

## MIT Open Access Articles

### *Messenger RNA modifications: Form, distribution, and function*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Gilbert, W. V., T. A. Bell, and C. Schaening. "Messenger RNA Modifications: Form, Distribution, and Function." *Science* 352.6292 (2016): 1408–1412.

**As Published:** <http://dx.doi.org/10.1126/science.aad8711>

**Publisher:** American Association for the Advancement of Science (AAAS)

**Persistent URL:** <http://hdl.handle.net/1721.1/105269>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.





Published in final edited form as:

*Science*. 2016 June 17; 352(6292): 1408–1412. doi:10.1126/science.aad8711.

## Messenger RNA modifications – Form, distribution, and function

Wendy V. Gilbert<sup>1,\*</sup>, Tristan A. Bell<sup>1,2,†</sup>, and Cassandra Schaening<sup>1,3,†</sup>

<sup>1</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>2</sup>Graduate Program in Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>3</sup>Graduate Program in Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

### Abstract

RNA contains more than 100 distinct modifications that promote the functions of stable non-coding RNAs in translation and splicing. Recent technical advances have revealed widespread and sparse modification of messenger RNAs with N<sup>6</sup>-methyladenosine (m<sup>6</sup>A), 5-methylcytosine (m<sup>5</sup>C) and pseudouridine (Ψ). Here we discuss the rapidly evolving understanding of the location, regulation and function of these dynamic mRNA marks, collectively termed the epitranscriptome. We highlight differences among modifications and between species that could instruct ongoing efforts to understand how specific mRNAs target sites are selected and how their modification is regulated. Diverse molecular consequences of individual m<sup>6</sup>A modifications are beginning to be revealed but the effects of m<sup>5</sup>C and Ψ remain largely unknown. Future work linking molecular effects to organismal phenotypes will broaden our understanding of mRNA modifications as cell and developmental regulators.

### Main Text

The first modified RNA nucleoside was identified almost 60 years ago by analyzing salt-soluble RNA from yeast (*1*). Since then, more than 100 chemically distinct modified nucleotides have been characterized, most of which were identified in tRNAs and other abundant non-coding RNAs from diverse organisms (*2*). N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) was the first internal mRNA modification discovered (*3,4*), and at ~1 to 3 m<sup>6</sup>A residues per message, it is abundant enough to be readily detected by bulk mRNA analysis. Next-generation sequencing approaches have allowed mapping of the locations of m<sup>6</sup>A and less-abundant modified nucleosides. Here we summarize these methods and the mRNA modification landscape they reveal. We then discuss the enzymes responsible for installing m<sup>6</sup>A, m<sup>5</sup>C, and pseudouridine (Ψ) at specific sites and emphasize questions regarding the basis for target specificity and regulation. Finally, we describe the remaining challenges in determining the functions of mRNA modifications, which have the potential to regulate genes with widespread consequences for development and disease.

\*Corresponding author. [wgilbert@mit.edu](mailto:wgilbert@mit.edu).

†These authors contributed equally to this work.

## Transcriptome-wide mapping of m<sup>6</sup>A, m<sup>5</sup>C, and Ψ

All high-throughput methods for locating modified nucleosides rely on one of two approaches: Either antibodies are used to isolate modified RNA fragments for sequencing or some modification-selective RNA chemistry is exploited (Fig. 1, A to C). The first genome-wide RNA modification maps were generated using antibodies against m<sup>6</sup>A to identify thousands of ~100 nucleotide RNA fragments containing the modification in mammalian cells (5,6). This approach has been adapted to give single-nucleotide-resolution m<sup>6</sup>A maps by cross-linking the antibody-RNA complexes and determining the sites of crosslink-induced mutations within enriched RNA fragments (7,8). Antibody-based modification profiling has the advantage of concentrating sequencing efforts on sites of interest.

There is some evidence for artifactual enrichment of mRNA fragments lacking m<sup>6</sup>A. In budding yeast, where the only known m<sup>6</sup>A-generating enzyme is not essential for cell viability, it was possible to rigorously determine the background association of unmethylated mRNA fragments with anti-m<sup>6</sup>A antibodies. Notably, almost half of the putative m<sup>6</sup>A peaks were found to be methyltransferase independent (9), suggesting a potentially widespread problem with the current m<sup>6</sup>A maps in many systems. Higher confidence m<sup>6</sup>A sites can be identified by performing methylome mapping after genetic manipulation of methyltransferases and/or demethylases or by overlaying m(6)A peak profiles with maps of methyltransferase interaction sites obtained by crosslinking and immunoprecipitation (CLIP) approaches.

Pseudouridine mapping relies on chemical strategies to selectively derivatize Ψ nucleosides with N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide p-tosylate (CMCT), a bulky covalent adduct that creates a block to reverse transcriptase (RT) (10). The use of a “click” chemistry-compatible CMC coupled to biotin allows pre-enrichment of Ψ-containing RNA fragments (11). Different computational strategies have been used to identify Ψ sites from analysis of CMC-dependent RT stops in yeast and human cells (11–14), which may explain some discrepancies between Ψ annotations.

