



Published in final edited form as:

Genomics. 2020 March ; 112(2): 1840–1846. doi:10.1016/j.ygeno.2019.10.017.

Advances in Methods and Software for RNA Cytosine Methylation Analysis

Xiguang Xu^{1,2}, Xiaoran Wei^{1,3}, Hehuang Xie^{1,2,3,*}

¹Fralin Life Sciences Institute at Virginia Tech, Blacksburg, VA 24061, USA

²Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA

³Department of Biomedical Sciences and Pathobiology, Virginia-Maryland College of Veterinary Medicine, Blacksburg, VA 24061, USA

Abstract

Our understanding of RNA modifications has been growing rapidly over the last decade. Epitranscriptomics has recently emerged as an exciting, new field for understanding the fundamental mechanisms underlying RNA modifications and their impact on gene expression. Among the over one hundred different kinds of RNA modifications, cytosine methylation in mRNA (5-mrC) is now recognized as an important epigenetic mark that modulates mRNA transportation, translation, and stability at the post-transcriptional level. Across plant and animal species, recent studies have revealed the roles of mRNA cytosine methylation in several fundamental biological processes. In mammals, genome-wide profiling has determined thousands of mRNA transcripts carrying the 5-mrC modification in a tissue specific manner. Here, we summarize the experimental techniques that were exploited to determine 5-mrC in mRNA and the computational procedures implemented for RNA bisulfite sequencing data analysis.

Keywords

RNA cytosine methylation; post-transcriptional regulation; RNA bisulfite sequencing; methylation data analysis

Background

“RNA epigenetics” or “epitranscriptomics” is an emerging new field in the study of RNA post-transcriptional modification [1–3]. Currently, around 170 distinct types of RNA modifications, including N6-methyladenosine, N1-methyladenosine, 5-methylcytosine, and

*Corresponding author: davidxie@vt.edu.

AUTHOR CONTRIBUTIONS

Conceptualization, X.X. and H.X.; original draft preparation and editing, X.W., X.X. and H.X.; funding acquisition, H.X.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

COMPETING INTERESTS

The authors declare no competing interests.

5-hydroxymethylcytosine, have been identified [4]. The N6-methyladenosine modification in poly(A) RNA has been extensively studied and was found to regulate messenger RNA (mRNA) splicing, stability, and translation efficiency in diverse biological processes [5–7]. RNA cytosine methylation (5-mrC) is another important form of RNA modification. In the 1960's, studies identified the presence of 5-mrC in ribosomal RNA [8]. Later studies showed that 5-mrC was not only found in rRNA and tRNA but was also found in mRNA and non-coding RNA from all three domains of life: Archaea, Bacteria, and Eukarya [9–14].

In recent years, several pivotal findings have been reported regarding the writers, erasers, and readers of 5-mrC in RNA. In mammalian cells, the addition of a methyl group on the fifth carbon of cytosine in RNA is catalyzed by a large protein family called the NOP2/Sun domain RNA methyltransferases (NSUN) and by DNA methyltransferase 2 [12, 15, 16]. Both Yanget. al and Huang et. al identified that NSUN2 is the major RNA methyltransferase mediating the formation of 5-mrC in mRNAs [14, 17]. Previous studies showed that the ten-eleven translocation (TET) family of Fe(II)- and 2-oxoglutarate-dependent dioxygenases function as DNA demethylases *via* sequential oxidation of 5-methylcytidine to yield 5-hydroxymethylcytidine, 5-formylcytidine, 5-carboxylcytidine, and eventually unmethylated cytosines [18–21]. Interestingly, 5-mrC in RNA can be oxidized by TET enzymes (TET1, TET2, TET3) to 5-hydroxymethylcytosine [22], and then further oxidized to 5-formylcytosine [23] and 5-carboxylcytosine [24]. The molecular mechanism that mediates the conversion of 5-carboxylcytosine to unmethylated Cs in RNA remains elusive. Very little information has been gained about 5-mrC reader proteins. Aly/REF Export Factor (ALYREF) was recently identified as a 5-mrC specific binding protein that mediates target mRNAs export from the nucleus to the cytoplasm [14], indicating the critical role of 5-mrC reader protein in RNA metabolism.

Advances in next generation sequencing (NGS) accelerated the development of high-throughput 5-mrC detection methods, which provided a comprehensive view of 5-mrC distribution across the transcriptome. Transcriptome-wide distribution of 5-mrC has been revealed in poly(A) RNAs from a broad range of mammalian cell lines and tissues [13, 14, 17, 25]. Almost all recent transcriptome-wide studies showed that methylated cytosines are preferentially enriched around the translation initiation sites (TIS) of mRNAs [13, 14, 17], indicating an important regulatory role of RNA cytosine methylation on the translation of mRNAs. Moreover, cytosine methylation in mRNA regulates systemic mRNA mobility and promotes mRNA nuclear export [14]. In *Arabidopsis thaliana*, 5-mrC are significantly enriched in graft-mobile mRNAs that can be transported over graft junctions to distinct plant parts [26]. Together with RNA-binding proteins, methylated RNA and RNA methyltransferases gain the ability to mediate the interactions between transcription factors and genomic DNA to participate in chromatin organization [27]. Despite these recent findings, the functional roles of 5-mrC in mRNA during biological processes and their relevance to human disease are just beginning to be understood.

