RNA-Seq for Bacterial Gene Expression

Poulsen, Line Dahl; Vinther, Jeppe

Published in: Current Protocols in Nucleic Acid Chemistry

DOI: 10.1002/cpnc.55

Publication date: 2018

Document version: Peer reviewed version

Significance Statement

In recent years, genome-wide studies have provided major contributions in understanding of gene expression and regulation in bacteria. The method of choice has become RNA sequencing (RNA-seq), where RNA molecules are converted into cDNA, which is used as input for library preparation for next-generation sequencing. The obtained sequences provide information both on which RNA transcripts are expressed in a cell population at a given time and on the expression level of the transcripts. Variations of RNA-seq protocols exist; however, many methods are expensive and time-consuming. We provide a detailed, easy-to-follow RNA-seq protocol which enables reproducible generation of high quality sequencing reads. The data obtained allows quantification of gene expression and identification of novel transcripts, thereby contributing to an increased understanding of bacterial populations.
RNA-seq for bacterial gene expression

Line Dahl Poulsen\textsuperscript{1} and Jeppe Vinther\textsuperscript{1,*}

\textsuperscript{1} Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark

* To whom correspondence should be addressed. Tel: +4535321264; Fax: +4535322128; Email: jvinther@bio.ku.dk

ABSTRACT

RNA sequencing (RNA-seq) has become the preferred method for global quantification of bacterial gene expression. With the continued improvements in sequencing technology and data analysis tools, the most labor-intensive and expensive part of an RNA-seq experiment is the preparation of sequencing libraries, which is also essential for the quality of the data obtained. Here, we present a straightforward and inexpensive protocol for preparation of strand specific RNA-seq libraries from bacterial RNA as well as a computational pipeline for the data analysis of sequencing reads. The protocol is based on the Illumina platform and allows easy multiplexing of samples and the removal of sequencing reads that are PCR duplicates.

Keywords:

RNA-seq; next-generation sequencing; library construction

INTRODUCTION

This unit describes the experimental steps required for preparation of RNA-seq libraries from total RNA isolated from bacteria. The method is based on a protocol for detection of reverse transcriptase termination sites (Lukasz J Kielbinski, Boyd, Sandelin, & Vinther, 2013), which we here have adapted for RNA-seq. As an example, we applied the method to RNA isolated from \textit{Bacillus subtilis}; however, the method is general and can be used to generate RNA-seq libraries from a wide variety of bacteria. Depletion of ribosomal RNA from the RNA sample can be achieved either with the Ribo-zero\textsuperscript{®} rRNA Removal Kit (Basic protocol 1) or by selective degradation with the 5\textsuperscript{-}monophosphate-dependent Terminator\textsuperscript{TM} exonuclease (Alternate protocol 1). While we provide a RNA-seq dataset and some general recommendations for data
analysis (See Internet Resources), more detailed pipelines have been described previously (Lukasz Jan Kielpinski, Sidiropoulos, & Vinther, 2015).

**STRATEGIC PLANNING**

The experiments should be carefully planned prior to sample collection. It is recommended to include at least three biological replicates for the statistical analysis, and also to include a technical replicate to allow estimation of the reproducibility of the protocol. Another important parameter to consider before performing a sequencing experiment is the sequence depth (or library size). The higher the number of reads in each sample, the more RNAs will be detected and quantified. For a typical bacterial transcriptome such as *Bacillus subtilis*, 1-2 million reads per replicate will be sufficient to achieve good coverage on the majority of expressed mRNAs when using input RNA depleted for ribosomal RNA with the Ribo-zero® rRNA Removal Kit or a similar kit (Basic protocol 1). In some cases, it can be cost-effective to use the 5'−monophosphate-dependent Terminator™ exonuclease (Alternate protocol 1) instead of the Ribo-zero® rRNA Removal Kit, and then increase sequence depth to obtain sufficient coverage on mRNAs. For *Bacillus subtilis*, which have around 90% of rRNA content, treatment with the exonuclease will result in a reduction of rRNA transcripts to around 60%, and around 4-5 million reads will be enough for this approach. If the goal of the experiment is to examine ribosomal RNAs, then much fewer reads are necessary.

