Chapter 13

MicroRNA Expression Analysis Using Small RNA Sequencing Discovery and RT-qPCR-Based Validation

Alan Van Goethem, Pieter Mestdagh, Tom Van Maerken, and Jo Vandesompele

Abstract

miRNAs are small noncoding RNA molecules that function as regulators of gene expression. Deregulated miRNA expression has been reported in various diseases including cancer. Due to their small size and high degree of homology, accurate quantification of miRNA expression is technically challenging. In this chapter, we present two different technologies for miRNA quantification: small RNA sequencing and RT-qPCR.

Key words miRNA, Small RNA sequencing, RT-qPCR, miRNA annotation, Normalization

1 Introduction

MicroRNAs are a large class of small noncoding RNAs that regulate gene expression at the posttranscriptional level. To date, over 28,000 hairpin miRNAs, giving rise to more than 35,000 mature miRNAs in 223 species, have been described in the miRBase sequence database (version 21, http://www.mirbase.org), including 2588 mature human miRNAs. The prominent role of miRNAs in virtually every aspect of cell biology and their involvement in disease have led to the development of both diagnostic and prognostic miRNA expression signatures as well as miRNA-based therapeutics. As miRNAs function as biological rheostats concurrently affecting several target genes, even subtle alterations in their abundance may have substantial impact. Unfortunately, their small size, low abundance, and the high degree of homology among miRNA family members make accurate quantification of mature miRNA expression levels technically challenging. Several platforms are available for assessing miRNA abundance, based on (micro-array) hybridization, reverse transcription qPCR (RT-qPCR), or (small RNA) sequencing.

Michael Kaufmann et al. (eds.), *Functional Genomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1654,
DOI 10.1007/978-1-4939-7231-9_13,

The miRQC study has comprehensively assessed different miRNA expression platforms using quantifiable performance metrics [1]. The result is an unbiased comparison of accuracy, specificity, sensitivity, and reproducibility among 12 different platforms from 9 different vendors. Each platform was found to have its strengths and weaknesses. The study outcome should aid researchers making an informed selection of platform corresponding to the experimental setting and the specific research question.

Since the start of the miRQC study, RNA sequencing has witnessed continuing technical and workflow improvements and decreasing costs. Combined with the possibility to assess a large number of small RNAs including the discovery of previously uncharacterized miRNAs, this has rendered small RNA sequencing as the gold standard method for miRNA discovery and quantification. Typically, and as recommended in the miRQC study conclusions, the initial sequencing based screening experiment is followed by validation of obtained results using RT-qPCR.

1.1 Small RNA Sequencing

Similar to most RNA sequencing approaches, small RNA sequencing requires the construction of cDNA libraries (Fig. 1). The initial step of library preparation is adapter ligation. The adapters serve as a template for primer-based RT, amplification, and sequencing. All RNA molecules containing a 5′ phosphate and 3′ hydroxyl group will be subjected to both 5′ and 3′ single stranded RNA adapter ligation. Adapter ligation is followed by reverse transcription of the adapter-ligated RNA into cDNA and PCR amplification of the cDNA libraries. During the PCR amplification step, each library is tagged with a unique index that enables identification of the library origin of individual reads when analyzing sequencing data, thus making it possible to simultaneously sequence a few dozen samples. Library preparation kits from different vendors are available for the preparation of small RNA libraries, they mainly differ in the process of adapter ligation and the presence or absence of adapter dimer removal [2, 3]. After PCR amplification, a library size selection step is performed to selectively enrich and select for the miRNA-containing fraction of the resulting libraries. Size selection involves a size-based separation of the library by agarose gel electrophoresis followed by DNA staining and the collection of the band containing the miRNA fraction. This can be performed manually or, alternatively, through the use of fully automated size-selection systems.

