Published in final edited form as:
Methods. 2008 January ; 44(1): 3–12. doi:10.1016/j.ymeth.2007.09.009.

Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing

Markus Hafner[1], **Pablo Landgraf**[1], **Janos Ludwig**[1], **Amanda Rice**[1], **Tolulope Ojo**[1], **Carolina Lin**[1], and **Thomas Tuschl**[1,*]

[1]Howard Hughes Medical Institute, Laboratory of RNA Molecular Biology, The Rockefeller University, 1230 York Avenue, Box 186, New York, NY, 10065, USA

Abstract

Distinct classes of small RNAs, 20 to 32 nucleotides long, play important regulatory roles for diverse cellular processes. It is therefore important to identify and quantify small RNAs as a function of development, tissue and cell type, in normal and disease states. Here we describe methods to prepare cDNA libraries from pools of small RNAs isolated from organisms, tissues or cells. These methods enable the identification of new members or new classes of small RNAs, and they are also suitable to obtain miRNA expression profiles based on clone count frequencies. This protocol includes the use of new deep sequencing methods (454/Roche and Solexa) to facilitate the characterization of diverse sequence pools of small RNAs.

Introduction

Small non-coding RNAs play a vital regulatory role in cells (reviewed in 1–7). The most abundant small RNAs in animals are 20 to 23 nucleotide (nt) long microRNAs (miRNAs). The first miRNA members were identified in *C. elegans* (8, 9). The discovery that double-stranded RNA (dsRNA) triggered RNA interference (RNAi) (10) was also mediated by similar-sized small RNA processing products, known as small interfering RNAs (siRNAs) (11–14), prompted the development of techniques to characterize naturally occurring small RNAs (15–17). These methods were based on small RNA cDNA library preparation and sequencing and ignited the discovery of new members and families of small RNAs (15–24).

Small RNAs, in association with their protein effector components, mediate sequence-specific posttranscriptional and transcriptional gene regulation. They control mRNA translation, stability and localization (reviewed in 25, 26) and feed into processes that control transposons (reviewed in 27, 28) and heterochromatin structure (reviewed in 4, 29). This wide range of functions stimulated great interest to identify and characterize the small RNAs expressed in different organisms, tissues and cell types, in normal and disease states.

Here we describe our protocols for the construction of small RNA libraries and their adaptation for various high throughput sequencing approaches. The protocols originate from methods

*corresponding author (ttuschl@rockefeller.edu, +1 212 327 7651, fax: +1 212, 327 7652).

described previously (30–32) and provide new details regarding the use of RNA ligases and the latest sequencing technology.

Description of Method

Overview

The experimental process is outlined in Figure 1, and includes the steps of small RNA isolation, cDNA library preparation, and sequencing. The annotation of the identified sequences is described in detail in an accompanying paper (Zavolan et al.).

We first isolate total RNA using the standard acidic guanidinium isothiocyanate/phenol/chloroform (GITC/phenol) extraction methods (33). Subsequently, we isolate small RNAs of the desired size ranges using denaturing polyacrylamide gel electrophoresis. Alternatively, classes of small RNAs may be isolated from lysates of fresh samples by immunoprecipitation using antibodies raised against the proteins associated with these specific classes of small RNAs (34–38).

To prepare cDNA from the isolated small RNAs, we first ligate synthetic oligonucleotide adapters of known sequence to the 3' and 5' ends of the small RNA pool using T4 RNA ligases. The adapters introduce primer-binding sites for reverse transcription and PCR-amplification. If desired, non-palindromic restriction sites present within the adapter/primer sequences can be used for generation of concatamers to increase the read length for conventional sequencing.

One of the characteristics of most classes of small regulatory RNAs is the presence of a 5' phosphate and a 3' hydroxyl group. RNA turnover products and RNase degradation products instead carry 5' hydroxyl groups and 2' or 3' phosphates. The protocol we describe is designed to specifically isolate small RNAs with 5' phosphate and 3' hydroxyl termini. However, precautions have to be taken to prevent circularization of 5' phosphate/3' hydroxyl small RNAs during adapter ligation (30). 1. We use chemically pre-adenylated 3' adapter deoxyoligonucleotides, which are blocked at their 3' ends to avoid their circularization. The use of pre-adenylated adapters eliminates the need for ATP during ligation, and thus minimizes the problem of adenylation of the pool RNA 5' phosphate that leads to circularization. 2. We use a truncated form of T4 RNA ligase 2, Rnl2(1–249), and more recently an improved mutant, Rnl2(1–249)K227Q, to minimize adenylate transfer from the 3' adapter 5' phosphate to the 5' phosphate of the small RNA pool and subsequent pool RNA circularization.

The recent introduction of massive parallel sequencing technology enabled the sequencing of hundreds of thousands to tens of millions of small RNA cDNA clones. This drastic technical improvement facilitated the identification of new small RNAs, and increasing clone counts allowed the determination of small RNA relative expression levels based on clone frequencies. These new methods include pyrosequencing (454-sequencing, Roche), which provides up to 400,000 sequences of up to 250 nt in length for a single read (39), and sequencing-by-synthesis (Solexa), which provides up to 30,000,000 sequences of up to 50 nt in length for a single read (40, 41). For cost-effective pilot sequencing, which is recommended for assessing the quality of a library preparation before expensive deep sequencing, it may be convenient to produce sequence concatamers from the PCR product of the library. Conventional sequencing from concatamer clones can yield more than a dozen different small RNA sequences. Data management and sequence analysis from small RNA cDNA libraries is best carried out in collaboration with an experienced computational biology laboratory (see accompanying paper, Zavolan at al.).

Step-by-Step protocol

Isolation of total RNA—Total RNA from tissue or cultured cells is isolated either by the GITC/phenol method, see below, or by using the related commercial Trizol (Invitrogen) reagent. There are column-based RNA-isolation kits available but not all are suitable to recover the small RNA fraction. Under all circumstances avoid procedures typically used to purify mRNAs including aqueous LiCl precipitation, as small RNAs, including tRNAs, are lost.

All reagents should be RNase free. RNA in solution is stored frozen at −20°C or below and kept on ice while reactions are being set up to minimize hydrolysis.

