nature cell biology

PERSPECTIVE

https://doi.org/10.1038/s41556-022-00880-5

Exploring the expanding universe of small RNAs

Junchao Shi[1], Tong Zhou[2]✉ and Qi Chen[1]✉

The world of small noncoding RNAs (sncRNAs) is ever-expanding, from small interfering RNA, microRNA and Piwi-interacting RNA to the recently emerging non-canonical sncRNAs derived from longer structured RNAs (for example, transfer, ribosomal, Y, small nucleolar, small nuclear and vault RNAs), showing distinct biogenesis and functional principles. Here we discuss recent tools for sncRNA identification, caveats in sncRNA expression analysis and emerging methods for direct sequencing of sncRNAs and systematic mapping of RNA modifications that are integral to their function.

Small noncoding RNAs (sncRNAs) are universally distributed in all kingdoms of life—from bacteria and archaea to various eukaryotic organisms[1–3]—and have not ceased to surprise us throughout the last two decades regarding their compositional and functional diversity. Although the definition of 'small' is relatively empirical and subjective in different contexts, in this paper we mainly discuss sncRNAs of 15–50 nucleotides (nt) in length, including the relatively well-characterized small interfering RNAs (siRNAs, 20–27 nt), microRNAs (miRNAs, 21–23 nt) and Piwi-interacting RNAs (piRNAs, 21–35 nt)[4–6], but with more focus on more recently discovered non-canonical sncRNAs (15–50 nt) that are derived from longer structured RNAs[7] such as transfer RNA (tRNA)[8,9], ribosomal RNA (rRNA)[10,11], Y RNA (yRNA)[11,12], small nuclear RNA (snRNA)[13,14], small nucleolar RNAs (snoRNA)[15,16], vault RNA (vtRNA)[17,18] and even messenger RNA (mRNA)[19,20]. Studies on non-canonical sncRNAs have recently gained momentum, exemplified by the new focus on tRNA-derived small RNA (tsRNA)[8], and are expected to expand to other categories with their systematic discovery. To facilitate communication and reduce confusion, we propose a unified naming system for these non-canonical sncRNAs (Box 1) when describing discoveries from different laboratories (usually using different names).

Like many noncoding RNAs in history, the emerging non-canonical sncRNAs were initially considered as merely random degradation products of RNA turnover/metabolism and thus neglected, yet increasing evidence has begun to put them in the spotlight as regulatory sncRNAs[8,21]. This is partly due to the revelation that they are regulated by both genetic and environmental factors[18,22–27] and that many of them are functional and associated with multiple diseases—including those linked to cancer[28–30], immunity[12,31], viral infection[32,33], neurological disorders[34,35], stem cells[26,36–39], retrotransposon control[40,41] and epigenetic inheritance[24,25,42–45]—as well as because in many cases, the exertion of their function depends on mechanisms that are distinct from those of well-studied siRNAs, miRNAs or piRNAs. Moreover, it was recently recognized that many non-canonical sncRNAs harbour various RNA modifications, some of which can prevent the detection of sncRNAs by traditional RNA sequencing (RNA-seq)[10,14,46,47]. This has promoted a recent wave of method improvements, leading to their comprehensive discovery and identification, which have in turn ignited interest in research centred on sncRNA modifications[48]. Here we briefly outline the biogenesis and functional principles of non-canonical sncRNAs and discuss recent methodological developments in promoting sncRNA discovery and accurate expression analyses as well as new techniques for direct multiplexed mapping of RNA modifications, which is necessary for decoding the full function of sncRNAs.

Distinct features of sncRNAs

The biogenesis and functions of siRNAs, miRNAs and piRNAs in eukaryotes have been extensively studied[5,6]. Both siRNA and miRNA are generated from double-stranded RNA precursors mainly by ribonuclease (RNase) III enzymes (for example, Dicer for siRNA, and Drosha and Dicer for miRNA)[4], whereas piRNA, found mainly in animal germline cells, is generated from single-stranded RNA precursors independently of Dicer and Drosha, involving a set of proteins for primary processing and the 'ping-pong cycle' for amplification[49]. The main functions of siRNAs, miRNAs and piRNAs all depend on base-pairing with their RNA and/or DNA targets, exerting RNA-silencing effects (for example, post-transcriptional mRNA cleavage, decay or translational repression and transcriptional silencing) via the Argonaute family proteins, where siRNAs and miRNAs are associated with the AGO sub-clade and piRNAs are associated with the PIWI sub-clade[50]. Notably, Argonaute-dependent RNA-silencing effects are generally believed to exist only in eukaryotes[50].

Compared with siRNA, miRNA and piRNA, the non-canonical sncRNAs bear several distinguishable characteristics regarding their evolutionary origin, cellular abundance, biogenesis and functional principles, which may update our traditional views on sncRNAs. For example, tsRNA and rRNA-derived small RNA (rsRNA) are predominantly found and dynamically regulated in ancient unicellular organisms (for example, bacteria, archaea, yeast and protozoa) where siRNA, miRNA and piRNA are absent[51–56]. This suggests that production of sncRNAs via the fragmentation or cleavage of longer structured RNAs (for example, tRNA, rRNA, snRNA, yRNA and vtRNA) may represent the most ancient pathway of sncRNA biogenesis that pre-date the emergence of siRNA, miRNA and piRNA[8]. In addition, the biogenesis of non-canonical sncRNAs involves the cleavage of their precursors (for example, tRNAs and rRNAs) by a range of ancient RNase families (for example, RNase P, RNase Z, RNase T2 and RNase A)[8] that pre-date the emergence of Dicer (which exists only in eukaryotes[50] and is responsible for generating siRNA and miRNA) and are profoundly affected by site-specific RNA modifications and related enzymes[8]. Finally, many non-canonical sncRNAs can exert versatile functions independent of Argonaute family proteins, exemplified in the recent emerging tsRNA studies[8], although our understanding of their full range of functionality is still in its infancy and awaits to be explored.

[1]Division of Biomedical Sciences, School of Medicine, University of California, Riverside, CA, USA. [2]Department of Physiology and Cell Biology, University of Nevada, Reno School of Medicine, Reno, NV, USA. ✉e-mail: tongz@med.unr.edu; qi.chen@medsch.ucr.edu

Box 1 | A unified naming system for sncRNAs derived from longer RNA precursors

Studies of non-canonical sncRNAs have been accumulating and have reached the critical mass to become a new branch of RNA biology. However, the lack of a unified naming system has led to a variety of naming styles. For example, sncRNAs derived from tRNAs have been reported by different laboratories in different contexts under different names, including tRNA-derived small RNAs (tsRNAs)[24,29,112], tRNA-derived small RNAs (tDRs)[56], tRNA-derived stress-induced RNAs (tiRNAs)[31,39,113] and tRNA fragments (tRFs)[28,43,114]. Here we propose a unified nomenclature for non-canonical sncRNAs that are derived from well-characterized longer RNA precursors, as shown in the table below, which is used throughout this paper to reduce confusion when describing discoveries from different laboratories and has the potential for further use in the research community. Although some laboratories may retain the initially reported names, it would be ideal to also include the new unified names in future publications to reduce confusion, especially for readers who are new to the field. More detailed naming criteria to label individual sncRNAs in each category (for example, tsRNAs) would need the group efforts of each community.

