Genome **Biology**

REVIEW

Characterizing 5-methylcytosine in the mammalian epitranscriptome

Shobbir Hussain[*†], Jelena Aleksic[†], Sandra Blanco, Sabine Dietmann and Michaela Frye[*]

Abstract

The post-transcriptional modification 5-methylcytosine (m^5C) occurs in a wide range of coding and non-coding RNAs. We describe transcriptome-wide approaches to capture the global m^5C RNA methylome. We also discuss the potential functions of m^5C in RNA and compare them to 6-methyladenosine modifications.

Introduction

Post-transcriptional changes in RNA processing are essential regulators for most, if not all, cellular responses. Although RNA modifications are more prevalent and diverse in their chemical nature than DNA modifications [1], our knowledge of their occurrence and function in RNA is generally limited. There are approximately 150 known ribonucleoside modifications: most of them have been found in tRNA and rRNA, but some occur in mRNA [1-4]. Post-transcriptional modifications are highly likely to add complexity to RNA-mediated functions.

The advent of next-generation sequencing (NGS) tools has enabled the identification of RNA modifications both globally and in a substrate-specific manner. 6-Methyladenosine (m^6A) was the first modification to be characterized, and is now known to be present in several types of RNA and, most notably, is highly enriched around stop codons in many mRNAs [2-4].

Another known modification in RNA is 5-methylcytosine (m^5C). Although m^5C is a well-characterized modification in DNA, its precise regulatory functions in RNA remain unclear [5,6]. Until recently, the detection of RNA methylation involved the digestion of highly purified RNA followed by separation techniques, such as high performance liquid chromatography (HPLC) and mass spectrometry, which only allowed the identification of m^5C in stable and highly

* Correspondence: sh465@cam.ac.uk; mf364@cam.ac.uk
†Equal contributors
Wellcome Trust - Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK

abundant tRNAs and rRNAs [7-10]. Labeling techniques with ^3H in living cells allowed detection of m^5C in mRNA and viral RNAs [11,12]. However, the deposition of m^5C into mRNA remained controversial [13-17]. The recent advances in high-throughput techniques, combined with NGS, have renewed interest in the field, and have led to the identification of m^5C as a widespread modification in coding and non-coding (nc)RNAs [18].

A fraction of these RNAs were found to be specifically methylated by the RNA methylase NSun2, including mRNAs, ncRNAs and several tRNAs [18]. NSun2 had been previously shown to methylate tRNAs at various positions [19-22]. Additional NSun2-methylated coding and ncRNAs were identified in two studies published in 2013 that used customized RNA immunoprecipitation approaches followed by NGS [23,24].

The regulatory functions of m^5C modifications in RNA are still not fully understood. In tRNAs, in vitro cytosine-5 methylation can affect Mg^{2+} binding to tRNA molecules, which in turn influences the anticodon stem loop conformation and stabilizes the secondary structure [25,26]. Cytosine-5 methylation alone, or in combination with other nonessential tRNA modifications, can also protect from degradation or cleavage [22,27-29]. In rRNA, m^5C is thought to play a role in translation [30]. Synthetic cytosine-5 methylated mRNAs exhibit increased stability, and loss of methylation in the 3' UTR of p16 has been reported to reduce its stability [31,32].

The biological functions of tRNA m^5C-methylation are linked to the regulation of protein translation in stress pathways and tissue differentiation in yeast, Drosophila, fish and mouse [19,22,29,33-36]. Mutations in the NSUN2 gene in humans cause an autosomal recessive syndrome characterized by intellectual disability, skin disorders and growth retardation [37-40]. These findings, together with studies carried out in NSun2-deficient mice and cell lines, suggest a wide-ranging role for m^5C modifications in RNA, including cellular signaling, tissue development and differentiation, and cancer [19,21,22,36,41-43].

In this review, we compare the current methods to identify m5C in the mammalian transcriptome with a particular focus on NSun2-mediated methylation. We further discuss the potential of these initial studies to comprehensively determine the global but enzyme-specific cytosine-5 RNA methylome. In addition, we compare studies focusing on m6A and m5C modifications and consider the likely benefits of characterizing and elucidating the functions of the mammalian epitranscriptome.
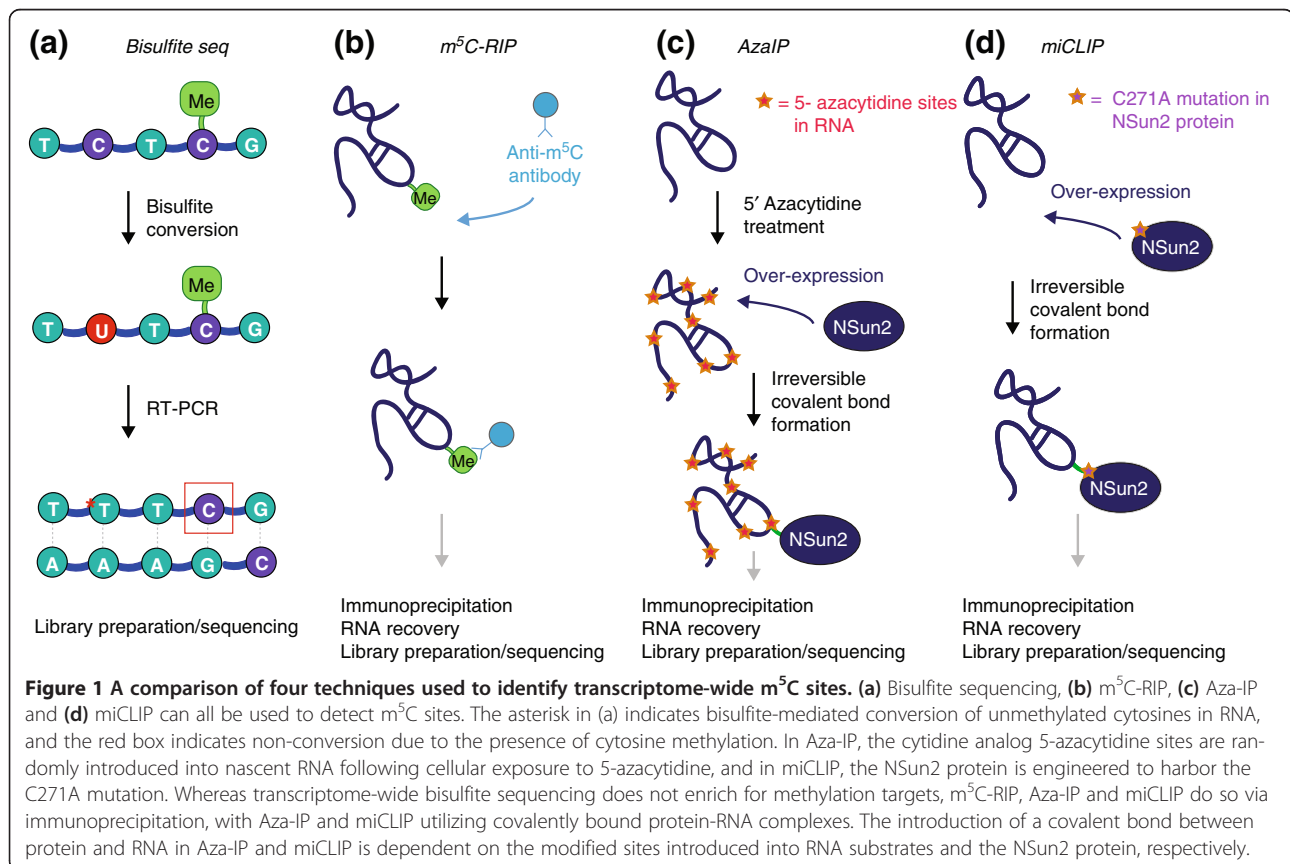
Mapping cytosine-5 methylation in the mammalian transcriptome

