



Published in final edited form as:

Curr Protoc Mol Biol. 2013 July ; 0 4: Unit–4.17. doi:10.1002/0471142727.mb0417s103.

Quantification of microRNA Expression with Next-Generation Sequencing

Seda Eminaga^{1,*}, Danos C. Christodoulou^{1,*}, Francois Vigneault^{1,2,3,*}, George M. Church^{1,2}, and Jonathan G. Seidman¹

¹Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

²Wyss Institute for Biologically Inspired Engineering, Boston, MA 02115, USA

³Ragon Institute of MGH, MIT and Harvard, Massachusetts, MA 02129, USA

Abstract

Rapid advancement of next generation sequencing technologies has made it possible to study expression profiles of microRNAs (miRNAs) comprehensively and efficiently. We have previously shown that multiplexing miRNA libraries by barcoding can significantly reduce sequencing cost per sample without compromising library quality [Alon et al. 2011, Vigneault et al 2012]. In this unit, we provide a step-by-step protocol to isolate miRNAs and construct multiplexed miRNA libraries. We also describe a custom computational pipeline designed to analyze the multiplexed miRNA library sequencing reads generated by Illumina-based technology.

INTRODUCTION

miRNAs are short, 17–25 nucleotide non-coding RNAs that have emerged as critical regulators of gene expression in various physiological and pathophysiological processes [Mendell and Olson, 2012] due to their ability to influence the stability/translation of a large number of RNAs. Our understanding of roles of miRNAs and their targets has been aided by miRNA expression profiling studies. While high-throughput platforms including multiplex PCR and microarrays have proven to be valuable tools, next generation sequencing technology has quickly emerged as the preferred platform for studying miRNA expression. One of the advantages of next-generation sequencing is the ability to pool and sequence multiple samples in one lane of a sequencer, significantly lowering costs, without compromising the ability to construct comprehensive expression profiles for every assessed sample.

A typical miRNA library construction protocol involves 3' and 5' adapter ligation to miRNAs, followed by reverse transcription to generate cDNA libraries which are then amplified by PCR and purified by gel extraction prior to sequencing [Motameny et al. 2010]. Multiplexing can be achieved by addition of barcodes (unique nucleotide tag sequences incorporated into the adapter or PCR primer) during library construction. However, barcodes on the adapter confer bias during ligation and thus must be avoided,

*These authors contributed equally to this work.

whereas barcodes/indexes introduced during the amplification step is a safe alternative (Alon et al., 2011).

One of the challenges of next generation sequencing is to sift through large data sets of millions of short reads generated by sequencers requiring computational analysis. Reads from miRNA sequencing are first processed to remove the 3' adapter sequences and then matched to the reference sequence for identification of mature and/or novel miRNAs. Identifying miRNAs with differential expression among samples by quantifying the number of reads per miRNA helps infer relevant biological processes.

The purpose of this unit is to provide a step-by-step protocol of total RNA extraction to generation of miRNA expression profiles in tissues of interest. First, we describe a protocol to isolate miRNAs and construct multiplexed miRNA libraries for Illumina sequencing (Alon et al 2011, Vigneault et al, 2012: CPHG Unit 11.12.1–10). Next, we describe a custom bioinformatics pipeline designed to construct comprehensive miRNA expression profiles and assess differential expression between samples. Any measured changes can be quantified with precision as each miRNA is expected to be sampled a large number of times during this process. [Although this pipeline is tested for Illumina reads, it can be adapted for any other platform.]

BASIC PROTOCOL 1

ISOLATION OF TOTAL RNA CONTAINING miRNAs

Total RNA containing miRNAs can be isolated from tissues either by Trizol/Phenol/Chloroform extraction or with commercially available kits. One important determinant when choosing a commercially available RNA isolation kit is whether they retain small RNAs (<200 nucleotides). We have successfully used Ambion's miRVana miRNA isolation kit to isolate good quality total RNA, specifically from mouse heart tissue. Before starting isolation, make sure to clean all equipment and bench-top with RNase Zap and frequently change gloves to prevent any RNase contamination.

Materials—RNase Zap (Ambion, cat. no AM9780)

RNAlater (Qiagen, cat. no. 76106)

RNase-free 1.5 ml microcentrifuge tubes

RNase-free tips

miRVana miRNA isolation kit (Ambion, cat. no. AM1560)

ACS grade 100% Ethanol

Microcentrifuge

Bioanalyzer 2100 (Agilent)

1. Isolate total RNA according to manufacturer's protocol (miRVana isolation kit or Trizol).

We recommend isolating RNA from fresh tissue or tissue stored in RNAlater. For good quality RNA, the time between tissue dissection and RNA isolation should be minimized. We prefer homogenizing with steel beads using TissueLyser (Qiagen) as it allows processing of several samples simultaneously, and prevents cross-contamination between samples. Alternatively, if TissueLyser is not available, we prefer homogenizing using rotor/stator type homogenizer; however, equipment should be thoroughly cleaned between processing of different samples to prevent contamination. If the tissue will not be immediately used for RNA isolation, it should be snap-frozen in liquid nitrogen and stored at -80°C . When isolating RNA from frozen tissue, tissue should first be grinded to powder in liquid nitrogen using prechilled mortar/pestle to prevent thawing of tissue before lysis. Basic Protocol 2 for miRNA library construction works well on total RNA, but if enrichment is desired, an aliquot of total RNA should be saved for quality control and miRVana kit can be used to enrich for miRNAs. Although we have not directly compared expression profiles prepared from miRNA-enriched vs. total RNA, there may be some differences and we recommend that when different samples are compared, they should be prepared with the same method to prevent any bias.

2. Confirm quality of RNA.

We strongly recommend that RNA quality is confirmed by Bioanalyzer (Agilent) and we have observed that RNA with RNA Integrity Number (RIN) >8 provides high quality libraries. Alternatively, RNA quality may be checked by running an agarose gel to visualize 28S and 18S RNA. A 2:1 ratio of 28S to 18S indicates good quality RNA.

