMT1* (28.55%). Here, gain refers to an increase in the copy number of the gene. The cancer type with the next highest gain of *TRMT1* is uterine carcinoma with 7 out of 56 cases affected (12.5%). Based on these data, I decided to study the effect of TRMT1 disruption in an ovarian cancer cell line as a model system.



TCGA abbreviation	Cancer name
OV	Ovarian serous cystadenocarcinoma
UCS	Uterine Carcinosarcoma
UCEC	Uterine Corpus Endometrial Carcinoma
SARC	Sarcoma
ESCA	Esophageal carcinoma
ACC	Adrenocortical carcinoma
LUSC	Lung squamous cell carcinoma
BRCA	Breast invasive carcinoma
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
BLCA	Bladder Urothelial Carcinoma
HNSC	Head and Neck squamous cell carcinoma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
MESO	Mesothelioma
STAD	Stomach adenocarcinoma
SKCM	Skin Cutaneous Melanoma
READ	Rectum adenocarcinoma
GBM	Glioblastoma multiforme
PAAD	Pancreatic adenocarcinoma
PRAD	Prostate adenocarcinoma

**Figure 5.1 - TRMT1 copy number alterations in cancer.** TCGA data showing the level of dysregulation TRMT1 in different cancer types. CNV refers to copy number variant.

### 5.3.2 The JHOC-5 ovarian cancer cell model

The JHOC-5 cell line is a human epithelial ovarian cancer cell line generated from a recurrent tumour of a clear cell adenocarcinoma, taken from a 47-year-old Japanese woman. This cell line demonstrates neoplastic and pleomorphic features and grows in multilayers. The cells are approximately 15-20µm in diameter and have a doubling time of 52 hours.

### 5.3.2 Generation of a TRMT1 CRISPR knock-out cell line

In order to study the effect of TRMT1 loss in JHOC-5 cells, the Lenti V2 CRISPR system was used to knockout the gene. Two guide RNA's were designed targeting the first and second methionine of exon 1 of *TRMT1* (figure 5.2A). They were cloned into the backbone vector, transfected into 293T cells and the virus produced was used to infect JHOC-5. To generate a control cell line, a non-targeting guide RNA was used. Knockout of *TRMT1* was confirmed by Western blotting (figure 5.2B).

**A** **sg2**

GGTGCTCTCTAGAGCCCGGTTTTTCGAGTGGCAGTCTCCAGGGCTGCC**GA**  
**ATACAGCAGCGATGGAGAA**CGGCACCCGGGCCCTACGGAGAAGAACGTCC  
ACGTGAAGTCCAGGAGACGACAGTCACCGAGGGGGCTGCCAAAATCGCCT

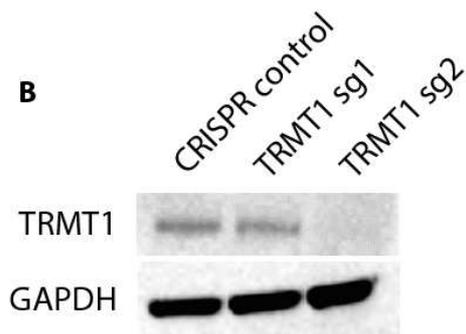
TRMT1



**sg1**

CGAACCAGGCTTGGCGGGCGGAGGCGCCAGCGGATGT**CTCAT**CAAGGAT  
**CGTCTCTGTGG**CTAAGCCTCACTTTCGCTCCGCCCGGGTGCTCTCTAGA  
GCCCGTTTTTCGAGTGGCAGTCTCCAGGGCTGCCGAATACAGCA

**B**



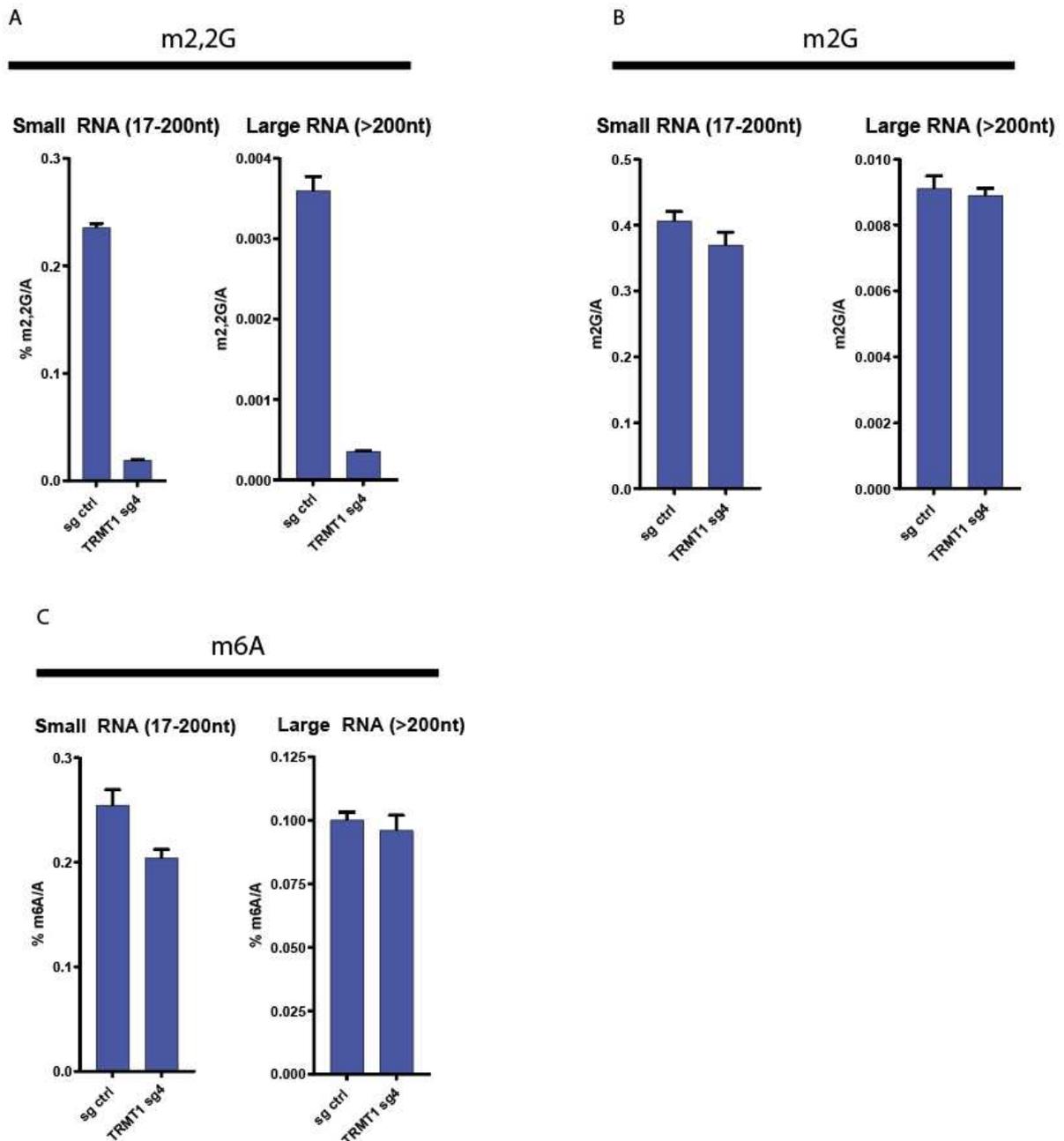
**Figure 5.2 - TRMT1 CRISPR knockout** **A)** Schematic showing the two guide RNAs for the TRMT1 CRISPR knockout. Sg1 targets the first methionine in the first codon and sg2 targets the second methionine in the first codon. Methionine codons are highlighted in green and the PAM sites are shown in red. The guide sequences themselves are shown in bold. **B)** Western blotting was carried out on CRISPR control cells and the two cell lines transfected with TRMT1 targeting guide RNAs. GAPDH was probed as a loading control.

Figure 5.2 shows that only one of the guide RNAs, sg2, successfully knocked out TRMT1. This guide RNA was taken forward for the work described below. Hereafter, the knockout cell line derived using sg2 will be referred to as *TRMT1* knockout.

### 5.3.3 *TRMT1* knockout RNA mass spectrometry

Total RNA was fractionated using size selection columns and analysed by RNA MS (Storm Therapeutics). RNA from 17nt to 200nt will hereafter be referred to as “small RNA” and RNA larger than 200nt will be referred to as “large RNA”. The small RNA fraction will contain tRNA and the large fraction will be mostly tRNA free. This is a useful separation because m<sub>2,2</sub>G is only known to exist on tRNAs in mammals, so the fractionation will aid the search for m<sub>2,2</sub>G in other larger RNA types.

Figure 5.3 shows the results from the RNA MS. Upon *TRMT1* knockout, m<sub>2,2</sub>G levels are dramatically decreased, but m<sub>2</sub>G levels remain constant. In addition, m<sub>6</sub>A levels were measured as a negative control and are also unchanged. This pattern occurs in both large and small RNA fractions. The small RNA fraction contains ~0.25% m<sub>2,2</sub>G in wild type cells, whereas the large RNA fraction contains only ~0.0035%. This difference in abundance was expected because m<sub>2,2</sub>G had only been detected on tRNA. However, the detectable presence of m<sub>2,2</sub>G in the large RNA fraction implies that it may exist on species of RNA other than tRNA. The abundance, however, is extremely low. This indicates there are very few targets, or RNAs that are modified with m<sub>2,2</sub>G are modified rarely. As a point of reference, m<sub>6</sub>A/A levels are consistently around 0.1% in both small and large RNA fractions.

