The Pivotal Regulatory Landscape of RNA Modifications

Sheng Li and Christopher E. Mason

Department of Physiology and Biophysics and HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY 10065; email: chm2042@med.cornell.edu

Annu. Rev. Genomics Hum. Genet. 2014. 15:127–50

First published online as a Review in Advance on June 2, 2014

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

This article's doi: 10.1146/annurev-genom-090413-025405

Keywords

epitranscriptome, RNA editing, ribonucleotide modifications, m^6A

Abstract

Posttranscriptionally modified nucleosides in RNA play integral roles in the cellular control of biological information that is encoded in DNA. The modifications of RNA span all three phylogenetic domains (Archaea, Bacteria, and Eukarya) and are pervasive across RNA types, including messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), and (less frequently) small nuclear RNA (snRNA) and microRNA (miRNA). Nucleotide modifications are also one of the most evolutionarily conserved properties of RNAs, and the sites of modification are under strong selective pressure. However, many of these modifications, as well as their prevalence and impact, have only recently been discovered. Here, we examine both labile and permanent modifications, from simple methylation to complex transcript alteration (RNA editing and intron retention); detail the models for their processing; and highlight remaining questions in the field of the epitranscriptome.

1. INTRODUCTION

Although nucleic acids were first discovered by Friedrich Miescher in 1869, whether the molecular substrate that stored hereditary information was RNA, DNA, or proteins was hotly debated until the 1940s and 1950s (27). Notably, RNA fell out of favor early because of its rapid turnover, instability, and perceived lack of the complexity that would be needed to represent all of life. Eventually, DNA was identified as the molecular basis for genetic inheritance, overcoming the previously favored protein-based theories. Yet it was clear that RNA had a central role in information transfer from DNA to protein, was an integral component of cellular dynamics, and even had autonomously functioning molecules. Indeed, RNA came in many forms that interacted with proteins and DNA, thus serving as a ubiquitous mediator of the information and transfer of states and functions within the cell. Owing to RNA's universal presence in biological information transfer—including in viruses that encode all of their instructions only in RNA—some scientists proposed ideas about an "RNA world" that may have existed before DNA became the predominant form of nucleic acid storage, in which primordial molecules on an early Earth could have been the fuel for the formation of complex life (115).

However, the techniques for characterizing RNA were limited until the 1970s. Eventually, reverse transcriptase was discovered and used to copy RNA into cDNA, making a more stable molecule that was easier to study. This led to the surprising discovery about adenoviruses in 1977 that showed that RNAs are not simple, linear copies of their genomic template DNA—instead, these molecules are sometimes reordered, or spliced, to create alternative isoforms of the original transcript (21). This introduced the concept that the RNAs encoded by DNA could be modified after being transcribed and opened up a large potential universe of posttranscriptional modifications. These data also thoroughly refuted the "one gene, one enzyme" hypothesis previously thought to represent the complexity of specific gene products (7). Thus, with better techniques, scientists began to unravel the complexity of RNA.

Similarly, the impact of methylation and other epigenetic-like modifications of RNA was not revealed until the 1960s and 1970s. The methylation of DNA was discovered in 1925 (54), but the methylation of RNA was not discovered until 1968, and even then, little was known about its precise role and subcellular localization (49, 50). However, recent work (discussed below) has shown that many variations of RNA occur, some of which directly impact the nucleotide content or ordering of the transcripts and others of which are epi-modifications of the bases, similar to the epigenetic modifications of DNA and histones that serve as regulators for the information encoded in DNA. These dynamic RNA epi-modifications, collectively called the epitranscriptome, are pervasive, conserved, and critical for many aspects of biology, including germline development, cellular signaling, and circadian rhythm control. Interestingly, these modifications may impact as many as ~16,000 human genes (91), and thus far they have been observed in almost all species, but their genome-wide prevalence has been discerned only in the past few years. Indeed, there are a total of 110 known RNA modifications across the three domains of life, but the roles and activities of the majority of them remain unknown.

Thus, to generate a complete picture of RNA modifications and their roles, we here review all known modifications and divide them into two main types: reversible and nonreversible. Some modifications, such as RNA editing, splicing, 5′ capping, and transcript-content modification (such as intron retention), are unidirectional and presumed to be nonreversible. Conversely, smaller-scale changes, such as ribose methylation and hydroxylation, are plastic and reversible, and these modifications serve a key role in regulating RNA function. Some enzymes that mediate this process of epitranscriptomic state reversal are already known, and an active field of research has begun to identify the rest of the binding partners of RNA that can modify their substrates. This expansive

catalog of reversible and nonreversible RNA modifications shows how RNA has moved from a transient intermediary of biological information to an active, pivotal regulator of cellular function. Indeed, the joint coordination of modifications along the transcriptome–epitranscriptome axis, much like the interplay of the genome and epigenome, now represents an essential component of the standard model of cellular and molecular biology.

2. REVERSIBLE mRNA MODIFICATIONS

We first examine RNA modifications that have been shown to be reversible or are likely reversible and that essentially comprise epitranscriptomic-type changes that do not fundamentally alter the nucleobase, thus maintaining the information content of the RNA (A, C, G, and U). Changes between purines and pyrimidines can be as simple as four atoms (e.g., a methyl group) or can sometimes be quite dramatic, including a doubling or tripling of the molecular mass of these RNAs (e.g., N^6-glycinylcarbamoyladenosine). These bases can be cataloged as a set of regulatory steps akin to DNA base modifications or histone modifications in epigenetics, insofar as they are plastic and often reversible, but their specific manifestation and control have key roles in basic aspects of gene regulation and cellular states.

2.1. RNA Modification Types

To date, 110 types of RNA nucleotide modifications have been documented, spanning RNA types including messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), transfer-messenger RNA (tmRNA), small nuclear RNA (snRNA), and chromosomal RNA, many of which are currently well cataloged at the RNA Modification Database (**http://mods.rna.albany.edu/mods**) (12). These modifications not only mark and regulate these transcripts, but also diversify and extend the proteins' functionality from the annotated genes (12, 25, 26, 67). This includes 12 types of modifications from eukaryotic mRNA (12, 25, 26, 67), where the methylation and modifications of all four main types of ribonucleotides (A, C, T, and U) enable flexible deviation from the simple genetic code in DNA. Notably, there are other, rarer modifications present in inosines and 7-deazaguanosines, but we focus here on the more recent discoveries in mRNA modifications that have shown a profound effect on biological systems.

The known mRNA modifications are found at varying levels in genes, and each has different functions (**Figure 1**). Starting at the 5′ end of the transcript, modifications include 2′-*O*-methylated ribonucleotides, such as N^6,2′-*O*-dimethyladenosine (m^6Am), N^6,N^6,2′-*O*-trimethyladenosine ($m^6{}_2Am$), and 3,2′-*O*-dimethyluridine (m^3Um) (4). These modifications are often found in the 5′ untranslated region (UTR) and define the beginning of the transcripts (103). In addition, during normal RNA processing, the 7-methylguanosine (m^7G) cap is added to the 5′ end of mRNA; this cap is essential for efficient gene expression, transcript stability, and cell viability (24) and is also a key stabilizing factor for the translation of most cellular mRNAs into proteins for eukaryotic organisms (24). The 5′ cap can also contain N^2,7-dimethylguanosine ($m^{2,7}G$) in viral mRNAs (48), and the guanosine variant N^2,N^2,7-trimethylguanosine ($m^{2,2,7}G$) has been observed in a subset of the 5′ end of *trans*-spliced snRNAs. This 5′ trimethylguanosine cap has been seen persistently throughout development on small, 100-nucleotide RNAs in *Caenorhabditis elegans*, and these rare RNAs are strongly associated with polysomes and may create a distinct class of trimethylguanosine-regulated RNAs (68). Also, $m^6{}_2Am$ and m^3Um have been reported to locate at the 5′ UTR of mRNA in kinetoplastid protozoa (4). Together, these 5′ modifications are important for stability, translation priming, and mRNA expression.