1.2 Processing Small RNA Sequencing Data

High-throughput sequencing of small RNAs leads to the generation of considerable amounts of data. The processing of these data for expression analysis can be roughly divided into two steps: preprocessing of raw sequencing data into a miRNA count table and differential gene expression analysis, including normalization of miRNA count data followed by statistical analyses. During

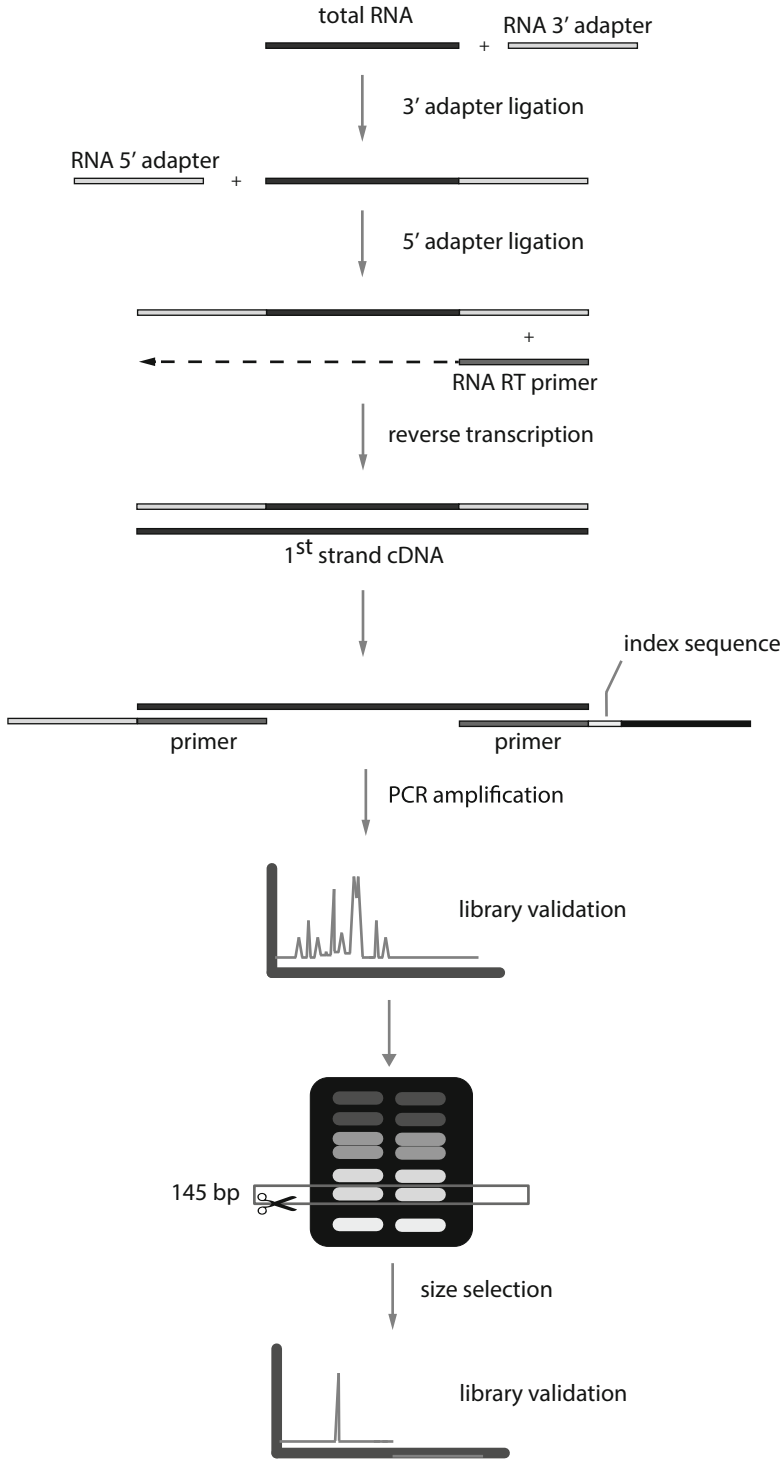

Fig. 1 Schematic overview of the TruSeq small RNA library preparation protocol

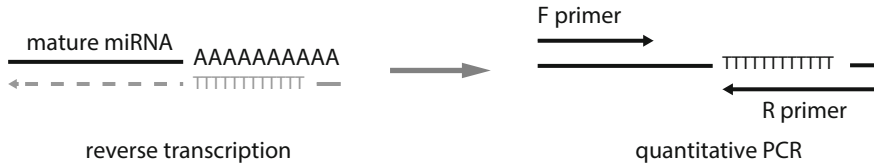

Fig. 2 Schematic overview of the universal PCR profiling platform

preprocessing adapter trimming, quality control and read mapping is performed, after which annotation information is retrieved from miRBase (and other reference databases if needed) to create a miRNA count table. As small RNA sequencing data are subject to various sources of technical variation (e.g., differences in library size or GC-content) it is necessary to perform some form of data normalization to correct for these variations. Several normalization strategies for (small) RNA-seq data have been developed (compared in [4]), but consensus on the optimal normalization method is currently lacking. The most widely used algorithm is the negative binomial-based approach DESeq2 [5] (*see* **Note 1**). DESeq2 applies the median of the ratios of each gene in one sample over the geometric mean of that gene across all samples as a scaling factor for all the genes in that sample [5].

1.3 RT-qPCR (Universal Primer)

As recommended in the miRQC study conclusions, any screening study should be followed by targeted validation using an independent method. Therefore, sequencing experiments are often followed by RT-qPCR validation of obtained results, typically a limited number of differentially expressed miRNAs. RT-qPCR-based quantification of miRNAs offers superior sensitivity and accuracy. To enable the detection of short RNA molecules like miRNAs by RT-qPCR, the reverse transcription (RT) reaction requires modification. The most widely used RT-qPCR platforms are based on either the use of stem-loop RT primers or polyadenylation of the mature miRNA to enable RT [6]. Here, we describe a protocol for universal RT-qPCR, meaning the cDNA can be used for the quantification of any miRNA. This approach is based on polyadenylation of the mature miRNA prior to oligo-dT primed cDNA synthesis (Fig. 2) [7].