As starting material for gel purification of small RNAs, we recommend to use 50 to 100 μg of total RNA, although we have repeatedly generated libraries from as little as 5 μg of total RNA using the same protocol. The total RNA is either isolated from freshly collected cultured cells, freshly harvested tissues, or flash-frozen samples, which were stored below −70°C. As a rule of thumb, 1 g of tissue or cells will yield about 1 mg of total RNA. The protocol we provide works well for isolating RNA from cultured cells and most tissues, though certain tissues, such as skin or fat, may require special procedures for rupturing the tissue or dealing with unusual amounts of lipids, respectively.

1. Cultured cells are detached from the culture plate, collected by centrifugation for 5 min at 500×g at 4°C and washed with phosphate-buffered saline (PBS). The cell pellet is resuspended in PBS and transferred to a preweighed 1.5 ml Eppendorf tube and centrifuged again for 5 min at 500×g. After removal of the supernatant, the weight of the cell pellet is determined. If RNA is extracted from tissue, determine the weight of the tissue slice. If you are isolating RNA from more than approximately 0.1–0.2 g of cells (expected yield about 0.1–0.2 mg total RNA) please adjust volumes and tube sizes.
2. Prepare the RNA extraction solution by mixing one volume of denaturing solution (4 M guanidinium isothiocyanate, 25 mM sodium citrate, 0.5 % N-lauroylsarcosine, 50 mM 2-mercaptoethanol) with one volume of acidic water-saturated phenol, pH 4.3. Add 5 ml extraction solution per g of cells to the cell pellet. Transfer the viscous mixture immediately to the homogenizer and homogenize completely keeping the sample on ice. Cells are easily homogenized using a Dounce glass homogenizer. Tissues are preferably homogenized using first a Polytron homogenizer (Kinematica), followed by douncing.
3. Add 1/10 volume of chloroform:isoamyl alcohol (24:1) and 1/20 volume of 2 M sodium acetate (pH 4.2). Homogenize the mixture by douncing as before until the mixture is turbid-white. Transfer the mixture to an Eppendorf tube and centrifuge for 5 min at 20,000×g. Typically a white interphase, composed of mostly DNA, forms between the lower organic and upper aqueous phase. Transfer the aqueous phase to a new tube with as little white interphase as possible. Extract with 1/2 volume of acid–buffered phenol:chloroform:isoamyl alcohol (25:24:1) and centrifuge. If necessary, repeat the extraction of the aqueous phase until there is no more white interphase and the aqueous phase remains clear. Finally, extract the aqueous phase once with 1/2 volume chloroform. Transfer the upper (aqueous) phase to a new tube.
4. Add 3 volumes of ethanol to precipitate the RNA for 1 h on ice or overnight at −20° C. Collect the RNA pellet at 4°C (15 min at 20,000×g) and remove the supernatant completely. Unless the pellet is very big and not all of the supernatant liquid can be removed easily, do not rinse the pellet with 300 μl 75% ethanol. If you have to wash, add the cold 75% ethanol, carefully invert the tube once to rinse the tube walls and

then collect the pellet by a 1 min spin at maximum speed. Remove the supernatant and dissolve the pellet in 100 µl water or formamide.

5. Determine the concentration of the RNA by measuring the absorbance of a 1:100 dilution at 260 nm. If using a quartz cuvette with 1 cm diameter, the concentration c of total RNA (in µg/µl) is roughly c = A(260 nm) * f * 0.04 µg/µl; whereby f is the dilution factor. The yield of total RNA should be about 1 mg per gram of tissue or cells, however, the yield may vary between different tissue types.

6. Control the quality of the RNA by separating 1 µg on a 1.5 % agarose gel using 0.5× TBE (50 mM Tris, 50 mM boric acid, 1 mM EDTA) buffer. If isolating RNA from eukaryotic cells the 28S and 18S rRNAs (5.0 and 1.9 kb, respectively) should be intact and visible as two sharp bands. Degradation will result in increased accumulation of rRNA and mRNA fragments in the library obstructing the characterization of regulatory small RNAs.

Gel purification of small RNAs—The total RNA is size-fractionated on a denaturing polyacrylamide gel and the small RNAs are then eluted from the excised gel slice. The size of the RNA is best determined by adding a trace amount of radioactively labeled RNA size markers to the total RNA before gel-separation. The radioactive bands of the marker oligonucleotides are visualized on a phosphorimager screen or X-ray film to identify the gel piece that contains the RNA of the desired length.

We use the following 5' ^{32}P-labeled oligoribonucleotide markers in our experiments:

19-nt size marker: CGUACGCGG*GUUUAAAC*GA

24-nt size marker: CGUACGCGGAAUA*GUUUAAAC*UGU

33-nt size marker: CAUCUUGGUCGUACGCGGAAUA*GUUUAAAC*UGU

35-nt size marker: CUCAUCUUGGUCGUACGCGGAAUA*GUUUAAAC*UGU

The size markers contain a 8-nt *PmeI* digestion site (shown in italic). To prevent the accumulation of marker sequences in the cDNA library, we digest the PCR-amplified library with *PmeI* before cloning and sequencing.

Importantly, use siliconized tubes (we use BioPlas 1.5 ml PP tubes, cat. no. 4165SL) for all manipulations of the small RNAs after size-fractionation of the total RNA. The minute amounts of small RNAs to be recovered after gel purification will readily adsorb to the walls of standard tubes.

7. Add one volume of formamide loading solution (50 mM EDTA, 0.05 % (w/v) bromophenol blue in formamide) to the solution of total RNA. Add 50 fmol of 5'-^{32}P-labeled RNA size markers cocktail. *Size markers are radiolabeled individually in a 10-µ l reaction at 1 µ M concentration by T4 polynucleotide kinase (NEB) using 50 µ Ci γ-^{32} P-ATP (see manufacturer's instructions). After gel purification of the markers, dissolve each of the marker oligoribonucleotides in 20 µl of formamide. Use 1-µl aliquots to prepare the size marker cocktail and dilute it 1:10 in formamide*. Denature the RNA by incubation for 30 s at 90°C and load the RNA on a 15% denaturing polyacrylamide gel. Avoid overloading the pocket and the gel to ensure good fractionation of the RNAs. We load 50 – 100 µg of total RNA per well of a 1.5 mm thick 15×17 cm size gel (10 well comb; total gel volume 50 ml). It is recommended to load 50 pmol of unlabeled size marker stock together with 50 fmol radioactively labeled size markers diluted in a total volume of 40 µl with formamide gel-loading solution on each side of the gel (20 µl each lane) and to carry these samples through the adapter ligation process and reverse transcription as positive controls.