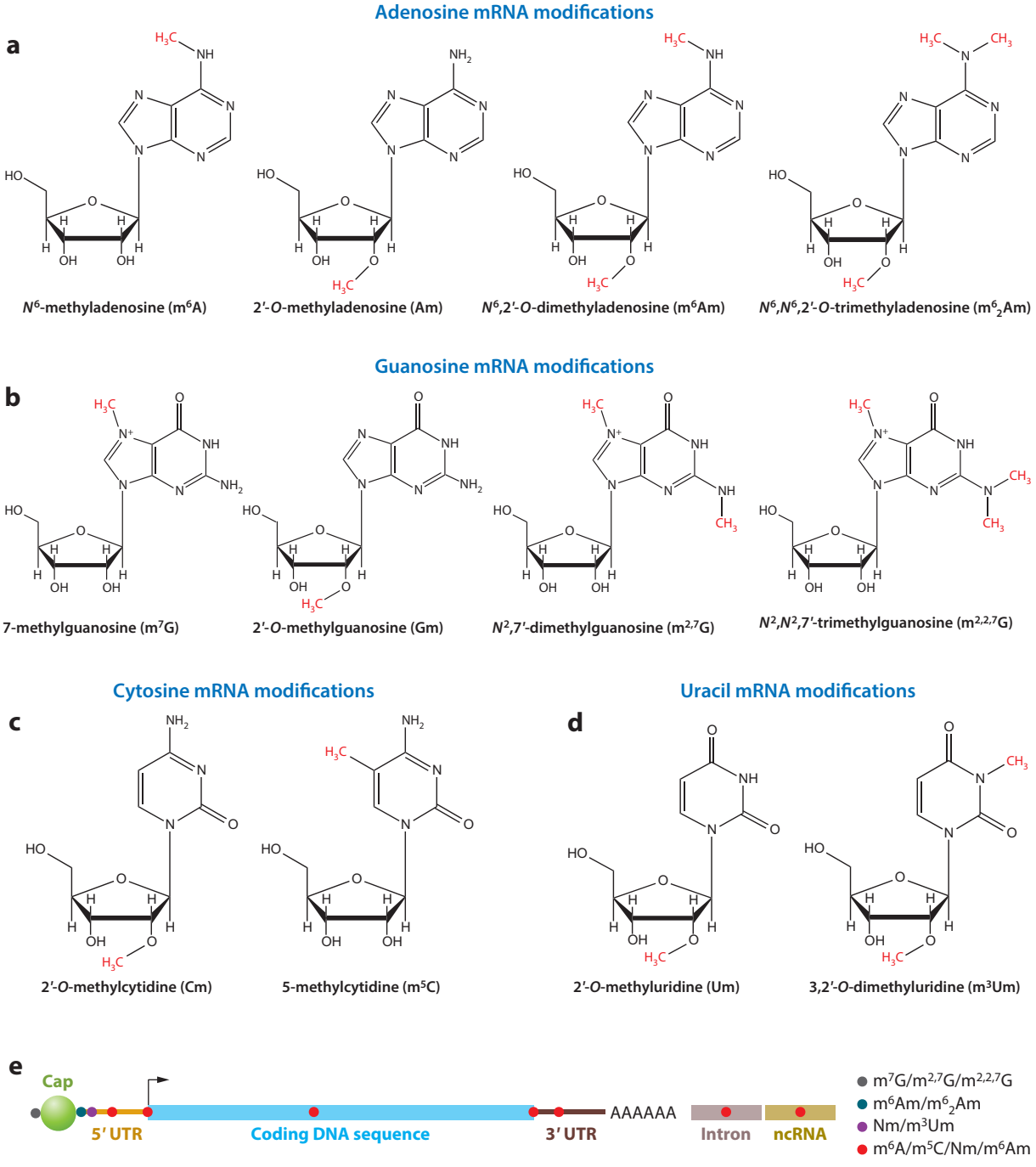

Figure 1

The structures of RNA modifications in eukaryote mRNA. (*a*) Adenosine mRNA modifications. (*b*) Guanosine mRNA modifications. (*c*) Cytosine mRNA modifications. (*d*) Uracil mRNA modifications. (*e*) Localization of mRNA modifications across the positions of most genes. Abbreviations: ncRNA, noncoding RNA; UTR, untranslated region.

The 3′ ends of transcripts also contain many modifications. A key feature of mature mRNAs is the addition of a variable-length poly(A) tail at the end of each transcript, created by polyadenylation enzymes such as polyadenylate polymerase and polyadenylate-binding protein (35). This 3′ modification can be theoretically reversed through the use of degradation enzymes like poly(A)-specific ribonuclease (42), but it is not currently considered one of the 110 types of epitranscriptomic modifications. We mention it here because it is key for transcript stability and the nuclear export of transcripts, and because it does have some dynamic states within cells and can be notably shorter in cancer cells (72).

Other modifications can exist in all regions of transcripts, even if localization spots do occur. Recent whole-transcriptome studies have shown frequent methylation of many cytosine sites across all regions of mRNAs and noncoding RNAs (94, 99), although enrichment was observed in the 5′ and 3′ UTRs. Also, recent work on N^6-methyladenosine (m^6A) has estimated this modification to be the most prevalent type in mRNAs, appearing in 5′ UTRs, coding DNA sequences (CDSs), and 3′ UTRs as well as noncoding RNAs and microRNAs (miRNAs) (32, 72). Yet there is a significant enrichment of m^6A in the last CDSs of most genes, and this modification is likely to have a widespread impact on gene expression, development, and (given the last-CDS enrichment) protein termination (74). These recent works have spurred the development of new methods and techniques to understand the impact of RNA methylation.

2.2. RNA Methylation and Demethylation

The methylation of RNA as m^6A was first observed in the 1970s, when the abundance of m^6A was estimated to be 0.1–0.4% of total adenosine residues in cellular RNA (33, 84, 114). Mammalian mRNA m^6A methylation sites were first found in 1984 (47), and measurements from mass spectrophotometry have shown that m^6A is the most abundant RNA modification in mammalian protein-coding (29) and long noncoding RNA. Yet not until 2012 were protocols for transcriptome-wide m^6A methylation detection developed and published, including methylated RNA immunoprecipitation and sequencing (74) and m^6A sequencing (32). Both of these protocols enable the discovery of regions, or peaks, of m^6A using antibody-enrichment methods. Both protocols found m^6A sites in all areas of transcripts but a strong enrichment in the ends of the last CDSs of genes, as well as some evidence of a 5′ UTR enrichment in some tissues, which has a potential impact on splicing. It is currently estimated that 15,020 human genes have methylated mRNAs, which means that the majority of human genes (based on RefSeq annotation) contain m^6A, and similar trends are seen in mice.

Although these methods revealed that most mammalian mRNAs appear to be regulated or modified by m^6A, the antibody-based methods did not enable single-base resolution of the precise base that was modified. Rather, they gave a region that was significantly enriched compared with a control RNA sample, and ascertainment of the exact adenosine that was modified was performed by looking at the mode of the reads' peak distribution in conjunction with a bioinformatics scan for the m^6A motif [G(AG)ACU]. Although these would often find the likely m^6A site in the center of a peak, the average peak width of 100 nucleotides meant that single- and multi-adenosine sites could not be distinguished, and also complicated any validation efforts for the specific modified adenosine.

Fortunately, recent technological and biochemical advances have now provided a way to localize precise modified bases in RNA. Single-molecule RNA sequencing was first employed in 2012 to study RNA base modifications, using a third-generation sequencer from Pacific Biosciences (91, 107). This method placed a reverse transcriptase in a well called a zero-mode waveguide, which allowed the observation of cDNA synthesis in real time. From these data, Saletore et al. (91) and

Vilfan et al. (107) have shown that the kinetic signature of the reverse transcriptase's speed while creating cDNA is sensitive enough to detect the presence of m^6A. This work was then expanded to reveal that structural variations of rRNAs and other RNAs could also be inferred from the kinetic data of the reverse transcriptase in the zero-mode waveguide (107). Finally, Liu et al. (70) recently used site-specific cleavage and radioactive labeling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET) to detect RNA modifications at single-base resolution.

2.3. Epitranscriptomic Enzymes and Mediators

There are at least three enzymes that guide the process of RNA methylation and demethylation (**Figure 2**). The methylation of m^6A is mediated by the *methyltransferase-like 3* (*METTL3*) gene (**Figure 2*a***), and its methylation mark can be oxidatively reversed by the Fe(II)- and oxoglutarate-dependent AlkB oxygenase family (52, 61). Human METTL3 was first discovered in HeLa cell nuclei as mRNA N^6-adenosine methyltransferase (9). Recently, Liu et al. (69) reported that METTL14 forms a stable heterodimer core complex with METTL3 and mediates mammalian nuclear RNA N^6-adenosine methylation, further revealing the complicated regulatory mechanisms in RNA m^6A methylation. One of the oxygenases is the obesity-associated protein alpha-ketoglutarate-dependent dioxygenase FTO (52) (**Figure 2*b***). FTO is known to play an important role in human obesity and energy homeostasis (22, 37, 96). It has been reported that FTO can oxidatively demethylate methylated uracil and thymidine in single-stranded RNA (ssRNA) (53). Jia et al. (52) recently found that FTO is an RNA demethylase that can oxidatively revert the m^6A modification to adenosine. Human cells with FTO knocked down by small interfering RNA (siRNA) had increased amounts of m^6A in mRNA, whereas overexpression of FTO resulted in decreased amounts of m^6A (52). The major FTO physiological substrate is nuclear RNA, as shown by partial colocalization of FTO with nuclear speckles (52), although there is also evidence of cytosolic activity of FTO.

Recently, another AlkB family protein, ALKBH5 (**Figure 2*c***), was also shown to act as a mammalian RNA demethylase both in vitro and in vivo (61, 120). The demethylation of m^6A via ALKBH5 significantly impacted mRNA export, RNA metabolism, and the assembly of mRNA processing factors in nuclear speckles (120). *Alkbh5*-knockout mice showed increased mRNA m^6A levels, and more than 1,500 genes were differentially expressed, including the p53 functional interaction network (65) and other genes that are known to regulate epigenetics. Thus, this work on ALKBH5 further showed that RNA methylation is important and that the enzymes that mediate their effects are far reaching.

It is highly likely that additional enzymes mediate RNA methylation or other epitranscriptomic states. For example, there are hundreds of RNA-binding proteins predicted in the human genome in addition to the eight in the AlkB family (ALKBH1–8). Indeed, recent work on the use of high-throughput sequencing kinetics to measure protein binding on RNA substrates has shown thousands of precise protein–RNA complexes that appear across the transcriptome (45). This work has also shown that binding, although extremely specific, functions over a wide range of affinities that can allow for some sequence divergence and mutation of the targets. These binding partners create a catalog of thousands of candidates that will need further examination in terms of their downstream impact on mRNA stability, nucleotide conversion rates, cellular regulation, and overall biological impact (69, 112).

Although the modifications of methylcytosine and hydroxylmethylcytosine have led to declarations of these as the "fifth base" and "sixth base" of DNA, we note that there are other intermediates in DNA between these states, such as formylcytosine and carboxylcytosine, that may also appear in RNA. TET enzymes can oxidatively catalyze the formation of these intermediates (58). mRNA