1.4 Identification of Stably Expressed miRNAs for Data Normalization

To distinguish technical variation from true biological difference, it is important to perform proper normalization of the obtained data. Using the geometric mean of multiple stable reference genes is widely accepted as the gold standard for the normalization of RT-qPCR data [8]. Using well-established algorithms such as geNorm, stably expressed reference genes are typically identified out of several candidate reference genes in a small pilot study. However, when it comes to normalization of microRNA data, there are no predefined sets of candidate reference microRNAs and all too often, small

nuclear or nucleolar RNAs (such as U6, U24, and U26) are used instead. We strongly advise against these internal controls as sn(o) RNAs are transcribed from a different RNA polymerase and have different functions than miRNAs. For the measurement of a large unbiased set of miRNAs we have instead published the use of the global mean miRNA expression for accurate normalization [9, 10]. However, in the context of focused validation experiments it is not possible to rely on whole-genome based normalization strategies as one is typically interested in the validation of a limited number of differentially expressed miRNAs. For this kind of experiments it is possible to identify miRNAs that resemble the global mean expression value and whose geometric mean can be used to mimic global mean normalization [9].

1.5 miRNA Annotation

The concluding step in most miRNA studies is reporting of experimental findings. From 2002, miRBase has emerged as the reference database of miRNA nomenclature. Since then miRBase underwent numerous additions and deletions of miRNA records, adaptations into more complex naming structures and changes in annotated miRNAs. These changes are the necessary consequence of increasing insights into the (mi)RNA world but they also give rise to substantial ambiguity concerning miRNA annotation in literature. Ignoring sequence annotation changes has led to erroneous interpretation, comparison, and integration of miRNA study results (*see* **Note 2**). To resolve these issues, our lab has developed miRBaseTracker (www.mirbasetracker.org), an online tool that enables comparison of all current and historical miRNA annotation data present in miRBase [11].

