Decreasing miRNA sequencing bias using single adapter and circularization approach

Sergio Barberan-Soler, Jenny M. Vo, Ryan E. Hogans, Anne Dallas, Quincy Okobi, Brian H. Johnston, Sergei A. Kazakov
SomaGenics, Inc., 2161 Delaware Ave., Santa Cruz, CA 95060 USA

Abstract

The ability to accurately quantify all the microRNAs (miRNAs) present in a sample is important for understanding miRNA biology as well as for development of new miRNA biomarkers and therapeutic targets. We have developed a new method for preparing miRNA sequencing libraries (RealSeq -AC) that involves ligating the miRNAs with a single adapter and circularizing the ligation products. When compared to other known methods, RealSeq -AC provided greatly reduced miRNA sequencing bias and allowed the identification of the largest variety of miRNAs in biological samples. This significantly reduced bias also allows robust quantification of miRNAs present in samples across a wide range of RNA input levels. RealSeq -AC sequencing libraries from as low as 1 ng of total RNA can be prepared using a gel-electrophoresis-free protocol. Recent findings show that cell-free miRNAs found in biofluids have great potential as biomarkers for several diseases. One of the main problems when detecting cell-free miRNAs is that their concentration in biofluids is several orders of magnitude lower than that of tissues or cells. To address this problem, we developed a new method, RealSeq -BF, focused in the detection of miRNAs and other small RNAs from biofluids. This method shares the same approach with RealSeq -AC allowing the accurate quantification of miRNAs from biofluid samples (plasma, serum and urine).

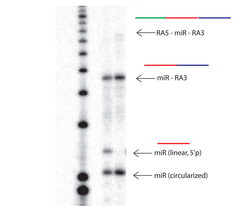
The problem

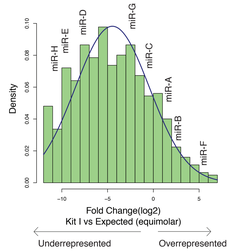
- Current market leaders in small RNA library preparation kits have large bias in detection of miRNAs due to inefficient ligation steps
- This bias in detection affects both absolute and relative quantification
- There is a need for an accurate small RNA library preparation kit

The solution

- By ligating miRNAs with a proprietary single-adapter and altered enzymatic steps RealSeq®-AC achieves a highly efficient ligation reaction that drastically reduces sequencing bias
- RealSeq®-AC is both accurate and sensitive, requiring fewer PCR cycles to achieve the same levels of detection

Problematic two-adapter ligation scheme





RealSeq®-AC protocol: Single-tube and gel-free

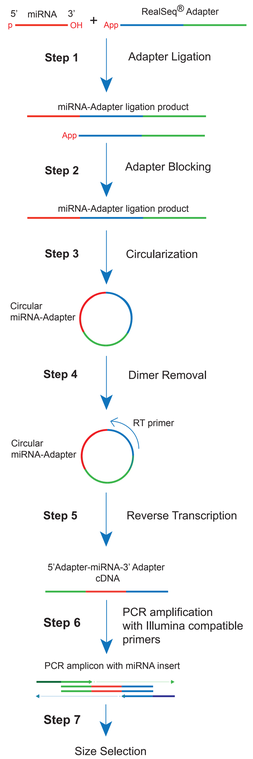


Figure 1. Two-adapter library preparation efficiency of ligation. Upper panel: efficiency of ligation obtained with Kit I using a two-adapter scheme for a single miRNA. Ligation of the second adapter (RA5) is highly inefficient, with less than 5% yield of ligated product. Using Kit I we prepared a sequencing library using 1 pmole of miRXPlore Universal Reference pool (Miltenyi Biotec) containing equimolar amounts of 963 synthetic miRNAs. This library was sequenced with a MiSeq instrument and the bias in detection of miRNAs was calculated (Lower panel).

