



Published in final edited form as:

RNA. ; 32(7): 1156–1181. doi:10.1261/rna.080861.125.

A method of comprehensive sequencing analysis of the small RNA fragmentome (RiboMarker®)

RACHEL C. CLARK^{1,2}, AIDAN C. MANNING^{1,2}, JONATHAN M. HOWARD¹, SERGIO BARBERÁN-SOLER¹, SERGEI A. KAZAKOV¹

¹RealSeq Biosciences, Inc., Santa Cruz, California 95060, USA.

²These authors contributed equally to this work

Abstract

Small RNAs and RNA fragments (sRNAs) found in blood and other biofluids have emerged as promising biomarkers for cancer and other pathologies. Sequencing analysis of sRNAs representing the entire RNA fragmentome could improve understanding of their roles in cancer development and be used for discovery of new biomarkers, cancer detection, and personalized treatment management. Conventional methods of sRNA-Seq library preparation are limited to detection of sRNAs with 5'-P and 3'-OH ends (sRNA Type 1) that represent only about 10% of the whole RNA fragmentome, whereas sRNA Types having other termini are hidden. Although recently developed sRNA-Seq methods provide detection of some or all hidden sRNAs, these methods cannot both detect all and distinguish between sRNAs of different RNA Types. Here we describe the RiboMarker® approach for preparation of sRNA sequencing libraries that addresses these shortcomings. It uses distinctive enzymatic pretreatment(s) providing conversion between the RNA termini that can enrich for or deplete specific sRNA Type(s) upfront of the universal method of library preparation. To monitor the efficacy of these pretreatments, we leveraged a pool of synthetic RNAs of different RNA Types and lengths spiked in brain or plasma RNA samples. This allows identification and analysis of relative abundance, sequencing profiles, and RNA Types of naturally occurring sRNAs representing different RNA classes. Using the RiboMarker® approach, we demonstrated its capability to enhance the capacity and sensitivity of distinguishing between plasma RNA samples from different donors (simulating healthy individuals and cancer patients) by selecting and using sRNAs of specific RNA Type(s).

Keywords

small RNA; RNA fragments; cell-free RNA; RNA circularization; RNA ends conversion; sequencing library preparation

Creative Commons License This article, published in RNA, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Corresponding author: skazakov@realseqbiosciences.com.

INTRODUCTION

Cell-free RNAs (cfRNAs) are emerging as a source for potent diagnostic, prognostic, and predictive biomarkers that can provide higher sensitivity and specificity for cancer detection compared to cell-free DNA that are currently considered as benchmark cancer biomarkers (Fish et al. 2018; Pardini et al. 2019; Ning et al. 2023; Tao et al. 2023). Tumor-associated cfRNAs (ctRNA) are released from malignant (both dead and alive) and normal cells surrounding the tumor and providing immune response to cancer (Cabús et al. 2022; Liu et al. 2023). The ctRNAs are transcribed in multiple copies from various genomic DNA regions including the ones that are only transcriptionally active in cancer cells (Larson et al. 2021; Vibert et al. 2022). Dysregulated RNA post-transcriptional events (including alternative splicing and formation of chimeric RNAs) that are detectable only in the transcriptome also contribute to the higher complexity of the ctRNA landscape (Cabús et al. 2022; Ning et al. 2023; Wang et al. 2024).

Most of the cfRNA molecules (~95%) found in circulation (e.g., in human plasma) are small RNAs and RNA fragments of <45 nucleotides (nt) in length, hereinafter together referred to as sRNA (Akat et al. 2019; Galvanin et al. 2019; Giraldez et al. 2019; Wang et al. 2024). sRNAs are processed from the longer, parent RNA transcripts of various RNA classes by one or more intracellular and/or extracellular ribonucleases (RNases) through one or more RNA cuts that can be site-specific or semi-random (Shi et al. 2022; Shigematsu and Kirino 2022; Chen and Zhou 2023; Lai et al. 2023; Tosar et al. 2024). Protective RNA-protein complexes, encapsulation into naturally occurring lipid extracellular vesicles (EVs), and/or RNA secondary structures help prevent further cleavage of processed sRNAs released from cells into circulation (Crocker et al. 2022; Shi et al. 2022; De Sota et al. 2024; Tosar et al. 2024). Most extracellular sRNAs are found outside EVs (Tosar et al. 2020; Jia et al. 2021). Furthermore, RNA molecules can undergo spontaneous cleavage *in vitro* during preparation, isolation, storage, handling and enzymatic reactions (AbouHaidar and Ivanov 1999; Chheda et al. 2024). Depending on the mechanisms of the parent RNA cleavage and naturally occurring parent RNA termini, the processed sRNAs may feature specific combinations of RNA 5' and 3' ends (hereafter called RNA Types) having different phosphorylation states (Fig. 1).

Most known sRNA sequences are assigned to the main RNA classes including ribosomal RNA (rRNA), transfer RNA (tRNA), microRNA (miRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), messenger RNA (mRNA) (including exons and UTRs), long non-coding RNA (lncRNA) and Piwi interacting RNA (piRNA) (Vickers et al. 2015; Hulstaert et al. 2020). Until recently, the biological analyses and clinical applications of sRNAs have been primarily focused on miRNA (Wang et al. 2018; Pardini et al. 2019; Wu et al. 2021). However, there are other sRNA classes that harbor similar or greater potential than miRNAs to influence and assess the state of a disease (Vickers et al. 2015; Akat et al. 2019; Yao et al. 2020; Cabús et al. 2022; Liu et al. 2022). For example, multiple overlapping sequences of sRNAs processed from mRNA and lncRNA contain information about biological phenotypes that inform about tissues of cancer origin and cancer subtypes (Akat et al. 2019; Giraldez et al. 2019; Larson et al. 2021; Chen et al. 2022; De Sota et al. 2024). Also, sRNAs with unique (defined) sequences processed from tRNAs (Wang et al.

2022; Chen and Zhou 2023; Di Fazio and Gullerova 2023), snRNAs (Pardini et al. 2019; Su et al. 2022), and snoRNAs (Gao et al. 2024) have been identified as potential cancer biomarkers.

Standard methods for sRNA-Seq library preparation capture only RNA Type 1 [5'-P and 3'-OH] including miRNAs and some other sRNAs having the same ends (Kawaji et al. 2008; Jackowiak et al. 2017; Shi et al. 2021b; Max et al. 2018; Wang et al. 2024), whereas RNA Type 2 [5'-OH and 3'-OH], RNA Type 3 [5'-OH and 3'-P] and RNA Type 4 [5'-P and 3'-P] are not incorporated into sequencing libraries (Crocker et al. 2022; Shi et al. 2022; Shigematsu and Kirino 2022). However, the sRNAs with 3'-OH ends (RNA Types 1 and 2) account for only ~10% of the human cellular RNA fragmentome, while sRNAs having 3'-phosphorylated ends (RNA Types 3 and 4) are “hidden” and cannot be detected by these methods (Lai et al. 2023). Moreover, some sRNAs (e.g., tRNA fragments) contain modified 5'-end (e.g., 5' cap) or internal nucleotide(s) that can interfere with reverse transcription, which prevents or allows only partial incorporation of these sRNA sequences into libraries (Crocker et al. 2022; Shi et al. 2022; Verwilt et al. 2023; Wang et al. 2024).

Current methods for detecting these hidden sRNAs by sequencing could be divided into two groups. The first group uses wild-type T4 polynucleotide kinase (PNK), which has both phosphatase and kinase activity in the presence of ATP, for treating sRNAs to erase differences between the phosphorylation states of the sRNA ends by converting them to the RNA Type 1 form (Akat et al. 2019; Galvanin et al. 2019; Giraldez et al. 2019; Crocker et al. 2022; Shi et al. 2022; Liu et al. 2022; Shigematsu and Kirino 2022; Solaguren-Beascoa et al. 2023). Although this approach allows simultaneous analysis of all RNA Types and RNA classes, the generated sRNA sequencing libraries contain very high proportions of rRNA fragments that overshadow the other RNA classes. While deep sequencing and/or commonly used rRNA depletion prior to sequencing library preparation could assist in the identification of low abundant sRNAs representing other RNA classes, such approaches significantly decrease the sensitivity and precision of their detection. The second group focuses exclusively on sRNAs with specific RNA ends such as 5'-OH; 3'-OH; 3'-P or 2',3'>P and cannot distinguish between the phosphorylation state at the opposite 5' or 3' end (Peach et al. 2015; Honda et al. 2016; Jia et al. 2021; Kugelberg et al. 2021; Crocker et al. 2022; Del Piano et al. 2022; Shi et al. 2022; Shigematsu and Kirino 2022; Lai et al. 2023). Except for RNA Type 1, none of these library preparation methods can specifically detect and compare individual sRNAs of Types 2, 3, and 4, representing the absolute majority of sRNAs as indicated above. Moreover, these known methods cannot detect all sRNA Types simultaneously and distinguish each individual Type specifically while using the same protocol for the preparation of sequencing libraries. Because of these method-specific biases (Zhuang et al. 2012a; Wright et al. 2019; Shi et al. 2021a), relative abundances of the different sRNA Types determined by different methods of library preparation cannot be accurately compared. Furthermore, all these reported methods require purification of intermediate reaction products, e.g. by gel-electrophoresis, phenol-chloroform or TRIzol extraction and ethanol precipitation (Akat et al. 2019; Giraldez et al. 2019; Kugelberg et al. 2021; Shi et al. 2021a; Del Piano et al. 2022; Lai et al. 2023; Solaguren-Beascoa et al. 2023), which leads to unavoidable loss of RNA during purification, limiting sensitivity and reproducibility of sRNA quantification by sequencing.

To address these shortcomings, we developed the RiboMarker[®] platform that enables comprehensive profiling of total sRNA content in biological samples, including the detection of all RNA Types simultaneously and discrimination of the individual Types. RiboMarker[®] uses distinctive enzymatic pretreatment(s) of RNA samples, determining the sRNA Type(s) to be sequenced upfront of a universal (the same for all different pretreatment protocols) library preparation protocol. It has the potential capability to both identify and then detect with enhanced sensitivity low abundance sRNAs (e.g., biomarkers for minimal residual disease or cancer at early stages).

RESULTS AND DISCUSSIONS

Design of RiboMarker[®] platform

The RiboMarker[®] platform has two key components. The first key component is a customizable group of RNA enzymatic pretreatment(s) that enriches for the specific sRNA Type(s) of interest while depleting other Type(s) in sequencing libraries. The RNA Type(s)-specific pretreatment protocols are schematically presented in Table 1. We applied column-based purification after each pretreatment step to concentrate reaction products, to exchange buffers, and to remove undesirable enzymes, which may interfere with the next reaction step.

Protocol T[1+2] provides the detection of RNA Types 1 and 2 simultaneously. This protocol is also used as a universal library preparation method in combination with all the other pretreatment steps described below.

Protocol T[2] has been designed to enrich for and detect RNA Type 2 with a single pretreatment by a mix of T4 RNA ligase 1 (Rnl1) and T4 RNA ligase 2 (Rnl2). This pretreatment results in the circularization of RNA Type 1 by ligating their 5'-P and 3'-OH ends and preventing incorporation of such RNA into the sequencing libraries.

Protocol T[1+2+3+4] has been designed to detect all RNA Types (Types 1, 2, 3, and 4) with a single pretreatment using T4 polynucleotide kinase (PNK) at pH 6.0 in the absence of ATP. Such pretreatment results in specific dephosphorylation of 3'-P or 2',3'-cyclic phosphate (2',3'>P) ends in RNA Types 3 and 4 without affecting the 5' ends of RNA Types 1 and 2.

Protocol T[3+4] has been designed to enrich for and detect RNA Types 3 and 4 with two sequential pretreatment steps, including: (i) simultaneously run in one mixture reaction with PNK mutant having no 3'-end phosphatase activity (PNK, 3'-minus) in the presence of ATP at pH 7.6 to phosphorylate 5' ends of RNA Types 2 and 3, converting them to RNA Types 1 and 4, respectively, and with Rnl1+Rnl2 mix to circularize RNA Type 1 excluding them from the sequencing library; and (ii) by PNK at pH 6.0 in the absence of ATP to dephosphorylate 3' ends of former RNA Type 3 (converted to RNA Type 4) and original RNA Type 4 converting both of them to RNA Type 1.

Protocol T[3] has been designed to enrich for and detect RNA Type 3 with three sequential treatment steps, including: (i) by Terminator 5'-Phosphate-dependent exonuclease, which specifically digests RNA Types 1 and 4 having 5'-P ends excluding them from sequencing libraries; (ii) simultaneously run in one mixture reaction with PNK, 3'-minus in the presence

of ATP at pH 7.6 to convert RNA Types 2 and 3, to RNA Types 1 and 4, respectively and Rnl1+Rnl2 mix to circularize RNA Type 1 excluding them from the sequencing library; and (iii) by PNK at pH 6.0 in the absence of ATP to dephosphorylate 3' ends of former RNA Type 3 (converted to RNA Type 4) converting these to RNA Type 1.

Protocol T[4] has been designed to enrich for and detect RNA Type 4 with three sequential pretreatment steps, including: (i) by RtcB ligase to circularize RNA Type 3 through ligation of its 5'-OH and 3'-phosphorylated ends and preventing their incorporation in the sequencing libraries; (ii) simultaneously run in one mixture reaction with PNK, 3'-minus in the presence of ATP at pH 7.6 to phosphorylate 5' ends of RNA Type 2, and Rnl1+Rnl2 mix to circularize RNA Type 1 and former RNA Type 2 converted to RNA Type 1; and (iii) by PNK at pH 6.0 in the absence of ATP to dephosphorylate of 3' ends of RNA Type 4.

The second key component of the RiboMarker[®] platform is the universal library preparation protocol T[1+2], compatible with all the pretreatment steps described above. This protocol is based on the original RealSeq platform technology for preparation of small RNA sequencing libraries (Barberán-Soler et al. 2018) that is commercially available as RealSeq[®]-Biofluids. The protocol T[1+2] workflow is schematically shown in supplementary Fig. S1, and its experimental details are described in MATERIALS and METHODS. A recent systematic evaluation of commercially available small RNA-Seq library preparation kits selected the RealSeq platform for further development of the small RNA sequencing pipeline for absolute quantitation based on the sensitivity, consistency of detection of miRNAs, low bias towards different miRNAs, as well as usability criteria (Khamina et al. 2022). This protocol can simultaneously capture RNA Types 1 and 2 (Shigematsu and Kirino 2022) which allows detection of a higher variety of specific small RNAs than the standard methods of small RNA-Seq library preparation (specific to RNA Type 1). The RiboMarker[®] approach could also be combined with these standard methods using alternative pretreatment protocols suggested in supplementary Table S1.

RNAs utilized for RiboMarker[®] development

In this study, we utilized synthetic RNAs and naturally occurring RNAs isolated from commercially available human brain and human plasma samples (see Materials and Methods).

To optimize and monitor the efficacy of conversion between the different RNA Types, we leveraged a pool of synthetic spike-in sRNAs corresponding to specific Types of sRNAs (stsRNAs). The stsRNA sequences are shown in Table 2. The stsRNA features are: (i) different lengths (20, 30, 40, 50, and 60 nt), (ii) 6-nt internal RNA Type-specific bar-codes, (iii) 4-nt randomized nucleotides at both ends, and (iv) different termini corresponding to sRNA Types 1 through 4 shown in Fig. 1. The core sequences of these stsRNAs (excluding the random and barcode sequences), which do not have homology to the human genome and do not form strong secondary structures, were adapted from Locati et al. (2015). A placement of four randomized nucleotides on both the 5' and 3' ends of stsRNAs were adapted from Lutzmayr et al. (2017). Equimolar pools of these 20 (4 Types x 5 lengths) individual stsRNAs were spiked in brain or plasma RNA samples at different inputs before pretreatments and/or preparation of sequencing libraries. We determined that the optimal

inputs of the stsRNA pool were from 1 to 5% sequencing reads relative to naturally occurring RNA samples, which allowed robust identification and quantification of stsRNAs along with the analysis of naturally occurring sRNA sequencing profiles. The developed pretreatment protocols result in stsRNA conversion between different RNA Types without shortening stsRNA lengths or causing their fragmentation.

For naturally occurring RNAs, we used a high input (50 ng) of brain total RNA and low input of total plasma RNA equivalent to 83 μ L of plasma volume (see more details in MATERIALS and METHODS). The brain and plasma RNA samples spiked with the stsRNAs were utilized for different aspects of the development of the RiboMarker[®] approach. For all tested conditions, technical triplicate sequencing libraries were made with the brain RNA. The amount of plasma RNA was selected based on the minimum volume of available plasma from individual donors that allowed the use of equal plasma RNA inputs across experiments. We used plasma samples from 10 different individuals that were grouped into four cohorts (A, B, C and D) based on the plasma source vendor and known pathology (supplementary Table S2). For the A, B and D cohorts, the plasma RNA samples from the three individuals of each cohort were considered biological replicates. A single plasma sample from the C cohort was available in a large enough volume to be used as technical replicates. Most figures present data for representative RNA samples displaying average sequencing profiles from the replicates shown in supplementary figures.

Implementation of RNA circularization in the RiboMarker[®] platform

The RiboMarker[®] approach exploits circularization reactions catalyzed by an RNA ligase for two different purposes. First, it is applied in the sequencing library preparation protocol T[1+2] (supplementary Fig. S1), where sRNAs are first ligated to the sequencing combo adapter (CAD) and then sRNA-CAD ligation products are circularized. In contrast to similar methods of sequencing library preparations based on circularization of cDNAs (Kwon 2011; Jackson et al. 2014; Heyer et al. 2015), our approach allows sequencing of full-length sRNAs rather than their RT cDNA products. The RNA circularization steps are also used in the pretreatment protocols (Table 1), where sRNAs are circularized to block the sRNA ends preventing adapter ligation and incorporation of “unwanted” RNA into the sequencing libraries.

Although rates of RNA intermolecular ligation reactions catalyzed by Rnl1 are strongly dependent on the identity of terminal nucleosides at 5'-P and 3'-OH ends involved in these reactions (England and Uhlenbeck 1978; Ohtsuka et al. 1980; Romaniuk et al. 1982), the RNA circularization (intramolecular ligation) reactions exhibit less bias to the identity of terminal nucleosides and a presence of ribose 2'-O-Methyl modification (2'-OMe) at the RNA 3'-end, depending mostly on RNA length and folding (Silber et al. 1972; Cranston et al. 1974; Kaufmann et al. 1974; Kumar et al. 2011; Del Piano et al. 2022). Although the generation of concatenated ligation products through intermolecular reactions may also occur concurrently with RNA circularization (Hafner et al. 2011; Chu et al. 2015), the circularization reaction greatly predominates for RNAs shorter than 60 nt (Uhlenbeck and Gumport 1982; Harrison and Zimmerman 1984) and can be additionally promoted by dilution of the ligation reaction mixtures (Obi and Chen 2021). Similar techniques were

adapted for circularization-based methods of sRNA-seq library preparation (Chu et al. 2015; Gout et al. 2017; Barberán-Soler et al. 2018; Bradley et al. 2024). In comparison to the intermolecular ligation of adapters to 5'-P ends of sRNAs in dual-adapter ligation methods, which has significant biases (Hafner et al. 2011; McCormick et al. 2011), the single adapter and circularization-based approach used in the RealSeq platform (supplementary Fig. S1) significantly reduces this bias (Barberán-Soler et al. 2018).