2 Materials

2.1 Small RNA Sequencing

1. TruSeq Small RNA library preparation kit V2 containing 10 mM ATP, HML (Ligation buffer), RNA 3′ adapter, RNA 5′ adapter, RNase Inhibitor, Stop solution, T4 RNA ligase, 25 mM dNTP mix, PCR mix, RNA PCR Primer, RNA PCR Primer Index, RNA RT Primer, RNase Inhibitor, 5× First Strand buffer, and 100 mM DTT.
2. T4 RNA ligase 2, deletion mutant, 200 U/μL.
3. Superscript II Reverse Transcriptase, 200 U/μL.
4. Agilent High Sensitivity DNA kit containing High Sensitivity DNA chips, High Sensitivity DNA Ladder, High Sensitivity DNA Markers (35–10,380 bp), High Sensitivity DNA dye concentrate, and High Sensitivity DNA gel Matrix.
5. Agilent 2100 Bio-Analyzer.
6. Nuclease-free water.

7. Pippin Prep cassettes (3% dye-free with legacy marker H) plus loading buffer and extra running buffer.
8. Glycogen, 20 μg/μL.
9. 3 M NaOAc.
10. 100% ethanol, −20 °C.
11. 70% ethanol.
12. 10 mM Tris–HCl, pH 8.5.
13. 5 μM Library quantification primer assays, Forward Primer: AATGATACGGCGACCACCGA, Reverse Primer: CAAGCAGAAGACGGCATACGA.
14. SsoAdvanced universal SYBR Green supermix.
15. Qubit DS DNA HS assay kit.
16. NextSeq 500 Mid/High Output V2 kit, 75 cycles.

2.2 Universal Primer RT-qPCR

1. miScript II RT kit containing miScript Reverse Transcriptase Mix, 10× miScript Nucleics mix, 5× miScript HiSpec Buffer, and 5× miScript HiFlex buffer.
2. Optional: miScript PreAMP kit containing 5× miScript PreAMP Buffer, HotStarTaq DNA Polymerase (2 U/μL), miScript PreAMP Universal Primer.
3. miScript SYBR Green PCR kit containing QuantiTect SYBR Green PCR Master Mix and 10× miScript Universal Primer.
4. 5 μM miScript Primer Assay.
5. Nuclease-free water.
6. TE buffer: pH 8.0, 10 mM Tris–HCl, 1 mM EDTA (for dissolving Primer assays).

3 Methods

3.1 Small RNA Sequencing

1. Add 5 μL of RNA sample (between 0.1 and 1 μg of total RNA derived from tissues or cells) to 1 μL of RNA 3′ adapter (*see* **Note 3**). Spin to collect the liquid. It is important to keep RNA on ice at all times. Do not vortex RNA.
2. Incubate the mixture at 70 °C for 2 min and immediately place on ice.
3. Combine 2 μL of Ligation buffer (HML), 1 μL of RNase inhibitor, and 1 μL of T4 RNA ligase, deletion mutant per sample (*see* **Note 4** for the preparation of mixtures when processing multiple samples). Pipet up and down and briefly spin. Add 4 μL of the ligation mix to the reaction tube from **step 2** and incubate for 1 h at 28 °C.
4. After 1 h, quickly spin, add 1 μL of Stop solution, and continue to incubate at 28 °C for 15 min.

5. Incubate RNA 5′ adapter at 70 °C for 2 min and place immediately on ice. Prepare the 5′ adapter ligation mix by combining 1 μL of 5′ RNA adapter, 1 μL of 10 mM ATP, and 1 μL T4 RNA ligase per sample. Pipet up and down and briefly spin. Add 3 μL of ligation mix to the reaction tube from **step 3** and incubate for 1 h at 28 °C.
6. Dilute the 25 mM dNTPs twofold to a 12.5 mM mix using nuclease-free water. Add 1 μL of RNA RT primer to 6 μL of 5′-3′ ligated RNA and incubate at 70 °C for 2 min. Immediately place on ice.
7. Prepare the RT mixture by combining 2 μL of 5× First Strand Buffer, 0.5 μL of 12.5 mM dNTP mix, 1 μL of 100 mM DTT, 1 μL of RNase Inhibitor, and 1 μL SuperScript II Reverse transcriptase per sample. Pipet up and down and briefly spin. Add 5.5 μL of RT mixture to the reaction tube from **step 5**.
8. Incubate at 50 °C for 1 h and immediately place on ice.
9. Prepare the PCR amplification mix by combining 2 μL of RNA PCR primer, 25 μL of PCR mix, and 8.5 μL of nuclease-free water per sample. Add 35.5 μL of the PCR amplification mix to 12.5 μL of RT product from **step 8**. Add 2 μL of RNA PCR Primer Index.
10. Run the PCR reaction as follows: 98 °C for 30 s, (98 °C for 10 s, 60 °C for 30 s, 72 °C for 15 s) × 11 (*see* **Note 5**), 72 °C for 10 min, 4 °C hold.
11. Dilute 1 μL of each library twofold using nuclease-free water and run on a High Sensitivity DNA chip to perform quality control analysis on the library. Figure 3 shows the typical