In this review, we focus on the experimental techniques and corresponding data analysis of mRNA cytosine methylation. We summarize current available approaches for detecting RNA cytosine-5 methylation at the global, transcriptome-wide and locus-specific levels. Additionally, we emphasize the bioinformatics data analysis of RNA bisulfite sequencing

datasets by comparing key features of three published packages, meRanTK, BS-RNA, and BisRNA [28–30], and discuss the major issues in the analysis of RNA bisulfite sequencing data.

Techniques for the detection of RNA Cytosine-5 methylation

Methylation at position 5 of cytosine in mRNA was discovered over 40 years ago [9, 10]. Most of the early studies on 5-mrC relied on radial labelling and paper chromatography [10]. Due to the lack of reliable and sensitive techniques for 5-mrC detection, the distribution and functional roles of 5-mrC in low abundance mRNA has remained largely unknown over the past four decades. Recent advances in NGS techniques have enabled a transcriptome-wide view of 5-mrC distribution in diverse biological processes, broadening our understanding of the functional roles of RNA cytosine methylation. Below, we summarized currently available approaches for detecting 5-mrC at the global, transcriptome-wide, and locus-specific levels (Table 1). The advantages and limitations of these techniques, including future directions, are discussed.

Global assessment of the 5-mrC level

The global level of 5-mrC modification in mRNA refers to the sum of all 5-mrC that can be identified in all mRNA transcripts from a given cell or tissue sample. Since tRNA and rRNA molecules are rich in 5-mrC modifications, one key step of the global approach for detecting 5-mrC in mRNA is to remove undesired RNA species. RNA dot blot and mass spectrometry are frequently used global approaches. Dot blot is a traditional technique that has been widely used to measure the level of protein expression [31]. This technique was later applied to detect base modifications, such as 5-mC, in DNA [19] and RNA [32]. RNA dot blot for 5-mrC utilizes the anti-5-mrC antibody to measure the levels of 5-mrC in RNA samples. The signal density captured represents the relative 5-mrC level. RNA dot blot results are regarded as qualitative or semi-quantitative data. Despite the straightforward signal provided, the RNA dot blot may not be able to detect slight changes in RNA methylation. Anti-5-mrC antibody has also been explored in Enzyme-Linked Immunosorbent Assay (ELISA)-based approaches [33, 34]. The standard curve, generated with controls at different methylation levels, allows the ELISA-based kit to accurately quantitate the global level of 5-mrC in RNA. Like dot blot, an ELISA-based kit accepts a wide range of input RNA samples from vertebrate, plant, and microbial sources.

Liquid chromatography coupled with tandem mass spectrometry (LC-MS) is an accurate, quantitative approach to assess the 5-mrC level globally [14, 22, 23]. Prior to the analysis, a critical step that should be taken is to completely digest the input RNA molecules into individual ribonucleotides. With a 5-mrC standard as a positive reference, LC-MS separates individual ribonucleotides to obtain the absolute 5-mrC level in a given RNA sample. RNA dot blot, ELISA and mass spectrometry can provide the global methylation level but not locus-specific methylation information. In other words, even if no change in 5-mrC level can be detected with these global approaches, some mRNA transcripts could have different levels of methylation modification at specific cytosines.

Transcriptome-wide approaches to generate 5-mrC profiles

A transcriptome-wide view of the 5-mrC profile may be achieved *via* antibody-based or bisulfite conversion-based approaches coupled with high-throughput sequencing. RNA immunoprecipitation of 5-mrC, followed by deep-sequencing (5-mrC-RIP-seq) utilizes 5-mrC-specific antibodies to enrich 5-mrC-modified RNAs [11, 35]. The use of antibodies enables the enrichment of mRNA transcripts with low 5-mrC levels, which may go undetected in a large pool of unmethylated RNA molecules. In addition, 5-mrC-RIP-seq allows distinction of RNA having the 5-mrC modification from RNA having other methylation modifications such as 5-hmrC. Not surprisingly, the specificity of such an approach is highly dependent on the antibody used. Non-specific bound RNA may be introduced in the immunoprecipitation process as well. The sequence reads generated for RNA pulled down by anti-5-mrC antibodies are usually 100-150 nt in length. Thus, the resolution of 5-mrC-RIP-seq for methylation detection is not at the single-nucleotide level.