Evaluation of circularization-based RiboMarker® sRNA sequencing library preparation method

We started development of the RiboMarker® platform with an assessment of the universal protocol T[1+2] and its capability to capture stsRNAs of different lengths by analyzing sequencing length profiles (sequencing read count vs length of sequences) while the stsRNA Types were identified by sequences of stsRNA barcodes.

Separate stsRNA libraries were prepared by protocol T[1+2] for stsRNAs in the absence of naturally occurring RNAs (mock controls) (Fig. 2A, left panel and supplementary Fig. S2A) or for stsRNAs spiked into total RNA isolated from either brain (Fig. 2B, left panel) or representative A and B cohorts' plasma (Figs. 2C-D, left panels) RNA samples. Our data demonstrated that protocol T[1+2] captured similar percentages of stsRNA Type 1 and Type 2, whereas only marginal percentages of stsRNA Type 3 and Type 4 were included in the sequencing libraries (Figs. 2A-D, left panels and supplementary Table S3). Triplicate sequencing length profiles demonstrated no significant variations between the ratios of the RNA Types for the stsRNAs of the same lengths when the stsRNAs were spiked into either brain (supplementary Fig. S3A) or into A or B cohorts of plasma RNA samples (supplementary Fig. S4). However, we noticed two phenomena related to differences of the sRNA profiles among the tested samples. The first phenomenon was small differing average ratios of stsRNA Type 1 and Type 2 of all lengths in the mock controls (Type 2 ~ Type 1) or when they were spiked into total RNA from brain (Type 1 > Type 2) and plasma (Type 2 > Type 1) samples (Figs. 2A-D, left panels and supplementary Table S3). In principle, a competition between naturally occurring sRNAs and synthetic stsRNAs of the same RNA Types present at different molecular ratios in brain and plasma samples could explain the Types ratio phenomenon. The second phenomenon was related to the differential capture of the stsRNAs of different lengths spiked into brain ($20 \geq 30 > 40 > 50 \gg 60$ nt) and plasma ($40 > 20 \sim 30 > 50 > 60$ nt) RNA samples (Figs. 2B-D, left panels). Although the differential capture of stsRNA lengths was minimal between the mock control and brain samples, the differences of stsRNA lengths between brain and plasma RNA samples were more pronounced. We hypothesized that opportunistic folding and co-folding of stsRNAs having randomized ends with natural sRNAs, which is length, structure, and concentration dependent, may affect enzymatic reactions of RNA molecules at their ends as previously proposed (Zhuang et al. 2012a; Maguire et al. 2020). Regardless of the mechanism of these phenomena, protocol T[1+2] provides the capability for robust quantification of sRNAs ≤ 50 nt in human plasma samples, where 95% of extracellular sRNAs are less than 45 nt in length (Akat et al. 2019; Galvanin et al. 2019; Giraldez et al. 2019; Wang et al. 2024).

Circularization of small RNAs prevents their incorporation into sequencing libraries

sRNA circularization has been previously used for the detection of sRNAs by RT-qPCR (Kumar et al. 2011) and sequencing (Chu et al. 2015; Gout et al. 2017; Barberán-Soler et al. 2018; Bradley et al. 2024). In this study, we applied it to exclude selected sRNA Types (Fig. 1) from sequencing libraries. Following the circularization of sRNA Type 1, the termini of selected non-circularized sRNA Types are converted into Types 1 and/or 2 in the next pretreatment step(s), and then these converted sRNAs are either also excluded by circularization or incorporated into sequencing libraries using the universal protocol. In doing so, the exclusion of highly abundant sRNA Types from the libraries allows for a more sensitive detection of rare sRNA of other Type(s). This new exclusion-by-circularization approach could also be used along with other methods of sequencing library preparation that rely on the ligation of adapter(s) to RNA end(s) or on 3'-end extension by a ribonucleotidyl transferase (Crocker et al. 2022; Shi et al. 2022; Shigematsu and Kirino 2022; Wang et al. 2024).

Protocol T[2] features a circularization step in which the abundance of RNA Type 1 is decreased while RNA Type 2 is increased. To find the most efficient conditions for this circularization step, we used the stsRNA pool spiked into brain RNA samples at 37°C in standard T4 RNA ligase reaction buffer supplemented with 10% PEG 8000, 20 U Murine RNase Inhibitor (all NEB) and 100 µM ATP with the following variables: 10 U Rnl1 for 1 hour (Rnl1_1h); 20 U Rnl1 for 1 hour (2xRnl1_1h); 20 U Rnl1 for 2 hours (2xRnl1_2h); and 10 U Rnl1 + 10 U Rnl2 for 1 hour (Rnl1+Rnl2_1hr). The Rnl1+Rnl2 combination was tested because of the ability of both Rnl1 and Rnl2 to circularize RNA Type 1 (Ho and Shuman 2002; Ho et al. 2004; Yin et al. 2004), and a potential synergy between these two enzymes having different structural and end-nucleotide biases for ligation of co-folded RNA ends (Yoshinari et al. 2017; Chen et al. 2020). Sequencing analysis of these different versions of protocol T[2] for the efficiency of RNA Type 2 enrichment revealed that the “Rnl1+Rnl2_1h” reaction condition provided 90.91±0.19% yield of stsRNA Type 2 in the brain RNA samples, and 93.50±0.51% in the A cohort and 93.90±0.24% in the B cohort plasma RNA samples (supplementary Table S4). The corresponding stsRNA length profiles demonstrated high reproducibility for the brain (supplementary Fig. S3B) and A and B cohorts' plasma (supplementary Fig. S5) RNA samples. The mock controls performed in the absence of naturally occurring RNA resulted in a 91.53±0.23% yield of stsRNA Type 2 (supplementary Tables S4 and S5).

Although two other circularization reaction conditions “Rnl1_1h” and “2xRnl1_1h” provided only marginally smaller yields (90.59±0.28% and 89.71±0.02%, respectively) of stsRNA Type 2 in the brain RNA samples in comparison to the selected condition, these conditions also produced higher percentages of Rnl1 reaction side products (stsRNA Types 3 and 4) than the “Rnl1+Rnl2_1h” condition (supplementary Table S4 and supplementary Fig. S6). The most prominent amounts of stsRNA Type 3 products (8.25%±0.03%) were produced under the “2xRnl1_2h” conditions (supplementary Table S4). Evidently, the stsRNA Types 3 and 4 were products of the known Rnl1 reverse reaction causing removal of single mononucleotide diphosphates (5'-P-N-3'-P or 5'-P-N-2',3'>P) from the 3'-end-phosphorylated RNA molecules (Krug and Uhlenbeck 1982; Uhlenbeck and Gumpert

1982). The sequencing length profiles confirmed that this side reaction indeed removes a single terminal nucleotide from the 3' ends of stsRNA Types 3 and 4 that makes them shorter by 1 nt and converts them to stsRNA Types 2 and 1, respectively, allowing their incorporation into sequencing libraries (supplementary Fig. S6). The higher percentage of sequencing reads for stsRNA Type 3 (converted by Rnl1 to a 1 nt shorter Type 2) in comparison to stsRNA Type 4 (converted to a 1 nt shorter Type 1) is because stsRNA Type 1 can then be circularized by Rnl1 and, therefore, excluded from the library.

Based on the above results, we selected the “Rnl1+Rnl2_1h” reaction condition to be used for the RNA Type 1 circularization in protocols T[2], T[3+4], T[3] and T[4]. In contrast to protocol T[1+2], the stsRNA sequencing length profiles for protocol T[2] were similar across the tested samples, including the mock controls of stsRNAs performed in the absence of naturally occurring RNA (Fig. 2A, right panel and supplementary Fig. S2B), brain (Fig. 2B, right panel) and representative A and B cohorts' plasma RNA samples (Figs. 2C-D, right panels). The average percentage of stsRNA Type 2 in the A cohort plasma RNA samples were found to slightly increase with increasing stsRNA length from 88.48±1.32% (20-nt) to 92.78±0.43% (30-nt) and 96.42±0.22% (40-nt), held at 95.21±0.44% (50-nt), and then dropped to 86.39±1.00% (60-nt) (supplementary Table S7). By comparing the percentage of stsRNA Type 1 reads of different lengths in the A cohort plasma RNA samples using protocols T[1+2] (supplementary Table S6) and T[2] (supplementary Table S7), we were able to evaluate both the extent of stsRNA Type 1 reduction from sequencing libraries and the corresponding efficiency of their circularization. We calculated the reduction for each stsRNA length for the A cohort plasma RNA samples by subtracting the corresponding stsRNA Type 1 percentages found with protocol T[2] from those with protocol T[1+2] and then dividing the differences by the Type 1 percentages for protocol T[1+2] (supplementary Table S8). The found reductions of stsRNA Type 1 of different lengths were: 79.85% (20-nt), 83.30% (30-nt), 89.05% (40-nt), 87.90% (50-nt), and 74.02% (60-nt). These results indicated that we were able to achieve high efficiency of stsRNA Type 1 reduction through circularization of stsRNA regardless of their lengths and termini sequences.

The efficiency of the sRNA circularization reaction could be further increased by optimizing the reaction buffer composition since Rnl1 and Rnl2 have different optimal concentrations of Mg²⁺ and reaction buffer pH (Ho et al. 2004; Yin et al. 2004; Viollet et al. 2011). Also, adding dimethyl sulfoxide (DMSO), a mild denaturant that disrupts secondary structures limiting the accessibility of RNA ends for the ligase, could enhance RNA circularization efficiency, similar to the DMSO effect on intermolecular RNA ligation (Romaniuk and Uhlenbeck 1983). Alternatively, thermostable RNA ligases like CircLigase could be used for circularizing heat-denatured RNA molecules at 60°C (Kumar et al. 2011). However, under this high-temperature reaction unfolding RNA structures can also lead to unintended sRNA fragmentation.

Detection of naturally occurring RNA Types 1 and 2

After our assessment of the synthetic stsRNAs, we analyzed sequencing length profiles for the major RNA classes of naturally occurring sRNAs in the same brain and plasma RNA samples processed using protocols T[1+2] and T[2]. For the brain RNA samples,

a comparison of the sRNA profiles showing all major RNA classes simultaneously for protocol T[1+2] (supplementary Fig. S7A, left panel) and protocol T[2] (supplementary Fig. S7A, right panel) confirmed the efficiency of depletion of sRNA Type 1 as shown by the reduction of the miRNA fraction in the protocol T[2] profiles. A subtraction of the normalized sequencing length profiles for protocol T[2], which is specific for RNA Type 2, from the profile for protocol T[1+2], which is specific for RNA Types 1 and 2, enables the identification of sequencing length profiles for the depleted sRNA Type 1. Using this approach, we found that most of the sRNA Type 1 in the brain sample were represented by miRNAs as well as fragments of tRNAs and rRNAs (supplementary Fig. S8). In comparison to the brain sRNA (supplementary Fig. S7A), the sequencing length profiles for representative A and B cohorts' plasma RNA samples were comprised of a larger fraction of ultrashort sequences (<20 nt) for both protocols T[1+2] and T[2] (supplementary Figs. S7B-C). Except for the depletion of miRNA sequences in the brain and plasma sRNA profiles for protocol T[2], there were no substantial differences between profiles for protocols T[1+2] and T[2] (supplementary Fig. S7) indicating that most sRNAs detected by protocol T[1+2] had RNA Type 2 ends. For protocol T[1+2], we also noticed significant biological variations of sRNA sequencing length profiles showing all major RNA classes simultaneously among the different A and B cohorts' plasma RNA samples, especially in the tRNA and miRNA classes (supplementary Fig. S9).

In contrast to sequencing length profiles for sRNA Type 1 for cellular (Kawaji et al. 2008; Jackowiak et al. 2017; Shi et al. 2021b) and plasma (Max et al. 2018; Wang et al. 2024) RNA samples, the profiles for naturally occurring sRNA Type 2 have not been previously described. Here we present the sequencing profiles for sRNA Type 2 of all major sRNA classes detected in brain and representative A and B cohorts' plasma RNA samples by protocol T[2] (Fig. 3). While length profiles for tRNA and mRNA-derived sRNAs were relatively consistent across the plasma samples (Figs. 3B and 3E), the profiles for snRNA, snoRNA, and lncRNA-derived sRNAs showed distinctive longer RNA fragments in the representative plasma sample from the B cohort than the A cohort (Figs. 3C, 3D, and 3F, respectively).

Furthermore, protocol T[2] revealed intriguing features of piRNA-related sequences. Like miRNAs, mature (processed) piRNAs could be classified as RNA Type 1; however, these contain a 2'-OMe modification at their 3'-end. It has been previously shown that this modification could reduce the circularization rate of sRNAs having 5'-P and 2'-OMe/3'-OH ends by Rnl1 (Kumar et al. 2011). For the brain RNA samples, we determined that protocol T[2] reduced the percentage of the miRNA-related reads in comparison to protocol T[1+2] by 11.24 ± 2.80 fold, whereas percentages of piRNA-related reads were reduced by only 1.30 ± 0.05 fold (supplementary Tables S9A-B). In similar comparisons of these protocols applied to the A cohort plasma samples, the observed reduction for miRNA- and piRNA-related reads were 27.38 ± 7.50 and 1.17 ± 0.05 fold, respectively (supplementary Table S9C-D). These results confirmed that the 2'-OMe modification substantially inhibited circularization of piRNA-related sRNAs and, therefore, reduced their depletion from sequencing libraries prepared by protocol T[2].

Detection of all four stsRNA Types after PNK pretreatment

PNK is commonly used to erase the differences between the phosphorylation states of sRNA ends by converting them simultaneously to RNA Type 1, followed by standard sRNA-Seq library preparation methods (Akat et al. 2019; Giraldez et al. 2019; Crocker et al. 2022; Shi et al. 2022, Shigematsu and Kirino 2022; Solaguren-Beascoa et al. 2023). This approach, also known as Phospho-RNA-Seq, uses the standard, one-step PNK treatment of RNA performed in Tris-HCl buffer at pH 7.6 in the presence of 1 mM ATP under conditions adopted from earlier studies (Pan et al. 1991; Busch et al. 2000; Schürer et al. 2002; Brown and Bevilacqua 2005). Although these reaction conditions are optimal for the phosphorylation of 5'-OH ends of RNA Types 2 and 3 (Cameron and Uhlenbeck 1977), the maximum rate of dephosphorylation of 3'-P and 2', 3'>P ends of RNA Types 3 and 4 is ~pH 5.9, whereas at pH 7.6 it decreases by ~10-fold (Cameron and Uhlenbeck 1977; Zhuang et al. 2012b). Also, PNK has significant 5'-P reverse (dephosphorylation and exchange) activities catalyzed by ATP (van de Sande et al. 1973). Additional anticipated problems include known bias of PNK to preferentially phosphorylate nucleic acids with certain 5'- and 3'-end nucleotides, sequences, lengths, and secondary structures (Székely and Sanger 1969; Lillehaug and Kleppe, 1975; van Houten et al. 1998; Lee et al. 2013).

In this study, we sought to enhance desirable and minimize unwanted PNK activities through an examination of the effects of different PNK pretreatment conditions on the sequencing output of our pool of stsRNA spiked into brain RNA samples. In contrast to the Phospho-RNA-Seq method, we aimed to perform the 3'-end dephosphorylation and 5'-end phosphorylation by PNK in two separate steps (rather than simultaneously) under the optimal conditions for each step. In the first step, 3'-dephosphorylation was performed as an RNA pretreatment as described in protocol T[1+2+3+4] (Table 1). For this purpose, we tested three buffer conditions adapted from Cameron and Uhlenbeck (1977), comprising 100 mM Tris-HCl at pH 7.6 in the presence of 1 mM ATP; and 100 mM Imidazole-HCl at pH 6.5 or 100 mM MES-NaOH at pH 6.0 (both in the absence of ATP) as described in MATERIALS and METHODS. In the second step, the 5'-phosphorylation of RNA Types 2 and 3 was carried out in Step 2b of the library preparation by PNK in standard NEB kinase reaction buffer at pH 7.6 in the presence of 100 μ M ATP which minimized the 5'-end dephosphorylation of RNA Types 1 and 4 (supplementary Fig. S1A). In contrast to Cameron and Uhlenbeck (1977) demonstrating that the rate of 3'-P dephosphorylation depends on the buffer pH, we did not observe significant differences among average percentages of stsRNA Types 3 and 4 in brain RNA samples pretreated under the described above conditions (supplementary Table S10).

Despite the different stsRNA Types having been premixed at an equimolar ratio of 25% per Type, we found that protocol T[1+2+3+4] applied to stsRNAs spiked into brain or plasma RNA samples did not produce the expected equal distribution for the stsRNA Types, with stsRNA Types 1 and 2 overrepresented and Type 4 slightly underrepresented, whereas Type 3 was substantially underrepresented (supplementary Table S10). Since stsRNA Types 3 and 4 both have 3'-P ends and were less represented than stsRNA Types 1 and 2 having 3'-OH ends, we hypothesized that the de-phosphorylation of 3'-P ends by commonly used 10 Units PNK (per reaction) might be incomplete. Therefore, we compared the outputs of

reactions containing either 10 or 20 Units of PNK for stsRNAs spiked into the C1 plasma RNA sample and found that the 2-fold increase of PNK input did not enhance the percentage of stsRNA Type 3 and 4 in the sequencing libraries (supplementary Table S10). Because the stsRNA Type 3 underrepresentation was also found with the mock stsRNA experiments under the selected PNK pretreatment condition while the other three stsRNA Types were more evenly represented (supplementary Table S5), we considered it as a bias related to the Type 3 stsRNAs negatively affecting their incorporation in sequencing libraries.

Although the PNK pretreatment with MES buffer at pH 6.0 did not produce higher yields of stsRNA Types 3 and 4 than using Tris-HCl at pH 7.6 (supplementary Table S10), we have selected it for 3'-P dephosphorylation in protocol T[1+2+3+4]. The rationale behind this selection is based on known factors leading to RNA fragmentation. To avoid an interference of PNK with the next enzymatic step, PNK should be either heat inactivated or removed by column purification prior to sequencing library preparation. Although the heat inactivation of PNK (and other thermolabile enzymes) is a common technique used in RNA biochemistry, we were concerned about additional sRNA fragmentation. As previously established, the heating of RNA molecules in select buffers can lead to their fragmentation, which is heavily influenced by RNA length, sequence, presence of divalent cations, buffer composition and pH (Chheda et al. 2024). In the MES buffer at pH 6.0, RNA molecules were expected to be more resistant to cleavage than in Tris buffer at pH 7.6 (Chheda et al. 2024). Also, both Imidazole and Tris in the presence of Mg^{2+} cations can catalyze RNA cleavage accelerated by high temperature (Breslow and Huang 1991; Vlassov et al. 1995; AbouHaidar and Ivanov 1999).