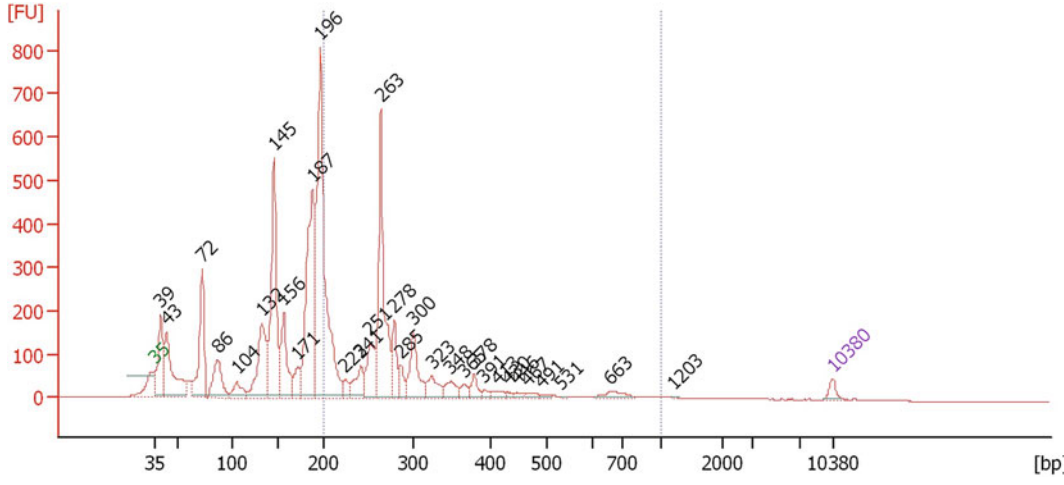

Fig. 3 Profile of a small RNA library prepared from 750 ng of tumor-derived RNA and run on an Agilent 2100 using a High Sensitivity DNA chip. The peak at 145 bp represents the miRNA fraction of the library

profile of a library prepared from 750 ng of tumor-derived RNA.

12. Load individual libraries in the lanes of a 3% agarose dye-free marker H cassette and run on a Pippin Prep with a specified collection range of 125–153 bp. Selecting this size range should maximize the collection of the miRNA fraction with minimal contaminant RNAs.
13. Collect 40 µL of resulting library from the collection well and add 2 µL of glycogen, 30 µL of 3 M sodium acetate, and 977 µL of 100% ethanol (−20 °C). Immediately centrifuge at 20,000 × *g* for 20 min at 4 °C in a fixed-angle centrifuge. Remove and discard the supernatant, leaving the pellet intact. Wash the pellet with 500 µL of 70% ethanol and centrifuge at 20,000 × *g* for 2 min at room temperature. Remove and discard the supernatant leaving the pellet intact. Dry the pellet by placing the tube, lid open, in a 37 °C heat block for 5–10 min or until dry. You will observe a shift from a white, opaque pellet to a transparent pellet if completely dry and pure. Dissolve the pellet in 20 µL 10 mM Tris–HCl, pH 8.5.
14. Dilute 1 µL of each library twofold and run on a High Sensitivity DNA chip to perform quality control analysis on the library. Figure 4 shows the typical profile of a library prepared from 750 ng of tumor-derived RNA after size selection.
15. Dilute 1 µL of each library 100,000-fold using nuclease-free water through serial dilution. Combine per sample 2.5 µL of SsoAdvanced universal SYBR Green supermix with 0.25 µL of each primer. Distribute PCR mixture in the PCR reaction and

