RealSeq®-AC, a novel method that greatly reduces bias in preparation of small-RNA sequencing libraries

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

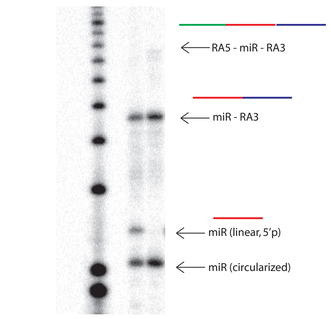
Abstract

Because dysregulation of miRNA expression has been implicated in cancer and other diseases, accurate expression profiling of all microRNAs (miRNA) and other small RNAs is important for understanding their biology and for development of new biomarkers and therapeutic targets. Small RNA-Seq is currently the most comprehensive approach for discovery and expression profiling of small RNAs. However, current methods for small RNA-Seq underestimate the abundance of most miRNAs in a sample, some by as much as 10,000-fold. Most of this misrepresentation results from low efficiency and sequence-specific bias in the intermolecular ligation of adapters to the two ends of the miRNA, the approach used in the most available kits for preparation of sequencing libraries. To address this problem, we have developed a new method, called RealSeq™-AC. By efficiently ligating a single adapter and circularizing miRNA-adapter ligation products, our method greatly reduces library preparation bias. In sequencing a reference pool containing equimolar concentrations of 962 synthetic miRNAs, we found that the percentage of miRNAs that appeared in their correct proportions (within 2-fold representation) in Illumina sequencing was ~72% for RealSeq ™-AC, significantly higher than other currently available kits for preparation of miRNA sequencing libraries. This reduction in bias allows for accurate quantification of miRNAs (and other small RNAs) and, by being more sensitive to the presence of otherwise poorly detected or undetected miRNAs, enables the discovery of novel small RNAs and the sequencing of lower RNA inputs (e.g., from plasma). Additionally, lower bias allows more samples to be sequenced in parallel (with barcoding) for the same cost. Profiling of miRNAs from a reference human brain RNA sample by Illumina sequencing showed that, among the tested small-RNA library preparation kits, the relative quantification by RealSeq™-AC provided the best correlation with RT-qPCR analysis. RealSeq ™-AC has a simple work-flow that does not require a gel-purification step. It could also be used to prepare sequencing libraries for fragmented large RNAs, so that small RNAs and large RNA could be sequenced simultaneously.

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

Problematic two-adapter ligation scheme



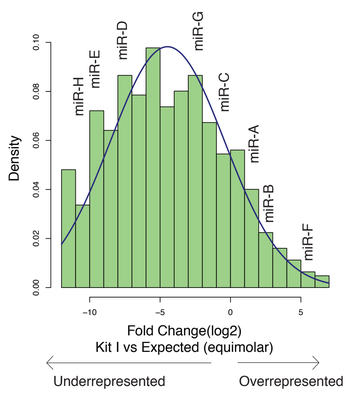
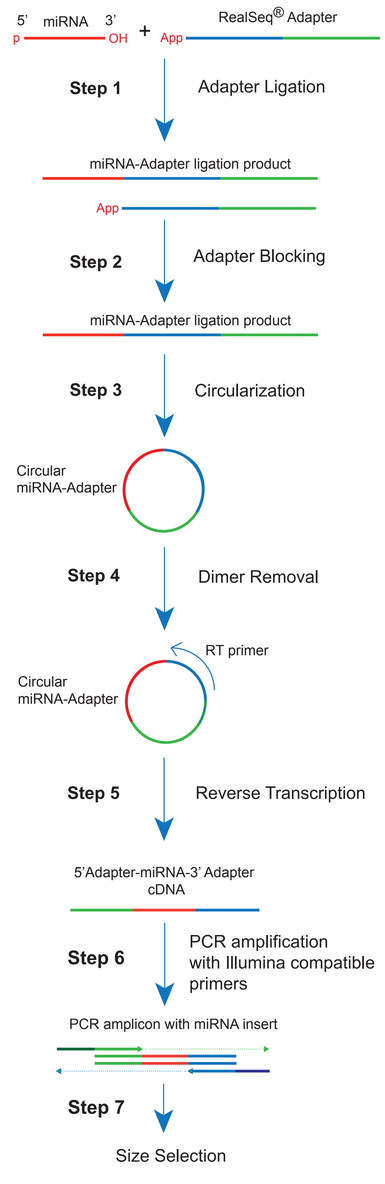


Figure 1. Two-adapter library preparation efficiency of ligation. Upper panel: efficiency of ligation obtained with Kit I using a two-adapter scheme for a single miRNA. Ligation of the second adapter (RA5) is highly inefficient, with less than 5% yield of ligated product. Using Kit I we prepared a sequencing 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).

