Published in final edited form as:

Methods Mol Biol. 2015 ; 1276: 211–228. doi:10.1007/978-1-4939-2392-2_12.

Preparation of cDNA libraries for high-throughput RNA sequencing analysis of RNA 5' ends

Irina O. Vvedenskaya, Seth R. Goldman, and Bryce E. Nickels

Department of Genetics and Waksman Institute, Rutgers University, Piscataway, NJ 08854, USA

Summary

We provide a detailed protocol for preparing cDNA libraries suitable for high throughput sequencing that are derived specifically from the 5' ends of RNA (5' specific RNA-seq). The protocol describes how cDNA libraries for 5' specific RNA-seq can be tailored to analyze specific classes of RNAs based upon the phosphorylation status of the 5' end. Thus, the analysis of cDNA libraries generated by these methods provides information regarding both the sequence and phosphorylation status of the 5' ends of RNAs. 5' specific RNA-seq can be used to analyze transcription initiation and post-transcriptional processing of RNAs with single base pair resolution on a genome-wide level.

Keywords

Transcription start site; RNA-seq; RNA degradation; cDNA library; Transcriptome; Post-transcriptional processing

1. Introduction

Conventional high throughput RNA sequencing (RNA-seq) uses massively parallel sequencing technology to analyze cDNA derived from total cellular RNA and is usually intended to survey the entire transcriptome. Often, however, it is preferable to derive sequence information from only a particular fraction of the transcriptome. In such cases, methods to generate cDNA libraries derived only from the RNA fraction of interest are required. Below we describe a protocol that facilitates the analysis of only the portion of the transcriptome comprising the 5' ends of RNAs. In particular, we describe a protocol for specifically generating cDNA libraries from RNA 5' ends. In addition, selective enzymatic treatments of RNA prior to cDNA library construction permits construction of cDNA libraries derived only from transcripts that carry a particular modification at the 5' end of the transcript. Our lab has developed these protocols for use with bacterial RNAs, which lack a 5' 7-methylguanosine mRNA cap. However, the methods described here are easily adaptable to the analysis of RNA derived from any organism.

Analysis of only the portion of the transcriptome comprising the 5' ends of RNAs provides useful insight into a number of biological processes. Perhaps the major utility of 5' specific

RNA-seq is the identification of positions in the genome that correspond to transcription start sites. In bacteria, the identification of transcription start sites is accomplished by analyzing the 5' ends of RNA transcripts that carry a 5' triphosphate moiety (5'-PPP) [see, for example (1–6)]. In contrast, the analysis of transcripts carrying a 5' monophosphate moiety (5'-P) or a 5' hydroxyl moiety (5'-OH) can provide insight into the post-transcriptional processing of RNA. We note that the analysis of 5' PPP ends is sometimes referred to as differential RNA-seq (dRNA-seq) (2).

Use of 5' specific RNA-seq can, in principle, facilitate the parallel analysis of the 5' ends associated with all of the RNAs in the transcriptome. Thus, cellular processes that alter the composition or distribution of RNA 5' ends can be examined at a genome-wide level. In addition, the single base pair resolution of 5' specific RNA-seq permits precise mapping of the 5' ends of individual RNAs. This single base pair resolution circumvents the inherent ambiguity associated with other methods for analyzing 5' ends of RNAs. Primer extension and S1 nuclease protection assays, for example, are often limited in resolution and cannot precisely define the position of the RNA 5' end. Furthermore, by providing sequence and positional information, an RNA-seq based approach enables a determination of whether RNA 5' ends carry bases that differ from the DNA template, i.e. “RNA-DNA differences”. This feature of 5' specific RNA-seq is especially useful for the analysis of transcripts initiating from promoters that are prone to “slippage” (i.e. reiterative transcription) during transcription initiation (7). Finally, the sensitivity of 5' specific RNA-seq enables one to interrogate the 5' ends of RNAs that would be otherwise inaccessible by primer extension analysis due to low abundance and/or the prevalence of extensive secondary structure.

2. Materials

2.1 Reagents

All chemicals used are of ACS grade or higher.

1. Nuclease-free water (non-DEPC treated).
2. T4 Polynucleotide Kinase
3. Alkaline Phosphatase
4. Optikinase
5. Terminator Exonuclease
6. RNA 5' Polyphosphatase
7. T4 RNA ligase 1, ssRNA Ligase
8. RNase H
9. Phusion HF DNA Polymerase
10. DNase I
11. Ribonuclease Inhibitor; RNase OUT
12. SuperScript III Reverse Transcriptase

