The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark

Lukas Trixl | Alexandra Lusser

Division of Molecular Biology, Biocenter, Medical University of Innsbruck, Innsbruck, Austria

Correspondence
Alexandra Lusser, Division of Molecular Biology, Biocenter, Medical University of Innsbruck, 6020 Innsbruck, Austria.
Email: alexandra.lusser@i-med.ac.at

Funding information
Austrian Science Fund, Grant/Award Number: P27024-BBL

It is a well-known fact that RNA is the target of a plethora of modifications which currently amount to over a hundred. The vast majority of these modifications was observed in the two most abundant classes of RNA, rRNA and tRNA. With the recent advance in mapping technologies, modifications have been discovered also in mRNA and in less abundant non-coding RNA species. These developments have sparked renewed interest in elucidating the nature and functions of those “epitranscriptomic” modifications in RNA. N6-methyladenosine (m6A) is the best understood and most frequent mark of mRNA with demonstrated functions ranging from pre-mRNA processing, translation, miRNA biogenesis to mRNA decay. By contrast, much less research has been conducted on 5-methylcytosine (m5C), which was detected in tRNAs and rRNAs and more recently in poly(A)RNAs. In this review, we discuss recent developments in the discovery of m5C RNA methylomes, the functions of m5C as well as the proteins installing, translating and manipulating this modification. Although our knowledge about m5C in RNA transcripts is just beginning to consolidate, it has become clear that cytosine methylation represents a powerful mechanistic strategy to regulate cellular processes on an epitranscriptomic level.

This article is categorized under:
- RNA Processing > RNA Editing and Modification
- RNA Interactions with Proteins and Other Molecules > Protein–RNA Interactions: Functional Implications
- RNA Processing > tRNA Processing
- RNA Turnover and Surveillance > Regulation of RNA Stability

KEYWORDS
5-methylcytosine, epitranscriptomic mark, miRNA, mRNA, N6-methyladenosine, RNA modification, rRNA, tRNA

1 | INTRODUCTION

Research on posttranscriptional modification of RNA species has commenced nearly 60 years ago with the description of pseudouridine (Ψ) (Cohn, 1960). Pseudouridine was characterized as a major modification in tRNA and rRNA that is very conserved throughout the prokaryotic and eukaryotic kingdoms (Goodman, Abelson, Landy, Brenner, & Smith, 1968; Zachau, Dütting, & Feldmann, 1966). In later research, the conserved H/ACA snoRNP proteins together with H/ACA snoRNAs were shown to be responsible for installing up to 100 Ψ modifications onto mammalian rRNA (McMahon, Contreras, & Ruggero,
The first modifications of mRNA molecules were discovered several decades ago, including the 5′ cap, which contributes among other things to mRNA stability and translation initiation, and the 3′ poly(A) tail that assists the progress of nuclear export, stability as well as translation (Edmonds & Abrams, 1960). Around the same time, the first internal mRNA modifications were described with N6-methyladenosine (m6A) as the most abundant and to date best understood mRNA modification (Dubin & Taylor, 1975; Perry & Kelley, 1975). m6A positions were mapped in mRNA upon introduction of an antibody-mediated detection strategy (Dominissini et al., 2012; Meyer et al., 2012). The discovery of specific “writer”, “reader”, and “eraser” proteins demonstrated that m6A is a dynamic modification that affects various aspects of RNA metabolism, including mRNA stability, translation or splicing (reviewed, e.g., in Cao, Li, Yin, & Flavell, 2016; Meyer & Jaffrey, 2017; Peer, Rechavi, & Dominissini, 2017; Schwartz, 2016; Song & Yi, 2017; Zhao, Roundtree, & He, 2016). Based on these findings, the idea was put forward that posttranscriptional modification of RNA might impose information on top of the sequence information contained in the RNA similar to what is known as epigenetic information for DNA and chromatin and prompted the coining of the terms “RNA epigenetics” (He, 2010) and “epitranscriptomics” (Saletore et al., 2012), respectively. Meanwhile, additional modifications have been studied in mRNA, including N1-methyladenosine (m1A) or 2′-O-methylribonucleosides, although their functional roles are only beginning to be elucidated (Song & Yi, 2017; Xiong, Li, & Yi, 2018; Zhao et al., 2016).

In this review, we put a spotlight on another modification of RNA that has gained increasing attention in recent years, the methylation of carbon 5 in cytosine (m5C). We will discuss methods for its detection, its distribution in different types of RNA, its effects on RNA function and the enzymes responsible for its deposition. Most of our current knowledge about m5C in RNA comes from research on the abundant rRNAs and tRNAs. Although we will briefly touch on these findings, in depth discussion of rRNA and tRNA methylation can be found in several excellent recent reviews (e.g., Bohnsack & Sloan, 2018; Sharma & Lafontaine, 2015; Sloan et al., 2017; Sokołowski, Klassen, Bruch, Schaffrath, & Glatt, 2018; Suzuki, Nagao, & Suzuki, 2011; Traube & Carell, 2017). Here, we will concentrate on m5C in mRNAs and in less well characterized non-coding RNA targets.

2 METHODS FOR DETECTING M5C IN RNA

Methylated cytosine has been discovered first in DNA (Hotchkiss, 1948; Wyatt, 1950) but soon after was also found in RNA (Amos & Korn, 1958). The detection of methylated cytosines in poly(A)RNA was demonstrated in the 1970 for the first time (Desrosiers, Friderici, & Rottman, 1974; Dubin & Taylor, 1975). For these studies, mainly chromatography-based methods were used to identify methylated nucleotides, including DEAE cellulose chromatography, thin-layer chromatography and liquid chromatography. Later on, mass spectrometry (MS) proved to be a highly accurate and sensitive method for the identification of RNA modifications especially in combination with liquid chromatography (LC–MS), and the latest techniques in this field allow for detection of RNA modifications in the femto- to attomol range. However, the sequence context is usually lost in such analyses because they require enzymatic digestion of RNA to nucleosides (Helm & Motorin, 2017; Kellner, Burhenn, & Helm, 2010). To circumvent this problem approaches have been developed that involve the targeted fragmentation of the RNA by specific enzymes coupled to LC–MS/MS analysis similar to the strategies applied in the proteomics field (Limbach & Paulines, 2017). Yet these approaches still suffer from the lack of suitable bioinformatics resources and tools and relatively low sensitivity that restricts their use to highly abundant RNA classes, such as tRNAs (Wetzel & Limbach, 2013). Top-down label-free MS in which RNA is not hydrolyzed was also shown to efficiently identify, localize and quantify methylated nucleobases at a relative level (Glaser, Riml, Micura, & Breuker, 2017). Additionally, studies on human and bacterial ribosomes have revealed high resolution (<3 Å) cryo-electron microscopy (cryo-EM) as a powerful tool to simultaneously monitor and localize hundreds of modifications on rRNA (Fischer et al., 2015; Natchiar, Myasnikov, Kratatz, Hazemann, & Klaholz, 2017; Polikanov, Melnikov, Söll, & Steitz, 2015; Shalev-Benami et al., 2016). However, these techniques are currently not suitable for the analysis of modifications, in particular methylation, in mRNAs, because they require uniform (i.e., one sequence) RNA (top-down MS) or RNA complexes (cryo-EM). Thus, current methods used for methylation mapping in mRNA typically are based on RNA sequencing coupled to prior chemical derivatization or enrichment strategies as will be discussed in the following sections.