Precursor RNA	Derivative sncRNA
tRNA	tRNA-derived-small RNA (tsRNA)
rRNA	rRNA-derived-small RNA (rsRNA)
yRNA	yRNA-derived small RNA (ysRNA)
vtRNA	vtRNA-derived small RNA (vtsRNA)
snRNA	snRNA-derived small RNA (snsRNA)
snoRNA	snoRNA-derived small RNA (snosRNA)
Long noncoding RNA (lncRNA)	lncRNA-derived small RNA (lncsRNA)
mRNA	mRNA-derived small RNA (msRNA)

Box 2 | Main sources of sequencing bias in sncRNA discovery and ways to conquer

Among the many sources of sequencing biases[60], one major aspect comes from the adaptor-ligation process during cDNA-library construction (Fig. 1a,b). The ligation process is designed to (ideally) add adaptor sequences to the termini of all sncRNAs in the pool; however, in reality, different sncRNAs harbour distinct termini generated by different enzymes and thus cannot be uniformly ligated. For example, sncRNAs generated by Dicer (for example, siRNA and miRNA) and RNase P or RNase Z (for example, a portion of tsRNAs) bear 5′-phosphate (5′-P) and 3′-hydroxyl (3′-OH) termini[112], whereas sncRNAs generated by RNase T2 or RNase A (for example, many tsRNAs and rsRNAs) bear 5′-hydroxyl (5′-OH) and 2′,3′-cyclic phosphate (2′,3′-cP)[8] termini, and the 2′,3′-cP can be further hydrolysed to a 3′-phosphate (3′-P)[115]. In practice, the most widely used sncRNA-sequencing protocol is optimized for those bearing 5′-P and 3′-OH termini, and thus, the sncRNAs with 2′,3′-cP or 3′-P and/or 5′-OH termini cannot be ligated and will not be included in the cDNA library[10]. Solutions to this problem include the use of enzymes to convert the termini, such as the use of T4PNK to convert 2′,3′-cP and 3′-P into 3′-OH and 5′-OH into 5′-P before the ligation process[109], or combining with a template-switching strategy to add a 5′ adaptor to the cDNA after the RT instead of directly adding a 5′ adaptor to the RNA[110,116], which can resolve most problems caused by 5′-terminal modifications.

The second major source of bias comes from the RT process, which converts the adaptor-ligated RNA into cDNA (Fig. 1c). Several RNA modifications (for example, N^1-methyladenosine (m^1A), N^3-methylcytosine (m^3C), N^1-methylguanosine (m^1G) and N^2,N^2-dimethylguanosine (m^2_2G)) can interfere with the RT process, either by preventing the passage of reverse transcriptase or generating misincorporation at the modified loci[47,117,118]. Under the traditional protocol, if the RT process is interrupted before reaching the 5′ terminus, this truncated cDNA will not be further amplified from the 5′ end during the following PCR and therefore will not be detected. The solution could be to either use enzymes to remove these RT-blocking modifications (for example, AlkB)[47,117,118] or use a high-processive reverse transcriptase (for example, TGIRT and BoMoC) to read through the modifications without being blocked[111,119]. The latter approach retains the misincorporation, which can be used to infer the nature of the modification[86].

However, before a full exploration of the expanding functions of sncRNAs, perhaps an even more urgent and pertinent question is whether we have discovered all sncRNAs. If not, what have we missed and how should we systematically identify them?

Improved methods lead to an updated landscape of sncRNAs

The wide use of next-generation sequencing has greatly advanced the discovery of sncRNAs. However, in the early days most of the small RNA-seq protocols aimed to discover miRNAs and siRNAs of approximately 20 nt by implementing a pre-size selection of <30 nt RNA (recovery from polyacrylamide gels following electrophoresis) to generate a complementary-DNA (cDNA) library for high-throughput sequencing, which prevented the discovery of sncRNAs that were >30 nt in length. The RNA size-selection was later extended to approximately 45 nt (with the aim of discovering more sncRNAs), which can cover the length of piRNAs (21–35 nt) and also lead to the discovery of other non-canonical sncRNAs under physiological conditions—for example, in mature sperm cells[57] and serum[58,59] where clear peaks of tsRNAs and/or yRNA-derived small RNAs (ysRNAs) are found at 30–40 nt.

However, unexplained phenomena were constantly observed when the size-selection was extended to 45 nt. For example, although RNA bands or smears at 30–40 nt can be clearly observed on a polyacrylamide gel, the sequencing results only show a sharp peak of miRNAs (approximately 20 nt), whereas the sequencing reads from the 30–40 nt fraction are usually very low[10]. This inconsistency strongly suggests that the widely used sncRNA-sequencing protocols have generated biased results and fail to capture a large portion of sncRNAs clearly present on the polyacrylamide gel.

Such sequencing bias has been found to be derived from two main issues during the cDNA-library preparation (Box 2). One is the terminal modifications in sncRNAs that prevent adaptor ligation (Fig. 1a,b) and the other is the internal RNA modifications in sncRNAs that interfere with the reverse transcription (RT) process that converts the RNA into cDNA (Fig. 1c). New methods (for example, panoramic RNA display by overcoming RNA modification aborted sequencing (PANDORA-seq) and Cap-Clip acid pyrophosphatase, PNK and AlkB-facilitated sncRNA sequencing (CPA-seq)) have recently been developed to overcome both problems through the use of consecutive enzymatic treatment to resolve RNA termini that block adaptor ligation and remove RT-blocking RNA modifications[10,14], which enabled the identification of many sncRNAs that were previously undetectable and revealed an updated sncRNA landscape. For example, PANDORA-seq has shown that tsRNA and rsRNA are more abundant than miRNA in many tissues and cells