To test whether the column-based removal of PNK or its heat inactivation in solution affected the stsRNA sequencing length profiles, we performed two sets of experiments using the C1 plasma RNA sample. The first experiment was to assess the effect of passing RNA through a column on the distribution of RNA lengths sequenced. We compared sequencing profiles of stsRNAs spiked into the plasma RNA samples passed through the Zymo RNA clean and concentrate columns and stsRNAs spiked into plasma RNA samples that were directly used as input for library preparation without a column clean-up for universal protocol T[1+2] (supplementary Fig. S10). We found that passing the sample through the column (supplementary Fig. S10B) did not significantly alter the resulting stsRNA length profiles as compared to the samples not passed through the columns (supplementary Fig. S10A). This result, which is universally applicable to all RiboMarker[®] pretreatment steps (Table 1), demonstrated that the column purification did not noticeably alter the stsRNA length distribution. In the second experiment, we assessed the effect of PNK heat inactivation after enzymatic pretreatment on the distribution of stsRNA lengths. We treated the stsRNA with PNK in the MES pH 6.0 buffer, and then PNK was either heat-inactivated at 65°C for 20 min in the presence of 20 mM EDTA pH 8.0 or removed by the columns without heating. In this case, the heat inactivation of PNK resulted in less robust capture of longer stsRNAs, with a significant increase in the relative abundance of shorter molecules (supplementary Fig. S11A) as compared to the column-purified stsRNAs (supplementary Fig. S11B). Given our previous observation that passing through the columns does not alter the inclusion ratio of different lengths and, therefore, does not artificially enrich for longer stsRNAs (supplementary Fig. S10), we concluded that the heat inactivation of PNK

leads to the partial fragmentation of our stsRNAs that evidently affects longer molecules more readily. Also, we found that PNK heat inactivation additionally reduced the average percentage of stsRNA Type 3 ($13.23 \pm 0.42\%$) of all stsRNA lengths as compared to the column-purified samples ($19.29 \pm 0.43\%$) (supplementary Table S10).

A comparison of stsRNA sequencing length profiles for the mock controls and stsRNAs spiked into the brain and representative A and B cohorts' plasma RNA samples indicates that the natural RNA compositions can significantly affect the ratios between stsRNA Types for protocol T[1+2+3+4] (Figs. 4A-E). Furthermore, the percentage of stsRNAs of different lengths were also affected. Although these phenomena are not fully understood, the putative competition and/or co-folding between naturally occurring sRNAs (whose contents can vary among different samples) and synthetic stsRNAs might be responsible for these effects.

We also compared sequencing length profiles of stsRNAs spiked into D cohort plasma RNA for protocol T[1+2+3+4] (supplementary Fig. S12A) and Phospho-RNA-Seq (supplementary Fig. S12B). For a fair comparison, we skipped the PNK heat inactivation step (which is commonly used for Phospho-RNA-Seq) but kept the column cleanup step which was common for protocol T[1+2+3+4]. We found substantial differences of the stsRNA distributions between these two protocols among the four RNA Types at different lengths. On average for all stsRNA lengths, protocol T[1+2+3+4] includes more stsRNA Type 2 ($35.15 \pm 1.51\%$) and less stsRNA Type 3 ($15.8 \pm 1.75\%$) than Phospho-RNA-Seq ($23.35 \pm 0.56\%$ and $19.02 \pm 0.32\%$, respectively), whereas Phospho-RNA-Seq includes more stsRNA Type 1 ($29.84 \pm 0.44\%$) and Type 4 ($27.79 \pm 0.49\%$) than protocol T[1+2+3+4] ($26.56 \pm 1.86\%$ and $22.49 \pm 1.52\%$, respectively) (supplementary Tables S11 and S12). Neither of these two protocols were perfect in capturing the intended 25% distribution of each stsRNA Type. There were significant variations from the average stsRNA Type distributions of different lengths for both protocol T[1+2+3+4] and Phospho-RNA-Seq (supplementary Tables S11 and S12). Furthermore, we noticed remarkable differences in efficiency of capturing stsRNAs of different lengths between these two protocols. Protocol T[1+2+3+4] provided almost even coverage of stsRNAs in the 20- to 50-nt range, with a slightly higher yield for 30-nt and a significantly lower yield for 60-nt species (Fig. 4E and supplementary Fig. S12A). In contrast, Phospho-RNA-Seq revealed a biased distribution of stsRNAs of different lengths, with a strong peak at 40-nt (10x yield) symmetrically surrounded by 30-nt and 50-nt (3x yield), and by 20-nt and 60-nt (1x yield) (Fig. 4F and supplementary Fig. S12B). The sequencing library preparation for Phospho-RNA-Seq with the NEBNext Small RNA library preparation kit for Illumina used a two-adapter ligation to RNA ends and (theoretically) should provide length-independent incorporation of the stsRNAs in the libraries. Even so, a co-folding between sRNAs and the adapters has been identified as a key determinant of ligation efficiency bias (Zhuang et al. 2012a; Maguire et al. 2020). A plausible explanation of this observed NEBNext bias would be an accidental co-folding between our stsRNAs and NEBNext sequencing adapter(s) that is either favorable or unfavorable to their ligation efficiency. However, profiling of stsRNAs to evaluate the efficiency of conversions of RNA ends represents only an approximate outline of profiling naturally occurring sRNAs that might have different abundances, lengths and non-random nucleotides at their ends than the stsRNA molecules.

Comparison of length profiles for naturally occurring RNA Types found with protocols T[1+2+3+4] and T[1+2]

After evaluating protocol T[1+2+3+4] with stsRNAs, we analyzed sequencing length profiles for naturally occurring sRNAs of all main RNA classes in the brain (supplementary Fig. S13) and plasma (A and B cohorts) (supplementary Fig. S14) RNA samples. In comparison to protocol T[1+2], the most distinctive differences were increased rRNA and decreased miRNA contents for both the brain and plasma RNA samples detected by protocol T[1+2+3+4] (supplementary Figs. S15A-B). This result was similar but less dramatic than the differences in the rRNA and miRNA content between common sRNA-Seq protocols detecting only sRNA Type 1 and Phospho-RNA-Seq detecting all four RNA Types simultaneously (Akat et al. 2019; Galvanin et al. 2019; Giraldez et al. 2019; Shi et al. 2021b; Solaguren-Beascoa et al. 2023).

In contrast to protocol T[2], which depletes RNA Type 1 (including miRNAs) from the sequencing libraries, protocol T[1+2+3+4] dilutes these reads by adding the “hidden” sRNA species of RNA Types 3 and 4 to the libraries after the pretreatment. If the 5'-P ends of miRNAs are not dephosphorylated by the PNK pretreatment, the reduction in the capture of miRNAs will be inversely proportional to the increase of new sRNA species. A comparison between these two protocols showed that the average inclusion into the resulting libraries for miRNA with protocol T[1+2+3+4] in comparison to protocol T[1+2] was reduced by 3.34 ± 0.09 fold in the brain (supplementary Fig. S15A), 8.08 ± 1.73 fold in the A cohort plasma, and 20.64 ± 9.93 fold in the B cohort plasma RNA samples (supplementary Fig. S15B). These results indicate that the “hidden” sRNA species that appeared after the PNK pretreatment by protocol T[1+2+3+4] can account for 95% of sRNA present in the B cohort plasma RNA samples. The relative abundance of sRNA across RNA classes indicates a high reproducibility for technical repeats for the brain (supplementary Fig. S15A) and significant biological variations across different plasma (supplementary Fig. S15B) RNA samples. In the brain sample, the percentage of sequencing reads for sRNAs derived from mRNA and lncRNA increased with protocol T[1+2+3+4] in comparison to protocol T[1+2], and for tRNA were about the same, whereas miRNA, snRNA and piRNA decreased with protocol T[1+2+3+4] (supplementary Fig. S15A). These differences for sRNAs derived from mRNA and lncRNA in the brain RNA sample were related to the inclusion of a bigger proportion of sRNA Types 3 and 4 than Types 1 and 2. Despite the expectation that most of the sequencing reads which align to mRNA and lncRNA found in circulation should belong to sRNA Type 3 due to the abundance of superfamily A RNases in human blood (Giraldez et al. 2019; Lee et al. 2019), we found that protocol T[1+2] detected about the same or larger fractions of these RNA classes than protocol T[1+2+3+4] in the A cohort and B cohort plasma RNA samples (supplementary Fig. S15B). This finding correlates with previous TGIRT-seq analyses, indicating that many of the mRNA fragments present in plasma have 3'-OH rather than 3'-P ends (Yao et al. 2020).

Furthermore, we analyzed the changes in the length profiles of the individual RNA classes obtained by protocols T[1+2] and T[1+2+3+4]. For the brain samples (supplementary Fig. S16), we detected comparatively long (>30 nt) fragments of snRNAs (supplementary Fig. S16D), snoRNAs (supplementary Fig. S16E), and lncRNAs (supplementary Fig. S16G)

whose relative abundances were decreased in protocol T[1+2+3+4] as compared to protocol T[1+2]. For mRNA, the abundance of both shorter and longer fragments was significantly increased by protocol T[1+2+3+4] (supplementary Fig. S16F). In the representative A cohort plasma RNA sample, most sRNA sequences that aligned to the individual RNA classes (except tRNA) were found to be <20 nt (supplementary Fig. S17). However, for the representative B cohort plasma RNA sample, we found relatively large fractions of sRNA sequences aligned to most individual RNA classes that were >20 nt for both protocols (supplementary Fig. S18). Particularly, we detected a relatively larger fraction of tRNA fragments of ≥ 30 nt for protocol T[1+2+3+4] than for protocol T[1+2] (supplementary Fig. S18C), suggesting their origin as RNA Types 3 and/or 4. In contrast, we detected a larger fraction of fragments of snRNA and snoRNA ≥ 30 nt for protocol T[1+2] than for protocol T[1+2+3+4] (supplementary Fig. S18D-E), suggesting that these fragments were RNA Types 1 and/or 2.

Comparison of length profiles for naturally occurring RNA Types found with protocols T[1+2+3+4] and Phospho-RNA-Seq

Using Phospho-RNA-Seq as a benchmark, we compared its capacity to simultaneously capture sRNAs with RiboMarker[®] protocol T[1+2+3+4]. Although the sequencing length profiles showing all RNA classes simultaneously for the D cohort plasma RNA samples appeared similar between these two protocols for each individual plasma sample (supplementary Fig. S19), the profiles of sRNA representing individual tRNA, snRNA, snoRNA and piRNA classes were found to be distinctively different as shown for a representative D cohort plasma RNA sample (supplementary Fig. S20). We then compared the percentage of sRNA sequencing reads uniquely mapped to the individual RNA classes as captured by these two protocols in the D cohort plasma RNA samples. Although both protocols have similar capacity in the detection of sRNAs mapped to rRNA, snoRNA and snRNA, protocol T[1+2+3+4] provided the detection of more sRNA sequences mapped to miRNA, piRNA and tRNA, whereas Phospho-RNA-Seq detected slightly more mRNA and lncRNA fragments (supplementary Fig. S21).

Detection of stsRNA Types 3 and 4

We reused single pretreatment steps from protocols T[2] and T[1+2+3+4] in protocols T[3+4], T[3] and T[4] which have multiple pretreatment steps (Table 1). First, we evaluated protocol T[3+4] to simultaneously detect RNA Types 3 and 4. Like protocol T[1+2] (Figs. 2A and 2D, left panels), the stsRNA sequencing length profiles for protocol T[3+4] significantly varied between the mock control (Fig. 5A, left panel and supplementary Fig. S2D) and stsRNAs spiked in a representative B cohort plasma RNA sample (Fig. 5A, right panel), while there were few differences between profiles of stsRNAs spiked in the plasma (A or B cohorts) RNA samples (supplementary Fig. S22A). For these profiles, the presence of naturally occurring sRNAs in the plasma samples provided almost even coverage of stsRNAs in the 20–50 nt range. Using protocol T[3+4] for stsRNAs spiked into the A cohort and B cohort plasma RNA samples, we successfully enriched for RNA Type 3 and Type 4, making up ~90% stsRNA sequences detected, with similar average values for RNA Type 3 (A cohort: $39.80 \pm 0.17\%$ and B cohort: $39.92 \pm 0.81\%$) and RNA Type 4 (A cohort:

49.39±0.94% and B cohort: 49.90±0.51%) (supplementary Tables S13A-B) with the ratio between stsRNA Type 3 to stsRNA Type 4 estimated as about 1 to 1.25.

Second, we evaluated protocol T[3]. The sequencing length profiles for this protocol found a steady decline of stsRNA capture in the 20–50 nt range for both the mock control (Fig. 5B, left panel and supplementary Fig. S2E) and stsRNAs spiked in the plasma (A and B cohorts) RNA samples (Fig. 5B, right panel; and supplementary Fig. S22B). In comparison to protocol T[3+4], the stsRNA sequencing profiles for protocol T[3] were less different among the mock controls and in the presence of the plasma RNA (Fig. 5B). However, the percentage of stsRNA Type 3 changed. In comparison to the mock control (53.07±1.61%; supplementary Table S5), the presence of the naturally occurring RNA significantly improved the stsRNA Type 3 specificity of this protocol for the stsRNAs spiked in the plasma RNA samples, where average percentages of the stsRNA Type 3 had similar values (76.96±1.68% and 75.09±1.42%) for the A and B cohorts' plasma RNA samples, respectively (supplementary Tables S14A-B).

Finally, we evaluated protocol T[4]. In the first pretreatment step of this protocol, we circularized sRNA Type 3 by RtcB ligase to prevent their incorporation into sequencing libraries (in a similar fashion to circularization of sRNA Type 1 by T4 RNA ligases). For this purpose, we adopted previously described RtcB ligase reaction conditions (Chakravarty et al. 2012; Petkovic and Müller 2015). As in protocol T[3+4], stsRNA sequencing length profiles for protocol T[4] varied between the mock control (Fig. 5C, left panel and supplementary Fig. S2F) and stsRNAs spiked in a representative B cohort plasma RNA sample (Fig. 5C, right panel), while there were few differences between profiles of stsRNAs spiked in the A and B cohorts' plasma RNA samples (supplementary Fig. S22C). Remarkably, both protocols T[3+4] and T[4] provided almost even coverage of stsRNA Type 4 in the 20–50 nt range in the presence of plasma RNA (A and B cohorts) (supplementary Figs. S22A and S22C). For protocol T[4], we achieved only moderate enrichment for stsRNA Type 4 over Type 3 (Fig. 5C) presumably because of the limited efficiency of RtcB on sRNA Type 3 circularization. For stsRNAs spiked into plasma RNA samples, the average ratio of RNA Type 4 to Type 3 for all stsRNA lengths were found to be 2.1 to 1 (Type 4: 58.98±1.01% vs. Type 3: 27.89±2.06%) for A cohort plasma RNA and 2.3 to 1 for stsRNAs (Type 4: 61.95±4.40% and Type 3: 26.88±4.60%) for B cohort plasma RNA (supplementary Tables S15A-B).

Detection of naturally occurring RNA Types 3 and 4

After assessing protocols T[3+4], T[3], and T[4] with the stsRNAs, we evaluated these protocols for detection of naturally occurring sRNAs in plasma RNA samples. A comparison of the percentage of normalized sequencing reads for the individual main sRNA classes in libraries of different plasma RNA samples detected was about the same for rRNA, snRNA, mRNA, and lncRNA by protocols T[3+4], T[3] and T[4], whereas sRNA Type 4 prevailed over sRNA Type 3 for tRNA and snoRNA classes (supplementary Fig. S15C). Although the differences in sRNA sequencing length profiles combining all RNA classes simultaneously were not easily notable for these three protocols in A and B cohorts' plasma RNA samples (supplementary Figs. S23A-C) because of high rRNA background, some individual RNA

classes had distinctive protocol-specific and sample-dependent features. For instance, the sRNA profiles for tRNA, snRNA, snoRNA, and piRNA classes for a representative B cohort plasma RNA sample contained more abundant longer sequences in comparison to the representative A cohort plasma RNA sample (Fig. 6). However, the profiles for other RNA classes are less distinctive (supplementary Figs. S24). The differences in sequencing length profiles for protocols T[3+4], T[3], and T[4] for the same plasma samples, such as the length of sRNA sequences and/or abundance of sRNA sequences having the same lengths, allows discrimination between sRNA Type 3 and Type 4. For example, in the representative B cohort plasma RNA sample, a subset of 35-nt tRNA fragments were identified as RNA Type 3 whereas other 30-nt fragments could be assigned to RNA Type 4 (Fig. 6A).

Analysis of sRNA transcriptomics using all RiboMarker® protocols

In contrast to sRNAs generated by random (or semi-random) fragmentation of RNAs, sRNAs derived from specific segments of larger RNA transcripts may not only have different RNA Types but also be differentially abundant in biological samples. Comparative analyses of sRNA sequencing profiles for all or selected RiboMarker® protocols could serve for the identification of sRNA sequences of biological and/or diagnostic interest, including their RNA class(es) and RNA Type(s), and thereby a method to select a specific protocol that would provide detection of these sRNA(s) with highest sensitivity. In addition to the sRNA sequencing length profiles, we have evaluated other analytical tools to compare the capacities of different protocols to detect sRNAs derived from the main RNA classes for plasma RNA samples.

First, to better understand the origin of sRNAs, we generated consensus mapping profiles of reads aligned for A and B cohorts' plasma RNA samples to all tRNA (Fig. 7A) and snRNA (Fig. 7B) transcripts for all RiboMarker® protocols. These data revealed that different RNA Type-specific protocols captured sRNAs derived from distinct regions of their parent transcript. For tRNA, protocol T[2] primarily captured sRNAs derived from the 3'-end of tRNA molecules, protocol T[4] from the 5'-end, and protocol T[3] from either end (Fig. 7A, left panels) for A cohort plasma RNA samples. In the case of snRNA, sRNA Type 2 were found to be derived from their internal regions, sRNA Type 3 from both 3'- and 5'-ends, while sRNA Type 4 mostly overlapped with sRNA Type 2 molecules (Fig. 7B, left panels) for A cohort plasma RNA samples. We observed notable differences in these consensus profiles between A and B cohorts' plasma RNA samples, particularly for sRNA Type 4 derived from tRNAs (Fig. 7A) and for sRNA Type 2 and Type 3 derived from snRNAs (Fig. 7B). Overall, our data demonstrated that the individual Type-specific protocols (e.g., T[2], T[3] or T[4]) could be more capable in detecting differences between the sRNA consensus profiles than the protocols detecting combinations of RNA Types (e.g., T[1+2+3+4], T[1+2] or T[3+4]).