13. 10 mM ATP
14. dNTP solution mix, 10 mM of each NTP
15. Acid phenol:chloroform (CHCl_3) pH 4.5
16. 3M Sodium Acetate, NaOAc, pH 5.5
17. 0.5M EDTA pH 8
18. Glycogen Ultrapure, 20 $\mu\text{g}/\mu\text{L}$
19. GlycoBlue, 15 mg/mL
20. 100% Ethanol
21. Isopropanol
22. Formamide, deionized
23. Ammonium Acetate, NH_4OAc
24. Sodium Chloride
25. Tris Base
26. Bromophenol Blue
27. Xylene Cyanol
28. Qubit dsDNA HS Assay kit
29. Qubit RNA Assay kit
30. Low Range ssRNA Ladder
31. Ultra Low Range DNA Ladder
32. 6X Orange DNA Loading Dye; contains xylene cyanol and orange G
33. Ethidium Bromide, 10 mg/mL
34. SYBR Gold nucleic acid gel stain
35. 1X TBE Electrophoresis Buffer (90 mM Tris-borate, 1 mM EDTA)
36. 1X TE Buffer (10mM Tris-HCl pH 8, 1 mM EDTA)
37. 1X Elution Buffer A: 0.3M NaCl in 1X TE buffer
38. 1X Elution Buffer B: 2.5M Ammonium Acetate in 1X TE buffer
39. 7.5M NH_4OAc in 1X TE buffer
40. 2X RNA Loading Dye (95% deionized formamide, 18 mM EDTA, 0.25% SDS, xylene cyanol, bromophenol blue)
41. Adaptors and primers; sequences are provided in Section 3

2.2 Supplies

1. Qubit assay tubes

2. Low-binding nuclease free microcentrifuge tubes, 1.7 mL
3. Low-binding nuclease free thin-wall PCR tubes, 0.2 mL
4. Low-binding P1000 Barrier Tips, sterile
5. Low-binding P200 Barrier Tips, sterile
6. Low-binding P10 Barrier Tips, sterile
7. Wide orifice P200 Barrier Tips, sterile
8. Sterile 50 mL Vacuum-driven Filtration System, 0.2 µm PES membrane
9. Stainless Steel Disposable Scalpels, No. 10 blade
10. Sterile needle 23 gauge
11. DNA LoBind Tube, 0.5 mL
12. Spin-X Centrifuge Tube Filter, 0.45 µm, RNase/DNase Free
13. Container for gel staining (we use a P10 tip box lid)
14. 6% TBE-Urea gels, 1mm × 10 wells
15. 10% TBE-Urea gels, 1mm × 10 wells
16. 10% TBE gels, 1mm × 10 wells
17. RNeasy Mini Kit (Qiagen)
18. RNeasy MinElute Clean Up Kit (Qiagen)

2.3 Equipment

1. UV transilluminator (preferably coupled to a Gel Documentation system such as GelDoc XR, Bio-Rad).
2. Instrument for high-precision nucleic acid quantification, e.g. Qubit Fluorometer (Life Technologies) or NanoDrop Spectrophotometer (Thermo Scientific).

3. Methods

3.1 Preparation of RNA for library construction

Prior to generating cDNA libraries it may be desirable to remove highly abundant tRNAs and rRNAs that might otherwise constitute the majority of the final library to the detriment of the detection of low abundance RNAs. We routinely remove tRNAs and other small RNAs <200 nt using an RNeasy Mini Kit (Qiagen) and remove rRNAs using magnetic bead based depletion kits such as MICROBExpress (Life Technologies) or Ribo-Zero (Epicentre). Use of these kits can be very costly and, in the case of RNA isolated from certain organisms, ineffective. In these cases, specialized approaches to remove rRNA may be required [*see*, for example, (8, 9)].

Removal of contaminating DNA by treatment with DNase I is essential prior to beginning the procedures described below. If RNeasy columns are used, DNase I treatment may be conveniently performed on-column.

All manipulations with RNAs should be performed on ice (or at 4°C in the case of centrifugation) unless otherwise specified and nuclease-free water should be used for all solutions. Due to the sensitivity of RNA-seq, even trace contaminants can compromise the quality of the final library. For that reason, we advise that anyone building libraries makes every effort to minimize distractions while working. Clean gloves should be worn at all times and changed frequently throughout. Pipet tips should not be re-used even when taking multiple aliquots from the same tube.

The following subsections describe enzymatic treatments to prepare cDNA libraries derived from 5' triphosphate ends (5' PPP ends) [section 3.1.1], 5' hydroxyl ends (5' OH ends) [section 3.1.2], or all 5' ends regardless of the phosphorylation status (all 5' ends) [section 3.1.3]. The first step in the construction of each cDNA library is the ligation of an adaptor to the 5' ends of input RNA (see below). To be ligated, the 5' end of an RNA must carry a 5' monophosphate (5' P). Thus, to derive cDNA libraries from only 5' P ends, proceed directly to section 3.2 below. In contrast, to prepare cDNA libraries derived from 5' PPP ends, 5' OH ends, or all 5' ends, the selective enzymatic treatments described in this section must first be performed (Fig. 1).