Sequencing statistics

	Kit I	Kit N	Kit B	Kit T	RealSeq®-AC
Usable reads	97%	97.8%	98.8%	96.7%	93.2%
miRNAs with >=5 reads	404	452	412	324	500
miRNAs with >=10 reads	328	365	352	239	385

Table 1. Reads statistics for human brain RNA sequencing libraries. Raw reads were trimmed of adapter dimers and inserts shorter than 5 nt to obtain usable reads. The average percentage of usable reads for triplicate libraries is shown. Trimmed reads were aligned against miRbase 21 and the number of miRNAs detected with coverage of ≥5 reads or ≥10 reads is shown. RealSeq®-AC identifies the greatest number of miRNAs with 500 at a coverage of ≥5 reads, and 385 with a coverage of ≥10.

RealSeq®-AC

Total RNA samples from 1000-1 ng

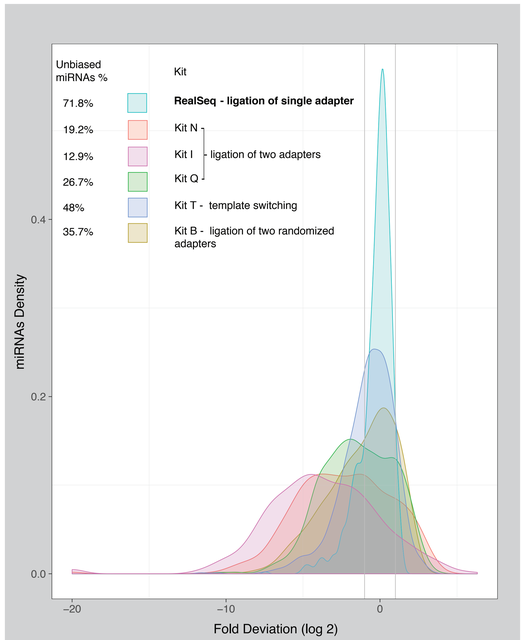


Figure 2. Distribution of sequencing reads from various small-RNA library preparation kits using a universal miRNA reference pool. All sequencing libraries were prepared in triplicate using 1 pmole of miRXPlore Universal Reference pool (Miltenyi Biotec), which contains equimolar amounts of 963 synthetic miRNAs. Purified libraries were sequenced on an Illumina MiSeq instrument. Sequencing reads were trimmed of adapter sequences using Cutadapt (Martin et al. 2011) and trimmed reads were aligned to a custom miRNA reference using Bowtie2 (Langmead et al. 2012). Reads mapping to miRNAs were counted using a custom script. Fold changes were calculated assuming an equimolar representation of miRNAs in the synthetic pool; fold changes are plotted as log2 values. The density of miRNAs within a two-fold deviation from the expected values (area between vertical lines) were considered unbiased according to Fuchs et al (2015)

Identification of novel miRNAs

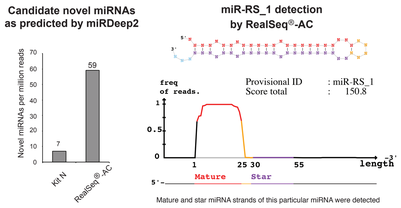


Figure 3. Discovery of novel miRNAs from MCF7 cells. We prepared sequencing libraries from 1 µg of total RNA from MCF7 cells with either Kit N or RealSeq®-AC. Fastq files were processed with CAP-miRSeq (Sun et al. 2014) and the identification of novel miRNAs by miRDeep2 (Friedlander et al. 2012) is shown. Results show that even in a well-characterized cell-line as MCF7, RealSeq®-AC identifies novel miRNAs not detected with two-adapter ligation schemes.

