Published in final edited form as:

Methods Mol Biol. 2015 ; 1299: 27–49. doi:10.1007/978-1-4939-2572-8_3.

Deep sequencing of cardiac microRNA-mRNA interactomes in clinical and experimental cardiomyopathy

Scot J. Matkovich and Gerald W. Dorn II

Center for Pharmacogenomics, Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA

Summary

MicroRNAs are a family of short (~21 nucleotide) noncoding RNAs that serve key roles in cellular growth and differentiation and the response of the heart to stress stimuli. As the sequence-specific recognition element of RNA-induced silencing complexes (RISCs), microRNAs bind mRNAs and prevent their translation via mechanisms that may include transcript degradation and/or prevention of ribosome binding. Short microRNA sequences and the ability of microRNAs to bind to mRNA sites having only partial/imperfect sequence complementarity complicates purely computational analyses of microRNA-mRNA interactomes. Furthermore, computational microRNA target prediction programs typically ignore biological context, and therefore the principal determinants of microRNA-mRNA binding: the presence and quantity of each. To address these deficiencies we describe an empirical method, developed via studies of stressed and failing hearts, to determine disease-induced changes in microRNAs, mRNAs, and the mRNAs targeted to the RISC, without cross-linking mRNAs to RISC proteins. Deep sequencing methods are used to determine RNA abundances, delivering unbiased, quantitative RNA data limited only by their annotation in the genome of interest. We describe the laboratory bench steps required to perform these experiments, experimental design strategies to achieve an appropriate number of sequencing reads per biological replicate, and computer-based processing tools and procedures to convert large raw sequencing data files into gene expression measures useful for differential expression analyses.

Keywords

microRNA; mRNA; Argonaute; Ago2; RNA-induced silencing complex (RISC); immunoprecipitation; deep sequencing; read alignment; gene annotation; differential expression

1. Introduction

Cellular homeostatic maintenance as well as responses to stress stimuli involve RNA transcription, not only of coding mRNAs but of regulatory noncoding RNAs, whose many potential functions are gradually coming to light. MicroRNAs are 21–22 nucleotide molecules in their mature, processed forms and comprise but one subfamily of noncoding

---

Corresponding author: Scot J. Matkovich, Ph.D, Assistant Professor, Department of Internal Medicine and Center for Pharmacogenomics, 660 S. Euclid Ave., Campus Box 8220, St. Louis, MO 63110., Phone: 314-747-3455, Fax: 314-362-8844, smatkovi@dom.wustl.edu.

The authors declare that they have no conflicts of interest relating to this manuscript.

RNAs. However, their importance to mammalian cellular function has been widely demonstrated, with dramatic effects on phenotype resulting from deletion of the enzymatic machinery required to process microRNA precursor transcripts into their active forms [1–4]. The mechanism of action of microRNAs is straightforward, in principle: a microRNA is incorporated into a protein complex including Argonaute family members and TNRC6 / GW182 (the RNA-induced silencing complex, or RISC), and selects mRNAs with partially complementary sequences for degradation and/or translational suppression [5–7]. Difficulties in understanding which microRNAs are actively repressing individual mRNAs arise from the presence of ~1,000 different microRNAs in the mammalian genome [8], ~300 of which are detectably expressed in the heart [9], and which each could theoretically interact with a large number of different mRNAs (from the ~10,000 detectably expressed in the heart [10,11]) considering that only partial sequence complementarity is needed for effective regulation. The simultaneous alteration of a large number of both microRNAs and mRNAs in clinical [12–16] and experimental cardiomyopathies [11,14,17] means that it is difficult, if not impossible, to determine which mRNAs in such contexts are undergoing regulation that is largely dependent on microRNAs, without knowledge of the changes induced in direct microRNA-mRNA interactions.

Our approach to this challenging question involves measurement of three distinct RNA fractions that underlie the microRNA regulation of mRNAs and their ultimate effect on the gene expression profile of the heart; that is, a) the microRNAs themselves; b) the mRNAs captured at the RISC via interaction via microRNAs; and c) the global cohort of mRNAs. We use next-generation / deep-sequencing methods to obtain quantitative RNA abundance information that cannot be determined using qPCR techniques or microarrays, and which have superior dynamic range and sensitivity compared to these approaches [10,18]. MicroRNAs and mRNAs are monitored in the same heart or cardiomyocyte preparation, together with the RISC-associated mRNAs that are targeted by microRNAs. Comparison of microRNAs, RISC-associated mRNAs and translationally competent mRNAs between control and diseased hearts permits the characterization of cardiomyopathic alterations in the microRNA-mRNA interactome. We do not utilize the CLIP approach (cross-linking with immunoprecipitation) [19] popular in a number of *in vitro* studies of mRNA interactions with mRNA-binding proteins. This is mostly due to the technical problems inherent in performing UV- [19] or photoactivatable-analog-based [20,21] crosslinking in intact heart preparations (although UV crosslinking has been performed in young mouse brain [22,23]), but partly out of concern that the contribution of RNAs that are only weakly associated with the RISC to the microRNA-mRNA interactome may be overestimated. Nonetheless, such methods are of undeniable importance for identifying microRNA binding sites on target mRNAs [23]. Our methods for Ago2 immunoprecipitation without crosslinking and subsequent comparison of RISC-bound and global mRNA content are based on studies by Karginov *et al.* [24], who evaluated microRNA targets in 293T cells using recombinantly overexpressed Ago2 and microarrays for differential expression analysis; our studies have shown it is possible to gain sufficient Ago2 and bound RNA by immunoprecipitating the Ago2 endogenous to mouse hearts.

In this article, we describe in detail laboratory procedures for obtaining the necessary RNA fractions and preparing libraries for deep sequencing analysis. Additionally, we describe the

computer-based workflows required to align the resulting sequence data to annotated transcriptomes, perform gene-level quantitation, and assess differential expression. Data visualization tools such as principal components analysis (PCA) and heatmaps, helpful in observing important trends in highly multidimensional data sets such as RNA-sequencing experiments, are not covered, but the interested reader can become familiar with our application of these graphic formats in published studies of cardiac RNA- and RISC-sequencing [9,11,25,26]. Likewise, advanced analytical procedures that build upon deep sequence microRNA and mRNA results to designate specific microRNA-mRNA pairing events in experimental context are beyond the scope of this chapter, but have been described [9,11].

2. Materials

2.1 miR-Seq library preparation

1. Trizol (Invitrogen)
2. Illumina TruSeq small RNA preparation kit (catalog # RS-200-00xx, with the xx digits referring to the particular set of 12 indexing primers supplied). There are a large number of reagents and specialized consumables required for completion of the protocol that need to be purchased separately, and a list can be found in the manual supplied with the kit.