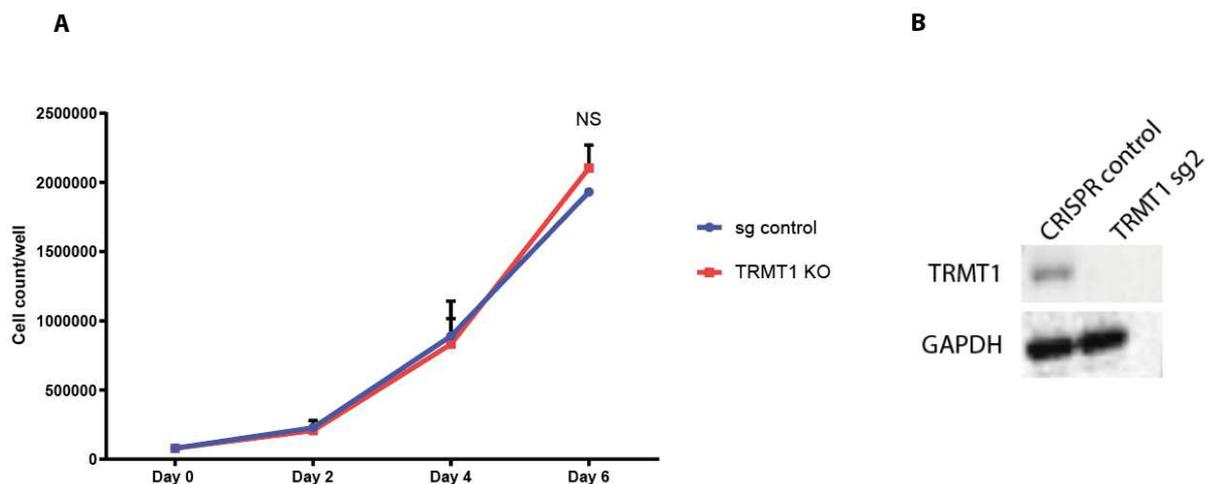


**Figure 5.3 - RNA MS analysis of *TRMT1* knockout.** Total RNA was extracted from both CRISPR control and *TRMT1* knockout JHOC-5 cells. The RNA was then fractionated into large and small fractions using RNA purification columns. **A)** m2,2G, **B)** m2G and **C)** m6A levels were measured using RNA MS and are shown here normalised to adenosine. Each sample was measured in triplicate and error bars show the standard deviation.

### 5.3.2 Biological effects of TRMT1 loss

#### Effect on cellular proliferation

To understand whether loss of TRMT1 affects cell growth of JHOC-5 cells, cell counts were carried out at time intervals on both wild type cells and *TRMT1* knockout cells (figure 5.4A). Successful knockout was confirmed by Western blotting for TRMT1 (figure 5.4B). The cells were grown using puromycin free media during this experiment to minimise antibiotic induced stress. The cells were seeded at equal densities in 6-well plates and counted after 2, 4 and 6 days of growth. These results show that TRMT1 loss has no effect on proliferation of JHOC-5 cells.



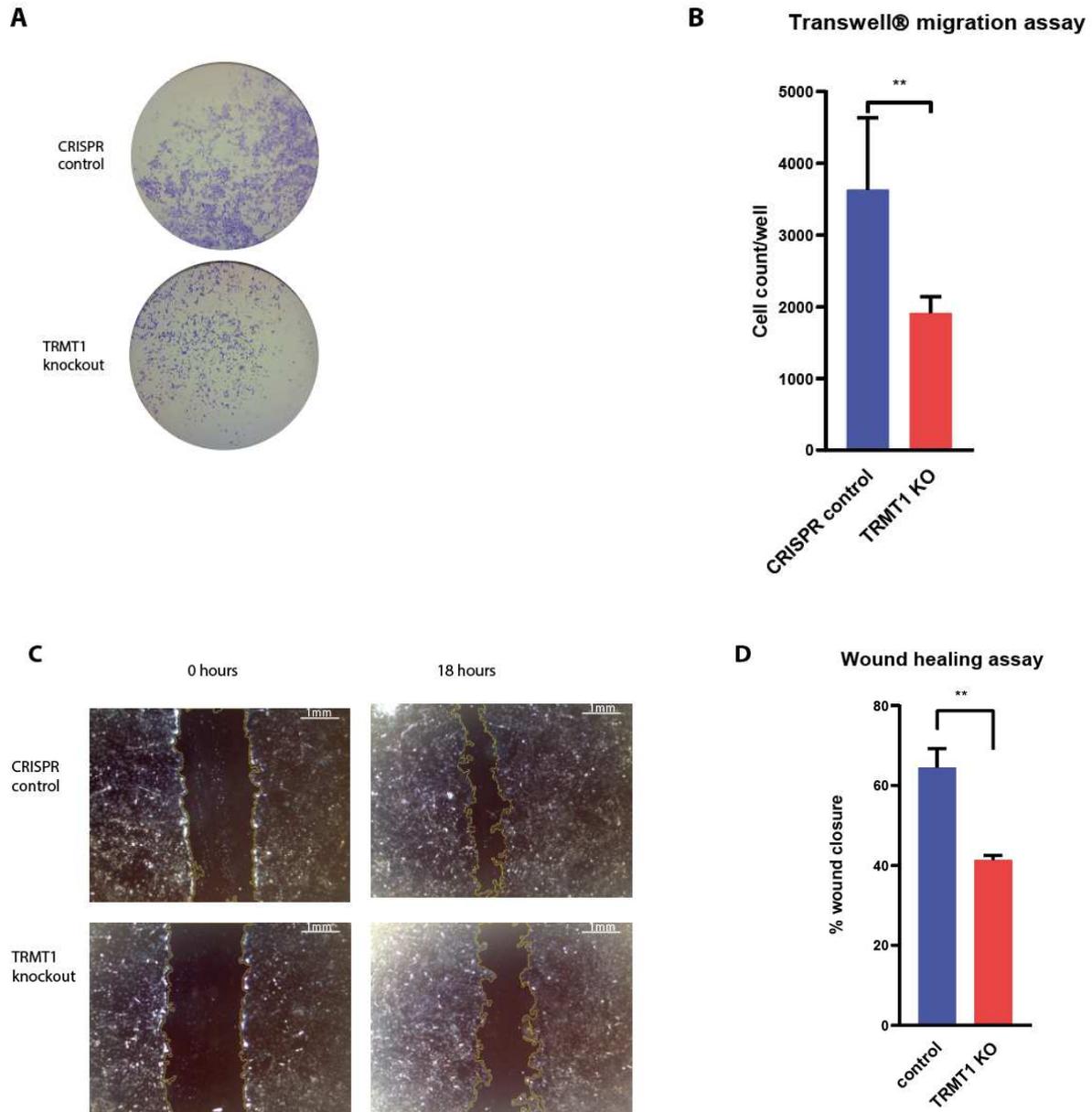
**Figure 5.4 - Proliferative effects of *TRMT1* knockout.** **A)** CRISPR control and *TRMT1* knockout JHOC-5 cells were grown in 6-well plates and cell counts were carried out at 2 day intervals as indicated.  $P = 0.77$  using Mann Whitney test. Error bars show the standard deviation. **B)** Western blot was carried out on cell extracts from the 6 day timepoint using anti-TRMT1 antibody to confirm knockout and anti-GAPDH to confirm equal loading.

### **Effect on cellular migration**

To study the migratory response of JHOC-5 cells towards a serum gradient, a standard membrane-based Transwell® assay was performed using CRISPR control cells and *TRMT1* knockout cells. The principle of Transwell® assay is explained in Chapter 2.7.9.

CRISPR control or *TRMT1* knockout JHOC-5 cells were seeded into Transwell® chambers and incubated for 24h. Successful *TRMT1* targeting was confirmed by Western blotting (figure 5.4B). Migrated cells were stained with crystal violet and imaged (figure 5.5A). The results show a significant reduction in the migratory capacity of JHOC-5 cells upon loss of *TRMT1*.

Next, wound healing scratch assays were performed with CRISPR control cells and *TRMT1* knockout JHOC-5 cells to assess cellular invasion. Cells were grown to confluence in 6-well dishes and a scratch was created using a pipette tip. Images were taken at regular intervals to monitor the closing of the lesion. Despite there being no difference in proliferative capacity of the two cell lines, the *TRMT1* depleted cells were much slower to repair the lesion (figure 5.5C).

