Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

CellPress

Trends in Biotechnology

Review

New Twists in Detecting mRNA Modification Dynamics

Ina Anreiter,[1,2] Quoseena Mir,[3] Jared T. Simpson,[1,2] Sarath C. Janga,[3,4,5] and Matthias Soller[6,*,@]

Modified nucleotides in mRNA are an essential addition to the standard genetic code of four nucleotides in animals, plants, and their viruses. The emerging field of epitranscriptomics examines nucleotide modifications in mRNA and their impact on gene expression. The low abundance of nucleotide modifications and technical limitations, however, have hampered systematic analysis of their occurrence and functions. Selective chemical and immunological identification of modified nucleotides has revealed global candidate topology maps for many modifications in mRNA, but further technical advances to increase confidence will be necessary. Single-molecule sequencing introduced by Oxford Nanopore now promises to overcome such limitations, and we summarize current progress with a particular focus on the bioinformatic challenges of this novel sequencing technology.

Highlights

Writers, readers, and erasers have now been discovered for many mRNA modifications.

Global topographic candidate maps have been generated for many modifications, but high error rates need to be addressed by technical improvements in detection and validation using orthogonal methods that apply rigid selection criteria.

Nanopore single-molecule direct RNA sequencing is progressing towards reliable detection of modified nucleotides in mRNA.

Epitranscriptomics – A Rapidly Developing Field

Chemical modifications on RNA are well-established and evolutionarily conserved features of structural RNAs such as rRNA and tRNA [1–3]. In the past few years the occurrence of these modifications on protein-coding mRNAs, long noncoding RNAs (lncRNAs), and small regulatory RNAs (srRNAs) has received renewed attention to determine their role in regulating gene expression, development, and health and disease (Box 1). The rapid evolution of transcriptome sequencing technologies has made it possible to develop methodologies that interrogate the topography of RNA modifications transcriptome-wide. This new field, termed **epitranscriptomics** (see Glossary), seeks to elucidate the role of RNA modifications in regulating gene expression, with a special focus on their biological functions in mRNA.

Major recent methodological advances in mass spectrometry (MS) and next-generation sequencing combined with immunoprecipitation and chemical or enzymatic conversion methods have increased the catalog of known mRNA modifications, and have led to new insights into the role of these modifications in regulating gene expression in humans and model organisms such as yeast, plants, *Drosophila*, and mice. However, the field still faces considerable methodological challenges because modifications in mRNA generally are not abundant. Moreover, many current methods for probing RNA modifications are hampered by high error rates, low specificity, and poor reproducibility. We summarize below current approaches and emerging technologies for assessing mRNA modifications. We aim to highlight the strengths and limitations of current methods regarding specificity, sensitivity, and reproducibility, with a particular focus on emerging single-molecule direct RNA sequencing by Nanopore.

Current Catalog of Modifications in mRNA

To date, 13 different chemical modifications have been identified in mRNA transcripts, and these can be divided into modifications of cap-adjacent nucleotides and internal modifications (Figure 1). These modifications are added by a variety of dedicated enzymes (Box 2 and Table 1). Modifications of cap-adjacent nucleotides are added to the 5′-ends of RNAs transcribed

[1]Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada
[2]Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada
[3]Department of BioHealth Informatics, School of Informatics and Computing, Indiana University–Purdue University Indianapolis, Indianapolis, IN 46202, USA
[4]Department of Medical and Molecular Genetics, Medical Research and Library Building, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[5]Center for Computational Biology and Bioinformatics, 5021 Health Information and Translational Sciences, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[6]School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

*Correspondence: m.soller@bham.ac.uk (M. Soller).
@Twitter: @MatthiasSoller

© 2020 Elsevier Ltd. All rights reserved.

Check for updates

Box 1. Discovery of RNA Modifications and Their Biological Functions

In 1957, Davis and Allen discovered pseudouridine as the first known chemically modified RNA nucleotide [121]. Since then, 162 more RNA modifications have been described and catalogued (RNA Modification Databases: http://mods.rna.albany.edu and http://modomics.genesilico.pl [115]). RNA modifications occur in all domains of life and in all types of RNA (tRNA, rRNA, mRNA, srRNA, snoRNA, snRNA, and lncRNA), and have been extensively linked to development, health, and disease. The high copy number of rRNAs and tRNAs in cells has greatly facilitated the study of modifications in these RNA species. rRNAs and tRNAs have complex 3D structures, which mediate ribosome function and protein translation. Modified nucleotides in rRNAs and tRNAs play essential roles in ribosome assembly and dynamics, and disruption of these modifications has been associated with lethality, severe growth defects, intellectual disability, diabetes, and cancer [32,33,122]. Although nucleotide modifications in mRNA have been known for >40 years, their functional interrogation has only become possible in the past decade owing to high-throughput sequencing-based mapping methodologies that overcome the low abundance of these modifications, as well as to the discovery of writer, reader, and eraser proteins (Box 2). Since then, may biological functions in development and disease have been associated to mRNA modifications, including a variety of cancers, mental disorders, and fertility and metabolic phenotypes [4,30,32,33].

by RNA polymerase II [mRNA, primary (pri-)miRNA transcript, lncRNA, small nucleolar (sno)RNA, and small nuclear (sn)RNA] [4]. The composition of the cap varies with the type of RNA molecule, and typically consists of a 7-methylguanosine (m^7G) moiety added in a characteristic 5′–5′ triphosphate linkage to the first transcribed nucleotide (Figure 1A). In some snRNAs and snoRNAs the cap guanosine is trimethylated to $m^{2'2'7}G$, and further alternative cap structures including NAD^+ have recently been identified [4–6]. The first and second nucleotides adjacent to the cap can be 2′-*O*-methylated at the ribose (cOMe) in animals, viruses, and protists. If the cap-adjacent nucleotide is an adenosine, it can be further methylated to N^6,2′-*O*-dimethyladenosine (m^6Am, Figure 1A) [4,7–11]. Of further special note are the extensive modifications in trypanosomes of cap-adjacent nucleotides in the splice leader RNAs that are *trans*-spliced onto the body of the mRNAs (Figure 1A) [12]. Modifications of cap-adjacent nucleotides are common, differ among tissues and transcripts, and regulate mRNA stability and translation [4,5,7,8,10,11,13].

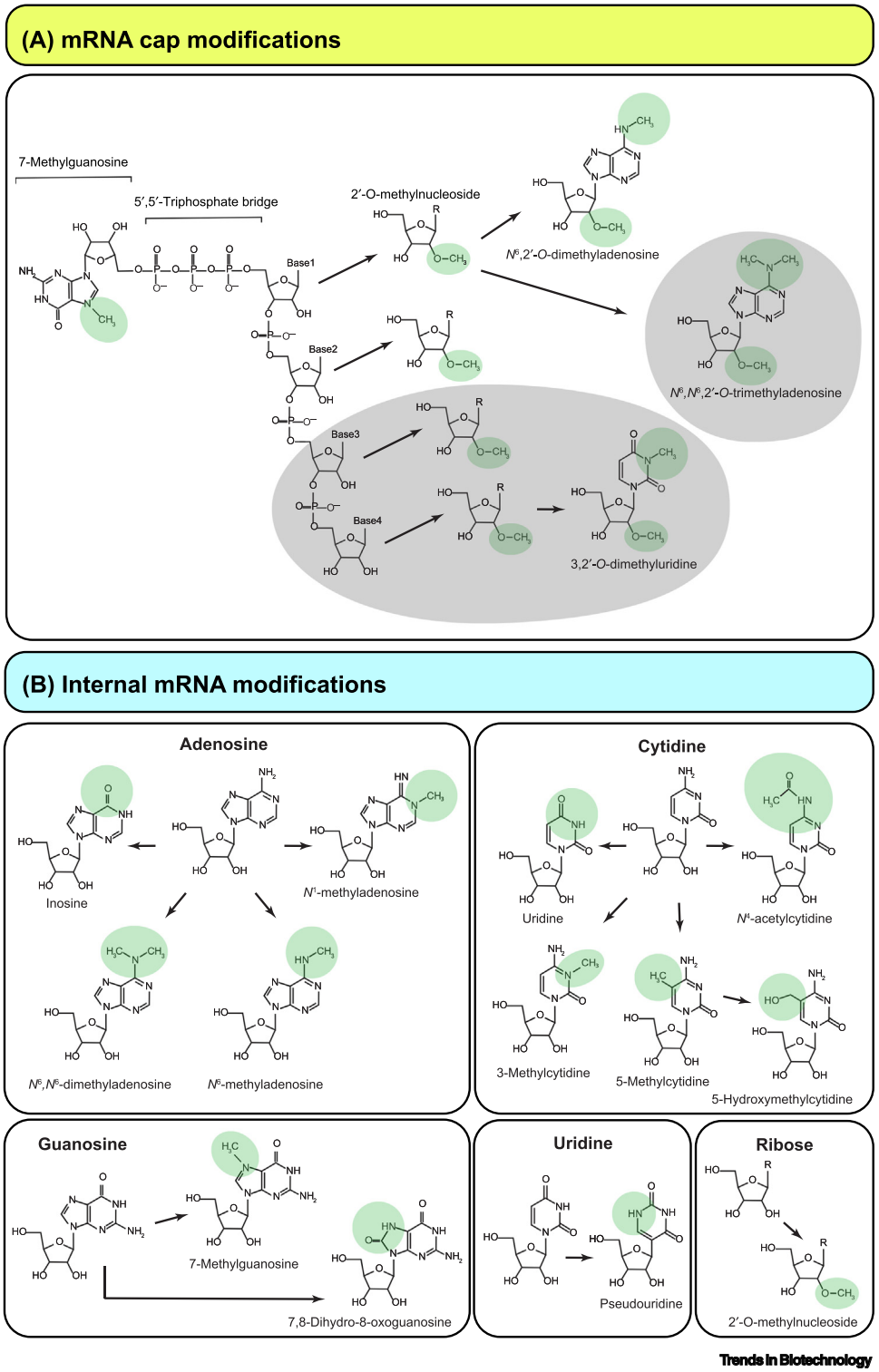
Internal modifications occur in 5′ and 3′ untranslated regions (UTRs), coding regions, and introns of mRNAs (Figure 1B). Mass spectrometric (MS) quantification indicates that N^6-methyladenosine (m^6A) and adenosine to inosine (A to I) editing are the most abundant internal mRNA modifications [14]. Less abundant internal mRNA modifications are 5-methylcytosine (m^5C) [15], pseudouridine (Ψ) [16–19], N^1-methyladenosine (m^1A) [20,21], N^4-acetylcytidine (ac^4C) [22], hydroxymethylcytosine (hm^5C) [23], 3-methylcytidine (m^3C) [24], cytosine to uridine (C to U) editing [25], m^7G [26,27], Nm [28], and 7,8-dihydro-8-oxoguanosine (8-oxoG, as a result of guanosine oxidation by reactive oxygen species) [29].