2.2 PolyA+ enrichment for mRNA-Seq library preparation

1. Invitrogen oligo(dT)₂₀ Dynabeads (Invitrogen #610-06) for polyA+ RNA selection. New England Biolabs also make an alternative magnetic bead purification system which we have used successfully ('Magnetic mRNA isolation kit', #S1550S).
2. Dynal magnet (Invitrogen) or other magnet suitable for pelleting magnetic beads in 1.5 mL microcentrifuge tubes.

2.3 Ago2 immunoprecipitation for RISC-Seq library preparation

1. Anti-mouse Ago2 monoclonal antibody (Wako Pure, clone #2D4)
2. Protein G-coupled Dynabeads (Invitrogen).
3. Lysis and immunoprecipitation: Ice-cold buffer solution comprised of 50 mM Tris.HCl, 5 mM EDTA, 5 mM EGTA, pH 7.5, with Roche Complete protease inhibitors, yeast tRNA (Invitrogen) (1 µg/µL), and SUPERnase-IN (Ambion) (1 U/µL); after initial homogenization, 0.5% Nonidet P-40 (Sigma IGEPAL CA-630).

2.4 RNA- & RISC-Seq library preparation

1. 5× fragmentation buffer, 50 mL: 1.21 g Tris base, 2.45 g potassium acetate, 1.61 g magnesium acetate tetrahydrate (200 mM Tris acetate, 500 mM potassium acetate, 150 mM magnesium acetate, DEPC H₂O). Adjust pH to 8.2 with 1 N acetic acid, and keep aliquots frozen at −20°C.

2. Zymo RNA Clean+Concentrator™-5 columns for RNA purification, catalog #R1015.
3. Invitrogen SuperScript III 1st-strand synthesis system, #18080-051 (50 reactions), which comprises SuperScript III reverse transcriptase, random hexamer solution, and buffer components including DTT and MgCl₂.
4. 5× 2nd-strand reaction buffer; 100 mM Tris.HCl pH 6.9, 450 mM KCl, 23 mM MgCl₂, 0.75 mM β-NAD⁺, 50 mM (NH₄)₂SO₄. Keep frozen aliquots at −20°C.
5. *E. coli* DNA ligase (New England Biolabs, catalog #M0205S) and DNA polymerase I (New England Biolabs, catalog #M0209S).
6. Qiagen QIAquick PCR purification kit (catalog #28104), MinElute kit (catalog #28004), gel purification kit (catalog #28704).
7. Adapter oligonucleotides for multiplexed Illumina library sequencing (Table 1).
8. Enzymes and reagents required for adapter ligation to cDNA fragments: End-It DNA End-Repair kit (Epicentre catalog #ER0720), Klenow 3′–5′ exo[−] (New England Biolabs M0212S), 1 mM dATP solution in aliquots, Promega LigaFast T4 DNA ligase and reaction buffer (catalog #M8221).
9. 2% low-melt agarose (e.g. Amresco Agarose II) to prepare gels for library size selection.
10. Indexing primers (Table 1) and post-size selection PCR reagent (Phusion High-Fidelity PCR master mix, New England Biolabs catalog #M0531S).
11. DNA-binding dye assay (such as the Invitrogen Qubit or PicoGreen assays)

2.5 Suggested software packages for data analysis

FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)

E-miR (http://www.lgtc.nl/services/general_services/software_tools/emir.php) [27]

miRdeep* (<http://www.australianprostatecentre.org/research/software/mirdeep-star>) [28]

Bowtie (<http://bowtie-bio.sourceforge.net/index.shtml>) [29]

Tophat (<http://tophat.cbcb.umd.edu/>) [30]

HTSeq (<http://www-huber.embl.de/users/anders/HTSeq/doc/index.html>). See especially the section on the included Python script *htseq-count*.

DESeq (<http://bioconductor.org/packages/release/bioc/html/DESeq.html>) [31]. Prior to publication of this article, an extensively revised version (DESeq2) has also been released (<http://www.bioconductor.org/packages/2.13/bioc/html/DESeq2.html>), but has not yet been evaluated in the Matkovich laboratory workflow.

3. Methods

3.1 Isolation of total RNA

Use Trizol following the manufacturer's directions precisely. Take the usual precautions for RNase-free benchtop work. Flash-frozen intact tissue should be homogenized (e.g. an Ika Ultra Turrax probe homogenizer or similar): keep tissue frozen (in liquid nitrogen), wash down homogenization probe with a few hundred μL of Trizol first, add e.g. 0.5 mL Trizol directly to tube as soon as it is removed from liquid nitrogen, and homogenize. For mouse heart apices, we use a total of 1 mL Trizol; the second aliquot of 0.5 mL is used to help wash tissue pieces out of the probe. This results in good extraction of the total RNA fraction. To enhance recovery of the small RNA fraction, increase the length of the isopropanol precipitation step at room temperature from the recommended 10 minutes to at least 30 minutes. Note 1. Resuspend the final precipitated RNA pellet in 50–100 μL DEPC-treated or nuclease-free H_2O and heat at 65°C for 5 minutes to ensure complete dissolution. Store RNA preparations at -80°C .

Quantitate RNA using a standard (cuvette-based) UV spectrometer or a NanoDrop-style instrument and expect high RNA purity; ($\text{OD}_{260} / \text{OD}_{280}$) 2.0. RNA integrity should also be analyzed; we find that electrophoresing 1 μg of total RNA on a 1% TBE-agarose gel is generally sufficient. Sharp 28S and 18S bands should be evident with a 28S:18S intensity ratio approaching 2.0. (Provided that the gel apparatus is treated beforehand with a suitable RNase inactivator such as RNaseZAP (Ambion) and that the gel and running buffer are prepared with DEPC-treated H_2O , there is no need to use a formaldehyde-agarose gel unless Northern blot transfer of the material is desired.) Another popular option, available in several laboratories or core facilities, is the use of an Agilent BioAnalyzer to size-separate and quantitate RNA species via capillary electrophoresis.

3.2 Preparation of small RNA (microRNA) sequencing libraries

To prepare small RNA (microRNA) sequencing libraries from total RNA aliquots, we use the exact reagents and protocol supplied by Illumina with their TruSeq small RNA preparation kits (catalog # RS-200-00xx, with the 'xx' digits referring to the particular set of indexing primers supplied; see below) and will not repeat the steps involved here. However, we note that we begin preparations using 1 μg of total cardiac RNA, as also recommended in Illumina's protocol, and have found that 14 cycles of PCR are optimal at the step that amplifies reverse-transcribed, adapter-ligated microRNA species prior to size selection on a 6% polyacrylamide gel.

