HHS Public Access

Author manuscript

Nat Cell Biol. Author manuscript; available in PMC 2022 October 12.

Published in final edited form as:

Nat Cell Biol. 2022 April ; 24(4): 415–423. doi:10.1038/s41556-022-00880-5.

Exploring the expanding universe of small RNAs

Junchao Shi[1], **Tong Zhou**[2,#], **Qi Chen**[1,#]

[1]Division of Biomedical Sciences, School of Medicine, University of California, Riverside, Riverside, CA, USA

[2]Department of Physiology and Cell Biology, University of Nevada, Reno School of Medicine, Reno, NV, USA

Abstract

The world of small non-coding RNAs (sncRNAs) is ever-expanding, from siRNAs, miRNAs, piRNAs to the recently emerging noncanonical sncRNAs derived from longer structured RNAs (e.g., tRNAs, rRNAs, YRNAs, snoRNAs, snRNAs 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 non-coding RNAs (sncRNAs) are universally distributed in all kingdoms of life: from bacteria, archaea to various eukaryotic lives[1-3], which have not ceased to surprise us throughout the last two decades regarding their compositional and functional diversity. While 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 noncanonical sncRNAs (15-50 nt) that are derived from longer structured RNAs[7] such as transfer RNAs (tRNAs)[8, 9], ribosomal RNAs (rRNAs)[10, 11], Y RNAs (yRNAs)[11, 12], small nuclear RNAs (snRNAs)[13, 14], small nucleolar RNAs (snoRNAs)[15, 16], vault RNAs (vtRNAs)[17, 18], and even mRNAs[19, 20]. Studies on noncanonical sncRNAs have recently gained momentum, exemplified by the new focus on tRNA-derived small RNAs (tsRNAs)[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 noncanonical sncRNAs (Box 1) when describing discoveries from different laboratories (usually using different names).

Like many non-coding RNAs in history, the emerging noncanonical 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 novel regulatory sncRNAs[8, 21]. This is partly due to the revelation that they are regulated by

[#]Corresponding author: T.Z. tongz@med.unr.edu or Q.C. qi.chen@medsch.ucr.edu.

Competing interests:
The authors have no competing interests

both genetic and environmental factors[18, 22-27], and that many of them are functional and related to multiple diseases, including 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], and because in many cases, the exertion of their function depends on mechanisms that are distinct from those of well-studied siRNAs/miRNAs/piRNAs. Moreover, it was recently recognized that many non-canonical sncRNAs harbor various RNA modifications, some of which can prevent the detection of sncRNAs by traditional 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 new interest in research centered on sncRNA modifications[48]. Here we briefly outline the biogenesis and functional principles of noncanonical sncRNAs, and discuss recent methodological developments in promoting sncRNA discovery and accurate expression analyses, and new techniques for direct multiplexed mapping of RNA modifications, which is much needed for decoding sncRNAs' full function.

Distinct features of sncRNAs

In eukaryotes, the biogenesis and functions of siRNAs, miRNAs and piRNAs have been extensively studied[5, 6]. Both siRNAs and miRNAs are generated from double-stranded RNA (dsRNA) precursors mainly by RNase III enzymes (e.g., Dicer for siRNAs, Drosha and Dicer for miRNAs)[4], while piRNAs, found mainly in animal germline cells, are 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 (e.g., posttranscriptional mRNA cleavage/decay/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 to siRNAs, miRNAs and piRNAs, the noncanonical 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, tsRNAs and rsRNAs are predominantly found and dynamically regulated in ancient unicellular organisms (e.g., Bacteria, Archaea, Yeast and Protozoa) where siRNAs, miRNAs and piRNAs are absent[51-56]. This suggests that producing sncRNAs via the fragmentation/cleavage of longer structured RNAs (e.g., tRNA, rRNA, snRNA, Y RNAs, and vault RNAs) may represent the most ancient pathway of sncRNA biogenesis that predate the emergence of siRNAs, miRNAs and piRNAs[8]. In addition, the biogenesis of noncanonical sncRNAs involves the cleavage of their precursors (e.g., tRNAs, rRNAs) by a range ancient ribonuclease (RNase) families (e.g., RNase P, RNase Z, RNase T2, RNase A)[8] that predate the emergence of Dicer (which exists only in eukaryotes[50], responsible for generating siRNA and miRNAs), and are profoundly affected by site-specific RNA modifications and related enzymes[8]. Finally, many noncanonical 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.

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 (NGS) 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 ~20 nt by implementing a pre-size selection of <30 nt RNA (recovery from PAGE gel) to generate a complementary DNA (cDNA) library for high-throughput sequencing, which prevented the discovery of sncRNAs >30 nt. Later, the RNA size-selection was extended to ~45 nt, aiming to discover more sncRNAs, which can cover the length of piRNAs (21-35 nt) and also lead to the discovery of other noncanonical sncRNAs under physiological conditions, for example, in mature sperm cells[57] and serum[58, 59] where clear peaks of tsRNAs and/or ysRNAs are found at 30-40 nt.

However, unexplained phenomena were constantly observed when size-selection is extended to ~45 nt. For example, although RNA bands or smears at 30-40 nt can be clearly observed on PAGE gel, the sequencing results only show a sharp peak of miRNAs (~20 nt), while the sequencing reads from the 30-40 nt 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 PAGE 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 adapter ligation (Fig.1a,b), and the other is the internal RNA modifications in sncRNAs that interfere with reverse transcription (RT) process that converts the RNA into cDNA (Fig.1c). Recently, new methods (e.g., PANDORA-seq (panoramic RNA display by overcoming RNA modification aborted sequencing) and CPA-seq (Cap-Clip acid pyrophosphatase, PNK, and AlkB-facilitated sncRNA sequencing)) have been developed to overcome both problems by using consecutive enzymatic treatment to resolve RNA termini that block adapter ligation and to 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 tsRNAs and rsRNAs are more abundant than miRNAs in many tissues and cells (e.g, spleen, embryonic stem cells, 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 ligation and RT process during cDNA library construction[10, 60], a possibility that awaits resolution.

