6-10-2022

Challenges Inherent to the Sequencing and Quantification of Short Non-Coding RNAs and How to Address Them

Justin Gumas
Thomas Jefferson University

Yohei Kirino
Thomas Jefferson University

Received: 27 May 2022 | Accepted: 29 May 2022

DOI: 10.1002/ctd2.93

CLINICAL AND TRANSLATIONAL DISCOVERY Open Access

WILEY

COMMENTARY

Challenges inherent to the sequencing and quantification of short non-coding RNAs and how to address them

Justin Gumas | **Yohei Kirino**

Computational Medicine Center, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania, USA

Correspondence
Yohei Kirino, 1020 Locust Street, JAH Suite Rm222, Philadelphia, PA 19107, USA.
Email: Yohei.Kirino@jefferson.edu

[Correction added on November 5, 2022 after first online publication: the Author Contributions, Acknowledgments, Funding, Conflicts of Interest, Data Availability Statement and Ethical approval has been updated].

In recent decades, advances in the understanding of RNA biology have increased appreciation of the role of short non-coding RNAs (sncRNAs) in the pathophysiology of human diseases. Coupled with the accumulation of publicly available RNA sequencing (RNA-seq) datasets, these advances have produced exciting opportunities to focus on and characterize specific sncRNA molecules as therapeutic targets and biomarkers. However, as sncRNA research grows, the need for greater appreciation of the diverse biochemical properties of sncRNAs and the challenges these pose for their proper analysis also arise. Cellular sncRNA molecules generally possess either a hydroxyl group (OH), a monophosphate (P), or a 2′,3′-cyclic phosphate (2′,3′-cP) at their termini. OH and P can be found at both the 5′- and 3′-ends of RNAs, while 2′,3′-cP presents only at the 3′-end of RNAs[1] (Figure 1**A**). For their characterization, there is no "one size fits all" approach that can shed light on the full range of sncRNAs. The selection of sequencing and quantification methods must be deliberate and tailored to the specific sncRNA molecules of interest, which requires understanding the properties of the sncRNAs.

Most of the RNA-seq data for sncRNAs have been obtained for the analyses of microRNAs (miRNAs), the best-studied sncRNAs so far, by using standard RNA-seq. Standard RNA-seq relies on 5′-P/3′-OH ends of sncRNAs for adaptor (AD) ligations, and thus, it cannot capture non-miRNA-sncRNAs without the 5′-P/3′-OH ends (Figure 1B). To address the issue of this invisibility of non-miRNA-sncRNAs, specific RNA-seq methods compatible with various terminal formations of sncRNAs have been developed,[2] enabling characterization of the previously unrecognized sncRNAs. For example, cP-RNA-seq can specifically amplify and sequence 2′,3′-cP-containing sncRNAs[3,4] (Figure 1B), revealing the expression profiles of 5′-transfer RNA (tRNA) halves and other 2′,3′-cP-containing sncRNAs[5] in diseases and germline development.[3,6,7] In contrast, a broader range of sncRNAs with all terminal phosphate states can be sequenced by Phospho-RNA-seq[8] (Figure 1B), in which RNA samples treated with T4 polynucleotide kinase (T4 PNK) are sequenced. Dephosphorylation of 3′-P/2′,3′-cP by T4 PNK treatment is crucial to capture the whole picture of extracellular sncRNAs (ex-sncRNAs) in samples such as plasma and serum, as the majority of ex-sncRNAs contain 3′-P or 2′,3′-cP ends, rendering them invisible to standard RNA-seq.[6,8] However, T4 PNK treatment erases the information of terminal phosphate states which are important to understanding the biogenesis and function of sncRNAs. It is also possible that the expression profiles of sncRNAs with relatively low abundance could be masked by increased amounts and variations of the sncRNAs in T4 PNK-treated RNA-seq data. Therefore, choosing the best sequencing method for sncRNA analyses depends on research purposes and the properties of the target sncRNA molecules, such as their terminal formation and abundance relative to the whole sncRNAome.