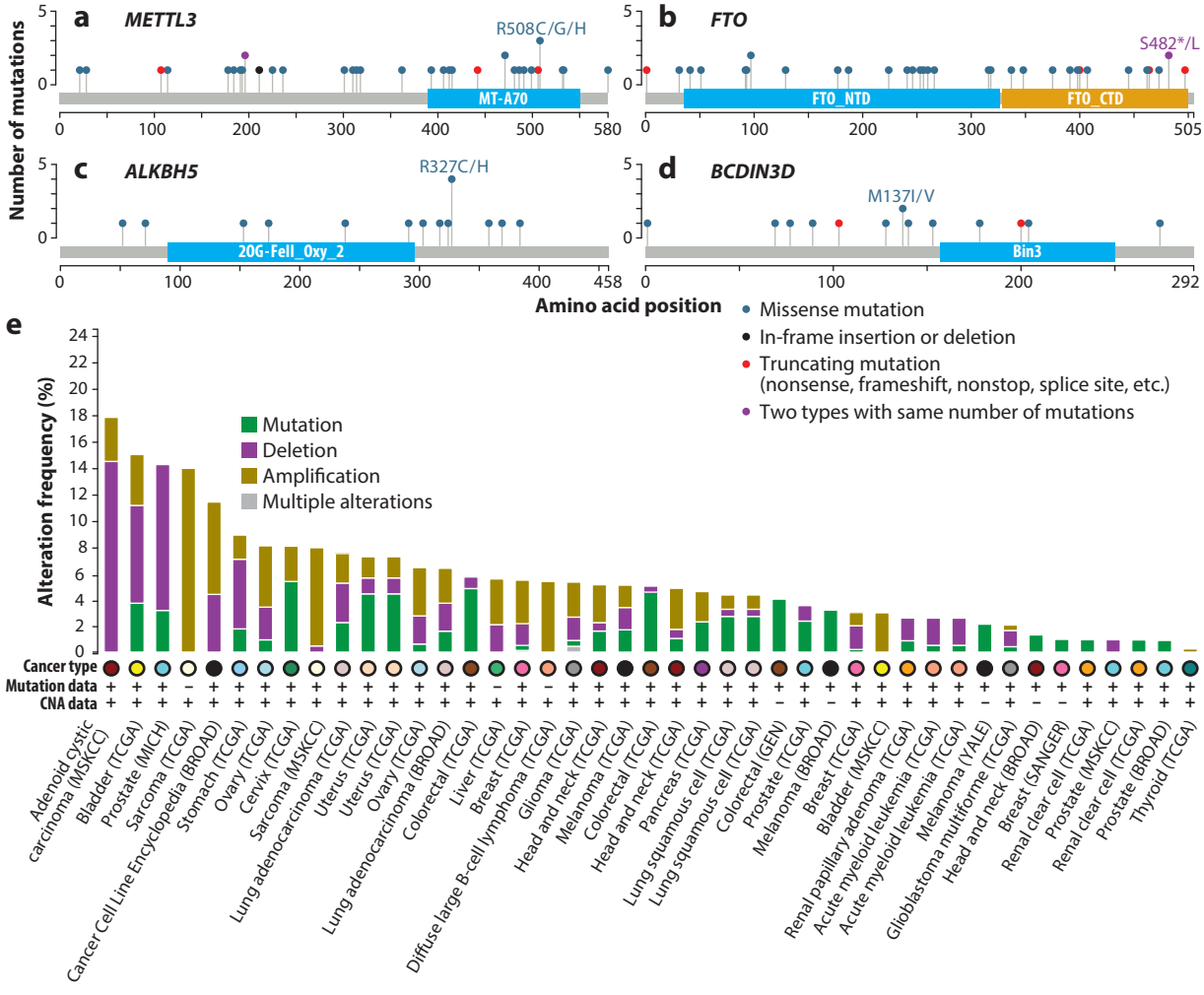

Figure 2

Epitranscriptomic enzymes and mediators. (*a*) METTL3. (*b*) FTO. (*c*) ALKBH5. (*d*) BCDIN3D. (*e*) Cross-cancer alteration summary for METTL3, FTO, ALKBH5, and BCDIN3D (52 studies covering four genes). Abbreviations: BROAD, Broad Institute of MIT and Harvard; CNA, copy-number alteration; GEN, Genentech; MICH, University of Michigan Medical School, Ann Arbor; MSKCC, Memorial Sloan-Kettering Cancer Center; SANGER, Wellcome Trust Sanger Institute; TCGA, The Cancer Genome Atlas; YALE, Yale University.

methylation also has similar intermediate states, such as N^6-hydroxylmethyladenosine (hm^6A) (38, 51) and N^6-formyladenosine (f^6A) (39), in humans and mice. We propose that there might be more intermediates waiting to be found (**Figure 3**) that can regulate mRNA methylation on adenosine (or cytosine), such as N^6-carboxyladenosine (ca^6A). Potentially, hm^6A, f^6A, and the proposed ca^6A may modulate RNA–protein interactions (39). These modifications may serve as independently regulated functional intermediates between modified bases in RNA, just as the cycle of DNA modifications does for epigenetics. They can provide various stop and check points in an RNA processing and regulation cycle that allows tight control of RNA function and localization.

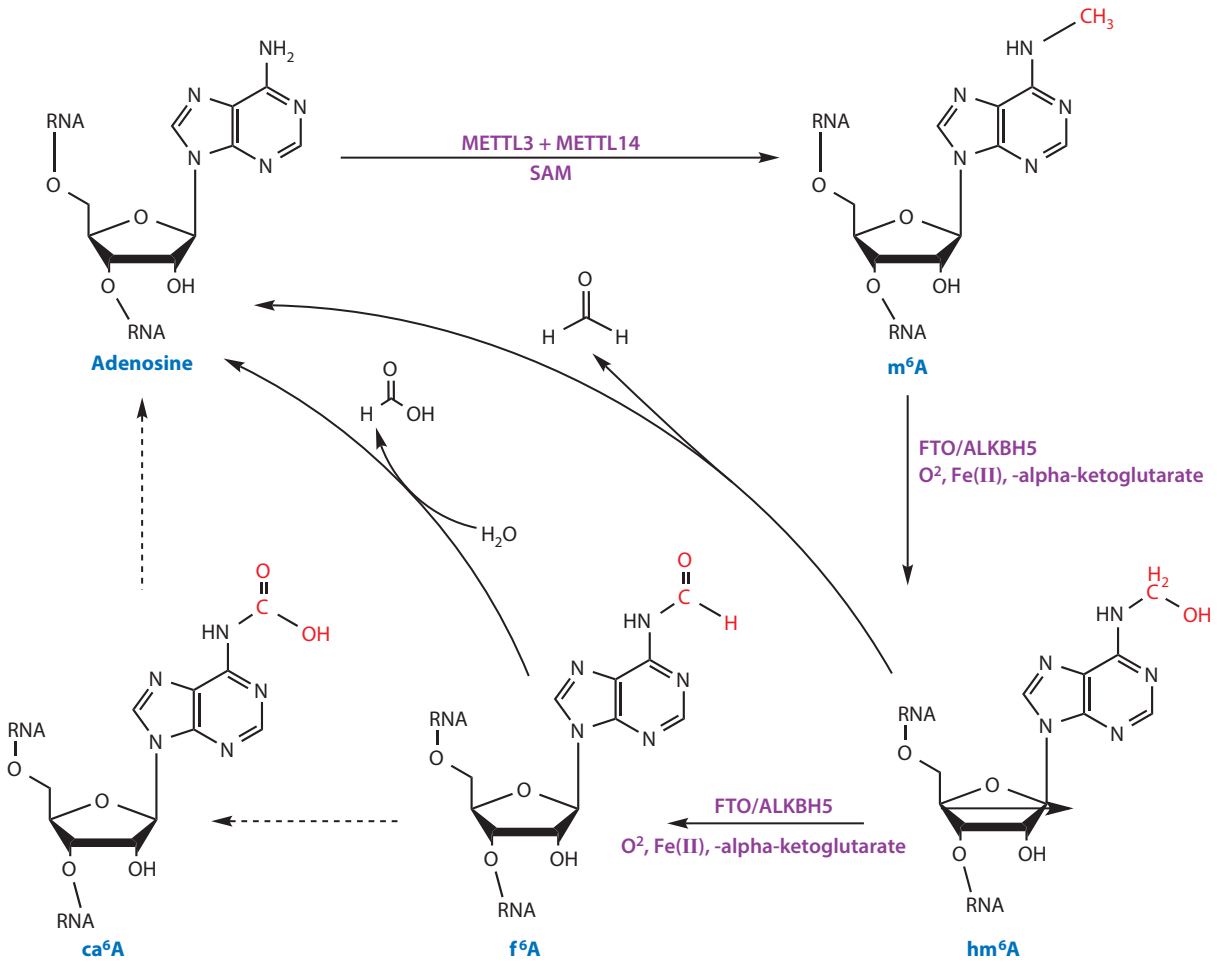

Figure 3

Reversible adenosine methylation and demethylation in mRNA and noncoding RNA. Abbreviations: ca^{6}A, N^6-carboxyladenosine; f^{6}A, N^6-formyladenosine; hm^{6}A, N^6-hydroxylmethyladenosine; m^{6}A, N^6-methyladenosine.

2.4. The Impact of Epitranscriptomic Changes

Although more than 100 modifications have been cataloged, the precise impact of all of these base modifications and their transcript localizations are still mostly unclear. For example, although m^{6}A has been predicted to affect protein translation and localization (90) or mRNA stability (95a, 112, 112a), the precise effects on translation rates or synthesis have not yet been shown. Also, there is evidence that RNA methylation is the long-sought balancing mechanism to prevent RNA editing (51), but this has been shown only in humans and mice, and RNA editing levels are extremely variable between species. Nonetheless, there is currently evidence for at least eight putative roles for epitranscriptomic states in RNA.

First, RNA demethylation regulates gene expression. The mRNA loss of m^{6}A demethylase affects gene expression. *Alkbh5*-knockout mice have altered gene expression in testis cells (120), although this could be a combination of a primary and secondary impact on gene expression.

Another intriguing aspect of this work is that the loss of RNA demethylation appeared to impact not only downstream RNA processing genes but also epigenetic modifying genes such as DNA methyltransferase 1 (DNMT1), Brd4, and Ell3. Thus, there is some evidence for a joint epigenetic–epitranscriptomic regulatory feedback loop and for some degree of activity coordination between these two regulatory layers (91).

Second, RNA methylation may help regulate nuclear mRNA exit and control. ALKBH5-deficient cells showed increased mRNA accumulation in the cytoplasm, indicating that normal m^6A methylation aids the transportation of mRNA from nuclear territory (120). The m^6A methylation also accelerated the nuclear exit of the mRNA of the mature circadian clock genes *Per2* and *Bmal1* (40). Thus, defects in METTL3 reduce m^6A methylation levels, delay nuclear RNA processing, and slow the circadian clock, and the inverse can also happen with defects in FTO and ALKBH5, owing to this fine-grained control of transcript export from the nucleus.

Third, RNA methylation has been implicated as a regulator of alternative splicing. When comparing *METTL3*-silenced HepG2 cells with control cells, the methylated genes were overrepresented in the genes whose constituent isoforms were differentially expressed, and differentially spliced exons and introns were significantly enriched with m^6A peaks (32). However, in terms of spatial localization, it seems that m^6A is not enriched at splice junctions in mouse brain RNA (74). These combined data indicate that RNA methylation may not directly impact the splicing machinery of the cell, and may instead serve to signal the choice of splicing partners in *trans*. There may also be species-specific differences in the regulatory aspects, but further research is needed to reveal the detailed mechanism of the functional role of m^6A methylation in alternative splicing.