Importantly, while 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 adapters can be utilized to reduce ligase bias in terminal ligation[61]. Further improvements, including adding terminal barcode sequences to resolve the PCR amplification bias (caused by intrinsic differences in amplification efficiency of cDNA templates)[62], can correct the number of reads with bioinformatic approaches, thus increasing the accuracy of sncRNA discovery. Additionally, the development of ultralow-input or single-cell level analyses[63, 64] based on improved bias-reducing protocols (e.g. PANDORA-seq) is needed to reveal the dynamic landscape of scarce biological samples, such as mammalian early embryos.

Caveats in analysing and interpreting of sncRNA sequencing data

With the discovery and bioinformatic annotation of major subcategories of sncRNAs (e.g., miRNAs, tsRNA, rsRNAs) in biological samples[65], new analytical difficulties have emerged, especially when trying to accurately measure the sncRNA expression changes between two (more 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 a sncRNA sequencing data (e.g., presented as reads per million (RPM)) represents the relative enrichment of this sequence in the sample, but not the absolute quantity. In this regard, the changes in the RPM value of certain sncRNAs does not necessarily reflect the changes in their net expression level, because the changes in RPM could result from very different scenarios. For example, if a cell expresses both miRNAs and tsRNAs (in real cases there could be more types of sncRNAs) (Fig.1d) and the deletion of a gene enhances the biogenesis of tsRNAs but does not affect the overall level of miRNAs, 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 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 the expression levels between different conditions, by using the same total RNA input as a loading control[10] (rather than using certain ‘housekeeping’ RNAs as internal control, as they may also change between the conditions). The results would provide the necessary additional information to evaluate the actual expression level of selected sncRNAs (e.g., miRNAs, tsRNAs and rsRNAs)[10] under different conditions and could be used as the ‘anchor points’ to correctly interpret the RPM value. Notably, Northern blot can have cross hybridization on sncRNAs that share very similar sequences, and thus cannot always separate them but provide combined signals of these similar sequences. 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 total RNA quantity will be problematic if the same quantity of total RNA between the two groups is contributed by different numbers of cells. 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 total RNA levels, this will lead to underestimation of the sncRNA expression level in cancer cells. Solution to such situation could be either performing Northern blots with or adding 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 RNAs 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 potentials and extracellular molecular interactions[48, 68-70]. This issue has become particularly significant for the emerging noncanonical sncRNAs that are derived from highly modified precursors such as tRNAs, which harbour more than 150 types of modifications[71]. However, by far many modifications of sncRNAs 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 only to a few well-known modifications (e.g., 5-Methylcytosine (m^5C), N^6-methyladenosine (m^6A), pseudouridine (ψ), inosine (I), N^1-methyladenosine (m^1A) and N^4-acetylcytidine (ac^4C)). Commonly used approaches included antibody-dependent methods, chemical conversion of the targeted modifications into a distinguishable base[72-80], and the newly developed nanopore-based direct RNA sequencing[81-83], but these methods usually analyze only one modification type at a time. Other methods, such as inferring the nucleotide misincorporation during reverse transcription, 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 currently no efficient methods for high-throughput, comprehensive, quantitative mapping of multiple types of modifications in sncRNAs, or RNAs in general.

While 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]. Currently, two classes of methods are being explored for direct RNA sequencing and quantitative multiplexed mapping of RNA modifications, either based on mass spectrometry (MS) or nanopore technology.

Mass spectrometry: old dog, new tricks

Liquid chromatography–tandem mass spectrometry (LC-MS/MS) has been widely used to analyze RNA modifications and is considered the 'gold standard' to quantify modifications in an RNA sample, because compared to 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 peptide sequence)[90]. However, when RNAs are digested into smaller fragments or single nucleotides before MS examination, the 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 NGS-based RNA-seq[91].

In theory, using MS to directly measure RNA sequences and RNA 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 by specific enzymatic cleavage[93].

In 2015, a landmark paper from the Jack Szostak lab overcame this challenge by developing an generalized and efficient way to fragment RNA in a controllable manner followed by 2D mass-retention time analysis of the resulting RNA fragments by LC separation, which permits the generation of perfect RNA mass ladders for direct RNA sequencing[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'-5' and 5'-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 optimizing the RNA degradation protocol to more evenly generate RNA fragments of different lengths and using 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 2D mass ladders and enabled the reading of the complete sequence of a given RNA from either the 3′- or the 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 (~80 nt) and advanced algorithms, the 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 upon the treatment with the dealkylating enzyme AlkB, and uncovered new RNA modifications that had not been reported for $tRNA^{Glu}$ [96]. MLC-seq will be particularly useful for the study of highly modified RNAs such as tRNAs/tsRNAs, and address open questions such as the tissue-specific differences in tRNAs/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 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 the recent decades, nanopore technology is now bringing a revolution in direct DNA/RNA sequencing due to its unique characteristics including label-free, amplification-free, and real-time detection of DNA/RNA at 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.