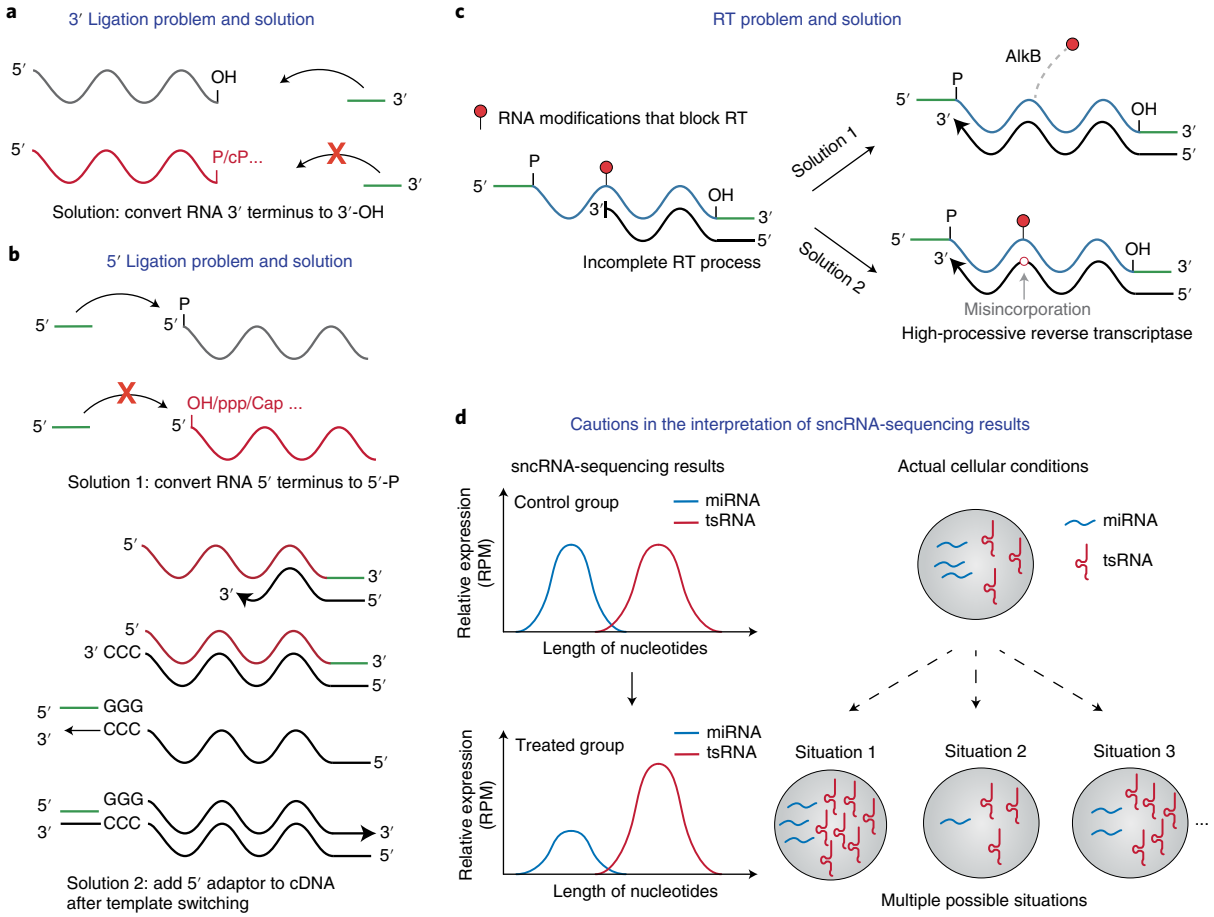

Fig. 1 | Methods to overcome biases in sncRNA discovery and cautions in the interpretation of sncRNA-sequencing results. a–c, Illustrations of the main sources of and solutions to sequencing bias in sncRNA discovery. **a**, Bias in 3′-adaptor (green line) ligation due to the existence of 3′-phosphate (3′-P),2′,3′-cyclic phosphate (2′,3′-cP) and so on. The solution involves using enzymes to convert the 3′ terminus into hydroxyl (3′-OH) before ligation. **b**, Bias in 5′-adaptor (green line) ligation due to the existence of 5′-OH, 5′-triphosphate group (5′-ppp) and 5′-m^{7}GpppN cap structure (5′-Cap) and so on. The solution involves either using enzymes to convert the 5′ terminus into a 5′-P before ligation or a template-switching strategy to add the adaptor to the intermediate cDNA rather than to the RNA. **c**, Bias in the RT process due to RNA modifications (for example, m^{1}G, m^{1}A, m^{3}C and m$^2_2$G). The solution involves the use of enzymes (for example, AlkB) to demethylate these RT-blocking modifications or high-processive reverse transcriptases (for example, TGIRT) to directly read through the modifications. Emerging methods such as PANDORA-seq[10] and CPA-seq[14] have started to resolve the above-mentioned three aspects of bias and have substantially improved panoramic sncRNA discovery. **d**, Schematic showing altered sncRNA profiles from sncRNA-sequencing results, which are based on the relative expression level (represented as RPM values) and could be derived from multiple intrinsic situations. Thus, the actual changes in the levels of sncRNA expression cannot be identified solely based on the sncRNA-sequencing results but will need additional analyses.

(for example, spleen, embryonic stem cells and HeLa cells), as validated by northern blot analyses[10]. However, it should be noted that even with the improved methods, we may still have not revealed the full landscape of sncRNAs (Box 3), as other terminal conditions and/or RNA modifications may exist to interfere with the ligation and RT process during cDNA-library construction[10,60], a possibility that awaits resolution.

Importantly, although different methods capture sncRNAs with specific properties regarding the termini and modification status (Table 1), a comparative analysis using different methods on one RNA sample can provide further information to deduce the compositional information of different types of sncRNAs[10]. In addition, pooled adaptors can be utilized to reduce ligase bias in terminal ligation[61]. Further improvements, including the addition of terminal barcode sequences to resolve the PCR amplification bias (caused by intrinsic differences in the amplification efficiency of cDNA templates)[62], can correct the number of reads with bioinformatic approaches, thus increasing the accuracy of sncRNA discovery. Moreover, the development of ultralow-input or single-cell-level analyses[63,64] based on improved bias-reducing protocols (for example, PANDORA-seq) is needed to reveal the dynamic landscape of scarce biological samples, such as mammalian early embryos.

Caveats to the analysis and interpretation of sncRNA-sequencing data

With the discovery and bioinformatic annotation of major subcategories of sncRNAs (for example, miRNA, tsRNA and rsRNA) in biological samples[65], new analytical difficulties have emerged, especially when trying to accurately measure the changes in sncRNA expression between two (or more) conditions, which concerns how to correctly interpret the sequencing results by considering the inherent nature and limitation of the RNA-seq data and the specific sample status. Here we dissect the main caveats in sncRNA data analyses and discuss potential solutions under different situations.

First and foremost, the reported expression level of a sequence from sncRNA-sequencing data (for example, presented as reads per million (RPM)) represents the relative enrichment of this sequence in the sample but not the absolute quantity. In this regard,

Box 3 | Blind men and the elephant

If the history of sncRNA research, or RNA research in general, has taught us anything, it would be that the old views and rules are constantly being overturned to forge new ones[120]. This may remind us of the old parable of 'The blind men with the elephant': we often have a tendency to be obsessed with the contemporary discoveries and try to use the existing knowledge to explain biological observations, yet every time new knowledge arrives, we realize that we have seen only part of the larger picture. It seems that the only question is when we might reach an end.

While in this Perspective we describe miRNA, siRNA and piRNA as canonical sncRNAs and describe other sncRNAs derived from longer RNA precursors as non-canonical, we may keep in mind that in principle, all RNA sequences (sometimes tuned by RNA modifications) harness base-pairing to bind to their DNA or RNA targets and their interactions with protein targets are based on their molecular structure. For example, earlier studies using cross-linking ligation and sequencing of hybrids (CLASH)—an experimental approach to identify RNA–RNA duplexes associated with Argonaute proteins in vivo—focused on revealing the RNA targets of miRNAs[121] and piRNAs[122]; however, more comprehensive analyses using these same datasets later revealed extensive tsRNA–mRNA[123,124] and rsRNA–mRNA[125] interactions, and even interactions between sncRNAs[125]. Further analyses extending to the potential interactions between other sncRNAs and long RNAs are highly possible and await discovery.

the changes in the RPM value of certain sncRNAs does not necessarily reflect the changes in their net expression levels because the changes in RPM could result from very different scenarios. For example, if a cell expresses both miRNA and tsRNA (in real cases there could be more types of sncRNAs; Fig. 1d), and the deletion of a gene enhances the biogenesis of the tsRNA but does not affect the overall level of miRNA, the sequencing result based on RPM would give the impression that the miRNAs are overall downregulated, a misinterpretation caused by the increased tsRNA reads that have consumed more of the relative RPM. The same RPM pattern change could result from other scenarios, such as that miRNAs are truly downregulated, while tsRNAs remain the same or both the miRNA and tsRNA levels are changed (Fig. 1d). Thus, simply using the RPM value to evaluate sncRNA expression changes is not sufficient and may cause systematic misinterpretation.

Northern blot analyses of multiple sncRNAs can be performed to normalize expression levels measured under different conditions by using equal quantities of total RNA input and loading controls[10] (rather than using certain 'housekeeping' RNAs as internal controls, as they may also change between the conditions). The results would provide the necessary additional information to evaluate the actual expression levels of selected sncRNAs (for example, miRNAs, tsRNAs and rsRNAs)[10] under different conditions and could be used as the 'anchor points' to correctly interpret the RPM value. Notably, cross hybridization on sncRNAs that share very similar sequences can occur in northern blotting; thus, sncRNAs cannot always be separated by northern blotting and instead combined signals of these similar sequences are obtained. Alternatively, spike-in RNA added during library construction can facilitate the quantification of sncRNAs in a sample[66] and can be used as internal controls to normalize the expression of sncRNAs between two samples.

However, it should be noted that adding spike-in RNA into RNA samples with the same quantity of total RNA will be problematic if different numbers of cells in the two groups contribute equal quantities of total RNA. For example, certain cancer cells generate 2–3 times more total RNA than normal cells[67]; if equal spike-in RNAs are added according to the total RNA levels, this will lead to underestimation of the sncRNA expression level in cancer cells. The solution to such situation could be to either perform northern blots with or add spike-in RNA into RNA samples extracted from an equal number of cells instead of based on equal RNA quantity. Ideally, future endeavours would aim to add spike-in RNA at the single-cell level and thus open the venue to absolute quantification of sncRNAs of individual cells when combined with improved protocols such as PANDORA-seq.

