



**Australian
National
University**

Investigating the enzymology and function of 5-methylcytosine RNA modification

A thesis submitted for the degree of Doctor of Philosophy of The
Australian National University

Marco Guarnacci

September 2023

© Copyright by Marco Guarnacci 2023

All Rights Reserved

Originality statement

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at ANU or any other educational institution, except where due acknowledgement is made in this thesis. Any contribution made to the research by others, with whom I have worked at ANU or elsewhere, is explicitly acknowledged in this thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project design and conception or in style, presentation and linguistic expression is acknowledged.

Abstract

5-methylcytosine (m⁵C) is a widespread RNA modification decorating diverse species of RNA. The re-discovery of m⁵C as an mRNA modification came to light by coupling bisulfite conversion of RNA with transcriptome-wide RNA sequencing (bsRNA-seq). Subsequent studies revealed prominence of deposition of m⁵C modification in regulating gene expression. Presently, bsRNA-seq experiments carried out in different cell types have generated several datasets for m⁵C-decorated mRNAs that are often different from each other. This disparity could be explained by differential methylation patterns in different cell types, the variation in the expression levels of the writer enzymes (NSUN family of proteins) and differences in data analysis pipelines. In an attempt to address this variability, we generated a comprehensive, transcriptome-wide human dataset of m⁵C sites by re-analysing existing bsRNA-seq datasets from 5 different human cell lines and 7 tissues, with a stringent pipeline for m⁵C site calling. This allowed us to compile a union list of 6,393 high-confidence m⁵C sites and to gain insights into the enzymology of different NSUN proteins (m⁵C writers). We also employed our union set to investigate the relationship between m⁵C modification and RNA-binding proteins (RBPs) binding to mRNA. Therefore, we overlapped mRNA m⁵C sites with publicly available enhanced Crosslinking Immunoprecipitation (eCLIP) data for >100 RBPs and identified several proteins whose footprint significantly overlaps with m⁵C modification, including AKAP1, DDX3X, DHX30, FTO, NCBP2, RPS3 and UPF1. We further generated a knockout HeLa cell line for NSUN2, the main m⁵C writer protein for mRNA, to investigate the effect of mRNA methylation level on the function of UPF1 RBP. By performing UPF1 knockdown experiments, in WT vs NSUN2 knockout HeLa cells, we showed that the lack of NSUN2 enzyme reduces the effect of UPF1 in regulating the steady state of a subgroup of its methylated mRNA targets. CLIP experiments also showed a differential binding of UPF1 RBP depending on its target methylation content. Overall, our data indicate an interplay between NSUN2 RNA methyltransferase activity and UPF1 function on mRNA.

Table of Contents

Originality statement.....	2
Abstract.....	3
1. Introduction	6
1.1 5-methylcytosine (m⁵C) is widespread RNA modification	7
1.1.1 m ⁵ C in transfer RNA	8
1.1.2 m ⁵ C in ribosomal RNA	10
1.1.3 m ⁵ C in other non-coding RNAs.....	11
1.1.4 m ⁵ C in messenger RNA.....	13
1.2 The protein players of m⁵C modification	15
1.2.1 m ⁵ C writer proteins.....	15
1.2.2 m ⁵ C eraser proteins.....	19
1.2.3 m ⁵ C reader proteins	20
1.3 The function of m⁵C.....	23
1.3.1 The role of m ⁵ C in molecular function	23
1.3.2 The role of m ⁵ C in cellular function.....	26
1.3.3 The role of m ⁵ C in cancer and immunotherapy	28
2. Aims of the thesis	31
3. Materials and Methods	32
3.1 Quality control and mapping of RNA BS-seq reads	32
3.2 m ⁵ C site calling and annotation for each dataset	32
3.3 Distribution and enrichment of m ⁵ C along mRNA.....	33
3.4 NSUNs dependent m ⁵ C sites.....	33
3.5 Comparison with RBP binding sites	33
3.6 Cell culture and transfection	34
3.7 Generation of KO cells.....	34
3.7.1 Guide sequence design	34
3.7.2 Ribonuclear protein preparation.....	35
3.7.3 Transfection	35
3.7.4 Single cell sorting.....	35
3.7.5 Sequencing	36
3.8 Protein Analysis	38
3.9 RNA Analysis	39
3.10 Bisulfite conversion and amplicon-bsRNA sequencing	39
3.11 miSeq analysis.....	41
3.12 UPF1 knockdown RNA-seq and differential expression analyses.....	41
3.14 Cross-linking immunoprecipitation.....	41
4. Results	43
4.1 Generation of a union list of m⁵C sites in human transcriptomes.....	43
4.1.1 Stringent and uniform re-analysis of bsRNA-Seq data	44

4.1.2 A union set of 6,393 human m ⁵ C sites	48
4.1.3 m ⁵ C distribution in the transcriptome	50
4.2 The spectrum of mRNA m⁵C writer enzymes	52
4.2.1 NSUN2- and NSUN6-dependent m ⁵ C sites	52
4.2.2 NSUN5 mediates mRNA methylation.....	54
4.2.3 Prediction and validation of NSUN5-dependent physiological mRNA sites	57
4.3 Multiple RBPs are enriched around m⁵C	60
4.3.1 Multiple RBPs are enriched around m ⁵ C	60
4.3.2 The function of UPF1 RBP is affected by the lack of NSUN2 methyltransferase.....	67
4.3.3 UPF1 binding of m ⁵ C-mRNAs is regulated by NSUN2 methyltransferase	77
5. Discussion	80
Abbreviation Table	86
6. References.....	88

1. Introduction

Among all the different layers of gene expression regulation, post-synthesis biochemical modification is a powerful and precise mechanism for regulating the functionality of macromolecules. This regulatory approach extends across various biological molecules, including DNA (Moosavi & Ardekani, 2016), RNA (Barbieri & Kouzarides, 2020) and proteins (Ramazi & Zahiri, 2021). Among these, RNA modifications have emerged as a pivotal focus within the intricate realm of gene expression regulation, cellular responses to environmental cues, and even the development of diseases and cancer (Cayir et al., 2020; Y. S. Chen et al., 2021; Roundtree et al., 2017; H. Song et al., 2022; Wilkinson et al., 2022). In this context, the revelation of RNA modifications marks a transformative era in RNA biology.

This transformational shift in our understanding of RNA biology and the tightly regulated process for its modification, has further consolidated the vital role of RNA.

RNA is a remarkably dynamic and functionally diverse molecule, featuring a multitude of post-transcriptional modifications. Current insights reveal that all four RNA bases are susceptible to modification, culminating in a repertoire of over 150 documented post-transcriptional biochemical modifications (Barbieri & Kouzarides, 2020). These modifications are not limited to a single class of RNA; instead, they span all known RNA species, with transfer RNA (tRNA) and ribosomal RNA (rRNA) being the most heavily modified (S. Li & Mason, 2014). Historically, the first glimpse of modified nucleotides within cellular RNAs emerged in the 1960s (Cohn, 1960). This pioneering observation paved the way for the seminal sequencing of biological RNA shortly thereafter, unveiling ten distinct modifications, including the now well-characterized pseudouridine (Holley et al., 1965). Fast forward to recent years, our understanding has dramatically expanded. The intricacy of the various modifications that an RNA molecule can carry is reflected in the complex regulation of RNA's biogenesis, structure, localization, stability, and function (Frye et al., 2018; Gilbert & Nachtergaele, 2023; Schaefer et al., 2017). As we delve deeper into the intricate world of epitranscriptomics, we uncover a fascinating landscape of chemical marks that dictate RNA's functions and its ability to finely tune the intricate orchestra of cellular biology.

Epitranscriptomics, a relatively recent field dedicated to investigating biochemical RNA modifications, has seen a rapid growth over the past decade. At present, ~170 RNA modifications have been identified (Boccaletto et al., 2022; Cantara et al., 2011). However, only a fraction of these modifications has been further studied and characterized (Linder &

Jaffrey, 2019). N⁶-methyladenosine and 5-methylcytosine are among those that have gained widespread recognition for their roles in regulating cellular mRNA fate (Murakami & Jaffrey, 2022; Trixl & Lusser, 2019; Wiener & Schwartz, 2021) and their impact on organismic development and disease (P. C. He & He, 2021; Motorin & Helm, 2022).

Despite its widespread presence across various RNA species, 5-methylcytosine modification remains relatively understudied. Fuelled by its intriguing yet relatively understudied nature, m⁵C was chosen as the focal point of this research.

1.1 5-methylcytosine (m⁵C) is widespread RNA modification

5-methylcytosine refers to a biochemical modification characterised by the addition of a methyl group to the carbon-5 position of cytidine residues (Johnson & Coghill, 1925). While the discovery of m⁵C as a DNA modification predates the elucidation of the double helix structure (Wyatt, 1950), it was only in the 1970s that researchers discovered its presence in RNA (Dubin & Taylor, 1975). Decades of research have now revealed that m⁵C is a widespread RNA modification found in all known RNA species, including coding (messenger RNA) and non-coding RNA, such as tRNA, rRNA, long non-coding RNA (lncRNA), and enhancer RNA (eRNA) (Boccalletto et al., 2022; Cantara et al., 2011).

While m⁵C modification of RNA has been reported in many species, including *Mus musculus*, *Danio rerio*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae* and *Caenorhabditis elegans* (J. Ma et al., 2022), its distribution can vary substantially. For example, tRNA and rRNA have more m⁵C modification compared to their bacterial counterparts (Cantara et al., 2011). The methylation of tRNA and rRNA has been the subject of extensive research, yielding valuable insights into their roles. In the case of tRNA, studies have revealed its pivotal role for m⁵C in a number of important functions, such as improving codon-anticodon pairing (Goll et al., 2006; Shanmugam et al., 2015), exerting control over stress responses (Chan et al., 2012) and finely regulating both translation accuracy and efficiency (Burgess et al., 2015; Goll et al., 2006; Shanmugam et al., 2015). Similarly, the absence of m⁵C modification in rRNA was demonstrated to result in a global reduction in translation in glioma, triggering the activation of an adaptive translational program, that is essential for the cells' survival under stress (Janin et al., 2019). Additionally, the absence of methylation renders these gliomas sensitive to bioactivatable substrates of the stress-related

enzyme NQO1, further highlighting the critical role of m⁵C modification in cellular stress responses (Janin et al., 2019).

Starting from the 1970s, when the presence of m⁵C modification was first reported in polyadenylated RNA (Dubin & Taylor, 1975), a substantial body of research has evolved. This research has delved into the regulation of various biological processes influenced by this modification in both mRNA and lncRNA, including mRNA stability (X. Chen et al., 2019; Y. Yang et al., 2019), nucleus-cytoplasm export (X. Yang et al., 2017a), cell proliferation and migration (S. Xue et al., 2019), DNA damage control (H. Chen et al., 2020) and stem cells development (Zou et al., 2020). Importantly, deregulation of m⁵C deposition in mRNA was associated with pathological conditions, such as bladder cancer (X. Chen et al., 2019), systemic lupus erythematosus (SLE) (Guo et al., 2020) and vascular inflammation and arteriosclerosis (Luo et al., 2016).

1.1.1 m⁵C in transfer RNA

Among all known RNA species, tRNA is the one that is most heavily modified, with an average eukaryotic tRNA containing 13 modifications per molecule (Roundtree et al., 2017). Historically, the bulk of research in m⁵C in RNA was prevalently focused on tRNA, highlighting its crucial role through demonstrating the necessity of proper methylation for its optimal processing and function (Brzezicha et al., 2006; Y. S. Chen et al., 2021; Goll et al., 2006; Long et al., 2016; Reid et al., 1999).

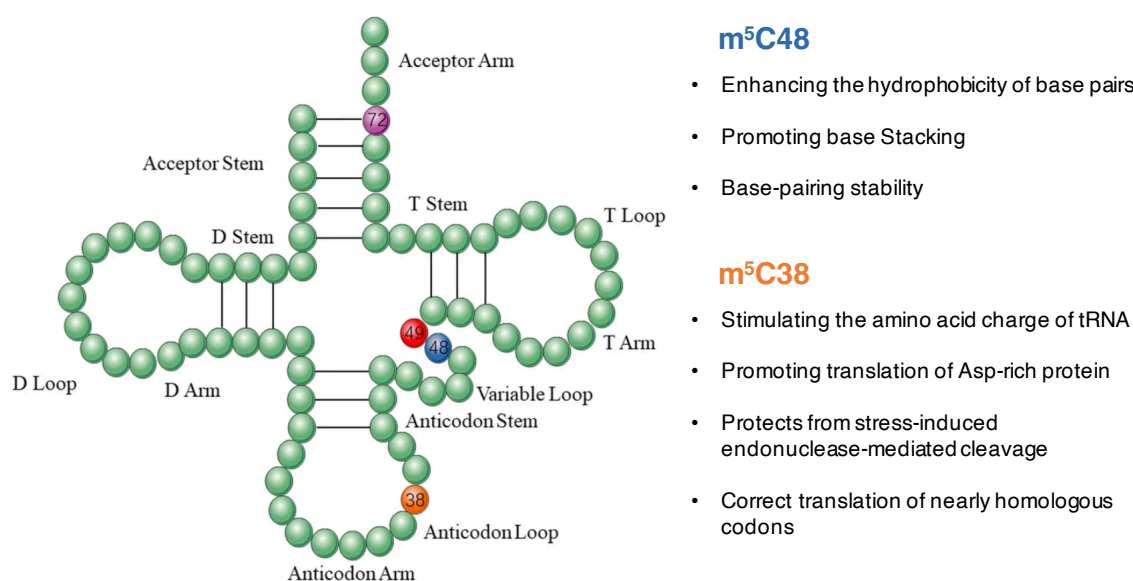


Figure 1 - Schematic summary of location and function of tRNA m⁵C sites. Adapted from (H. Song et al., 2022).

m⁵C modification of tRNA predominantly occurs at specific cytosine positions, such as positions spanning C:47 to C:50 (Reid et al., 1999). In particular, m⁵C48 and m⁵C49 modifications are strategically positioned at the junction sequence of the V-loop (VL) and T-stem-loop (TSL). These modifications contribute significantly to tRNA secondary structure stability, influencing its ability to recognize codons and promote efficient protein translation (Reid et al., 1999; Van Haute et al., 2019). The impact of tRNA methylation on protein synthesis is particularly noteworthy (L. Xu et al., 2017). The methylation of C:48 serves to enhance hydrophobic interactions between base pairs, promoting the stacking of bases and reinforcing the structural stability of tRNA (Brzezicha et al., 2006). Notably, the same C:48 is the one base-pairing with the nucleotide in position 15 on the D ring, also known as a Levitt pair, responsible for the formation of a distinct L-shaped three-dimensional structure (Brzezicha et al., 2006).

Another frequently methylated site in tRNA is C:38, within the anticodon loop. Methylation of C:38 in mouse tRNA^{Asp} plays a pivotal role in stimulating the charging of tRNA with L-aspartate, both *in vivo* and *in vitro*. This stimulation results in enhanced translation efficiency of proteins harbouring multiple aspartic acid residues (Goll et al., 2006). Moreover, the methylation of C:38 serves as a protective mechanism for tRNAs, guarding them against stress-induced endonuclease-mediated cleavage and ensuring the precise translation of nearly homologous codons (Goll et al., 2006). The absence of m⁵C38 in tRNA^{Met} leads to a moderated level of translational fidelity during the decoding of aspartate codons (Y. Zhang et al., 2018). Additionally, DNMT2-mediated m⁵C38 tRNA^{Asp} plays a pivotal role in distinguishing homologous and nearly homologous codons, such as tRNA^{Asp} and tRNA^{Glu}, thereby enhancing translation accuracy (Burgess et al., 2015).

Research centred on mitochondria has unveiled that mitochondrial 5-formylcytidine (f⁵C) modification originates from m⁵C34 of mt-tRNA^{met}. This modification plays a critical role in mitochondrial translation, particularly in the decoding of unconventional AUA codons (Van Haute et al., 2017). In yeast, the m⁵C modification of C:34 in tRNA^{Leu} was shown to be closely linked to translation regulation. Its methylation levels significantly increase during oxidative stress responses, leading to enhanced translation of mRNAs rich in UUG codons, including those coding for the ribosomal protein El22A (Brzezicha et al., 2006).

Furthermore, during the late stages of tRNA^{Thr} and tRNA^{Cys} transcription, a CCA sequence is added at the 3' end, followed by confirmed methylation at C:72. Despite this well-established process, the precise function of m⁵C:72 modification remains unknown (Long et al., 2016).

1.1.2 m⁵C in ribosomal RNA

Ribosomes serve as intricate molecular machinery responsible for the synthesis of all cellular proteins, a tightly controlled biochemical process known as translation. In eukaryotic cells, the 80S ribosome is composed of two main components: a smaller 40S subunit and a larger 60S subunit (Baßler & Hurt, 2019). The 40S subunit contains the 18S ribosomal RNA (rRNA) and 33 distinct ribosomal proteins (r-proteins). Meanwhile, the 60S subunit comprises 25S, 5.8S, and 5S rRNA, along with 47 r-proteins (Baßler & Hurt, 2019). Ribosomal RNA in the nucleus, cytoplasm and mitochondria can all be decorated with m⁵C modification (Gao & Fang, 2021). This modification is known to play a significant role in regulating rRNA stability, consequently influencing protein synthesis (H. Song et al., 2022).

A cluster of methylation in the fourth domain of the 25S rRNA was found to be necessary for the correct folding and stabilisation of the 60S subunit of the ribosome in Yeast (Sharma et al., 2013). The methylation of the same position in 25S rRNA in yeast was found to be important during oxidative stress conditions for the selective recruitment and subsequent translation of specific mRNAs (Sharma et al., 2013). Consequently, deletion of the m⁵C modification in position C:2278 resulted in the impaired stability of the 60S subunit and reduced translation efficiency (Sharma et al., 2013). In addition, the methylation of C:2278 is further suggested to be important for ribosome function, as the deletion of Rcm1, the methyltransferase responsible for the deposition of such modification, was found to lead to the incorrect reading of premature stop codons in reporter assays (Schosserer et al., 2015). Once mature, m⁵C:2278 makes direct contact with ribosomal protein eL41 (RPL41), serving as a pivotal point for facilitating small subunit (SSU) rotation during the translation process (Sharma & Lafontaine, 2015). This suggests a mechanistic role for m⁵C modification in regulating translation.

In human, 28S ribosomal RNA is characterised by the m⁵C modification in position C:3782 (Janin et al., 2019). Lack of methylation on this residue was found to drive an overall depletion of protein synthesis which leads to the activation of an adaptive translational program for survival under conditions of cellular stress (Janin et al., 2019). 28S rRNA also contains m⁵C modifications at positions C:3761 and C:4413, which were reported to be necessary for rRNA correct folding and translation fidelity (Arzumanian et al., 2022; Schosserer et al., 2015).

In mitochondria, the deposition of m⁵C modification on position C:911 of 12S rRNA was found to be important for the correct assembly of mitochondrial ribosomes. Disruption of m⁵C911 level was observed to result in loss of mitochondrial protein synthesis (Metodieiev et al., 2014).

1.1.3 m⁵C in other non-coding RNAs

In addition to tRNA and rRNA, other classes of non-coding RNAs (ncRNAs) were reported to be modified with m⁵C modification (Cusenza et al., 2023; Del Valle-Morales et al., 2022; Squires et al., 2012).

Vault RNAs (vtRNAs) are a class of short non-coding RNAs found in eukaryotic cells, of about 100 nucleotides in length (Stadler et al., 2009). While the exact function of this type of ncRNA has been poorly studied, vtRNAs have recently been reported to have a role in cell differentiation and development (Hussain, Sajini, et al., 2013; Sajini et al., 2019).

m⁵C deposition on position 69 of the vtRNA1.1 was found to determine its transition into small-vault RNAs, which have a role in coordinating epidermal differentiation (Sajini et al., 2019). Interestingly, it was reported that the absence of m⁵C methylation in vault RNA leads to an abnormal processing of the RNA into Argonaute-associated regulatory small RNA fragments and will function similarly to microRNAs (Hussain, Sajini, et al., 2013).

The discovery of m⁵C modification on microRNAs (miRNAs) is also fascinating, and sheds light on an additional layer of regulation in gene expression. For instance, the methylation of pri-miR-125b2, as observed in HeLa cells, has been found to have important functional implications (S. Yuan et al., 2014). In this context, methylation of pri-miR-125b2 inhibits its normal processing into pre-miR-125b2 and, subsequently, the cleavage into mature miR-125. This disruption in the processing pathway results in the attenuation of the recruitment of the RNA-Induced Silencing Complex (RISC), which is essential for the function of miRNAs in gene silencing (S. Yuan et al., 2014). As a consequence, the methylation of miRNA can lead to the repression of its function in silencing gene expression, highlighting the dynamic and intricate regulatory mechanisms that control gene expression at the post-transcriptional level.

The discovery of m⁵C modification on the mature sequences of miRNAs miR-21-3p and miR-200c-3p (Konno et al., 2019) presents insights into the role of epitranscriptomic modifications in cancer biology. The methylation levels of miRNAs miR-21-3p and miR-200c-3p were found to be increased in pancreatic and colorectal cancer tissues when compared to normal samples, despite miRNAs' expression level remained unaltered. In addition, these miRNAs m⁵C levels were higher in serum samples from pancreatic and colorectal cancer patients than

in those from normal subjects (Konno et al., 2019), indicating a potential role for these modifications as biomarkers for cancer detection.

miR-181a-5p, a miRNA with a known role as a tumour suppressor, was also found to carry a m⁵C modification (Cheray et al., 2020). However, the removal of m⁵C from miR-181a-5p results in the loss of its tumour-suppressing capabilities (Y. Li et al., 2015). miR-181a-5p's modification was found to inhibit the formation of miRNA/mRNA complex, therefore impeding the miRNA's ability to bind its target mRNAs. Interestingly, miR-181a-5p is often found downregulated in glioblastoma (Brower et al., 2014), and its methylation is a hallmark for poor prognosis in glioblastoma patients (Cheray et al., 2020).

These findings add a new layer of complexity to our understanding of how m⁵C modification can influence the roles and functions of crucial regulatory molecules like miRNAs in cancer and beyond. Overall, these studies underscore the importance of investigating epitranscriptomic modifications for a deeper understanding of cancer biology and the development of novel diagnostic and therapeutic strategies.

Lastly, the presence of m⁵C modifications on long non-coding RNAs (lncRNAs) has unveiled a fascinating layer of epitranscriptomic regulation. While several long non-coding RNAs were reported to carry m⁵C modifications (Squires et al., 2012), its molecular function was described only on a small number of them (Amort et al., 2013; Cusenza et al., 2023).

Among the well-known lncRNAs, HOTAIR and XIST were found to bear m⁵C modifications (Amort et al., 2013). Position C:1683 of HOTAIR lncRNA was found to have a remarkably high methylation level in multiple cell lines. Although the methylation level in this position was found to remain unaltered in breast and ovarian cancer cells (Amort et al., 2013), it is noteworthy that this m⁵C site resides within the H1 region of the lncRNA, known for its interaction with the LSD1 complex (Tsai et al., 2010). This rises the intriguing question of whether the m⁵C content of HOTAIR could influence its binding affinity with LSD1.

In the same study, several m⁵C sites have been identified in XIST lncRNA, in positions 701, 702, 703, 711 and 712 (Amort et al., 2013). While the methylation content in these sites was calculated to be ~20% in HEK293 cell line, no methylation was observed in RNA samples derived from mouse embryos. *In vitro* RNA-binding assays revealed that the deposition of m⁵C modification impaired the binding of XIST to PRC2 complex (Amort et al., 2013).

Lastly, the m⁵C content in several other lncRNAs was demonstrated to have prognostic significance (Cusenza et al., 2023) in pathological conditions such as Pancreatic Ductal Adenocarcinoma (H. Yuan et al., 2021), Lung Adenocarcinoma (Pan et al., 2021) and others (Bai & Sun, 2022; Y. Li et al., 2020).

Given that lncRNAs and their methylation levels can be detected in different bodily fluids (Chandra Gupta & Nandan Tripathi, 2017), lncRNAs hold the potential to serve as valuable biomarkers for prognosis in a wide range of tumours.

This emerging field of m⁵C-modified lncRNAs opens exciting avenues for further research into their roles, mechanisms, and clinical applications in cancer and other diseases.

1.1.4 m⁵C in messenger RNA

The first observation of m⁵C modification in polyadenylated fractions of RNA was obtained about 5 decades ago, by incubating hamster (BHK-21) cells with [methyl-³H]-methionine and ³²P_i (Dubin & Taylor, 1975). For many years following this first discovery, research on m⁵C in mRNA remained dormant. Research on m⁵C as an mRNA modification started accelerating when a method for m⁵C detection and quantification became available for transcriptome-wide studies. In 2010, Squires and colleagues published a method for a quantitative and single-nucleotide detection of m⁵C, by coupling bisulfite conversion of RNA and high-throughput sequencing (bsRNA sequencing) (Squires & Preiss, 2010).

From a biochemical point of view, the treatment of RNA with bisulfite leads to deamination of cytosines into uracil. However, the extra methyl group in 5-methylcytosine protects the nucleoside from said deamination. As a consequence, upon cDNA generation via reverse transcription, library preparation and sequencing, m⁵C modification can be detected as non-converted cytosines, while unmodified cytosines will be detected as uracil (thymine in cDNA) (Sibbritt et al., 2022).

The application of this method on RNA samples derived from human cells lead to the discovery of approximately 10,000 m⁵C sites. Of these, ~8,500 were located on protein coding genes and showed a stoichiometry over 20% (Squires et al., 2012).

Subsequently, similar studies were published revealing comparable numbers of m⁵C modifications in other cell lines or tissues. In 2017, a murine methylome was published, featuring approximately 7,500 m⁵C sites with a non-conversion level above 20% found on 1650 mRNAs derived from mouse embryonic stem cells and about 2,000 sites located on almost 500 mRNAs derived from mouse brain tissues. (Amort et al., 2017). These early investigations established m⁵C as a widespread mRNA modification, primarily found in the coding sequence, with a pronounced methylation peak surrounding the start codon. Another study published on the same year found 5,399 m⁵C sites in HeLa cells, 5,036 of which were located in almost 2,000 mRNAs (X. Yang et al., 2017a). In 2018, a different group reported a

similar number of m⁵C sites in mRNA, highlighting an enrichment of methylation in the 5' untranslated regions (5' UTRs), in both mouse and human cells (Wei et al., 2018).

The bsRNA sequencing approach was also applied to other species, such as *Arabidopsis thaliana*, where researchers detected approximately 100 m⁵C sites on coding genes (David et al., 2017).

In addition to the significant body of work surrounding the methylome in mRNA, the bioinformatic approach used for m⁵C site calling in bsRNA sequencing data has undergone continuous refinement over the years. In recent works (Huang et al, 2019; Schumann et al, 2020; Liu et al, 2021), a more refined pipeline for m⁵C sites calling was adopted, which features a more stringent sets of filters that aim to ensure that only high-confidence sites are called and to filter out false positives. Some of these revised criteria include (a) a cut-off of 10% as a minimum non conversion level required for a given site to be considered convincingly methylated, as well as other filters like (b) signal-to-noise ratio and (c) the removal of reads with more than three non-converted cytosine, which account for highly structured regions of the RNA that are inaccessible to the bisulfite conversion and therefore potentially driving the number of false positive sites.

With this revisited approach in bsRNA sequencing data analysis, it became clear that the previously reported number of m⁵C sites in the transcriptome was likely an overestimate. In particular, bsRNA sequencing performed in HeLa cells revealed either 1,215 m⁵C sites (of which 686 in poly(A)-enriched RNA) (Huang et al, 2019), or 1,034 m⁵C sites (of which, 846 being in protein coding genes) (Schumann et al, 2020).

To date, it appears that for any given cell type, a number of m⁵C sites of approximately 1,000 or less seems more plausible than the initial estimates of 10,000. These numbers were also reiterated in other studies, whereby bsRNA sequencing experiments were performed in other human cell lines (Liu et al, 2021) or various human tissues (Huang et al, 2019).

Interestingly, the number of m⁵C sites can swell to several tens of thousands when analysing maternal mRNAs during early animal embryonic development (J. Liu et al., 2022). These surprising results seem to re-introduce the idea of m⁵C as an abundant RNA modification, despite the more conservative numbers proposed by recent studies (Huang et al, 2019; Schumann et al, 2020; Liu et al, 2021). However, the high number of m⁵C sites discovered during development seem to be in line with the dramatic changes in expression level of the many m⁵C methyltransferase enzymes during different developmental stages (Chi & Delgado-Olguín, 2013). These discoveries could pave the way towards a better understanding of the

distribution and function of m⁵C modification, which may vary depending on specific physiological or pathological conditions throughout the lifespan of an organism.

1.2 The protein players of m⁵C modification

In a similar fashion to other RNA modifications, such as m⁶A (Y. Yang et al., 2018), there are a number of proteins that revolve around the molecular deposition and function of m⁵C. The three categories of protein partners are writer (RNA methyltransferases), erasers (dioxxygenase enzymes) and reader proteins (RNA-binding proteins that interact with RNA via recognition of m⁵C modification).

1.2.1 m⁵C writer proteins

The deposition of m⁵C modification on RNA can be mediated by over ten different RNA methyltransferases (Y. S. Chen et al., 2021). Among these, there are proteins belonging to the NOL1/NOP2/SUN domain (NSUN) family, the homologue of the DNA methyltransferases (DNMT) protein, DNMT2, and the tRNA methyltransferase TRDMT (Bohnsack et al., 2019).

1.2.1.1 NSUN family proteins

The NSUN family of proteins features seven m⁵C methyltransferase enzymes, NSUN1 to NSUN7 (Bohnsack et al., 2019). While NSUN1, NSUN2, and NSUN5 are conserved throughout eukaryotes, the remaining NSUN proteins are only present in higher eukaryotes (Bohnsack et al., 2019). All the NSUN enzymes can deposit m⁵C modification to one or more RNA species, thereby playing important roles in cellular physiology.

NSUN1 is a nucleolar protein that can directly bind and methylate ribosomal RNA in position C:4447 of 28S rRNA in human and C: 2870 of 25S rRNA in yeast (Bourgeois et al., 2015). Early studies showed that NSUN1 expression was associated with enhanced proliferation of cancer cells in human (Trerè et al., 2000). NSUN1 was also shown to deposit m⁵C modification on TAR RNA, which suppresses HIV-1 transcription (Kong et al., 2020).

NSUN2 is perhaps the most studied among m⁵C RNA methyltransferases. This writer protein was shown to be able to deposit m⁵C modifications on several eukaryotic RNA species, such as mRNA (Huang et al, 2019; Schumann et al, 2020; Yang et al, 2017), tRNA (Auxilien et al., 2012; Blanco et al., 2014) and non-coding RNA (Sajini et al., 2019; Squires et al., 2012).

Firstly characterised for its methyltransferase activity on tRNA, NSUN2 catalyses the addition of m⁵C modifications in position C:48/49/50 on the variable loop of several tRNAs (Bohnsack et al., 2019), where they enhance tRNA stability and protein synthesis (Blanco et al., 2014). NSUN2 is believed to be the main methyltransferase also for mRNA, where it can decorate both the coding sequence and untranslated regions with m⁵C modification (Huang et al, 2019; Schumann et al, 2020; Yang et al, 2017). NSUN2-dependent m⁵C sites are typically characterised by a m⁵CNNGG sequence motif and a secondary structure that mimics the sequence context around the NSUN2-dependent m⁵C site located in the variable loop of tRNA (Huang et al, 2019; Schumann et al, 2020), and therefore classed as “Type I” sites.

Dysregulations of NSUN2 levels have been associated with different pathological conditions. For example, lack of NSUN2 protein can lead to mitotic disorders, aneuploidy and cell death (Hussain, Sajini, et al., 2013). NSUN2 has regulatory role also in neuroepithelial and epidermal stem cells differentiation (Blanco et al., 2014; Hussain, Sajini, et al., 2013), and deficiency of NSUN2 during brain development was shown to delay neural differentiation and migration, triggering the onset of Dubowitz-like syndrome (Flores et al., 2017; Martinez et al., 2012).

In addition, NSUN2 was found to be differentially expressed in various cancer cells, where it is involved in pathogenesis. In gallbladder cancer, lack of NSUN2 leads to reduced proliferation and tumorigenesis (Gao et al., 2019). Similarly, NSUN2-dependent methylation promotes the pathogenesis of human bladder cancer by stabilising mRNAs (X. Chen et al., 2019).

It's interesting to note that NSUN2 plays a role in viral metabolism, particularly in the context of HIV-1 RNA (Courtney, Tsai, et al., 2019) and murine leukemia virus (MLV) (Courtney, Chalem, et al., 2019). By depositing m⁵C modifications on these viral RNAs, NSUN2 impacts various aspects of viral metabolism, such as viral mRNA alternative splicing and translation in HIV-1 virus (Courtney, Tsai, et al., 2019) and gene expression and replication of MLV virus (Courtney, Chalem, et al., 2019). This dual role of NSUN2 in both cellular processes and viral interactions highlights its multifaceted functions and the potential significance of its regulation in different biological contexts.

NSUN3 is responsible for the methylation of the mitochondrial tRNA^{Met} (mt-tRNA^{Met}) in position C:34 (Trixl et al., 2018), which in turn can be oxidised into f⁵C34 by the human and mouse dioxygenase ABH1/ALKBH1 (Haag et al., 2016). This modification is located in the wobble position of mt-tRNA^{Met}, it is necessary for efficient mitochondrial translation and enables the recognition of the three codons for methionine used in the non-universal genetic code of mitochondria (Haag et al., 2016). Upon NSUN3 knockout, human cells showed a dramatic

reduction translation of mitochondrial protein and a reduced oxygen consumption, leading to deficient mitochondrial activity (Nakano et al., 2016). In mouse embryonic stem cells (mESCs), the knockout of NSUN3 methyltransferase triggered differentiation towards endoderm and mesoderm instead of neuroectoderm (Trixl et al., 2018).

NSUN4 is the methyltransferase responsible for the modification of rRNA (Cámara et al., 2011). NSUN4 was shown to deposit m⁵C modification in position C:911 of 12S mitochondrial rRNA, upon recruitment by MTERF4 (Metodiev et al., 2014). NSUN4 can regulate mitochondrial ribosome assembly and protein synthesis (Cámara et al., 2011). The knockout of NSUN4 protein in the heart was shown to cause cardiomyopathy and respiratory defects, probably caused by defects in mitochondrial assembly and function (Metodiev et al., 2014).