Indeed, 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 tRNAs. A recent attempt using Oxford Nanopore MinION to comparatively sequence purified biological tRNAs (from *E. coli*) versus corresponding synthetic non-modified tRNAs 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. Additionally, the reading accuracy of synthetic non-modified tRNAs is lower than that of mRNAs[99], suggesting that the current method is not well-adapted for short RNAs (e.g., tRNAs and sncRNAs) and awaits improvement, such as ligating the tRNA/sncRNA to longer adapter RNAs with optimized sequences.

One major difficulty in accurately mapping RNA modification using nanopores is that the presence of 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/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/tsRNAs where the effects of different RNA modifications may overlap and generate complicated situations. In theory, this problem might be conquered by synthesizing

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 currently cannot be readily synthesized. 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 (e.g., site-specific mutation) either the main pore or the motor protein of the existing nanopores, or both, or to choose completely different pores (e.g., 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 substantially increasing, 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) reverse transcription of RNA also has the potential to directly detect multiple RNA modifications from the RNA template through analysing the kinetics of the reverse transcriptase using Zero-mode waveguides (ZMWs)[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. In fact, the past few years have witnessed 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. However, these methods are mostly optimized for long RNAs such as mRNAs, while the short length of sncRNAs has limited the options in designing nucleic acid probes, and the probe may bind to multiple targets (e.g., both sncRNAs and their precursors); thus, the locations of sncRNAs would be difficult to determine with accuracy. Additionally, 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 within 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 RNAs, DNA, and RNA-binding proteins, as well as the social behaviour of sncRNAs within (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 programs to systematically predict RNA[107] and protein[102] 3D structures, which should also make the systematic prediction of RNA-protein interactions only be 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.

Acknowledgements

We thank Paul Schimmel (The Scripps Research Institute), Xuemei Chen (UC Riverside) and our lab members for critical discussions on the contents of the manuscript. Research in the Q.C. lab is in part supported by the National Institutes of Health (NIH, R01HD092431, R01ES032024 and P50HD098593). The T.Z. lab is in part supported by the NIH (R01ES032024). The authors apologize for not being able to cite all related literatures due to length limitation.

References:

1. Grosshans H & Filipowicz W Molecular biology: the expanding world of small RNAs. Nature 451, 414–416 (2008). [PubMed: 18216846]
2. Storz G, Vogel J & Wassarman KM Regulation by small RNAs in bacteria: expanding frontiers. Molecular cell 43, 880–891 (2011). [PubMed: 21925377]
3. Babski J et al. Small regulatory RNAs in Archaea. RNA biology 11, 484–493 (2014). [PubMed: 24755959]
4. Carthew RW & Sontheimer EJ Origins and Mechanisms of miRNAs and siRNAs. Cell 136, 642–655 (2009). [PubMed: 19239886]
5. Bartel DP Metazoan MicroRNAs. Cell 173, 20–51 (2018). [PubMed: 29570994]
6. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D & Zamore PD PIWI-interacting RNAs: small RNAs with big functions. Nature reviews. Genetics 20, 89–108 (2019).
7. Seal RL et al. A guide to naming human non-coding RNA genes. The EMBO journal 39, e103777 (2020). [PubMed: 32090359]
8. Chen Q, Zhang X, Shi J, Yan M & Zhou T Origins and evolving functionalities of tRNA-derived small RNAs. Trends in biochemical sciences (2021).
9. Schimmel P The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. Nature reviews. Molecular cell biology 19, 45–58 (2018). [PubMed: 28875994]
10. Shi J et al. PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications. Nature cell biology 23, 424–436 (2021). [PubMed: 33820973]
11. Gu W et al. Peripheral blood non-canonical small non-coding RNAs as novel biomarkers in lung cancer. Mol Cancer 19, 159 (2020). [PubMed: 33176804]
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). [PubMed: 28167565]
13. Chen CJ & Heard E Small RNAs derived from structural non-coding RNAs. Methods 63, 76–84 (2013). [PubMed: 23684746]
14. Wang H et al. CPA-seq reveals small ncRNAs with methylated nucleosides and diverse termini. Cell Discov 7, 25 (2021). [PubMed: 33867522]
15. Taft RJ et al. Small RNAs derived from snoRNAs. Rna 15, 1233–1240 (2009). [PubMed: 19474147]
16. Ender C et al. A human snoRNA with microRNA-like functions. Molecular cell 32, 519–528 (2008). [PubMed: 19026782]