Make sure to leave at least one well empty between the different samples to avoid cross-contamination. We run our 50 ml gels at 30 W using 0.5× TBE buffer until the bromophenol blue dye of the gel-loading solution appears within the lower third of the gel. It is not recommended to run the gel further than necessary to separate the size markers. This ensures that the sample is contained within as small as a gel slice as possible.

8. Dismantle the gel, leaving it mounted on one glass plate. To facilitate the alignment of the gel to the phosphorimager paper printout, we recommend implanting three tiny radioactive gel pieces asymmetrically at three of the four corners of the gel. Radioactive gel pieces can be collected from the gel that was used to purify the size markers after ^{32}P-labeling. Wrap the gel in plastic film (e.g. Saran wrap) to avoid contamination and expose it to a phosphorimaging screen for 45 min. Print out a 100%-scaled image of the gel, align the gel on top of the printout according to the position of the three radioactive gel pieces. Excise the bands defined by the mobility of the RNA size markers. Cut the gel slice in smaller pieces so they can fit into a preweighed siliconized 1.5 ml tube. Determine the weight of the gel slices. Add 2 to 3 volumes (v/w) of RNase-free 0.4 M NaCl and elute the small RNAs from the gel by incubating the tube overnight at 4°C under constant agitation. Collect the supernatant and precipitate the small RNAs for at least 2 h on ice or overnight at −20° C after the addition of 3 to 4 volumes of absolute ethanol

9. Collect the small RNA pellet after ethanol precipitation in a tabletop centrifuge for 15 min at maximum speed (~14,000×g) at 4°C. Remove the supernatant and collect the residual liquid at the bottom of the tube by an additional 5 s centrifuge spin. Remove the residual liquid completely using a small pipette tip without perturbing the pellet. The additional spin is needed to collect all residual liquid. If necessary air-dry the RNA pellet to evaporate residual ethanol. Be sure that the ethanol has been evaporated as it may inhibit the subsequent enzymatic steps. All subsequent RNA precipitations (see below) are performed similarly.

Adapter ligations—Adapters need to be joined to the small RNA pool to allow for RT/PCR amplification. The adapters used to introduce the constant regions and their corresponding PCR primers vary, depending on the sequencing method that is being used.

For the concatamerization approach we use the following adapter and primer sequences:

Adapter set	3' adapter	AppTTTAACCGCGAATTCCAG-L
	5' adapter	ACGGAATTCCTCACTrArArA
First PCR	5' primer	CAGCCAACGGAATTCCTCACTAAA
	3'primer	GACTAGCTGGAATTCGCGGTTAAA
second PCR	5' primer	CAGCCAACAGGCACCGAATTCCTCACTAAA
	3'primer	GACTAGCTTGGTGCCGAATTCGCGGTTAAA

A,C,G,T, DNA residues; rA, rU, rC, rG, RNA residues; L, 3'OH blocking group; underlined: non-palindromic *BanI* recognition site

For pyrosequencing (39) we use the same adapter set and first PCR primer set as described above and introduce the recognition tag in the second PCR with the following primers:

second PCR	5' primer	GCCTCCCTCGCGCCA**TCAG**CGGAATTCCTCACTAAA
	3'primer	GCCTTGCCAGCCCGC**TCAG**TGGAATTCGCGGTTAAA

Bold: 454 recognition tag

The adapter and the primer sets for Solexa high-throughput sequencing-by-nucleotide are different because this method does not yet allow for reads similar in size to 454. Therefore, the sequencing primer-binding site has to be already present in the 5' adapter.

Adapter set	3' adapter	AppTCGTATGCCGTCTTCTGCTTG-L
	5' adapter	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC
PCR	5' primer	AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA
	3'primer	CAAGCAGAAGACGGCATACGA

A,C,G,T, DNA residues; rA, rU, rC, rG, RNA residues; L, 3'OH blocking group

It is also possible to convert a Solexa library for 454 sequencing by introducing 454 sequencing primer binding sites by a 2nd PCR analogous to the conversion of our standard library to 454 sequencing described above. For concatamerization of a Solexa PCR product, BanI restriction sites can to be placed analogous to the 2nd PCR described for our standard library preparation.

3' Adapter ligation: The 3' adapter ligation is performed using chemically pre-adenylated oligodeoxynucleotides. The adenylation reaction is described below in detail in the section "Synthesis of the 5' adenylated 3' adapter oligonucleotide". It is adapted from the original synthesis by Lau et al. (17) following methods previously published by Mukaiyama (42) and Orgel (43).

10. Dissolve the RNA pellet in 10.5 µl water.
11. Prepare a reaction mixture for ligation of the 3' adenylated adapter by combining the following components: 2 µl of 10× RNA ligase buffer without ATP (0.5 M Tris-HCl, pH 7.6; 0.1 M $MgCl_2$; 0.1 M 2-mercaptoethanol; 1 mg/ml acetylated BSA (Sigma, B-8894)), 6 µl 50% aqueous DMSO, 0.5 µl 100 µM adenylated 3' adapter oligodeoxynucleotide. Add the 8.5 µl of the mixture to the sample.
12. Denature the RNA by incubating the tube for 30 s at 90°C. Place the tube immediately on ice for 20 s.
13. Add 1 µl of Rnl2(1–249) (1 µg/µl) or Rnl2(1–249)K227Q (1 µg/µl), mix gently, and incubate overnight at 0°C. Rnl2(1–249) is commercially available from New England Biolabs; our plasmid for expression of the his-tagged mutant Rnl2(1–249)K227Q, pET16b-Rnl2(1–249)K227Q, can be obtained from www.addgene.com.
14. Add 20 µl of gel loading solution (see above) and load the samples in two adjacent wells of a 20-well 15% acrylamide gel (15 cm×17 cm×0.8 mm; 30 ml gel volume). Make sure to space different samples appropriately, typically at a two-well distance, to avoid cross contamination. Run the gel for 1 h at 30 W using 0.5× TBE buffer until the bromophenol blue dye is close to the bottom of the gel. Image the gel as described in step 8 and excise the ligation product. Elute the ligation product from the gel slices overnight at 4°C with constant agitation with at least 3 to 4 volumes of 0.4 M NaCl.
15. Ethanol-precipitate and collect the RNA as described in step 9.