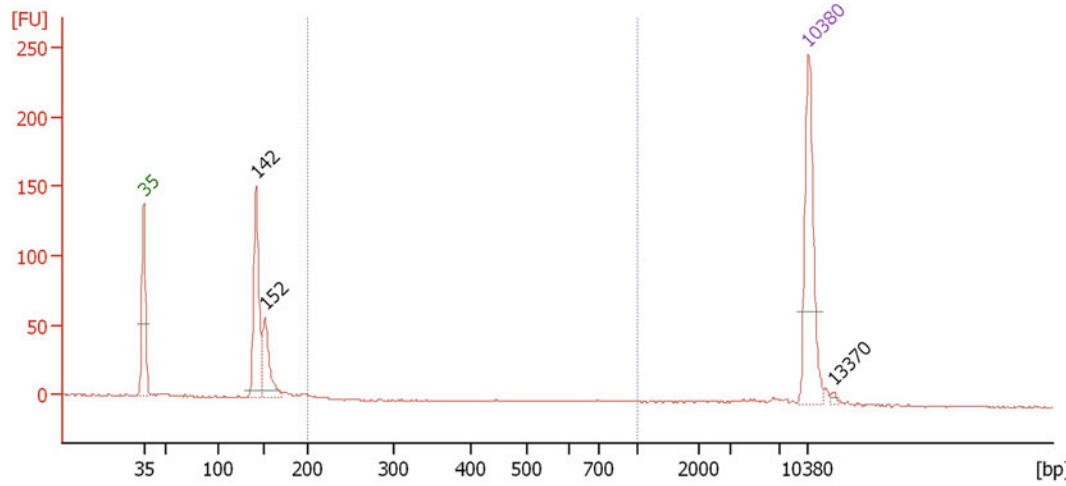

Fig. 4 Profile of a small RNA library prepared from 750 ng of tumor-derived RNA and run on an Agilent 2100 using a High Sensitivity DNA chip after library size selection. Only the library fraction containing miRNAs is retained, represented by the peak at 142 bp

add 2 μL of diluted library (triplicate reactions). Run the PCR reaction as follows: 95 °C for 15 min, (95 °C for 5 s, 60 °C for 30 s, 72 °C for 1 s) × 45 cycles, followed by melting curve analysis.

16. Prepare an equimolar library pool based on relative qPCR concentrations of the individual libraries.
17. Quantify the resulting library using the Qubit DS DNA HS assay. An additional ethanol precipitation step may be required as for accurate Qubit measurement the concentration should be above 1.5–2 ng/μL.
18. Sequence the library at a final concentration of 1.2 pM on a NextSeq 500 using a NextSeq 500 mid or high output V2 kit (*see* **Note 6**).

3.2 RT-qPCR

1. Thaw RNA and reverse transcriptase mix on ice. Thaw other RT II kit components at room temperature (15–25 °C). It is important to keep RNA on ice at all times to prevent degradation. Mix each solution by flicking tubes and centrifuge briefly to collect liquid from the sides and then store on ice. Do not vortex RNA.
2. Dilute RNA samples to a concentration of 100 ng/μL (*see* **Note 7**). Sensitivity can be improved by increasing the amount of input RNA.
3. Prepare RT mix by combining 2 μL of HiFlex or HiSpec buffer with 1 μL of 10× miScript Nucleics mix, 1 μL miScript Reverse Transcriptase Mix, and 4 μL of nuclease-free water (*see* **Notes 8** and **9**). If processing multiple samples, distribute 8 μL of this mixture in separate tubes and add 2 μL of RNA to each tube.
4. Incubate the reverse transcription mixture at 37 °C for 60 min followed by 5 min at 95 °C to inactivate miScript Reverse Transcriptase Mix. Place on ice.
5. Dilute the RT product 22-fold by adding 210 μL nuclease-free water (*see* **Note 10**).
6. Thaw 2× QuantiTect SYBR Green PCR Master Mix, 10× miScript Universal Primer, 10× miScript Primer Assay and template cDNA at room temperature (15–25 °C). Mix the individual solutions and place on ice.
7. Prepare the PCR mix by combining 5 μL 2× Quantitect SYBR Green PCR Master mix, 1 μL of 10× Universal Primer, 1 μL of 10× Primer assay, and 2 μL of nuclease-free water per reaction. If preparing for multiple reactions simply multiply by the number of reactions +10%. Mix by pipetting up and down and briefly spin.
8. Dispense 9 μL PCR mix in the wells of the reaction plate and add 1 μL of diluted cDNA to each reaction well. Seal reaction