3. Store RNA in small aliquots at -80°C .

Repeated freeze-thaw of RNA should be avoided to minimize RNA degradation, as degraded fragments can adversely affect multiple steps and possibly introduce biases.

BASIC PROTOCOL 2

MULTIPLEX microRNA LIBRARY CONSTRUCTION FOR ILLUMINA SEQUENCING

[Basic Protocol 2 has previously been published and incorporated from Vigneault et al in Current Protocols in Human Genetics, CPHG Unit 11.12.1–10]

This procedure describes a method for constructing multiplexed miRNA libraries. A miRNA library is made (figure 1) from each RNA sample by 3' adapter ligation, 5' RT primer annealing, 5' adapter ligation, reverse transcription, and PCR amplification. Although the forward PCR primer is the same, a different reverse PCR primer with a unique barcode is used for each RNA sample. The different libraries can then be pooled into a single sequencing reaction at the end of the library construction. The following instructions are for the preparation of one sample, so users must scale-up according to the specific number of samples they are preparing. All incubations are conducted in a thermal cycler.

Materials—RNase Zap (Ambion, AM9780)

Nuclease-free water (Ambion, AM9937)

Starting RNA

10× T4 RNA Ligase 2tr Buffer (Enzymatics, L607)

3' rApp-adapter (see Table 1)

100% DMSO (Sigma, D9170)

RNase Inhibitor (Enzymatics, Y924L)

RT primer (see Table 1)

5' RNA adapter (see Table 1)

ATP (Enzymatics, N207-10-L)

T4 RNA Ligase 1 (Enzymatics, L605L)

dNTPs (Enzymatics, N205L)

Superscript III First-Strand Synthesis System (Invitrogen, 180080-051)

Phusion High-Fidelity DNA Polymerase (NEB, M0530S)

BCmiRNA_PCR1 (see Table 1)

BCmiRNA_PCR2_BC (see Table 1)

AgencourtAMPure XP 5 mL Kit (Beckman Coulter Genomics, A63880)

E-Gel EX Gel, 2% (Invitrogen, G4020-02)

70% (v/v) ethanol

25 bp Ladder (Invitrogen, 10597-011)

100 bp Ladder (Invitrogen, 15628-019)

MinElute Reaction Cleanup Kit (Qiagen, 28204)

Agilent High Sensitivity DNA Kit (Agilent, 5067-4626)

Thermal Cycler (for all incubations)

E-Gel I-Base Power System (Invitrogen, G6400)

E-Gel Safe Imager Real-Time Transilluminator (Invitrogen, G6500)

Dynamag-2 Magnet (Invitrogen, 123-21D)

Recommended: Iceless Cold Pack (Eppendorf 022510509)

Recommended: Agilent 2100 Bioanalyzer

Optional: Nanodrop Spectrophotometer 2000

PROCEDURE

Ligation of 3' adenylated adapter

Make sure to clean surfaces and instruments with RNase Zap and maintain RNase-free conditions throughout the protocol. While as few as 100 ng of total RNA is sufficient, we recommend starting with at least 1 µg of total RNA (one can also use the equivalent fraction of enriched for small RNAs if desired). We recommend verifying RNA quality using the Agilent Bioanalyzer RNA nano or pico chip and using samples of RIN value of 7 or above.

- 1| Dilute the starting RNA to 200 ng/µL in nuclease-free dH₂O, if possible.
- 2| Set up a ligation reaction in a 200 µl PCR tube.

Component	Volume (µl)	Final Concentration
200 ng/µlRNA in dH ₂ O	5	1 µg total
10× T4 RNA Ligase 2tr Buffer	1	1×
10 µM 3'rApp-adapter	1	10 pmoles total
100% DMSO	1	10%

- 3| Denature for 30 sec at 90°C, then for at least 30 sec at 4°C.
- 4| Add the following directly to the ligation reactions on ice:

Component	Volume (µl)	Final Concentration
RNase Inhibitor (40 U/µl)	0.5	2 U/µl
T4 RNA Ligase 2tr (200 U/µl)	1.5	30 U/µl

- 5| Incubate for 1 h at 22°C.

We recommend using Enzymatics buffer, as its composition gave us significantly higher yield than other commercially available T4 RNA ligase 2 truncated buffers.

Annealing of RT primer

- 6| Add the following directly to each reaction on ice:

Component	Volume (μl)	Final Concentration
10μM RT Primer	1	10 pmoles total

The final amount of RT primer must be at equimolar ratio (10 pmoles) with the starting amount of 3' rApp-adapter (10 pmoles) for each sample.

- 7| Incubate for 30 sec at 90°C, then for 5 min at 65°C, then for at least 30 sec at 4°C.

Ligation of 5' RNA adapter

- 8| Prepare the 5' RNA adapter by incubating ~5μl in a 200 ul PCR tube at 70°C for 2 min, then 4°C for at least 30 sec.

An excess of volume is prepared to account for evaporation and facilitate the pipetting of the proper volume at the next step.

- 9| Spin down the ligation mixture by centrifuging 10 sec at ~2000×g, room temperature, using a microcentrifuge and add the following reagents directly to it:

Component	Volume (μl)	Final Concentration
10 mM ATP	1.5	1μM
10 μM 5' RNA Adapter	1	10 pmoles total
T4 RNA Ligase 1 (20 U/μl)	1.5	2 U/μl

- 10| Incubate for 1 h at 20°C.

Reverse transcription of captured MicroRNAs

The previous steps result in a reaction volume of 15 μl. Only 5μl is used in the subsequent RT-PCR step, and so the remaining can be stored (−80°C) as a backup (highly recommended) for two more runs. However, the rest of the protocol below can be scaled up 3 times and the full 15μl may be processed at once if you need to achieve higher yield (for example, when starting with lower amounts of RNA).