Previous methods for detecting m5C in RNA have required exceedingly high amounts of RNA and only reproducibly identified methylated sites in highly abundant RNAs, such as tRNAs and rRNAs. Over the last four years, NGS has allowed researchers to successfully develop more sensitive techniques. Bisulfite sequencing for the detection of m5C in RNA was first described in 2009 [44]. In 2012, the method was combined with NGS and provided the first transcriptome-wide view of the human cytosine-5 RNA methylome (Figure 1A) [18]. Earlier this year, three independent studies used RNA immunoprecipitation (RIP) followed by deep sequencing to detect m5C globally (Figures 1B-D) [23,24,45]. Khoddami and Cairns, and our own study [24], developed similar but

technically distinct approaches to detect enzyme-specific deposition of m5C: 5-azacytidine-mediated RNA immunoprecipitation (Aza-IP) (Figure 1C) [23] and methylation-individual nucleotide resolution crosslinking immmunoprecipitation (miCLIP) (Figure 1D) [24]. Both Aza-IP and miCLIP rely on covalent bond formation between the RNA methylase and substrate but differ in the way in which the stable covalent bond formation is achieved. We now describe these techniques in detail and discuss their specific merits and disadvantages.

Bisulfite sequencing

Bisulfite sequencing, a method based on the chemical deamination of cytosines and originally developed to detect m5C in DNA, was previously adapted for use with RNA (Figure 1A) [44]. The technique is based on the differential chemical reactivity of m5C compared with cytosine. Sodium bisulfite causes the deamination of unmethylated cytosines into uridines in single-stranded DNA or RNA, while m5C remains unconverted [6]. A crucial parameter is the fraction of cytosine that is converted to uridine, but a high conversion rate is only achieved by prolonged incubation under consecutive acidic and alkaline conditions, which also causes RNA degradation. This degradation can compromise the

Figure 1 A comparison of four techniques used to identify transcriptome-wide m5C sites. (a) Bisulfite sequencing, **(b)** m5C-RIP, **(c)** Aza-IP and **(d)** miCLIP can all be used to detect m5C sites. The asterisk in (a) indicates bisulfite-mediated conversion of unmethylated cytosines in RNA, and the red box indicates non-conversion due to the presence of cytosine methylation. In Aza-IP, the cytidine analog 5-azacytidine sites are randomly introduced into nascent RNA following cellular exposure to 5-azacytidine, and in miCLIP, the NSun2 protein is engineered to harbor the C271A mutation. Whereas transcriptome-wide bisulfite sequencing does not enrich for methylation targets, m5C-RIP, Aza-IP and miCLIP do so via immunoprecipitation, with Aza-IP and miCLIP utilizing covalently bound protein-RNA complexes. The introduction of a covalent bond between protein and RNA in Aza-IP and miCLIP is dependent on the modified sites introduced into RNA substrates and the NSun2 protein, respectively.

subsequent reverse transcription and PCR amplification steps [18,46].

Using this approach, the occurrence of m^{5}C can be reproducibly and quantitatively detected in tRNA and rRNA [44]. Coupling bisulfite conversion of RNA with NGS has led to detection of m^{5}C in coding and non-coding RNAs, in addition to tRNAs [18,45]. Squires *et al.* [18] used polyA-enriched RNA and identified m^{5}C as a widespread modification in coding and non-coding RNAs.

m^{5}C-RNA immunoprecipitation

In contrast to Squires *et al.*, bisulfite sequencing of size-selected mRNA was used by Edelheit *et al.* [45] to avoid tRNA and rRNA contamination from the prokaryote *Sulfolobus solfataricus* and, despite covering most coding genes, this approach detected low methylation levels in only 15 coding genes. However, all m^{5}C sites were confirmed by RIP using a monoclonal antibody raised against m^{5}C (m^{5}C-RIP) (Figure 1B) [45]. m^{5}C-RIP has not been performed in mammalian cells as yet.

5-Azacytidine-mediated RNA immunoprecipitation

Aza-IP exploits the formation of a covalent bond between RNA and specific RNA methylases (Figure 1C) [23]. Both DNA and RNA m^{5}C methyltransferases form a temporary covalent bond between the catalytic cysteine in the enzymatic active site and carbon 6 of the targeted cytosine. This catalytic intermediate complex is later resolved to release the methylated cytosine at carbon 5 and to regenerate the free enzyme [47,48]. In the first step of the Aza-IP method, 5-azacytidine, a cytidine analog, is incorporated into nascent RNA molecules by RNA polymerases (Figure 1C) [23,49]. The only structural difference between cytidine and 5-azacytidine is a nitrogen substitution at carbon 5 that inhibits the release of the methylated RNA from its enzyme.

In a study by Khoddami and Cairns [23], this covalently bound protein-RNA complex was immunoprecipitated for the RNA methylases NSun2 and Dnmt2. The identification of the precise cytosine targeted by the enzymes was possible due to a specific transversion of cytosine to guanosine only at Dnmt2- and NSun2-methylated sites. Although it is unclear why and how this conversion occurs, it has been confirmed that there are Dnmt2-dependent m^{5}C sites in the three tRNAs Asp, Gly and Val [22,23,29]. Putative novel targets for Dnmt2 were the type I cytokeratin *KRT18* mRNA and *KRT18* pseudogene mRNA [23]. NSun2-dependent methylation sites occurred in tRNAs as well as several other ncRNAs. Although a potential site in the Src homology 2 domain containing F (SHF) mRNA was identified, statistically significant sites in other mRNAs were absent [23].