The potential effect of an mRNA modification depends on both the molecular consequences and the percentage of transcripts that are modified. For example, a modification that leads to accelerated mRNA decay is unlikely to have much biological effect if only 1% of transcripts are modified, whereas a modification that caused an alternative protein variant to be produced could be functionally important even at very low levels. A limitation of current m<sup>6</sup>A and Ψ profiling methods is the lack of quantitative information about the extent of modification. Changes in the relative enrichment of a particular sequence in m<sup>6</sup>A pull-downs from different growth states have been used to infer regulation of modification, but the absolute fraction of mRNA that is modified cannot be determined from these data. Similarly, differences in the abundance of CMC-dependent reads can indicate relative changes in pseudouridylation when comparing the same mRNA site in different conditions but cannot be compared between different sites due to sequence-dependent capture biases in library preparation. A method has been developed to enable absolute quantitation of modified nucleosides at specific mRNA sites (15), but this technique does not scale to allow parallel

measurements at many sites. High-throughput methods to quantify site-specific m<sup>6</sup>A and Ψ would considerably advance the field.

In contrast to m<sup>6</sup>A and Ψ, the level of m<sup>5</sup>C at specific sites in RNA or DNA can be quantitatively determined by bisulfite sequencing. For transcriptome-wide analysis of m<sup>5</sup>C, isolated RNA is treated with bisulfite to convert unmodified cytosines to uracils before cDNA synthesis. The extent of modification at each cytosine residue is then determined by observing the rate of nonconversion – the fraction of reads that do not show the expected C to T sequence change – assuming complete conversion of unmodified C nucleosides. The first map of m<sup>5</sup>C claimed more than 8000 candidate sites in mRNAs from human (HeLa) cells (16); however, these sites could include other cytosine modifications known or suspected to interfere with bisulfite conversion (17) and may include false positives from stochastic non-conversion events. More targeted high-throughput methods exploit the catalytic mechanism of m<sup>5</sup>C methyltransferases to trap covalent intermediates formed between these enzymes and their target sites which are then identified by RNA sequencing after immunoprecipitation of the methyltransferase (18). Targeted bisulfite sequencing at candidate m<sup>5</sup>C sites allows verification and quantitation of modification, and could be scaled to monitor hundreds of sites using microfluidics-based multiplex polymerase chain reaction and deep sequencing (19).

Although non-coding RNAs are extensively modified in all organisms, the modification landscape of prokaryotic mRNAs has barely begun to be explored. m<sup>6</sup>A has been reported in mRNA from *Escherichia coli* and *Pseudomonas aeruginosa* (20), and m<sup>5</sup>C has been mapped in *Sulfolobus solfataricus* mRNAs (21). Because there are substantial differences between prokaryotic and eukaryotic mRNA metabolism, the functional consequences of mRNA modifications are likely to differ as well. It will be interesting to see how these ancient and conserved RNA modifications are harnessed for post-transcriptional gene regulation in organisms with widely differing lifestyles.

Reproducibility and orthogonal validation give confidence in mark annotations. The m<sup>6</sup>A methylomes of different human cell lines are highly overlapping (5,22), genetic manipulation of methyltransferases and demethylases produces the expected changes in m<sup>6</sup>A signals (9,22–24), and thousands of m<sup>6</sup>A target RNAs have been shown to crosslink to the modifying enzymes in vivo (24,25). In contrast, for m<sup>5</sup>C and Ψ there are still few transcriptome-wide profiles and almost no independent analysis of similar cell types or growth states by different groups. Because functionally important sites are likely to be identified in multiple studies, a critical next phase will be the determination of m<sup>5</sup>C and Ψ sites whose detection is robust to technical variation.

## mRNA modifying enzymes and target specificity

In mammalian mRNA, m<sup>6</sup>A is primarily produced by the methyltransferases METTL14 and METTL3, orthologs of yeast *IME4*. METTL3 and METTL14 associate with the regulatory subunit WTAP (Wilms' tumor1-associating protein) to form a 200 kDa methyltransferase complex (24,25). Knockout mammalian embryonic stem cell lines lacking METTL3 or METTL14 display up to 99% reductions in bulk mRNA methylation and reduced m<sup>6</sup>A

signal at thousands of sites (22,26). Most mammalian m<sup>6</sup>A sites are found within the consensus sequence Rm<sup>6</sup>ACH (R = G or A, H = A, C, or U), which is consistent with the enriched binding motifs observed in CLIP studies of METTL14, METTL3 and WTAP (GGAC, GGAC and GACU, respectively) (24).