- 11| Prepare the following reaction in a 200 ul PCR tube:

Component	Volume (μl)	Final Concentration
Ligated miRNAs	5	-
5× First strand buffer	2	1×
12.5mM dNTP mix	0.5	625 μM
100mM DTT	1	10 mM

Component	Volume (μ l)	Final Concentration
RNase Inhibitor (40 U/ μ l)	0.5	2 U/ μ l
Superscript III (200 U/ μ l)	1	20 U/ μ l

- 12| Incubate for 30 min at 48°C.

As the RT primer was annealed earlier, do not denature or conduct an annealing cycle at this stage but go directly to the reverse transcriptase incubation temperature (48°C, as shown above).

Limited PCR Amplification

- 13| Prepare the PCR reaction in a 200 μ l PCR tube:

Component	Volume (μ l)	Final concentration
dH ₂ O	27	To 50 μ l total
Reverse Transcribed-miRNAs	10	-
5 \times HF buffer	10	1 \times
25mM dNTPs	0.5	0.5 mM
25 μ M BCmiRNA_PCR1	1	0.5 μ M
25 μ M BCmiRNA_PCR2_BC*	1	0.5 μ M
Phusion DNA pol. (2 U/ μ l)	0.5	1 U

BCmiRNA_PCR2_BC* refers to the bar-coded primer, where for each unique starting RNA sample a unique bar-code primer needs to be used (see Table 1). To limit bar-code / samples aerosol contamination, it is recommended to only open and close one tube of primer at a time.

- 14| Cycle the PCR reaction as follows in a thermal cycler:
1. 98°C for 30 sec
 2. 98°C for 10 seconds
 3. 60°C for 20 seconds
 4. 72°C for 20 seconds go to step 2, 11 more times
 5. 72°C for 5 min
 6. 4°C pause

The number of cycles can be varied according to the amount of microRNA present in the starting sample. In our hands, a total of 12 cycles generally results in the best yield while limiting unnecessary cycling. We recommend not exceeding 15 cycles as this will increase non-specific background amplification and reduce optimal yield of the desired products. Instead, additional starting RNA should be prepared in parallel and combined at the final stage to increase yield.

PCR Clean-Up with AMPure XP beads

- 15| Transfer PCR reactions to a new 1.5 mL tube.
- 16| Vigorously mix the AMPure XP beads and then add 90 μ L of beads to each 50 μ L PCR reaction. Pipet the beads slowly.
- 17| Vortex for 30 seconds, and then incubate on bench for 5 min.
- 18| Quick spin, and place on the magnetic rack for 5 min.
- 19| With the tubes still on the magnet, aspirate and discard the liquid from the reaction.
- 20| With the tubes still on the magnet, add 400 μ L 70% EtOH to the beads and leave for 30 sec. Then discard the 70% EtOH.
- 21| Repeat the previous step for a second wash.
- 22| Quick spin on a microfuge to collect last traces of EtOH. Put tubes back on magnet and remove any last drops of EtOH at the bottom of the tube.
- 23| Leave the tube open to air dry for 2 min.
- 24| Remove the tube from the magnet and add 45 μ L of nuclease-free water.
- 25| Vortex for 30 sec.
- 26| Place the tubes on the magnet and leave for 1 min.
- 27| With the tubes on the magnet, transfer 42 μ L to new 1.5 mL tubes.

Gel extraction of microRNA library

- 28| Prepare a 2% Agarose Gel EX following the manufacturer's protocol.
- 29| Dilute the 25 bp and 100 bp ladders 1:20 in water and load 20 μ L of each.
- 30| Split each microRNA library prepared above across 2 lanes by loading 20 μ L per well.
- 31| Run the gel for 14 minutes on the Invitrogen I-Base using the 2% E-Gel settings.
- 32| When the run is complete, take a picture of the gel.

The migration patterns of DNA on E-gels are affected by the total amount and salts present in the loaded sample, and sometime one may observe a shift in migration of the expected product in relation to the ladder.
- 33| Pry open the E-Gel by cracking open each side.
- 34| Using a clean razor blade for each sample, cut between 125 and 175 bp to capture the two dominant miRNA bands.
- 35| Gel extract using the Mini Elute Qiagen Gel-Extraction Kit following the manufacturer's protocol, conducting the final elution in 15 μ L of dH₂O

Melt the gel bands at 37°C instead of the recommended 55 °C. The MinElute columns have a tendency to trap residual EtOH from the wash steps. To avoid

this issue, dry spin the column for 1 minute at maximum speed and then turn the column 180 degrees and repeat the spin for another 1 minute. Then transfer the column to a recovery tube and leave the column open for 3 min to air dry prior to adding the elution buffer.

Library QC and Mixing

- 36] The library can now be mixed at equimolar concentration, prior to submission for sequencing. We strongly recommend analyzing the quality and concentration of each final library using the Agilent Bioanalyzer DNA high sensitivity chip in order to combine the different libraries at equimolar ratios into a single multiplexed library. Although less accurate, a Nanodrop spectrophotometer would also work to a decent degree for this step if an Agilent Bioanalyzer is inaccessible. For high throughput project with high amount of samples, the Bioanalyzer can be used to combine the libraries prior to gel extraction.
- 37] Submit library for sequencing using standard Illumina genomic primer or Truseq primer with 75 bp single-read. Alternatively, a custom indexing primer can also be used if desired (Table 1).

BASIC PROTOCOL 3

BIOINFORMATICS ANALYSIS OF MULTIPLEXED miRNA LIBRARY SEQUENCING DATA

Analysis pipeline presented here aims at measuring miRNA expression and assessing differential expression between samples. It also provides an efficient way to construct comprehensive miRNA expression profiles. Here, we use annotated mature miRNAs deposited regularly in miRBase as a reference. (<http://www.mirbase.org>). In our analysis pipeline, first, the 3' adapter sequences are removed and then the reads are assigned to separate output files based on different barcodes used. Next, miRNAs are identified by aligning the reads to miRBase (version18) (Kozomara A and Griffiths-Jones S 2011) and the reads are tallied to generate total counts for each miRNA. Finally, statistical significance (p-value) between 2 or more samples is calculated to generate differential expression profiles. The analysis programs are written in Perl and the steps described can be performed through the command line. The workflow of analysis pipeline is shown in Figure 2.

