2015-1

An optimized kit-free method for making strand-specific deep sequencing libraries from RNA fragments

Erin E. Heyer
University of Massachusetts Medical School

Et al.

Published online 12 December 2014

Nucleic Acids Research, 2015, Vol. 43, No. 1 e2
doi: 10.1093/nar/gku1235

An optimized kit-free method for making strand-specific deep sequencing libraries from RNA fragments

Erin E. Heyer[1,2,3], Hakan Ozadam[1,2,3], Emiliano P. Ricci[1,2,3], Can Cenik[1,2,3,4] and Melissa J. Moore[1,2,3,*]

[1]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA, [2]RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA, [3]Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA and [4]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

Received February 03, 2014; Revised October 09, 2014; Accepted November 10, 2014

ABSTRACT

Deep sequencing of strand-specific cDNA libraries is now a ubiquitous tool for identifying and quantifying RNAs in diverse sample types. The accuracy of conclusions drawn from these analyses depends on precise and quantitative conversion of the RNA sample into a DNA library suitable for sequencing. Here, we describe an optimized method of preparing strand-specific RNA deep sequencing libraries from small RNAs and variably sized RNA fragments obtained from ribonucleoprotein particle footprinting experiments or fragmentation of long RNAs. Our approach works across a wide range of input amounts (400 pg to 200 ng), is easy to follow and produces a library in 2–3 days at relatively low reagent cost, all while giving the user complete control over every step. Because all enzymatic reactions were optimized and driven to apparent completion, sequence diversity and species abundance in the input sample are well preserved.

INTRODUCTION

In cells, all RNA molecules interact with RNA binding proteins (RBPs) to form ribonucleoprotein particles (RNPs). An ever-increasing number of methodologies employ deep sequencing to map these protein-RNA interaction sites transcriptome-wide. Such techniques include ultraviolet-crosslinking methods (e.g. CLIP, PAR-CLIP; (1,2)) to map the ribonucleotides directly in contact with an individual RBP and RNP footprinting (e.g. Ribo-Seq, RIPiT-Seq; (3,4)) to map the occupancy sites of larger complexes. Many projects in our laboratory are focused on transcriptome-wide RNP footprint analysis (5–7). Depending on the complex being examined and the RNA fragmentation method utilized (e.g. RNase or sonication), bound RNA fragments can range from 10 to 200 nucleotides (nts). Therefore, we require a strand-specific library generation method that works for diverse RNA lengths, faithfully preserves their relative abundances in the original sample and excludes any contaminating DNA fragments.

Multiple commercial kits currently exist for strand-specific library preparation, but most are intended to capture either long RNAs (e.g. RNA-Seq) or short RNAs (e.g. miRNA-Seq), but not both. Further, commercial kits are regularly updated with new preparation methods. Because preparation method is the primary source of variability between deep sequencing libraries (8), quantitative comparisons are best done between identically generated libraries (i.e. with a single commercial kit version). However, the expense of commercial kits (and remaking libraries as new kits appear and older versions are phased out) is cost prohibitive for many academic laboratories. We therefore set out to develop an optimized, strand-specific RNA library preparation protocol that utilizes commonly available reagents and works over a wide range of input amounts. We also wanted an approach that can be used to capture full-length RNP footprints as well as map sites of reverse transcriptase stalling (e.g. sites of RNA-protein crosslinking from CLIP experiments or abasic/alkylated sites).

All current library preparation methods utilize enzymes to capture nucleic acid fragments by appending 5′ and 3′ adaptor sequences. Enzymes have inherent substrate preferences that are most significant at low substrate concentrations (k_{cat}/K_m conditions) and at short reaction times (9). For ligation reactions, low temperatures can favor capture of sequences capable of base pairing with the adaptor (10). Low temperatures can also disfavor capture of sequences containing internal secondary structures. Many published library preparation protocols are suboptimal for

*To whom correspondence should be addressed. Tel: +1 508 856 8014; Fax: +1 508 856 1002; Email: melissa.moore@umassmed.edu

one or more of these factors, resulting in differential capture of small RNAs (e.g. miRNA-Seq; (10–12)) and highly non-uniform ('peaky') coverage of long RNAs (e.g. RNA-Seq of RNA Pol II transcripts; (13)). For these reasons, we decided to re-examine 5′- and 3′-end capture conditions, with the goal of driving every reaction to completion.

Here, we present the detailed protocol for strand-specific RNA library preparation currently in use in our laboratory, as well as the titration and time course data we used to optimize each step. Also presented are deep sequencing data on (i) the effects of time and temperature on initial 3′-end capture and (ii) capture uniformity analysis for an equimolar pool of 29 miRNAs. Taken together, these data show that our method faithfully preserves fragment diversity and abundance in complex starting mixtures and is minimally affected by fragment sequence or folding potential.

MATERIALS AND METHODS

Gel analysis

All acrylamide gels were prepared using AccuGel reagents (National Diagnostics). Ligation samples were prepared in an equal volume of 2× denaturing load buffer (12% Ficoll Type 400-DL, 7 M Urea, 1× TBE, 0.02% Bromophenol Blue, 0.02% Xylene Cyanol), denatured for 5 min at 95°C and cooled on ice prior to loading on denaturing 15% polyacrylamide (19:1)-8 M Urea-1× TBE gels. Reverse transcription (RT) samples were diluted in one-third volume of 3× denaturing load buffer (18% Ficoll Type 400-DL, 10.5 M Urea, 1.5× TBE, 0.02% Bromophenol Blue, 0.02% Xylene Cyanol), denatured for 5 min at 95°C, and analyzed on 10% denaturing polyacrylamide gel electrophoresis (PAGE) gels. Circularization reactions were prepared similarly to ligation reactions and analyzed on 10% denaturing PAGE gels. Polymerase chain reaction (PCR) products for gel analysis were mixed with 5× non-denaturing load buffer (15% Ficoll Type 400-DL, 1× TBE, 0.02% Bromophenol Blue, 0.02% Xylene Cyanol) before separation on native 8% PAGE gels. PCR products to be sequenced were similarly prepared and analyzed on the Double Wide Mini-Vertical system (C.B.S. Scientific) to limit the amount of heat denaturation. Gels were either exposed to a phosphorimager screen (Amersham Biosciences) or stained with SYBR Gold (Invitrogen) prior to visualization on a Typhoon Trio (Amersham Biosciences). Quantifications were performed with ImageQuant (GE Healthcare).

3′-adaptor ligation

Indicated amounts of either 5′-^{32}P-labeled N24 RNA oligonucleotide (Dharmacon) or 28-mer oligonucleotide (5′-AUGUACACGGAGUCGACCCGCAACGCGA-3′; IDT) were ligated to preadenylated adaptor mirCat-33 (5′-rApp TGGAATTCTCGGGTGCCAAGGddC-3′; IDT) or EH-preaden (5′-rApp NNNNTGGAATTCTCGGGTG CCAAGGddC-3′; IDT) using T4 RNL2 Tr.K227Q (NEB) with the conditions described in this paper. Due to the high viscosity of 50% PEG 8000, we found that low retention filter tips aided consistent pipetting while simultaneously preventing sample cross-contamination. Ligation efficiencies were calculated by dividing the quantified pixel signal of ligated RNA by the total amount of RNA signal (bands corresponding to both ligated and unligated RNA) in each lane, and multiplying by 100.

Reverse transcription

RT was performed with gel purified RT primers 5′-pGG-B-AGATCGGAAGAGCGTCGTGTAGGGAAAGAG TGT-SP18-CTCGGCATTCCTGCTGAACCGCTC TTCCGATCT-CCTTGGCACCCGAGAATTCCA-3′, where B indicates a 5-nt barcode of sequence ATCAC, CGATG, TAGCT, GCTCC, ACAGT, CAGAT, TCCCG, GGCTA, AGTCA, CTTGT, TGAAT or GTAGA. RT products were detected by incorporating $^{-32}$P-dCTP in the reaction. RT products intended for circularization were gel purified. For the data in Figures 4 and 5, we eluted the cDNA from crushed gel pieces in 300 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA) during an overnight incubation at room temperature with constant rotation; eluted material was ethanol precipitated before circularization. We have since modified our approach to increase elution yield by eluting in TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) and incubating at 37°C overnight with constant rotation. With this buffer, we can concentrate the eluate (either by butanol extraction or SpeedVac) before precipitating the sample in a single tube.

Circularization efficiency and PCR amplification

Circularization reactions were performed on gel-purified RT product as described in the text. The single-stranded DNA input was either body-labeled with $^{-32}$P-dCTP in the RT reaction or end-labeled in an exchange reaction with ^{32}P-$^{-}$ATP. Circularized RT product was separated from non-reactive, linear RT product on 10% denaturing PAGE gels, and the gels were exposed and quantified as described. The amount of circularization was determined by quantifying the pixel signal corresponding to the circularized product and dividing that value by the total pixel signal corresponding to the circularized product plus the remaining linear input, and multiplying by 100.

PCR amplification from the circularized RT product was performed with KAPA HiFi Library Amplification Kit (Kapa Biosystems) according to manufacturer's instructions, except where otherwise noted. All PCR products were analyzed on native 8% PAGE gels and quantified as described above. Samples to be sequenced were excised and gel extracted as described for RT products, precipitated and quantified by gel analysis before sample submission.

N24 library construction and analysis

N24 libraries were constructed from 2 pmol of N24 RNA oligo using the optimized conditions shown in Supplementary Table S1, except for the described variations in 3′ ligation conditions. In one case (22°C 6 hr library), a minute amount of 28-mer oligo was added. All libraries were amplified with 7 PCR cycles and gel purified prior to sequencing on a single Illumina HiSeq2000 lane (Genewiz).