**BASIC PROTOCOL 1**

**LIBRARY PREPARATION FROM TOTAL RNA**

This protocol describes bacterial RNA sequencing from total RNA. The protocol for library preparation includes the following steps (Figure 1): first, the ribosomal RNA (rRNA) is depleted using the Ribo-zero® rRNA Removal Kit (Bacteria), which is based on selective binding of biotinylated probes to rRNAs using a hybridization/bead capture procedure. The remaining mRNA is then fragmented and reverse transcribed using a random primer with an Illumina adapter overhang, and an adapter is ligated to the 3' end of the cDNA. The cDNA is PCR amplified using an indexed primer to allow multiplexing. Finally, the DNA libraries are quantified and pooled, and 200-600 bp fragments are selected for sequencing. The libraries are strand specific and compatible with the different Illumina sequencing platforms, including miSeq,
HighSeq and NextSeq and can be distinguished by their unique single 6 base index in the P7 region of the adapter.

**Materials**

RNA of interest (1-5 µg of total RNA from bacteria; see Critical Parameters)

Ribo-zero® rRNA Removal Kit (Bacteria)

Glycogen (20 mg/ml)

Sodium acetate pH 5.2 (3 M)

70 % and 100 % ethanol

RNase-free water

2x Fragmentation buffer (see Reagents and Solutions)

EDTA (50 mM)

Agencourt® RNAClean® XP (Beckman Coulter)

Reverse transcription primer (RT_random_primer, table 1)

dNTPs (10 mM)

PrimeScript® Reverse Transcriptase (Takara)

5X PrimeScript Buffer (Takara)

Agencourt® AMPure® XP (Beckman Coulter)

CircLigase™ 10X Reaction Buffer

Ligation adapter (3'-Ligation_Adapter, table 1)

ATP (1 mM)

MnCl₂ (50 mM)

50% PEG 6000

CircLigaseTM ssDNA Ligase (100 U/µl) (Epicentre)

PCR forward primer (10 µM) (PCR_FW, table 1)

PCR reverse primer (10 µM) (PCR_RV_index#, table 1)
Phusion® High-Fidelity DNA Polymerase (2000 U/ml) (New England BioLabs)
5x Phusion HF buffer (New England BioLabs)

E-gel™ SizeSelect™ II Agarose gel, 2% (or similar gel for separation of DNA fragments)
Nanodrop ND-1000 Spectrophotometer (or similar equipment)
Agilent 2100 Bioanalyzer
Thermocycler for PCR tubes
Magnetic stand
Savant™ DNA SpeedVac® DNA120 (optional)
E-Gel® iBase™ Power System (or similar system for size-selection of DNA-fragments)

**rRNA depletion with ribozero kit**

1. Deplete rRNA with the Ribo-zero® rRNA Removal Kit (Bacteria) (Use a total of 1-5 µg total RNA as input, and follow the manufacturer’s instructions).
2. Clean up the rRNA depleted RNA by precipitation. Add the following:
   - RNase-free water to 200 µl
   - 20 µl 3 M sodium acetate (pH 5.2)
   - 1 µl 20 mg/ml glycogen
   - 600 µl 100 % ethanol
3. Vortex and incubate at -20°C for at least an hour.
4. Centrifuge at 12,000 g for 30 min at 4 °C.
5. Discard the supernatant, and add 1 ml 70 % ethanol to wash the pellet.
6. Centrifuge at 12,000 g for 30 min at 4 °C.
7. Discard the supernatant, and centrifuge briefly to collect and remove residual ethanol.
8. Resuspend the pellet in 12 µl RNase-free water.
9. Measure the RNA concentration on a fluorometer or a spectrophotometer to assess the yield.
10. To evaluate the RNA quality, run 1 µl Ribo-zero-treated RNA (with a concentration around 5 ng/µl) on an Agilent 2100 Bioanalyzer using a RNA 6000 Pico Chip (Figure 2).
RNA fragmentation

