

Detecting RNA modifications in the epitranscriptome: predict and validate

Mark Helm, Yuri Motorin

▶ **To cite this version:**

Mark Helm, Yuri Motorin. Detecting RNA modifications in the epitranscriptome: predict and validate. Nature Reviews Genetics, 2017, 18 (5), pp.275-291. 10.1038/nrg.2016.169 . hal-02956837

HAL Id: hal-02956837

https://hal.univ-lorraine.fr/hal-02956837

Submitted on 12 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Detecting RNA modifications in the epitranscriptome: predict and validate

Mark HELM[1] *and Yuri MOTORIN*[2]

[1]Institute of Pharmacy and Biochemistry, Johannes Gutenberg University Mainz, Staudingerweg 5, 55128 Mainz, Germany

[2]IMoPA UMR7365 CNRS-UL, BioPole Lorraine University, 9 avenue de la Foret de Haye, 54505 Vandoeuvre-les-Nancy, France

Correspondence to M.H: or Y.M.
e-mail : mhelm@uni-mainz.de or yuri.motorin@univ-lorraine.fr

Abstract

RNA modifications are emerging players in the field of posttranscriptional regulation of gene expression, and are attracting a comparable degree of research interest to DNA and histone modifications in the field of epigenetics. The true potential of only a handful out of more than 100 RNA modifications is currently emerging as the consequence of a leap in detection technology, principally associated with high-throughput sequencing. This review outlines the major developments in this field by a structured discussion of detection principles, lays out advantages and drawbacks of new high-throughput approaches, and presents conventional biophysical identification of modifications as meaningful ways for validation.

Introduction

Perception of the field of RNA modifications had been in a near dormant state until a few years ago when innovations in detection techniques unlocked several technology barriers, giving access to big data. On par with genomics and transcriptomics in terms of volume and impact, these data quickly uncovered numerous new layers of information involved in regulation of gene expression. The field of epigenetics is discovering an impact of RNA modifications that easily rivals that of modifications of proteins and DNA in terms of research activity and breakthrough results [1] . Consequently, the term "epitranscriptomics" was coined [2-4] to include posttranscriptional alterations that do not affect the RNA sequence *per se* [5, 6]. In quick succession, numerous completely unsuspected features were discovered that gave impetus to the field. In particular, the dynamic nature of modifications in many noncoding RNAs and in mRNAs gave rise to the notion that we are about to reveal a previously invisible code that resides in nucleic acids, yet outside their sequence.

The detection of RNA modifications has a longstanding history, just like that of DNA modifications, reaching back to the early days of molecular biology. Indeed, pseudouridine (Ψ), the most frequent posttranscriptional modification and therefore frequently called the fifth nucleoside [7], was only discovered 2 years after the first modified DNA nucleoside, 5 methylcytosine (5mC). Since then, a stream of newly detected modifications has provided 100-150 structures of naturally occurring and chemically distinct ribonucleosides [8]. The rapid speed of these advances is clearly due to technical progress in detection methodologies, and there are no indications of an end to this development. Early discoveries relied on separation by chromatography and organic chemistry for structural elucidation, but were soon supplemented with electrophoresis and mass spectrometry. This simultaneously drove progress in related areas such as sequencing and structural probing of nucleic acids. The richest source of modifications, both old and new, is tRNA with up to 25% of its nucleotides modified, which also features the largest chemical variety and sophistication in its modifications [8, 9]. Modifications range from simple methylations to sophisticated multistep transformations, including the incorporation of a

variety of low-molecular-weight metabolites known from other pathways; a few examples are shown in Figure 1A. The limited size and relative abundance of tRNA allowed the determination of a sequence context for these modifications, by chromatographic separation of small fragments after nuclease digestion or selective chemical cleavage [10]. Detection, identification, and especially determination of a sequence context of modifications in the much larger rRNAs was consequently much harder, and its progress was a consequence of the development of sequencing techniques using e.g. slab gel electrophoresis and capillary electrophoresis [11]. In comparison to the high diversity and large number of modifications in these highly abundant noncoding RNAs, relatively few modifications were discovered (and most discoveries were made by serendipity), in other, less-abundant species including small nuclear RNAs (snRNAs), small nucleolar RNAs, snoRNAs and microRNAs (miRNAs) [12-15]. Of note, mRNA was known to contain various modifications such as *N6*-methyladenosine (m^6A), 5-methylcytidine (m^5C), and ribose methylations for decades [15-18], but despite the obvious potential of mRNA modifications for regulation of gene expression, advancement of the field stalled at this stage for decades, until recently mapping of these and other modifications in mRNA by high-throughput sequencing techniques led to a resurgence in the field [19, 20], and are ultimately responsible for its current highly visible state.

What were the problems of old, and why has the field been revolutionized by techniques based on deep sequencing? The present review will address these questions by discussing established technology as well as recent breakthroughs. Importantly, whereas biophysical methods such as chromatography and mass spectrometry used to be the principal means for detection and quantification of RNA modifications on a case-by-case basis, current studies report datasets of hundreds to thousands of sites, predicted from high-throughput sequencing data to contain RNA modifications. Because different high-throughput detection techniques are error-prone to varying degrees, and because the field is seeing an ever increasing number of published datasets, it is clear that an important number of published modification sites are error fraught, or, in other words, plain wrong. To avoid damage to biological interpretations, we posit to treat sites predicted from big data as candidate sites to avoid widespread over-interpretation of modification data, until such time, when a

dataset has been validated with at least one additional independent method. To this end, we discuss principles, advantages, and weaknesses of methods based high-throughput sequencing, as well as the possibilities for validation of candidate sites by chemical and biophysical methods.

Sequencing RT signatures of RNA modifications

Traces of modifications left in cDNA.

During reverse transcription (RT), modifications may disturb the faithful generation of the inverse carbon copy that is cDNA. Confronted with a non-canonical substrate, the enzyme may still incorporate the complementary deoxy-ribonucleoside triphosphate (dNTP) that would conform to Watson-Crick pairing with an unmodified template, as is for example the case for pseudouridine, ribothymidine (T), or m^5C, which are RT-silent, and leave no or very faint traces in cDNA. However, larger modifications or modifications on the Watson-Crick face tend to either lead to reverse transcription arrest, or to misincorporation of a non-complementary dNTP. The former case leaves information in the form of abortive cDNA ending at or near the site of modification, and the latter type of events manifests as apparent mutations in the cDNA sequences [21-23] (see Figure 1B). Depending on the nature of modified nucleotide, the proportions of correct cDNA synthesis, misincorporation and abortive products vary. For example, inosine (I), which is derived from A by deamination, base-pairs with C and leads to incorporation of dCTP, rather than dTTP, into cDNA. The faithfully resulting misincorporation signal, which can be distinguished from an actual genomically encoded mutation by comparison with genomic DNA, was the first to be investigated in several transcriptome-wide studies [24, 25] because of its ease of detection.

The characteristics of RT-signatures consisting of misincorporation and abortive cDNA products are not well studied for the vast majority of modified residues. Our recent investigations uncovered the consequences of 1-methyladenosine (m^1A) in template RNA [26, 27] for cDNA synthesis and their use in machine-learning-based detection of respective patterns in high-throughput sequencing data. Because comprehensive studies including both misincorporation and abortive RT

events are not yet available for the large majority of modifications, their application to the detection of naturally occurring modifications is limited [28]. However, the RT-blocking and misincorporation properties of certain modifications are frequently used for detection, e.g. in m1A mapping after antibody-mediated enrichment [29, 30], or in structural probing applications where modifications such as m^1A and 3-methylcytidine (m^3C) are artificially created by treatment with alkylating agents [31, 32].

Enhancing RT-signatures by chemical treatment.

An additional dimension of information in high-throughput sequencing data is accessible after treatment of the RNA template with reagents that specifically react with modifications in a way that changes their RT-signature in terms of either reverse transcription arrest or misincorporation. For example, Inosine-specific CyanoEthylation by acrylonitrile (ICE-Seq) [33, 34] generates strong RT-stop, allowing real A→I conversion sites to be distinguished from simple sequencing errors. More recently, the well-known reaction specificity for Ψ residues [35] of N-Cyclohexyl-N′-β-(4-Methylmorpholinium)ethylCarbodiimide Toluensulfonate (CMCT), a water soluble diimide, (see Figure 1C) was successfully applied by several independent studies for pseudouridine mapping in yeast and human transcriptomes [36-39]. After CMCT treatment of RNA, significant amounts of RT-arresting adducts occur at G, U and Ψ residues, but while the former two types can be hydrolyzed during an additional mild alkaline treatment, the relevant adduct in pseudouridine resists hydrolysis under these conditions and therefore can be detected as RT-arrest signals. Another approach is the recent adaptation of bisulfite-based detection of m^5C in RNA [40-45]. In the course of this technique, which was adapted from the well-established detection of 5mC [46] and 5hmC [47] in DNA, modification of cytosine to m^5C or hm5C [48] (and also probably m^4C and ac^4C) protects these sites from the effects of bisulfite treatment, whereas unmodified cytosine are deaminated to uridines. Guanosines in the resulting cDNA thus stem from modified cytidines, thus allowing facile detection in DeepSeq data. A number of further reagents with specificities for modifications (shown in Table 1), known from decades of research, await their application to transcriptome-wide analyses [49]. An approach used for the validation of MeRIP data on m^1A modifications in mRNA makes use of the Dimroth

rearrangement undergone by m^1A under alkaline conditions, which results in the formation of m^6A and thus an alteration of the RT-signature upon treatment [30]. An interesting new perspective is the use of enzymes with demethylating activity, that alter the RT-signature by removing a defined subset of modifications, thereby enabling their detection by differential analysis [50, 51].

Modification calling, drawbacks and pitfalls.

The term modification "calling", is used in the field for the identification of a position that is, according to the method in question, a good modification candidate. High-throughput sequencing based modification calling in an entire transcriptome must inevitably face problems rooted in statistics. Even extremely well-defined RT-signatures like that of inosine become less predictive when confronted with a strong background noise. The larger the query transcriptome, the higher the chances are that noise produces a signal that resembles that of a positive instance and thus leads to a false positive identification. This problem appears to be particularly pronounced in pseudouridine-calling: an analysis of four cardodiimide-based transcriptome analyses termed Pseudo-seq [38], Ψ-seq [36] , PSI-seq [37], and CeU-seq [39] found disturbingly little overlap [52], although further studies might still reveal that this could have been due to different biological settings including e.g. stress conditions.

RNA bisulfite sequencing also suffers from noise, in particular from incomplete deamination of unmodified cytidines, thus producing false positives. This is a consequence of the reaction conditions for deamination, which must be conducted at less stringent, i.e. less alkaline conditions for RNA than for DNA, since RNA is prone to phosphodiester backbone hydrolysis under alkaline conditions. DNA contaminations, which survive longer under these conditions, must therefore be rigorously excluded. Excessive deamination and the concurrent degradation of RNA lead to insufficient coverage especially of low abundant RNA species. This effect is exacerbated because deamination of cytidines reduces the information content of a given sequence, and thus a bisulfite sequencing read needs to be correspondingly longer than a conventional read to allow reliable mapping. As with pseudouridine, prediction of m^5C modification sites in abundant tRNA and rRNA is robust [44], but modification calling in low-abundance RNAs is fraught with high error. For example, the vast majority of an

initially reported >10000 m^5C sites[41] outside tRNA and rRNA lack reproduction or confirmation by independent approaches [53].

Ways for improvement of the signal-to-noise ratio include in particular an increase in coverage, either by augmenting the sequencing depth, or by enrichment of RNA sequences containing modifications (see below). Additionally, as the number of studies profiling RNA modifications increases, it is becoming ever more clear that at least a subset of called modification candidates must be verified by independent methods to ensure reliability [1].