Despite the strong consensus, only a small fraction of RACH sites are detectably methylated in vivo, arguing that the sequence motif is not sufficient to determine the distribution of m<sup>6</sup>A. Furthermore, m<sup>6</sup>A sites are strongly biased toward the 3' ends of transcripts in organisms as diverse as yeast and humans (5,6,9). This distribution suggests a functional coupling between modification and other RNA processing steps. There is also substantial enrichment of m<sup>6</sup>A sites in the vicinity of stop codons in a variety of mammalian cell types (5,6,22), which hints at a role for the ribosome. A simple model can predict which RGAC sites will be methylated in yeast using just three features: the sequence flanking the site, the proximity of the site to the 3' end, and the predicted secondary structure at the site (unstructured favored). Expanding the consensus sequence to ANRGACNNU yielded the greatest predictive power; nevertheless, only ~10% of sites that matched this consensus were observed to be methylated (9). It remains unclear how much of the observed site-specificity of m<sup>6</sup>A modifications is due to the intrinsic substrate preferences of these enzymes, and it would be useful to systematically and quantitatively test modification of diverse methyltransferase substrates.

Additional methyltransferase enzymes may also modify mRNAs with m<sup>6</sup>A. METTL4 is closely related to METTL14 and METTL3 and is predicted to have catalytic activity (27). Although knockdown of METTL4 had no detectable effect on bulk m<sup>6</sup>A levels in mRNA from HeLa cells (24), the sensitivity of this assay is limited and would not detect changes in modification of a small sub-population of mRNA substrates. Human cells express at least one active m<sup>6</sup>A methyltransferase in addition to METTL14 and METTL3, as suggested by a study of an m<sup>6</sup>A site on the human U6 small nuclear RNA (snRNA). The sequence context of this site does not match the RACH consensus, and methylation of this site cannot be outcompeted by an mRNA target of METTL3 or METTL14 (28).

Pseudouridine sites in mRNA are much less abundant than m<sup>6</sup>A yet the enzymology is more complex (Table 1). There is evidence for mRNA pseudouridylation by 8 out of 10 yeast Ψ synthases, whose canonical targets are mostly sites in tRNA, but also include snRNA and ribosomal RNA sites (12–14). A plurality of identified mRNA substrates in yeast has been genetically assigned to two tRNA modifying enzymes: Pus1 (12), whose human ortholog also modifies mRNAs (11), and Pus7 (13). The basis for Pus1's target site specificity has long puzzled those studying tRNA modification. Pus7 modifies mRNA sites found in the specific sequence context UGΨAR in agreement with its known tRNA target site preferences. However the structural context of this motif is critical for pseudouridylation of non-coding RNAs (29), which may explain why only a tiny fraction of RNAs with UGUAR motifs are modified. Pseudouridylation by all Pus enzymes likely involves recognition of RNA secondary structures, the details of which remain to be discovered. Transcriptome-wide RNA structure probing techniques should facilitate identification of the structural determinants for mRNA pseudouridylation, similar to work on m<sup>6</sup>A (30). mRNA pseudouridylation has been determined in only a few eukaryotic cell types thus far. Because

the Ψ synthases that modify mRNAs in yeast and human cells have homologues in all domains of life, mRNA pseudouridylation is likely to be widely distributed.

Comparatively little is known about the enzymology of m<sup>5</sup>C deposition in mRNA. Although m<sup>5</sup>C is common in non-coding RNAs from all domains of life, there are marked species differences in the reported prevalence of m<sup>5</sup>C in mRNA: ~8000 m<sup>5</sup>C sites in human mRNAs compared to a single mRNA m<sup>5</sup>C site in budding yeast (16,21). *Saccharomyces cerevisiae* expresses three known m<sup>5</sup>C methyltransferases: Ncl1, which modifies multiple tRNAs at several positions and Nop2 and Rcm1, which target the ribosomal RNA. In human cells, the methyltransferases Dnmt2 and Nsun2 have been shown to modify certain mRNAs (16,18), but their verified targets do not include most reported m<sup>5</sup>C sites, suggesting that additional enzymes may be active towards mRNA substrates. Humans have seven known proteins in the Nsun methyltransferase family (Nop2 and Nsun2 to 7), of which five have been shown to have catalytic activity, and all seven have the predicted active site cysteine residues. No sequence motifs are common among reported m<sup>5</sup>C sites in human mRNAs, which is expected if multiple enzymes are responsible. In contrast to humans, all known m<sup>5</sup>C sites in *S. solfataricus* exactly match the sequence of a ribosomal m<sup>5</sup>C site suggesting a common enzyme (21).

## Regulation of mRNA modifications

m<sup>6</sup>A is a truly dynamic mRNA modification that can be enzymatically removed by demethylase “eraser” enzymes. Fat mass and obesity-associated protein, FTO, was the first mammalian RNA demethylase implicated in the dynamics of m<sup>6</sup>A modifications (23). ALKBH5 is a second, conserved eraser of m<sup>6</sup>A marks that is highly expressed in the testes and is required for spermatogenesis and fertility in mice (31). The sequence or structural preferences of the demethylases may account for some or all of the observed specificity in the distribution of m<sup>6</sup>A among individual RACH sites.