17. Persson H et al. The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. Nature cell biology 11, 1268–1271 (2009). [PubMed: 19749744]
18. Hussain S et al. NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell reports 4, 255–261 (2013). [PubMed: 23871666]
19. Pircher A, Bakowska-Zywicka K, Schneider L, Zywicki M & Polacek N An mRNA-derived noncoding RNA targets and regulates the ribosome. Molecular cell 54, 147–155 (2014). [PubMed: 24685157]
20. Reuther J et al. A small ribosome-associated ncRNA globally inhibits translation by restricting ribosome dynamics. RNA biology, 1-16 (2021).
21. Tuck AC & Tollervey D RNA in pieces. Trends Genet 27, 422–432 (2011). [PubMed: 21741109]
22. Schaefer M et al. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes & development 24, 1590–1595 (2010). [PubMed: 20679393]
23. Tuorto F et al. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. Nature structural & molecular biology 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). [PubMed: 26721680]
25. Zhang Y et al. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. Nature cell biology 20, 535–540 (2018). [PubMed: 29695786]
26. Guzzi N et al. Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. Cell 173, 1204–1216 e1226 (2018). [PubMed: 29628141]
27. Natt D et al. Human sperm displays rapid responses to diet. PLoS biology 17, e3000559 (2019). [PubMed: 31877125]
28. Goodarzi H et al. Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. Cell 161, 790–802 (2015). [PubMed: 25957686]
29. Kim HK et al. A transfer-RNA-derived small RNA regulates ribosome biogenesis. Nature 552, 57–62 (2017). [PubMed: 29186115]
30. Balatti V et al. tsRNA signatures in cancer. Proceedings of the National Academy of Sciences of the United States of America 114, 8071–8076 (2017). [PubMed: 28696308]
31. Yue T et al. SLFN2 protection of tRNAs from stress-induced cleavage is essential for T cell-mediated immunity. Science 372 (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). [PubMed: 23183536]
33. Liu YM et al. Exosome-delivered and Y RNA-derived small RNA suppresses influenza virus replication. J Biomed Sci 26, 58 (2019). [PubMed: 31416454]
34. Hogg MC et al. Elevation in plasma tRNA fragments precede seizures in human epilepsy. The Journal of clinical investigation 129, 2946–2951 (2019). [PubMed: 31039137]
35. Zhang X et al. Small RNA modifications in Alzheimer's disease. Neurobiol Dis 145, 105058 (2020). [PubMed: 32835860]
36. Blanco S et al. Stem cell function and stress response are controlled by protein synthesis. Nature 534, 335–340 (2016). [PubMed: 27306184]
37. Sajini AA et al. Loss of 5-methylcytosine alters the biogenesis of vault-derived small RNAs to coordinate epidermal differentiation. Nature communications 10, 2550 (2019).
38. Krishna S et al. Dynamic expression of tRNA-derived small RNAs define cellular states. EMBO reports 20, e47789 (2019). [PubMed: 31267708]
39. Kfoury YS et al. tiRNA signaling via stress-regulated vesicle transfer in the hematopoietic niche. Cell stem cell (2021).
40. Schorn AJ, Gutbrod MJ, LeBlanc C & Martienssen R LTR-Retrotransposon Control by tRNA-Derived Small RNAs. Cell 170, 61–71 e11 (2017). [PubMed: 28666125]
41. Martinez G, Choudury SG & Slotkin RK tRNA-derived small RNAs target transposable element transcripts. Nucleic acids research 45, 5142–5152 (2017). [PubMed: 28335016]

42. Sarker G et al. Maternal overnutrition programs hedonic and metabolic phenotypes across generations through sperm tsRNAs. Proceedings of the National Academy of Sciences of the United States of America 116, 10547–10556 (2019). [PubMed: 31061112]
43. Sharma U et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science 351, 391–396 (2016). [PubMed: 26721685]
44. Wahba L, Hansen L & Fire AZ An essential role for the piRNA pathway in regulating the ribosomal RNA pool in C. elegans. Developmental cell 56, 2295–2312 e2296 (2021). [PubMed: 34388368]
45. Zhang Y et al. Angiogenin mediates paternal inflammation-induced metabolic disorders in offspring through sperm tsRNAs. Nature communications 12, 6673 (2021).
46. Honda S et al. Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers. Proceedings of the National Academy of Sciences of the United States of America 112, E3816–3825 (2015). [PubMed: 26124144]
47. Cozen AE et al. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. Nature methods 12, 879–884 (2015). [PubMed: 26237225]
48. Zhang X, Cozen AE, Liu Y, Chen Q & Lowe TM Small RNA Modifications: Integral to Function and Disease. Trends in molecular medicine 22, 1025–1034 (2016). [PubMed: 27840066]
49. Huang X, Fejes Toth K & Aravin AA piRNA Biogenesis in Drosophila melanogaster. Trends Genet 33, 882–894 (2017). [PubMed: 28964526]
50. Shabalina SA & Koonin EV Origins and evolution of eukaryotic RNA interference. Trends in ecology & evolution 23, 578–587 (2008). [PubMed: 18715673]
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 (2021).
52. Lee SR & Collins K Starvation-induced cleavage of the tRNA anticodon loop in Tetrahymena thermophila. The Journal of biological chemistry 280, 42744–42749 (2005). [PubMed: 16272149]
53. Thompson DM, Lu C, Green PJ & Parker R tRNA cleavage is a conserved response to oxidative stress in eukaryotes. Rna 14, 2095–2103 (2008). [PubMed: 18719243]
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). [PubMed: 23326205]
55. Garcia-Silva MR et al. Extracellular vesicles shed by Trypanosoma cruzi are linked to small RNA pathways, life cycle regulation, and susceptibility to infection of mammalian cells. Parasitology research 113, 285–304 (2014). [PubMed: 24241124]
56. Fricker R et al. A tRNA half modulates translation as stress response in Trypanosoma brucei. Nature communications 10, 118 (2019).
57. Peng H et al. A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. Cell research 22, 1609–1612 (2012). [PubMed: 23044802]
58. Dhahbi JM 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). [PubMed: 23638709]
59. Zhang Y et al. Identification and characterization of an ancient class of small RNAs enriched in serum associating with active infection. Journal of molecular cell biology 6, 172–174 (2014). [PubMed: 24380870]
60. Raabe CA, Tang TH, Brosius J & Rozhdestvensky TS Biases in small RNA deep sequencing data. Nucleic acids research 42, 1414–1426 (2014). [PubMed: 24198247]
61. Jayaprakash AD, Jabado O, Brown BD & Sachidanandam R Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic acids research 39, e141 (2011). [PubMed: 21890899]
62. Saunders K et al. Insufficiently complex unique-molecular identifiers (UMIs) distort small RNA sequencing. Scientific reports 10, 14593 (2020). [PubMed: 32884024]
63. Faridani OR et al. Single-cell sequencing of the small-RNA transcriptome. Nature biotechnology 34, 1264–1266 (2016).