Enriching selected RNA populations for sequencing

Cellular RNA is predominantly composed of abundant rRNA and tRNA species (> 90%), depending on cell type. This means that an enrichment step prior to high-throughput analysis is generally highly beneficial, as otherwise the overwhelming majority of sequencing reads typically map to well-known rRNA and/or tRNA species. The most popular approach is to isolate the so-called $polyA^+$ fraction, which is composed of mRNA and other polyadenylated RNAs, by affinity purification with poly-dT. However the resulting fraction still contains some rRNA, and many non-coding RNA species (such as snRNAs, snoRNAs and miRNAs) are excluded from the analysis. Alternatively, or in addition, rRNA depletion protocols are available that remove that vast majority of rRNA molecules using complementary DNA oligonucleotides, albeit leaving tRNA contaminations that still affect final results.

Role and impact of antibodies against modifications.

Beyond these widespread and commercially supported protocols for polyA+ enrichment and/or rRNA depletion, more specific pre-selection procedures have been developed for modified nucleotide mapping, that rely an affinity based enrichment of sequences containing a given modification.

Except for highly modified eucaryotic tRNA species, populations of cellular RNAs generally contain an estimated 0.1-1% of modified nucleotides. Thus, specific enrichment of modified fragments is necessary to reduce sequencing requirements and map modified nucleotides to a given RNA regions.

Antibodies are highly useful tools that are known for their potential to recognize molecular structures with very high affinity and specificity. Applications to nucleic acids include analytics of DNA modifications [54, 55], and efforts to raise specific antibodies against modified nucleotides in RNA have a long-standing history [56-59]. That history was low key until commercial antibodies enabled the enrichment of m^6A-containing mRNA fragments [60, 61], whose mapping revealed pronounced preferential occurrence in sequence elements crucial to mRNA translation, such as stop codons, splice sites, and 3′ untranslated regions (3′ UTRs). Further currently available antibodies are available against m^1A [29, 30], m^5C [62], and hm^5C [63].

However, even for those antibodies that are efficient against the synthetic haptens employed in their generation (typically an IO_4^- oxidized ribonucleoside coupled to a neutral protein such as bovine serum albumin (BSA) or keyhole limpet hemocyanin (KLH) [58]), many studies reported that specificity and affinity was poor for RNA molecules containing the same modified residue [62]. Arguably, the structural context of an RNA chain, which includes in particular multiple negative charges of the backbone phosphates, might be sufficiently different from the single modified nucleoside used to raise the antibody that binding and recognition of the antigen is substantially impeded in the context of an RNA chain. Thus, among the various antibodies of variegated specificity available on the market, including multipüle different directed against m^6A, only a few are highly specific [64]. Although some are by now characterized [65], e.g. m^6A-targeted antibodies, and m^5C-targeted antibodies, the absence of a rigorous routine characterization is keenly felt in the field .

Regardless of potential biases, antibodies have clearly revolutionized the field in numerous m^6A-specific methylated RNA immunoprecipitation (MeRIP)[66] applications, which were developed in analogy to MeDIP from the DNA modification field [67]. MeRIP-based mapping of m^6A, m^5C, and m^1A residues have revealed the presence and (approximate) location of these residues in mRNA and other poorly characterized RNA species (Figure 2A).

All antibody-based enrichment techniques allow mapping of specific RNA modifications to a given region (generally 100 nt or so in length), potentially containing modified residues. More precise localization is generally based on the search for specific consensus sequence for respective

modification enzyme (if known), or on various cross-linking techniques allowing to narrow down the detection window.

Covalent crosslink with cognate proteins.

As antibody binding provides only non-covalent complexes with modified RNA, the stringency of washing steps and thus enrichment is limited. A significant improvement was obtained by a UV-induced crosslinking step after formation of non-covalent complexes between modified RNAs and an antibody. In addition, covalent cross-linking of specific antibodies to modified site in RNA leaves specific RT-signatures in the corresponding sequencing profile, thus allowing approximate or fully single-nucleotide resolution, depending on the degree to which the antibody has been characterized [64]. Several variants of this technique have been proposed. For example, combining methylated RNA immunoprecipitation (MeRIP) with individual-nucleotide resolution crosslinking and immunoprecipitation (m^6A individual-nucleotide resolution crosslinking and immunoprecipitation" (miCLIP))[64] leverages the natural reactivity of RNA nucleobases excited at UV 254 nm, and combining MeRIP with photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (photo-crosslinking-assisted m^6A sequencing strategy (PA-m^6A-seq))[61] involves the prior incorporation of 4-thiouridine (s^4U) in to the RNA *in vivo* to increase the efficiency and specificity of the crosslink at UV 365 nm (Figure 2B).

Approaches similar to antibody crosslinking have also been developed for other proteins interacting with modifications, in particular the enzymes that themselves catalyze the RNA modifications. Of note, simple crosslinks of a given modification enzyme are likely to reflect an enzyme's affinity to the RNA, but not necessarily enzymatic turnover. In particular, tRNA modifying enzymes such as yeast Pus1 [68] display similar affinities for substrate and non-substrate tRNAs alike. For example, iCLIP experiments with DnmA/Dnmt2, a tRNA m^5C-methyltransferase (MTase) from *Dictyostelium discoideum,* led to the isolation of U2 snRNA containing a tRNA-like sequence; however this sequence was subsequently found not to be modified *in vivo*, as interrogated by bisulfite sequencing [53]. Actual substrates of m^5C-MTases were caught by trapping the covalent intermediate

that is an obligatory part of the catalytic mechanism [69]. In certain m^5C-MTases, one of two cysteine residues in the active site is required for resolving this intermediate and thus for enzyme release. Mutation of this so-called "regeneration" cysteine residue leads to a suicide reaction and formation of an irreversible covalent bond with the target cytidine residue. This technique was successfully applied for characterization of RNA substrates of the human Nsun2 and Nsun4 proteins [48, 70, 71] (Figure 2C).

As a related concept, alteration of the RNA substrate can also stall the enzymatic reaction and lead to stable enzyme–RNA complexes that are suitable for pull-down and mapping. When 5-azaC, an inhibitor known to form irreversible covalent intermediates similar to those described above, is added to cell culture growth medium it is randomly incorporated into cellular RNA. Following 5-azaC incorporation, immunoprecipitation of m^5C-MTases Nsun2, Nsun6 and Dnmt2 and then deep sequencing of the attached RNA has been used to identify RNAs modified by these MTases[72-74] (Figure 2C insert). Interestingly, the sequences obtained showed frequent nucleotide misincorporations at the site of crosslink, which is a phenomenon known from CLIP experiments with s^4U [75], as well as from crosslinking with certain m^6A-targeted antibodies [64]. These misincorporation sites point out the neighbouring nucleotides around of modification site and thus provide approximate single-nucleotide resolution. Of note, formation of similar covalent complexes between certain pseudouridine-synthases and 5-fluorouridine containing RNA substrates was also observed in the past, but not yet explored for coupling to next-generation sequencing [76, 77].

Clickable or biotinylated specific chemicals.

Small-molecule reagents present a viable alternative to antibodies for affinity-based enrichment of modified RNA fragments. In contrast to antibody but similar to CLIP-type approaches, they selectively form covalent adducts with modified residues. For subsequent purification, an affinity tag such as biotin, or a functional group for the conjugation of an affinity tag are required. In contrast to antibodies, which have a recognition domain for noncovalent interaction that is considerably larger than the modified residues themselves, small-molecule reagents achieve selective enrichment via differential chemical reactivity. This feature typically resides in the electronic property of the

nucleobases, thus representing a fundamentally different concept for discrimination. Successful applications of small-molecule reagents include the tagging of nicotinamide residues at the 5′ end of bacterial RNA [78] and an interesting improvement of the CMCT-based pseudouridine detection [39]. Here, the use of an azide-containing CMCT derivative in so-called CeU-Seq allowed subsequent attachment of an alkynylated biotin via click chemistry, and a subsequent enrichment of pseudouridine-containing RNAs prior to high-throughput sequencing analysis (Figure 2D). This study, which also included validations of several new sites in rRNA and mRNA by the SCARLET method (see below), reported 4-5 times more sites than each of the other CMCT based approaches.

Sequencing library preparation methods.

High-throughput sequencing using current technologies typically requires the attachment of particular sequences to the nucleic acid molecules destined to be sequenced. The protocol by which the addition of these so-called adapters is implemented, is generally called library preparation. In the case of RNA analysis, library preparation consists in conversion to dsDNA using reverse transcription and second strand synthesis, concomitant with attachment of appropriate adapters, which may include primer binding sites for later amplification, as well as barcodes. The latter are short 6 or 8 nucleotide sequences allowing sample multiplexing, meaning the use of the mix for several samples in the same sequencing lane. After sequencing, the obtained reads are sorted according to barcode and attributed to a given sample. Another common meaning for barcode is a sequence of random nucleotides (designated e.g. N6 for six consecutive degenerated positions) which aid during the bioinformatic analysis e.g. to eliminate multiple redundant copies of the same cDNA. The choice of a library preparation protocol depends on the sequencing method used for RNA modification detection. All RT-based techniques applied to long RNAs, with or without previous chemical modification, require single-base precision in mapping of the 3′ end of the cDNA (where the cDNA synthesis aborted opposite the modification site), and thus only a few popular protocols used for gene expression analysis are suitable for this purpose (Figure 3A). The first option is a direct single-stranded ligation

of a 3′-blocked DNA oligonucleotide to the cDNA 3′ end, using T4 RNA ligase [36]. Another established approach is based on GG (or CC) tailing of the cDNA 3′ end, followed by ligation of a DNA duplex with a corresponding CC (or GG) overhang, for second strand synthesis [26] (Figure 3B left). Yet another approach uses the so-called CircLigase protocol initially developed for ribosome profiling [79] (Figure 3B middle). Finally, an alternative method is the templated extension of the cDNA 3′ end with a random (N6) terminal extension oligonucleotide[1] (Terminal-Tagging Oligonucleotide, TTO, Figure 3B right). In all cases, these steps are followed by PCR amplification, typically including barcoding.

Approaches that do not require single-base precision include MeRIP protocols, which ignore abortive cDNA or approaches of the Ribose Methylation Sequencing type (RiboMethSeq) [80, 81] (Figure 3D), where the precise definition of fragments resides in the 5′- and 3′-adapter ligation protocols. Both published RiboMethSeq protocols are based on reduced cleavage of RNA phosphodiester bond 3'-adjacent to 2'-O-Me residue in RNA. Random alkaline cleavage produces collection of fragments starting and ending at all RNA positions, except at position +1 to 2'-O-Me residue. Fragments are thus converted to the library, sequenced and a 5'-end (or cumulated 5'-and 3'-end) counting profile is used for detection of modification. In these cases standard library preparation protocols already developed for small RNA (typically miRNA) analysis are appropriate [80], or non-conventional ligation techniques have been invented [81]. Ribose 2'-O-methylation in yeast rRNA was quantified by both published RiboMethSeq approaches and compared with the data obtained by direct rRNA analysis by mass-spectrometry [82]. Even though exact methylation ratios vary between the approaches used, the overall results are very coherent for fully and partially modified rRNA sites.