New era for direct and multiplexed mapping of all RNA modifications in sncRNAs

Beyond the primary RNA sequence, the complex modifications on sncRNAs were previously neglected but increasing evidence has now demonstrated that RNA modifications represent an additional layer of information that is integral to the function of sncRNAs by regulating RNA stability, structure, binding potential and extracellular molecular interactions[48,68–70]. This issue has become particularly important for the emerging non-canonical sncRNAs that are derived from highly modified precursors such as tRNAs, which harbour more than 150 types of modifications[71]. However, substantially more sncRNA modifications remain undetectable or underexplored because the current mainstream RNA-seq methods are in fact sequencing the cDNA intermediate of RNAs and the conversion of RNA to cDNA has resulted in the loss of most RNA-modification information. The existing tools for site-specific high-throughput mapping of RNA modifications are mainly for long RNAs and are limited to only a few well-known modifications (for example, 5-methylcytosine (m^5C), N^6-methyladenosine (m^6A), pseudouridine (ψ), inosine (I), m^1A and N^4-acetylcytidine (ac^4C)). Commonly used approaches include antibody-dependent methods, chemical conversion of the targeted modifications into a distinguishable base[72–80] and the newly developed nanopore-based direct RNA-seq[81–83]; however, these methods usually analyse only one modification type at a time. Other methods, such as inferring the nucleotide misincorporation during RT, can simultaneously deduce the distribution of multiple RNA modifications[84–86] but only in a qualitative, and not quantitative, manner and suffer from false-positive calling due to multiple factors, including the selection of the RT enzyme, the reaction conditions and the accuracy of the algorithm[87]. In short, there are no efficient methods available at present for high-throughput, comprehensive, quantitative mapping of multiple types of modifications in sncRNAs or RNAs in general.

Although different methods are continuously being developed or improved based on sequencing of cDNA intermediates to identify RNA modifications[88], it has become an imminent concern that the intrinsic nature of complex RNA modifications has made the cDNA-based approaches inefficient and inadequate to resolve the full scope of RNA modifications; thus, the field urgently needs transformative methods that can directly sequence RNA and simultaneously identify all modifications[89]. Two classes of methods are currently being explored for direct RNA-seq and quantitative multiplexed mapping of RNA modifications, based on either mass spectrometry (MS) or nanopore technology.

MS: old dog, new tricks. Liquid chromatography with tandem MS (LC-MS/MS) has been widely used to analyse RNA modifications and is considered the 'gold standard' to quantify modifications in an RNA sample because compared with other indirect methods—such as antibody-based and cDNA conversion-based modification detection—MS directly measures a specific RNA fragment (or a single nucleotide) based on its physical properties, such as retention time and molecular mass (similar to the use of MS to determine the peptide sequence)[90]. However, when RNAs are digested into smaller fragments or single nucleotides before MS examination, the

Table 1 | Recent methods to improve sncRNA sequencing (next-generation sequencing) by overcoming specific RNA modifications

Method	Resolving terminal modifications to improve ligation	Resolving internal modifications to improve RT	Other features and concerns
ARM-seq[47]	Unresolved	• AlkB treatment to remove RNA modifications that block RT	• Potential degradation of longer RNAs (for example, tRNA) during AlkB treatment would generate RNA fragments that will be sequenced as artifacts[10]
cP-RNA-seq[108]	• A series of treatments by alkaline phosphatase, calf intestinal (CIP), periodate and then T4PNK to selectively capture the RNAs with 2′,3′-cP at their 3′ termini	Unresolved	• Selectively sequence the 2′,3′-cP-containing sncRNAs • sncRNA containing both 2′,3′-cP and other RT-blocking modifications could be missed
Improved RNA-seq[109]	• T4PNK treatment converts 3′-P and 2′,3′-cP at 3′ termini into 3′-OH, and 5′-OH at 5′ termini into 5′-P	Unresolved	
5′ XP sRNA-seq[110]	• Simultaneously captures 5′-P and non-5′-P RNAs with the 5′-P RNA tagged with a sequence to be distinguished during bioinformatic analyses • 3′-P, 2′,3′-cP unresolved	Unresolved	• Enables comparative analysis of 5′-P and non-5′-P sncRNA • Analyses are limited to sncRNAs that have a 3′-OH and do not have RT-blocking modifications
TGIRT-seq[111]	• T4PNK treatment converts 3′-P and 2′,3′-cP at 3′ termini into 3′-OH • Template-switching activity by TGIRT adds an adaptor to the 3′ end of cDNA instead of the 5′ end of RNA, thus resolving 5′ RNA modifications	• Highly processive reverse transcriptase TGIRT to read through RNA modifications	• Simultaneous profiling of longer RNAs (for example, mRNA and lncRNA) • TGIRT cannot always read though RNA modifications and needs reaction-condition optimization[86]
AQRNA-seq[66]	• Alkaline phosphatase treatment to convert 3′-P into 3′-OH and 5′-P into 5′-OH • Add adaptor to the 3′ end of cDNA instead of the 5′ end of RNA • 2′,3′-cP unresolved	• AlkB treatment to remove RNA modifications that block RT	• Quantification of sncRNA by adding spike-in RNA to the sample
CPA-seq[14]	• Cap-Clip treatment removes the 5′ cap and 5′-triphosphate group to generate 5′-P • T4PNK treatment converts 3′-P and 2′,3′-cP at 3′ termini into 3′-OH and 5′-OH at 5′ termini into 5′-P	• AlkB treatment to remove RNA modifications that block RT • Highly processive reverse transcriptase TGIRT to read through RNA modifications	• Deacylation buffer (pH 9.0) to remove aminoacyl residues in 3′ tsRNAs • Potential degradation of longer RNAs (for example, tRNA) during AlkB treatment would generate RNA fragments that will be sequenced as artifacts[10]
PANDORA-seq[10]	• T4PNK treatment converts 3′-P and 2′,3′-cP at 3′ termini into 3′-OH and 5′-OH at 5′ termini into 5′-P	• AlkB treatment to remove RNA modifications that block RT	• Pre-size selection (<50 nt RNA) eliminates fragmentation of longer RNAs (for example, tRNA) by AlkB treatment • Data analysis by SPORTS[65] to improve non-canonical sncRNA identification and characterization

Different experimental strategies are used to resolve and reduce biases during cDNA-library construction of sncRNAs that are caused by adaptor-ligation bias and RT blocking, along with other improvements.

positional information is lost. Thus, obtaining the RNA-modification information within an RNA sequence context usually relies on the complementary methods, such as reference sequences provided by next-generation sequencing-based RNA-seq[91].

In theory, using MS to directly measure RNA sequences and modifications is possible and attractive[92]; if an RNA can be uniformly degraded into a mass ladder, the RNA sequence and the modification information can be directly 'read' according to the mass shift along the ladder, which is conceptually similar to the Sanger sequencing strategy in regard to the formation of a DNA ladder (Fig. 2a). However, a high-quality RNA mass ladder cannot be easily generated by random RNA degradation or specific enzymatic cleavage[93].

In 2015 a landmark paper from the Jack Szostak laboratory overcame this challenge by developing a generalized and efficient way to fragment RNA in a controllable manner, followed by two-dimensional mass-retention time analysis of the resulting RNA fragments by liquid-chromatography separation, which permits the generation of perfect RNA mass ladders for direct RNA-seq[93] (Fig. 2a). The key success of the method is the application of a time-controlled protocol for RNA degradation by formic acid, generating RNA fragments of different lengths to form perfect mass ladders in both the 3′ to 5′ and 5′ to 3′ directions, which enables de novo bidirectional sequencing of the RNA sample along with the site-specific RNA modifications.