3.1.1 Enzymatic treatments to selectively derive cDNAs from 5' PPP ends—

Preparation of this library is accomplished through selective degradation of 5' P RNA using Terminator 5'-Phosphate-Dependent Exonuclease followed by conversion of 5' PPP RNA to ligation competent 5' P RNA via RNA 5'-Polyphosphatase which removes the 5'- and 5'-phosphate groups of tri-phosphorylated RNAs leaving 5' monophosphate groups. 5' OH RNA in the sample is unligatable due to lack of 5' P. Note that RNA 5'-Polyphosphatase does not remove 5' 7-methylguanosine caps and therefore should be replaced with Tobacco Acid Pyrophosphatase for the analysis of 5' 7-methylguanosine RNAs isolated from eukaryotes.

3.1.1.1. Terminator Exonuclease treatment

1. In a total reaction volume of 20 µl, combine 1–2 µg of RNAs (*see* Note 1), 2 µl of Terminator Exonuclease 10X Reaction Buffer A, 1 µl (40U) of RNase OUT and 1 µl (1U) of 5'-Terminator Exonuclease.
2. Incubate for 1 hr at 30°C.
3. Add 180 µl nuclease-free water and 200 µl of acid-phenol-chloroform. Vortex.
4. Centrifuge 10 min at 21,000 × g at 4°C.
5. Carefully transfer 180 µl of upper phase into a new tube (*see* Note 2).

¹Because enzymatic treatments and precipitations can result in a significant reduction in the amount of material available to begin the cDNA library construction, we recommend beginning the procedures to prepare libraries derived from 5' PPP ends, 5' OH ends, or all 5' ends with at least 1 µg of starting RNA.

6. Add 20 μ l of 3M NaOAc and 2 μ l of 10 mg/mL glycogen. Mix well. Add 600 μ l of ice-cold 100% EtOH. Vortex.
7. Precipitate RNAs at -80°C overnight or at least 4 h. Following precipitation, centrifuge samples 30 min at $21,000 \times g$ at 4°C .
8. Carefully remove the supernatant with a pipet and wash the pellet two times each by adding 1 mL of 75% ice-cold EtOH, vortexing to disengage the pellet from the side of the tube, and centrifuging for 5 min at $21,000 \times g$ at 4°C .
9. Air-dry the pellet 5 min at room temperature.
10. Resuspend RNAs in 25.5 μ l nuclease-free water; incubate for at least 10 min on ice to enable full resuspension (pipet sample to resuspend, do not vortex) (*see* Note 3).

Optional step: use a 0.5 μ l aliquot to analyze by gel electrophoresis.

3.1.1.2. RNA 5'-Polyphosphatase treatment

1. Add 3 μ l of RNA 5'-Polyphosphatase 10X Reaction Buffer, 1 μ l (40U) of RNase OUT, and 1 μ l (20U) of RNA 5'-Polyphosphatase to the resuspended Terminator-treated RNAs.
2. Gently mix the reaction by pipetting.
3. Incubate 1h at 37°C .
4. Bring the total volume of the reaction mixture to 200 μ l with nuclease-free water and extract with acid-phenol-chloroform and precipitate RNA as described above.
5. Resuspend RNAs in 10.5 μ l nuclease-free water.

Optional step: use a 0.5 μ l aliquot to analyze by gel electrophoresis.

3.1.2 Enzymatic treatments to selectively derive cDNAs from 5' OH ends—

Preparation of this library is accomplished through selective removal of 5' P ends by use of Terminator 5'-Phosphate-Dependent Exonuclease followed by treatment with Optikinase to add a 5' phosphate group to RNAs that carry a 5' hydroxyl thus making them competent for ligation.

3.1.2.1. Terminator Exonuclease treatment

Treat 1–2 μ g as described in section 3.1.1.1 with the exception that the treated

RNAs should be resuspended in 34.5 μ l nuclease-free water after precipitation.

Optional step: use a 0.5 μ l aliquot to analyze by gel electrophoresis.

²Following centrifugation, the mixture separates into the lower phenol-chloroform phase, inter-phase, and the colorless upper aqueous phase. RNA remains exclusively in the aqueous phase whereas proteins and DNA are in the inter-phase and organic phase. Be careful not to disturb the inter-phase.

³An RNA pellet after precipitation and centrifugation may appear tiny and be difficult to visualize. We recommend use of a colored glycogen, GlycoBlue, to aid in the visualization of the pellet.

3.1.2.2. Optikinase treatment

1. Add 5 µl of Optikinase 10X Reaction Buffer, 5 µl of 10 mM ATP, 1 µl (40U) of RNase OUT, 5 µl (50U) of Optikinase to the Terminator-treated RNAs (total reaction volume 50 µl).
2. Gently mix the reaction and incubate 60 min at 37°C.
3. Bring the total volume of the reaction mixture to 200 µl with nuclease-free water and extract with acid-phenol-chloroform and precipitate RNA as described in section 3.1.1.1, steps 3–9.
4. Resuspend RNAs in 10.5 µl nuclease-free water.