Fourth, the mRNA targets of most highly expressed miRNAs have shown a higher level of m^6A methylation, which suggests some interplay between the miRNA-regulated transcripts and the methylation levels of those transcripts (74). Also, whereas m^6A sites are enriched for the 5′ ends of 3′ UTRs, most miRNAs are enriched for the 3′ ends of 3′ UTRs, indicating spatial anticorrelation of the miRNA-binding sites vis-à-vis the m^6A methylation sites. These data raise the questions of how much miRNA sites might be affected by RNA methylation [5-methylcytidine (m^5C) or m^6A, and others] and how the methylated RNA sites in both the miRNA and the mRNA may interact. Xhemalce et al. (118) have already shown that the maturation of miRNAs is somewhat dependent on the methylation of the pre-miRNAs by BCDIN3D, and it is possible that the substrate specificity of miRNA binding and efficacy are regulated by methylation and epitranscriptomic changes.

Fifth, the methylation of m^6A exerts a strong effect on mouse fertility and likely serves critical roles in germline development and maturity. Knocking out the RNA demethylase Alkbh5 in mice led to dysfunctional spermatogenesis, low sperm motility, smaller testicle size, and increased sperm death, resulting in compromised male fertility (120). Also, the meiotic metaphase-stage spermatocytes were impaired by apoptosis because of the increased m^6A mRNA in Alkbh5-deficient male mice (120). Thus, germline development and fertilization itself are strongly affected and regulated by RNA methylation.

Sixth, Fustin et al. (40) demonstrated that the circadian clock is highly sensitive to global m^6A methylation inhibition. These data support the model in which m^6A methylation regulates RNA processing and serves as a circadian pacesetter. This work also showed that RNA methylation helps to determine the period and oscillatory stability of mammalian circadian clockwork and that defects in this regulation quickly disrupt the normal processing of the timekeeping framework of the cells. Together, these functions all point toward the epitranscriptome as a key regulator in the cell cycle.

Seventh, mRNA modification might impact RNA-binding proteins. Di Ruscio et al. (30) reported that DNMT1 can interact with RNA arising from the *CEBPA* gene locus and can play an

important role in local DNA methylation profiles. The interaction of the RNA with DNMT1 inhibits *CEBPA* gene locus methylation (30). Furthermore, Marín-Béjar et al. (71) recently found that long intergenic noncoding RNA interacts with Polycomb repressive complex 2 (PRC2) and is required for PRC2 targeting of specific genes for H3K27 trimethylation and repression. These data show that RNA plays an important role in regulating epigenetic modifiers. RNA methylation may possibly modulate or switch the RNA interaction with RNA-binding proteins. Taken together, these data further point to possible crosstalk between the epitranscriptome and epigenetics for gene regulation.

Finally, data from The Cancer Genome Atlas (TCGA) has shown that a trend of underexpression of the two mRNA demethylases and overexpression of mRNA methyltransferase occurs in many cancers. FTO (**Figure 2*b***) has been mostly reported as underexpressed in cancer, including osteosarcoma (55), non-small-cell lung cancer (92), metastatic prostate cancer (104), and human mucosa-associated lymphoid tissue (MALT) lymphoma (106), when compared with healthy donors. ALKBH5 (**Figure 2*c***) has shown decreased expression levels in osteosarcoma (55), papillary thyroid carcinoma (46), prostate cancer (15, 104), and acute myeloid leukemia (36). Also, METTL3 (**Figure 2*a***) is upregulated in prostate carcinoma (14) and prostate cancer (15, 102, 104). Cross-cancer genetic alterations of these three m^6A methylation modifiers as well as BCDIN3D (**Figure 2*d–e***) have been detected across multiple cancer types from cBioPortal (**http://www.cbioportal.org**) (13, 41). These data show that enzymes for RNA methylation, including m^6A regulation, may have a significant role in various cell and cancer types, and these genes might potentially contribute to cancer risk and evolution via m^6A methylation levels or other RNA modifications.

3. NONREVERSIBLE mRNA MODIFICATIONS

Although small chemical side chains (e.g., $-CH_3$) are relatively easy to modify and often exist within a predicted range of efficiency and stoichiometry for biochemical reactions, there are other posttranscriptional changes to RNA that are more permanent and unidirectional. Most of the changes directly alter the information content of the targeted transcript, adding to the complexity of the encoding by the genome and allowing far greater plasticity and diversity of the molecular roles of these modified RNAs.

3.1. RNA Editing Types and Prevalence

RNA editing is a posttranscriptional process that modifies the primary RNA and miRNA transcripts, thus creating new information in the RNA that is not encoded directly in the DNA. RNA editing was first discovered in trypanosome mitochondrial mRNA in 1986 (8). The mitochondrial cytochrome oxidase subunit II gene in trypanosomes contains a frameshift at amino acid 170 that is not encoded in the DNA (8). The first mammalian nucleus-encoded mRNA editing was found in intestinal mRNA (85). This editing introduces a single C-to-T base difference from DNA and results in a translational stop at codon 2153 (85). Since then, RNA editing events have been found in tRNA (64), mRNA (85), rRNA (28), and miRNA (20, 43).

3.1.1. RNA editing types and prevalence. The primary canonical type of RNA editing is the conversion of adenosine into inosine (A-to-I editing) (6, 108), after which the newly formed inosine is recognized by the translational machinery as guanine. Another type is the conversion of cytosine into uracil (C-to-U editing), which is mediated by the members of the cytidine deaminase (AID/APOBEC) family of proteins (18, 23). The canonical mammalian RNA editing includes the

prevalent A-to-I editing and the rare C-to-U editing, but the existence of noncanonical RNA editing types has been proposed, even though validation rates are low (65).

A-to-I conversion is the most common form of editing and is mediated by the ADAR protein and its isoforms (ADAR1–3 in humans) (5, 87). Data from human cell lines have shown that the RNA editing events in coding sequences are far less frequent than those in noncoding sequences such as UTRs and introns (86). In exonic RNA editing in *Drosophila*, the gene expression increases with the number of edited sites in the transcript, which is not the case in intronic RNA editing (100). Other work in *Drosophila* has shown that the coding-type RNA editing that does occur takes place more often in the genes that have more isoforms per gene. However, editing is likely a consequence rather than a cause of splicing complexity, because this trend does not change in *ADAR*-null mutants compared with the wild type (100). These enrichments of editing sites in noncoding areas are likely due to the danger of RNA editing in protein-coding regions, which potentially leads to nonsynonymous amino acid substitutions, alternative splicing, and nuclear retention of mRNA. RNA editing in non-protein-coding regions may lead to alterations of miRNA seed regions (77) or regulatory changes, but both types of editing are established and utilized by cells.

Multiple large-scale RNA sequencing projects have revealed widespread RNA editing sites in the human transcriptome. The majority of sites are A-to-I editing, and other types of editing have very low validation rates, owing to either false positives or cellular rarity (or both). A much smaller number of editing sites were detected in miRNAs, but some were still found, which suggested a potential link between RNA editing and miRNA-mediated regulation (82). Wang et al. (110) recently identified more than 60,000 A-to-I editing sites (~10,000 of which were known and ~50,000 of which were novel) and several thousand genes whose expression levels are influenced by ADARs. Also, the Encyclopedia of DNA Elements (ENCODE) project examined RNA editing events in whole-transcriptome sequencing data from 15 human cell lines, which provided a great resource to study editing events. This analysis confirmed that only a small fraction of RNA editing events occur in protein-coding sequence (79). It also showed that lymphoblastoid cell lines have more diversified RNA editing events than brain cell lines (79). The Database of RNA Editing (DARNED; **http://beamish.ucc.ie**), which catalogs all RNA editing events, has been widely used for comparisons of known and novel A-to-I editing sites (59, 60). RNA editing events are also prevalent in human Alu repeats, with some estimates indicating that as many as 97% of Alus undergo some editing (86). The Rigorously Annotated Database of A-to-I RNA Editing (RADAR; **http://www.rnaedit.com**) also has a catalog of manually curated annotations on A-to-I RNA editing for humans, mice (*Mus musculus*), and fruit flies (*Drosophila melanogaster*) (86). For humans, only 2,411 out of 1,379,403 annotated A-to-I editing sites are in protein-coding sequence (86). Combining these curated and confirmed RNA editing sites will be helpful to navigate putative RNA editing events as they are discovered.

3.1.2. Regulation of RNA editing. RNA editing is under heavy regulatory influence to maintain an optimal degree of editing for normal biological function (98). The specificity of editing and the extent of the editing performed by ADARs are usually determined by local primary sequence and secondary structural imperfections in duplex RNA or other variations of RNA secondary structure. Tertiary structural elements determine the specificity and extent of mRNA editing (88), and accessing RNA in three-dimensional structures is the means by which ADARs function best.

Regulation of alternative splicing by RNA editing was first reported for ADAR2 in 1999, in a study showing that ADAR2 modulates its own alternative splicing by editing at 3′ splicing junctions (89). Up to 80% of all Adar2 pre-mRNAs extracted from whole rat brain are self-edited and subsequently alternatively spliced (89). However, the CTD domain of RNA polymerase II is required for

efficient editing (63), which means that ADAR2 does require partners for ideal function. Finally, the inhibition of editing at alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor GluA2 Q/R sites closely correlates with intronic editing efficiency, which is linked to splicing efficiency. Edited GluA2 Q/R mRNA shows higher splicing efficiency (83), indicating that the editing efficacy and specificity are under fairly tight control through these cofactors.

In addition, A-to-I RNA editing has the capacity to tune protein function in response to external stimuli such as temperature, spatial localization, and ionic gradients. The AMPA receptor GluA2 R/G RNA editing site is cell-type specific and dynamically regulated in CA1 hippocampal neurons. These changes are bidirectional and reversible and correlate with levels of the editase Adar2. R/G editing is dynamically regulated by neuronal activity, which is linked to Ca^{2+} influx (3). Another example of A-to-I RNA editing responding to the external stimuli was discovered in the K^+ channels from polar octopuses (extensively edited) when compared with tropical species (mostly unedited). This creates protein functional diversity with greatly accelerated gating kinetics by destabilizing the open state of delayed rectifier K^+ channel genes upon adaptation on temperature (44), showing the remarkable plasticity and sensitivity of RNA editing.