2.1 RNA-bisulfite-sequencing

With the development of the bisulfite sequencing technique in 1994, it was possible to study cytosine methylation in DNA in a sequence-specific manner (Frommer et al. 1992; Clark, Harrison, Paul, & Frommer, 1994). In single stranded DNA, HSO3− reacts with cytosine (C) in acidic pH resulting in deamination and formation of uracil-sulfonate which converts to uracil
(U) under basic pH conditions thus causing a C-to-U conversion that can be detected by sequencing (Figure 1). In principle, this reaction can also occur on m5C, yet it is much slower, which allows for selective distinction between C and m5C. In the case of RNA, sodium bisulfite found use in the investigation of amino-acylation and amino acid acceptance of tRNAs (Chakraburtty, 1975; Sabban & Bhanot, 1982). For RNA methylation studies, however, bisulfite conversion had not been considered as a tool of choice, because of the harsh reaction conditions (denaturation at 95°C, alkali conditions) required, which cause strong degradation of RNA. In 2009, Schäfer et al. reported that by lowering denaturing temperatures and extending incubation times, bisulfite sequencing could also be applied for the detection of m5C in RNA which was demonstrated by the identification of m5C sites in tRNA and rRNA (Schaefer, Pollex, Hanna, & Lyko, 2009). Unlike any of the other available methods for m5C detection (described below), RNA-bisulfite-sequencing (RNA-BS-seq) allows for the determination of the extent of methylation of a specific C position in RNA. However, one big caveat of RNA-BS-seq is its failure to react with Cs in a base-paired conformation. Considering the low overall m5C occurrence in RNA and specifically in mRNA (0.03–0.1% of all Cs; Huber et al., 2015; Legrand et al., 2017), achievement of high C-to-U conversion rates is crucial. Thus, efficient denaturation of RNA secondary structures and concomitant best possible preservation of RNA integrity are critical factors determining the robustness of data sets generated by RNA-BS-seq (Schaefer, 2015). Recent improvements addressing this issue are the addition of the double-strand destabilizing agent formamide to the reaction or random fragmentation of the RNA prior to the treatment (Khodddami, Yerra, & Cairns, 2015).

2.2 | Methylated-RNA-immunoprecipitation

Methylated-RNA-immunoprecipitation (meRIP-seq) was originally used to identify the m6A methylome in randomly fragmented RNA transcripts (Dominissini et al., 2012; Meyer et al., 2012). This method takes advantage of a highly specific m6A antibody for immunoprecipitation and massive parallel sequencing in order to obtain transcriptome wide methylation data. meRIP with an antibody against m5C coupled to Illumina sequencing (Figure 1) was used to verify RNA-BS-seq results in the archaean Sulfolobus solfataricus confirming all of the RNA and 71% of the mRNA sites detected by RNA-BS-seq (Edelheit, Schwartz, Mumbach, Wurtzel, & Sorek, 2013) demonstrating its suitability to detect m5C in RNA. However, a drawback to this method is that the antibody reacts preferentially with single stranded nucleic acids (Weber et al., 2005). Thus, RNA secondary structure may obscure potential m5C sites from detection.

2.3 | Aza-IP and miCLIP

While RNA-BS-seq and meRIP-seq probe m5C occurrence without previous knowledge of the nature of the responsible methylating enzyme, aza-IP and miCLIP were designed to identify the target sites of RNA cytosine methyltransferases (RCMTs). Both techniques take advantage of the catalytic mechanism of RNA methyltransferases which includes a transient covalent linkage of the enzyme to the methylation target cytosine. In aza-IP, 5-azacytidine (5-azaC) is incorporated into RNA by feeding cells with the modified nucleoside. If this occurs at an RCMT target site, it traps the RCMT because the covalent adduct with the target C cannot be resolved. Thus, antibodies against the RCMT (or a tagged version) can be used to immunoprecipitate the protein along with the covalently bound RNA, which is sequenced. Enrichment analysis is then performed to identify
methylated RNAs, and sites with increased C-to-G transversion signatures (resulting from a ring opening of 5-azaC during the protocol) correspond to the sites of methylation (Figure 1) (Khoddami & Cairns, 2013). While aza-IP appears to identify m^5C sites with high specificity, it may be hampered by limited sensitivity. 5-azaC is toxic to cells (Flatau, Gonzales, Michalowsky, & Jones, 1984; Jüttermann, Li, & Jaenisch, 1994) necessitating short labeling periods, in which only a small portion of C is replaced by 5-azaC thereby reducing the probability of being incorporated at the site of a modification. This may be particularly critical for RNAs with low-expression levels. Along the same lines, some methylation sites may escape detection in enriched sequences since C-to-G transversion is not quantitative (Khoddami & Cairns, 2013).

Rather than incorporating a RCMT “suicide” inhibitor, such as 5-azaC, methylation-individual nucleotide resolution cross-linking and immunoprecipitation (miCLIP) is based on the use of a modified RCMT in which the cysteine that is responsible for releasing the RCMT from the RNA substrate is mutated resulting in a covalently linked RNA-protein complex. Similar to aza-IP, immunoprecipitation with an antibody against the RCMT is used to enrich bound RNA, which is then subjected to deep sequencing (Figure 1). Because enzyme-RNA crosslinking leads to termination of reverse transcription at this site, cytosine methylation positions are detected at the +1 site of the sequencing reads rather than by C-to-G transversion signatures as in aza-IP (Hussain et al., 2013). This method does not require incorporation of a modified nucleotide into RNA, but it relies on the overexpression of a mutant RCMT, which may cause changes in methylation patterns. Also, the multistep procedure for detecting the RNAs may lead to decreased data output, which may negatively affect sensitivity.

Taken together, all available methods suffer from different drawbacks caused by a variety of reasons ranging from the biological features of the RNA itself to the experimental procedures and the data analysis approaches (Grozhik & Jaffrey, 2018). Nevertheless, at this point, the complementary use of different methods and carefully designed control experiments should allow for the generation of reliable m^5C data sets.

### 3 | 5-METHYLCYTOSINE IN ABUNDANT RNA SPECIES

Over the years, m^5C was detected in tRNA (transfer RNA), rRNA (ribosomal RNA), mRNA (messenger RNA), snRNA (small nuclear RNA), miRNA (microRNA), IncRNA (long noncoding RNA) or eRNAs (enhancer RNA) from many species and in all three domains of life. There are differences, however, in the occurrence of m^5C in specific RNA types in different species. For example, m^5C appears not to be present in tRNA and mRNA from bacteria, while it has been found in eukaryal and archeal tRNA and mRNA (Amort et al., 2017; Cui et al., 2017; David et al., 2017; Dubin & Taylor, 1975; Edelheit et al., 2013; Fu et al., 2014; Huang et al., 2016; Huber et al., 2015; Motorin, Lyko, & Helm, 2010; Salditt-Georgieff et al., 1976; Squires et al., 2012; Yang et al., 2017).

#### 3.1 | tRNA

Most information about m^5C comes from its study in tRNAs. Methylation occurs most often at cytosines at the junction of the variable loop and the T stem and loop at one, two or three Cs spanning positions 47–50 (Figure 2). It has been proposed that methylation of C48, which forms an unusual “Levitt pair” with nucleoside 15 in the D-loop to generate the characteristic L-shape three-dimensional structure, stabilizes this interaction by increasing the hydrophobicity of the base
pair and contributing to base stacking (Väre, Eruysal, Narendran, Sarachan, & Agris, 2017). Another site that is frequently methylated in animals is C38 in the anticodon loop (Figure 2). Methylation of C38 in mouse tRNA^Leu^CAA and mitochondrial (mt) tRNA^Met^ in mice (Blanco et al., 2014; Trixl et al., 2018) and for mt-tRNA^Met^ in humans (Haag et al., 2016; Nakano et al., 2016; Van Haute et al., 2016) (Figure 2). C34 methylation of mt-tRNA^Met^ precedes further modification of this site by oxidation to 5-formyl-cytosine (f5C), which is important for the decoding of AUA methionine codons (Tuorto et al., 2015). Furthermore, it was shown that in *Schizosaccharomyces pombe*, replacement of guanosine at the wobble position 34 in the anticodon loop by queuosine promotes methylation of C38 providing an example for crosstalk of different tRNA modifications although the functional significance of this cross-talk is not yet clear (Ehrenhofer-Murray, 2017; Jeltsch et al., 2017; Müller et al., 2015). The existence of C34 methylation was also confirmed for C34 in tRNA^Leu^CAA and mitochondrial (mt) tRNA^Met^ in mice (Blanco et al., 2014; Trixl et al., 2018) and for mt-tRNA^Met^ in humans (Haag et al., 2016; Nakano et al., 2016; Van Haute et al., 2016) (Figure 2). C34 methylation of mt-tRNA^Met^ precedes further modification of this site by oxidation to 5-formyl-cytosine (f5C), which is important for the decoding of AUA methionine codons during mitochondrial translation (Takemoto et al., 2009). Finally, C72 in human tRNA^Thr^ and tRNA^Cys^ has been shown to be methylated at a late step of tRNA biogenesis, since methylation is dependent on the presence of the posttranscriptionally added CCA sequence at the 3′end (Haag et al., 2015) but so far no specific function has been ascribed to this modification (Figure 2).