The Illumina TruSeq system allows for individual sequencing libraries to be prepared with nucleotide 'indexes', such that multiple libraries from different samples can be loaded on a single sequencing lane and the sequencing reads so obtained can be demultiplexed. We further discuss considerations for the number of samples appropriate to load on a single lane (i.e. the number of sequencing reads appropriate for each individual sample) in section 3.7, for RNA- and RISC-sequencing libraries. However, our current experience is that 5–8 million aligned sequencing reads are more than enough for accurate microRNA quantitation and comparison. Typically we eliminate all microRNAs that represent less than one

millionth of the total sequencing reads obtained (~5) from later analytic steps, as we regard these species to be too rare to be of likely biological significance in our experiments. Whether or not other noncoding RNA species will be present in the sequencing libraries is highly dependent on the accuracy of gel-based size selection during the preparation process. For this reason, we also recommend combining and size-selecting multiple microRNA sequencing libraries for a particular analytical project on the same gel or lane.

3.3 Poly(A)+-RNA enrichment for mRNA-Seq

1. Use oligo(dT)-coupled magnetic beads to purify polyadenylated RNA from total RNA samples. We typically use the Dynabeads mRNA purification kit and associated Dynal magnet. Use no more than 100 μ L of bead suspension per 5 μ g total RNA sample. Note 2. The following steps briefly summarize the steps required for polyA+ RNA enrichment using the Invitrogen kit.
2. Follow the protocol given with the kit to perform the first round of isolation. Keep the supernatant (after initial binding of RNA to beads) on ice, and after washing beads, elute polyA+ RNA in 15 μ L 10 mM Tris, pH 8.2.
3. Re-prepare Dynabeads for binding by washing once in 1 $\times$ 'binding buffer'; it will be necessary to prepare a dilution of the 2 $\times$ buffer supplied with the kit.
4. Incubate the supernatant preserved from the initial binding step at 65°C, 2' and then chill on ice. Perform an additional round of RNA binding to beads.
5. Wash beads and elute further polyA+ RNA in another 15 μ L 10 mM Tris, and add to the RNA eluted in the first isolation. The aim is to maximize the amount of polyA+ RNA obtained more so maximally eliminating non-polyA+ species; we find that there is minimal carryover of rRNA in our final sequenced libraries.
6. Quantitate polyA+ RNA using a NanoDrop or low-volume spectrophotometer. Store purified RNA at -80°C.

3.4 Ago2 immunoprecipitation without crosslinking and RNA extraction for RISC-Seq

For the analysis of RISC-bound RNA, Ago2 immunoprecipitates are prepared from intact tissue or cells and RNA is directly extracted from these immunoprecipitates. Further rRNA depletion is not performed on the very small amount of RNA obtained from these procedures.

Flash-frozen mouse heart bases (approximately 2/3rd of the heart) are homogenized in 500 μ L ice-cold 50 mM Tris.HCl, 5 mM EDTA, 5 mM EGTA, pH 7.5, with Roche Complete protease inhibitors, and yeast tRNA and SUPERnase-IN added to final concentrations of 1 μ g/ μ L and 1 U/ μ L, respectively. Following homogenization, unbroken cellular material is removed at 100g, 5', 4°C and Nonidet P-40 is added to a final concentration of 0.5% (w/v) to solubilize proteins (15', 4°C, end-over-end rotating) followed by removal of insoluble material at 10000g, 15', 4°C. The supernatant is added to a previously prepared solution of 50 μ L protein G-coupled Dynabeads, to which 5 μ g anti-mouse Ago2 monoclonal antibody has been previously bound according to the supplied Dynabead protocol. Note 3. Following

1 h rotational incubation at 4°C, the beads are washed 3× with Dynabead washing buffer, transferring the suspension to a fresh tube for the last wash.

The last wash is performed in 200 µL of Dynabead washing buffer. Prior to pelleting beads and removing supernatant, extract 20 µL to a new tube and add an appropriate volume of Laemmli denaturing sample buffer to prepare a protein extract for SDS-PAGE. This will allow later SDS-PAGE analysis to ensure that Ago2 was successfully immunoprecipitated.

Pellet beads from the remaining suspension, remove and discard the supernatant, and add 500 µL of Trizol directly to the Dynabeads and retained immunoprecipitated material to extract RNA. Proceed with total RNA extraction as described under section 3.1, but see [Note 4](#).

While there are various rRNA reduction procedures and reagents available, e.g. RiboMinus™ from Invitrogen or Ribo-Zero™ from Epicentre, we have not yet tested whether these leave a sufficient yield of Ago2-RNA to allow successful library construction. Despite the availability of methodologies for RNA/cDNA amplification (e.g. SMARTer® from Clontech), we do not attempt to amplify the Ago2-immunoprecipitated RNA prior to fragmentation and cDNA synthesis. We noticed significant disruption to RNA abundances measured after the use of such a technique [32].

3.5 For RNA-Seq and RISC-Seq: RNA fragmentation and cDNA synthesis

This step chemically shears full-length RNA transcripts into 200–500 nt fragments, permitting uniform reverse transcription during subsequent stages. An important initial paper on RNA-Seq methodology by Mortazavi *et al.* [33] documents full-length RNA + oligo(dT) reverse transcription *vs* fragmented RNA + random primer reverse transcription, and concludes that fragmentation followed by random primer-based reverse transcription improves the coverage of exonic elements with reduction of 3' bias in final sequencing libraries.

1. Take 100–500 ng polyA+ RNA, or the Ago2-immunoprecipitated RNA; typically we use one-half of the RNA obtained, thus preserving an aliquot should there be a processing error in subsequent steps, and prepare a solution of 20 µL in '1× fragmentation buffer' (5× fragmentation buffer is diluted appropriately with DEPC H₂O).
2. Heat at 94°C, 2 min 30 s, then chill on ice. [Note 6](#).
3. Purify the fragmented RNA from the acetate buffer using e.g. Zymo RNA Clean +Concentrate™ columns, eluting in 8–9 µL for maximum input into the SuperScript III 1st-strand cDNA reaction mixture used at the next step.

All the following steps for cDNA synthesis and later sequencing library preparation are the same, regardless of whether the starting material is polyA+-selected mRNA or Ago2-immunoprecipitated RNA (Figure 1).

1st-strand cDNA—Use the SuperScript III 1st-strand synthesis system,. Each reaction can accommodate 8 µL of fragmented polyA⁺ RNA. Follow the directions supplied, but do not treat with supplied RNase H at the end of this procedure.

