REVIEW ARTICLE

RNA epigenetics

Q13 **NIAN LIU, and TAO PAN**

CHICAGO, ILL

Mammalian messenger RNA (mRNA) and long noncoding RNA (lncRNA) contain tens of thousands of posttranscriptional chemical modifications. Among these, the N^6-methyl-adenosine (m^6A) modification is the most abundant and can be removed by specific mammalian enzymes. m^6A modification is recognized by families of RNA binding proteins that affect many aspects of mRNA function. mRNA/lncRNA modification represents another layer of epigenetic regulation of gene expression, analogous to DNA methylation and histone modification. (Translational Research 2014;■:1–8)

Abbreviations: lncRNA = long noncoding RNA; m^1A = N^1-methyl-A; m^1G = N^1-methyl-G; m^5C = 5-methyl cytosine; m^6A = N^6-methyl adenosine; METTL14 = Methyltransferaselike 14; mRNA = Messenger RNA; Nm = 2′-*O*-methyl nucleotides; Ψ = pseudouridine; RT = reverse transcriptase; tRNA = transfer RNA

Q2 More than 100 types of posttranscriptional modifications have been identified in cellular RNA, starting during the 1950s Q3 (http://mods.rna.albany.edu/). For example, the human ribosomal RNA contains more than 200 modifications consisting of 3 major types[1]: ~100 2′-*O*-methyl nucleotides (Nm), ~100 pseudouridines (Ψ), and ~10 base Q4 methylations (eg, 5-methyl cytosine [m^5C]). Each human transfer RNA (tRNA) contains, on average, 14 modifications consisting of various base methylations, Ψ, Nm, and chemically elaborate, modified wobble bases that require catalysis by multiple enzymes.[2,3] Ribosomal RNA modifications are generally used as quality control checkpoints in ribosome assembly.[4] tRNA modifications outside the anticodon loop are generally used to maintain tRNA stability or to modulate tRNA folding, whereas modifications in the anticodon loop are generally used to tune decoding capacity and to control decoding accuracy.[5]

Up until 2 years ago, internal modifications in messenger RNA (mRNA) and long noncoding RNA (lncRNA) were very much neglected. Discovered during the 1970s,[6-9] the most abundant internal mRNA/lncRNA modification is made of N^6-methyl adenosine (m^6A), present, on average, in more than 3 sites per mRNA molecule[10-13] (Fig 1, *A*). Other types of modifications, such as m^5C or Nm, have also been indicated to occur internally in mRNA[9,14] (Fig 1, *B* and *C*), and many m^5C modification sites have now been identified.[15,16] A common feature of these modifications is that their presence cannot be detected by the commonly used reverse transcriptases in complementary DNA Q5

From the Department of Chemistry, University of Chicago, Chicago, Ill; Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Ill; Institute of Biophysical Dynamics, University of Chicago, Chicago, Ill.

Conflicts of Interest: All authors have read the journal's policy on disclosure of potential conflicts of interest and have none to declare.

This work was funded by a National Institutes of Health EUREKA grant (GM88599 to TP).

Submitted for publication January 14, 2014; revision submitted April 1, 2014; accepted for publication April 1, 2014.

Reprint requests: Tao Pan, Department of Biochemistry and Molecular Biology, University of Chicago, 929 East 57th Street, GCIS, W134, Chicago, IL 60637; e-mail: taopan@uchicago.edu. Q1

1931-5244/$ - see front matter

http://dx.doi.org/10.1016/j.trsl.2014.04.003

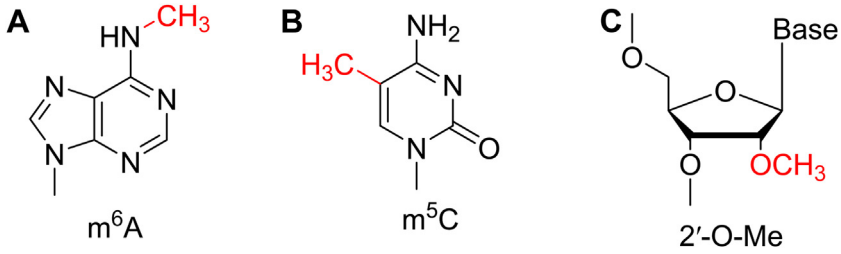

Fig 1. Chemical structure of internal messenger RNA/long noncoding RNA modifications. (**A**) N^6-methyl-adenosine (m^6A). (**B**) 5-Methyl cytosine (m^5C). (**C**) 2′′-*O*-methyl nucleotides (2′-*O*-Me). Q11