Internal mRNA modifications are involved in a variety of gene-regulatory functions in the cell, where m^6A is thought to be the most diverse because of the presence of multiple *bona fide* readers of the **YTH protein** domain family (Box 2). Internal m^6A can regulate all aspects of gene expression of an mRNA, including mRNA splicing, 3′-end processing, export, stability, localization, and translation [30–35]. Other internal modifications tend to be more process-specific and can affect translation efficiency (e.g., inosine [36], m^1A [37], ac^4C [22], m^6Am [9], m^3C [24]), codon recoding (e.g., inosine [36], 8-oxoG [29]), mRNA stability (e.g., ac^4C [22]), and nuclear export (e.g., m^6A [38]). Among these mRNA modifications, m^6A, m^6Am, and m^1A have been proposed to function dynamically because they can be reversed by eraser proteins [13,30,39,40] (Box 2).

Although the abundance of mRNA modifications are generally low, with a prevalence below 0.5% of all nucleotides [14,24,26], they play important functional roles in mRNA regulation, cell function, development, the immune system, and homeostasis (Box 1). In addition, many mRNA modifications play key roles in viral infection and replication, and in response to stress [4,25,34,35]. For

Glossary

AlkAniline-Seq: RNA alkaline hydrolysis and aniline cleavage sequencing.
ALKBH5: ALKB homolog 5, an m^6A mRNA demethylase that removes the methyl group associated with N^6-methylated adenosine
Epitranscriptomics: the study of transcriptome-wide distribution and function of RNA modifications.
FTO protein: fat mass and obesity associated protein, an m^6A mRNA demethylase that removes the methyl group of N^6-methylated adenosine at the first position after the cap.
LAIC-seq: m^6A-level and isoform-characterization sequencing.
m^1A-ID-seq: meRIP for m^1A that leverages the inherent stalling of reverse transcriptases at modified nucleotides to give improved resolution.
m^1A-MAP: misincorporation-assisted profiling of m^1A.
m^6A-REF-seq: m^6A-sensitive RNA endoRNase-facilitated sequencing.
MAZTER-seq: RNA digestion via m^6A-sensitive RNase and sequencing.
meRIP-seq: methylated RNA immunoprecipitation and sequencing.
METTL3: methyltransferase-like 3, the catalytic subunit of the m^6A methyltransferase complex that is involved in m^6A deposition on mRNA, also named MTA-70, MTA, and IME4.
METTL14: methyltransferase-like 14, a catalytically inactive component of the m^6A methyltransferase complex that serves as a support protein for METTL3.
miCLIP-seq: methylated individual nucleotide resolution crosslinking immunoprecipitation and sequencing.
NsunC271A-CLIP: miCLIP for 5-methylcytosine (m^5C) using covalent binding of the methyltransferase Nsun2 to its RNA targets.
PA-m^6A-seq: photo-crosslinking-assisted m^6A sequencing
RiboMethSeq: ribose methylation sequencing.
RibOxi-seq: site-specific identification of 2′-*O*-methylation sites using ribose oxidation sequencing.
RNMT: mRNA cap guanine-7 methyltransferase, an enzyme that methylates the inverted guanosine at the mRNA cap to m^7G.
TRIBE: targets of RNA-binding proteins identified by editing.
YTH proteins: a class of proteins that specifically recognize m^6A in mRNA.



(See figure legend at the bottom of the next page.)

Box 2. Writers, Readers, and Erasers: The Complex Machinery Underlying mRNA Modifications

The enzymes that deposit, remove, and bind to mRNA modifications have been termed 'writers', 'erasers', and 'readers', respectively [31,34,35] (see Table 1 in main text). Writers bind to short consensus sequence motifs in target mRNAs; however, these motifs are far more common than the modifications themselves, suggesting that additional factors determine writer binding. Other writers, such as CMTrs, PUS1, and METTL16, bind to mRNA based on mRNA structure motifs [31,123]. Erasers and readers recognize the modifications themselves; however, given the large variety of different reader proteins, additional factors are probably involved in targeting readers to their mRNA targets. The machinery of writers, erasers, and readers is complex, often consisting of large protein complexes with a variety of cofactors. The writer complex for m^6A, for instance, consists of a ~900 kDa complex with two core methyltransferases (METTL3 and **METTL14**) and several auxiliary cofactors [31]. Although loss of the methyltransferases does not result in lethality in *Drosophila*, loss of several of the cofactors does, suggesting that these enzymes exercise alternative functions unrelated to m^6A-methylation. Some writer proteins can also act as reader proteins, as is the case for METTL16 under particular cellular conditions [31]. In addition, some modifications have alternative writers, erasers, and readers. METTL3/METTL14, METTL4, and METTL16 writers for m^6A act in different complexes, and both FTO and **ALKBH5** have been shown to erase m^6A [31]. A varying number of m^6A readers have been identified across organisms, with a record of 13 different readers in plants [31]. There is also crosstalk between modifications, and some enzymes target more than one type of modification. For instance, FTO was first described as an eraser for internal m^6A [124], but was later shown to also demethylate m^6Am at the cap [125]. Furthermore, mRNA modifications might influence the ability of mRNA binding proteins to bind to mRNA targets. For instance, interferon-induced proteins with tetratricopeptide repeats (IFITs), mRNA binding proteins that are involved in innate immunity and viral response, show low affinity for mRNA transcripts with ribose methylation at the cap [47–49]. Whether alternative writers, erasers, and readers of mRNA modifications act in different tissues, in different cellular/physiological contexts, or on different mRNA targets remains mostly an open question. Furthermore, although the machinery acting on m^6A has been well characterized in recent years, the factors acting on other modifications are less well explored, and new enzymes or functions could be discovered in years to come. Further identification of these enzymes will aid our understanding of the biological functions of mRNA modifications and yield prime new targets for pharmaceutical targeting of mRNA modification-associated diseases such as cancer and obesity.

instance, m^6A and/or m^6Am levels are altered in specific brain regions of acutely stressed mice and by glucocorticoid administration [41]. Human patients with major depressive disorder have lower m^6A and/or m^6Am levels in their blood and do not respond to glucocorticoid stimulation [41]. Human cells under serum starvation and peroxide treatment show stress-specific induced m^1A sites [20]. Internal mRNA m^7G is enriched in coding sequences (CDS) and 3′-UTRs, and is depleted in 5′-UTRs in mammalian cells and brain tissues in response to heat shock and oxidative stress [42]. Similar responses to stress have been described for Ψ [17,19], 8-oxoG [29], and C to U editing [43].

Another important function of mRNA modifications resides in the immune system. Of particular importance here is A to I editing, which in vertebrates tunes down the autoimmune response to double-stranded RNA [44,45]. Furthermore, 2′-*O*-ribose methylation of cap-adjacent nucleotides (cOMe) has been proposed as a mechanism to distinguish self from non-self mRNAs because it repels binding of interferon-induced proteins with tetratricopeptide repeats (IFITs) and RIG-I, the

Figure 1. Catalog of Currently Known mRNA Modifications. (A) RNA modifications at the mRNA cap. The first transcribed base (base 1) is linked first to an inverted guanosine by a 5′–5′-triphosphate linkage and subsequently becomes methylated to 7-methylguanosine (m^7G) by RNMT (RNA guanine 7-methyltransferase). The first and second nucleotides of the mRNA can be ribose 2′-*O*-methylated (cOMe) by the cap methyltransferases CMTr1 and CMTr2. If base 1 is a ribose 2′-*O*-methylated adenosine it can be further methylated to N^6,2′-*O*-dimethyladenosine (m^6Am) by PCIF1 (phosphorylated CTD-interacting factor 1). In the splice leader sequence added to all mRNAs in trypanosomes (shaded in grey) the third and fourth nucleotides are also 2′-*O*-ribose methylated, the first adenosine is trimethylated to N^6N^6,2′-*O*-trimethyladenosine ($m^6{}_2Am$) instead of m^6Am, and the fourth base is methylated to 3,2′-*O*-dimethyluridine. (B) Internal mRNA nucleotides carry different modifications depending on the nucleotide (Box 2 for enzymes). Adenosine can be converted to inosine by adenosine deaminases (ADARs), or methylated to N^1-methyladenosine (m^1A), N^6-methyladenosine (m^6A), or N^6,N^6-dimethyladenosine. Cytidine can be converted to uridine (U), acetylated N^4-acetylcytidine (ac^4C), or methylated to 3-methylcytidine (m^3C) or 5-methylcytidine (m^5C). m^5C can be converted to 5-hydroxymethylcytidine (hm^5C). Guanosine can be methylated to 7-methylguanosine (m^7G) or oxidized to 7,8-dihydro-8-oxoguanosine (8-oxoG). Uridine can be converted to pseudouridine (Ψ). Finally, the ribose sugars of all nucleotides can be 2′-*O*-methylated (Nm).