wells carefully and centrifuge for 1 min at 1000 × *g* to remove bubbles.

9. Run the PCR reaction as follows: 95 °C for 15 min, (94 °C for 15 s, 55 °C for 30 s, 70 °C for 30 s) × 40 cycles, melting curve analysis.

3.3 Normalization of RT-qPCR Data

1. The normalized relative quantity for miRNA i in sample j is defined as:

$$\mathrm{NRQ}_{i,j} = 2^{\left(\mathrm{Cq}_{i,j} - \mu_j\right)},$$

with μ corresponding to either the global mean expression value or the arithmetic mean of multiple stable reference miRNAs (*see* **Note 11**), assuming 100% PCR efficiency. The qbase + software (http://www.qbaseplus.com) is particularly well suited for qPCR data analysis, including global mean or multiple reference gene normalization, and PCR efficiency correction if needed.

2. To identify stably expressed reference miRNAs: import normalized miRNA expression data into a spreadsheet like MS Excel. Calculate the standard deviation for each miRNA and select candidate miRNAs that have the lowest standard deviation, expressed in all samples and do not belong to the same miRNA family (*see* **Note 12**). Select between five and eight miRNAs as candidate reference genes. Verify in an RT-qPCR experiment that these candidate reference miRNAs are stably expressed. This means they should have low M values when using the geNorm algorithm (*see* **Note 13**).

3. In case you do not have access to miRNA-profiling data, we recommend to sequence a few representative samples followed by the procedure described above.

4. In case you do not have access to miRNA-profiling data and **step 2** is not an option, you can set up a classic geNorm pilot experiment using published candidate miRNA reference genes. Typically, eight candidate references small RNAs are evaluated in at least ten representative samples. Use the geNorm algorithm to identify the most stably expressed reference genes.

4 Notes

1. DESeq2 is available as an R package at the Bioconductor depository (www.bioconductor.org).

2. We strongly encourage using the following annotation schemes when reporting miRNA findings: the miRNA sequence itself, the miRNA name in combination with the miRBase version, or

the miRNA accession number in combination with the miRBase version.

3. To guarantee successful miRNA quantification, it is evident that the small RNA fraction is retained after RNA isolation. Several commercial kits are available that enable the extraction of total RNA including the small RNA fraction. Make use of microfluidics-based electrophoresis systems such as the Bioanalyzer or the Experion to evaluate the presence of the small RNA fraction. We strongly advise to only include RNA samples of sufficient quality. When performing miRNA quantification of total cell-free RNA present in serum or plasma samples, designated RNA isolation kits are available from different vendors. We have good experience with the miRNeasy serum/plasma kit (Qiagen).
4. When preparing a mixture for multiple samples simultaneously we advise to always prepare mixture for an additional 10%.
5. In the case of very-low input samples (e.g., serum or plasma), the number of PCR cycles can be further increased up until 16.
6. To determine optimal sequencing depth, it is possible to perform a saturation analysis in a pilot experiment by sequencing a small number of representative samples. After standard data processing the R package subSeq (available at www.bioconductor.org) can be used to determine whether enough reads were generated to detect all relevant biological information, or whether it's possible to multiplex more samples and thus work with fewer reads [12]. We typically aim for ten million reads for fresh tissue or cellular RNA, and 15 million reads for FFPE tissue or body fluid samples.
7. Concentration is dependent on the abundance of the mature miRNA target, ensure between 10 ng and 2 μg of RNA as input for the RT reaction. After RT, samples should be diluted to ensure between 25 pg and 1.5 ng of cDNA per PCR reaction.
8. In the miScript II RT kit, two buffers are included: the miScript HiSpec Buffer and the miScript HiFlex Buffer. The HiSpec Buffer is specifically formulated to facilitate the selective conversion of mature miRNAs into cDNA. The HiFlex buffer promotes the conversion of all RNA species into cDNA to enable combined study of miRNA and other RNA species like mRNA.
9. Besides dNTPs, rATP, and oligo-dT primers, the miScript Nucleics Mix contains an internal synthetic RNA control, the miRNA reverse transcription control (miRTC), that can be used to assess reverse transcription performance (i.e., absence of inhibitors).
10. To enable miRNA profiling studies of single cells and body fluids we advise to include a limited-cycle preamplification step