synthesis. It was, therefore, extremely difficult to map these modifications at single-nucleotide resolution. Global m^6A modification was shown to be important functionally because siRNA knockdown of a known human m^6A methyltransferase (METTL3) led to apoptosis in cell culture.[17] Suggested functions for m^6A modification include effects on mRNA splicing, transport, stability, and immune tolerance.[17,18]

Interest in mRNA/lncRNA modification was revived in 2011 upon the discovery that m^6A modification is the cellular substrate for the human enzyme FTO.[19] FTO belongs to a family of human genes that are homologous to the *Escherichia coli* AlkB protein, which catalyzes oxidative reversal of methylated DNA and RNA bases.[20,21] In genomewide association studies, the human FTO gene is associated greatly with Q6 diabetes and obesity in the human population.[22,23] FTO knockout mice are much leaner than the wild-type mice, presumably as a result of perturbations in controlling cellular metabolism.[24] The discovery of FTO acting on m^6A in mRNA/lncRNA indicates that m^6A modification is subject to sophisticated cellular control.

The discovery of this first RNA demodification enzyme also highlights the idea that RNA modifications may act as epigenetic markers and controls akin to DNA methylation and histone modification.[25,26] Three groups of proteins are needed for epigenetic control that maintains specific modification patterns in cell type- and cell state-dependent manners. "Writers" catalyze chemical modifications at specific sites, "erasers" remove modifications at specific sites, and "readers" recognize the modified sites in DNA or histones (Fig 2, *A*). For m^6A in mRNA/lncRNA, members in all 3 groups of proteins have now been found in mammalian cells (Fig 2, *B*). However, the current list of these proteins likely represents just the beginning. In particular, the number of reader proteins that recognize m^6A modified mRNA/lncRNA sites will certainly expand greatly in the coming years. As of today, only the m^6A modification has been shown to exhibit all signatures of epigenetic regulation. This review therefore focuses on m^6A modifications in mRNA/lncRNA, with an emphasis on its effect on human health and disease.

TECHNIQUES USED TO STUDY m^6A IN mRNA/lncRNA

A prerequisite for mRNA/lncRNA transcriptome studies is the copying of RNA into complementary DNA by reverse transcriptase (RT). m^6A modification does not affect Watson-Crick base pairing, and it behaves like an unmodified adenosine for the commonly used RTs. A widely applied method for m^6A study is to use immunoprecipitation with a commercial m^6A antibody followed by high-throughput sequencing (m^6A-seq or MeRIP-seq[27,28]). The mRNA/lncRNA mixture is first fragmented chemically to produce suitable-size RNA segments for deep sequencing and to increase the resolution of m^6A detection. The fragmented RNA is split in 2. One is used for m^6A antibody immunoprecipitation to enrich RNA segments that contain m^6A; the other is used as the reference. The location of m^6A modification is obtained by comparing the sequencing read profiles of both samples. This method could identify readily tens of thousands of candidate m^6A modification sites in mammalian mRNA/lncRNA at an average resolution of ~100 nucleotides.[27,28] Studies before the advent of high-throughput sequencing have determined a consensus sequence for mammalian m^6A modification consisting of RRACH (R = A, G, H = A, C, U, m^6A site underlined[13]). Indeed, this consensus sequence is present Q7 in a majority of m^6A/MeRIP-seq peaks. Peaks without this consensus sequence are likely m^6A antibody binding artifacts, as demonstrated in a yeast m^6A study.[29]

To map transcriptomewide m^6A sites at or near single-nucleotide resolution, a combination of high-coverage sequencing and bioinformatics was used in the yeast m^6A study for ~1300 m^6A sites.[29] This approach may not be readily applicable to mammalian RNA, in which the number of m^6A sites is at least 1 order of magnitude greater and the context of m^6A modification is much more diverse. It was shown recently that the human immunodeficiency virus RT is sensitive to the presence of m^6A in RNA using