Methylation-individual nucleotide resolution crosslinking immunoprecipitation

In an alternative approach, we utilized covalently bound catalytic intermediate RNA-protein complexes to identify methylation substrates of NSun2 (Figure 1D). NSun2 releases methylated RNA from the enzyme-substrate complex through a highly conserved cysteine residue in the catalytic domain [50,51]. Mutating this cysteine to alanine results in a covalently bound stable RNA-protein complex [41,50,51]. In the human NSun2 protein, this cysteine is represented by C271 and the mutation to alanine (C271A) allowed capturing of stable NSun2-RNA complexes [24]. Combined with a customized individual-nucleotide resolution cross-linking immunoprecipitation (iCLIP) approach where the UV-crosslinking step is omitted, and followed by NGS, we named this method methylation-iCLIP (miCLIP). This approach identified NSun2-mediated cytosine-5 methylation at the nucleotide resolution [24,52]. miCLIP not only confirmed known tRNA and non-tRNA methylation sites of NSun2, but also identified novel sites in coding and non-coding RNAs [24].

Challenges of detecting site-specific 5-methylcytosines

Of the four system-wide techniques that have been used to identify m^{5}C sites in the epitranscriptome (Figure 1) [18,23,24,45], RNA bisulfite sequencing remains the gold standard. This is because it directly detects the enzymatic conversion of cytosines to 5-methylcytosines in RNA. However, RNA bisulfite sequencing has at least three major practical disadvantages: (1) cytosines are resistant to conversion when in double-stranded RNA and several modifications other than m^{5}C inhibit the C to U conversion, leading to false positives [53]; (2) the chemical conversion of the nucleotides leads to degradation of the RNA and requires extreme deep sequencing to reveal methylated RNAs of low abundance; and (3) it reveals no information about substrate specificity when, for instance, specific RNA methylases are knocked out. For instance, it cannot distinguish direct from indirect targets or capture the cross-reaction of additional RNA methylases.

m^{5}C-RIP, Aza-IP and miCLIP do not rely on chemical base conversion and avoid harsh chemical and thermal conditions that can lead to the degradation of target RNA. They also enrich for RNA substrates during the immunoprecipitation step. Therefore, RNAs with lower abundances can also be captured with standard deep sequencing protocols.

A key future goal in the m^{5}C epitranscriptomics field is to gain a methylase-specific view of the mammalian m^{5}C RNA methylome. The m^{5}C-RIP approach has yet to be performed in a mammalian system and, like bisulfite

sequencing, analysis of enzyme-specific m^{5}C modification requires genetic deletion or knockdown of the RNA methylase being investigated. Aza-IP and miCLIP not only allow detection of enzyme-specific methylation, but also have the advantage that the covalently bound methylase-RNA complexes can be purified under higher stringency conditions, thus reducing the background of non-specifically bound RNAs.

The mechanism of cytosine-5 RNA methylation is highly conserved, and both Aza-IP and miCLIP are applicable, in principle, for most, if not all, m^{5}C RNA methylases. Aza-IP relies on the chemical incorporation of 5-azacytidine into the entire transcriptome prior to isolation of protein-RNA complexes, which may compromise RNA stability and integrity. In addition, the analog will be incorporated into DNA and may alter transcriptional processes. Although Aza-IP was developed using overexpressed NSun2 and Dnmt2 proteins, it can in principle be performed on the endogenous methylases using the appropriate antibodies. In contrast, miCLIP depends on the overexpression of the mutant methylase protein, yet it does not involve any chemical modification of RNA, thus offering a distinct advantage. Since miCLIP only utilizes the last step of the catalytic process, in which the methylated substrate is released from the enzyme, it does not affect the methyl-transfer step, or the endogenous RNA structure, stability and integrity.

Particular consideration should also be given to the best bioinformatics approach to analyze different m^{5}C sites. For instance, mapping the methylation sites in tRNAs is particularly challenging because of their abundance and repetitive nature [54,55]. One possibility is to only use uniquely mapping reads, which may result in the omission of methylation sites in particularly repetitive tRNA sequences to allow for more accurate mapping. However, since tRNAs are highly repetitive sequences in the genome, this often leads to discarding a large proportion of the available data. Another strategy is to use uniquely mapping reads to simulate the distribution of reads in a given area, thereby assigning multiple mapping reads proportionally. Mapping m^{5}C sites obtained by bisulfite sequencing is hampered by the bisulfite conversion itself and RNA degradation, which may alter the base content.

AzaIP also presents some difficulties for identifying the complete set of methylation sites - a large number of targets often do not contain C to G transversions, and so the methylation sites in those targets are not mapped. Mapping methylation sites in miCLIP, on the other hand, does not depend on such artificial modifications, but rather on the stalling of the reverse transcriptase at the cross-link/methylation site during the reverse transcription step. This phenomenon was utilized to efficiently map methylation sites.

While all four methods have their individual drawbacks, together they have the power to reliably detect m^{5}C modifications. We will now turn our attention to comparing the cytosine-5 methylated RNAs that have been identified in the mammalian transcriptome using the different techniques.

A comparison of RNA m^{5}C methylation sites identified by different approaches

To better understand the ability of the different techniques to detect m^{5}C, we compared the data produced by the studies focused on the mammalian system [18,23,24]. Data used for the comparison are available in Additional file 1. We compared global cytosine-5 methylation sites identified by RNA bisulfite sequencing with NSun2-dependent RNA methylation identified by Aza-IP and our own developed method miCLIP (Figure 2) [18,23,24]. Each of the three studies identified m^{5}C sites at nucleotide resolution. RNA bisulfite sequencing enriched for the mRNA fraction, identifying about 10,500 m^{5}C sites [18]. The finding that m^{5}C was a common modification in protein coding RNAs is surprising (Figure 2A) because cytosine-5 methylation has not yet been detected in mRNAs using alternative approaches such as mass spectrometry [1]. NSun2-specific Aza-IP [23] and miCLIP [24] using total RNA determined about 600 and 1,100 sites, respectively, suggesting that less than 10 % of the methylome is directly dependent on NSun2. Most RNAs identified by both Aza-IP and miCLIP corresponded to tRNAs (Figure 2A), where approximately one-third of the tRNA methylation sites were commonly identified by both methods (Figure 2B; tRNA). In addition to tRNAs, all three studies identified m^{5}C in non-coding RNAs. While bisulfite sequencing and miCLIP also identified several protein coding RNAs, only one statistically significant protein coding RNA was identified by Aza-IP. Aza-IP also detected additional rRNAs (Figure 2A). There are ten non-tRNA m^{5}C sites found in at least two of the three studies, so these can be said to be true NSun2 non-tRNA targets with high confidence (Figure 2C).