Conclusions RealSeq-AC

- Currently available small RNA library preparation kits introduce a bias in detection of miRNAs of up to 10,000x
- Bias is introduced during ligation steps, with most kits including a highly inefficient 5' adapter ligation
- By modifying the ligation scheme to use a single-adapter and altered enzymatic steps, RealSeq®-AC minimizes the bias in detection
- This highly accurate quantification is evident when profiling both synthetic reference pools and when quantifying total RNA samples
- RealSeq®-AC protocol is single-tube, gel-free procedure and can be performed in under six hours
- Its high efficiency reduces the number of PCR cycles needed to only five cycles for an input of 1 µg total RNA
- RealSeq®-AC is optimized for inputs between 1 µg to 1 ng of total RNA; however when available 100 ng input is recommended

Quantification of human brain reference RNA

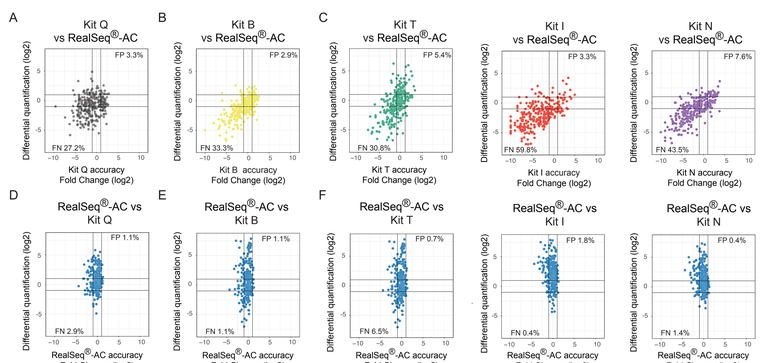
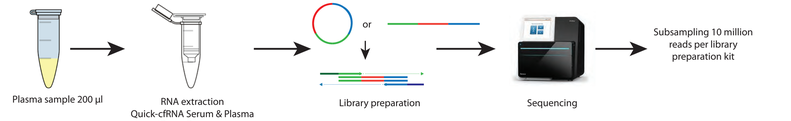


Figure 4. Quantification accuracy for brain miRNAs for each tested library preparation kit. A reference brain RNA sample was used to prepare libraries in triplicate for each kit. Purified libraries were sequenced on an Illumina MiSeq instrument. Sequencing reads were processed as described in Figure 2, except that read counts were only determined for a set of high-confidence human miRNAs (Kozomara et al 2014). This set of human miRNAs is also included in the pool of synthetic miRNAs used for Figure 2. This inclusion allows us to use the determined accuracy of detection for a synthetic miRNA (Figure 3 data) to compare the different quantification for a brain sample between the library preparation kits. The scatterplot for each kit shows the differential quantification values between each kit and a reference kit (Y axis) and also includes the accuracy of detection for each kit as calculated for the synthetic pool of miRNAs (X axis). False positive/negative miRNAs were determined as miRNAs with an increase/decrease in signal for a given kit (>(+/-) 2 fold compared to RealSeq®-AC), when that same miRNA was also shown to have a significant bias (>(+/-)2 fold) with the given kit.

RealSeq®-Biofluids

Gel -free extracellular small RNA profiling from as low as 50 µl plasma



- Plasma and other biofluid samples contain extremely low concentrations of cf-miRNAs
- Accurate and sensitive quantification of cf-miRNAs from biofluids requires different reaction conditions compared to tissue samples
- Gel-free detection is a must for reproducible and automatable biomarker discovery pipelines
- RealSeq-Biofluids capitalizes on the accuracy of RealSeq-AC while sensitive enough to allow gel-free detection of cf-miRNAs from biofluids