Alternatively, the m<sup>6</sup>A “writing” and “erasing” enzymes may have only minimal sequence requirements (for RACH motifs), allowing the accessibility of potential sites and/or the availability of competing RNA-binding proteins to shape the landscape of m<sup>6</sup>A modification. An example of this mode of regulating m<sup>6</sup>A comes from studies of the heat shock response in mammalian cells (MEF and HeLa) (32). It is plausible that m<sup>6</sup>A modification is somehow coupled to co-transcriptional RNA processing events, given the pronounced 3′ bias of modified sites. Moreover, shifts in the relative abundance of m<sup>6</sup>A in 5′ compared with 3′ UTRs (3′ untranslated regions) [e.g. upon heat shock (32,33)] could arise from a global change in recruitment of methyltransferases or demethylases to promoters.

Changes in enzyme or substrate localization may also play an important role in regulating mRNA modifications. In yeast, heat shock induces relocalization of the Ψ synthase Pus7 from the nucleus to the cytoplasm. This shift correlates with a 10-fold increase in the number of Pus7-dependent pseudouridylated mRNA target sites (13), which could reflect an increased window of opportunity for Pus7 to interact with potential mRNA substrates. Similarly, conditions that decrease the rate of mRNA export, either globally or for specific

messages, could increase mRNA modification by nuclear enzymes. In *S. cerevisiae*, the Ime4-Mum2-Slz1 m<sup>6</sup>A methyltransferase complex localizes to nucleoli at the stage of meiosis in which m<sup>6</sup>A mRNA levels peak, and mutations in *SLZ1* that prevent nucleolar accumulation of Ime4 also reduce m<sup>6</sup>A levels by about a factor of 3 (9).

Another apparently localization dependent regulatory mechanism involves heat shock-induced nuclear accumulation of the mammalian m(6)A “reader” protein, YTHDF2, which competes with the FTO demethylase for binding to m(6)A sites (32). The observation that heat shock causes a substantial shift in the distribution of m(6)A from 5′ to 3′ sites implies that 5′ sites are more efficiently targeted by demethylases during normal growth and may be more inherently dynamic.

Pseudouridylation is thought to be irreversible, so its observed dynamics are likely mediated by the production or degradation of pseudouridylated messages. It is also possible that the molecular consequences of an irreversible modification such as Ψ could be functionally mitigated by additional chemical transformations of the modified nucleoside (i.e. by affecting RNA-protein interactions and/or RNA structure). Ψ can be further modified by N1 methylation (34), raising the possibility that the effects of mRNA pseudouridylation can be modulated if not reversed. Similarly, m(5)C in RNA can be further modified to 5-hydroxymethylcytidine, 5-formylcytidine, and 5-carboxylcytidine by ten-eleven translocation (Tet) family enzymes (35). The direct reversibility of m<sup>5</sup>C in RNA is an open question, but there are multiple pathways for demethylation of m<sup>5</sup>C in DNA. However, given that all of the known mRNA demethylases act in the nucleus and that mRNAs can be turned over rapidly, reversible and irreversible mRNA modifications may in fact have similar regulatory potential. Consistent with this possibility, m<sup>6</sup>A shows highly dynamic changes during meiosis in *S. cerevisiae* despite the apparent lack of an m<sup>6</sup>A demethylase (9).

It is notable that most known mRNA modifying enzymes are nuclear during normal growth. This localization allows m<sup>6</sup>A, and likely other modifications, to influence mRNA biogenesis from the earliest stages and may enable regulation of mRNA modification state to augment transcriptional control circuits.

## Functional consequences

At the molecular level, mRNA modifications have the potential to affect most post-transcriptional steps in gene expression. Some broad regulatory themes have emerged from mechanistic studies of mRNA regulation by m<sup>6</sup>A, which has been linked to control of mRNA stability (26,36,37), splicing (38,39), and translational efficiency (32,33,40) and to pri-microRNA (miRNA) processing (41) (Fig. 2). m<sup>6</sup>A, and likely other modifications, can mediate diverse effects on mRNA metabolism by affecting interactions with RNA binding proteins. YTHDF2, the first modification-specific RNA binding protein, or ‘reader’, identified (5), increases turnover of m<sup>6</sup>A modified mRNA by promoting co-localization with decay factors (36). This destabilizing function of m<sup>6</sup>A plays an important biological role in stem cell differentiation by regulating key pluripotency factors (26,37). Individual YTH proteins interact with distinct subsets of m<sup>6</sup>A sites and produce different effects on gene

expression when perturbed in different cellular contexts (32,36,40), suggesting combinatorial control by uncharacterized co-regulators.

Dedicated reader proteins have not yet been identified for m<sup>5</sup>C or Ψ, but the ribosome has the potential to function as a universal reader of mRNA modifications. Incorporation of single m<sup>6</sup>A, m<sup>5</sup>C, or Ψ modifications in specific codon contexts reduced protein production in *E. coli* by 20 to 70% (42), a level of repression exceeding that of many conserved microRNA target sites (43). There is also the possibility of encoding truncated proteins through site-specific ribosome stalling at modified codons, depending on how the stall is resolved. Perhaps the most exciting potential for mRNA modifications to affect gene function is through regulated re-wiring of the genetic code, and there are limited but intriguing observations supporting this possibility. Insertion of m<sup>5</sup>C led to 4% re-coding of proline as leucine in *E. coli* (42), and Ψ-containing stop codons were efficiently mis-translated as specific amino acids in budding yeast (44). The effects of modified nucleosides on translation have been tested in very limited contexts, and the mechanisms responsible for non-canonical decoding events are unknown. If coding sequence modifications lead to the production of alternative protein variants, even low occupancy modifications could have substantial biological effects.