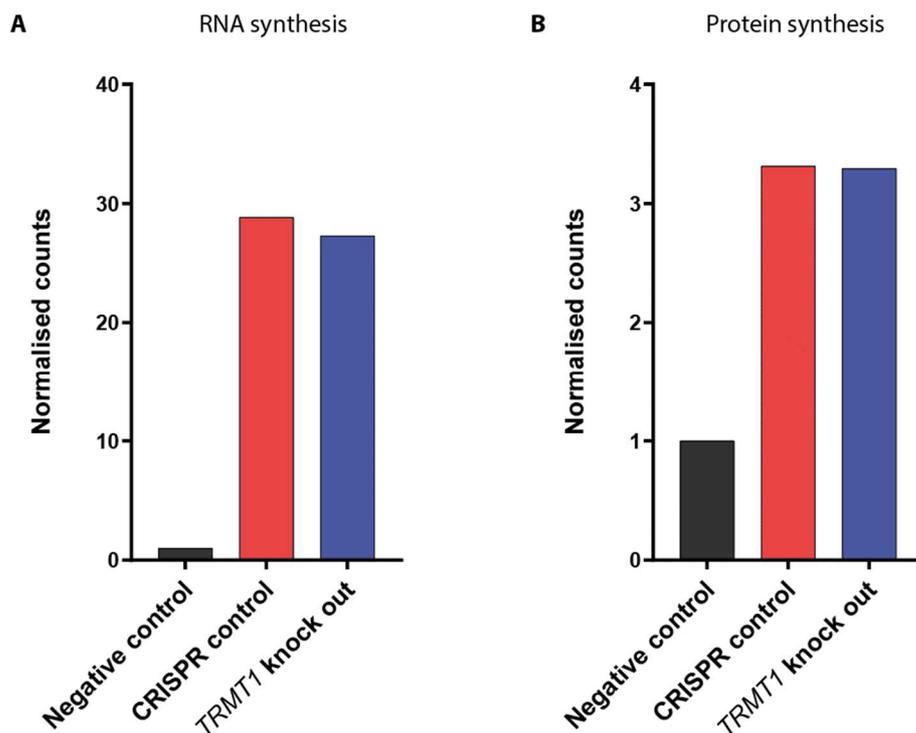


**Figure 5.5 - Effect of *TRMT1* knockout on migration in JHOC-5 cells. A-B)** CRISPR control and *TRMT1* knockout JHOC-5 cells were used in a Transwell® migration assay. Migrated cells were stained with crystal violet and imaged using ZEISS Stemi 305 microscope. The migrated cells were quantified using ImageJ colony area plugin. Mean and standard deviation are shown ( $p = 0.0022$  using Mann Whitney test,  $n=4$ ). **C-D)** CRISPR control and *TRMT1* knockout JHOC-5 cells were grown to confluency in 6-well plates. A wound/scratch was then introduced, and the plates were imaged at 0 and 16 hours. Wound area was calculated using ImageJ wound healing plugin. Mean and standard deviation are shown ( $p = 0.0012$  using unpaired t-test,  $n=4$ ).

### 5.3.5 Transcriptional and translational effect of TRMT1 loss from JHOC-5 cells

To measure the global RNA transcription levels in *TRMT1* knockout and CRISPR ctrl JHOC-5 cells, an RNA synthesis kit (ab228561, Abcam) was utilised. This kit relies on the incorporation of 5-ethynyl uridine into nascent RNA. The RNA with incorporated 5-ethynyl uridine is then labelled with fluorescent azide-containing dye by click chemistry. Flow cytometry is then used as a read out for staining to measure the amount of nascent RNA in a given sample. Using this method, there was no observable change in global RNA synthesis in cells lacking TRMT1 (figure 5.6A).

Next, a global protein synthesis kit (ab235634, Abcam) was used, followed by flow cytometry to compare the levels of protein synthesis in *TRMT1* knockout cells to CRISPR ctrl JHOC-5 cells. This kit relies on a click chemistry reaction, whereby fluorescent azide labels nascent polypeptide chains. The counts detected by flow cytometry are then proportional to the global protein synthesis levels. As TRMT1 is known to methylate tRNA, it is expected to play a role in translation. Unexpectedly, by this method there was no obvious difference in global translation rates upon knockout of *TRMT1* (figure 5.6B).



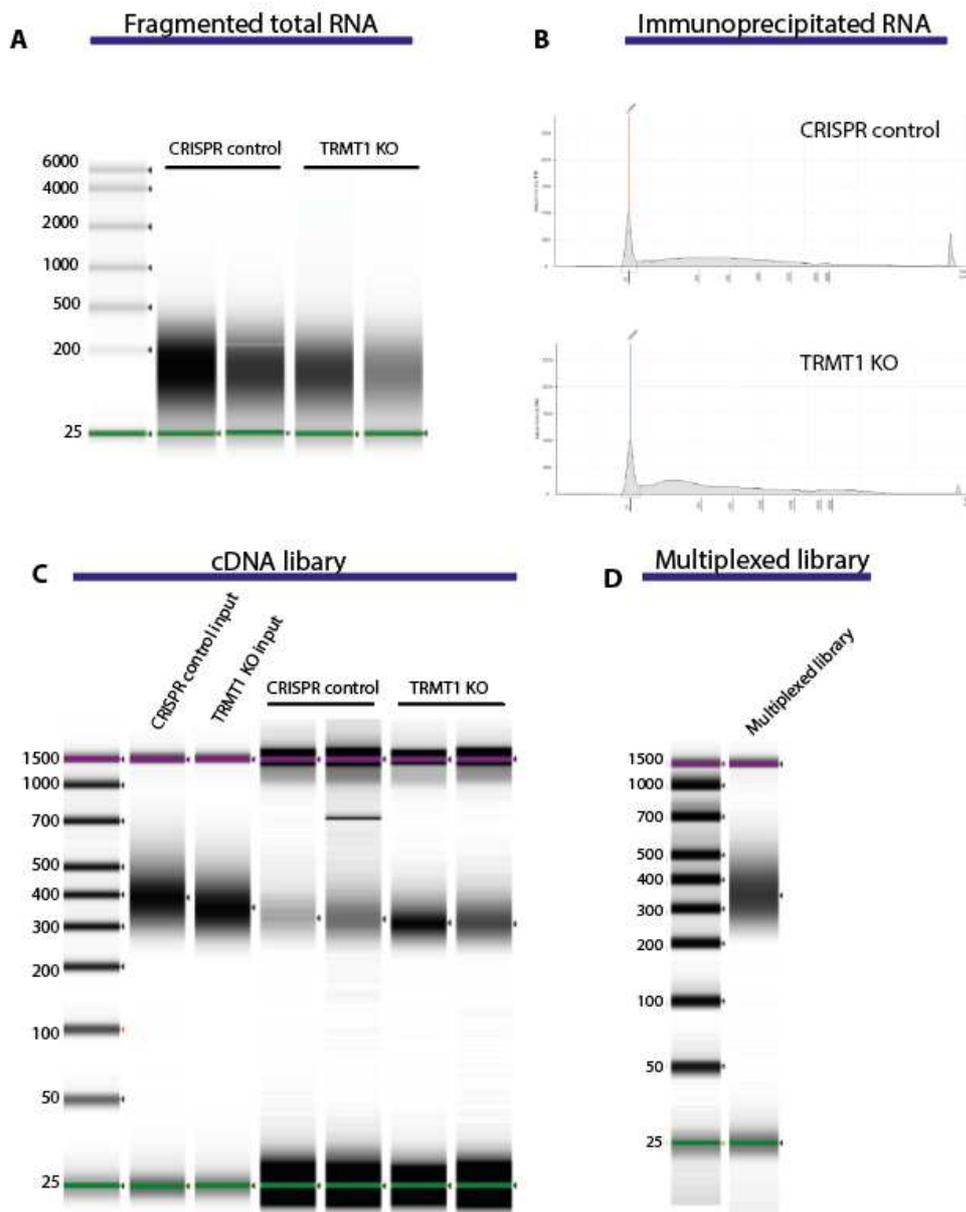
**Figure 5.6 - Global transcription and translation changes upon *TRMT1* knockout.** CRISPR control and *TRMT1* knockout JHOC-5 cells were grown in 10cm dishes without puromycin selection to ~70% confluency before using **A)** an RNA synthesis kit and **B)** a global protein synthesis kit to measure the rate of transcription and translation respectively. Signal was measured using Flow cytometry and the number of counts is shown. Negative control represents cells that were not treated with the fluorescent azide (n=1).

### 5.3.6 Mapping m<sup>2</sup>,2G within the transcriptome

#### **RNA immunoprecipitation (MeRIP)**

meRIP-seq was carried out using the anti-m<sup>2</sup>,2G antibody described in Chapter 3. Although the antibody was not found to enrich the modification as determined by RNA MS, meRIP-seq was performed regardless. It was reasoned that the high sensitivity of next generation sequencing might extract information that RNA MS could not. As reported in a recent paper regarding m<sup>1</sup>A, there may be a very small number of mRNA transcripts modified, if any at all<sup>36</sup>. If this were the case with m<sup>2</sup>,2G on mRNA, meRIP-seq would provide a chance of identifying potential candidate mRNAs.

Total RNA was fragmented to ~100nt and incubated with the anti-m<sup>2</sup>,2G antibody overnight as described in Materials and Methods (Chapter 2). The antibody was subsequently captured with a mixture of protein G and protein A magnetic beads, before washing and eluting with SDS and proteinase K. A small amount of RNA was detectable by tapestation after elution (figure 5.7B), so libraries were prepared and sequenced by 50bp paired-end sequencing on an Illumina HiSeq 1500<sup>®</sup> system.



**Figure 5.7 - meRIP and cDNA library preparation using the m2,2G RabMAb®.** **A)** Total RNA from CRISPR control and *TRMT1* knockout JHOC-5 cells was fragmented to ~100 nucleotides. The fragmented RNA was analysed on RNA tapestation. **B)** Following meRIP with the m2,2G RabMAb®, the immunopurified RNA was analysed on RNA tapestation HS, showing size distribution of RNA fragments as electropherogram image. **C)** Immunoprecipitated RNA and input RNA was used to produce cDNA libraries for sequencing. The individual cDNA libraries were analysed on DNA tapestation HS. **D)** The six individually barcoded libraries were multiplexed at an equimolar ratio using concentration derived from Qubit DNA HS and average size from tapestation. The multiplexed library was analysed again for size distribution using DNA tapestation HS.