This first success was followed by further methodological improvements, including optimization of the RNA degradation protocol to more evenly generate RNA fragments of different lengths and use of a hydrophobic end-labelling strategy to add different chemical labels at the 3′ and 5′ ends of the fragmented products, which enhanced the identification of the differentially labelled two-dimensional mass ladders and enabled the reading

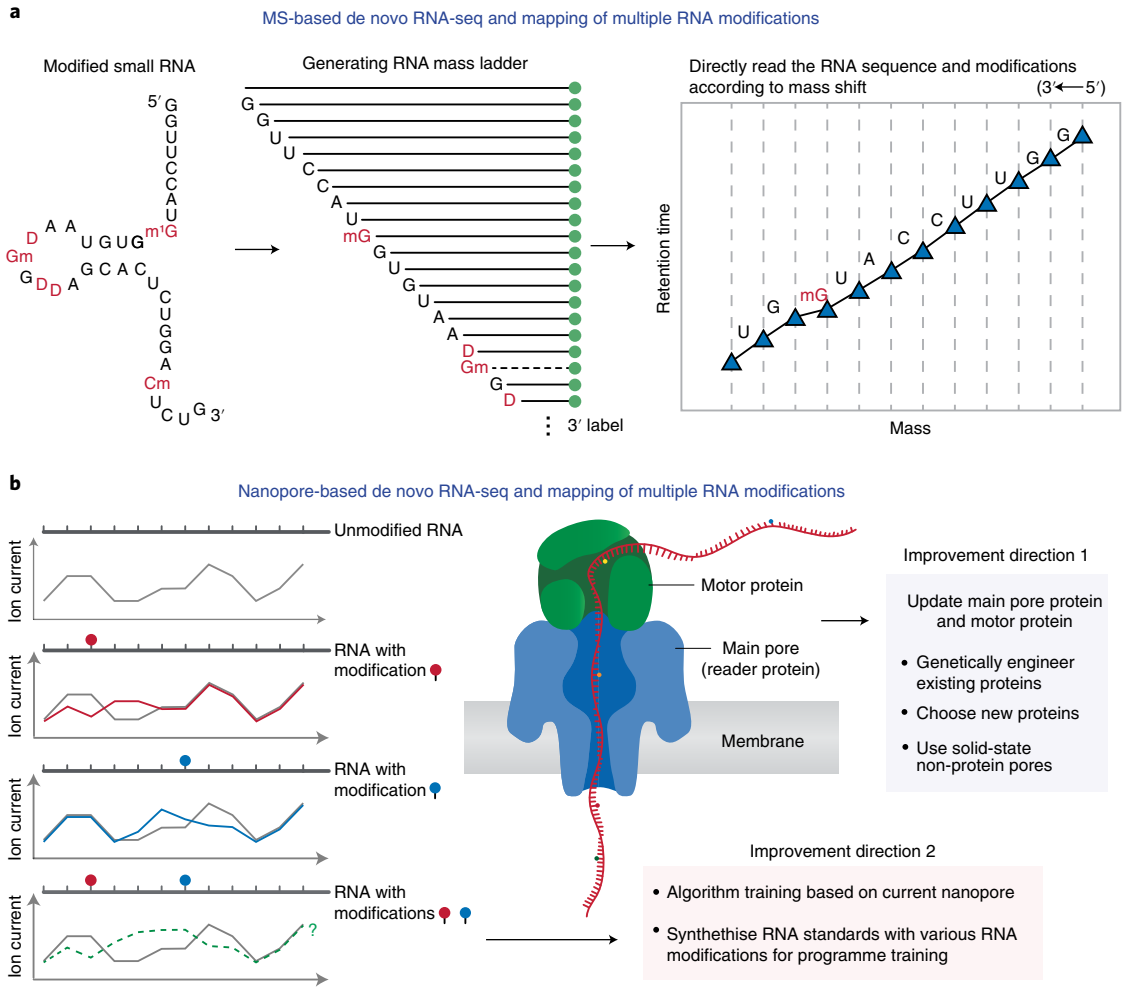

Fig. 2 | Two methods for future direct sequencing of RNA and multiplexed mapping of RNA modifications without cDNA intermediates. a, Main concept and workflow for MS-based de novo sequencing of modified sncRNA, which involves controlled fragmentation of RNA (by formic acid) into ladder fragments, followed by measurement of the resultant RNA fragments using LC-MS/MS, generating sequences of both canonical and modified nucleosides based on the mass signature. Note that additional methods are needed to distinguish modified nucleotides with the same mass shift. For example, the sensitivity to AlkB treatment can be used to distinguish between m^1A and m^6A, m^3C and m^5C, or m^1G and m^2G, where m^1A, m^1G and m^3C can be demethylated by AlkB[96]; nucleotides with 2-O′-methylation (Am, Um, Cm and Gm) can prevent the acid hydrolyzation and thus generate a mass gap in the mass ladder[93,94]; and chemical conversion of ψ to CMC-ψ (by reaction with *N*-cyclohexyl-*N*′-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulfonate (CMC)) to distinguish ψ from U[94]. **b**, Schematic showing that some RNA modifications will change not only the ion current of the modified nucleotide but also that of the adjacent unmodified nucleotides and the combinatorial effect of two modifications on the ion current of adjacent nucleotides remains largely unexplored. Two main directions for future improvements of nanopore-based direct sequencing are shown, and ideally will be applied together.

of the complete sequence of a given RNA from either the 3′ or 5′ end, rather than requiring paired-end sequences from both directions[94] (Fig. 2a). With the proper algorithm and automated analysis, the improved method has been used to de novo sequence a complete purified yeast $tRNA^{phe}$ with all eleven RNA modifications[95]. Through further improvements involving increased MS read length (approximately 80 nt) and advanced algorithms, MS ladder complementation sequencing (MLC-seq) was developed to assemble full MS ladders from partial ladders with missing ladder components, making it possible to de novo sequence RNAs with relatively low abundance[96]. In a recent application, MLC-seq analysis of $tRNA^{Glu}$ extracted from mouse liver accurately pinpointed the location of modifications in $tRNA^{Glu}$ and their stoichiometric changes following treatment with the dealkylating enzyme AlkB and uncovered new RNA modifications that had not been reported for $tRNA^{Glu}$ (ref. [96]). MLC-seq will be particularly useful for the study of highly modified RNAs such as tRNAs and tsRNAs, and to address open questions such as the tissue-specific differences in tRNAs and tsRNAs in regard to both sequence and modifications under normal and disease conditions.

These series of MS-based methodological developments have unleashed a path to simultaneously identify the sncRNA sequence and RNA modifications with single-nucleotide and stoichiometric precision, although they need further development to reach high-throughput. Future development of a comprehensive MS reference database of various types of tRNAs (or other sncRNAs), along with optimized bioinformatic tools, would enable a path to increase scalability and thus to sequence RNA mixtures with increased complexity.

Nanopore technology: a vigorous teenager to be trained. Nanopore technology is inspired by and derived from the elegant structures of natural membrane ion channels and was first utilized in 1996 to detect and identify single-stranded DNA and RNA

based on the alterations in ionic current as they pass through the channel pore[97]. With continuous improvements in recent decades, nanopore technology is now bringing a revolution in direct DNA and RNA sequencing due to its unique characteristics, including label-free, amplification-free and real-time detection of DNA or RNA at the single-molecule level with long-read capacity[98], which also holds great promise to directly determine the identity of the associated RNA modifications if they generate distinguishable ionic currents.

Nanopore-based direct sequencing has recently enabled the direct mapping of several RNA modifications, including m^6A, ψ and 2′-O-methylation[81–83], achieved by machine learning-based 'base-calling' algorithms for each specific modification. However, the simultaneous detection of multiple RNA modifications on a single RNA strand remains extremely difficult, especially for highly modified RNAs such as tRNA. A recent attempt using Oxford Nanopore MinION to comparatively sequence purified biological tRNA (from *Escherichia coli*) versus corresponding synthetic non-modified tRNA has revealed systematic miscalls at or adjacent to the positions of known modified nucleotide positions when sequencing biological tRNA samples[99]. These miscalls could not be correctly assigned to specific modifications by current algorithms. In addition, the reading accuracy of synthetic non-modified tRNA is lower than that of mRNA[99], suggesting that the current method is not well-adapted for short RNAs (for example, tRNA and sncRNA) and awaits improvement, such as ligating the tRNA or sncRNA to longer adaptor RNAs with optimized sequences.