Deep sequencing data were analyzed with custom scripts unless otherwise noted. Data were parsed into individual

libraries by 5′ barcode, allowing 1 mismatch. The 3′ adaptor sequence was removed from all libraries allowing 3 mismatches. Once individual sequence reads were identified, read lengths were calculated. All subsequent analysis utilized only 24 nt reads. For each library, we calculated the observed nt frequencies at each of the 24 positions. To determine expected values, we used the data across positions 5–20 from all libraries and fitted least squares lines to the frequency pattern for each nt. The equations for the lines fits yielded the expected nt frequencies at all 24 positions. The chi-square statistic was calculated for each library by summing [(observed nt count - expected nt count)^2/(expected nt count)] across all four nts at each N 24 position.

PhiX reads were identified if they mapped to the PhiX 174 genome with a maximum of 6 errors within the 51 sequenced nts. Mismatches were identified and counted if the sequenced nt was different than the PhiX 174 genome sequence. Mismatch frequencies were calculated by dividing the mismatch counts at each position by the total number of PhiX reads. For analysis of nt distribution across ribosome footprints (6), all 26–30 nt reads were selected and aligned by their 3′ ends; nt frequencies were calculated by dividing the observed nt count at each position by the total number of reads.

miRNA library construction and analysis

Libraries were constructed from either 1 pmol or 50 fmol of an equimolar mix of 29 miRNAs (14) according to the optimized conditions shown in Supplementary Table S1. For each input amount, the ligation was performed with either the fixed or N4 preadenylated 3′-adaptor. Libraries were pooled and sequenced on a single MiSeq lane. Deep sequencing data were parsed into individual libraries by 5′ barcode using cutadapt version 1.3 (15), allowing 1 mismatch. Reads were mapped to reference sequences using a custom script which (i) required that the 3′ adaptor be present in the read and (ii) only counted reads mapping to reference miRNA sequences with 0 mismatches. Additionally, we counted the reads with 5′ or fewer non-templated 5′ terminal additions and 5′ or fewer 5′-terminal deletions. Observed miRNA frequencies (F_{obs}) were calculated using the total number of reads for each miRNA (including 5′ terminal additions and subtractions). The expected frequency (F_{exp}) for each miRNA is 1/29 or 0.0345. Coefficients of variation (CV) were calculated by dividing standard deviation (miRNA counts) by the mean (miRNA counts). Terminal transferase activity was assessed by dividing total miRNA reads in each 5′ addition bin by the total full-length miRNA reads in each library. Free energy values from in silico folding were calculated using the Vienna RNA Package v. 2.1.7 using the -T 30 parameter to obtain structure predictions at 30°C (16).

RESULTS

Protocol design

To generate strand-specific deep sequencing libraries, both ends of the captured RNA must be appended to fixed sequences (adaptors) to enable primer hybridization for amplification and sequencing. These adaptors generally correspond to the forward and reverse primer sequences used for clonal cluster amplification on the desired sequencing platform. All strand-specific RNA-Seq and small RNA library preparations published to date capture the 3′-end in one of the following ways: (i) RT of full length or fragmented RNAs with oligo-dT and/or random hexamers, or a longer DNA primer containing a 3′ randomized region (17–21); (ii) polyA tailing of RNA fragments followed by RT with an anchored oligo-dT 3′-end sequence (3,8); or (iii) direct 3′-end adaptor ligation (22–24). Disadvantages of random hexamer RT include the introduction of mutations at the point of primer hybridization plus capture biases resulting from differential hybridization efficiencies on different sequences (25). Random hexamer RT is also not an option for small RNAs. In our hands, polyA tailing of fragmented RNA samples proved inconsistent (data not shown). Therefore, we decided to adopt a 3′-end adaptor ligation approach widely used in the small RNA field (23) - direct ligation of a preadenylated DNA adaptor to the 3′-end of RNA fragments using RNA ligase (Figure 1, Step 1). We chose to use a truncated and mutant form of T4 RNA Ligase 2 (RNL2 Tr. K227Q) because published reports indicated it has less substrate bias and produces fewer side products than the full-length wild-type enzyme (12,26), and RNL2 is known to be less affected by nt identity at the ligation site than T4 RNA Ligase 1 (27). Following 3′ adaptor ligation, a highly efficient method for appending the 5′ adaptor is to reverse transcribe the RNA from the 3′ adaptor with an RT primer containing the 5′ adaptor sequence at the other end and then circularize the resulting single-stranded cDNA using CircLigase (3) (Figure 1, Steps 2 and 4). A long flexible linker (Spacer 18, an 18-atom hexa-ethyleneglycol spacer) is placed between the fixed adaptor sequences to minimize structural constraints for circularization and preclude the possibility of rolling circle PCR (28).

A common strategy for reducing deep sequencing costs is to 'barcode' individual libraries so that they can be mixed together and sequenced in a single lane. Barcodes consist of 2–10 unique nts appended either 5′ or 3′ to the captured sequences (29), and ideally differ by more than 2 nts so as to minimize incorrect library identification due to sequencing errors. Barcodes can be placed in one of the adaptors (30,31) or in the reverse PCR primer (30), or they can be ligated to the double-stranded library post-PCR amplification (32). Barcode incorporation immediately downstream of the forward sequencing primer hybridization site allows both the barcode and the adjacent captured fragment to be decoded in one single-end sequencing reaction. In theory, barcodes can be appended to either end of the captured fragment. However, RNL2 ligation efficiency is significantly affected by the 3′ adaptor sequence - therefore, placement of the barcode at the 5′-end of the 3′ adaptor can result in significant and different sequence biases dependent on the barcode (11,33). Because we were able to find conditions under which cDNA circularization is quantitative (see below), we chose to place our barcodes at the 3′-end of the 5′ adaptor (i.e. between the forward primer sequence and the captured sequences). Nonetheless, to minimize any confounding effects of varying the nt composition at the site of circularization, we introduced two guanine residues at the 5′-end of

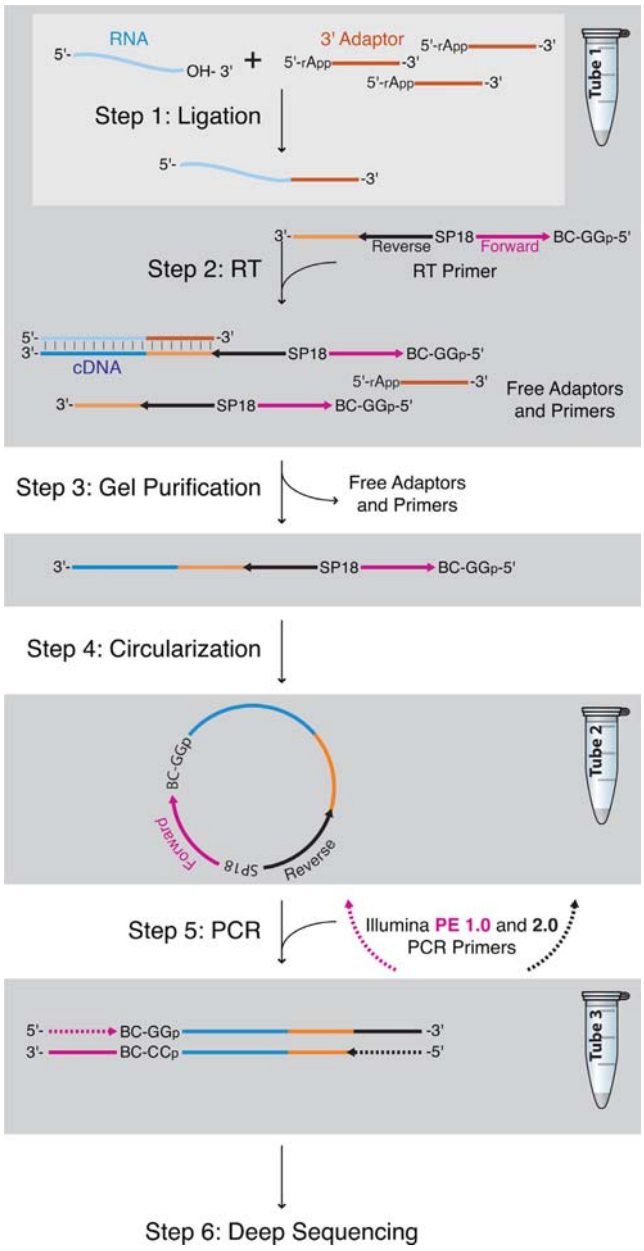

Figure 1. Method overview. Step 1: Ligation. RNA, shown in blue, is ligated to a preadenylated DNA adaptor to form a RNA–DNA hybrid. In the same tube, RT is performed (Step 2). The RT primer contains both the reverse and forward priming sequences for Illumina sequencing, as well as a barcode to uniquely identify the sample. Step 3: The RT product is gel purified, removing unligated adaptors and unextended RT primers from the sample. Step 4: The gel purified RT product is circularized, forming a template for PCR (Step 5). The PCR product is then purified and used for deep sequencing (Step 6).

each RT primer so that the nts interacting with CircLigase would be the same regardless of barcode.

A final consideration for making strand-specific cDNA libraries is the quantity of starting material required. Major factors leading to material loss during library preparation are the number of gel purification steps and the number of different surfaces (i.e. tips and tubes) with which the sample comes in contact. Thus, we opted for a protocol wherein the ligation (Step 1) and RT (Step 2) were carried out in a single tube without any cleanup or buffer exchange in between, and the sample is only subjected to a single gel purification (Step 3) after RT.

Protocol optimization

For optimization of each step, we used a pool of randomized RNA 24mers (N24) to mimic the diversity of sequences in a biological sample. Ligation reactions were visualized using 5′-end ^{32}P-labeled RNAs. RT products were visualized by including α-^{32}P-dCTP in the RT reaction. Circularization reactions were visualized using either body-labeled or 5′-end-labeled RT products.

Step 1: preadenylated 3′ adaptor ligation. When we initiated this project, the manufacturer's (NEB) suggested conditions for RNL2 Tr. K227Q ligation reactions were 500 nM single-stranded RNA, 1 μM 3′ adaptor, 10 U/μl enzyme and 15% w/v PEG 8000 in 1× reaction buffer at 16°C overnight. As our goal was to create a robust protocol that could be successfully employed over a wide range of RNA input concentrations, we set out to explore the limits of these parameters (Figure 2). For all experiments below, we pre-mixed the RNA and 3′-adaptor in water and incubated this mixture at 65°C for 10 min prior to enzyme addition.