Despite having successfully created libraries from both CRISPR control and *TRMT1* knockout eluted RNA (figure 5.7), bioinformatical analysis of the resulting sequencing data did not identify any significant peaks.

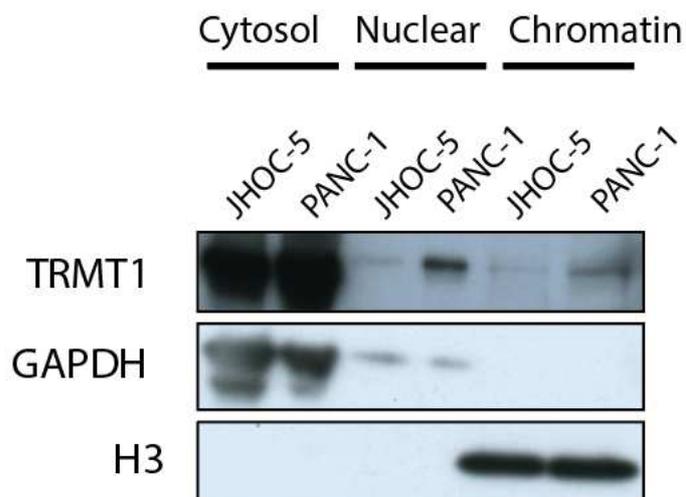
### 5.3.6.2 Nanopore direct RNA sequencing

A recent paper published by the Kouzarides lab shows how Oxford Nanopore direct-RNA sequencing (DRS) can be employed to identify RNA modification sites<sup>65</sup>. In this paper the analytical framework, Nanocompore was used to predict the presence of RNA modifications. Using this method, no “training” of the system is needed. Instead an experimental RNA sample is compared to a control sample with fewer or no modifications, through knockdown or complete loss of writer enzymes respectively. Despite only being validated with the abundant m6A modification, the authors suggest that this method can potentially be used for any modification, provided that the control sample is sufficiently depleted of the modification of interest. Given the significant loss of m2,2G upon *TRMT1* knockout (figure 5.2), I decided to use this technique in an attempt to map m2,2G within mRNAs. I prepared total RNA from JHOC-5 cells with CRISPR control and *TRMT1* knockout in triplicate for analysis by nanopore DRS. Total RNA was extracted using QIAzol reagent as described in Materials and Methods (Chapter 2). Analysis using Nanocompore did not detect any significant differences in signal between RNA isolated from CRISPR control versus *TRMT1* knockout cells (data not shown).

### 5.3.7 Chromatin association of TRMT1 in JHOC-5 ovarian cancer cell line

As described previously in the literature, mRNA modifying enzymes have been shown to bind to chromatin and co-transcriptionally modify nearby transcripts<sup>120</sup>. It is possible that other RNA modifying enzymes also regulate gene expression through their association with chromatin.

Using chromatin extraction techniques followed by Western blotting, I investigated which cellular compartments contain TRMT1, and in which proportions (figure 5.8). In addition to JHOC-5 cells, the same protocol was carried out on PANC-1 cells. Interestingly, TRMT1 is found in small amounts in both the nuclear and chromatin fractions of both cell lines. However, TRMT1 is generally more highly expressed in all 3 cellular fractions of PANC-1 than in the corresponding fractions of JHOC-5 cells. Notwithstanding, JHOC-5 cells were not selected as the model system based on expression levels, but on the TCGA data described above (figure 5.1).



**Figure 5.8 - Cellular localisation of TRMT1.** Cell fractionation was carried out on JHOC-5 and PANC-1 cells, to determine the subcellular localisation of TRMT1. Cytosolic, nuclear and chromatin extracts were Western blotted for TRMT1, GAPDH and histone H3 as indicated.

Although this experiment does not provide solid evidence that TRMT1 is present on chromatin, the results were sufficient to justify a subsequent ChIP-seq experiment.

### 5.3.8 Genome-wide mapping of TRMT1 using chromatin immunoprecipitation (ChIP)

ChIP-seq was subsequently performed on chromatin isolated from *TRMT1* knockout and CRISPR control JHOC-5 cells to search for TRMT1 binding sites. ChIP was performed using a nuclear extraction method as described in Chapter 2.11. In the absence of directly detecting m<sup>2</sup>,2G sites using meRIP-seq or Nanopore DRS, this ChIP-seq approach could act as a proxy to determine potential substrates. Despite producing successful ChIP-seq libraries, bioinformatic analysis did not identify any TRMT1 associated peaks (data not shown).

## 5.4 TRMT1 discussion

TRMT1 is highly expressed in many cancers and especially ovarian cancers. However, its role in cancer is yet to be determined. The phenotypic changes in migratory behaviour shown here suggest that TRMT1 overexpression may be enhancing the ability of cancer cells to metastasise. The m<sup>2</sup>,2G deposited by TRMT1 on tRNAs is likely necessary for their stability and correct folding. If the nucleotide sequences decoded by these tRNAs are overrepresented in cancer promoting mRNAs, it is feasible that there is an increased dependence on TRMT1 in some cancers. For example, metastasis associated transcripts could be especially dependent on TRMT1, which could explain the migratory phenotype observed above. More work is needed to understand the biology behind TRMT1 function, but these preliminary results highlight the potential role of TRMT1 in ovarian cancer.

The lack of an apparent effect of TRMT1 loss on transcription or translation (figure 5.6) was unexpected. It is difficult to explain how the loss of a highly conserved tRNA modification could have no effect on these fundamental biological processes. The drastic decrease in m<sup>2</sup>,2G levels as seen by RNA MS demonstrates that there is not any m<sup>2</sup>,2G enzyme redundancy. But perhaps an alternative modification could occur in its place. It could be informative to perform a nuclear run-on assay such as GRO-seq (Global run-on sequencing) in addition to Ribo-seq to more thoroughly characterise any potential transcriptional and translational effects of TRMT1 loss. Ribo-seq would allow a much more specific identification of translational changes and may uncover small effects that were not discernible at a more global level. Likewise, GRO-seq would allow more granular interrogation of the effect of TRMT1 loss on transcription. As mentioned previously, RNA modifications and their effects appear to be highly context dependent. Speaking to this, a recent paper by Dewe *et al*<sup>182</sup> also investigated the loss of TRMT1. Conversely though, the authors found phenotypic alterations that I did not. For example, in their 293T cell line, loss of TRMT1 produced a proliferation defect and reduction in total global protein synthesis. However, in agreement with the work presented in this thesis, the authors observed that TRMT1 disruption greatly affected m<sup>2</sup>,2G levels in total RNA, but did not change m<sup>2</sup>G levels. 293T cells are a human embryonic kidney cell line and are significantly different to the ovarian cancer cell line JHOC-5 used in this thesis. There may be redundancy or mutations in the cancer cell line that circumvent the reliance on TRMT1 for proliferation. Furthermore, it would be interesting to search for alternate conditions where TRMT1 loss may affect proliferation. For example, various stress conditions or serum concentrations may uncover specific dependencies on TRMT1.

In section 5.3.7, I demonstrated that TRMT1 is present in the chromatin fraction of JHOC-5 cells. However, it is obvious that PANC-1 cells have a higher overall expression of TRMT1 than JHOC-5 cells, and correspondingly higher proportion of TRMT1 located on chromatin. This is a crude experiment and not quantitative, but this does suggest that PANC-1 cells, and by extension, other cell lines may be better suited for detecting TRMT1 acting on chromatin. Alternatively, there may be more effective ways at detecting TRMT1 on chromatin. For instance, double cross-linking using both EGS (ethylene glycol bis (succinimidyl succinate)) and formaldehyde could enhance ChIP efficiency. Moreover, using an exogenous tagged version of TRMT1 may result in ChIP-seq peaks due to the higher concentration of TRMT1 in the reaction. Additionally, anti-tag antibodies may perform better in ChIP experiments than the anti-TRMT1 antibody used here.

The data presented in this thesis demonstrate the necessity of TRMT1 for m<sup>2</sup>,2G presence in JHOC-5 cells. This is not to say that TRMT1 is the methyltransferase directly responsible, but it does appear to be required for m<sup>2</sup>,2G deposition. It is possible that TRMT1 is in fact a co-factor or part of a larger complex responsible for the modification. More work needs to be carried out to express recombinant TRMT1 for *in vitro* methyltransferase assays. Upon knockout of *TRMT1*, m<sup>2</sup>,2G levels are significantly reduced as expected. Surprisingly however, m<sup>2</sup>G levels were not affected. This is in contradiction to the Modomics RNA modification database and other papers, which suggest TRMT1 is responsible for m<sup>2</sup>,2G and m<sup>2</sup>G modification in humans<sup>178</sup>. Although of course, there may be more redundancy in mammalian m<sup>2</sup>G pathways compared to m<sup>2</sup>,2G pathways, where another m<sup>2</sup>G catalysing enzyme(s) compensates for loss of TRMT1 activity.