NSUN5 protein can target rRNA and can modify position C:3782 in human and C:3438 in mouse (Heissenberger et al., 2019; Janin et al., 2019). Human m⁵C3438 has a major role in rRNA function (Janin et al., 2019). The epigenetic silencing of NSUN5 gene in glioma was shown to drive an overall depletion of protein synthesis, and leads to the emergence of an adaptive translational program for survival under conditions of cellular stress (Janin et al., 2019). In addition, such NSUN5 deficiency also renders these gliomas sensitive to bioactivatable substrates of the stress-related enzyme NQO1 (Janin et al., 2019).

NSUN5 methyltransferase is found under-expressed in 95% of patients affected by Williams-Beuren syndrome (P. Chen et al., 2019), as well as cognitive deficit disease (T. Zhang et al., 2019). Fibroblasts with NSUN5 haploinsufficiency extracted from Williams-Beuren syndrome patients showed a substantial reduction of m⁵C content at position C:3782. Altogether, this suggests that the methyltransferase activity of NSUN5 protein can play a major role in regulating the development of the nervous system.

NSUN5 can also mediate the deposition of m⁵C in position C:2268 of 25S rRNA in *Arabidopsis Thaliana* (Burgess et al., 2015), while its yeast homolog modifies 25S rRNA in position C:2278 (Schosserer et al., 2015).

NSUN6 can deposit m⁵C modification in position C:72 of the acceptor stem of human tRNA^{Cys} and tRNA^{Thr} (Haag et al., 2015). Importantly, in order for NSUN6 to recognise its tRNA targets and correctly methylate them, tRNAs have to be full length, properly folded and added with the extra CCA sequence at their 3' end (Haag et al., 2015).

The homolog of NSUN6 in archaeal *Pyrococcus horikoshii* can also methylate tRNA^{Cys} in position C:72, thereby conferring this tRNA a moderately improved thermal stability (J. Li et al., 2019).

NSUN6 can also act as an m⁵C methyltransferase for mRNA, where it accounts for a considerable amount of m⁵C sites, after NSUN2 (Selmi et al, 2021; Liu et al, 2021). NSUN6-dependent m⁵C sites on mRNA are classed as “Type II” sites and are typically characterised by the sequence (m⁵CUCCA) and secondary structure motif found in its tRNA target (Liu et al, 2021).

The last of the member of the NSUN family of protein, NSUN7, acts as the methyltransferase for enhancer RNAs (eRNAs), where it is suggested to function as a regulator of adaptive metabolism response (Aguilo et al., 2016). NSUN7 expression level is abundant during mouse embryonic development, during which a deficient expression of NSUN7 can lead to sperm mobility issues and infertility (Harris et al., 2007), as well as intellectual disability (Chi & Delgado-Olguín, 2013).

A recent study discovered that NSUN7 can mediate the deposition of m⁵C on several tens of mRNAs (Ortiz-Barahona et al., 2023). Among these, the mRNA of the coiled-coil domain containing 9B (CCDC9B) gene (chr15:40335427) requires NSUN7-mediated methylation for transcript stability. NSUN7 is often epigenetically silenced in liver malignant cells; as a consequence, CCDC9B mRNA gets destabilised, leading to the upregulation of its protein partner IVNS1ABP, involved in MYC signalling. This cascade of events ultimately leads to liver cancer cells to be more sensitive towards the growth-inhibitory actions of bromodomain inhibitors (Ortiz-Barahona et al., 2023).

1.2.1.2 TRDMT1

tRNA-aspartic acid methyltransferase 1 (TRDMT1), originally known as DNA methyltransferase 2 (DNMT2), is a conserved m⁵C writer protein responsible for the deposition of m⁵C modification in position C:38 of tRNA^{Asp} (Goll et al., 2006). This modification was shown to be improve the translation of Asp-rich proteins (Shanmugam et al., 2015) as well as to ensure translation fidelity (Tuorto et al., 2015).

TRDMT1 is necessary to trigger a normal immune response in *Drosophila*, and its deficiency was reported to lead to accumulation of *Drosophila* C virus (Durdevic et al., 2013). The lack of TRDMT1 protein in *Drosophila* was also found to increase cell sensitivity to heat and oxidative stress, as a result of stress-induced cleavage of tRNA (Schaefer et al., 2010).

Finally, TRMDT1 was found to deposit m⁵C modification on mRNA, and its depletion was reported to inhibit HEK293 cell proliferations and migration (S. Xue et al., 2019).

1.2.1.3 TRM4A/B

Both TRM4A and TRM4B catalyse the addition of m⁵C modification on tRNA in *Arabidopsis thaliana* (J. Hu et al., 2019). In addition to tRNA, TRM4B (but not 4A) can also methylate mRNA and other non-coding RNAs, such as lncRNA and snRNA (Cui et al., 2017; David et al., 2017). In particular, TRM4B-mediated deposition of m⁵C was reported to be an important regulator of *Arabidopsis*' cell proliferation and sensitivity to oxidative stress (David et al., 2017).

1.2.2 m⁵C eraser proteins

Unlike other epitranscriptomic marks, such as m⁶A modification, that has designated proteins responsible for the removal of the methyl group from the modified adenosine (Jia et al., 2011; Zheng et al., 2013), m⁵C eraser proteins present a more complex scenario.

Recent research showed that m⁵C does not seem to have a protein responsible for demethylation, but it rather presents a more intricate landscape. There are two proposed m⁵C “erasers”, but instead of removing the methylation they further modify the target RNA (Y. S. Chen et al., 2021).

TET1/2/3 are a group of dioxygenase proteins that belong to the ten-eleven translocation (TET) family. They were originally characterised for their ability to oxidise m⁵C into 5-hydroxymethylcytosine (hm⁵C) on DNA (Kohli & Zhang, 2013; Tahiliani et al., 2009). However, a recent study discovered that These set pf proteins can also target RNA, thereby oxidising RNA m⁵C into hm⁵C (Fu et al., 2014). The overexpression of TET proteins in HEK293 cells resulted in the overall increase of cellular hm⁵C levels (Fu et al., 2014). Notably, TET2 protein was found to target the oxidation of m⁵C modification in mRNA, which triggers mRNA degradation (Shen et al., 2018). This indicates an important role for both modification in regulating mRNA gene expression, and the deregulation of mRNA stability upon disruption of m⁵C-hm⁵C balance.

Similarly, alpha-ketoglutarate-dependent dioxygenase ABH1 (ALKBH1) protein is also dioxygenase enzyme that can target both DNA (Wu et al., 2016) and RNA (Westbye et al.,

2008). In the context of m⁵C modification in RNA, ALKBH1 can catalyse the oxidation of m⁵C34 to hm⁵C34 of tRNA^{Cys} in the cytoplasm and tRNA^{Met} in mitochondria (Haag et al., 2016; Kawarada et al., 2017). The lack of ALKBH1 protein was associated with deficient translation of mitochondrial proteins and oxygen consumption (Kawarada et al., 2017). Therefore, the data collected in these studies indicate the important of m⁵C for the correct function of mitochondrial metabolism.

m ⁵ C writer proteins	
MTase	target RNA species
NSUN1	28S rRNA
NSUN2	tRNA, vtRNA, mRNA
NSUN3	mt-tRNA
NSUN4	mt-12S rRNA
NSUN5	28S rRNA, mRNA
NSUN6	tRNA, mRNA
NSUN7	eRNA, mRNA
TRDMT1	tRNA, mRNA
TRM4A	tRNA, mRNA, lncRNA
TRM4B	tRNA

Table 1 – m⁵C writer proteins.

1.2.3 m⁵C reader proteins

The way in which epitranscriptomic modifications exert their function in regulating RNA metabolism is often related to their ability to interact with RNA-binding proteins (RBPs). Proteins that bind RNA specifically through the interaction with their m⁵C modifications are called m⁵C readers. Recent literature has revealed a handful of RBPs that work as m⁵C readers: Aly/REF export factor (ALYREF) (X. Yang et al., 2017a), Y-box-binding protein 1 (YBX1) (X. Chen et al., 2019; Y. Yang et al., 2019; Zou et al., 2020), Y-box-binding protein 2 (YBX2) (X. Wang et al., 2022) and serine and arginine rich splicing factor 2 (SRSF2) (H. L. Ma et al., 2023).

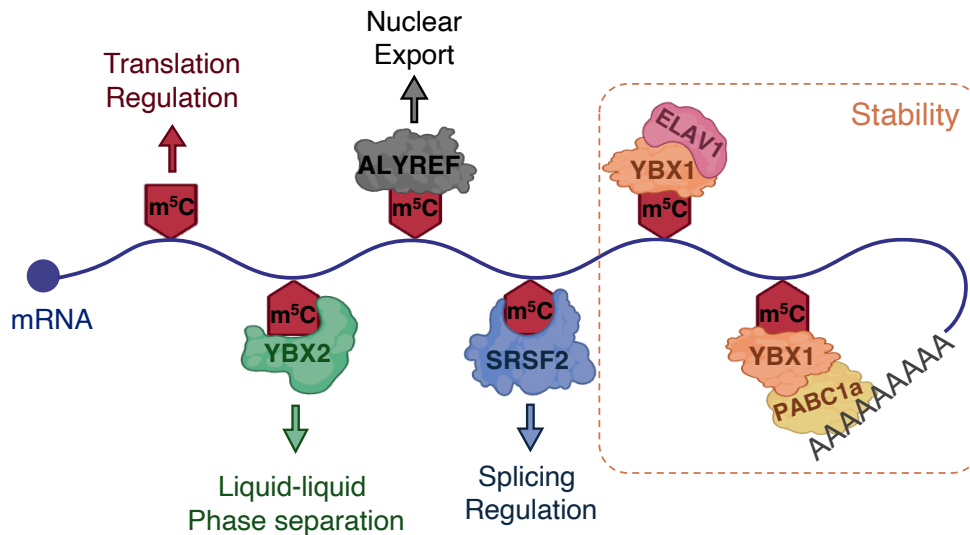


Figure 2 – The function of m⁵C:mRNA reader proteins.

Adapted from Guarnacci & Preiss, (2024)

ALYREF is the very first identified m⁵C reader protein (X. Yang et al., 2017a). Being a key component of the TREX complex, a well-known protein complex for mRNA export (Dufu et al., 2010), ALYREF is a nuclear protein. In the nucleus, ALYREF is able to binds m⁵C-decorated mRNA via interaction with its m⁵C recognition site on K171 and mediate their transport to the cytoplasm (X. Yang et al., 2017a). Knockdown experiments of NSUN2 methyltransferase resulted in a significant accumulation of ALYREF reader in the nuclear speckles, despite its unchanged overall abundance in the cells. This suggests that m⁵C modification contributes to the regulation of mRNA export to the cytoplasm (X. Yang et al., 2017a). Photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation (PAR-CLIP) sequencing experiments revealed that almost 50% of m⁵C-decorated mRNAs in HeLa and T24 cells are bound by ALYREF reader protein (X. Chen et al., 2019), further indicating the extent and significance of ALYREF as an m⁵C reader protein.

ALYREF plays an important role as an m⁵C reader also in viral biology (Dufu et al., 2010). Triple-quadrupole (QQQ) liquid chromatography-mass spectrometry (LC-MS) analyses revealed that the RNA genome of murine leukemia virus (MLV) is highly modified with m⁵C. ALYREF knockdown and overexpression experiments revealed a m⁵C/ALYREF-dependent regulation of viral protein production and replication, specifically via regulating the levels of MLV Gag protein. Notably, overexpression of a mutant version of ALYREF protein (K171A) that is unable to bind m⁵C-RNA failed to rescue the phenotype (Dufu et al., 2010).

YBX1 is another m⁵C reader protein discovered in human cells (X. Chen et al., 2019). YBX1 is a cytoplasmic protein, and can bind and stabilise m⁵C-decorated RNA via interaction with its indole ring of W65 in its cold-shock domain (CSD) (X. Chen et al., 2019).

In bladder cancer (human urothelial carcinoma of the bladder - UCB), YBX1 can stabilise several oncogenes via recruiting ELAVL1 protein onto methylated mRNA. Among these, the oncogene HDGF (chr1:156743366) is a target of YBX1, and its stabilisation was linked to cancer pathogenesis and progression (X. Chen et al., 2019). Notably, in the same study, PAR-CLIP sequencing experiments revealed that 60% of m⁵C-decorated mRNAs in HeLa and T24 cells are bound by YBX1 reader protein and 394 of them had a reduced RNA level upon NSUN2 methyltransferase knockdown (X. Chen et al., 2019).

YBX1 is an m⁵C reader also in zebrafish, where over 85% of m⁵C-mRNAs were recognised by YBX1, and 397 of them had a reduced RNA level upon YBX1 knockdown (Y. Yang et al., 2019). In the same study, researchers demonstrated that YBX1 has an important role during zebrafish maternal-to-zygotic transition (MZT) by stabilising maternal mRNAs. In particular, once bound to a methylated mRNA, YBX1 can recruit Pabpc1 (poly(A) binding protein cytoplasmic 1), and therefore co-ordinately regulate the stability of m⁵C-modified maternal mRNA (Y. Yang et al., 2019).

The homolog protein of YBX1 in *Drosophila*, Ypsilon schachte (YPS), also works as a m⁵C reader protein in *Drosophila* ovary (Zou et al., 2020). YPS protein was demonstrated to regulate the homeostasis, proliferation and differentiation of *Drosophila* germline stem cells (GSCs), and mutations in the CSD of this protein lead to defective GSC development (Zou et al., 2020).

A second member of the YBX family found to function as an m⁵C reader protein is the testis-enriched YBX2 (X. Wang et al., 2022). YBX2 contains an intrinsically disordered domain and is inclined to undergo liquid-liquid phase separation. Research has shown that m⁵C:RNA facilitates the formation of YBX2 droplets in vitro. In cultured HeLa cells, YBX2 binding was notably enriched around the 5' region m⁵C sites introduced by NSUN2, and the formation of YBX2 droplets was reduced when NSUN was knocked out (X. Wang et al., 2022).

Finally, the most recent m⁵C reader protein was identified to be the serine and arginine rich splicing factor 2 (SRSF2) (H. L. Ma et al., 2023). SRSF2 footprints were found to colocalise with m⁵C modifications on mRNA, and its knockdown results in splicing deregulation (H. L. Ma et al., 2023).

m ⁵ C reader proteins	
RBP	Function
ALYREF	Nuclear Export
SRSF2	Splicing Regulation
YBX1	mRNA Stability
YBX2	Liquid-liquid Phase Separation

Table 2 - m⁵C reader proteins.

1.3 The function of m⁵C

The accumulation of studies from the past decade has brought to the light how m⁵C is a widespread RNA modification, present in all known species of RNA (Boccaletto et al., 2022). The combined efforts of different research groups have also revealed how m⁵C can have a broad effect on regulating different steps of coding and non-coding RNA metabolism, including stability (X. Chen et al., 2019; Y. Yang et al., 2019), export (X. Yang et al., 2017a) and translation (Schumann et al, 2020).

From a cellular biology point of view, m⁵C modification and several of its methyltransferase proteins have been implicated in different cellular contexts, including cell proliferation (W. Wang, 2016), migration and development (P. Song et al., 2021).

Lastly, m⁵C levels of RNA is often found to be deregulated in different cancer types, and m⁵C writer, reader and eraser proteins were all found to have important roles in cancer and cancer microenvironment (H. Song et al., 2022).

1.3.1 The role of m⁵C in molecular function

The regulation of gene expression is an intricate and multifaceted process. It is now clear how the regulation of mRNA processing plays a considerable part in a wide array of cellular processes. Building upon this, the discovery and study of post-transcriptional modifications of RNA have revealed a new and powerful tool that provides versatility and adaptability to many different molecular and biological processes, making epitranscriptomic an additional layer of gene expression regulation.

Among other mRNA modifications, m⁵C was found to be an important regulator for several steps of RNA metabolism (Gilbert & Nachtergaele, 2023).

To date, it is unclear at what stage of mRNA processing m⁵C is deposited. A small proportion of m⁵C sites have been reported to be localised within introns both in human (X. Yang et al., 2017a) and in mouse embryonic stem cells (mESCs) and brain tissue (Amort et al., 2017). This suggests that the deposition of m⁵C may happen on pre-mRNA, prior to splicing events. This hypothesis would also be in agreement with the nuclear presence of some of the m⁵C methyltransferase enzymes (Bohnsack et al., 2019). However, the majority of these intronic sites were found from the sequencing of poly(A)-enriched RNA samples, therefore it cannot be excluded that the m⁵C deposition happened on mature mRNAs with retained introns as opposed to nascent pre-mRNAs. Nucleus-cytoplasm fractionation followed by bsRNA sequencing experiments revealed that intronic m⁵C sites in poly(A)-selected RNA were much more abundant in the nucleus (44% to 70%) compared to the total cellular poly(A) RNA (3% to 11%) (Amort et al., 2017). These findings could suggest that these transcripts with retained introns are either retained and degraded in the nucleus or are promptly degraded upon export to the cytoplasm. In both hypotheses, it is possible that m⁵C could work as a mark for improperly spliced mRNAs.

Additionally, enrichment analyses performed by Amort and colleagues (2017), revealed that the footprint of SRSF3 and SRSF4 splicing factors on target mRNAs was significantly overlapping with m⁵C sites, indicating the possibility for m⁵C modifications to take part in splicing regulation.

mRNA export to the cytoplasm is a tightly regulated process (Adams & Wente, 2013; Reed & Hurt, 2002). m⁵C modification has been implicated in mRNA export regulation via the interaction of the TREX complex (X. Yang et al., 2017a). This process requires the direct interaction between m⁵C sites on mRNA and ALYREF RBP, an m⁵C reader protein that functions as a link to the TREX complex, which facilitates mRNA export. Upon NSUN2 knockout, the shuttling of ALYREF between the nucleus and the cytoplasm was affected, therefore affecting m⁵C-mRNA subcellular localisation. The lack of NSUN2 methyltransferase resulted in the retention of m⁵C-mRNAs in the nucleus, while no effect was observed in non-methylated transcripts, therefore indicating a critical role for m⁵C modification in regulating mRNA export (X. Yang et al., 2017a).

In 2019, two different studies were published providing evidence for m⁵C being a regulator for mRNA stability in human (X. Chen et al., 2019) and zebrafish (Y. Yang et al., 2019). Notably, while both groups linked m⁵C modification to mRNA stability, the mechanism behind this regulation in the two different species relied on different pathways.

The maternal-to-zygotic transition (MZT) represents a crucial phase in development where the overall expression of maternal mRNAs decreases, and the activation of zygotic genes becomes essential (Yartseva & Giraldez, 2015). In the case of zebrafish, the NSUN2 methyltransferase is responsible for introducing the m⁵C modifications, which serve to enhance the stability of a specific group of maternal mRNAs involved in mRNA processing and cell cycle regulation (Y. Yang et al., 2019). This phenomenon is dependent on the YBX1 recognising and binding m⁵C-containing mRNAs, and subsequently recruiting PABPC1, thereby stabilising these mRNA transcripts. Notably, the absence of YBX1 in zebrafish leads to a halt in development at 6 hours post-fertilization and eventual fatality, underscoring the pivotal role of m⁵C-mediated maternal mRNA stabilization in ensuring proper development (Y. Yang et al., 2019).

In human, m⁵C-dependent mRNA stabilisation was observed in urothelial carcinoma of the bladder (UCB) (X. Chen et al., 2019). RNA samples derived from UCB patients were found to have significantly more m⁵C sites compared to their relative control, and this higher m⁵C content correlated with increased stability of several oncogenes previously observed to drive tumorigenesis and metastasis (X. Chen et al., 2019). Chen and colleagues demonstrated that YBX1 protein, upon binding m⁵C-decorated oncogenes could recruit ELAVL1 protein, thereby improving stability.

m⁵C is also a stability regulator for other classes of RNA molecules. In tRNA for example, NSUN2 dependent m⁵C48 and m⁵C49 modifications are strategically positioned at the junction sequence of the V-loop (VL) and T-stem-loop (TSL). These modifications contribute significantly to tRNA stability and promote efficient protein translation (Reid et al., 1999; Van Haute et al., 2019). Similarly, m⁵C modification in position C:2278 on 25S rRNA in yeast has an important role in stabilising the molecule, and its deletion results in the impaired stability of the 60S subunit and reduced translation efficiency (Sharma et al., 2013).

Lastly, m⁵C modification of mRNA as well as tRNA and rRNA has a role in regulating protein synthesis. m⁵C sites on tRNA are predominantly located on the variable loop and anticodon arm (**Fig. 1**) and are important for tRNA stability and codon recognition (H. Song et al., 2022). In mice, the deletion of m⁵C site in position C:38 of tRNA^{Met} leads to reduced translation fidelity (Tuorto et al., 2015). Similarly, m⁵C site on C:38 of tRNA^{Asp} is important to distinguish between cognate and near-cognate codons, such as differentiating tRNA^{Asp} from tRNA^{Glu} (Tuorto et al., 2015). In yeast, m⁵C:34 of tRNA^{Leu} is also important for translation regulation. m⁵C:34 levels are significantly increased upon oxidative stress, which leads to an enhanced translation of

UUG-codons abundant mRNAs, such as the ribosomal protein eL22a (Rpl22a) (Chan et al., 2012).

m⁵C deposition on rRNA is relevant for ribosome and protein synthesis. The methylation of C:2278 in yeast 25S rRNA maintains correct rRNA folding upon oxidative stress, which in turn facilitates the selective recruitment and translation of mRNAs in response to distinct cellular signals (Schosserer et al., 2015). rRNA methylation is also critical for mitochondrial translation: with m⁵C:911 on 12S rRNA being necessary for the correct assembly of the mitochondrial ribosome and therefore protein synthesis (Metodieiev et al., 2014).

m⁵C deposition on mRNA also has an effect on their translation efficiency. As m⁵C modification on the anticodon loop of tRNA affects the codon-anticodon recognition (Shanmugam et al., 2015), it is plausible to hypothesise that the methylation of mRNA could have similar effects. As NSUN2 and NSUN6 methyltransferases can target both mRNA and tRNA, a global assessment of mRNA translation using knockdown or knockout approaches would make it challenging to distinguish between direct and indirect effects. However, Schumann and colleagues used a different approach to assess the role of effects of m⁵C modification on mRNA on translation efficiency. Polysome profiling experiments in HeLa cells, followed by bsRNA sequencing revealed a broad reverse correlation between m⁵C content on mRNA and their association with (poly)ribosomes (Schumann et al., 2020). This indicates that the methylation levels of mRNA could affect the efficiency at which ribosome could process mRNAs for protein synthesis.

1.3.2 The role of m⁵C in cellular function

A collection of recent literature has brought to light numerous links between m⁵C modification, and its methyltransferase enzymes, with various cellular pathways (Wilkinson et al., 2022).

In 2021, Jian and colleagues discovered a link between m⁵C and cell death. BsRNA sequencing performed on a neuronal model for oxygen-glucose deprivation and reoxygenation (OGD/R) revealed an overall increase of methylation upon OGD/R (Jian et al., 2021). In particular, a subset of transcripts were found hypermethylated and they were functionally associated with apoptosis (Jian et al., 2021). Studies about the involvement of m⁵C modification in cell death are limited, and while the functional significance of this association needs to be further investigated, it suggests a role for m⁵C in initiating apoptosis upon OGD/R-mediated stress conditions.

NSUN2-mediated RNA methylation was found to promote cellular senescence in different contexts. In HeLa cells, the m⁶A writer complex METTL3/METTL14 and m⁵C methyltransferase NSUN2 can synergistically deposit m⁶A and m⁵C on the 3' UTR of p21 gene (C:2079) (Q. Li et al., 2017). Together, these two methylation marks can increase the translation level of p21 protein, which ultimately induces cellular senescence (Q. Li et al., 2017). In human umbilical vein endothelial cells, NSUN2-dependent methylation of SHC mRNA (chr1:154963697) enhances the translation of the three proteins, p66SHC, p52SHC, and p46SHC. As a result, an accelerated oxidative stress- and high-glucose-induced senescence of human vascular endothelial cells (HUVEC) was observed (Cai et al., 2016).

Conversely, NSUN2 was shown to negatively affect senescence in human diploid fibroblasts (Tang et al., 2015). NSUN2 deposits an m⁵C modification in position C:64 on the 5' UTR of p27, an inhibitor of CDK, which lowers the translation efficiency of p27. A decrease of p27 protein level leads to an increase in expression of CDK1, which in turn inhibits cellular senescence. Therefore, by indirectly promoting CDK1 protein synthesis, NSUN2-dependent m⁵C can promote cell proliferation and inhibit cellular senescence (Tang et al., 2015).

Overall, the apparently contrasting roles of NSUN2 to regulate cellular senescence may be context specific and relying on different pathways to either promote or contrast cellular senescence.

NSUN2 protein is also involved in regulating the delicate balance between cellular self-renewal and differentiation, by methylating tRNA. Upon NSUN2 knockout in epidermal stem cells, the anagen phase (the active growth phase of hair follicles) cannot be initiated correctly and significantly delays the accurate differentiation of hair lineages (Blanco et al., 2011).

In male gonads, NSUN2-mediated m⁵C modification on tRNAs is not essential for spermatogonial stem cells but become specifically necessary for the meiotic progression of germ cells into the pachytene stage (Hussain, Tuorto, et al., 2013). Similarly, during early human brain development, NSUN2 is expressed in neuroepithelial progenitors and is required for their differentiation potential. The loss of NSUN2 results in increased tRNA fragmentation, impairing the migration and differentiation of neuroepithelial stem cells (Flores et al., 2017).

In a similar fashion to NSUN2, NSUN7 has also been implicated in cell differentiation. This connection is highlighted by the fact that loss-of-function mutations in NSUN7 have been associated with male infertility, suggesting that the NSUN7-mediated m⁵C methylation pathway could play a role in germ cell differentiation (Harris et al., 2007).

NSUN2 protein was also shown to have a role in regulating cell migration in an m⁵C-dependent manner in the human glioma cell line U87 (X. Xu et al., 2020). NSUN2 deposits an m⁵C

modification at C:2756 of the 3' UTR of the autotaxin (ATX) gene, which is then bound by ALYREF reader protein, transported into the cytoplasm and translated. ATX is a glycoprotein recognised as a key factor for the conversion of lysophosphatidylcholine into lysophosphatidic acid (LPA), a lipid which promotes cellular migration and tumorigenicity (Valdés-Rives & González-Arenas, 2017; X. Xu et al., 2020). Interestingly, knockdown of NSUN2 decreased ATX protein levels and suppressed the migration of U87 cells, whereas overexpression of NSUN2 elevated ATX protein levels (X. Xu et al., 2020).

NSUN2 was also found to promote cell migration in gastric cancer (GC), possibly via the deposition of m⁵C sites on PIK3R1 and PCYT1A mRNAs, genes known to be involved in multiple cancer-related signalling pathways (Y. Hu et al., 2021).

Notably, mRNA can be decorated with m⁵C modification at DNA damage sites. This modification is mediated by TRDMT1 methyltransferase, which is recruited to these sites to deposit m⁵C modifications on RNA molecule in a DNA:RNA hybrid (H. Chen et al., 2020). The absence of TRDMT1 compromises homologous recombination (HR) and heightens cellular sensitivity to DNA double-strand breaks (DSBs). These findings reveal an unexpected TRDMT1-m⁵C relationship that is necessary for HR, suggesting that m⁵C RNA modifications can also function as DNA damage codes, regulating DNA repair processes (H. Chen et al., 2020).

1.3.3 The role of m⁵C in cancer and immunotherapy

The overall level of m⁵C content, as well as the expression levels of m⁵C writers, readers or erasers, are often related to different cancer types to varying degrees (H. Song et al., 2022). For instance, writer proteins like NSUN1 and NSUN2 have shown potential as predictive markers for the onset and progression of various cancer types, including colorectal, breast, and lung cancers (Chellamuthu & Gray, 2020). Similarly, the expression levels of writer protein DNMT1 and the reader protein ALYREF could predict morbidity and mortality in patients affected with head and neck squamous cell carcinoma (HNSCC) (M. Xue et al., 2020).

Another recent study compared the mutation and copy number variations (CNVs) rates of several m⁵C writers, readers and erasers between several different cancer types and their adjacent normal tissue controls (Y. He et al., 2021). This pan-cancer analysis revealed that aberrant alterations of m⁵C regulatory genes significantly correlated to poor clinical outcomes among various tumours, which may drive cancer pathogenesis and/or survival (Y. He et al., 2021).

Two recent studies have proposed m⁵C content and its methyltransferase enzymes expression level to be a prognostic biomarker for Hepatocellular Carcinoma (HCC) (Y. He et al., 2020; Z. Sun et al., 2020). Aberrant regulation of m⁵C regulatory genes was linked to HCC patients' survival rate. In particular, elevated expression of NSUN4 and ALYREF is significantly associated with low survival rates among HCC patients (Y. He et al., 2020). Moreover, NSUN2-dependent methylation of H19 lncRNA was found to increase its stability and therefore interaction with the oncoprotein G3BP1, thereby leading to poor HCC differentiation (Z. Sun et al., 2020).

Methylation levels of mRNAs has also been linked to cancer pathogenesis. In human bladder urothelial carcinoma (UCB), NSUN2-dependent m⁵C modification on the 3'UTR of HDGF (chr1:156743366) oncogene and recognition by YBX1 reader protein drives pathogenesis of the urothelial carcinoma (X. Chen et al., 2019). Moreover, overexpression of NUSN2, YBX1, or HDGF predicted a poorer survival rate in UCB patients (C. Zhang et al., 2019).

Given the established relation between m⁵C modification biology and various types of cancers, researchers have started investigating the potential role of m⁵C modification and m⁵C writers and readers in cancer immunotherapy.

One of the hallmarks of cancer is the cell evasion from the immune system, with T-cell exhaustion being one of the possible mechanisms by which human cancer can escape immune surveillance (Lu et al., 2020). The interaction between NSUN2 expression and T-cell activation status was reported to correlate to HNSCC patients' survival, which indicates that NSUN2 could serve as a precision marker for immune-checkpoint blockade and a potential target for therapy (Lu et al., 2018, 2020). Moreover, NSUN2 was shown to promote the translation of the intracellular adhesion molecule 1 (ICAM-1) mRNA via m⁵C deposition, which functionally leads to the inhibition of M2-macrophage polarisation and the suppression of tumour metastasis (Luo et al., 2016).

Other m⁵C regulator proteins are also reported to be possible targets for cancer immunotherapy. YBX1 m⁵C reader protein is one of the most common tumour-associated antigen and is known to induce T cell response (Trevarton et al., 2019), which could make it a good target of effector immunity and a possible vaccine candidate.

YBX1 also exhibits upregulation in chemo-resistant HCC cells, and its knockout has been shown to reverse chemoresistance through two mechanisms: by inhibiting PD-L1 expression and by enhancing T-cell activation within the tumour microenvironment (Maurya et al., 2017).

This reversal involves an increase in functional cytotoxic CD8⁺ T cells and a decrease in myeloid-derived suppressor cells and regulatory T cells, ultimately overcoming the immunosuppressive environment and immune escape observed in HCC (Maurya et al., 2017). Furthermore, it was demonstrated that targeting the YBX1 signalling cascade can effectively address both tumour immune evasion and multidrug resistance, offering a promising therapeutic approach against chemo-resistant tumours (Tao et al., 2019). In summary, these findings highlight the potential utility of modulating m⁵C regulators as a viable strategy to develop therapy.

2. Aims of the thesis

This thesis is driven by the overarching goal of unravelling the intricate landscape of 5-methylcytosine (m⁵C) modifications within RNA molecules and deciphering their multifaceted roles in cellular processes. The primary aims of this research can be summarized into three key objectives:

1. Compiling a union list of human m⁵C sites

The first aim of this thesis is to provide a high confidence unified dataset of m⁵C modifications across various RNA species within the human transcriptome, comprehensive of multiple cell types for which m⁵C detection data is available. This involves the systematic re-analysis of publicly available RNA Bisulfite-Sequencing (BsRNA-Seq) data to generate a union list of human m⁵C sites. By delineating the distribution and prevalence of m⁵C modification on RNA, we aim to clarify the “methylome” of diverse human cell types in a way that is unbiased by transcriptome differences of individual cells in the human body, with regard to different expression levels of m⁵C writer proteins and their targets. Additionally, the resulting union set will provide a high statistical power that can be used to perform various enrichment analyses.

2. Investigating the enzymology of different NSUN proteins

The second major objective of this thesis revolves around investigating the diverse array of writer proteins responsible for the deposition of m⁵C modification. This includes the exploration of specific m⁵C writers, such as NSUN2, NSUN6, and NSUN5, and their roles in mediating mRNA methylation. With this, we aim to elucidate the specific RNA features that are recognised by these proteins and are responsible for m⁵C modification distribution pattern in the human transcriptome.

3 Investigating the interplay between m⁵C modifications and RBPs

The third aim of this research focuses on the interplay between m⁵C modifications and RNA-binding proteins (RBPs). We seek to identify and analyse the enrichment of known RBPs around m⁵C sites, trying to unveil the possible regulatory mechanisms behind these interactions. Lastly, we aim to elucidate how the function of UPF1 RBP is influenced by the methylation content of its target mRNAs, thereby understanding the regulatory role of m⁵C modification in the context of RNA metabolism.

3. Materials and Methods

3.1 Quality control and mapping of RNA BS-seq reads

For each public dataset, raw reads were subjected to FastQC (v0.11.9). Low-quality bases and adaptor sequences were removed using Trimmomatic (v0.38) with options ILLUMINCLIP:Adapter.fa:2:30:10:8:true LEADING:3 TRAILING:3 SLIDINGWINDOW: 4:20 MINLEN:50. We checked the conversion level for each dataset using spike-in sequence. Clean reads were mapped on the 45S rRNA, pre-tRNA, and predicted mature tRNA using meRanT tool (align bsRNA-seq reads to reference using Bowtie2 v2.3.5) in MeRanTK (v1.2.1b) with options (-k 10). Then, paired reads were obtained from unmapped reads using seqkit. Finally, unmapped reads pairs were mapped to the hg38 genome using the meRanGh tool (align bsRNA-seq reads to reference using HISAT2 v2.1.0) in MeRanTK with options (-fmo). Only uniquely mapped reads were retained and used for m⁵C sites detection.

3.2 m⁵C site calling and annotation for each dataset

The discovery of m⁵C sites for each dataset was performed similarly to Schumann et al, 2020. Read coverage at each cytosine position in the genome was obtained using the 'mpileup' function in samtools (v1.9). The C-to-T mismatches on forward reads and G-to-A mismatches on reverse reads were regarded as non-converted cytosines, and called using a custom script with parameters '-minBQ 30 --overhang 6', where reads were filtered by minimum base quality score 30 and removed terminal 6 nt to avoid overestimation of non-conversion. For ribosome profiling data, data from individual fraction libraries was combined with their respective biological replicate.