11. On ice, add 5 µl 2 x fragmentation buffer to 5 µl RNA sample.
12. Incubate at 95°C for 120 s (see critical parameters and troubleshooting).
13. Place the samples back on ice immediately after incubation, and add 2.5 µl 50 mM EDTA.
14. Add 18 µl Agencourt® RNAClean® XP beads to the samples and incubate at room temperature for 10 min, mixing after 5 min by vortexing.
15. Set the tube on the magnetic stand for 5 min and discard the supernatant.
16. Keep the sample on the magnetic stand and wash the beads with ethanol by pouring 150 µl of 85% EtOH. Wait 30 s and then remove the supernatant. Repeat this washing step once more.
17. Add 10 µl water to the beads and pipette at least 60 times to elute the RNA.
18. Set the tube on the magnetic stand for 5 min to separate the beads. Collect the supernatant and keep it on ice.

First-strand synthesis

19. In a PCR tube, mix 9.5 µl RNA and 0.5 µl of 10 µM reverse transcription primer (RT_random_primer).
20. Incubate at 65°C for 5 min and cool on ice.
21. Prepare master mix. For one reaction, add:
   - 4 µl 5X PrimeScript Buffer
   - 1 µl 10 mM dNTPs
   - 4 µl water
   - 1 µl PrimeScript Reverse Transcriptase
22. Mix gently and spin down.
23. Add 10 µl master mix to the RNA-reverse transcription primer solution, mix gently and spin down.
24. Incubate in a thermal cycler as follows: 30°C, 10 min; 42°C, 60 min; 70°C, 15 min; and keep at 4°C until next step.

cDNA purification and up-concentration

25. Add 36 µl Agencourt® AMPure® XP beads to the samples and incubate at room temperature for 20 min, mixing every 5 min by vortexing.
26. Set the tube on the magnetic stand for 5 min and discard the supernatant.
27. Keep the sample on the magnetic stand and wash the beads with ethanol by pouring 150 µl of 75% EtOH. Wait 30 sec and then remove the supernatant. Repeat this washing step once more.
28. Add 20 µl water to the beads and pipette at least 60 times to elute the cDNA.
29. Set the tube on the magnetic stand for 5 min to separate the beads. Collect the supernatant.
30. Use a SpeedVac to decrease the sample volume to 4 µl (alternatively, the DNA can be ethanol-precipitated and resuspended in 4 µl water).

**Adapter ligation and purification**

31. Prepare master mix. For one reaction, add:
   - 1 µl CircLigase™ 10X Reaction Buffer
   - 0.5 µl 100 µM adapter (3’-ligation_adapter)
   - 0.5 µl 1 mM ATP
   - 0.5 µl 50 mM MnCl₂
   - 2 µl 50% PEG 6000
   - 2 µl water
   - 0.5 µl CircLigaseTM ssDNA Ligase (100 U/µl)
32. Mix 7 µl master mix with 3 µl cDNA.
33. Incubate 2 hours at 60°C, 1 hour at 68°C, and 10 min at 80°C; and keep at 4°C until next step.
34. Purify with Agencourt® AMPure® XP beads as in step 15-19, but this time add 18 µl beads to the 10 µl cDNA solution and elute in 10 µl water.

**Library amplification and indexing**

35. Prepare master mix on ice without the reverse PCR primer. For one reaction, add:
   - 2.5 µl 10 µM PCR forward primer (PCR_FW, table 1)
   - 10 µl 5xHF buffer
   - 1 µl 10 mM dNTPs
   - 28 µl water
   - 1 µl Phusion® High-Fidelity DNA Polymerase (2000 U/ml)
36. Mix and split the master mix into 42.5 µl aliquots in PCR tubes
37. Add 5 µl cDNA template and 2.5 µl PCR reverse primer (10 µM) (PCR_RV_index#, table 1) (use different index primers for multiplexing)
38. Mix gently and spin down
39. Transfer the PCR tubes to a PCR machine with the block preheated to 98°C and start thermocycling immediately: 98°C, 3 min; (98°C, 80 s; 68°C, 30 s; 72°C, 30 s) x 12; 72, 5 min; hold on 4°C
40. Purify with Agencourt® AMPure® XP beads as in step 15-19, but this time add 90 µl beads to the 50 µl PCR reaction and elute in 50 µl water.
41. Run 1 µl purified PCR product on an Agilent 2100 Bioanalyzer using a DNA High Sensitivity chip (Figure 3A).