One major difficulty in accurately mapping RNA modifications using nanopores is that the presence of a modification at a specific location will change not only the ion current of the modified nucleotide but also that of the unmodified nucleotides nearby[100,101] (due to the chemical and physical nature of the nanopore protein; Fig. 2b). This has created substantial difficulties in the training of algorithms, especially for highly modified sequences such as tRNA and tsRNA, where the effects of different RNA modifications may overlap and generate complicated situations. In theory this problem might be conquered by synthesising thousands of different standard RNA sequences with single and/or multiple modifications (either the same or mixed types) inserted at different positions, followed by intensive deep-learning algorithm training (Fig. 2b). However, this direction faces another practical difficulty as many standard RNA modifications cannot be readily synthesized at present. This problem may require intensive technical investments as it represents a major hurdle for future experimental design and algorithm development.

Another direction for improving the capacity and accuracy of nanopore-based RNA-modification detection is to genetically redesign or engineer (for example, site-specific mutation) the main pore or the motor protein of the existing nanopores, or both, or to choose completely different pores (for example, new membrane proteins or solid-state non-protein pores made of novel nanomaterials) and/or motor proteins that may recognize and distinguish RNA modifications with better resolution (Fig. 2b). Notably, the previous lack of protein-pore candidates is due largely to the lack of knowledge on the crystal structures of many membrane proteins, but now with the aid of Alphafold, which provides open access to protein-structure predictions of thousands of membrane proteins[102], the candidate pool is increasing substantially, which may lead to the selection of more specific pores that would be optimal for the sensitive detection of both RNAs and RNA modifications.

Finally, PacBio's Single-molecule real-time (SMRT) RT of RNA also has the potential to directly detect multiple RNA modifications from the RNA template through the analysis of the kinetics of the reverse transcriptase using zero-mode waveguides[103], which represents another direction for future exploration.

Conclusion and perspectives

The systematic capture of all sncRNA sequences with all modifications is a grand dream but even its accomplishment would represent only a first step. Another major challenge concerns the subcellular spatial compartmentalization of sncRNAs. Great advances in the spatial mapping of the transcriptome at the single-cell level based on in situ hybridization, either through multiplexed imaging[104] or by sequencing[105] approaches, have been witnessed in the past few years. However, these methods are mostly optimized for long RNAs such as mRNA, whereas the short length of sncRNAs limits nucleic-acid-probe design options and the probe may bind to multiple targets (for example, both sncRNAs and their precursors); thus, the locations of sncRNAs would be difficult to determine with accuracy. In addition, many RNA modifications and RNA structures in sncRNAs can prevent efficient hybridization in situ. These are among the practical issues that must be resolved before the systematic spatial mapping of sncRNAs at subcellular resolution.

A deeper and long-standing question posed regarding the expanding universe of sncRNAs is about their function and the versatile ways to achieve it, especially when they are spatially condensed and compartmentalized in the cell. We have chosen to use the word 'RNA code' to describe the complex information represented by the whole repertoire of sncRNAs[106], which includes, but is not limited to, their linear sequence and site-specific RNA modifications; their interaction potential with target RNA, DNA and RNA-binding proteins as well as the social behaviour of sncRNAs in (and between) cells, such as the competition of and synergistic effects on mutual targets. How to systematically decode this information of astronomical complexity remains extremely challenging, even with decades of experimental and computational approaches, especially when considering the physiological relevance under normal and disease conditions. However, paradigm-changing tools are constantly emerging such as the recent use of deep-learning programmes to systematically predict the three-dimensional structures of RNA[107] and protein[102], which should also make the systematic prediction of RNA–protein interactions only a matter of time. These fast-evolving tools would bring new excitement to cracking the RNA code enabled by the complexity of the sncRNA universe, which represents an endless frontier worthy of deep exploration by new generations of human (and machine) intelligence.

Received: 20 October 2021; Accepted: 2 March 2022;
Published online: 12 April 2022

References

1. Grosshans, H. & Filipowicz, W. Molecular biology: the expanding world of small RNAs. *Nature* **451**, 414–416 (2008).
2. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* **43**, 880–891 (2011).
3. Babski, J. et al. Small regulatory RNAs in Archaea. *RNA Biol.* **11**, 484–493 (2014).
4. Carthew, R. W. & Sontheimer, E. J. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).
5. Bartel, D. P. Metazoan microRNAs. *Cell* **173**, 20–51 (2018).
6. Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D. & Zamore, P. D. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20**, 89–108 (2019).
7. Seal, R. L. et al. A guide to naming human non-coding RNA genes. *EMBO J.* **39**, e103777 (2020).
8. Chen, Q., Zhang, X., Shi, J., Yan, M. & Zhou, T. Origins and evolving functionalities of tRNA-derived small RNAs. *Trends Biochem. Sci.* **46**, 790–804 (2021).
9. Schimmel, P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat. Rev. Mol. Cell Biol.* **19**, 45–58 (2018).
10. Shi, J. et al. PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications. *Nat. Cell Biol.* **23**, 424–436 (2021).
11. Gu, W. et al. Peripheral blood non-canonical small non-coding RNAs as novel biomarkers in lung cancer. *Mol. Cancer* **19**, 159 (2020).