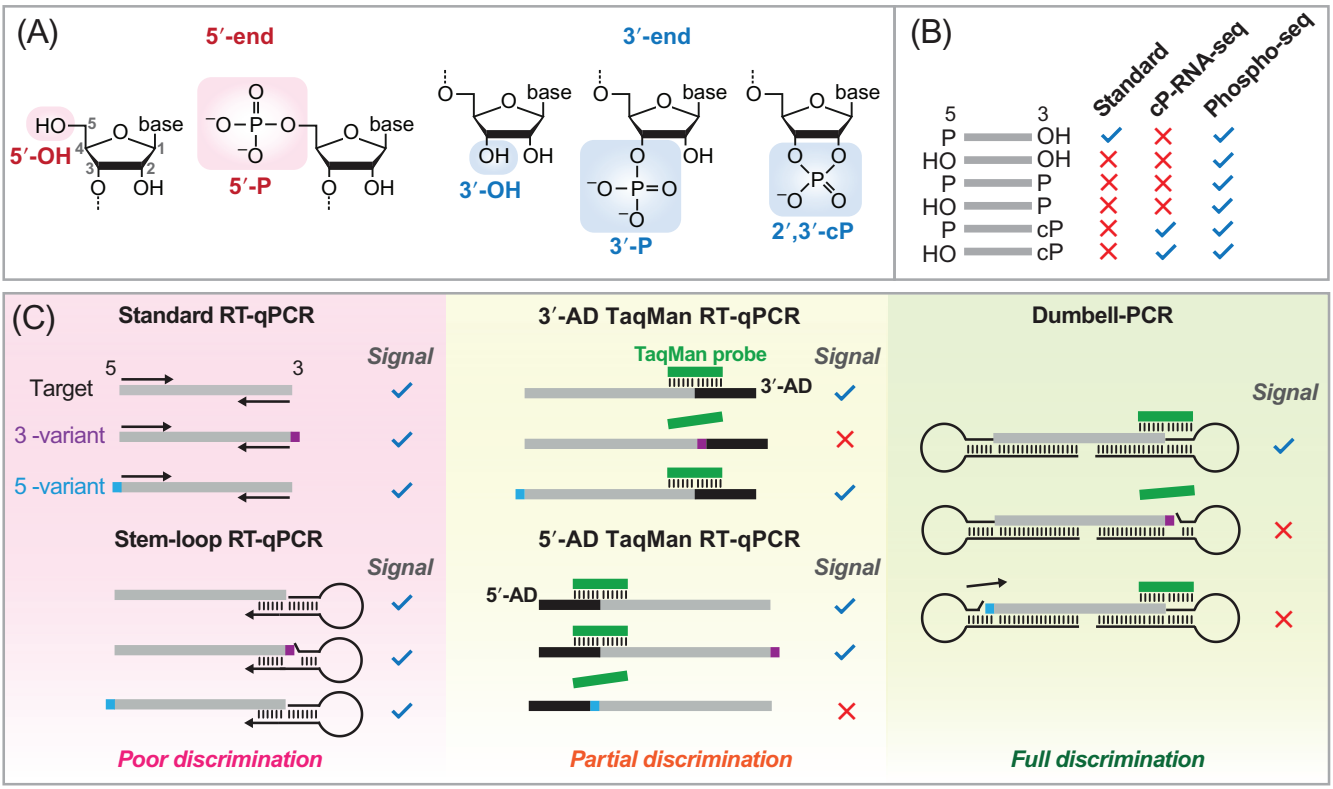

FIGURE 1 Specific sequencing and quantification of short non-coding RNAs (sncRNAs). (A) Chemical structures of sncRNA termini (modified from our previous review[2]). (B) Targeted sncRNAs in standard RNA-seq, cP-RNA-seq, and phospho-seq. (C) Discrimination of sncRNA terminal variants in standard RT-qPCR, stem-loop RT-qPCR, TaqMan RT-qPCR for AD-ligated sncRNAs, and Db-PCR

Although appropriately selected RNA-seq methods can capture the entire repertoire of targeted sncRNAs, the required cost, time, and bioinformatic analyses could preclude the use of RNA-seq for examining only specific sncRNAs in numerous samples. Accurate and convenient quantification of individual sncRNA molecules is thus imperative in sncRNA research. Cellular sncRNAs are often expressed as multiple isoforms with complex heterogeneity in both length and terminal sequences. For example, identical miRNA genes encode mature isoforms, termed isomiRs, that vary in size by one or more nucleotides (nt) at the 5′- and/or 3′-end of the miRNA. The sncRNAs derived from tRNAs and ribosomal RNAs (rRNAs) also exhibit complex heterogeneity,[5] and those sncRNAs co-exist with their much more abundant precursor molecules (i.e., mature tRNAs/rRNAs and their precursors). Conventional RT-qPCR, microarray, and northern blot are all insufficient to distinguish target sncRNA with its precursor and/or variants. A stem-loop RT-qPCR,[9] widely used to quantify miRNAs and other sncRNAs, is also unable to clearly discriminate the amplification signals of target sncRNA from those of their variants[10] (Figure 1C). In response to the failure of these methods to specifically quantify sncRNAs, TaqMan RT-qPCR for AD-ligated RNAs has been developed originally for quantification of tRNA half molecules[3] and has since been further utilized for other sncRNAs such as those from rRNAs, messenger RNAs (mRNAs), or transposons[5,7] (Figure 1C). In this method, linear AD is ligated to the 5′- or 3′-ends of target RNA, and the TaqMan probe targets the boundary of the ligation product with single-nucleotide resolution. Because this method cannot simultaneously ensure the specificity of both 5′- and 3′-ends of the target RNA, an extended version of this method, termed Dumbbell-PCR (Db-PCR), has further been developed[10] (Figure 1C). In Db-PCR, both the 5′- and 3′-ends of RNAs are ligated to 5′- and 3′-stem-loop ADs, respectively, followed by TaqMan RT-qPCR to ensure specific quantification of target RNA with single-nucleotide resolution at both ends.[10]