### m^5^C in mRNA

**3.2 | rRNA**

Ribosomal RNA from all kingdoms of life is subject to cytosine methylation at carbon 5, and the modified positions are fairly conserved throughout evolution (Motorin et al., 2010) (Figure 2). In human and yeast 28S rRNA two methylated cytosines, m^5^C2278 and m^5^C2270, are known (Bourgeois et al., 2015; Motorin et al., 2010; Schaefer et al., 2009; Schosserer et al., 2015; Sharma, Yang, Watzinger, Kötter, & Entian, 2013; Squires et al., 2012), while m^5^C has not been detected in eukaryal 18S RNA (Edelheit et al., 2013). m^5^C2270 resides close to the peptidyltransferase center in the yeast crystal structure of the ribosome, while m^5^C2278 is located at the interface between large and small subunit (Sharma & Lafontaine, 2015). Loss of m^5^C2278 methylation in yeast 25S rRNA resulted in transient changes of rRNA folding upon oxidative stress treatment and it promoted translational read-through in a reporter construct with a premature stop codon (Schosserer et al., 2015). In the structure of the yeast ribosome, m^5^C2278 along with three other modified bases in the large subunit is contacted by the ribosomal protein eL41, which forms a bridge (eB14) between large and small subunit. This contact was proposed to be important during translation, when eL41 acts as a rotation center for 40S subunit movement (Sharma & Lafontaine, 2015). m^5^C was also detected in mouse mitochondrial 12S but not 16S rRNA (Metodiev et al., 2014) (Figure 2). The function of this modification is currently unknown.

**3.3 | m^5^C in mRNA**

**3.3.1 | Prevalence of m^5^C sites in mRNA**

The existence of m^5^C in eukaryotic mRNA has been known since the 1970s when it was detected in mRNA from HeLa and hamster cells (Dubin & Taylor, 1975; Salditt-Georgieff et al., 1976), although some earlier studies had failed to detect m^5^C in mammalian cells (Desroisiers et al., 1974; Perry, Kelley, Friderici, & Rottman, 1975). In recent years, improved liquid chromatography-mass spectrometry (LS-MS) methods showed that mRNAs indeed possess methylation as well as hydroxymethylation on internal cytosines (Fu et al., 2014; Huang et al., 2016; Huber et al., 2015). Since in LC–MS analysis information regarding the position of m^5^C is lost, the adaptation of bisulfite sequencing for use with RNA unlocked new possibilities to map m^5^C at nucleotide resolution in mRNA (Schaefer et al., 2009). Using this method, in 2012, Preiss and colleagues published the first cytosine methylome for human cells (Squires et al., 2012). They found ~10,000 sites showing >20% methylation and mapping to ~8,500 mRNAs resulting in a rate of 0.43% m^5^C of all sequenced Cs. Recently, the cytosine methylome of mouse embryonic stem cells (ESC) and of the brain was reported (Amort et al., 2017). In this study, ~7,500 m^5^C sites (>20% methylation) mapping to 1,650 mRNAs were detected in ESCs and 2,075 m^5^C sites mapping to 486 mRNAs in the brain. Another recent publication on HeLa cell and mouse cytosine methylomes identified ~3,600 sites within ~2,000 mRNAs in HeLa cells and 2,500–4,400 sites (1,000–1,655 mRNAs) in different mouse tissues (Yang et al., 2017). In the plant Arabidopsis thaliana, a few 100 m^5^C sites in mRNA were detected using bisulfite sequencing (David et al., 2017), while another study using meRIP-seq found 6,045 peaks mapping to 4,465 expressed genes (Cui et al., 2017). meRIP-seq was also used to examine m^5^C levels in budding yeast and the archaeabacterium Sulfolobus solfataricus and revealed a single site in yeast and 14 methylated mRNAs in *S. solfataricus* (Edelheit et al., 2013). Finally, a recent study in mouse ESCs applying bisulfite sequencing reported 745 m^5^C sites (Legrand et al., 2017).

These divergent findings by different studies raise two obvious questions: What is the reason for the differences between studies? Does m^5^C in mRNA exist at all? With respect to the first question, it is clear that some of the differences can be
attributed to the different methods used. As detailed above, none of the existing techniques to map m$^5$C in RNA are without caveats. Secondary structures, sensitivity and specificity issues or data analysis differences, but also different RNA sources (organisms, cell types) may all influence the outcome of m$^5$C methylome analyses. Hence, further improvement of the methodology is clearly needed to obtain highly reliable data sets. This will also address the second question of whether mRNA actually contains m$^5$C. However, although it is possible that the number of true m$^5$C sites in poly(A)RNA in the end will be considerably lower than suggested now, multiple lines of evidence suggest that mRNA is indeed subject to cytosine methylation (Amort et al., 2017; Edelheit et al., 2013; Fu et al., 2014; Huber et al., 2015; Hussain, Aleksic, Blanco, Dietmann, & Frye, 2013; Legrand et al., 2017).

3.3.2 | Distribution of m$^5$C sites in mRNA

Comparing the findings from different studies available so far revealed that the distribution of m$^5$C within mRNA is not random (Figure 2). In HeLa and mouse cells, methylated cytosines were found to be enriched in 5’ and 3’ UTRs and depleted in coding regions (Amort et al., 2017; Squires et al., 2012; Yang et al., 2017). In particular, a pronounced peak of m$^5$C was detected in the vicinity of the translational start codon (Amort et al., 2017; Yang et al., 2017) of the cells/tissues examined, while enrichment in the 3’UTR appeared to be rather cell/tissue type specific (Amort et al., 2017). The results from Arabidopsis are more conflicting, with one study finding enrichment in the 3’UTR but not in CDS and 5’UTR (David et al., 2017) and the other reporting enrichment in the CDS and depletion in 5’ and 3’UTRs (Cui et al., 2017).

At the level of the individual m$^5$C site there is consistent evidence that the modification occurs in a cell/tissue-type specific manner. In the mouse, the majority of sites detected in different samples was unique to the specific cell/organ (Amort et al., 2017; Yang et al., 2017). Interestingly, the relatively small overlap of sites between ESCs and brain was due to differential methylation rather than differential expression of the target mRNAs at least for the majority of sites detected only in ESCs. In other words, mRNAs that were found to be methylated only in ESCs were expressed but not methylated in the brain. On the other hand, sites detected in the brain samples mapped predominantly to mRNAs that are poorly or not expressed in ESCs (Amort et al., 2017). Likewise, a considerable number of m$^5$C sites was specific to a particular mouse tissue although the corresponding mRNA was expressed in all tissues (Yang et al., 2017). In Arabidopsis, m$^5$C sites that were present in multiple tissues showed differences with respect to the degree of methylation in those tissues (David et al., 2017). Together, these different methylation patterns point towards regulation of cytosine methylation in mRNAs dependent on differentiation type and/or developmental stage.

3.3.3 | Function of m$^5$C in mRNA

As we are only beginning to uncover the levels and distribution of m$^5$C in mRNA, not much is known about the potential functions of this modification. Methylation of cytosine occurs at the Hoogsteen edge of the base and therefore does not affect Watson-Crick base pairing. Because methylation increases the hydrophobicity of the major groove of the RNA, it may have an effect on base stacking (Harcourt, kietrys, & Kool, 2017; Wang & Kool, 1995). The different methylome analyses have attempted to gain information about m$^5$C function by performing GO term enrichment analyses of methylated mRNAs or by correlating m$^5$C sites with regulatory elements and protein binding sites. In the mouse and in Arabidopsis, enrichment of pathways characteristic to the specific cell type analyzed but also of basic cellular and metabolic pathways was detected (Amort et al., 2017; Cui et al., 2017; Yang et al., 2017). By contrast, no such enrichment was found in HeLa cells (Squires et al., 2012). Generally, there was also no correlation between m$^5$C occurrence and overall transcript levels (Cui et al., 2017; Yang et al., 2017), own unpublished observation). On the other hand, m$^5$C sites have been found to overlap to a certain degree with binding sites of several RNA regulatory proteins, such as Argonaute (Squires et al., 2012) or splicing-and mRNA decay-associated factors, such as SRSF3 or UPF (Amort et al., 2017) by correlating m$^5$C data with PAR-CLIP-data from public databases. However, although enrichment is statistically significant, typically the fraction of overlapping protein binding sites is small and experimental analyses are largely lacking so far. Nevertheless, there are some studies describing specific functions for m$^5$C in mRNA metabolism which will be discussed in the following.