A most important caveat with all high-throughput approaches is related to various biases, which affect library preparation, sequencing and data analysis. At the library preparation step, one has to consider possible biochemical biases depending on the sequence content, including nucleotide biases like AU *versus* GC-rich regions of RNA. A relatively well understood example for the preference of enzymes for certain sequences or nucleotides are ligases, whose efficiency is

[1] http://www.nature.com/app_notes/nmeth/2010/101310

particularly affected by the nature of nucleotides at 5′- and 3′-ends of RNA fragment [83-85]. Sequence context is also known to affect the efficiency of random priming of RT-reaction using N6-DNA oligonucleotides, whereas the presence of modified nucleotides at the 5′- and/or 3′-end of the RNA considerably affects ligation efficiency [86] and also the PNK-catalyzed 5'-phosphorylation step required to convert free 5'-OH to 5'-phosphate prior to ligation step [87] (Figure 3C).

Bioinformatics and data mining.

Bioinformatics play an essential role in mapping and prediction of modified nucleotides from high-throughput sequencing data. Particular approaches have been developed (and more will have to be developed in the future), depending on the nature of an RT-signature in the sequencing profile, or a specific chemical treatment [34, 40]. Method development and adaptation is also required for mining data from enrichment techniques, which may differ substantially between e.g. antibody-based and small-molecule-based enrichment.

As a note of caution, every method and every algorithm relies on threshold parameters for modification calling that must ultimately be defined by the user and may thus contain a measure of ambiguity. Noise in the sequencing data creates fluctuations in every such parameter, and is more likely to produce e.g. false positives with increasing numbers of positions to be evaluated. Given the huge numbers of such sites in a transcriptome, all deep sequencing based methods are liable to produce false positives as well as false negatives in significant numbers that will depend on the settings of the threshold parameters.

Typical parameters for evaluating the performance of a method include accuracy, selectivity, specificity, false positive and false negative rates, positive and negative predictive values, as well as false discovery rate and the area under the curve (AUC) in a receiver operating characteristics (ROC) curve. Assessment of these parameters requires representative test data sets that include known, and ideally, independently validated sites of modification. Therefore, method development for data mining from transcriptome-wide RNA modification profiles begins and ends with databases. Until

recently, only highly specialized databases were available with data sorted by RNA type (e.g. tRNA, rRNA, snRNA or snoRNA), which considerably limits possibilities for automatic data mining (see URL links at the end of the manuscript). The availability of large transcriptomic-wide datasets for m^5C, m^6A and pseudouridine stimulated the creation of more general databases such as the RNA Modification Base (RMBase), which now includes all this information [88]. The field would certainly profit from an organized collection and possibly curation of new datasets according to common standards. The overall performance of a method thus not only depends on chemical treatment and library preparation steps, but to a significant degree also on the performance of bioinformatics pipelines and the robustness of training dataset used for parameters optimization.

Biophysical validation of candidate modification sites

Given the uncertainties and pitfalls outlined above, confidence in a dataset of modification candidates can and should be significantly improved by validation using an independent method. Beyond an indirect confirmation of a candidate site e.g. through the absence of a corresponding signal in knockout organisms (such as those lacking the likely enzyme catalyzing the addition of that mark), biophysical methods that provide the most confidence include those based on chromatography and/or and mass spectrometry. These methods, which have been mentioned in the introduction as the tools of trade before the advent of deep sequencing, now serve as powerful validation tools. Biophysical properties that distinguish modified nucleosides from canonical ones include molecular mass changes as analyzed by mass spectrometry, as well as altered lipophilicity as exploited through chromatographic separation.

Mass spectrometry.

The apparently most straightforward approach to the high-throughput and sequence-specific analysis of transcriptomes comes from adapting methodologies that have been extensively used in proteomics, in which complete hydrolysis after defined residues, typically effected by treatment with the protease trypsin, generates a peptide fragment library, which is subsequently separated by chromatography.

The eluting fragments are subjected to mass spectrometry analysis, which can detect modifications from fragmentation patterns and by comparison to the calculated mass of the unmodified fragment [89]. The concept as such, has been applied to RNA (Figure 4A), but suffers from two major drawbacks. First, although nucleases such as RNase T1 and MC1 [90] allow the preparation of fragment libraries, the protocols, tools and databanks that are commonplace in proteomics are mostly lacking in the RNA field [91]. Essentially, because the high attention paid to RNA modifications is still a recent phenomenon, only a few laboratories have developed relevant expertise [92]. Second, RNA, as a polyanion is much harder to transfer into the vacuum that is mandatory for mass spectrometry than is a neutral molecule. As the jargon goes, RNA "flies" less effectively than proteins (and less effectively even than the slightly less polar DNA), and consequently, the amount of RNA sample required for a comprehensive or even partial analysis of an mRNA population exceeds that achievable by current experimental protocols by several orders of magnitude. Hence, this type of analysis is currently restricted to abundant RNA species [93]. As an excellent example for the state of the art, the tRNA transcriptome of *E. coli* has recently been sequenced with >80% coverage by mass spectrometry [94].

A widespread method to detect and quantify RNA modifications involves complete enzymatic degradation of an RNA sample to nucleosides (Figure 4B), which "fly" much easier than oligonucleotides in mass spectrometry analytics, and allow detection and quantification in the femtomol to attomol range. The nucleoside mixture is typically resolved on a reverse phase chromatography column, from which the eluting species are transferred to a mass spectrometer for analysis. Biophysical parameters that allow the identification of a given nucleoside include its chromatographic retention time, its molecular mass, and its fragmentation pattern. While numerous laboratories have developed expertise for precise quantification [92, 95-100], the method does not provide any sequence information. The wider sequence context must therefore be introduced by other means, typically involving oligonucleotides such as biotinylated DNA for hybridization-based affinity purification of selected RNA sequences [26, 101, 102]. In an alternative approach, nuclease-protection was achieved by duplex formation with complementary DNA oligonucleotides (50-75 nucleotides in length), such that unhybridized RNA was then degraded by Mung bean or other nucleases [103], and the

remaining duplexes subjected to nucleoside analysis by liquid chromatography – mass spectrometry (LC–MS) [80, 104]. Other methods that have been applied to excision of defined fragments from larger RNAs include application to site-specific cleavage of RNase H and short DNA oligonucleotides (or DNA/2′-O-Me-RNA chimeric oligonucleotides), or by DNAzymes [105] (Figure 5A).

Chromatography.

Retention behavior, which is exploited in LC–MS approaches, reflects a biophysical property that also forms the basis for thin-layer chromatography (TLC) approaches for the identification and quantification of RNA modifications. Methods have been developed for nucleosides as well as nucleotides [106, 107], but detection of minute amounts relies on a ^{32}P radioactive label. Of the different solid phases, microcrystalline cellulose is most widely used in a TLC system that employs pairwise combinations of up to three different solvent systems [107]. The retention values of around 70 modified nucleotides have been mapped in three dimensions [108]. As outlined in the following paragraph, increasingly sophisticated methods have been developed to introduce some measure of sequence information within the labeling protocol, which typically comprises phosphorylation with [γ-^{32}P]-ATP.

Pre- and post-labeling techniques.

Pre-labeling designates techniques in which the ^{32}P-label is introduced to the RNA prior to the modification event (Figure 5B). *In vitro*, this is typically achieved *via* transcription by the T7 RNA polymerase in the presence of [α-^{32}P]-NTPs [107]. After modification, a protocol named nearest neighbor analysis, including digestion with RNase T2 and subsequent TLC analysis, identifies the nucleotide to the 3′ of a detected modification, and thus reveals a limited measure of sequence information on the modification site. Labeling *in vivo*, or in cell culture, is typically conducted by growth in ^{32}P-containing medium, and leads to uniform labeling [18]. Analysis of a nucleotide digest of total RNA, of subfractionated RNA, or of an isolated RNA species allows an assessment of all modifications present, as well as their relative quantification [18, 101], but does not provide sequence information.

Due to the above mentioned lack of sequence information from pre-labeling techniques, the validation of modification sites in cellular RNA preparations typically relies on post-labeling approaches, which are designed to selectively label a given position in a sequence context. Current methods in principle derive from the Stanley–Vassilenko approach [109], which combines random hydrolysis to generate free 5′-OH groups, post-labeling by 5′-phosphorylation using [γ-^{32}P]ATP, subsequent size fractionation, release of the labeled 5′-terminal nucleotide monosphosphate, and finally TLC analysis (Figure 5C). This results in a powerful method, which was particularly successful in its application to tRNA sequencing [110]. However, because the random hydrolysis step necessitates pure and precisely size fractionated RNA, the method was further developed to include hydrolysis that is sequence specific rather than random. Several molecular tools have been adapted to sequence-specific RNA cleavage in this context, including RNase H [111, 112] and DNAzymes [113, 114]. The most recent and sophisticated development is the site-Specific Cleavage And Radioactive-Labeling followed by ligation-assisted Extraction and Thin-layer chromatography (SCARLET) variant (Figure 5C). In this case, the RNA of interest is first cleaved at a desired site with RNAse H whose cleavage is directed by a chimeric oligonucleotide comprising DNA and 2′-O-Me-RNA. The free 5′-extremity is then ^{32}P-labeled and connected to the 3′ of a long DNA oligonucleotide by splinted ligation. The RNA part is subsequently degraded by RNases, and the ^{32}P-labeled residue is purified along with the DNA oligonucleotide, to which it is attached. The labeled nucleotide is then released by nuclease P1 mediated hydrolysis prior to TLC analysis [115-117]. In contrast to other methods, SCARLET involves two, rather than one, sequence dependent enrichment steps (Figure 5C), which is likely the reason for its successful application in complex mixtures like total RNA or mRNA preparations. In several such applications it was used to validate high-throughput data on m^{6}A [115] and Ψ [39] residues in mRNAs and lncRNAs .

Oligonucleotide-based validation techniques

Methods for validation of modification candidates that input sequence information *via* oligonucleotides are not limited to the sequence hydrolysis approaches outlined above. One other

technique uses a ligase, which is somewhat sensitive to the nature of 3′- and 5′-nucleotides in ligation reactions, and employed specifically modified splint DNA [118, 119]. This method has so far been developed for validation of m^6A, 2′-O-Me and pseudouridine.

The presence of modifications blocking conventional basepairing — such as m^1A, 1-methylguanosine (m^1G) and *N2*,*N2*-dimethylguanosine (m^2_2G) — lowers the stability of duplexes formed with cDNA oligonucleotides. On the basis of this principle, a DNA-chip based technique was developed that detects modifications in a microarray [120].

With some limitations, primer extension by RT can also be used as a validation method of candidates from high-throughput mapping of RT-arrest signals. This applies especially if the conditions of the RT-based validation differ substantially from the original library preparation protocol, e.g. in terms of enzyme, buffer, or Mg^{2+} concentrations, or if the generation of the RT-signal is based on a principle different from the original screening method. A pertinent example is modification calling after alkaline hydrolysis according to the RiboMethSeq method [80, 81], validated by RT at low concentrations of dNTPs. Abortive cDNA that is preferably generated at sites of 2′-O-methylations can either be directly analyzed by polyacrylamide gel electrophoresis (PAGE) [121], or in semi-quantitative PCR or qPCR assays [122, 123].

Designing RNA modification mapping experiments

Depending on the nature of modified nucleotide to be studied, the exact steps of a mapping project may vary, but several essential points are always conserved. As a general rule, thorough evaluation of the pros and cons (table 2) should include the potential for applying different methods to the detection of the same modification.