Materials

Hardware pre-requisites:

Linux, Unix or Mac OS X installed computer or Cygwin with Windows.

A computer cluster may not be necessary although recommended.

No significant memory requirements.

Software pre-requisites:

Perl v5.10.0

BioPerl module for SAGE comparison

Text-editor program (e.g. TextWrangler)

Analysis package:

Analysis programs are all written in Perl and can be found at http://seidman.med.harvard.edu/fgs/software/mirna_soft/mirna_soft.tar.gz

Files needed:

Sequence file in FASTQ format (e.g. reads.fastq)

A FASTQ file from a sequencing run includes 4 lines per read. The first line starts with “@” followed by a unique identifier, second line includes the sequencing read, third line may contain additional sequencing run information and fourth line includes the quality scores for each nucleotide in the sequencing read.

mature.fa.gz (<http://www.mirbase.org/ftp.shtml>)

reads.fastq is provided in the package for a test run.

1. Download and install the following programs:

- a. Perl: <http://www.perl.org>
- b. BioPerl module for SAGE comparison by algorithm described by Audic and Claverie (1997) <http://search.cpan.org/~scottzed/Bio-SAGE-Comparison-1.00/> Once downloaded, unpack the module. Type:

```
tar xvfz Bio-SAGE-Comparison-1.00.tar.gz
```

Then, add the module’s path to the PERL library environment variable. Use PERL5LIB for Perl 5. Type:

```
export
PERL5LIB=/location_to_downloaded_module/Bio-SAGE-
Comparison-1.00/lib
```

- c. The analysis package: http://seidman.med.harvard.edu/fgs/software/mirna_soft/mirna_soft.tar.gz which includes the following scripts: separatebarcodes.pl, countreads_mirna.pl and cmp_mirna.pl. To unpack the modules, type:

```
tar xvfz mirna_soft.tar.gz
```

2. Download the latest mature.fa.gz from <http://www.mirbase.org/ftp.shtml> containing all the mature miRNA sequences.
3. Remove the 3’ adapter sequence and separate reads into individual files by barcode. Type:

```
perl separatebarcodes.pl reads.fastq
```

A section of `separatebarcodes.pl` is shown below and MUST be modified according to the samples/barcodes/adapters used by using a text editor program such as TextWrangler (Mac) or equivalent.

```
my $adaptor="ACGGGCTAATATTTATCGGTGGAGC"; ## specific adapter sequence
my %barcodes = (
  CGTGAT => "Sample1", #BC1 ## these are associating each barcode with
  ACATCG => "Sample2", #BC2 ## corresponding sample
  GCCTAA => "Sample3", #BC3
  TGGTCA => "Sample4", #BC4
  CACTGT => "Sample5", #BC5
  ATTGGC => "Sample6", #BC6
  GATCTG => "Sample7", #BC7
  TCAAGT => "Sample8", #BC8
);
```

This perl script is designed to analyze the sequencing reads obtained from multiplexed miRNA libraries constructed using the adapters and PCR amplification primers shown in Table 1. Usually, the first part in the sequence corresponds to the miRNA sequence. The miRNA sequence is then followed by the adapter sequence “ACGGGCTAATATTTATCGGTGGAGC”, which is followed by a 6-nucleotide barcode (Table 1, **bold-underlined**). First, the script opens the sequencing file `reads.fastq`, removes the 3’ adapter, identifies each sample by barcode and creates one output file per sample based on the barcodes. The user is advised to confirm the output file by ‘less’ or ‘more’ command or by opening with any text-editor program. The output file should look like `SampleName.fq`. If Illumina TruSeq small RNA adapters/primers are used, the user can easily modify the script to reflect the appropriate adapter and barcode sequences.

4. Uncompress `mature.fa.gz` downloaded from miRBase. Type:

```
gunzip mature.fa.gz
```

Open the file `mature.fa` and select the entries corresponding to the appropriate species (e.g. *mus musculus* for mouse). Save as `species_mature.fa` in a separate file to use in the next step. Alternatively, from the command line type (for mouse):

```
grep -Al musculus mature.fa > mouse_mature.fa
```

Open and verify final file.

- Match the reads to reference miRBase and tally the counts of reads per microRNA to generate an expression profile per sample. Here, output files generated in step 3 are used as input. Type:

```
perl countreads_mirna.pl species_mature.fa
OUTPUT_STEP3
```

This perl script is applied to each output file from step 3 (OUTPUT_STEP3, e.g. `SampleName.fq`) after processing with the first script. First, the sequences shorter than 17 nucleotides are eliminated. Then, the script aligns the sequence to reference mature miRNA sequences, downloaded from miRBase. There are 2 output files created per sample: The first file `SampleName.m` contains information for aligned miRNAs: “miRNA name”, “Sequence of the miRNA”, “Unique matches”, “Multi-matches”. “Unique matches” are reads matching specifically to that miRNA out of the total list. “Multi-matches” includes reads with multiple reference miRNA sequences and are not subsequently used. To maximize read counts and incorporate reads from putative isomiRs, an unmatched full length read is trimmed at the ends before another matching attempt. The second file `SampleName.unm` contains the reads that did not match to any miRNA in the miRBase. These unmatched reads often contain small inserts that may not qualify as miRNAs such as fragmented RNAs or bad quality reads, and any unannotated novel miRNAs. To identify novel miRNAs, `SampleName.unm` can be matched to other species’ miRNA sequences using the method described here, and they can also be aligned to the reference genome depending on the uniqueness and length of the sequence.

The output file can be confirmed by using ‘less’ or ‘more’ commands or can be opened by any text-editor program.