Reads containing more than three unconverted cytosines were considered as conversion failure and removed from the bam files ('3C filter') (Schumann et al, 2020). An icSHAPE value (L. Sun et al., 2021) was calculated for HeLa, HEK293 and HepG2 cells and used to evaluate RNA structuredness, with highly-structured regions having a value close to 0. Finally, candidate sites with a signal-to-noise ratio < 0.9 (3C/raw; 'S/N ≥ 0.9') were dropped to reduce false positives. To retain high-confidence non-conversion sites, the following criteria were applied: (1). Minimum total read coverage was set as 20 ('20RC'); (2). non-converted C ≤ 3 ('3C'); (3). C + T coverage ≥ 80% ('80CT'); (4). Non-conversion of ≥ 10% ('10MM'); (5). For replicates integration, we referred to Huang et al., (2019). We select m⁵C candidates detected in at least 2 replicates, and set non-converted C ≥ 5 for sites detected in only one replicate.

Candidate m⁵C sites were annotated to gene locus and transcript types according to the GENCODE v38 annotation (UCSC) using 'intersectBed' in bedtools and mapped to six features simultaneously: 5'UTR, CDS, 3'UTR, ncRNA_exonic, intronic and intergenic.

3.3 Distribution and enrichment of m⁵C along mRNA

Only m⁵C sites on exonic protein-coding transcripts were used for this part analysis. For the distribution analysis, mRNAs with m⁵C sites were divided into three segments (5'UTR, CDS, and 3'UTR). Each segment was normalized according to its average length. All C positions on the transcripts with m⁵C candidates were selected as background control. The relative position of each candidate site and background C in the corresponding segment were noted. We identified the relative position of m⁵C sites and background C around the start codon and stop codon separately to calculate the spatial enrichment for each bin.

3.4 NSUNs dependent m⁵C sites

We collected NSUN2 (Huang et al, 2019, and Yang et al, 2017) and NSUN6 depletion datasets (Liu et al., 2021a) for NSUN2/6-dependent m⁵C sites analysis. The analysis method was referred to Chen et al, 2019. In details, m⁵C sites which show reduction in m⁵C stoichiometry > 0.05 and a methylation level reduced to > 0.1 in cells depleted of either NSUN2 or NSUN6 cells were considered to be catalysed by NSUN2 or NSUN6, respectively. In particular, for NSUN2-dependent m⁵C we used the union set of Huang et al, 2019, and Yang et al, 2017. The five nucleotide sequence downstream of NSUNs-dependent m⁵C (positions 0 to 5, with m⁵C centred at 0) were used to calculate the position weight matrixes (PWM) of consensus motifs. NSUNs-dependence of m⁵C sites in union set can be further predicted by PWM through FIMO (Find Individual Motif Occurrences) in the MEME suite (v 5.4.1). For prediction of NSUN5-dependent candidate sites, we used NSUN5-overexpression and epigenetically silenced NSUN5 data in LN229 (Janin et al., 2019). Sites were predicted to be NSUN5-dependent when they showed a non-conversion level > 0.1 and an increase in m⁵C stoichiometry ≥ 0.05 upon NSUN5 overexpression. All the consensus motifs are plotted by ggseqlogo.

3.5 Comparison with RBP binding sites

Binding sites of RNA binding proteins (RBP) reported by Van Nostrand et al., (2020) were downloaded from ENCODE. The intersections of two biological replicates with fold-enrichment ≥ 4 and $p \leq 10^{-3}$, were filtered as significant peaks and the middle sites of the peaks were

regarded as RBP binding sites. Genes with both RBP binding sites and m⁵C sites were considered for enrichment analysis. Bins were divided around RBP binding sites with the number of m⁵C sites and total C within the bins. Then Fisher's exact test was applied to calculate the enrichment significance. We set up a gradient region as ± 20 nt, ± 30 nt, ± 50 nt, and ± 70 nt, to test the influence of bin size. Finally, ± 50 nt was used for further analysis. The relative distance of m⁵C sites to RBP binding sites was identified as well as the background C within the same region to get the distribution of m⁵C around RBP binding sites.

3.6 Cell culture and transfection

HeLa cells (human cervical cancer) were obtained from ATTC and confirmed with CellBank Australia. Cells were grown in DMEM medium (Gibco™) supplemented with 10% FBS and 1% antibiotic-antimycotic solution (Sigma) and passaged when 70–100% confluent.

For knockdown experiments, 150×10^3 cells were plated in 35mm plates and transfected 6–12hrs later with the siRNA against the target selected (FlexiTube siRNA UPF-1, Cat. No. GS5976; Qiagen) or the negative control (non-silencing siRNA, Cat. No. 1022076; Qiagen) with a final concentration 30nM, using 5 μ L of Lipofectamine RNAiMAX Reagent (Thermo Fisher Scientific) and 300 μ L of Opti-MEM (Thermo Fisher Scientific). The medium was replaced 12hrs later and cells were harvested 48hrs later.

3.7 Generation of KO cells

NSUN2 and NSUN5 KO HeLa cell lines were generated using Crispr/Cas9 technology.

3.7.1 Guide sequence design

NSUN2 and NSUN5 knockout HeLa cell lines were achieved using CRISPR-Cas9 technologies, via non-homologous end joining (NHEJ) mechanism, by generating large deletions at their genomic locations. Two crRNAs were designed for each gene, targeting the 5'- and 3'- regions. Briefly, crRNAs were designed by inputting target sequences from Ensembl^{ASIA} into CCTop (Stemmer et al., 2015) to check for efficacy and predict potential off-target cleavage effects. Sequences with high predicted efficacy, and minimal off-target effects were selected as crRNAs and ordered as Alt-R® CRISPR-Cas9 crRNA from Integrated DNA Technologies (IDT); sequences can be found in **Table 3**.

Gene	Exon	Sequence
NSUN2	Exon 2	AGGCUACCCCGAGAUCGUCA
NSUN2	Exon 19	AAUGAGAGUGCAGCCAGCAC
NSUN5	Exon 2	GAUGCUGUGAUCGCCAGCGC
NSUN5	Exon 10	CCGGGCUGACUCCUCCUGG

Table 3 - crRNA sequences used for large gene knockouts for NSUN2 and NSUN5.

3.7.2 Ribonuclear protein preparation

2.5 μ M of NSUN2 exon2 crRNA was combined with equimolar amounts of NSUN2 exon 19 crRNA. These crRNA were then ligated to 5 μ M Alt-R® CRISPR-Cas9 tracrRNA, ATTO™ 550 (IDT) in 10 μ L of IDT Duplex Buffer. The same method was used for NSUN5 gene knockout. Each ribonuclear protein (RNP) reaction consisted of 0.575 μ M of annealed crRNA:tracrRNA complexed with 30.5 pmol of IDT Alt-R® S.p. Cas9 Nuclease V3, in 2.2 μ L Invitrogen Neon™ Transfection System ‘R’ resuspension buffer. The crRNA:tracrRNA and Cas9 were complexed for at 37°C for 5 mins, and kept at room temperature hereafter.

3.7.3 Transfection

Electroporation was conducted using Invitrogen Neon™ Transfection System and following the manufacturer’s protocol, with the following modifications: HeLa cells were resuspended in Invitrogen Neon™ Transfection System ‘R’ resuspension buffer to a concentration of 2.8×10^7 /ml. For each electroporation reaction, 2×10^5 cells prepared above were incubated with 1xRNP at 37°C for 5 mins, before being electroporated at 1005volts, 35milliseconds with 2 pulses. Two reactions were seeded per well of a 24-well plate. Cells were recovered in complete medium at 37°C with 5% CO₂ for 24 - 36 hours.

3.7.4 Single cell sorting

Cells were sorted for singlets and ATTO™550 positivity on a BD FACSAria™ II Cell Sorter by the Flow Cytometry Facility at the John Curtin School of Medical Research, the Australian National University. Although all singlets were positive when compared with negative controls, only cells with high intensity ATTO™550 ($>10^{33}$) were sorted into 96-well plates for culture. Representative FACS plots are indicated in **Figure 3**. Cells were maintained in complete media and expanded to 6-well plates for genomic DNA (gDNA) extraction when they reached 70% confluency in smaller culture vessels.

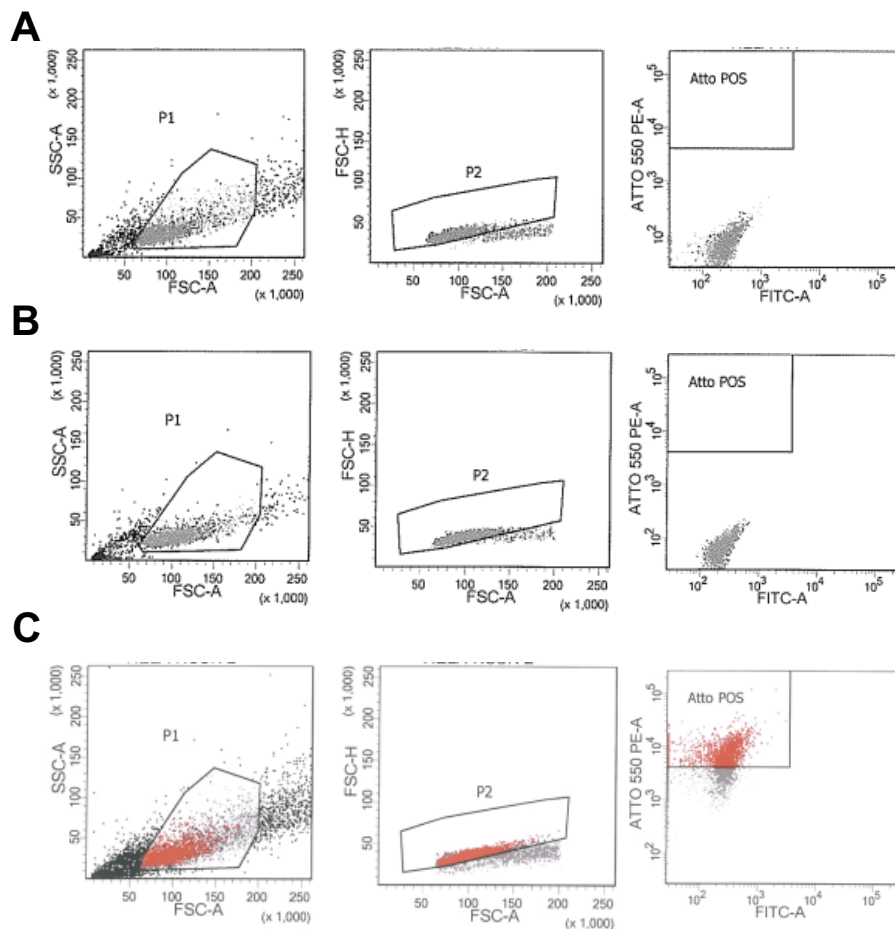


Figure 3 – Single cell sorting

Representative FACS plots for single cell sorting showing (A) negative control, no electroporation, (B) negative control with electroporation, no RNP, and (C) ATTOTM550 against FITC to determine auto-fluorescent cells; single cells were selected based on ATTOTM550 positivity above threshold of 10^{33} .

3.7.5 Sequencing

The gDNA was extracted by incubating cell pellets with 30 μ L of in-house rapid lysis buffer (40 μ g Proteinase K in Tris 10 mM, EDTA 1mM, 0.1% Tween-20, pH8.0) at 56°C for 1 hr followed by denaturation at 95°C for 10 mins. Amplification of NSUN2 and NSUN5 genes was conducted with standard protocols of 35 cycles in a Mastercycler® Nexus (Eppendorf) using Q5® High-Fidelity DNA Polymerase (New England BioLabs), 5 μ L of extracted gDNA in a PCR reaction and primer pairs listed in **Table 4**. Amplicons were purified with ExoSAP-IT (Applied BiosystemsTM) and sequenced on an AB 3730xl DNA Analyzer, by the ACRF Biomolecular Resource Facility (The John Curtin School of Medical Research, Australian

National University) following the manufacturer's protocol (Applied Biosystems 2002). Sequencing data was analysed manually using SnapGene software (from Insightful Science; available at snapgene.com). Large genomic deletions were detected using primer pairs NSUN2-F1/NSUN2-R2 and NSUN5-F1/NSUN5-R2. Sequences for the clones used in this paper are in **Figure 4**.

Gene	Primer Name	Sequence
NSUN2	NSUN2-F1	5'- TTCGCTGTTTTCCCTGTCAC -3'
	NSUN2-R1	5'- CGCCTTTAACCTCCAGCATC -3'
	NSUN2-F2	5'- GGAAACTTGAGGTGAGCTGC -3'
	NSUN2-R2	5'- GGCTCCTATGATCCCACCAG -3'
NSUN5	NSUN5-F1	5'- GTTTTGCCACAGAGCCGTAA -3'
	NSUN5-R1	5'- ACTTCCTTACAGTCTCCGCC -3'
	NSUN5-F2	5'- GGAGAATGAAGACGTGGTGC -3'
	NSUN5-R2	5'- TAGGATTGGAGGGTCTGGGT -3'

Table 4 - Primer pairs used to amplify and sequence genomic deletions in NSUN2^{-/-} and NSUN5^{-/-}.

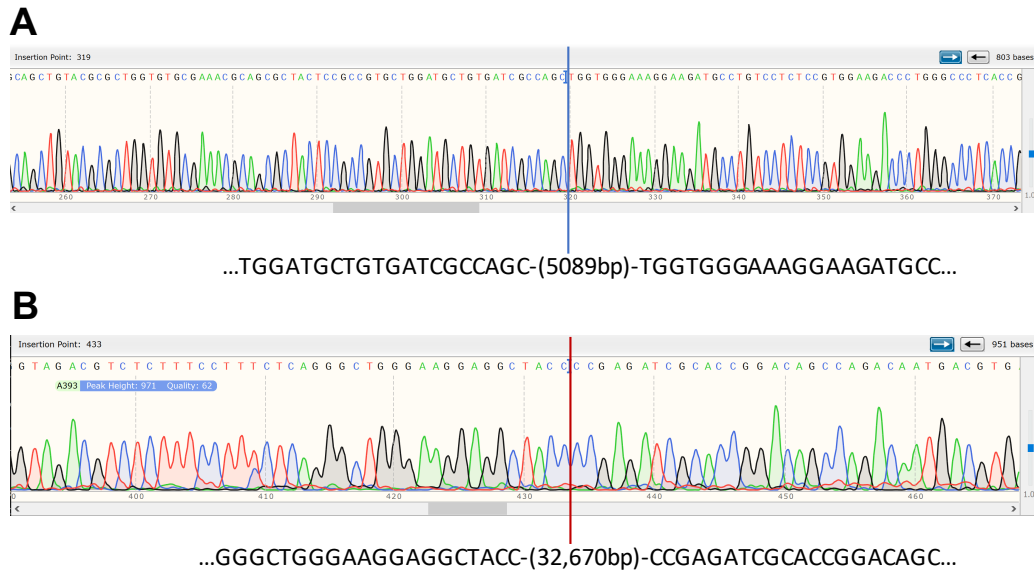


Figure 4 – Large deletion in NSUN2 and NSUN5 KO cell lines

Chromatograms for **(A)** NSUN5^{-/-} showing homology repair site with deletion of 5089 bp, demarcated by the blue line, and **(B)** NSUN2^{-/-} homology repair site with deletion of 32,670 bp demarcated by the red line.

3.8 Protein Analysis

Cells were harvested with 200-500 μ L of Protein Extraction Buffer (50 mM Tris pH7.5, 5 mM EDTA, 150 mM NaCl, 21.5 mM MgCl₂, 10% glycerol, 1% Triton, 1X PIC [cOmplete™, EDTA-free Protease Inhibitor Cocktail, Sigma]) and incubated 10min on ice, then incubated on a rotator for 30min at 4°C and centrifuged at 13000xrpm for 10min at 4°C.

The supernatant was transferred to a clean tube, used or stored at -80°C. Total protein concentration was measured through the Qubit™ Protein Assay Kit (Thermo Fisher Scientific) following manufacturer's instructions. 30 μ g of proteins were loaded on NuPage™ 4–12% Bis-Tris Protein Gels (Invitrogen) followed by transfer onto nitrocellulose or PVDF (for IR-Dye detection) membrane. The membrane was blocked in Odyssey Blocking Buffer (for IR-Dye detection; LI-COR 927-40000) and probed with primary antibodies: (anti-NSUN2 (1:1000; Proteintech 20854-1-AP), anti-NSUN5 (1:1000; Proteintech 15449-1-AP), anti-UPF1 (1:1000; Abcam ab86057), anti-ACTB (1:1000; SantaCruz sc-4778 AF790). The membranes were probed with a secondary antibody, either anti-mouse-IR-Dye800 (1:10,000; LI-COR 926-32210) or anti-rabbit-IR-Dye680 (1:10,000; LI-COR 925-68071), and scanned using the Odyssey® CLx Imaging System (LI-COR).

3.9 RNA Analysis

Extraction of total RNA was performed using the Direct-zol RNA Miniprep (Zymo Research) kit according to the manufacturer's instructions. Reverse transcription reactions were performed with PrimeScript RT Master Mix (Takara Bio) on 0.5-1 µg of total RNA in a 10 µL reaction. RT-qPCR analyses were performed with 15-20 ng equivalent of cDNA, 5 µL of 2X SYBR Mastermix (QIAGEN), 1 µL of 5 µM primers and water to a final volume of 10 µL. DNA amplification (followed by melting curve analysis) was monitored with a QuantStudio 12k Flex Realtime PCR instrument. Primers' sequence used for RT-qPCR are listed in **Table 5**.

Relative RNA quantity was calculated as the fold change ($\Delta\Delta C_t$) with respect to the experimental control sample set as 1 and normalised over ACTB. The ratio of each sample versus its experimental control was tested by two-tailed Student's t-test. One asterisk indicates a p-value < 0.05, two asterisks indicate a p-value < 0.01, and three asterisks a p-value < 0.001.

Gene	Primer	Sequence
FADS2	FW	CAACAAGGTGATCGAAGGCC
	RV	ACACCAGCTCCACCTCAAAC
FKBP10	FW	CAACAAGGTGATCGAAGGCC
	RV	ACACCAGCTCCACCTCAAAC
NIBAN2	FW	AATACGACTACGACAGCAG
	RV	GTCCTCGAAGATCAGCTC
OSMR	FW	TTCATATCCTACTGCCCATGG
	RV	TTATTAATGACAGGATGCTGCTC
PEA15	FW	ACATCCCCAGCGAAAAGAG
	RV	ACCATAGTGAGTAGGTCAGGAC
DNAL4	FW	TGGAGCTATGTGTACAGC
	RV	TCTCAAACCCAAAGCCCTC
ITM2B	FW	ACATCAAGGCTGGAACCTATTTG
	RV	TGCGAAACAATTGCTGGCTT
ACTB	FW	CTGGAACGGTGAAGGTGACA
	RV	AAGGGACTTCCTGTAACAATGCA
SPIKE-IN	FW	CGGCAAAAGCACTCTGATTGAC
	RV	CCTTGTCGTATCCCTGGAAGATG

Table 5 - Primers used for RT-qPCR

3.10 Bisulfite conversion and amplicon-bsRNA sequencing

RNA extracted from three biological replicates of WT and NSUN5 KO HeLa cells was treated with bisulfite conversion as described in Sibbritt et al., (2022). Bisulfite-converted samples

were then reverse-transcribed using SuperScript™ IV Reverse Transcriptase (Invitrogen) with random hexamers and subjected to amplicon-specific PCR amplification. PCR conditions were optimised for each primer set and carried out in technical triplicates (bs-converted primer sequences are shown in **Table 6**, below). The obtained products were analysed by 2% agarose gel electrophoresis, purified using the MinElute® PCR Purification Kit (Qiagen) and the technical replicates combined. PCR products from the same RNA sample were pooled and subjected to library preparation using the TruSeq® DNA Nano Library Kit (Illumina) according to the manufacturer's instructions. Libraries were mixed with 40% PhiX Control Library (Illumina) and sequenced on the MiSeq System (Illumina), acquiring 151bp read length.

Gene	Primer	Sequence
ALG8	FW	TTTTTTTTTTGTATGGTTTGAG
	RV	AATGGTTAAAGTAAAATAATTC
NUP133	FW	GTATTTAGTAGATAATTTATGTT
	RV	TTAAATAACTTCTACAATAACTA
PPP2R5D	FW	ATAAATTGAAAAAGGAGAAGGAG
	RV	TTCTTCACCTCCACCATCCT
PAAF	FW	TTTAGTAAGATTTAGATTATG
	RV	CATCTCAACAACATATATAA
CRY	FW	ATATATTTATGATTTTTTGAATG
	RV	TTCAAACAACCTACCTCAA
CORO1C	FW	AATTAAATGAAGATAAGATATTA
	RV	ATACAATCATCAAAATAACCA
AJUBA	FW	GTTGTTGTTGTTTTTTTTTGG
	RV	CTACAAAATAACAACAACAAC
PIK3R3	FW	GTTATTTGTGTGGTTTGATAG
	RV	CTATACCATTTCACAAACT
28S rRNA (C:3782)	FW	GTGATTTTTGTTTAGTGTGTTTGAATGTT
	RV	TTCATCCATTTCATACACATCAC
tRNA ^{Thr}	FW	GTTTTATAGTTTAGGGGTTA
	RV	CCCCAACAAAAATTAAACTC
tRNA ^{Gly}	FW	GTATGGGTGGTTTAGTGGTA
	RV	CATAAACCAAAAATCAAACCC

Table 6 - Primers used for amplicon-bsRNA sequencing.

3.11 miSeq analysis

The target RNA regions of all amplicons were combined into a single reference and the forward strand C-to-T converted. Reads from 'in-depth' sequencing were subjected to FastQC (v0.11.5); adapter removal and low-quality base trimming was performed using cutadapt (v1.18) with options (-q 20,20 -m 50 --trim-n -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT). Clean reads were aligned to the reference using the MeRanT tool in meRanTK in both directions, as sequencing was not strand-specific, and the resulting bam files merged. Non-conversion level was determined for each C position within the amplicons.

3.12 UPF1 knockdown RNA-seq and differential expression analyses

The raw reads were subjected to FastQC (v0.11.9). Low-quality bases and adaptor sequences were removed using Trimmomatic (v0.38) with options (ILLUMINCLIP:Adapter.fa:2:30:10:8:true LEADING:3 TRAILING:3 SLIDINGWINDOW: 4:20 MINLEN:50). Clean reads were mapped to 4S rRNA using bowtie2 (v2.3.5) with parameters ('-q --sensitive --reorder --no-unal --un-conc-gz'). Unmapped reads were mapped to hg38 genome using HISAT2 (v2.1.0) with parameters ('--no-softclip --score-min L,-16,0 --mp 7,7 --rfg 0,7 --rdg 0,7 --max-seeds 20 -k5 --dta'). The mapped reads were processed to featureCounts (v2.0.1) for feature counting.

To quantify RNA, fragment abundance in genes with ≥ 50 mapped reads were selected and normalized using the size factors estimated by the median of all genes implemented in the DESeq2 (v1.30.1) Bioconductor package. Differential expression analysis was performed by DESeq2. Genes with $\log_2(\text{fold change}) \geq 1.2$ and adjusted p-value ≤ 0.05 were regarded as differentially expressed genes.

3.14 Cross-linking immunoprecipitation

1 million HeLa cells were washed twice with 1X PBS and irradiated once at $150 \text{ mJ} \cdot \text{cm}^{-2}$ at 254 nm using the Stratalinker® UV crosslinker. Cells were then scraped in CLIP Lysis Buffer [50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1% Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate, 1X PIC (cOmplete™, EDTA-free Protease Inhibitor Cocktail, Sigma), and RNasin® Plus Ribonuclease Inhibitor (Promega)], incubated on ice for 5 min and passed through a 21G needle. Lysates were treated with 3 μL of TURBO™ DNase (Thermo Fisher Scientific) for 3 min at 37 °C while shaking at 1,100 rpm. Samples were then centrifuged for

10 min at 17,000 g at 4 °C to clear the lysate. The supernatant was collected and quantified using Qubit™ Protein Assay Kit (Thermo Fisher Scientific) following manufacturer's instructions.

50 µL of Dynabeads™ Protein G (Thermo Fisher Scientific) were previously washed twice in CLIP Lysis Buffer and then resuspended in the same buffer with 9 µg of anti-UPF1 goat polyclonal antibody (Cat. No. A300-038A), or 9 µg of Abcam mouse IgG2a Isotype Control antibody (Cat. No. ab18413) in a final volume of 100 µL and incubated in rotation at room temperature for 1 hr for antibody conjugation. The conjugated beads were then incubated overnight on a rotator at 4°C with 500 µg of lysate. 50 µg of lysates was also rotated to be used as inputs. The next day beads were washed four times in High-Salt Buffer (50 mM Tris-HCl, pH 7.4, 1 M NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate) and twice in PK buffer (100 mM Tris-HCl pH 7.4, 50 mM NaCl, 10 mM EDTA). Beads were resuspended in 50 µL of PK buffer, of which 10 µL were used for Western Blotting analysis and 40 µL were digested with Proteinase K Solution (Thermo Fisher Scientific, Cat. No. 25530049) in shaking at 1,100 rpm for 30 min at 37 °C to RNA. RNA was purified by phenol/chloroform extraction and precipitated for further analysis.

4. Results

Note: Part of the findings described in this thesis have also been published in (Guarnacci et al., 2024). Bioinformatic analyses in this thesis were performed by Pei-Hong Zhang.

4.1 Generation of a union list of m⁵C sites in human transcriptomes

The speed at which epitranscriptomic research progresses is often determined by the availability of methods available for a modification detection and quantification. This is particularly true for m⁵C. Despite early evidence of m⁵C in mammalian mRNA surfaces in the 1970's (Dubin & Taylor, 1975), further research in this field remained dormant during the subsequent decades. However, a turning point in the field was coupling bisulfite conversion of RNA (Schaefer et al., 2009) with high-throughput sequencing about ten years ago (Squires et al., 2012). The establishment of this new method for m⁵C detection and quantification marked the beginning of the rediscovery of this modification and paved the way towards a decade of fruitful research. Initially, ~10,000 m⁵C sites were reported for both human coding and non-coding genes (Squires et al., 2012), followed by subsequent studies that reported a similar number of sites (Amort et al., 2017; X. Yang et al., 2017a). As research exploration of m⁵C progressed, it became apparent that these numbers for individual cell types were likely overestimates. This inconsistency could be due to an incomplete bisulfite conversion on highly structured parts of RNA (Huang et al, 2019; Schumann et al, 2020). More recently, thanks to a more refined and sophisticated pipeline for m⁵C identification, site calling and more stringent set of filters to avoid false positives, it was possible to confirm the existence of < 1,000 sites per cell type, the majority of which had a stoichiometry lower than 20% (Huang et al, 2019; Schumann et al, 2020).

Presently, bsRNA-seq experiments have been carried out in different human cell types and tissues (**Table 7**). This generation of m⁵C sites lists, although prolific, led to the publication of datasets that are often different from each other, in terms of both number of m⁵C sites and their stoichiometry. Different factors should be considered to elucidate this disparity, such as (a) the possibility for differential methylation patterns in different cell types, (b) the variation in the expression levels of the writer enzymes (NSUN proteins), and (c) technical differences in data analysis pipelines.

One of the first aims of this thesis work was work towards enhancing coherence, we embarked on re-examining and re-analyse publicly available datasets with a stringent and unified pipeline for m⁵C site calling and filtering. This yielded a union set of high confidence m⁵C sites.

Upon perusing the literature around the biology of m⁵C modification on RNA, several publications were identified in which the abundance and localisation of m⁵C was assessed transcriptome wide (**Table 7**). We were particularly interested in bisulfite-converted RNA sequencing (bsRNA-Seq) data generated from human cells.

BsRNA-Seq datasets from 7 human tissues (brain, liver, lung, heart, spleen and kidney) and 5 human cell lines (LN229, HepG2, HeLa, HEK293 and T24) were collected from 6 published studies (Chen et al, 2019b; Huang et al, 2019; Janin et al, 2019; Liu et al, 2021; Schumann et al, 2020; Yang et al, 2017) and selected for re-analysis. These datasets are typically replicated and include several conditions in which one or more NSUN protein was altered, either via knockdown or knockout (NSUN2 and NSUN6) or via epigenetic silencing and subsequent overexpression (NSUN5). These conditions, in conjunction with WT or untreated samples, allowed us to confirm and gain further knowledge around the enzymology of specific methyltransferase enzymes.

Human Bisulfite RNA-sequencing data used in the study		
* Samples used for the generation of the union list		
Sample	# of replicates	Studies
Heart	2*	Huang <i>et al.</i> , 2019
Liver	2*	
Lung	2*	
Muscle	2*	
Spleen	2*	
Testis	2*	
Frontal Cortex	2*	
HeLa	4 (2x WT*, 2x NSUN2 KD)	Yang <i>et al.</i> , 2017
	4 (2x WT*, 2x NSUN2 KO, 2x NSUN6 KO)	Huang <i>et al.</i> , 2019
	12* (ribosome profiling fractions)	Schumann <i>et al.</i> , 2020
HEK293T	1*	Huang <i>et al.</i> , 2019
	4 (2x WT*, 2x NSUN6 KO)	Liu <i>et al.</i> , 2021
LN229	6 (3x epigenetically silenced NSUN5* and 3x NSUN5 overexpression)	Janin <i>et al.</i> , 2019
T24	1*	Chen <i>et al.</i> , 2019
HepG2	2*	Liu <i>et al.</i> , 2021

Table 7 – Collection of human bsRNA sequencing datasets used in this thesis.

4.1.1 Stringent and uniform re-analysis of bsRNA-Seq data

The strategy we decided to adopt for the generation of our union set of human m⁵C sites was to re-analyse each independent sample individually, from reads quality control to m⁵C calling and filtering, using the same stringent pipeline and only at the end pull together all the high-confidence sites into one union list.

The pipeline used for this analysis is based on Schumann et al., (2020), with minor modifications (described in details in the method section [3.2](#)). Briefly, after data pre-processing and mapping to the C-to-T and G-to-A converted genome (plus spike-in sequences as appropriate), non-conversion sites were called as variations by a custom script, e.g., T-to-C in read2 and A-to-G in read1 (**Fig. 5A**). Read distribution and overall cytosine conversion rate were checked to assess quality of the 50 datasets. For seven of the primary tissue samples, one of the two replicates was excluded as metagene analysis showed strongly biased coverage towards mRNA 3' ends suggesting excessive degradation (**Fig. 5B**). The remaining 43 datasets each showed high cytosine conversion rates, e.g., $\geq 99.7\%$ conversion rate for all annotated genes, $\geq 99.8\%$ for protein coding genes and RNA spike-ins (depending on sample, ERCC, *dhfr* or R-Luc, **Fig. 6**, top panel) all consistent with the original reports, e.g. Schumann et al., (2020).

A set of filters were then applied to identify high confidence m⁵C candidate sites: reads with more than 3 non-converted cytosines were discarded via a custom script to remove clustered non-conversion likely caused by RNA secondary structure (3C filter) (**Fig 6**, bottom panel). Furthermore, we also excluded non-converted cytosine positions for which less than 90% of reads passed the 3C filter (S/N90 filter). We then retained 'C' positions with ≥ 20 reads coverage (20RC filter), at ≥ 3 reads supporting the non-conversion, and with a non-conversion ratio $\geq 10\%$ (10MM). Sites that passed all filters in at least 2 biological replicates were called as high confidence m⁵C sites. Sites passing the filtering steps in only one of the replicates had to satisfy a higher threshold of ≥ 5 reads supporting non-conversion.

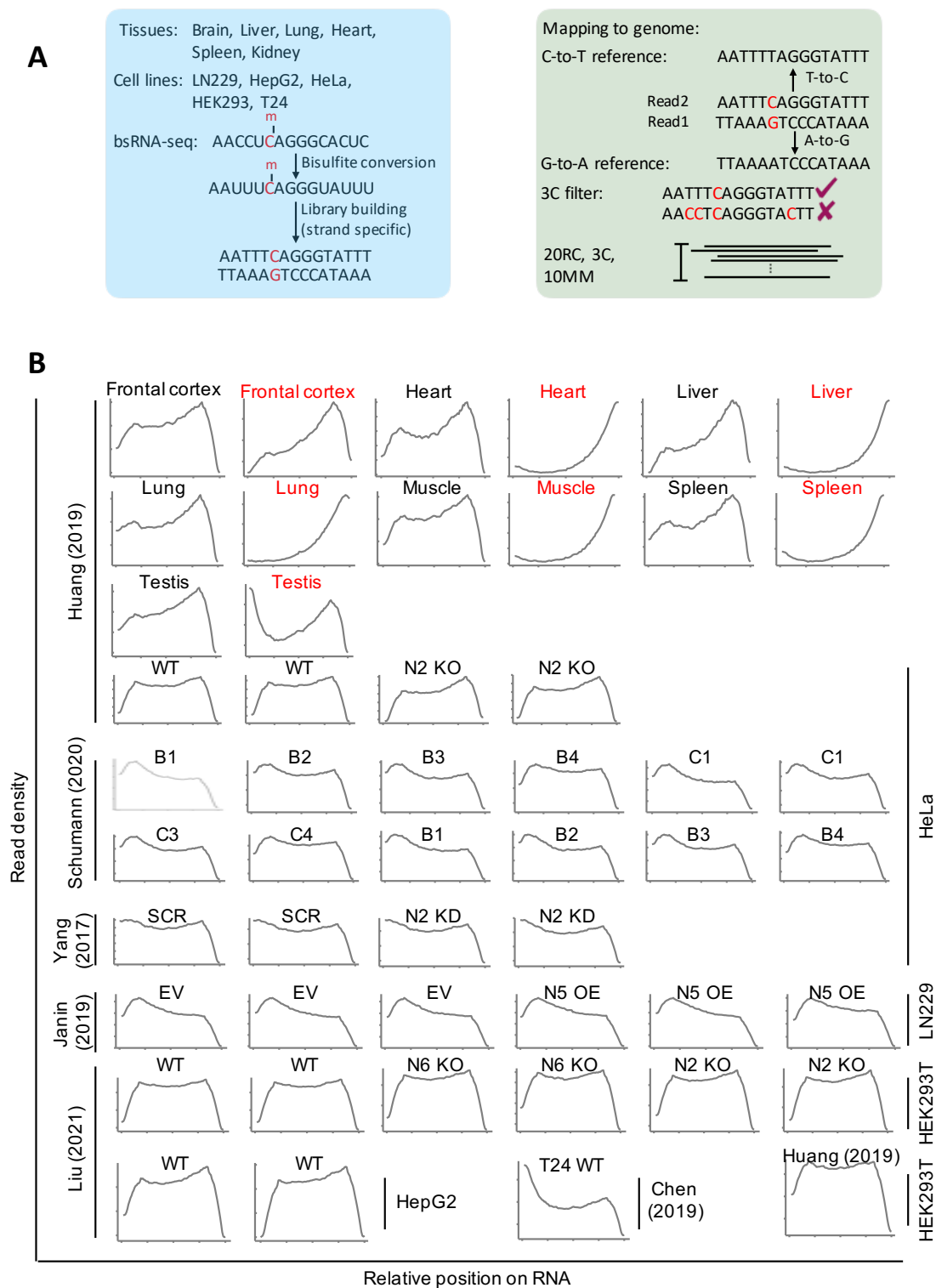


Figure 5 - (A) Schematic summary of the strategy used for the generation of union set of human m⁵C sites: selection of publicly available bsRNA-Seq data and high confidence m⁵C sites calling with stringent unified pipeline. **(B)** Metagene distribution of read coverage of bisulfite-treated RNA samples. Seven samples (in red) showed a strong bias in read coverage and were therefore excluded from further analyses.