64. Yang Q et al. Single-cell CAS-seq reveals a class of short PIWI-interacting RNAs in human oocytes. Nature communications 10, 3389 (2019).
65. Shi J, Ko EA, Sanders KM, 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). [PubMed: 29730207]
66. Hu JF et al. Quantitative mapping of the cellular small RNA landscape with AQRNA-seq. Nature biotechnology 39, 978–988 (2021).
67. Loven J et al. Revisiting global gene expression analysis. Cell 151, 476–482 (2012). [PubMed: 23101621]
68. Ji L & Chen X Regulation of small RNA stability: methylation and beyond. Cell research 22, 624–636 (2012). [PubMed: 22410795]
69. Frye M, Harada BT, Behm M & He C RNA modifications modulate gene expression during development. Science 361, 1346–1349 (2018). [PubMed: 30262497]
70. Flynn RA et al. Small RNAs are modified with N-glycans and displayed on the surface of living cells. Cell 184, 3109–3124 e3122 (2021). [PubMed: 34004145]
71. Suzuki T The expanding world of tRNA modifications and their disease relevance. Nature reviews. Molecular cell biology 22, 375–392 (2021). [PubMed: 33658722]
72. Schaefer M, Pollex T, Hanna K & Lyko F RNA cytosine methylation analysis by bisulfite sequencing. Nucleic acids research 37, e12 (2009). [PubMed: 19059995]
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). [PubMed: 21370043]
74. Schwartz S et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell 159, 148–162 (2014). [PubMed: 25219674]
75. Carlile TM et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature 515, 143–146 (2014). [PubMed: 25192136]
76. Hussain S, Aleksic J, Blanco S, Dietmann S & Frye M Characterizing 5-methylcytosine in the mammalian epitranscriptome. Genome biology 14, 215 (2013). [PubMed: 24286375]
77. Dominissini D et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 485, 201–206 (2012). [PubMed: 22575960]
78. Meyer KD et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell 149, 1635–1646 (2012). [PubMed: 22608085]
79. Sas-Chen A et al. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. Nature 583, 638–643 (2020). [PubMed: 32555463]
80. Li X et al. Base-Resolution Mapping Reveals Distinct m(1)A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. Molecular cell 68, 993–1005 e1009 (2017). [PubMed: 29107537]
81. Begik O et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. Nature biotechnology 39, 1278–1291 (2021).
82. Liu H et al. Accurate detection of m(6)A RNA modifications in native RNA sequences. Nature communications 10, 4079 (2019).
83. Parker MT et al. Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m(6)A modification. Elife 9 (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 research 48, 3734–3746 (2020). [PubMed: 32095818]
85. Khoddami V et al. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. Proceedings of the National Academy of Sciences of the United States of America 116, 6784–6789 (2019). [PubMed: 30872485]
86. Behrens A, Rodschinka G & Nedialkova DD High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. Molecular cell 81, 1802–1815 e1807 (2021). [PubMed: 33581077]
87. Sas-Chen A & Schwartz S Misincorporation signatures for detecting modifications in mRNA: Not as simple as it sounds. Methods 156, 53–59 (2019). [PubMed: 30359724]

88. Owens MC, Zhang C & Liu KF Recent technical advances in the study of nucleic acid modifications. Molecular cell 81, 4116–4136 (2021). [PubMed: 34480848]

89. Alfonzo JD et al. A call for direct sequencing of full-length RNAs to identify all modifications. Nature genetics 53, 1113–1116 (2021). [PubMed: 34267373]

90. Ross RL, Cao X & Limbach PA Mapping Post-Transcriptional Modifications onto Transfer Ribonucleic Acid Sequences by Liquid Chromatography Tandem Mass Spectrometry. Biomolecules 7 (2017).

91. Kimura S, Dedon PC & Waldor MK Comparative tRNA sequencing and RNA mass spectrometry for surveying tRNA modifications. Nature chemical biology 16, 964–972 (2020). [PubMed: 32514182]

92. Sample PJ, Gaston KW, Alfonzo JD & Limbach PA RoboOligo: software for mass spectrometry data to support manual and de novo sequencing of post-transcriptionally modified ribonucleic acids. Nucleic acids research 43, e64 (2015). [PubMed: 25820423]

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). [PubMed: 26495937]

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 research 47, e125 (2019). [PubMed: 31504795]

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). [PubMed: 32364699]

96. Yuan X MLC-Seq: de novo Sequencing of Full-Length tRNAs and Quantitative Mapping of Multiple RNA Modifications. Researchsquare Doi: 10.21203/rs.3.rs-1090754/v1 (2021).

97. Kasianowicz JJ, Brandin E, Branton D & Deamer DW Characterization of individual polynucleotide molecules using a membrane channel. Proceedings of the National Academy of Sciences of the United States of America 93, 13770–13773 (1996). [PubMed: 8943010]

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). [PubMed: 29232619]

99. Thomas NK et al. Direct Nanopore Sequencing of Individual Full Length tRNA Strands. ACS Nano (2021).

100. Garalde DR et al. Highly parallel direct RNA sequencing on an array of nanopores. Nature methods 15, 201–206 (2018). [PubMed: 29334379]

101. Smith AM, Jain M, Mulroney L, Garalde DR & Akeson M Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. PLoS One 14, e0216709 (2019). [PubMed: 31095620]

102. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). [PubMed: 34265844]

103. Vilfan ID et al. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. J Nanobiotechnology 11, 8 (2013). [PubMed: 23552456]

104. Zhuang X Spatially resolved single-cell genomics and transcriptomics by imaging. Nature methods 18, 18–22 (2021). [PubMed: 33408406]

105. Larsson L, Frisen J & Lundeberg J Spatially resolved transcriptomics adds a new dimension to genomics. Nature methods 18, 15–18 (2021). [PubMed: 33408402]

106. Zhang Y, Shi J, Rassoulzadegan M, Tuorto F & Chen Q Sperm RNA code programmes the metabolic health of offspring. Nature reviews. Endocrinology 15, 489–498 (2019).