Optional step: use a 0.5 µl aliquot to analyze by gel electrophoresis.

3.1.3 Enzymatic treatments to derive cDNAs from all 5' ends—Preparation of this library is accomplished through conversion of RNAs carrying 5' OH ends to 5' P ends by treatment with T4 Polynucleotide Kinase in the presence of ATP followed by treatment with RNA 5'-Polyphosphatase to remove the 5'γ- and 5'β-phosphate groups of tri-phosphorylated RNAs leaving ligation competent 5' monophosphate groups (*see* Note 4 for an alternative procedure to build this library).

3.1.3.1. T4 Polynucleotide Kinase treatment

1. In a total reaction volume of 50 µl, combine 1–2 µg of RNAs, 5 µl of T4 Polynucleotide Kinase 10X Reaction Buffer, 5 µl of 10 mM ATP and 1 µl (10U) of T4 Polynucleotide Kinase.
2. Mix well by pipetting and incubate for 1h at 37°C.
3. Use RNeasy Mini Kit or RNeasy MinElute Clean Up Kit (Qiagen) to remove excess ATP from the reaction.
4. Elute RNAs from the filter column with 25.5 µl nuclease-free water.

Optional step: use a 0.5 µl aliquot to analyze by gel electrophoresis.

3.1.3.2. RNA 5'-Polyphosphatase treatment

Perform as described in section 3.1.1.2.

3.2 cDNA library construction

For construction of cDNA libraries derived from 5' PPP ends, 5' OH ends, or all 5' ends, use the material generated as described in section 3.1 as input. For construction of cDNA libraries derived from 5' P ends use 500 ng of RNA that has not undergone any of the enzymatic treatments in 3.1; *see* Fig. 2.

⁴Alternatively, the preparation of this library can be accomplished through the treatment of RNAs with Alkaline Phosphatase to convert 5' PPP and 5'P ends to 5' OH ends followed by treatment with Optikinase to convert 5' OH ends to ligatable 5' P ends (*see* 5, 6).

3.2.1. Preparation of the 5' adaptor –ligated RNAs

1. In a total volume of 30 μ l combine 10 μ l of 5'-P RNAs (prepared as in 3.1; *see* Note 5), 7.5 μ l of 5'-Adaptor (10 pmol/ μ l), 3 μ l of T4 RNA ligase 1 10X Reaction Buffer, 1 μ l (40U) of RNase OUT and 1 μ l (20U) T4 RNA ligase; gently mix the reaction.
2. Incubate at 16°C overnight.
3. Add 30 μ l of 2X RNA loading dye; incubate at 95°C for 5 min, then place on ice for 2 min.
4. Load 20 μ l aliquots in 3 consecutive wells of a 6% TBE-Urea gel (1mm $\times$ 10 wells) that has been pre-run for at least 10 min at 200V in 1X TBE buffer. Load 5 μ l of Low Range ssRNA Ladder diluted 1:5 with loading dye in an adjacent lane.
5. Allow the sample to enter the gel at 85V, then increase voltage to 200V and run until the bromophenol blue reaches the center of the gel (*see* Note 6).
6. Remove gel from the cassette and incubate in EtBr staining solution (1:10,000 dilution of 10 mg/mL EtBr stock solution in 1X TBE) for 10 min.
7. Remove the well dividers and place gel on a UV transilluminator to visualize the 5' adaptor. Use a sterile scalpel to excise the region starting just above the 5' adaptor and extending to the top of the gel (Fig. 3, *see* Note 7).
8. Cut the large gel piece into ~8 smaller pieces of equivalent size.
9. Punch 3 holes in the bottom of 4 0.5 mL centrifuge tubes using a 23-gauge needle.
10. Place 2 gel slices into each tube, close the lids, and place the tubes into a 1.7 mL centrifuge tube (the hinge of the 0.5 mL tube will rest on the lip of the 1.7 mL tube). Centrifuge for at least 1.5 min at 21,000 $\times$ g at room temperature until all of the gel has been extruded into the bottom of the 1.7 mL tubes through the needle holes in the 0.5 mL tube.
11. Add 0.6 mL 1X Elution Buffer A to each tube (*see* Note 8). Elute RNAs from gel at 4°C with slow agitation for 16 hours.
12. Use a wide orifice P200 barrier tip to transfer ~300 μ l of the gel slurry to the top of a Spin-X cellulose acetate filter column and centrifuge 1.5 min at 15,000 $\times$ g (*see* Note 9).

⁵To analyze 5' P ends use 500 ng of RNA that has not undergone any of the enzymatic treatments in 3.1 in 10 μ l of water.