Table 1. Mammalian Writers, Readers, and Erasers of mRNA Modifications

mRNA modification	Writer	Eraser	Reader
Cap modifications			
m^7G	**RNMT** [4]		
cOMe	CMTr1, CMTr2 [4,122]		
m^6Am	PCIF1[a] [7,10,11]	FTO [125]	
Internal modifications			
m^6A	METTL3[b], METTL14[b] METTL4[c], METTL16 [31]	FTO, ALKBH5 [31]	YTHDF1–3, YTHDC1–2 [31]
m^1A	TRMT6/61A, TRMT61B [37,63]	ALKBH1/3 [37]	YTHDF1–3 [139]
m^5C	NSUN2 [73], DNMT2[c] [140]	TET (through conversion to hm^5C) [141]	YBX1 [142–145]
hm^5C	TET1–3[c,d] [141]	TET (through further oxidation)	
m^3C	METTL8 [24]	ALKBH1 [146]	
ac^4C	NAT10 [22]		
Ψ	DKC1, TRUB1, PUS7 [32,81]		
m^7G	METTL1 [26,59]		
8-OxoG	Nonenzymatic [29]		
Nm	FTSJ3[c] [147]		
A to I	ADAR 1–3 [148]		
C to U	APOBEC family [25]		

[a]Requires ribose methylation introduced by CMTr to methylate adenosine the at N^6 position.
[b]Part of a 900 kDa holoenzyme.
[c]Remains to be confirmed for mRNA.
[d]Requires m^5C.

main inducers of antiviral and inflammatory responses [46–49]. However, cOMe likely has other functions because only a fraction of host mRNAs carry cOMe, and levels differ between tissues and transcripts in mice [8,13]. Moreover, IFIT1 in some viruses does not inhibit translation of non-ribose methylated transcripts [50].

Current Methods for Global Assessment of mRNA Modifications

Although many recent studies outline the importance of mRNA modifications, transcriptome-wide mapping of modifications still faces significant challenges. Because the abundance of mRNA modifications is generally low and variable (in the range 5–88% at a given site or transcript [51,52], and only 2–5% of cellular RNA is mRNA), current methods (Table 2) require large amounts of mRNAs and/or very deep sequencing. Furthermore, current methods have yielded highly variable results, and suffer from many false positives and lack of specificity and/or single-base resolution.

Assessing the Presence of Modifications Using MS

Liquid chromatography coupled to tandem MS (LC-MS/MS) has been widely applied to detect and quantify the relative abundances of RNA modifications [5,14,26,27,53,54]. Generally, the approach involves digesting total RNA or purified mRNA into individual nucleotides, followed by separation by LC and quantification by MS. The presence and amount of all nucleotide modifications in an RNA sample can then be assessed by comparison of the MS peaks from the sample with

Table 2. Methods To detect Transcript-Specific mRNA Modifications Genome-Wide by Short-Read Sequencing

Method name	mRNA modification	Method principle	Refs
meRIP-seq, miCLIP-seq, PA-m6A-seq, m1A-ID-seq, m1A-MAP, LAIC-seq	m^6A, m^1A, m^7G, m6Am hm^5C, ac^4C, m^5C[a]	Immunoprecipitation with modification-specific antibody	[9,20–23,26,37,40,52,61–63,67,70,149,150]
m^6A-label-seq	m^6A	Metabolic labeling to N^6-allyladenosine for further chemical modification	[68]
Writer-CLIP, Reader-CLIP, DARTseq, TRIBE	All modifications	Crosslinking or APOBEC or ADAR editing nearby modifications	[91,92,94]
mRNA editing	A to I and C to U mRNA editing	Variant analysis (I reads as G)	[72]
Bs-seq, RNA-BisSeq	m^5C	Deamination of nonmethylated cytosines to uridines using bisulfite	[15,73,74]
Aza-IP, NsunC271A-CLIP	m^5C	Crosslinking	[75–77]
AlkAniline-Seq	m^7G, m^3C	Alkaline hydrolysis and aniline cleavage at modified sites	[71]
m^7G-MaP-seq	m^7G	Reduction of modified sites to abasic sites	[60]
m^6A-SEAL-seq	m^6A	FTO-assisted oxidation to hm^6A for biotin labeling to IP	[90]
Nm-seq, RibOxi-seq, RiboMethSeq, 2′-OMe-Seq	Nm	Iterative removal of nonmethylated bases by oxidation and dephosphorylation	[28,78–80]
Ψ-seq, pseudo-seq, CeU-seq	Pseudouridine	Chemical modification with CMC, blocking reverse transcription at modified nucleotides	[16,17,19]
MAZTER-seq, m^6A-REF-seq	m^6A	Restriction with endonucleases that cut methylated but not unmethylated nucleotides	[88,89]

[a]Only applied to unicellular organisms so far.

the MS peaks of standards [5,14,26,53]. Provided that the nucleotides can be sufficiently separated, LC-MS/MS measurements are quantitative and concordant between studies even if the levels of the modified nucleotides are low [14,26,27,53]. This approach has also been combined with enzymatic digestion by nucleases which have differential activity towards cap and internal m^7G modifications to demonstrate the prevalence of internal m^7G in mRNA [27] or enrichment of caps by ion-exchange (CapQuant and CAP-MAP) [5,12,55]. However, LC-MS/MS requires large amounts of input RNA (>0.65 µg polyA mRNA for internal modifications, and >20 µg polyA mRNA for modifications of cap-adjacent nucleotides) and is limited for low-abundance nucleotides such as caps (100- to 10 000-fold less sensitive than radioactive labeling) [5,12,14,27]. In addition, oligo-dT-purified polyA mRNA carries over rRNA, leading to misannotations [56].

LC-MS/MS does not provide information about the location of the modification in a transcript, but this can be circumvented by using oligonucleotides obtained by digestion with different RNases or from tRNAs [54,57,58]. Digestion with RNase, however, can introduce errors in mapping fragments to the genome. For example, the m^7G site in let-7 miRNA is in fact a fragment of rRNA with a well-known Gm site [59,60]. Although the complement of modifications has been mapped in cellular tRNAs, the amount of input and the computational challenges are significant [54,57].

Antibody-Based Methods To Detect Modifications

The most widely used methods for transcriptome-wide mapping of mRNA modifications rely on RNA immunoprecipitation (RIP) by commercially available antibodies that recognize modified nucleotides (Box 2), but the specificity of these antibodies varies. For profiling transcripts carrying modifications, RIP is generally followed by whole-transcriptome sequencing (RIP-seq or **meRIP-seq**) to identify modified regions ~100–200 nt in length. To refine these regions to single-base resolution, methods have been developed that take advantage of nucleotide mismatches or truncation signatures induced by crosslinking antibodies to modified nucleotides before reverse transcription and sequencing (e.g., **PA-m^{6}A-seq** [61], **miCLIP-seq** [11,62], **m^{1}A-MAP** [63], and m^{7}G-MeRIP-Seq [26]). **LAIC-seq** assesses m^{6}A modifications without prior fragmentation to distinguish methylation levels between mRNA isoforms, but at the cost of lost positional information [52].

Results from antibody-based mRNA modification profiling are highly variable because, for example, antibodies raised against m^{6}A can also recognize adenosine, and thus IP will not enrich only for m^{6}A-carrying transcripts [64]. Early studies using meRIP-seq reported ~7000 mRNA targets [40,65], whereas studies applying miCLIP for single-base resolution mapping reported ~3500 m^{6}A-containing mRNA targets [62]. It is unclear whether these discrepancies are due to high false-positive rates in meRIP-seq, high false-negative rates in miCLIP (owing to weak misincorporation and truncation signatures), or reflect true biological variance in different cells and conditions. However, reproducibility across m^{6}A meRIP-seq datasets has been found to be low (~30 to 60%), even in the same cell type and between biological replicates [61,66]. Furthermore, commonly used m^{6}A antibodies crossreact with m^{6}Am [62]. Thus, the discordance among antibody-based identification methods is largely a result of high false-positive rates. Higher confidence in antibody-based identification methods can be gained by including a mutant control. In yeast, 1308 high-confidence m^{6}A sites in 1183 transcripts were identified by comparing wild-type with **METTL3** mutant (lacking m^{6}A) cells [67]. A recent study used metabolic labeling (m^{6}A-label-seq) to substitute m^{6}A with *N*^{6}-allyladenosine (a^{6}A) to map m^{6}A sites, and obtained a number of mRNA targets similar to that obtained with miCLIP (2480 -4512) [68].

Antibody-based detection of m^{1}A initially reported >4000 mRNA targets [21]. Reanalysis of m^{1}A-induced misincorporation and truncation signatures from stalling of the reverse transcriptase in later studies reduced the number of targets to ~450–600 (**m^{1}A-ID-seq** [20] and m^{1}A-MAP [63]). Further, reanalysis of the 474 m^{1}A sites previously identified by m^{1}A-MAP [63] showed that 89% of the identified sites were false positives originating from genetic variation, sequencing errors, and misannotations [69]. In fact, the most recent studies using m^{1}A-seq or LC-MS/MS suggest that virtually all mapping of m^{1}A is caused by crossreactivity of commonly used m^{1}A antibodies to the m^{7}G cap, and that m^{1}A sites in mRNA are in fact exceedingly rare [14,37,70], although they have been validated in transcripts of the mitochondrial *ND5* gene [37,70].

For several other modifications, the available mapping methods remain very limited. Two studies using meRIP and miCLIP for m^{7}G identified several thousands of internal m^{7}G sites in different mouse and human cell lines [26,42]. The overlap of targets identified in the different cell lines was relatively high in one study (73.5%) [42], but low in the other (27%) [26]. Only one study has used antibody-based methods to map ac^{4}C [22]. It is thus not clear whether RIP for these modifications is hampered by similar false-positive rates and antibody crossreactivities to other modifications. Orthogonal methods to map m^{7}G and m^{3}C at single-base resolution by alkaline hydrolysis (**AlkAniline-Seq**) [71] have so far failed to detect these modifications in mRNA [71].

Chemical Methods To Detect Modifications

Chemical reactions specific to a given RNA modification, followed by short-read sequencing, provide an alternative to antibody-based detection of RNA modifications. In the case of A to I and C to U mRNA editing by deamination, no conversion step is necessary because nucleotide changes (I reads as G) can be directly assessed from RNA sequencing data by applying carefully controlled variant analysis [72].

The most common chemical conversion for mapping m^5C by bisulfite sequencing (Bs-seq or RNA-BisSeq) relies on chemical deamination of cytidines, but not of m^5C, to uridine by sodium bisulfite [15,73,74]. Bs-seq is straightforward and cost-effective, but the efficiency of chemical conversion sets a limit to the detection of rare modifications [73,75].