RNA modifications may broadly influence mRNA metabolism through their effects on RNA structures. m<sup>6</sup>A destabilizes RNA duplexes in vitro (45), and m(6)A sites in mRNA tend to be unstructured in vivo (30). Both m<sup>5</sup>C and Ψ impact tRNA folding and are likely to affect mRNA structures as well (46). By changing the accessibility of binding sites for regulatory factors, the effect of RNA structure may be direct (e.g. inhibiting splicing or translation initiation by blocking access to functional sites) or indirect, as was recently shown for m<sup>6</sup>A structural switches affecting splicing (39). Given the stabilizing effects of Ψ on RNA structure and the central role of RNA structure in prokaryotic gene regulation, it will be particularly interesting to see if eubacterial and archaeal mRNAs are also pseudouridylated. The effect of individual modifications is likely to be highly dependent on context. Therefore, it will be important to balance unifying descriptions of the effects of particular modifications against the need to understand sites that deviate from global trends. The incorporation of context effects has been critical for progress towards elucidating the ‘splicing code’ (47), which offers an instructive model for investigating the regulatory effects encoded in mRNA modifications.

As the mRNA modification field develops further, the most promising future work will directly relate individual modifications to cellular and organismal phenotypes. To date, most mRNA modification studies have employed sequencing-based assays to explore the epitranscriptomic landscape in vivo. Although this approach has rapidly expanded our knowledge of the frequency and diversity of cellular mRNA modifications, the importance of most modification sites remains enigmatic. Detailed investigations of individual modification sites are sorely needed. A key step is determining the molecular effects of preventing or introducing individual modifications at physiologically relevant sites, as recently determined for m<sup>6</sup> methylation of A103 of the 5′ UTR of Hsp70. This modification is necessary and sufficient to promote non-canonical cap-independent translation (32),



though the cellular importance of this methylation event in the heat shock response remains unclear.

The most substantial barrier to detailed studies of molecular function is identifying which modifications sites are the most biologically relevant. Conservation analysis is an under-utilized and cost-effective tool for prioritizing specific mRNA modification sites for in-depth characterization. Hundreds of m<sup>6</sup>A sites are conserved between human and mouse embryonic stem cells (22). Likewise, dozens of orthologous sites become modified with m<sup>6</sup>A during meiosis in yeast species that last shared a common ancestor more than 5 million years ago (9). There is substantially more conservation of modified genes than sites in both yeast and mammals. This might be expected if the relevant (conserved) functional consequence of m<sup>6</sup>A addition is recruitment of a trans-acting factor that works similarly from anywhere within the 3' UTR. On the other hand, modification sites that control alternative splicing or protein re-coding should be conserved at the nucleotide level. Thus, examining the patterns of conservation could illuminate the likely molecular function of a modification as well as its biological importance.

With the use of CRISPR-based genome engineering, it is currently possible to directly test the influence of altering any modification site in many organisms. CRISPR multiplexing strategies could also potentially permit interrogation of many sites in parallel and hasten functional discoveries. As a complementary approach, it would be valuable to promote de novo modification at specific sites by engineering fusions between catalytic domains (e.g. methyltransferase, pseudouridine synthase) and a programmable RNA-binding domain scaffold such as Pumilio.

## The expanding epitranscriptome

Just how complex is the epitranscriptome? High-throughput sequencing methods were required to reveal the presence of sparse Ψ modifications in mRNA, suggesting that new approaches will discover still more previously unseen mRNA modifications. In fact, while this Review was in preparation, the first maps of 5-hydroxymethylcytidine and N1-methyladenosine sites were reported using new sequencing techniques to examine mRNA from flies, yeast, mice and human cells (48–50). Given the enormous diversity of RNA modifications found in tRNA, and the demonstrations that many tRNA modifying enzymes also have mRNA substrates (11–14, 16, 18), the mRNA modification landscape is likely to be very rich indeed.

The past 5 years have witnessed a revolution in our understanding of the extent and diversity of mRNA modifications. Many clever enrichment and chemical modification strategies coupled to high-throughput sequencing have produced maps of m<sup>6</sup>A, m<sup>5</sup>C, and Ψ across the transcriptomes of many organisms and demonstrated the dynamics of these modifications across different cellular growth states. Now that diverse molecular functions of mRNA modifications are beginning to emerge, the field is poised for breakthroughs in understanding the most important cellular and organismal functions of mRNA modification.