5-azacytidine-mediated RNA immunoprecipitation (Aza-IP) utilizes 5-azacytidine, a cytidine analog that traps its target RNA methyltransferase by forming a stable RNA methyltransferase-RNA adduct. Covalently bound enzyme-RNA complexes may be immunoprecipitated with either tag- or enzyme-specific antibodies. The target RNA with 5-Aza-C is eventually read as a guanine during reverse transcription and sequencing [36]. The most significant advantage is that this technique allows for identification of enzyme-specific cytosine substrates at single-nucleotide resolution. Due to stable covalent binding between the RNA methyltransferase and the 5-azacytidine, the enzyme-RNA substrate complexes can be immunoprecipitated with highly stringent washes, thus largely reducing the non-specific binding of unmethylated RNA. However, efficient enrichment of the enzyme-RNA complexes depends highly on the specific antibodies against the target enzymes or the expression of epitope-tagged enzymes in the target cells. The incorporation efficiency of the cytidine analog 5-Aza is also a concern. The methylation targets in nascent RNA molecules sites but without 5-Aza incorporation will be missed. Furthermore, genomic DNA in somatic tissues are heavily methylated and 5-Aza may incorporate into DNA molecules, particularly in proliferating cells. Such altered DNA methylation profiles may lead to differential gene expression and, thus, may influence the transcription profile.

Methylation-individual nucleotide resolution crosslinking and immunoprecipitation (miCLIP) is a customized technique derived from the individual-nucleotide-resolution crosslinking and immunoprecipitation (iCLIP) method, which allows the detection of RNA methyltransferase-specific substrate sites at nucleotide resolution [37]. This technique has been used to identify NSUN2 and NSUN3 substrates [38, 39]. The point mutation of the conserved cysteine that is needed within the catalytic domain of RNA methyltransferases for the release of methylated RNA from the enzyme results in the irreversible formation of covalent RNA-enzyme complex at the methylation sites. Covalent crosslinking of the RNA-protein complex leaves a short peptide at the target 5-mrC site, which stalls the reverse transcription during library construction. As a result, all sequences end at the methylation site [38, 39]. Despite its robustness and high specificity, miCLIP requires the generation of mutant enzymes, which is expensive and time-consuming.

Bisulfite sequencing was originally developed to detect the 5-mC sites in genomic DNA [40]. In the presence of sodium bisulfite, unmethylated cytosines are converted to uracils, which are later replaced by thymines during subsequent PCR amplification, while methylated cytosines remain unchanged. In recent years, bisulfite sequencing has been modified to identify the 5-mrC profile in RNAs on a transcriptome-wide scale [41]. After the initial development of RNA bisulfite sequencing, this technique has been commercialized and various RNA bisulfite conversion kits are available, including the EZ RNA Methylation Kit from ZymoResearch and Methylamp RNA Bisulfite Conversion Kit from Epigentek [42]. The primary advantage of this technique is that it can provide a transcriptome-wide view of 5-mrC deposition at single-nucleotide resolution. However, bisulfite sequencing has the limitation that it cannot differentiate 5-methylcytosine from 5-hydroxymethylcytosine, as both are resistant to deamination, but the level of 5-hmrC is very low in human and mouse mRNAs [23, 43]. The ratio between 5-hmrC:5-mrC is estimated to be around 1:5,000 [22], making RNA bisulfite sequencing an attractive approach to generate the 5-mrC profile. Bisulfite treatment results in significant degradation of RNA, making it difficult to detect 5-mrC in low expressed mRNA molecules [44]. To protect RNA integrity, RNA bisulfite conversion is usually performed at a relatively low temperature compared to DNA bisulfite conversion. Bisulfite conversion can also be encumbered by the secondary structures of RNAs, such as double-strand RNA (dsRNA) and stem-loop structures. Thus, incomplete denaturation of RNA secondary structure may introduce cytosines resistant to bisulfite conversion, which end as false positive signals. Despite these disadvantages, RNA bisulfite sequencing has been increasingly applied to study RNA cytosine methylation in recent years [13, 14, 17, 25, 30, 45].

Locus-specific approaches to determine methylation within a given mRNA

Locus-specific approaches have been developed to measure the methylation level of specific 5-mrC sites in mRNA. The most common approach is to use 5-mrC RIP, followed by RT-qPCR [13, 35]. In this procedure, RNA molecules are fragmented and pulled down by the 5-mrC antibody and then reverse transcribed to cDNA. Real-time qPCR is then performed to measure the relative fold changes for specific transcripts. With appropriate controls, such as normal IgG control, this approach has been used to validate the 5-mrC sites identified by RNA bisulfite sequencing [13]. RNA bisulfite conversion combined with either cloning-based [41, 45] or PCR amplicon-based [11, 14] Sanger sequencing are another two locus-specific methylation assays commonly used in the validation of 5-mrC sites. The cDNA template derived from bisulfite-converted RNA was used for cloning into vectors or PCR amplification with primers fused with consensus sequences, and then subject to Sanger sequencing. RNA bisulfite pyrosequencing may be developed as an alternative approach to determine the 5-mrC levels for multiple cytosines in a short stretch of RNA molecule. Similar to the pyrosequencing of bisulfite-converted DNA [46], RNA molecules may be subjected to bisulfite conversion first prior to cDNA generation. After reverse transcription, cDNA molecules are used as templates for PCR and pyrosequencing.