The catalytic mechanism of m^5C methylation involves transient covalent binding of a cytosine to the enzyme. This can be exploited by incorporation of azacytidine into nascent mRNA followed by IP with an antibody against the methyltransferase and sequencing of the target RNA [75,76]. Alternatively, mutating one of the two cysteines in the catalytic center of the methyltransferase (C271A for Nsun2) will covalently bind the methyltransferase to its RNA targets [77]. Intriguingly, the overlap between Bs-seq, 5-azacytidine immunoprecipitation (Aza-IP), and **NsunC271A-CLIP** is surprisingly low [75]. Because most m^5C sites in tRNAs were detected by these approaches, m^5C potentially differs widely between cell types or is highly dynamic, but the lack of overlap could also be due to rapid degradation of protein–RNA adducts.

Chemical methods to map m^7G and m^3C have used alkaline hydrolysis of structural RNAs (AlkAniline-Seq) [71] and reduction to abasic sites with sodium borohydride (m^7G-MaP-seq) [60], but this approach did not detect these modifications in mRNA owing to the limits of chemical conversion for detecting rare modifications [71].

To map internal Nm, nonmodified nucleotides are removed from the 3′ end of RNA with iterative oxidation–elimination–dephosphorylation cycles, but Nm blocks this process. After ligation of linkers to the Nm-modified nucleotide at the 3′-end, these sites can be mapped by sequencing (Nm-seq or **RibOxi-seq**) [28,78]. In mammalian mRNA, Nm sites in 1267 transcripts were identified [28]. Determination of Nm levels in mRNA by MS suggested that ~3% of nucleotides are ribose-methylated (18 Nm/mRNA), but this is likely a massive overestimate as a result of rRNA carry-over [14,64]. Although not yet used for mRNA, Nm can be mapped based on resistance to alkaline hydrolysis resulting in Nm nucleotides being depleted from the start of sequencing reads (**RiboMethSeq**) [79]. In addition, Nm stalls reverse transcriptase under low dNTP concentrations, resulting in Nm-specific truncation signatures during reverse transcription (2′-O-methylation sequencing, 2′-OMe-Seq [80]).

Ψ can be mapped by chemical conversion with *N*-cyclohexyl-*N*′-(2-morpholinoethyl) carbodiimide metho-*p*-toluenesulphonate (CMC), which then blocks reverse transcription (Ψ-seq or pseudo-seq) [16,19]. Ψ-seq detected 89 [16] and 353 [67] modified mRNA transcripts in different human cell lines. As with Nm-seq, the number of sites identified with Ψ-seq is low compared with the prevalence of Ψ reported by MS [14,17], and the overlap between the different studies is very modest [81]. Furthermore, a method employing a chemical pulldown enrichment step for Ψ (N3-CMC-enriched pseudouridine sequencing, CeU-seq) identified 1929 modified mRNAs [17], but whether most of these sites were missed by Ψ-seq or are false positives remains to be determined.

Enzymatic Methods To Detect Modifications

The earliest methods to map modifications in mRNA relied on the cleavage specificity of RNases, in particular RNase T1, which cleaves after guanosine [82,83]. Using this approach in combination with radioactive labeling and separation of individual nucleotides, m^6A was first mapped in prolactin and Rous sarcoma virus transcripts [84–86]. This approach remains the most sensitive and can distinguish between mRNA and rRNA because m^6A is not in a GA context in rRNA [71]. Moreover, cOMe can be analyzed in a similar sensitive way [8].

Recently, endoRNases have been discovered that are blocked by methylation [87]. In **MAZTER-seq** [88] and **m^6A-REF-seq** [89], the endoRNases MazF and ChpBK cut unmethylated RNA at ACA and UAC motifs, respectively, leaving m^6A methylated RNA intact, resulting in transcriptome-wide m^6A-dependent restriction profiles [88,89]. Although this method requires little input RNA and is highly specific, the specificity for ACA restricts MAZTER-seq to ~16% of all m^6A sites in mammals [88].

Conversion of m^6A to hm^6A by FTO protein can be exploited by its reactivity towards dithiothreithol that allows selective chemical labeling (m^6A-SEAL) [90]. This approach revealed a greater specificity and sensitivity than antibody-based (e.g., meRIP, miCLIP) and enzymatic methods (e.g., deamination adjacent to RNA modification targets, DART-seq) [90].

The APOBEC (apolipoprotein B mRNA editing enzyme) family of cytidine deaminases has been successfully used for transcriptome-wide mapping of m^6A and hm^5C. hm^5C converted by T4 bacteriophage β-glucosyltransferase (T4-BGT) is protected from deamination by APOBEC3A, allowing mapping of hm^5C in RNA (APOBEC-coupled epigenetic sequencing, ACE-seq) [91]. Similarly to hm^5C, m^6A can be mapped by using an APOBEC1–YTH fusion protein in which the m6A reader YTH targets m^6A sites, and APOBEC1 deaminates adjacent cytidines (DART-seq) [92]. DART-seq was shown to identify ~60% of m^6A meRIP targets, suggesting higher specificity than meRIP [92]. However, the efficiency of purified APOBEC1–YTH is limited *in vitro*, and exogenous expression of APOBEC1–YTH in cells is necessary, which could introduce errors owing to altered expression levels of the YTH RNA-binding protein [64,92,93]. A concept similar to DART-seq underlies m^6A mapping with **TRIBE** (targets of RNA-binding proteins identified by editing). In TRIBE, the YTH (or METTL3) RNA-binding domains are fused to the catalytic domain of ADARs. Binding of this fusion protein to m^6A sites then results in A to I editing by the ADAR catalytic domain in the vicinity [94]. TRIBE and DART-seq identify a subset of targets and also require a complementary DNA sequencing dataset to identify RNA editing events.

Site-Specific Validation of Modifications

The limitations in the sensitivity and accuracy of transcriptome-wide mRNA modification mapping techniques make validation by an orthogonal method essential. The most reliable technique to validate the presence and methylation level of m^6A is site-specific cleavage and radioactive labeling of the modified nucleotide followed by ligation-assisted extraction and thin-layer chromatography (SCARLET). Mapping of m^6A is accomplished by chimeric oligonucleotides consisting of ribo- and deoxyribonucleotides to direct site-specific cleavage of mRNA transcripts by RNase H [51,88,95], but deoxyribozymes are a valid alternative to RNase H-directed cleavage [96,97]. If the candidate site is not known at nucleotide resolution, *in vitro* transcribed substrate RNAs can be methylated in nuclear extracts and the modifications can be mapped by molecular methods [64,84,86,98]

Direct RNA Sequencing by Nanopore

Oxford Nanopore Technologies™ (ONT) is a long-read sequencing platform that generates long reads of up to several kilobases from single mRNA molecules in real time (Box 3) [99,100]. By

Box 3. Nanopore Sequencing

Oxford Nanopore Technologies™ (ONT) sequencers are available as small hand-held devices (MinION) and larger high-throughput sequencers (PromethION). The underlying technology is the same between sequencers, and the largest difference is sequencing output. ONT is a rapidly developing technology, and frequent improvements are released for sequencing reagents, sequencing flow cells, and bioinformatics. These technological updates have resulted in the rapid improvement of ONT sequencing error rates (Box 4). Although bioinformatic updates can be applied to previous datasets (Box 4), reagent and flow-cell updates might make it difficult to compare data from older and newer studies. ONT also relies heavily on user-developed solutions. For direct RNA sequencing specifically, ONT only offers sequencing library kits for mRNA or polyA RNA, and studies investigating other types of RNA (e.g., *in vitro* transcribed RNA, rRNA, tRNA) have developed their own custom adaptations [126,127].

For direct mRNA sequencing, polyA mRNA is enriched with oligo-dT beads and a cDNA strand is synthesized using an oligo-dT primer (see Figure 2A in main text). The cDNA strand is not sequenced, but prevents RNA from forming secondary structures and protects against degradation by most RNases, and thus increases throughput. To prepare nucleic acids for nanopore sequencing, an RNA sequencing adapter (RMX) is ligated to the 3′ end. This adaptor carries a 'motor protein' at the 3′ end and a 'tether' protein at the 5′ end. The motor protein moves the nucleic acid strand (~400 bases/s for DNA and ~70 bases/s for RNA) through the pore while the tether protein prevents the cDNA from passing through the pore and being sequenced.

The current nanopore (R9.4 or R9.5) consists of an engineered protein derived from *Escherichia coli* Curlin sigma S-dependent growth (CsgG) pore [128] embedded in an electrically resistant membrane made from a synthetic polymer [129–131] (see Figure 2B in main text). MinKNOW™, the software made by ONT to control the MinION device, performs several core tasks such as assignment of run parameters, data acquisition, and feedback on how the experiment is progressing. The membrane in a MinION flow cell holds individually addressable 2048 pores controlled in groups of 512 channels. MinKNOW assigns four pores per channel through mux scan (a process to check pore activity), thus allowing simultaneous sequencing of 512 molecules [131]. Nanopores are immersed in an ionic solution, and when voltage is applied a ionic current passes through the nanopore that is individually recorded by a sensor [131–133]. The sensor measures the current several thousand times per second, and the collected data are sent to MinKNOW [131,133].

When a nucleic acid strand traverses the nanopore from one chamber to the other, the current changes in a characteristic way such that the nucleic acid sequence can be identified. The raw signals (current measurements) plotted over time are known as 'squiggle' plots (see Figure 2C in main text). Unlike fluorescence-based sequencing, where individual nucleotides are recognized, the measured signal is affected by multiple nucleotides that reside in the pore, which is modeled using 6 nt subsequences (k-mers) for DNA and 5 nt for RNA. These 5 nt subsequences from RNA result in 4^5 different signal configurations [101,134–136]. These signal configurations are modeled using Gaussian distributions where the mean represents the level of the signal and the standard deviation represents the variation from measurement and intrinsic noise (see Figure 2D in main text).

contrast to other long-read sequencing technologies (e.g., PacBio), ONT allows direct RNA sequencing (DRS) without prior amplification or cDNA conversion. Therefore, ONT has potential in the direct detection of RNA modifications, opening a completely new approach to the study of epitranscriptomic modifications, but accurate detection faces significant challenges.