The degree of overlap between the three studies was surprisingly low, and this may be due to several technical differences, as well as the use of different cell lines. Despite this low degree of overlap, the analyses found identical m^{5}C sites in tRNAs and non-tRNAs and confirmed NSun2-dependent methylation sites in several ncRNAs. Our comparison further identified methylated cytosines in mRNAs commonly found in at least two of the three studies. However, we noted that these mRNA methylation sites often shared a genomic location with tRNA genes (see section below). Further optimization of the approaches is clearly required to achieve a
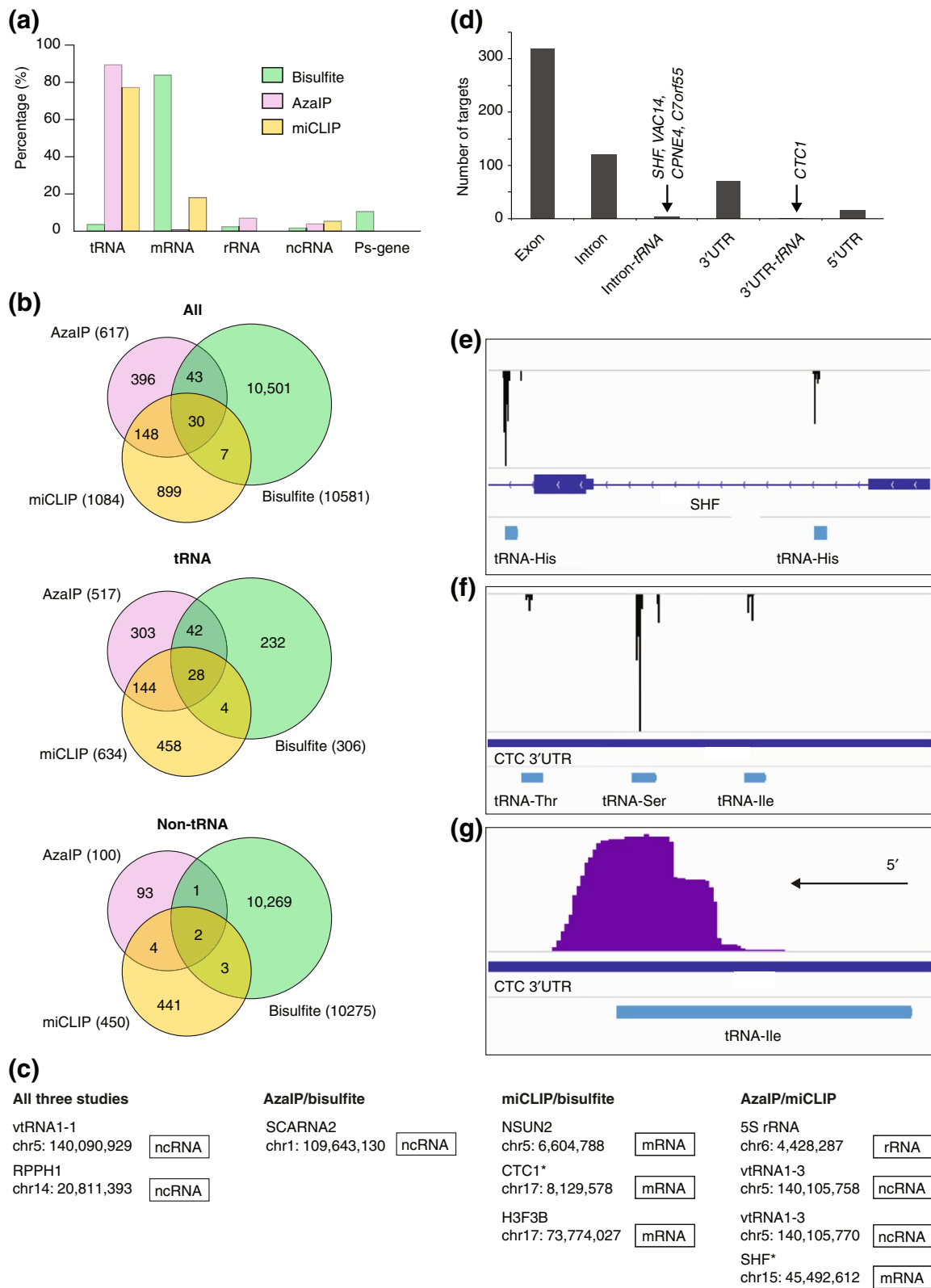
Figure 2 (See legend on next page.)

(See figure on previous page.)

Figure 2 A comparison of m⁵C target sites identified by three transcriptome-wide approaches (Additional file 1) [18,23,24]. The bisulfite study identifies the global pattern of m⁵C methylation in the genome, while the AzaIP and miCLIP data shown focus specifically on Nsun2-mediated methylation. **(a)** Types of transcripts found to be methylated by each technique. **(b)** Comparison of the genomic locations of individual methylation sites found by each of the three approaches. The Venn diagrams show the overlap between all sites (upper panel), the sites that map to tRNAs (middle panel), and non-tRNA sites (lower panel). **(c)** Non-tRNA target sites found in at least two of the three studies, which represent high-confidence sites. The genomic location (hg19) of methylated cytosines is given. * = target sites that overlap with tRNA locations. (d) The proportion of miCLIP-identified targets occurring in exons/introns/UTRs of mRNAs, including those sites overlapping with tRNA genes. **(e)** Genome browser view of the miCLIP-identified SHF mRNA methylation. Predicted methylation sites are indicated by the black peaks. Note that methylation may occur either in an SHF mRNA intron or within a tRNA, which is transcribed from the same genomic DNA strand. **(f)** miCLIP data showing identified sites within the CTC1 3′ UTR. The CTC1 3′ UTR contains three tRNA genes predicted to be transcribed from the same genomic DNA strand. The miCLIP predicted methylation sites (black peaks) overlapped specifically with these three regions. **(g)** An example showing that all miCLIP sequence reads, indicated by purple bars, only extend slightly beyond the annotated tRNA genes. It is most likely that these represent extensions only into the 3′ trailer sequences of tRNAs, and that tRNA rather than mRNA methylation is being detected.

comprehensive identification of high-confidence methylation sites.

Methylation of mRNA sites that overlap with tRNA genes

tRNA and tRNA-like sequences are sometimes found embedded within the intronic sequence of an mRNA, or more rarely in its UTR regions. The significance of this is unknown, but in addition to being present in the pre-mRNA/mRNA sequence, it is also possible that these tRNA sequences are independently transcribed into functionally mature tRNAs. Our analysis showed that m⁵C sites mapping to commonly identified mRNAs can overlap with the genomic location of tRNA genes (Figure 2D). We determined how often miCLIP-identified mRNA methylation sites overlapped with the presence of tRNA genes. Most miCLIP-identified mRNA m⁵C sites occurred within exons, fewer within introns and more rarely in 5′ and 3′ UTRs (Figure 2D). Of those, the vast majority of m⁵C sites did not coincide with the occurrence of tRNA or tRNA-like sequences, and only five miCLIP-identified mRNAs contained tRNA sequences (Figure 2D). For example, Aza-IP identified *SHF* as the only statistically significant mRNA containing an NSun2-dependent m⁵C site [23], and this site was also identified by miCLIP (Figure 2C) [24]. However, the m⁵C sites occurred in introns of the *SHF* mRNA and overlap with tRNA sequences (Figure 2E).