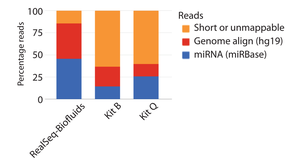
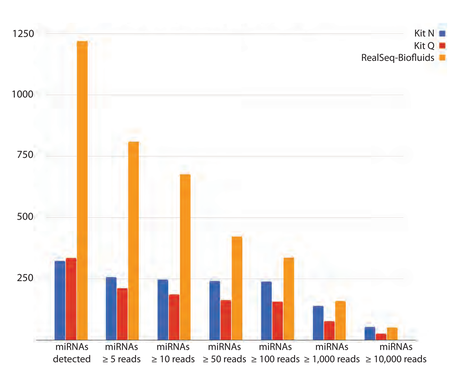
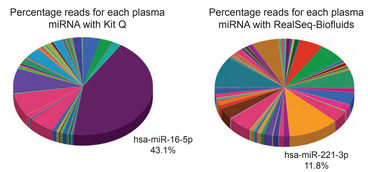


Figure 5. Outline of library preparation from plasma samples. The same plasma sample was used to prepare sequencing libraries with three different kits. Lower right panel shows sequencing metrics of libraries prepared with each kit. Short reads correspond to reads <15 nt after adapter trimming; reads passing this filter are then align to a reference file with all human miRNAs (miRBase 21), reads that do not align to miRNAs are then aligned to the human genome (hg19).

Detection of plasma miRNAs with different library preparation kits





- Kit Q identifies miR-16-5p as the most abundant plasma miRNA with 43.1% of the reads
- Profiling of miR-16-5p with miRXplore pool (Figure 2) shows that Kit Q over-represents this miRNA with >5 fold bias
- Comparatively, RealSeq-Biofluids identifies miR-221-3p as the most abundant plasma miRNA with 11.8% of the reads
- Profiling of miR-221-3p with miRXplore pool (Figure 2) shows that RealSeq detects this miRNA with no bias

Figure 6. Profiling of plasma miRNAs with three different library preparation kits. 200 ul of plasma sample from a healthy donor was used to extract RNA with Quick-cfRNA Plasma/Serum kit (Zymo research) following manufacturer recommendations. RNA from three extractions was pooled and used to prepare sequencing libraries with the three kits following manufacturer recommendations for gel-free libraries. To normalize for sequencing coverage reads were subsampled to 10 million reads per kit. Sequencing reads were processed as Figures 2-3, except that reads were aligned to a reference that includes all human miRNAs in miRBase 21. The left panel shows the number of miRNAs detected at different coverage for each library preparation kit. The right panel shows the percentage of plasma reads for each miRNA with kits Q and RealSeq-Biofluids.

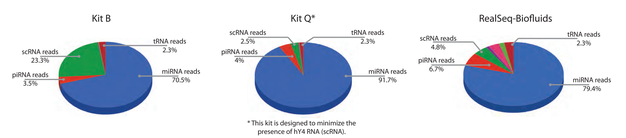


Figure 7. Percentage of reads that map to different classes of ncRNAs for each library preparation kit. Kit Q, according to the manufacturer, is specifically designed to remove reads mapping to HY4 RNA (scRNA) impeding its quantification.

Conclusions RealSeq-Biofluids

- RealSeq-Biofluids allows preparation of gel-free sequencing libraries with an RNA input obtained from only 50 µl of plasma
- RealSeq-Biofluids delivers the highest percentage of usable reads (>15 nt and that align to either miRBase or genome) of the 3 kits tested
- Highly accurate profiling allows the identification of a larger set of cf- miRNAs (Figure 6, left panel)
- Detection bias inherent in the two-adapter platforms reduces the number of miRNAs identified (Figure 6, left panel)
- Detection bias also results in the overrepresentation of a few miRNAs that consume the majority of sequencing reads (Figure 6)
- scRNAs are overrepresented in libraries prepared with the two-adapter ligation scheme (Figure 7), while they represent only 4% of the reads for RealSeq-Biofluids (single-adapter and circularization)
- RealSeq-Biofluids allows accurate and sensitive quantification of cell-free miRNAs with a gel-free protocol

• Acknowledgement: This project was supported in part by NIH R43HG007788, R44HG007788 to S.A.K.; and R43GM115124, R44GM115124 to S.B.S.