Know the reagent

The selectivity of an envisaged chemical reagent or antibody is often incompletely characterized, especially in the context of high-throughput sequencing. Indeed, the large numbers of potential modification sites in a typical transcriptome can render a chemical reagent prone to produce false

positives in large numbers. For example, a spike-in control in bisulfite sequencing of RNA was reportedly converted with 99.5-99.8% efficiency [41], the remaining 0.2-0.5% giving rise to up to 10^5 incomplete deamination signals in a hypothetical transcriptome of $2x10^7$. In such instances, multiple coverage combined with thorough statistical analysis is required, to determine the probability of multiple deamination hits on the same site, which would erroneously be called a modification. In the case at hand, a minimum coverage of 10 reads and a bisulfite conversion of 80% were set as threshold values for modification calling. Note that quantitative numbers for selectivity are not available for the majority of reagents described as "selective" for a modification in literature (Table 1), because selectivity was typically determined by biochemical methods without numeric evaluation [49]. Both reactivity and selectivity of a chemical reagent towards modified nucleotides vary as a function of diverse reaction parameters including temperature, solvent, and in particular, pH [124]. An optimization of these parameters by comparing modified versus unmodified RNA in a suitable readout is strongly advisable. A further point to consider is that not all sites of a structured RNA are equally accessible to a reagent, a feature exploited in so-called structural probing experiments [31, 32]. Note that conducting a reaction in denaturing solvent such as DMSO is liable to affect selectivity [124, 125].

Similar considerations apply to antibodies, some of which were apparently used in good faith, *i.e.* without an experimental verification of the suppliers description of selectivity [42]. Testing for specificity e.g. with synthetic modified *versus* unmodified RNA [63, 64, 66] should be considered before engaging in MeRIP-type experiments.

Know your stats

As all other high-throughput approaches, RNA modification mapping is prone to generate multiple false-positive hits. Among the numerous recent publications revolving around modification calling, few are designed to raise awareness about the actual or presumed fraction of wrongly called sites (false positive rate, fallout,) and the non-detected true sites (false negative, miss rate). A full set of statistical parameters, including specificity, selectivity and also a Receiver operating characteristics (ROC) curve, should become standard to facilitate an assessment of a given method's performance.

Read mapping, while being a field onto itself, should observe some basic concepts. Changes in mapping results upon variation of such parameters as multiple mapping and mismatch allowance should be critically surveyed. Multiple copies of reads originating from the same RNA molecule, arising from amplification of the same cDNA, can be avoided by barcoding concepts implemented in adapters and primers.

Development of new methods typically includes testing against a training set of known modification sites. Since the methods performance characteristics will be more significant with higher instance numbers, the training set should be conceived as large as possible. Also, higher read coverage will improve the signal-to-noise ratio, making modification sites in highly abundant tRNA and rRNA "too easy" to serve as a valid confirmation for modification calls on less abundant RNA species. Analysis of at least three biological replicates should become an obligate requirement. We posit that entire sets, issued from modification site calling, should be considered, and continue to be addressed as, "candidates" until some sites from less abundant RNA species, i.e. presumably of low coverage, can be validated by an orthogonal method [39].

Validate

Techniques used for the validation of selected sites from a modification calling based on high-throughput sequencing should employ a different detection principle. Confidence will increase, the more the validation approach differs in its nature, e.g. an MS-based approach or a SCARLET assay will be more convincing for validation of a polymerase-based modification calling than a qPCR approach, because the pitfalls and error sources differ more strongly (Table 2). Less biophysical approaches to increased confidence include spike-in experiments akin to the use of an internal standard, and the parallel investigation of organisms that are functionally depleted of relevant modification enzymes. Here, genome-editing techniques i.e. true knock-outs are preferable over RNAi based knockdowns, which are frequently incomplete and may lead to unknown levels of residual modification activity.

Conclusions

With the field of RNA modification having received a tremendous boost from innovations in detection technology, scientists are on the lookout for methods with even better performance. Apart from better detection efficacy and higher throughput, a feature that is currently missing in most of the detection technologies is the simultaneous detection of different modifications in the same biological sample or, preferably, in a single RNA molecule. Suitable technologies would have to processively screen entire single molecules, and would circumvent amplification events that erase traces of modifications. Because they might reveal several modifications in the same molecule, such data would, among other things, be able to inform on the combinatorial status of those different modifications in a given molecule, and thus yield highly interesting insight into the sequence of multiple modification events [126], as well as an interplay between different modification enzymes.

Again, innovative sequencing technologies hold the most promise. Using the technology employed in SMRT sequencing, the molecular events during reverse transcription that eventually lead to the RT-signature can be analyzed by single molecule spectroscopy in nanoscale wells, and can provide direct evidence of the efforts of the polymerase to bypass a modification such as m^6A [2, 127]. However, the presented data suggest that, while an influence of the modification is clearly visible, the inverse route, i.e. methods to derive the presence of a modification from a specific signal pattern, may take some time in coming.

Taking a cue from the DNA modification field, where Nanopore technology can detect multiple and different modification in a single chain [128], a similar development in RNA sequencing seems promising, although currently still in the development stage.

As opposed to the surprising and sudden influx of deep sequencing into the field in 2012 [60, 66], today's community is watching every development with anticipation.

Figure legends

Figure 1 | **Reverse-transcription based techniques for detection of modified nucleotides.** Legend. **a** | Chemical structures of common modified residues discussed in this review. **b** | Components of Reverse Transcription (RT) signatures depend on the type of RNA modification. Top – RT-silent residue (m^5C), middle and bottom – RT-blocking residue (m^1A). Exemplary coverage profiles and corresponding misincorporation sites are shown on the bottom **c** | Induction or enhancement of an RT signature using modification-specific chemicals.

Figure 2 | **Enrichment strategies in analysis of RNA modifications.** Legend. **a** | Methylated RNA ImmunoPrecipitation (MeRIP). RNA is fragmented to pieces of 100-150 nucleotides, modification-containing fragments are enriched by Antibody (Ab) pulldown, and converted into a sequencing library. **b** | Combination of MeRIP and individual-nucleotide resolution Cross-Linking and ImmunoPrecipitation (iCLIP) or Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP). After fragmentation, RNA is cross-linked at reactive residues (indicated by red bars) with a specific Ab at either 254 nm UV-light and natural RNA (iCLIP) or at 365 nm UV-light and s^4U residues incorporated into RNA (PAR-CLIP). The Ab is then removed and an RT profile reveals misincorporation and stops corresponding to cross-linked residues. **c** | An active-site mutated modification enzyme makes a covalent link to its cognate target nucleotide in RNA. Protein is then removed by proteinase K digestion and the cross-link position is determined by primer extension and sequencing. Alternative technique consists in inclusion of 5-azacytosine in RNA substrate, in this case the enzyme makes irreversible covalent link to target cytosine (boxed). **d** | Selection of modified RNAs (or fragments) using modification-specific chemicals that are "clickable" or biotinylated for affinity purification. The method employs a derivative of a modification–specific reagent containing biotin or a reactive group to which biotin can be selectively attached in a secondary reaction. Enrichment of biotin-modified fragments is effected with avidin beads, followed by library preparation and sequencing.

Figure 3 | **Library preparation issues.** Legend. **a** | To detect RT-arrest signals, mapping of cDNA 3′-ends is a crucial step to be included in the library preparation protocol (top). Initiation of RT primer extension for long and short RNAs. Short RNAs require ligation of 3-adapter for RT priming, while for long RNAs random priming with a DNA oligonucleotide containing 6 degenerated nucleotides (NNNNNN) is appropriate (bottom). **b** | Treatment of cDNA ″-end. Exact definition of cDNA 3′-ends can be achieved by three alternative techniques, namely by (i) homonucleotide tailing with terminal transferase (NN-tailing), (ii) the CircLigation protocol using 5'-phosphorylated DNA primer and CircLigase, and (iii) the use of templated cDNA extension with a Terminal-Tagging Oligonucleotide, (TTO) blocked at its 3'-end. In all cases, library preparation protocols end with PCR amplification and appropriate barcoding for multiplexed sequencing. **c** | Possible sources of ligation biases at 5'- and 3'-extremities of modified RNA. Ligation efficiency and degree of phosphorylation (if used) depend on the type of canonical or modified nucleotides at either end. **d** | RiboMethSeq approach for mapping of 2′-O-Me residues in RNAs. The approach is based on a bias in fragmentation, since 2′-O-Me residues protect adjacent phosphodiester bonds from alkaline cleavage. The major steps of the RiboMethSeq protocol show how this bias is exploited for detection and quantification.

Figure 4 | **Mass-spectrometry approaches for validation of RNA modification.** Legend. **a** | RNA fragments are prepared by complete digestion with RNase T1, which generates a catalogue of fragments ending with a G residue at the 3′-end. These fragments separated by HPLC, their molecular mass determined by MS and the presence or absence of modified residues deduced by MS/MS fragmentation or comparison with the theoretical molecular mass of unmodified fragments. **b** | Liquid chromatography coupled to classical or tandem mass-spectrometry (LC-MS/MS). Isolated RNA or RNA fragment are degraded to the nucleosides, which are then separated by reverse-phase HPLC and further analyzed by mass-spectrometry. Parameters providing information for the identification of a given modification include retention time, molecular mass, and fragmentation patters from MS/MS.

Figure 5 | **Methods for validation of modified RNA nucleotide.** Legend. **a** | Post-labeling of RNAs and their analysis. RNA is metabolically labeled with ^{32}P, followed by isolation of individual RNA species, hydrolysis to 5'-NMPs and 2D-TLC analysis. **b** | Techniques used for site-specific RNA cleavage by DNAzymes or RNase H include 5'-end radioactive labeling of fragments with [γ-^{32}P]ATP and polynucleotide kinase (PNK). Labeled fragments are size fractionated and analyzed by TLC after digestion to 5'-NMPs, or submitted to the SCARLET protocol. This includes an additional sequence selection step during a splint ligation, followed by size fractionation, digestion, and TLC analysis. **c** |Specific RNA fragments can be isolated after protection from nuclease digestion in a DNA-RNA duplex, by use of specific DNAzymes or by directed RNAse H cleavage with complementary DNA oligonucleotides composed of DNA/2′-O-Me RNA residues. Modifications in thus obtained fragments can be detected and quantified by LC-MS/MS.

Table 1 | **Small molecule reagents specific for RNA modifications**

A	Reagents applied to detection by deep sequencing			
reagent	modification detected	reaction type	detection	reference
bisulfite	m^5C	deamination	m^5C remains unconverted	40;41;42
N-cyclohexyl-N'-β-(4-methylmorpholinium) ethylcarbodiimide p-tosylate (CMCT)	pseudouridine	*N*-acylation	RT- arrest	35;129;38;36, 37
acryolnitrile	inosine	*N*-alkylation	RT- arrest	34
methylthiosulfonate-biotin (MTS-biotin)	s^4U	disulfide formation	affinity enrichment	130
B	Reagents not yet applied to detection by deep sequencing			
acid	Wybutosine	depurination	AICS; RT-arrest	131, 132
Sodium borohydride	m^7G	reduction, depurination	AICS; RT-arrest	133, 134
mild alkali treatment	dihydrouridine	cycle opening	RT-arrest	135
hydrazine	m^3C	nucleophilic addition	AICS; RT-arrest	133
isothiocyanate-derivatives, NHS-derivatives	primary amines	*N*-alkylation	tagging	136
iodoacetamido-compounds	Thiol-containing nucleotides, e.g. s^2U,s^4U,	*S*-alkylation	tagging	137
methylvinylsulfonate	pseudouridine	*N*-alkylation	mass spectrometry	129
acryolnitrile	pseudouridine	*N*-alkylation	mass spectrometry	129
N-acryloyl-3-aminophenylboronic acid (APB)	queuosine	affinity : chelation by cis-diols	retardation in gel electrophoresis	138
acrylo-phenylmercury (APM)	thiolated nucleotides, incl. s^2U,s^4U,	affinity : mercury-sulfur	retardation in gel electrophoresis	139

- AICS- Aniline induced chain scission

Table 2 | **Pros and cons of techniques for detection and validation of RNA modifications**