Ligation efficiency depends on successful collision of multiple components. Such collisions can be increased by molecular crowding agents (e.g. PEG) and/or dehydrating co-solutes (e.g. dimethylsulfoxide (DMSO)), and published 3′ adaptor ligation protocols vary with regard to PEG 8000 and DMSO inclusion (34–38). Consistent with a recent report that 25% PEG 8000 enhances ligation efficiency (see Figure 4B in (38)), we found that 25% PEG 8000 resulted in near complete N24 ligation at 16°C O/N (Figure 2A). However, increasing DMSO had no effect, regardless of PEG 8000 absence or presence (Figure 2B). Thus, all subsequent ligation reactions included 25% PEG 8000 but no DMSO.

We next titrated preadenylated 3′-adaptor, N24 and enzyme concentrations. Using two different N24 concentrations, near complete ligation was observed at all adaptor concentrations above 130 nM (Figure 2C). At 470 nM adaptor, ligation was highly efficient with N24 concentrations above 50 nM (Figure 2D) and enzyme concentrations above 6 U/μl (Figure 2E). A greater dependence of ligation efficiency on enzyme concentration at 10 nM N24 does suggest, however, that additional enzyme will increase yields for very dilute RNA samples (39).

Published reports using T4 RNA ligases for library preparation employ a wide range of reaction times (1 h to overnight) and temperatures (5°C–37°C) (1,23,34,37,40–43). However, colder temperatures should stabilize both intra- and inter-molecular secondary structures, potentially biasing ligations against internally structured RNAs and toward RNA sequences that partially base pair with the 3′-adaptor (10–11,27). Higher temperatures should alleviate these issues, but could decrease enzyme stability and increase RNA degradation. Using our N24 pool, we assessed ligation efficiencies across a range of incubation times and temperatures (Figure 2F). Both 4°C and 37°C yielded poor

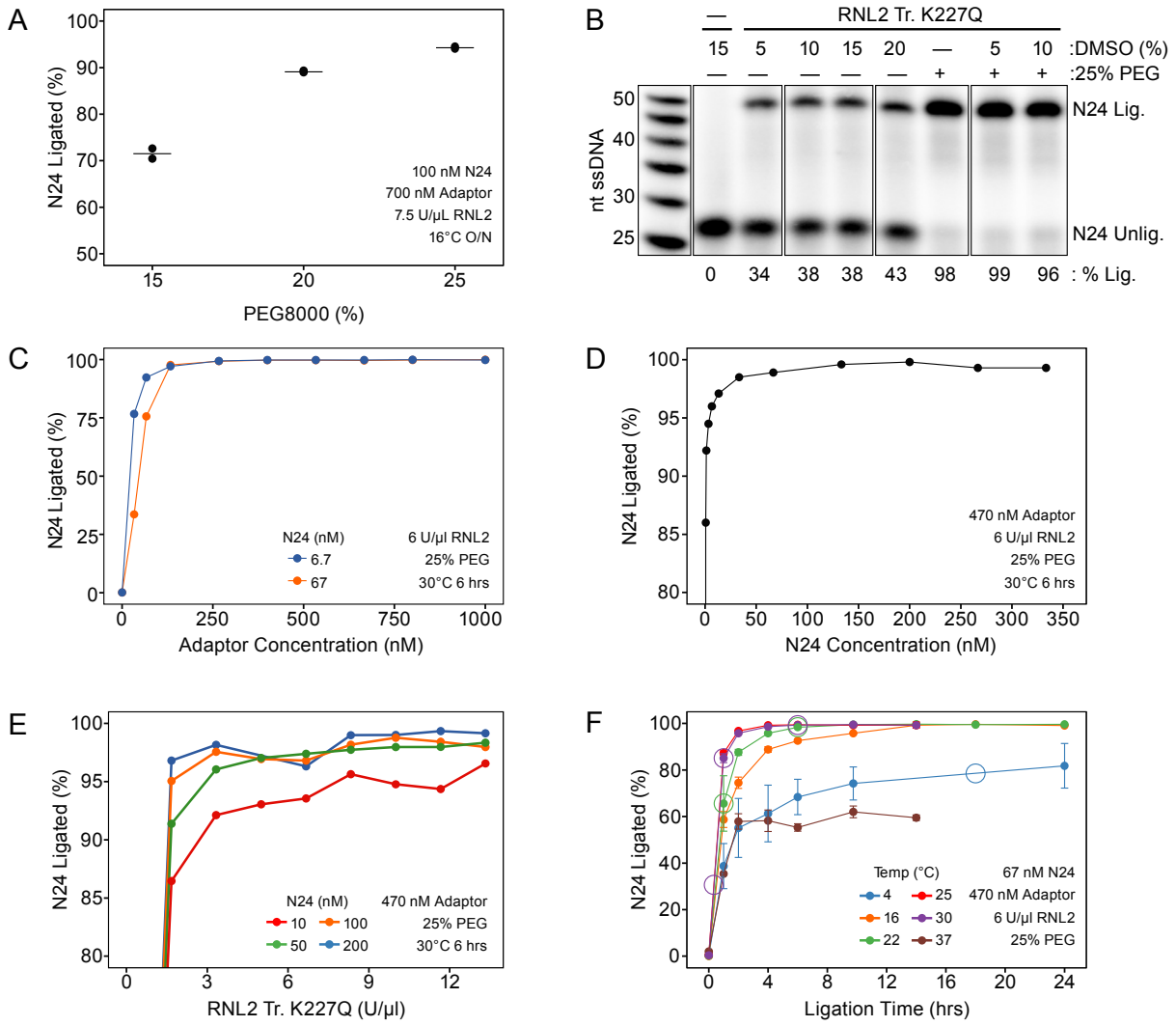

Figure 2. 3′ adaptor ligation optimization. (A) Ligation efficiency versus % PEG 8000 (w/v) (n = 2; black line, mean). (B) Comparison of DMSO and PEG as ligation enhancers. Absence or presence of indicated species are indicated by – and +; ligation efficiencies are indicated below each lane. N24 RNA was 5′-end labeled with ^{32}P- -ATP. (C) Ligation efficiency versus 3′-adaptor concentration (n = 1). (D) Ligation efficiency versus N24 concentration (n = 1). (E) Ligation efficiency versus RNL2 concentration at four different N24 RNA concentrations (n = 1). (F) Ligation efficiency versus time and temperature (n = 3; error bars, standard deviation). Circles indicate ligation conditions for N24 libraries. In all panels, data were generated by quantification of denaturing polyacrylamide gels similar to that shown in panel B; ligation efficiency = (ligated RNA–DNA product)/(unligated RNA + ligated RNA–DNA product) in each lane.

ligation efficiencies at all incubation times. Using radioactively labeled RNA, we determined that the lower yields at 37°C were not due to increased RNA degradation (data not shown); rather, the plateau reached after 2 h suggests that enzyme is unstable at 37°C. All reactions incubated between 16°C and 30°C ultimately resulted in near complete ligation. However, the 16°C and 22°C reactions took longer to reach completion (10–14 h) than did the 25°C and 30°C reactions (4–6 h).

Based on all of the above data, we adopted the following as our standard ligation reaction conditions: 470 nM adaptor, 50–330 nM RNA, 6 U/µl RNL2 K227Q, 1× RNL2 reaction buffer (from NEB: 50 mM Tris-HCl, pH 7.5 @ 25°C, 10 mM $MgCl_2$, 1 mM DTT) plus an additional 1 mM DTT to ensure a reducing environment, incubated for 6 h at 30°C and then 20 min at 65°C (to heat inactivate the enzyme). These conditions yield efficient ligation over the wide range of RNA fragment lengths we generally obtain when footprinting endogenous RNP complexes (4–6).

Step 2: reverse transcription. A number of high fidelity reverse transcriptases are commercially available. For our purposes, we wanted an enzyme that produced a high yield of full-length product with minimal side products when added directly to the heat-inactivated/diluted 3′-adaptor ligation reaction from Step 1. We tested Accuscript (Agilent), AMV RT (Finnzymes), Superscript III (Invitrogen) and Transcriptor (Roche) (Figure 3A). In all cases, ligation reactions were diluted and supplemented with either (i) the appropriate amount of manufacturer-supplied 5× or 10× RT buffer or (ii) the same buffer minus $MgCl_2$ (as the Step 1 reaction already contains $MgCl_2$, and concentrations of