3.3.4 | m$^5$C as a nuclear export regulator

A recent report discovered that the activity of the nuclear export factor ALYREF/THOC4 is strongly affected by the methylation status of its target mRNAs (Yang et al., 2017). mRNAs bound to ALYREF were found to be enriched in m$^5$Cs in the vicinity of the translational start codon and in a CG sequence context. Moreover, knock-down of ALYREF in HeLa cells resulted in increased nuclear retention of m$^5$C-modified mRNA which could be rescued by expression of wild-type ALYREF but not of a mutant version that was unable to bind m$^5$C. In contrast, non-m$^5$C bearing mRNA showed no nuclear export defects upon ALYREF knockdown (Yang et al., 2017). Thus, ALYREF appears to be a bona fide m$^5$C “reader” protein with
the ability to regulate mRNA fate dependent on its m\(^5\)C status, similar to what has been shown for other proteins in the context of the m\(^6\)A modification (Dominissini & Rechavi, 2017).

### 3.3.5 m\(^5\)C as a modulator of protein translation

Investigating the molecular mechanisms governing the increase of the cyclin-dependent kinase inhibitor p2\(^7\)KIP1 during replicative senescence, Tang et al. (2015) showed that p2\(^7\)KIP1 is subject to cytosine methylation in the 5\(^\prime\)UTR and that m\(^5\)C is progressively lost during cell aging. Moreover, it was found that in a cellular reporter gene assay, reporter activity was significantly inhibited upon overexpression of the RNA methyltransferase NSUN2 and that the opposite was true for a NSUN2 knock-down suggesting that m\(^5\)C introduced by NSUN2 inhibits translation. The latter was also observed in an in vitro translation system arguing against an indirect effect via NSUN2-mediated tRNA methylation (see later section in this review). Conversely, methylation of specific cytosines in the 3\(^\prime\)UTRs of the cell cycle regulators CDK1 and p21, respectively, was shown to promote translation of these mRNAs in vitro and in a reporter gene system in vivo (Li et al., 2017; Xing et al., 2015). Similarly, an m\(^5\)C site within interleukin-17A mRNA was observed to promote translation of IL-17A (Wang, Tang, Wang, Wang, & Feng, 2017). However, the precise mechanism by which m\(^5\)C affects translation are presently unknown. In the case of CDK1, NSUN2 knockdown resulted in the association of CDK1 mRNA with a ribosomal fraction containing smaller polysomes, which suggests that m\(^5\)C-mediated translation regulation might occur at the initiation rather than the elongation level (Xing et al., 2015). Nevertheless, m\(^5\)C may also have the potential to affect ribosomal translation efficiency and to affect decoding potential. Using a bacterial in vitro translation system, Hoernes et al. showed in a systematic analysis of the effect of m\(^5\)C at different codon positions that m\(^5\)C at any position reduced translation efficiency and altered codon specificity when it was inserted at the second position (Hoernes et al., 2016). However, it remains to be shown, if this also holds true for eukaryotic translation.

Together the findings from these studies focusing on m\(^5\)C in specific mRNAs suggest that the position of m\(^5\)C within the mRNA (5\(^\prime\)UTR, coding region, 3\(^\prime\)UTR) has differential effects on mRNA function.

### 3.3.6 m\(^5\)C IN OTHER RNA TYPES

Besides mRNA, tRNA and rRNA, m\(^5\)C has also been detected in other RNA types, specifically in long non-coding RNAs (lncRNA) and smaller non-coding RNAs, such as enhancer associated RNAs (eRNAs) or vault RNAs (vtRNAs). For lncRNAs, m\(^5\)C was found near a protein binding domain in human HOTAIR in various cancer cell lines as well as in the functionally important A-region of the lncRNA XIST. Importantly, m\(^5\)C can interfere with binding of XIST to the chromatinerelative PRC2 (polycomb repressive complex 2) complex in vitro (Amort et al., 2013). Moreover, the various transcriptome-wide m\(^5\)C mapping studies also revealed several lncRNAs as methylation targets (Amort et al., 2017; David et al., 2017; Yang et al., 2017). Furthermore, Ribonuclease P RNA component H1 (RPPH1), 5S rRNA, the snoRNA small cajal body specific RNA 2 (SCARN2A), RNY1 and signal recognition particle RNA (7SL RNA) were shown to contain m\(^5\)C by various methods (Hussain, Sajini, et al., 2013; Khoddami & Cairns, 2013; Squires et al., 2012). The functional significance of m\(^5\)C in those RNAs, however, remains elusive so far. m\(^5\)C sites were also identified in vtRNAs using BS-sequencing, miCLIP and aza-IP (Hussain, Sajini, et al., 2013; Khoddami & Cairns, 2013; Squires et al., 2012). vtRNAs are RNA components of the vault ribonucleoproteincomplexes the function of which is not yet understood (Kedersha & Rome, 1986). Frye and colleagues showed that the three vtRNAs vtRNA1.1, vtRNA1.2 and vtRNA1.3 contain cytosine methylation sites generated by the methyltransferase NSUN2 and that methylation of vtRNA1.1 affected its processing into smaller fragments (svRNAs). One of these fragments, svRNA4, may have miRNA-like functions as its increase correlated with a decrease of two potential target mRNAs (Hussain, Sajini, et al., 2013). Another ncRNA target for cytosine methylation that was characterized in some detail are enhancer RNAs (eRNAs) associated with the regulatory regions of several target genes of the transcriptional coregulator peroxisome proliferator-activated receptor-gamma coactivator 1 alpha (PGC-1\(\alpha\)) in mouse hepatocytes. PGC-1\(\alpha\) was found to interact with the methyltransferase Nsun7, and depletion of Nsun7 caused a decrease of the presence of m\(^5\)C as well as of the levels of the respective eRNAs suggesting that m\(^5\)C may influence the stability of these eRNAs (Aguilo et al., 2016). As Nsun7 expression in the liver of mice was elevated upon prolonged fasting, concomitant with an increase in overall RNA m\(^5\)C levels, methylation was suggested as a stress-response mechanism to stabilize eRNA thereby offering a possibility to fine-tune gene expression under these conditions (Aguilo et al., 2016). Although the number of ncRNA methylation targets that have been studied so far is still small, available data suggest that the m\(^5\)C mark can act as a versatile tool to fine-tune RNA processing, stability, translation, as well as RNA-protein interaction.

### 3.3.7 RNA m\(^5\)C METHYLTRANSFERASES

RNA m\(^5\)C methyltransferases belong to the superfamily of Rossman fold-containing enzymes that use S-adenosyl-L-methionine (SAM) as a methyl group donor. Members of this group of proteins can be found in all domains of life targeting proteins, DNA, RNA, lipids and small molecules for methylation. To date all confirmed m\(^5\)C-specific RCMTs belong to either
the DNMT2 or the NOL1/NOP2/sun (Nsun) subgroups of methyltransferases (Bujnicki, Feder, Ayres, & Redman, 2004; Motorin et al., 2010). In mammals, the Nsun family of enzymes comprises seven genes that include \textit{Nsun1}, \textit{Nsun2}, \textit{Nsun3}, \textit{Nsun4}, \textit{Nsun5}, \textit{Nsun6} and \textit{Nsun7}. RCMTs drive the transfer of a methyl group onto cytosine residues of diverse RNA species. All enzymes characterized so far harbor a similar structural core containing the catalytic domain and the SAM binding site. Two conserved cysteine residues that are located in the so-called motives IV and VI catalyze the methylation reaction by the Nsun family members. The thiol group of the cysteine within motive VI, which is found as a conserved dipeptide motive with threonine (TC), enables the enzyme to covalently bind the target cytosine by attacking the carbon 6 atom of the pyrimidine ring, thus activating the non-nucleophilic carbon 5 (Liu & Santi, 2000). Extensive protonation leads to a nucleophilic carbon 5 atom that is now susceptible for methylation by electrophilic SAM (Cheng & Roberts, 2001). Beta-elimination then generates the methylated product and the free enzyme involving the cysteine in motive IV embedded in a dipeptide formation with proline (PC; (King & Redman, 2002). By contrast, DNMT2 lacks the PC motive and, like DNA methyltransferases, uses a one-cysteine catalytic mechanism (Jeltsch et al., 2017; Jurkowski et al., 2008). Below we will discuss current findings with respect to the substrates and biological roles of eukaryotic RCMTs (Figure 3).