Method	**Application**		**Pros**	**Cons**
RT-signature from RT-arrest and misincorporation	detection	m1A	● straightforward protocol ● precise single nucleotide mapping of modification sites ● adaptable to different types of modifications	● negative control requires knockout strain of known modification enzyms ● Semi-quantitative
chemical reagent + RT-stop coupled to deep sequencing	detection	Ψ I	● Single nucleotide resolution ● Precise mapping of modification site	● reagents available only for few modified residues ● false positives from RNA structure and/or sequence ● detects only one type of modification
Ab-based enrichment coupled to deep sequencing	detection	m^6A m^5C m^1A	● enrichment by Ab reduces required sequencing depth	● no single nucleotide resolution ● ill-defined specificity of various commercial antibodies
Ab-enrichment and crosslink coupled to deep sequencing	detection	m^6A	● single nucleotide resolution ● enrichment by Ab reduces required sequencing depth	● ill-defined specificity of various commercial antibodies ● low crosslink efficiency
RiboMethSeq	detection/ quantification	2'-O-Me	● single nucleotide resolution ● Relative quantification of methylation rate ● ng amount of input RNA	● false positives due to RNA structure ● requires high sequencing depth for analysis ● currently restricted to abundant RNA species
bisulfite RNA sequencing	detection & quantification	m^5C	● single nucleotide resolution ● Quantification is possible	● noise from incomplete deamination in structured RNA regions ● requires significant amount of input material ● requires high sequencing depth for analysis ● mapping of reads is difficult (degeneracy)
mass-spectrometry of RNAse T1 digests	detection, validation	various	● includes sequence information ● single nucleotide resolution	● requires significant amount of input material ● requires highly specialized equipment ● requires highly developed expertise for data mining
LC-MS & LC-MS/MS of nucleosides	quantification, validation	various	● highly accurate quantification ● expertise is reasonably	● no sequence information ● requires specialized equipment

			widespread	
SCARLET and related methods	validation, quantification	m^6A Ψ	• includes sequence information • quantification possible • no specialized equipment	• single-site query – no high throughput • labor intense
RT-stop	validation	2'-O-Me	• includes sequence information • experimentally simple, e.g. PCR	• single-site query – moderate throughput only • indirect validation without physicochemical properties
DNA chip / hybridization efficiency	validation	various	• includes sequence information • multiple-site queries, at least medium throughput	• indirect validation • requires some technical expertise and equipment

Glossary

BARCODING – identification of samples in sequencing run by ligation of short specific sequences (often 6-8 degenerated nucleotides) attached to sequencing adapters. Allows unambiguous assignment of samples from the same sequencing lane, and/or identification of redundant, PCR-amplified reads that originate from the same RT event.

CLICK CHEMISTRY – is a class biocompatible (bioorthogonal) reactions, generally involving azide and alkyne moieties.

CIRCLIGATION – Circularization of cDNA ssDNA templates having a 5´-phosphate and a 3´-hydroxyl group with CircLigase™

DNAZYME – Artificially generated DNA sequence that contains catalytic activity, e.g. RNase activity.

HAPTENS -- small, otherwise immunosilent molecules that elicit an immune response only exposed to the immune system in the form of a covalent conjugate to a large carrier such as a protein.

LIPOPHILICITY -- “fat-loving” character of a compound, reflected in its ability to dissolve in non-polar solvents

NANOPORE TECHNOLOGY --technology that funnels molecules, including nucleic acids, through nanosized pores, thereby opening the possibility to determine their sequence. In the field of nucleic acids often synonymously used for High throughput sequencing through protein pores.

PULLDOWN – technique for selective isolation of RNA (or protein) species from complex biological sample. Generally relies on affinity capture with antibodies or another high-affinity interaction.

NUCLEOTIDE - smallest structural unit of a nucleic acid, composed of a nucleobase connected to a ribose (or deoxyribose) *via* a glycosidic bond, and a phosphate esterified to a hydroxylgroup of the sugar moiety. The combination of only nucleobase and sugar is referred to as a nucleoside.

REVERSE TRANSCRIPTION – the synthesis of single-stranded DNA (complementary DNA, or cDNA) by reverse transcriptase (RNA-dependent DNA polymerase) using single-stranded RNA as a template.

RECEIVER OPERATING CHARACTERISTICS (ROC) CURVE -- A graphical plot illustrating the performance of a binary classifier system as a function of variegated thresholds for true positives and false positives. Plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings leads to the ROC curve, whose area under the curve (AUC) is a measure of the performance. The true-positive rate is also known as sensitivity, The false-positive rate can be calculated as 1 − specificity.

RNA MODIFICATION – Posttranscriptional step of RNA maturation, resulting in chemically altered (modified) nucleotides in mature RNA chain.

RT-SIGNATURE – specific pattern composed of misincorporation in cDNA and RT-arrest of primer extension, induced by modified residues in an RNA chain

SPLINT LIGATION - technique to biochemically join to RNA fragments by enzymatic ligation. Both fragments are brought in spatial proximity by hybridization to a so called splint DNa, an oligodeoxynucleotide that is essentially a cDNA of the target RNA molecule.

1D AND 2D-TLC – mono (or two) dimensional thin-layer chromatography. Technique for separation performed in a thin layer of adsorbent material, called also stationary PHASE.

Acknowledgements

This work was supported by joint ANR-DFG grant HTRNAMod (ANR-13-ISV8-0001/HE 3397/8-1) and by the DFG priority programme SPP1784.

Online links:

http://modomics.genesilico.pl/

http://mirlab.sysu.edu.cn/rmbase/

http://trnadb.bioinf.uni-leipzig.de/

http://mods.rna.albany.edu/

https://people.biochem.umass.edu/fournierlab/3dmodmap/main.php;

FURTHER INFORMATION

Link 1: MODOMICS [http://modomics.genesilico.pl/]

Link 2: RMB database [http://mirlab.sysu.edu.cn/rmbase/]

Link 3: RNA modification database [http://mods.rna.albany.edu/]

Access to this interactive links box is free online.

References

1. Frye, M., Jaffrey, S.R., Pan, T., Rechavi, G. & Suzuki, T. RNA modifications: what have we learned and where are we headed? *Nat Rev Genet* **17**, 365-72 (2016).
2. Saletore, Y. et al. The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol* **13**, 175 (2012).
3. Schwartz, S. Cracking the epitranscriptome. *RNA* **22**, 169-74 (2016).
4. Witkin, K.L. et al. RNA editing, epitranscriptomics, and processing in cancer progression. *Cancer Biol Ther* **16**, 21-7 (2015).
5. Roundtree, I.A. & He, C. RNA epigenetics--chemical messages for posttranscriptional gene regulation. *Curr Opin Chem Biol* **30**, 46-51 (2016).
6. Liu, N. & Pan, T. RNA epigenetics. *Transl Res* **165**, 28-35 (2015).
7. Spenkuch, F., Motorin, Y. & Helm, M. Pseudouridine: still mysterious, but never a fake (uridine)! *RNA Biol* **11**, 1540-54 (2014).
8. Machnicka, M.A. et al. MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res* **41**, D262-7 (2013).
9. Helm, M. & Alfonzo, J.D. Posttranscriptional RNA Modifications: playing metabolic games in a cell's chemical Legoland. *Chem Biol* **21**, 174-85 (2014).
10. Holley, R.W. et al. Structure of a Ribonucleic Acid. *Science* **147**, 1462-5 (1965).
11. Ban, E. & Song, E.J. Recent developments and applications of capillary electrophoresis with laser-induced fluorescence detection in biological samples. *J Chromatogr B Analyt Technol Biomed Life Sci* **929**, 180-6 (2013).
12. Karijolich, J. & Yu, Y.T. Spliceosomal snRNA modifications and their function. *RNA Biol* **7**, 192-204 (2010).
13. Zinshteyn, B. & Nishikura, K. Adenosine-to-inosine RNA editing. *Wiley Interdiscip Rev Syst Biol Med* **1**, 202-9 (2009).
14. Yu, B. et al. Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932-5 (2005).
15. Motorin, Y. & Helm, M. RNA nucleotide methylation. *Wiley Interdiscip Rev RNA* **2**, 611-31 (2011).
16. Moss, B., Gershowitz, A., Stringer, J.R., Holland, L.E. & Wagner, E.K. 5'-Terminal and internal methylated nucleosides in herpes simplex virus type 1 mRNA. *J Virol* **23**, 234-9 (1977).
17. Nichols, J.L. & Welder, L. The modified nucleotide constituents of human prostatic cancer cell (MA-160) poly(A)-containing RNA. *Biochim Biophys Acta* **608**, 1-18 (1980).
18. Rottman, F.M., Desrosiers, R.C. & Friderici, K. Nucleotide methylation patterns in eukaryotic mRNA. *Prog Nucleic Acid Res Mol Biol* **19**, 21-38 (1976).
19. Limbach, P.A. & Paulines, M.J. Going global: the new era of mapping modifications in RNA. *Wiley Interdiscip Rev RNA* (2016).
20. Song, C.X., Yi, C. & He, C. Mapping recently identified nucleotide variants in the genome and transcriptome. *Nat Biotechnol* **30**, 1107-16 (2012).
21. Ebhardt, H.A. et al. Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. *Nucleic Acids Res* **37**, 2461-70 (2009).

22. Findeiss, S., Langenberger, D., Stadler, P.F. & Hoffmann, S. Traces of post-transcriptional RNA modifications in deep sequencing data. *Biol Chem* **392**, 305-13 (2011).
23. Ryvkin, P. et al. HAMR: high-throughput annotation of modified ribonucleotides. *RNA* **19**, 1684-92 (2013).
24. Alon, S. et al. Systematic identification of edited microRNAs in the human brain. *Genome Res* **22**, 1533-40 (2012).
25. Dominissini, D., Moshitch-Moshkovitz, S., Amariglio, N. & Rechavi, G. Adenosine-to-inosine RNA editing meets cancer. *Carcinogenesis* **32**, 1569-77 (2011).
26. Hauenschild, R. et al. The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Res* **43**, 9950-64 (2015).
27. Tserovski, L. et al. High-throughput sequencing for 1-methyladenosine (m1A) mapping in RNA. *Methods* (2016).
28. Gustafsson, C. & Persson, B.C. Identification of the rrmA gene encoding the 23S rRNA m1G745 methyltransferase in Escherichia coli and characterization of an m1G745-deficient mutant. *J Bacteriol* **180**, 359-65 (1998).
29. Li, X. et al. Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat Chem Biol* **12**, 311-6 (2016).
30. Dominissini, D. et al. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441-6 (2016).
31. Talkish, J., May, G., Lin, Y., Woolford, J.L., Jr. & McManus, C.J. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* **20**, 713-20 (2014).
32. Weeks, K.M. Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol* **20**, 295-304 (2010).
33. Sakurai, M., Yano, T., Kawabata, H., Ueda, H. & Suzuki, T. Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. *Nat Chem Biol* **6**, 733-40 (2010).
34. Suzuki, T., Ueda, H., Okada, S. & Sakurai, M. Transcriptome-wide identification of adenosine-to-inosine editing using the ICE-seq method. *Nat Protoc* **10**, 715-32 (2015).
35. Bakin, A. & Ofengand, J. Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* **32**, 9754-62 (1993).
36. Schwartz, S. et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-62 (2014).
37. Lovejoy, A.F., Riordan, D.P. & Brown, P.O. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PLoS One* **9**, e110799 (2014).
38. Carlile, T.M. et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143-6 (2014).
39. Li, X. et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol* **11**, 592-7 (2015).
40. Schaefer, M., Pollex, T., Hanna, K. & Lyko, F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res* **37**, e12 (2009).
41. Squires, J.E. et al. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* **40**, 5023-33 (2012).
42. Edelheit, S., Schwartz, S., Mumbach, M.R., Wurtzel, O. & Sorek, R. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet* **9**, e1003602 (2013).
43. Burgess, A.L., David, R. & Searle, I.R. Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. *BMC Plant Biol* **15**, 199 (2015).
44. Militello, K.T., Chen, L.M., Ackerman, S.E., Mandarano, A.H. & Valentine, E.L. A map of 5-methylcytosine residues in Trypanosoma brucei tRNA revealed by sodium bisulfite sequencing. *Mol Biochem Parasitol* **193**, 122-6 (2014).