1. Combine 8 µL RNA, 1 µL 50 ng/µL random hexamers, 1 µL 10 mM dNTP mix.
2. Incubate at 65°C for 5 min, then place on ice for 1 min.
3. Add 10 µL ‘cDNA synthesis mix’, prepared as follows:

10× supplied RT buffer	2 µL
25 mM MgCl ₂	4 µL
0.1 M DTT	2 µL
RNaseOUT (40 U/µL)	1 µL
SuperScript III RT (200 U/µL)	1 µL

4. Incubate 10 min at 25°C, 50 min at 50°C, 5 min at 85°C, and chill on ice.

2nd-strand cDNA—To the 20 µL 1st-strand reaction, add on ice:

DEPC-treated H ₂ O	91 µL
5× 2 nd -strand reaction buffer	30 µL
10 mM dNTP mix	3 µL
E. coli DNA ligase (10 U/µL)	1 µL
E. coli DNA polymerase I (10 U/µL)	4 µL
E. coli RNase H (2 U/µL)	1 µL

Vortex gently to mix, and incubate for 2 h at 16°C. Do not allow the temperature to rise above 16°C. Purify and concentrate the synthesized cDNA via binding to DNA-binding and purification columns (such as the Qiagen QIAquick PCR purification kit), and elute the purified DNA in a volume of 34 µL of 10 mM Tris.HCl, pH 8.5, for optimal input into the first step of section 3.6. Note 6.

For RNase H, use the enzyme supplied by Invitrogen with SuperScript III reverse transcription kits, or purchase separately from New England Biolabs.

3.6 For RNA-Seq and RISC-Seq: adapter ligation to cDNA fragments

The protocol given below for generating sequencing libraries for the Illumina platform is based on the ‘ChIP sequencing protocol’ available at <http://bioinfo.mbb.yale.edu/array/resources.html> (Yale Center for Excellence in Genomic Science). The detailed steps we describe in section 3.6 and 3.7 are specific for the Illumina platform and will be different for different sequencing platforms, requiring consultation of the manufacturers’ documentation or other validated protocols.

3.6.1. End-repair—Use the ‘End-It DNA End Repair Kit’..

1. Combine and mix the following kit components in a microfuge tube:

Double-stranded cDNA	x μ L
10 $\times$ End-Repair Buffer	5 μ L
2.5 mM dNTP Mix	5 μ L
10 mM ATP	5 μ L
sterile water	y μ L
End-Repair Enzyme Mix	1 μ L
Total reaction volume:	50 μ L

2. Incubate at room temperature for 45 minutes.
3. Purify on one QIAquick column using the QIAquick PCR Purification Kit and protocol, eluting in 34 μ L of the supplied buffer EB (10 mM Tris.HCl, pH 8.5).

3.6.2. Addition of 'A' base to 3' ends—Use Klenow (3' $\rightarrow$ 5' exo⁻), *not* the large Klenow fragment typically used to fill in DNA ends.

1. Combine and mix the following components:

DNA from Step A	34 μ L
Klenow buffer (NEB Buffer 2)	5 μ L
1 mM dATP	10 μ L
Klenow fragment (3' to 5' exo ⁻)	1 μ L
Total reaction volume:	50 μ L

2. Incubate for 30 min at 37°C.
3. Purify on one QIAquick MinElute column using the Qiagen MinElute PCR Purification Kit and protocol. Elute in 10 μ L buffer EB (10 mM Tris.HCl, pH 8.5).

3.6.3. Ligation of common 'core' adapters to cDNA fragments—The adapter mix is a pre-annealed blend of Illumina-Index-AdapterA and 5'-phosphorylated Illumina-Index-AdapterB (see Table 1).

For optimal performance of the ligation step, a reasonable estimate of the cDNA fragment molarity should be calculated. Moles of starting cDNA = $x \text{ g} / (660 \text{ g mol}^{-1} \text{ bp}^{-1} * y \text{ bp})$, where bp = typical length of heat+acetate-sheared fragments (200–300 bp). For e.g. 1 pmol starting ds cDNA, plan to add 10 pmol annealed adapter mix (10-fold molar excess). The aim is to ensure complete ligation of cDNA fragments to core adapters, without introducing an unnecessarily large quantity of these adapters. The maximum yield of transcriptome-aligned sequencing reads can be compromised if unligated or self-ligated adapters are not completely removed.

1. Combine and mix the following components in a microfuge tube:

DNA from Step B	10 μ L
2 $\times$ DNA ligase buffer (Promega)	15 μ L
RNase, DNase-free water	2 μ L
Adapter mix	1 μ L
DNA ligase (Promega)	2 μ L
Total reaction volume:	30 μ L

2. Incubate for 15 min at room temperature.
3. Purify on one QIAquick MinElute column using the Qiagen MinElute PCR Purification Kit and protocol. Elute in 10 μ L buffer EB (10 mM Tris.HCl pH 8.5).

3.7 For RNA-Seq and RISC-Seq: size selection, PCR recovery, sample indexing and preparation for cluster formation

Considerations for sample indexing and the number of separate samples to be combined and sequenced on a single lane—For standard mRNA work, it is likely that sample ‘indexing’ (conceptually similar to DNA barcoding) will be desired so that 8–12 indexed samples can be combined into 1 sequencing lane of ~240 million raw 50 nucleotide single-end reads – this represents the current capacity of the Illumina HiSeq 2000-family sequencers available at our core facility. Indexes are read separately and do not detract from the 50 nucleotides available for sequence alignment. The nature of RISC-Seq analysis involves a large amount of rRNA is present in the Ago2 immunoprecipitate, which dilutes the reads useful for determining the abundance of bound mRNAs. For this reason we submit approximately half of the number of RISC-Seq samples on a single lane in comparison to what we submit for RNA-Seq on a single lane; in the final analysis, the number of sequencing reads that map to mRNA species is largely similar.

Some examples of indexing primers are given in Table 1. Particular core sequencing facilities may recommend particular 7-nucleotide indexes over others. Provided that the nucleotides are chosen are varied, without >3 of the same nucleotide occurring in series, we have observed no particular bias or difficulty in recovering sequence reads resulting from the use of one index sequence compared to another.

A common question in work of this nature relates to the number of reads that are appropriate in order to perform accurate quantitation of the starting RNAs and consequently, to perform differential expression analyses of a suitable power. This is a complex and user-dependent issue, and will partly depend on the rarity of genes of interest, especially if analysis of splice variants is required. Some recommendations for appropriate sequencing depth can be found in the following references: [34,35]. For analysis of cardiac global mRNA, we allocate 20 million raw sequencing reads to each library and typically find that 75% of these can be aligned to the transcriptome. For RISC-bound RNA (including unwanted rRNA, as described above) 40 million sequencing reads are allocated to each library, with an alignment to the transcriptome of 40–50%. Individual RNAs with less than 10 aligned reads are typically eliminated from downstream analyses.

3.7.1. Size selection of sequencing libraries via agarose gel purification

1. Gel-purify the DNA (using a 2% low-melting agarose gel) by cutting a gel slice that does NOT include any DNA from a potential adapter-adapter (self-ligated) band migrating at ~120 bp. The ligation product may not be readily visible at this stage if the cDNA input was small (0.5 µg or less), and will appear as a smear of fragments of different sizes in any case. Isolate cDNA in the 150–300 bp range. If purifying multiple libraries on a single gel, be very careful to avoid cross-contamination by leaving at least two empty lanes between each sample (and preferably three to four).
2. Purify the DNA from the agarose slice using a Qiagen Gel Extraction Kit. Note 8. Elute in 30 µL EB.