**Quantification and Size selection**

42. For multiplexing: after running PCR product on an Agilent 2100 Bioanalyzer using a DNA High Sensitivity chip, perform smear analysis to quantify the DNA. This is performed by using the 2100 Expert software and choosing: Global (in the side panel) > advanced > regions (under smear analysis) > set regions to 200 bp to 600 bp. Choose the individual traces to estimate the molarity in the given region in the ‘region table’ below the trace. Use these estimates to mix the DNA samples in the molar ratios relevant for your experiment (typically equimolar ratios for obtaining approximately the same number of reads for each sample).
43. After mixing the PCR products, use an E-gel™ SizeSelect™ II Agarose gel (or a similar system) to select the 200-600 bp region. Alternatively, size selection can be performed by gel electrophoresis followed by excision of the desired bands.
44. Use a SpeedVac to decrease the sample volume to 40 µl (alternatively, the DNA can be ethanol-precipitated and resuspended in 40 µl water).
45. Purify with Agencourt® AMPure® XP beads as in step 14-18, but this time add 72 µl beads to the 40 µl collected sample and elute in 50 µl water.
46. Run 1 µl purified PCR product on an Agilent 2100 Bioanalyzer using a DNA High Sensitivity chip or a DNA1000 chip, to validate successful size-selection of the library (Figure 3B). After ensuring correct size selection, the libraries are ready for sequencing on an Illumina sequencing platform.
**ALTERNATE PROTOCOL 1**

**DEPLETION OF RIBOSOMAL RNA WITH TERMINATOR 5’-PHOSPHATE-DEPENDENT EXONUCLEASE**

This alternate protocol uses Terminator™ 5'-phosphate-dependent exonuclease (TEX) to enrich for bacterial mRNAs, an approach that is fundamentally different from the Ribo-Zero method, which uses probes that hybridize to ribosomal RNAs. Instead, TEX selectively degrades RNAs with a 5'-monophosphate. In bacteria, the genes encoding 23S rRNA, 16S rRNA and 5S rRNA are typically co-transcribed and then processed post-transcriptionally, resulting in 23S rRNA and 16S rRNA transcripts with 5'-monophosphates that are accessible for TEX degradation. It is important to note that 5S rRNA and tRNAs are not degraded by TEX, and if these are not removed by additional purification steps, they will contribute to a substantial fraction of the sequencing reads. Furthermore, some RNAs are protected from TEX digestion by secondary structures that protect their 5'-monophosphates against the exonuclease. It is essential that the input RNA is of high quality when using TEX to enrich for mRNAs, as rRNA degradation products with 5’-OH groups are also not degraded by TEX.

**Materials**

RNA of interest (2 µg total RNA from bacteria; see Critical Parameters)

RNase-free water

Terminator 10X Reaction Buffer A (Epicentre)

Terminator™ 5'-phosphate-dependent exonuclease (1 U/µl) (Epicentre)

RiboGuard™ RNase Inhibitor

EDTA (100 mM, pH 8.0)

Agencourt® RNAClean® XP (Beckman Coulter)

ThermoMixer

Magnetic stand
**rRNA depletion with Terminator 5'-Phosphate Dependent exonuclease**

1. Adjust volume of RNA to 15.5 µl with RNase-free water.
2. On ice, add the following:
   - 2 µl Terminator 10X Reaction Buffer A
   - 0.5 µl RiboGuard RNase Inhibitor
   - 2 µl Terminator 5'-phosphate-dependent exonuclease (1 U/ µl)
3. Mix gently and spin down.
4. Incubate 30°C for 60 min in a ThermoMixer.
5. Add 1 µl of 100 mM EDTA (pH 8.0) to terminate the reaction.