5' Adapter ligation

16. Dissolve the pellet in 9 µl water.

17. Prepare the following reaction mixture by combining 1 µl of 100 µM 5' adapter oligonucleotide (see sequences above), 2 µl of 10× RNA ligase buffer with ATP (0.5 M Tris-HCl, pH 7.6; 0.1 M $MgCl_2$; 0.1 M 2-mercaptoethanol; 1 mg/ml acetylated BSA (Sigma, B-8894); 2 mM ATP) and 6 µl 50% aqueous DMSO. Add the 9 µl of this mixture to the sample.

18. Denature the RNA by incubation for 30 s at 90°C. Place the tube immediately on ice for 20 s.

19. Add 2 µl of T4 RNA ligase 1 (Rnl1) (Fermentas), mix gently, and incubate for 1 h at 37°C.

20. Add 20 µl of gel loading solution (see above) and load the samples in two adjacent wells of a 20-well 15% acrylamide gel (15 cm×17 cm×0.8 mm; 30 ml gel volume). Make sure to space different samples appropriately, typically at a two-well distance, to avoid cross contamination. Run the gel for 1 h at 30 W using 0.5× TBE buffer until the bromophenol blue dye is close to the bottom of the gel. Image the gel as described in step 8 and excise the new ligation product. Elute the ligation product from the gel slices overnight at 4°C with constant agitation using at least 3 to 4 volumes of 0.4 M NaCl. Add 1 µl of 100 µM 3' primer (see above) during the elution as carrier to facilitate the recovery of the ligation product.

21. Ethanol-precipitate and collect the RNA as described in step 9.

Reverse transcription

22. Dissolve the pellet in 5.6 µl water. Prepare the following reaction mix: 1.5 µl 0.1 M DTT, 3 µl 5× first-strand buffer (provided by the manufacturer), 4.2 µl 10× dNTPs (10× concentrations are 2 mM each).

23. Denature the RNA by incubating the tube for 30 s at 90°C and transfer the tube to a 50°C incubator.

24. Add 8.7 µl of the reaction mix to each sample and incubate for 3 min at 50°C. Add 0.75 µl of Superscript III reverse transcriptase and incubate for 30 min at 42°C.

25. To hydrolyze the RNA, add 40 µl of 150 mM KOH/20 mM Tris base and incubate for 10 min at 90°C.

26. Neutralize the solution by addition of 40 µl of 150 mM HCl and check the pH value of the mixture by spotting 1 µl on pH paper. It should be between 7.0 and 9.5 so that the subsequent PCR is not inhibited. If necessary, readjust the pH by adding more base or acid.

PCR amplification

27. Perform a standard 100 µl PCR with *Taq* polymerase (10 µl from the neutralized cDNA solution, 0.5 µM primers, 2 mM $MgCl_2$). To determine the necessary number of cycles for amplifying the cDNA library, remove 12 µl aliquots every other cycle starting with cycle number 15. To remove aliquots from the PCR tube, temporarily pause the PCR cycler at the end of the 72°C step. For PCR, use the following cycle conditions: 45 s at 94°C, 85 s at 50°C, 60 s at 72°C. Do not amplify for more than 30 cycles because it will result in distortion of the small RNA ratios. Analyze the samples on a 2.5 % agarose gel. The PCR-product might appear as a double band with the higher band running at the expected length of about 70–80 nt and a lower band

corresponding to the 3' adapter to 5' adapter ligation products running at about 50 nt. Do not forget to perform a negative control to check for DNA contamination in the reaction mixture. Define the optimal cycle number for cDNA amplification, which has to be within the exponential amplification phase of the PCR, about 5 cycles away from reaching the saturation level of PCR amplification.

28. Perform a 300 µl PCR with the optimal cycle number and analyze the product again on an agarose gel. Purify the PCR-product by phenol/chloroform extraction and subsequent ethanol precipitation.

PmeI digestion—Radioactively labeled size markers that were used as internal standards for small RNA library preparation are removed by a *PmeI* digestion step after the PCR. Be careful not to denature the double-stranded PCR product before or during the *PmeI* digestion. Denaturation and subsequent re-annealing of a complex sequence pool will result in imperfect rehybridization and formation of DNA duplexes with internal bulges that might compromise *PmeI* digestion. The *PmeI* digestion removes all marker sequences. As control, the PCR product obtained from the ligation of adapters to the marker oligonucleotides alone (marker control sample) must be digested completely.

29. Collect the DNA-pellet by centrifugation. In the meantime, prepare a *PmeI* digestion mixture: 2 µl 10× buffer (NEB), 0.2 µl of BSA (10 mg/ml, heat inactivated or acetylated), 0.5 µl of *PmeI* (NEB) and 17.3 µl of water. Remove all of the supernatant but do not dry the pellet completely, as this will cause the DNA to denature. Also, never dissolve the pellet in water in the absence of buffer or salt, as this will denature the duplex DNA. Dissolve the DNA pellet in 20 µl of the *PmeI* digestion mixture and incubate for at least 2 h at 37°C.