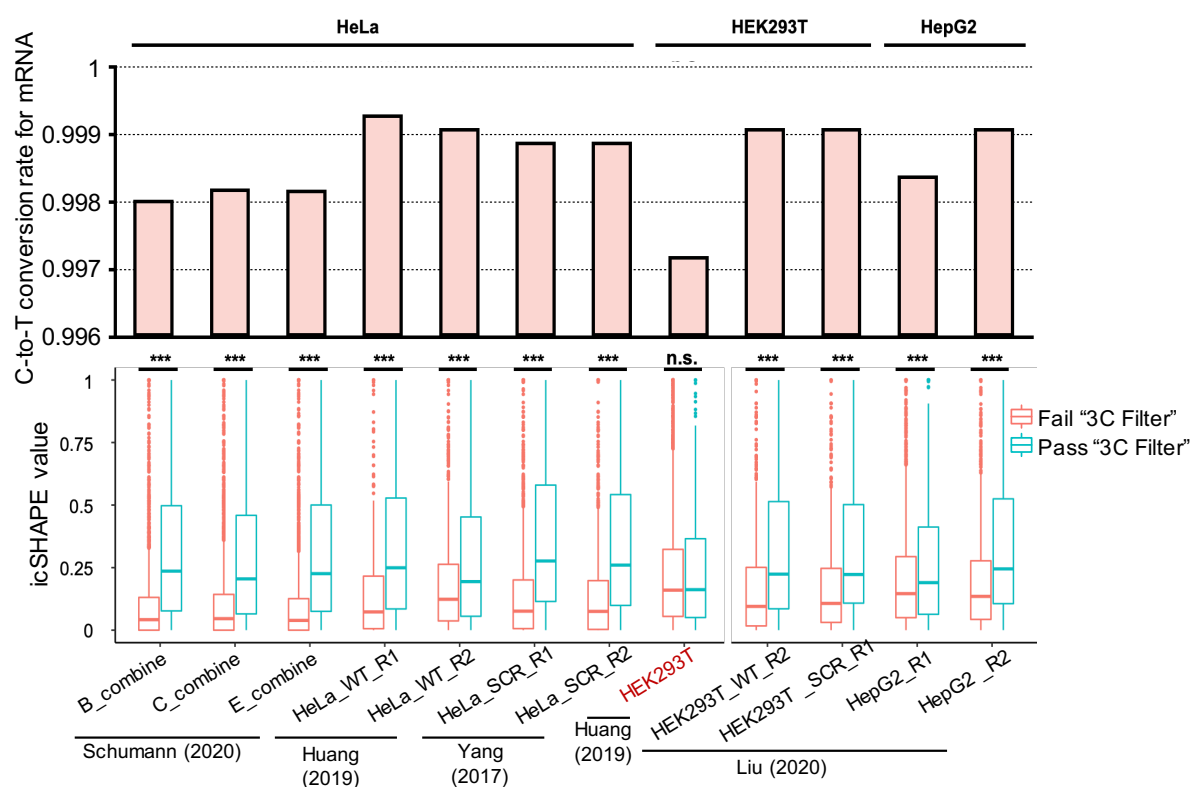


Figure 6 - the global conversion ratio of these datasets (top); icSHAPE value of genome sites that passed “3C filter” and failed “3C filter (bottom).

4.1.2 A union set of 6,393 human m⁵C sites

After merging called sites from all datasets of non-manipulated NSUN gene expression, a collective compilation union set of 6,393 m⁵C sites was obtained. Upon examining the overlaps among these collective sites across distinct samples, we observed a modest level of overlap, in line with our expectations. (**Fig. 7A**). There are several plausible reasons as to why that is the case for starters, the samples used for compiling the union list are diverse. This diversity leads to variation in the expression levels of m⁵C methyltransferase enzymes like NSUN2 and NSUN6 that target mRNA. This diversity is substantially pronounced across different tissues and cell lines (**Fig. 7B**). Read coverage also varies strongly between datasets for many sites (**Fig. 7C**), likely due to biological and technical reasons such as differences in gene expression profile between samples and depth of library sequencing. Moreover, as most called sites show a non-conversion rate just above the, although commonly used, empirically chosen $\geq 10\%$ cut-off, there is a considerable random fluctuation as to whether a site is called as 'present' or 'absent' in a given sample. Therefore, when comparing any two samples, the majority m⁵C sites that are not shared, either did not meet the ≥ 20 read coverage threshold or exhibited stoichiometry below threshold (e.g., between 1%-10%) in one of the two samples (**Fig. 7D and 7E**).

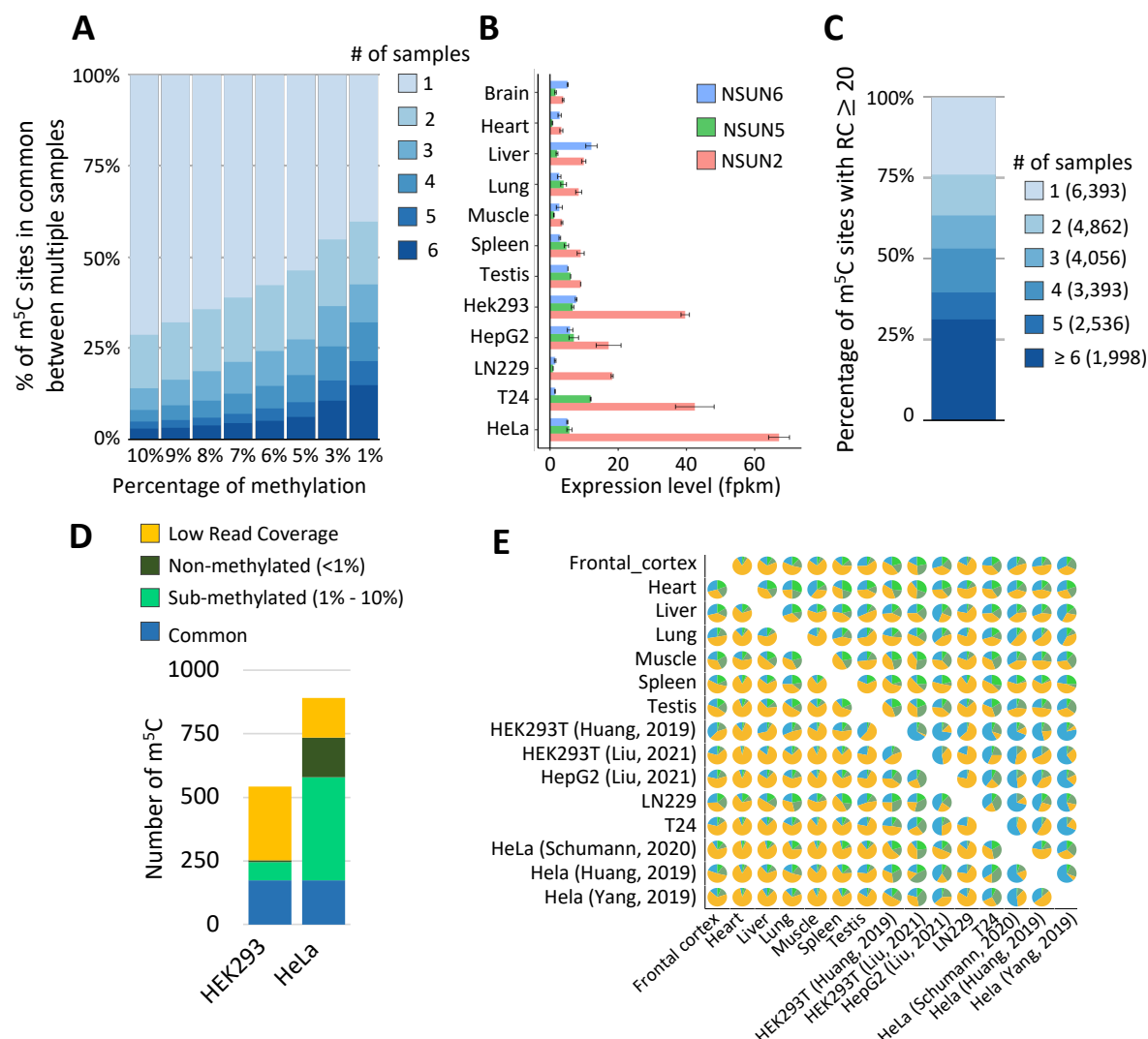


Figure 7 – (A) Percentage of the 6,393 m⁵C sites that are in common between 2 or more datasets re-analysed for the generation of the union set. The x-axis refers to the minimum methylation level of a given candidate site to be called as methylated. (B) Expression level of NSUN2, NSUN5 and NSUN6 in multiple tissues/cell lines. (C) Overlaps of sites covered by more than 20 reads. (D) The expression levels and methylation rates of m⁵Cs identified as unique to HEK293 (Liu et al, 2021) or HeLa (Huang et al, 2019). (E) Paired-comparison shows the expression level and methylation level of m⁵Cs in each sample (in row) when they are calculated in another sample (in column), the tissue/cell specific m⁵Cs are classified into 3 groups: Low read coverage (RC) in another sample, indicate differential expression gene; Sub-stoichiometric, which means the non-conversion level of putative m⁵C in bsRNA-seq is lower than 10% but higher than 1%, and pass binomial test ($p < 0.05$); Non-methylated, represent sites have lower than 1% non-conversion level or couldn't pass binomial test. Pie charts in the red box show a higher non-methylation ratio.

4.1.3 m⁵C distribution in the transcriptome

When looking at the distribution of the union sites across the human transcriptome, we found that a large majority of m⁵C sites are deposited on protein coding genes, in both coding sequence as well as untranslated regions (**Fig. 8A**). Moreover, these sites appear to be distributed nonrandomly along mature mRNAs. In particular, we found a marked peak of methylation surrounding the start codon (**Fig. 8B**). This is a pattern that was repeatedly found by others in single source material in published work, such as (Chen et al, 2019b; Huang et al, 2019; Schumann et al, 2020; Yang et al, 2017). This phenomenon was also confirmed when investigating methylation hotspots in individual samples (**Fig. 8C**), which allows us to support the previously established link between m⁵C content of mRNA and translation regulation.

Overall, we have established utility of our union set as a larger yet still prototypical list of human transcriptomic sites across multiple cell types for further exploration of m⁵C enzymology and function.

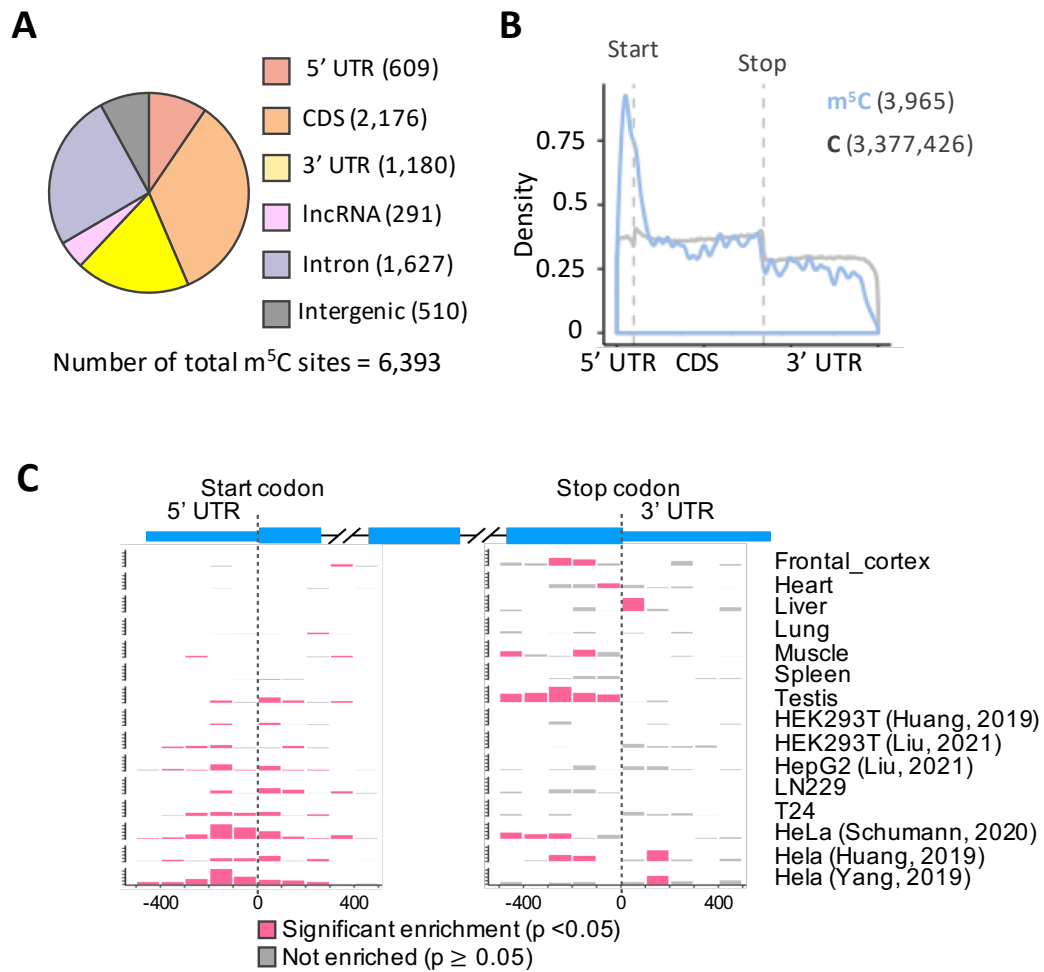


Figure 8 - (A) Distribution of m⁵C sites across the transcriptome **(B)** Distribution of m⁵C sites along mRNA. mRNA regions were scaled on the mean of their length. m⁵C distribution is shown in blue, background cytosine distribution in grey. **(C)** Metagene enrichment around start codon and stop codon of m⁵C in multiple datasets.

4.2 The spectrum of mRNA m⁵C writer enzymes

Several proteins are known to have m⁵C methyltransferase activity on RNA, particularly NSUN2 and NSUN6 are both known to target mRNA in addition to their canonical tRNA substrate (Liu et al, 2021; Schumann et al, 2020; Yang et al, 2017; Huang et al, 2019). Therefore, we decided to extend our analysis to several of those bsRNA-Seq dataset in which one or more NSUN enzyme was altered.

4.2.1 NSUN2- and NSUN6-dependent m⁵C sites

The pool of datasets we collected for our analysis includes samples from (a) cell lines depleted for NSUN2, such as NSUN2 knockdown (KD) and knockout (KO) HeLA cells from (Huang et al, 2019; Yang et al, 2017), and (b) NSUN6 knockout (KO) HEK293T cells from (Liu et al, 2021).

In order to assess the extent of these two enzymes in the deposition of m⁵C on human mRNA, we first decided to compare the methylation level of mRNA in the aforementioned (a and b) NSUN2 and NSUN6 KD/KO samples to the corresponding controls within the same cell line. Using this approach, we classed 1,368 NSUN2-dependent and 275 NSUN6-dependent m⁵C sites, based on experimental evidence (**Fig. 9A**). Both sets show typical sequence and secondary structural features known to be associated with their respective NSUN enzyme (Liu et al, 2021). NSUN2-dependent m⁵C sites are of “type I”, characterized by a downstream G-rich triplet at the 5’ end of a hairpin, also found in other tRNA sites methylated by NSUN2 (**Fig. 9B, 9C – top**). NSUN6-dependent m⁵C sites are instead of “type II”, located in the loop of a hairpin followed by a downstream “UCCA” motif as found at the 3’ end of tRNAs (**Fig. 9B, 9C – middle**). The remaining sites (44) that could not be attributed to either NSUN2 or NSUN6 showed no strong discernible commonalities (**Fig. 9B, 9C – bottom**).

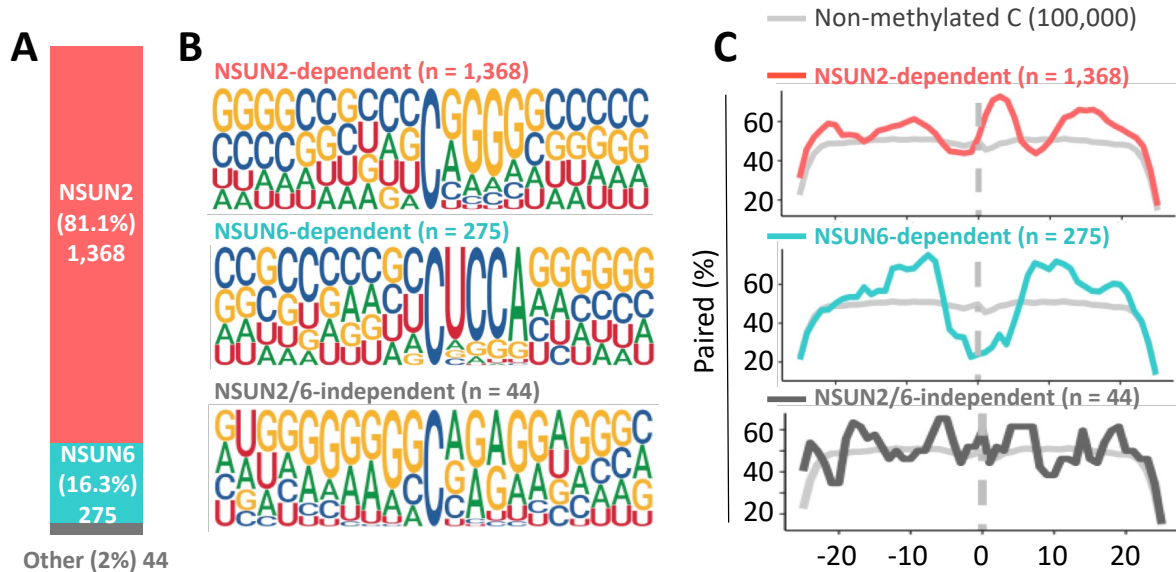


Figure 9 - (A) Composition of NSUNs-dependent m⁵C sites from experimentally validated data. (B) and (C) Consensus motifs and base-pairing propensity meta-profile of regions surrounding different NSUNs-dependent m⁵C sites, separately (NSUN2 in pink, NSUN6 in turquoise, non-methylated C in grey).

Our next step was to extend this type of classification to our union set. As there are no available NSUN KO/KD data for many human cell lines, we decided to develop a way to make a prediction on the human transcriptome based on the available data in HeLa and HEK293T cells. Using the sequence characteristics at position 0 to 5, downstream of these experimentally identified NSUN2 and NSUN6 targets, we calculated the position weight matrix (PWM) of their consensus motifs, which allowed us to predict the enzymes responsible for the deposition of m⁵C sites in our union set (see method in 3.4). Such an approach revealed that in the wider human transcriptome around two-thirds of m⁵C sites are likely deposited by NSUN2 (4,281), while one-sixth are predicted to be deposited by NSUN6 (1,037) (**Fig. 10A**). In addition, these sites still comply with type I and type II characteristics, both in terms of sequence motifs (**Fig. 10B**) and secondary structure (**Fig. 10C**).

Around 1,000 of our sites (~16%) were found to be independent of either NSUN2 or NSUN6 and therefore left unassigned. This confirms our general expectations that NSUN2 and -6 are the main m⁵C writers in mRNA, but also leaves scope for other enzymes to be responsible for the deposition of the remain one-sixth of unassigned m⁵C sites on mRNA.

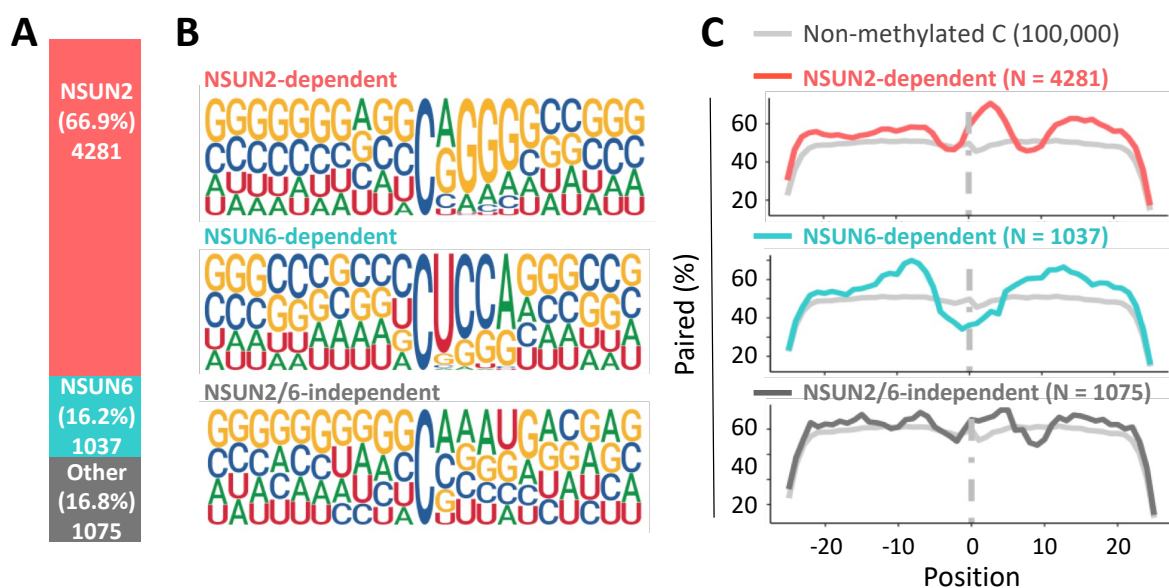


Figure 10 - (A) Composition of NSUNs-dependent m⁵C sites from union set. (B) and (C) Consensus motifs and base-pairing propensity meta-profile of regions surrounding different NSUNs-dependent m⁵C sites from our union set, separately. (NSUN2 in pink, NSUN6 in turquoise, non-methylated C in grey).

4.2.2 NSUN5 mediates mRNA methylation

m⁵C methyltransferase enzymes responsible for mRNA methylation, such as NSUN2 and NSUN6, were originally identified as writer proteins for tRNA (Frye & Watt, 2006; Haag et al., 2015). We therefore wondered whether these ~1,000 unassigned m⁵C sites we found on mRNA could be attributed to another enzyme from the NSUN family of proteins that is already known for having methyltransferase activity on other RNA species.

This prompted our decision to investigate the potential role of NSUN5 protein in methylating mRNA. NSUN5 has already an established RNA methyltransferase activity on ribosomal RNA, which in human targets position C:3782 of 28S rRNA (Janin et al., 2019). The methylation content of this site is well known to be conserved also across species, such as *Drosophila melanogaster*, *C. elegans* and *Saccharomyces cerevisiae* (Schosserer et al., 2015). In order to assess if NSUN5 deposits m⁵C modification on mRNA, we decided to re-analyse publicly available bsRNA-seq data from cell lines where NSUN5 was either epigenetically silenced (hereafter referred to as “control”) or overexpressed (OE) (Janin et al., 2019).

Using our custom pipeline, we observed a disparity in the count of m⁵C sites under two distinct conditions. Within the control sample we identified 446 mRNA sites in the and upon the overexpression (OE) of NSUN5 3,921 sites identified (**Fig. 11A**). On investigating the sequence motifs that were enriched in the different conditions, we found an interesting variation (**Fig. 11B**). In the control condition the downstream G-rich triplet typical of NSUN2-mediated sites were found to be the most abundant, and the sites were mainly enriched around the start codon (**Fig. 11B**, left panel). Conversely, In NSUN5 OE condition, we found a GUXGCm⁵CAXXU motif to be enriched, with a different distribution pattern along mRNA, that was more scattered throughout the coding sequence (**Fig. 11B**, right panel). This specific sequence motif we found in NUSN5 OE samples clearly resembles the sequence surrounding the m⁵C site found on human 28S rRNA (C:3782) mediated by NSUN5 (**Fig. 11C**), also conserved in other species such as *M. musculus* and *S. cerevisiae* (Heissenberger et al., 2019).

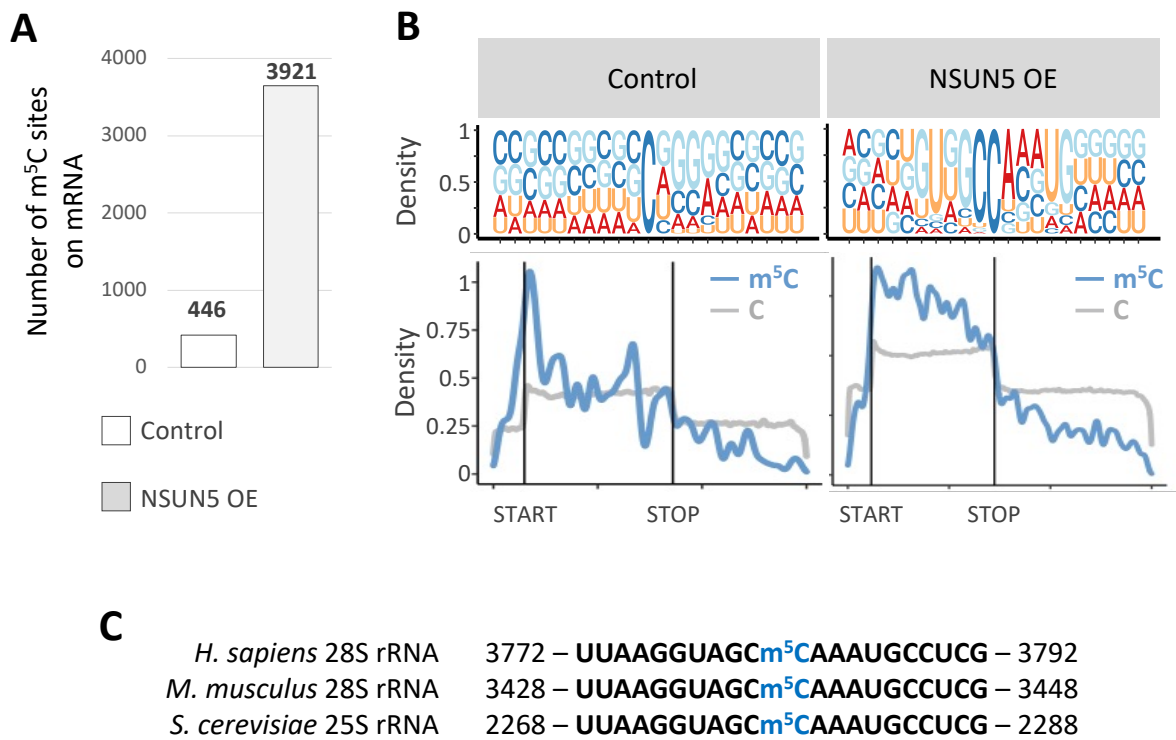


Figure 11 - (A) Number of m⁵C found on mRNA in epigenetically silenced NSUN5 (Control) and NSUN5 overexpression (OE) LN229 cells. (B) Consensus motif surrounding m⁵C sites in epigenetically silenced NSUN5 (Control) and NSUN5 OE LN229 cells (top) and their metagene distribution over mRNA (bottom). (C) Alignment of rRNA sequence carrying an m⁵C modification (in blue) across three different species.

This publicly available dataset we used is missing a WT or untreated sample, therefore it is impossible to make a direct comparison with the samples with unaltered NSUN5 expression level (and methyltransferase activity). This complicates the direct attribution of these m⁵C sites solely to NSUN5. However, we made additional considerations as outlined below.

When we clustered these sites depending on their sequence motif, we could see how those sites that can be clearly attributed to either NSUN2 or NSUN6 did not vary between control and NSUN5 OE conditions. On the other end, sites that had a strong increase of methylation upon NSUN5 OE were observed to have an enriched sequence motif very similar to the NSUN5-dependent C:3782, despite no difference in read coverage (**Fig. 12A**). In addition, these same sites were found to have a specific pattern of secondary structure prediction (**Fig. 12B**), which resembles the immediate structural context surrounding NSUN5-dependent C:3782 in 28S rRNA (**Fig. 12C**).

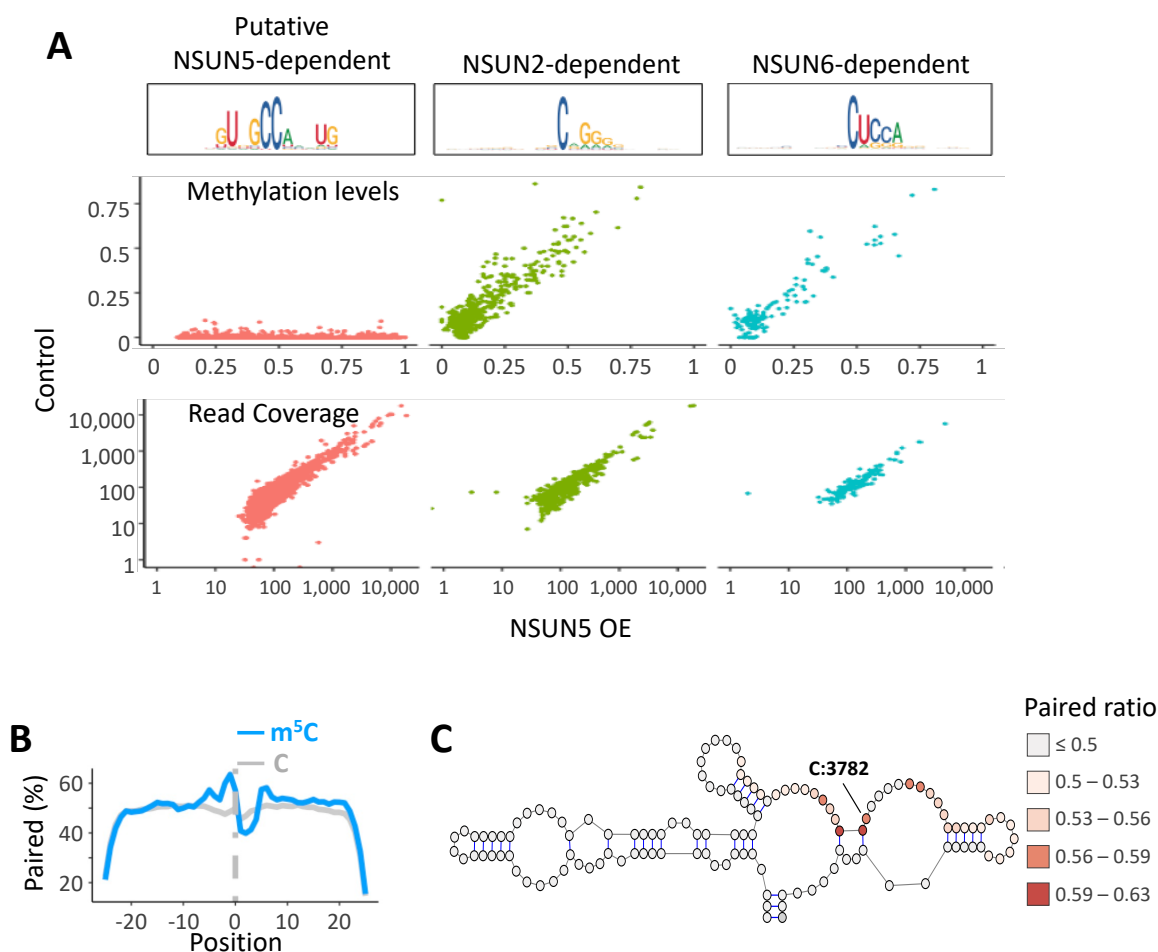


Figure 12 - (A) Metagene distribution of m^5C in epigenetically silenced NSUN5 (Control) and NSUN5-OE LN229 cell line. **(B)** Base-pairing propensity meta-profile of regions surrounding putative NSUN5-dependent m^5C sites. **(C)** Schematic representation of the structure of human 28S rRNA juxtaposed with the secondary structure prediction of putative NSUN5-dependent m^5C sites. Paired ratio values indicate the propensity to form base pairing, with values close to 1 being highly structured.

4.2.3 Prediction and validation of NSUN5-dependent physiological mRNA sites

While it remains challenging to confidently attribute these ~1,000 unassigned m^5C sites to NSUN5, we managed to utilise what we know about the enzymology of different NSUN enzymes to make a prediction about NSUN5 methylation of mRNA.

We decided to start with the total number of m^5C sites detected upon NSUN5 OE and discard those sites we know are deposited by either NSUN2 or NSUN5 using experimental data

(NSUN2 KO from Yang et al., (2017) and epigenetically silenced NSUN5 from Janin et al., (2019)), which left us with 29 sites (**Fig. 13A**). We noticed that of these 29 sites, 21 showed the CUCCA sequence motif characteristic of NSUN6-dependent sites and were therefore discarded (**Fig. 13A**). We were then left with 8 candidate mRNA sites which we predict could be methylated by NSUN5. During our examination of the sequence context surrounding these m⁵C sites, an important observation emerged: all of them exhibited a sequence motif remarkably akin to both the NSUN5-dependent C:3782 motif in 28S rRNA (as illustrated in **Fig. 13B**) as well as the sequence motif that experienced enrichment upon NSUN5 OE, highlighted in **Figure 11B**. This finding provides additional support for the notion of NSUN5's involvement in the methylation of these mRNAs.