Second, we examined heatmaps visualizing patterns of RNA transcripts and/or fragmented sequences derived from different transcripts of all main RNA classes that were detected by the different RiboMarker® protocols for A cohort (Fig. 8A) and B cohort (Fig. 8B) plasma RNA samples. Rather than identifying the best protocol(s) for detecting transcripts of individual RNA classes with higher sensitivity, these analyses facilitate the identification

of the best protocol(s) for specific detection of transcripts of interest within each RNA class. In the heatmaps, the larger width of RNA class-specific segments corresponds to larger numbers of unique RNA sequences while higher Z-scores are related to the abundance of specific RNA sequences detected. Combined, transcript presence and abundance correspond to the stronger capability of different RiboMarker[®] protocols to enrich for sRNA sequences aligned to the indicated RNA classes. The numbers of unique RNA sequences were compared and ranked (from higher to lower) for each RiboMarker[®] protocol applied to the indicated RNA classes (Fig. 8A-B). For tRNA sequences, the protocols were ranked as $T[1+2] \sim T[2] > T[4] > T[3+4] > T[1+2+3+4] > T[3]$ in A cohort plasma (Fig. 8A) and as $T[1+2] \sim T[2] > T[4] > T[1+2+3+4] \sim T[3+4] > T[3]$ in B cohort plasma (Fig. 8B) RNA samples. However, the number of unique fragments derived from tRNAs detected in the A cohort plasma RNA samples was ~2 times greater in comparison to the B cohort plasma RNA samples. For miRNA sequences, the protocols were ranked as $T[1+2] > T[1+2+3+4] > T[2] > T[4] \sim T[3+4] > T[3]$ for detection in both A and B cohorts' plasma RNA samples (Fig. 8A-B). Although the dominance of protocol T[1+2] for detecting miRNAs (RNA Type 1) was obvious and the presence of small fractions of miRNAs, which were detected by both protocols T[1+2] and T[2], could be explained by incomplete circularization of miRNAs, the uncovering of significant fractions of sRNA Type 4 aligned to miRNA sequences by protocol T[4] was intriguing. We also clearly identified short (e.g., 17-nt in length) fragments of miRNAs in sequencing length profiles for protocol T[4] (supplementary Fig. S24B). Previously, small numbers of 3'-end phosphorylated miRNAs have also been detected by Phospho-RNA-Seq in human plasma (Giraldez et al. 2019) and in cells (Lai et al. 2023).

In other sRNA classes, we also noticed significant differences in enrichment of their different RNA Types. For instance, the ratio between the numbers of unique sRNA sequences derived from snoRNAs in comparison to snRNAs were about 1 to 4 in A cohort plasma (Fig. 8A) and 2 to 1 in the B cohort plasma (Fig. 8B) RNA samples. For snoRNA, the protocols were ranked as $T[2] \sim T[1+2] > T[1+2+3+4] > T[4] \sim T[3+4] > T[3]$ in the A cohort plasma (Fig. 8A) and as $T[2] \sim T[1+2] > T[1+2+3+4] > T[4] \sim T[3+4] > T[3]$ in the B cohort plasma (Fig. 8B) RNA samples. For snRNA sequences, the protocols were ranked as $T[2] \sim T[1+2] \sim T[1+2+3+4] > T[3] \sim T[3+4] > T[4]$ in the A cohort plasma (Fig. 8A) and as $T[2] \sim T[1+2] > T[3+4] \sim T[1+2+3+4] > T[4] \sim T[3]$ in the B cohort plasma RNA samples (Figs. 8B).

For piRNA, the number of detected transcripts was ~2.5 times smaller in the A cohort plasma (Fig. 8A) than in the B cohort plasma (Fig. 8B) RNA samples, and the protocols were ranked as $T[4] \sim T[3+4] \geq T[3] > T[1+2] \sim T[2] \geq T[1+2+3+4]$ in the A cohort plasma (Fig. 8A) and as $T[2] \sim T[1+2] > T[3] \sim T[4] \sim T[3+4] > T[1+2+3+4]$ in the B cohort plasma RNA samples (Fig. 8B). Remarkably, we found that sRNA sequences aligned to piRNA transcripts were represented by all RNA Types rather than predominantly by RNA Type 1, common for the mature form of this RNA class. The latter results agree with the previously suggested piRNA biogenesis pathway indicating that all four sRNA Types could be generated from precursor piRNAs (Shigematsu et al. 2021).

Although the Phospho-RNA-Seq approach was originally developed to increase the overall recovery of unique mRNA and lncRNA fragments (Akat et al. 2019; Giraldez et al. 2019; Yao et al. 2020), we found that, in D cohort plasma RNA samples, both Phospho-RNA-Seq and protocol T[1+2+3+4] underperformed in this task in comparison to protocols T[2], T[3] and T[3+4] (supplementary Fig. S21). For mRNA sequences, the protocols were ranked as $T[4] > T[3] > T[3+4] \gg T[2] > T[1+2] \gg T[1+2+3+4]$ in the A cohort plasma (Fig. 8A) and as $T[3] \sim T[4] > T[3+4] \sim T[2] \sim T[1+2] \gg T[1+2+3+4]$ in the B cohort plasma (Fig. 8B) RNA samples. For lncRNA sequences, the RiboMarker[®] protocols were ranked as $T[4] > T[3] \sim T[3+4] > T[2] \sim T[1+2] \gg T[1+2+3+4]$ in the A cohort plasma (Fig. 8A) and as $T[2] > T[1+2] \sim T[3] \sim T[3+4] > T[4] \gg T[1+2+3+4]$ in the B cohort plasma (Fig. 8B) RNA samples.

Additionally, we found significant differences among the transcript patterns visualized in heatmaps for all featured RNA classes within the A cohort (Fig. 8A) and B cohort (Fig. 8B) plasma RNA samples. In general, protocols T[1+2] and T[3+4] are complementary for detection of sRNAs derived from all RNA classes, providing in combination, more sensitive and comprehensive coverage of the entire RNA fragmentome than “all inclusive” protocol T[1+2+3+4]. Moreover, we did not find any individual RNA class for which RiboMarker[®] protocol T[1+2+3+4] outperformed the Type-specific protocols in variety and sensitivity of detecting transcripts for all RNA classes featured in this study (Figs. 8A-B). A comparison of the heatmaps for libraries prepared using either protocol T[1+2+3+4] or Phospho-RNA-Seq for D cohort plasma RNA samples (supplementary Fig. S25A) and Pearson correlation analysis comparing the presence and abundance of specific RNA transcripts for each RNA class between these protocols (supplementary Fig. S25B) further confirmed that these two protocols preferentially detected different groups of sRNA transcripts within every analyzed RNA class.

Discrimination between A cohort and B cohort plasma samples as a prototype for sRNA biomarker discovery

Above in this study, we identified differences among sequencing profiles for the A and B cohorts' plasma RNA samples to highlight both technical aspects as well as biological variation observed using different RiboMarker[®] protocols. These cohorts simulate plasma samples collected from healthy individuals (A cohort) and breast cancer patients (B cohort) (supplementary Table S2). Using the RiboMarker[®] approach, we demonstrated its potential capability to enhance the capacity and sensitivity of distinguishing between these or any other plasma RNA samples, by selecting and using sRNAs of specific RNA Type(s). Our study might serve as a prototype for utilizing the RiboMarker[®] platform for sRNA biomarker discovery, followed by validation and translational research before clinical application.

The notable differences among the sequencing profiles of A and B cohorts' plasma RNA samples for some sRNA classes have already been demonstrated above for selected RiboMarker[®] protocols (see Figs. 3, 6 and 7). However, these profiles and the differences among them were determined for average data covering entire sRNA classes, and rather individual, selected sRNA sequences within these RNA classes could prove to be more

sensitive and precise, as previously shown using miRNAs (Wang et al. 2018). To demonstrate the feasibility of this approach, we first compared sRNA heatmaps for different RiboMarker[®] protocols and identified the most significant differences in transcript patterns among the A cohort and B cohort plasma RNA samples for the main RNA classes (Fig. 8C). Although the enrichment of certain sRNAs in either B cohort or A cohort plasma samples could be used for selecting sRNAs of interest, perhaps sRNAs that are unique for the “disease” B cohort samples over the “healthy” A cohort samples could be better biomarker candidates.

We noticed that sRNAs derived from snoRNAs and snRNAs could provide the largest number of potential sRNA biomarker candidates (Fig. 8C). Without any specific preferences, we selected a group of fragments of RNU2-1 snRNA, hereby referred to as RNU2-1_3p, which we have marked in Fig. 8C, and whose sequences are shown in supplementary Fig. S26, as an example. Protocols T[1+2] and T[2] were found to provide the best enrichment of RNU2-1_3p among the B cohort plasma to discriminate these versus the A cohort plasma RNA samples (Fig. 8C). Differential expression analyses confirmed that RNU2-1_3p was indeed among the most significantly enriched sRNAs among the B cohort plasma versus the A cohort plasma RNA samples using both T[1+2] and T[2] protocols (Fig. 9A). These two protocols also excel in the sensitivity of detection of 24-nt, 52-nt, and 65-nt long RNU2-1 fragments that were found to be unique for both B1 and B2 plasma RNA samples (Fig. 9B). The profile of the one B3 plasma RNA sample had only minor differences as compared to the A cohort plasma RNA samples, which may be related to a different breast cancer subtype in the B3 sample in comparison to the B1 and B2 plasma RNA samples. Finally, we compared read coverage profiles mapping to the RNU2-1 snRNA transcript found in these A and B cohorts’ plasma RNA samples for libraries prepared by protocols T[1+2] and T[2] (Fig. 9C). These profiles identified two areas, 01 and 02, comprising sequences of the detected RNU2-1 snRNA fragments, where area 01 was common for both A and B cohorts’ plasma RNA samples while area 02 was specific to the B cohort plasma RNA samples (Fig. 9C). Also, the finding that protocol T[2] outperformed protocol T[1+2] in the detection of the RNU2-1 fragments in area 02 for the B1 and B2 plasma samples (two bottom panels in Fig. 9C) indicated that most of these RNU2-1 fragments had RNA Type 2 ends. The top 20 most abundant sRNA sequences aligned to the RNU2-1 snRNA that we detected in the selected B1 and B2 plasma RNA samples for libraries prepared by protocol T[2] are shown in supplementary Fig. S26.

Although short 17–25-nt fragments derived from RNU2-1 snRNA area 01 were previously identified as RNA Type 1 miR-U2-1 species in human blood using standard sRNA-Seq methods (Mazières et al. 2013; Akat et al. 2019), the existence of sRNA Type 2 of longer 24–30-nt fragment sequences corresponding to area 01 and those 52–65-nt fragment sequences mapping to area 02 had not been observed. Additionally, the miR-U2-1 species were found in the blood of both cancer patients and healthy donors and have been implicated as potential multi-cancer diagnostic and prognostic biomarkers through their overexpression in non-small-cell lung, pancreatic, colorectal, and breast cancer (Mazières et al. 2013; Hahn et al. 2015; Köhler et al. 2016; Pantazi et al. 2022). However, the RNA Type 2 RNU2-1 fragments could conceivably be alternative, more sensitive biomarkers than the miR-U2-1 fragments, as we observed higher disease specificity. This example further emphasizes

the general power of the RiboMarker[®] platform for the identification and utilization of alternative Type-specific sRNAs to serve as novel, more potent biomarkers for cancer and other diseases.

The RiboMarker[®] platform enables a powerful capability to accurately quantify the complete RNA fragmentome. This work demonstrates that distinctive enzymatic pretreatments can be applied to RNA samples prior to universal sequencing library preparation. These pretreatments, which include enzymatic conversions of RNA termini and circularization, make it possible to either detect all or specifically enrich for individual sRNAs. The groundbreaking protocols can be used individually or in combination to enrich for sequencing reads of the RNA Type of greatest interest. The RiboMarker[®] platform protocols developed here offer a foundation to advance customizable RNA fragmentomics, empowering the search for novel sRNA biomarkers.

MATERIALS AND METHODS

Synthetic spike-in sRNA oligonucleotides

We designed 20 custom synthetic spike-in RNA oligonucleotides called stsRNAs. The core sequences of these stsRNAs, which do not have homology to the human genome, were adapted from Locati et al. (2015). The stsRNAs comprise 5'- and 3'-end combinations corresponding to four main RNA Types (1 through 4) shown in Fig. 1. For each RNA Type, there were five stsRNAs of different lengths: 20, 30, 40, 50, and 60 nt. The five stsRNAs for each RNA Type contain a 6-nt barcode that is RNA Type-specific in their middle section. Also, all stsRNAs feature 4 randomized nucleotides at both ends. The complete sequences of these stsRNAs are shown in Table 2. The stsRNAs were diluted to working concentrations with IDTE pH 7.5 buffer (IDT) and equimolar pooled while working on ice to minimize RNA degradation. The resulting equimolar pool was aliquoted and stored immediately at -80°C . Sequences of combo adapter (CAD), blocking oligonucleotides, adapter dimer capture probes, RT blocking oligonucleotide, RT and PCR primers (schematically shown in supplementary Fig. S1) were the same as previously described in Barberán-Soler et al. (2018) except the CAD sequence having an extra A ribonucleotide at its 3'-end. All synthetic oligonucleotides were synthesized and quantified spectrophotometrically by IDT.

Biological samples

Commercially available reference human brain total RNA samples (Thermo Fisher Scientific) were diluted to 100 ng/ μL with IDTE pH 7.5 buffer (IDT), aliquoted, and stored at -80°C . Commercially available human plasma samples collected in EDTA plasma tubes from various individuals are described in supplementary Table S2. Total RNA was extracted from plasma using Zymo Quick cfRNA Serum and Plasma kit (Zymo Research) according to the manufacturer's protocol. Briefly, 1 mL of plasma was extracted per column. Columns were eluted with 12 μL of nuclease-free water so that 1 μL of extracted RNA is equivalent to the amount of RNA found in 83 μL of plasma. Plasma and extracted RNA was stored at -80°C .

Experimental details of RiboMarker® protocols

The RiboMarker® protocols include RNA Type-specific enzymatic pretreatments of RNA samples and a universal library preparation protocol for untreated and pretreated RNA samples (Table 1). The RNA sample inputs for each of these protocols was 50 ng of human brain total RNA or the total human plasma RNA amount equivalent to 83 μ L of plasma. Each input of naturally occurring RNA was supplemented with 16 picomoles of synthetic RNA pool spike-ins (stsRNAs). This amount of stsRNAs was determined by testing multiple stsRNA input amounts into the 50 ng human brain total RNA to end up with stsRNA reads making up ~1-5% of the total trimmed sequencing reads. When applied to the RNA from 83 μ L of plasma, the stsRNA reads made up <0.01% to ~20% of the trimmed sequencing reads depending on the individual that the plasma sample came from and the RiboMarker® protocol. Mock control RNA sample input comprises 16 picomoles of the stsRNA pool without naturally occurring RNA.

Protocol T[1+2] is used here for specific capture of RNA Types 1 and 2 as well as the universal method for sRNA sequencing library preparation method for all RiboMarker® protocols (see Library preparation and sequencing section below). This is the only protocol that does not have a pretreatment step. All other RiboMarker® protocols have pretreatment step(s) before the library preparation that are described below.

To select an optimal reaction condition for protocol T[2], we used the stsRNA pool spiked into human brain RNA samples incubated in standard T4 RNA ligase reaction buffer supplemented with 10% PEG 8000, 20 U Murine RNase Inhibitor (all NEB) and 100 μ M ATP at 37°C with the following variables: (i) 10 U Rnl1 for 1 hour; (ii) 20 U Rnl1 for 1 hour; (iii) 20 U Rnl1 for 2 hours; or (iv) 10 U Rnl1 + 10 U Rnl2 for 1 hour. The selected protocol T[2] consisted of a single pretreatment reaction with a mix of T4 RNA ligase 1 and T4 RNA ligase 2 under condition (iv). The same reaction condition was also used for one of several pretreatment steps in protocols T[3+4], T[3] and T[4] as described in Table 1.

To select an optimal reaction condition for protocol T[1+2+3+4], we used the stsRNA pool spiked into human brain RNA samples incubated with 10 U T4 PNK (NEB), 20 U Murine RNase Inhibitor (NEB), 5 mM DTT (Thermo Fisher Scientific), 10 mM MgCl₂ at 37°C for 30 minutes in various buffers, including: (a) 100 mM Tris-HCl at pH 7.6 in the presence of 1 mM ATP; (b) 100 mM Imidazole-HCl at pH 6.5 in the absence of ATP; or (c) 100 mM MES-NaOH at pH 6.0 in the absence of ATP. The selected protocol T[1+2+3+4] had a single pretreatment reaction under condition (c) and was used in experiments with human plasma samples.

Protocol T[3+4] involved two sequential pretreatment reactions. The first reaction is 10 U T4 PNK 3' minus, 10 U T4 RNA ligase 1, 10 U T4 RNA ligase 2, 20 U Murine RNase Inhibitor, 1 mM ATP, 10% PEG8000, and 1X T4 PNK reaction buffer pH 7.6 (all NEB), incubated at 37°C for 1 hour. The second reaction is 10 U T4 PNK (NEB), 20 U Murine RNase Inhibitor (NEB), 5 mM DTT (ThermoFisher), 10 mM MgCl₂, and 100 mM MES pH 6.0 (ThermoFisher), incubated at 37°C for 30 minutes.

Protocol T[3] involved three sequential pretreatment reactions. The first reaction is 1 U Terminator 5'-Phosphate-dependent exonuclease (LCG Biosearch Technologies), 1X Terminator Reaction Buffer A (LCG Biosearch Technologies), and 10 U RNase Inhibitor (NEB), incubated at 37°C for 1 hour. This reaction was terminated by adding 1 µL of 100 mM EDTA pH 8.0 with a final reaction concentration of 5 mM EDTA. The second reaction is 10 U T4 PNK 3' minus, 10 U T4 RNA ligase 1, 10 U T4 RNA ligase 2, 20 U Murine RNase Inhibitor, 1 mM ATP, 10% PEG8000, and 1X T4 PNK reaction buffer pH 7.6 (all NEB), incubated at 37°C for 1 hour. The third reaction is 10 U T4 PNK (NEB), 20 U Murine RNase Inhibitor (NEB), 5 mM DTT (ThermoFisher), 10 mM MgCl₂, and 100 mM MES pH 6.0 (ThermoFisher), incubated at 37°C for 30 minutes.