30. Run all of the digested sample in one or two wells of a 3% low melting agarose (NuSieve) gel containing 0.4 µg/ml of ethidium bromide in 0.5× TBE running buffer along with a 25 base pair (bp) DNA ladder (Invitrogen) for approximately 1 h at 100 V until the marker bands are sufficiently resolved. Visualize the DNA in the gel using a 360 nm UV transilluminator and excise the band of approximately 70–80 bp size. Transfer the gel slice to a 1.5 ml reaction tube and weigh it. Add 0.4 M NaCl to a total volume of 500 µl (w/v). Incubate at 70°C for 10 min to melt the agarose. Add 1 volume of pre-heated buffered water-saturated phenol (pH 7.8). Vortex the solutions vigorously and immediately separate the phases centrifuging for 5 min at maximum speed in a tabletop centrifuge at room temperature. The agarose will accumulate at the interphase. Collect the aqueous upper phase and extract it once again with 1 volume buffered (pH 8) phenol/chloroform/isoamylalcohol and then with 1 volume chloroform. Precipitate the DNA by addition of 2 volumes of absolute ethanol and incubation for at least 1 h on ice or overnight at −20°C.

Preparation of the library for sequencing—The library is now ready for concatamerization if traditional cloning and sequencing is performed. After an additional PCR amplification the library is restriction digested with *BanI* and the fragments are concatamerized, ligated into a TOPO-TA vector and transformed into bacteria to isolate single colonies and clonal DNA.

With the availability of deep sequencing methods it is possible to avoid the bacterial cloning steps and directly use PCR for clonal amplification. For sequencing by the 454 method the 454 recognition tag as well as the sequencing primer binding sites are introduced by a second PCR using primers that overlap with the first PCR primer sequences. For Solexa sequencing, we cannot simply add a second PCR, as the sequencing read length is limited. Instead, the Solexa adapter and PCR primer set has to be used for ligation and PCR amplification.

Solexa and 454-sequencing: Perform the following PCRs with lower primer concentrations (100 nM) to eliminate the need for removal of unincorporated primer oligodeoxynucleotides.

31. Collect the DNA pellet from step 30 and dissolve it in 30 µl of water. Prepare a 1:100 dilution of the DNA.
32. Use 10 µl of the dilution for a 100 µl pilot standard PCR using *Taq* polymerase (100bnM primers, 2 mM $MgCl_2$) with the following cycle conditions: 45 s at 94°C, 1 min 25 s at 50°C, 1 min at 72°C. Take aliquots after 6, 8, 10, 12, 14 cycles and analyze them on a 2 % agarose gel. Select the number of cycles before reaching saturation, typically after 8 cycles.
33. Perform a 100 µl (10 µl sample, 100 nM primers) PCR with optimal number of cycles and analyze it on an agarose gel.

The sample is now ready for entering the 454 emulsion PCR step (see 454/Roche manufacturer's protocols). If unincorporated primers are detectable as faster mobility bands on the agarose gel, an gel filtration purification step using a G50 size exclusion spin column (Roche or GE) can be performed.

Concatamerization and Sanger sequencing: For the concatamerization reaction approximately 5 times more *PmeI*-digested PCR DNA is required compared to deep sequencing, which is obtained by re-amplifying the gel-purified and *PmeI*-digested PCR product.

34. Collect the DNA-pellet from step 30 and dissolve it in 40 µl water.
35. Use 2 µl for a 100 µl pilot PCR (500 nM primers, 2 mM $MgCl_2$) to determine the optimal cycle number with the exponential amplification phase. Use the following cycle conditions: 45 s at 94°C, 1 min 25 s at 50°C, 1 min at 72°C. Take aliquots after 6, 8, 10 and 12 cycles and analyze them on a 2.5 % agarose gel.
36. For the second PCR, prepare a reaction mixture of a final volume of 600 µl. Distribute the mixture over 6 PCR tubes, and use the number of PCR cycles as determined above.
37. Purify the DNA using a phenol/chloroform extraction, followed by a chloroform extraction, and an ethanol precipitation.

BanI digestion and concatamerization

38. Collect the PCR DNA pellet and dissolve it in 194 µl of a solution containing 20 µl of 10× NEB buffer 4 and 174 µl water. Set aside a 4 µl aliquot for later gel analysis. Add 10 µl of *BanI* restriction enzyme and incubate at 37°C for 3 h.
39. Add 12 µl 5 M NaCl and perform a phenol/chloroform- and chloroform extraction. Precipitate the DNA with ethanol.
40. Collect the DNA by centrifugation and dissolve the pellet in 96 µl of a cocktail consisting of 10 µl 10× T4 DNA ligase buffer (NEB), 3 µl of 100 µM PCR 5' primer and 3 µl of 100 µM PCR 3' primer, and 80 µl water. Incubate for 10 min at 65°C to denature the small fragments created by the *BanI* digestion and for strand exchange with the primers
41. Cool the sample to room temperature and add 4 µl of T4 DNA ligase and incubate for 5 h at 22°C. Alternatively the incubation can be carried out at 16°C overnight.
42. Analyze 4 µl of the reaction mixture before and after the *BanI* digestion on a 2 % agarose gel. Also load a 100 bp DNA ladder. The concatamerization product appears as a fuzzy ladder over a size range of about 60 bp (the monomer) to 1 or 2 kb.

43. Add 5 µl of 5× loading dye to 50 µl of the concatamers and load the sample into two wells of an 12-welll ethidium bromide containing 1.5 % NuSieve agarose gel. In an adjacent well, also run a 100 bp ladder for identification of the desired size range of the concatamers. Run the gel for approximately 1 h at 100 V until the marker bands are sufficiently resolved. If the concatamerized DNA is barely visible, perform a phenol/chloroform and chloroform extraction followed by an ethanol precipitation to concentrate the sample before loading it on a single well of the agarose gel.

44. Visualize the DNA in the gel using a 360 nm UV transilluminator. Excise the DNA band between 700–1200 bp. Extract DNA as indicated in step 30.

Cloning into TOPO-TA vector

45. Collect the DNA pellet in a tabletop centrifuge spinning for 15 minutes at maximum speed. Dissolve the pellet in 15 µl of a PCR mixture containing 1× dNTPs, 1× PCR buffer and 0.15 µl of *Taq* polymerase.

46. Incubate the reaction mixture for 30 min at 72°C to fill in the overhangs and to add a 3' untemplated adenosine required for T/A cloning.