In order to experimentally confirm our predictions, we decided to perform an amplicon-based validation of these 8 candidate sites using bisulfite conversion technique. The first step was to generate a NSUN5 KO cell line starting from our WT cells, to directly compare the methylation levels of these mRNAs with and without NSUN5 protein (for details, see materials and methods section, 3.7). We therefore used CRISPR/Cas9 technology to delete a big portion of NSUN5 gene and generate a viable KO cell line (**Fig. 13C**).

We then collected biologically replicated RNA samples from both WT and NSUN5 KO HeLa cells and treated them with bisulfite. Then, following the protocol established in our lab and explained in Sibbritt et al., (2022), we produced short amplicons of approximately 100 nucleotides surrounding each of our candidate sites and sequenced them on the Illumina miSeq platform. Preliminary analyses showed promising results. When comparing the non-conversion levels of NSUN5-dependent C:3782 in 28S rRNA, used as positive control, we find that its methylation level disappears upon NSUN5 KO when comparing to WT condition (**Fig. 13D, top**). This further confirms the successful deletion of NSUN5 gene as well as the effective conversion obtained with bisulfite treatment. Similarly, when looking at one of the candidate sites in PIK3R3 gene, we find that the methylation level remains constant at around 20% in WT condition across the three biological replicates and drops down to nearly 0 upon NSUN5 KO in two out of three replicates (**Fig.3, bottom**). It is worth noticing how the third NSUN5 KO replicate (N5_KO_C), whilst being an outlier also shows abnormal non-conversion in the positions surrounding the m⁵C of interest. This is likely due to an incomplete conversion at the bisulfite treatment step of the experiment, a technical issue that will be addressed during the preparation of the next set of libraries. Our results, although preliminary and only restricted to one candidate mRNA, are supportive of the role of NSUN5 in methylating mRNA.

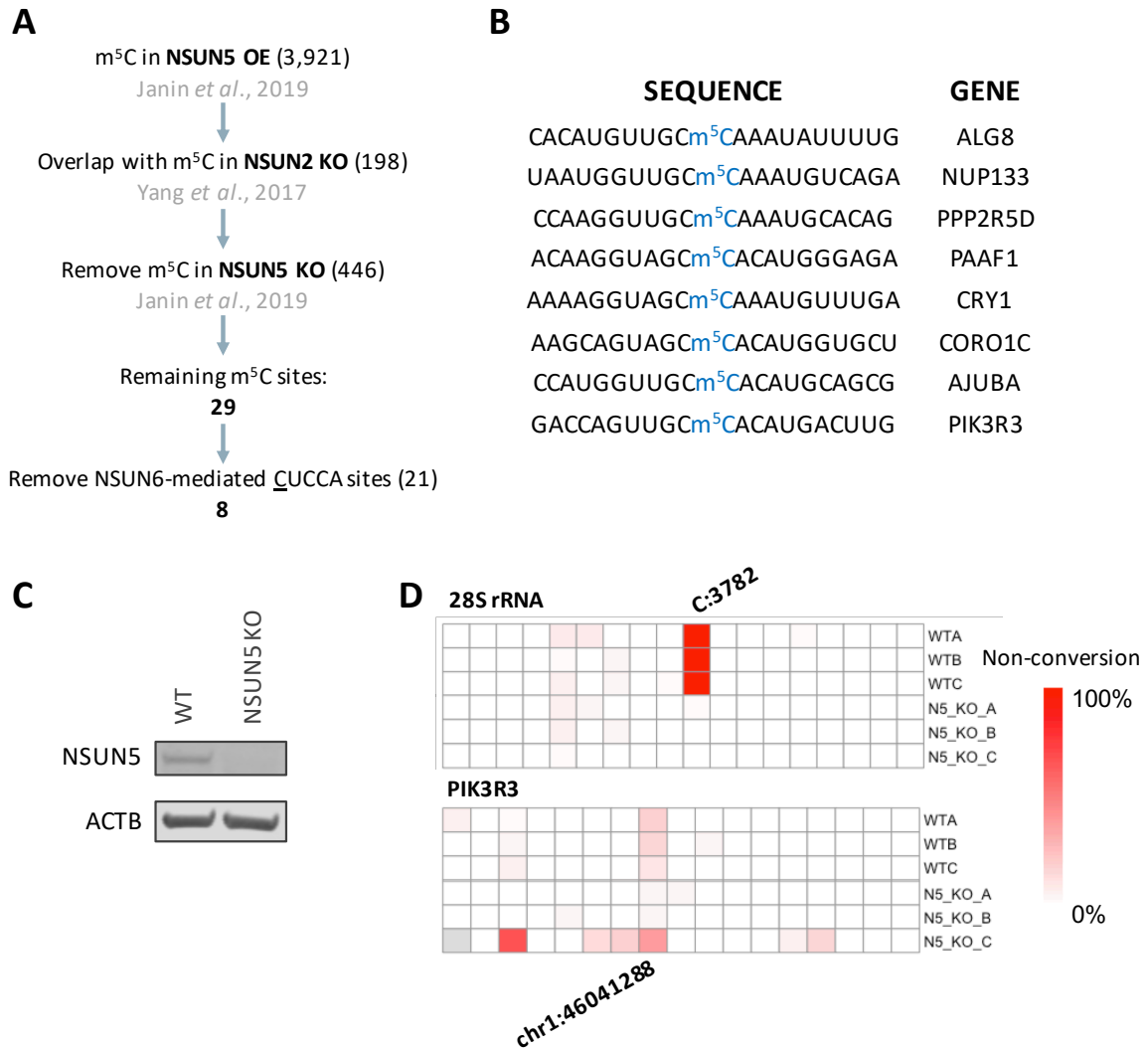


Figure 13 - (A) Strategy used to predict NSUN5-mediated candidate sites on mRNA. (B) 8 predicted NSUN5-methylated genes. (C) Representative replicate of Western blot showing the NSUN5 protein in WT and NSUN5 KO HeLa cells. BACT was used as a loading control. (D) Validation and NSUN2-dependence of candidate m⁵C sites. Grids showing cytosine position along the analysed transcript section, with each square coloured by observed cytosine non-conversion (white-to-red colour scale). Genomic coordinates of the m⁵C sites are shown on the figure.

4.3 Multiple RBPs are enriched around m⁵C

One way RNA methylation is known to exert its function in regulating RNA metabolism is through the interaction with RNA-binding protein. For instance, m⁶A-decorated mRNAs are specifically recognised and bound by reader proteins containing a YTH (YT521B homology) domain, among which two sub-families can be identified: YTHDF (1, 2, 3) and YTHDC (1, 2) (Meyer & Jaffrey, 2017). Different reader protein can bind m⁶A modifications at different positions, and regulate a different aspect of these RNA molecules, including splicing (Di Timoteo et al., 2020), translation (X. Wang et al., 2015), stability (Kennedy et al., 2017), and degradation (Du et al., 2016).

In a similar fashion, two different m⁵C reader proteins were identified to be regulating mRNA metabolism. The Aly/REF export factor (ALYREF) was shown to bind m⁵C-containing mRNAs and selectively promote their export from the nucleus (X. Yang et al., 2017a), with recognised roles in controlling cell migration (X. Xu et al., 2020) and adipogenesis (Liu et al., 2021b). The Y-box binding protein 1 (YBX1) preferentially binds m⁵C-decorated mRNAs through its cold shock domain (X. Chen et al., 2019; X. Yang et al., 2017a). In human urothelial carcinoma of the bladder, YBX1 recruits the ELAV-like 1 (ELAVL1) RNA-binding protein to increase the stability of the mRNA for the oncogene heparin-binding growth factor (HDGF), thereby promoting pathogenesis of bladder cancer (X. Chen et al., 2019). In early zebrafish embryos, YBX1 regulates the maternal-to-zygotic transition by promoting the stability of m⁵C-containing maternal mRNAs through recruitment of the poly(A) binding protein cytoplasmic 1a (PABC1a) (Y. Yang et al., 2019). This resonates with intricate waves of maternal mRNA methylation mediated by NSUN2 and NSUN6, seen in both vertebrate and invertebrate species including flies, fish, frogs, mice and humans (J. Liu et al., 2022).

4.3.1 Multiple RBPs are enriched around m⁵C

Given the established role that RNA-binding proteins can have as reader proteins in regulating gene expression of methylated RNA, we decided to utilise the strong statistical power derived from our union set to investigate the potential function of known RBPs in relation to m⁵C modification. We therefore established a pipeline to investigate spatial proximity of known RBPs to our union list of 6,393 human m⁵C sites.

In order to accomplish a large-scale enrichment analysis of RBPs around m⁵C modification, we downloaded public eCLIP-seq data from ENCODE (Van Nostrand et al., 2020). These

datasets contain information about the footprint of >100 known RBPs on their target mRNAs from two different cell lines: HepG2 and K562 lymphoblast cells from myelogenous leukemia patient and human hepatoma, respectively. We also included similar binding data for the known m⁵C reader ALYREF (Y. C. T. Yang et al., 2015), used as a positive control.

We selected the middle region of each eCLIP peak for each of the proteins as the middle nucleotide in their footprint and regarded it as the RBP binding site. Then, a gradient of regions surrounding the each RBPs' binding site ($\pm 20/30/50/70$ nucleotides) was selected as 'RBP adjacent region'. We filtered those mRNAs that contained both m⁵C modifications and RBP peaks and calculated the number of m⁵C and C within and outside the selected regions. Finally, we calculated the odds ratio and p-value using proportion test and adjusted the p-value using the 'fdr' method. Details of the pipeline is outlined in the materials and methods section [3.5](#).

This aforementioned method was applied to the eCLIP-seq data downloaded from ENCODE, which include the binding regions of 122 RBPs from K562 cells and 104 from HepG2 cells, and using our union set of 6,393 human m⁵C sites. The enriched p-value for gradient regions of these RBPs are shown in **Figure 14 A and B**. m⁵C sites located within the 50 nucleotides region around an RBP binding site were selected for the subsequent analyses. Interestingly, several RBPs were found to have their footprint on mRNA significantly enriched near m⁵C sites. Among these, AKAP1, DDX3X, DHX30, FTO, NCBP2, RPS3 and UPF1 proteins were found to be top hits in both cell lines (**Fig. 14 C and D**).

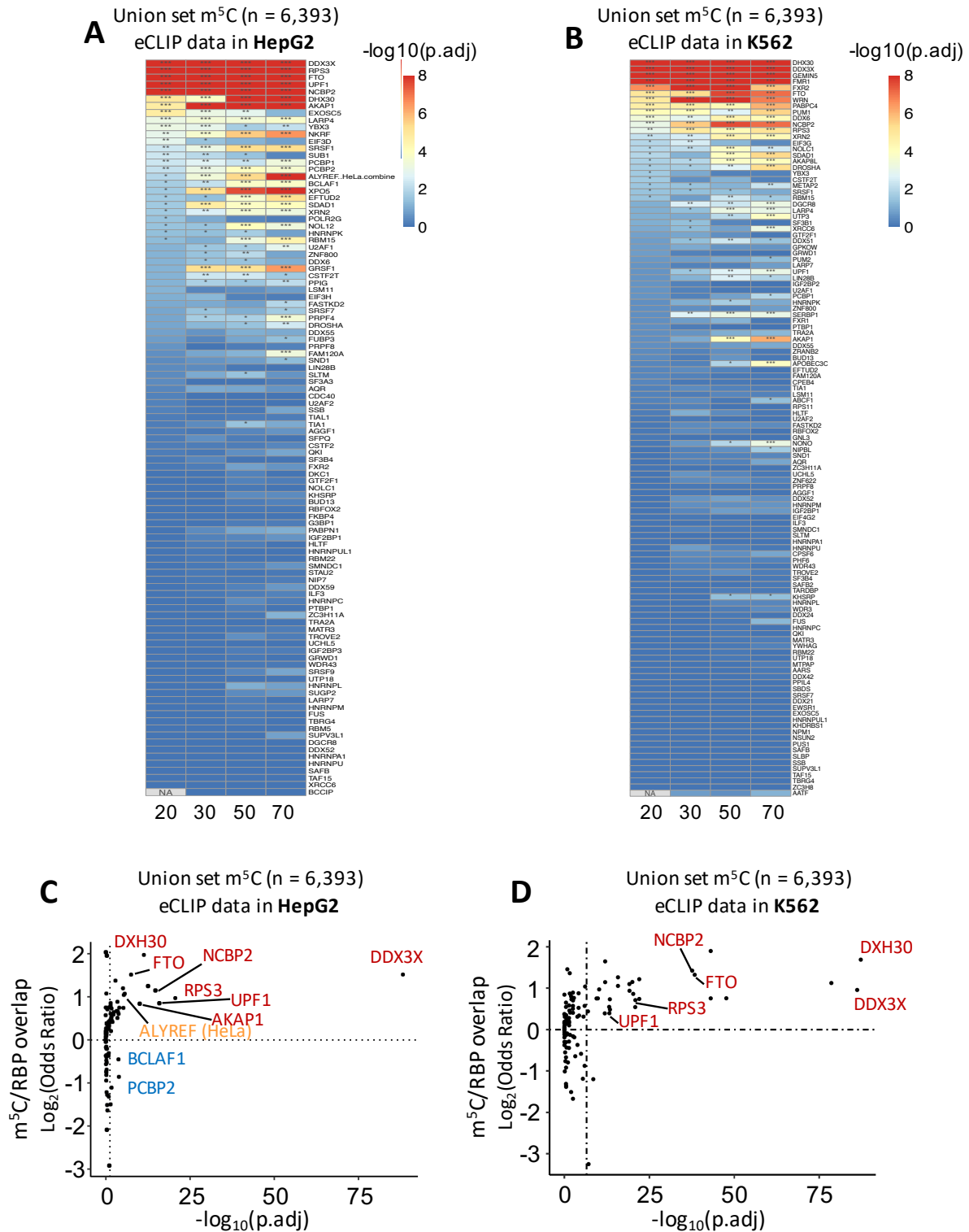


Figure 14 - (A and B) Enrichment analysis of RBPs using union set of m⁵C sites. eCLIP data from ENCODE in HepG2 cells (A) and K562 cells (B). The regions around RBP binding sites were set as intervals of ± 20 nt, ± 30 nt, ± 50 nt or ± 70 nt, with the middle position of the footprint set as 0. P-values were calculated as Chi-Squared Test: *p < 0.05; **p < 0.01; ***p <

0.001 (**C**) Enrichment analysis for m⁵C sites from union set with RBPs eCLIP binding-sites in HepG2 cell line. (**D**) Enrichment analysis for m⁵C sites from union set with RBPs eCLIP binding-sites in K562 cell line.

We next decided to check whether these results are the same when restricting our enrichment analyses to only m⁵C sites found in individual cell lines. We therefore repeated the analysis using only m⁵C sites found in either HeLa or HepG2 cell lines. 2,342 m⁵C sites were detected on mRNA from five HeLa bsRNA-Seq datasets from three studies (Yang et al, 2017; Schumann et al, 2020; Huang et al, 2019). In addition, 708 m⁵C sites were detected on mRNA from two HepG2 bsRNA-Seq datasets from one study (Liu et al, 2021). The enriched p-value for gradient regions of these RBPs are shown in **Figure 15 A and B**. Similarly to the results obtained using our union list, the top RBPs were found to significantly overlap with m⁵C sites detected in HeLa or HepG2 cells only (**Fig. 15 C and D**).

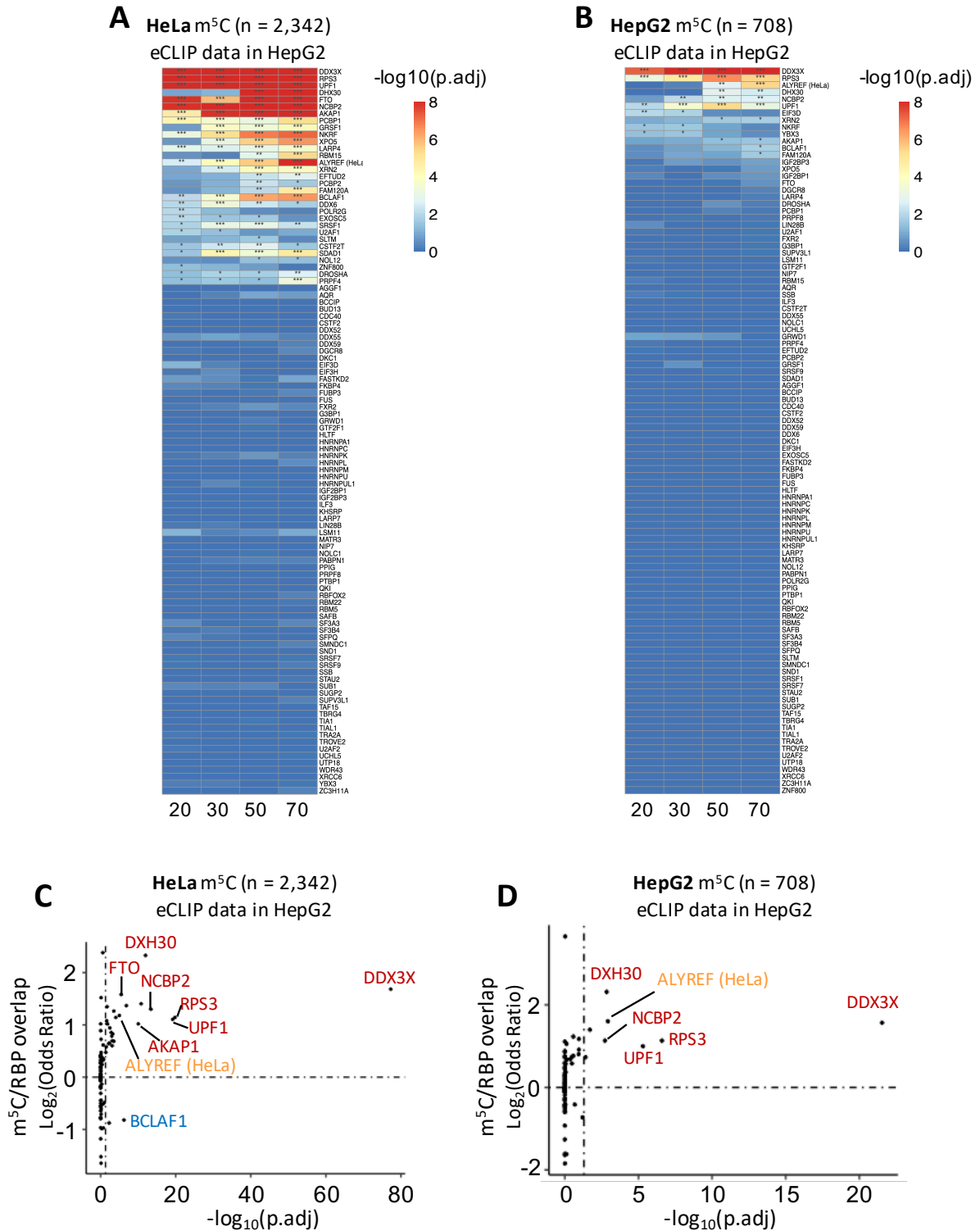


Figure 15 - (A and B) Enrichment analysis of RBPs (from ENCODE in HepG2 and known m⁵C reader in HeLa) and m⁵C from HeLa (A) and HepG2 (B). The regions around RBP binding sites were set as intervals of ± 20 nt, ± 30 nt, ± 50 nt or ± 70 nt, with the middle position of the footprint set as 0. P-values were calculated as Chi-Squared Test: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (C) Enrichment analysis for m⁵C sites from HeLa cells with RBPs eCLIP binding-sites

in HepG2 cell line. **(D)** Enrichment analysis for m⁵C sites from HepG2 cells with RBPs eCLIP binding-sites in HepG2 cell line.

Encouraged by the recurring enrichment of these proteins around m⁵C sites, we decided to further characterize the extent of such overlap. To accomplish this, we opted to examine the frequency of a given RNA-binding protein (RBP) binding to an m⁵C-modified mRNA within both a 50-nucleotide region surrounding the modification site and outside this region. We focused on the 2,342 m⁵C sites identified in HeLa cells, as this cell line would serve as the target for any probable experimental validation approach, as previously mentioned. Once again, we found that the top enriched RBPs identified from our previous analyses clustered together (in red), as opposed to depleted RBPs (in blue) (**Fig. 16A**). This data confirms that within this dataset, a subgroup of proteins, upon binding to methylated mRNAs, exhibits an affinity to interact with their target mRNAs in close proximity to an m⁵C modification site. In addition, when looking at the number and distribution of m⁵C modifications along the mRNAs targeted by these proteins, we observed that in many cases, the hotspots of methylation overlapped with the RBPs' footprint (**Fig. 16B**). This was particularly noticeable for DDX3X, RPS3 and NCBP2. For the sake of comparison, we conducted a similar analysis on other RBPs from the same datasets, which exhibited a contrasting behavior involving the location of their footprints and the presence of the m⁵C modifications within the same mRNA targets. As expected, our findings revealed a lack of correlation between the binding sites of these RBPs on their target mRNAs and the locations of methylation within these mRNAs (**Fig. 16C**).

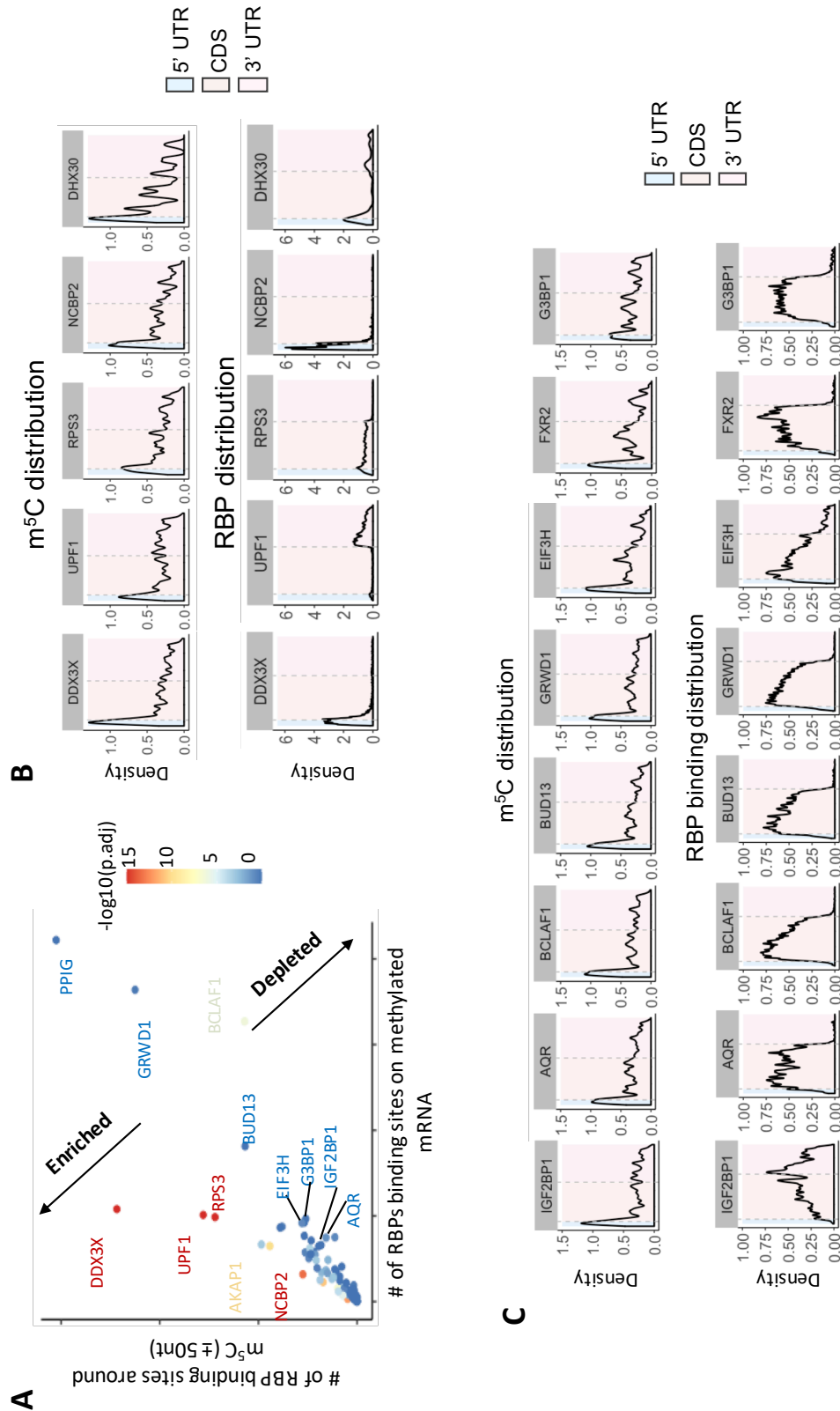


Figure 16 - (A) Top enriched RBPs in A and depleted RBPs shown by comparing the number of RBP binding sites on HeLa m⁵C methylated mRNA and around m⁵C methylation (± 50 nt). Metagene distribution of enriched **(B)** and depleted **(C)** RBPs (bottom panel) and m⁵C (top panel) around mRNA.

To further appreciate the extent of the methylation peaks over the binding region of our RBPs, we decided to set the middle point of each protein's footprint as 0, over which we overlaid both the m⁵C and C density of their target mRNAs. What we observed was a marked peak of methylation that coincided with the central region of their binding sites (**Fig. 17**). Altogether, our data points towards a different binding pattern of some RBPs that could be influenced by the methylation content of their target mRNAs.

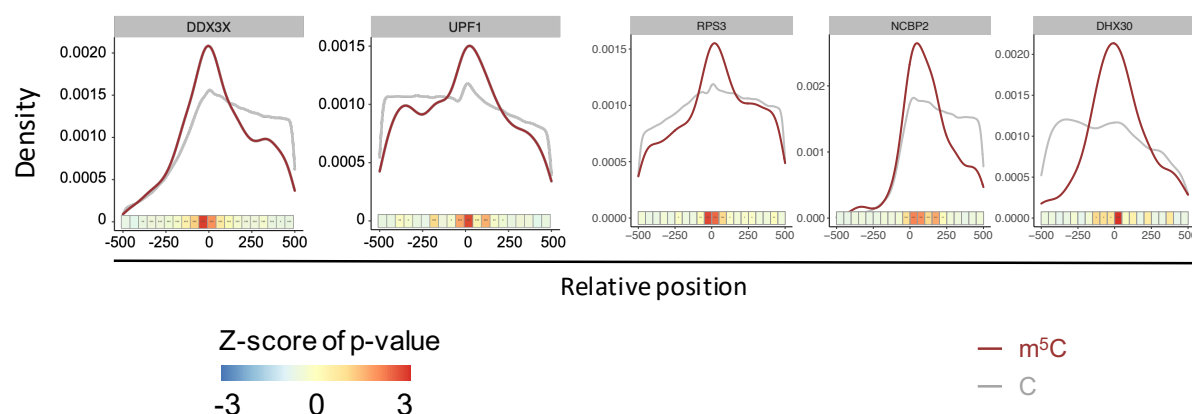


Figure 17 - Distribution of m⁵C (in red) and background C (in grey) overlaid on RBP binding sites.

4.3.2 The function of UPF1 RBP is affected by the lack of NSUN2 methyltransferase

Our findings were particularly intriguing that sparked our interest in further investigating the relationship and the functional interaction between m⁵C modifications and RBPs.

Among the proteins we found in our enrichment analysis, we decided to follow up on some of them, including UPF1.

UPF1 is a known RNA-binding protein with ATP-dependent RNA helicase activity that can bind non-specifically to mRNA. UPF1 is also the core component of nonsense mediated decay (NMD), a translation-dependent mechanism that cells use to trigger degradation of transcript then they contain a premature termination codon (PTC) (Kurosaki et al. 2019). These PTCs are often the result of genomic mutations or splicing aberrations, which generate aberrant open reading frames (ORFs) that could be deleterious for the cell (Lloyd, 2018). The NMD process is triggered by the phosphorylation of UPF1 mediated the SMG1 kinase, which is in

turn stimulated by UPF2 and UPF3B proteins (Lee et al., 2015). Once in its hyperphosphorylated state, UPF1 can then recruit SMG6 protein, an endonuclease that cleaves the target RNA, as well as SMG5 and SMG7, which will induce de-capping and de-adenylation (Lavysh & Neu-Yilik, 2020). Together, they will ultimately lead to transcript degradation.

While there is a consensus in the community for NMD being a translation-dependent mechanism (Lavysh & Neu-Yilik, 2020), it was shown that also small nucleolar RNAs (snoRNAs) and long non-coding RNAs (lncRNAs) can be targeted by NMD (Lykke-Andersen et al., 2014). This implies that either these lncRNAs do interact with the ribosome or that NMD is not a purely translation-dependent mechanism.

In 2020, UPF1 was reported to initiate mRNA degradation independently from translation termination (Fischer et al., 2020). In synergy with the RNA-binding protein G3BP1, UPF1 can initiate a decay pathway that degrades mRNAs that have a highly structured 3' UTR. Rather than scanning for particular regions or sequences, this pathway recognises high level of general structuredness and is referred to as structure-mediated decay (SRD) (Fischer et al., 2020).

Moreover, in addition to proofreading mRNAs for PTCs, NMD was also discovered to regulate the steady-state of physiological mRNAs which encodes for seemingly functional, full-length proteins as a way to regulate embryonic development and stress response (Nasif et al., 2018). These mRNAs were found to be enriched in UPF1 binding via CLIP experiments and sensitive to UPF1-induced degradation thanks to UPF1 KD RNA sequencing experiments, but yet they did not present any of the features typically necessary to trigger NMD or SRD, like PTCs or highly structured 3' UTRs (Nasif et al., 2018).

Overall, UPF1 function in mRNA decay appears to not be limited to degradation of PTC-containing mRNAs, but rather a powerful regulator of gene expression regulation, including mRNAs encoding functional ORFs. As a unified mechanism for UPF1 function is lacking, further investigation is necessary to fully comprehend the extent of UPF1-mediated regulation of mRNA decay and additional mRNA features that regulate UPF1-dependent steady-state regulation.

4.3.2.1 m⁵C-mRNAs are differentially (de)regulated upon UPF1 KD

Given the significant overlap between UPF1 footprint on mRNA and m⁵C modifications, we wondered whether there could be a connection between UPF1 function on its targets and their methylation content. We assessed whether the regulatory influence exerted by UPF1 on the steady state of m⁵C-decorated mRNA could be affected by varying their methylation state. In order to set up experimental conditions in which m⁵C levels were reliably and consistently reduced, we employed the Crispr/Cas9 technology to generate a NSUN2 KO HeLa cell line (details are outlined in the materials and methods section, [3.7](#)).

The reason we decided to target NSUN2 among the different m⁵C methyltransferase is because both in our data (**Fig. 9A and 10A**) as well as in the published literature (Yang et al, 2017; Amort et al, 2017; Schumann et al, 2020), NSUN2 was found to be the main contributor to mRNA m⁵C methylation. Given that the majority of m⁵C sites rely on NSUN2, the logical deduction follows that knocking out NSUN2 would impact a large proportion of m⁵C sites. This approach specifically targets a single m⁵C methyltransferase while affecting the greatest number of sites. In addition, having a KO cell line over performing a knockdown via siRNA boasts an advantage of being unaffected by fluctuations in the expression level of NSUN2 protein, a downside that can be observed when transiently downregulating RNA methyltransferases.

Once we had NSUN2 KO cell line, we knocked down UPF1 protein in both WT and NSUN2 KO HeLa cells, which allowed us to directly compare the effect of the regulation mediated by UPF1 with (WT) and without (NSUN2 KO) m⁵C modification. After ensuring that the protein levels of both UPF1 and NSUN2 were appropriately affected (**Fig. 18A**) via Western blot analysis, we performed total RNA sequencing on the RNA derived from three biological replicates (**Fig. 18B**).

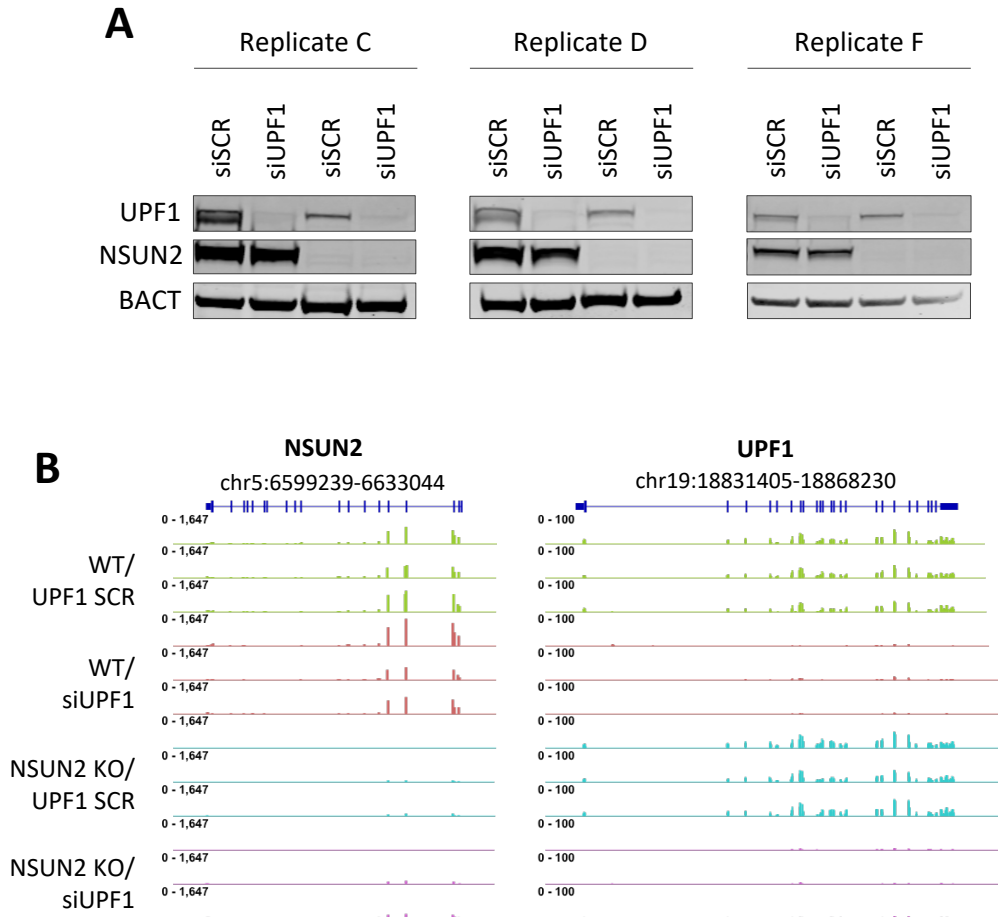


Figure 18 - (A) Three biological replicates of the proteins derived from control condition (siSCR) and the knockdown of UPF1 protein in wild-type and NSUN2 KO HeLa cells. ACTB was used as loading control. **(B)** Genome tracks showing the reduction in expression level of both NSUN2 and UPF1 in four experimental groups in the three biological replicates.