3.7.2 Recovery and amplification of size-selected libraries, with simultaneous addition of index sequences—Illumina Index PCR 1.0

Illumina Index PCR 2.0

Illumina Index reverse primers

As for the ‘core’ adapters described in section 3.6C, sequences are available in Table 1.

Components for PCR:

DNA from Step A	10 µL

(This will retain 20 µL of the gel-purified material from step A, should there be a future need to add a different index sequence, or if optimization of the number of PCR cycles is required. See Note 8.)

Phusion HF PCR master mix	25 µL
Index PCR 1.0, 25 µM	1 µL
Index PCR 2.0, 0.5 µM	1 µL
PCR index-specific reverse primer, 25 µM	1 µL
Nuclease-free H ₂ O	12 µL
Total reaction volume:	50 µL

Amplify using the following PCR protocol:

30 sec at 98°C

[10 sec at 98°C, 30 sec at 65°C, 30 sec at 72 °C], 12 cycles total for mRNA-Seq, 16 cycles total for RISC-Seq. Note 8.

5 min at 72°C

Hold at 4°C

3.7.3 Quality check of final library preparations, and preparation for

sequencing—Purify the amplified product on a QIAquick PCR purification column, similarly to the preceding steps in library generation, and elute in 30 µL buffer EB. Run 3–5 µL on a 2% agarose gel (no need for low-melting agarose) to verify that the amplified fragment sizes are similar to what was excised from the purification gel in step A. Alternatively, an Agilent BioAnalyzer trace could be gained on the amplified material. It is advisable to electrophorese the gel sufficiently to allow good visualization of any material in the 100–120 bp range. Typically, detection of sharp bands in this range indicates the presence of PCR primer-dimers or PCR amplification of core adapters that were not sufficiently removed after the ligation step.

If these low molecular-weight products represent a significant fraction of the total sequencing library, then it is usually desirable to perform a second size-selection step to remove them. Our experience has been that these products tend to be more evident when beginning with a lower overall RNA input, such as occurs with RISC-Seq. If the low molecular-weight products are not removed, they will compete with the desired cDNA fragments for binding to the Illumina sequencing flowcell and decrease the number of useful sequencing reads obtained as a result. One option is to perform another gel purification round, this time on the PCR-amplified material. Our core sequencing facility offers AMPure XP bead purification (Beckman Coulter), on the basis of differential binding of differently-sized fragments, to achieve the same effect.

Quantitate the final libraries using a fluorescent DNA-binding dye protocol (such as the Qubit or PicoGreen assays); we have repeatedly found that this gives more accurate loading than NanoDrop or other UV-based measurement for sequencing library quantitation. Prepare an equimolar mixture of the desired, indexed libraries and dilute as appropriate for the chosen sequencing instrument. As an example, we submit library mixtures at 10 nM to our sequencing facility, which are then further diluted to 5 pM for cluster formation on Illumina flowcells.

3.8 Sequence data generation

Typically, sequencing reactions and determination of nucleotide sequences (basecalling) are performed using highly specialized reagents and equipment which are particular to the major platforms (e.g. Illumina HiSeq, Applied Biosystems SOLiD, Life Technologies Ion Torrent, and the like). These steps are often performed at a core facility of an academic institution or a commercial service. As such, it is beyond the scope of this chapter to discuss detailed protocols for generation of raw sequence data.

However, it is worthwhile to consider the format in which basecalled sequencing read data are provided to the end-user. It is very helpful if demultiplexing (allocating reads according to indexes) is performed prior to receipt of the data, as this process can be very time-consuming on a desktop-class workstation. A common format for distribution of sequencing read data, with accompanying probabilities of basecalling accuracy, is the fastq (.fq) format. We maintain archives of the original fastq data for our RNA-seq experiments together with

tables describing the experimental design and index allocation. Ensuring that original data are preserved at this level means that they are readily available for alignment against new transcriptome versions (approved genome releases) in the future.

3.9 Sequence data quality control

In high-throughput analyses where hundreds to thousands of RNA species are being measured simultaneously, it can be challenging to identify whether a particular biological or technical replicate is behaving so differently from its counterparts that it should be classed as an outlier. While this topic is covered in more detail under section 3.12, some simple determinations can help to pinpoint problematic samples.

Analyzing fastq files with the freely available program FastQC can point out anomalies such as a high percentage of self-ligated or unligated sequencing adapters or indexing primers in the sequencing read data, or an unduly large number of identical sequencing reads. The program is written in Java and can thus be used in Windows, Mac and Linux/Unix operating environments.

It is also prudent to monitor the number of raw sequencing reads present in each library, and the number which were successfully aligned to the transcriptome. These data are gained from the procedures outlined in sections 3.10 and 3.11, but should be assessed before proceeding with downstream analyses. Inconsistencies in these parameters reveal samples that were insufficiently loaded on to the sequencer or that suffered degradation or other problems during library preparation.

3.10 microRNA-Seq data alignment and quantitation

In this section and the following ones, a general workflow will be described, together with recommendations for specific software packages that we currently use. However, many of the programs used for read alignment and quantitation are under continuous development, or are superseded by more powerful alternatives, and due to their 'open-source' nature may become unavailable or relocated to different web addresses without notice. The majority of these programs are designed to be run in a Linux environment, either on a native Linux computer, a Linux emulator for Windows (such as Cygwin), or the 'Terminal' of Mac OS X. Biological scientists that have not had previous exposure to Unix-style operating systems may experience some level of discomfort with the use of these programs at first, and some initial assistance from a computational biologist or bioinformatician that is fluent in the environment may be helpful in overcoming the initial learning curve. A detailed description of Unix file management and commands, read alignment procedures, and use of the popular statistical software R and its associated Bioconductor programs is beyond the scope of this chapter, and would require a chapter of its own. We have previously published example programs (Unix 'shell scripts') that accomplish the steps used in our RNA-sequencing pipelines (supplemental material of reference [25]). However, we encourage biologists not to be dismayed by the unfamiliarity of these procedures, and to at least develop a strong understanding of the conceptual underpinnings, even if the bulk of the computational work is outsourced. One helpful resource for the beginner, and which allows many of the

following procedures to be performed via a friendly web interface, is the Galaxy project (<http://galaxyproject.org/>).

microRNA sequencing alignment and quantitation require the following:

1. Trimming of adapter sequences from sequencing reads and alignment of trimmed reads to annotated microRNAs and other small RNAs
2. Separation of microRNA-specific reads from other small RNA reads
3. Quantitation of microRNAs, typically as a proportion of the total microRNA-specific reads, e.g. RpM (Reads per Million microRNA reads).