to increase the sensitivity of the RT-qPCR reaction. In the preamplification procedure, miRNA specific forward primers and universal reverse primers are used to amplify the cDNA template in a limited-cycle PCR, typically 12–15 cycles.

11. Baseline and threshold settings should be carefully evaluated when determining Cq-values. Typically, the baseline should be set to the cycle interval where no amplification takes place. The threshold is set, with the Υ-axis in log-scale, where all assays are in log linear phase.
12. miRNA families can be inspected in a dedicated miRBase file (ftp://mirbase.org/pub/mirbase/CURRENT/miFam.dat.gz).
13. Besides the tissue or disease type, the stability of candidate reference miRNAs also depends on the experimental conditions (e.g., treatment of the cells with siRNA or compound). We therefore advise to verify the stability of reference miRNAs when changing experimental conditions by measuring their expression on a representative selection of samples followed by geNorm analysis.

References

1. Mestdagh P, Hartmann N, Baeriswyl L et al (2014) Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. Nat Methods 11:809–815. doi:10.1038/nmeth.3014
2. Tian G, Yin X, Luo H et al (2010) Sequencing bias: comparison of different protocols of MicroRNA library construction. BMC Biotechnol 10:64. doi:10.1186/1472-6750-10-64
3. Baran-Gale J, Kurtz CL, Erdos MR et al (2015) Addressing bias in small RNA library preparation for sequencing: a new protocol recovers MicroRNAs that evade capture by current methods. Front Genet 6:1506. doi:10.1093/nar/gkr1263
4. Garmire LX, Subramaniam S (2012) Evaluation of normalization methods in mammalian microRNA-Seq data. RNA 18:1279–1288. doi:10.1261/rna.030916.111
5. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:31. doi:10.1186/s13059-014-0550-8
6. Benes V, Castoldi M (2010) Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. Methods 50:244–249. doi:10.1016/j.ymeth.2010.01.026
7. Shi R, Chiang V (2005) Facile means for quantifying microRNA expression by real-time PCR. Biotech 39:519–525. doi:10.2144/000112010
8. Vandesompele J, De Preter K, Pattyn F et al (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:RESEARCH0034
9. Mestdagh P, Van Vlierberghe P, De Weer A et al (2009) A novel and universal method for microRNA RT-qPCR data normalization. Genome Biol 10:R64. doi:10.1186/gb-2009-10-6-r64
10. D'haene B, Mestdagh P, Hellemans J, Vandesompele J (2011) miRNA expression profiling: from reference genes to global mean normalization. In: Methods in molecular biology. Humana Press, Totowa, NJ, pp 261–272
11. Van Peer G, Lefever S, Anckaert J et al (2014) miRBase tracker: keeping track of microRNA annotation changes. Database 2014:bau080. doi:10.1093/database/bau080
12. Robinson DG, Storey JD (2014) subSeq: determining appropriate sequencing depth through efficient read subsampling. Bioinformatics 30:3424–3426. doi:10.1093/bioinformatics/btu552