⁶It is important to not run the gel too far. Running the gel too far will increase the size of the gel that needs to be excised and lead to more difficulty in isolating the desired material from the larger gel slice. Two separate libraries may be loaded on the same gel to perform gel size selection steps provided that at least two empty wells are between them. Furthermore, after electrophoresis the gel should be cut in half prior to incubation in the EtBr staining solution and each half of the gel should be stained in separate containers to avoid library cross-contamination.

⁷Use new sterile scalpel for every new sample. Do not reuse a scalpel unless it is used to cut the same gel. EtBr and UV light degrade and damage an RNA library. Minimize the time of staining procedure to 10 min and UV exposure to 0.5 min. If possible, use low intensity UV light. Use Barrier Tip P10 box lids as containers to stain a gel. If reusing lids for gel staining, prior to procedure rinse 3–5 times with deionized water, spray with 100% EtOH and dry with paper towel to remove any traces of alcohol. Label containers while processing multiple libraries.

⁸Buffers used to elute nucleic acids from acrylamide gels should be purified by filtration using sterile 50 mL vacuum-driven 0.2 μ m filtration system.

13. Pool the eluates (there will be ~2.5 mL total) and aliquot 330 μ L into 7×1.7 mL microfuge tubes. Add 33 μ L of 3M NaOAc and 3.3 μ L of 10 mg/mL glycogen, mix well. Add 1100 μ L of ice-cold 100% EtOH. Vortex.
14. Precipitate at -80°C 12–16 hours. Following the precipitation, centrifuge samples 30 min at $21,000 \times g$ at 4°C .
15. Wash with 1 mL of 75 % ice-cold EtOH.
16. Air dry pellets 5 min at room temperature.
17. To resuspend RNA, add 3 μ L of nuclease-free water directly to each pellet, rehydrate 5 min at room temperature. Pool the resuspended material (~21 μ L total) into a single tube (*see* Note 10).

3.2.2. First strand cDNA synthesis

1. Mix 10 μ L of 5'-Adaptor-ligated material (*see* Note 11) with 0.3 μ L of RT primer (100 pmol/ μ L). As with the 5' adaptor, the sequence of the RT primer will vary depending upon the sequencing platform that will be used (*see* below for sequence information). We have constructed libraries compatible for sequencing with either the Illumina HiSeq platform or the SOLiD (Sequencing by Oligonucleotide Ligation and Detection) genome analyzer.
2. Heat the RT primer/5'-Adapter ligated RNA mixture 5 min at 65°C to denature the RNA and place on ice for at least 1 min to allow the RT primer to anneal.
3. To this mixture, add 9.7 μ L of enzyme cocktail [4 μ L of 5X First-Strand Buffer, 1 μ L of 10 mM dNTP mix, 1 μ L of 100 mM DTT, 1 μ L (40U) RNase OUT, 1 μ L (200U) of SuperScript III Reverse Transcriptase and 1.7 μ L of nuclease-free H_2O]. The total reaction volume will be 20 μ L. Gently mix.
4. Incubate in a thermal cycler with a heated lid for 5 min at 25°C followed by 60 min at 55°C , followed by 15 min at 70°C (to inactivate the enzyme).
5. Bring the temperature of the reaction to 37°C and then add 1 μ L (10U) of RNase H, and incubate an additional 20 min at 37°C (*see* Note 12).

3.2.3. Isolation of cDNAs

1. Add 20 μ L of 2X RNA loading dye to the sample and incubate at 95°C for 5 min then cool the sample on ice for 2 min.

⁹We routinely load three ~300 μ L aliquots of gel slurry onto the same column sequentially. Before reloading the column first remove the flow-through to a new tube and remove the gel pieces from the top of the column. This can be accomplished by dislodging the gel pieces with a clean P200 pipet tip but care must be taken not to damage the membrane.

¹⁰If desired, one can estimate the recovery at this step. We would recommend use of either a NanoDrop or Qubit. The typical recovery of the 5' Adaptor-ligated libraries is 25 to 50% of the amount taken for library construction.

¹¹Only half of the material generated so far is used for library construction, the remainder is held back as a precaution in case something goes wrong in the remainder of the procedure.

¹²The sample can be stored at -20°C prior to the size selection step.