**RNA purification**

6. Add 50 µl Agencourt® RNAClean® XP beads to the samples and incubate at room temperature for 10 min, mixing after 5 min by vortexing.
7. Set the tube on the magnetic stand for 5 min and discard the supernatant.
8. Keep the sample on the magnetic stand and wash the beads with ethanol by pouring 150 µl of 85% EtOH. Wait 30 sec and then remove the supernatant. Repeat this washing step once more.
9. Add 20 µl water to the beads and pipette at least 60 times to elute the cDNA.
10. Set the tube on the magnetic stand for 5 min to separate the beads. Collect the supernatant and keep it on ice.
11. Measure the RNA concentration on a fluorometer or a spectrophotometer to assess the yield.
12. To evaluate the RNA, run 1 µl TEX-treated RNA (with a concentration around 5 ng/µl) on an Agilent 2100 Bioanalyzer using a RNA 6000 Pico Chip (Figure 2). The RNA is now ready for library preparation (from step 11 of Basic protocol 1).

**REAGENTS AND SOLUTIONS**

**2x Fragmentation buffer**

10 mL 1 M Tris-HCl pH 8.0
1 mL 1 M MgCl₂
Bring to 100 mL with RNase-free water
Autoclave and store at room temperature.

**COMMENTARY**

**Background Information**

Methods for quantification of gene expression have been central for our understanding of cellular properties and how cells interact with and adapt to their environment. Over the years, the methodologies have improved, thereby allowing gene expression of many thousands of genes to be analyzed in one experiment. In the 1990s, microarrays allowed high throughput identification of candidate transcripts, thereby revolutionizing the approach to study gene expression (Schena, Shalon, Davis, & Brown, 1995), but recently with the invention of massive parallel sequencing, RNA-seq has become the method of choice to study gene expression. Compared with DNA arrays, RNA-seq has an improved dynamic range for quantification of gene expression levels and and allows the identification novel RNA species (Wang, Gerstein, & Snyder, 2009). In bacteria, both RNA-seq and more specialized protocols based on sequencing have revolutionized the understanding of bacterial transcriptomes (Croucher & Thomson, 2010; Sharma & Vogel, 2014).

The library preparation is essential for a successful RNA-seq experiment and many different commercial kits are available. However, these kits remain relatively expensive and in some cases contain secret reagents or compositions, which create the need for alternative RNA-Seq protocols that are robust and in addition time and cost effective. The key point for construction of a sequencing library, which can be subjected to massive parallel sequencing on the Illumina platform, is the requirement for addition of adapters to the ends of DNA that are sequenced. Over the years, different strategies have been developed for this. In the initial RNA-seq experiments, adapters were attached to double-stranded cDNAs, and this resulted in a loss of directionality (Nagalakshmi et al., 2008). Later, several strategies have been developed to preserve strand specificity, including attaching adapter sequences to RNA molecules prior to reverse transcription (Mamanova et al., 2010) and incorporation of dUTP instead of dTTP in the second strand cDNA synthesis to allow selective degradation of the second strand following adapter ligation (Zhang, Theurkauf, Weng, & Zamore, 2012). Also, the SMART-seq strategy
takes advantage of the template switching mechanism of reverse transcriptase for adding a 3’ adapter prior to PCR amplification (Zhu, Machleder, Chenchik, Li, & Siebert, 2001).

Here, we present a protocol that is based on the ligation of an adapter to the 3’ end of cDNAs produced by reverse transcription of an RT-primer carrying an Illumina adapter sequence overhang (Li & Weeks, 2006). We previously used this strategy for preparing libraries for RNA probing experiments (Lukasz J Kielpinski et al., 2013; Poulsen, Kielpinski, Salama, Krogh, & Vinther, 2015), but it also works nicely for RNA-seq that is based on sequencing of fragmented purified mRNAs. Our strategy resembles the Ligation Mediated RNA sequencing protocol developed by Thomson and co-workers for eukaryotic RNA-seq (Hou et al., 2015) and has many of the same advantages, including low costs and time consumption. In addition, our protocol includes the possibility to recognise PCR duplicates through the use of a barcode in the adapter ligated to the 3’ end of the cDNA, which is especially valuable for samples with limited input material.