To address the question of whether methylation is taking place in the pre-mRNA/mRNA sequence or an independently transcribed tRNA, we focused on the m⁵C site within the CTS telomere maintenance complex component 1 (*CTC1*) mRNA as an example. The *CTC1* m⁵C sites, one of which was also detected by bisulfite sequencing (Figure 2C), localized to the 3′ UTR of the mRNA (Figure 2F). We asked whether the miCLIP sequence reads extended further into the mRNA sequence beyond the tRNA regions. If the sequence reads did not extend further into the mRNA sequence, but were confined to the tRNA sequence, it would suggest that we had

detected methylation in independently transcribed tRNAs. We found that miCLIP reads usually extended only slightly beyond the annotated tRNA genes (Figure 2G). Since the annotated tRNA genes do not include 5′ leader or 3′ trailer sequences, it is most likely that these small extensions overlap with tRNA 3′ trailer sequences, which are removed post-transcriptionally. None of the sequence reads extended beyond these tRNA-trailer sequence regions, indicating that these examples represent tRNA methylation rather than methylation of pre-mRNAs/mRNAs.

Other RNA methylases

The existence of at least six more putative m⁵C RNA methylases in mammals may contribute to the general modest overlap between the NSun2- and Dnmt2-specific cytosine-5 methylation sites and the transcriptome-wide methylome [18,23,24]. The additional enzymes NSun1 and NSun3-7 [56] are predicted to methylate RNA based on sequence conservation of key catalytic residues. Although the substrate specificities of these enzymes are unknown, NSun1 and NSun5, in addition to NSun2, have been identified as mRNA-binding proteins [57]. The biological functions of the NSun protein family is largely unknown, although all of them are expressed during mouse embryogenesis [58]. The transcripts of NSun2-7 are enriched in the developing brain, which is consistent with a proposed role in neurocognitive development [58].

The limited information we do have about the functional roles of NSun proteins indicates an essential role in a wide range of biological processes and human diseases. NSun1 (also called NOP2 or p120) is a nucleolar protein implicated in rRNA methylation and biogenesis. It has been shown to regulate the cell cycle and is involved in tumorigenesis [59-64]. Whether the enzymatic activity of NSun1 is directly required to mediate these cellular processes remains unclear. NSun4 interacts with mitochondrial transcription termination factor 4 (MTERF4) to control mitochondrial ribosomal biogenesis and translation [65,66].

The *NSUN5* gene is located in a genomic region deleted in patients with Williams-Beuren syndrome, a rare neurodevelopmental disorder [67]. Patients with Williams-Beuren syndrome display intellectual disability and developmental delay, as well as craniofacial and cardiovascular abnormalities [68]. Whether deletion of *NSUN5* directly contributes to these symptoms is unknown. Mutations in the *NSUN7* gene cause infertility in mice due to impaired sperm motility [69]. Nothing is known regarding the functions of NSun3 and NSun6 proteins.

Together, RNA bisulfite sequencing, m5C-RIP, Aza-IP and miCLIP provide powerful tools to determine methylated substrates of all NSun protein family members. This will help to provide a more detailed and informative picture of a methylome-wide m5C landscape.

Delineating the epitranscriptome: comparing m6A and m5C modifications

Although there are several types of RNA methylation [1], so far, transcriptome-wide approaches determining site-specific RNA methylation have only been performed for m6A and m5C. As already discussed, the m5C modification is most prominently associated with various ncRNAs, but recent studies suggest that it may also be a common modification of mRNAs [18,24]. By contrast, most m6A sites were found in mRNAs and only rarely occurred in ncRNAs, though they were, for instance, found in various long intergenic non-coding RNAs (lincRNAs) [4]. In mRNAs, both m6A and m5C modifications are commonly found in coding regions and UTRs rather than in introns [4,24]. Enrichment of m6A sites is most pronounced near stop codons [3,4], whereas m5C is not enriched at stop codons. Both m6A and m5C can occur in 3′ UTRs [4,18], suggesting a potential role in regulating microRNA responses. It is intriguing to speculate that m6A and m5C modifications may cooperate in regulating RNA function and this should be a focus for future studies.

A key difference between the m6A and m5C transcriptome-wide studies is that m5C sites in RNA were experimentally determined at the nucleotide resolution, whereas the precise locations of m6A sites have to be computationally predicted [3,4]. The chemistry of bisulfite sequencing makes it unsuitable for the detection of m6A and the lack of understanding of how the m6A modification is catalyzed makes experimental determination at the nucleotide resolution difficult. Identification of m5C sites at nucleotide resolution relies on a covalently bound enzyme-RNA catalytic intermediate, but it is not known whether m6A modification also occurs via the formation of a temporary covalent complex. Thus, the methods available to study m6A and m5C RNA modifications differ technically. The two studies analyzing m6A in the transcriptome used almost identical

approaches, resulting in high similarity between two datasets [3,4]. The development of novel techniques to identify RNA m6A sites may be beneficial to identify additional targets not yet detected by the RIP-seq approach. However, it is clear that the currently available techniques detecting m6A and m5C have vastly increased our knowledge regarding RNA methylation and have provided a solid platform for functional studies of these modifications.

Regulatory and biological functions of m6A and m5C modifications

Little is known about m6A RNA methyltransferases, except that they consist of several subunits and include the methyltransferase like 3 (METTL3) enzymatic component [70]. No other enzyme has been identified that catalyzes RNA m6A modification in humans. A recent study showed that METTL3-dependent m6A RNA methylation in circadian transcripts is crucial for regulating their own negative transcription-translation feedback loop [71]. Thus, the authors demonstrated a novel mechanism whereby RNA methylation regulates the circadian period and the function of the biological clock.