45. Schaefer, M. RNA 5-Methylcytosine Analysis by Bisulfite Sequencing. *Methods Enzymol* **560**, 297-329 (2015).
46. Krueger, F., Kreck, B., Franke, A. & Andrews, S.R. DNA methylome analysis using short bisulfite sequencing data. *Nat Methods* **9**, 145-51 (2012).
47. Booth, M.J., Marsico, G., Bachman, M., Beraldi, D. & Balasubramanian, S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat Chem* **6**, 435-40 (2014).
48. Van Haute, L. et al. Deficient methylation and formylation of mt-tRNA(Met) wobble cytosine in a patient carrying mutations in NSUN3. *Nat Commun* **7**, 12039 (2016).
49. Behm-Ansmant, I., Helm, M. & Motorin, Y. Use of specific chemical reagents for detection of modified nucleotides in RNA. *J Nucleic Acids* **2011**, 408053 (2011).
50. Zheng, G. et al. Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods* **12**, 835-7 (2015).
51. Clark, W.C., Evans, M.E., Dominissini, D., Zheng, G. & Pan, T. tRNA base methylation identification and quantification via high-throughput sequencing. *RNA* (2016).
52. Zaringhalam, M. & Papavasiliou, F.N. Pseudouridylation meets next-generation sequencing. *Methods* (2016).
53. Jeltsch, A. et al. Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. *RNA Biol*, 1-16 (2016).
54. Poirier, M.C. Antisera specific for carcinogen-DNA adducts and carcinogen-modified DNA: applications for detection of xenobiotics in biological samples. *Mutat Res* **288**, 31-8 (1993).
55. Strickland, P.T. & Boyle, J.M. Immunoassay of carcinogen-modified DNA. *Prog Nucleic Acid Res Mol Biol* **31**, 1-58 (1984).
56. Sawicki, D.L., Erlanger, B.F. & Beiser, S.M. Immunochemical detection of minor bases in nucleic acids. *Science* **174**, 70-2 (1971).
57. Vold, B.S., Longmire, M.E. & Keith, D.E., Jr. Thiolation and 2-methylthio- modification of Bacillus subtilis transfer ribonucleic acids. *J Bacteriol* **148**, 869-76 (1981).
58. Milstone, D.S., Vold, B.S., Glitz, D.G. & Shutt, N. Antibodies to N6-(delta2-isopentenyl) adenosine and its nucleotide: interaction with purified tRNAs and with bases, nucleosides and nucleotides of the isopentenyladenosine family. *Nucleic Acids Res* **5**, 3439-55 (1978).
59. Horowitz, S., Horowitz, A., Nilsen, T.W., Munns, T.W. & Rottman, F.M. Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc Natl Acad Sci U S A* **81**, 5667-71 (1984).
60. Dominissini, D. et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-6 (2012).
61. Chen, K. et al. High-resolution N(6) -methyladenosine (m(6) A) map using photo-crosslinking-assisted m(6) A sequencing. *Angew Chem Int Ed Engl* **54**, 1587-90 (2015).
62. Mishima, E. et al. Immuno-Northern Blotting: Detection of RNA Modifications by Using Antibodies against Modified Nucleosides. *PLoS One* **10**, e0143756 (2015).
63. Delatte, B. et al. RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **351**, 282-5 (2016).
64. Linder, B. et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12**, 767-72 (2015).
65. Schwartz, S. et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**, 1409-21 (2013).
66. Meyer, K.D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635-46 (2012).
67. Jacinto, F.V., Ballestar, E. & Esteller, M. Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. *Biotechniques* **44**, 35, 37, 39 passim (2008).
68. Arluison, V., Buckle, M. & Grosjean, H. Pseudouridine synthetase Pus1 of Saccharomyces cerevisiae: kinetic characterisation, tRNA structural requirement and real-time analysis of its complex with tRNA. *J Mol Biol* **289**, 491-502 (1999).

69. Motorin, Y., Lyko, F. & Helm, M. 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Res* **38**, 1415-30 (2010).
70. Hussain, S. et al. NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep* **4**, 255-61 (2013).
71. Metodiev, M.D. et al. NSUN4 is a dual function mitochondrial protein required for both methylation of 12S rRNA and coordination of mitoribosomal assembly. *PLoS Genet* **10**, e1004110 (2014).
72. Khoddami, V. & Cairns, B.R. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat Biotechnol* **31**, 458-64 (2013).
73. Khoddami, V. & Cairns, B.R. Transcriptome-wide target profiling of RNA cytosine methyltransferases using the mechanism-based enrichment procedure Aza-IP. *Nat Protoc* **9**, 337-61 (2014).
74. Haag, S. et al. NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. *RNA* **21**, 1532-43 (2015).
75. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-41 (2010).
76. Gu, X., Liu, Y. & Santi, D.V. The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. *Proc Natl Acad Sci U S A* **96**, 14270-5 (1999).
77. Spedaliere, C.J. & Mueller, E.G. Not all pseudouridine synthases are potently inhibited by RNA containing 5-fluorouridine. *RNA* **10**, 192-9 (2004).
78. Cahova, H., Winz, M.L., Hofer, K., Nubel, G. & Jaschke, A. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **519**, 374-7 (2015).
79. Kielpinski, L.J., Boyd, M., Sandelin, A. & Vinther, J. Detection of reverse transcriptase termination sites using cDNA ligation and massive parallel sequencing. *Methods Mol Biol* **1038**, 213-31 (2013).
80. Marchand, V., Blanloeil-Oillo, F., Helm, M. & Motorin, Y. Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res* **44**, e135 (2016).
81. Birkedal, U. et al. Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew Chem Int Ed Engl* **54**, 451-5 (2015).
82. Taoka, M. et al. The complete chemical structure of Saccharomyces cerevisiae rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Res* (2016).
83. Hafner, M. et al. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**, 1697-712 (2011).
84. Munafo, D.B. & Robb, G.B. Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. *RNA* **16**, 2537-52 (2010).
85. Raabe, C.A. et al. The rocks and shallows of deep RNA sequencing: Examples in the Vibrio cholerae RNome. *RNA* **17**, 1357-66 (2011).
86. Shen, Y., Zheng, K.X., Duan, D., Jiang, L. & Li, J. Label-free microRNA profiling not biased by 3' end 2'-O-methylation. *Anal Chem* **84**, 6361-5 (2012).
87. Mourato, L.L., Beland, F.A. & Marques, M.M. 32P-Postlabeling of N-(deoxyguanosin-8-yl)arylamine adducts: a comparative study of labeling efficiencies. *Chem Res Toxicol* **12**, 661-9 (1999).
88. Sun, W.J. et al. RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res* **44**, D259-65 (2016).
89. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389**, 1017-31 (2007).
90. Addepalli, B., Lesner, N.P. & Limbach, P.A. Detection of RNA nucleoside modifications with the uridine-specific ribonuclease MC1 from Momordica charantia. *RNA* **21**, 1746-56 (2015).
91. Cao, X. & Limbach, P.A. Enhanced detection of post-transcriptional modifications using a mass-exclusion list strategy for RNA modification mapping by LC-MS/MS. *Anal Chem* **87**, 8433-40 (2015).

92. Wetzel, C. & Limbach, P.A. Mass spectrometry of modified RNAs: recent developments. *Analyst* **141**, 16-23 (2016).
93. Ross, R., Cao, X., Yu, N. & Limbach, P.A. Sequence mapping of transfer RNA chemical modifications by liquid chromatography tandem mass spectrometry. *Methods* **107**, 73-8 (2016).
94. Wetzel, C. & Limbach, P.A. The global identification of tRNA isoacceptors by targeted tandem mass spectrometry. *Analyst* **138**, 6063-72 (2013).
95. Kellner, S. et al. Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res* **42**, e142 (2014).
96. Sakaguchi, Y., Miyauchi, K., Kang, B.I. & Suzuki, T. Nucleoside Analysis by Hydrophilic Interaction Liquid Chromatography Coupled with Mass Spectrometry. *Methods Enzymol* **560**, 19-28 (2015).
97. Fu, L. et al. Simultaneous Quantification of Methylated Cytidine and Adenosine in Cellular and Tissue RNA by Nano-Flow Liquid Chromatography-Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal Chem* **87**, 7653-9 (2015).
98. Cai, W.M. et al. A Platform for Discovery and Quantification of Modified Ribonucleosides in RNA: Application to Stress-Induced Reprogramming of tRNA Modifications. *Methods Enzymol* **560**, 29-71 (2015).
99. Basanta-Sanchez, M., Temple, S., Ansari, S.A., D'Amico, A. & Agris, P.F. Attomole quantification and global profile of RNA modifications: Epitranscriptome of human neural stem cells. *Nucleic Acids Res* **44**, e26 (2016).
100. Huang, W. et al. Determination of DNA and RNA Methylation in Circulating Tumor Cells by Mass Spectrometry. *Anal Chem* **88**, 1378-84 (2016).
101. Helm, M., Florentz, C., Chomyn, A. & Attardi, G. Search for differences in post-transcriptional modification patterns of mitochondrial DNA-encoded wild-type and mutant human tRNALys and tRNALeu(UUR). *Nucleic Acids Res* **27**, 756-63 (1999).
102. Suzuki, T. & Suzuki, T. Chaplet column chromatography: isolation of a large set of individual RNAs in a single step. *Methods Enzymol* **425**, 231-9 (2007).
103. Sharma, S., Watzinger, P., Kotter, P. & Entian, K.D. Identification of a novel methyltransferase, Bmt2, responsible for the N-1-methyl-adenosine base modification of 25S rRNA in Saccharomyces cerevisiae. *Nucleic Acids Res* **41**, 5428-43 (2013).
104. Buchhaupt, M. et al. Partial methylation at Am100 in 18S rRNA of baker's yeast reveals ribosome heterogeneity on the level of eukaryotic rRNA modification. *PLoS One* **9**, e89640 (2014).
105. Bourgeois, G. et al. Eukaryotic rRNA Modification by Yeast 5-Methylcytosine-Methyltransferases and Human Proliferation-Associated Antigen p120. *PLoS One* **10**, e0133321 (2015).
106. Kellner, S., Burhenne, J. & Helm, M. Detection of RNA modifications. *RNA Biol* **7**, 237-47 (2010).
107. Grosjean, H., Droogmans, L., Roovers, M. & Keith, G. Detection of enzymatic activity of transfer RNA modification enzymes using radiolabeled tRNA substrates. *Methods Enzymol* **425**, 55-101 (2007).
108. Keith, G. Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* **77**, 142-4 (1995).
109. Stanley, J. & Vassilenko, S. A different approach to RNA sequencing. *Nature* **274**, 87-9 (1978).
110. Juhling, F. et al. tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* **37**, D159-62 (2009).
111. Yu, Y.T., Shu, M.D. & Steitz, J.A. A new method for detecting sites of 2'-O-methylation in RNA molecules. *RNA* **3**, 324-31 (1997).
112. Zhao, X. & Yu, Y.T. Detection and quantitation of RNA base modifications. *RNA* **10**, 996-1002 (2004).