**Critical Parameters and Troubleshooting**

The integrity of the input RNA is critical for obtaining high quality RNA-seq data. While RNA isolation is not covered in this protocol, numerous methods are available in the literature and finding the most appropriate for a given bacteria species is essential for performing a successful experiment. Regardless of which method is chosen, the RNA quality should be measured before RNA-seq. The RNA quality can be measured with an Agilent Bioanalyzer which will produce a RNA Integrity number (RIN). The RIN number is between 1 and 10, where 10 corresponds to the highest quality with no RNA degradation. Low RIN values may result in incorrect biological conclusions, and we recommend using RNA that has a RIN value of at least 7. Alternatively, the RNA can be run on an agarose gel to infer RNA integrity, but this method is less sensitive.

Since prokaryotic RNAs generally are non-polyadenylated, a major issue in library preparation is depletion of ribosomal RNAs, which are the most abundant RNA species in the cells. Different approaches have been used to solve this challenge, for instance mRNA polyadenylation followed by binding to poly(dT) beads (Amara & Vijaya, 1997) or capture of rRNAs with sequence-specific biotinylated probes. Here, we have used the Ribo-Zero rRNA removal Kit (Bacteria) (Basic protocol 1) which resulted in almost complete depletion of rRNA however,
several other kits are commercially available. An alternative is treatment with Terminator™ 5'-phosphate-dependent exonuclease (TEX) (Alternate protocol 1) which selectively degrades RNAs containing 5'-monophosphates such as rRNAs. The enrichment for mRNAs is not as efficient, as degradation can be blocked by secondary structures in the substrate RNA. Also, it is important to remember that treatment with TEX results in enrichment of primary transcripts, as processed transcripts often contain 5'-monophosphates.

The presence of a band with a length corresponding to around 145 bp indicates a problem in library preparation. This size is equal to direct ligation of the 3'-ligation adapter to the random RT primer followed by PCR amplification, but without insert (Figure 3A). Observing this band confirms successful 3'-adapter ligation and PCR amplification, but also indicates either too low input amount or a high degree of fragmentation of the input RNA. If fragments of larger sizes are also observed in the library, then the problem may be solved with a size selection step as described in basic protocol 1. However, if no other bands are observed, it will be necessary to redo the library preparation with more input RNA, and it might also be advantageous to test different RNA fragmentation incubation times.

PCR duplicates are sequence reads that arise from amplification of the same template RNA molecule, and these represent a major issue in next-generation sequencing experiments. When the RNA starting material is limited, PCR amplification is necessary to obtain enough material for sequencing; however, increasing the number of cycles in the PCR also increases the risk of PCR duplicates. In general, it is recommended to perform a small-scale PCR with different number of cycles to choose the lowest number of cycles that are visible either on a gel or on a Bioanalyzer DNA chip. To accommodate potential problems with PCR duplicates, our protocol introduces a random barcode in the 3’ cDNA ligation step prior to PCR amplification. Observing multiple reads with identical barcodes on the same DNA fragment indicate that these are PCR duplicates and should be collapsed to one read.

**Anticipated Results**

RNA-seq results in raw sequencing reads that must be processed before biological interpretation. The analysis of an RNA-seq experiment has many variations, and the exact
pipeline depends on the aims of a given experiment. The main focus of this unit is preparation of high quality RNA-seq libraries from bacteria; however, we will briefly describe the framework for analyzing the obtained data and the expected results. We provide a dataset that has been generated with the protocol and we also provide command line scripts for preprocessing, mapping and barcode collapsing (see Internet Resources).

The first step in any data analysis pipeline is quality control, and this can be done with FASTQC (Andrews, 2010) or similar programs. Examining plots from the quality control is important for detecting and subsequently dealing with potential problems in the libraries. Following satisfying quality control, the adapter sequence should be removed from the reads. One of the tools developed for this purpose is Cutadapt (Martin, 2011). Here, default settings can be used, and with the quality cutoff (-q option) set to 20. For sequences obtained for Illuminas two-dye systems (MiniSeq, NextSeq and NovaSeq), use the --nextseq-trim=20 option to remove 3’ terminal Gs stemming from dark cycles on short fragments. If the reads are paired-end reads instead of single-end, then Cutadapt should be run once for each FASTQ file with the respective adapter sequence. Following adapter removal, the preprocessing tool developed in our group (Lukasz Jan Kielpinski et al., 2015) can be used to remove the barcode introduced at the 3’-end of the cDNA during library preparation. The preprocessed reads can now be aligned to the investigated RNAs or genomes, and this can be done with a number of alignment programs. In the provided example bowtie2 is used (Langmead & Salzberg, 2012). Following mapping, PCR duplicates can be removed with the collapse script which will output a .sam file. The code in this script collapses reads that mapped to the same position, if they also contain identical barcodes.