3.1.3. Biological roles of RNA editing. There are many established biological roles for RNA editing and ADARs, spanning many species and implicated in several diseases. But, first, it is worth noting that ADARs do have a biological role independent of editing: maintaining transcript stability via interaction with HuR proteins in human B cells (110). However, the focus for ADARs is often on the way they modify RNAs, and there are many places where these events have a large impact. For example, ADAR1 has been implicated in the regulation of stem cell pluripotency and maintenance through A-to-I RNA editing of Alu sequences. Also, human fibroblasts can be reprogrammed to induced pluripotent stem cells when they downregulate their ADAR1 level, indicating that RNA editing helps maintain lineage choices for cells. Finally, there is some evidence of evolutionary selection pressure to drive higher rates of editing-site loss relative to gain, and the strength of selection against editing has become increasingly stringent over the course of angiosperm evolution (75). In mammals, however, humans are among the most edited of primates, with almost 10 times as much editing as is seen in mice, indicating that selection pressures can also work to increase the rates of editing (81).

Based on TCGA data, Burns et al. (11) proposed that APOBEC3B-catalyzed genomic uracil lesions are responsible for a large proportion of both dispersed and clustered mutations in multiple distinct cancers, including bladder, cervix, lung (adenocarcinoma and squamous cell carcinoma), head and neck, and breast cancers. Further studies in breast cancer have shown that APOBEC3B is an enzymatic source of mutation, and a mechanistic linkage was found between increased APOBEC3B and inactivated TP53 in primary tumor data and cell lines (10). In addition, miRNA-mediated loss of ADAR1 by miR-17 and miR-432 in metastatic melanoma promotes tumor growth, and these two miRNAs are frequently highly expressed in melanoma (76).

These changes in RNA editing levels can also impact cancer therapies. Nonsynonymous RNA editing in the mRNA coding regions in these cancer genes can lead to deleterious amino acid changes that may result in conformational changes and differences in protein functions. For example, AZIN1 is an antizyme inhibitor, and its S367G RNA editing event can lead to a conformational change that results in higher AZIN1 protein affinity for antizyme. The sequestration of antizyme suppresses the ubiquitin-independent degradation of ornithine decarboxylase and cyclin D1, which promotes cell proliferation and has been proposed to be a potential driver for human hepatocellular carcinoma pathogenesis (17).

Another function of RNA editing is to mediate miRNA function. One of the prerequisites for efficient RNA editing is double-stranded RNA (dsRNA), which is also the structure of miRNA

during the maturation process. So far, 44 editing sites in miRNAs have been detected (82), and a comparative whole-genome expression microarray analysis revealed that ADAR1 controls the expression of more than 100 miRNAs that regulate many genes associated with the observed phenotypes (76). This indicates that RNA editing may interfere with miRNA-mediated regulation, in either a direct or an indirect way. The A-to-I editing of miRNA in mammalian brain increases during development, indicating that RNA editing may influence brain maturation via the miRNA repertoire (34).

However, precise mechanistic examples by which this regulation occurs have only recently been elucidated. First, primary miRNA editing leads to early degradation inside the nucleus by Tudor-SN or suppression of Dicer/TRBP cleavage. Tudor-SN has been reported to degrade mir142 (93, 119). Pre-miRNA editing can suppress Dicer/TRBP cleavage, which then allows for certain primary transcripts of miR151 subject to A-to-I RNA editing. The edited precursor of miR151 blocks the cleavage by the Dicer/TRPB complex and leads to the accumulation of edited pre-miR-151 RNAs (57), thus attenuating its function.

Second, miRNA editing has the potential to add another layer of complexity to targets of gene regulation pathways, especially if editing occurs within the miRNA–mRNA recognition site. Specifically, RNA editing of miRNA seed sequences can potentially redirect their target specificity. One case is miR-376, where the primary miR-376a1 is edited by ADAR2 in normal whole brain tissue, targets AMFR, and decreases its expression level. However, dysregulated editing in high-grade gliomas results in the accumulation of unedited miR-376a*, which targets RAP2A. These low levels of RAP2A and high levels of AMFR lead to glioma cell migration and invasion (31). To track these changes, a database of predicted A-to-I-edited miRNA-binding sites has been released (miREdiTar; **http://microrna.osumc.edu/mireditar**) (62). The database contains predicted miRNA-binding sites that could be affected by A-to-I editing and sites that could become miRNA-binding sites as a result of A-to-I editing.

It is also notable that editing can indirectly modify miRNA function and RNA interference (RNAi) pathways as well. RNAi is a gene-silencing phenomenon caused by the recognition and degradation of dsRNA, although the hindrance of a gene's translation or transcription can also be a means to suppress gene function. ADAR1 interacts with endogenous RNAi pathways such as Dicer and the RNA-induced silencing complex (RISC), which contains the catalytic RISC component Argonaute. The normal A-to-I RNA editing has two types of functional roles when interacting with endogenous RNAi. The first is as an antagonist: ADAR1 homodimers form and edit the siRNA precursors, leading to less siRNA and miRNA (19). The second, recently reported by Ota et al. (78), is in the Dicer/ADAR1 heterodimer complex, which facilitates RISC loading and enhances cleavage of miRNA and siRNA. The expression of miRNAs is globally inhibited in *ADAR1*$^{-/-}$ mouse embryos (78). Nemlich et al. (76) reported another example of ADAR1 regulating miRNA expression in an RNA editing–independent manner, in which ADAR1 controlled Dicer expression via let-7 at the translational level.

Large-scale studies have begun to reveal the substantial extent to which modifications of RNA editing levels play key roles in cellular regulation. RNA editing can modify regulator miRNAs as well as the miRNA target sites. Such editing can create partial pairing of the miRNA complex to target 3′ UTR sites, which then results in deadenylation and degradation of the mRNA (80). These mechanisms have been thoroughly characterized and compared through an analysis of the human nuclear and cytosolic "editomes" from the ENCODE project. Specifically, Chen (16) found that RNA editing is globally associated with the modification of miRNA regulation in 3′ UTRs, whereas editing events in coding regions are rare and tend to be synonymous. The 3′ UTR editing sites significantly overlap with miRNA targets, with the outcomes including disruption of original miRNA targets (16.1%), creation of new miRNA targets (19.8%), and changes

in miRNA targets through the destruction of the original targets and the creation of new ones (16).

Beyond the global analysis of the RNA editing regulation of miRNA targeting sites, Wang et al. (111) investigated ARHGAP26 mRNA. ADAR1 performs extensive A-to-I editing on the 3′ UTR of ARHGAP26 that removes the repression of this mRNA's expression caused by miR-30b-3p and miR-357 (111). The authors found that ADAR1 expression positively correlates with ARHGAP26 expression. Because these two miRNAs target the ARHGAP26-encoding gene and the editing disrupts their targeting site, the authors further concluded that ARHGAP26 escapes miRNA repression via RNA editing. Because the miRNA can regulate gene expression and either inhibit expression/translation or repress translation (80), the catalog of regulators of RNA editing in the context of miRNA targeting may have many components and needs further study for clarification.

Systematic identification of edited miRNAs in humans using next-generation sequencing data enables whole-miRNA profile editing and potentially changes of their targets (1, 2). Quantification of A-to-I editing of miRNAs using a conventional method involves reverse-transcription polymerase chain reaction (PCR) amplification of regions containing the editing sites followed by subcloning of the PCR products and sequencing (56). Although it is not as precise as the subcloning method, this procedure enables studies of RNA editing events in many samples (56). Also, recent work on single-cell RNA sequencing has aroused significant interest in the field following the rapid development of next-generation sequencing technology for sorted cells. Currently, studies of single-cell-level transcriptome dynamics (117) can be readily accomplished with sorting protocols or specific instruments, and a recent study in *Drosophila* demonstrated genome-wide A-to-I RNA editing at the single-cell level (100). These data showed that targeted mRNAs were regulated by editing and that their changes in editing strongly correlated with alternative splicing (100). Also, the editing sites in *Drosophila* exhibited sequence-motif preferences and tended to be concentrated within a small subset of total RNAs (100).

4. A MODEL OF RNA EDITING AND RNA METHYLATION

Powell et al. (85) suggested that RNA methylation can prevent or reduce the impact of RNA editing, and there is substantial evidence that this "RERM" (RNA editing and RNA methylation) regulatory pathway may guide transcript functions. ssRNA can be methylated by METTL3 and can gain the m^6A modification. On average, five adenosines per transcript are methylated in the human transcriptome (95), and at least 15,020 genes have m^6A modifications present across humans and mice (32). This event is reversible by FTO and ALKBH5, which can oxidatively remove methyl groups from ssRNA (52, 120). Although FTO is able to act on dsDNA and dsRNA, it shows only 40% and 24%, respectively, of the enzyme's 100% activity on ssRNA. Conversely, the m^6A demethylase ALKBH5 has almost no detectable demethylation activity on dsRNA (120); it also strongly prefers to target ssRNA. This provides two avenues to modify the methylation state of transcripts, depending on whether they are in their folded state, are in their bound state, or are nascent, single-stranded transcripts. In addition, RNA methylation blocks the majority of RNA editing for ADAR2, resulting in a level of activity that is only 2% of that of normal adenosine (105).

Given these data, we propose a coordinated model that mediates the crosstalk between the two posttranscriptional pathways of RNA editing and RNA methylation (**Figure 4**). RNA methylation and the entire catalog of epitranscriptomic changes may work to prevent or accelerate RNA folding, which then may impact the ability of ADAR to reach its target. Then, once dsRNA is formed, ADAR can be prevented from working on the methylated (or otherwise modified) adenosine site.