3.3.8 | \textit{Nsun 1, Nsun 4 and Nsun 5: methylating ribosomal RNA}

\textit{Nsun1:} The role of Nsun1 (or Nop2p in \textit{S. cerevisae}) has been investigated already two decades ago, showing a critical involvement in 60S biogenesis in yeast (Sato et al., 1999). Human NSUN1 (also termed NOP2/ nucleolar antigen p120) was found to be strongly overexpressed in multiple human cancers, while normal cell types exhibit low expression levels. Thus, NSUN1 is considered as an effective prognostic marker for cancer development (Jhiang, Yaneva, & Busch, 1990;
Uchiyama et al., 1997; Ueki et al., 1997). Nsun1 localizes predominantly to nucleoli but weaker cytoplasmic staining was also observed in the early developing mouse embryo (Cui et al., 2016). Although long suspected as an RCMT, it was only recently shown to catalyze the transfer of a methyl group onto C2870 of the 25S rRNA in yeast, which strongly affected 60S biogenesis (Sharma et al., 2013). In addition, human NSUN1 was found to be able to complement a yeast nopaline mutant indicating functional conservation (Bourgeois et al., 2015). Depletion of Nop2p in yeast also compromised the processing of 35S and 27S pre-rRNAs resulting in lower levels of 25S and 5.8S rRNA. It is not clear yet if this phenotype is related to Nop2p's RCMT function (Hong, Brockenbrough, Wu, & Aris, 1997). In a recent report, human NSUN1 was found to interact with the chromatin regulator bromodomain containing protein 4 (BRD4) and actively elongating RNA polymerase II, and this association was increased in 5-azacytidine-resistant versus sensitive leukemia cell lines and in patient derived samples (Cheng et al., 2018) suggesting a role for NSUN1 in transcriptional regulation. It remains unknown at this point if NSUN1 participates in this process by way of its RCMT capacity and what its potential methylation substrates are or if NSUN1 can act as an RNA binding protein in the context of nascent transcription.

Nsun5: The enzyme responsible for modifying the second m^5C position in eukaryotic rRNA is Nsun5 (or Rcm1p in yeast). Rcm1p, which is localized in nucleoli, was identified as the enzyme modifying C2278 in 25S rRNA (Gigova, Duggimpudi, Pollex, Schafer, & Kos, 2014; Sharma et al., 2013). Loss of rcm1 resulted in strongly decreased stability of the 60S ribosomal subunit due to impaired association of several ribosomal proteins (Schosserer et al., 2015). Interestingly, knock-down of Nsun5 in flies, worms and in yeast resulted in extended life span that was dependent on low-energy nutrition. Moreover, it was observed that reduced Nsun5 levels correlated with decreased translational fidelity and promoted the recruitment of stress-associated mRNAs to the ribosomal fraction. The physiological downregulation of Nsun5 during senescence and the concomitantly reduced C2278 rRNA methylation might therefore represent a mechanism to cope with aging-related effects of stress (Schosserer et al., 2015). In humans, the NSUN5 gene locus lies in a large chromosomal deletion region that is associated with the multisystemic genetic disorder Williams-Beuren Syndrome suggesting a contribution of NSUN5 to the disease phenotype. In this region, two additional copies of NSUN5 (NSUN5B, C) can be found. These genes are transcribed but give rise to shorter open reading frames, and it is currently not clear if the corresponding transcripts are translated (Schubert, 2009).

Nsun4: The third rRNA-specific RCMT is Nsun4. It is imported into mitochondria via an N-terminal 26 amino acid motif, which is cleaved after its import (Cámara et al., 2011). Nsun4 forms a stoichiometric complex with the mitochondrial regulatory factor MTERF4, which is required for the recruitment of Nsun4 to the large subunit of the mitochondrial ribosome (Cámara et al., 2011). In the absence of MTERF4 (and Nsun4) from the large subunit, mitochondrial ribosome assembly is strongly impaired (Cámara et al., 2011). However, the methylation target of Nsun4 is C911 on the small subunit 12S rRNA and methylation of this site is independent of MTERF4 (Metodiev et al., 2014). Nsun4 inactivation in mice resulted in embryonic lethality, and conditional knock-out in the heart caused cardiomyopathy with mitochondrial dysfunction underscoring its pivotal role in the regulation of mitochondrial ribosome assembly (Metodiev et al., 2014).

3.3.9 | Nsun2: an enzyme with versatile target specificity

Nsun2 was first studied two decades ago in yeast, showing that the disruption of the Nsun2 gene had neither an effect on ribosome synthesis nor on cell growth, but that cells exhibited increased sensitivity for the antibiotic paromomycin and that Nsun2 localized to the nucleolus in yeast (Wu, Brockenbrough, Paddy, & Aris, 1998). In human cells, NSUN2 was found at different cellular locations in the course of the cell cycle: in G1, it was mostly detected in the nucleolus, during S phase, it was distributed between nucleoli and nucleoplasm, in G2, it localized to the cytoplasm and during M-phase, it was detected at the centrioles (Frye & Watt, 2006). Site-specific phosphorylation of Nsun2 at Ser139 by the cell cycle related kinase Aurora B leads to release of Nsun2 from complexes with nucleolar proteins at the onset of mitosis and also significantly reduces its methyltransferase activity (Sakita-Suto et al., 2007). The first identified Nsun2 methylase substrates were tRNAs. Brzezicha and colleagues showed that NSUN2 knockout mice were resistant to tRNA methylation. Brzezicha et al. (2006). Further research revealed that Nsun2 methylates most of the transcribed tRNAs at the variable loop region (C47-C50) (Blanco et al., 2014). In Nsun2−/− knockout mice, methylation at these positions was lost in tRNA^Gly, tRNA^Leu, tRNA^Asp, and tRNA^Val, while C38 methylation installed by Dnmt2 was unaffected (Tuorto et al., 2012). Furthermore, Dnmt2−/−/Nsun2−/− double knockout-mice showed a reduction in total tRNA methylation by more than 90% illustrating that the majority of actively transcribed tRNAs are substrates for Nsun2 and Dnmt2 (Blanco et al., 2014). Accordingly, overall protein synthesis was markedly reduced in Dnmt2−/−/Nsun2−/− mouse embryonic fibroblasts, while translation in single knockout cells was not affected (Tuorto et al., 2012). Besides tRNAs, however, Nsun2 was found to methylate several other ncRNAs (Hussain, Sajini, et al., 2013; Khoddami & Cairns, 2013). For instance, Nsun2 methylates specific vtRNAs, and this modification affects processing of a precursor vtRNA into svRNAs (Hussain, Sajini, et al., 2013). Last but not least,
Nsun2 was described in several studies to target mRNAs (Hussain, Sajini, et al., 2013; Li, Li, et al., 2017; Squires et al., 2012; H. Tang et al., 2015; Xing et al., 2015). Recently, it was shown that only overexpression/suppression of NSUN2 but not of any other NSUN enzyme, affected overall m\(^5\)C levels in mRNA from HeLa cells (Yang et al., 2017).

Nsun2 has been implicated in a variety of biological pathways. It was identified as a direct target gene for activation by the transcription factor Myc (Frye & Watt, 2006). Knockdown of Nsun2 resulted in the reduction of Myc-induced proliferation as well as Myc-induced terminal differentiation of primary human keratinocytes. Consistent with that, Nsun2 was found to be overexpressed in papillomas and squamous cell carcinomas in the mouse (Frye & Watt, 2006), as well as in a variety of human cancer types (Okamoto et al., 2012). Dnmt2\(^{-/-}\)/Nsun2\(^{-/-}\) double knockout mice revealed various developmental defects. The mice appeared consistently smaller and lighter than their wild-type littermates and died before P3. The organization of the cerebral cortex was impaired, and they presented with an immature skeleton with incomplete ossification. Furthermore, altered cellular lipid storage with a strong reduction of brown adipose tissue was observed (Tuorto et al., 2012). Nsun2 was further implicated in testis differentiation (Hussain et al., 2013). Male Nsun2\(^{-/-}\) mice were sterile due to decreased testes size and a severe reduction in spermatids numbers. Deletion of Nsun2 leads to the absence of chromatoid bodies, which contain various RNA processing factors, and to abortion of spermatogenesis in the pachytene stage by blocking the progression of prophase I of male meiosis (Hussain, Tuorto, et al., 2013).