113. Hengesbach, M., Meusburger, M., Lyko, F. & Helm, M. Use of DNAzymes for site-specific analysis of ribonucleotide modifications. *RNA* **14**, 180-7 (2008).
114. Pyle, A.M., Chu, V.T., Jankowsky, E. & Boudvillain, M. Using DNAzymes to cut, process, and map RNA molecules for structural studies or modification. *Methods Enzymol* **317**, 140-6 (2000).
115. Liu, N. et al. Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19**, 1848-56 (2013).
116. Liu, N. & Pan, T. Probing RNA Modification Status at Single-Nucleotide Resolution in Total RNA. *Methods Enzymol* **560**, 149-59 (2015).
117. Liu, N. & Pan, T. Probing N(6)-methyladenosine (m(6)A) RNA Modification in Total RNA with SCARLET. *Methods Mol Biol* **1358**, 285-92 (2016).
118. Saikia, M. et al. A systematic, ligation-based approach to study RNA modifications. *RNA* **12**, 2025-33 (2006).
119. Dai, Q. et al. Identification of recognition residues for ligation-based detection and quantitation of pseudouridine and N6-methyladenosine. *Nucleic Acids Res* **35**, 6322-9 (2007).
120. Hiley, S.L. et al. Detection and discovery of RNA modifications using microarrays. *Nucleic Acids Res* **33**, e2 (2005).
121. Maden, B.E., Corbett, M.E., Heeney, P.A., Pugh, K. & Ajuh, P.M. Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie* **77**, 22-9 (1995).
122. Aschenbrenner, J. & Marx, A. Direct and site-specific quantification of RNA 2'-O-methylation by PCR with an engineered DNA polymerase. *Nucleic Acids Res* **44**, 3495-502 (2016).
123. Dong, Z.W. et al. RTL-P: a sensitive approach for detecting sites of 2'-O-methylation in RNA molecules. *Nucleic Acids Res* **40**, e157 (2012).
124. Kellner, S., Kollar, L.B., Ochel, A., Ghate, M. & Helm, M. Structure-function relationship of substituted bromomethylcoumarins in nucleoside specificity of RNA alkylation. *PLoS One* **8**, e67945 (2013).
125. Kellner, S., Seidu-Larry, S., Burhenne, J., Motorin, Y. & Helm, M. A multifunctional bioconjugate module for versatile photoaffinity labeling and click chemistry of RNA. *Nucleic Acids Res* **39**, 7348-60 (2011).
126. Müller, M. et al. Dynamic modulation of Dnmt2-dependent tRNA methylation by the micronutrient queuine. *Nucleic Acids Res* **43**, 10952-62 (2015).
127. Vilfan, I.D. et al. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J Nanobiotechnology* **11**, 8 (2013).
128. Wescoe, Z.L., Schreiber, J. & Akeson, M. Nanopores discriminate among five C5-cytosine variants in DNA. *J Am Chem Soc* **136**, 16582-7 (2014).
129. Durairaj, A. & Limbach, P.A. Mass spectrometry of the fifth nucleoside: a review of the identification of pseudouridine in nucleic acids. *Anal Chim Acta* **623**, 117-25 (2008).
130. Duffy, E.E. et al. Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. *Mol Cell* **59**, 858-66 (2015).
131. Ladner, J.E. & Schweizer, M.P. Effects of dilute HCl on yeast tRNAPhe and E. coli tRNA1fMet. *Nucleic Acids Res* **1**, 183-92 (1974).
132. Schleich, H.G., Wintermeyer, W. & Zachau, H.G. Replacement of wybutine by hydrazines and its effect on the active conformation of yeast tRNAPhe. *Nucleic Acids Res* **5**, 1701-13 (1978).
133. Peattie, D.A. Direct chemical method for sequencing RNA. *Proc Natl Acad Sci U S A* **76**, 1760-4 (1979).
134. Wintermeyer, W. & Zachau, H.G. Tertiary structure interactions of 7-methylguanosine in yeast tRNA Phe as studied by borohydride reduction. *FEBS Lett* **58**, 306-9 (1975).
135. Xing, F., Hiley, S.L., Hughes, T.R. & Phizicky, E.M. The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. *J Biol Chem* **279**, 17850-60 (2004).

136. Plumbridge, J.A., Baumert, H.G., Ehrenberg, M. & Rigler, R. Characterisation of a new, fully active fluorescent derivative of E. coli tRNA Phe. *Nucleic Acids Res* **8**, 827-43 (1980).
137. Watson, B.S. et al. Macromolecular arrangement in the aminoacyl-tRNA.elongation factor Tu.GTP ternary complex. A fluorescence energy transfer study. *Biochemistry* **34**, 7904-12 (1995).
138. Igloi, G.L. & Kossel, H. Affinity electrophoresis for monitoring terminal phosphorylation and the presence of queuosine in RNA. Application of polyacrylamide containing a covalently bound boronic acid. *Nucleic Acids Res* **13**, 6881-98 (1985).
139. Igloi, G.L. Interaction of tRNAs and of phosphorothioate-substituted nucleic acids with an organomercurial. Probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry* **27**, 3842-9 (1988).