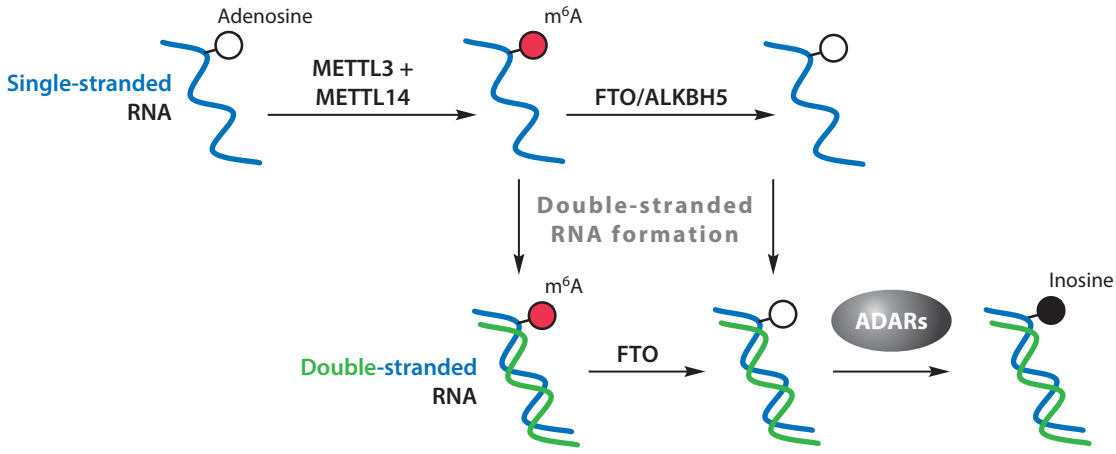

Figure 4

A model of RNA editing and RNA methylation. Abbreviation: m^{6}A, N^6-methyladenosine.

This means that RNA methylation could impact RNA editing both before and during ADAR interaction, and that m^{6}A demethylase may function as a direct or indirect regulator for RNA editing.

Moreover, these changes could also modify mRNA interactions with miRNAs. In an indirect role, the activity of ADAR2 could be significantly reduced by miRNA methylation changes such that no editing event would be observed. However, in the direct crosstalk model, after FTO oxidates m^{6}A on dsRNA (even with 24% enzyme activity), ADAR2 could convert adenosine to inosine, which would eventually change the protein structure/function. Also, m^{6}A could be demethylated by FTO or ALKHB5 as an ssRNA and then form a dsRNA and be A-to-I edited by ADAR2. These models can apply to both miRNA and mRNA, but miRNA is more likely to be affected by the direct model in pre-miRNA because it often interacts with molecules in the cell as dsRNA. These models may help explain new mechanisms underlying diseases that previously have been connected to RNA editing, such as cancer and neuronal diseases.

4.1. Splicing

Another nonreversible RNA modification is splicing. Alternative splicing was initially thought to be a rare event, perhaps limited to viruses or only a few organisms, but it is now known to occur in all eukaryotic organisms and likely in all multiexon genes (109). This complexity is extremely important for variable biological function and response, particularly because of the very large number of possible splice junctions. The vast repertoire of splicing forms is beyond the scope of this review, and they have been well summarized elsewhere (113). However, recent deep RNA sequencing experiments have raised questions about the upper limit of complexity for the number of potential transcripts in the human genome, and here we address these questions with a mathematical proof.

For each gene with n exons, the total number of possible combinations of that gene's exons is the sum of the combination of all possible junctions (r), creating a large total number of potential splice variants (V):

$$V = \left(\sum_{1}^{r} \frac{n!}{r!(n-r)!} \right) - 1.$$

Table 1 Splicing complexity by gene size

Exons	Transcript variants	Splice junctions	Total variants
1	1	0	1
2	3	1	4
3	7	3	11
4	15	6	26
5	31	10	57
6	63	15	120
7	127	21	247
8	255	28	502
9	511	36	1,013
10	1,023	45	2,036
50	1.13×10^{15}	1,225	1.13×10^{15}
100	1.27×10^{30}	4,950	1.27×10^{30}
362	9.39×10^{108}	65,341	9.39×10^{108}

The total number of possible splice variants includes the number of variants for each iteration preceding the current one (**Table 1**). For example, a gene with three exons has 7 possible combinations for alternative splicing, including splicing out single exons for an expression event. A gene with four exons has 8 more opportunities, summing to 15 variants total. For five exons, the total is 31, and so on. The number of splice variants then becomes the sum of a row of Pascal's triangle (minus 1 because "no splice form" is not considered a splice form). Therefore, for a gene with n exons, the total number of possible splice variants is $2^n - 1$.

Interestingly, the number of potential transcripts encoded by the human genome is greater than the current estimate of the number of atoms in the universe (**Figure 5**, **Table 1**), totaling approximately 9.4×10^{108} combinations. This comes from the simple combinatorial statistics of the GENCODE (version 18) gene annotation set of the human genome, which contains 57,445 genes. However, at the same time, even the gene with the largest number of exons (*TTN*, $n = 362$) generates a fairly small number of transcripts ($n = 15$), indicating that there are evolutionary and functional constraints that direct the specific synthesis and regulation of very selective transcript isoforms.

Clearly, the number of total potential transcript variants (V) is extremely high, and this is without including variable transcription start or stop sites or alternative promoters. However, the number of possible splice junctions for a given gene with n exons increases much more slowly (**Figure 5**) and also is in accordance with Pascal's mathematical series. This is a factorial addition, where the number of times that a dual, nonreversible union (r, where here $r = 2$ for two exons conjoined) is created with n exons is equal to

$$\frac{n!}{r!\,(n-r)!}.$$

For example, for a gene with five exons,

$$\frac{5!}{2!(5-2)!} = \frac{120}{2 \cdot 6} = \frac{120}{12} = 10 \text{ junctions},$$

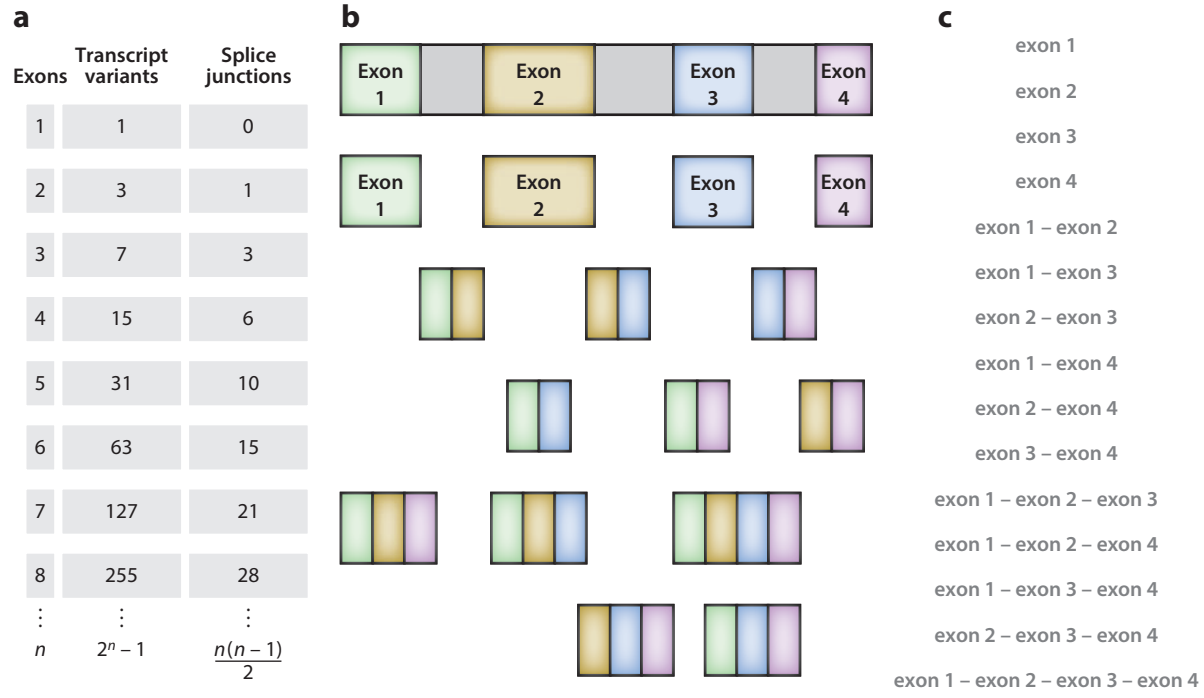

Figure 5

Splicing increases the complexity of the transcriptome encoded by the human genome. (*a*) Given a linear increase in the number of exons in a transcript, the number of transcript combinations increases at a rate of $2^n - 1$, whereas the number of splice junctions increases at a slower rate of $n(n-1)/2$. (*b*) A visual representation of the combinations for a gene with three (*left*) or four (*right*) exons.

and for a gene with six exons,

$$\frac{6!}{2!(6-2)!} = \frac{720}{2 \cdot 24} = \frac{720}{48} = 15 \text{ junctions.}$$

This can also be expressed more simply once one realizes that we are always counting a union of only two exons (making $r = 2$). Thus,

$$\frac{n!}{r!(n-r)!} = \frac{n!}{2!(n-2)!} = \frac{n!}{2 \cdot (n-2)!} = \frac{n \cdot (n-1) \cdot (n-2)!}{2 \cdot (n-2)!} = \frac{n \cdot (n-1) \cdot \cancel{(n-2)!}}{2 \cdot \cancel{(n-2)!}} = \frac{n(n-1)}{2}.$$

So, for any gene with n exons, the number of exon junctions is $n(n-1)/2$.

Because the number of splice junctions is a far more manageable number, and the large potential transcript space is scarcely utilized, this helps demonstrate the importance of having a robust catalog of all splice junctions rather than one of all splice variants.

4.2. Intron Retention

One other notable category of nonreversible RNA modification is intron retention. Recent advances in deep RNA sequencing have allowed an unprecedented look at the frequency and types of intron retention, and their role in disease has also emerged. Specifically, instead of the usual method of enriching for polyadenylated RNAs, protocols now exist that perform rRNA depletion from the sample. This has the benefit of removing the rRNA component that constitutes the

majority of the RNA in a cell while allowing for lower-expression and nonpolyadenylated RNA molecules to be found. However, this also means that pre-mRNAs will be caught when samples are prepared, including transcripts that had not yet completed splicing and processing. Finally, new methods of capturing RNA for targeted RNA expression analysis allow the targeting of specific types of RNA.