- Normalize and compare expression profiles of miRNAs. Type:

```
perl cmp_mirna.pl SampleName_1.m SampleName_2.m
...SampleName_8.m
```

This script can be applied to two or more output files from step 5 to make a comparison. The read counts are normalized to counts per million by dividing the total read counts of a miRNA by the total read counts of the sample and multiplying this number by 10^6 . The script uses a Bayesian comparison based on Audic and Claverie (1997) to calculate the p-value. The output file for two samples will include “miRNA ID”, “Tag Sequence”, “Normalized Counts for Sample1”, “Normalized Counts for Sample2”, “Raw reads for Sample 1”, “Raw reads for Sample2”, and “p-value”, while for more samples, it will include corresponding additional fields.

- Analyze the final output file.

We generally consider p-value = 0.01 as an appropriate cut-off given the list of miRNAs. However, more stringent cut-off may be used.

COMMENTARY

Background Information

Owing to the development of high-throughput methods to study miRNA expression profiles, there has been an exponential increase in the amount of data generated in the miRNA field over the last decade. Compared to microarray or qPCR-based miRNA expression profiling techniques, next generation sequencing technology offers several advantages, including high sensitivity to measure miRNA levels over a wide dynamic range, ability to identify novel miRNAs and to detect miRNA expression levels in species for which complete genomes are not yet available. In addition, next-generation sequencing is able to detect miRNAs that differ by just one nucleotide [Pritchard et al. 2012]. Last but not the least, next-generation sequencing allows multiplexing of samples by tagging libraries with barcodes during library preparation.

Even though different sequencing platforms require different protocols and adapters/primers, they typically follow similar steps. First step of miRNA library construction is to capture miRNAs using specific adapters. miRNAs have a 5' phosphate and a 3' hydroxyl group as a result of RNase III activity of Droscha and Dicer. To prevent self ligation and circularization of miRNAs during adapter ligation, pre-adenylated 3' adapters are used with truncated T4 RNA ligase 2 (which does not require ATP). Next, miRNAs are ligated to a 5' adapter, which contains the binding site for sequencing primer. The resulting adapter-captured miRNAs are reverse transcribed with a primer complementary to the 3' adapter to generate cDNA library. To provide enough yield for sequencing, cDNA library then needs to be PCR-amplified. Finally, the amplified library is gel extracted and quality-checked prior to sequencing. To sequence multiple libraries in one lane of a flow cell, each miRNA library can be barcoded by including a unique tag sequence as part of the adapter or PCR amplification primer during construction. However, we and others have reported previously that including barcodes in the adapter sequence creates significant ligation bias [Alon et al. 2011, Hafner et al. 2011], e.g. same biological sample tagged with 2 different barcodes show significant differences in miRNA expression levels, presumably as a result of bias by T4 RNA ligase-mediated ligation [Zhuang et al. 2012, Jayaprakash et al. 2011]. In the current protocol, we add barcode to each library during PCR amplification step, which helps avoid any ligation bias and therefore, provides a significant advantage (Alon et al 2011, Vigneault et al. 2012 CPHG Unit 11.12.1–10). With this method, up to 12 libraries can be prepared in parallel and sequenced in one lane of Illumina Genome Analyzer /HiSeq.

Several bioinformatics tools have been developed to analyze miRNA sequencing reads (Li et al. 2012) and they generally follow similar steps with some variations. Typically, after the first step of 3' adapter removal, the sequences are aligned against a reference sequence to identify annotated miRNAs as well as novel ones. One of the important parameters for bioinformatics analysis of miRNA sequencing reads is the alignment criteria. Several groups have identified isomiRs (variants from the reference) for a given mature miRNA and reported that the most abundant isomiR may not always be same as the mature miRNA

[Morin et al. 2008, Wyman et al. 2011, Lee et al. 2010]. IsomiRs, which differ mainly at the 3' end, and to a lesser extent at the 5' end, may also include nucleotide substitutions or 3' non-template addition of nucleotides. While some studies have quantified the most abundant isomiR as the most representative of the miRNA level [Morin et al. 2008], others have quantified miRNAs with 100% length and sequence match with mature miRNA [Wang et al. 2009]. In a recent study, expression levels of isomiRs were reported to highly correlate with that of annotated mature miRNAs [Cloonan et al. 2011]. While isomiRs may well be biologically relevant, abundance of some isomiRs above that of mature miRNA may also result from T4 RNA ligase bias in capturing small RNAs [Hafner et al. 2011, Jayaprakash et al. 2011, Zhuang et al. 2012]. Since the biological significance of isomiRs is largely unknown and their biogenesis is not understood, currently, our analysis pipeline counts miRNAs with exact matches as well as those with some variation to the annotated mature miRNA in miRBase (version 18). Our analysis pipeline also keeps a separate file of reads not aligning to miRBase, and these reads can be mapped against the reference genome to identify putative novel miRNAs.

Critical Parameters and Troubleshooting

One of the most critical parameters for a high quality library is the quality of starting RNA. Therefore, it is important to maintain an RNase-free workspace throughout RNA isolation (Basic Protocol 1) and library construction protocol (Basic Protocol 2: steps 1–12). Another critical parameter is to avoid contaminating pre-amplified library with amplified library, as even the slightest contamination will have detrimental effects for downstream analysis. Therefore, we recommend designating separate equipment, reagents and bench-space for all the steps prior to PCR-amplification of libraries (Basic Protocol 2: Step 14) as “pre-PCR” and for all the steps after PCR-amplification as “post-PCR”. We strongly recommend including a “no-ligase” control reaction (where T4 RNA ligase is omitted) and processing in parallel to the experimental samples to assess for any possible contamination, which will be evident after PCR amplification step. If an amplified library is detected in the control, the prepared libraries should be discarded and new reagents should be used to start over. Therefore, to prevent wasting of reagents, we recommend that oligos, and all other library construction reagents are stored in small aliquots.