In humans, mutations within the Nsun2 gene have been linked to autosomal-recessive intellectual disability (Khan et al., 2012). Genotyping of an affected consanguineous Pakistani family revealed a homozygosity-by-descent (HBD) locus within region 5p15.32, and a homozygous base substitution was mapped to exon 19 of Nsun2, resulting in a glycine to arginine change (G679R). Overall, all individuals showed significant delay in development and speech. Height and weight were below the 5% percentile but muscle tone was increased in all limbs. Although patients had smaller heads, computer tomography of the brain did not show any abnormalities. Expression of mutant Nsun2(G679R) protein in mouse brain prevented its localization to the nucleolus in Purkinje cells of the cerebellum (Khan et al., 2012). A different mutation in the Nsun2 gene affecting splicing of exons 5–7 was correlated with symptoms of the Dubowitz syndrome, characterized by small stature, intellectual disability, mild microcephaly and a distinct facial appearance. Aberrant splicing resulted in severe reduction of protein levels of the enzyme (Martinez et al., 2012). A recent report provides evidence that the decrease in brain size upon loss of Nsun2 in mice might derive from the inability to generate sufficient amounts of differentiated neurons during neurogenesis (Flores et al., 2017). Developing Nsun2\(^{-/-}\) mouse cerebral cortex exhibited a thicker layer and higher numbers of intermediate neuronal progenitors but a decrease in upper-layer neurons. tRNAs within Nsun2\(^{-/-}\) brains were found to be hypomethylated and cleaved by angiogenin, which led to an accumulation of 5′ derived tRNA fragments. Indeed, injection of an angiogenin inhibitor into pregnant Nsun2\(^{+/+}\) mice rescued the brain development phenotype of the progeny. It was further shown that Nsun2-depleted human neuroepithelial stem cells exhibited migration defects. Thus, it was suggested that Nsun2-dependent tRNA methylation is required for differentiation and migration of neural progenitor cells during brain development (Flores et al., 2017). Nsun2 has also been implicated in vascular endothelial inflammation and atherosclerosis. Dynamic methylation of the mRNA of intercellular adhesion molecule 1 (ICAM-1), which is a critical factor for inflammatory and immune responses of the endothelium, upregulated expression of ICAM-1 at the translational level. Elevated ICAM-1 levels partly mediated TNF-α or homocysteine induction of the endothelial inflammatory response, which in turn led to the adhesion of leukocytes to the endothelial cells (Luo, Feng, Xu, Wang, & Wang, 2016).

### 3.3.10 | Dnmt2, Nsun3 and Nsun6: methylating tRNA

*Dnmt2*: In multicellular organisms, Dnmt2 was among the first confirmed RNA m\(^5\)C methyltransferases (Brzezicka et al., 2006; Goll et al., 2006). It was originally thought to act as a DNA methyltransferase, since it shows all sequence and structural characteristics of a DNA methyltransferase, except for a specific nucleic acid binding domain (Dong et al., 2001; Wilkinson, Bartlett, Nurse, & Bird, 1995). However, overall genomic DNA methylation was unaltered in Dnmt2-deficient mice, *Drosophila* and *Arabidopsis*, and Dnmt2 localized predominantly to the cytoplasm (Goll et al., 2006; Okano, Xie, & Li, 1998). Goll et al. showed in vitro experiments that recombinant Dnmt2 was not able to methylate genomic DNA but instead targeted tRNA\(^{\text{Asp}}\) for methylation at C38 in the anticodon stem loop (Goll et al., 2006). Later, two additional tRNAs, namely tRNA\(^{\text{Gly}}\) and tRNA\(^{\text{Val}}\), were found to be methylated by Dnmt2 at position C38 (Schaefer et al., 2010; Tuorto et al., 2012). Interestingly, all three tRNAs share the same sequence around the methylated C38 (5′ CA-m\(^5\)C-GGC 3′), suggesting target recognition by the enzyme in a sequence-specific manner. Dnmt2-deficient mice, flies and plants were viable and fertile and were morphologically indistinguishable from wild-type organisms (Goll et al., 2006). Closer inspection of Dnmt2-deficient mice, however, revealed tissue specific differentiation defects reflected in delayed endochondral ossification and hematopoiesis. This phenotype was proposed to be caused by decreased translational fidelity during the decoding of aspartate codons due to the absence of C38 methylation in the responsible tRNA (Tuorto et al., 2015). Another study reported that Dnmt2-deficient mouse cells exhibited increased amounts of uncharged tRNA\(^{\text{Asp}}\) and impaired synthesis of poly-asp-containing proteins (Shanmugam et al., 2015).
Thus, Dnmt2-mediated methylation of tRNAAsp C38 appears to be required for proper recognition by its cognate tRNA synthetase, and C38 methylation may be furthermore involved in the decoding of asp codons (Jeltsch et al., 2017). Another phenotype detected in Dnmt2-deficient mice is cardiac hypertrophy. As a potential defect-mediating mechanism the non-coding RNA Rn7sk was proposed (Ghanbarian et al., 2016). Rn7sk associates with the transcription elongation regulator P-Tefb inhibiting its phosphorylation of the C-terminal domain of RNA polymerase II. In Dnmt2-deficient embryonic stem cells, the association between Rn7sk with P-Tefb was strongly reduced, and merRIP analysis revealed severely decreased cytosine methylation of Rn7sk isolated from Dnmt2-deficient hearts suggesting that methylation of Rn7sk by Dnmt2 might regulate its interaction with P-Tefb. Consequently, overactive P-Tefb might lead to enhanced transcription and cardiac hypertrophy (Ghanbarian et al., 2016). Interestingly, DNMT2 was recently also identified in a complex with P-Tefb, NSUN3 (see below) and hnRNPK at phospho-serine 2-containing active RNA pol II in human leukemia cells. However, the role for DNMT2 in this complex is unknown (Cheng et al., 2018).

While overall Dnmt2-deficient animals show a relatively mild phenotype under standard laboratory conditions, under stress conditions (heat, oxidative, arsenite stress) Dnmt2-deficient flies exhibited shortened life-span, while dDnmt2 over-expression resulted in increased stress resistance (Lin, Tang, Reddy, & Shen, 2005; Schaefer et al., 2010). These stress-associated phenotypes of Dnmt2 mutants might be linked to their translation defects, particularly of poly-asp containing proteins that might be required for adequate stress response (Shanmugam et al., 2015). Alternatively or in addition, the ability of Dnmt2 to protect tRNA from cleavage by methylating C38 may also play a role in this process. tRNA fragments can impair translation (Gebetsberger & Polacek, 2013; Sobala & Hutvagner, 2013), and in Drosophila, they can serve as Dicer 2 (Dcr-2) substrates and inhibit the action of Dcr-2 on long double stranded RNAs thus regulating siRNA pathways (Durdevic, Mobin, Hanna, Lyko, & Schaefer, 2013). Drosophila mutants of Dnmt2 have also been connected to de-repression of transposable elements (TE) in the genome (Phalke et al., 2009). A recent report confirmed and expanded these findings and showed that Nsun2 mutation, too, leads to increased genomic instability. It was suggested that the mechanism by which this occurs is via a disturbance of tRNA metabolism (stability, fragmentation), thus affecting translation of chromatin regulatory proteins involved in the silencing of mobile elements. However, it is also interesting to note that the Dnmt2-deletion phenotype could be rescued by expression of a catalytically inactive Dnmt2 transgene, which raises the possibility that Dnmt2 might also engage in an RCMT-independent function in this process (Genenncher et al., 2018). Finally, Dnmt2 was found to be critically required for RNA-mediated transgenerational inheritance of phenotypic variations. These phenomena are exemplified by the epigenetic modulation of the Kit gene, resulting in altered fur coloration of mice, and the modulation of the Sox9 gene, resulting in an overgrowth phenotype. In Dnmt2-deficient mice the paramutation phenotypes were not transmitted to the next generation. The authors proposed a model in which Dnmt2-dependent methylation of small paramutation inducing RNAs in the sperm or in the early embryo would protect those RNAs from cleavage thus preserving them to elicit the mutant phenotype (Kiani et al., 2013).