Since each technique for RNA methylation detection has its own features, the combination of these approaches may provide more comprehensive understanding on multiple levels. For example, RNA dot blot and mass spectrometry can be used as the initial steps to explore the

changes of 5-mrC at the global level in a specific biological process [35]. Aza-IP and miCLIP can be used to study the substrates of a specific RNA methyltransferase. As the sequencing cost continues to decrease, RNA bisulfite sequencing becomes even more attractive for gaining a transcriptome-wide view of RNA methylation at single nucleotide resolution. Although 5-mrC-RIP cannot provide methylation information at single nucleotide resolution, it may serve as an alternative approach to validate bisulfite sequencing results and to eliminate the false-positive 5-mrC sites resulting from an incomplete bisulfite conversion.

Data analysis for RNA cytosine-5 methylation studies

The methods used for 5-mrC data analysis depend on the types of data results obtained with different 5-mrC detection approaches. For techniques used to detect 5-mrC at the global level (i.e., ELISA) or at the locus-specific level (i.e., RNA bisulfite pyrosequencing), each measurement provides a numerical number. A typical experiment often includes multiple biological or technical replicates as one group and the research goal may embrace the determination of group differences. A two-tailed paired Student's t-test is frequently used to determine the significance of the methylation differences between two groups, while ANOVA can be used to compare the methylation levels among two or more groups.

For transcriptome-wide approaches, the data analysis strategies vary depending on the principle of each technique. The analysis of datasets generated using antibody-based techniques follows the same principle as ChIP-seq for the identification of transcription factor binding sites. One frequently used tool is Model-based Analysis of ChIP-Seq (MACS), which adopts a dynamic Poisson distribution for peak calling [47]. Peaks, ranked by p-value, indicate the local biases of read coverage in the genome. The primary goal of both Aza-IP-seq and miCLIP-seq techniques is to identify the direct RNA substrates of cytosine-5 RNA methyltransferases. The data analysis of Aza-IP-seq includes sequence alignment, enrichment analysis and signature analysis. After the sequences alignment, enrichment analysis is performed using the open-source USeq package to identify transcripts that are enriched in replicate samples compared to IgG control sample. Signature analysis is then performed using the VarScan package [48] to scan the enriched transcripts for significant C to G transversion sites that are caused by Aza-IP but not SNPs or indel. These transversion sites are then determined as the cytosine targets of a specific methyltransferase [49]. Despite differential methylation analysis is not desired, meRanTK toolkit provides functions of mapping, methylation calling and enrichment comparison for Aza-IP data. Similarly, the analysis of miCLIP-seq data is to identify enzyme-specific target sites. After sequence alignment, the miCLIP read stop positions will be determined and read counts are normalized to per thousand reads in the replicates. To perform differential methylation analysis for 5-mrC-RIP-seq, miCLIP-seq, and Aza-IP-seq results, both the enrichment of peaks/sites and RNA expression level will be required. Therefore, additional RNAseq data has to be generated. With the reduced cost of NGS, RNA bisulfite sequencing is becoming the prevailing approach to study 5-mrC profiles at single nucleotide resolution. However, the data analysis for RNA bisulfite sequencing is a challenging task. Below, we summarize the key features for several bioinformatics packages dealing with RNA bisulfite sequencing data.

Shared steps for RNA bisulfite sequencing data analysis

Like regular RNA-seq data processing, RNA bisulfite sequencing data analysis involves steps for quality control and read alignment to references. Due to bisulfite conversion, unmethylated cytosines in mRNA will end up as thymines after cDNA conversion. Given that, the level of methylated cytosine in mRNA is much lower than that in genomic DNA [13] and the frequency of C (or G in the cDNA) is extremely low in mRNA bisulfite sequencing data. For Illumina sequencing the sequence quality deteriorates along the read, particularly for bisulfite sequencing reads with low GC content. Prior to sequence alignment, low quality bases should be trimmed off from the raw RNA bisulfite sequencing reads along with adaptor sequences. Clean reads may be obtained using software tools such as Cutadapt [50], Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore), or Trimmomatic [51] to eliminate low-quality bases.

Either an annotated genome or a transcriptome may be used as a reference for the alignment of bisulfite sequencing reads. A step that should not be skipped is to prepare an *in silico* bisulfite-converted reference. If a transcriptome is chosen as the reference, Bowtie 2 is recommended, which is a memory-efficient, highly sensitive and accurate alignment algorithm [52]. Mapping with the transcriptome as a reference may have the issue that a sequence read may be aligned to multiple transcripts derived from the same gene. To address this issue, the longest transcript with the highest mapping score were usually selected as the top candidate [28]. Using a large set of small indexes, HISAT2 is a fast and sensitive splicing aware program with alignment strategies that manage reads spanning multiple exons [53]. Thus, it is a great tool to align reads to the genome. Either using the transcriptome or the genome as a reference, the mapping efficiency is expected to be around 70-80%. To achieve a higher mapping rate, genome and transcriptome references may be used in sequential order. For instance, sequence reads may be mapped to the genome first, and then aligned against the transcriptome for the reads that cannot be mapped to the genome [14].