Many RNA modifications, so far only identified on tRNA and rRNA, may indeed be present on many other RNA species. However, regarding m<sup>2</sup>,2G I find no compelling evidence of m<sup>2</sup>,2G existing beyond tRNA. The low m<sup>2</sup>,2G signal detected on the large RNA fraction (significantly depleted of tRNAs) could still be due to contaminating tRNAs or their fragments. I attempted to address this by performing meRIP-seq with the anti-m<sup>2</sup>,2G antibody, where I hoped to identify peaks of m<sup>2</sup>,2G in mRNAs, whose signals were diminished by TRMT1 loss. Unfortunately, the meRIP-seq (and the nanopore DRS) did not identify any m<sup>2</sup>,2G candidate positions. If another anti-m<sup>2</sup>,2G specific antibody is developed that can demonstrably immunopurify the modification, it would be well worth repeating meRIP-seq to search for novel target RNAs. However, with the current tools at hand, I cannot make any claims about the existence of m<sup>2</sup>,2G in mammalian mRNA. In future work, I believe it would be useful to perform highly stringent mRNA purification using the same method used in section 4.3.2. This would allow us to be more confident about the existence of m<sup>2</sup>,2G on mRNA.

The data presented in this chapter rely heavily on a single clone of *TRMT1* knockout. It is possible, although unlikely, that the phenotypic effects observed are due to CRISPR off-target effects. In future, I plan to construct additional guide RNAs and *TRMT1* knockouts to corroborate the results shown with the one guide RNA presented here.

## Chapter 6 - Final conclusions and overall discussion

It is abundantly clear that RNA modifications are not merely the structural necessities confined to highly abundant RNA species, they were once thought to be. Accumulating evidence is linking RNA modifications to a wide range of biological mechanisms and therefore diseases. Moreover, as our detection techniques improve, we are finding them in low abundance RNA species. Some modifications on low abundance RNA species, such as m6A, have been well characterised. However, such modifications could be the 'low hanging fruit'. m6A is abundant, widely spread throughout the transcriptome and modified RNAs are frequently modified. We hypothesised that such marks are the 'tip of the iceberg'. Other, rarer modifications may exist beyond our current ability to detect.

### 6.1 In depth characterisation of novel RNA modification antibodies

Producing antibodies to RNA modifications pushes the limits of what can be done with such binders. However, previous studies demonstrate that they can be very useful and are, in principal, possible to produce. Here I have presented the validation of eight RabMAb<sup>®</sup> antibodies provided by Abcam Plc. Two binding assays on modified RNA oligonucleotides were used to assess specificity, and an immunoprecipitation coupled to MS was used to determine if the antibody could pull-down the modification. The combination of these assays highlighted that many RNA modification antibodies can be specific but not capable of immunoprecipitating modification containing RNA. Binding of the antibody to an immobilised oligonucleotide does not appear to be totally representative of the quality of the antibody as a pull-down reagent. Furthermore, the number of known modifications exceeds 170, but checking cross-reactivity to all of them is obviously not feasible. Because of this, antibodies to RNA modifications must be seen as a first step approach to identify candidate modification sites, rather than infallible proof of a novel mark. Follow up validation involving enzyme disruption, or orthogonal approaches must be used before we can be confident of a modification's existence.

The success rate of these antibodies is relatively low. Initial screening for binding against the antigen, or modified oligonucleotides gives promise of success, but the candidates frequently fail to work in immunoprecipitation, as seen in this work. Further investigation could look at a higher throughput immunoprecipitation RNA MS assay to select antibody clones earlier on in development. This would allow earlier identification of potentially useful antibodies and could be followed up with specificity screening.

Additionally, another validation step that could be employed in future work would be using knock out/down of writer enzymes in lower organisms. For example, total RNA could be produced from knock out strains of yeast representing known writer enzymes for particular modifications. Yeast generally has less redundancy of enzymes and is a much simpler tool for producing knock out strains. For example, to validate the anti-m1G antibody on yeast RNA, only Trm5 and Trm10 knock out strains are needed and are readily available. Such screening as this could be used to give an early indication of specificity during development. However, in situations where writer enzymes are not known for a given modification, this method would be unavailable.

Epitranscriptomics is in its infancy and claims around RNA modifications are controversial and often contested. For example, the overestimation of m1A targets was due to antibody cross-reactivity and demonstrates how cautious researchers need to be when studying a new or uncharacterised modification. Antibody characterisation is of critical importance, and ideally orthogonal techniques are employed to validate antibody-based discoveries. The work described here has produced three potentially useful RNA modification antibodies (m1G, m6,6A and m1A). The anti-m1G antibody has been utilised in a subsequent chapter, and further characterised using enzyme disruption.

## 6.2 TRMT5 and TRMT1 final discussion

In the work presented in this thesis, I show evidence that the canonical modification of tRNA by TRMTs may be accompanied by modification of a small number of low abundance RNAs, such as *nestin* mRNA by TRMT5. However, these events appear to be incredibly rare. Jaffrey *et al* suggest that if there are yet-to-be discovered mRNA modifications, they are likely very uncommon, especially hard stop modifications<sup>36</sup>. The relative incompatibility of these modifications with translation machinery could induce degradation under such circumstances<sup>183</sup>. It may be that the cell treats aberrant modifications in the same way as modifications arising from RNA damage. This would explain the rarity of such modifications if they are only transiently present on mRNAs. However, if m1G on *nestin* were causing mRNA decay we would expect to see *nestin* protein levels increase with the loss of TRMT5, but the opposite is true. The data presented here generally agrees with the proposition that novel modifications are rare on mRNA.

Another pressing question is that if TRMTs are depositing modifications away from their canonical targets, is this a bug or a feature of their activity? Has this evolved as a fine-tuning mechanism for a small number of mRNAs, or under specific conditions, or does it occur erroneously and is selected against? If mutations in specific cancers cause increased TRMT expression, does the incidence of such modifications climb too? In the case of METTL3, the expression pattern and subsequent modification pattern is highly contextual. It is reasonable to assume that other, albeit rarer, mRNA modifications could be following a similar pattern. In future work, it could be informative to carry out the same meRIP-seq on normal cells where TRMT5 levels are much lower. Along the same line, Dewe *et al* found loss of TRMT1 to have significant effects on the biology of their cell line, where in JHOC-5 cells I observed no changes.

There are now numerous studies focussing on individual modifications on isolated RNA. However, to date there has been little research on the cross talk between multiple modifications. One recent paper highlighted the interaction between TRMT10A and FTO<sup>184</sup>. In this paper, TRMT10A ablation alters the m6A methylome in mRNA. Interestingly, the authors also found that the transcripts with increased m6A upon TRMT10A knockdown contained an overrepresentation of m1G9-containing tRNAs that are modified by TRMT10A. This implies that there is cross talk between modifications on different RNA families (tRNA and mRNA). It would be interesting in future work to analyse the effect of TRMT5 loss on other modifications, which could be done on the current MS datasets available. Alternatively, m6A-seq could be carried out on RNA derived from TRMT5 depleted cells, to search for cross talk between m1G and m6A.

In summary, my PhD characterised a new group of anti-RNA modification antibodies and highlights the importance of thorough characterisation for their use in future. I have identified a novel role for TRMT5 through the deposition of m1G on a small group of mRNAs, which may implicate more TRMTs as mRNA modifying enzymes. Furthermore, I have shown that TRMTs can be necessary for proliferation and migration of cancer cells both *in vitro* and *in vivo*, underlining their potential as drug targets.

## Bibliography

1. Boccaletto, P. *et al.* MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).
2. Mishima, Y. & Tomari, Y. Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Mol. Cell* **61**, 874–885 (2016).
3. Tushev, G. *et al.* Alternative 3' UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments. *Neuron* **98**, 495–511.e6 (2018).
4. Leppke, K., Das, R. & Barna, M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology* **19**, 158–174 (2018).
5. Cabili, M. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
6. Xiu Cheng Quek, Daniel W. Thomson, Jesper L.V. Maag, Nenad Bartonicek, Bethany Signal, Michael B. Clark, Brian S. Gloss, and M. E. D. lncRNADB v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Research* (2015). Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4384040/#B6>. (Accessed: 10th July 2020)
7. Lina Ma, Ang Li, Dong Zou, Xingjian Xu, Lin Xia, Jun Yu, Vladimir B. Bajic, Z. Z. lncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs | Nucleic Acids Research | Oxford Academic. *Nucleic Acids Research* (2014). Available at: <https://academic.oup.com/nar/article/43/D1/D187/2438893>. (Accessed: 10th July 2020)
8. Sakai, Y., Miyauchi, K., Kimura, S. & Suzuki, T. Biogenesis and growth phase-dependent alteration of 5-methoxycarbonylmethoxyuridine in tRNA anticodons. *Nucleic Acids Res.* **44**, 509–523 (2016).
9. Lorenz, C., Lünse, C. & Mörl, M. tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules* **7**, 35 (2017).
10. Windows, M., Corporation, M., Hori, K. & Sakajiri, A. Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* (1957).
11. BROMBERG, P. A., GUTMAN, A. B. & WEISSMANN, B. *The purine bases of human urine. I. Separation and identification. The Journal of biological chemistry* **224**, (1957).
12. Perry, R. P. & Kelley, D. E. Existence of methylated messenger RNA in mouse L cells. *Cell* **1**, 37–42 (1974).
13. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 3971–3975 (1974).
14. Dubin, D. T. & Taylor, R. H. The methylation state of poly A-containing-messenger RNA from