Nsun3: One of the latest additions to the group of tRNA-methylating enzymes is NSUN3. Similar to its closest relative in the Nsun family of RCMTs, it localizes to the mitochondrial matrix in human cells and mouse embryonic stem cells, where it introduces an m5C at the “wobble position” C34 of mitochondrial (mt)-tRNAMet (Haag et al., 2016; Nakano et al., 2016; Trixl et al., 2018; Van Haute et al., 2016). It was shown that the m5C modification can be further oxidized by the alpha-ketoglutarate and Fe(II)-dependent dioxygenase ALKBH1/ABH1 to generate 5-formylcytidine (f5C) at this position (Haag et al., 2016; Kawarada et al., 2017). This modification is critical for the translation of methionine-encoding AUA codons in mitochondria (Takemoto et al., 2009). Consistently, mutation of NSUN3 results in reduced mitochondrial protein translation and mitochondrial respiration (Haag et al., 2016; Nakano et al., 2016; Trixl et al., 2018; Van Haute et al., 2016). Catalytic inactivation of Nsun3 in mouse embryonic stem cells further caused impaired differentiation into the neuroectodermal lineage (Trixl et al., 2018). In order to methylate C34, NSUN3 requires an intact anticodon stem loop (Haag et al., 2016) which may explain why two mutations in this region that are associated with mitochondrial disease in humans substantially reduced C34 methylation by NSUN3 in vitro (Nakano et al., 2016). In a recent report, NSUN3 was found to form a complex with hnRNPK, DNMT2 and P-TEFb at elongating RNA polymerase II sites in leukemia cells implying a role for NSUN3 in transcriptional regulation in the nucleus (Cheng et al., 2018). This is surprising as NSUN3 was so far found to be a mitochondrial protein (Haag et al., 2016; Trixl et al., 2018; Van Haute, Powell, & Miniczuk, 2017). Although it was not shown if NSUN3 relocates to the nucleus specifically in leukemia cells or if a minor portion of NSUN3 might show nuclear localization in all cell types, it will be interesting to see, whether the proposed nuclear role of NSUN3 involves methylation of nascent mRNA or whether its function is methylation-independent.

Nsun6: This enzyme was shown to reside in the cytoplasm and to partially localize to the Golgi apparatus. Using UV-crosslinking and analysis of cDNA (CRAC) and aza-IP Nsun6 was identified as a tRNA-specific RCMT with specificity for tRNAThr and tRNACys in human cells. It introduces an m5C at position 72 in the 3′ acceptor stem. Nsun6 binding to its tRNA substrates requires the presence of a 3′-CCA sequence in order to carry out methylation, because tRNACys and tRNAThr
mediates on the way to DNA demethylation, hm5C has been demonstrated to have regulatory potential on its own. For RNA, a be the end of the story. In DNA m5C is known to be further oxidized by the TET enzymes to 5-hydroxymethyl cytosine. In this area will greatly contribute towards an understanding of the biological functions of m5C on RNA. Finally, m5C may not be the end of the story. In DNA m5C is known to be further oxidized by the TET enzymes to 5-hydroxymethyl cytosine (hm5C), 5-formylcytosine (f5C) and eventually to 5-carbonylcytosine (ca5C). While f5C and ca5C are mostly considered intermediates on the way to DNA demethylation, hm5C has been demonstrated to have regulatory potential on its own. For RNA, a recent report has shown that hm5C can be found in mRNAs from *Drosophila*, especially from the fly brain. The *Drosophila* Tet enzyme is responsible for generation of this mark and *Tet*-mutant flies exhibit impaired brain development and reduced hm5C levels (Delatte et al., 2016). Existence of hm5C has also been shown for cells of mouse and human origin (Fu et al., 2014; Huber et al., 2015), although these findings could not be confirmed in another study (Legrand et al., 2017). Taken together, it is clear that large areas on the m5C epitranscriptomic map are still obscure awaiting discovery in the years to come.

3.3.11 | Nsun7: specialized for enhancer RNAs?

According to the Stanford SOURCE database, 50% of all existing NSUN7 protein is testis-derived. Microarray analysis of mouse testis revealed that Nsun7 expression is the highest in spermatocytes and haploid spermatids (Shima, McLean, McCarrey, & Griswold, 2004). During embryogenesis, however, Nsun7 is broadly expressed in different mouse tissues (Chi & Delgado-Olguin, 2013). In a forward genetic screen in mice a mutation in exon 7 of Nsun7, which causes a premature stop codon and severe truncation of the protein was linked to impaired male fertility due to reduced motility and aberrant swimming behavior of the sperm (Harris, Marquez, Suarez, & Schimenti, 2007). However, the molecular basis for this defect remains unknown. In a recent report, Nsun7 was found to interact with the transcriptional coactivator PGC-1α, which contributes to metabolic response of a cell, and to colocalize with PGC-1α at various target genes. Specifically, knock-down of Nsun7 reduced enhancer RNA (eRNA) transcripts of PGC-1α-controlled genes and this led to a decrease in transcript levels of these genes. Knock-down of Nsun7 also apparently reduced cytosine methylation in eRNAs suggesting that m5C in eRNAs enhances their stability (Aguilo et al., 2016).

4 | CONCLUSIONS AND PERSPECTIVES

Research on RNA modifications, particularly in mRNA and other lower abundance RNA species, is an emerging and highly dynamic field with new discoveries reported almost on a monthly basis. Yet despite all the exciting new additions to our knowledge, the field is still in its infancy. With respect to m5C, this fact is illustrated by seemingly incongruent findings on the nature and distribution of m5C in transcriptomes from different sources. Doubtlessly, the refinement and further development of the methodology to map m5C on RNA will be an important issue to solve in the future in order to generate high confidence m5C methylome data. In this respect, the ongoing improvement of next generation sequencing techniques, such as nanopore sequencing, may hold great potential. Moreover, further development of robust data analysis tools and statistics approaches will be required to deal with large amounts of generated data and to minimize miscalling of m5C sites. The elucidation of “writers” and in particular “readers” and “erasers” of the m5C epitranscriptomic code is only at its beginning, and discoveries in this area will greatly contribute towards an understanding of the biological functions of m5C on RNA. Finally, m5C may not be the end of the story. In DNA m5C is known to be further oxidized by the TET enzymes to 5-hydroxymethyl cytosine (hm5C), 5-formylcytosine (f5C) and eventually to 5-carbonylcytosine (ca5C). While f5C and ca5C are mostly considered intermediates on the way to DNA demethylation, hm5C has been demonstrated to have regulatory potential on its own. For RNA, a recent report has shown that hm5C can be found in mRNAs from *Drosophila*, especially from the fly brain. The *Drosophila* Tet enzyme is responsible for generation of this mark and *Tet*-mutant flies exhibit impaired brain development and reduced hm5C levels (Delatte et al., 2016). Existence of hm5C has also been shown for cells of mouse and human origin (Fu et al., 2014; Huber et al., 2015), although these findings could not be confirmed in another study (Legrand et al., 2017). Taken together, it is clear that large areas on the m5C epitranscriptomic map are still obscure awaiting discovery in the years to come.

ACKNOWLEDGMENTS

Research in the author’s lab was supported by the Austrian Science Fund (FWF) P27024-BBL to A.L.
CONFLICT OF INTEREST
The authors have declared no conflicts of interest for this article.