Nucleotide modifications in ONT sequencing reads were first demonstrated on CpG-methylated DNA based on deviations of ONT raw current signals from a model of canonical nucleotides (Box 4 for bioinformatics of ONT basecalling) [101,102]. CpG-methylation calling software is now an integrated feature of the ONT analysis software Nanopolish. This methylation caller is a hidden Markov model that was trained on synthetically methylated DNA to distinguish between methylated and unmethylated cytosines using raw current signals [101], and has been applied to the study of DNA methylation patterns in the human genome and epigenetic imprinting in mice [103,104].

In 2018, ONT introduced a direct RNA sequencing protocol and showed raw signal differences between modified m^6A, m^5C, and unmodified bases using a synthetically designed firefly luciferase (*FLuc*) transcript with embedded modifications [105]. Subsequently, ONT DRS was applied to *Escherichia coli* full-length 16S rRNA, showing that ONT signals are sensitive to m^7G and Ψ via corresponding errors in basecalling [106]. The first study applying ONT DRS to the human

Box 4. Bioinformatics of Nanopore Sequencing

To obtain the sequences of the input molecules, the electrical signal must be translated into individual bases; a process called basecalling. MinKNOW has a built-in basecaller for 'live' basecalling during the sequencing run, or the user can turn this function off and instead basecall the raw files later.

Early versions of the sequencers from ONT used a basecaller that ran on the Metrichor™ cloud-based platform. This basecaller used a hidden Markov model (HMM), a probabilistic model for decoding an unknown sequence of 'states' based on an observed sequence. For nanopore basecalling the unknown states were k-mers of the DNA or RNA molecule, and the observed sequence was a segmentation of the raw signal. This approach to basecalling has been superseded by neural network-based approaches (reviewed in [137]), as comprehensively benchmarked by Wick and colleagues [134].

The latest ONT basecaller, called Guppy, uses graphical processing units (GPUs) to accelerate this crucial signal-to-sequence step to facilitate high-throughput sequencing experiments. Guppy has two different basecalling modes – a 'fast' basecaller that aims to keep up with live data generation, and a 'high-accuracy' model that lowers the sequencing error rate at the expense of longer processing times. Guppy (V3.4.5) has a basecalling error rate of 4–6% for DNA and 7–12% for RNA in high-accuracy mode [134]. Guppy's latest version can also directly call modified bases in DNA (m^5C and m^6A), although only in limited contexts (CpG and CCWGG for m^5C, and GATC for m^6A). Both HMM and recurrent neural network (RNN) models depend on training signals, and the performance of a basecaller therefore depends on having a good training data set. With each update to the basecaller and further reduction in error rates, it is possible to re-basecall old datasets as long as the raw signal files are retained.

transcriptome showed that m^6A could be detected within the methylation consensus motif (GGACU) in the human cell line GM12878. As a proof of concept, the study showed a raw ONT signal difference at a known m^6A site (chr19:3,976,327, GRCh38 human genome reference) in eukaryotic elongation factor 2 (*EEF2*) transcripts compared with an *in vitro* transcribed unmodified transcript. The study then identified 86 genes that showed signal differences at the same GGACU site between different transcripts of the same gene [107]. The ONT software package Tombo can also identify modified bases in RNA based on signal deviations. This has been used to call m^5C methylation patterns in the viral RNA of human coronavirus [108]. These initial studies provided the first evidence that ONT DRS sequencing is sensitive to RNA modifications and thus might be used for transcriptome-wide mapping of epitranscriptomic marks.

Although direct RNA sequencing holds a great promise for the identification of RNA modifications at single-base resolution, the challenge, however, lies in the interpretation of raw signals corresponding to modified and unmodified base sequence contexts. Currently available computational algorithms to map RNA modifications using ONT DRS datasets are either based on modification-induced basecalling errors (differential error calling) or machine-learning models that look at differences in raw signal levels. Both approaches generally require matching low- and high-modification datasets, such as lack-of-function mutants for METTL3 [109,110], *in vitro* synthesized RNA [105,109,111], or reverse transcribed RNA (cDNA) [112]. In addition, most studies generally use antibody-based data, known modified sites, and/or and filtering for known motifs (such as the DRACH motif for m^6A) to refine and/or validate results [109,111–113].

Differential Error Calling

The first category of algorithms to call RNA modifications from ONT DRS reads operates on the assumption that RNA modifications cause a deviation of ONT signal that results in bases being miscalled. These modification-induced calling errors are used as a surrogate to map modified nucleotides on transcripts [112]. For instance, ELIGOS (epitranscriptional landscape inferring from glitches of ONT signals) works on the percentage error of specific bases (%ESB) between native RNA sequencing data and cDNA data for the same sample (because modifications are removed from cDNA), and has been used to identify rRNA methylation sites in *E. coli*, yeast, and human cells [112]. However, this approach is subject to differences arising from the different nanopore motors, sequencing directions, and basecalling models between RNA and cDNA.

An elegant approach to identify m^{6}A-induced ONT sequencing errors used matched sets of synthetic RNA molecules from *MALAT1* lncRNA with and without m^{6}A at known locations. This approach was then extrapolated to use the differences in error calling between wild-type, METTL3-defective mutant, and METTL3 genetic-rescue *Arabidopsis* to identify m^{6}A sites transcriptome-wide [109]. In the same system, m^{6}A sites were further analyzed by miCLIP to show that 66% of called nanopore DRS differential error sites fall within miCLIP peaks [109]. A similar approach was taken to identify base-resolution and isoform-specific m^{6}A sites in adenovirus RNA by comparing DRS errors in m^{6}A wild-type and METTL3 knockout datasets [113].

Likewise, the algorithm EpiNano was trained on 100% methylated or unmethylated synthetic RNA sequences to predict m^{6}A modifications based on sequencing errors [111]. EpiNano predicted m^{6}A sites with 90% accuracy in synthetic reads and 87% accuracy in yeast *in vivo*, but prediction accuracy depends on the extent of methylation. The prediction accuracy of EpiNano was validated on DRS reads from a yeast METTL3 knockout strain lacking m^{6}A modifications.

Machine-Learning Methods Using DRS Raw Signals

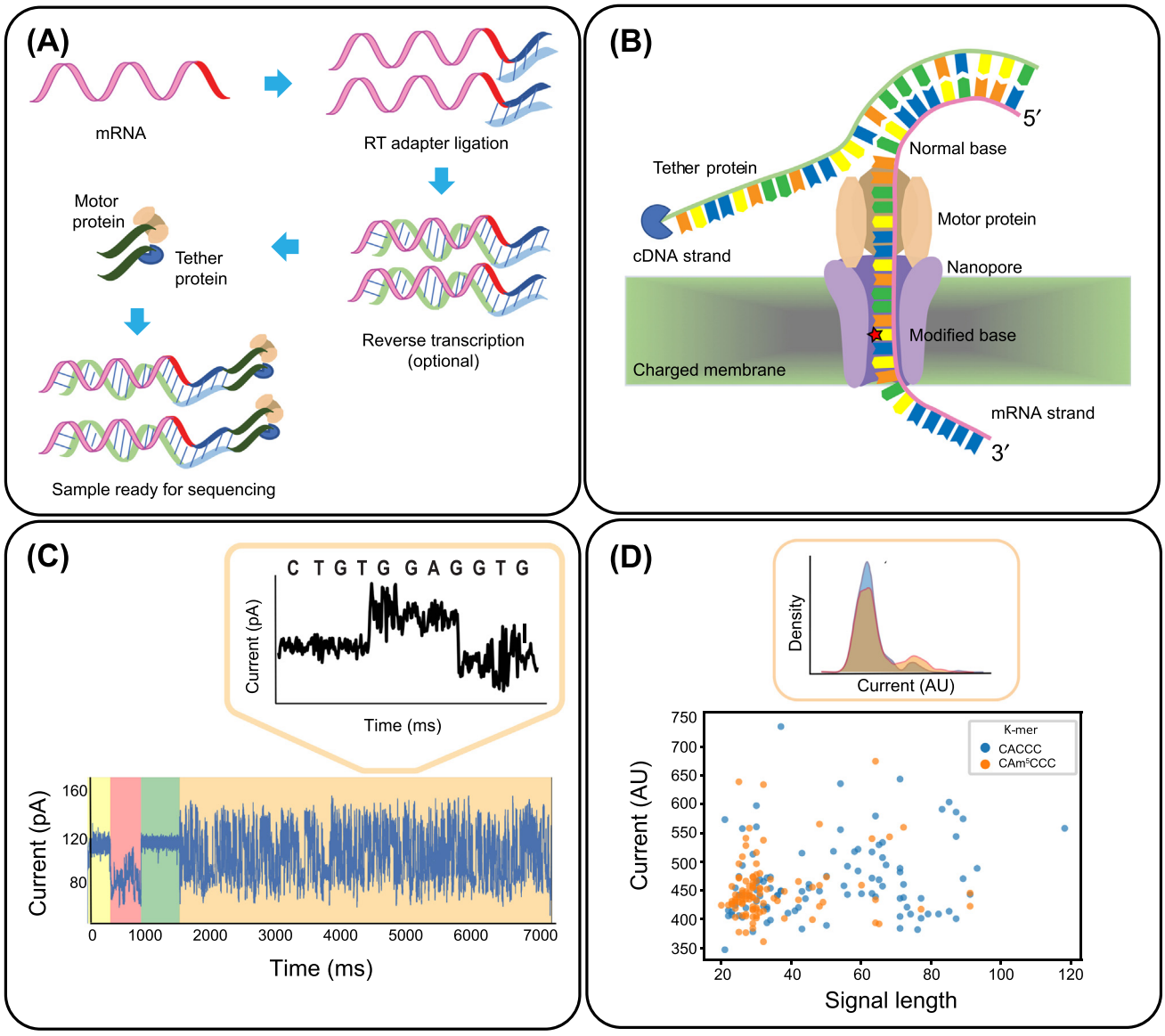
The second category of methods to identify RNA modifications is based on training machine-learning models using modifications identified by antibody-based methods. These maps are then used to develop modification-specific basecallers for direct RNA-sequencing data to predict modifications [110]. This approach was used to develop MINES (m^{6}A identification using Nanopore sequencing), a random forest classifier trained on miCLIP m^{6}A sites within DRACH motifs in HEK293T cells. MINES was able to predict m^{6}A miCLIP modification sites with 80% accuracy and identified a significant set of previously unannotated sites. The authors validated MINES by showing the sensitivity of identified sites to METTL3 knockdown and *ALKB5* overexpression. Importantly, sites identified by miCLIP, but not predicted by MINES, did not show sensitivity of METTL3 or *ALKB5* manipulations, as would be expected from false-positive miCLIP sites [110].