47. Perform TOPO TA cloning as described by the manufacturer (Invitrogen, with pCR2.1-TOPO vector).

PCR screening of colonies for concatamer inserts

48. Fill the desired number of wells of a 96-well microtiter plate with 30 µl of a PCR solution, containing 3 µl 10× PCR buffer, 3 µl 10× dNTPs, 0.5 µl 100 µM primer M13 (−20) F, 0.5 µl 100 µM primer M13 R (for sequences see manufacturer's manual) and 0.3 µl *Taq* polymerase.

49. Transfer individual white colonies to the wells of the microtiter plate filled with PCR mixture. PCR amplify the insert with the following cycle conditions: 28 cycles: 2 min at 95°C, 1 min at 50°C, 1 min at 72°C. Final elongation 5 min at 72°C Analyze 6 µl of the PCR products on a 2 % standard agarose gel, using the 100 bp DNA ladder as a size marker. PCR products from empty vectors are about 200 bp long.

50. Purify the remaining PCR with the QIAquick PCR purification kit according to the manufacturer's instructions. Submit the purified PCR product for automated sequencing using the T7 primer as sequencing primer.

Synthesis of the 5' adenylated 3' adapter oligonucleotides

Synthesis of adenosine-5'-phosphoimidazolide (ImpA)

Reagents

- 5' AMP free acid (Sigma cat. # A2252)
- Triphenylphosphine (Aldrich cat. # T84409)
- 2,2'-Dipyridyldisulfide (aldrithiol-2, Aldrich cat. # 143049)
- Imidazole (molecular biology grade, Sigma cat. # I5513)
- Sodium perchlorate
- Anhydrous solvents: dimethylformamide, triethylamine, acetone, diethyl ether.

Perform all experiments under a fume hood

1. Prepare two dry 50 ml glass flasks. Flasks are dried overnight in a drying oven at 140° C under vacuum. While the vacuum pump is still running, the oven is allowed to cool to room temperature and the vacuum is then released by shutting off the pump and venting the drying oven with argon from an argon tank. The flasks are then immediately sealed with rubber septa. Instead of a drying oven, one can also flame dry the flask using a Bunsen burner and chasing the hot humid air out by a constant argon flow from an argon-filled balloon connected to the flask by a needle through a rubber septum; a second needle is plucked into the septum to release the humid air and incoming argon. Once the humid air has been replaced by dry argon, the second needle is removed while the argon filled balloon remains connected while the flask is allowed to cool to room temperature.
2. Suspend 174 mg (0.5 mmol) of 5' AMP free acid in 15 ml of anhydrous dimethylformamide in one of the dried round-bottom flasks. Keep the flask closed with a rubber septum. The AMP will not dissolve entirely.
3. In the other flask, dissolve 262 mg (1 mmol) of triphenylphosphine, 220 mg (1 mmol) of 2,2'-dipyridyldisulfide, and 170 mg (2.5 mmol) of imidazole in 15 ml of dimethylformamide and 0.9 ml (2.5 mmol) of triethylamine. Remove a 20 µl aliquot and label as sample A.
4. Add the AMP solution/suspension dropwise to the vigorously stirred triphenylphosphine-containing solution. Retain a 20 µl aliquot of the AMP solution from the flask for subsequent thin layer chromatography (TLC) analysis (labeled as sample B). Stir the reaction mixture for another 2.5 h at room temperature; keep the flask closed with the rubber septum. Remove a 20 µl aliquot and label as sample C before proceeding to the next step. The 5'AMP from the DMF solution/suspension will dissolve completely and should turn to a clear yellow-green color.
5. Precipitate the ImpA by adding the reaction mixture dropwise into a thin and tall 500-ml beaker containing a vigorously stirred solution of 1.1 g (9 mmol) of sodium perchlorate, 110 ml acetone and 55 ml anhydrous diethyl ether. Precipitation will begin immediately and the solution will turn progressively cloudier as the entire reaction mixture is added.
6. Turn off the stirrer to allow the precipitate to settle to the bottom of the beaker. After approximately 1 h, decant as much as possible of the supernatant without perturbing the precipitate or use a large glass pipette connected to a pipetting aid to aspirate off the clear supernatant.
7. Once the volume has been reduced to about 20 ml, resuspend the precipitate in the residual supernatant and transfer to two 30 ml Corex glass centrifugation tubes. Rinse the beaker with small volumes (5 ml) of acetone and combine the wash solutions with the suspension already transferred to Corex tubes. Collect the precipitate by centrifugation at 5000 rpm (3,000×g) for 10 min.
8. Pour off the supernatant and wash the pellet twice by resuspending it with 20 ml acetone in the Corex tubes followed by 5 min centrifugation at 5000 rpm. The pellet should be white and the supernatant should be clear.
9. Resuspend the pellet in 10 ml diethylether and collect it again at 5000 rpm for 20 min. Pour off the ether supernatant. Seal the centrifugation tubes with a septum or parafilm and place 2 needles or poke small holes, so that the pellet in the tubes can be dried overnight under reduced pressure at 40°C. After releasing the pressure by flushing

the oven with argon, remove the needles so that the dry powder is protected from exposure to the air.

10. The dried powder may be stored in a sealed bottle for up to one week at −80°C protected from humidity. The yield of ImpA is approximately 100 mg. The molecular weight of ImpA is 396.3 g/mol.

Quality control of synthesized ImpA

11. For quality control, spot 1 μl unreacted mixture (sample A), AMP (sample B), and the mixture after 2.5 h reaction (sample C) on fluorescence-indicator coated Silica gel 60 pre-coated (TLC) plates (EMD, 5719-2).

12. Dry the plate with a hair dryer and develop the TLC in a chromatography chamber using isopropanol/water/25% ammonia (7/2/1) as solvent system. Develop the TLC until the solvent front is about 1 cm away from the top of the plate. Mark the solvent front with a pencil. Dry the plate with the hair dryer

13. Visualize the spots at 254 nm with a UV hand lamp. The starting material should be completely consumed. With these conditions the retention factors are 0.14 for AMP and 0.46 for ImpA.