107. Townshend RJL et al. Geometric deep learning of RNA structure. Science 373, 1047–1051 (2021). [PubMed: 34446608]

108. Haussecker D et al. Human tRNA-derived small RNAs in the global regulation of RNA silencing. Rna 16, 673–695 (2010). [PubMed: 20181738]

109. Yamasaki S, Ivanov P, Hu GF & Anderson P Angiogenin cleaves tRNA and promotes stress-induced translational repression. The Journal of cell biology 185, 35–42 (2009). [PubMed: 19332886]
110. Lee YS, Shibata Y, Malhotra A & Dutta A A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes & development 23, 2639–2649 (2009). [PubMed: 19933153]
111. 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). [PubMed: 30538719]
112. Akat KM et al. Detection of circulating extracellular mRNAs by modified small-RNA-sequencing analysis. JCI Insight 5 (2019).
113. Dai H & Gu W Strategies and Best Practice in Cloning Small RNAs. Gene Technol 9 (2020).
114. 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 biology, 1–12 (2020).
115. Zheng G et al. Efficient and quantitative high-throughput tRNA sequencing. Nature methods 12, 835–837 (2015). [PubMed: 26214130]
116. Dai Q, Zheng G, Schwartz MH, Clark WC & 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 Engl 56, 5017–5020 (2017). [PubMed: 28371071]
117. Xu H, Yao J, Wu DC & Lambowitz AM Improved TGIRT-seq methods for comprehensive transcriptome profiling with decreased adapter dimer formation and bias correction. Scientific reports 9, 7953 (2019). [PubMed: 31138886]
118. Upton HE et al. Low-bias ncRNA libraries using ordered two-template relay: Serial template jumping by a modified retroelement reverse transcriptase. Proceedings of the National Academy of Sciences of the United States of America 118 (2021).
119. Cech TR & Steitz JA The noncoding RNA revolution-trashing old rules to forge new ones. Cell 157, 77–94 (2014). [PubMed: 24679528]
120. Helwak A, Kudla G, Dudnakova T & Tollervey D Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153, 654–665 (2013). [PubMed: 23622248]
121. Shen EZ et al. Identification of piRNA Binding Sites Reveals the Argonaute Regulatory Landscape of the C. elegans Germline. Cell 172, 937–951 e918 (2018). [PubMed: 29456082]
122. Kumar P, Anaya J, Mudunuri SB & 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 biology 12, 78 (2014). [PubMed: 25270025]
123. Guan L, Karaiskos S & Grigoriev A Inferring targeting modes of Argonaute-loaded tRNA fragments. RNA biology 17, 1070–1080 (2020). [PubMed: 31613177]
124. Guan L & Grigoriev A Computational meta-analysis of ribosomal RNA fragments: potential targets and interaction mechanisms. Nucleic acids research 49, 4085–4103 (2021). [PubMed: 33772581]
125. Honda S, Morichika K & Kirino Y Selective amplification and sequencing of cyclic phosphate-containing RNAs by the cP-RNA-seq method. Nature protocols 11, 476–489 (2016). [PubMed: 26866791]

Box 1.

A unified naming system for sncRNAs derived from longer RNA precursors

Studies of noncanonical sncRNAs have been constantly 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 labs in different contexts under different names including tRNA-derived small RNAs (tsRNAs)[24, 29, 108], tRNA-derived small RNAs (tDRs)[56], tRNA-derived stress-induced RNAs (tiRNAs)[31, 39, 109] and tRNA fragments (tRFs)[28, 43, 110]. Here, we propose a unified nomenclature for noncanonical 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 labs and has the potential for further use in the research community. While some labs 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 (e.g., tsRNAs) would need the group efforts of each community.

Precursor RNAs	**Derivative sncRNAs**
Transfer RNA (tRNA)	tRNA-derived small RNA (tsRNA)
Ribosomal RNA (rRNA)	rRNA-derived small RNA (rsRNA)
Y RNA (yRNA)	yRNA-derived small RNA (ysRNA)
Vault RNA (vtRNA)	vtRNA-derived small RNA (vtsRNA)
Small nuclear RNA (snRNA)	snRNA-derived small RNA (snsRNA)
Small nucleolar RNA (snoRNA)	snoRNA-derived small RNA (snosRNA)
Long non-coding RNA (lncRNA)	lncRNA-derived small RNA (lncsRNA)
Messenger RNA (mRNA) ……	mRNA-derived small RNA (msRNA) ……

Box 2.

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

Among many sources of sequencing biases[60], one major aspect comes from adapter ligation process during cDNA library construction (Fig.1a,b). The ligation process is designed to (ideally) add adapter 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 (e.g., siRNAs and miRNAs) and RNase P/RNase Z (e.g., a portion of tsRNAs) bear a 5′-phosphate (5′-P) and a 3′-hydroxyl (3′-OH) termini[108], whereas sncRNAs generated by RNase T2/RNase A (e.g., 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)[111]. 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/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/3’-P into 3’-OH, and 5′-OH into 5’-P before the ligation process[112], or combining with a template-switching strategy to add a 5’ adapter to the cDNA after the reverse transcription, instead of directly adding a 5’ adapter to the RNA[113, 114], which can resolve most problems caused by 5’ terminal modifications.