Our analysis revealed that compared to siSCR (non-targeting) control condition, the knockdown of UPF1 in WT and NSUN2 KO cells resulted in a similar number of downregulated genes (~2,000). However, the number of upregulated genes upon UPF1 KD was lower in NSUN2 KO (1,908) compared to WT (2,365) cells, a difference of more than 450 genes (**Fig. 19A**). In order to assess whether this differential regulation we observed could be attributed to the different m⁵C content in mRNA in WT vs NSUN2 KO cells, we used bisulfite sequencing data of WT untreated cells and NSUN2 KO/KD HeLa cells (Huang et al, 2019; Yang et al, 2017) to compile a list of 1,633 NSUN2-dependent m⁵C sites in HeLa cells (**Table 7**). Next, we selected those m⁵C sites located on mRNAs that are also enriched in UPF1 binding. We then decided to further separate these candidate genes into two subgroups, depending on whether their m⁵C site was found within or outside the 50-nucleotide region centered on the middle point of UPF1 footprint. We then looked at their expression levels upon UPF1 knockdown, in WT vs NSUN2 KO. We selected those m⁵C sites located on mRNA that are also enriched in UPF1 binding and looked at their expression levels upon UPF1 knockdown, in WT vs NSUN2 KO. We found that the 'in region' candidate mRNAs (**Fig. 19B, top panel**) were found to significantly accumulate in WT cells upon UPF1 KD. This accumulation however was significantly reduced when UPF1 was knocked down in NSUN2 KO cells (**Fig. 19B**). An intriguing observation emerged when examining the second subgroup, where the m⁵C site was situated beyond the 50-nucleotide span around the midpoint of the UPF1 footprint (as depicted in **Figure 19C**, top panel). We noted that not only did these mRNAs exhibit substantially lower accumulation upon UPF1 knockdown in WT cells, but they also lost or drastically reduced the distinctive differential regulation we had previously observed in NSUN2 knockout (KO) cells for the "in region" genes (as shown in **Figure 19C**).

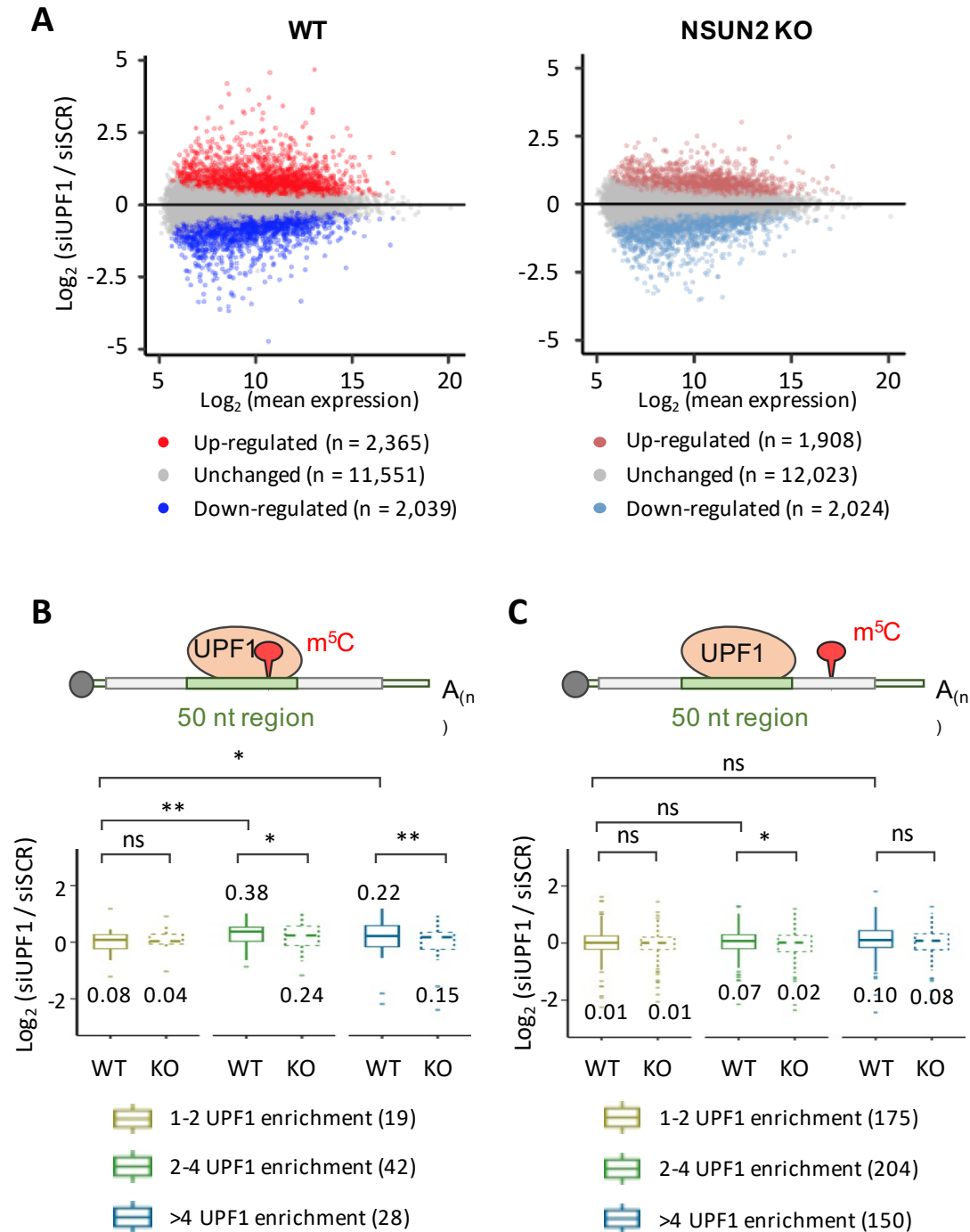


Figure 19 - (A) Changes in RNA expression levels upon knockdown of UPF1 in wild-type and NSUN2 KO HeLa cells. (B and C) RNA levels of UPF1 target genes with m⁵C within (B) or outside of (C) the 50 nucleotides region (top, in green) from the middle point of UPF1 footprint. Different colours indicate different extent of UPF1 enrichment (in fold change). P-values were calculated using student's t test: *p < 0.05; **p < 0.01; ***p < 0.001.

As an additional control, we looked at the RNA levels of those mRNAs with UPF1 binding overlapping with NSUN2-independent m⁵C sites. In particular, we selected sites that had the m⁵CUCCA sequence motif that is typically associated with NSUN6 methyltransferase (Liu et al, 2021), which showed no changes in expression level upon NSUN2 KO/KD (data not shown). As expected, the downregulation of UPF1 had the same effect on the steady state of these transcripts in both WT and NSUN2 KO cells and no differential regulation was observed (**Fig. 20**).

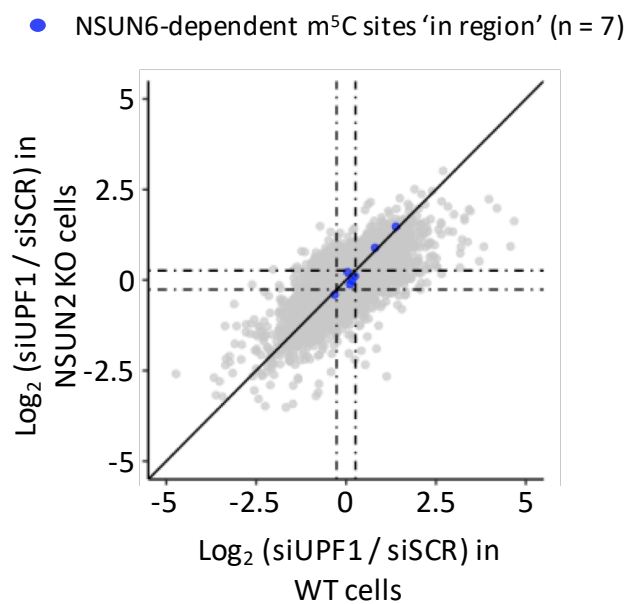


Figure 20 - Scatterplot showing the change in RNA levels upon UPF1 KD, in WT and NSUN2 KO cells. In blue, 'in region' NSUN6-dependent m⁵C sites characterised by the m⁵CUCCA sequence motif.

We then decided to further validate the results obtained from our RNA-seq experiment. We selected 4 mRNAs from the 'in region' subgroup shown in **Figure 19B**: FADS2, FKBP10, NIBAN2 and OSMR (**Fig. 21A**). Together with a positive and negative control gene, performed RT-qPCR using cDNA derived from the same RNA samples sent for sequencing as a template. Our validation experiment confirmed a significant accumulation of these mRNA upon UPF1 knockdown in WT cells and the lack of such regulation upon UPF1 knockdown in NSUN2 KO HeLa cells (**Fig. 21B**).

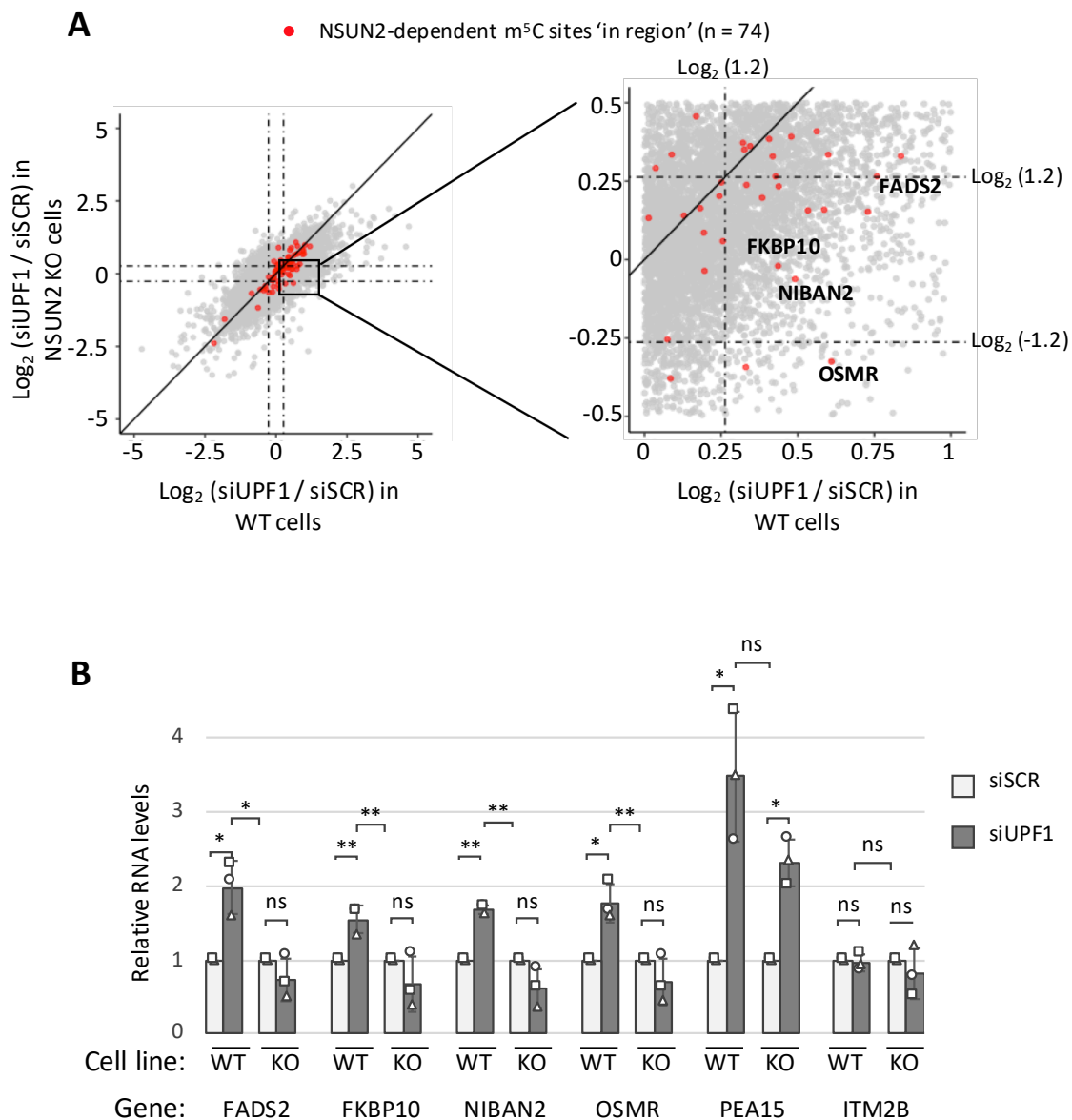


Figure 21 - (A) Scatterplot showing the change in RNA levels upon UPF1 KD, in WT and NSUN2 KO cells. In red, 'in region' NSUN2-dependent m⁵C sites. candidates chosen for RT-qPCR validation. (B) Relative RNA levels of genes selected from (A) showing upregulation

upon UPF1 knockdown in wild-type but not in NSUN2 KO HeLa cells. PEA15 and ITM2B were used as positive and negative controls, respectively. Values were normalised over ACTB and expressed as relative quantity with respect to siSCR treatment set as 1. Individual biological replicates are shown in figure on each bar as different shapes. N = 3. P-values were calculated using two-tailed Student's t test. Student's t test-derived p values: *p < 0.05; **p < 0.01; ***p < 0.001.

4.3.2.2 UPF1-targeted mRNAs are differentially (de)regulated upon NSUN2 KO

We were interested delving deeper into comprehending the convoluted interplay between the activity of NSUN2-dependent methylation and the role played by UPF1 in adjusting mRNA stability. We established that the lack of NSUN2 methyltransferase, and therefore a reduced methylation content in mRNA, affects the function of UPF1 on some of its target mRNAs. We then wondered if the opposite was also true. In particular, we were curious whether a similar differential regulation could be observed when comparing the steady-state level of mRNA in NSUN2 KO compared to WT cells, upon UPF1 KD.

We decided to divide the transcriptome into five different groups (**Fig 22A**): targets of neither NSUN2 nor UPF1 – 'non targets' (group I); NSUN2 targets only (group II); UPF1 targets only (group III); UPF1 and NSUN2 targets (group IV) – same as **Figure 19C**; UPF1 binding around m⁵C (group V) – same as **Figure 19B**. We then looked at the RNA levels of each group upon NSUN2 KO, with (siSCR) and without (siUPF1) UPF1 protein. Interestingly, we found that mRNAs belonging to group V, not only had the highest degree of accumulation upon NSUN2 KO in siSCR conditions, but also had the highest significant difference between siSCR and UPF1 KD, compared to group I (**Fig. 22B**). Indeed, in UPF1 KD condition, no accumulation was observed upon NSUN2 KO. Again, the observed differential regulation observed upon NSUN2 KO resembles the differential regulation observed in the previous section upon UPF1 KD. This data successfully recapitulates the effects of UPF1 knockdown and links the function of UPF1 protein with the methyltransferase activity of NSUN2 enzyme.

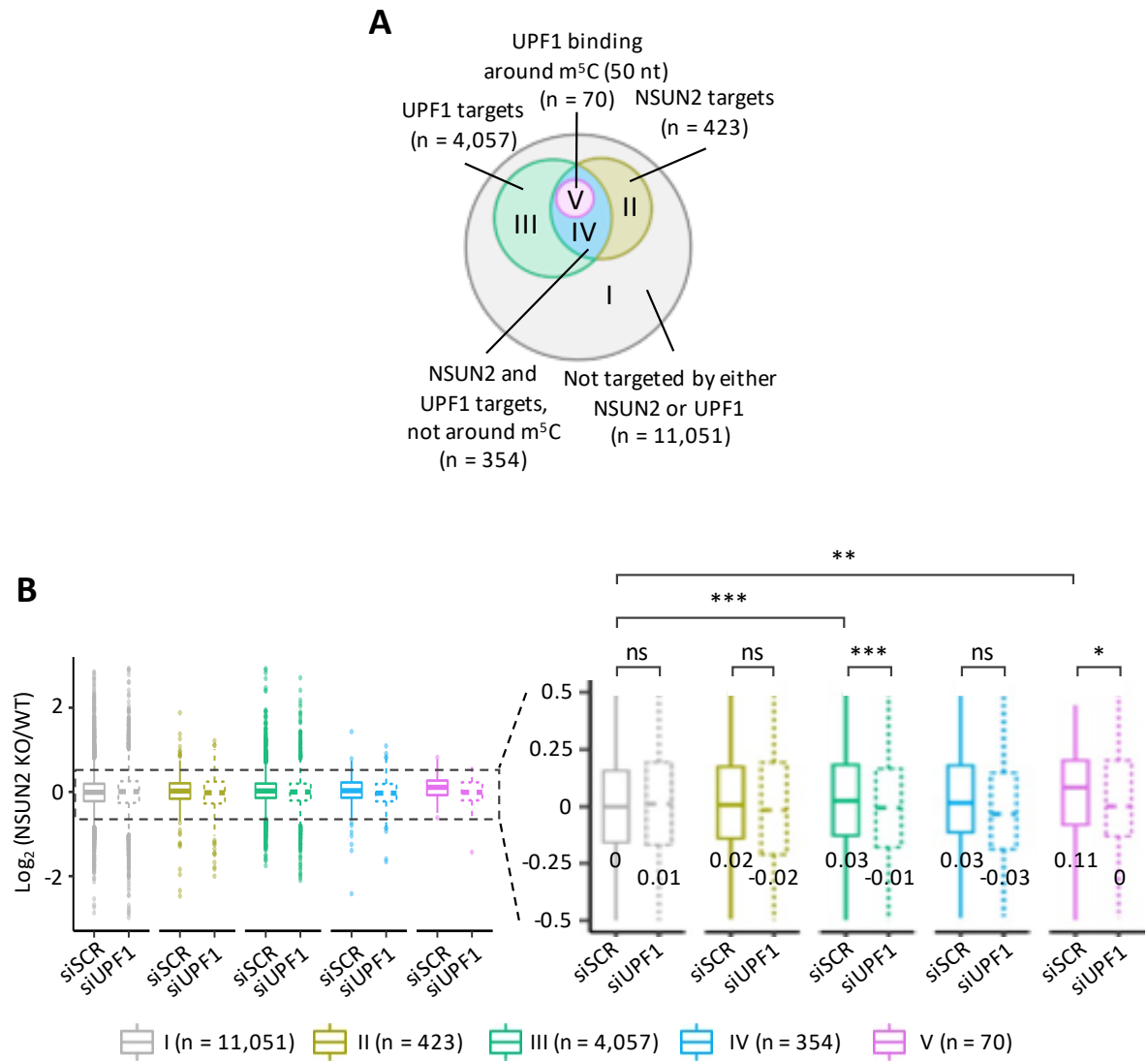


Figure 22 - (A) Transcriptome subdivision into 5 categories. **(B)** Boxplot showing RNA levels upon knockout of NSUN2 proteins in siSCR and siUPF1 conditions. Different RNA populations are defined in the Venn diagram (A). P-values were calculated using student's t test: *p < 0.05; **p < 0.01; ***p < 0.001.

4.3.3 UPF1 binding of m⁵C-mRNAs is regulated by NSUN2 methyltransferase

Considering the functional connection between NSUN2 expression levels and UPF1 activity on m⁵C-decorated mRNAs, we wondered if the distinct regulatory patterns observed in the preceding section could be attributed to alterations in UPF1 binding resulting from varying levels of mRNA methylation.

We wanted to know whether the decrease in methylation content on certain mRNAs could impact UPF1's ability to bind to them. To address this, we decided to perform CLIP experiments. We optimised the immunoprecipitation of UPF1 protein after UV-crosslinking and performed the experiment in both WT and NSUN2 KO HeLa cells. This enabled direct comparison between the enrichment of UPF1 protein on specific mRNAs with (WT) and without (NSUN2 KO) m⁵C modification.

CLIP experiments were performed in 5 independent biological replicates and checked for the relative amount of UPF1 protein recovered using Western blot analysis (**Fig. 23A**).

We managed to achieve a similar enrichment of the target protein in both cell lines, compared to their input sample. This was an important goal, as it makes the direct comparison between the two conditions reasonable. We then extracted the RNA fraction associated with the immunoprecipitated protein and performed RT-qPCR for the same candidate genes as in **Figure 21B**: FADS2, FKBP10, NIBAN2, OSMR, together with a positive and negative control. The analysis of the RT-qPCR experiment revealed a significant increase in binding activity for UPF1 upon NSUN2 KO. These target genes were found to be significantly more enriched in NSUN2 KO samples compared to their WT control (**Fig. 23B**), despite similar levels of enrichment were observed for the positive control DNAL4 (**Fig. 23C**).

DNAL4 gene was selected as a positive control for this CLIP experiment because, like the other target genes, it was enriched in UPF1 binding, making DNAL4 a good control to test the efficiency of the immunoprecipitation. However, DNAL4 lacks any NSUN2-dependent m⁵C modification, which makes it a good control to ensure that the efficiency of the immunoprecipitation was comparable between the WT and NSUN2 KO regardless of any m⁵C levels, and its enrichment remained unchanged in the two conditions.

Unlike DNAL4, whose binding to UPF1 was unaltered upon NSUN2 KO, the other four candidates we tested (FADS2, FKBP10, NIBAN2 and OSMR) showed a strong and significant > 2-fold increase in binding to UPF1 in NSUN2 KO condition (**Fig. 23B**), further linking the activity of NSUN2 m⁵C methyltransferase to the function of UPF1.

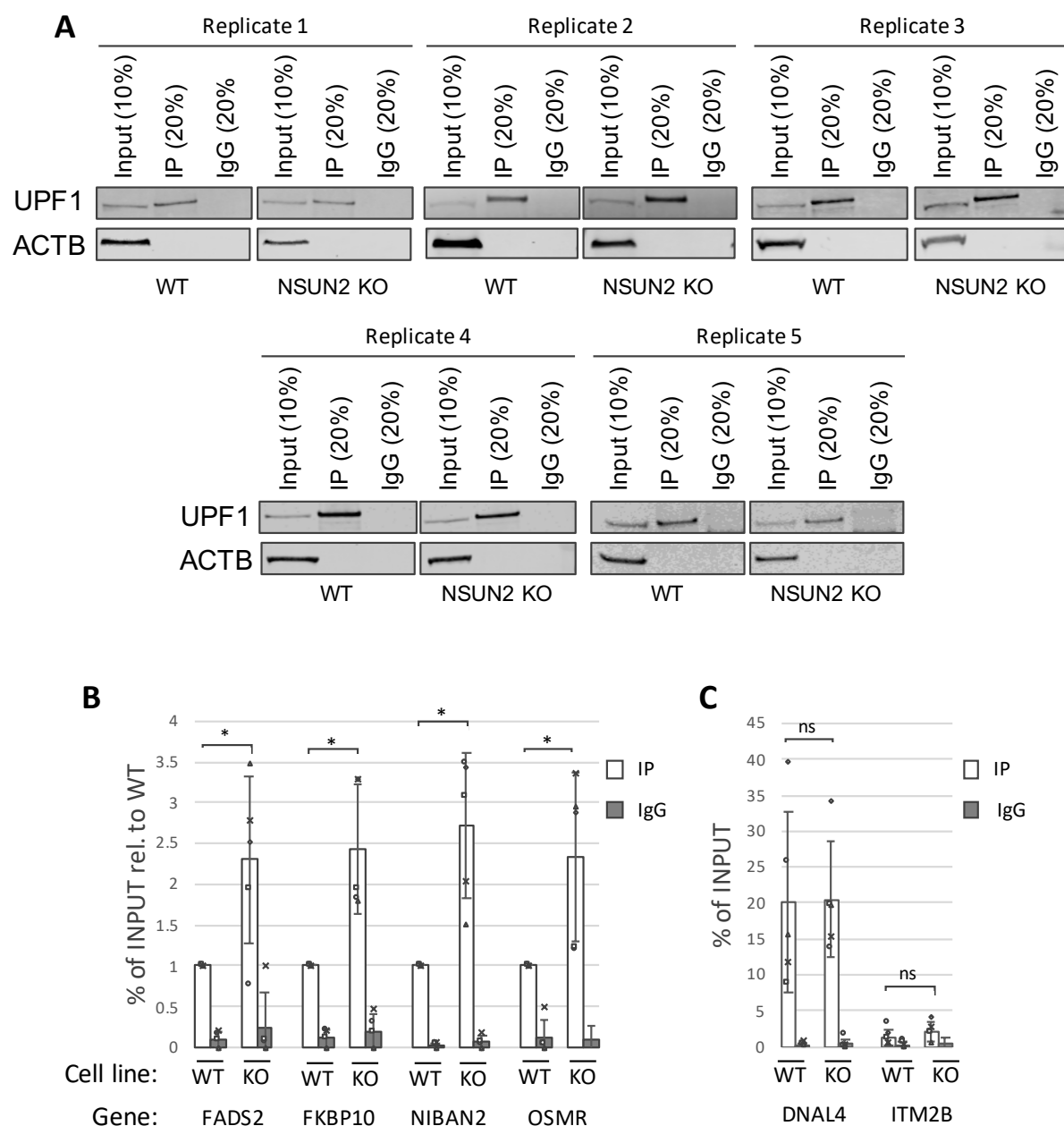


Figure 23 - (A) Five biological replicates of proteins derived from UPF1 CLIP in WT and NSUN2 KO experiment. ACTB was used as a control. N = 5. (B and C) Enrichment levels of mRNA candidates in WT and NSUN2 KO HeLa cells upon UPF1 CLIP. Values were normalised over RNA SPIKE-IN and expressed as % of input relative to WT (B) or % of input (C). DNAL4 and ITM2B were used as positive and negative controls, respectively. Individual biological replicates are shown in figure on each bar as different shapes. N = 5. P-values were calculated using two-tailed Student's t test. Student's t test-derived p values: *p < 0.05; **p < 0.01; ***p < 0.001.

Overall, our data indicate a functional association between m⁵C modification of mRNA and UPF1-mediated regulation of mRNA steady state, gaining further insights into an additional mechanism of action for UPF1 function. We propose that the portion of mRNAs that are regulated by UPF1, despite lacking the features typically associated with either NMD (PTCs) or SRD (structured 3' UTRs), can be dependent on NSUN2-dependent m⁵C content.

5. Discussion

A growing collection of studies indicates a key role for RNA modification as an additional layer of gene expression regulation (Amort et al., 2017; Huang et al., 2019; Liu et al., 2021a, 2022; Schumann et al., 2020; Squires et al., 2012; Yang et al., 2017, 2019). 5-methylcytosine, once only regarded as a tRNA modification (Reid et al., 1999), is now known as a widespread RNA modification that can be found on all known RNA species, including mRNA (Huang et al., 2019; Squires et al., 2012; Schumann et al., 2020; Yang et al., 2017). Therefore, m⁵C could be an important biochemical mark, which has the potential for regulating several different steps of RNA metabolism and function (Gilbert & Nachtergaele, 2023).

This broad rediscovery of m⁵C as an important mRNA modification was possible thanks to the development of a method for the detection and quantification of this modification in a transcriptome-wide manner. This method became available with the coupling of bisulfite conversion of RNA with high-throughput sequencing techniques (Squires et al., 2012).

There are tens of publications now that have used this approach and reported several different lists of m⁵C modification on RNA derived from a variety of different species, including human (**Table 7**). In particular, bsRNA sequencing technique has been applied to several different human cell lines and tissues, which resulted in a deeper understanding of the distribution and abundance of m⁵C modification in the human transcriptome.

However, upon a close inspection of the lists of m⁵C sites that have been reported during the past 10 years, it seems clear that the number of sites generally reported by different groups do not always match. In particular, early studies that published bsRNA sequencing results have reported a number of m⁵C sites in various human cell lines between 5,000 and 10,000 (Amort et al., 2017; Squires et al., 2012; Wei et al., 2018; X. Yang et al., 2017a). More recent publications, instead, report a more conservative number of sites of 1,000 or less (Liu et al., 2021; Huang et al., 2019; Schumann et al., 2020).

There are several reasons as per why that is the case, which include technical differences in the data analysis pipeline used by different groups, as well as substantial difference in the transcriptomes of different cell type, which include different expression levels of the m⁵C writer proteins and the very mRNA molecules that are decorated with m⁵C modification.

In this thesis we proposed to resolve this disparity by generating a union dataset of m⁵C site across the human transcriptome. The way we achieved this goal started with the collection and re-analysis of several different bsRNA sequencing datasets that are publicly available in the literature. We collected several bsRNA sequencing data from 5 different human cell line and 7 human tissues and re-analysed them all using a stringent and unified pipeline for m⁵C

site calling (**Table 7**). We used our custom pipeline independently on all the selected datasets. Once we obtained a list of high-confidence m⁵C sites from each of these datasets, we then pulled these together to compile the most comprehensive (to date) union set of 6,393 human m⁵C site.

While the number of sites in our union set appears to be closer to the early papers reporting m⁵C sites on RNA (Amort et al., 2017; Squires et al., 2012; Wei et al., 2018; X. Yang et al., 2017a), our 6,393 sites are the result of a union set generated by the thorough re-analysis of available datasets from various sources. While we expanded our union set to include sites from several different human cell lines and tissues, we retained the high confidence derived from a stringent m⁵C site calling method. We are therefore confident that these sites are as high confidence as more recent studies reporting hundreds of m⁵C sites in individual cell types, while not being constricted to the RNA methylome of specific cell line.

Our union set of human m⁵C is useful and important for a number of reasons. For starters, our dataset contains all those sites that can be confidently called as m⁵C sites in at least one of the cell types we have bsRNA sequencing data for. This means that differences in the expression level of different NSUN enzymes, as well as the expression level of the RNA molecules that can be targeted by these writer proteins, are accounted for. This is an advantage over an “overlap” datasets, that only includes m⁵C sites present in all the cell types in the analyses. This is because if an mRNA is confidently methylated in one sample (e.g. HeLa cells) but not in another (e.g. HEK293 cells), it does not exclude a possible biological significance for that site in the first sample. In fact, we have demonstrated in this thesis that the differential methylation observed for most sites between different samples is mainly due to (1) low read coverage (differential expression of m⁵C-mRNAs) or (2) sub-methylation level (differential expression of m⁵C writer proteins or different pipelines used for the data analysis). Lastly, our union set contains a high number of high-confidence m⁵C sites (6,393), which translates into a high statistical power to perform a number of enrichment analyses that would otherwise be challenging, including the investigation of possible interaction between m⁵C modification and RBPs.

In the second chapter of this thesis (section 4.2), we focused on better understanding the enzymology of different NSUN methyltransferase proteins, such as NSUN2, NSUN5 and NSUN6.

We started by re-analysing publicly available bsRNA sequencing data whereby one of these m⁵C writers was interfered with. In particular, by re-analysing NSUN2 KD and KO samples (Yang et al, 2017; Huang et al, 2019), as well as NSUN6 KO samples (Liu et al, 2021), we

were able to attribute the majority of m⁵C sites in HeLa and HEK293 cells to either NSUN2 (81.1%) or NSUN6 (16.3%). In addition, we could confirm the previously proposed categorisation of m⁵C sites into either Type I (which mimic sequence and structural context around tRNA variable loops, a canonical substrate of NSUN2) (Huang et al, 2019; Schumann et al, 2020) or Type II (which carry mixed features recognised by NSUN6 in the tRNA acceptor stem) (Liu et al, 2021). However, using these established features that NSUN2 and NSUN6 writer proteins recognise in mRNA molecules to trigger m⁵C deposition, we could further categorise m⁵C sites into Type I or II in our union set, even in those cell types where no KO or KD sample was available. We therefore expanded our knowledge on the enzymology of NSUN2 and NSUN6 proteins, by confidently predicting that 4281 m⁵C sites of our union set (66.9%) are mediated by NSUN2 (hence type I), while 1037 sites are mediated by NSUN6 (hence type II).

Moreover, we identified thousands of sites that are methylated with m⁵C modification upon the overexpression of NSUN5 protein. While previously identified as a m⁵C methyltransferase targeting rRNA (Heissenberger et al., 2019; Janin et al., 2019), we demonstrated that NSUN5 can also deposit m⁵C modifications on mRNA.

By comparing the sequence and structure motifs observed around these NSUN5-dependent m⁵C sites we observed a strong similarity to NSUN5 canonical target on the 28S rRNA (C:3782). We therefore propose a third class of m⁵C sites, Type III, which is characterised by a GUXGm⁵CAXXUG sequence logo a secondary structure that mimics the immediate surrounding of C:3782 in 28S rRNA. Looking at the distribution of these m⁵C sites along the mRNA sequence, we noticed how the overexpression of NSUN5 leads to an unusual density of methylation along the CDS. This phenomenon is likely to be the result of the abnormally high expression level of NSUN5 and therefore not reflecting the physiological distribution of sites seen in other samples. However, elevated levels of NSUN5 were observed in pathological condition such as colorectal cancer, where it promotes cell proliferation (Jiang et al., 2020). The deregulation of NSUN5 expression level and the consequent aberrant deposition of m⁵C modification on other transcriptomic sites that resemble, in sequence or structure, its canonical sites on rRNA might therefore lead to disease, a concept that has been poorly studied. In the future, it will be interesting to investigate whether the aberrant expression of NSUN5 and its methyltransferase activity targeting mRNA may have a role in pathogenesis,

In the third chapter of this thesis (section 4.3), we used the high statistical power derived from our union set of human sites to investigate the possible extent and function of m⁵C interaction with known RBPs. Indeed, by overlapping our list of m⁵C sites on mRNA with eCLIP dataset

for >100 different RBP, we identified several RBPs whose footprint on mRNA significantly overlaps with m⁵C modifications.