Adenylation of 3' adapter oligodeoxynucleotides

Reagents

- Freshly prepared adenosine-5'-phosphoimidazolide (ImpA)
- 1 M $MgCl_2$
- 5' phosphorylated 3' aminolinker-modified 3' adapter oligodeoxynucleotide, commercially available from oligonucleotide custom synthesis companies.

1. The protocol is given for a conversion of 50 nmol adapter. Dry down the equivalent of 50 nmol of 5' phosphorylated 3' adapter oligonucleotide solution in a 1.5 ml Eppendorf tubes in a Speedvac. Prepare 150 μl of a solution containing 100 mM ImpA (MW 396.3 g/mol) and 50 mM $MgCl_2$. Add 100 μl of the ImpA/$MgCl_2$ solution to the dried down oligonucleotide. Incubate the solution at 50°C for 1.5 h. Add the remaining 50 μl of the ImpA/$MgCl_2$ solution and incubate for another 1.5 h. Reduce the reaction volume to about half the final volume in a Speedvac, and then add approximately one volume (75 μl) urea gel-loading solution (8 M urea, 50 mM EDTA, pH 8, 0.5 mg/ml bromophenol blue). Proceed to the next step or store the solution at −20°C.

2. We recommend the use of sequencing size, preparative gels (40 cm×25 cm×1.5 mm, gel volume about 200 ml), to properly resolve the adenylated product from the non-adenylated adapter starting material. Load the material in one well of an 8-well 20% polyacrylamide gel and run the gel for 6 h at 50 W until the bromophenol blue dye is about to exit the bottom of the gel.

3. Dismantle the gel and wrap it in plastic foil (Saran wrap). Place it onto a 254 nm fluorescence-indicator coated silica gel plate and visualize the oligodeoxynucleotides by shadowing with a 254 nm UV lamp. Mark the UV-absorbing dark bands, excise the product band and elute the adenylated oligodeoxynucleotide from the gel slices overnight at 4°C in at least 3 volumes of 0.4 M NaCl. The intensity of the band corresponding to the adenylated product is similar to the intensity of the non-adenylated starting material.

4. Precipitate the adenylated oligodeoxynucleotide by addition of 3 volumes of absolute ethanol and by incubation for at least 1 h at 0°C or overnight at −20°C.
5. Collect the pellet by centrifugation for 15 min at 4°C (14,000×g). Dissolve the pellet in 20 μl of water and determine the concentration by UV absorbance. The overall yield of the adenylation reaction is approximately 20%.

Concluding Remarks

Cloning and sequencing of small RNA libraries remains important until a comprehensive cell- and tissue-specific clone-based small RNA gene expression database is available for each species. The protocol is also essential for defining the small RNA binding partners of the individual members of the Ago/Piwi protein family using immunopurification protocols with specific antibodies. Furthermore, small RNA cloning is useful for identifying the binding sites of mRNA binding proteins (44) by isolating small RNAs from immunopurified and partially nuclease digested mRNPs. The methods can also be useful to provide an overview of general gene expression if total RNA or polyA+ mRNA are partially nuclease digested, converted into clone libraries and deep sequenced.

References

1. Seto AG, Kingston RE, Lau NC. Mol Cell 2007;26:603–609. [PubMed: 17560367]
2. Nakayashiki H. FEBS Lett 2005;579:5950–5957. [PubMed: 16137680]
3. Vaucheret H. Genes Dev 2006;20:759–771. [PubMed: 16600909]
4. Grewal SI, Elgin SC. Nature 2007;447:399–406. [PubMed: 17522672]
5. Zaratiegui M, Irvine DV, Martienssen RA. Cell 2007;128:763–776. [PubMed: 17320512]
6. Meister G, Tuschl T. Nature 2004;431:343–349. [PubMed: 15372041]
7. Bartel DP. Cell 2004;116:281–297. [PubMed: 14744438]
8. Lee RC, Feinbaum RL, Ambros V. Cell 1993;75:843–854. [PubMed: 8252621]
9. Reinhart BJ, Slack FJ, Basson M, et al. Nature 2000;403:901–906. [PubMed: 10706289]
10. Fire A, Xu S, Montgomery MK, et al. Nature 1998;391:806–811. [PubMed: 9486653]
11. Hamilton AJ, Baulcombe DC. Science 1999;286:950–952. [PubMed: 10542148]
12. Hammond SM, Bernstein E, Beach D, Hannon GJ. Nature 2000;404:293–296. [PubMed: 10749213]
13. Zamore PD, Tuschl T, Sharp PA, Bartel DP. Cell 2000;101:25–33. [PubMed: 10778853]
14. Elbashir SM, Lendeckel W, Tuschl T. Genes Dev 2001;15:188–200. [PubMed: 11157775]
15. Lee RC, Ambros V. Science 2001;294:862–864. [PubMed: 11679672]
16. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Science 2001;294:853–858. [PubMed: 11679670]
17. Lau NC, Lim LP, Weinstein EG, Bartel DP. Science 2001;294:858–862. [PubMed: 11679671]
18. Aravin AA, Lagos-Quintana M, Yalcin A, et al. Dev Cell 2003;5:337–350. [PubMed: 12919683]
19. Aravin A, Gaidatzis D, Pfeffer Sb, et al. Nature 2006;442:203–207. [PubMed: 16751777]
20. Lau NC, Seto AG, Kim J, et al. Science 2006;313:363–367. [PubMed: 16778019]
21. Saito K, Sakaguchi Y, Suzuki T, et al. Genes Dev 2007;21:1603–1608. [PubMed: 17606638]
22. Hartig JV, Tomari Y, Forstemann K. Genes Dev 2007;21:1707–1713. [PubMed: 17639076]
23. Carmell MA, Girard A, van de Kant HJ, et al. Dev Cell 2007;12:503–514. [PubMed: 17395546]
24. Ruby JG, Jan C, Player C, et al. Cell 2006;127:1193–1207. [PubMed: 17174894]
25. Parker R, Sheth U. Mol Cell 2007;25:635–646. [PubMed: 17349952]
26. Pillai RS, Bhattacharyya SN, Filipowicz W. Trends Cell Biol 2007;17:118–126. [PubMed: 17197185]
27. O'Donnell KA, Boeke JD. Cell 2007;129:37–44. [PubMed: 17418784]
28. Vastenhouw NL, Plasterk RH. Trends Genet 2004;20:314–319. [PubMed: 15219396]
29. Henderson IR, Jacobsen SE. Nature 2007;447:418–424. [PubMed: 17522675]