The second major source of bias comes from the reverse transcription (RT) process, which converts the adapter-ligated RNA into cDNA (Fig.1c). Several RNA modifications (e.g., m^1A, m^3C, m^1G, and $m^2{}_2G$) can interfere with the RT process, either by preventing the passage of reverse transcriptase or generating misincorporation at the modified loci[47, 115, 116]. 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 either using enzymes to remove these RT blocking modifications (e.g. AlkB)[47, 115, 116], or using a high-processive reverse transcriptase (e.g. TGIRT, BoMoC) to read through the modifications without being blocked[117, 118]. The latter approach retains the misincorporation, which can be used to infer the nature of the modification[86].

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[119]. 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 when new knowledge arrives, we realized 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 noncanonical, we may keep in mind that in principle, all RNA sequences (sometimes tuned by RNA modifications) harness base-pairing to bind to their DNA/RNA targets, and their interactions with protein targets are based on their molecular structure. For example, earlier studies using CLASH, an experimental approach to identify RNA-RNA duplexes associated with Argonaute proteins in vivo, focused on revealing the RNA targets of miRNAs[120] or piRNAs[121]; however, later, more comprehensive analyses using these same datasets have revealed extensive tsRNA-mRNA interactions[122, 123], rsRNA-mRNA interactions[124], and even interactions between sncRNAs[124]. Further analyses extending to the potential interactions between other sncRNAs and long RNAs are highly possible and await discovery.

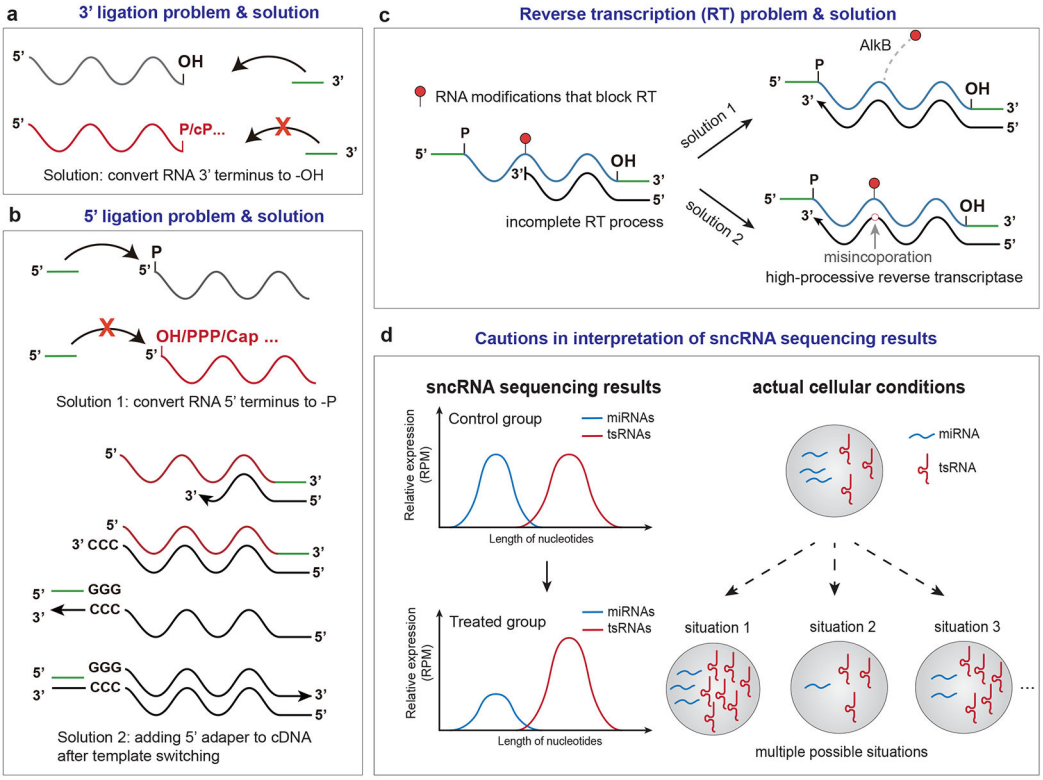

Fig.1. Methods to overcome biases in sncRNA discovery and cautions in interpretation of sncRNA sequencing results.