During library construction, 3' and 5' adapters are used in excess to ensure efficient capture of miRNAs and to prevent their self-circularization. As a result, the adapters can ligate to each other and the undesired adapter-adapter dimer band is often observed in the agarose gels during library preparations. In this protocol, annealing the RT primer to the 3' adapter-ligated product, prior to ligation of 5' adapter, can significantly reduce the adapter dimer formation [as also shown by NEB, cat. no. E6120, and Bioscientific, cat.no. 5132-02] and therefore, we do not expect to see the dimer band (114bp) on an agarose gel. However, if an adapter dimer band is observed, one or two rounds of denaturing PAGE extractions should be performed (as described in Alon et al. 2011) to remove the adapter dimer and to prevent unnecessary sequencing of this undesired ligation product. Denaturing gel extraction ensures that any dimer fraction that may have annealed to the full-length library fragments is removed.

Another critical parameter is to quantify the final multiplexed library yield (Basic Protocol 2: step 36) right before submission, as low yield library may fail to cluster. If low yield libraries are obtained, we recommend starting with either more RNA, or preparing multiple libraries from the same RNA in parallel. In addition, PCR amplification cycles can be increased (but not more than 15, as this may result in low complexity library).

The multiplexed miRNA libraries prepared according to our protocol can be sequenced on Illumina Genome Analyzer II or HiSeq using either the standard Illumina primer for genomic libraries or primer for Truseq. It is critical to perform a >75bp single-read sequencing to ensure that insert miRNA (<30bp), 3' adapter (25bp) and barcode (6bp) are all sequenced. Alternatively, the libraries can be sequenced using a custom indexing primer (Table 1).

For bioinformatics analysis of sequencing reads, we highly recommend that for the first-time analysis, the user runs each script successively to generate the final output file. This will ensure that each module is running successfully, and if there is a problem, it will be easier to identify. If an extremely high-throughput is desired, the steps can be automated for more efficient processing of the sequencing data.

Even though we expect strong correlation ($r > 0.99$) in technical replicates, in order to minimize differences due to technical variation, we recommend that the libraries that are to be directly compared, be handled and prepared in parallel. It is also expected that the biological replicates show a strong correlation; however, if they do not, this can be due to several factors including poor RNA quality (we recommend $RIN > 8$), low complexity library (we recommend increasing the library yield before the PCR amplification step by starting with more RNA or processing multiple libraries from the same RNA and keep PCR amplification to no more than 15 cycles) or contamination with another library. We have also observed that there is a strong correlation between libraries prepared from “pooled” RNA samples (pooling RNA from 3 biological replicates) and average of 3 libraries prepared separately from RNA of 3 biological replicates. We believe that “pooled” RNA approach can help reduce the biological variation, allow user to pool RNA when sources are limited and also reduce the cost of sequencing.

As a first follow-up experiment, we highly recommend that the sequencing data is confirmed by quantitative RT-PCR. There are several commercially available qPCR kits to quantify miRNAs, and in our hands, Applied Biosystems' stem-loop qPCR primers (Taqman) work well.

Anticipated Results

Technical and/or biological replicates are expected to have a strong correlation ($r > 0.99$) with sequence counts ranging from 10 to $100000 <$ counts per million. We advise ignoring miRNAs with less than 10 counts per million in all the groups being compared, since they can be due to sequencing errors.

Time Considerations

Total RNA isolation can be completed in under 1 hour for few to several samples. The library construction protocol of up to 12 samples can be completed in one day. The bioinformatics analysis can be completed within one to several hours depending on the number of total sequences.

Acknowledgments

FUNDING

This work is supported by the Fondation LeDucq, NIH U54 Syscode PL1 HL092552 and NHLBI 1U01HL098166 (JG Seidman) and Canadian Institutes of Health Research and Ragon Institute Fellowship (F Vigneault).

REFERENCES

- Audic S, Claverie JM. The significance of digital gene expression profiles. *Genome Res.* 1997 Oct; 7(10):986–995. [PubMed: 9331369]
- Alon S, Vigneault F, Eminaga S, Christodoulou DC, Seidman JG, Church GM, Eisenberg E. Barcoding bias in high-throughput multiplex sequencing of miRNA. *Genome Res.* 2011 Sep; 21(9): 1506–1511. [PubMed: 21750102]
- Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, Barbacioru C, Steptoe AL, Martin HC, Nourbakhsh E, Krishnan K, Gardiner B, Wang X, Nones K, Steen JA, Matigian NA, Wood DL, Kassahn KS, Waddell N, Shepherd J, Lee C, Ichikawa J, McKernan K, Bramlett K, Kuersten S, Grimmond SM. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.* 2011 Dec 30.12(12):R126. [PubMed: 22208850]
- Hafner M, Renwick N, Brown M, Mihailovi A, Holoch D, Lin C, Pena JT, Nusbaum JD, Morozov P, Ludwig J, Ojo T, Luo S, Schroth G, Tuschl T. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA.* 2011 Sep; 17(9):1697–1712. [PubMed: 21775473]
- Jayaprakash AD, Jabado O, Brown BD, Sachidanandam R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* 2011 Nov. 39(21):e141. Epub 2011 Sep 2. [PubMed: 21890899]
- Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 2011 Jan; 39(Database issue):D152–D157. [PubMed: 21037258]
- Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, Wang K. Complexity of the microRNA repertoire revealed by next-generation sequencing. *RNA.* 2010 Nov; 16(11):2170–2180. [PubMed: 20876832]
- Li Y, Zhang Z, Liu F, Vongsangnak W, Jing Q, Shen B. Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis. *Nucleic Acids Res.* 2012 May 1; 40(10):4298–4305. [PubMed: 22287634]
- Mendell JT, Olson EN. MicroRNAs in stress signaling and human disease. *Cell.* 2012 Mar 16; 148(6): 1172–1187. Review. [PubMed: 22424228]
- Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* 2008 Apr; 18(4):610–621. [PubMed: 18285502]
- Motameny S, Wolters S, Nurnberg P, Schumacher B. Next Generation Sequencing of miRNAs – Strategies, Resources and Methods. *Genes.* 2010; 1(1):70–84. [PubMed: 24710011]
- Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. *Nat Rev Genet.* 2012 Apr 18; 13(5):358–369. [PubMed: 22510765]
- Vigneault F, Sismour AM, Church GM. Efficient microRNA capture and bar-coding via enzymatic oligonucleotide adenylation. *Nat Methods.* 2008 Sep; 5(9):777–779. [PubMed: 19160512]