2. Load the sample into 2 consecutive wells (load 20 μ l in each well) of a 10% TBE-Urea gel (1 mm $\times$ 10 wells) that has been pre-run for at least 10 min at 200V in 1X TBE buffer. Load 5 μ l of ssRNA Ladder next to the sample.
3. Allow the sample to enter the gel at 85V, then increase voltage to 200V and run until the bromophenol blue reaches the center of the gel (*see* Note 6). Remove gel from the cassette and incubate in EtBr staining solution for 10 min.
4. Use a UV transilluminator to visualize the ladder and a sterile scalpel to excise the portion of the gel that is between ~110–550 nt as estimated by comparison with the ladder (*see* Note 7). *See* Note 13 for an alternative procedure to size select these products.
5. Use a scalpel to cut the large gel piece into 4 smaller pieces of equivalent size. Use the extrusion procedure described in section 3.2.1 (steps 9–10) to crush these gel slices into 2 $\times$ 1.7 mL microcentrifuge tubes.
6. Add 400 μ l 1X Elution Buffer A to each tube. Elute DNA for 2 h at room temperature, then 1 h at 37°C with gentle agitation (or at 4°C for 16 h).
7. Use wide orifice P200 barrier tips to transfer ~300 μ l of the gel slurry to the top of a Spin-X cellulose acetate filter column (*see* Note 9).
8. Centrifuge 1.5 min at 15,000 $\times$ g. Divide the eluates into two new 1.7 mL tubes (there will be ~400 μ l in each tube) and add 400 μ l of 7.5M NH_4OAc in 1X TE, 8 μ l of glycogen (10 mg/mL) and 560 μ l of isopropanol (*see* Note 14). Vortex.
9. Precipitate at –20°C for at least 2 hours (or overnight).
10. Centrifuge for 30 min at 21,000 $\times$ g at 4°C. Wash pellets with 1 mL of 75% room temperature EtOH. Repeat wash step 2 more times.
11. Air dry pellets 5 min at room temperature.
12. To resuspend DNA, add 5 μ l of nuclease-free water directly to each pellet, rehydrate for at least 5 min at room temperature. Pool the resuspended material (~10 μ l total) into a single tube.

3.2.4. PCR amplification—This step generates the second strand of cDNA, introduces a barcode or index and amplifies the library by PCR. It is essential to use only the minimal number of amplification cycles needed to obtain sufficient material for sequencing to avoid over-amplification of the libraries, a major source of bias in the results. For this reason, we first perform a test PCR using 12 cycles and a small portion of the sample then, if necessary, adjust the number of cycles used for the final PCR with the rest of the material (Fig. 3).

¹³Size selections can also be performed using automated fractionation devices such as the Pippin Prep (Sage Science) or LabChip XT (Perkin Elmer). We have successfully used the Pippin Prep to size select cDNAs after the first strand synthesis as well as the final cDNA library after PCR amplification.

¹⁴High salt and low temperature are used here to maximize DNA recovery from the very dilute solutions used in this protocol. Typically, isopropanol precipitations are performed at room temperature to minimize co-precipitation of salt. To ensure maximal recovery, we use a –20°C incubation for our isopropanol precipitations and rely on washes with room temperature 75% ethanol to remove excess salts and to replace the less volatile isopropanol with ethanol to facilitate drying of the pellet.

3.2.4.1. Test PCR: Use 10% of the volume of the cDNA recovered after gel extraction to perform a test PCR to optimize the number of amplification cycles.

1. Mix 1 µl of cDNA with 2 µl of 5X Phusion HF PCR Buffer, 1 µl of each PCR Primer (0.25 µM final; use a 2.5 µM stock) (*see* Note 15), 0.2 µl dNTP mix (0.2 mM final), 0.1 µl (0.2 U) of Phusion HF DNA polymerase. Add RNase-free water to the total volume 10 µl.
2. Perform PCR using the following amplification conditions:

Step	Action	Temp	Time
1	Hold	98°C	30 sec
2	12 cycles steps 2 to 4	98°C	10 sec
3	—	62°C	20 sec
4	—	72°C	10 sec (at 15 sec/Kb)
5	Hold	72°C	7 min
6	Hold	4°C	forever

3. Add 2 µl of 6X Orange DNA Loading Dye and load the entire sample into one well of a 10% TBE gel (1 mm × 10 wells) that has been pre-run for at least 10 min at 200V in 1X TBE buffer. Load 1 µl of DNA Ladder diluted 1:5 with 6X Orange DNA loading dye next to the sample.
4. Allow the sample to enter the gel at 85V, then increase voltage to 200V and run until the orange G dye front reaches the bottom of the gel.
5. Remove gel from the cassette and incubate in SYBR Gold nucleic acid gel staining solution (1:10,000 dilution of the stock in 1X TBE) (*see* Note 16). The goal here is to select an appropriate number of amplification cycles to produce a distribution of bands in the desired size range (~180 to ~350 bp) and to avoid over amplification of one or more prominent bands (Fig. 4).

3.2.4.2. Final PCR: Once the appropriate number of PCR cycles has been selected, perform PCR using the remainder of the size-selected cDNAs. Mix the cDNAs with 18 µl of 5X Phusion HF PCR Buffer, 9 µl of 5' PCR Primer (0.25 µM final, use 2.5 µM stock), 9 µl of 3' PCR index (0.25 µM final, use 2.5 µM stock), 1.8 µl of dNTP mix (0.2 mM final), 0.9 µl (1.2U) of Phusion HF DNA polymerase. Add RNase-free water to bring the total volume of the PCR reaction to 90 µl. Perform PCR using amplification conditions described above and the number of amplification cycles determined for each individual library during the test PCR.