It is worth noting that MINES extracts raw DRS signals using Tombo, which does not support spliced mapping and only allows assessment of sites in the 3′-UTRs of transcripts. Comparing the performance of Tombo with the error-based algorithm EpiNano revealed that ~50% of reads were discarded by Tombo and that EpiNano achieved higher accuracy than Tombo (87% vs 59%), but reduced sensitivity (32% vs 59%) [111].

Nanocompore, another machine-learning algorithm, uses Nanopolish to align raw DRS signals to the reference genome, which allows spliced alignment of RNA reads [114]. Nanocompore runs an automated pipeline for data preprocessing including basecalling, alignment using Minimap2, and signal alignment using Nanopolish. It operates by comparing DRS signals between high- and low-modification samples to detect transcriptome-wide m^{6}A modification events. The authors stipulate that, given an appropriate low-modification control, Nanocompore can be used to detect a large variety of different RNA modifications [114].

Current Limitations of ONT RNA Modification Prediction

The methods mentioned here demonstrate that ONT sequencing harbors the potential to revolutionize the transcriptome-wide prediction of RNA modifications. However, each of the available solutions currently has limitations. Differential error algorithms require matched low- and high-modification samples with identical genetic backgrounds because differences in DNA sequence (e.g., SNPs) will result in errors that are unrelated to RNA modifications. This limits the current use of these algorithms to systems where a low-modification control can be obtained on the exact genetic background of the sample of interest.



Trends in Biotechnology

Figure 2. Nanopore Sequencing. (A) Schematic of the library preparation procedure for Nanopore direct RNA sequencing. PolyA RNA is enriched using oligo-dT primers and a reverse transcription (RT) adaptor is ligated. After second-strand synthesis, the sequencing adapter RMX, which is preloaded with motor protein and tether protein, is then ligated. (B) Schematic of Nanopore direct RNA sequencing. The motor protein feeds the RNA molecule through the nanopore in the 3′–5′ direction. The five bases passing through a nanopore cause a characteristic disruption in the current which is stored as raw signal. (C) A current trace (squiggle plot) showing the raw signal generated by nanopore sequencing of a single mRNA molecule. Leader and adapter sequences are shaded yellow and pink, the polyA tail is shaded green, and the mRNA body is shaded orange. The inset (top right) illustrates how the nucleotide sequence is inferred from the raw current trace originating from a sliding window of five nucleotides (k-mer). Machine-learning algorithms are then used to calculate the probability that a signal corresponds to a given k-mer, thus inferring the nucleotide sequence from the calculated probabilities. (D) The two features recorded by Oxford Nanopore Technologies (ONT) sequencers are the current signal (in arbitrary units, AU) and the time that a given k-mer takes to transverse the pore (signal length, retention time or 'dwell'). The scatter plot depicts the distribution of mean current and signal length for 100 reads each in a different sequence context of the unmodified k-mer CACCC (blue) and the modified k-mer CAm^{5}CCC (orange, identified by parallel bisulfite sequencing, where m^{5}C is 5-methylcytosine). Note that, despite an identical k-mer, the signal varies as a result of different measurements and intrinsic noise in different reads, and possibly also by the different surrounding sequence of a given k-mer. This variability can be represented as a signal density plot for each k-mer, depicted in the top-right inset (density distribution for raw current signal). RNA modifications can affect raw current reads as well as signal length, resulting in a shift in signal distributions (e.g., divergence between blue and orange). However, these signal shifts can be modest, as shown by the largely overlapping density plots for CACCC and CAm^{5}CCC, making accurate prediction of modified bases a computational challenge. Plots were generated with Sequoia, an interactive visual analytics platform for interpretation and feature extraction from ONT sequencing datasets [138].

The prediction of RNA modifications in ONT sequencing is further complicated by the fact that each sequencing signal originates from a group (k-mer) of 5 nt rather than from a single nucleotide. Moreover, the signal for each k-mer is variable, leading to a distribution of possible signals (Figure 2D). RNA modifications can result in a shift in the signal distribution for a given k-mer, and this shift can be used to predict the presence of modifications. However, these shifts can be relatively small, and the distribution of signals generated by modified k-mers largely overlaps with the distribution of unmodified k-mers (Figure 2D).

Modification prediction platforms that employ machine-learning models depend on the quality of the training datasets (miCLIP or similar), and are thus limited by the quality of these datasets. Currently >160 RNA modifications have been reported in the literature [115]. Generating training models that encompass these diverse modifications will require very large and complex training datasets.

Concluding Remarks and Future Directions

The emerging field of epitranscriptomics has brought global insights into the topography of nucleotide modifications in mRNA and their impact on gene expression. The prevalent links of RNA modifications to lifestyle diseases such as obesity, as well as to many neurological diseases and cancer, have made RNA modification enzymes prime targets for the pharmaceutical industry in employing their classic repertoire to develop inhibitors. To better understand the biological roles of RNA modifications, and how to pharmacologically interfere with them, accurate epitranscriptomic maps are required.

Recent technical advances at many levels have increased the catalog of modifications in mRNA and promise to overcome technical challenges to detect mRNA modifications with high confidence to address key questions about their role in the regulation of gene expression. However, it is important to consider the inherent limitations of antibody- and chemical conversion-based methods to detect modifications that are generally not abundant. The reliability of results can be improved by higher sample sizes (most studies currently only report 2–3 replicates), increased sequencing depth [70], and by considering variability due to different reaction conditions, antibodies, and enzymes [116]. A major improvement in obtaining higher accuracy in mapping modifications is to include a knockout of the writer and verification of specific targets by orthogonal methods.

One big open question concerns whether the variability observed in global levels of mRNA modifications reflects limitations in current methods, cell-to-cell variability, or cell type-specific features (see Outstanding Questions). In addition, current methods can only assess a single type of modification at a time, but it is likely that a combination of modifications is present on a given transcript. Thus, to truly understand how mRNA modifications regulate gene expression, it will be important to develop methodologies to determine the combinations of modifications that are present in individual transcripts and from small amounts of input such as individual cells.

Single-molecule sequencing by ONT now promises the technological advances to eventually measure low and dynamic changes in mRNA modifications. Moreover, ONT will permit combinations of modifications to be assessed on single transcripts in the near future. To address key questions about the function of modifications, single-cell RNA sequencing will inevitably be required. Because direct RNA sequencing by ONT excludes cDNA amplification steps, sequencing mRNAs from single cells presents a particular challenge, but solutions to this conundrum could include using smaller devices and more replicates.

Outstanding Questions

Which nucleotides are modified in mRNA?

What fraction of mRNAs carry modified nucleotides?

Why do only a fraction of mRNAs carry a specific nucleotide modification?

Are mRNA modifications dynamic?

Do mRNAs carry multiple different modifications, and is there a modification code?

How do mRNA modifications contribute to development and cause disease?

For ONT to become the leading technology, further improvements at multiple levels are required, with a particular focus on the major challenge in computational interpretation of sequencing signals. Several machine-learning algorithms have recently been developed to predict modifications from ONT data. These algorithms are generally random forest models that predict modified sites based on a combination of RIP-seq or ONT training data, but they also incorporate sequence motifs and other sequence features that have been developed from RIP-seq data [117–120]. The main limitation here is that the ability of computational algorithms to predict modifications relies on the quality of the training data. High false-positive rates in RIP-seq data and high rates of sequencing errors in ONT data thus limit the accuracy of such algorithms.

Given rapid technological progress, we anticipate soon being able to more comprehensively understand how the modification of nucleotides is dynamically instructed to impact on mRNA processing, and how its misregulation results in human disease.

Acknowledgments

We apologize to colleagues whose work was not cited owing to space limitations. We thank I. Haussmann and R. Arnold for comments on the manuscript. We thank R. Koonchanok and S. Daulatabad for help with preparation of Figure 2. For this work, we acknowledge funding from the UK Biotechnology and Biological Sciences Research Council (BBSRC) and the Leverhulme trust to M.S. I.A. is supported by the Schmidt Science Fellows in partnership with the Rhodes Trust. J.T.S. is supported by the Ontario Institute for Cancer Research through funds provided by the Government of Ontario and the Government of Canada via Genome Canada and Ontario Genomics (OGI-136). S.C.J. acknowledges support from the National Institute of General Medical Sciences of the National Institutes of Health (grant R01GM123314) and the National Science Foundation (1940422 and 1908992).