- cultured hamster cells. *Nucleic Acids Res.* **2**, 1653–1668 (1975).
15. Jia, G. *et al.* N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
  16. Mauer, J. *et al.* Reversible methylation of m6Am in the 5' cap controls mRNA stability. *Nature* **541**, 371–375 (2017).
  17. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).
  18. Meyer, K. D. *et al.* Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. *Cell* **149**, 1635–1646 (2012).
  19. Li, X. *et al.* Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. *Nat. Chem. Biol.* **12**, 311–316 (2016).
  20. Liu, F. *et al.* ALKBH1-Mediated tRNA Demethylation Regulates Translation. *Cell* **167**, 816–828.e16 (2016).
  21. Zheng, G. *et al.* ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* **49**, 18–29 (2013).
  22. Hussain, S. *et al.* NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep.* **4**, 255–261 (2013).
  23. Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).
  24. Hussain, S., Aleksic, J., Blanco, S., Dietmann, S. & Frye, M. Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol.* **14**, (2013).
  25. Chen, K. *et al.* High-resolution N6-methyladenosine (m6A) map using photo-crosslinking-assisted m6A sequencing. *Angew. Chemie - Int. Ed.* **54**, 1587–1590 (2015).
  26. Liu, N. & Pan, T. Probing N 6-methyladenosine (m6A) RNA Modification in Total RNA with SCARLET. in 285–292 (Humana Press, New York, NY, 2016). doi:10.1007/978-1-4939-3067-8\_17
  27. Molinie, B. *et al.* m6A-LAIC-seq reveals the census and complexity of the m6A epitranscriptome. *Nat. Methods* **13**, 692–698 (2016).
  28. Schaefer, M., Pollex, T., Hanna, K. & Lyko, F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.* **37**, (2009).
  29. Khoddami, V. & Cairns, B. R. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* **31**, 458–464 (2013).
  30. Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O. & Sorek, R. Transcriptome-Wide Mapping of 5-methylcytidine RNA Modifications in Bacteria, Archaea, and Yeast Reveals m5C

- within Archaeal mRNAs. *PLoS Genet.* **9**, e1003602 (2013).
31. Delatte, B. *et al.* Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science (80-. ).* **351**, 282–285 (2016).
  32. Carlile, T. M., Rojas-Duran, M. F. & Gilbert, W. V. Pseudo-Seq: Genome-Wide Detection of Pseudouridine Modifications in RNA. in *Methods in Enzymology* **560**, 219–245 (Academic Press Inc., 2015).
  33. Lovejoy, A. F., Riordan, D. P. & Brown, P. O. Transcriptome-wide mapping of pseudouridines: Pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One* **9**, (2014).
  34. Li, X. *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **11**, 592–597 (2015).
  35. Dominissini, D. *et al.* The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).
  36. Grozhik, A. V. *et al.* Antibody cross-reactivity accounts for widespread appearance of m1A in 5'UTRs. *Nat. Commun.* **10**, 5126 (2019).
  37. Li, X., Peng, J. & Yi, C. Transcriptome-wide mapping of N1-methyladenosine methylome. in *Methods in Molecular Biology* **1562**, 245–255 (Humana Press Inc., 2017).
  38. Pandolfini, L. *et al.* METTL1 Promotes let-7 MicroRNA Processing via m7G Methylation. *Mol. Cell* (2019). doi:10.1016/J.MOLCEL.2019.03.040
  39. Arango, D. *et al.* Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell* **175**, 1872-1886.e24 (2018).
  40. Kadumuri, R. V. & Janga, S. C. Epitranscriptomic Code and Its Alterations in Human Disease. *Trends Mol. Med.* **24**, 886–903 (2018).
  41. Horowitz, S., Horowitz, A. & Nilsen, T. W. Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 5667–5671 (1984).
  42. Squires, J. E. *et al.* Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* **40**, 5023–5033 (2012).
  43. Levin, J. Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7**, 709–715 (2010).
  44. Schwartz, S. *et al.* High-Resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**, 1409–1421 (2013).
  45. Grozhik, A. V. *et al.* Antibody cross-reactivity accounts for widespread appearance of m1A in 5'UTRs. *Nat. Commun.* **10**, 5126 (2019).
  46. Gillen, A. E., Yamamoto, T. M., Kline, E., Hesselberth, J. R. & Kabos, P. Improvements to the HITS-CLIP protocol eliminate widespread mispriming artifacts. *BMC Genomics* **17**, 1–11 (2016).

47. Legrand, C. *et al.* Statistically robust methylation calling for wholetranscriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res.* **27**, 1589–1596 (2017).
48. Gustafsson, C. & Persson, B. C. Identification of the *rrmA* gene encoding the 23S rRNA m1G745 methyltransferase in *Escherichia coli* and characterization of an m1G745- deficient mutant. *J. Bacteriol.* **180**, 359–365 (1998).
49. Ryvkin, P. *et al.* HAMR: High-throughput annotation of modified ribonucleotides. *RNA* **19**, 1684–1692 (2013).
50. Maden, B. E. H. Mapping 2'-O-methyl groups in ribosomal RNA. in *Methods* **25**, 374–382 (Academic Press Inc., 2001).
51. Dong, Z.-W. *et al.* RTL-P: a sensitive approach for detecting sites of 2'-O-methylation in RNA molecules. *Nucleic Acids Res.* **40**, e157–e157 (2012).
52. Aschenbrenner, J. *et al.* Engineering of a DNA Polymerase for Direct m<sup>6</sup>A Sequencing. *Angew. Chemie Int. Ed.* **57**, 417–421 (2018).
53. Cozen, A. E. *et al.* ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat. Methods* **12**, 879–884 (2015).
54. Zheng, G. *et al.* Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* **12**, 835–837 (2015).
55. Safra, M. *et al.* The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **551**, 251 (2017).
56. Hauenschild, R. *et al.* The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Res.* **43**, 9950–9964 (2015).
57. Gaston, K. W. & Limbach, P. A. The identification and characterization of non-coding and coding RNAs and their modified nucleosides by mass spectrometry. *RNA Biology* **11**, 1568–1585 (2014).
58. Chen, P., Crain, P. F., Näsvall, S. J., Pomerantz, S. C. & Björk, G. R. A 'gain of function' mutation in a protein mediates production of novel modified nucleosides. *EMBO J.* **24**, 1842–1851 (2005).
59. Chan, C. T. Y., Dyavaiah, M., Demott, M. S., Taghizadeh, K. & Dedon, P. C. A Quantitative Systems Approach Reveals Dynamic Control of tRNA Modifications during Cellular Stress. *PLoS Genet* **6**, 1001247 (2010).
60. Quantitative measurement of dihydrouridine in RNA using isotope dilution liquid chromatography-mass spectrometry (LC/MS). Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146067/>. (Accessed: 23rd April 2020)

61. Pomerantz, S. C. & McCloskey, J. A. Analysis of RNA hydrolyzates by liquid chromatography-mass spectrometry. *Methods Enzymol.* **193**, 796–824 (1990).
62. Zhou, S., Sitaramaiah, D., Pomerantz, S. C., Crain, P. F. & McCloskey, J. A. Tandem Mass Spectrometry for Structure Assignments of Wye Nucleosides from Transfer RNA. *Nucleosides, Nucleotides and Nucleic Acids* **23**, 41–50 (2004).
63. Bond, A. *et al.* Analysis of urinary nucleosides. V. Identification of urinary pyrimidine nucleosides by liquid chromatography/electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.* **20**, 137–150 (2006).
64. Kullolli, M., Knouf, E., Arampatzidou, M., Tewari, M. & Pitteri, S. J. Intact MicroRNA analysis using high resolution mass spectrometry. *J. Am. Soc. Mass Spectrom.* **25**, 80–87 (2014).
65. Leger, A. *et al.* RNA modifications detection by comparative Nanopore direct RNA sequencing. *bioRxiv* 843136 (2019). doi:10.1101/843136
66. Smith Id, A. M., Jain, M., Mulrone y Id, L., Garalde, D. R. & Akesonid, M. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. (2019). doi:10.1371/journal.pone.0216709
67. Workman, R. E. *et al.* Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
68. Garalde, D. R. *et al.* Highly parallel direct RN A sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
69. Tuorto, F. *et al.* RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* **19**, 900–905 (2012).
70. Bakin, A. & Ofengand, J. 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).
71. Sprinzl, M., Steegborn, C., Hübel, F. & Steinberg, S. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research* **24**, 68–72 (1996).
72. Pereira, M. *et al.* Impact of tRNA Modifications and tRNA-Modifying Enzymes on Proteostasis and Human Disease. *Int. J. Mol. Sci.* **19**, 3738 (2018).
73. Barbieri, I. & Kouzarides, T. Role of RNA modifications in cancer. *Nat. Rev. Cancer* 1–20 (2020). doi:10.1038/s41568-020-0253-2
74. Björk, G. R. *et al.* A primordial tRNA modification required for the evolution of life? *EMBO J.* **20**, 231–9 (2001).
75. Brulé, H., Elliott, M., Redlak, M., Zehner, Z. E. & Holmes, W. M. Isolation and characterization of the human tRNA-(N1G37) methyltransferase (TRM5) and comparison to the Escherichia coli