Several insights into the function of m⁵C in mRNA have emerged from its consistent co-enrichment with the footprints of certain RNA-binding proteins. For instance, FTO primarily demethylates N⁶,2'-O-dimethyladenosine in the 5' cap structure of mRNA, affecting its stability (Mauer et al., 2017), potentially linking m⁵C with m⁶A metabolism. NCBP2, a subunit of the nuclear cap-binding complex, is another interesting candidate; it promotes various steps in mRNA metabolism, including nuclear export in conjunction with ALYREF (Kataoka, 2024). ALYREF is known to be a direct m⁵C reader protein that aids in the export of bound mRNA (X. Yang et al., 2017b); hence, the proximity of NCBP2 footprints to mRNA m⁵C sites is likely indirect but further implicates m⁵C in mRNA export. The cap-binding complex also plays a role in mRNA translation, as do several other enriched RBPs. RPS3 is part of the entry channel of the 40S ribosomal subunit, where it facilitates mRNA binding (Dong et al., 2017), whereas AKAP1 is a versatile protein involved in binding mRNA for local protein synthesis at the mitochondrial outer membrane (Cohen et al., 2024). Additionally, three RNA helicases (Bohnsack et al., 2023), DHX30, DDX3X, and UPF1, were persistently enriched. Two of these are linked to translation; a cytoplasmic form of DHX30 is involved in coordinating cytoplasmic translation and mitochondrial function (Bosco et al., 2021). DDX3X is strongly associated with the initiation stage of translation (Ryan & Schroder, 2022) and has been shown to bind the NRAS mRNA 5' UTR similarly to NSUN5 (Herdy et al., 2018). These findings suggest potential explanations for the known connection between m⁵C and mRNA translation (Huang et al, 2019; Schumann et al, 2020; Liu et al, 2021). Lastly, UPF1 is a known factor triggering the decay of m⁶A:mRNA via interacting with the m⁶A reader YTHDF2 (Boo et al., 2022), providing further links between m⁵C and m⁶A.

While all these RBPs are interesting candidates for further investigation into possible new m⁵C reader proteins, we focused our attention on one of them, UPF1. A systematic screening of the other top candidate RBPs will be part of future projects.

UPF1 is an ATP-dependent helicase with a known role in regulating mRNA decay in various mechanism. The most studied pathway in which UPF1 has a central role is NMD, a translation-dependent mechanism that aims to degrade those mRNA with a premature stop codon in their coding sequence (Kurosaki et al., 2019). These PTCs are often the result of genomic mutations or intron retention, and can lead to the synthesis of a truncated protein potentially deleterious for the cell. Once NMD gets activated, UPF1 is hyperphosphorylated and initiates the degradation of target mRNAs via de-capping, de-adenylation and endonuclease cleavage

of the RNA molecule (Kurosaki et al., 2019). A second pathway for mRNA degradation mediated by UPF1 is dependent of overall structuredness of the 3' UTR of mRNA (SRD) (Fischer et al., 2020). In this pathway, UPF1 can recognise and bind to highly structured 3' UTRs and, in complex with G3BP1 protein, can initiate mRNA decay. The molecular mechanism of SRD is currently unknown, although it has been proposed that UPF1/G3BP1 complex can recruit an endonuclease enzyme that degrades the mRNA target (Fischer et al., 2020). Despite the body of literature studying the complex mechanisms in which UPF1 can regulate mRNA steady state, a number a mRNAs have been reported to be targeted by UPF1 binding and can undergo UPF1-mediated decay, despite not having any of the features typically recognised in NMD and SRD (Nasif et al., 2018). In this gap of knowledge, we propose that m⁵C modification can alter UPF1 function in regulating mRNA steady state (**Figure 24**).

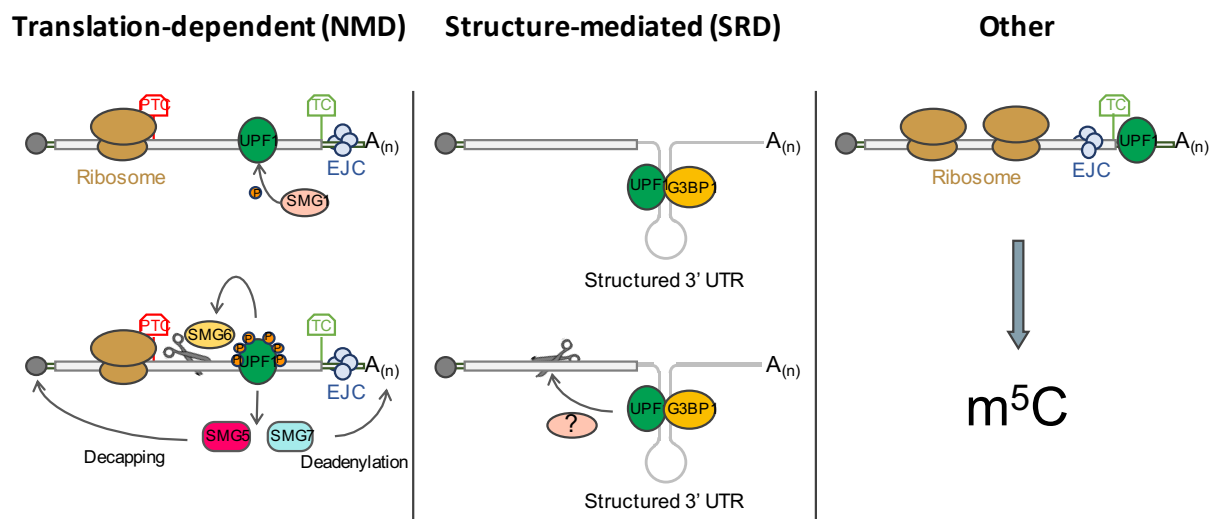


Figure 24 - Schematic representation of different pathways by which UPF1 regulates mRNA decay.

We decided to analyse the steady state level of a subgroup of UPF1 targets that carry an m⁵C modification within a 50-nucleotide region from the middle of UPF1 footprint. By performing UPF1 knockdown experiment in WT vs NSUN2 KO cells, we found that the steady state of these mRNAs was differentially regulated by UPF1 depending on NSUN2 expression level and therefore their methylation content. In particular, m⁵C:mRNAs turnover

was triggered by UPF1 only in WT cells, while upon NSUN2 KO the same non-methylated mRNAs were not affected by UPF1 dependent decay.

In addition, UPF1 CLIP experiments performed in WT and NSUN2 KO HeLa cells revealed that the binding affinity of the same mRNAs to UPF1 protein was dependent on their m⁵C content. In particular, we found that UPF1 could bind to these mRNAs more strongly when they lack methylation (NSUN2 KO), compared to when they are physiologically methylated (WT). While UPF1 does not act as a canonical m⁵C reader, as its binding affinity to mRNA increases upon NSUN2 KO (as opposed to decreasing), our data indicate a functional association between NSUN2-dependent m⁵C modification and UPF1 regulation of mRNA decay. The mechanism by which m⁵C modification affects UPF1 is currently unknown, therefore further research is needed. It is known that m⁵C modification can stabilise RNA structures by promoting base stacking and by increasing the thermal stability of hydrogen bonding in C:G pairing (Hayrapetyan et al., 2009; Motorin & Helm, 2010). We hypothesise that a change in methylation content obtained upon NSUN2 KO could lead to a change in the secondary structure of the area recognised and bound by UPF1, which could lead to UPF1 staying in an inactive state. When inactive, UPF1's helicase and ATPase centres are sterically inhibited, resulting in UPF1 clamping to RNA (Chakrabarti et al., 2011; Franks et al., 2010). This hypothesised situation can explain the lack of regulation by UPF1 but increased binding upon NSUN2 KO. In the future, further experiments are required to confirm the validity of this hypothesis and to elucidate the potential mechanism behind m⁵C dependent UPF1 function, perhaps by assessing the interaction of UPF1 with other core proteins regulating mRNA decay in WT and NSUN2 KO conditions.

Abbreviation Table

Abbreviation Table	
Definition	Abbreviation
Bisulfite RNA Sequencing	bsRNA-Seq
Complementary DNA	cDNA
Coding Sequence	CDS
Crosslinking Immunoprecipitation	CLIP
Copy Number Variation	CNV
CRISPR RNA	crRNA
Dulbecco's Modified Eagle Medium	DMEM
Double-stranded Break	DSB
Ethylenediaminetetraacetic Acid	EDTA
Enhancer RNA	eRNA
Fetal Bovine Serum	FBS
Genomic DNA	gDNA
Hepatocellular Carcinoma	HCC
5-hydroxymethylcytosine	hm ⁵ C
Head and Neck Squamous Cell Carcinoma	HNSCC
Knockdown	KD
Knockout	KO
Liquid Chromatography-Mass Spectrometry	LC-MS
Long non-coding RNA	lncRNA
5-methylcytosine	m ⁵ C
N ⁶ -methyladenosine	m ⁶ A
mouse Embryonic Stem Cells	mESCs
micro-RNA	miRNA
messenger RNA	mRNA
mitochondrial RNA	mtRNA
Maternal-to-Zygotic Transition	MZT
non-coding RNA	ncRNA
Non-Homologous End Joining	NHEJ
Nonsense-Mediated Decay	NMD
Overexpression	OE
Oxygen-Glucose Deprivation and Reoxygenation	OGD/R
Open Reading Frame	ORF
Photoactivatable Ribonucleoside enhanced	PAR-CLIP

Crosslinking and Immunoprecipitation	
Premature Termination Codon	PTC
Position Weight Matrix	PWM
RNA-Binding Proteins	RBP
RNA-Induced Silencing Complex	RISC
Ribosomal RNA	rRNA
Short interfering RNA	siRNA
Structure-Mediated Decay	SRD
transfer RNA	tRNA
Urothelial Carcinoma of the Bladder	UCB
Untranslated Region	UTR
vault RNA	vtRNA
Wild Type	WT

6. References

- Adams, R. L., & Wente, S. R. (2013). Uncovering nuclear pore complexity with innovation. *Cell*, 152(6), 1218–1221. <https://doi.org/10.1016/j.cell.2013.02.042>
- Aguilo, F., Li, S. De, Balasubramaniyan, N., Sancho, A., Benko, S., Zhang, F., Vashisht, A., Rengasamy, M., Andino, B., Chen, C. hung, Zhou, F., Qian, C., Zhou, M. M., Wohlschlegel, J. A., Zhang, W., Suchy, F. J., & Walsh, M. J. (2016). Deposition of 5-Methylcytosine on Enhancer RNAs Enables the Coactivator Function of PGC-1 α . *Cell Reports*, 14(3), 479–492. <https://doi.org/10.1016/j.celrep.2015.12.043>
- Amort, T., Rieder, D., Wille, A., Khokhlova-Cubberley, D., Riml, C., Trixl, L., Jia, X. Y., Micura, R., & Lusser, A. (2017). Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biology*, 18(1), 1. <https://doi.org/10.1186/s13059-016-1139-1>
- Amort, T., Soulière, M. F., Wille, A., Jia, X. Y., Fiegl, H., Wörle, H., Micura, R., & Lusser, A. (2013). Long non-coding RNAs as targets for cytosine methylation. *RNA Biology*, 10(6), 1002–1008. <https://doi.org/10.4161/rna.24454>
- Arzumanyan, V. A., Dolgalev, G. V., Kurbatov, I. Y., Kiseleva, O. I., & Poverennaya, E. V. (2022). Epitranscriptome: Review of Top 25 Most-Studied RNA Modifications. *International Journal of Molecular Sciences*, 23(22), 13851. <https://doi.org/10.3390/IJMS232213851/S1>
- Auxilien, S., Guérineau, V., Szweykowska-Kulińska, Z., & Golinelli-Pimpaneau, B. (2012). The human tRNA m5C methyltransferase Misu is multisite-specific. *RNA Biology*, 9(11), 1331–1338. <https://doi.org/10.4161/rna.22180>
- Bai, M., & Sun, C. (2022). M5C-Related lncRNA Predicts Lung Adenocarcinoma and Tumor Microenvironment Remodeling: Computational Biology and Basic Science. *Frontiers in Cell and Developmental Biology*, 10(3), 885568. <https://doi.org/10.3389/fcell.2022.885568>
- Barbieri, I., & Kouzarides, T. (2020). Role of RNA modifications in cancer. In *Nature Reviews Cancer* 20(6), 303–322. <https://doi.org/10.1038/s41568-020-0253-2>
- Baßler, J., & Hurt, E. (2019). Annual Review of Biochemistry Eukaryotic Ribosome Assembly. *Annual Review of Biochemistry*, 88(20), 281–306. <https://doi.org/10.1146/annurev-biochem-013118>
- Blanco, S., Dietmann, S., Flores, J. V., Hussain, S., Kutter, C., Humphreys, P., Lukk, M., Lombard, P., Treps, L., Popis, M., Kellner, S., Hölter, S. M., Garrett, L., Wurst, W., Becker, L., Klopstock, T., Fuchs, H., Gailus-Durner, V., Hrabě de Angelis, M., ... Frye, M. (2014). Aberrant methylation of t RNA s links cellular stress to neuro-developmental disorders. *The EMBO Journal*, 33(18), 2020–2039. <https://doi.org/10.15252/emj.201489282>
- Blanco, S., Kurowski, A., Nichols, J., Watt, F. M., Benitah, S. A., & Frye, M. (2011). The RNA-methyltransferase misu (NSun2) poises epidermal stem cells to differentiate. *PLoS Genetics*, 7(12), 1002403. <https://doi.org/10.1371/journal.pgen.1002403>
- Boccaletto, P., Stefaniak, F., Ray, A., Cappannini, A., Mukherjee, S., Purta, E., Kurkowska, M., Shirvanizadeh, N., Destefanis, E., Groza, P., Avşar, G., Romitelli, A., Pir, P., Dassi, E., Conticello, S. G., Aguilo, F., & Bujnicki, J. M. (2022). MODOMICS: A database of RNA modification pathways. 2021 update. *Nucleic Acids Research*, 50(D1), D231–D235. <https://doi.org/10.1093/nar/gkab1083>
- Bohnsack, K. E., Höbartner, C., & Bohnsack, M. T. (2019). Eukaryotic 5-methylcytosine (M5C) RNA methyltransferases: Mechanisms, cellular functions, and links to disease. In *Genes* 10(2), 102. <https://doi.org/10.3390/genes10020102>

- Bohnsack, K. E., Yi, S., Venus, S., Jankowsky, E., & Bohnsack, M. T. (2023). Cellular functions of eukaryotic RNA helicases and their links to human diseases. *Nature Reviews Molecular Cell Biology* 24:10, 24(10), 749–769. <https://doi.org/10.1038/s41580-023-00628-5>
- Boo, S. H., Ha, H., Lee, Y., Shin, M. K., Lee, S., & Kim, Y. K. (2022). UPF1 promotes rapid degradation of m6A-containing RNAs. *Cell Reports*, 39(8), 110861. <https://doi.org/10.1016/J.CELREP.2022.110861>
- Bosco, B., Rossi, A., Rizzotto, D., Hamadou, M. H., Bisio, A., Giorgetta, S., Perzolli, A., Bonollo, F., Gaucherot, A., Catez, F., Diaz, J. J., Dassi, E., & Inga, A. (2021). DHX30 coordinates cytoplasmic translation and mitochondrial function contributing to cancer cell survival. *Cancers*, 13(17), 4412. <https://doi.org/10.3390/CANCERS13174412/S1>
- Bourgeois, G., Ney, M., Gaspar, I., Aigueperse, C., Schaefer, M., Kellner, S., Helm, M., & Motorin, Y. (2015). Eukaryotic rRNA modification by yeast 5- methylcytosine-methyltransferases and human proliferation-associated antigen p120. *PLoS ONE*, 10(7), e0133321. <https://doi.org/10.1371/journal.pone.0133321>
- Brower, J. V., Clark, P. A., Lyon, W., & Kuo, J. S. (2014). MicroRNAs in cancer: Glioblastoma and glioblastoma cancer stem cells. In *Neurochemistry International* 77(68), 77. Pergamon. <https://doi.org/10.1016/j.neuint.2014.06.002>
- Brzezicha, B., Schmidt, M., Makołowska, I., Jarmołoski, A., Pieńkowska, J., & Szweykowska-Kulińska, Z. (2006). Identification of human tRNA: m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the pre-tRNA(CAA)Leu. *Nucleic Acids Research*, 34(20), 6034–6043. <https://doi.org/10.1093/nar/gkl765>
- Burgess, A. L., David, R., & Searle, I. R. (2015). Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. *BMC Plant Biology*, 15(1), 1–17. <https://doi.org/10.1186/s12870-015-0580-8>
- Cai, X., Hu, Y., Tang, H., Hu, H., Pang, L., Xing, J., Liu, Z., Luo, Y., Jiang, B., Liu, T., Gorospe, M., Chen, C., & Wang, W. (2016). RNA methyltransferase NSUN2 promotes stress-induced HUVEC senescence. *Oncotarget*, 7(15), 19099–19110. <https://doi.org/10.18632/oncotarget.8087>
- Cámara, Y., Asin-Cayuela, J., Park, C. B., Metodiev, M. D., Shi, Y., Ruzzenente, B., Kukat, C., Habermann, B., Wibom, R., Hultenby, K., Franz, T., Erdjument-Bromage, H., Tempst, P., Hallberg, B. M., Gustafsson, C. M., & Larsson, N. G. (2011). MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. *Cell Metabolism*, 13(5), 527–539. <https://doi.org/10.1016/j.cmet.2011.04.002>
- Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., Vendeix, F. A. P., Fabris, D., & Agris, P. F. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Research*, 39(SUPPL. 1), D195-D201. <https://doi.org/10.1093/nar/gkq1028>
- Cayir, A., Byun, H. M., & Barrow, T. M. (2020). Environmental epitranscriptomics. In *Environmental Research* 189, 109885). <https://doi.org/10.1016/j.envres.2020.109885>
- Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., & Conti, E. (2011). Molecular Mechanisms for the RNA-Dependent ATPase Activity of Upf1 and Its Regulation by Upf2. *Molecular Cell*, 41(6), 693–703. <https://doi.org/10.1016/j.molcel.2011.02.010>
- Chan, C. T. Y., Pang, Y. L. J., Deng, W., Babu, I. R., Dyavaiah, M., Begley, T. J., & Dedon, P. C. (2012). Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nature Communications*, 3(1), 1–9. <https://doi.org/10.1038/ncomms1938>

- Chandra Gupta, S., & Nandan Tripathi, Y. (2017). Potential of long non-coding RNAs in cancer patients: From biomarkers to therapeutic targets. In *International Journal of Cancer* (Vol. 140, Issue 9, pp. 1955–1967). <https://doi.org/10.1002/ijc.30546>
- Chellamuthu, A., & Gray, S. G. (2020). The RNA Methyltransferase NSUN2 and Its Potential Roles in Cancer. In *Cells* 9(8), 1758. Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/cells9081758>
- Chen, H., Yang, H., Zhu, X., Yadav, T., Ouyang, J., Truesdell, S. S., Tan, J., Wang, Y., Duan, M., Wei, L., Zou, L., Levine, A. S., Vasudevan, S., & Lan, L. (2020). m5C modification of mRNA serves a DNA damage code to promote homologous recombination. *Nature Communications*, 11(1), 2834. <https://doi.org/10.1038/s41467-020-16722-7>
- Chen, P., Zhang, T., Yuan, Z., Shen, B., & Chen, L. (2019). Expression of the RNA methyltransferase Nsun5 is essential for developing cerebral cortex. *Molecular Brain*, 12(1), 1–14. <https://doi.org/10.1186/s13041-019-0496-6>
- Chen, X., Li, A., Sun, B. F., Yang, Y., Han, Y. N., Yuan, X., Chen, R. X., Wei, W. S., Liu, Y., Gao, C. C., Chen, Y. S., Zhang, M., Ma, X. D., Liu, Z. W., Luo, J. H., Lyu, C., Wang, H. L., Ma, J., Zhao, Y. L., ... Yang, Y. G. (2019). 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nature Cell Biology*, 21(8), 978–990. <https://doi.org/10.1038/s41556-019-0361-y>
- Chen, Y. S., Yang, W. L., Zhao, Y. L., & Yang, Y. G. (2021). Dynamic transcriptomic m5C and its regulatory role in RNA processing. *Wiley Interdisciplinary Reviews: RNA* 12(4), e1639. <https://doi.org/10.1002/wrna.1639>
- Cheray, M., Etcheverry, A., Jacques, C., Pacaud, R., Bougras-Cartron, G., Aubry, M., Denoual, F., Peterlongo, P., Nadaradjane, A., Briand, J., Akcha, F., Heymann, D., Vallette, F. M., Mosser, J., Ory, B., & Cartron, P. F. (2020). Cytosine methylation of mature microRNAs inhibits their functions and is associated with poor prognosis in glioblastoma multiforme. *Molecular Cancer*, 19(1), 1–16. <https://doi.org/10.1186/s12943-020-01155-z>
- Chi, L., & Delgado-Olguín, P. (2013). Expression of NOL1/NOP2/sun domain (Nsun) RNA methyltransferase family genes in early mouse embryogenesis. *Gene Expression Patterns*, 13(8), 319–327. <https://doi.org/10.1016/j.gep.2013.06.003>
- Cohen, B., Golani-Armon, A., & Arava, Y. S. (2024). Emerging implications for ribosomes in proximity to mitochondria. *Seminars in Cell & Developmental Biology*, 154(Pt B), 123–130. <https://doi.org/10.1016/J.SEMCDB.2023.01.003>
- Cohn, W. E. (1960). Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *The Journal of Biological Chemistry*, 235(5), 1488–1498. [https://doi.org/10.1016/s0021-9258\(18\)69432-3](https://doi.org/10.1016/s0021-9258(18)69432-3)
- Courtney, D. G., Chalem, A., Bogerd, H. P., Law, B. A., Kennedy, E. M., Holley, C. L., & Cullen, B. R. (2019). Extensive epitranscriptomic methylation of A and C residues on murine leukemia virus transcripts enhances viral gene expression. *MBio*, 10(3), e01209-19. <https://doi.org/10.1128/mBio.01209-19>
- Courtney, D. G., Tsai, K., Bogerd, H. P., Kennedy, E. M., Law, B. A., Emery, A., Swanstrom, R., Holley, C. L., & Cullen, B. R. (2019). Epitranscriptomic Addition of m5C to HIV-1 Transcripts Regulates Viral Gene Expression. *Cell Host and Microbe*, 26(2), 217–227. <https://doi.org/10.1016/j.chom.2019.07.005>
- Cui, X., Liang, Z., Shen, L., Zhang, Q., Bao, S., Geng, Y., Zhang, B., Leo, V., Vardy, L. A., Lu, T., Gu, X., & Yu, H. (2017). 5-Methylcytosine RNA Methylation in Arabidopsis Thaliana. *Molecular Plant*, 10(11), 1387–1399. <https://doi.org/10.1016/j.molp.2017.09.013>

- Cusenza, V. Y., Tameni, A., Neri, A., & Frazzi, R. (2023). The lncRNA epigenetics: The significance of m6A and m5C lncRNA modifications in cancer. In *Frontiers in Oncology* 13(9), 1063636. Frontiers Media SA. <https://doi.org/10.3389/fonc.2023.1063636>
- David, R., Burgess, A., Parker, B., Li, J., Pulsford, K., Sibbritt, T., Preiss, T., & Searle, I. R. (2017). Transcriptome-wide mapping of RNA 5-methylcytosine in arabidopsis mRNAs and noncoding RNAs. *Plant Cell*, 29(3), 445–460. <https://doi.org/10.1105/tpc.16.00751>
- Del Valle-Morales, D., Le, P., Saviana, M., Romano, G., Nigita, G., Nana-Sinkam, P., & Acunzo, M. (2022). The Epitranscriptome in miRNAs: Crosstalk, Detection, and Function in Cancer. *Genes* 13(7), 1289. Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/genes13071289>
- Di Timoteo, G., Dattilo, D., Centrón-Broco, A., Colantoni, A., Guarnacci, M., Rossi, F., Incarnato, D., Oliviero, S., Fatica, A., Morlando, M., & Bozzoni, I. (2020). Modulation of circRNA Metabolism by m6A Modification. *Cell Reports*, 31(6), 107641. <https://doi.org/10.1016/j.celrep.2020.107641>
- Dong, J., Aitken, C. E., Thakur, A., Shin, B. S., Lorsch, J. R., & Hinnebusch, A. G. (2017). Rps3/uS3 promotes mRNA binding at the 40S ribosome entry channel and stabilizes preinitiation complexes at start codons. *Proceedings of the National Academy of Sciences of the United States of America*, 114(11), E2126–E2135. <https://doi.org/10.1073/pnas.1620569114>
- Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., & Wu, L. (2016). YTHDF2 destabilizes m 6 A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nature Communications*, 7(1), 12626. <https://doi.org/10.1038/ncomms12626>
- Dubin, D. T., & Taylor, R. H. (1975). The methylation state of poly A-containing-messenger RNA from cultured hamster cells. *Nucleic Acids Research*, 2(10), 1653–1668. <https://doi.org/10.1093/nar/2.10.1653>
- Dufu, K., Livingstone, M. J., Seebacher, J., Gygi, S. P., Wilson, S. A., & Reed, R. (2010). ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes and Development*, 24(18), 2043–2053. <https://doi.org/10.1101/gad.1898610>
- Durdevic, Z., Hanna, K., Gold, B., Pollex, T., Cherry, S., Lyko, F., & Schaefer, M. (2013). Efficient RNA virus control in Drosophila requires the RNA methyltransferase Dnmt2. *EMBO Reports*, 14(3), 269–275. <https://doi.org/10.1038/embor.2013.3>
- Fischer, J. W., Busa, V. F., Shao, Y., & Leung, A. K. L. (2020). Structure-Mediated RNA Decay by UPF1 and G3BP1. *Molecular Cell*, 78(1), 70–84. <https://doi.org/10.1016/j.molcel.2020.01.021>
- Flores, J. V., Cordero-Espinoza, L., Oeztuerk-Winder, F., Andersson-Rolf, A., Selmi, T., Blanco, S., Tailor, J., Dietmann, S., & Frye, M. (2017). Cytosine-5 RNA Methylation Regulates Neural Stem Cell Differentiation and Motility. *Stem Cell Reports*, 8(1), 112–124. <https://doi.org/10.1016/j.stemcr.2016.11.014>
- Franks, T. M., Singh, G., & Lykke-Andersen, J. (2010). Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense-mediated mRNA decay. *Cell*, 143(6), 938–950. <https://doi.org/10.1016/j.cell.2010.11.043>
- Frye, M., Harada, B. T., Behm, M., & He, C. (2018). RNA modifications modulate gene expression during development. *Science*, 361(6409), 1346–1349. <https://doi.org/10.1126/science.aau1646>
- Frye, M., & Watt, F. M. (2006). The RNA Methyltransferase Misu (NSun2) Mediates Myc-Induced Proliferation and Is Upregulated in Tumors. *Current Biology*, 16(10), 971–981. <https://doi.org/10.1016/j.cub.2006.04.027>

- Fu, L., Guerrero, C. R., Zhong, N., Amato, N. J., Liu, Y., Liu, S., Cai, Q., Ji, D., Jin, S. G., Niedernhofer, L. J., Pfeifer, G. P., Xu, G. L., & Wang, Y. (2014). Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *Journal of the American Chemical Society*, 136(33), 11582–11585. <https://doi.org/10.1021/ja505305z>
- Gao, Y., & Fang, J. (2021). RNA 5-methylcytosine modification and its emerging role as an epitranscriptomic mark. In *RNA Biology* 18(S1), 117–127. <https://doi.org/10.1080/15476286.2021.1950993>
- Gao, Y., Wang, Z., Zhu, Y., Zhu, Q., Yang, Y., Jin, Y., Zhang, F., Jiang, L., Ye, Y., Li, H., Zhang, Y., Liang, H., Xiang, S., Miao, H., Liu, Y., & Hao, Y. (2019). NOP2/Sun RNA methyltransferase 2 promotes tumor progression via its interacting partner RPL6 in gallbladder carcinoma. *Cancer Science*, 110(11), 3510–3519. <https://doi.org/10.1111/cas.14190>
- Gilbert, W. V., & Nachtergaele, S. (2023). mRNA Regulation by RNA Modifications. *Annual Review of Biochemistry*, 92(1), 175–198. <https://doi.org/10.1146/annurev-biochem-052521-035949>
- Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., Golic, K. G., Jacobsen, S. E., & Bestor, T. H. (2006). Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science*, 311(5759), 395–398. <https://doi.org/10.1126/science.1120976>
- Guarnacci, M., & Preiss, T. (2024). The je ne sais quoi of 5-methylcytosine in messenger RNA. *RNA*, 30(5), 560–569. <https://doi.org/10.1261/RNA.079982.124>
- Guarnacci, M., Zhang, P.-H., Kanchi, M., Hung, Y.-T., Lin, H., Shirokikh, N. E., Yang, L., & Preiss, T. (2024). Substrate diversity of NSUN enzymes and links of 5-methylcytosine to mRNA translation and turnover. *Life Science Alliance*, 7(9), e202402613. <https://doi.org/10.26508/LSA.202402613>
- Guo, G., Wang, H., Shi, X., Ye, L., Yan, K., Chen, Z., Zhang, H., Jin, Z., & Xue, X. (2020). Disease Activity-Associated Alteration of mRNA m5 C Methylation in CD4⁺ T Cells of Systemic Lupus Erythematosus. *Frontiers in Cell and Developmental Biology*, 8(5), 430. <https://doi.org/10.3389/fcell.2020.00430>
- Haag, S., Sloan, K. E., Ranjan, N., Warda, A. S., Kretschmer, J., Blessing, C., Hübner, B., Seikowski, J., Dennerlein, S., Rehling, P., Rodnina, M. V., Höbartner, C., & Bohnsack, M. T. (2016). NSUN 3 and ABH 1 modify the wobble position of mt-t RNA Met to expand codon recognition in mitochondrial translation. *The EMBO Journal*, 35(19), 2104–2119. <https://doi.org/10.15252/embj.201694885>
- Haag, S., Warda, A. S., Kretschmer, J., Günnigmann, M. A., Höbartner, C., & Bohnsack, M. T. (2015). NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. *RNA*, 21(9), 1532–1543. <https://doi.org/10.1261/rna.051524.115>
- Harris, T., Marquez, B., Suarez, S., & Schimenti, J. (2007). Sperm motility defects and infertility in male mice with a mutation in Nsun7, a member of the sun domain-containing family of putative RNA methyltransferases. *Biology of Reproduction*, 77(2), 376–382. <https://doi.org/10.1095/biolreprod.106.058669>
- Hayrapetyan, A., Grosjean, H., & Helm, M. (2009). Effect of a quaternary pentamine on RNA stabilization and enzymatic methylation. *Biological Chemistry*, 390(9), 851–861. <https://doi.org/10.1515/BC.2009.096>
- He, P. C., & He, C. (2021). m6A RNA methylation: from mechanisms to therapeutic potential. *The EMBO Journal*, 40(3), e105977. <https://doi.org/10.15252/embj.2020105977>

- He, Y., Yu, X., Li, J., Zhang, Q., Zheng, Q., & Guo, W. (2020). Role of m5C-related regulatory genes in the diagnosis and prognosis of hepatocellular carcinoma. *American Journal of Translational Research*, 12(3), 912–922.
- He, Y., Yu, X., Zhang, M., & Guo, W. (2021). Pan-cancer analysis of m5C regulator genes reveals consistent epigenetic landscape changes in multiple cancers. *World Journal of Surgical Oncology*, 19(1), 1–12. <https://doi.org/10.1186/s12957-021-02342-y>
- Heissenberger, C., Liendl, L., Nagelreiter, F., Gonskikh, Y., Yang, G., Stelzer, E. M., Krammer, T. L., Micutkova, L., Vogt, S., Kreil, D. P., Sekot, G., Siena, E., Poser, I., Harreither, E., Linder, A., Ehret, V., Helbich, T. H., Grillari-Voglauer, R., Jansen-Dürr, P., ... Schosserer, M. (2019). Loss of the ribosomal RNA methyltransferase NSUN5 impairs global protein synthesis and normal growth. *Nucleic Acids Research*, 47(22), 11807–11825. <https://doi.org/10.1093/nar/gkz1043>
- Herdy, B., Mayer, C., Varshney, D., Marsico, G., Murat, P., Taylor, C., D'Santos, C., Tannahill, D., & Balasubramanian, S. (2018). Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Research*, 46(21), 11592–11604. <https://doi.org/10.1093/NAR/GKY861>
- Holley, R. W., Everett, G. A., Madison, J. T., & ZAMIR, A. (1965). Nucleotide sequences in the yeast alanine transfer ribonucleic acid. *The Journal of Biological Chemistry*, 240(5), 2122–2128. [https://doi.org/10.1016/s0021-9258\(18\)97435-1](https://doi.org/10.1016/s0021-9258(18)97435-1)
- Hu, J., Manduzio, S., & Kang, H. (2019). Epitranscriptomic RNA methylation in plant development and abiotic stress responses. In *Frontiers in Plant Science* 10(17), 500. Frontiers Media S.A. <https://doi.org/10.3389/fpls.2019.00500>
- Hu, Y., Chen, C., Tong, X., Chen, S., Hu, X., Pan, B., Sun, X., Chen, Z., Shi, X., Hu, Y., Shen, X., Xue, X., & Lu, M. (2021). NSUN2 modified by SUMO-2/3 promotes gastric cancer progression and regulates mRNA m5C methylation. *Cell Death and Disease*, 12(9), 495. <https://doi.org/10.1038/s41419-021-04127-3>
- Huang, T., Chen, W., Liu, J., Gu, N., & Zhang, R. (2019). Genome-wide identification of mRNA 5-methylcytosine in mammals. *Nat Struct Mol Biol*, 26(26), 380–388. <https://doi.org/10.1038/s41594-019-0218-x>
- Hussain, S., Sajini, A. A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., Paramor, M., Gleeson, J. G., Odom, D. T., Ule, J., & Frye, M. (2013). NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Reports*, 4(2), 255–261. <https://doi.org/10.1016/j.celrep.2013.06.029>
- Hussain, S., Tuorto, F., Menon, S., Blanco, S., Cox, C., Flores, J. V., Watt, S., Kudo, N. R., Lyko, F., & Frye, M. (2013). The Mouse Cytosine-5 RNA Methyltransferase NSun2 Is a Component of the Chromatoid Body and Required for Testis Differentiation. *Molecular and Cellular Biology*, 33(8), 1561–1570. <https://doi.org/10.1128/mcb.01523-12>
- Janin, M., Ortiz-Barahona, V., de Moura, M. C., Martínez-Cardús, A., Llinàs-Arias, P., Soler, M., Nachmani, D., Pelletier, J., Schumann, U., Calleja-Cervantes, M. E., Moran, S., Guil, S., Bueno-Costa, A., Piñeyro, D., Perez-Salvia, M., Rosselló-Tortella, M., Piqué, L., Bech-Serra, J. J., De La Torre, C., ... Esteller, M. (2019). Epigenetic loss of RNA-methyltransferase NSUN5 in glioma targets ribosomes to drive a stress adaptive translational program. *Acta Neuropathologica*, 138(6), 1053–1074. <https://doi.org/10.1007/S00401-019-02062-4>
- Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y. G., & He, C. (2011). N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chemical Biology*, 7(12), 885–887. <https://doi.org/10.1038/nchembio.687>