References

1. Machnicka, M.A. *et al.* (2014) Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol.* 11, 1619–1629
2. Sloan, K.E. *et al.* (2017) Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152
3. Soller, M. and Fray, R. (2019) RNA modifications in gene expression control. *Biochim. Biophys. Acta Gene Regul. Mech.* 1862, 219–221
4. Galloway, A. and Cowling, V.H. (2019) mRNA cap regulation in mammalian cell function and fate. *Biochim. Biophys. Acta Gene. Regul. Mech.* 1862, 270–279
5. Wang, J. *et al.* (2019) Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. *Nucleic Acids Res.* 47, e130
6. Walters, R.W. *et al.* (2017) Identification of NAD^+ capped mRNAs in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 114, 480–485
7. Sun, H. *et al.* (2019) Cap-specific, terminal N6-methylation by a mammalian m6Am methyltransferase. *Cell Res.* 29, 80–82
8. Kruse, S. *et al.* (2011) A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA. *Sci. Rep.* 1, 126
9. Akichika, S. *et al.* (2019) Cap-specific terminal N6-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science* 363, eaav0080
10. Sendinc, E. *et al.* (2019) PCIF1 catalyzes m6Am mRNA methylation to regulate gene expression. *Mol. Cell* 75, 620–630
11. Boulias, K. *et al.* (2019) Identification of the m6Am methyltransferase PCIF1 reveals the location and functions of m6Am in the transcriptome. *Mol. Cell* 75, 631–643
12. Bangs, J.D. *et al.* (1992) Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J. Biol. Chem.* 267, 9805–9815
13. Mauer, J. *et al.* (2017) Reversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nature* 541, 371–375
14. Dong, C. *et al.* (2016) tRNA modification profiles of the fast-proliferating cancer cells. *Biochem. Biophys. Res. Commun.* 476, 340–345
15. Squires, J.E. *et al.* (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* 40, 5023–5033
16. Carlile, T.M. *et al.* (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515, 143
17. Li, X. *et al.* (2015) Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* 11, 592–597
18. Lovejoy, A.F. *et al.* (2014) Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One* 9, e110799
19. Schwartz, S. *et al.* (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159, 148–162
20. Li, X. *et al.* (2016) Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. *Nat. Chem. Biol.* 12, 311–316
21. Dominissini, D. *et al.* (2016) The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* 530, 441–446
22. Arango, D. *et al.* (2018) Acetylation of cytidine in mRNA promotes translation efficiency. *Cell* 175, 1872–1886
23. Delatte, B. *et al.* (2016) Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* 351, 282–285
24. Xu, L. *et al.* (2017) Three distinct 3-methylcytidine (m3C) methyltransferases modify tRNA and mRNA in mice and humans. *J. Biol. Chem.* 292, 14695–14703
25. Salter, J.D. *et al.* (2016) The APOBEC protein family: united by structure, divergent in function. *Trends Biochem. Sci.* 41, 578–594
26. Zhang, L.S. *et al.* (2019) Transcriptome-wide mapping of internal N7-methylguanosine methylome in mammalian mRNA. *Mol. Cell* 74, 1304–1316

27. Chu, J.M. *et al.* (2018) Existence of internal N7-methylguanosine modification in mRNA determined by differential enzyme treatment coupled with mass spectrometry analysis. *ACS Chem. Biol.* 13, 3243–3250
28. Dai, Q. *et al.* (2017) Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat. Methods* 14, 695–698
29. Dai, D.-P. *et al.* (2018) Transcriptional mutagenesis mediated by 8-oxoG induces translational errors in mammalian cells. *Proc. Natl. Acad. Sci.* 115, 4218–4222
30. Roignant, J.Y. and Soller, M. (2017) m6A in mRNA: an ancient mechanism for fine-tuning gene expression. *Trends Genet.* 33, 380–390
31. Balacco, D.L. and Soller, M. (2019) The m6A writer: rise of a machine for growing tasks. *Biochemistry* 58, 363–378
32. Dezi, V. *et al.* (2016) Nucleotide modifications in messenger RNA and their role in development and disease. *Biochem. Soc. Trans.* 44, 1385–1393
33. Frye, M. *et al.* (2018) RNA modifications modulate gene expression during development. *Science* 361, 1346–1349
34. Shi, H. *et al.* (2019) Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Mol. Cell* 74, 640–650
35. Zaccara, S. *et al.* (2019) Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.* 20, 608–624
36. Licht, K. *et al.* (2019) Inosine induces context-dependent recoding and translational stalling. *Nucleic Acids Res.* 47, 3–14
37. Safra, M. *et al.* (2017) The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* 551, 251–255
38. Lesbirel, S. and Wilson, S.A. (2018) The m6A methylase complex and mRNA export. *Biochim. Biophys. Acta Gene Regul. Mech.* 1862, 319–328
39. Liu, F. *et al.* (2016) ALKBH1-mediated tRNA demethylation regulates translation. *Cell* 167, 816–828
40. Dominissini, D. *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206
41. Engel, M. *et al.* (2018) The Role of m6A/m-RNA methylation in stress response regulation. *Neuron* 99, 389–403
42. Malbec, L. *et al.* (2019) Dynamic methylome of internal mRNA N7-methylguanosine and its regulatory role in translation. *Cell Res.* 29, 927–941
43. Sharma, S. *et al.* (2015) APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat. Commun.* 6, 6881
44. Mannion, N.M. *et al.* (2014) The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.* 9, 1482–1494
45. Shulman, Z. and Stern-Ginossar, N. (2020) The RNA modification N6-methyladenosine as a novel regulator of the immune system. *Nat. Immunol.* 21, 501–512
46. Schuberth-Wagner, C. *et al.* (2015) A conserved histidine in the RNA sensor RIG-I controls immune tolerance to N1-2'O-methylated self RNA. *Immunity* 43, 41–51
47. Devarkar, S.C. *et al.* (2016) Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Proc. Natl. Acad. Sci. U. S. A.* 113, 596–601
48. Abbas, Y.M. *et al.* (2017) Structure of human IFIT1 with capped RNA reveals adaptable mRNA binding and mechanisms for sensing N1 and N2 ribose 2'-O methylations. *Proc. Natl. Acad. Sci. U. S. A.* 114, 2106–2115
49. Johnson, B. *et al.* (2018) Human IFIT3 modulates IFIT1 RNA binding specificity and protein stability. *Immunity* 48, 487–499
50. Young, D.F. *et al.* (2016) Human IFIT1 inhibits mRNA translation of Rubulaviruses but not other members of the Paramyxoviridae family. *J. Virol.* 90, 9446–9456
51. Liu, N. *et al.* (2013) Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* 19, 1848–1856
52. Molinie, B. *et al.* (2016) m6A-LAIC-seq reveals the census and complexity of the m6A epitranscriptome. *Nat. Methods* 13, 692–698
53. Yuan, B.-F. (2017) Liquid chromatography–mass spectrometry for analysis of RNA adenosine methylation. In *RNA Methylation: Methods and Protocols* (Lusser, A., ed.), pp. 33–42, Springer New York
54. Jora, M. *et al.* (2019) Detection of ribonucleoside modifications by liquid chromatography coupled with mass spectrometry. *Biochim. Biophys. Acta Gene Regul. Mech.* 1862, 280–290
55. Galloway, A. *et al.* (2020) CAP-MAP: cap analysis protocol with minimal analyte processing, a rapid and sensitive approach to analysing mRNA cap structures. *Open Biol.* 10, e190306
56. Legrand, C. *et al.* (2017) Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res.* 27, 1589–1596
57. Jiang, T. *et al.* (2019) Oligonucleotide sequence mapping of large therapeutic mRNAs via parallel ribonuclease digestions and LC-MS/MS. *Anal. Chem.* 91, 8500–8506
58. Yu, N. *et al.* (2019) tRNA modification profiles and codon-decoding strategies in *Methanocaldococcus jannaschii*. *J. Bacteriol.* 201, e00690-18
59. Pandolfini, L. *et al.* (2019) METTL1 promotes let-7 microRNA processing via m7G methylation. *Mol. Cell* 74, 1278–1290
60. Enroth, C. *et al.* (2019) Detection of internal N7-methylguanosine (m7G) RNA modifications by mutational profiling sequencing. *Nucleic Acids Res.* 47, e126
61. Chen, K. *et al.* (2015) High-resolution N6-methyladenosine (m6A) map using photo-crosslinking-assisted m6A sequencing. *Angew. Chem. Int. Ed.* 54, 1587–1590
62. Linder, B. *et al.* (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* 12, 767–772
63. Li, X. *et al.* (2017) Base-resolution mapping reveals distinct m1A methylome in nuclear- and mitochondrial-encoded transcripts. *Mol. Cell* 68, 993–1005
64. Haussmann, I.U. *et al.* (2016) m6A potentiates Sxl alternative pre-mRNA splicing for robust *Drosophila* sex determination. *Nature* 540, 301–304
65. Meyer, K.D. *et al.* (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635–1646
66. McIntyre, A.B.R. *et al.* (2020) Limits in the detection of m6A changes using MeRIP/m6A-seq. *Sci. Rep.* 10, 6590
67. Schwartz, S. *et al.* (2013) High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* 155, 1409–1421
68. Shu, X. *et al.* (2020) A metabolic labeling method detects m6A transcriptome-wide at single base resolution. *Nat. Chem. Biol.* Published online April 27, 2020. https://doi.org/10.1038/s41589-020-0526-9
69. Schwartz, S. (2018) m1A within cytoplasmic mRNAs at single nucleotide resolution: a reconciled transcriptome-wide map. *RNA* 24, 1427–1436
70. Grozhik, A.V. *et al.* (2019) Antibody cross-reactivity accounts for widespread appearance of m1A in 5'UTRs. *Nat. Commun.* 10, 5126
71. Marchand, V. *et al.* (2018) AlkAniline-seq: profiling of m7G and m3C RNA modifications at single nucleotide resolution. *Angew Chem. Int. Ed. Engl.* 57, 16785–16790
72. Tan, M.H. *et al.* (2017) Dynamic landscape and regulation of RNA editing in mammals. *Nature* 550, 249–254
73. Yang, X. *et al.* (2017) 5-methylcytosine promotes mRNA export – NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell Res.* 27, 606–625
74. Amort, T. *et al.* (2017) Distinct 5-methylcytosine profiles in poly (A) RNA from mouse embryonic stem cells and brain. *Genome Biol.* 18, 1
75. Hussain, S. *et al.* (2013) Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol.* 14, 215
76. Khoddami, V. and Cairns, B.R. (2013) Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* 31, 458–464
77. Hussain, S. *et al.* (2013) NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep.* 4, 255–261
78. Zhu, Y. *et al.* (2017) High-throughput and site-specific identification of 2'-O-methylation sites using ribose oxidation sequencing (RibOxi-seq). *RNA* 23, 1303–1314