A comparison of existing tools for RNA bisulfite sequencing data analysis

Several bioinformatics tools have been developed to aid in mapping the clean reads and subsequent methylation calling processes [17, 28, 29]. Methylated RNA analysis ToolKit (meRanTK) is the first publicly available software specialized for high-throughput RNA cytosine methylation data analysis [28]. Written in the Perl language, it utilizes splice-aware bisulfite sequencing read mapping to either the genome or the transcriptome. The toolkit allows for methylation calling and the identification of differentially methylated cytosines with statistical analysis. In addition, a package is provided by meRanTK to annotate candidate 5-mrC sites with genomic features such as gene or transcript names and positional metrics. Worthy of mention, MeRanTK can be used to handle Aza-IP data as well.

Similar to meRanTK, BS-RNA is another efficient and highly automated mapping and annotation tool developed in the Perl language [29]. BS-RNA only supports RNA bisulfite sequencing data generated from directional libraries. Yet, the mapping speed of BS-RNA is much faster than that of meRanTK. By calling the HISAT2 program, BS-RNA can finish the mapping of 80 M 100 bp paired-end reads to the reference genome within five hours. The same job takes over 35 hours to perform for meRanGs using STAR [54] or 101 hours for

meRanGt using TopHat2 [55], which are the two variants of aligners provided by meRanTK. Similar to meRanTK, BS-RNA can also manage “dovetailing” reads generated with paired-end sequencing, where one or both reads seem to extend past the start of the mate read. Such “dovetailing” reads often result from the sequence reads that have their 5'-ends trimmed.

BisRNA is a statistical modeling method for methylation calling [30]. This software integrates tailored filtering to address sequencing and alignment artifacts and data-driven statistical modeling to eliminate the artifacts associated with bisulfite sequencing. Using BisRNA, Legrand et. al reported that very sparse methylated Cs, or possibly none at all, can be found in mRNAs [30]. This result raises awareness for developing more reasonable and statistically reliable data analysis strategies for RNA bisulfite sequencing datasets. BisRNA software can only be used for methylation calling. meRanTK and BS-RNA toolkits have similar functions on handling the processes of mapping, methylation calling, and annotation. Liang et al. performed a comparison between BS-RNA, meRanGs and meRanGt [29]. They concluded that BS-RNA has a better performance than both meRanGs and meRanGt when dealing with simulated reads in the mapping process. Both BS-RNA and meRanGs performed better than meRanGt when mapping published single-end bisulfite sequencing reads. In the methylation calling process, although there is no significant difference in precision among these tools, BS-RNA has a significant higher recall rate than meRanGt and meRanGs.

Several methods have been taken to eliminate false positive sites. Most of them adopted statistical methods to avoid false positive sites and set strict filters during methylation calling [11, 13, 14, 17]. In addition, low quality and unconverted reads were excluded [11, 14, 17] and RNA secondary structure prediction tools were used to filter bisulfite conversion-resistant sites [13, 56]. Furthermore, databases including dbSNP for single nucleotide polymorphisms (SNPs) and REDIdb for RNA editing sites may be explored to filter candidate methylated cytosines overlapping SNPs or RNA editing sites [57]. A recent published paper integrates some of these filters together to exclude the noise that occurs during the generation of RNA bisulfite sequencing data [17]. First, it sets filters in the methylation calling process for read coverage, methylation level, and methylated cytosine depth of sites. Then the Gini coefficient is used to determine the C-cutoff to remove the reads that have too many unconverted cytosines. A signal ratio filter is used to further remove sites in regions that are resistant to bisulfite conversion. P-value is calculated for the gene-specific conversion rate and genes with low conversion rates are discarded. Lastly, Stouffer's method is adopted to calculate the combined P value for biological replicates. A comparison of mapping procedures and filtering steps used in recent publications is summarized in Table 2.

Conclusions and Future Perspectives

In the past decade, technology advancements in methylation detection has reignited interest in the dynamics and biological impacts of 5-mrC in mRNA. However, several issues should be taken into consideration when undertaking RNA methylation studies. mRNA molecules are prone to heat degradation and are more chemically labile than DNA. To avoid RNA degradation, the less aggressive conditions that are adopted in bisulfite conversion will lead

to a large number of false positive sites. Thus, it is critical to ensure successful bisulfite conversion, i.e., by monitoring the bisulfite conversion rate of spike-in RNA controls. On the other hand, over 60% of cytosines in mRNA have methylation levels of less than 20% in mammals [14, 17]. This poses a challenge to accurately determining all the methylation sites in a given sample. The multiple filtering steps during analytical procedures may result in a significant number of false negative calls. Development of novel techniques and associated bioinformatics tools is driven by the needs to address specific biological questions. For instance, determination of co-methylated mRNA transcripts in a single cell may reveal gene pathways sharing a same regulatory mechanism. Finally, future techniques and associated analytical procedures are desired to generate and analyze more sophisticated data to determine the association of mRNA methylation with other important biological phenomena, such as RNA splicing, RNA editing, and other kinds of RNA modifications.