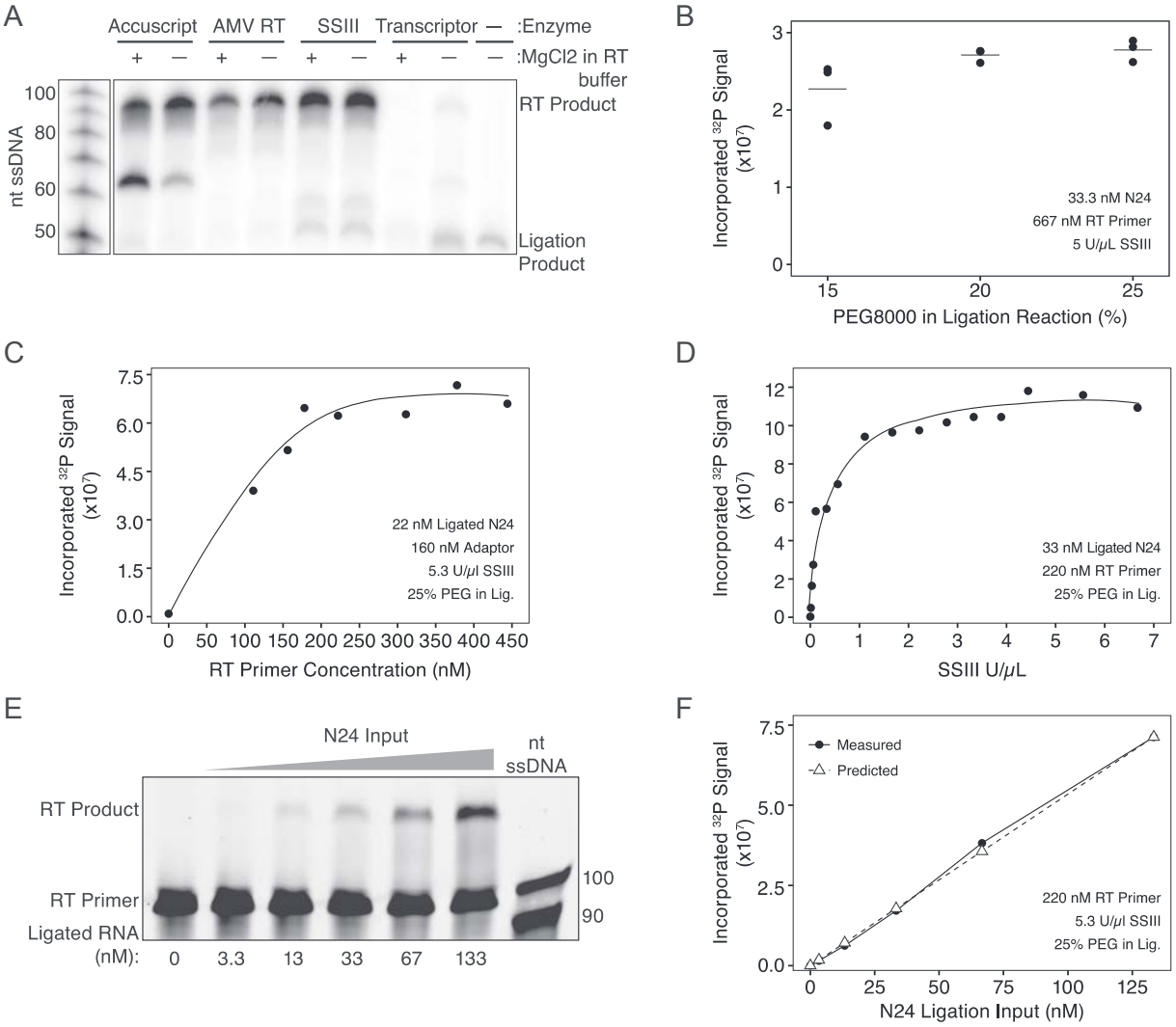

Figure 3. RT optimization. (A) Comparison of high-fidelity reverse transcriptases for the amount of RT product generated ± $MgCl_2$ in the RT buffer. Absence or presence of indicated species are indicated by – and +. (B) RT product signal versus % PEG 8000 (w/v) in the ligation reaction (n = 3; black line, mean). (C) RT product signal versus RT primer concentration (n = 1). (D) RT product signal versus SSIII concentration (n = 1). (E) RT product signal varies with RNA input concentration, ranging from 3.3 nM (lane 2) to 133 nM (lane 6). (F) RT product signal versus RNA input concentration (n = 1). Replicate of panel E, incorporating ^{32}P in the RT for quantification. In panels B, C, D and F, data were generated by quantification of denaturing polyacrylamide gels similar to panels A and E.

$MgCl_2$ above 3 mM can inhibit RT (44)). For all four enzymes (tested at the manufacturer's recommended concentration), we observed more full-length RT product when no Mg^{2+} was added beyond that supplied by the diluted ligation reaction. As SuperScript III gave the highest RT product yield, we chose it for subsequent optimization. By varying the amount of the heat-inactivated Step 1 reaction in the Step 2 reaction, we determined that maximal RT product yield was obtained when the ligation reaction constituted one-third of the final volume of the RT reaction (data not shown). This resulted in a final $MgCl_2$ concentration of 3.3 mM. At this 3-fold dilution, we found no inhibitory effect on RT by the PEG 8000 present in the Step 1 reaction; rather, Step 1 reactions containing 25% PEG 8000 gave the highest Step 2 yields (Figure 3B).

We next varied RT primer, enzyme and RNA input amounts. To maximize RT product yield, it is important that the RT primer concentration be greater than the 3′-adaptor concentration but not excessively so, as this would favor empty circle formation in the subsequent circularization reaction (Step 4). We observed no advantage for RT yield when the RT primer:3′-adaptor ratio was significantly higher than 1.3:1 (Figure 3C). Further, all SuperScript III concentrations above 3 U/μl gave comparable product yields (Figure 3D). Varying the temperature (50°C, 55°C and 60°C) and time (30 min and 1 h) of the RT reactions revealed 55°C for 30 min to be optimal (data not shown). When the input RNA was varied between 3.3 and 133 nM, the yield of RT product increased linearly across this range (Figure 3E and F). Thus, like the ligation reaction, the RT reaction proved highly robust and amenable to library construction over a wide range of input amounts.

Based on the above data, we adopted the following as our standard Step 2 reaction conditions: 3-fold dilution of the

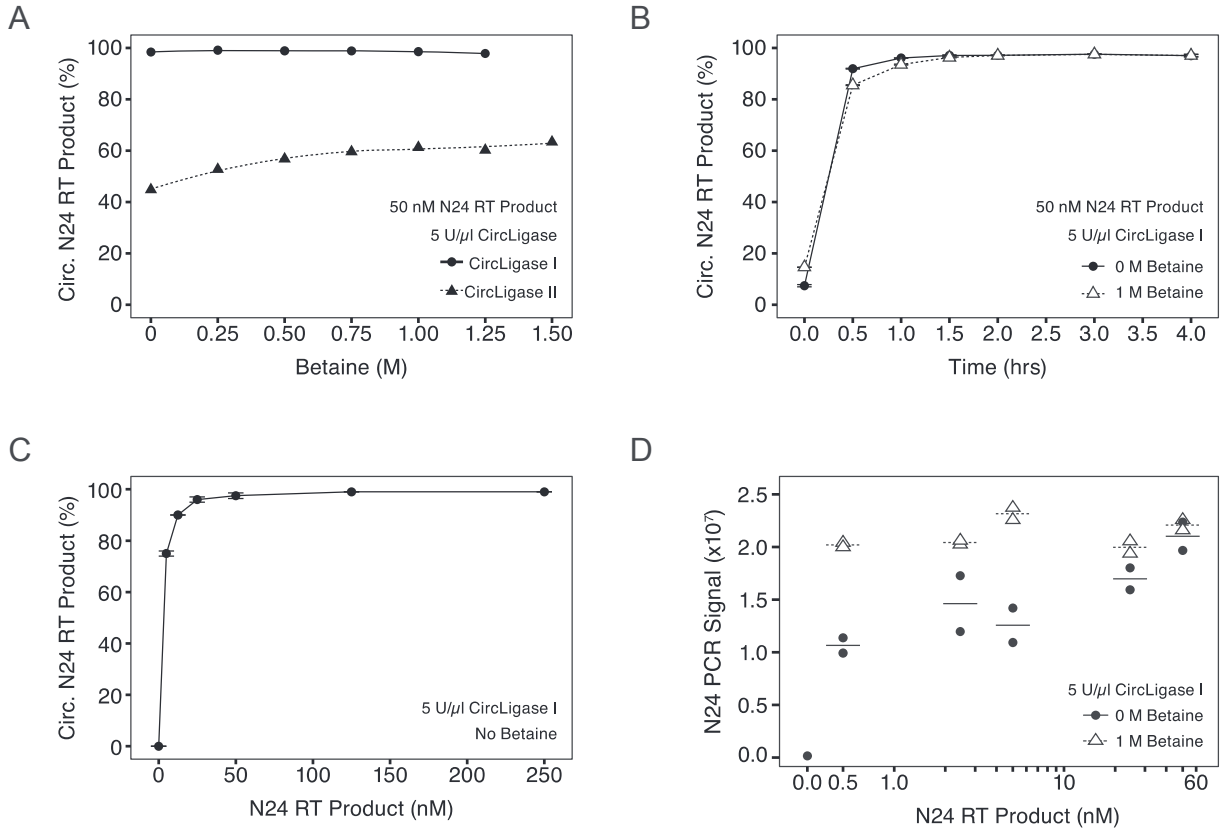

Figure 4. Circularization optimization. (A) Circularization efficiency versus betaine concentration (n = 1) for CircLigase I and II (n = 1). (B) Circularization efficiency versus time and betaine concentration (n = 1). (C) Circularization efficiency versus N24 RT product concentration (n = 2). (D) N24 PCR signal versus N24 RT product concentration prior to circularization (n = 2; line, mean) at 0M and 1M betaine. In all panels, data were generated by quantification of polyacrylamide gels (denaturing, panels A–C; non-denaturing, panel D). Circularization efficiency = (circularized RT product)/(linear RT product + circularized RT product) in each lane. N24 PCR signal = intensity of N24 PCR product band.

heat-denatured ligation reaction from Step 1, supplemented with 333 nM RT primer, 5.33 U/μl SuperScript III (to ensure consistent results and allow for some variability in nucleic acid concentration determination and enzyme activity), 50 mM Tris-HCl (pH 8.3 at room temperature), 75 mM KCl and 5 mM DTT. This mixture is incubated at 55°C for 30 min followed by heat inactivation at 75°C for 15 min.

Step 3: gel purification. See Materials and Methods.

Step 4: circularization. There are currently two commercially available enzymes for ssDNA circularization: CircLigase I and II (Epicentre). We tested both at 50 nM input ssDNA and found that CircLigase I gave much higher circularization efficiencies (98–99%) than CircLigase II (45–61%) (Figure 4A). Betaine, a compound commonly used in PCR reactions to eliminate the energy difference between A-T and G-C base pairs, is recommended by Epicentre for use with CircLigase II. However, as no amount of betaine improved CircLigase II efficiency to that obtained with CircLigase I, we decided to proceed with CircLigase I.