## Acknowledgments

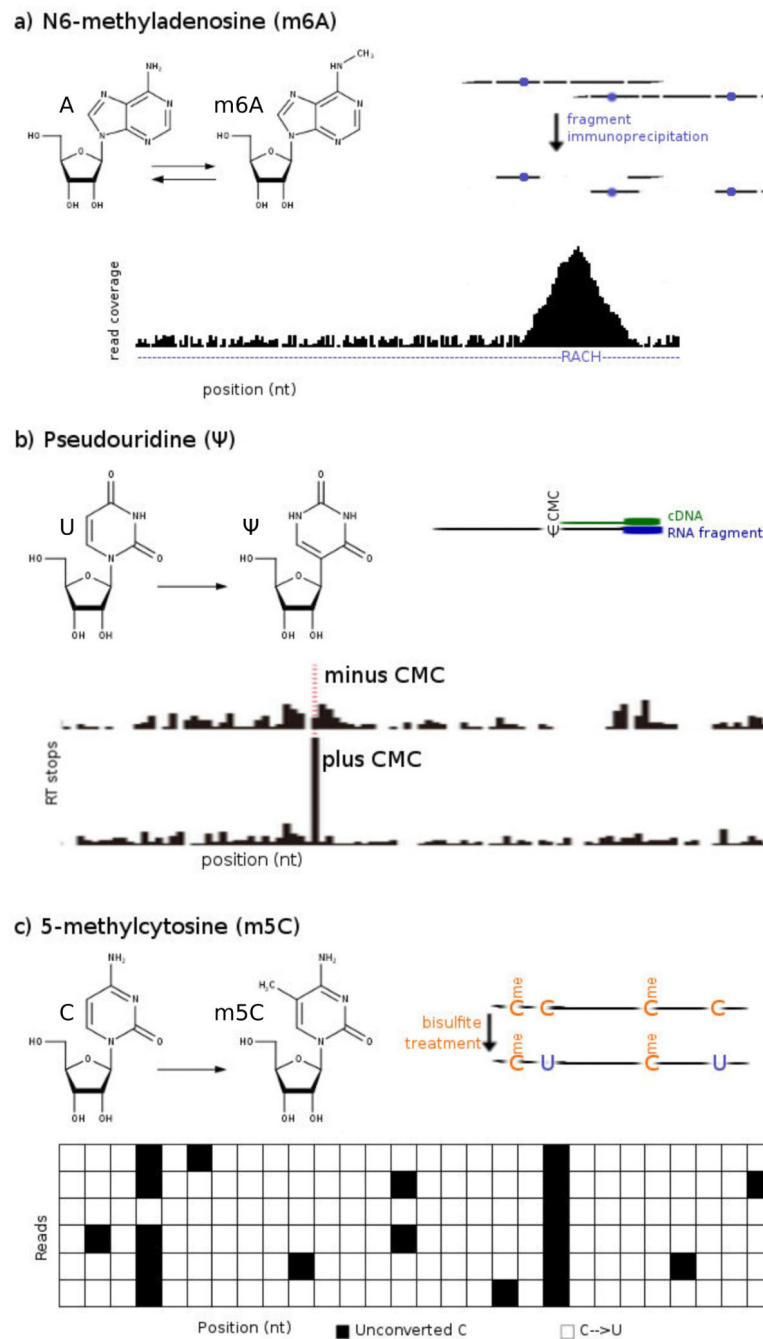
We thank members of the Gilbert lab for discussion. This work was supported by grants from the NIH (GM101316 and CA187236) and the American Cancer Society (RSG-13-396-01-RMC) to W.V.G., a NIH Pre-Doctoral Training grant (T32GM007287) to T.A.B., and an NSF Graduate Research Fellowship to C.S.

## References

1. DAVIS FF, ALLEN FW. Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* 1957; 227:907–15. [PubMed: 13463012]
2. Machnicka MA, et al. MODOMICS: a database of RNA modification pathways–2013 update. *Nucleic Acids Res.* 2013; 41:D262–7. [PubMed: 23118484]
3. Perry RP, Kelley DE. Existence of methylated messenger RNA in mouse L cells. *Cell.* 1974; 1:37–42.
4. 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; 71:3971–5. [PubMed: 4372599]
5. Dominissini D, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature.* 2012; 485:201–6. [PubMed: 22575960]
6. Meyer KD, et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell.* 2012; 149:1635–46. [PubMed: 22608085]
7. Chen K, et al. High-resolution N(6)-methyladenosine (m(6)A) map using photo-crosslinking-assisted m(6)A sequencing. *Angew. Chem. Int. Ed. Engl.* 2015; 54:1587–90. [PubMed: 25491922]
8. Linder B, et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods.* 2015; 12:767–72. [PubMed: 26121403]
9. Schwartz S, et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell.* 2013; 155:1409–21. [PubMed: 24269006]
10. Bakin A, Ofengand J. Four newly located pseudouridylylation residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry.* 1993; 32:9754–62. [PubMed: 8373778]
11. Li X, et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* 2015; 11:592–7. [PubMed: 26075521]
12. Carlile TM, et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature.* 2014; 515:143–6. [PubMed: 25192136]
13. Schwartz S, et al. Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell.* 2014; 159:148–62. [PubMed: 25219674]
14. Lovejoy AF, Riordan DP, Brown PO. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One.* 2014; 9:e110799. [PubMed: 25353621]
15. Liu N, Pan T. Probing RNA Modification Status at Single-Nucleotide Resolution in Total RNA. *Methods Enzymol.* 2015; 560:149–59. [PubMed: 26253970]
16. Squires JE, et al. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* 2012; 40:5023–33. [PubMed: 22344696]
17. Schaefer M, Pollex T, Hanna K, Lyko F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.* 2009; 37:e12. [PubMed: 19059995]
18. Khoddami V, Cairns BR. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* 2013; 31:458–64. [PubMed: 23604283]
19. Zhang R, et al. Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nat. Methods.* 2014; 11:51–4. [PubMed: 24270603]
20. Deng X, et al. Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Res.* 2015; 43:6557–67. [PubMed: 26068471]

21. Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet.* 2013; 9:e1003602. [PubMed: 23825970]
22. Batista PJ, et al. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell.* 2014; 15:707–19. [PubMed: 25456834]
23. Jia G, et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 2011; 7:885–7. [PubMed: 22002720]
24. Liu J, et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* 2014; 10:93–5. [PubMed: 24316715]
25. Ping X-L, et al. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* 2014; 24:177–89. [PubMed: 24407421]
26. Geula S, et al. Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science.* 2015; 347:1002–6. [PubMed: 25569111]
27. Bujnicki JM, Feder M, Radlinska M, Blumenthal RM. Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. *J. Mol. Evol.* 2002; 55:431–44. [PubMed: 12355263]
28. Shimba S, Bokar JA, Rottman F, Reddy R. Accurate and efficient N-6-adenosine methylation in spliceosomal U6 small nuclear RNA by HeLa cell extract in vitro. *Nucleic Acids Res.* 1995; 23:2421–6. [PubMed: 7630720]
29. Urban A, Behm-Ansmant I, Branlant C, Motorin Y. RNA sequence and two-dimensional structure features required for efficient substrate modification by the *Saccharomyces cerevisiae* RNA:Psi-synthase Pus7p. *J. Biol. Chem.* 2009; 284:5845–58. [PubMed: 19114708]
30. Spitale RC, et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature.* 2015; 519:486–90. [PubMed: 25799993]
31. Zheng G, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell.* 2013; 49:18–29. [PubMed: 23177736]
32. Zhou J, et al. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature.* 2015; 526:591–4. [PubMed: 26458103]
33. Meyer KD, et al. 5' UTR m6A Promotes Cap-Independent Translation. *Cell.* 2015; 163:999–1010. [PubMed: 26593424]
34. Meyer B, et al. The Bowen-Conradi syndrome protein Nep1 (Emg1) has a dual role in eukaryotic ribosome biogenesis, as an essential assembly factor and in the methylation of Ψ1191 in yeast 18S rRNA. *Nucleic Acids Res.* 2011; 39:1526–37. [PubMed: 20972225]
35. Shen L, Song C-X, He C, Zhang Y. Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu. Rev. Biochem.* 2014; 83:585–614. [PubMed: 24905787]
36. Wang X, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature.* 2014; 505:117–20. [PubMed: 24284625]
37. Wang Y, et al. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* 2014; 16:191–8. [PubMed: 24394384]
38. Zhao X, et al. FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.* 2014; 24:1403–19. [PubMed: 25412662]
39. Liu N, et al. N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature.* 2015; 518:560–564. [PubMed: 25719671]
40. Wang X, et al. N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell.* 2015; 161:1388–1399. [PubMed: 26046440]
41. Alarcón CR, Lee H, Goodarzi H, Halberg N, Tavazoie SF. N6-methyladenosine marks primary microRNAs for processing. *Nature.* 2015; 519:482–5. [PubMed: 25799998]
42. Hoernes, TP, et al. Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res.* 2015. <<http://www.ncbi.nlm.nih.gov/pubmed/26578598>>
43. Eichhorn SW, et al. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol. Cell.* 2014; 56:104–15. [PubMed: 25263593]

44. Karijolic J, Yu Y-T. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature*. 2011; 474:395–398. [PubMed: 21677757]
45. Roost C, et al. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *J. Am. Chem. Soc.* 2015; 137:2107–15. [PubMed: 25611135]
46. Helm M. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* 2006; 34:721–33. [PubMed: 16452298]
47. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*. 2008; 14:802–13. [PubMed: 18369186]
48. Delatte B, et al. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* (80-. ). 2016; 351:282–285.
49. Dominissini D, et al. The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature*. 2016 doi:10.1038/nature16998.
50. Li X, et al. Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat. Chem. Biol.* 2016 doi:10.1038/nchembio.2040.