- Jian, H., Zhang, C., Qi, Z. Y., Li, X., Lou, Y., Kang, Y., Deng, W., Lv, Y., Wang, C., Wang, W., Shang, S., Hou, M., Zhou, H., & Feng, S. (2021). Alteration of mRNA 5-Methylcytosine Modification in Neurons After OGD/R and Potential Roles in Cell Stress Response and Apoptosis. *Frontiers in Genetics*, *12*, 633681. <https://doi.org/10.3389/fgene.2021.633681>
- Jiang, Z., Li, S., Han, M.-J., Hu, G.-M., & Cheng, P. (2020). High expression of NSUN5 promotes cell proliferation via cell cycle regulation in colorectal cancer. *Am J Transl Res*, *12*(7), 3858–3870.
- Johnson, T. B., & Coghill, R. D. (1925). Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *Journal of the American Chemical Society*, *47*(11), 2838–2844. <https://doi.org/10.1021/ja01688a030>
- Kataoka, N. (2024). The Nuclear Cap-Binding Complex, a multitasking binding partner of RNA polymerase II transcripts. *The Journal of Biochemistry*, *175*(1), 9–15. <https://doi.org/10.1093/JB/MVAD081>
- Kawarada, L., Suzuki, T., Ohira, T., Hirata, S., Miyauchi, K., & Suzuki, T. (2017). ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications. *Nucleic Acids Research*, *45*(12), 7401–7415. <https://doi.org/10.1093/nar/gkx354>
- Kennedy, E. M., Bogerd, H. P., Kornepati, A. V. R., Kang, D., Ghoshal, D., Marshall, J. B., Poling, B. C., Tsai, K., Gokhale, N. S., Horner, S. M., & Cullen, B. R. (2017). Erratum: Posttranscriptional m6A Editing of HIV-1 mRNAs Enhances Viral Gene Expression. *Cell Host and Microbe* *22*(6), 830. <https://doi.org/10.1016/j.chom.2017.11.010>
- Kohli, R. M., & Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* *502*(7472), 472–479. <https://doi.org/10.1038/nature12750>
- Kong, W., Biswas, A., Zhou, D., Fiches, G., Fujinaga, K., Santoso, N., & Zhu, J. (2020). Nucleolar protein NOP2/NSUN1 suppresses HIV-1 transcription and promotes viral latency by competing with Tat for TAR binding and methylation. *PLoS Pathogens*, *16*(3), e1008430. <https://doi.org/10.1371/journal.ppat.1008430>
- Konno, M., Koseki, J., Asai, A., Yamagata, A., Shimamura, T., Motooka, D., Okuzaki, D., Kawamoto, K., Mizushima, T., Eguchi, H., Takiguchi, S., Satoh, T., Mimori, K., Ochiya, T., Doki, Y., Ofusa, K., Mori, M., & Ishii, H. (2019). Distinct methylation levels of mature microRNAs in gastrointestinal cancers. *Nature Communications*, *10*(1), 1–7. <https://doi.org/10.1038/s41467-019-11826-1>
- Kurosaki, T., Popp, M. W., & Maquat, L. E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. In *Nature Reviews Molecular Cell Biology* *20*(7), 406–420. <https://doi.org/10.1038/s41580-019-0126-2>
- Lavysh, D., & Neu-Yilik, G. (2020). UPF1-mediated RNA decay—danse macabre in a cloud. In *Biomolecules* *10*(7), 1–34. <https://doi.org/10.3390/biom10070999>
- Lee, S. R., Pratt, G. A., Martinez, F. J., Yeo, G. W., & Lykke-Andersen, J. (2015). Target Discrimination in Nonsense-Mediated mRNA Decay Requires Upf1 ATPase Activity. *Molecular Cell*, *59*(3), 413–425. <https://doi.org/10.1016/j.molcel.2015.06.036>
- Li, J., Li, H., Long, T., Dong, H., Wang, E. D., & Liu, R. J. (2019). Archaeal NSUN6 catalyzes m5C72 modification on a wide-range of specific tRNAs. *Nucleic Acids Research*, *47*(4), 2041–2055. <https://doi.org/10.1093/nar/gky1236>
- Li, Q., Li, X., Tang, H., Jiang, B., Dou, Y., Gorospe, M., & Wang, W. (2017). NSUN2-Mediated m5C Methylation and METTL3/METTL14-Mediated m6A Methylation Cooperatively Enhance p21 Translation. *Journal of Cellular Biochemistry*, *118*(9), 2587–2598. <https://doi.org/10.1002/jcb.25957>

- Li, S., & Mason, C. E. (2014). The pivotal regulatory landscape of RNA modifications. *Annual Review of Genomics and Human Genetics*, 15, 127–150. <https://doi.org/10.1146/annurev-genom-090413-025405>
- Li, Y., Jiang, T., Zhou, W., Li, J., Li, X., Wang, Q., Jin, X., Yin, J., Chen, L., Zhang, Y., Xu, J., & Li, X. (2020). Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nature Communications*, 11(1), 1000. <https://doi.org/10.1038/s41467-020-14802-2>
- Li, Y., Kuscus, C., Banach, A., Zhang, Q., Pulkoski-Gross, A., Kim, D., Liu, J., Roth, E., Li, E., Shroyer, K. R., Denoya, P. I., Zhu, X., Chen, L., & Cao, J. (2015). miR-181a-5p inhibits cancer cell migration and angiogenesis via downregulation of matrix metalloproteinase-14. *Cancer Research*, 75(13), 2674–2685. <https://doi.org/10.1158/0008-5472.CAN-14-2875>
- Linder, B., & Jaffrey, S. R. (2019). Discovering and mapping the modified nucleotides that comprise the epitranscriptome of mRNA. *Cold Spring Harbor Perspectives in Biology*, 11(6), a032201. <https://doi.org/10.1101/cshperspect.a032201>
- Liu, J., Huang, T., Chen, W., Ding, C., Zhao, T., Zhao, X., Cai, B., Zhang, Y., Li, S., Zhang, L., Xue, M., He, X., Ge, W., Zhou, C., Xu, Y., & Zhang, R. (2022). Developmental mRNA m5C landscape and regulatory innovations of massive m5C modification of maternal mRNAs in animals. *Nature Communications*, 13(1), 2484. <https://doi.org/10.1038/s41467-022-30210-0>
- Liu, J., Huang, T., Zhang, Y., Zhao, T., Zhao, X., Chen, W., & Zhang, R. (2021a). Sequence- and structure-selective mRNA m5C methylation by NSUN6 in animals. *National Science Review*, 8(6), nwaa273. <https://doi.org/10.1093/NSR/NWAA273>
- Liu, Y., Zhao, Y., Wu, R., Chen, Y., Chen, W., Liu, Y., Luo, Y., Huang, C., Zeng, B., Liao, X., Guo, G., Wang, Y., & Wang, X. (2021b). mRNA m5C controls adipogenesis by promoting CDKN1A mRNA export and translation. *RNA Biology*, 18(S2), 711–721. <https://doi.org/10.1080/15476286.2021.1980694>
- Lloyd, J. P. B. (2018). The evolution and diversity of the nonsense-mediated mRNA decay pathway. *F1000Research*, 7(15), 1299. <https://doi.org/10.12688/F1000RESEARCH.15872.1>
- Long, T., Li, J., Li, H., Zhou, M., Zhou, X. L., Liu, R. J., & Wang, E. D. (2016). Sequence-specific and shape-selective RNA recognition by the human RNA 5-methylcytosine methyltransferase NSun. *Journal of Biological Chemistry*, 291(46), 24293–24303. <https://doi.org/10.1074/jbc.M116.742569>
- Lu, L., Gaffney, S. G., Cannataro, V. L., & Townsend, J. (2020). Transfer RNA methyltransferase gene NSUN2 mRNA expression modifies the effect of T cell activation score on patient survival in head and neck squamous carcinoma. *Oral Oncology*, 101, 104554. <https://doi.org/10.1016/j.oraloncology.2019.104554>
- Lu, L., Zhu, G., Zeng, H., Xu, Q., & Holzmam, K. (2018). High tRNA Transferase NSUN2 Gene Expression is Associated with Poor Prognosis in Head and Neck Squamous Carcinoma. *Cancer Investigation*, 36(4), 246–253. <https://doi.org/10.1080/07357907.2018.1466896>
- Luo, Y., Feng, J., Xu, Q., Wang, W., & Wang, X. (2016). NSun2 deficiency protects endothelium from inflammation via mRNA methylation of ICAM-1. *Circulation Research*, 118(6), 944–956. <https://doi.org/10.1161/CIRCRESAHA.115.307674>
- Lykke-Andersen, S., Chen, Y., Ardal, B. R., Lilje, B., Waage, J., Sandelin, A., & Jensen, T. H. (2014). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes and Development*, 28(22), 2498–2517. <https://doi.org/10.1101/gad.246538.114>

- Ma, H. L., Bizet, M., Soares Da Costa, C., Murisier, F., de Bony, E. J., Wang, M. K., Yoshimi, A., Lin, K. T., Riching, K. M., Wang, X., Beckman, J. I., Arya, S., Droin, N., Calonne, E., Hassabi, B., Zhang, Q. Y., Li, A., Putmans, P., Malbec, L., ... Fuks, F. (2023). SRSF2 plays an unexpected role as reader of m5C on mRNA, linking epitranscriptomics to cancer. *Molecular Cell*, 83(23), 4239–4254. <https://doi.org/10.1016/j.molcel.2023.11.003>
- Ma, J., Song, B., Wei, Z., Huang, D., Zhang, Y., Su, J., De Magalhães, J. P., Rigden, D. J., Meng, J., & Chen, K. (2022). M5C-Atlas: A comprehensive database for decoding and annotating the 5-methylcytosine (m5C) epitranscriptome. *Nucleic Acids Research*, 50(D1), D196–D203. <https://doi.org/10.1093/nar/gkab1075>
- Martinez, F. J., Lee, J. H., Lee, J. E., Blanco, S., Nickerson, E., Gabrie, S., Frye, M., Al-Gazali, L., & Gleeson, J. G. (2012). Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. *Journal of Medical Genetics*, 49(6), 380–385. <https://doi.org/10.1136/jmedgenet-2011-100686>
- Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D. P., Linder, B., Pickering, B. F., Vasseur, J. J., Chen, Q., Gross, S. S., Elemento, O., Debart, F., Kiledjian, M., & Jaffrey, S. R. (2017). Reversible methylation of m6Am in the 5' cap controls mRNA stability. *Nature*, 541(7637), 371–375. <https://doi.org/10.1038/NATURE21022>
- Maurya, P. K., Mishra, A., Yadav, B. S., Singh, S., Kumar, P., Chaudhary, A., Srivastava, S., Murugesan, S. N., & Mani, A. (2017). Role of Y box protein-1 in cancer: As potential biomarker and novel therapeutic target. In *Journal of Cancer*, 8(10), 1900–1907. Ivyspring International Publisher. <https://doi.org/10.7150/jca.17689>
- Metodiev, M. D., Spåhr, H., Loguercio Polosa, P., Meharg, C., Becker, C., Altmueller, J., Habermann, B., Larsson, N. G., & Ruzzenente, B. (2014). NSUN4 Is a Dual Function Mitochondrial Protein Required for Both Methylation of 12S rRNA and Coordination of Mitoribosomal Assembly. *PLoS Genetics*, 10(2), e1004110. <https://doi.org/10.1371/journal.pgen.1004110>
- Meyer, K. D., & Jaffrey, S. R. (2017). *Annual Review of Cell and Developmental Biology Rethinking m 6 A Readers, Writers, and Erasers*. 33(6), 319–342. <https://doi.org/10.1146/annurev-cellbio-100616>
- Moosavi, A., & Ardekani, A. M. (2016). Role of epigenetics in biology and human diseases. In *Iranian Biomedical Journal*, 20(5), 246–258. <https://doi.org/10.22045/ibj.2016.01>
- Motorin, Y., & Helm, M. (2010). tRNA stabilization by modified nucleotides. *Biochemistry* 49(24), 4934–4944. American Chemical Society. <https://doi.org/10.1021/bi100408z>
- Motorin, Y., & Helm, M. (2022). RNA nucleotide methylation: 2021 update. *Wiley Interdisciplinary Reviews: RNA* 13(1), 2236–51. John Wiley and Sons Inc. <https://doi.org/10.1002/wrna.1691>
- Murakami, S., & Jaffrey, S. R. (2022). Hidden codes in mRNA: Control of gene expression by m6A. *Molecular Cell* 82(12), 2236–2251. <https://doi.org/10.1016/j.molcel.2022.05.029>
- Nakano, S., Suzuki, T., Kawarada, L., Iwata, H., Asano, K., & Suzuki, T. (2016). NSUN3 methylase initiates 5-formylcytidine biogenesis in human mitochondrial tRNA Met. *Nature Chemical Biology*, 12(7), 546–551. <https://doi.org/10.1038/nchembio.2099>
- Nasif, S., Contu, L., & Mühlemann, O. (2018). Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression. *Seminars in Cell and Developmental Biology*, 75, 78–87. <https://doi.org/10.1016/j.semcdb.2017.08.053>
- Ortiz-Barahona, V., Soler, M., Davalos, V., García-Prieto, C. A., Janin, M., Setien, F., Fernández-Rebollo, I., Bech-Serra, J. J., De La Torre, C., Guil, S., Villanueva, A., Zhang, P. H., Yang, L., Guarnacci, M., Schumann, U., Preiss, T., Balaseviciute, U., Montal, R., Llovet, J. M., & Esteller, M. (2023). Epigenetic inactivation of the 5-

- methylcytosine RNA methyltransferase NSUN7 is associated with clinical outcome and therapeutic vulnerability in liver cancer. *Molecular Cancer*, 22(1), 83.
<https://doi.org/10.1186/s12943-023-01785-z>
- Pan, J., Huang, Z., & Xu, Y. (2021). m5C-Related lncRNAs Predict Overall Survival of Patients and Regulate the Tumor Immune Microenvironment in Lung Adenocarcinoma. *Frontiers in Cell and Developmental Biology*, 9, 671821.
<https://doi.org/10.3389/fcell.2021.671821>
- Ramazi, S., & Zahir, J. (2021). Post-translational modifications in proteins: Resources, tools and prediction methods. *Database (Oxford)* 2021, baab012.
<https://doi.org/10.1093/database/baab012>
- Reed, R., & Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell*, 108(4), 523–531. [https://doi.org/10.1016/S0092-8674\(02\)00627-X](https://doi.org/10.1016/S0092-8674(02)00627-X)
- Reid, R., Greene, P. J., & Santi, D. V. (1999). Exposition of a family of RNA m5C methyltransferases from searching genomic and proteomic sequences. *Nucleic Acids Research*, 27(15), 3138–3145. <https://doi.org/10.1093/nar/27.15.3138>
- Roundtree, I. A., Evans, M. E., Pan, T., & He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. *Cell* 169(7), 1187–1200.
<https://doi.org/10.1016/j.cell.2017.05.045>
- Sajini, A. A., Choudhury, N. R., Wagner, R. E., Bornelöv, S., Selmi, T., Spanos, C., Dietmann, S., Rappsilber, J., Michlewski, G., & Frye, M. (2019). Loss of 5-methylcytosine alters the biogenesis of vault-derived small RNAs to coordinate epidermal differentiation. *Nature Communications*, 10(1), 1–13.
<https://doi.org/10.1038/s41467-019-10020-7>
- Schaefer, M., Kapoor, U., & Jantsch, M. F. (2017). Understanding RNA modifications: The promises and technological bottlenecks of the ‘epitranscriptome’. *Open Biology*, 7(5), 170077. <https://doi.org/10.1098/rsob.170077>
- Schaefer, M., Pollex, T., Hanna, K., & Lyko, F. (2009). RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Research*, 37(2), 12.
<https://doi.org/10.1093/nar/gkn954>
- Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., & Lyko, F. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes and Development*, 24(15), 1590–1595.
<https://doi.org/10.1101/gad.586710>
- Schossere, M., Minois, N., Angerer, T. B., Amring, M., Dellago, H., Harreither, E., Calle-Perez, A., Pircher, A., Gerstl, M. P., Pfeifenberger, S., Brandl, C., Sonntagbauer, M., Kriegner, A., Linder, A., Weinhäusel, A., Mohr, T., Steiger, M., Mattanovich, D., Rinnerthaler, M., ... Grillari, J. (2015). Methylation of ribosomal RNA by NSUN5 is a conserved mechanism modulating organismal lifespan. *Nature Communications* 2015 6:1, 6(1), 1–17. <https://doi.org/10.1038/ncomms7158>
- Schumann, U., Zhang, H. N., Sibbritt, T., Pan, A., Horvath, A., Gross, S., Clark, S. J., Yang, L., & Preiss, T. (2020a). Multiple links between 5-methylcytosine content of mRNA and translation. *BMC Biology*, 18(1), 1–23. <https://doi.org/10.1186/S12915-020-00769-5/FIGURES/6>
- Selmi, T., Hussain, S., Dietmann, S., Heiß, M., Borland, K., Flad, S., Carter, J. M., Dennison, R., Huang, Y. L., Kellner, S., Bornelöv, S., & Frye, M. (2021). Sequence- And structure-specific cytosine-5 mRNA methylation by NSUN6. *Nucleic Acids Research*, 49(2), 1006–1022. <https://doi.org/10.1093/nar/gkaa1193>
- Shanmugam, R., Fierer, J., Kaiser, S., Helm, M., Jurkowski, T. P., & Jeltsch, A. (2015). Cytosine methylation of tRNA-Asp by DNMT2 has a role in translation of proteins

- containing poly-Asp sequences. *Cell Discovery*, 1(9), 15010. <https://doi.org/10.1038/celldisc.2015.10>
- Sharma, S., & Lafontaine, D. L. J. (2015). ‘View From A Bridge’: A New Perspective on Eukaryotic rRNA Base Modification. *Trends in Biochemical Sciences* 40(10), 560–575. <https://doi.org/10.1016/j.tibs.2015.07.008>
- Sharma, S., Yang, J., Watzinger, P., Kötter, P., & Entian, K. D. (2013). Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively. *Nucleic Acids Research*, 41(19), 9062–9076. <https://doi.org/10.1093/nar/gkt679>
- Shen, Q., Zhang, Q., Shi, Y., Shi, Q., Jiang, Y., Gu, Y., Li, Z., Li, X., Zhao, K., Wang, C., Li, N., & Cao, X. (2018). Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. *Nature*, 554(7690), 123–127. <https://doi.org/10.1038/nature25434>
- Sibbritt, T., Schumann, U., Shafik, A., Guarnacci, M., Clark, S. J., & Preiss, T. (2022). Target-Specific Profiling of RNA m5C Methylation Level Using Amplicon Sequencing. *Methods in Molecular Biology*, 2404, 375–392. https://doi.org/10.1007/978-1-0716-1851-6_21
- Song, H., Zhang, J., Liu, B., Xu, J., Cai, B., Yang, H., Straube, J., Yu, X., & Ma, T. (2022). Biological roles of RNA m5C modification and its implications in Cancer immunotherapy. *Biomarker Research* 10(1), 1–15. <https://doi.org/10.1186/s40364-022-00362-8>
- Song, P., Tayier, S., Cai, Z., & Jia, G. (2021). RNA methylation in mammalian development and cancer. *Cell Biology and Toxicology* 37(6), 811–831. <https://doi.org/10.1007/s10565-021-09627-8>
- Squires, J. E., Patel, H. R., Nousch, M., Sibbritt, T., Humphreys, D. T., Parker, B. J., Suter, C. M., & Preiss, T. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Research*, 40(11), 5023–5033. <https://doi.org/10.1093/nar/gks144>
- Squires, J. E., & Preiss, T. (2010). Function and detection of 5-methylcytosine in eukaryotic RNA. *Epigenomics* 2(5), 709–715. <https://doi.org/10.2217/epi.10.47>
- Stadler, P. F., Chen, J. J. L., Hackermüller, J., Hoffmann, S., Horn, F., Khaitovich, P., Kretzschmar, A. K., Mosig, A., Prohaska, S. J., Qi, X., Schutt, K., & Ullmann, K. (2009). Evolution of vault RNAs. *Molecular Biology and Evolution*, 26(9), 1975–1991. <https://doi.org/10.1093/molbev/msp112>
- Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., & Mateo, J. L. (2015). CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS ONE*, 10(4), e0124633. <https://doi.org/10.1371/journal.pone.0124633>
- Sun, L., Xu, K., Huang, W., Yang, Y. T., Li, P., Tang, L., Xiong, T., & Zhang, Q. C. (2021). Predicting dynamic cellular protein–RNA interactions by deep learning using in vivo RNA structures. *Cell Research*, 31(5), 495–516. <https://doi.org/10.1038/s41422-021-00476-y>
- Sun, Z., Xue, S., Zhang, M., Xu, H., Hu, X., Chen, S., Liu, Y., Guo, M., & Cui, H. (2020). Aberrant NSUN2-mediated m5C modification of H19 lncRNA is associated with poor differentiation of hepatocellular carcinoma. *Oncogene*, 39(45), 6906–6919. <https://doi.org/10.1038/s41388-020-01475-w>
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324(5929), 930–935. <https://doi.org/10.1126/science.1170116>
- Tang, H., Fan, X., Xing, J., Liu, Z., Jiang, B., Dou, Y., Gorospe, M., & Wang, W. (2015). NSun2 delays replicative senescence by repressing p27 (KIP1) translation and elevating CDK1 translation. *Aging*, 7(12), 1143–1158. <https://doi.org/10.18632/aging.100860>

- Tao, Z., Ruan, H., Sun, L., Kuang, D., Song, Y., Wang, Q., Wang, T., Hao, Y., & Chen, K. (2019). Targeting the YB-1/PD-L1 axis to enhance chemotherapy and antitumor immunity. *Cancer Immunology Research*, 7(7), 1135–1147. <https://doi.org/10.1158/2326-6066.CIR-18-0648>
- Trerè, D., Migaldi, M., Montanaro, L., Pession, A., & Derenzini, M. (2000). *p120 Expression provides a reliable indication of the rapidity of cell duplication in cancer cells independently of tumour origin*. *Journal of Pathology*, 192(2), 216–220. [https://doi.org/10.1002/1096-9896\(2000\)9999:9999<::AID-PATH695>3.0.CO;2-L](https://doi.org/10.1002/1096-9896(2000)9999:9999<::AID-PATH695>3.0.CO;2-L)
- Trevarton, A. J., Zhou, Y., Yang, D., Rewcastle, G. W., Flanagan, J. U., Braithwaite, A., Shepherd, P. R., Print, C. G., Wang, M. W., & Lasham, A. (2019). Orthogonal assays for the identification of inhibitors of the single-stranded nucleic acid binding protein YB-1. *Acta Pharmaceutica Sinica B*, 9(5), 997–1007. <https://doi.org/10.1016/j.apsb.2018.12.011>
- Trixl, L., Amort, T., Wille, A., Zinni, M., Ebner, S., Hechenberger, C., Eichen, F., Gabriel, H., Schoberleitner, I., Huang, A., Piatti, P., Nat, R., Troppmair, J., & Lusser, A. (2018). RNA cytosine methyltransferase Nsun3 regulates embryonic stem cell differentiation by promoting mitochondrial activity. *Cellular and Molecular Life Sciences*, 75(8), 1483–1497. <https://doi.org/10.1007/s00018-017-2700-0>
- Trixl, L., & Lusser, A. (2019). The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark. *Wiley Interdisciplinary Reviews: RNA* 10(1), e1510. Blackwell Publishing Ltd. <https://doi.org/10.1002/wrna.1510>
- Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E., & Chang, H. Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, 329(5992), 689–693. <https://doi.org/10.1126/science.1192002>
- Tuorto, F., Herbst, F., Alerasool, N., Bender, S., Popp, O., Federico, G., Reitter, S., Liebers, R., Stoecklin, G., Gröne, H.-J., Dittmar, G., Glimm, H., & Lyko, F. (2015). The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis. *The EMBO Journal*, 34(18), 2350–2362. <https://doi.org/10.15252/emboj.201591382>
- Valdés-Rives, S. A., & González-Arenas, A. (2017). Autotaxin-Lysophosphatidic Acid: From Inflammation to Cancer Development. *Mediators of Inflammation* 2017(2017), 9173090. <https://doi.org/10.1155/2017/9173090>
- Van Haute, L., Lee, S. Y., McCann, B. J., Powell, C. A., Bansal, D., Vasiliauskaitė, L., Garone, C., Shin, S., Kim, J. S., Frye, M., Gleeson, J. G., Miska, E. A., Rhee, H. W., & Minczuk, M. (2019). NSUN2 introduces 5-methylcytosines in mammalian mitochondrial tRNAs. *Nucleic Acids Research*, 47(16), 8720–8733. <https://doi.org/10.1093/NAR/GKZ559>
- Van Haute, L., Powell, C. A., & Minczuk, M. (2017). Dealing with an unconventional genetic code in mitochondria: The biogenesis and pathogenic defects of the 5-formylcytosine modification in mitochondrial tRNAMet. *Biomolecules* 7(1), 8720–33. <https://doi.org/10.3390/biom7010024>
- Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J. Y., Cody, N. A. L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C., Bouvrette, L. P. B., Bergalet, J., Duff, M. O., Garcia, K. E., Gelboin-Burkhart, C., ... Yeo, G. W. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature*, 583(7818), 711–719. <https://doi.org/10.1038/s41586-020-2077-3>
- Wang, W. (2016). mRNA methylation by NSUN2 in cell proliferation. *Wiley Interdisciplinary Reviews: RNA* 7(6), 838–842. Wiley Interdiscip Rev RNA. <https://doi.org/10.1002/wrna.1380>

- Wang, X., Wang, M., Dai, X., Han, X., Zhou, Y., Lai, W., Zhang, L., Yang, Y., Chen, Y., Wang, H., Zhao, Y. L., Shen, B., Zhang, Y., Huang, Y., & Yang, Y. G. (2022). RNA 5-methylcytosine regulates YBX2-dependent liquid-liquid phase separation. *Fundamental Research*, 2(1), 48–55. <https://doi.org/10.1016/J.FMRE.2021.10.008>
- Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., & He, C. (2015). N6-methyladenosine modulates messenger RNA translation efficiency. *Cell*, 161(6), 1388–1399. <https://doi.org/10.1016/j.cell.2015.05.014>
- Wei, Z., Panneerdoss, S., Timilsina, S., Zhu, J., Mohammad, T. A., Lu, Z. L., De Magalhães, J. P., Chen, Y., Rong, R., Huang, Y., Rao, M. K., & Meng, J. (2018). Topological characterization of human and mouse M5C epitranscriptome revealed by bisulfite sequencing. *International Journal of Genomics*, 2018(3), 1351964. <https://doi.org/10.1155/2018/1351964>
- Westbye, M. P., Feyzi, E., Aas, P. A., Vågbo, C. B., Talstad, V. A., Kavli, B., Hagen, L., Sundheim, O., Akbari, M., Liabakk, N. B., Slupphaug, G., Otterlei, M., & Krokan, H. E. (2008). Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *Journal of Biological Chemistry*, 283(36), 25046–25056. <https://doi.org/10.1074/jbc.M803776200>
- Wiener, D., & Schwartz, S. (2021). The epitranscriptome beyond m6A. *Nature Reviews Genetics* 22(2), 119–131. <https://doi.org/10.1038/s41576-020-00295-8>
- Wilkinson, E., Cui, Y. H., & He, Y. Y. (2022). Roles of RNA Modifications in Diverse Cellular Functions. *Frontiers in Cell and Developmental Biology* 10(8), 828683. <https://doi.org/10.3389/fcell.2022.828683>
- Wu, T. P., Wang, T., Seetin, M. G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S. D., Mackintosh, S. G., Zhong, M., Tackett, A., Wang, G., Hon, L. S., Fang, G., Swenberg, J. A., & Xiao, A. Z. (2016). DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature*, 532(7599), 329–333. <https://doi.org/10.1038/nature17640>
- Wyatt, G. R. (1950). Occurrence of 5-methyl-cytosine in nucleic acids. *Nature* 166(4214), 237–238. <https://doi.org/10.1038/166237b0>
- Xu, L., Liu, X., Sheng, N., Oo, K. S., Liang, J., Chionh, Y. H., Xu, J., Ye, F., Gao, Y. G., Dedon, P. C., & Fu, X. Y. (2017). Three distinct 3-methylcytidine (m3C) methyltransferases modify tRNA and mRNA in mice and humans. *Journal of Biological Chemistry*, 292(35), 14695–14703. <https://doi.org/10.1074/jbc.M117.798298>
- Xu, X., Zhang, Y., Zhang, J., & Zhang, X. (2020). NSun2 promotes cell migration through methylating autotaxin mRNA. *Journal of Biological Chemistry*, 295(52), 18134–18147. <https://doi.org/10.1074/jbc.RA119.012009>
- Xue, M., Shi, Q., Zheng, L., Li, Q., Yang, L., & Zhang, Y. (2020). Gene signatures of m5C regulators may predict prognoses of patients with head and neck squamous cell carcinoma. *American Journal of Translational Research*, 12(10), 6841–6852. PMID: PMC7653571
- Xue, S., Xu, H., Sun, Z., Shen, H., Chen, S., Ouyang, J., Zhou, Q., Hu, X., & Cui, H. (2019). Depletion of TRDMT1 affects 5-methylcytosine modification of mRNA and inhibits HEK293 cell proliferation and migration. *Biochemical and Biophysical Research Communications*, 520(1), 60–66. <https://doi.org/10.1016/J.BBRC.2019.09.098>
- Yang, X., Yang, Y., Sun, B.-F., Chen, Y.-S., Xu, J.-W., Lai, W.-Y., Li, A., Wang, X., Bhattarai, D. P., Xiao, W., Sun, H.-Y., Zhu, Q., Ma, H.-L., Adhikari, S., Sun, M., Hao, Y.-J., Zhang, B., Huang, C.-M., Huang, N., ... Yang, Y.-G. (2017). 5-methylcytosine promotes mRNA export-NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell Research*, 27(5), 606–625. <https://doi.org/10.1038/cr.2017.55>

- Yang, Y. C. T., Di, C., Hu, B., Zhou, M., Liu, Y., Song, N., Li, Y., Umetsu, J., & Lu, J. J. (2015). CLIPdb: A CLIP-seq database for protein-RNA interactions. *BMC Genomics*, 16(1), 1–8. <https://doi.org/10.1186/s12864-015-1273-2>
- Yang, Y., Hsu, P. J., Chen, Y. S., & Yang, Y. G. (2018). Dynamic transcriptomic m6A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Research* 2018 28:6, 28(6), 616–624. <https://doi.org/10.1038/s41422-018-0040-8>
- Yang, Y., Wang, L., Han, X., Yang, W. L., Zhang, M., Ma, H. L., Sun, B. F., Li, A., Xia, J., Chen, J., Heng, J., Wu, B., Chen, Y. S., Xu, J. W., Yang, X., Yao, H., Sun, J., Lyu, C., Wang, H. L., ... Yang, Y. G. (2019). RNA 5-Methylcytosine Facilitates the Maternal-to-Zygotic Transition by Preventing Maternal mRNA Decay. *Molecular Cell*, 75(6), 1188–1202. <https://doi.org/10.1016/j.molcel.2019.06.033>
- Yartseva, V., & Giraldez, A. J. (2015). The Maternal-to-Zygotic Transition During Vertebrate Development: A Model for Reprogramming. *Current Topics in Developmental Biology* 113, 191–232). <https://doi.org/10.1016/bs.ctdb.2015.07.020>
- Yuan, H., Liu, J., Zhao, L., Wu, P., Chen, G., Chen, Q., Shen, P., Yang, T., Fan, S., Xiao, B., & Jiang, K. (2021). Prognostic Risk Model and Tumor Immune Environment Modulation of m5C-Related LncRNAs in Pancreatic Ductal Adenocarcinoma. *Frontiers in Immunology*, 12(8), 800268. <https://doi.org/10.3389/fimmu.2021.800268>
- Yuan, S., Tang, H., Xing, J., Fan, X., Cai, X., Li, Q., Han, P., Luo, Y., Zhang, Z., Jiang, B., Dou, Y., Gorospe, M., & Wang, W. (2014). Methylation by NSun2 Represses the Levels and Function of MicroRNA 125b. *Molecular and Cellular Biology*, 34(19), 3630–3641. <https://doi.org/10.1128/mcb.00243-14>
- Zhang, C., Chang, X., Chen, D., Yang, F., Li, Z., Li, D., Yu, N., Yan, L., Liu, H., & Xu, Z. (2019). Downregulation of HDGF inhibits the tumorigenesis of bladder cancer cells by inactivating the PI3K-AKT signaling pathway. *Cancer Management and Research*, 11, 7909–7923. <https://doi.org/10.2147/CMAR.S215341>
- Zhang, T., Chen, P., Li, W., Sha, S., Wang, Y., Yuan, Z., Shen, B., & Chen, L. (2019). Cognitive deficits in mice lacking Nsun5, a cytosine-5 RNA methyltransferase, with impairment of oligodendrocyte precursor cells. *GLIA*, 67(4), 688–702. <https://doi.org/10.1002/glia.23565>
- Zhang, Y., Zhang, X., Shi, J., Tuorto, F., Li, X., Liu, Y., Liebers, R., Zhang, L., Qu, Y., Qian, J., Pahima, M., Liu, Y., Yan, M., Cao, Z., Lei, X., Cao, Y., Peng, H., Liu, S., Wang, Y., ... Chen, Q. (2018). Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nature Cell Biology*, 20(5), 535–540. <https://doi.org/10.1038/s41556-018-0087-2>
- Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., Vågbø, C. B., Shi, Y., Wang, W. L., Song, S. H., Lu, Z., Bosmans, R. P. G., Dai, Q., Hao, Y. J., Yang, X., Zhao, W. M., Tong, W. M., Wang, X. J., Bogdan, F., ... He, C. (2013). ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Molecular Cell*, 49(1), 18–29. <https://doi.org/10.1016/j.molcel.2012.10.015>
- Zou, F., Tu, R., Duan, B., Yang, Z., Ping, Z., Song, X., Chen, S., Price, A., Li, H., Scott, A., Perera, A., Li, S., & Xie, T. (2020). Drosophila YBX1 homolog YPS promotes ovarian germ line stem cell development by preferentially recognizing 5-methylcytosine RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 117(7), 3603–3609. <https://doi.org/10.1073/pnas.1910862117>