The above considerations on sncRNA sequencing and quantification speak to the importance of understanding the biogenesis pathways and resultant properties of target sncRNAs before characterizing them. Appropriate choice of sequencing and quantification methods for specific sncRNAs, as well as accurate interpretation of experimental results, depends on the molecules' terminal formations, abundances relative to their precursors/variants and other sncRNAs, post-transcriptional modifications, etc. What makes things more complicated is that these properties of sncRNAs can be heterogeneous and vary in different

samples, tissues, and/or disease settings. Although the realm of sncRNA is expanding and various forms of RNA-seq have been revealing greater numbers of sncRNA species, the biogenesis mechanisms and responsible ribonucleases remain unknown for a large part of sncRNAs. Further basic biology research to clarify those are necessary to utilize sncRNA molecules in clinical and translational contexts, such as to use circulating ex-sncRNAs in serum or plasma as disease biomarkers. In the meantime, current RNA biology knowledge accumulated on each class of sncRNAs should be fully leveraged in determining methodologies for sncRNA detection and quantification in clinical research. The full realization of the clinical potential of sncRNAs will occur insofar as the insights of basic RNA biology are recognized and utilized in the translational context.

AUTHOR CONTRIBUTIONS

Mr. Justin Gumas and Dr. Yohei Kirino contributed to the preparation and collection of original literatures and figures and the writing and editing of manuscript.

ACKNOWLEDGEMENTS

We are grateful to Drs. Megumi Shigematsu and Takuya Kawamura (Thomas Jefferson University) for helpful discussions.

FUNDING INFORMATION

Work in the Kirino lab on this topic is currently supported by the National Institutes of Health Grant (GM106047, HL150560, AI151641, and AI168975 to Yohei Kirino) and the American Cancer Society Research Scholar Grant (RSG-17-059-01-RMC to Yohei Kirino).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ETHICAL APPROVAL

Not applicable.

ORCID

Yohei Kirino https://orcid.org/0000-0001-5232-4742

REFERENCES

1. Shigematsu M, Kawamura T, Kirino Y. Generation of 2',3'-cyclic phosphate-containing RNAs as a hidden layer of the transcriptome. *Front Genet*. 2018;9:562.
2. Shigematsu M, Kirino Y. Making invisible RNA visible: discriminative sequencing methods for RNA molecules with specific terminal formations. *Biomolecules*. 2022;12:611.
3. Honda S, Loher P, Shigematsu M, et al. Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers. *Proc Natl Acad Sci USA*. 2015;112:E3816-25.
4. Honda S, Morichika K, Kirino Y. Selective amplification and sequencing of cyclic phosphate-containing RNAs by the cP-RNA-seq method. *Nat Protoc*. 2016;11:476-489.
5. Shigematsu M, Morichika K, Kawamura T, Honda S, Kirino Y. Genome-wide identification of short 2',3'-cyclic phosphate-containing RNAs and their regulation in aging. *PLoS Genet*. 2019;15:e1008469.
6. Pawar K, Shigematsu M, Sharbati S, Kirino Y. Infection-induced 5'-half molecules of tRNAHisGUG activate toll-like receptor 7. *PLoS Biol*. 2020;18:e3000982.
7. Shigematsu M, Kawamura T, Morichika K, et al. RNase kappa promotes robust piRNA production by generating 2',3'-cyclic phosphate-containing precursors. *Nat Commun*. 2021;12:4498.
8. Giraldez MD, Spengler RM, Etheridge A, et al. Phospho-RNA-seq: a modified small RNA-seq method that reveals circulating mRNA and lncRNA fragments as potential biomarkers in human plasma. *EMBO J*. 2019;38:e101695.
9. Chen C, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res*. 2005;33:e179.
10. Honda S, Kirino Y. Dumbbell-PCR: a method to quantify specific small RNA variants with a single nucleotide resolution at terminal sequences. *Nucleic Acids Res*. 2015;43:e77.

How to cite this article: Gumas Justin, Kirino Yohei Challenges inherent to the sequencing and quantification of short non-coding RNAs and how to address them. *Clin Transl Disc*. 2022;2:e93. https://doi.org/10.1002/ctd2.93