Enzymes associated with regulating both m6A and m5C modifications have also been associated with human diseases. Genetic polymorphisms causing upregulation of the m6A-demethylase fat mass and obesity-associated (*FTO*) gene are associated with increased body mass index and an increased risk of obesity [72]. In addition, *Fto* knockout mice exhibit impaired synaptic transmission in dopaminergic neurons [73]. Interestingly, m6A RIP-seq analyses on the mid-brain and striatum of *Fto* knockout mice revealed increased levels of m6A modification in a subset of mRNAs involved in dopamine signaling pathways. Given the intellectual disability and epilepsy phenotypes observed in NSun2-deficient patients, it is intriguing to speculate that the interplay of different RNA modifications may regulate neuronal signaling pathways. For example, one possibility is that mRNA methylation may regulate the translation rates of proteins that have been associated with synaptic plasticity. This hypothesis may, at least in part, explain the memory and learning deficits associated with the loss of NSun2 [37]. In addition, spatiotemporal control of protein translation at the growth cones of terminally differentiated neurons during neuronal circuit formation is known to be an important regulator of axonal pathfinding and synapse formation. It may also be worth considering whether mRNA methylation plays a regulatory role in such events.

A regulatory role for NSun2 in protein translation is supported by the fact that we showed no correlation of potentially methylated mRNAs to their turnover or stability [24]. In addition, in testicular round spermatids,

NSun2 is a component of the chromatoid body, which is known to be a cytoplasmic subdomain for localized protein translation control [36]. Simultaneous deletion of NSun2 and Dnmt2 impair protein synthesis and tRNA stability [22]. The systematic co-ordination of protein translation steps occur during differentiation processes of embryonic stem cells [74], and aberrant translational control may contribute to the neurodevelopmental disease phenotypes in humans lacking functional NSun2 [19,38,39,75]. In addition, mis-regulation of protein translation has been associated with other intellectual disability associated genes, such as fragile-X mental retardation 1 (FMR1) [76].

The most consistent NSun2-dependent m^5C site identified in a non-tRNA is C69 in the vault RNA, vtRNA1.1 [24]. Using NSun2-deficient patient-derived fibroblasts, we showed that methylation of vtRNA1.1 influenced its processing into microRNA-like molecules, which regulated the calcium channel, voltage-dependent, gamma subunit (CACNG)7 and CACNG8 mRNAs [24]. Both CACNG7 and CACNG8 encode voltage-gated calcium channels, and genes encoding ion channels are often mutated in disorders associated with epilepsy and intellectual disability [77]. We previously observed that the NSun2 protein is expressed in Purkinje neurons of the cerebellum as well as other neurons throughout the mouse brain [38]. Mis-regulation of synaptic transmission in neurons of NSun2-deficient patients may contribute to the human disease.

Conclusions

Despite all the technical advances, further optimization of current methods is required to achieve much closer agreement of the site-specific deposition of m^5C, especially in the non-tRNA targets identified by each method. Only then will we have a comprehensive characterization of the global distribution of m^5C in RNA that will help to further define the functional roles of this modification.

The recently adapted single-molecule, real-time (SMRT) sequencing technologies for the detection of RNA sequences, together with the successful engineering of novel reverse transcriptases more sensitive to the different RNA modifications occurring in RNAs, may enable the identification of m^5C transcriptome-wide without the need of any RNA processing step, not even cDNA conversion, which is usually biased and causes the loss of information about structure and base modifications [78,79].

In summary, the current transcriptome-wide approaches for determining m^5C methylation sites are furthering our understanding of this RNA modification in terms of regulatory functions and disease. Indeed, it may not be long before our expectations from understanding

RNA m^5C modification matches, or even exceeds, that of DNA m^5C modification.

Additional file

Additional file 1: Nucleotide-resolution positions of m5C sites detected by AzaIP, miCLIP and bisulfite sequencing.

References
1. Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, Helm M, Bujnicki JM, Grosjean H: MODOMICS: a database of RNA modification pathways - 2013 update. *Nucleic Acids Res* 2013, **41**:D262–D267.
2. Bodi Z, Button JD, Grierson D, Fray RG: Yeast targets for mRNA methylation. *Nucleic Acids Res* 2010, **38**:5327–5335.
3. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, Sorek R, Rechavi G: Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 2012, **485**:201–206.
4. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR: Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 2012, **149**:1635–1646.
5. Suzuki MM, Bird A: DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 2008, **9**:465–476.
6. Motorin Y, Lyko F, Helm M: 5-Methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Res* 2010, **38**:1415–1430.
7. Maden BE: Locations of methyl groups in 28 S rRNA of Xenopus laevis and man. Clustering in the conserved core of molecule. *J Mol Biol* 1988, **201**:289–314.
8. Noon KR, Bruenger E, McCloskey JA: Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile Sulfolobus solfataricus. *J Bacteriol* 1998, **180**:2883–2888.
9. Smith JE, Cooperman BS, Mitchell P: Methylation sites in Escherichia coli ribosomal RNA: localization and identification of four new sites of methylation in 23S rRNA. *Biochemistry* 1992, **31**:10825–10834.
10. Veldman GM, Klootwijk J, de Regt VC, Planta RJ, Branlant C, Krol A, Ebel JP: The primary and secondary structure of yeast 26S rRNA. *Nucleic Acids Res* 1981, **9**:6935–6952.
11. Dubin DT, Stollar V, Hsuchen CC, Timko K, Guild GM: Sindbis virus messenger RNA: the 5'-termini and methylated residues of 26 and 42 S RNA. *Virology* 1977, **77**:457–470.
12. Dubin DT, Taylor RH: The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic Acids Res* 1975, **2**:1653–1668.
13. Adams JM, Cory S: Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA. *Nature* 1975, **255**:28–33.
14. Desrosiers R, Friderici K, Rottman F: Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc Natl Acad Sci U S A* 1974, **71**:3971–3975.
15. Furuichi Y, Morgan M, Shatkin AJ, Jelinek W, Salditt-Georgieff M, Darnell JE: Methylated, blocked 5 termini in HeLa cell mRNA. *Proc Natl Acad Sci U S A* 1975, **72**:1904–1908.

16. Lavi S, Shatkin AJ: **Methylated simian virus 40-specific RNA from nuclei and cytoplasm of infected BSC-1 cells.** *Proc Natl Acad Sci U S A* 1975, **72**:2012–2016.

17. Salditt-Georgieff M, Jelinek W, Darnell JE, Furuichi Y, Morgan M, Shatkin A: **Methyl labeling of HeLa cell hnRNA: a comparison with mRNA.** *Cell* 1976, **7**:227–237.

18. Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, Parker BJ, Suter CM, Preiss T: **Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA.** *Nucleic Acids Res* 2012, **40**:5023–5033.

19. Blanco S, Kurowski A, Nichols J, Watt FM, Benitah SA, Frye M: **The RNA-methyltransferase Misu (NSun2) poises epidermal stem cells to differentiate.** *PLoS Genet* 2011, **7**:e1002403.

20. Brzezicha B, Schmidt M, Makalowska I, Jarmolowski A, Pienkowska J, Szweykowska-Kulinska Z: **Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the pre-tRNA Leu (CAA).** *Nucleic Acids Res* 2006, **34**:6034–6043.