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

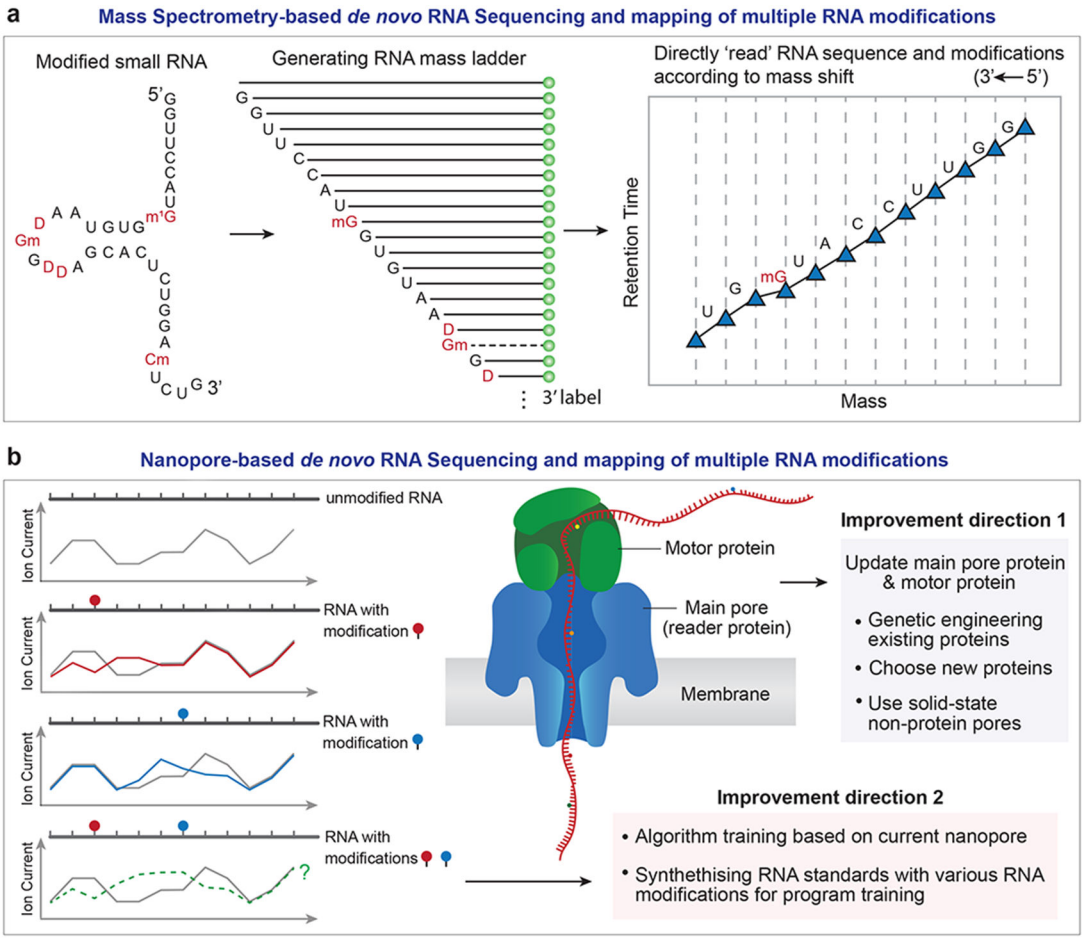

Fig.2. Two methods for future direct sequencing of RNA and multiplexed mapping of RNA modifications without cDNA intermediates.

(**a**) Main concept and workflow for mass spectrometry-based *de novo* sequencing of modified sncRNA, which involves the controlled fragmentation of RNA (by formic acid) into ladder fragments, followed by LC-MS measurement of the resultant RNA fragments, generating sequence of both canonical and modified nucleosides based on 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 for distinguishing between m^1A and m^6A, or between 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 hydrolysation 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**) Illustrative figure 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 in the figure, and ideally will be applied together.

Table 1

Recent methods to improve sncRNA sequencing (NGS) by overcoming specific RNA modifications

Method	Resolving terminal modifications to improve ligation	Resolving internal modifications to improve reverse transcription (RT)	Other features and concerns
ARM-Seq[47]	Unresolved	• AlkB treatment to remove RNA modifications that block RT	• Potential degradation of longer RNAs (e.g., tRNAs) during AlkB treatment would generate RNA fragments which will be sequenced as artifacts[10]
cP-RNA-seq[125]	• 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 • sncRNAs containing both 2′,3′-cP and other RT blocking modifications could be missed
improved RNA-seq[112]	• T4PNK treatment converts 3’-P and 2’,3’-cP at 3’ terminal into 3-OH; converts 5’-OH at 5’ terminal into 5’-P	Unresolved	
5′XP sRNA-seq[114]	• 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 sncRNAs • The analyses limited to sncRNAs that have a 3’-OH and do not have RT blocking modifications
TGIRT-Seq[117]	• T4PNK treatment converts 3’-P and 2’,3’-cP at 3’ terminal into 3-OH • Template-switching activity by TGIRT adds adapter to 3’ end of cDNA instead of 5’end of RNA, thus resolve 5’ RNA modifications	• Highly processive reverse transcriptase TGIRT to read through RNA modifications	• Simultaneous profiling of longer RNAs (e.g., mRNAs and lncRNAs) • TGIRT cannot always read though RNA modifications and need reaction condition optimization[86]
AQRNA-seq[66]	• Alkaline Phosphatase treatment to convert 3’-P into 3’-OH, 5’-P into 5’-OH • Add adapter to 3’end of cDNA instead of 5’end of RNA • 2’,3’-cP unresolved	• AlkB treatment to remove RNA modifications that block RT	• Quantification of sncRNAs by adding spike-in RNAs in the sample
CPA-seq[14]	• Cap-Clip treatment removes the 5′-cap and 5′-ppp to generate 5′-P • T4PNK treatment converts 3’-P and 2’,3’-cP at 3’ terminal into 3-OH; converts 5’-OH at 5’ terminal into 5’-P	• AlkB treatment to remove RNA modifications that block RT • Highly processive reverse transcriptase TGIRT to read	• Deacylation buffer (pH = 9.0) to remove aminoacyl residues in 3-’tsRNAs • Potential degradation of longer RNAs (e.g., tRNAs) during AlkB treatment would generate RNA fragments which will be sequenced as artifacts[10]

Method	Resolving terminal modifications to improve ligation	Resolving internal modifications to improve reverse transcription (RT)	Other features and concerns
		through RNA modifications	
PANDORA-seq[10]	• T4PNK treatment converts 3'-P and 2',3'-cP at 3' terminal into 3-OH; converts 5'-OH at 5' terminal into 5'-P	• AlkB treatment to remove RNA modifications that block RT	• Pre-size selection (<50 nt RNA) eliminate fragmentation of longer RNA (e.g., tRNAs) by AlkB treatment • Data analysis by *SPORTS*[65] to improve noncanonical 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.