These protocols have enabled the discovery of a new landscape of intron retention, noncoding RNAs, and nonpolyadenylated transcripts in normal and cancerous samples. For example, mutations in histone modifiers in human kidney cancer created widespread defects in RNA processing that resulted in significant enrichment of transcripts with intron retention (97). These data show how RNA processing defects can help drive cancer phenotypes. In addition, Mercer et al. (73) showed that thousands of noncoding RNAs are present below the normally detectable levels of current expression measures and that these transcripts played roles in modifying key genes like *p53* and *HOX*. Wong et al. (116) recently showed that intron retention coupled with nonsense-mediated decay is important in the regulation of normal granulocyte differentiation. Physiological intron retention may potentially control the dynamic level of gene expression in an energetically favorable manner prior to sustained gene translation (116). Finally, work from the Sequencing Quality Control Consortium (101) and the Association of Biomolecular Research Facilities study (66) on next-generation RNA sequencing has shown that these ribo-depletion methods can reveal thousands of genes with low expression levels, but the overall agreement with traditional methods for RNA profiling is quite high.

5. SUMMARY

Recent advances in the field of RNA sequencing and epitranscriptome characterization have opened an entirely new layer of regulation and function to study and apply to development and disease. This includes aspects of life as basic as cellular specification and lineage maintenance and those as critical as circadian rhythm and germline development. Also, our TCGA summary data clearly show that the enzymes for RNA modifications are strongly implicated in several types of cancer and provide an exciting new avenue to study and potentially treat these diseases.

Yet much work remains in the field. Although myriad RNA modifications have been cataloged and described, the molecular functions, cellular localizations, and biological roles of these epitranscriptomic changes are essentially unknown for most of the modified ribonucleotides. Also, given the putative impact of these modifications on RNA editing, splicing, and intron retention, a coordinated model of RNA regulation of these factors is emerging that needs to be tested. Finally, the interplay of RNA modifications, epigenetics, and histone modifications further suggests an expanding role of RNA regulation in the epigenome, but work on this is also at an early stage.

Nonetheless, current data support the hypothesis that a dual axis of coordinated regulation and cellular mechanisms exists between the genome and epigenome and between the transcriptome and epitranscriptome, because they often use similar types of substrates in their function and work along common pathways. This returns RNA to a role as a more central mediator of information within the cell, serving as an information carrier, modifier, and attenuator for many biological processes. Future work will likely continue to expand the many roles of RNA and its modifications in the basic function and regulation of the cell, and from these expanded roles, one could argue that we are still living in an "RNA world."

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank the members of the Mason laboratory, specifically Yogesh Saletore and Paul Zumbo, along with Eric Schadt, Jonas Korlach, and Steve Turner for informal discussions. This work was supported by funding from the National Institutes of Health (including R01HG006798 and R01NS076465) as well as by funding from the Irma T. Hirschl and Monique Weill-Caulier Charitable Trusts and the Starr Consortium (I7-A765). We also thank entropy.

LITERATURE CITED

1. Alon S, Eisenberg E. 2013. Identifying RNA editing sites in miRNAs by deep sequencing. *Methods Mol. Biol.* 1038:159–70
2. Alon S, Mor E, Vigneault F, Church GM, Locatelli F, et al. 2012. Systematic identification of edited microRNAs in the human brain. *Genome Res.* 22:1533–40
3. Balik A, Penn AC, Nemoda Z, Greger IH. 2013. Activity-regulated RNA editing in select neuronal subfields in hippocampus. *Nucleic Acids Res.* 41:1124–34
4. Bangs JD, Crain PF, Hashizume T, McCloskey JA, Boothroyd JC. 1992. Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J. Biol. Chem.* 267:9805–15
5. Bass BL, Weintraub H. 1987. A developmentally regulated activity that unwinds RNA duplexes. *Cell* 48:607–13
6. Bass BL, Weintraub H. 1988. An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* 55:1089–98
7. Beadle GW, Tatum EL. 1941. Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci. USA* 27:499–506
8. Benne R, Van den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC. 1986. Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46:819–26
9. Bokar JA, Rath-Shambaugh ME, Ludwiczak R, Narayan P, Rottman F. 1994. Characterization and partial purification of mRNA N^6-adenosine methyltransferase from HeLa cell nuclei: Internal mRNA methylation requires a multisubunit complex. *J. Biol. Chem.* 269:17697–704
10. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, et al. 2013. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494:366–70
11. Burns MB, Temiz NA, Harris RS. 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* 45:977–83
12. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, et al. 2011. The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.* 39:D195–201
13. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, et al. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2:401–4
14. Chambers KF, Bacon JR, Kemsley EK, Mills RD, Ball RY, et al. 2009. Gene expression profile of primary prostate epithelial and stromal cells in response to sulforaphane or iberin exposure. *Prostate* 69:1411–21
15. Chandran UR, Ma C, Dhir R, Bisceglia M, Lyons-Weiler M, et al. 2007. Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. *BMC Cancer* 7:64
16. Chen L. 2013. Characterization and comparison of human nuclear and cytosolic editomes. *Proc. Natl. Acad. Sci. USA* 110:E2741–47
17. Chen L, Li Y, Lin CH, Chan TH, Chow RK, et al. 2013. Recoding RNA editing of *AZIN1* predisposes to hepatocellular carcinoma. *Nat. Med.* 19:209–16
18. Chester A, Somasekaram A, Tzimina M, Jarmuz A, Gisbourne J, et al. 2003. The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *EMBO J.* 22:3971–82
19. Cho DS, Yang W, Lee JT, Shiekhattar R, Murray JM, Nishikura K. 2003. Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *J. Biol. Chem.* 278:17093–102

20. Choudhury Y, Tay FC, Lam DH, Sandanaraj E, Tang C, et al. 2012. Attenuated adenosine-to-inosine editing of microRNA-376a* promotes invasiveness of glioblastoma cells. *J. Clin. Investig.* 122:4059–76
21. Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 12:1–8
22. Church C, Moir L, McMurray F, Girard C, Banks GT, et al. 2010. Overexpression of *Fto* leads to increased food intake and results in obesity. *Nat. Genet.* 42:1086–92
23. Conticello SG. 2008. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* 9:229
24. Cowling VH. 2010. Regulation of mRNA cap methylation. *Biochem. J.* 425:295–302
25. Crain PF, McCloskey JA. 1996. The RNA modification database. *Nucleic Acids Res.* 24:98–99
26. Crain PF, McCloskey JA. 1997. The RNA modification database. *Nucleic Acids Res.* 25:126–27
27. Dahm R. 2008. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* 122:565–81
28. Decatur WA, Fournier MJ. 2003. RNA-guided nucleotide modification of ribosomal and other RNAs. *J. Biol. Chem.* 278:695–98
29. Desrosiers RC, Friderici KH, Rottman FM. 1975. Characterization of Novikoff hepatoma mRNA methylation and heterogeneity in the methylated 5′ terminus. *Biochemistry* 14:4367–74
30. Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, et al. 2013. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503:371–76
31. Dominissini D, Amariglio N, Rechavi G. 2012. Micro-editing mistake translates into a devastating brain tumor. *J. Clin. Investig.* 122:3842–45
32. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, et al. 2012. Topology of the human and mouse m^6A RNA methylomes revealed by m^6A-seq. *Nature* 485:201–6
33. Dubin DT, Taylor RH. 1975. The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic Acids Res.* 2:1653–68
34. Ekdahl Y, Farahani HS, Behm M, Lagergren J, Ohman M. 2012. A-to-I editing of microRNAs in the mammalian brain increases during development. *Genome Res.* 22:1477–87
35. Elkon R, Ugalde AP, Agami R. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* 14:496–506
36. Figueroa ME, Wouters BJ, Skrabanek L, Glass J, Li Y, et al. 2009. Genome-wide epigenetic analysis delineates a biologically distinct immature acute leukemia with myeloid/T-lymphoid features. *Blood* 113:2795–804
37. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, et al. 2007. A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316:889–94
38. Fu Y, He C. 2012. Nucleic acid modifications with epigenetic significance. *Curr. Opin. Chem. Biol.* 16:516–24
39. Fu Y, Jia G, Pang X, Wang RN, Wang X, et al. 2013. FTO-mediated formation of N^6-hydroxymethyladenosine and N^6-formyladenosine in mammalian RNA. *Nat. Commun.* 4:1798
40. Fustin JM, Doi M, Yamaguchi Y, Hida H, Nishimura S, et al. 2013. RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell* 155:793–806
41. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, et al. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6:pl1
42. Gao M, Fritz DT, Ford LP, Wilusz J. 2000. Interaction between a poly(A)-specific ribonuclease and the 5′ cap influences mRNA deadenylation rates in vitro. *Mol. Cell* 5:479–88
43. García-López J, Hourcade JDD, Del Mazo J. 2013. Reprogramming of microRNAs by adenosine-to-inosine editing and the selective elimination of edited microRNA precursors in mouse oocytes and preimplantation embryos. *Nucleic Acids Res.* 41:5483–93
44. Garrett S, Rosenthal JJ. 2012. RNA editing underlies temperature adaptation in K^+ channels from polar octopuses. *Science* 335:848–51
45. Guenther UP, Yandek LE, Niland CN, Campbell FE, Anderson D, et al. 2013. Hidden specificity in an apparently nonspecific RNA-binding protein. *Nature* 502:385–88