- Vigneault F, Ter-Ovanesyan D, Alon S, Eminaga S, C Christodoulou D, Seidman JG, Eisenberg E, M Church G. High-throughput multiplex sequencing of miRNA. *Curr Protoc Hum Genet*. 2012 Apr; 11(Unit 11.12):1–10.
- Wang WC, Lin FM, Chang WC, Lin KY, Huang HD, Lin NS. miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics*. 2009 Oct 12.10:328. [PubMed: 19821977]
- Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, Krouse MA, Webster PJ, Tewari M. Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res*. 2011 Sep; 21(9):1450–1461. [PubMed: 21813625]
- Zhuang F, Fuchs RT, Sun Z, Zheng Y, Robb GB. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res*. 2012 Apr.40(7)

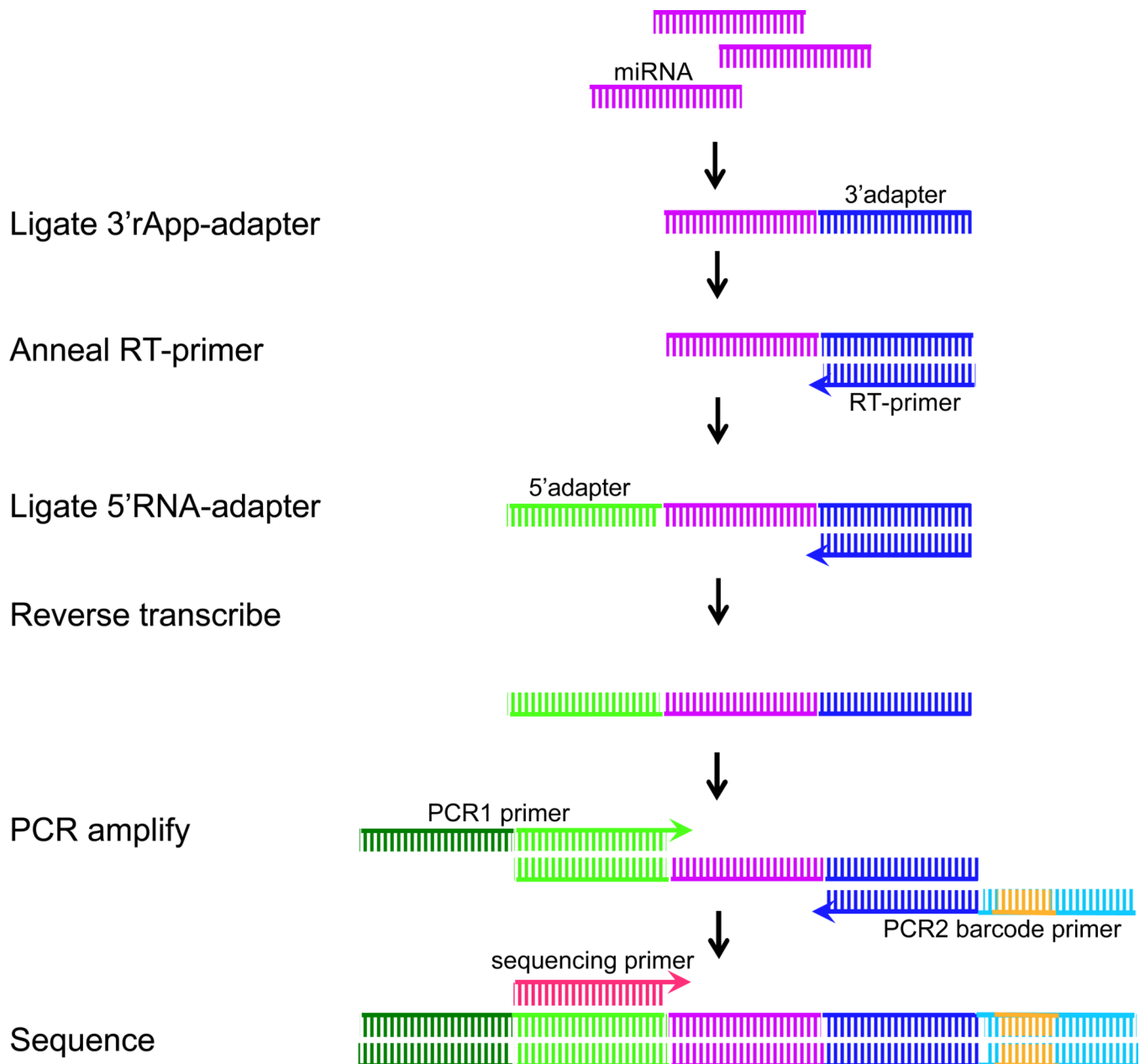


Figure 1. Construction of multiplex miRNA libraries. Small RNAs are first captured by ligating a 3rApp adapter, annealing an RT primer, and then ligating a 5 RNA adapter. The resulting ligation product is reverse transcribed to obtain cDNA, which is amplified using a PCR1 forward primer and PCR2 barcoded reverse primers to generate an ~135-bp product containing miRNAs (~22 bp). The final product is purified prior to sequencing. Adapted from Vigneault et al. (2012).