79. Marchand, V. *et al.* (2016) Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res.* 44, e135
80. Incarnato, D. *et al.* (2017) High-throughput single-base resolution mapping of RNA 2-O-methylated residues. *Nucleic Acids Res.* 45, 1433–1441
81. Safra, M. *et al.* (2017) TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. *Genome Res.* 27, 393–406
82. Keith, G. (1995) Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* 77, 142–144
83. Motorin, Y. and Helm, M. (2011) RNA nucleotide methylation. *Wiley Interdiscip. Rev. RNA* 2, 611–631
84. Horowitz, S. *et al.* (1984) Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc. Natl. Acad. Sci. U. S. A.* 81, 5667–5671
85. Kane, S.E. and Beemon, K. (1985) Precise localization of m6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. *Mol. Cell. Biol.* 5, 2298–2306
86. Harper, J.E. *et al.* (1990) Sequence specificity of the human mRNA N6-adenosine methylase in vitro. *Nucleic Acids Res.* 18, 5735–5741
87. Imanishi, M. *et al.* (2017) Detection of N6-methyladenosine based on the methyl-sensitivity of MazF RNA endonuclease. *Chem. Commun. (Camb.)* 53, 12930–12933
88. Garcia-Campos, M.A. *et al.* (2019) Deciphering the 'm6A Code' via antibody-independent quantitative profiling. *Cell* 178, 731–747
89. Zhang, Z. *et al.* (2019) Single-base mapping of m6A by an antibody-independent method. *Sci. Adv.* 5, eaax0250
90. Wang, Y. *et al.* (2020) Antibody-free enzyme-assisted chemical approach for detection of N6-methyladenosine. *Nat. Chem. Biol.* Published online April 27, 2020. https://doi.org/10.1038/s41589-020-0525-x
91. Schutsky, E.K. *et al.* (2018) Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. *Nat. Biotechnol.* 36, 1083–1090
92. Meyer, K.D. (2019) DART-seq: an antibody-free method for global m6A detection. *Nat. Methods* 16, 1275–1280
93. Zaharieva, E. *et al.* (2015) Concentration and localization of coexpressed ELAV/Hu proteins control specificity of mRNA processing. *Mol. Cell. Biol.* 35, 3104–3115
94. Worpenberg, L. *et al.* (2019) Identification of methylated transcripts using the TRIBE approach. In *Epitranscriptomics: Methods and Protocols* (Wajapeyee, N. and Gupta, R., eds), pp. 89–106, Springer New York
95. Liu, N. and Pan, T. (2016) Probing N6-methyladenosine (m6A) RNA modification in total RNA with SCARLET. *Methods Mol. Biol.* 1358, 285–292
96. Sednev, M.V. *et al.* (2018) N6-methyladenosine-sensitive RNA-cleaving deoxyribozymes. *Angew Chem. Int. Ed. Engl.* 57, 15117–15121
97. Bujnowska, M. *et al.* (2020) Deoxyribozyme-based method for absolute quantification of N6-methyladenosine fractions at specific sites of RNA. *J. Biol. Chem.* 295, 6992–7000
98. Narayan, P. and Rottman, F.M. (1988) An in vitro system for accurate methylation of internal adenosine residues in messenger RNA. *Science* 242, 1159–1162
99. Banerjee, A. *et al.* (2010) Molecular bases of cyclodextrin adapter interactions with engineered protein nanopores. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8165–8170
100. Stoddart, D. *et al.* (2010) Multiple base-recognition sites in a biological nanopore: two heads are better than one. *Angew Chem. Int. Ed. Engl.* 49, 556–559
101. Simpson, J.T. *et al.* (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* 14, 407–410
102. Rand, A.C. *et al.* (2017) Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* 14, 411–413
103. Jain, M. *et al.* (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345
104. Gigante, S. *et al.* (2019) Using long-read sequencing to detect imprinted DNA methylation. *Nucleic Acids Res.* 47, e46
105. Garalde, D.R. *et al.* (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* 15, 201–206
106. Smith, A.M. *et al.* (2019) Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS One* 14, e0216709
107. Workman, R.E. *et al.* (2019) Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* 16, 1297–1305
108. Viehweger, A. *et al.* (2019) Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res.* 29, 1545–1554
109. Parker, M.T. *et al.* (2020) Nanopore direct RNA sequencing maps the complexity of *Arabidopsis* mRNA processing and m6A modification. *eLife* 9, e49658
110. Lorenz, D.A. *et al.* (2019) Direct RNA sequencing enables m6A detection in endogenous transcript isoforms at base specific resolution. *RNA* 26, 19–28
111. Liu, H. *et al.* (2019) Accurate detection of m6A RNA modifications in native RNA sequences. *Nat. Commun.* 10, 4079
112. Wongsurawat, T. *et al.* (2018) Decoding the epitranscriptional landscape from native RNA sequences. *BioRxiv* Published online December 17, 2018. https://doi.org/10.1101/487819
113. Price, A.M. *et al.* (2019) Direct RNA sequencing reveals m6A modifications on adenovirus RNA are necessary for efficient splicing. *BioRxiv* Published online December 5, 2019. https://doi.org/10.1101/865485
114. Leger, A. *et al.* (2019) RNA modifications detection by comparative Nanopore direct RNA sequencing. *BioRxiv* Published online November 14, 2019. https://doi.org/10.1101/843136
115. Boccaletto, P. *et al.* (2018) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* 46, D303–D307
116. Werner, S. *et al.* (2020) Machine learning of reverse transcription signatures of variegated polymerases allows mapping and discrimination of methylated purines in limited transcriptomes. *Nucleic Acids Res.* 48, 3734–3746
117. Zhou, Y. *et al.* (2016) SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res.* 44, e91
118. Zhao, W. *et al.* (2019) PACES: prediction of N4-acetylcytidine (ac4C) modification sites in mRNA. *Sci. Rep.* 9, 11112
119. Li, J. *et al.* (2018) RNAm5Cfinder: a web-server for predicting RNA 5-methylcytosine (m5C) sites based on random forest. *Sci. Rep.* 8, 17299
120. Zhou, Y. *et al.* (2019) NmSEER V2.0: a prediction tool for 2'-O-methylation sites based on random forest and multi-encoding combination. *BMC Bioinformatics* 20, 690
121. Davis, F.F. and Allen, F.W. (1957) Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* 227, 907–915
122. Dimitrova, D.G. *et al.* (2019) RNA 2'-O-methylation (Nm) modification in human diseases. *Genes (Basel)* 10, 117
123. Carlile, T.M. *et al.* (2019) mRNA structure determines modification by pseudouridine synthase 1. *Nat. Chem. Biol.* 15, 966–974
124. Jia, G. *et al.* (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7, 885–887
125. Mauer, J. *et al.* (2019) FTO controls reversible m6Am RNA methylation during snRNA biogenesis. *Nat. Chem. Biol.* 15, 340–347
126. Keller, M.W. *et al.* (2018) Direct RNA sequencing of the coding complete influenza a virus genome. *Sci. Rep.* 8, 14408
127. Gilpatrick, T. *et al.* (2020) Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* 38, 433–438
128. Goyal, P. *et al.* (2014) Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* 516, 250–253
129. Ayub, M. and Bayley, H. (2016) Engineered transmembrane pores. *Curr. Opin. Chem. Biol.* 34, 117–126
130. Henley, R.Y. *et al.* (2016) Studies of RNA sequence and structure using nanopores. *Prog. Mol. Biol. Transl. Sci.* 139, 73–99
131. Ip, C.L.C. *et al.* (2015) MinION analysis and reference consortium: phase 1 data release and analysis. *F1000Res.* 4, 1075

132. Jain, M. *et al.* (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17, 239
133. Deamer, D. *et al.* (2016) Three decades of nanopore sequencing. *Nat. Biotechnol.* 34, 518–524
134. Wick, R.R. *et al.* (2019) Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 20, 129
135. Bayley, H. (2015) Nanopore sequencing: from imagination to reality. *Clin. Chem.* 61, 25–31
136. Zorkot, M. *et al.* (2016) The power spectrum of ionic nanopore currents: the role of ion correlations. *Nano Lett.* 16, 2205–2212
137. Rang, F.J. *et al.* (2018) From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* 19, 90
138. Koonchanok, R. *et al.* (2019) Sequoia: an interactive visual analytics platform for interpretation and feature extraction from nanopore sequencing datasets. *BioRxiv* Published online October 11, 2019. https://doi.org/10.1101/801811
139. Dai, X. *et al.* (2018) Identification of YTH domain-containing proteins as the readers for N1-methyladenosine in RNA. *Anal. Chem.* 90, 6380–6384
140. Genenncher, B. *et al.* (2018) Mutations in cytosine-5 tRNA methyltransferases impact mobile element expression and genome stability at specific DNA repeats. *Cell Rep.* 22, 1861–1874
141. Fu, L. *et al.* (2014) Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* 136, 11582–11585
142. Yang, Y. *et al.* (2019) RNA 5-methylcytosine facilitates the maternal-to-zygotic transition by preventing maternal mRNA decay. *Mol. Cell* 75, 1188–1202
143. Chen, X. *et al.* (2019) 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nat. Cell Biol.* 21, 978–990
144. Zou, F. *et al.* (2020) *Drosophila* YBX1 homolog YPS promotes ovarian germ line stem cell development by preferentially recognizing 5-methylcytosine RNAs. *Proc. Natl. Acad. Sci. U. S. A.* 117, 3603–3609
145. Yang, L. *et al.* (2019) m5C methylation guides systemic transport of messenger RNA over graft junctions in plants. *Curr. Biol.* 29, 2465–2476
146. Ma, C.J. *et al.* (2019) AlkB homologue 1 demethylates N3-methylcytidine in mRNA of mammals. *ACS Chem. Biol.* 14, 1418–1425
147. Ringeard, M. *et al.* (2019) FTSJ3 is an RNA 2'-O-methyltransferase recruited by HIV to avoid innate immune sensing. *Nature* 565, 500–504
148. Nishikura, K. (2016) A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* 17, 83–96
149. Dominissini, D. *et al.* (2013) Transcriptome-wide mapping of N6-methyladenosine by m6A-seq based on immunocapturing and massively parallel sequencing. *Nat. Protoc.* 8, 176–189
150. Edelheit, S. *et al.* (2013) Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet.* 9, e1003602