46. He H, Jazdzewski K, Li W, Liyanarachchi S, Nagy R, et al. 2005. The role of microRNA genes in papillary thyroid carcinoma. *Proc. Natl. Acad. Sci. USA* 102:19075–80
47. Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. 1984. Mapping of N^6-methyladenosine residues in bovine prolactin mRNA. *Proc. Natl. Acad. Sci. USA* 81:5667–71
48. HsuChen CC, Dubin DT. 1976. Di-and trimethylated congeners of 7-methylguanine in Sindbis virus mRNA. *Nature* 264:190–91
49. Iwanami Y, Brown GM. 1968. Methylated bases of ribosomal ribonucleic acid from HeLa cells. *Arch. Biochem. Biophys.* 126:8–15
50. Iwanami Y, Brown GM. 1968. Methylated bases of transfer ribonucleic acid from HeLa and L cells. *Arch. Biochem. Biophys.* 124:472–82
51. Jia G, Fu Y, He C. 2013. Reversible RNA adenosine methylation in biological regulation. *Trends Genet.* 29:108–15
52. Jia G, Fu Y, Zhao X, Dai Q, Zheng G, et al. 2011. N^6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7:885–87
53. Jia G, Yang CG, Yang S, Jian X, Yi C, et al. 2008. Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett.* 582:3313–19
54. Johnson TB, Coghill RDW. 1925. Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *J. Am. Chem. Soc.* 47:2838–44
55. Jones KB, Salah Z, Del Mare S, Galasso M, Gaudio E, et al. 2012. miRNA signatures associate with pathogenesis and progression of osteosarcoma. *Cancer Res.* 72:1865–77
56. Kawahara Y. 2012. Quantification of adenosine-to-inosine editing of microRNAs using a conventional method. *Nat. Protoc.* 7:1426–37
57. Kawahara Y, Zinshteyn B, Chendrimada TP, Shiekhattar R, Nishikura K. 2007. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Rep.* 8:763–69
58. Kellinger MW, Song CX, Chong J, Lu XY, He C, Wang D. 2012. 5-Formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 19:831–33
59. Kiran AM, Baranov PV. 2010. DARNED: a DAtabase of RNa EDiting in humans. *Bioinformatics* 26:1772–76
60. Kiran AM, O'Mahony JJ, Sanjeev K, Baranov PV. 2013. Darned in 2013: inclusion of model organisms and linking with Wikipedia. *Nucleic Acids Res.* 41:D258–61
61. Kurowski MA, Bhagwat AS, Papaj G, Bujnicki JM. 2003. Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics* 4:48
62. Lagana A, Paone A, Veneziano D, Cascione L, Gasparini P, et al. 2012. miR-EdiTar: a database of predicted A-to-I edited miRNA target sites. *Bioinformatics* 28:3166–68
63. Laurencikiene J, Kallman AM, Fong N, Bentley DL, Ohman M. 2006. RNA editing and alternative splicing: the importance of co-transcriptional coordination. *EMBO Rep.* 7:303–7
64. Lavrov DV, Brown WM, Boore JL. 2000. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. *Proc. Natl. Acad. Sci. USA* 97:13738–42
65. Li M, Wang IX, Li Y, Bruzel A, Richards AL, et al. 2011. Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 333:53–58
66. Li S, Tighe SW, Nicolet CM, Grove D, Levy S, et al. 2014. Multi-platform and cross-methodological reproducibility of transcriptome profiling by RNA-seq in the ABRF Next-Generation Sequencing Study (ABRF-NGS). *Nat. Biotechnol.* In press
67. Limbach PA, Crain PF, McCloskey JA. 1994. Summary: the modified nucleosides of RNA. *Nucleic Acids Res.* 22:2183–96
68. Liou RF, Blumenthal T. 1990. *Trans*-spliced *Caenorhabditis elegans* mRNAs retain trimethylguanosine caps. *Mol. Cell. Biol.* 10:1764–68
69. Liu J, Yue Y, Han D, Wang X, Fu Y, et al. 2014. A METTL3–METTL14 complex mediates mammalian nuclear RNA N^6-adenosine methylation. *Nat. Chem. Biol.* 10:93–95

70. Liu N, Parisien M, Dai Q, Zheng G, He C, Pan T. 2013. Probing N^6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* 19:1848–56
71. Marín-Béjar O, Marchese FP, Athie A, Sánchez Y, González J, et al. 2013. *Pint* lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biol.* 14:R104
72. Mayr C, Bartel DP. 2009. Widespread shortening of 3′ UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138:673–84
73. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, et al. 2012. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* 30:99–104
74. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. 2012. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* 149:1635–46
75. Mower JP. 2008. Modeling sites of RNA editing as a fifth nucleotide state reveals progressive loss of edited sites from angiosperm mitochondria. *Mol. Biol. Evol.* 25:52–61
76. Nemlich Y, Greenberg E, Ortenberg R, Besser MJ, Barshack I, et al. 2013. MicroRNA-mediated loss of ADAR1 in metastatic melanoma promotes tumor growth. *J. Clin. Investig.* 123:2703–18
77. Nishikura K. 2010. Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* 79:321–49
78. Ota H, Sakurai M, Gupta R, Valente L, Wulff BE, et al. 2013. ADAR1 forms a complex with Dicer to promote microRNA processing and RNA-induced gene silencing. *Cell* 153:575–89
79. Park E, Williams B, Wold BJ, Mortazavi A. 2012. RNA editing in the human ENCODE RNA-seq data. *Genome Res.* 22:1626–33
80. Pasquinelli AE. 2012. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.* 13:271–82
81. Paz-Yaacov N, Levanon EY, Nevo E, Kinar Y, Harmelin A, et al. 2010. Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. *Proc. Natl. Acad. Sci. USA* 107:12174–79
82. Peng Z, Cheng Y, Tan BC, Kang L, Tian Z, et al. 2012. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.* 30:253–60
83. Penn AC, Balik A, Greger IH. 2013. Steric antisense inhibition of AMPA receptor Q/R editing reveals tight coupling to intronic editing sites and splicing. *Nucleic Acids Res.* 41:1113–23
84. Perry RP, Kelley DE, Friderici K, Rottman F. 1975. The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5′ terminus. *Cell* 4:387–94
85. Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J. 1987. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* 50:831–40
86. Ramaswami G, Li JB. 2014. RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* 42:D109–113
87. Rebagliati MR, Melton DA. 1987. Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity. *Cell* 48:599–605
88. Rieder LE, Staber CJ, Hoopengardner B, Reenan RA. 2013. Tertiary structural elements determine the extent and specificity of messenger RNA editing. *Nat. Commun.* 4:2232
89. Rueter SM, Dawson TR, Emeson RB. 1999. Regulation of alternative splicing by RNA editing. *Nature* 399:75–80
90. Saletore Y, Chen-Kiang S, Mason CE. 2013. Novel RNA regulatory mechanisms revealed in the epitranscriptome. *RNA Biol.* 10:342–46
91. Saletore Y, Meyer K, Korlach J, Vilfan ID, Jaffrey S, Mason CE. 2012. The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* 13:175
92. Sanchez-Palencia A, Gomez-Morales M, Gomez-Capilla JA, Pedraza V, Boyero L, et al. 2011. Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer. *Int. J. Cancer.* 129:355–64
93. Scadden AD. 2005. The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat. Struct. Mol. Biol.* 12:489–96
94. Schaefer M, Pollex T, Hanna K, Lyko F. 2009. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.* 37:e12
95. Schibler U, Kelley DE, Perry RP. 1977. Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. *J. Mol. Biol.* 115:695–714

95a. Schwartz S, Agarwala SD, Mumbach MR, Jovanovic M, Mertins P, et al. 2014. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* 155:14091–21
96. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, et al. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–45
97. Simon JM, Hacker KE, Singh D, Brannon AR, Parker JS, et al. 2014. Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. *Genome Res.* 24:241–50
98. Smith RM, Webb A, Papp AC, Newman LC, Handelman SK, et al. 2013. Whole transcriptome RNA-Seq allelic expression in human brain. *BMC Genomics* 14:571
99. Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, et al. 2012. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* 40:5023–33
100. St Laurent G, Tackett MR, Nechkin S, Shtokalo D, Antonets D, et al. 2013. Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in *Drosophila*. *Nat. Struct. Mol. Biol.* 20:1333–39
101. Su Z, Łabaj P, Li S, Thierry-Mieg J, Thierry-Mieg D, et al. 2014. Power and limitations of RNA-seq: findings from the SEQC (MAQC-III) consortium. *Nat. Biotechnol.* In press
102. Traka M, Gasper AV, Melchini A, Bacon JR, Needs PW, et al. 2008. Broccoli consumption interacts with *GSTM1* to perturb oncogenic signalling pathways in the prostate. *PLoS ONE* 3:e2568
103. Tschudi C, Ullut E. 2002. Unconventional rules of small nuclear RNA transcription and cap modification in trypanosomatids. *Gene Expr.* 10:3–16
104. Varambally S, Yu J, Laxman B, Rhodes DR, Mehra R, et al. 2005. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* 8:393–406
105. Veliz EA, Easterwood LM, Beal PA. 2003. Substrate analogues for an RNA-editing adenosine deaminase: mechanistic investigation and inhibitor design. *J. Am. Chem. Soc.* 125:10867–76
106. Vicente-Dueñas C, Fontán L, Gonzalez-Herrero I, Romero-Camarero I, Segura V, et al. 2012. Expression of *MALT1* oncogene in hematopoietic stem/progenitor cells recapitulates the pathogenesis of human lymphoma in mice. *Proc. Natl. Acad. Sci. USA* 109:10534–39
107. Vilfan ID, Tsai YC, Clark TA, Wegener J, Dai Q, et al. 2013. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J. Nanobiotechnol.* 11:8
108. Wagner RW, Smith JE, Cooperman BS, Nishikura K. 1989. A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and *Xenopus* eggs. *Proc. Natl. Acad. Sci. USA* 86:2647–51
109. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, et al. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456:470–76
110. Wang IX, So E, Devlin JL, Zhao Y, Wu M, Cheung VG. 2013. ADAR regulates RNA editing, transcript stability, and gene expression. *Cell Rep.* 5:849–60
111. Wang Q, Hui H, Guo Z, Zhang W, Hu Y, et al. 2013. ADAR1 regulates *ARHGAP26* gene expression through RNA editing by disrupting miR-30b-3p and miR-573 binding. *RNA* 19:1525–36
112. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, et al. 2014. N^6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505:117–20
112a. Wang Y, Li Y, Toth JI, Petrosky MD, Zhang Z, Zhao JC. 2014. N^6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* 16:191–98
113. Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14:802–13
114. Wei CM, Gershowitz A, Moss B. 1975. Methylated nucleotides block 5′ terminus of HeLa cell messenger RNA. *Cell* 4:379–86
115. Woese CR. 1967. *The Genetic Code: The Molecular Basis for Genetic Expression*. New York: Harper and Row
116. Wong JJ, Ritchie W, Ebner OA, Selbach M, Wong JW, et al. 2013. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* 154:583–95
117. Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, et al. 2014. Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* 11:41–46

118. Xhemalce B, Robson SC, Kouzarides T. 2012. Human RNA methyltransferase BCDIN3D regulates microRNA processing. *Cell* 151:278–88
119. Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, et al. 2006. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat. Struct. Mol. Biol.* 13:13–21
120. Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, et al. 2013. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* 49:18–29