ACKNOWLEDGEMENTS

This work was supported by the Center for One Health Research at the Virginia-Maryland, College of Veterinary Medicine and The Edward Via College of Osteopathic Medicine, NIH grant NS094574, and the Fralin Life Sciences Institute faculty development fund for H.X., and VT's Open Access Subvention Fund. We recognize The Center for Engineered Health and the Virginia-Maryland College of Veterinary Medicine at Virginia Tech. We thank Dr. Janet Webster for English language editing.

REFERENCES

1. He C, Grand challenge commentary: RNA epigenetics? *Nat Chem Biol*, 2010 6(12): p. 863–5. [PubMed: 21079590]
2. Saletore Y, et al., The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol*, 2012 13(10): p. 175. [PubMed: 23113984]
3. Song J and Yi C, Chemical Modifications to RNA: A New Layer of Gene Expression Regulation. *ACS Chem Biol*, 2017 12(2): p. 316–325. [PubMed: 28051309]
4. Boccaletto R, et al., MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res*, 2018 46(D1): p. D303–d307. [PubMed: 29106616]
5. Zhao X, et al., FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res*, 2014 24(12): p. 1403–19. [PubMed: 25412662]
6. Wang X, et al., N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, 2014 505(7481): p. 117–20. [PubMed: 24284625]
7. Wang X, et al., N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell*, 2015 161(6): p. 1388–99. [PubMed: 26046440]
8. Iwanami Y and Brown GM, Methylated bases of ribosomal ribonucleic acid from HeLa cells. *Arch Biochem Biophys*, 1968 126(1): p. 8–15. [PubMed: 5671075]
9. Dubin DT and Stollar V, Methylation of Sindbis virus 26S messenger RNA. *Biochem Biophys Res Commun*, 1975 66(4): p. 1373–9. [PubMed: 1191298]
10. Dubin DT and Taylor RH, The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic Acids Res*, 1975 2(10): p. 1653–68. [PubMed: 1187339]
11. Edelheit S, et al., Transcriptome-wide mapping of 5-methylcytosine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet*, 2013 9(6): p. e1003602.
12. Squires JE, et al., Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res*, 2012 40(11): p. 5023–33. [PubMed: 22344696]
13. Amort X, et al., Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biol*, 2017 18(1): p. 1. [PubMed: 28077169]
14. Yang X, et al., 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader *Cell Res*, 2017 27(5): p. 606–625. [PubMed: 28418038]

15. Go 11 MG., et al., Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science*, 2006 311(5759): p. 395–8. [PubMed: 16424344]
16. Tuorto F, et al., RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat Struct Mol Biol*, 2012 19(9): p. 900–5. [PubMed: 22885326]
17. Huang X, et al., Genome-wide identification of mRNA 5-methylcytosine in mammals. *Nat Struct Mol Biol*, 2019 26(5): p. 380–388. [PubMed: 31061524]
18. Tahiliani M, et al., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 2009 324(5929): p. 930–5. [PubMed: 19372391]
19. Ito S, et al., Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 2010 466(7310): p. 1129–33. [PubMed: 20639862]
20. He YF., et al., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, 2011 333(6047): p. 1303–7. [PubMed: 21817016]
21. Ito S, et al., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, 2011 333(6047): p. 1300–3. [PubMed: 21778364]
22. Fu L, et al., Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J Am Chem Soc*, 2014 136(33): p. 11582–5. [PubMed: 25073028]
23. Huber SM, et al., Formation and abundance of 5-hydroxymethylcytosine in RNA. *Chembiochem*, 2015 16(5): p. 752–5. [PubMed: 25676849]
24. Basanta-Sanchez M, et al., TET1-Mediated Oxidation of 5-Formylcytosine (5fC) to 5-Carboxycytosine (5caC) in RNA. *Chembiochem*, 2017 18(1): p. 72–76. [PubMed: 27805801]
25. Shen Q, et al., Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. *Nature*, 2018 554(7690): p. 123–127. [PubMed: 29364877]
26. Yang L, et al., m(5)C Methylation Guides Systemic Transport of Messenger RNA over Graft Junctions in Plants. *Curr Biol*, 2019.
27. Cheng JX, et al., RNA cytosine methylation and methyltransferases mediate chromatin organization and 5-azacytidine response and resistance in leukaemia. *Nat Commun*, 2018 9(1): p. 1163. [PubMed: 29563491]
28. Rieder D, et al., meRanTK: methylated RNA analysis ToolKit. *Bioinformatics*, 2016 32(5): p. 782–5. [PubMed: 26543174]
29. Liang F, et al., BS-RNA: An efficient mapping and annotation tool for RNA bisulfite sequencing data. *Comput Biol Chem*, 2016 65: p. 173–177. [PubMed: 27647160]
30. Legrand C, et al., Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res*, 2017 27(9): p. 1589–1596. [PubMed: 28684555]
31. Vera-Cabrera L, et al., Dot blot assay for detection of anti-diacyltrehalose antibodies in tuberculous patients. *Clin Diagn Lab Immunol*, 1999 6(5): p. 686–9. [PubMed: 10473518]
32. Miao Z, et al., 5-hydroxymethylcytosine is detected in RNA from mouse brain tissues. *Brain Res*, 2016 1642: p. 546–552. [PubMed: 27117867]
33. Lewinska A, et al., Downregulation of methyltransferase Dnmt2 results in condition-dependent telomere shortening and senescence or apoptosis in mouse fibroblasts. *J Cell Physiol*, 2017 232(12): p. 3714–3726. [PubMed: 28177119]
34. Lewinska A, et al., Reduced levels of methyltransferase DNMT2 sensitize human fibroblasts to oxidative stress and DNA damage that is accompanied by changes in proliferation-related miRNA expression. *Redox Biol*, 2018 14: p. 20–34. [PubMed: 28843151]
35. Cui X, et al., 5-Methylcytosine RNA Methylation in Arabidopsis Thaliana. *Mol Plant*, 2017 10(11): p. 1387–1399. [PubMed: 28965832]
36. Khoddami Y and R Cairns B, Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat Biotechnol*, 2013 31(5): p. 458–64. [PubMed: 23604283]
37. George H, Ule J, and Hussain S, Illustrating the Epitranscriptome at Nucleotide Resolution Using Methylation-iCLIP (miCLIP). *Methods Mol Biol*, 2017 1562: p. 91–106. [PubMed: 28349456]
38. Hussain S, et al., NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep*, 2013 4(2): p. 255–61. [PubMed: 23871666]