- TrmD protein. *Biochemistry* (2004). doi:10.1021/bi049671q
76. Pütz, J., Florentz, C., Benseler, F. & Giegé, R. A single methyl group prevents the mischarging of a tRNA. *Nature Structural Biology* **1**, 580–582 (1994).
  77. Lee, C., Kramer, G., Graham, D. E. & Appling, D. R. Yeast mitochondrial initiator tRNA is methylated at guanosine 37 by the Trm5-encoded tRNA (guanine-N1)-methyltransferase. *J. Biol. Chem.* **282**, 27744–27753 (2007).
  78. Powell, C. A. *et al.* TRMT5 Mutations Cause a Defect in Post-transcriptional Modification of Mitochondrial tRNA Associated with Multiple Respiratory-Chain Deficiencies. *Am. J. Hum. Genet.* **97**, 319–328 (2015).
  79. Camper, S. A., Albers, R. J., Coward, J. K. & Rottmani, F. M. *Effect of Undermethylation on mRNA Cytoplasmic Appearance and Half-Life.* *MOLECULAR AND CELLULAR BIOLOGY* (1984).
  80. Lavi, U., Fernandez-mufioz, R. & Darnell, J. E. *Content of N-6 methyl adenylic acid in heterogeneous nuclear and messenger RNA of HeLa cells.* *Nucleic Acids Research* **4**, (1977).
  81. Wei, C. M., Gershowitz, A. & Moss, B. 5'-Terminal and Internal Methylated Nucleotide Sequences in HeLa Cell mRNA. *Biochemistry* **15**, 397–401 (1976).
  82. Ke, S. *et al.* A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev.* **29**, 2037–53 (2015).
  83. Liu, N. *et al.* Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19**, 1848–56 (2013).
  84. Edupuganti, R. R. *et al.* N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* **24**, 870–878 (2017).
  85. Wang, X. *et al.* N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388–1399 (2015).
  86. Shi, H. *et al.* YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Res.* **27**, 315–328 (2017).
  87. Du, H. *et al.* YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat. Commun.* **7**, 1–11 (2016).
  88. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* **61**, 507–519 (2016).
  89. Roundtree, I. A. *et al.* YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *Elife* **6**, (2017).
  90. Shima, H. *et al.* S-Adenosylmethionine Synthesis Is Regulated by Selective N6-Adenosine Methylation and mRNA Degradation Involving METTL16 and YTHDC1. *Cell Rep.* **21**, 3354–3363 (2017).

91. Hsu, P. J. *et al.* Ythdc2 is an N6 -methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Res.* **27**, 1115–1127 (2017).
92. Yang, X. *et al.* 5-methylcytosine promotes mRNA export — NSUN2 as the methyltransferase and ALYREF as an m5C reader. *nature.com* (2017).
93. Huber, S. M. *et al.* Formation and abundance of 5-hydroxymethylcytosine in RNA. *ChemBioChem* **16**, 752–755 (2015).
94. Nishikura, K. Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu. Rev. Biochem.* **79**, 321–349 (2010).
95. Furuichi, Y., LaFiandra, A. & Shatkin, A. J. 5'-Terminal structure and mRNA stability. *Nature* **266**, 235–239 (1977).
96. Lin, S. *et al.* Mettl1/Wdr4-Mediated m<sup>7</sup>G tRNA Methylome Is Required for Normal mRNA Translation and Embryonic Stem Cell Self-Renewal and Differentiation. *Mol. Cell* **71**, 244-255.e5 (2018).
97. Zhang, L. S. *et al.* Transcriptome-wide Mapping of Internal N7-Methylguanosine Methylome in Mammalian mRNA. *Mol. Cell* **74**, 1304-1316.e8 (2019).
98. Stern, L. & Schulman, L. H. *The Role of the Minor Base IV4-Acetylcytidine in the Function of the Escherichia coli Noninitiator Methionine Transfer RNA\**.
99. Sendinc, E. *et al.* PCIF1 Catalyzes m<sup>6</sup>A mRNA Methylation to Regulate Gene Expression. *Mol. Cell* **75**, 620-630.e9 (2019).
100. Hess, M. E. *et al.* The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry. *Nat. Neurosci.* **16**, 1042–1048 (2013).
101. Grozhik, A. V. & Jaffrey, S. R. Epitranscriptomics: Shrinking maps of RNA modifications. *Nature* **551**, 174 (2017).
102. Dominissini, D. *et al.* The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).
103. Gray, M. C. M. W. Pseudouridine in RNA: What, Where, How, and Why. *IUBMB Life (International Union Biochem. Mol. Biol. Life)* **49**, 341–351 (2000).
104. Stabilization of RNA stacking by pseudouridine | Nucleic Acids Research | Oxford Academic. Available at: <https://academic.oup.com/nar/article-abstract/23/24/5020/2400691>. (Accessed: 18th May 2020)
105. Liu, B. *et al.* A potentially abundant junctional RNA motif stabilized by m<sup>6</sup>A and Mg<sup>2+</sup>. *Nat. Commun.* **9**, 1–10 (2018).
106. Roost, C. *et al.* Structure and thermodynamics of N6-methyladenosine in RNA: A spring-loaded base modification. *J. Am. Chem. Soc.* **137**, 2107–2115 (2015).

107. Base pairing and structural insights into the 5-formylcytosine in RNA duplex | Nucleic Acids Research | Oxford Academic. Available at: <https://academic.oup.com/nar/article/44/10/4968/2516588>. (Accessed: 18th May 2020)
108. Zhou, H. *et al.* M1A and m1G disrupt A-RNA structure through the intrinsic instability of Hoogsteen base pairs. *Nat. Struct. Mol. Biol.* **23**, 803–810 (2016).
109. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).
110. Lin, S., Choe, J., Du, P., Triboulet, R. & Gregory, R. I. The m6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol. Cell* **62**, 335–345 (2016).
111. Wang, X. *et al.* N6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**, 1388–1399 (2015).
112. Choi, J. *et al.* N6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. *Nat. Struct. Mol. Biol.* **23**, 110–115 (2016).
113. Jonkhout, N. *et al.* The RNA modification landscape in human disease. *RNA* **23**, 1754–1769 (2017).
114. Paul, M. S. & Bass, B. L. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA | The EMBO Journal. *EMBO J* **17**: 1120–1127 (1998). Available at: <https://www.embopress.org/doi/full/10.1093/emboj/17.4.1120>. (Accessed: 18th May 2020)
115. Chi, L. & Delgado-Olguín, P. Expression of NOL1/NOP2/sun domain (Nsun) RNA methyltransferase family genes in early mouse embryogenesis. *Gene Expr. Patterns* **13**, 319–327 (2013).
116. Yi, J. *et al.* Overexpression of NSUN2 by DNA hypomethylation is associated with metastatic progression in human breast cancer. *Oncotarget* **8**, 20751–20765 (2017).
117. Endres, L., Fasullo, M. & Rose, R. tRNA modification and cancer: potential for therapeutic prevention and intervention. *Future Med. Chem.* fmc-2018-0404 (2019). doi:10.4155/fmc-2018-0404
118. Vu, L. P. *et al.* The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat. Med.* **23**, 1369–1376 (2017).
119. Lin, X. *et al.* RNA m6A methylation regulates the epithelial mesenchymal transition of cancer cells and translation of Snail. *Nat. Commun.* **10**, 1–13 (2019).
120. Barbieri, I. *et al.* Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. *Nature* **552**, 126 (2017).
121. Liu, J. *et al.* m6A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. *Nat. Cell Biol.* **20**, 1074–1083 (2018).

122. Yang, S. *et al.* m6A mRNA demethylase FTO regulates melanoma tumorigenicity and response to anti-PD-1 blockade. *Nat. Commun.* **10**, 1–14 (2019).
123. Yanyan Ping†, Yulan Deng†, Li Wang, Hongyi Zhang, Yong Zhang, Chaohan Xu, Hongying Zhao, Huihui Fan, Fulong Yu, Y. X. and X. L. Identifying core gene modules in glioblastoma based on multilayer factor-mediated dysfunctional regulatory networks through integrating multi-dimensional genomic data | *Nucleic Acids Research* | Oxford Academic. Available at: <https://academic.oup.com/nar/article/43/4/1997/2411558>. (Accessed: 19th May 2020)
124. Cui, Q. *et al.* m6A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. *Cell Rep.* **18**, 2622–2634 (2017).
125. Jonkhout, N. *et al.* The RNA modification landscape in human disease. *RNA* **23**, 1754–1769 (2017).
126. Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature* **1** (2019). doi:10.1038/s41586-019-1103-9
127. Schapira, M. Structural Chemistry of Human RNA Methyltransferases. *ACS Chemical Biology* **11**, 575–582 (2016).
128. Stein, E. M. *et al.* The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia. *Blood* **131**, 2662–2669 (2018).
129. Bedi, R. K. *et al.* Small-Molecule Inhibitors of METTL3, the Major Human Epitranscriptomic Writer. *ChemMedChem* **15**, 744–748 (2020).
130. Wang, R. *et al.* Identification of Natural Compound Radicol as a Potent FTO Inhibitor. *Mol. Pharm.* **15**, 4092–4098 (2018).
131. Rapino, F., Delaunay, S., Zhou, Z., Chariot, A. & Close, P. tRNA Modification: Is Cancer Having a Wobble? *Trends in Cancer* **3**, 249–252 (2017).
132. Rapino, F. *et al.* Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature* **558**, 605–609 (2018).
133. Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **6**, pl1–pl1 (2013).
134. Okamoto, M. *et al.* tRNA Modifying Enzymes, NSUN2 and METTL1, Determine Sensitivity to 5-Fluorouracil in HeLa Cells. *PLoS Genet.* **10**, e1004639 (2014).
135. Gokhale, N. S. & Horner, S. M. RNA modifications go viral. *PLOS Pathog.* **13**, e1006188 (2017).
136. Stoltzfus, C. M. & Dane, R. W. Accumulation of Spliced Avian Retrovirus mRNA Is Inhibited in S-Adenosylmethionine-Depleted Chicken Embryo Fibroblasts. *J. Virol.* **42**, (1982).
137. Lichinchi, G. *et al.* Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. *Cell Host Microbe* **20**, 666–673 (2016).