¹⁵For most experiments it is useful to pool cDNA libraries from different samples in order to save time and money on sequencing costs. This is accomplished by adding a unique identifier (i.e. index/barcode) to each library. In the protocol provided here, barcoding occurs during the PCR amplification step and the barcode is contained within the 3' PCR primer. Avoiding cross-contamination of index/barcode stocks is absolutely essential for successful multiplexing. To minimize this risk, only a single tube containing index/barcode should be opened at any time.

¹⁶SYBR Gold nucleic acid gel staining solution is used here in place of EtBr due to its superior sensitivity.

3.2.5. Isolation of PCR amplified cDNAs

1. Add 18 μ l of 6X Orange DNA Loading Dye and load the entire sample into 4 consecutive wells (load $\sim$ 27 μ l in each well) of a 10% TBE gel (1mm $\times$ 10 wells) that has been pre-run for at least 10 min at 200V in 1X TBE buffer.
2. Load 5 μ l of DNA Ladder diluted 1:2 with 6X DNA loading dye next to the sample. Allow the sample to enter the gel at 85V, then increase voltage to 140V and run until the orange G dye front reaches the bottom of the gel.
3. Prepare EtBr staining solution. Remove gel from the cassette and incubate in staining solution for 10 min.
4. Use a UV transilluminator to visualize the libraries and a sterile scalpel to excise the portion of the gel that is between $\sim$ 180 to 350 bp as estimated by comparison with the ladder (Fig. 4) (*see* Note 7 and Note 17); *see* Note 13 for an alternative procedure to size select the PCR products.
5. Use a scalpel to cut the gel piece into 4 pieces of equivalent size. Use the extrusion procedure described in section 3.2.1 (steps 9–10) to crush these gel slices into 2 $\times$ 1.7 mL microcentrifuge tubes.
6. Add 400 μ l 1X Elution Buffer B to each tube. Elute DNA from gel for 2 h at room temperature, then 1 h at 37°C with gentle agitation (or at 4°C for 16h).
7. Use wide orifice P200 barrier tips to transfer $\sim$ 300 μ l of the gel slurry to the top of a Spin-X cellulose acetate filter column (*see* Note 9).
8. Centrifuge 1.5 min at 15,000 $\times$ g. Divide the eluate into two new 1.7 mL tubes (there will be $\sim$ 400 μ l in each tube) and add 8 μ l of glycogen (10 mg/mL) and 560 μ l of isopropanol (*see* Note 14). Vortex.
9. Precipitate libraries at -20°C for at least 2 hours to overnight.
10. Centrifuge samples 30 min at 21,000 $\times$ g at 4°C. Wash pellets with 1 mL of 75% room temperature EtOH. Repeat wash step 2 more times.
11. Air dry pellets 5 min at room temperature. To resuspend DNA, add 5 μ l of nuclease-free water directly to each pellet, rehydrate for at least 5 min at room temperature. Pool the resuspended material ($\sim$ 10 μ l total) into a single tube.

3.3 Adaptor and primer sequences for cDNA library construction

3.3.1. Illumina HiSeq—Sequences of the adaptors and primers to use for generating libraries compatible with Illumina HiSeq platform (*see* Note 18):

Illumina RNA 5' Adaptor sequence (based upon RA5)

¹⁷5' and 3' adaptor sequences account for 93 bp of the final PCR product in libraries generated for SOLiD sequencing and 124 bp of the final PCR product in libraries generated for Illumina.

¹⁸The design of the oligos for construction of cDNA libraries compatible with the Illumina HiSeq platform were based the TruSeq Small RNA primers described in the Illumina customer sequence letter dated Sept 7, 2012: RA5 (5' adaptor), RA3 (3' adaptor), PCR primers [RP1 + Indexes RPI1–8]. We have had success using PAGE purified versions of these adaptors obtained from commercial oligo vendors such as Integrated DNA Technologies (IDT).

5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNN-3' (32 nt) (*see* Note 19)

Illumina RT Primer (note the nine random bases at the 3' end)

5'-GCCTTGGCACCCGAGAATTCCANNNNNNNNN-3' (31 nt)

Illumina 5' PCR Primer (based upon RP1)

5'-

AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGA-3'

Illumina 3' PCR Primers (based upon RPI series) include indexes (i.e. barcodes for multiplexing). Below we have listed 4 of primers; sequence of the index (barcode) is **underlined and in bold**.