Protocol T[4] consisted of three sequential pretreatment reactions. The first reaction is 30 pmol RtcB, 20 U Murine RNase Inhibitor, 100 µM MnCl₂, 10% PEG8000, 1x RtcB Reaction Buffer (all NEB), incubated at 37°C for 1 hour as previously described (Chakravarty et al. 2012; Petkovic and Müller 2015). The second reaction is 10 U T4 PNK 3' minus, 10 U T4 RNA ligase 1, 10 U T4 RNA ligase 2, 20 U Murine RNase Inhibitor, 1 mM ATP, 10% PEG8000, and 1X T4 PNK reaction buffer pH 7.6 (all NEB), incubated at 37°C for 1 hour. The third reaction is 10 U T4 PNK (NEB), 20 U Murine RNase Inhibitor (NEB), 5 mM DTT (ThermoFisher), 10 mM MgCl₂, and 100 mM MES pH 6.0 (ThermoFisher), incubated at 37°C for 30 minutes.

After each step of enzymatic pretreatments of the input RNA, the reactions were stopped by immediately processing the pretreated RNA through RNA Clean and Concentrator Kit columns (Zymo Research). The column washing protocol was modified to retain sRNAs of ≥ 15 nt in length. Briefly, 20 µL pretreatment reactions were mixed with 40 µL Binding Buffer. The resultant 60 µL mix was then mixed with 150 µL 95% ethanol. The final mix was applied to the column and washed according to the manufacturer's instructions. All pretreated RNA was eluted into 6 µL nuclease-free water and used as input for the next pretreatment reaction or directly into the sequencing library preparation as described below.

Library preparation and sequencing

All enzymatically pretreated RNA samples were prepared for sequencing using universal library preparation protocol T[1+2], which was also applied to input RNA samples that had not been enzymatically pretreated. Protocol T[1+2] was a modified version of commercially available RealSeq[®]-Biofluids protocol, which is based on the original RealSeq[®]-AC method for preparation of small RNA sequencing libraries (Barberán-Soler et al. 2018). The protocol T[1+2] workflow is schematically shown in Supplemental Figure S1 and includes the following modifications to the RealSeq[®]-Biofluids protocol: (i) adding 100 µM ATP to the adapter blocking reaction of step 2; (ii) decreasing the concentration of ATP from 1 mM to 100 µM during circularization step 3; and (iii) decreasing the concentration of additionally added dNTPs in PCR step 6 from 3 mM to 1.2 mM. The RNA inputs for protocol T[1+2] were either: (i) 16 picomoles of stsRNAs (for mock controls); (ii) 50 ng of human brain total RNA with 16 picomoles of stsRNAs; (iii) RNA volume equivalent to 83 µL of human plasma with 16 picomoles of the stsRNAs; or (iv) 6 µL of pretreated RNA sample with an initial pretreatment input as in (i), (ii), and (iii). For all library preparations, technical

triplicate libraries were made with the brain RNA while the RNA from the three individuals of a single cohort of plasma samples were considered biological replicates. The exception is C1 plasma sample, which was available in a large enough volume to be used as technical triplicates, and is not in a cohort with plasma from other individuals.

Phospho-RNA-Seq libraries were prepared with the sRNA input of the extracted total RNA volume equivalent to 83 μ L of human plasma with 16 picomoles of the stsRNA pool. The RNA was pretreated in a reaction of 10 U T4 PNK, 20 U Murine RNase Inhibitor, 1 mM ATP, and 1X T4 PNK Reaction Buffer pH 7.6 (all NEB), incubated at 37°C for 30 minutes. Following the pretreatment incubation, the reactions were stopped by immediately processing the pretreated RNA using RNA Clean and Concentrator Kit columns (Zymo Research). Similar to RiboMarker[®] protocol T[1+2+3+4], the column-washing protocol was modified to retain sRNAs of ≥ 15 nt in length. The pretreated RNA was eluted into 6 μ L nuclease-free water and used as input directly into the NEBNext Small RNA Library Prep Set for Illumina (NEB). The libraries were prepared according to the manufacturer's protocol through PCR amplification. To more accurately compare Phospho-RNA-Seq and RiboMarker[®] prepared libraries, 50 μ L of the Phospho-RNA-Seq PCR reaction was cleaned up following the same SPRI bead size selection protocol used in RiboMarker[®] protocol T[1+2].

During the preparation of sequencing libraries, PCR was performed with indexed primers to enable identification of reads unique to each sample during multiplex sequencing. The libraries were analyzed with an Agilent D1000 ScreenTape on a 4150 TapeStation instrument (Agilent) and then quantified with the TapeStation and the Qubit dsDNA BR Assay kit (Thermo Fisher Scientific). Libraries were pooled and then sequenced in multiplex using the Singular Genomics G4 as 130 bp single end runs, aiming for 10 million reads per library.

Computational methods

stsRNA quantification: To robustly assess the efficiency of capture of differentially processed naturally occurring sRNAs based on end chemistries and fragment length, we used 20 unique synthetic RNAs (stsRNAs) which were described above. For each of the four different stsRNA Types 1 through 4 (Fig. 1), a unique RNA Type-specific barcode was encoded in the internal part of the sequence (Table 2) to distinguish between the different Types during mapping. Mapping to these sequences was done after adapter trimming using bowtie2 (Martin 2011) [`--no-1mm-upfront --mp 6 --rdg 6,3 --rfg 6,3 --nrc --n-ceil L,5,0.15 -k 1 -D 20 -R 2 -N 0 -L 10 -i S,1,0.40 --score-min L,0,-0.6 --np 1`] to account for the 4-nt randomized ends, internal type-specific barcodes, and differing stsRNA lengths. A custom Python script was used to parse the alignments for each sample and quantify individual stsRNAs based on these mappings.

Read processing and feature counting: All sequencing libraries were subsampled to the maximum number of reads available based on the sample with the fewest number of reads to accurately compare samples within an experiment. Sequencing reads were trimmed of the adapter sequence using cutadapt (Martin 2011) and then further filtered to remove any

reads less than 15 nt in length. Reads were then sequentially aligned using either bowtie or bowtie2 (Langmead and Salzberg 2012) to vector contaminants (UniVec), rRNAs, tRNAs (with the addition of the - CCA tail) identified using tRNAScan-SE (Chan et al. 2021), miRNAs from miRbase (Kozomara et al. 2019), snRNAs/snoRNAs from Ensembl (Dyer et al., 2025), piRNAs from piRbase (Wang et al. 2022), processed mRNA transcripts including 5'UTR, CDS and 3'UTR (mRNA/protein_coding) from Ensembl (Dyer et al. 2025), and lastly to the genome. Reads associated with these features were then quantified using a custom Python script, choosing the feature annotation with the largest degree of overlap. Multimapping ncRNA reads with more than one highest-scoring alignment were randomly selected, while multimapping mRNA reads were not included in the downstream analysis. For tRNA, snRNA, and snoRNA mapping reads, a read was assigned as either a '5p', '3p', or an 'internal' feature based on where it mapped in the parent transcript. For '5p' features, it must map within 10 nt of the 5' end, '3p' features must map within 10 nt of the 3' end, and an 'internal' feature does not map within 10 nt of either the 5' or 3' end of the parent transcript. The resulting raw read counts underwent normalization and differential expression analysis using DESeq2 (Love 2014). Visualizations were generated using custom Python scripts. Raw fastq files are available on the NCBI Sequence Read Archive database under bioproject accession: PRJNA1348206.

stsRNA and natural RNA read length profiles: The proportion (P) of a given stsRNA Type, or RNA class (T) (for natural RNAs, excluding unmapped reads), at a specified read length (L) was calculated by dividing its raw read count ($C_{T,L}$) by the total number of reads across all classes and lengths for its respective category (i.e., all stsRNA or all naturally occurring RNAs). This is shown by the following equation:

$$P_{T,L} = \frac{C_{T,L}}{Total_{stsRNA} \text{ or } Total_{Natural}}$$

Consensus read mapping profiles: The distribution of read coverage across transcript bodies was calculated from the 5'-end transcript start to the 3' transcript end by first generating a vector of read coverage at each position in a given RNA transcript. To normalize for different transcript lengths, each coverage vector was partitioned into 100 equal-sized bins, and the total counts within each bin were summed. These 100-point vectors were then aggregated by summing them for all transcripts within a given RNA class (e.g., tRNA, snRNA) for each sample. To create a final profile for each experimental protocol, these aggregated vectors were averaged across all biological replicates. For visualization, the resulting average profiles were min-max scaled to a range of [0,1] to emphasize the shape of the read distribution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Dr. Denise McGrath for helpful discussions, comments and technical editing of the manuscript. This work was supported in part by NIH grants 1R43HG013284-01 and 2R44HG013284-02A1 to S.A.K.

REFERENCES

- Akat KM, Lee YA, Hurley A, Morozov P, Max KE, Brown M, Bogardus K, Sopeyin A, Hildner K, Diacovo TG, Neurath MF, Borggrete M, Tuschl T. 2019. Detection of circulating extracellular mRNAs by modified small-RNA-sequencing analysis. *JCI Insight* 5: e127317. doi:10.1172/jci.insight.127317. [PubMed: 30973829]
- AbouHaidar MG, Ivanov IG. 1999. Non-enzymatic RNA hydrolysis promoted by the combined catalytic activity of buffers and magnesium ions. *Z Naturforsch C J Biosci* 54: 542–8. doi:10.1515/znc-1999-7-813. [PubMed: 10488562]
- Barberán-Soler S, Vo JM, Hogans RE, Dallas A, Johnston BH, Kazakov SA. 2018. Decreasing miRNA sequencing bias using a single adapter and circularization approach. *Genome Biol* 19: 105. doi:10.1186/s13059-018-1488-z. [PubMed: 30173660]
- Bradley CC, Wang C, Gordon AJE, Wen AX, Luna PN, Cooke MB, Kohrn BF, Kennedy SR, Avadhanula V, Piedra PA, Lichtarge O, Shaw CA, Ronca SE, Herman C. 2024. Targeted accurate RNA consensus sequencing (tARC-seq) reveals mechanisms of replication error affecting SARS-CoV-2 divergence. *Nat Microbiol* 9: 1382–1392. doi:10.1038/s41564-024-01655-4. [PubMed: 38649410]
- Breslow R, Huang DL. 1991. Effects of metal ions, including Mg^{2+} and lanthanides, on the cleavage of ribonucleotides and RNA model compounds. *Proc Natl Acad Sci USA* 88: 4080–4083. doi:10.1073/pnas.88.10.4080. [PubMed: 1709734]
- Brown TS, Bevilacqua PC. 2005. Method for assigning double-stranded RNA structures. *Biotechniques* 38: 368–372. doi:10.2144/05383BM04. [PubMed: 15786803]
- Busch S, Kirsebom LA, Notbohm H, Hartmann RK. 2000. Differential role of the intermolecular base-pairs G292-C(75) and G293-C(74) in the reaction catalyzed by Escherichia coli RNase P RNA. *J Mol Biol* 299: 941–951. doi:10.1006/jmbi.2000.3789. [PubMed: 10843849]
- Cabús L, Lagarde J, Curado J, Lizano E, Pérez-Boza J. 2022. Current challenges and best practices for cell-free long RNA biomarker discovery. *Biomarker Res* 10: 62. doi:10.1186/s40364-022-00409-w.
- Cameron V, Uhlenbeck OC. 1977. 3'-Phosphatase activity in T4 polynucleotide kinase. *Biochemistry* 16: 5120–5126. doi:10.1021/bi00642a027. [PubMed: 199248]
- Chakravarty AK, Subbotin R, Chait BT, Shuman S. RNA ligase RtcB splices 3'-phosphate and 5'-OH ends via covalent RtcB-(histidinyl)-GMP and polynucleotide-(3')pp(5')G intermediates. 2012. *Proc Natl Acad Sci USA* 109: 6072–6077. doi:10.1073/pnas.1201207109. [PubMed: 22474365]
- Chan PP, Lin BY, Mak AJ, Lowe TM. 2021. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res* 49: 9077–9096. doi:10.1093/nar/gkab688. [PubMed: 34417604]
- Chen H, Cheng K, Liu X, An R, Komiyama M, Liang X. 2020. Preferential production of RNA rings by T4 RNA ligase 2 without any splint through rational design of precursor strand. *Nucleic Acids Res* 48: e54. doi:10.1093/nar/gkaa181. [PubMed: 32232357]
- Chen S, Jin Y, Wang S, Xing S, Wu Y, Tao Y, Ma Y, Zuo S, Liu X, Hu Y, Chen H, Luo Y, Xia F, Xie C, Yin J, Wang X, Liu Z, Zhang N, Zech Xu Z, Lu ZJ, Wang P. 2022. Cancer type classification using plasma cell-free RNAs derived from human and microbes. *Elife* 11: e75181. doi:10.7554/eLife.75181. [PubMed: 35816095]
- Chen Q, Zhou T. 2023. Emerging functional principles of tRNA-derived small RNAs and other regulatory small RNAs. *J Biol Chem* 299: 105225. doi:10.1016/j.jbc.2023.105225. [PubMed: 37673341]
- Chheda U, Pradeepan S, Esposito E, Strezsak S, Fernandez-Delgado O, Kranz J. 2024. Factors affecting stability of RNA - temperature, length, concentration, pH, and buffering species. *J Pharm Sci* 113: 377–385. doi:10.1016/j.xphs.2023.11.023. [PubMed: 38042343]

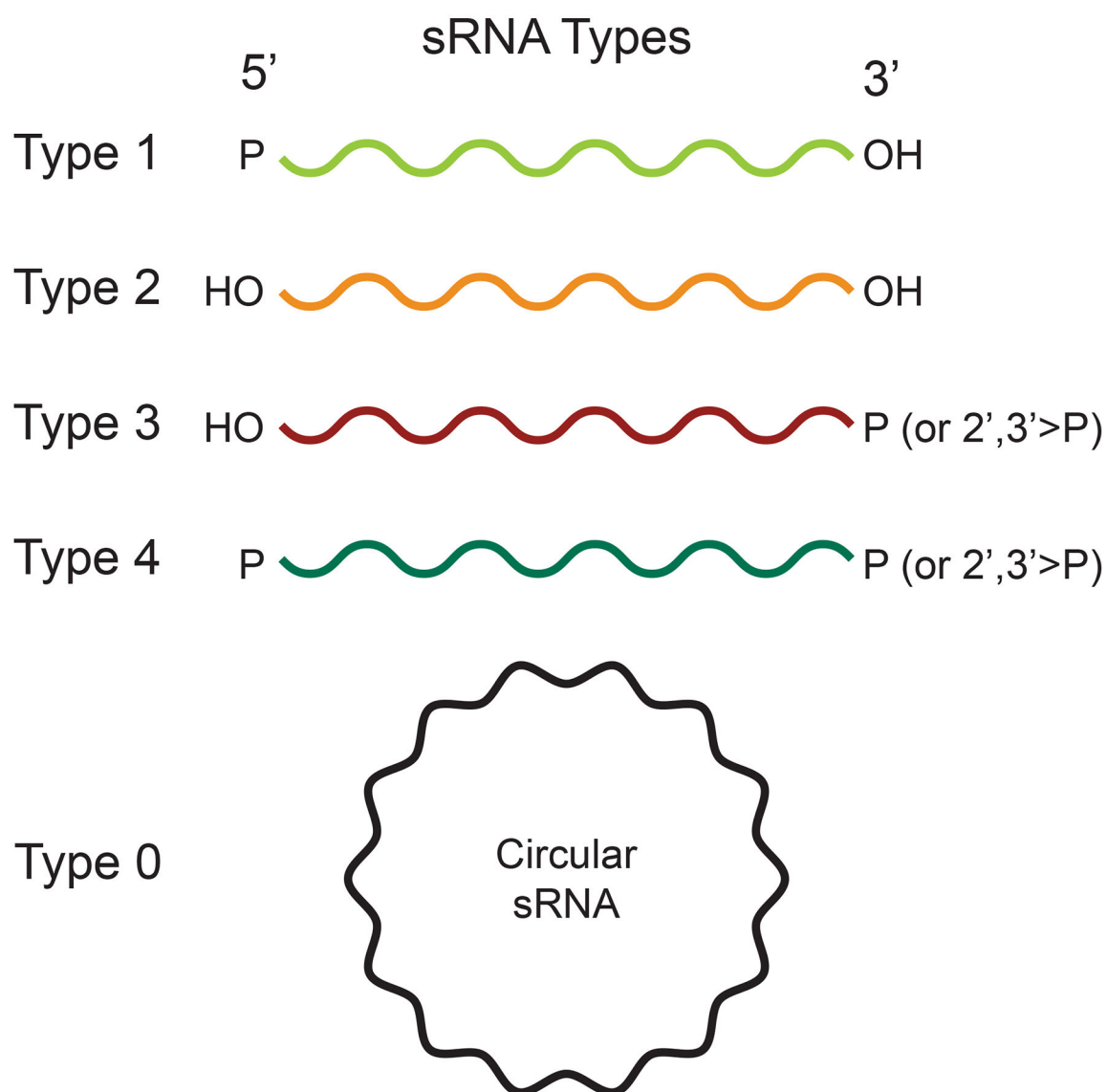
- Chu Y, Wang T, Dodd D, Xie Y, Janowski BA, Corey DR. 2015. Intramolecular circularization increases efficiency of RNA sequencing and enables CLIP-Seq of nuclear RNA from human cells. *Nucleic Acids Res* 43: e75. doi:10.1093/nar/gkv213. [PubMed: 25813040]
- Cranston JW, Silber R, Malathi VG, Hurwitz J. 1974. Studies on ribonucleic acid ligase: Characterization of an adenosine triphosphate-inorganic pyrophosphate exchange reaction and demonstration of an enzyme-adenylate complex with T4 bacteriophage-induced enzyme. *J Biol Chem* 249: 7447–7456. doi:10.1016/S0021-9258(19)81259-0. [PubMed: 4373468]
- Crocker OJ, Trigg NA, Conine CC. 2022. Cloning and sequencing eukaryotic small RNAs. *Curr Protoc* 2: e495. doi:10.1002/cpz1.495. [PubMed: 35926113]
- De Sota RE, Quake SR, Sninsky JJ, Toden S. 2024. Decoding bioactive signals of the RNA secretome: the cell-free messenger RNA catalogue. *Expert Rev Mol Med* 26: e12. doi:10.1017/erm.2024.12. [PubMed: 38682644]
- Del Piano A, Kecman T, Schmid M, Barbieri R, Brocchieri L, Tornaletti S, Firrito C, Minati L, Bernabo P, Signoria I, Lauria F, Gillingwater TH, Viero G, Clamer M. 2022. Phospho-RNA sequencing with circAID-p-seq. *Nucleic Acids Res* 50: e23. doi:10.1093/nar/gkab1158. [PubMed: 34850942]
- Di Fazio A, Gullerova M. 2023. An old friend with a new face: tRNA-derived small RNAs with big regulatory potential in cancer biology. *Br J Cancer* 128: 1625–1635. doi:10.1038/s41416-023-02191-4. [PubMed: 36759729]
- Dyer SC, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, Barrera-Enriquez VP, Becker A, Bennett R, Beracochea M, Berry A, et al. 2025. Ensembl 2025. *Nucleic Acids Res* 53: D948–D957. doi:10.1093/nar/gkae1071. [PubMed: 39656687]
- England TE, Uhlenbeck OC. 1978. Enzymatic oligoribonucleotide synthesis with T4 RNA ligase. *Biochemistry* 17: 2069–76. doi:10.1021/bi00604a008. [PubMed: 667012]
- Fish L, Zhang S, Yu JX, Culbertson B, Zhou AY, Goga A, Goodarzi H. 2018. Cancer cells exploit an orphan RNA to drive metastatic progression. *Nat Med* 24: 1743–1751. doi:10.1038/s41591-018-0230-4. [PubMed: 30397354]
- Galvanin A, Dostert G, Ayadi L, Marchand V, Velot É, Motorin Y. 2019. Diversity and heterogeneity of extracellular RNA in human plasma. *Biochimie* 164: 22–36. doi:10.1016/j.biochi.2019.05.011. [PubMed: 31108123]
- Gao L, Fan J, He J, Che X, Wang X, Han C. 2024. Small nucleolar RNAs as diagnostic and prognostic biomarkers in cancer: a systematic review and meta-analysis. *Technol Cancer Res Treat* 23: 15330338241245939. doi:10.1177/15330338241245939. [PubMed: 38752263]
- Giraldez MD, Spengler RM, Etheridge A, Goicochea AJ, Tuck M, Choi SW, Galas DJ, Tewari M. 2019. Phospho-RNA-seq: a modified small RNA-seq method that reveals circulating mRNA and lncRNA fragments as potential biomarkers in human plasma. *EMBO J* 38: e101695. doi:10.15252/embj.2019101695. [PubMed: 31053596]
- Gout JF, Li W, Fritsch C, Li A, Haroon S, Singh L, Hua D, Fazelinia H, Smith Z, Seeholzer S, Thomas K, Lynch M, Vermulst M. 2017. The landscape of transcription errors in eukaryotic cells. *Sci Adv* 3: e1701484. doi:10.1126/sciadv.1701484. [PubMed: 29062891]
- Hafner M, Renwick N, Brown M, Mihailović A, Holoch D, Lin C, Pena JT, Nusbaum JD, Morozov P, Ludwig J, Ojo T, Luo S, Schroth G, Tuschl T. 2011. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* 17: 1697–712. doi:10.1261/rna.2799511. [PubMed: 21775473]
- Hahn SA, Zoellner H, Maghnouj A, Vangala DB. 2015. U2 small nuclear RNA as a biomarker in cancer. In: Preedy V, Patel V (eds.) *Biomarkers in Cancer. Biomarkers in Disease: Methods, Discoveries and Applications*. Springer, Dordrecht. pp. 233–250. doi:10.1007/978-94-007-7681-4_6.
- Harrison B, Zimmerman SB. 1984. Polymer-stimulated ligation: enhanced ligation of oligo- and polynucleotides by T4 RNA ligase in polymer solutions. *Nucleic Acids Res* 12: 8235–8251. doi:10.1093/nar/12.21.8235. [PubMed: 6504699]
- Heyer EE, Ozadam H, Ricci EP, Cenik C, Moore MJ. 2015. An optimized kit-free method for making strand-specific deep sequencing libraries from RNA fragments. *Nucleic Acids Res* 43: e2. doi:10.1093/nar/gku1235. [PubMed: 25505164]