The provided example contains libraries generated from: 1) total RNA with no rRNA depletion, 2) rRNA depleted RNA using the Ribo-Zero rRNA removal Kit (Basic protocol 1), and 3) Terminator™ 5’-phosphate-dependent exonuclease (TEX) treated RNA (Alternate protocol 1). Following adapter removal and preprocessing, the reads have been aligned to the reference genome for the strain used in this experiment (Bacillus subtilis str. 168 (ASM904v1) from Ensembl). The number of reads aligning to the genome is 93.1-98.7% (Table 2). Aligning only to rRNA sequences shows that 90.5% of all reads from the total RNA sample maps to rRNAs, and as expected, rRNA depletion with TEX lowers this number to 64.2%. In contrast, rRNA depletion using the Ribo-Zero rRNA removal Kit as described in Basic protocol 1 result in only 0.1% mapping to the ribosomal RNAs. Using the provided collapse script to remove PCR duplicates, we can see that including a barcode may be beneficial to avoid PCR duplicates (Table 2).
Downstream analysis will depend on the specific scientific questions that are being investigated. In a standard RNA-seq setup, the reads mapping to the different genes are counted with a tool such as HT-seq. Finally, gene counts from replicate experiments are analyzed with specialized tools such as EdgeR or DESeq2 based on the negative binominal distribution to produce lists of differentially expressed genes with associated fold changes and significance estimates.

**Time Considerations**

Preparation of RNA-seq libraries from isolated RNA can be completed in 4-5 days. The use of a SpeedVac instead of ethanol precipitation for up-concentration of RNA or DNA molecules significantly reduce the time needed to generate libraries. The time needed to perform data analysis can vary from days to weeks depending on which analyses are performed.

**ACKNOWLEDGEMENT**

The research was funded by Innovation Fund Denmark [5163-00010B].

**LITERATURE CITED**


INTERNET RESOURCES

The RNA-seq dataset and an example of the data analysis workflow are available for download at http://people.binf.ku.dk/~jvinther/data/RNA-seq
### TABLES

**Table 1.** Oligonucleotides used for library preparation. The oligonucleotide sequences of the Illumina genomic DNA adapters are copyrighted by Illumina, Inc. 2006. All rights reserved. Index sequences for multiplexing are shown in bold. A table of additional PCR index primers can be found in (Lukasz J Kielpinski et al., 2013).

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
</table>
| RT_random_primer      | AGACGTCGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

**Table 2.** Mapping statistics for the example data provided.
The starting material is total RNA from bacteria. The rRNA is depleted from this sample and RNA is then fragmented. Following purification, RNA is reverse transcribed to cDNA using a random primer containing a 5’ adapter overhang. After ligation of an adaptor to the 3’-end, the cDNA serves as a template for PCR amplification adding Illumina overhangs, which are necessary for sequencing.
Figure 2: Bioanalyzer electropherogram of RNA molecules. (A) Total RNA which has an RIN value of 9.6. (B) RNA sample following rRNA depletion with the Ribo-zeroR rRNA Removal Kit (bacteria). (C) RNA sample following rRNA depletion with TerminatorTM 5'-phosphate-dependent exonuclease (TEX; Alternate Protocol). (D) An example of an RNA sample after fragmentation (shown here is fragmented total RNA). RIN, RNA integrity number.
Figure 3 DNA libraries. (A) Example of a bioanalyzer electropherogram of a library before size selection. The peak around 145 bp (black arrow) represents ligation of the 3'-ligation adapter to the random RT primer followed by PCR amplification but without insert. (B) Bioanalyzer electropherogram of a library after size selection.