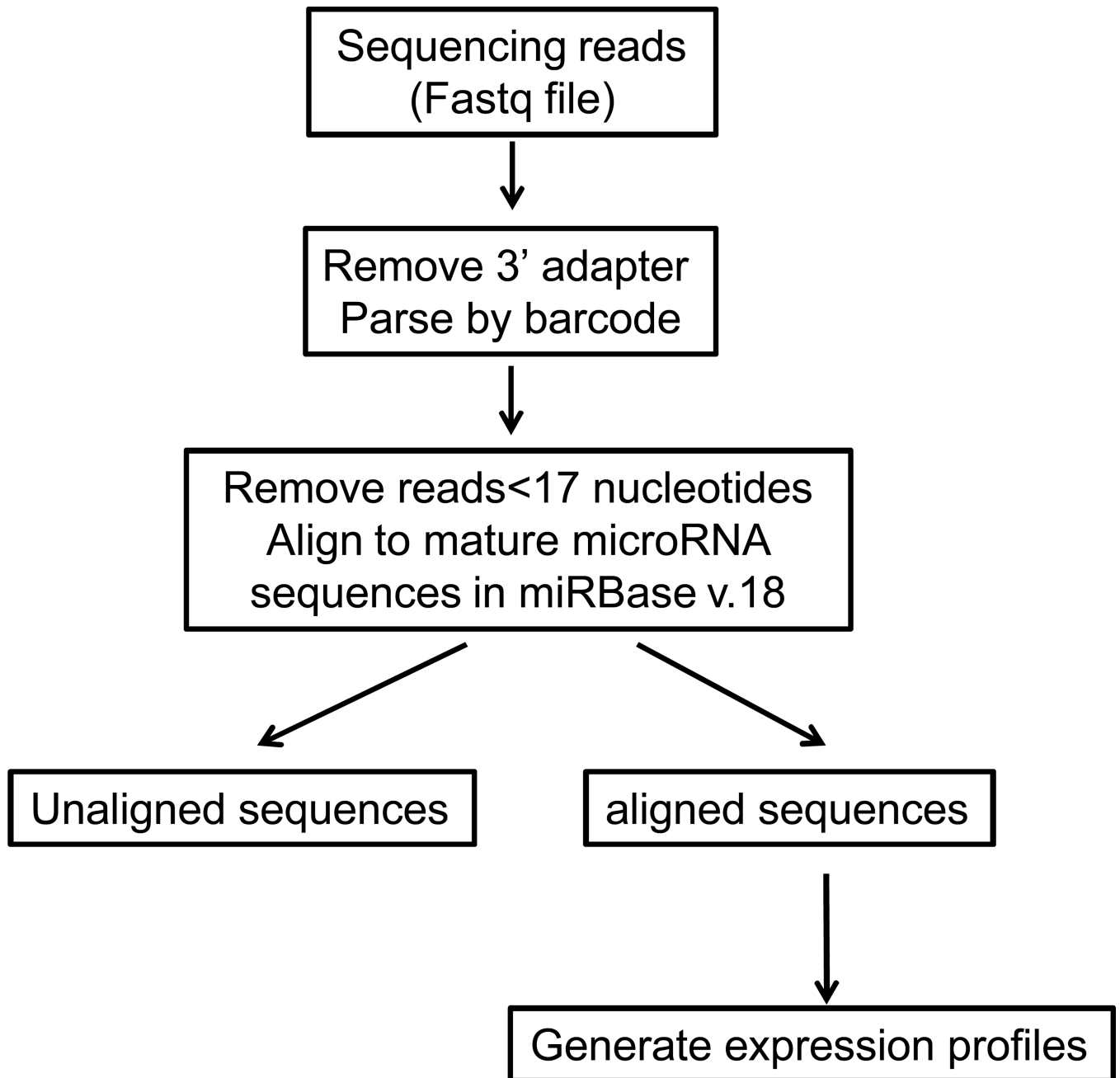


Figure 2.

Analysis pipeline. Bioinformatics analysis involves removal of the 3' adapter from the reads and parsing of reads by barcode. Reads with < 17 nt are then removed, and the remaining reads are aligned to mature miRNA sequences in miRBase. Finally, the total number of matches to each mature miRNA sequence is used to generate expression profiles.

Table 4.17.1Oligos for Multiplexed miRNA Library Preparation for Illumina Sequencing^a

Oligo name ^b	Sequence (5'-3') ^c
BCPCR_3'rAppadapter	/5rApp/ACGGGCTAATATTTATCGGTGG/3SpC3/
BCPCR_5'RNA-adapter	rUrCrCrCrUrArCrArCrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrC
BCPCR RT primer	GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
BCPCR_PCR2-BC1	CAAGCAGAAGACGGCATAACGAGAT CGTGAT GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC2	CAAGCAGAAGACGGCATAACGAGAT ACATCGG GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC3	CAAGCAGAAGACGGCATAACGAGAT GCCTAAG CTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC4	CAAGCAGAAGACGGCATAACGAGAT TGGTCA GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC5	CAAGCAGAAGACGGCATAACGAGAT CACTGT GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC6	CAAGCAGAAGACGGCATAACGAGAT ATTGGC GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC7	CAAGCAGAAGACGGCATAACGAGAT GATCTG GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC8	CAAGCAGAAGACGGCATAACGAGAT TCAAGT GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC9	CAAGCAGAAGACGGCATAACGAGAT CTGATC GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC10	CAAGCAGAAGACGGCATAACGAGAT AAGCTA GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC11	CAAGCAGAAGACGGCATAACGAGAT GTAGCC GCTCCACCGATAAAATATTAGCCCGT
BCPCR_PCR2-BC12	CAAGCAGAAGACGGCATAACGAGAT TACAAG GCTCCACCGATAAAATATTAGCCCGT
BC_Custom Indexing (optional)	ACGGGCTAATATTTATCGGTGGAGC

^a All oligonucleotides can be ordered through Integrated DNA Technologies (<http://www.idtdna.com>) and should be ordered with HPLC purification. For a less costly option, the adenylated adapter can be made as described by Vigneault et al. (2008).

^b Oligos from Alon et al. (2011).

^c Bold, underlined bases represent 6-nt barcodes.