21. Frye M, Watt FM: **The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors.** *Curr Biol* 2006, **16**:971–981.

22. Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, Frye M, Helm M, Stoecklin G, Lyko F: **RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis.** *Nat Struct Mol Biol* 2012, **19**:900–905.

23. Khoddami V, Cairns BR: **Identification of direct targets and modified bases of RNA cytosine methyltransferases.** *Nat Biotechnol* 2013, **31**:458–584.

24. Hussain S, Sajini AA, Blanco S, Dietmann S, Lombard P, Sugimoto Y, Paramor M, Gleeson JG, Odom DT, Ule J, Frye M: **NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs.** *Cell Rep* 2013, **4**:255–261.

25. Chen Y, Sierzputowska-Gracz H, Guenther R, Everett K, Agris PF: **5-Methylcytidine is required for cooperative binding of Mg2+ and a conformational transition at the anticodon stem-loop of yeast phenylalanine tRNA.** *Biochemistry* 1993, **32**:10249–10253.

26. Motorin Y, Helm M: **tRNA stabilization by modified nucleotides.** *Biochemistry* 2010, **49**:4934–4944.

27. Alexandrov A, Chernyakov I, Gu W, Hiley SL, Hughes TR, Grayhack EJ, Phizicky EM: **Rapid tRNA decay can result from lack of nonessential modifications.** *Mol Cell* 2006, **21**:87–96.

28. Chernyakov I, Whipple JM, Kotelawala L, Grayhack EJ, Phizicky EM: **Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1.** *Genes Dev* 2008, **22**:1369–1380.

29. Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M, Lyko F: **RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage.** *Genes Dev* 2010, **24**:1590–1595.

30. Chow CS, Lamichhane TN, Mahto SK: **Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications.** *ACS Chem Biol* 2007, **2**:610–619.

31. Warren L, Manos PD, Ahfeldt T, Loh YH, Li H, Lau F, Ebina W, Mandal PK, Smith ZD, Meissner A, Daley GQ, Brack AS, Collins JJ, Cowan C, Schlaeger TM, Rossi DJ: **Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA.** *Cell Stem Cell* 2010, **7**:618–630.

32. Zhang X, Liu Z, Yi J, Tang H, Xing J, Yu M, Tong T, Shang Y, Gorospe M, Wang W: **The tRNA methyltransferase NSun2 stabilizes p16INK(4) mRNA by methylating the 3'-untranslated region of p16.** *Nat Commun* 2012, **3**:712.

33. Chan CT, Pang YL, Deng W, Babu IR, Dyavaiah M, Begley TJ, Dedon PC: **Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins.** *Nat Commun* 2012, **3**:937.

34. Rai K, Chidester S, Zavala CV, Manos EJ, James SR, Karpf AR, Jones DA, Cairns BR: **Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish.** *Genes Dev* 2007, **21**:261–266.

35. Wei CM, Gershowitz A, Moss B: **Methylated nucleotides block 5' terminus of HeLa cell messenger RNA.** *Cell* 1975, **4**:379–386.

36. Hussain S, Tuorto F, Menon S, Blanco S, Cox C, Flores JV, Watt S, Kudo NR, Lyko F, Frye M: **The mouse cytosine-5 RNA methyltransferase NSun2 is a component of the chromatoid body and required for testis differentiation.** *Mol Cell Biol* 2013, **33**:1561–1570.

37. Abbasi-Moheb L, Mertel S, Gonsior M, Nouri-Vahid L, Kahrizi K, Cirak S, Wieczorek D, Motazacker MM, Esmaeeli-Nieh S, Cremer K, Weißmann R, Tzschach A, Garshasbi M, Abedini SS, Najmabadi H, Ropers HH, Sigrist SJ, Kuss AW: **Mutations in NSUN2 cause autosomal-recessive intellectual disability.** *Am J Hum Genet* 2012, **90**:847–855.

38. Khan MA, Rafiq MA, Noor A, Hussain S, Flores JV, Rupp V, Vincent AK, Malli R, Ali G, Khan FS, Ishak GE, Doherty D, Weksberg R, Ayub M, Windpassinger C, Ibrahim S, Frye M, Ansar M, Vincent JB: **Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability.** *Am J Hum Genet* 2012, **90**:856–863.

39. Martinez FJ, Lee JH, Lee JE, Blanco S, Nickerson E, Gabriel S, Frye M, Al-Gazali L, Gleeson JG: **Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome.** *J Med Genet* 2012, **49**:380–385.

40. Fahiminiya S, Almuriekhi M, Nawaz Z, Staffa A, Lepage P, Ali R, Hashim L, Schwartzentruber J, Abu Khadija K, Zaineddin S, Gamal H, Majewski J, Ben-Omran T: **Whole exome sequencing unravels disease-causing genes in consanguineous families in Qatar.** *Clin Genet* 2013, : . doi: 10.1111/cge.12280.

41. Hussain S, Benavente SB, Nascimento E, Dragoni I, Kurowski A, Gillich A, Humphreys P, Frye M: **The nucleolar RNA methyltransferase Misu (NSun2) is required for mitotic spindle stability.** *J Cell Biol* 2009, **186**:27–40.

42. Sakita-Suto S, Kanda A, Suzuki F, Sato S, Takata T, Tatsuka M: **Aurora-B regulates RNA methyltransferase NSUN2.** *Mol Biol Cell* 2007, **18**:1107–1117.

43. Frye M, Dragoni I, Chin SF, Spiteri I, Kurowski A, Provenzano E, Green A, Ellis IO, Grimmer D, Teschendorff A, Zouboulis CC, Caldas C, Watt FM: **Genomic gain of 5p15 leads to over-expression of Misu (NSUN2) in breast cancer.** *Cancer Lett* 2009, **289**:71–80.

44. Schaefer M, Pollex T, Hanna K, Lyko F: **RNA cytosine methylation analysis by bisulfite sequencing.** *Nucleic Acids Res* 2009, **37**:e12.

45. Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R: **Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m(5)C within archaeal mRNAs.** *PLoS Genet* 2013, **9**:e1003602.

46. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL: **A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.** *Proc Natl Acad Sci U S A* 1992, **89**:1827–1831.

47. Wu JC, Santi DV: **Kinetic and catalytic mechanism of HhaI methyltransferase.** *J Biol Chem* 1987, **262**:4778–4786.

48. Liu Y, Santi DV: **m5C RNA and m5C DNA methyl transferases use different cysteine residues as catalysts.** *Proc Natl Acad Sci U S A* 2000, **97**:8263–8265.

49. Santi DV, Garrett CE, Barr PJ: **On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs.** *Cell* 1983, **33**:9–10.

50. King MY, Redman KL: **RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine.** *Biochemistry* 2002, **41**:11218–11225.