3' PCR Primer, Index 1 (RPI1)

5'CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCCTTGGC
ACCCGAGAATTCCA-3'

3' PCR Primer, Index 2 (RPI2)

5'CAAGCAGAAGACGGCATACGAGAT**ACATCG**GTGACTGGAGTTCCTTGGC
ACCCGAGAATTCCA-3'

3' PCR Primer, Index 3 (RPI3)

5'CAAGCAGAAGACGGCATACGAGAT**GCCTAAG**TGACTGGAGTTCCTTGGC
ACCCGAGAATTCCA-3'

3' PCR Primer, Index 4 (RPI4)

5'-

CAAGCAGAAGACGGCATACGAGAT**TGGTCAG**TGACTGGAGTTCCTTGGCAC
CCGAGA ATTCCA-3'

3.3.2. SOLiD—Sequences of the adaptors and primers to use for generating libraries compatible with SOLiD genome analyzer:

SOLiD 5' adaptor sequence (note this is an RNA oligo)

5'-CCACUACGCCUCCGCUUCCUCUCUAUGGGCAGUCGGUGAU -3' (41 nt)

SOLiD, RT primer (note the nine random bases at the 3' end)

5'-CTGCTGTACGGCCAAGGCGNNNNNNNN-3' (28 nt)

SOLiD 5' PCR Primer

5'-CCACTACGCCTCCGCTTTCCTCTCTATG-3'

SOLiD 3' PCR primers include indexes (i.e. barcodes for multiplexing). Below we have listed 4 of primers; sequence of the index (barcode) is **underlined and in bold**.

¹⁹To potentially minimize bias during the ligation step (10), random nucleotides can be inserted at the 3' end of the 5' adaptor. For example, for the Illumina 5' adaptor: 5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNN-3' (32 nt).

Barcode- 001

5'-
CTGCCCCGGGTTTCCTCATTCTCTGTGTAAGAGGCTGCTGTACGGCCAAGGC
G-3'

Barcode- 002

5'-
CTGCCCCGGGTTTCCTCATTCTCTAGGGAGTGGTCTGCTGTACGGCCAAGGC
G-3'

Barcode- 003

5'-
CTGCCCCGGGTTTCCTCATTCTCTATAGGTTATACTGCTGTACGGCCAAGGCG
-3'

Barcode- 004

5'-
CTGCCCCGGGTTTCCTCATTCTCTGGATGCGGTCCTGCTGTACGGCCAAGGC
G-3'

Acknowledgments

This work is supported by NIH grants GM096454 and GM088343.

References

1. Cho BK, Zengler K, Qiu Y, et al. The transcription unit architecture of the *Escherichia coli* genome. *Nat Biotechnol.* 2009; 27:1043–1049. [PubMed: 19881496]
2. Sharma CM, Hoffmann S, Darfeuille F, et al. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature.* 2010; 464:250–255. [PubMed: 20164839]
3. Dornenburg JE, Devita AM, Palumbo MJ, et al. Widespread antisense transcription in *Escherichia coli*. *MBio.* 2010;1.
4. Mitschke J, Georg J, Scholz I, et al. An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci U S A.* 2011; 108:2124–2129. [PubMed: 21245330]
5. Goldman SR, Sharp JS, Vvedenskaya IO, et al. NanoRNAs prime transcription initiation in vivo. *Mol Cell.* 2011; 42:817–825. [PubMed: 21700226]
6. Vvedenskaya IO, Sharp JS, Goldman SR, et al. Growth phase-dependent control of transcription start site selection and gene expression by nanoRNAs. *Genes & Dev.* 2012; 26:1498–1507. [PubMed: 22751503]
7. Turnbough CL Jr. Regulation of gene expression by reiterative transcription. *Curr Opin in Microbiol.* 2011; 14:142–147.
8. Miller DF, Yan PS, Buechlein A, et al. A new method for stranded whole transcriptome RNA-seq. *Methods.* 2013; 63:119–125. [PubMed: 23583680]
9. Adiconis X, Borges-Rivera D, Satija R, et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods.* 2013; 10:623–629. [PubMed: 23685885]
10. Zhuang F, Fuchs RT, Sun Z, et al. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res.* 2012; 40:e54. [PubMed: 22241775]

Fig. 1. Enzymatic treatments to prepare RNA for cDNA library construction. Species of interest are shown in purple.

Fig. 2.
cDNA library construction

Fig. 3. Ligation of 5' adaptor. Material that has been ligated to the 5' adaptor after electrophoresis on a 6% TBE-Urea gel. Black box shows the position of the excess 5' adaptor. The dashed yellow box indicated the portion of the gel that should be excised for gel extraction. M: Ultra Low Range DNA ladder.

Fig. 4.

Test PCR analysis. cDNA libraries analyzed on a 10% TBE gel after test PCR. M: Ultra Low Range DNA ladder, Lane 1: library has undergone the appropriate number of PCR cycles. Lane 2: library is under amplified (increase PCR cycles by at least 1 for final PCR). Lane 3: library is over amplified (decrease PCR cycles by at least 1 for final PCR). Red stars indicate prominent bands indicative of library over amplification. The dashed yellow box indicated the portion of the gel that will be excised for gel extraction after the final PCR is performed. Also indicated are prominent non-specific products that appear in libraries generated for use with the SOLiD genome analyzer. These products migrate around ~150 bp and should be avoided in the final library.