- Ho CK, Shuman S. 2002. Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc Natl Acad Sci USA* 99: 12709–12714. doi:10.1073/pnas.192184699. [PubMed: 12228725]
- Ho CK, Wang LK, Lima CD, Shuman S. 2004. Structure and mechanism of RNA ligase. *Structure* 12: 327–339. doi:10.1016/j.str.2004.01.011. [PubMed: 14962393]
- Honda S, Morichika K, Kirino Y. 2016. Selective amplification and sequencing of cyclic phosphate-containing RNAs by the cP-RNA-seq method. *Nat Protoc* 11: 476–89. doi:10.1038/nprot.2016.025. [PubMed: 26866791]
- Hulstaert E, Morlion A, Avila Cobos F, Verniers K, Nuytens J, Vanden Eynde E, Yigit N, Anckaert J, Geerts A, Hindryckx P, et al. 2020. Charting extracellular transcriptomes in the human biofluid RNA atlas. *Cell Rep* 33: 108552. doi:10.1016/j.celrep.2020.108552. [PubMed: 33378673]
- Jackson TJ, Spriggs RV, Burgoyne NJ, Jones C, Willis AE. 2014. Evaluating bias-reducing protocols for RNA sequencing library preparation. *BMC Genomics* 15: 569. doi:10.1186/1471-2164-15-569. [PubMed: 25001197]
- Jackowiak P, Hojka-Osinska A, Philips A, Zmienko A, Budzko L, Maillard P, Budkowska A, Figlerowicz M. 2017. Small RNA fragments derived from multiple RNA classes - the missing element of multi-omics characteristics of the hepatitis C virus cell culture model. *BMC Genomics* 8: 502. doi:10.1186/s12864-017-3891-3.
- Jia J, Yang S, Huang J, Zheng H, He Y, Wang L. 2021. Distinct extracellular RNA profiles in different plasma components. *Front Genet* 12: 564780. doi:10.3389/fgene.2021.564780. [PubMed: 34234804]
- Kaufmann G, Klein T, Littauer UZ. 1974. T4 RNA ligase: substrate chain length requirements. *FEBS Lett* 46: 271–275. doi:10.1016/0014-5793(74)80385-6. [PubMed: 4609429]
- Kawaji H, Nakamura M, Takahashi Y, Sandelin A, Katayama S, Fukuda S, Daub CO, Kai C, Kawai J, Yasuda J, Carninci P, Hayashizaki Y. 2008. Hidden layers of human small RNAs. *BMC Genomics* 9: 157. doi:10.1186/1471-2164-9-157. [PubMed: 18402656]
- Khamina K, Diendorfer AB, Skalicky S, Weigl M, Pultar M, Krammer TL, Fournier CA, Schofield AL, Otto C, Smith AT, et al. 2022. A MicroRNA next-generation-sequencing discovery assay (miND) for genomescale analysis and absolute quantitation of circulating microRNA biomarkers. *Int J Mol Sci* 23: 1226. doi:10.3390/ijms23031226. [PubMed: 35163149]
- Köhler J, Schuler M, Gauler TC, Nöpel-Dünnebacke S, Ahrens M, Hoffmann AC, Kasper S, Nensa F, Gomez B, Hahnemann M, et al. 2016. Circulating U2 small nuclear RNA fragments as a diagnostic and prognostic biomarker in lung cancer patients. *J Cancer Res Clin Oncol* 142: 795–805. doi:10.1007/s00432-015-2095-y. [PubMed: 26687686]
- Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Res* 47: D155–D162. doi:10.1093/nar/gky1141. [PubMed: 30423142]
- Krug M, Uhlenbeck OC. 1982. Reversal of T4 RNA ligase. *Biochemistry* 21: 1858–1864. doi:10.1021/bi00537a024. [PubMed: 7082652]
- Kugelberg U, Nätt D, Skog S, Kutter C, Öst A. 2021. 5'XP sRNA-seq: efficient identification of transcripts with and without 5' phosphorylation reveals evolutionary conserved small RNA. *RNA Biol* 18: 1588–1599. doi:10.1080/15476286.2020.1861770. [PubMed: 33382953]
- Kumar P, Johnston BH, Kazakov SA. 2011. miR-ID: a novel, circularization-based platform for detection of microRNAs. *RNA* 17: 365–380. doi:10.1261/rna.2490111. [PubMed: 21169480]
- Kwon YS. 2011. Small RNA library preparation for next-generation sequencing by single ligation, extension and circularization technology. *Biotechnol Lett* 33: 1633–41. doi:10.1007/s10529-011-0611-y. [PubMed: 21479628]
- Larson MH, Pan W, Kim HJ, Mauntz RE, Stuart SM, Pimentel M, Zhou Y, Knudsgaard P, Demas V, Aravanis AM, Jamshidi A. 2021. A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat Commun* 12: 2357. doi:10.1038/s41467-021-22444-1. [PubMed: 33883548]
- Lai H, Feng N, Zhai Q. 2023. Discovery of the major 15-30 nt mammalian small RNAs, their biogenesis and function. *Nat Commun* 14: 5796. doi:10.1038/s41467-023-41554-6. [PubMed: 37723159]

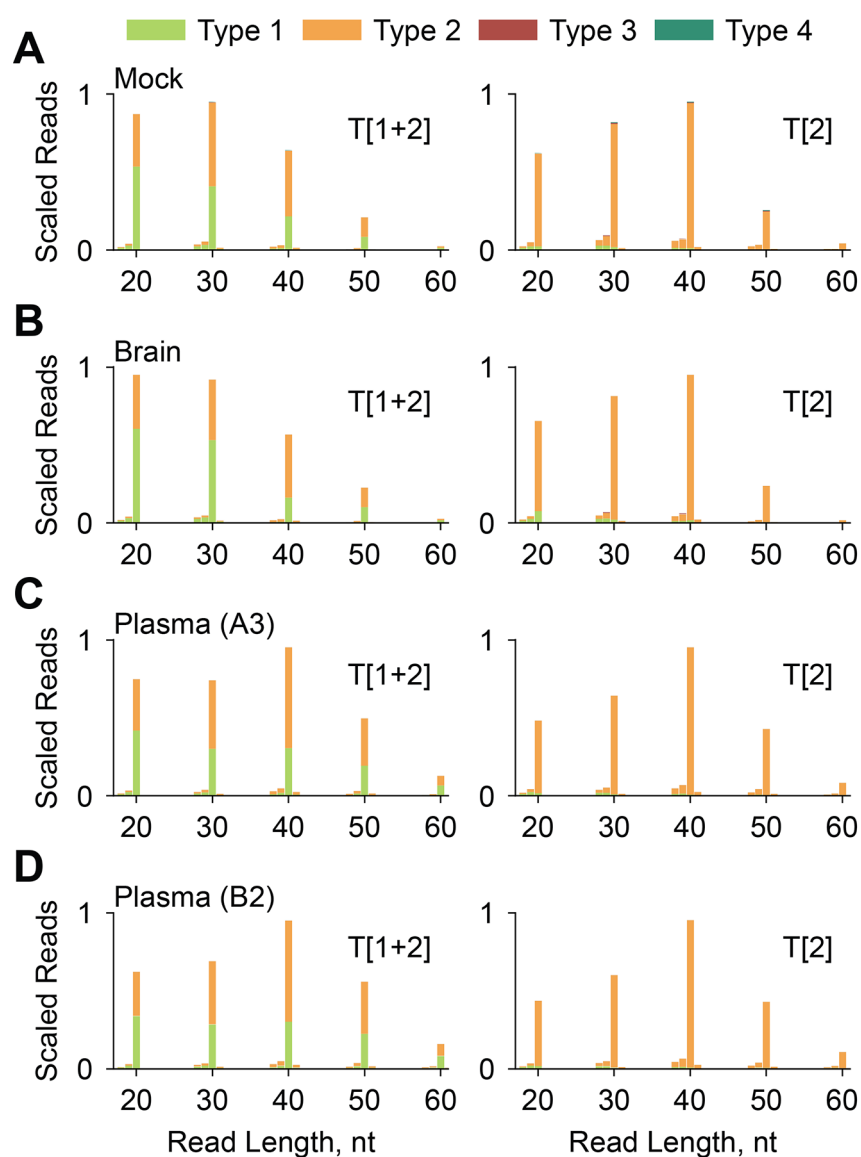
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–9. doi:10.1038/nmeth.1923. [PubMed: 22388286]
- Lee C, Harris RA, Wall JK, Mayfield RD, Wilke CO. 2013. RNase III and T4 polynucleotide Kinase sequence biases and solutions during RNA-seq library construction. *Biol Direct* 8: 16. doi:10.1186/1745-6150-8-16. [PubMed: 23826734]
- Lee HH, Wang YN, Hung MC. 2019. Functional roles of the human ribonuclease A superfamily in RNA metabolism and membrane receptor biology. *Mol Aspects Med* 70: 106–116. doi:10.1016/j.mam.2019.03.003. [PubMed: 30902663]
- Lillehaug JR, Kleppe K. 1975. Kinetics and specificity of T4 polynucleotide kinase. *Biochemistry* 14:1221–1225. doi:10.1021/bi00677a020. [PubMed: 235284]
- Liu Z, Wang T, Yang X, Zhou Q, Zhu S, Zeng J, Chen H, Sun J, Li L, Xu J, Geng C, Xu X, Wang J, Yang H, Zhu S, Chen F, Wang WJ. 2022. Polyadenylation ligation-mediated sequencing (PALM-Seq) characterizes cell-free coding and non-coding RNAs in human biofluids. *Clin Transl Med* 12: e987. doi:10.1002/ctm2.987. [PubMed: 35858042]
- Liu Y, Liang Y, Li Q, Li Q. 2023. Comprehensive analysis of circulating cell-free RNAs in blood for diagnosing non-small cell lung cancer. *Comput Struct Biotechnol J* 21: 4238–4251. doi:10.1016/j.csbj.2023.08.029. [PubMed: 37692082]
- Locati MD, Terpstra I, de Leeuw WC, Kuzak M, Rauwerda H, Ensink WA, van Leeuwen S, Nehrdich U, Spaink HP, Jonker MJ, Breit TM, Dekker RJ. 2015. Improving small RNA-seq by using a synthetic spike-in set for size-range quality control together with a set for data normalization. *Nucleic Acids Res* 43: e89. doi:10.1093/nar/gkv303. [PubMed: 25870415]
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550. doi:10.1186/s13059-014-0550-8. [PubMed: 25516281]
- Lutzmayr S, Enugutti B, Nodine MD. 2017. Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-Seq data. *Sci Rep* 7: 5913. doi:10.1038/s41598-017-06174-3. [PubMed: 28724941]
- Maguire S, Lohman GJS, Guan S. 2020. A low-bias and sensitive small RNA library preparation method using randomized splint ligation. *Nucleic Acids Res* 48: e80. doi:10.1038/s41598-017-06174-3. [PubMed: 32496547]
- Mazières J, Catherine C, Delfour O, Gouin S, Rouquette I, Delisle MB, Prévot G, Escamilla R, Didier A, Persing DH, Bates M, Michot B. 2013. Alternative processing of the U2 small nuclear RNA produces a 19–22nt fragment with relevance for the detection of non-small cell lung cancer in human serum. *PLoS One* 8: e60134. doi:10.1371/journal.pone.0060134. [PubMed: 23527303]
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17: 10–12. doi:10.14806/ej.17.1.200.
- Max KEA, Bertram K, Akat KM, Bogardus KA, Li J, Morozov P, Ben-Dov IZ, Li X, Weiss ZR, Azizian A, et al. 2018. Human plasma and serum extracellular small RNA reference profiles and their clinical utility. *Proc Natl Acad Sci USA* 115: E5334–E5343. doi:10.1073/pnas.1714397115. [PubMed: 29777089]
- McCormick KP, Willmann MR, Meyers BC. 2011. Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. *Silence* 2: 2. doi:10.1186/1758-907X-2-2. [PubMed: 21356093]
- Ning C, Cai P, Liu X, Li G, Bao P, Yan L, Ning M, Tang K, Luo Y, Guo H, Wang Y, Wang Z, Chen L, Lu ZJ, Yin J. 2023. A comprehensive evaluation of full-spectrum cell-free RNAs highlights cell-free RNA fragments for early-stage hepatocellular carcinoma detection. *EBioMedicine* 93: 104645. doi:10.1016/j.ebiom.2023.104645. [PubMed: 37315449]
- Obi P, Chen YG. 2021. The design and synthesis of circular RNAs. *Methods* 196: 85–103. doi:10.1016/j.ymeth.2021.02.020. [PubMed: 33662562]
- Ohtsuka E, Doi T, Uemura H, Taniyama Y, Ikehara M. 1980. Comparison of substrate base sequences for RNA ligase reactions in the synthesis of a tetradecanucleotide corresponding to bases 21–34 of *E. coli* tRNA^{fMet} 1. *Nucleic Acids Res* 8: 3909–3916. doi:10.1093/nar/8.17.3909. [PubMed: 7003540]
- Pan T, Gutell RR, Uhlenbeck OC. 1991. Folding of circularly permuted transfer RNAs. *Science* 254: 1361–1364. doi:10.1126/science.1720569. [PubMed: 1720569]