RELATED WIREs ARTICLES
mRNA methylation by NSUN2 in cell proliferation
Mapping and significance of the mRNA methylome
RNA methylation in nuclear pre-mRNA processing

ORCID
Alexandra Lusser https://orcid.org/0000-0002-2226-9081

REFERENCES
Aguilo, F., Li, S., Balasubramaniyan, N., Sancho, A., Benko, S., Zhang, F., ... Walsh, M. J. (2016). Deposition of 5-Methylcytosine on enhancer RNAs enables the coactivator function of PGC-1α. Cell Reports, 14(3), 479–492. http://doi.org/10.1016/j.celrep.2015.12.043
Amort, T., Rieder, D., Wille, A., Khokhlova-Cubberley, D., Riml, C., Trixl, L., ... Lusser, A. (2017). Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. Genome Biology, 18(1), 1. http://doi.org/10.1186/s13059-016-1139-1
Amort, T., Souldière, M. F., Wille, A., Jia, X.-Y., Fieg, H., Wöhrle, H., ... Lusser, A. (2013). Long non-coding RNAs as targets for cytosine methylation. RNA Biology, 10(6), 1003–1008. http://doi.org/10.4161/rna.24454
Amos, H., & Korn, M. (1958). 5-Methyl cytosine in the RNA of Escherichia coli. Biochimica et Biophysica Acta, 29(2), 444–445.
Blanco, S., Dietmann, S., Flores, J. V., Hussain, S., Kutter, C., Humphreys, P., ... Frye, M. (2014). Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. The EMBO Journal, 33(18), 2020–2039. http://doi.org/10.15252/embj.201489282
Bohnsack, M. T., & Sloan, K. E. (2018). The mitochondrial epitranscriptome: The roles of RNA modifications in mitochondrial translation and human disease. Cellular and Molecular Life Sciences: CMLS, 75(2), 241–260. http://doi.org/10.1007/s00018-017-2598-6
Bourgeois, G., Ney, M., Gaspar, I., Aigueperse, C., Schaefer, M., Kellner, S., ... Humphreys, P., Amos, H., & Korn, M. (1958). 5-Methyl cytosine in the RNA of Escherichia coli. Biochimica et Biophysica Acta, 29(2), 444–445.
Cohn, W. E. (1960). Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: Isolation, structure, and chemical characteristics.

Related Wires Articles
Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O., & Sorek, R. (2013). Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genetics*, 9(6), e1003602. http://doi.org/10.1371/journal.pgen.1003602

Edmonds, M., & Abrams, R. (1960). Polynucleotide biosynthesis: Formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei. *The Journal of Biological Chemistry*, 235, 1142–1149.

Ehrenhofer-Murray, A. E. (2017). Cross-talk between Dnmt2-dependent RNA methylation and Queuosine modification. *Biomoecules*, 7(1), 14. http://doi.org/10.3390/biom7010014

Fischer, N., Neumann, P., Konevega, A. L., Bock, L. V., Ficner, R., Rednina, M. V., & Stark, H. (2015). Structure of the E. coli ribosomal EF-Tu complex at <3 Å resolution by cryo-EM. *Nature*, 520(7548), 567–570. http://doi.org/10.1038/nature14275

Flatau, E., Gonzales, F. A., Michalowsky, L. A., & Jones, P. A. (1984). DNA methylation in 5-aza-2′-deoxycytidine-resistant variants of C3H 10T1/2 C18 cells. *Molecular and Cellular Biology*, 4(10), 2098–2102.

Flores, J. V., Cordero-Espinoza, L., Oeztuerk-Winder, F., Andersson-Rolf, A., Selmi, T., Blanco, S., … Frye, M. (2017). Cytosine-5 RNA methylation regulates neural stem cell differentiation and mobility. *Stem Cell Reports*, 8(1), 112. http://doi.org/10.1016/j.stemcr.2016.11.014

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., … Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, 89(5), 1827–1831. http://doi.org/10.1073/pnas.89.5.1827

Frye, M., & Watt, F. M. (2006). The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors. *Current Biology*, CB, 16(10), 971–981. http://doi.org/10.1016/j.cub.2006.04.027

Fu, L., Guerrero, C. R., Zhong, N., Amato, N. J., Liu, Y., Liu, S., … Wang, Y. (2014). Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *Journal of the American Chemical Society*, 136(33), 11582–11585. http://doi.org/10.1021/ja505305z

Gebetsberger, J., & Polacek, N. (2013). Slicing tRNAs to boost functional ncRNA diversity. *Ehrenhofer-Murray, A. E.* (2017). Cross-talk between Dnmt2-dependent RNA methylation and Queuosine modification. *Biomoecules*, 7(1), 14. http://doi.org/10.3390/biom7010014

Hussain, S., Tuorto, F., Menon, S., Blanco, S., Cox, C., Flores, J. V., … Frye, M. (2013). Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biology*, 14(11), 215–210. http://doi.org/10.1186/gb1413

Hussain, S., Sajini, A. A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., … 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. http://doi.org/10.1016/j.celrep.2013.06.029

Hussain, S., Tuorto, F., Menon, S., Blanco, S., Cox, C., Flores, J. V., … Frye, M. (2013). The mouse cytosine-5 RNA methyltransferase NSun2 is a component of the chromatin body and required for testis differentiation. *Molecular and Cellular Biology*, 33(8), 1561–1570. http://doi.org/10.1128/MCB.01523-12

Jeltsch, A., Ehrenhofer-Murray, A., Jurkowski, T. P., Lyko, F., Reuter, G., Ankri, S., … Helm, M. (2017). Mechanism and biological role of Dnmt2 in nucleic acid methylation. *RNA Biology*, 14(9), 1108–1123. http://doi.org/10.1089/rna.2016.1191737

Jiang, S. M., Yaneva, M., & Busch, H. (1990). Expression of a human mRNA-association nucleotide antigen p120. *Cell Growth & Differentiation*, 1(7), 319–324.

Jurkowski, T. P., Meusburger, M., Phalke, S., Helm, M., Nellen, W., Reuter, G., & Jeltsch, A. (2008). Human Dnmt2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. *RNA (New York, NY)*, 14(8), 1663–1670. http://doi.org/10.1261/rna.970408

Jüttnermann, R., Li, E., & Jaenisch, R. (1994). Toxicity of 5-aza-2′-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proceedings of the National Academy of Sciences of the United States of America*, 91(25), 11797–11801.

Kawarada, L., Suzuki, T., Ohira, T., Hirata, S., Miyaiuchi, K., & Suzuki, T. (2017). ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications. *Nucleic Acids Research*, 45(12), 7401–7415. http://doi.org/10.1093/nar/gkx354
Wilkinson, C. R., Bartlett, R., Nurse, P., & Bird, A. P. (1995). The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. *Nucleic Acids Research, 23*(2), 203–210.

Wu, P., Brockenbrough, J. S., Paddy, M. R., & Aris, J. P. (1998). NCL1, a novel gene for a non-essential nuclear protein in *Saccharomyces cerevisiae*. *Gene, 220*(1–2), 109–117.

Wyatt, G. R. (1950). Occurrence of 5-methylcytosine in nucleic acids. *Nature, 166*(4214), 237–238.

Xing, J., Yi, J., Cai, X., Tang, H., Liu, Z., Zhang, X., … Wang, W. (2015). NSun2 promotes cell growth via elevating cyclin-dependent kinase 1 translation. *Molecular and Cellular Biology, 35*(23), 4043–4052. http://doi.org/10.1128/MCB.00742-15

Xiong, X., Li, X., & Yi, C. (2018). N1-methyladenosine methylome in messenger RNA and non-coding RNA. *Current Opinion in Chemical Biology, 45*, 179–186. http://doi.org/10.1016/j.cbpa.2018.06.017

Yang, X., Yang, Y., Sun, B.-F., Chen, Y.-S., Xu, J.-W., Lai, W.-Y., … Yang, Y. G. (2017). 5-methylcytosine promotes mRNA export—NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Research, 27*(5), 606–625. http://doi.org/10.1038/cr.2017.55

Zachau, H. G., Dütting, D., & Feldmann, H. (1966). Nucleotidesequenzen zweier serinspezifischer Transfer-Ribonucleinsäuren. *Angewandte Chemie, 78*(7), 392–393. http://doi.org/10.1002/ange.19660780714

Zhao, B. S., Roundtree, I. A., & He, C. (2016). Post-transcriptional gene regulation by mRNA modifications. *Nature Reviews Molecular Cell Biology, 18*(1), 31–42. http://doi.org/10.1038/nrm.2016.132

---

**How to cite this article:** Trixl L, Lusser A. The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark. *WIREs RNA*. 2019;10:e1510. [https://doi.org/10.1002/wrna.1510](https://doi.org/10.1002/wrna.1510)