We use the E-miR program [27], which is well-documented and runs in a Linux environment, to perform step 1. E-miR aligns reads to all known small or noncoding RNAs in Ensembl databases. It requires a separate alignment software module to be installed, and for this we use Bowtie, a well-established and efficient program which runs easily on desktop or laptop-class workstations [29]. Notes 9 and 10. Trimming of adapter sequences is necessary because the length of the sequencing read is greater than that of the mature microRNA sequence. We manually filter the resulting data for microRNAs (step 2) by importing the output files from E-miR into a spreadsheet program such as Microsoft Excel. Finally, we add together all of the sequence reads for each sample and use this as the denominator when calculating RpM.

Many other possibilities exist for analysis of small RNA sequencing data, such as the software miRdeep* [28].

3.11 mRNA- and RISC-Seq data alignment and quantitation

The successful alignment and quantitation of relatively short sequencing reads to mRNA sequences is substantially more complex than for microRNAs, in which the entire sequence of the mature microRNA is encompassed by the length of the read. Sequences aligning to different sections of the same mRNA must be gathered together, and the alignment program must be aware that sequences may span exon-exon boundaries.

Our workflow is as follows, and is the same for both mRNA and RISC RNA sequencing (Figure 2):

1. Alignment of sequencing reads to the desired genome / transcriptome. We use the program TopHat [30,36], which employs the Bowtie aligner [29] at its core with specific consideration for the splicing out of introns. TopHat can use a genome-wide, annotated transcriptome database for its alignment procedure. Annotated genome-wide transcriptomes from Ensembl or UCSC sources often contain multiple noncoding RNA species in addition to mRNAs. While this is useful in some circumstances, we prefer the alternative allowed by TopHat of defining a more limited transcriptome and annotating only to this. An advantage of such an approach is that it is easy to limit alignments to mRNAs only (without inclusion of noncoding RNAs), which are the only data we require for parallel mRNA- and RISC-Seq analyses, and thus discarding all reads which map to e.g. the rRNA that is inevitably contained in RISC-Seq libraries.

2. Conversion of aligned sequencing reads to gene expression measures. The calculation of gene expression values (with or without provision for determining isoform fractions and/or new alternatively-spliced transcripts) and the use of normalization procedures to compensate for differences library sequencing depth and/or gene length [37,38] has been the subject of much discussion. These issues cannot be divorced from debate over the most appropriate method for determination of significant differences in gene expression between treatments, with three principal scientific groups involved in the development of methods to achieve this. All have released packages to calculate expression values and associated statistics (HTSeq and DESeq [31]; Cufflinks and Cuffdiff [39]; and edgeR [40]) and the strengths and weaknesses of each approach have been the subject of energetic discussions in online forums (e.g. <http://seqanswers.com/forums/index.php>).

HTSeq/DESeq, and edgeR, use the actual number of reads aligned to a given mRNA, normalized only to the total reads obtained for each individual sequence library, in their statistical calculations. They can use either local (per-gene) or global (per gene-set) dispersion ($\sim$ variance) to determine statistical significance, and rely on a negative binomial distribution. A variant of DESeq, DEXSeq [41], can calculate exon-level differences for analysis of separate gene isoforms. Cufflinks and Cuffdiff use a gene expression measurement that adjusts for both library depth and gene length, the FPKM (Fragments per kb of exon per million mapped reads), and offer to investigate isoform-level differences, but estimate variances in a different manner [39]. A workflow for the full ‘Tuxedo’ suite (Bowtie, Tophat, Cufflinks and Cuffdiff) has been described by the programs’ authors [36].

Our internal workflow currently uses the HTSeq/DESeq method for normalization of differences in sequencing depth between individual samples, and reports gene expression as the normalized number of reads for each gene (without any attempt to compensate for gene length, or any attempt to discover novel transcripts). For convenience, we do use Cufflinks to process TopHat-aligned data into FPKM values, which give a sense of the concentration (molarity) of each gene in the preparation that is not possible when using HTSeq reads alone, but do not use Cuffdiff for differential expression analyses. By monitoring HTSeq reads and Cufflinks FPKM values, we eliminate genes from our data sets that have insufficient reads for reliable determination, and/or are likely expressed at cellular levels of limited biological significance. Indeed, we perform similar filtering of raw microRNA data prior to statistical analysis, which includes elimination of all RNAs in the output files for which zero reads were obtained. Thus, false discovery rates (which are based partly on the number of input RNAs) are calculated in accordance with the number of detected RNA species, rather than in accordance with the (greater) number of RNA species present in an annotation file.

3.12 Statistical tools to determine differential gene expression

In many of our earlier publications using deep sequencing methodologies, we used a fairly straightforward method of importing Cufflinks-derived FPKM values into statistical software such as Partek Genomics Suite (<http://www.partek.com/>), log-transformed the data (achieving a better fit to the normal distribution) and calculated p-values and associated false

discovery rates for comparisons between treatment conditions. However, in the light of studies comparing various strategies for read depth normalization across different individual samples [42] and differential gene (mRNA) expression methods [37], we now use the DESeq package [31], to compute differential microRNA and mRNA expression. DESeq uses a negative binomial distribution in comparison to the Gaussian distribution assumed by other tools for log-transformed data, and is designed to be more robust in contexts where relatively few genes are highly abundant and there is a majority of less abundant genes. edgeR is based on similar concepts [40].

DESeq comprises a series of data import and manipulation commands for the R statistical environment. The software is freely available and includes thorough documentation. One potential stumbling block for the use of this and many related data analysis packages involves correct formatting of the read data from each sample (which is specific to each package). For small numbers of samples, careful import into a spreadsheet program with any necessary manual rearrangement may suffice. However, for import and grouping of large numbers of samples, with the attendant possibilities for operator error, it may prove valuable to employ a user-friendly relational database such as Microsoft Access to import and correctly group data in a more automated fashion. While there is insufficient space to give detailed procedures here, mastering the concepts of 'crosstab query' and 'lookup tables' should prove to be useful in converting the output from programs such as HTSeq and Cufflinks into the data tables required for statistical analysis.

An additional consideration is the number of biological replicates that are necessary to measure a given differential gene expression event (i.e., the experiment's power). Once again, the multidimensional nature of the data obtained from these assays means that classical methods of analyzing experimental power are not well suited to these tasks. In practical terms, we aim to obtain six biological replicates of any given treatment condition, which is readily achievable for most studies involving mouse hearts. Our experience has been that the reduction in variance obtained by using six (vs three or four replicates), and subsequent increase in power, far outweighs the additional considerations of cost, especially with continued decreases in the cost of sequencing per base.