To explore the limits of CircLigase I performance, we tested a range of conditions. Changing the enzyme concentration and doubling or reducing by half the reaction volume had no significant effect on circularization efficiency (data not shown), so we continued to use the manufacturer's suggested conditions. A time course revealed that complete circularization with 5 U/μl enzyme and 50 nM input N24 RT product required at least 2 h at 60°C (Figure 4B). Titration of the N24 RT product indicated that ligation efficiencies dropped off precipitously below 25 nM ssDNA (Figure 4C). This drop-off was unaffected by either increasing or decreasing the enzyme concentration (data not shown), but was substantially rescued by the inclusion of 1 M betaine in the circularization reaction (Figure 4D). In this case, as circularization of < 5 nM N24 RT product could not be detected by direct observation of the ^{32}P-labeled substrate and product on a gel, relative PCR product yields served as a proxy for circularization yields, with cycle number adjusted for RNA input amount. In order to exclude the possibility of betaine stimulating the yield of the PCR reaction instead of the circularization reaction, we added betaine subsequent to heat inactivation of CircLigase I; under these conditions, no betaine-dependent increase in PCR signal was observed (data not shown).

Based on the above data, we adopted the following as our standard Step 4 reaction conditions: 1× CircLigase buffer (Epicentre), 1 M betaine, 50 μM adenosine triphosphate, 2.5 mM $MnCl_2$ and 5 U/μl CircLigase I in 20 μl containing all of the ssDNA isolated in Step 3. This mixture is incu-

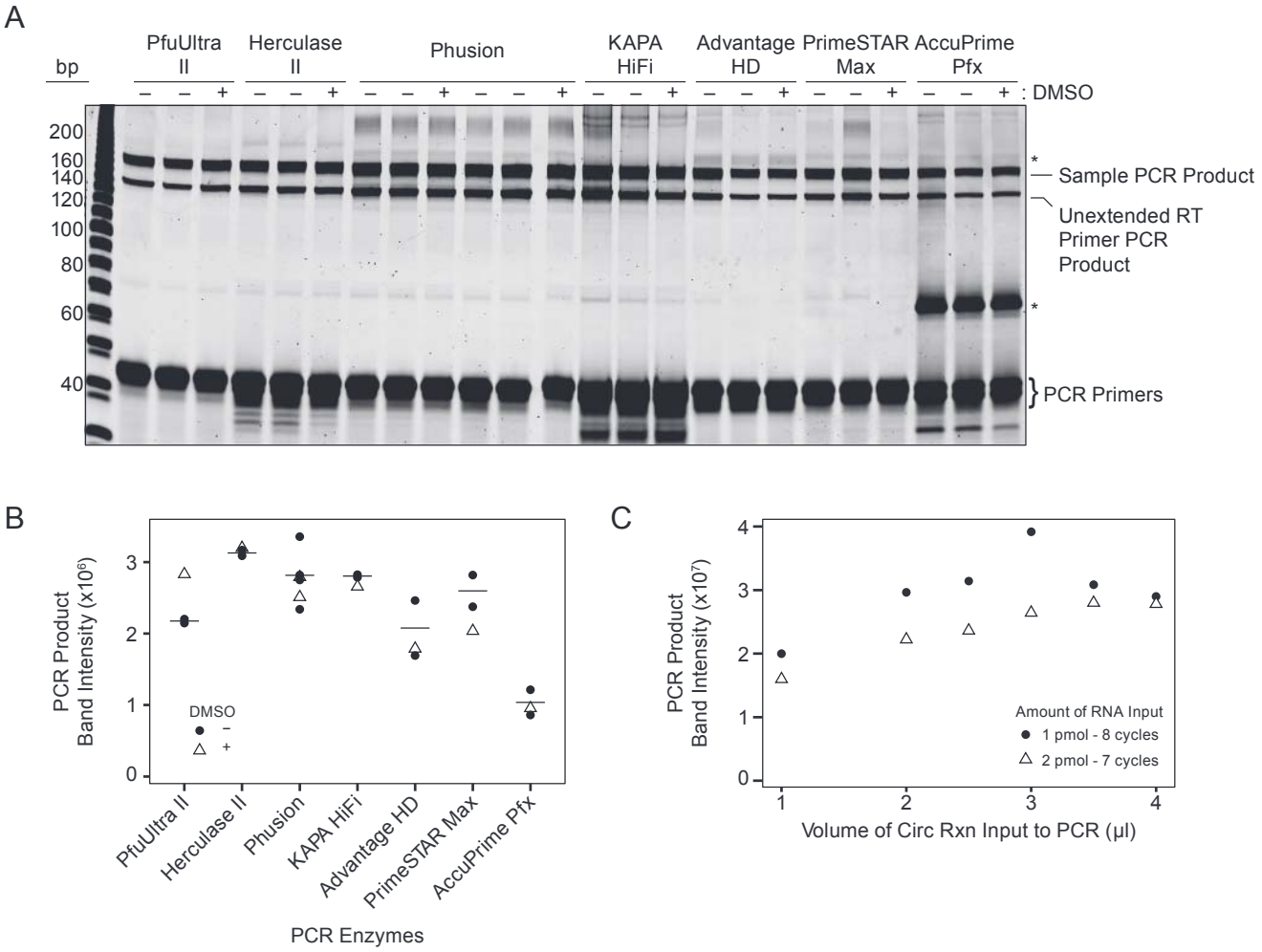

Figure 5. PCR optimization. (A) Comparison of proofreading PCR enzymes for the amount of sample PCR product ± DMSO. *, PCR by-products. (B) PCR product signal versus enzyme; quantification of panel A; black line, mean. (C) PCR product signal versus circularization reaction input volume (n = 1) for 1 pmol and 2 pmol RNA starting material. In panels B and C, data were generated by quantifying sample PCR product band on non-denaturing polyacrylamide gels.

bated at 60 C for 3 h followed by heat inactivation at 80 C for 10 min.

Step 5: PCR. To eliminate another gel purification step, we decided to use a portion of the completed and inactivated circularization reaction as direct input to PCR amplification. Adding 1.5 µl of a heat-inactivated circularization reaction containing ~88 nM input RT product directly to a 25 µl (final volume) PCR reaction, we tested the following high-fidelity polymerases, each using their respective manufacturer's supplied buffer and recommended cycling conditions (i.e. times and temperatures) for 8 cycles: PfuUltra II (Stratagene), Herculase II (Stratagene), Phusion (Finnzymes), KAPA HiFi (Kapa Biosystems), Advantage HD (Clontech), PrimeSTAR Max (Clontech) and AccuPrime Pfx (Invitrogen). Addition of DMSO, a PCR enhancing agent, did not significantly increase PCR amplification with any enzyme, perhaps with the exception of PfuUltra II (Figure 5A and B). PfuUltra II, Herculase II, Phusion, PrimeSTAR Max and KAPA HiFi all gave comparable product yields, but KAPA HiFi generated the least amount of slower migrating side products (indicated by *) just above the desired product (Figure 5A and B). Because of this and an independent report demonstrating its robustness with regard to GC content (45), we decided to proceed with KAPA HiFi.

When preparing deep sequencing libraries, higher amounts of input DNA and low cycle numbers are desirable to amplify the greatest number of unique species. However, as with the RT reaction (Step 2), we were concerned that the diluted circularization buffer might affect PCR efficiency. Therefore, we titrated the volume of CircLigase reaction included in each PCR reaction. When this volume was varied from 0.5 to 3.5 µl in a 15 µl PCR reaction, the PCR band intensity increased with increasing input, but not to scale (i.e. a 2-fold increase in input from 1 to 2 µl produced only a 1.5-fold increase in output; Figure 5C), likely indicating some inhibitory effect of the CircLigase reaction on PCR efficiency. We therefore limit the amount of added CircLigase reaction to one-fifth of the total PCR reaction volume.

Consequences of incomplete 3′ adaptor ligation

Having optimized each step in the protocol (Supplementary Table S1), we next wanted to assess the quality of libraries it generates. Because many published protocols use lower 3′ adaptor ligation temperatures and/or shorter incubation times than our optimized conditions (Figure 2F), we also wanted to test the effects of these variables. Therefore, we prepared seven different libraries using our synthetic N24 pool. All libraries were prepared identically except for the 3′-adaptor ligation step, for which the conditions are shown in Figure 2F and Supplementary Figure S2A. In one library, we also included four randomized nts at the 5′-end of the 3′ adaptor (N4 adaptor) to assess whether this would reduce 3′-end capture bias, as has been previously suggested (10,14,33). To eliminate possible sequencing variability, all libraries were barcoded, mixed together and sequenced to similar depth within a single Illumina HiSeq 2000 lane (Supplementary Figure S2A). Also included in this lane was a library of random ~500 nt fragments generated from the PhiX 174 genome (~15% of total sequences); PhiX inclusion increases the nt diversity at every position, thereby increasing the base calling accuracy (46).

To address the concern that long incubation times at higher temperatures could lead to significant RNA hydrolysis, we first examined the lengths of the captured sequences (Figure 6A). In all libraries, the majority of captured sequences were 24 nts. As expected, however, incubation at 22°C or 30°C for 6 h did result in a small decrease (<7%) in the fraction of full-length species compared to the 20 min and 1 h incubation times (Figure 6A, inset I). Also as expected, this effect was somewhat less apparent at 4°C. Nonetheless, the impact of this material loss must be weighed against the higher capture variability introduced by shorter ligation times and lower temperatures (see below).

For further analysis we focused solely on full-length (24 nt) reads. Because the number of possible sequences in a 24-nt random oligo (>10^{14}) so vastly outnumbers the reads obtained per library (~10^{7}), unique species constituted >99.5% of each library and >99.6% of the entire pooled data set (Supplementary Figure S2A). Because each library captured a unique sequence set, it was not possible to calculate the capture frequency for individual species. Therefore, to assess capture bias driven by nt identity, we measured nt frequency at each position in our captured fragments (Figure 6B). Across all libraries, there was a notable enrichment in G that decreased linearly in the 5′ → 3′ direction. To determine the extent to which this might be due to base misincorporation/miscalling at the sequencing level, we determined the mismatch frequency in the PhiX fragments sequenced alongside our N24 libraries (Supplementary Figure S2B). Across all positions corresponding to our N24 inserts, the PhiX mismatch frequency was no greater than 0.00049 for any of the 4 nts, with G being the least frequently miscalled base (<0.00021). Additionally, when analyzing the nt frequency per position in ribosome footprinting libraries made with our optimized ligation conditions, we see no 3′–5′ trend toward G enrichment (Supplementary Figure S2C). Thus, the most likely explanation for the overabundance of G in the N24 libraries was guanosine phosphoramidite overincorporation during oligonucleotide synthesis (47).

Examination of Figure 6B reveals that the majority of interlibrary variance occurred at the 3′ termini of captured RNAs (positions 21–24). To estimate expected nt frequencies (F_{exp}) at these terminal positions, we used the observed frequency (F_{obs}) data from all libraries to generate four best-fit lines (one for each nt) through positions 5–20 (Figure 6B), as these internal positions should be least affected by enzyme preference during 3′ adaptor ligation and circularization. We then used these best-fit lines to calculate expected nt counts at every nt position for each library. Calculating the chi-square statistic allowed us to quantify the deviation in observed nt count from expected nt count (Figure 6C). This analysis revealed that the chi-square statistic at positions 21–24 decreased in the following order: 30°C–20 min > 4°C–18 h > 22°C–1 h > 30°C–1 h > (30°C–6 h ≈ 30°C–6 h-N4 ≈ 22°C–6 h). That is, the libraries exhibiting the greatest deviation from expected were those wherein 3′ adaptor ligation was only ~30–85% complete (Figure 2F), either because of insufficient incubation time or a suboptimal ligation temperature. For reactions that did proceed to apparent completion (the three 6-h libraries), inclusion of four randomized nts at the 5′-end of the 3′ adaptor (5′ N4) had no additional benefit in reducing position 21–24 deviation compared to the fixed-sequence 3′ adaptor (although see miRNA data below).

Unexpectedly, position 22 exhibited equal or greater deviation than position 24 in all seven libraries. When comparing F_{obs}–F_{exp} for each nt, another feature readily observable in the 30°C–20 min library, and to a lesser extent in the 30°C–1 h library, is a tendency toward higher GC content at positions 11–15 (Supplementary Figure S3). Currently, we have no clear explanations for either of these effects (see Discussion), but both strengthen the point that uneven capture is accentuated by short ligation times.

Method validation

To assess how our optimized protocol performs on a known RNA sample, we made libraries from 50 fmol or 1 pmol of an equimolar 29 miRNA pool previously used to benchmark small RNA library preparation (SRR899527 and SRR899530; 14). Barcoded libraries were generated using either the fixed or N4 preadenylated 3′-adaptor, then pooled and sequenced on a single MiSeq lane (Table 1). Plotting F_{obs} versus F_{exp} (where F_{exp} = 1/29 = 0.0345) revealed no recurring over- or underrepresentation pattern for any individual miRNA across our four libraries (Figure 7A). Importantly, all four of four libraries exhibited less variability than both the previous benchmark (14) (Figure 7B) and a new library preparation protocol for capturing scarce miRNAs (39). In our libraries, the lowest CV in F_{obs} were obtained with the fixed adaptor at 1 pmol input and the N4 adaptor at 50 fmol and 1 pmol input. At 50 fmol input, however, the fixed adaptor did result in somewhat higher variability. Therefore, the N4 adaptor may be preferable when using our protocol to construct libraries from very low input RNA.

It has previously been noted that both secondary structure internal to individual miRNAs and the ability of in-

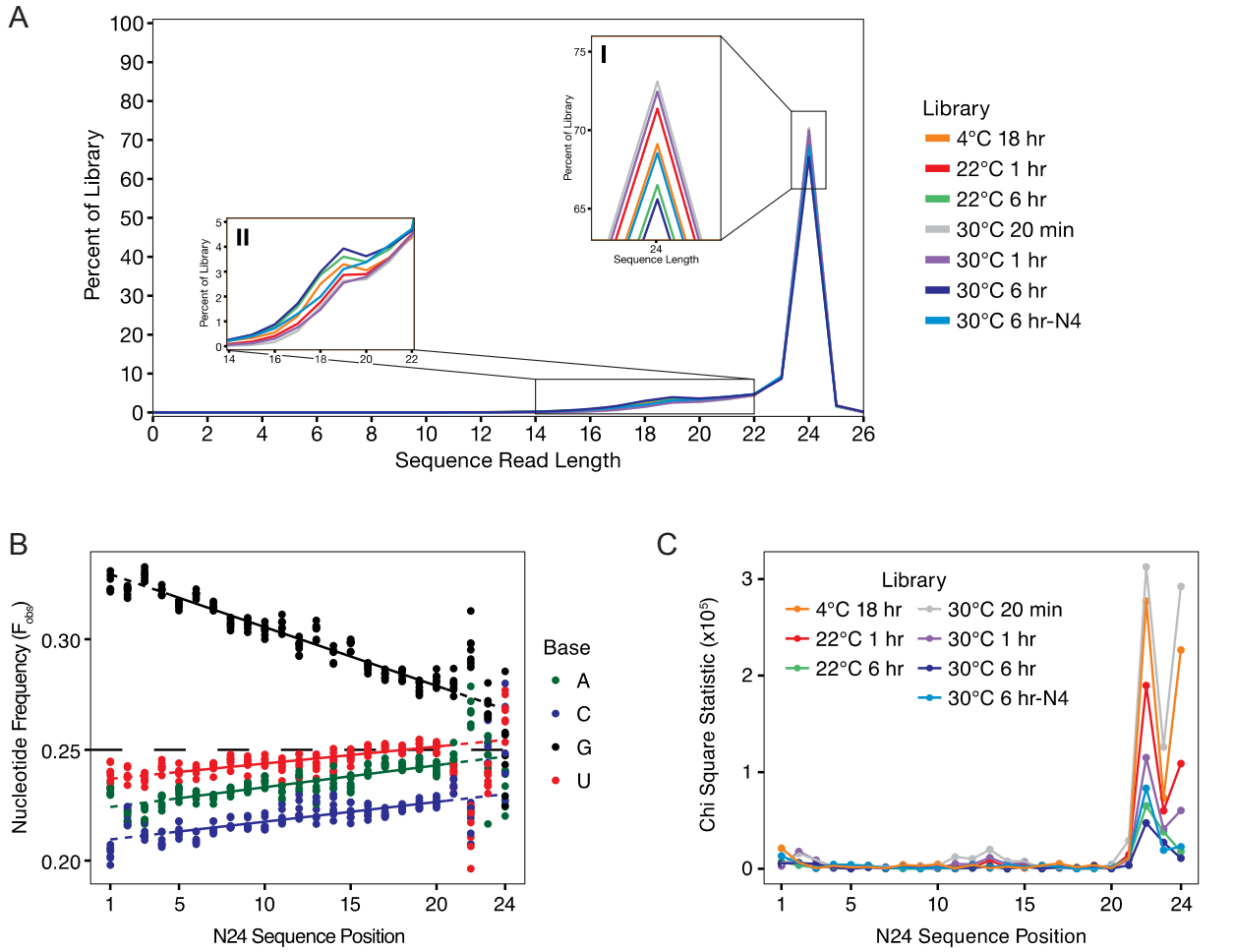

Figure 6. N24 length and bias analysis. (A) Distribution of read lengths, shown as a percent of the total sequences. (B) Nt frequency versus N24 sequence position. Dashed line indicates ideal 25% incorporation and capture of all four nts. (C) Total bias at each N24 sequence position.

Table 1. miRNA libraries

Input	Adaptor	Sequencing platform	Mapped reads
1 pmol	Fixed N4	MiSeq	1 044 234 1 393 238
50 fmol	Fixed N4		1 389 911 676 609
SRR 899527		HiSeq 2000	715 728
SRR 899530			1 424 004

dividual miRNAs to hybridize to the 3′-adaptor can affect capture efficiency (10,27). To address this possibility, we made scatter plots of read frequency versus individual miRNA features and calculated both slope and -value for the line best fitting the data (Supplementary Figure S4). (We note that a slope other than 0 is potentially indicative of bias, with the magnitude of the slope indicating the strength of the bias dependent on the particular feature being plotted. The -value indicates only how well the line fits the data.) These plots revealed no correlation with a |-value| > 0.5 between F_{obs} and GC-content, or between F_{obs} and the calculated folding energies (G) for each miRNA alone or each miRNA co-folded with the adaptor in any of our four libraries. We could also detect no apparent folding energy effects in the previous benchmark libraries. With the latter samples, however, there were readily observable trends with regard to nt composition, the most significant being a negative correlation (mean slope m = −0.058; mean = −0.72) between F_{obs} and the number of U's in the last 10 nts of each miRNA (Supplementary Figure S5). This is consistent with our N24 data showing an increased bias against U's in the last few nts when ligation reactions conditions are suboptimal (Supplementary Figure S3). The absence of the same trend in our miRNA libraries highlights the more even coverage provided by our optimized ligation conditions.

Under some conditions, reverse transcriptases can exhibit terminal transferase (TdT) activity, resulting in non-templated nt addition to cDNA 3′ ends (48). Examination of our miRNA libraries revealed that, while some untemplated addition did occur, extensions were generally limited to a single nt and these extended species were 20- to 50-fold less abundant than full-length species (Figure 7C).

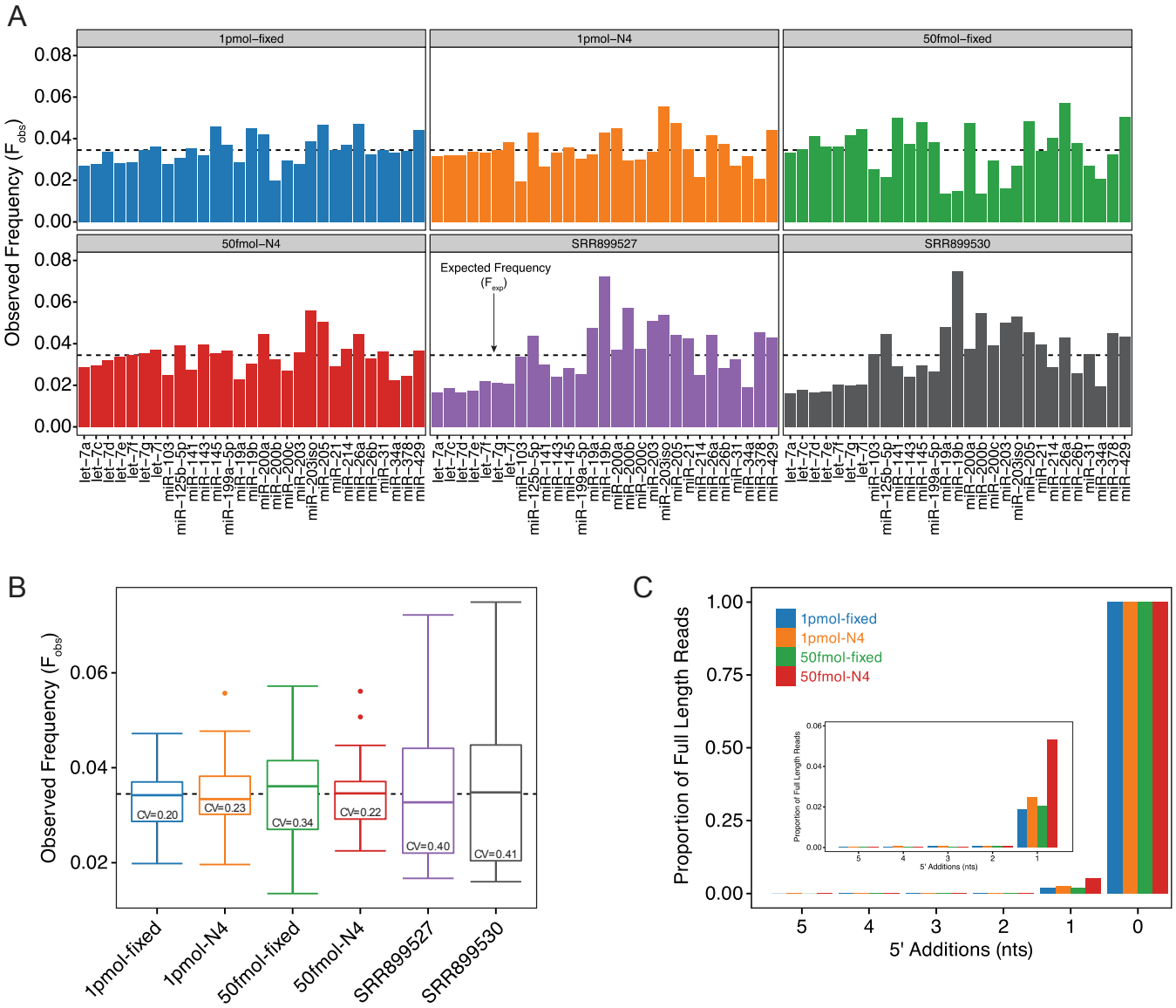

Figure 7. miRNA pool libraries. (A) Proportion of each miRNA in each library. Line represents perfectly even capture with each miRNA representing 1/29th of the reads. (B) Boxplot showing the distribution of proportions. CV = standard deviation (miRNA counts)/mean (miRNA counts). (C) Terminal transferase activity. Barchart showing percent of 5 additions and subtractions as a percentage of full-length reads.

During preparation, these samples were immediately gel-purified after RT (Supplementary Table S1). With one set of libraries, we observed more extensive TdT activity when the RT reaction was maintained at 4 C overnight following the heat inactivation step (data not shown). This suggests that Superscript III is not completely inactivated by the manufacturer's suggested heat inactivation regimen and will continue to add untemplated nts during long, low temperature incubations.

DISCUSSION

In this study, we set out to develop a method that yields robust strand-specific deep sequencing libraries from diverse RNA inputs. Our method involves 3 ligation of a preadenylated adaptor followed by RT, circularization and PCR. This approach combines features of several previously published protocols (3,23,43), with modifications to enhance capture efficiency and minimize sample loss. Our method works across a range of input amounts, is easy to follow, and produces a library in 2–3 days at relatively low reagent cost (< $25 per sample), all while giving the user complete control over every step. Because the input to our method is generic single-stranded RNA with a 3 hydroxyl, it can be used to capture many different sized RNA footprints. Our approach can also be used to map sites of RNA-protein crosslinking (e.g. from CLIP experiments) and other base modifications that cause reverse transcriptase to either stall (e.g. abasic or alykylated sites) or incorporate the wrong base (e.g. PAR-CLIP). To date, various members of our laboratory have used this method to generate multiple footprinting libraries for Ribo-Seq and other RNA-protein

complexes, as well as RNA-Seq libraries (6 and unpublished results). Input fragment sizes have ranged from 20 to 200 nts, input amounts have ranged from 400 pg to 200 ng RNA and all resulted in highly complex libraries. Our method is highly reproducible, with both read counts and RPKM for Ribo-Seq and RNA-Seq biological replicates having correlation coefficients of 0.93–0.99 (6 and unpublished results).

One of our major goals in developing this protocol was to minimize capture biases. We did so by identifying conditions wherein both the RNL2 and CircLigase reactions were driven to apparent completion, thereby minimizing ligase sequence preferences and any intra- and inter-molecular secondary structure effects. Our analysis of the effects of time and temperature on 3-adaptor ligation clearly indicates that incomplete ligation exacerbates capture bias (Figures 2, 6B and 6C and Supplementary Figure S3). Nonetheless, even under conditions where the ligation reaction appeared to proceed to completion, apparent 3-end biases were not fully eliminated (Figure 6C). Three recent papers reported that 3-end capture bias can be reduced by including a short (2–4 nt) randomized region at the 5-end of the 3 adaptor (10,14,33). Inclusion of degenerate nts in the adaptor also allows for identification of species that are preferentially amplified during the PCR reaction (49). Although we observed no advantage of the N4 adaptor over our fixed sequence adaptor with 1–2 pmol N24 or miRNA pool input (Figures 6C, 7A and B and Supplementary Figure S3), the N4 adaptor was clearly superior when the miRNA pool input was lowered to 50 fmol (Figure 7A and B). Therefore, using a 5 randomized adaptor is recommended.

Contrary to expectation (10,27), we could detect no effects on N24 or miRNA capture efficiency that could be attributed to either internal secondary structure forming propensity or the ability of captured sequences to hybridize with the adaptor (Supplementary Figure S4 and data not shown). In our N24 data, however, we did detect an unexpected nt identity bias at the -3 position relative to the 3-adaptor ligation site (Figure 6C). This is consistent with a previous report demonstrating -3 substrate bias by both RNL1 and RNL2 (27). Currently, there is no clear explanation for this effect, as a crystal structure of RNL2 bound to substrate suggests that RNL2 substrate specificity is dictated solely by the nts at positions -1 and -2 (50). Nonetheless, our N24 data highlight the importance of driving the 3-ligation reaction as close to completion as possible.

Following ligation, RT of the captured RNA attaches a sequence tag to the 3-end of the RNA, allowing for PCR amplification and deep sequencing. Although the adaptor sequences used here are for sequencing on Illumina platforms, libraries can be prepared for any deep sequencing platform by simply modifying the 5 and 3 adaptor sequences. Our method employs a variety of RT primers that differ only by their 5 barcode, allowing multiple samples to be sequenced on the same flow cell lane. Barcoding the samples during the RT step minimizes opportunities for accidental mixing or cross-contamination of samples. We currently use a set of twelve 5-nt barcodes (see Materials and Methods) that were chosen such that the first position is balanced (to increase initial base calling accuracy by Illumina platforms) and there is no possibility for barcode misidentification, even with two sequencing errors. After circularization, the barcode is positioned 5 to the captured cDNA sequence, allowing for barcode identification and fragment sequencing all in one single-end sequencing run.

Following circularization, one must determine the optimal number of PCR cycles for each sample. Cycle number is highly dependent on the original RNA input amount. Our current approach is to empirically determine the correct number of PCR cycles by gel analysis; too few cycles will result in product yield below the sequencing input requirement; too many cycles will result in PCR jackpots that can overwhelm the library and introduce significant bias. A recently published qPCR approach for identifying the correct number of cycles can easily be applied to our method (17).

Two similar protocols for making strand-specific libraries were recently published (51,52), speaking to the overall strength of this strategy. Nonetheless, the modifications we describe here (i.e. inclusion of 25% PEG in the 3-adaptor ligation reaction; no additional $MgCl_2$ in the RT reaction; a single gel purification step; inclusion of 1M betaine in the CircLigase I reaction; and optimized times and temperatures to ensure completion of all reactions) offer significant improvements over similar methods. To assist the reader in implementing our protocol, we have included a short summary of the conditions (Supplementary Table S1) and placed a detailed protocol at http://www.umassmed.edu/moorelab/resources/protocols/.

ACCESSION NUMBERS

High-throughput sequencing data have been deposited in the GEO database under accession number GSE63606.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Catherine Sterling, Victor Ambros and Phillip Zamore for helpful discussions related to method development. Rui Yi kindly provided the 29 miRNA mix. Fred Hyde and Hank Daum at Epicentre aided in troubleshooting the circularization reaction.

FUNDING

M.J.M. is a Howard Hughes Medical Institute Investigator. Funding for open access charge: Howard Hughes Medical Institute.
Conflict of interest statement. None declared.

REFERENCES

1. Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A. and Darnell, R.B. (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science, 302, 1212–1215.
2. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M. Jr., Jungkamp, A.-C., Munschauer, M. et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell, 141, 129–141.

3. Ingolia,N.T., Ghaemmaghami,S., Newman,J.R.S. and Weissman,J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science, 324, 218–223.
4. Singh,G., Ricci,E.P. and Moore,M.J. (2014) RIPiT-Seq: a high-throughput approach for footprinting RNA:protein complexes. Methods, 65, 320–332.
5. Singh,G., Kucukural,A., Cenik,C., Leszyk,J.D., Shaffer,S.A., Weng,Z. and Moore,M.J. (2012) The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell, 151, 750–764.
6. Ricci,E.P., Kucukural,A., Cenik,C., Mercier,B.C., Singh,G., Heyer,E.E., Ashar-Patel,A., Peng,L. and Moore,M.J. (2014) Staufen1 senses overall transcript secondary structure to regulate translation. Nat. Struct. Mol. Biol., 21, 26–35.
7. Chen,W., Shulha,H.P., Ashar-Patel,A., Yan,J., Green,K.M., SUery,C.C., Rhind,N., Weng,Z. and Moore,M.J. (2014) Endogenous U2·U5·U6 snRNA complexes in S. pombe are intron lariat spliceosomes. RNA, 20, 1–13.
8. Linsen,S.E.V., de Wit,E., Janssens,G., Heater,S., Chapman,L., Parkin,R.K., Fritz,B., Wyman,S.K., de Bruijn,E., Voest,E.E. et al. (2009) Limitations and possibilities of small RNA digital gene expression profiling. Nat. Methods, 6, 474–476.
9. Fersht,A. (1985) Enzyme Structure and Mechanism, 2nd ed. W H Freeman & Co., New York.
10. Sorefan,K., Pais,H., Hall,A.E., Kozomara,A., Griffiths-Jones,S., Moulton,V. and Dalmay,T. (2012) Reducing ligation bias of small RNAs in libraries for next generation sequencing. Silence, 3, 4.
11. Hafner,M., Renwick,N., Brown,M., Mihailovic,A., Holoch,D., Lin,C., Pena,J.T.G., Nusbaum,J.D., Morozov,P., Ludwig,J. et al. (2011) RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. RNA (New York, N.Y.), 17, 1697–1712.
12. Bissels,U., Wild,S., Tomiuk,S., Holste,A., Hafner,M., Tuschl,T. and Bosio,A. (2009) Absolute quantification of microRNAs by using a universal reference. RNA (New York, N.Y.), 15, 2375–2384.
13. Levin,J.Z., Yassour,M., Adiconis,X., Nusbaum,C., Thompson,D.A., Friedman,N., Gnirke,A. and Regev,A. (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Publ. Group, 7, 709–715.
14. Zhang,Z., Lee,J.E., Riemondy,K., Anderson,E.M. and Yi,R. (2013) High-efficiency RNA cloning enables accurate quantification of miRNA expression by deep sequencing. Genome Biol., 14, R109.
15. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput reads. EMBnet.journal, 17, 10–12.
16. Lorenz,R., Bernhart,S.H., Höner Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. Algorithms for Molecular Biology, 6, 26.
17. Langevin,S.A., Bent,Z.W., Solberg,O.D., Curtis,D.J., Lane,P.D., Williams,K.P., Schoeniger,J.S., Sinha,A., Lane,T.W. and Branda,S.S. (2013) Peregrine: a rapid and unbiased method to produce strand-specific RNA-Seq libraries from small quantities of starting material. RNA Biol., 10, 502–515.
18. Kwok,C.K., Ding,Y., Sherlock,M.E., Assmann,S.M. and Bevilacqua,P.C. (2013) A hybridization-based approach for quantitative and low-bias single-stranded DNA ligation. Anal. Biochem., 435, 181–186.
19. Zhang,Z., Theurkauf,W.E., Weng,Z. and Zamore,P.D. (2012) Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. Silence, 3, 9.
20. Armour,C.D., Castle,J.C., Chen,R., Babak,T., Loerch,P., Jackson,S., Shah,J.K., Dey,J., Rohl,C.A., Johnson,J.M. et al. (2009) Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. Nat. Methods, 6, 647–649.
21. Cloonan,N., Forrest,A.R.R., Kolle,G., Gardiner,B.B.A., Faulkner,G.J., Brown,M.K., Taylor,D.F., Steptoe,A.L., Wani,S., Bethel,G. et al. (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat. Methods, 5, 613–619.
22. Elbashir,S.M., Lendeckel,W. and Tuschl,T. (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev., 15, 188–200.
23. Lau,N.C., Lim,L.P., Weinstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science, 294, 858–862.
24. Pan,T. and Uhlenbeck,O.C. (1992) In vitro selection of RNAs that undergo autolytic cleavage with lead (2+). Biochemistry, 31, 3887–3895.
25. Hansen,K.D., Brenner,S.E. and Dudoit,S. (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res., 38, e131.
26. Viollet,S., Fuchs,R.T., Munafo,D.B., Zhuang,F. and Robb,G.B. (2011) T4 RNA ligase 2 truncated active site mutants: improved tools for RNA analysis. BMC Biotechnol., 11, 72.
27. Zhuang,F., Fuchs,R.T., Sun,Z., Zheng,Y. and Robb,G.B. (2012) Structural bias in T4 RNA ligase-mediated 3′-adapter ligation. Nucleic Acids Res., 40, e54.
28. Ingolia,N.T. (2010) Genome-wide translational profiling by ribosome footprinting. Methods Enzymol., 470, 119–142.
29. Parameswaran,P., Jalili,R., Tao,L., Shokralla,S., Gharizadeh,B., Ronaghi,M. and Fire,A.Z. (2007) A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. Nucleic Acids Res., 35, e130–e130.
30. Alon,S., Vigneault,F., Eminaga,S., Christodoulou,D.C., Seidman,J.G., Church,G.M. and Eisenberg,E. (2011) Barcoding bias in high-throughput multiplex sequencing of miRNA. Genome Res., 21, 1506–1511.
31. Hafner,M., Renwick,N., Farazi,T.A., Mihailovic,A., Pena,J.T.G. and Tuschl,T. (2012) Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. Methods, 58, 164–170.
32. Van Nieuwerburgh,F., Soetaert,S., Podshivalova,K., Ay-Lin Wang,E., Schaffer,L., Deforce,D., Salomon,D.R., Head,S.R. and Ordoukhanian,P. (2011) Quantitative bias in Illumina TruSeq and a novel post amplification barcoding strategy for multiplexed DNA and small RNA deep sequencing. PloS ONE, 6, e26969.
33. Jayaprakash,A.D., Jabado,O., Brown,B.D. and Sachidanandam,R. (2011) Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic Acids Res., 39, e141.
34. Vivancos,A.P., Güell,M., Dohm,J.C., Serrano,L. and Himmelbauer,H. (2010) Strand-specific deep sequencing of the transcriptome. Genome Res., 20, 989–999.
35. Eminaga,S., Christodoulou,D.C., Vigneault,F., Church,G.M. and Seidman,J.G. (2013) Quantification of microRNA Expression with Next-Generation Sequencing. Curr. Protoc. Mol. Biol., Chapter 4, Unit 4.17.
36. Mamanova,L. and Turner,D.J. (2011) Low-bias, strand-specific transcriptome Illumina sequencing by on-flow cell reverse transcription (FRT-seq). Nat. Protoc., 6, 1736–1747.
37. Pfeffer,S., Lagos-Quintana,M. and Tuschl,T. (2005) Cloning of small RNA molecules. Curr. Protoc. Mol. Biol., Chapter 26, Unit 26.4.
38. Munafo,D.B. and Robb,G.B. (2010) Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. RNA, 16, 2537–2552.
39. Sterling,C.H., Veksler-Lublinsky,I. and Ambros,V. (2014) An efficient and sensitive method for preparing cDNA libraries from scarce biological samples. Nucleic Acids Res., doi:10.1093/nar/gku637.
40. Morin,R.D., O'Connor,M.D., Griffith,M., Kuchenbauer,F., Delaney,A., Prabhu,A.-L., Zhao,Y., McDonald,H., Zeng,T., Hirst,M. et al. (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res., 18, 610–621.
41. Lee,R.C. and Ambros,V. (2001) An extensive class of small RNAs in Caenorhabditis elegans. Science, 294, 862–864.
42. Ingolia,N.T., Brar,G.A., Rouskin,S., McGeachy,A.M. and Weissman,J.S. (2013) Genome-wide annotation and quantitation of translation by ribosome profiling. Curr. Protoc. Mol. Biol., Chapter 4, Unit 4.18.
43. Lui,W.-O., Pourmand,N., Patterson,B.K. and Fire,A. (2007) Patterns of known and novel small RNAs in human cervical cancer. Cancer Res., 67, 6031–6043.
44. Gerard,G.F., Fox,D.K., Nathan,M. and D'Alessio,J.M. (1997) Reverse transcriptase. The use of cloned Moloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA. Mol. Biotechnol., 8, 61–77.
45. Quail,M.A., Otto,T.D., Gu,Y., Harris,S.R., Skelly,T.F., McQuillan,J.A., Swerdlow,H.P. and Oyola,S.O. (2012) Optimal enzymes for amplifying sequencing libraries. Nat. Methods, 9, 10–11.

46. Illumina (2013) Technical note: using a PhiX control for HiSeq sequencing runs. http://res.illumina.com/documents/products/technotes/technote_phixcontrolv3.pdf. (October 2013, date last accessed).
47. Bartel, D.P. and Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences. Science, 261, 1411–1418.
48. Chen, D. and Patton, J.T. (2001) Reverse transcriptase adds nontemplated nucleotides to cDNAs during 5′-RACE and primer extension. BioTechniques, 30, 574–582.
49. König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat. Struct. Mol. Biol., 17, 909–915.
50. Nandakumar, J., Shuman, S. and Lima, C.D. (2006) RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. Cell, 127, 71–84.
51. Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M. and Weissman, J.S. (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat. Protoc., 7, 1534–1550.
52. Epicentre Technologies Corporation (2012) ARTseq Ribosome Profiling Kit. http://www.epibio.com/applications/rna-sequencing/ribosome-profiling/artseq-ribosome-profiling-kits?protocols. (June 2013, date last accessed).