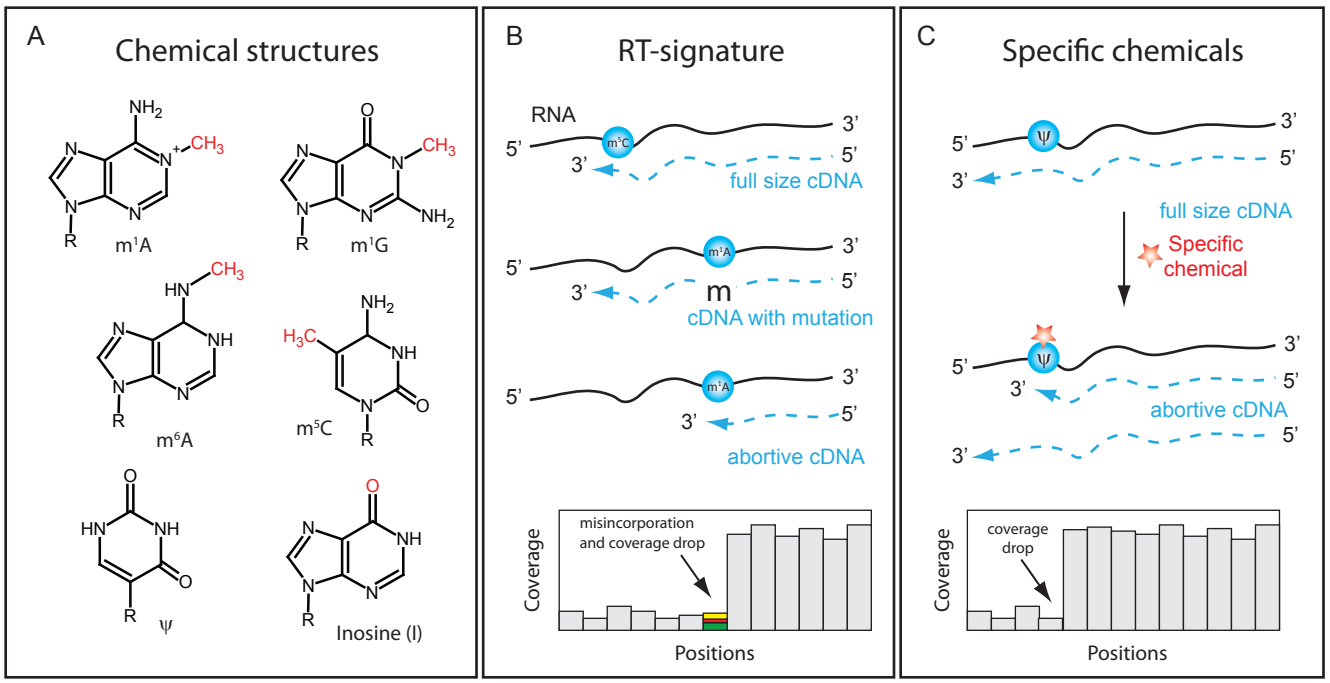


Figure 1 Helm & Motorin

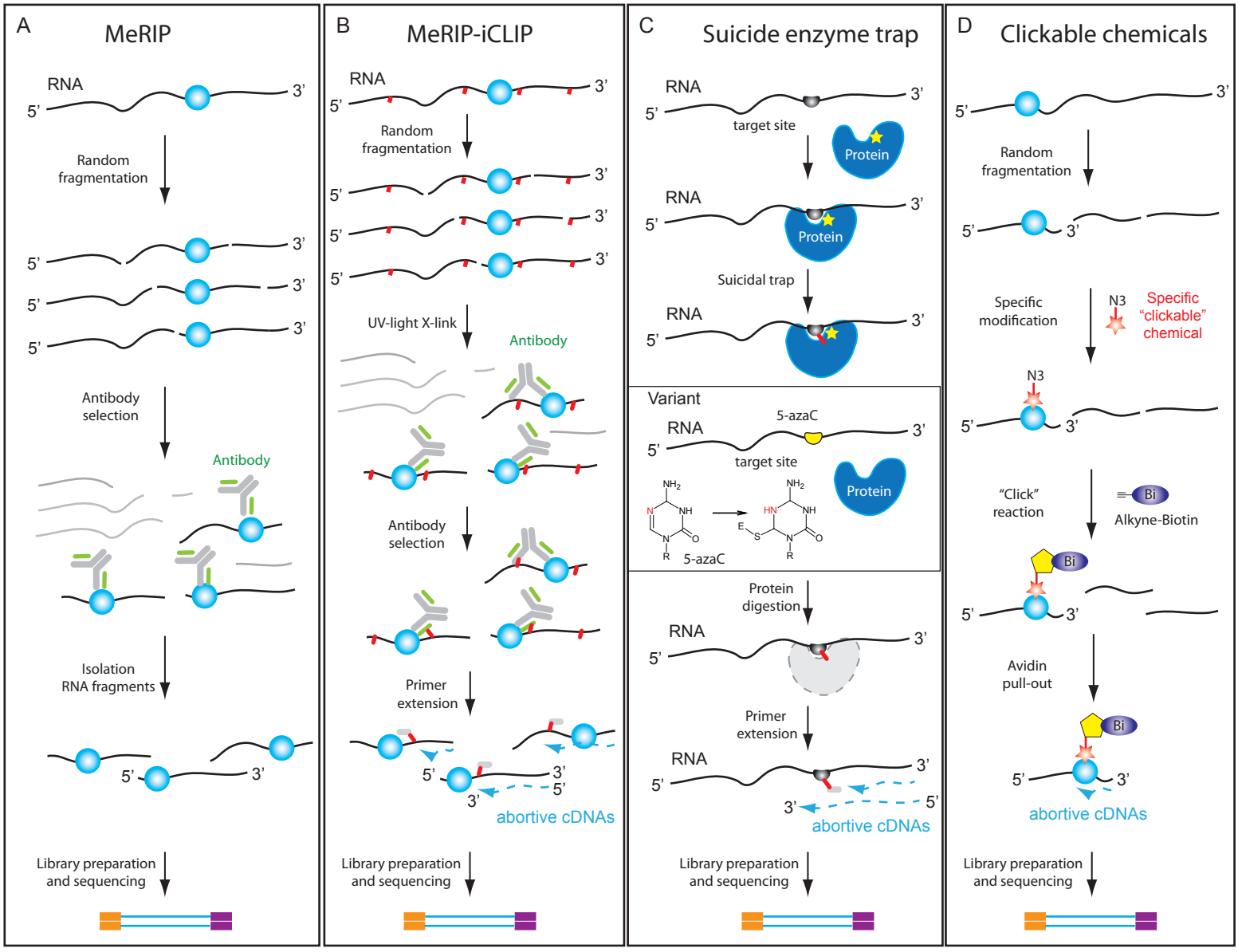


Figure 2 Helm & Motorin

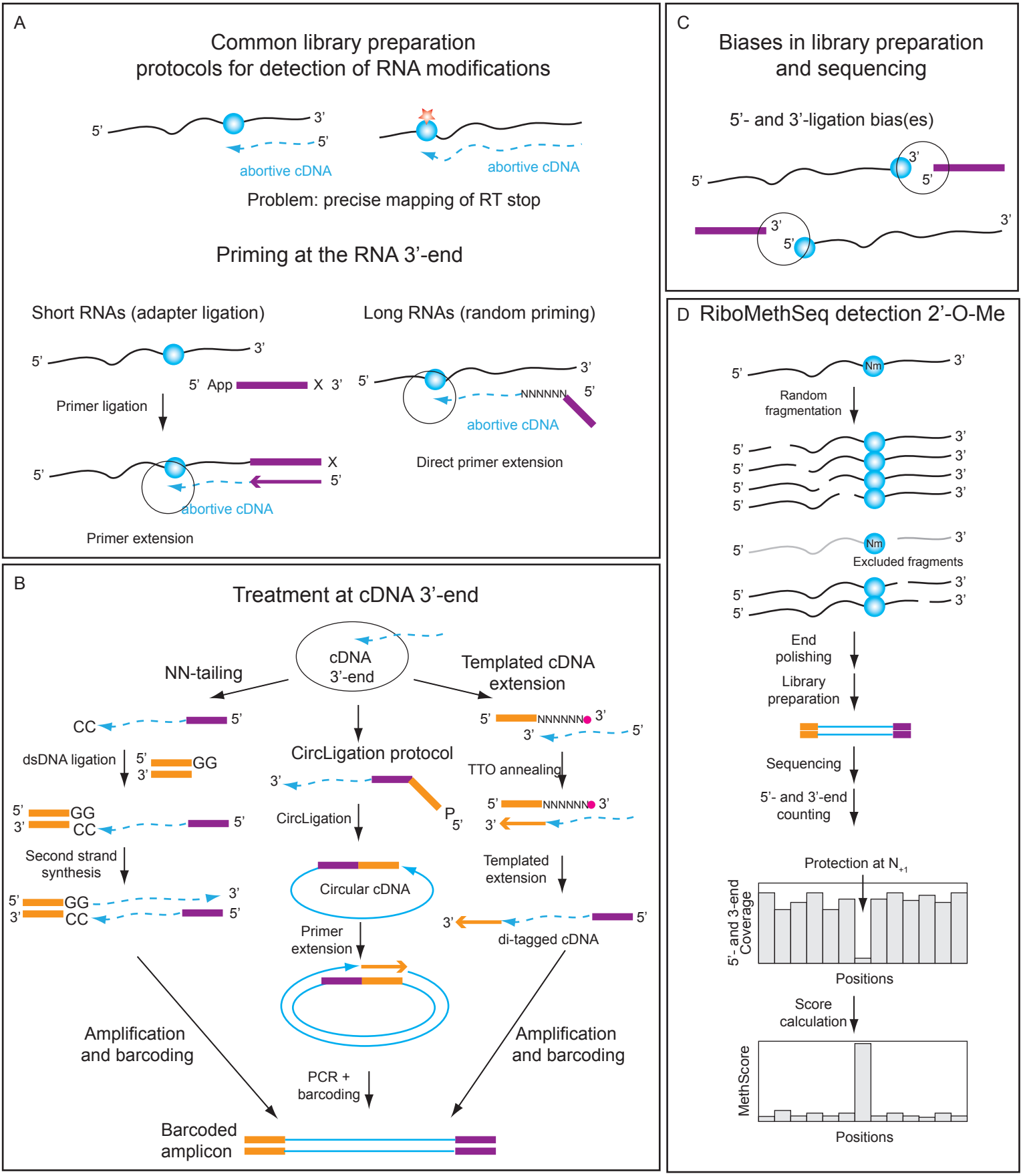


Figure 3 Helm & Motorin

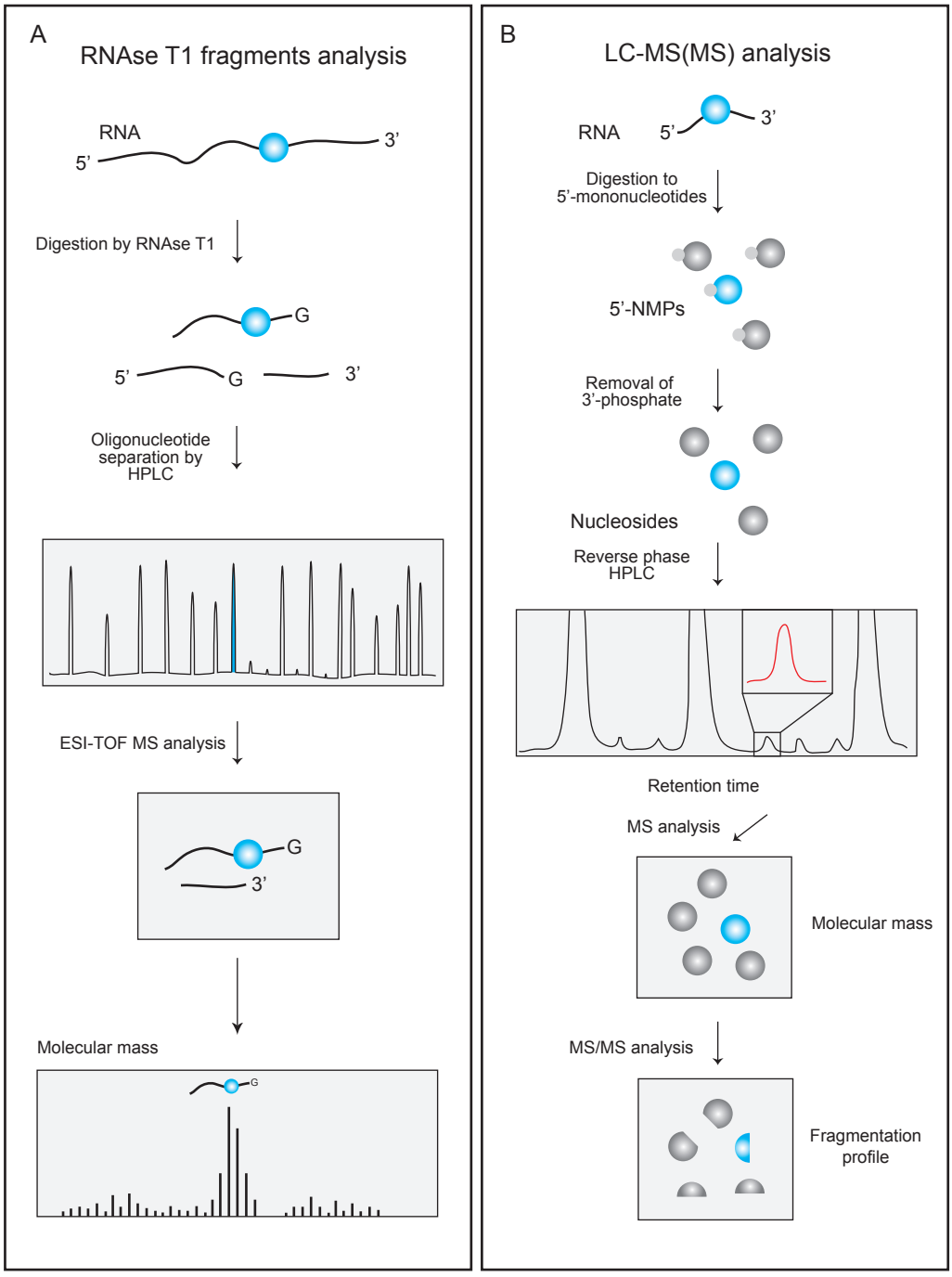


Figure 4 Helm & Motorin

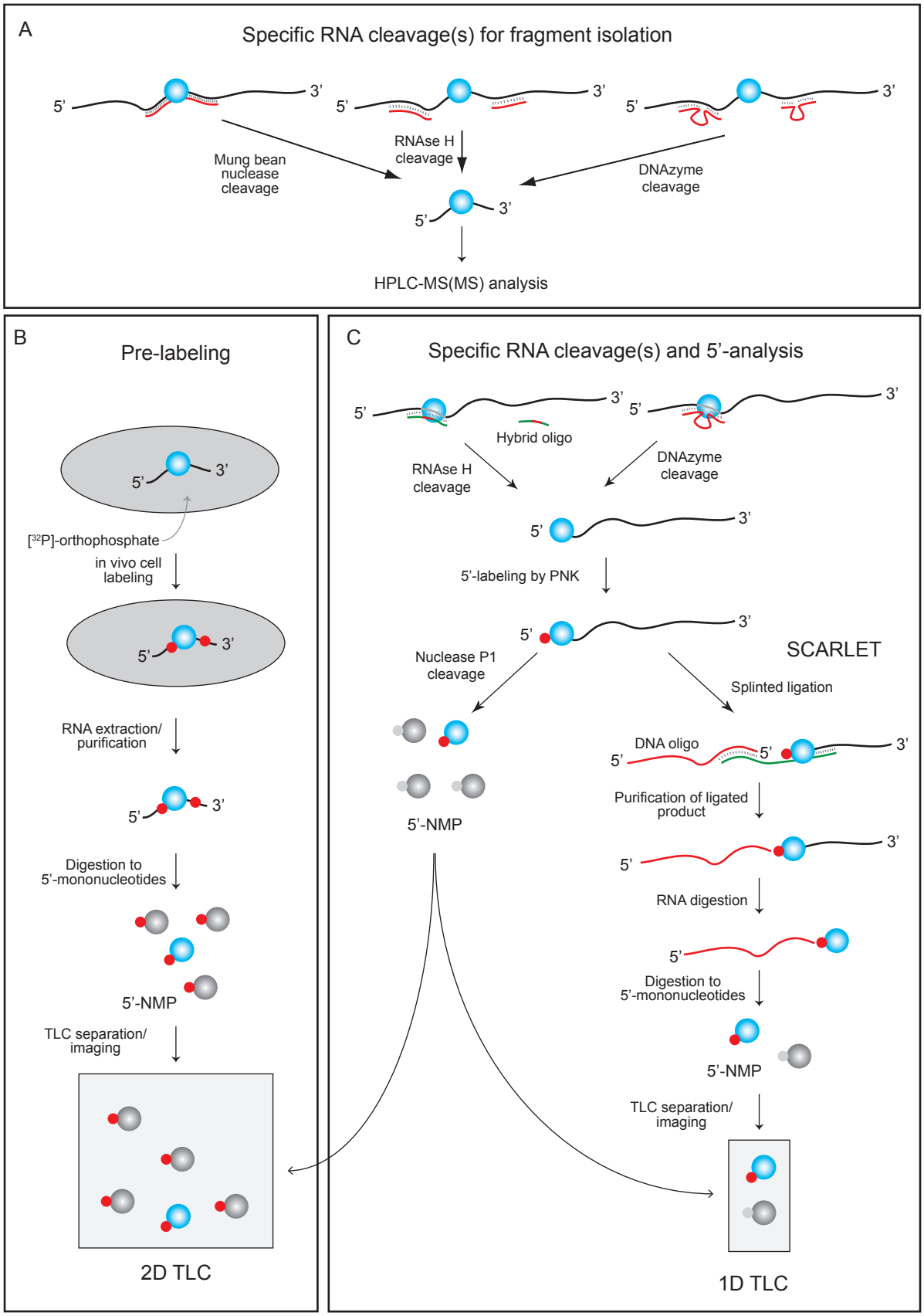


Figure 5 Helm & Motorin