4. Notes

1. On the issue of effective precipitation of small RNA using the Trizol method, see reference [43], which suggests that some small RNAs may be selectively lost during Trizol extraction. While the title may not sound especially hopeful, in this important technical paper the authors only use a 10 min precipitation in isopropanol with Trizol when comparing to microRNA extraction with Ambion's miRvana kit using a column-binding step, and the 'small number of cells' referred to in the manuscript as a difficulty are those sufficient for a total RNA yield of only ~1 µg. Our RNA extraction protocol using Trizol calls for at least a 30 min isopropanol precipitation step. The extraction deficit appears to be minimized at 10 µg yield and absent at 50 µg (see figure 1E of [43]), and considering the approximate total RNA yield of at least 50 µg from even a mouse heart apex (1/3rd of the heart) this should be of no or only minor concern.

2. This quantity of oligo(dT)-Dynabeads (100 μ L) has a binding capacity of 1 μ g of polyA⁺ RNA. Considering that 5 μ g of total RNA is unlikely to contain more than 100–150 ng of polyA⁺ RNA (2–3%), it is likely possible to reduce the amount of Dynabead suspension further if desired.
3. The Wako Pure monoclonal anti-Ago2 antibody recommended here is targeted toward mouse, rat and hamster Ago2. While we have not used it, an alternate monoclonal antibody is available from Wako Pure is available for anti-human Ago2 immunoprecipitation. For smaller quantities of tissue, less anti-Ago2 antibody should be required for maximal RNA yield, although this will need to be optimized by the user.
4. Since only small quantities of RNA are present in the Ago2 immunoprecipitate, it will not be possible to visualize an RNA pellet after isopropanol precipitation and centrifugation. To avoid unintentional loss of RNA at this stage, add 20 μ g RNase-free glycogen immediately prior to isopropanol precipitation, and precipitate for 30 min at room temperature. After washing with 75% ethanol, resuspend the final dried pellet in 30 μ L 10 mM Tris, pH 8.2. Store RNA at –80 C.
5. Time the length of heat treatment exactly to ensure reproducibility of chemical shearing between libraries prepared on different occasions.
6. Out of an abundance of caution to protect against RNA degradation, we perform RNA fragmentation, column purification, and 1st- and 2nd-strand cDNA synthesis in a single working day.
7. Perform the steps listed as ‘optional’ in the Qiagen booklet by adding 1 gel volume of isopropanol to the melted agarose, and an additional wash in 0.5 mL buffer QG prior to washing in buffer PE.
8. 12 cycles of PCR at this step is optimal when starting with 100–200 ng of polyA⁺ + RNA as input into sections 3.5–3.7, and using 1/3rd of the size-selected, gel-purified cDNA as input into PCR. The lower input of RNA derived from Ago2 immunoprecipitates necessitates further amplification; our rule-of-thumb has been that an immunoprecipitation from 2/3rd of a mouse heart, from which a sequencing library is prepared in the same way as for mRNA-Seq, requires at least an additional 4 PCR cycles in order to be visualized and accurately quantitated.

Over-amplification and introduction of a degree of PCR bias is possible at this stage, and the use of too many PCR cycles is typically indicated by the formation of distinct bands within the desired size range, rather than an even ‘smear’ of fragments across the range. If this is seen to occur, it is desirable to begin PCR with a new aliquot of 10 μ L of the size-selected, gel-purified library and adjust the number of PCR cycles.
9. The E-miR version using Bowtie as its alignment module appears to have a bug (at least as of late 2011) in which mature microRNAs which originate from multiple genomic loci fail to map correctly. We have overcome this by manually

altering the line in the E-miR Perl program which invokes the Bowtie command, using appropriate parameters for Bowtie to allow alignment of such ‘multiply-mapped’ reads [29].

10. Specifically, we modify line 563 of the EmiR_Bowtie.pl file as follows:

```
(from line 551)
#####
#####
##      POST ALIGNMENT SUBS
#####
#####
#####
sub load_bowtie_output_files # and process
{
    & time; print "Running Bowtie alignment using $processors
processors start =
$theTime...\n";
    # system("cat *_TBDL | awk '{print >\n\"$1\"\\n\\n\"$1}'>
bowtie_temp.fa ");
    $outfile = "bowtie_temp.fa_OUT";
(line 563) system("$bowtie_location -p $processors -n 1 -l 32 -f -
m 1 --best --strata
$genome_location bowtie_temp.fa $outfile");
(change line 563 to) system("$bowtie_location -p $processors -n 1 -
l 32 -f
$genome_location bowtie_temp.fa $outfile");
```

11. E-miR aligns reads against genomic regions in which microRNA precursors are embedded. While we do find the program useful, we have noticed using the Bowtie version of the program (modified as described in [note 9](#)) that mature microRNAs which originate from more than 1 genomic locus (e.g. miR-9 and miR-133a) can be randomly assigned to one locus or the other in output tables. We manually add together the microRNA reads from multiple loci to arrive at a single value for the number of reads pertaining to that microRNA.

Acknowledgments

Related work in the authors' laboratories is supported by the NIH-sponsored Diabetes Research Center at Washington University, grant 5 P30 DK020579 (to S.J.M.) and NIH grant R01 HL108943 (to G.W.D.).

References

1. Lynn FC, Skewes-Cox P, Kosaka Y, et al. MicroRNA expression is required for pancreatic islet cell genesis in the mouse. *Diabetes*. 2007; 56:2938–2945. [PubMed: 17804764]
2. Kuehbachner A, Urbich C, Zeiher AM, et al. Role of Dicer and Drosha for endothelial microRNA expression and angiogenesis. *Circ Res*. 2007; 101:59–68. [PubMed: 17540974]