39. Van Haute L, et al., Deficient methylation and formylation of mt-tRNA(Met) wobble cytosine in a patient carrying mutations in NSUN3. *Nat Commun*, 2016 7: p. 12039. [PubMed: 27356879]
40. Frommer M, et al., 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(5): p. 1827–31. [PubMed: 1542678]
41. Schaefer M, et al., RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res*, 2009 37(2): p. e12.
42. Chen YS., et al., 5-Methylcytosine Analysis by RNA-BisSeq. *Methods Mol Biol*, 2019 1870: p. 237–248. [PubMed: 30539560]
43. Foss-Feig JH, et al., Searching for Cross-Diagnostic Convergence: Neural Mechanisms Governing Excitation and Inhibition Balance in Schizophrenia and Autism Spectrum Disorders. *Biol Psychiatry*, 2017 81(10): p. 848–861. [PubMed: 28434615]
44. Hussain S, et al., Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol*, 2013 14(11): p. 215. [PubMed: 24286375]
45. Amort T, et al., Long non-coding RNAs as targets for cytosine methylation. *RNA Biol*, 2013 10(6): p. 1003–8. [PubMed: 23595112]
46. Tost J and Gut IG, DNA methylation analysis by pyrosequencing. *Nat Protoc*, 2007 2(9): p. 2265–75. [PubMed: 17853883]
47. Feng J, et al., Identifying ChIP-seq enrichment using MACS. *Nat Protoc*, 2012 7(9): p. 1728–40. [PubMed: 22936215]
48. Koboldt DC, et al., VarScarr: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*, 2009 25(17): p. 2283–5. [PubMed: 19542151]
49. Khoddami V and Cairns BR, Transcriptome-wide target profiling of RNA cytosine methyltransferases using the mechanism-based enrichment procedure Aza-IP. *Nat Protoc*, 2014 9(2): p. 337–61. [PubMed: 24434802]
50. Martin M, Cut adapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 2011 17:p. 10–12.
51. Bolger AM, Lohse M, and Usadel B, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014 30(15): p. 2114–20. [PubMed: 24695404]
52. Langmead B and Salzberg SL, Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 2012 9(4): p. 357–9. [PubMed: 22388286]
53. Kim D, Langmead B, and Salzberg SL, HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*, 2015 12(4): p. 357–60. [PubMed: 25751142]
54. Dobin A, et al., STAR: ultrafast universal RNA-seq aligner *Bioinformatics*, 2013 29(1): p. 15–21. [PubMed: 23104886]
55. Kim D, et al., TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 2013 14(4): p. R36. [PubMed: 23618408]
56. Wei Z, et al., Topological Characterization of Human and Mouse m(5)C Epitranscriptome Revealed by Bisulfite Sequencing. *Int J Genomics*, 2018 2018: p. 1351964. [PubMed: 30009162]
57. Parker BJ, Statistical Methods for Transcriptome-Wide Analysis of RNA Methylation by Bisulfite Sequencing. *Methods Mol Biol*, 2017 1562: p. 155–167. [PubMed: 28349460]

Highlights

1. Epitranscriptomics is an exciting, new field for understanding the fundamental mechanisms underlying RNA modifications and their impact on gene expression.
2. Cytosine methylation in mRNA (5-mrC) is an important epigenetic mark that modulates mRNA transportation, translation, and stability at the post-transcriptional level.
3. This short review summarized the experimental techniques that were exploited to determine 5-mrC in mRNA and the computational procedures implemented for RNA bisulfite sequencing data analysis.