12. Cambier, L. et al. Y RNA fragment in extracellular vesicles confers cardioprotection via modulation of IL-10 expression and secretion. *EMBO Mol. Med.* **9**, 337–352 (2017).
13. Chen, C. J. & Heard, E. Small RNAs derived from structural non-coding RNAs. *Methods* **63**, 76–84 (2013).
14. Wang, H. et al. CPA-seq reveals small ncRNAs with methylated nucleosides and diverse termini. *Cell Discov.* **7**, 25 (2021).
15. Taft, R. J. et al. Small RNAs derived from snoRNAs. *RNA* **15**, 1233–1240 (2009).
16. Ender, C. et al. A human snoRNA with microRNA-like functions. *Mol. Cell* **32**, 519–528 (2008).
17. Persson, H. et al. The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nat. Cell Biol.* **11**, 1268–1271 (2009).
18. Hussain, S. et al. NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep.* **4**, 255–261 (2013).
19. Pircher, A., Bakowska-Zywicka, K., Schneider, L., Zywicki, M. & Polacek, N. An mRNA-derived noncoding RNA targets and regulates the ribosome. *Mol. Cell* **54**, 147–155 (2014).
20. Reuther, J. et al. A small ribosome-associated ncRNA globally inhibits translation by restricting ribosome dynamics. *RNA Biol.* **18**, 2617–2632 (2021).
21. Tuck, A. C. & Tollervey, D. RNA in pieces. *Trends Genet.* **27**, 422–432 (2011).
22. Schaefer, M. et al. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* **24**, 1590–1595 (2010).
23. Tuorto, F. et al. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* **19**, 900–905 (2012).
24. Chen, Q. et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351**, 397–400 (2016).
25. Zhang, Y. et al. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nat. Cell Biol.* **20**, 535–540 (2018).
26. Guzzi, N. et al. Pseudouridylation of tRNA-derived fragments steers translational control in stem cells. *Cell* **173**, 1204–1216 (2018).
27. Natt, D. et al. Human sperm displays rapid responses to diet. *PLoS Biol.* **17**, e3000559 (2019).
28. Goodarzi, H. et al. Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. *Cell* **161**, 790–802 (2015).
29. Kim, H. K. et al. A transfer-RNA-derived small RNA regulates ribosome biogenesis. *Nature* **552**, 57–62 (2017).
30. Balatti, V. et al. tsRNA signatures in cancer. *Proc. Natl Acad. Sci. USA* **114**, 8071–8076 (2017).
31. Yue, T. et al. SLFN2 protection of tRNAs from stress-induced cleavage is essential for T cell-mediated immunity. *Science* **372**, eaba4220 (2021).
32. Wang, Q. et al. Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. *Mol. Ther.* **21**, 368–379 (2013).
33. Liu, Y. M. et al. Exosome-delivered and Y RNA-derived small RNA suppresses influenza virus replication. *J. Biomed. Sci.* **26**, 58 (2019).
34. Hogg, M. C. et al. Elevation in plasma tRNA fragments precede seizures in human epilepsy. *J. Clin. Invest.* **129**, 2946–2951 (2019).
35. Zhang, X. et al. Small RNA modifications in Alzheimer's disease. *Neurobiol. Dis.* **145**, 105058 (2020).
36. Blanco, S. et al. Stem cell function and stress response are controlled by protein synthesis. *Nature* **534**, 335–340 (2016).
37. Sajini, A. A. et al. Loss of 5-methylcytosine alters the biogenesis of vault-derived small RNAs to coordinate epidermal differentiation. *Nat. Commun.* **10**, 2550 (2019).
38. Krishna, S. et al. Dynamic expression of tRNA-derived small RNAs define cellular states. *EMBO Rep.* **20**, e47789 (2019).
39. Kfoury, Y. S. et al. tiRNA signaling via stress-regulated vesicle transfer in the hematopoietic niche. *Cell Stem Cell* **28**, 2090–2103 (2021).
40. Schorn, A. J., Gutbrod, M. J., LeBlanc, C. & Martienssen, R. LTR-retrotransposon control by tRNA-derived small RNAs. *Cell* **170**, 61–71 (2017).
41. Martinez, G., Choudury, S. G. & Slotkin, R. K. tRNA-derived small RNAs target transposable element transcripts. *Nucleic Acids Res.* **45**, 5142–5152 (2017).
42. Sarker, G. et al. Maternal overnutrition programs hedonic and metabolic phenotypes across generations through sperm tsRNAs. *Proc. Natl Acad. Sci. USA* **116**, 10547–10556 (2019).
43. Sharma, U. et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391–396 (2016).
44. Wahba, L., Hansen, L. & Fire, A. Z. An essential role for the piRNA pathway in regulating the ribosomal RNA pool in *C. elegans*. *Dev. Cell* **56**, 2295–2312 (2021).
45. Zhang, Y. et al. Angiogenin mediates paternal inflammation-induced metabolic disorders in offspring through sperm tsRNAs. *Nat. Commun.* **12**, 6673 (2021).
46. Honda, S. et al. Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers. *Proc. Natl Acad. Sci. USA* **112**, E3816-25 (2015).
47. Cozen, A. E. et al. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat. Methods* **12**, 879–884 (2015).
48. Zhang, X., Cozen, A. E., Liu, Y., Chen, Q. & Lowe, T. M. Small RNA modifications: integral to function and disease. *Trends Mol. Med.* **22**, 1025–1034 (2016).
49. Huang, X., Fejes Toth, K. & Aravin, A. A. piRNA biogenesis in *Drosophila melanogaster*. *Trends Genet.* **33**, 882–894 (2017).
50. Shabalina, S. A. & Koonin, E. V. Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol.* **23**, 578–587 (2008).
51. Raad, N., Luidalepp, H., Fasnacht, M. & Polacek, N. Transcriptome-wide analysis of stationary phase small ncRNAs in *E. coli*. *Int. J. Mol. Sci.* **22**, 1703 (2021).
52. Lee, S. R. & Collins, K. Starvation-induced cleavage of the tRNA anticodon loop in *Tetrahymena thermophila*. *J. Biol. Chem.* **280**, 42744–42749 (2005).
53. Thompson, D. M., Lu, C., Green, P. J. & Parker, R. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **14**, 2095–2103 (2008).
54. Gebetsberger, J., Zywicki, M., Kunzi, A. & Polacek, N. tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in *Haloferax volcanii*. *Archaea* **2012**, 260909 (2012).
55. Garcia-Silva, M. R. et al. Extracellular vesicles shed by *Trypanosoma cruzi* are linked to small RNA pathways, life cycle regulation, and susceptibility to infection of mammalian cells. *Parasitol. Res.* **113**, 285–304 (2014).
56. Fricker, R. et al. A tRNA half modulates translation as stress response in *Trypanosoma brucei*. *Nat. Commun.* **10**, 118 (2019).
57. Peng, H. et al. A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Res.* **22**, 1609–1612 (2012).
58. Dhahbi, J. M. et al. 5′ tRNA halves are present as abundant complexes in serum, concentrated in blood cells, and modulated by aging and calorie restriction. *BMC Genomics* **14**, 298 (2013).
59. Zhang, Y. et al. Identification and characterization of an ancient class of small RNAs enriched in serum associating with active infection. *J. Mol. Cell Biol.* **6**, 172–174 (2014).
60. Raabe, C. A., Tang, T. H., Brosius, J. & Rozhdestvensky, T. S. Biases in small RNA deep sequencing data. *Nucleic Acids Res.* **42**, 1414–1426 (2014).
61. Jayaprakash, A. D., Jabado, O., Brown, B. D. & Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* **39**, e141 (2011).
62. Saunders, K. et al. Insufficiently complex unique-molecular identifiers (UMIs) distort small RNA sequencing. *Sci. Rep.* **10**, 14593 (2020).
63. Faridani, O. R. et al. Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.* **34**, 1264–1266 (2016).
64. Yang, Q. et al. Single-cell CAS-seq reveals a class of short PIWI-interacting RNAs in human oocytes. *Nat. Commun.* **10**, 3389 (2019).
65. Shi, J., Ko, E. A., Sanders, K. M., Chen, Q. & Zhou, T. SPORTS1.0: a tool for annotating and profiling non-coding RNAs optimized for rRNA- and tRNA-derived small RNAs. *Genomics Proteomics Bioinformatics* **16**, 144–151 (2018).
66. Hu, J. F. et al. Quantitative mapping of the cellular small RNA landscape with AQRNA-seq. *Nat. Biotechnol.* **39**, 978–988 (2021).
67. Loven, J. et al. Revisiting global gene expression analysis. *Cell* **151**, 476–482 (2012).
68. Ji, L. & Chen, X. Regulation of small RNA stability: methylation and beyond. *Cell Res.* **22**, 624–636 (2012).
69. Frye, M., Harada, B. T., Behm, M. & He, C. RNA modifications modulate gene expression during development. *Science* **361**, 1346–1349 (2018).
70. Flynn, R. A. et al. Small RNAs are modified with *N*-glycans and displayed on the surface of living cells. *Cell* **184**, 3109–3124 (2021).
71. Suzuki, T. The expanding world of tRNA modifications and their disease relevance. *Nat. Rev. Mol. Cell Biol.* **22**, 375–392 (2021).
72. Schaefer, M., Pollex, T., Hanna, K. & Lyko, F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.* **37**, e12 (2009).
73. Sakurai, M. & Suzuki, T. Biochemical identification of A-to-I RNA editing sites by the inosine chemical erasing (ICE) method. *Methods Mol. Biol.* **718**, 89–99 (2011).
74. Schwartz, S. et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).
75. Carlile, T. M. et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143–146 (2014).