138. Lichinchi, G. *et al.* Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells. *Nat. Microbiol.* **1**, 16011 (2016).
139. Gordon, D. E. *et al.* A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. *bioRxiv* 2020.03.22.002386 (2020). doi:10.1101/2020.03.22.002386
140. van Delft, P. *et al.* The Profile and Dynamics of RNA Modifications in Animals. *ChemBioChem* **18**, 979–984 (2017).
141. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods* **11**, 783–784 (2014).
142. Guzmán, C., Bagga, M., Kaur, A., Westermarck, J. & Abankwa, D. ColonyArea: An ImageJ Plugin to Automatically Quantify Colony Formation in Clonogenic Assays. *PLoS One* **9**, e92444 (2014).
143. Gillotin, S., Davies, J. D. & Philpott, A. Subcellular localisation modulates ubiquitylation and degradation of Ascl1. *Sci. Rep.* **8**, 1–13 (2018).
144. Van Haute, L. *et al.* METTL15 introduces N4-methylcytidine into human mitochondrial 12S rRNA and is required for mitoribosome biogenesis. *Nucleic Acids Res.* **47**, 10267–10281 (2019).
145. Jackman, J. E., Montange, R. K., Malik, H. S. & Phizicky, E. M. Identification of the yeast gene encoding the tRNA m1G methyltransferase responsible for modification at position 9. *RNA* **9**, 574–585 (2003).
146. Gillis, D. *et al.* TRMT10A dysfunction is associated with abnormalities in glucose homeostasis, short stature and microcephaly. *J. Med. Genet.* **51**, 581–586 (2014).
147. Yew, T. W., McCreight, L., Colclough, K., Ellard, S. & Pearson, E. R. tRNA methyltransferase homologue gene TRMT10A mutation in young adult-onset diabetes with intellectual disability, microcephaly and epilepsy. *Diabet. Med.* **33**, e21–e25 (2016).
148. Cosentino, C. *et al.* Pancreatic  $\beta$ -cell tRNA hypomethylation and fragmentation link TRMT10A deficiency with diabetes. *Nucleic Acids Res.* **46**, 10302–10318 (2018).
149. Elisa Vilardo, Fabian Amman, Ursula Toth, Annika Kotter, M. H. and W. R. Functional Characterization of the Human tRNA Methyltransferases TRMT10A and TRMT10B - PubMed. Available at: <https://pubmed.ncbi.nlm.nih.gov/32392304/>. (Accessed: 16th June 2020)
150. Clark, W. C., Evans, M. E., Dominissini, D., Zheng, G. & Pan, T. tRNA base methylation identification and quantification via high-throughput sequencing. *RNA* **22**, 1771–1784 (2016).
151. Li, X. *et al.* Base-Resolution Mapping Reveals Distinct m1A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Mol. Cell* **68**, 993-1005.e9 (2017).
152. Christian, T., Gamper, H. & Hou, Y. M. Conservation of structure and mechanism by Trm5 enzymes. *RNA* **19**, 1192–1199 (2013).

153. Zhou, M. *et al.* A hypertension-associated mitochondrial DNA mutation introduces an m1G37 modification into tRNAMet, altering its structure and function. *J. Biol. Chem.* **293**, 1425–1438 (2018).
154. Yao, W., Maitra, A. & Ying, H. Recent insights into the biology of pancreatic cancer. *EBioMedicine* **53**, 102655 (2020).
155. Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015).
156. Tu, B. *et al.* YAP1 oncogene is a context-specific driver for pancreatic ductal adenocarcinoma. *JCI Insight* **4**, (2019).
157. Heterogeneous Stromal Signaling within the Tumor Microenvironment Controls the Metastasis of Pancreatic Cancer | Cancer Research. Available at: <https://cancerres.aacrjournals.org/content/77/1/41.full-text.pdf>. (Accessed: 17th June 2020)
158. Hezel, A. F., Kimmelman, A. C., Stanger, B. Z., Bardeesy, N. & DePinho, R. A. Genetics and biology of pancreatic ductal adenocarcinoma. *Genes and Development* **20**, 1218–1249 (2006).
159. Canon, J. *et al.* The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **575**, 217–223 (2019).
160. Witkiewicz, A. K. *et al.* Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat. Commun.* **6**, 1–11 (2015).
161. Sun, J. *et al.* A systematic analysis of FDA-approved anticancer drugs. *BMC Syst. Biol.* **11**, (2017).
162. Matsuda, Y. *et al.* Nestin is a novel target for suppressing pancreatic cancer cell migration, invasion and metastasis. *Cancer Biol. Ther.* **11**, 512–523 (2011).
163. Björk, G. R. Genetic Dissection of Synthesis and Function of Modified Nucleosides in Bacterial Transfer RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **50**, 263–338 (1995).
164. Allnér, O. & Nilsson, L. Nucleotide modifications and tRNA anticodon-mRNA codon interactions on the ribosome. *RNA* **17**, 2177–2188 (2011).
165. Helm, M. & Alfonzo, J. D. Posttranscriptional RNA modifications: Playing metabolic games in a cell's chemical legoland. *Chemistry and Biology* **21**, 174–185 (2014).
166. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
167. Zhang, C. & Jia, G. Reversible RNA Modification N1-methyladenosine (m1A) in mRNA and tRNA. *Genomics. Proteomics Bioinformatics* **16**, 155–161 (2018).
168. Howell, N. W., Jora, M., Jepson, B. F., Limbach, P. A. & Jackman, J. E. Distinct substrate specificities of the human tRNA methyltransferases TRMT10A and TRMT10B. *RNA* rna.072090.119 (2019). doi:10.1261/rna.072090.119

169. Mauer, J. & Jaffrey, S. R. FTO, m<sup>6</sup>A<sub>m</sub>, and the hypothesis of reversible epitranscriptomic mRNA modifications. *FEBS Lett.* **592**, 2012–2022 (2018).
170. Helm, M. & Alfonzo, J. D. Posttranscriptional RNA modifications: Playing metabolic games in a cell's chemical legoland. *Chemistry and Biology* **21**, 174–185 (2014).
171. Guo, Q. *et al.* Arabidopsis TRM5 encodes a nuclear-localised bifunctional tRNA guanine and inosine-N1-methyltransferase that is important for growth. (2019). doi:10.1371/journal.pone.0225064
172. Kawamoto, M. *et al.* Nestin expression correlates with nerve and retroperitoneal tissue invasion in pancreatic cancer. *Hum. Pathol.* **40**, 189–198 (2009).
173. Cho, S.-H. *et al.* Direct determination of nucleosides in the urine of patients with breast cancer using column-switching liquid chromatography–tandem mass spectrometry. *Biomed. Chromatogr.* **20**, 1229–1236 (2006).
174. Phillips, J. H. & Kjellin-Stråby, K. Studies on microbial ribonucleic acid: Two mutants of *saccharomyces cerevisiae* lacking N2-dimethylguanine in soluble ribonucleic acid. *J. Mol. Biol.* **26**, 509–518 (1967).
175. Martin, N. C. & Hopper, A. K. How single genes provide tRNA processing enzymes to mitochondria, nuclei and the cytosol. *Biochimie* **76**, 1161–1167 (1994).
176. Copela, L. A., Chakshusmathi, G., Sherrer, R. L. & Wolin, S. L. The La protein functions redundantly with tRNA modification enzymes to ensure tRNA structural stability. *RNA* **12**, 644–654 (2006).
177. Pallan, P. S., Kreutz, C., Bosio, S., Micura, R. & Egli, M. Effects of N2,N2-dimethylguanosine on RNA structure and stability: Crystal structure of an RNA duplex with tandem m22G:A pairs. *RNA* **14**, 2125–2135 (2008).
178. Towns, W. L. & Begley, T. J. Transfer RNA methyltransferases and their corresponding modifications in budding yeast and humans: activities, predications, and potential roles in human health. *DNA Cell Biol.* **31**, 434–54 (2012).
179. The human tRNA(m22G26)dimethyltransferase: functional expression and characterization of a cloned hTRM1 gene. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC110725/>. (Accessed: 25th March 2020)
180. Davarniya, B. *et al.* The role of a novel TRMT1 gene mutation and rare GRM1 gene defect in intellectual disability in two azeri families. *PLoS One* **10**, (2015).
181. Vauti, F. *et al.* The mouse Trm1-like gene is expressed in neural tissues and plays a role in motor coordination and exploratory behaviour. *Gene* **389**, 174–185 (2007).
182. Dewe, J. M., Fuller, B. L., Lentini, J. M., Kellner, S. M. & Fu, D. TRMT1-Catalyzed tRNA

- Modifications Are Required for Redox Homeostasis To Ensure Proper Cellular Proliferation and Oxidative Stress Survival. *Mol. Cell. Biol.* **37**, (2017).
183. Simms, C. L. & Zaher, H. S. Quality control of chemically damaged RNA. *Cellular and Molecular Life Sciences* **73**, 3639–3653 (2016).
184. Jordan Ontiveros, R. *et al.* Coordination of mRNA and tRNA methylations by TRMT10A. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7782–7791 (2020).