Table 1. Summary of techniques for the detection of RNA cytosine-5 methylation (5-mrC)

Level	Techniques	Principle	Advantages	Disadvantages	Resolution	Data analysis
Global level	RNA dot blot	antibody-antigen reaction	Quick to obtain result	Qualitative or semi-quantitative data	Relative global 5-mrC level	t-test
	ELISA-based 5-mrC detection	antibody-antigen reaction	Quantitative measurement of 5-mrC level; commercially-available kit	Unable to detect locus-specific 5-mrC level	Absolute global 5-mrC abundance	t-test
	LC-MS/MS	mass spectrometry	Accurate quantitative measurement of 5-mrC abundance	Unable to detect locus-specific 5-mrC level	Absolute global 5-mrC abundance	t-test
Transcriptome-wide level	5-mrC-RIP-seq	RNA immunoprecipitation	Enable to enrich low abundance RNAs with 5-mrC modification	IP induces background noise; not single nucleotide resolution	100nt-200nt resolution	MACS for peak calling; compare peak intensity between groups
	Aza-IP-seq	Protein immunoprecipitation	Enable to identify specific enzyme target 5-mrC sites	IP induces background noise; 5-Azais toxic to cells	enzyme-specific nucleotide resolution	meRanTK: mapping, methylation calling (C to G transversion), enrichment comparison
	miCLIP-seq	Protein immunoprecipitation	Enable to identify specific enzymetarget 5-mrC sites	The generation of mutant enzymes is time-consuming and costly	enzyme-specific nucleotide resolution	mapping, methylation calling (end of sequencing reads)
	RNA Bisulfite sequencing	Bisulfite conversion of unmethylated cytosine to uracil while methylated cytosine remains unchanged	Provide unbiased transcriptome-wide single nucleotide view	Difficult to detect 5-mrC modification in low abundance RNAs due to degradation during bisulfite conversion process	single-nucleotide resolution	Bioinformatic tools: meRanTK, BS-RNA; BisRNA
	5-mrC-RIP followed by RT-qPCR	RNA immunoprecipitation	Enable to measure the relative methylation level of specific 5-mrC sites.	IP induces background noise	Locus-specific level	t-test
Locus-specific level	RNA bisulfite conversion combined with PCR amplicon-based or cloning-based Sanger sequencing	Sanger sequencing	Enable to detect the methylation level of specific 5-mrC sites	sequencing depth	Locus-specific level	t-test
	RNA bisulfite PCR combined with pyrosequencing	pyrosequencing	Enable to detect the methylation level of specific 5-mrC sites	Primer design is sequence context-dependent.	Locus-specific level	t-test

Abbreviations: IP: immunoprecipitation

Table 2: A comparison of mapping procedures and filters used in mRNA methylation studies

Mapping tool	Edelheit, S., et al., PLoS Genet, 2013 ^[11]	Amort, T., et al., Genome Biol, 2017 ^[13]	Yang, X., et al., Cell Res, 2017 ^[14]	Huang, T., et al., 2019 ^[17]
Reference	Novoalign Genome/Transcriptome	meRanGs (STAR) Genome	Bismark (Bowtie2) Genome → Transcriptome → exon-exon junctions	HISTA2/Bowtie2 Genome → Transcriptome
Reads filters	Identical reads were considered as a single read to eliminate PCR duplicate in the genome-based analysis; 40-nt long reads with 3 unconverted cytosines were eliminated	Potential PCR duplicates were filtered by defining at most five identical reads	Reads with > 30% unconverted cytosines were eliminated	Gini coefficient was used to determine C-cutoff to remove the reads with unconverted cytosines
Sites filters	Coverage depth 5; Methylation level threshold; Base quality >20; P value < 0.01	Coverage depth 10; Methylation level 0.2; Base quality 35 for single-end reads; Base quality 30 for paired-end reads; FDR < 0.01	Coverage depth 30; Methylation level 0.1; Methylated cytosine depth 5	Coverage depth 20; Methylation level 0.1; Methylated cytosine depth 3; Base quality 30; P value < 0.001
Other filters	Candidate methylation sites within 10 nt from an additional candidate site were discarded	10 bases on the 5' end of forward reads and 7 bases on the 5' end of reverse reads were excluded from methylation calling; RNA secondary structure was conducted with RNA fold to discard base-paired sites; The presence of candidate site in all three replicates was required	The presence of candidate site in two replicates was required	Signal ratio filter was used to further remove sites in conversion resistant regions; Excluded genes with low conversion rate; The presence of candidate site in the biological replicates was required