**Fig. 1. Nucleotide modifications and detection strategies**

For each modification, the chemical structure (left), detection strategy (right), and sample output of mapped sequencing reads (bottom) are shown. (A) For detection of m<sup>6</sup>A, antibodies are used to select methylated RNA fragments. A typical broad peak in read coverage overlaps an m<sup>6</sup>A site. nt, nucleotide. (B)  $\Psi$  are detected as CMC-dependent reverse transcriptase stops at the 3' end of the U site. (C) m<sup>5</sup>C is protected from conversion into U during bisulfite treatment, and putative sites are identified by their high rates of

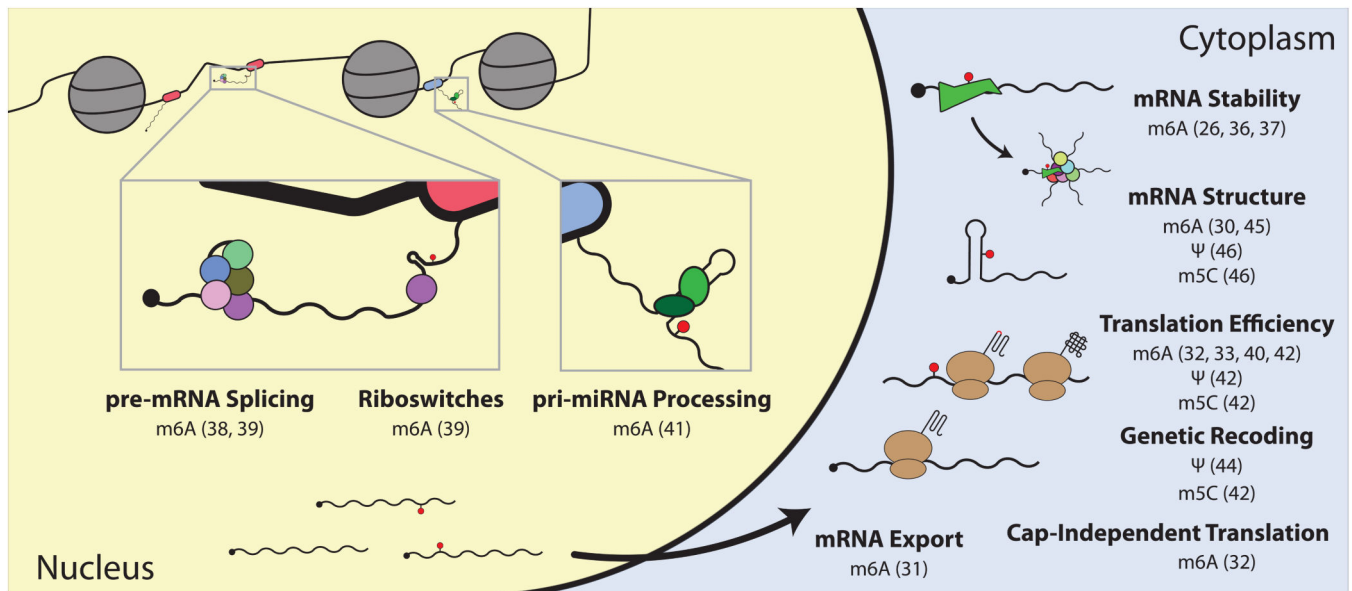
nonconversion. Each row represents one sequencing read, and filled squares indicate unconverted Cs. me, methylated site.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 2. Diverse molecular functions of m<sup>6</sup>A, Ψ, and m<sup>5</sup>C in coding RNAs**

Nascent RNA transcripts in eukaryotic cells are chemically modified (red dot) by m<sup>6</sup>A, Ψ, and m<sup>5</sup>C “writer” enzymes. In the nucleus, m<sup>6</sup>A, and potentially other modifications, alters processing of pre-mRNA and pri-miRNA, through both direct recognition and induced changes in RNA secondary structure (38,39,41). After export to the cytoplasm, which is enhanced by m<sup>6</sup>A (31), mRNA modifications alter the efficiency and fidelity of translation (32,33,40,42,44) and turnover of transcripts in the actively translating mRNA pool (26,36,37).

**Table 1**

mRNA modifying enzymes and modification readers

	<b>m<sup>6</sup>A</b>	<b>Ψ</b>	<b>m<sup>5</sup>C</b>
<b>Writers</b>	METTL14 (m) (22, 26) METTL3 (m) (22, 26) METTL4? (m) (27)	Pus1 (m, y) (11, 12, 13) Pus2 (y) (12, 13) Pus3 (y) (12, 13) Pus4 (y) (11, 12, 13) Pus6 (y) (12, 13) Pus7 (m, y) (11, 12, 13) Pus9 (y) (12, 13) Cbf5 (y) (12, 13)	Dnmt2 (m) (16, 18) Nsun2 (m) (16, 18)
<b>Erasers</b>	FTO (m) (23) ALKBH5 (m) (31)	None Identified	None Identified
<b>Readers</b>	YTH Family (m) (32, 36, 40) YTHDF2 (m) (5, 32)	None Identified	None Identified

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript