Removing bias against short sequences enables northern blotting to better complement RNA-seq for the study of small RNAs

Yun S. Choi, Lanelle O. Edwards, Aubrey DiBello and Antony M. Jose*

Department of Cell Biology and Molecular Genetics, University of Maryland, College, Park, MD 20742, USA

Received October 16, 2015; Revised January 12, 2017; Editorial Decision January 31, 2017; Accepted February 02, 2017

ABSTRACT

Changes in small non-coding RNAs such as micro RNAs (miRNAs) can serve as indicators of disease and can be measured using next-generation sequencing of RNA (RNA-seq). Here, we highlight the need for approaches that complement RNA-seq, discover that northern blotting of small RNAs is biased against short sequences and develop a protocol that removes this bias. We found that multiple small RNA-seq datasets from the worm Caenorhabditis elegans had shorter forms of miRNAs that appear to be degradation products that arose during the preparatory steps required for RNA-seq. When using northern blotting during these studies, we discovered that miRNA-length probes can have ~1000-fold bias against detecting even synthetic sequences that are 8 nt shorter. By using shorter probes and by performing hybridization and washes at low temperatures, we greatly reduced this bias to enable nearly equivalent detection of 24 to 14 nt RNAs. Our protocol can discriminate RNAs that differ by a single nucleotide and can detect specific miRNAs present in total RNA from C. elegans and pRNAs in total RNA from bacteria. This improved northern blotting is particularly useful to analyze products of RNA processing or turnover, and functional RNAs that are shorter than typical miRNAs.

INTRODUCTION

Small non-coding RNAs such as microRNAs (miRNAs) regulate much of our genome (1) and thus can be broad indicators of health and disease. Next generation RNA sequencing (RNA-seq) (2) has quickly developed to become a powerful method to measure the expression of small RNAs. Because changes in miRNA expression have been observed in multiple disease states (3), there is great interest in measuring miRNA levels accurately for use in clinical diagno-

© The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
One of the best approaches for validating the expression of a new RNA without the need for ligation or reverse transcription and irrespective of terminal modifications on the RNA is northern blotting. Therefore, substantial effort has been devoted to improving the sensitivity of this technique for the analysis of small RNAs. These include the use of locked nucleic acid (LNA)-modified probes (27), the use of chemical crosslinking of RNA to the membrane (28) and the demonstration that digoxigenin-labeled probes (29, 30) can be more sensitive than radioactive probes (31).

Together, improvements in sensitivity have made it possible to detect as little as 50 attomoles of a miRNA (31). Furthermore, the recent development of a technique to unmask small RNAs bound to complementary RNAs (25) and an approach to multiplex probes (32) continue to improve the ease of use and applicability of northern blotting.

Here we detect unexpected forms of well-studied C. elegans miRNAs in some RNA-seq datasets, discover that northern blotting can have drastic biases against short RNAs and develop an improved northern blotting protocol that reduces this bias to enable specific detection of RNAs in the 14 to 24 nt size range including endogenous miRNAs in C. elegans and product RNAs (pRNAs) in bacteria.

MATERIALS AND METHODS

Analysis of RNA-seq datasets

All datasets analyzed in this study are listed in Supplementary Table S1.

Detection of length variants of miRNAs in RNA-seq datasets.

FASTQ or FASTA datasets were aligned using Bowtie to the C. elegans genome (Ce6) after removing adaptor sequences in the Galaxy interface (33–36). The bowtie parameters used were: bowtie -q -p 8 -S -n 2 -e 70 -i 28 – maxbts 125 -k 4 -phred33-quals <ce6 index> <fastq>. For adapter trimming, a minimum sequence length after trimming was set to 15 nt and all sequences with unknown (N) bases were discarded. Reads that aligned to genomic intervals of known miRNA genes (miRBase WBCel235) were filtered and used to calculate the average length of reads per miRNA gene. All reads (if total <1000) or 1000 randomly selected reads (if total >1000) that aligned to miRNAs that showed shortening were included to determine the coverage per nucleotide using Matlab (Mathworks).

Detection of miR-52 variants.

Wild-type (N2) datasets that were deposited as compressed FASTA files or as SRA files were retrieved from the NCBI Sequence Read Archive. SRA datasets were converted to FASTQ format using SRA Toolkit V.2.4.2 and used regardless of quality scores. Prevalence of miR-52 with missing 5′ end nucleotides (5′-truncated) in N2 datasets was calculated by comparing the number of reads that exactly matched the full, annotated length (24 nt) to the number of reads that included the 3′-most 15 bases of miR-52 (ATATGTTTTCGCTGCT) using grep function in unix. This 15-mer showed only one match to the C. elegans genome (miR-52) when searched with BLASTn. Only reads that exactly matched the reference genome were kept and reads with 3′ end variation were excluded. The number of reads required to consider a variant was scaled with the size of the RNA-seq dataset. In our detailed examination of miR-52 variants, we counted only those forms of miR-52 that had a perfect match to the annotated 3′ end of miR-52. In addition to these 5′ truncated versions of miRNAs, there were additional forms that varied at both the 5′ and 3′ ends as expected because of the known 3′ end heterogeneity of miRNAs (12). Such miRNA variants with 3′ truncation and/or 3′ nucleotide addition were excluded from our counts.

Circularly permuted forms of miR-52 were identified by searching for reads (using grep) that exactly matched all the rearranged full-length miR-52 sequences. These sequences were used to search the nucleotide database with BLASTn and no alignments other than miR-52 were found in the C. elegans genome. All plots were created using Excel (Microsoft), Matlab (Mathworks) and Illustrator (Adobe).

Generation of mut-2(jam9)

The mut-2 locus (37) was edited by homology-based repair using the method described earlier (38, 39). Briefly, plasmids that separately express Cas9, sgRNA, a homology repair template and mCherry markers were co-injected into worms. Transgenic worms that expressed the mCherry marker were screened in the mut-2 locus.

The sgRNA targeting mut-2 was designed using E-CRISP (40) and cloned into the pU6::unc-119 sgRNA expression vector (Addgene 46169, (39)). The unc-119 sgRNA was removed from the plasmid by restriction digest (MfeI and XbaI; all restriction enzymes were from NEB). The large fragment of the plasmid digest was gel purified (28706, Qiagen) and used as the vector backbone for cloning of the new sgRNA. To prepare the insert containing the sgRNA targeting mut-2, two overlapping polymerase chain reaction (PCR) products were amplified with Phusion DNA polymerase (NEB, used for all PCRs except as indicated) in separate reactions using pU6::unc-119 as a template. Primer pair YC001 and YC128 was used for the first reaction and primer pair YC127 and YC020 was used for the second reaction. Both PCR products were used as templates and the fused product was amplified using YC001 and YC020 as primers. The amplified product was column purified (28106, Qiagen), digested (MfeI and XbaI), gel purified (28706, Qiagen) and cloned into the prepared pU6::unc-119 sgRNA expression vector with T4 DNA ligase (NEB). The homology repair template was prepared by initially cloning the mut-2 locus and then replacing the mut-2 coding regions with that of gfp. The mut-2 coding sequence and ~1.4 kb of flanking sequences were amplified from wild-type (N2) genomic DNA using the primers YC087 and YC085. The PCR product was column purified (28106, Qiagen), digested (ApaI and PmlI) and used as the insert. The plasmid pCFJ-151 (Addgene 19330, (41)) was digested (ApaI and PmlI) and the ~2.3 kb fragment was gel purified, and used as vector. The resultant cloned plasmid (pmut-2, genomic) was used to amplify the 5′ (primers YC039 and YC134) and 3′ (primers YC137 and YC138) flanking sequences of mut-2 in separate reactions. The gfp coding region was then amplified from pGCC305 (Addgene 19646, (42)) with primers (YC135 and YC136) that also overlap...
mut-2 flanking sequences. The three fragments were combined by overlap extension PCR using primers YC039 and YC138. The resultant ~1.8 kb fusion product was gel purified, digested (PstI and BamHI), column purified (28106, Qiagen) and used as insert. The plasmid pmut-2 genomic was digested (PstI and BamHI), ~3.8 kb fragment was gel purified (28706, Qiagen) and used as vector to generate pmut-2_to_eGFP_HRT.

Plasmids for microinjection were suspended in 10 mM Tris–Cl, pH 8.5 to the following final concentrations: 50 ng/µl Peptide-cas9-SV40-NLS::ttb-2 3′UTR (Addgene 46168, (39)); 22.5 ng/µl PU6::mut-2_sgRNA (based on Addgene 46169, (39)); 10 ng/µl pmut-2_to_eGFP_HRT; 10 ng/µl pMA122 (Addgene 34873, (43)); 10 ng/µl pH8 (Addgene 19359, (41)); 2.5 ng/µl pCFJ90 (Addgene 19327, (41)); 5.0 ng/µl pCFJ104 (Addgene 19328, (41)).

F1 progeny of injected worms that expressed the mCherry coinjection markers were selected and allowed to have progeny. Twenty sets of five F2 progeny that did not express mCherry coinjection markers were moved to a plate. Pools of worms from each plate were screened for editing by PCR (primers YC176 and YC171). Individuals with successful editing were identified, homozygosed, confirmed by genomic DNA sequencing and designated as mut-2 (jam9).

RNA purification
Worms grown on 35 mm plates of LB agar and fed *Escherichia coli* (OP50) were collected from starved plates by rinsing with M9 buffer (22 mM KH2PO4, 42 mM Na2HPO4, 86 mM NaCl and 1 mM MgSO4). Worms were pelleted (3 min at 750 × g tabletop microcentrifuge) and excess M9 buffer was aspirated leaving behind ~50 µl 1.0 ml TRIzol (LifeTechnologies) was added to the worm pellet and frozen at −80°C. After thawing, the samples were processed according to manufacturer’s instructions. RNA was resuspended in nuclease free water (IDT) and its concentration was measured by ultraviolet (UV) spectrophotometry (Nanovue GE).

Induction and purification of pRNA
*Escherichia coli* strains were grown in LB for 24 h at 37°C in a shaking incubator. pRNAs were induced by transferring 10 ml of saturated culture to 40 ml of pre-warmed LB and placed back in the shaking incubator. Aliquots (5 ml) of induced culture were removed after 3 and 5 min of shaking, placed on ice and then collected in a 4°C centrifuge (3 min, 1500 × g). The growth medium was removed and cells were resuspended in 1.0 ml of Trizol and then transferred to 1.7 ml microcentrifuge tubes. Cells in Trizol were disrupted by freeze-thaw cycles of flash freezing in liquid N2 and then incubated at 65°C for 5 min. Total RNA was then collected as per manufacturer’s protocol (Life Technologies).

RT-PCR
A total of 250 ng of total RNA was reverse transcribed with SuperScript III (LifeTechnologies) using 500 ng oligo-dT18 primer (IDT) according to manufacturer’s protocol. Gene specific primers and Taq polymerase were used to amplify target genes from 1 µl of the cDNA synthesis reaction. Cycling conditions were an initial incubation at 95°C for 30 s followed by 30 cycles of: 95°C, 30 s; annealing temperature (52°C, GFP; 55°C, mut-2; and 55°C, *ttb-2*); 30 s; 72°C, 30 s; and a final 72°C incubation of 5 min.

Northern blotting
Northern blotting was done as described earlier (31) with some essential modifications as noted below. In brief, 5 µg (mut-2/jam9)) or 10 µg (miRNA) of total worm RNA or 100 fmol of each synthetic RNA was separated by denaturing 20% polyacrylamide gel electrophoresis (PAGE) (7M urea) and transferred (25 V, 1.0 A, 30 min, BioRad Transblot-Turbo) to a positively charged nylon membrane (11209299001, Roche). RNA was crosslinked to the membrane using UV irradiation (1200 µJ/m2, VWR UV Crosslinker). Synthetic DNA (IDT) probes were made with Roche DIG Oligonucleotide Tailing kit following manufacturer’s protocol or bought as 5′-monodigoxigenin-labeled (mono-DIG) oligonucleotides (IDT). Probes were hybridized to membranes (2.5 pmol probe in 5 ml ULTRA-hyb buffer, Ambion LifeTechnologies) by incubating overnight (with rotation at 8 RPM in a VWR Hybridization Oven) at 42°C for full-length probes and at room temperature (RT, ~22°C) for short probes. Membranes were then washed (2 × 15 min at 37°C for full-length miR-52 probe; 2 × 15 min at 42°C for full-length miR-53 probe; and 2 × 15 min at RT for all short probes). Membranes were blocked (31), incubated with DIG antibodies (31), washed (4 × 15 min at RT for long probes and 2 × 5 min for short probes), developed (31) and imaged (Fujifilm LAS-3000). Blots were stripped by incubation in ~5–10 ml of 5% sodium dodecyl sulphate at 80°C for 15 min and washed twice with ~5–10 ml 2× Saline-Sodium Citrate (SSC) at 80°C for 5 min. The total RNA blot was probed first for miR-52, stripped and then probed for U6. For the miR-52 and miR-53 blot, the probes were used in the following order: short miR-52, short miR-53, full-length miR-52 and full-length miR-53. The synthetic RNA blot was probed in the following order: miR-1, let-7, miR-58 and miR-52. The miRNA in total RNA preparations blot was probed in the following order: miR-52, miR-58 and U6. The blot for comparing mono-DIG labeling with DIG tailing was probed in the following order: 15 nt DIG-tailed, 15 nt mono-DIG, 24 nt DIG-tailed, 24 nt mono-DIG and U6.

Northern blotting of miR-1 from total RNA was done as above with 15 and 17 nt DIG-tailed DNA oligos. Northern blotting of pRNA was done in a similar manner using 10 µg of total RNA purified from *E. coli* strains. The membrane was first probed with a 15 nt DIG-tailed DNA oligo for pRNAs, stripped and repoared with a 22 nt DIG-tailed DNA oligo for 5S rRNA.

Image analysis
Northern blot images were superimposed as separate layers in Photoshop (Adobe) and arranged so that the same area of the membrane was cropped for analysis. Cropped images were analyzed with ImageJ64 1.48v (44). Background was subtracted (50 pixels radius) before measuring band inten-
Some miRNAs show substantial 5’ truncation in multiple RNA-seq datasets

A growing body of work shows that processes that change the length of a miRNA can be regulated (17–19) and are biologically important. To look for such variations in the length of miRNAs from *C. elegans*, we examined some published small RNA-seq datasets (Supplementary Table S1). Because non-templated nucleotide addition to 3’ ends of miRNAs is a common source of size heterogeneity (7), we compared datasets prepared from wild-type worms and from worms lacking a putative nucleotidyltransferase, MUT-2, which is predicted to add nucleotides to the 3’ ends of RNA (37). Surprisingly, we found that the average lengths of many miRNAs were substantially shorter in a dataset obtained from *mut-2(ne3364)* animals (Figure 1A), but to a lesser extent in a dataset obtained from *mut-2(ne298)* animals (data not shown). We also observed a similar shortening of miRNAs in a dataset from wild-type animals undergoing *pos-1* RNAi (Figure 1A) but not in a dataset from animals lacking MUT-16 (Figure 1A), which promotes formation of perinuclear foci where MUT-2 is thought to extend small RNAs (46,47). Despite the two datasets that showed an abundant population of shorter miRNAs being generated in different labs (see Supplementary Table S1), the datasets shared a subset of miRNAs that were substantially truncated. These included all members of an essential miRNA family in *C. elegans*—miR-51 to miR-56 (48) (Figure 1B for *mut-2(ne3364)* and wild-type animals undergoing RNAi). Furthermore, miRNA truncation was specific to the 5’ end of the more frequently sequenced arm (miRNA-3p or miRNA-5p) of the miRNA precursor (Figure 1C and Supplementary Figure S1).

An abundant 5’-truncated miRNA observed in the *mut-2(ne3364)* RNA-seq dataset is not detectable by northern blotting in newly generated *mut-2(jam9)* null mutants

To test if the wild-type activity of MUT-2 is required to protect miR-51 to miR-56 family from 5’ truncation, we examined levels of the most highly expressed and most severely truncated member, miR-52, using northern blotting of total RNA prepared from *mut-2(-)* animals. To control for mutations induced by unregulated transposon activity in *mut-2*
results suggest that MUT-2 is not required for the expression that was observed in RNA-seq datasets. This result is in contrast to the 9 nt truncated miR-52 variants that were not detectable in mut-2(null mutant) animals (50). We analyzed independent passages of a newly generated mut-2 null mutant. Specifically, we replaced the entire coding portion of mut-2 with DNA encoding gfp (Supplementary Figure S2A), which resulted in the expression of gfp instead of mut-2 (Supplementary Figure S2B).

When miR-52 was examined in three independent passages of the mut-2 null mutant using northern blotting, no evidence for shortening was detectable (Figure 2). For each passage, all detectable miR-52—from wild-type and from mut-2(null mutant) animals—was ~24 nt long, which is the annotated length of miR-52. This result is in contrast to the 9 nt truncation that was observed in ~94% of the miR-52 reads by RNA-seq of mut-2(null mutant) animals (Supplementary Figure S2B). These results suggest that MUT-2 is not required for the expression of full-length miR-52.

5′ truncated miRNAs are likely degradation products that arose during RNA-seq

Our failure to detect 5′-truncated miR-52 in mut-2 null mutants raised the possibility that the apparent 5′ truncation observed in RNA-seq datasets results from the preparatory procedures required for RNA-seq. Consistently, we observed 5′-truncated miR-52 in 17 of 59 datasets with prevalence ranging from 79 to 19% of the miR-52 reads (Supplementary Figure S3A). Such 5′ truncation of miR-52 was not restricted to datasets from any lab, developmental stage, sex or cloning method (Supplementary Figure S3A and Table S1). Additionally, all variations of circularly permuted miR-52 could also be detected (Supplementary Figure S3B). The simplest explanation for the presence of these circularly permuted sequences could be that the miRNA was circularized and then randomly cleaved along the phosphate backbone (Supplementary Figure S3C). If the linearized forms had the necessary 5′- and 3′-functional groups, it could then be captured in the RNA-seq library cloning protocol. Permuted miR-52 variants were rare compared to the total number of miR-52 reads but some forms were nevertheless sequenced more frequently than many other annotated miRNAs. Taken together, these observations suggest that RNA-seq datasets contain 5′-truncated and other processed versions of miRNAs whose provenance is currently unclear.

Northern blotting using miRNA-length probes is biased against detecting short RNAs

To complement RNA-seq data, additional methods that effectively detect small RNAs are needed. Although the average size of miRNAs is ~23 nt, miRbase (V.21) contains a wide range of miRNA sizes (from the 15 nt miR-7238-3p from Mus musculus to the 34 nt miR-5971 from Haemonchus contortus) with varying degrees of functional validation. Therefore, a method that can effectively detect RNAs in the range of 15 to 34 nt would be useful to independently evaluate the expression of some miRNAs and other short RNAs like the truncated variants that we detected in RNA-seq datasets.

When examining the levels of the 24 nt long miR-52, we made an observation that suggests that typical northern blotting has a strong bias against shorter sequences. The 24 nt probe was able to detect 10 fmol of 24 nt DNA analogue of miR-52 effectively but barely detected 50 fmol of 16 nt DNA representing 5′-truncated miR-52 (Figure 2, DNA oligos lane). Because both these DNA sequences were synthetic, the difference in their detection is likely due to differences in the length of the sequence and not due to any chemical changes to the nucleotides that could have occurred in vivo. Thus, northern blotting needs to be modified for effectively detecting short RNAs.

Northern blotting can be improved to reduce bias against short RNAs and maintain target specificity

A previous study reported the detection of a specific ~14 nt RNA called pRNA in samples prepared from two species of bacteria using a modified northern blotting protocol (51). This approach used LNA probes that are 5′ digoxigenin-labeled, chemical crosslinking of RNA to nylon membranes using 1-ethyl-3(3-dimethylaminopropyl)-carbodiimide (EDC), native PAGE and hybridization at 50 to 68°C. While this approach is powerful, the authors noted a few shortcomings: (i) LNA probes are expensive and detection is affected by the distribution of LNA modifications on the probe; (ii) EDC linking requires a terminal phosphate, in the absence of which, the signal was reduced by ~10-fold; and (iii) to avoid EDC crosslinking to urea, RNA samples have to be run on native PAGE, which can affect the migration of small RNAs with secondary structure or that interact with complementary sequences (e.g. stable 6S RNA:pRNA hybrids). Therefore, we wondered whether the detection of miRNAs as well as truncated forms of miRNAs or other such shorter RNAs using cheaper DNA probes and under denaturing conditions could be achieved by UV crosslinking and performing hybridization at lower temperatures.

To begin testing this possibility, we used two different length probes to detect progressively 5′-shortened versions of synthetic miR-52 and miR-53. In addition to the 24 nt.
probes, we designed shorter probes (15 nt for miR-52 and 16 nt for miR-53) that target the 3' end of both miRNAs and include the single nucleotide polymorphism that distinguishes miR-52 from miR-53 (Figure 3A). To account for the differential base pairing potential of different length probes, we hybridized at 42°C for 24 nt probes and at room temperature (~22°C) for short probes. The 24 nt probe could only detect the 24 to 18 nt versions of the miRNAs, with rapidly diminishing signal intensity for shorter RNAs (Figure 3A and B, black). However, the short probes were able to detect all of the different length RNAs with comparable efficiency (Figure 3A and B, red). The stark bias against shorter RNAs when 24 nt probes were used (e.g. ~30-fold for 18 nt miR-53 and ~1000-fold for 16 nt miR-52) was greatly reduced when 16 or 15 nt probes were used (a maximum of ~7-fold difference across 24 to 14 nt).

Another important consideration for northern blotting is specific recognition of a target RNA. We measured the specificity of short and long probes by comparing their ability to detect perfectly complementary RNA (target) with that to detect RNA that differed by a single nucleotide. The 24 nt miR-52 probe showed almost equal reactivity to 24 nt miR-52 and miR-53 (Figure 3A), consistent with previous reports of cross-reactivity (52). The 24 nt miR-53 probe, which was subjected to washes at higher temperature (42°C washes compared to 37°C used for miR-52), showed reduced cross-reactivity. However, the increased stringency of the washes increased the bias against detecting shorter RNAs as evidenced by the exponential decrease in signal as the length of target RNA decreased (Figure 3A and B). In contrast, short probes were able to detect all targets with specificity comparable to full-length probes (Figure 3A–C; Supplementary Figures S4 and 5). Finally, short probes were able to specifically detect different small RNAs of varying GC content (Figure 3D). A 15 nt probe with only 27% GC could still bind specifically to its miR-1 target with no cross-reactivity to three other abundant synthetic RNAs on the same membrane (Figure 3D).

Thus, short probes and hybridization at low temperatures can reduce the bias against the detection of short RNAs that arise when long probes and typical hybridization temperatures are used.

**Fifteen nucleotide probes can specifically detect miRNAs in total RNA from C. elegans and pRNAs in total RNA from E. coli**

Although our method demonstrates the specific detection of synthetic RNAs, it will be of limited utility unless it can be applied to detect specific sequences in complex mixtures of RNA. Therefore, we attempted the detection of three specific miRNAs (miR-1, miR-42 and miR-58) with 15 nt probes of varying GC content in preparations of C. elegans total RNA. These miRNAs were chosen based on the availability of viable deletion mutants that can be used as essential controls and to cover a range of GC content (miR-1, 27% GC; miR-52, 40% GC and miR-58, 47% GC). Total RNA from wild-type worms and from mutant worms with a deletion for a miRNA gene was purified and used for northern blotting. Both miR-58 and miR-52 were detected in total RNA from wild-type worms but not in total RNA from the corresponding mutants (Figure 4). However, miR-1 could not be detected in any case despite seven attempts (data not shown) and could only be detected using a 17 nt probe (Supplementary Figure S6), possibly because the low GC content of the 15 nt probe used (27% GC) is below the threshold for stable binding at room temperature.
Figure 4. Short probes can specifically detect miRNAs in total RNA. Total RNA prepared from wild-type and miRNA deletion (mir-58(+/n4640), mir-52(+/n4114) and mir-1(+/n4101)) worm strains was used for northern blotting with short (15 nt) probes against mir-58 or mir-52. Synthetic RNAs that matched the sequence of full length (10 fmol of 24 nt, weakly detected) and shortened forms (50 fmol of 16 nt, clearly detected) of miR-58 and miR-52 were run as size markers. U6 was probed to indicate loading.

Finally, we set out to test if our method can be applied to detect endogenous RNAs that are smaller than typical miRNAs in complex biological samples. A well-studied example of such biologically important small RNAs is pRNAs (reviewed in 53 and 54). These small RNAs are transcribed using 6S RNA as a template and can be detected shortly after starved cells are shifted to abundant nutrient conditions. Total RNA was purified from wild-type and 6S RNA (ssrS1) null E. coli 3 and 5 min after transferring stationary cells to fresh growth media. A DIG-tailed 15 nt DNA oligo hybridized at room temperature was able to specifically detect pRNAs from 14 nt elongated to 23 nt with no signal seen in the 6S RNA deletion strain (Figure 5). These results demonstrate the utility of the modifications we propose — shorter probes and lower hybridization temperatures—for the effective detection of a specific small RNA in total RNA.

DISCUSSION

We detected unexpected forms of miRNAs in some RNA-seq datasets and found that northern blotting using miRNA-length probes can be biased against detecting shorter RNAs. We demonstrate that adjusting probe size and hybridization conditions can allow equivalent detection of 24 to 14 nt RNAs with single-nucleotide sequence discrimination using affordable DNA oligos as probes.

Possible origins of miRNA variants in RNA-seq datasets

We detect two unusual miRNA variants in RNA-seq datasets—5′ truncated forms of miRNA and circularly permuted forms of miRNA.

The truncated forms of miRNA could arise during sample preparation before RNA extraction. We observed that in most cases truncated miRNAs were missing sequences from the 5′ end (Figure 1 and Supplementary Figure S1). Such truncations were particularly pronounced in all members of the miR-51 family. These observations suggest possible degradation of miRNAs by a 5′ to 3′ exoribonuclease. Consistently, C. elegans has a 5′ to 3′ exoribonuclease called XRN-2 that can act on miRNAs in general (55) and for which the miR-51 family could be an especially good substrate (56). Such mechanisms that can degrade miRNAs raise concerns about the accuracy of inferring the levels of miRNAs that function in vivo from RNA-seq and suggest the use of truncated versions of miR-51 family members as indicators of poor RNA quality.

Circularly permuted forms of miRNA could arise during ligation steps required for library preparation. The number of reads that we count as circularly permuted is likely an underestimation. The ~0.16% of miR-52 detected as circularly permuted in Supplementary Figure S3B reflect only the sub-population of circularized RNAs that became linear and were then successfully joined to sequencing adaptors. Those RNAs that remained as circles or did not get ligated to adaptors would escape detection. Although circular RNAs that result from RNA splicing have been detected in C. elegans (57,58), there is no evidence yet for the generation of circular miRNAs in vivo.

Technical considerations for the use of northern blotting to complement RNA-seq

Our results suggest that the hybridization temperature and GC content of probes are critical variables for the detection of short RNAs. Long probes hybridized at 22°C had off-target binding and retained a bias against short RNAs (Supplementary Figure S4), while short probes hybridized...
at 37–42°C did not have any detectable signal (data not shown). The significance of GC content can be seen in the miRNA northerns. The miR-52 probe with 40% GC had a markedly weaker signal compared to the miR-58 probe with 47% GC (Figure 4)—a difference of one GC base pair. We failed to detect miR-1 using a 15 nt probes with only four GC bases (27%). By using a 17 nt probe while still keeping four GC bases, we were able to successfully detect miR-1 in total RNA (Supplementary Figure S6), suggesting a minimum nucleotide composition and length requirement to use this northern blotting method. We also noted that incubation below 20°C results in components of the hybridization buffer becoming insoluble limiting how low we can adjust temperature to promote stable probe hybridization.

Low incubation temperatures and appropriate washing conditions can enable short sequences to specifically detect complementary RNA. This specificity was despite the addition of a long tail (>100 nt) of random dA and DIG-conjugated dU to generate the DIG-tailed probe. Mono-DIG-labeled 15 nt DNA probes failed to detect miR-52 and mono-DIG-labeled 24 nt DNA probes detected miR-52 with a weaker signal than DIG-tailed probes (Supplementary Figure S4). Thus, using DIG-tailed probes reduces cost (~20× compared with mono-DIG probes and ~150× compared to DIG-LNA probes) and enhances the detectable signal without apparent off-target interactions from the long tail.

Applications of the modified northern blotting described in this study

The drastic bias against short RNAs observed when miRNA-length probes are used (Figure 3A), suggest that the careful study of isomiRs (12) will require designing probes that match the shortest form of the miRNA. Furthermore, this bias raises the possibility that RNA quantified by northern blotting in previous studies might have under-reported RNAs that are shorter than the length of the probe. Such shorter RNAs include biologically relevant RNAs such as pRNAs (~14 nt) in bacteria (59), NU RNAs (<18 nt) in C. elegans (60) or unusually small RNAs (~17 nt) that were reported to associate with Argonaute proteins in mammalian cells (61). Unlike a previous method that detected pRNAs (51), our approach does not need to sacrifice specificity for sensitivity. The ability to better discriminate RNAs that differ by single nucleotides when using short probes (Figure 3C) could help resolve expression from two different alleles without the amplification necessary for RNA-seq (62) or PCR (63) approaches. Finally, reducing the bias against shorter RNAs could enable tracking intermediates of RNA processing or small RNA turnover (64) in mechanistic studies and is thus an essential consideration when probing the realm of tiny RNAs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank members of the Jose lab for critical reading of the manuscript, the C. elegans Genetic stock Center for worm strains and Karen Wasserman (University of Wisconsin, Madison) for bacterial strains. This manuscript has been improved with comments from anonymous reviewers who helped place this work in the context of other efforts underway to improve the detection of small RNAs.

FUNDING

National Institutes of Health [R01GM111457 to A.M.J.]; UMD Postdoctoral Fellowship [CMNS/CBMG to Y.S.C.]. Funding for open access charge: National Institutes of Health [R01GM111457].

Conflict of interest statement. None declared.

REFERENCES

1. Hammond,S.M. (2015) An overview of microRNAs. Adv. Drug Deliv. Rev., 87, 3–14.
2. Oszolak,F. and Milos,P. (2011) RNA sequencing: advances, challenges and opportunities. Nat. Rev. Genet., 12, 87–98.
3. Li,Y. and Kowdley,K. (2012) MicroRNAs in common human diseases. Genomics Proteomics Bioinformatics, 10, 246–253.
4. Williams,Z., Ben-Dov,I., Elias,R., Mihailovic,A., Brown,M., Rosewaks,Z. and Tuschl,T. (2013) Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. Proc. Natl. Acad. Sci. U.S.A., 110, 4255–4260.
5. Cheng,L., Sharples,R., Siciluna,B. and Hill,A. (2014) Exosomes provide a protective and enriched source of miRNA for biomarker profiling compared to intracellular and cell-free blood. J. Extracell. Vesicles, 3, 23743.
6. Pritchard,C.C., Cheng,H.H. and Tewari,M. (2012) MicroRNA profiling: approaches and considerations. Nat. Rev. Genet., 13, 358–369.
7. Landgraf,P., Rusu,M., Sheridan,R., Sewer,A., Iovino,N., Aravin,A., Pfeffer,S., Rice,A., Kamphorst,A., Landthaler,M. et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. Cell, 129, 1401–1414.
8. Morin,R., O’Connor,M., Griffith,M., Kuchenbauer,F., Delaney,A., Prabhu,A., Zhao,Y., McDonald,H., Zeng,T., Hirst,M. et al. (2008) Application of massively parallel sequencing to miRNA profiling and discovery in human embryonic stem cells. Genome Res., 18, 610–621.
9. Lagos-Quintana,M., Rauhut,R., Lendeckel,W. and Tuschl,T. (2001) Identification of novel genes coding for small expressed RNAs. Science, 294, 853–858.
10. 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.
11. Lee,R.C. and Ambros,V. (2001) An extensive class of small RNAs in Caenorhabditis elegans. Science, 294, 862–864.
12. Neilson,C.T., Goodall,G.J. and Bracken,C.P. (2012) IsomiRs—the overlooked repertoire in the dynamic microRNAsome. Trends Genet., 28, 544–549.
13. Lee,M., Choi,Y., Kim,K., Jin,H., Lim,J., Nguyen,T.A., Yang,J., Jeong,M., Giraldez,A.J., Yang,H. et al. (2014) Adenylation of maternally inherited microRNAs by Wispy. Mol. Cell, 56, 696–707.
14. Katoh,T., Hojo,H. and Suzuki,T. (2015) Destabilization of microRNAs in human cells by 3' deadenylation mediated by PARN and CUGBP1. Nucleic Acids Res., 43, 7521–7534.
15. Marzi,M.J., Ghini,F., Cerruti,B., de Pretis,S., Bonetti,P., Giacomelli,C., Gorski,M.M., Kress,T., Pelizola,M., Muller,H. et al. (2016) Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. Genome Res., 26, 554–565.
16. Lewis,B.P., Shih,I.H., Jones-Rhoades,M.W., Bartel,D.P. and Burge,C.B. (2003) Prediction of mammalian microRNA targets. Cell, 115, 787–798.
17. Fukunaga,R., Han,B., Hung,J., Xu,J., Weng,Z. and Zamore,P. (2012) Dicer partner proteins tune the length of mature miRNAs in flies and mammals. Cell, 151, 533–546.
21. Raabe, C.A., Tang, T.-H., Brosius, J. and Rozhdestvensky, T.S. (2014) Monouridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. Cell, 151, 521–532.

22. Maran, M., Polakowski, M., Sarnik, A., Rea, D., Yu, Y. and Grzegorczyk, I. (2014) S. mormorium isomiR variation is functional and evolutionary important. Nucleic Acids Res., 42, 9424–9435.

23. Válczi, A., Hornyik, C., Varga, N., Burgyán, J., Kauppinen, S. and Hafner, M., Renwick, N., Brown, M., Mihailović, A., Holoch, D., Hafner, M., Renwick, N., Brown, M., Mihailović, A., Holoch, D.,GIardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, B., Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, B., Holloway, J., Nusbaum, J., Morozov, P., Ludwig, J. et al. (2011) RNA-ligase-dependent biases in microRNA representation in deep-sequenced small RNA cDNA libraries. RNA, 17, 1697–712.

24. Hafner, M., Polakowski, M., Sarnik, A., Rea, D., Yu, Y. and Grzegorczyk, I. (2014) S. mormorium isomiR variation is functional and evolutionary important. Nucleic Acids Res., 42, 9424–9435.

25. Kim, Y., Yeo, J., Kim, B., Ha, M. and Kim, V.N. (2012) Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. Cell, 151, 521–532.

26. Maran, M., Polakowski, M., Sarnik, A., Rea, D., Yu, Y. and Grzegorczyk, I. (2014) S. mormorium isomiR variation is functional and evolutionary important. Nucleic Acids Res., 42, 9424–9435.
61. Li, Z., Kim, S.W., Lin, Y., Moore, P.S., Chang, Y. and John, B. (2009) Characterization of viral and human RNAs smaller than canonical MicroRNAs. *J. Virol.*, **83**, 12751–12758.

62. Wang, X. and Clark, A.G. (2014) Using next-generation RNA sequencing to identify imprinted genes. *Heredity (Edinb)*., **113**, 156–166.

63. Strerath, M., Detmer, I., Gaster, J. and Marx, A. (2007) Modified oligonucleotides as tools for allele-specific amplification. *Methods Mol. Biol.*, **402**, 317–328.

64. Rüegger, S. and Großhans, H. (2012) MicroRNA turnover: when, how, and why. *Trends Biochem. Sci.*, **37**, 436–446.