3. Chen JF, Murchison EP, Tang R, et al. Targeted deletion of Dicer in the heart leads to dilated cardiomyopathy and heart failure. *Proc Natl Acad Sci USA*. 2008; 105:2111–2116. [PubMed: 18256189]
4. Melkman-Zehavi T, Oren R, Kred-Russo S, et al. miRNAs control insulin content in pancreatic β -cells via downregulation of transcriptional repressors. *EMBO J*. 2011; 30:835–845. [PubMed: 21285947]
5. van Rooij E, Olson EN. MicroRNAs: powerful new regulators of heart disease and provocative therapeutic targets. *J Clin Invest*. 2007; 117:2369–2376. [PubMed: 17786230]
6. Liu N, Olson EN. MicroRNA regulatory networks in cardiovascular development. *Dev Cell*. 2010; 18:510–525. [PubMed: 20412767]
7. Chen CY, Zheng D, Xia Z, et al. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat Struct Mol Biol*. 2009; 16:1160–1166. [PubMed: 19838187]
8. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 2011; 39:D152–157. [PubMed: 21037258]
9. Matkovich SJ, Hu Y, Dorn GW 2nd. Regulation of cardiac microRNAs by cardiac microRNAs. *Circ Res*. 2013; 113:62–71. [PubMed: 23625950]
10. Matkovich SJ, Zhang Y, Van Booven D, et al. Deep mRNA sequencing for in vivo functional analysis of cardiac transcriptional regulators. Application to G α_q . *Circ Res*. 2010; 106:1459–1467. [PubMed: 20360248]
11. Hu Y, Matkovich SJ, Hecker PA, et al. Epitranscriptional orchestration of genetic reprogramming is an emergent property of stress-regulated cardiac microRNAs. *Proc Natl Acad Sci USA*. 2012; 109:19864–19869. [PubMed: 23150554]
12. Matkovich SJ, Van Booven DJ, Youker KA, et al. Reciprocal regulation of myocardial microRNAs and messenger RNA in human cardiomyopathy and reversal of the microRNA signature by biomechanical support. *Circulation*. 2009; 119:1263–1271. [PubMed: 19237659]
13. Ikeda S, Kong SW, Lu J, et al. Altered microRNA expression in human heart disease. *Physiol Genomics*. 2007; 31:367–373. [PubMed: 17712037]
14. van Rooij E, Sutherland LB, Liu N, et al. A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. *Proc Natl Acad Sci USA*. 2006; 103:18255–18260. [PubMed: 17108080]
15. Putt ME, Hannonhalli S, Lu Y, et al. Evidence for coregulation of myocardial gene expression by MEF2 and NFAT in human heart failure. *Circ Cardiovasc Genet*. 2009; 2:212–219. [PubMed: 20031589]
16. Margulies KB, Matiwala S, Cornejo C, et al. Mixed messages: transcription patterns in failing and recovering human myocardium. *Circ Res*. 2005; 96:592–599. [PubMed: 15718504]
17. van Rooij E, Sutherland LB, Thatcher JE, et al. Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc Natl Acad Sci USA*. 2008; 105:13027–13032. [PubMed: 18723672]
18. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*. 2011; 12:87–98. [PubMed: 21191423]
19. Jensen KB, Darnell RB. CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *Methods Mol Biol*. 2008; 488:85–98. [PubMed: 18982285]
20. Hafner M, Landthaler M, Burger L, et al. PAR-CLIP—a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J Vis Exp*. 2010
21. Hafner M, Landthaler M, Burger L, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010; 141:129–141. [PubMed: 20371350]
22. Licatalosi DD, Mele A, Fak JJ, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*. 2008; 456:464–469. [PubMed: 18978773]
23. Chi SW, Zang JB, Mele A, et al. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*. 2009; 460:479–486. [PubMed: 19536157]
24. Karginov FV, Conaco C, Xuan Z, et al. A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci USA*. 2007; 104:19291–19296. [PubMed: 18042700]

25. Matkovich SJ, Hu Y, Eschenbacher WH, et al. Direct and indirect involvement of microRNA-499 in clinical and experimental cardiomyopathy. *Circ Res.* 2012; 111:521–531. [PubMed: 22752967]
26. Dorn GW 2nd, Matkovich SJ, Eschenbacher WH, et al. A human 3' miR-499 mutation alters cardiac mRNA targeting and function. *Circ Res.* 2012; 110:958–967. [PubMed: 22374132]
27. Buermans HP, Ariyurek Y, van Ommen G, et al. New methods for next generation sequencing based microRNA expression profiling. *BMC Genomics.* 2010; 11:716. [PubMed: 21171994]
28. An J, Lai J, Lehman ML, et al. miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic Acids Res.* 2013; 41:727–737. [PubMed: 23221645]
29. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10:R25. [PubMed: 19261174]
30. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009; 25:1105–1111. [PubMed: 19289445]
31. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010; 11:R106. [PubMed: 20979621]
32. Matkovich SJ, Van Booven DJ, Eschenbacher WH, et al. RISC RNA sequencing for context-specific identification of in vivo microRNA targets. *Circ Res.* 2011; 108:18–26. [PubMed: 21030712]
33. Mortazavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008; 5:621–628. [PubMed: 18516045]
34. Toung JM, Morley M, Li M, et al. RNA-sequence analysis of human B-cells. *Genome Res.* 2011; 21:991–998. [PubMed: 21536721]
35. Labaj PP, Leparic GG, Linggi BE, et al. Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. *Bioinformatics.* 2011; 27:i383–391. [PubMed: 21685096]
36. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* 2012; 7:562–578. [PubMed: 22383036]
37. Bullard JH, Purdom E, Hansen KD, et al. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics.* 2010; 11:94. [PubMed: 20167110]
38. Oshlack A, Wakefield MJ. Transcript length bias in RNA-seq data confounds systems biology. *Biol Direct.* 2009; 4:14. [PubMed: 19371405]
39. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and abundance estimation from RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010; 28:511–515. [PubMed: 20436464]
40. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26:139–140. [PubMed: 19910308]
41. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* 2012; 22:2008–2017. [PubMed: 22722343]
42. Dillies MA, Rau A, Aubert J, et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform.* 2013; 14:671–683. [PubMed: 22988256]
43. Kim YK, Yeo J, Kim B, et al. Short structured RNAs with low GC content are selectively lost during extraction from a small number of cells. *Mol Cell.* 2012; 46:893–895. [PubMed: 22749402]

Figure 1.
Workflow for laboratory processing of input RNA to final sequencing libraries.

Figure 2.

Workflow for processing sequence reads to gene abundance data and differential expression.

Table 1

Adapter oligonucleotide and PCR primer sequences.

Name	Sequence 5'-3'
Index Adapter A	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Index Adapter B	/5phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Index PCR 1.0	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Index PCR 2.0	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Indexing primer X	CAAGCAGAAGACGGCATACGAGAT <u>ACGCCT</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGA
Indexing primer Y	CAAGCAGAAGACGGCATACGAGAT <u>AGTTAAA</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGA
Indexing primer Z	CAAGCAGAAGACGGCATACGAGAT <u>GAGGACC</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGA

Nucleotide sequences for adapters and index primers are proprietary designs of Illumina Corporation. Adapter B requires 5' phosphorylation (shown by the symbol /5phos/, and adapters A and B are mixed in equimolar proportions and annealed prior to use. The underlined nucleotides in indexing primers X, Y and Z reflect sequences which can be varied to produce different indexing primers; these sequences are the *reverse complement* of what will be observed in raw sequencing reads of the index region (i.e. primer X will be read as GAGGCGT, primer Y as TTAACT, etc.). We order Adapters A and B and Index PCR 1.0 and 2.0 as 100 nmole scale, HPLC-purified oligonucleotides from Integrated DNA Technologies (<http://www.idtdna.com>) and indexing primers as 4 nmole scale 'Ultramers'.