76. Hussain, S., Aleksic, J., Blanco, S., Dietmann, S. & Frye, M. Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol.* **14**, 215 (2013).
77. Dominissini, D. et al. Topology of the human and mouse m^{6}A RNA methylomes revealed by m^{6}A-seq. *Nature* **485**, 201–206 (2012).
78. Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
79. Sas-Chen, A. et al. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature* **583**, 638–643 (2020).
80. Li, X. et al. Base-resolution mapping reveals distinct m^{1}A methylome in nuclear- and mitochondrial-encoded transcripts. *Mol. Cell* **68**, 993–1005 (2017).
81. Begik, O. et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat. Biotechnol.* **39**, 1278–1291 (2021).
82. Liu, H. et al. Accurate detection of m^{6}A RNA modifications in native RNA sequences. *Nat. Commun.* **10**, 4079 (2019).
83. Parker, M. T. et al. Nanopore direct RNA sequencing maps the complexity of *Arabidopsis* mRNA processing and m^{6}A modification. *eLife* **9**, e49658 (2020).
84. Werner, S. et al. 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 (2020).
85. Khoddami, V. et al. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc. Natl Acad. Sci. USA* **116**, 6784–6789 (2019).
86. Behrens, A., Rodschinka, G. & Nedialkova, D. D. High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Mol. Cell* **81**, 1802–1815 (2021).
87. Sas-Chen, A. & Schwartz, S. Misincorporation signatures for detecting modifications in mRNA: not as simple as it sounds. *Methods* **156**, 53–59 (2019).
88. Owens, M. C., Zhang, C. & Liu, K. F. Recent technical advances in the study of nucleic acid modifications. *Mol. Cell* **81**, 4116–4136 (2021).
89. Alfonzo, J. D. et al. A call for direct sequencing of full-length RNAs to identify all modifications. *Nat. Genet.* **53**, 1113–1116 (2021).
90. Ross, R.L., Cao, X. & Limbach, P.A. Mapping post-transcriptional modifications onto transfer ribonucleic acid sequences by liquid chromatography tandem mass spectrometry. *Biomolecules* **7**, 21 (2017).
91. Kimura, S., Dedon, P. C. & Waldor, M. K. Comparative tRNA sequencing and RNA mass spectrometry for surveying tRNA modifications. *Nat. Chem. Biol.* **16**, 964–972 (2020).
92. Sample, P. J., Gaston, K. W., Alfonzo, J. D. & Limbach, P. A. RoboOligo: software for mass spectrometry data to support manual and de novo sequencing of post-transcriptionally modified ribonucleic acids. *Nucleic Acids Res.* **43**, e64 (2015).
93. Bjorkbom, A. et al. Bidirectional direct sequencing of noncanonical RNA by two-dimensional analysis of mass chromatograms. *J. Am. Chem. Soc.* **137**, 14430–14438 (2015).
94. Zhang, N. et al. A general LC-MS-based RNA sequencing method for direct analysis of multiple-base modifications in RNA mixtures. *Nucleic Acids Res.* **47**, e125 (2019).
95. Zhang, N. et al. Direct sequencing of tRNA by 2D-HELS-AA MS Seq reveals its different isoforms and dynamic base modifications. *ACS Chem. Biol.* **15**, 1464–1472 (2020).
96. Zhang, S. et al. MLC-Seq: de novo sequencing of full-length tRNAs and quantitative mapping of multiple RNA modifications. Preprint at *Researchsquare* https://doi.org/10.21203/rs.3.rs-1090754/v1 (2021).
97. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl Acad. Sci. USA* **93**, 13770–13773 (1996).
98. Wang, S., Zhao, Z., Haque, F. & Guo, P. Engineering of protein nanopores for sequencing, chemical or protein sensing and disease diagnosis. *Curr. Opin. Biotechnol.* **51**, 80–89 (2018).
99. Thomas, N.K. et al. Direct nanopore sequencing of individual full length tRNA strands. *ACS Nano* **15**, 16642–16653 (2021).
100. Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
101. Smith, A. M., Jain, M., Mulroney, L., Garalde, D. R. & Akeson, M. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS ONE* **14**, e0216709 (2019).
102. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
103. Vilfan, I. D. et al. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J. Nanobiotechnol.* **11**, 8 (2013).
104. Zhuang, X. Spatially resolved single-cell genomics and transcriptomics by imaging. *Nat. Methods* **18**, 18–22 (2021).
105. Larsson, L., Frisen, J. & Lundeberg, J. Spatially resolved transcriptomics adds a new dimension to genomics. *Nat. Methods* **18**, 15–18 (2021).
106. Zhang, Y., Shi, J., Rassoulzadegan, M., Tuorto, F. & Chen, Q. Sperm RNA code programmes the metabolic health of offspring. *Nat. Rev. Endocrinol.* **15**, 489–498 (2019).
107. Townshend, R. J. L. et al. Geometric deep learning of RNA structure. *Science* **373**, 1047–1051 (2021).
108. Honda, S., Morichika, K. & Kirino, Y. Selective amplification and sequencing of cyclic phosphate-containing RNAs by the cP-RNA-seq method. *Nat. Protoc.* **11**, 476–489 (2016).
109. Akat, K. M. et al. Detection of circulating extracellular mRNAs by modified small-RNA-sequencing analysis. *JCI Insight* **5**, e127317 (2019).
110. Kugelberg, U., Natt, D., Skog, S., Kutter, C. & Ost, A. 5′ XP sRNA-seq: efficient identification of transcripts with and without 5′ phosphorylation reveals evolutionary conserved small RNA. *RNA Biol.* **18**, 1588–1599 (2021).
111. Xu, H., Yao, J., Wu, D. C. & Lambowitz, A. M. Improved TGIRT-seq methods for comprehensive transcriptome profiling with decreased adapter dimer formation and bias correction. *Sci. Rep.* **9**, 7953 (2019).
112. Haussecker, D. et al. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* **16**, 673–695 (2010).
113. Yamasaki, S., Ivanov, P., Hu, G. F. & Anderson, P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J. Cell Biol.* **185**, 35–42 (2009).
114. Lee, Y. S., Shibata, Y., Malhotra, A. & Dutta, A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* **23**, 2639–2649 (2009).
115. Shigematsu, M., Kawamura, T. & Kirino, Y. Generation of 2′,3′-cyclic phosphate-containing RNAs as a hidden layer of the transcriptome. *Front. Genet.* **9**, 562 (2018).
116. Dai, H. & Gu, W. Strategies and best practice in cloning small RNAs. *Gene Technol.* **9**, 151 (2020).
117. Zheng, G. et al. Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* **12**, 835–837 (2015).
118. Dai, Q., Zheng, G., Schwartz, M. H., Clark, W. C. & Pan, T. Selective enzymatic demethylation of N^2,N^2-dimethylguanosine in RNA and its application in high-throughput tRNA sequencing. *Angew. Chem. Int. Ed.* **56**, 5017–5020 (2017).
119. Upton, H. E. et al. Low-bias ncRNA libraries using ordered two-template relay: serial template jumping by a modified retroelement reverse transcriptase. *Proc. Natl Acad. Sci. USA* **118**, e2107900118 (2021).
120. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
121. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665 (2013).
122. Shen, E. Z. et al. Identification of piRNA binding sites reveals the Argonaute regulatory landscape of the *C. elegans* germline. *Cell* **172**, 937–951 (2018).
123. Kumar, P., Anaya, J., Mudunuri, S. B. & Dutta, A. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol.* **12**, 78 (2014).
124. Guan, L., Karaiskos, S. & Grigoriev, A. Inferring targeting modes of Argonaute-loaded tRNA fragments. *RNA Biol.* **17**, 1070–1080 (2020).
125. Guan, L. & Grigoriev, A. Computational meta-analysis of ribosomal RNA fragments: potential targets and interaction mechanisms. *Nucleic Acids Res.* **49**, 4085–4103 (2021).

Acknowledgements

We thank P. Schimmel (The Scripps Research Institute), X. Chen (UC Riverside) and our laboratory members for critical discussions on the contents of the manuscript. Research in the Q.C. laboratory is in part supported by the National Institutes of Health (NIH grant nos R01HD092431, R01ES032024 and P50HD098593). The T.Z. laboratory is in part supported by the NIH (grant no. R01ES032024).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Tong Zhou or Qi Chen.

Peer review information *Nature Cell Biology* thanks Ravi Sachidanandam and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.