- Pantazi P, Clements T, Venø M, Abrahams VM, Holder B. 2022. Distinct non-coding RNA cargo of extracellular vesicles from M1 and M2 human primary macrophages. *J Extracell Vesicles* 11: e12293. doi:10.1002/jev2.12293. [PubMed: 36544271]
- Pardini B, Sabo AA, Birolo G, Calin GA. 2019. Noncoding RNAs in extracellular fluids as cancer biomarkers: the new frontier of liquid biopsies. *Cancers (Basel)* 11: 1170. doi:10.3390/cancers11081170. [PubMed: 31416190]
- Peach SE, York K, Hesselberth JR. 2015. Global analysis of RNA cleavage by 5'-hydroxyl RNA sequencing. *Nucleic Acids Res* 43: e108. doi:10.1093/nar/gkv536. [PubMed: 26001965]
- Petkovic S, Müller S. 2015. RNA circularization strategies in vivo and in vitro. *Nucleic Acids Res* 43: 2454–2465. doi:10.1093/nar/gkv045. [PubMed: 25662225]
- Romaniuk E, McLaughlin LW, Neilson T, Romaniuk PJ. 1982. The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. *Eur J Biochem* 125: 639–643. doi:10.1111/j.1432-1033.1982.tb06730.x. [PubMed: 7117259]
- Romaniuk PJ, Uhlenbeck OC. 1983. Joining of RNA molecules with RNA ligase. *Methods Enzymol* 100: 52–59. doi: 10.1016/0076-6879(83)00045-2. [PubMed: 6194411]
- Shi H, Zhou Y, Jia E, Pan M, Bai Y, Ge Q. 2021a. Bias in RNA-seq library preparation: current challenges and solutions. *Biomed Res Int* 2021: 6647597. doi:10.1155/2021/6647597. [PubMed: 33987443]
- Shi J, Zhang Y, Tan D, Zhang X, Yan M, Zhang Y, Franklin R, Shahbazi M, Mackinlay K, Liu S, et al. 2021b. PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications. *Nat Cell Biol* 23: 424–436. doi:10.1038/s41556-021-00687-w. [PubMed: 33820973]
- Shi J, Zhou T, Chen Q. 2022. Exploring the expanding universe of small RNAs. *Nat Cell Biol* 24: 415–423. doi:10.1038/s41556-022-00880-5. [PubMed: 35414016]
- Shigematsu M, Kawamura T, Morichika K, Izumi N, Kiuchi T, Honda S, Pliatsika V, Matsubara R, Rigoutsos I, Katsuma S, Tomari Y, Kirino Y. 2021. RNase κ promotes robust piRNA production by generating 2',3'-cyclic phosphate-containing precursors. *Nat Commun* 12: 4498. doi:10.1038/s41467-021-24681-w. [PubMed: 34301931]
- Shigematsu M, Kirino Y. 2022. Making invisible RNA visible: Discriminative sequencing methods for RNA molecules with specific terminal formations. *Biomolecules* 12: 611. doi:10.3390/biom12050611. [PubMed: 35625540]
- Schürer H, Lang K, Schuster J, Mörl M. 2002. A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic Acids Res* 30: e56. doi:10.1093/nar/gnf055. [PubMed: 12060694]
- Silber R, Malathi VG, Hurwitz J. 1972. Purification and properties of bacteriophage T4-induced RNA ligase. *Proc Natl Acad Sci USA* 69: 3009–3013. doi:10.1073/pnas.69.10.3009. [PubMed: 4342972]
- Solaguren-Beascoa M, Gámez-Valero A, Escaramís G, Herrero-Lorenzo M, Ortiz AM, Minguet C, Gonzalo R, Bravo MI, Costa M, Martí E. 2023. Phospho-RNA-Seq highlights specific small RNA profiles in plasma extracellular vesicles. *Int J Mol Sci* 24: 11653. doi:10.3390/ijms241411653. [PubMed: 37511412]
- Su Z, Monshaugen I, Klungland A, Ougland R, Dutta A. 2022. Characterization of novel small non-coding RNAs and their modifications in bladder cancer using an updated small RNA-seq workflow. *Front Mol Biosci* 9: 887686. doi:10.3389/fmolb.2022.887686. [PubMed: 35923465]
- Székely M, Sanger F. 1969. Use of polynucleotide kinase in fingerprinting non-radioactive nucleic acids. *J Mol Biol* 43: 607–617. doi:10.1016/0022-2836(69)90362-3. [PubMed: 5401237]
- Tao Y, Xing S, Zuo S, Bao P, Jin Y, Li Y, Li M, Wu Y, Chen S, Wang X, et al. 2023. Cell-free multi-omics analysis reveals potential biomarkers in gastrointestinal cancer patients' blood. *Cell Rep Med* 4: 101281. doi:10.1016/j.xcrm.2023.101281. [PubMed: 37992683]
- Tosar JP, Segovia M, Castellano M, Gámbaro F, Akiyama Y, Fagúndez P, Olivera Á, Costa B, Possi T, Hill M, Ivanov P, Cayota A. 2020. Fragmentation of extracellular ribosomes and tRNAs shapes the extracellular RNAome. *Nucleic Acids Res* 48: 12874–12888. doi:10.1093/nar/gkaa674. [PubMed: 32785615]
- Tosar JP, Castellano M, Costa B, Cayota A. 2024. Small RNA structural biochemistry in a post-sequencing era. *Nat Protoc* 19: 595–602. doi:10.1038/s41596-023-00936-2. [PubMed: 38057624]

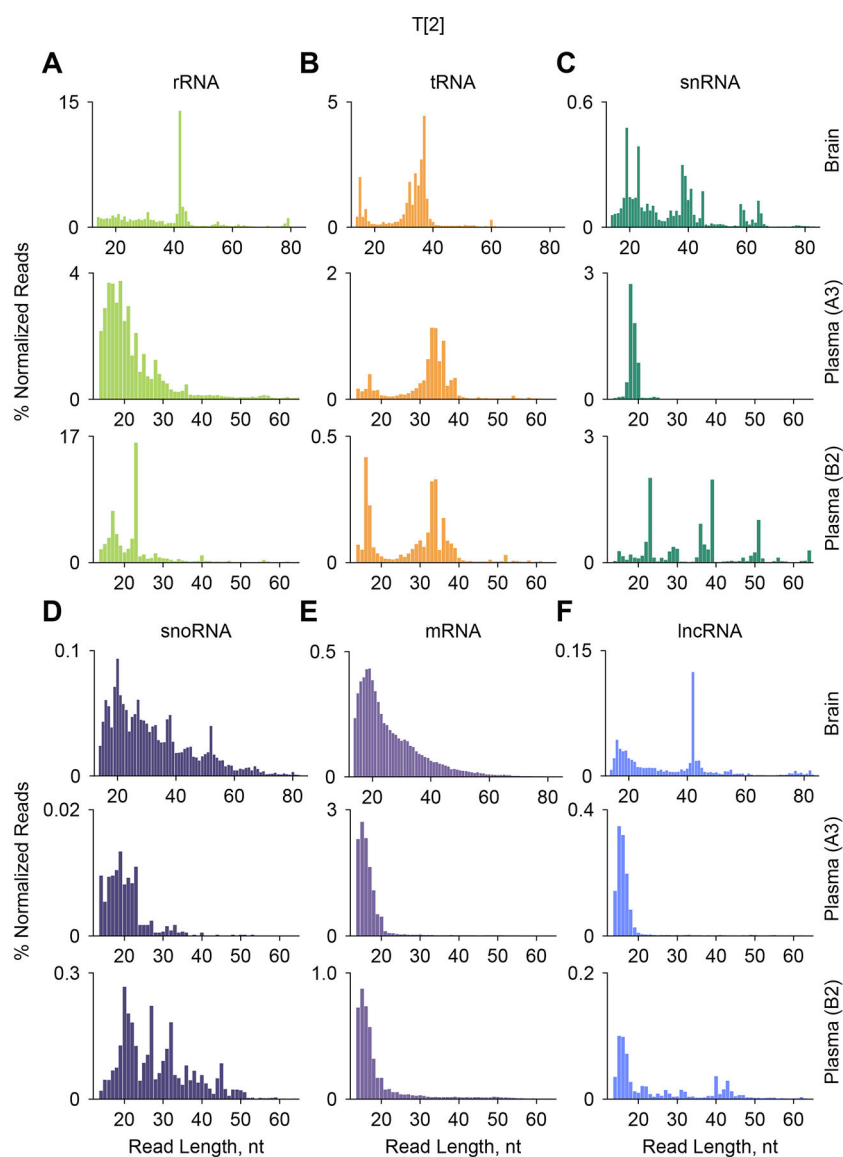
- Uhlenbeck OC, Gumport RI. 1982. T4 RNA ligase. The Enzymes 15, Nucleic Acids, part B, Boyer PD, ed. (New York: Academic Press), pp. 31–60. doi:10.1016/S1874-6047(08)60274-7.
- van Houten V, Denkers F, van Dijk M, van den Brekel M, Brakenhoff R. 1988. Labeling efficiency of oligonucleotides by T4 polynucleotide kinase depends on 5'-nucleotide. *Anal Biochem* 265: 386–389. doi:10.1006/abio.1998.2900.
- van de Sande JH, Kleppe K, Khorana HG. 1973. Reversal of bacteriophage T4 induced polynucleotide kinase action. *Biochemistry* 12: 5050–5055. doi:10.1021/bi00749a004. [PubMed: 4366306]
- Verwilt J, Mestdagh P, Vandesompele J. 2023. Artifacts and biases of the reverse transcription reaction in RNA sequencing. *RNA* 29: 889–897. doi:10.1261/rna.079623.123. [PubMed: 36990512]
- Vibert J, Saulnier O, Collin C, Petit F, Borgman KJE, Vigneau J, Gautier M, Zaidi S, Pierron G, Watson S, et al. 2022. Oncogenic chimeric transcription factors drive tumor-specific transcription, processing, and translation of silent genomic regions. *Mol Cell* 82: 2458–2471.e9. doi:10.1016/j.molcel.2022.04.019. [PubMed: 35550257]
- Vickers KC, Roteta LA, Hucheson-Dilks H, Han L, Guo Y. 2015. Mining diverse small RNA species in the deep transcriptome. *Trends Biochem Sci*. 40: 4–7. doi:10.1016/j.tibs.2014.10.009. [PubMed: 25435401]
- Viollet S, Fuchs RT, Munafo DB, Zhuang F, Robb GB. 2011. T4 RNA ligase 2 truncated active site mutants: improved tools for RNA analysis. *BMC Biotechnol* 11: 72. doi:10.1186/1472-6750-11-72. [PubMed: 21722378]
- Vlassov VV, Zuber G, Felden B, Behr JP, Giegé R. 1995. Cleavage of tRNA with imidazole and spermine imidazole constructs: a new approach for probing RNA structure. *Nucleic Acids Res* 23: 3161–3167. doi: 10.1093/nar/23.16.3161. [PubMed: 7667092]
- Wang H, Peng R, Wang J, Qin Z, Xue L. 2018. Circulating microRNAs as potential cancer biomarkers: the advantage and disadvantage. *Clin Epigenetics* 10: 59. doi:10.1186/s13148-018-0492-1. [PubMed: 29713393]
- Wang Y, Weng Q, Ge J, Zhang X, Guo J, Ye G. 2022. tRNA-derived small RNAs: Mechanisms and potential roles in cancers. *Genes Dis* 9: 1431–1442. doi:10.1016/j.gendis.2021.12.009. [PubMed: 36157501]
- Wang J, Huang J, Hu Y, Guo Q, Zhang S, Tian J, Niu Y, Ji L, Xu Y, Tang P, et al. 2024. Terminal modifications independent cell-free RNA sequencing enables sensitive early cancer detection and classification. *Nat Commun* 15: 156. doi:10.1038/s41467-023-44461-y. [PubMed: 38168054]
- Wright C, Rajpurohit A, Burke EE, Williams C, Collado-Torres L, Kimos M, Brandon NJ, Cross AJ, Jaffe AE, Weinberger DR, Shin JH. 2019. Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *BMC Genomics* 20: 513. doi:10.1186/s12864-019-5870-3. [PubMed: 31226924]
- Wu Y, Li Q, Zhang R, Dai X, Chen W, Xing D. 2021. Circulating microRNAs: Biomarkers of disease. *Clin Chim Acta* 516: 46–54. doi:10.1016/j.cca.2021.01.008. [PubMed: 33485903]
- Yao J, Wu DC, Nottingham RM, Lambowitz AM. 2020. Identification of protein-protected mRNA fragments and structured excised intron RNAs in human plasma by TGIRT-seq peak calling. *Elife* 9: e60743. doi:10.7554/eLife.60743. [PubMed: 32876046]
- Yin S, Kiong Ho C, Miller ES, Shuman S. 2004. Characterization of bacteriophage KVP40 and T4 RNA ligase 2. *Virology* 319: 141–51. doi:10.1016/j.virol.2003.10.037. [PubMed: 14967495]
- Yoshinari S, Liu Y, Gollnick P, Ho CK. 2017. Cleavage of 3'-terminal adenosine by archaeal ATP-dependent RNA ligase. *Sci Rep* 7: 11662. doi:10.1038/s41598-017-11693-0. [PubMed: 28912583]
- Zhuang F, Fuchs RT, Sun Z, Zheng Y, Robb GB. 2012a. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res* 40: e54. doi:10.1093/nar/gkr1263. [PubMed: 22241775]
- Zhuang F, Fuchs RT, Robb GB. 2012b. Small RNA expression profiling by high-throughput sequencing: implications of enzymatic manipulation. *J Nucleic Acids* 2012: 360358. doi:10.1155/2012/360358. [PubMed: 22778911]

**FIGURE 1.**

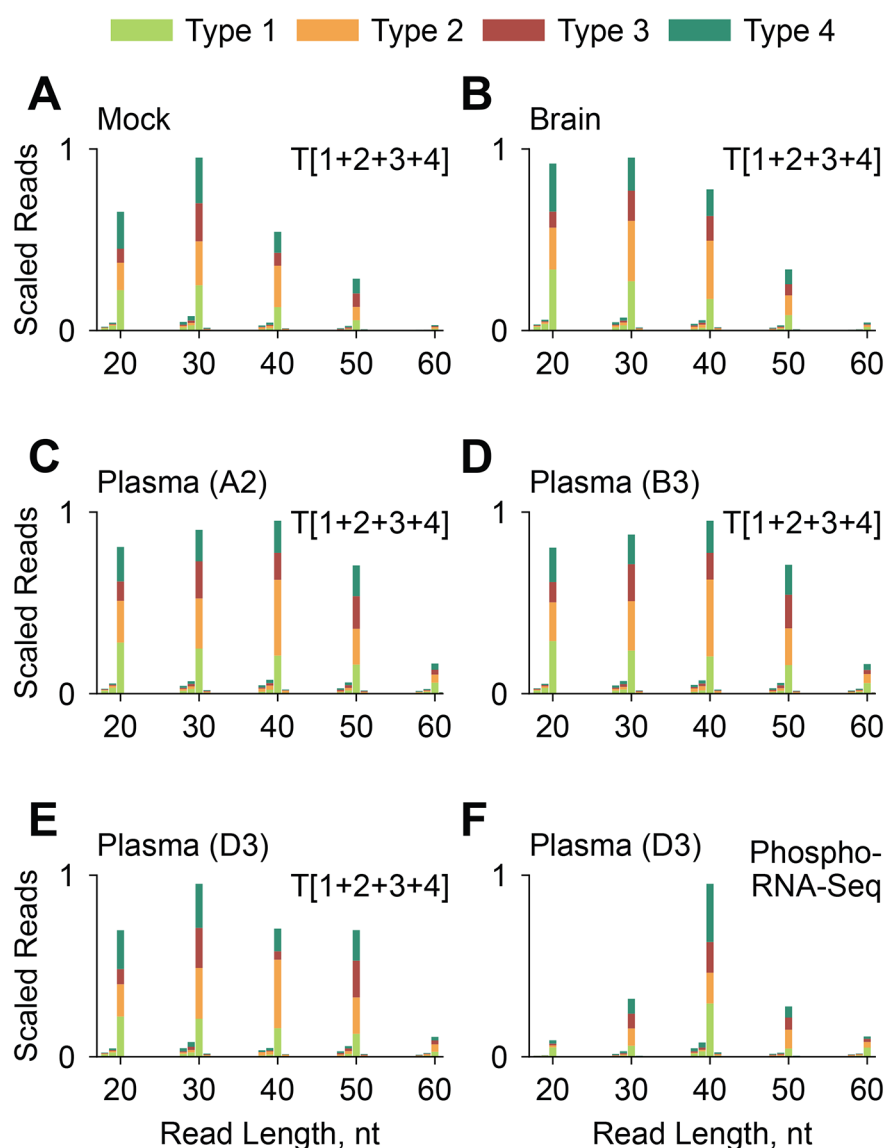
RNA Types representing small RNA molecules and RNA fragments (sRNAs) having different phosphorylation states of their ends (Types 1 through 4) or their circular form (Type 0) with no ends.

**FIGURE 2.**

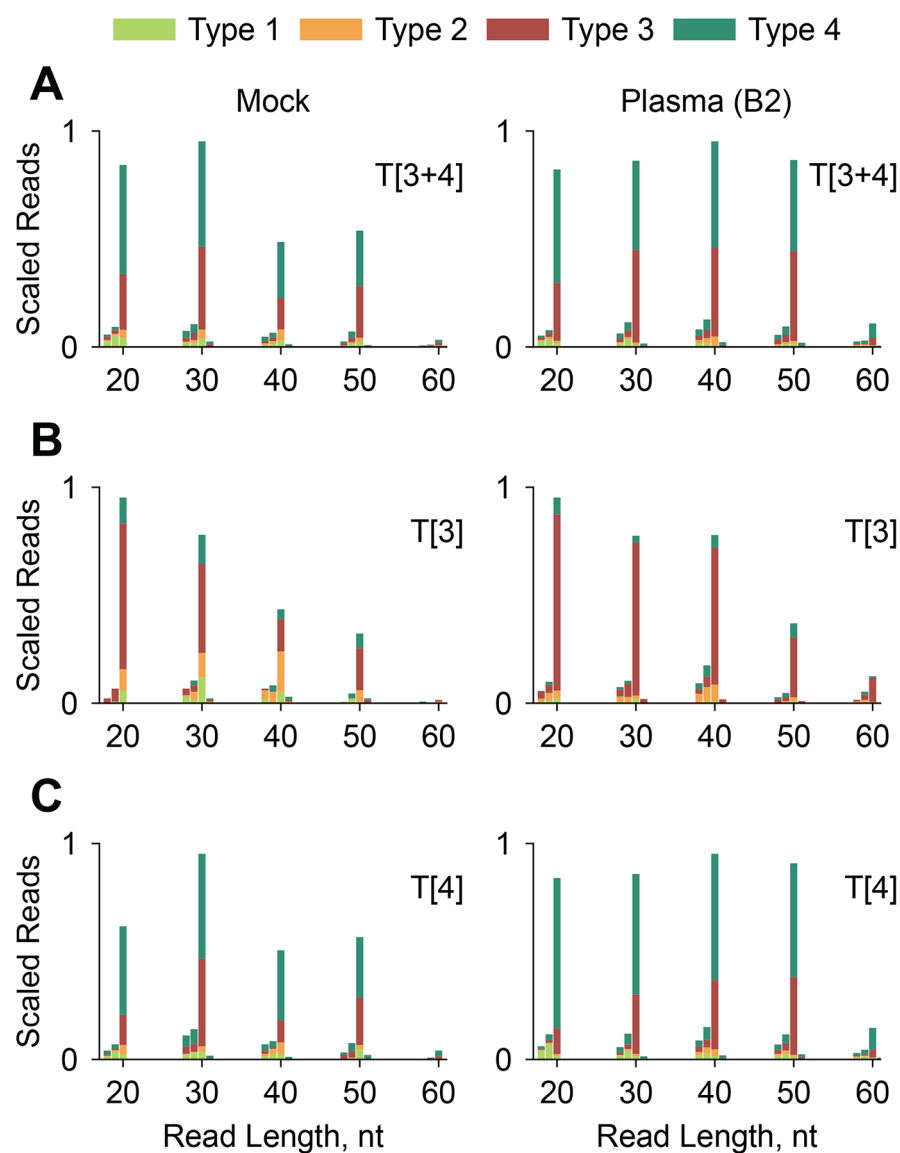
Sequencing length and RNA Type profiles for (A) stsRNAs in the absence of naturally occurring RNAs (mock controls); (B) stsRNAs spiked into brain RNA; and (C–D) stsRNAs spiked into representative A and B cohorts' plasma RNA samples for the libraries prepared by protocols T[1+2] (left panels) and T[2] (right panels).