30. Pfeffer S, Sewer A, Lagos-Quintana M, et al. Nat Methods 2005;2:269–276. [PubMed: 15782219]
31. Landgraf P, Rusu M, Sheridan R, et al. Cell 2007;129:1401–1414. [PubMed: 17604727]
32. Pfeffer S, Lagos-Quintana M, Tuschl T. Current Protocols in Molecular Biology 2005:26.4.1–26.4.18.
33. Chomczynski P, Sacchi N. Anal Biochem 1987;162:156–159. [PubMed: 2440339]
34. Mourelatos Z, Dostie J, Paushkin S, et al. Genes Dev 2002;16:720–728. [PubMed: 11914277]
35. Maniataki E, De Planell Saguer MD, Mourelatos Z. Methods Mol Biol 2005;309:283–294. [PubMed: 15990408]
36. Beitzinger M, Peters L, Zhu JY, et al. RNA Biol 2007:4.
37. Brennecke J, Aravin AA, Stark A, et al. Cell 2007;128:1089–1103. [PubMed: 17346786]
38. Nelson PT, De Planell-Saguer M, Lamprinaki S, et al. Rna 2007:13.
39. Margulies M, Egholm M, Altman WE, et al. Nature 2005;437:376–380. [PubMed: 16056220]
40. Bentley DR. Current Opinion in Genetics & Development 2006;16:545–552. [PubMed: 17055251]
41. Bennett S. Pharmacogenomics 2004;5:433–438. [PubMed: 15165179]
42. Mukaiyama T, Hashimoto M. Bulletin of the Chemical Society of Japan 1971;44:2284.
43. Lohrmann R, Orgel LE. Tetrahedron 1978;34:853–855.
44. Ule J, Jensen K, Mele A, Darnell RB. Methods 2005;37:376–386. [PubMed: 16314267]

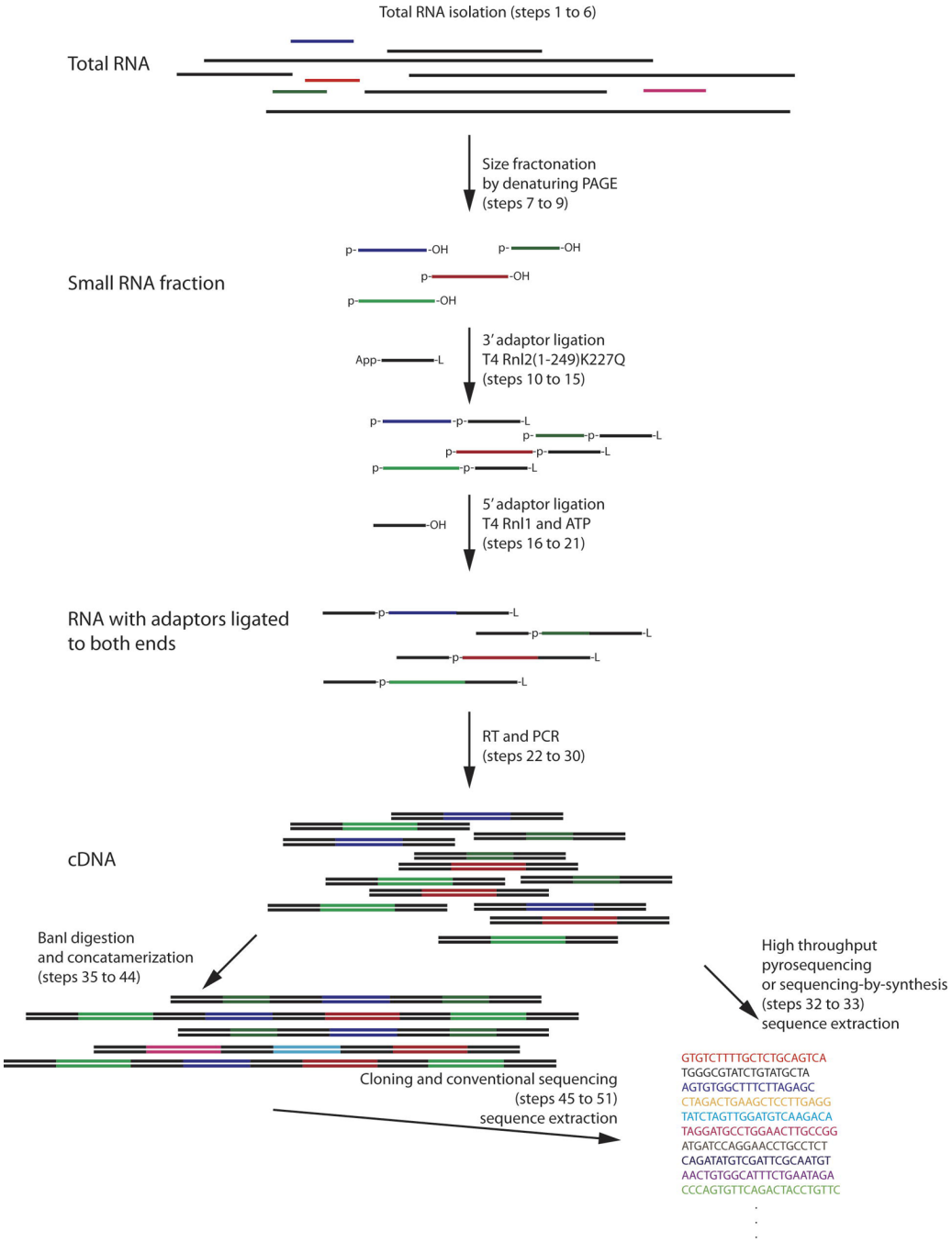

Figure 1. Schematic representation of small RNA cDNA library preparation

The step numbers refer to the step-by-step protocol. Abbreviations: p, 5' phosphate; OH, 3'-hydroxyl; L, 3' aminolinker.