51. Redman KL: **Assembly of protein-RNA complexes using natural RNA and mutant forms of an RNA cytosine methyltransferase.** *Biomacromolecules* 2006, **7**:3321–3326.

52. König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J: **iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution.** *Nat Struct Mol Biol* 2010, **17**:909–915.

53. Behm-Ansmant I, Helm M, Motorin Y: **Use of specific chemical reagents for detection of modified nucleotides in RNA.** *J Nucleic Acids* 2011, **2011**:408053.

54. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25**:955–964.

55. Chan PP, Lowe TM: **GtRNAdb: a database of transfer RNA genes detected in genomic sequence.** *Nucleic Acids Res* 2009, **37**:D93–D97.

56. Reid R, Greene PJ, Santi DV: **Exposition of a family of RNA m(5)C methyltransferases from searching genomic and proteomic sequences.** *Nucleic Acids Res* 1999, **27**:3138–3045.

57. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW: **Insights into RNA biology from an atlas of mammalian mRNA-binding proteins.** *Cell* 2012, **149**:1393–1406.

58. Chi L, Delgado-Olguin P: **Expression of NOL1/NOP2/sun domain (Nsun) RNA methyltransferase family genes in early mouse embryogenesis.** *Gene Expr Patterns* 2013, **13**:319–327.

59. de Beus E, Brockenbrough JS, Hong B, Aris JP: **Yeast NOP2 encodes an essential nucleolar protein with homology to a human proliferation marker.** *J Cell Biol* 1994, **127**:1799–1813.
60. Hong B, Brockenbrough JS, Wu P, Aris JP: **Nop2p is required for pre-rRNA processing and 60S ribosome subunit synthesis in yeast.** *Mol Cell Biol* 1997, **17**:378–388.
61. Gustafson WC, Taylor CW, Valdez BC, Henning D, Phippard A, Ren Y, Busch H, Durban E: **Nucleolar protein p120 contains an arginine-rich domain that binds to ribosomal RNA.** *Biochem J* 1998, **331**:387–393.
62. Freeman JW, Busch RK, Gyorkey F, Gyorkey P, Ross BE, Busch H: **Identification and characterization of a human proliferation-associated nucleolar antigen with a molecular weight of 120,000 expressed in early G1 phase.** *Cancer Res* 1988, **48**:1244–1251.
63. Ochs RL, Reilly MT, Freeman JW, Busch H: **Intranucleolar localization of human proliferating cell nucleolar antigen p120.** *Cancer Res* 1988, **48**:6523–6529.
64. Perlaky L, Valdez BC, Busch RK, Larson RG, Jhiang SM, Zhang WW, Brattain M, Busch H: **Increased growth of NIH/3 T3 cells by transfection with human p120 complementary DNA and inhibition by a p120 antisense construct.** *Cancer Res* 1992, **52**:428–436.
65. Cámara Y, Asin-Cayuela J, Park CB, Metodiev MD, Shi Y, Ruzzenente B, Kukat C, Habermann B, Wibom R, Hultenby K, Franz T, Erdjument-Bromage H, Tempst P, Hallberg BM, Gustafsson CM, Larsson NG: **MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome.** *Cell Metab* 2011, **13**:527–539.
66. Spåhr H, Habermann B, Gustafsson CM, Larsson NG, Hallberg BM: **Structure of the human MTERF4-NSUN4 protein complex that regulates mitochondrial ribosome biogenesis.** *Proc Natl Acad Sci U S A* 2012, **109**:15253–15258.
67. Doll A, Grzeschik KH: **Characterization of two novel genes, WBSCR20 and WBSCR22, deleted in Williams-Beuren syndrome.** *Cytogenet Cell Genet* 2001, **95**:20–27.
68. Martens MA, Wilson SJ, Reutens DC: **Williams syndrome: a critical review of the cognitive, behavioral and neuroanatomical phenotype.** *J Child Psychol Psychiatry* 2008, **49**:576–608.
69. Harris T, Marquez B, Suarez S, Schimenti J: **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.** *Biol Reprod* 2007, **77**:376–382.
70. Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM: **Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase.** *RNA* 1997, **3**:1233–1247.
71. Fustin JM, Doi M, Yamaguchi Y, Hida H, Nishimura S, Yoshida M, Isagawa T, Morioka MS, Kakeya H, Manabe I, Okamura H: **RNA-methylation-dependent RNA processing controls the speed of the circadian clock.** *Cell* 2013, **155**:793–806.
72. Church C, Moir L, McMurray F, Girard C, Banks GT, Teboul L, Wells S, Brüning JC, Nolan PM, Ashcroft FM, Cox RD: **Overexpression of Fto leads to increased food intake and results in obesity.** *Nat Genet* 2010, **42**:1086–1092.
73. Hess ME, Hess S, Meyer KD, Verhagen LA, Koch L, Brönneke HS, Dietrich MO, Jordan SD, Saletore Y, Elemento O, Belgardt BF, Franz T, Horvath TL, Rüther U, Jaffrey SR, Kloppenburg P, Brüning JC: **The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry.** *Nat Neurosci* 2013, **16**:1042–1048.
74. Ingolia NT, Lareau LF, Weissman JS: **Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes.** *Cell* 2011, **147**:789–802.
75. Kuss AW, Garshasbi M, Kahrizi K, Tzschach A, Behjati F, Darvish H, Abbasi-Moheb L, Puettmann L, Zecha A, Weissmann R, Hu H, Mohseni M, Abedini SS, Rajab A, Hertzberg C, Wieczorek D, Ullmann R, Ghasemi-Firouzabadi S, Banihashemi S, Arzhangi S, Hadavi V, Bahrami-Monajemi G, Kasiri M, Falah M, Nikuei P, Dehghan A, Sobhani M, Jamali P, Ropers HH, Najmabadi H: **Autosomal recessive mental retardation:homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots.** *Hum Genet* 2011, **129**:141–148.
76. Darnell JC, Van Driesche SJ, Zhang C, Hung KY, Mele A, Fraser CE, Stone EF, Chen C, Fak JJ, Chi SW, Licatalosi DD, Richter JD, Darnell RB: **FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism.** *Cell* 2011, **146**:247–261.
77. Poduri A, Lowenstein D: **Epilepsy genetics - past, present, and future.** *Curr Opin Genet Dev* 2011, **21**:325–332.
78. Roberts RJ, Carneiro MO, Schatz MC: **The advantages of SMRT sequencing.** *Genome Biol* 2013, **14**:405.
79. Vilfan ID, Tsai YC, Clark TA, Wegener J, Dai Q, Yi C, Pan T, Turner SW, Korlach J: **Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription.** *J Nanobiotechnol* 2013, **11**:8.