**FIGURE 3.**

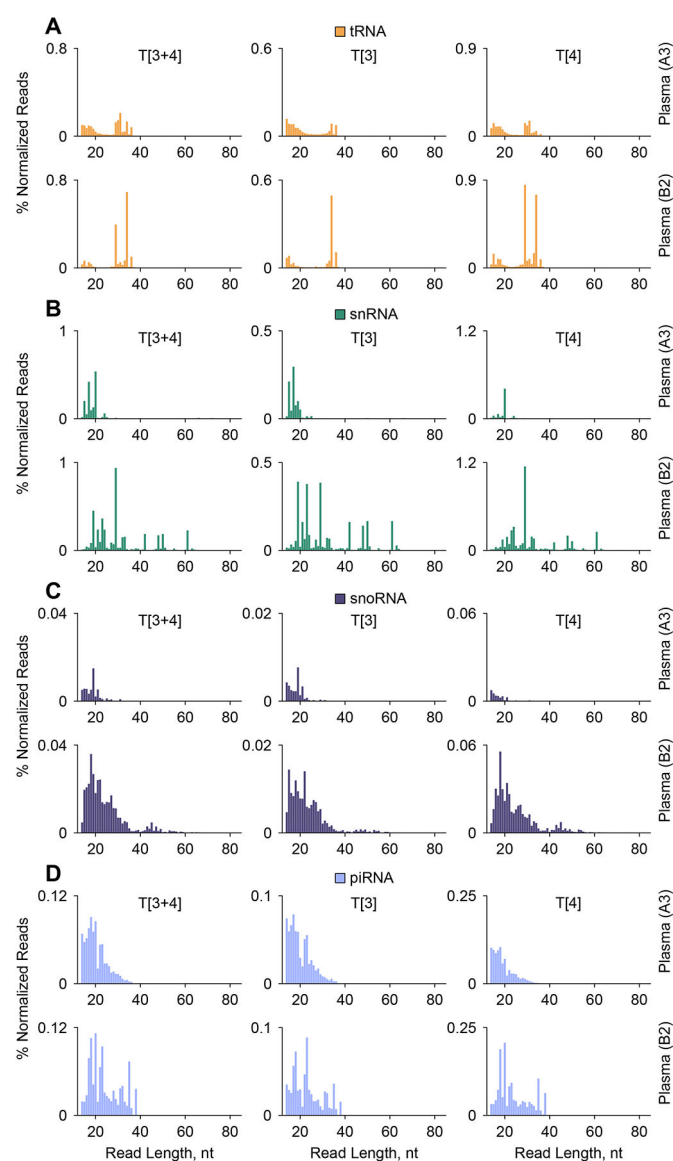
Sequencing length profiles for the main RNA classes of naturally occurring sRNAs found in brain RNA and in representative A and B cohorts' plasma RNA samples for libraries prepared by protocol T[2].

**FIGURE 4.**

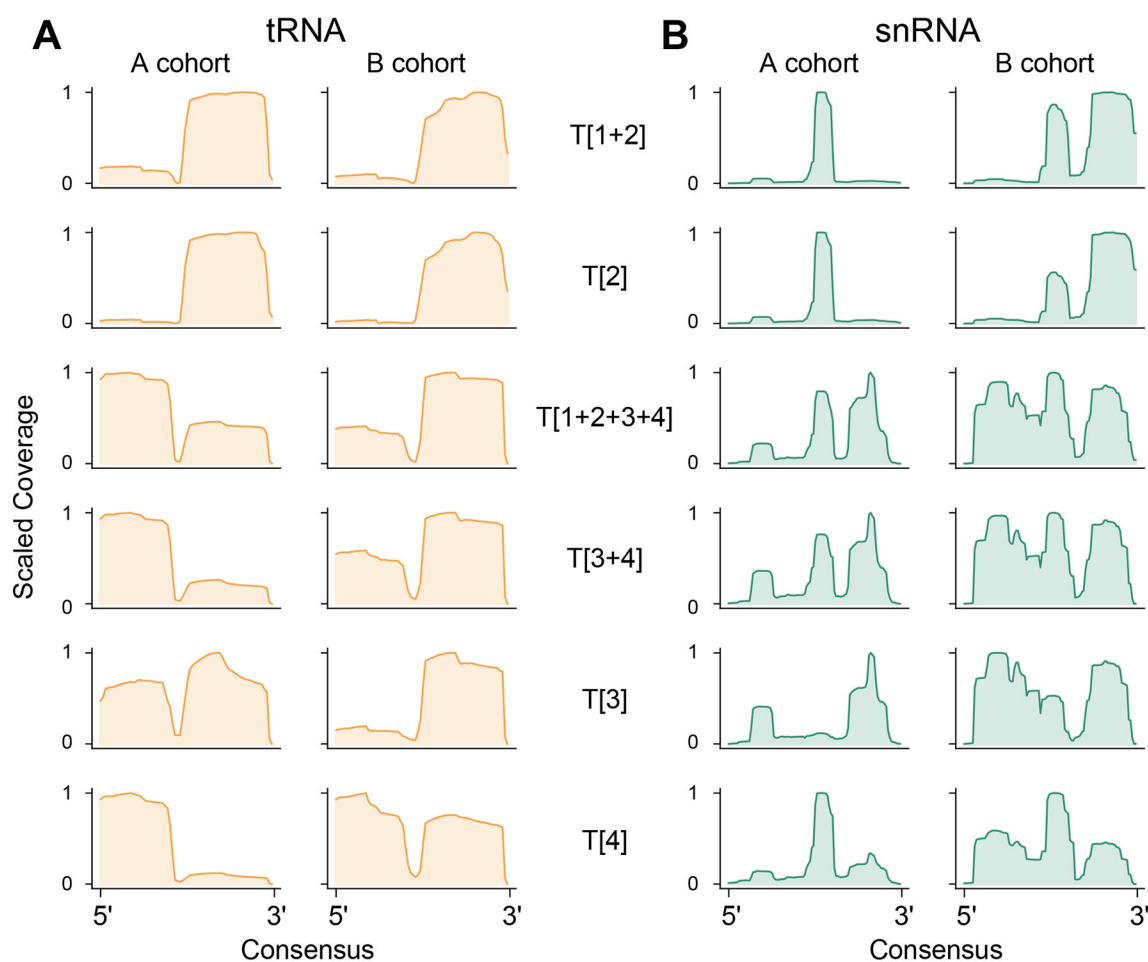
Representative sequencing length and RNA Type profiles for different stsRNAs contexts, including: (A) mock control lacking a naturally occurring sRNA background; and stsRNAs spiked into either (B) brain RNA or (C – F) representative A, B, and D cohorts' plasma RNA samples for libraries prepared either by (A – E) RiboMarker[®] protocol T[1+2+3+4] or (F) Phospho-RNA-Seq.

**FIGURE 5.**

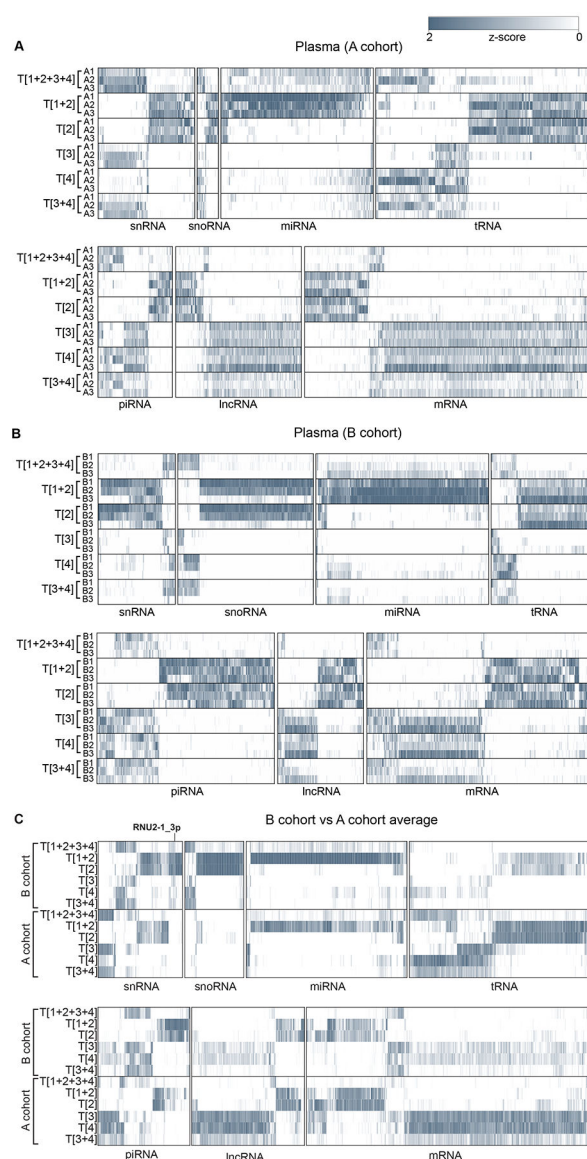
Representative sequencing length and RNA Type profiles for different stsRNA contexts, including: a mock control lacking a naturally occurring sRNA background (left column) and stsRNAs spiked into representative plasma (from B cohort as indicated) RNA sample (right column), for libraries prepared by RiboMarker[®] protocols (A) T[3+4], (B) T[3], or (C) T[4].

**FIGURE 6.**

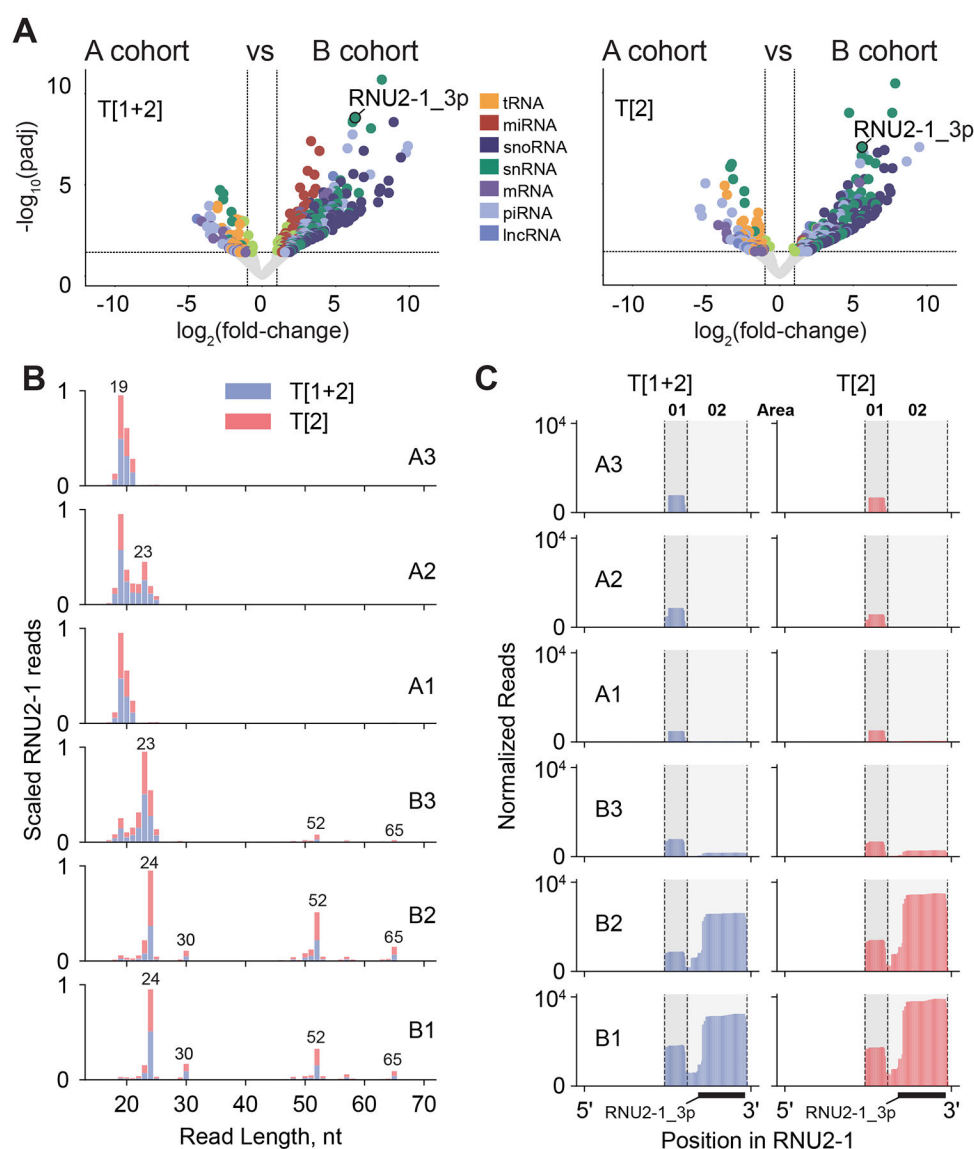
Comparison of sequencing length profiles for sRNAs of selected RNA classes (A) tRNA, (B) snRNA, (C) snoRNA, and (D) piRNA for representative A and B cohorts' plasma RNA samples prepared by RiboMarker® protocols T[3+4] (left column), T[3] (middle column), or T[4] (right column).

**FIGURE 7.**

Comparison of consensus sequencing read coverage profiles for all **(A)** tRNA and **(B)** snRNA molecules, using the average distribution in A cohort and B cohorts' plasma RNA samples for libraries prepared by all RiboMarker[®] protocols (indicated for each row).

**FIGURE 8.**

Hierarchically clustered heatmaps showing the relative abundance (row-wise z-score) of sRNA sequences aligned to significantly enriched RNA transcripts and/or their fragments of main RNA classes among RiboMarker® protocols (DESeq2 LRT; padj < 0.05) detected in the individual plasma RNA samples of (A) A cohort; (B) B cohort; and (C) comparison of average z-scores calculated for all individual plasma RNA samples among each A and B cohort for libraries prepared by indicated RiboMarker® protocols. Darker rows indicate a higher relative abundance of the corresponding RNA transcripts.

**FIGURE 9.**

Comparison of differentially enriched sRNAs detected in the plasma RNA samples from A and B cohorts. **(A)** Volcano plots showing the significant changes ($|\log_2\text{fold-change}| > 1$, $\text{padj} < 0.05$) in the abundance of sRNAs indicated by dots, whose colors correspond to transcripts of the indicated RNA classes they were derived from. Average data calculated for all individual plasma RNA samples among each A and B cohort for libraries prepared by protocols T[1+2] and T[2] are shown in the left and right panels, respectively. **(B)** Overlapped sequencing length profiles for sRNAs derived from the RNU2-1 snRNA for libraries prepared using either protocols T[1+2] (blue) or T[2] (red) for each individual plasma RNA sample from A and B cohorts. **(C)** Read coverage profiles for those reads mapping to the RNU2-1 snRNA using either protocols T[1+2] (blue; left) or T[2] (red; right) for individual plasma RNA samples of A and B cohorts. The region corresponding to the counts associated with the

RNU2-1_3p feature in panel (A) is underlined, as well as the designated areas “01” and “02” of the RNU2-1 snRNA transcript whose sequences are shown in supplementary Fig. S26.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 1.

Selected RNA Type-specific protocols for the preparation of RiboMarker® sequencing libraries

RiboMarker® Protocols	T[1+2]	T[2]	T[1+2+3+4]	T[3+4]	T[3]	T[4]
RNA Types in the sample	1, 2, 3, 4	1, 2, 3, 4	1, 2, 3, 4	1, 2, 3, 4	1, 2, 3, 4	1, 2, 3, 4
Pretreatment # 1	–	Rnl1 + Rnl2 ligases (+ATP)	PNK (–ATP) pH 6.0	PNK, 3'-minus (+ATP) pH 7.6	Terminator 5'-p - dependent exonuclease	RtcB ligase
RNA Types conversion	–	1 → 0	3 → 2 4 → 1	2 → 1 3 → 4	Degrade 1, 4	3 → 0
Pretreatment # 2	–	–	–	Rnl1 + Rnl2 ligases (+ATP)	PNK, 3'-minus (+ATP) pH 7.6	PNK, 3'-minus (+ATP) pH 7.6
RNA Types conversion	–	–	–	1 → 0	2 → 1 3 → 4	2 → 1
Pretreatment # 3	–	–	–	PNK (–ATP) pH 6.0	Rnl1 + Rnl2 ligases (+ATP)	Rnl1 + Rnl2 ligases (+ATP)
RNA Types conversion	–	–	–	4 → 1	1 → 0	1 → 0
Pretreatment # 4	–	–	–	–	PNK (–ATP) pH 6.0	PNK (–ATP) pH 6.0
RNA Types conversion	–	–	–	–	4 → 1	4 → 1
Library preparation Protocol	T[1+2]	T[1+2]	T[1+2]	T[1+2]	T[1+2]	T[1+2]
RNA Types depleted from library	3, 4	1, 3, 4	None	1, 2	1, 2, 4	1, 2, 3
RNA Types enriched in library	1, 2 (3'-OH Types)	2	1, 2, 3, 4 (All Types)	3, 4 (3'-P Types)	3	4

TABLE 2.

Sequences of spike-in RNAs (stsRNAs) having barcodes (underlined> specific to each RNA Type.

RNA Type	Length (nt)	Sequences and Ends (5' → 3')
1	20	pNNNNGAGCGUGAU <u>CAG</u> NNNN
1	30	pNNNNUAUCAGUGCGUGAU <u>GUCUCAG</u> NNNN
1	40	pNNNNCAUCGAUCUGAUCCGUGAU <u>ACGAAGUCUACUG</u> NNNN
1	50	pNNNNAUGUCACUAUGAGUCACACGUGAU <u>AUCACACUCUACUACUC</u> NNNN
1	60	pNNNNGACAGUGAUGACAACACACUCAGCGUGAU <u>GUACUCAUAGCAAGCGACACUA</u> NNNN
2	20	NNNNGAGACAUCG <u>CAG</u> NNNN
2	30	NNNNUAUCAGUGACAUCG <u>GUCUCAG</u> NNNN
2	40	NNNNCAUCGAUCUGAUCACAUCGACGAAGUCUACUGNNNN
2	50	NNNNAUGUCACUAUGAGUCACAACAUCGAUCACACUCUACUACUCUNNNN
2	60	NNNNGACAGUGAUGACAACACACUCAGACAUCGGUACUCAUAGCAAGCGACACUAUNNNN
3	20	NNNNGAGGCCUA <u>ACAG</u> NNNNp
3	30	NNNNUAUCAGUGGCCUA <u>AGUCUCAG</u> NNNNp
3	40	NNNNCAUCGAUCUGAUCGCCUA <u>ACGAAGUCUACUG</u> NNNNp
3	50	NNNNAUGUCACUAUGAGUCACAGCCUA <u>AUCACACUCUACUACUC</u> NNNNp
3	60	NNNNGACAGUGAUGACAACACACUCAGGCCUA <u>AGUACUCAUAGCAAGCGACACUA</u> NNNNp
4	20	pNNNNGAGUGGUC <u>ACAG</u> NNNNp
4	30	pNNNNUAUCAGUGUGGUC <u>AGUCUCAG</u> NNNNp
4	40	pNNNNCAUCGAUCUGAUCUGGUC <u>ACGAAGUCUACUG</u> NNNNp
4	50	pNNNNAUGUCACUAUGAGUCACAUGGUC <u>AUCACACUCUACUACUC</u> NNNNp
4	60	pNNNNGACAGUGAUGACAACACACUCAGUGGUC <u>AGUACUCAUAGCAAGCGACACUA</u> NNNNp