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

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



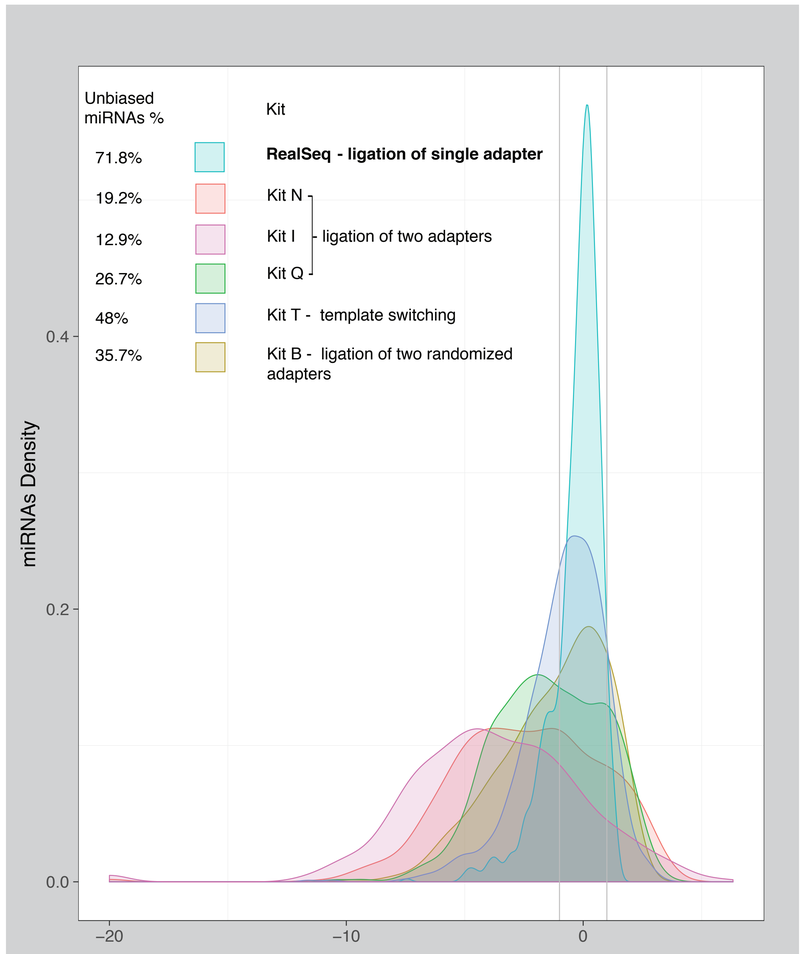


Figure 2. Distribution of sequencing reads from various small-RNA library preparation kits using a universal miRNA reference pool. All sequencing libraries were prepared in triplicate using 1 pmole of miRXPlore Universal Reference pool (Miltenyi Biotec), which contains equimolar amounts of 963 synthetic 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

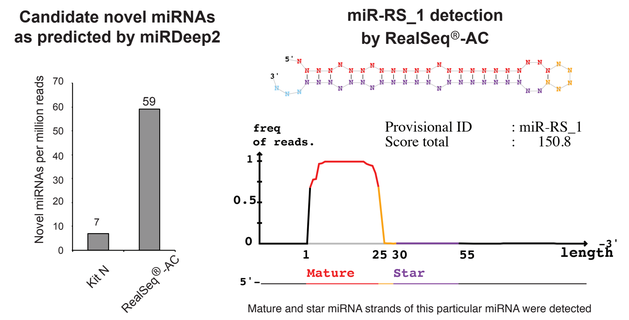


Figure 3. Discovery of novel miRNAs from MCF7 cells. We prepared sequencing libraries from 1 µg of total RNA from MCF7 cells with either Kit N or RealSeq®-AC. Fastq files were processed with CAP-miRSeq (Sun et al. 2014) and the identification of novel miRNAs by miRDeep2 (Friedlander et al. 2012) is shown. 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

- 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 100 ng of total RNA; however a lower input version is under development

- Acknowledgement: This project was supported in part by NIH R43HG007788-01A1 and 2R44HG007788-02A1 to S.A.K.
- RealSeq® technology is covered by issued and pending patents.

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.

Quantification of human brain reference RNA

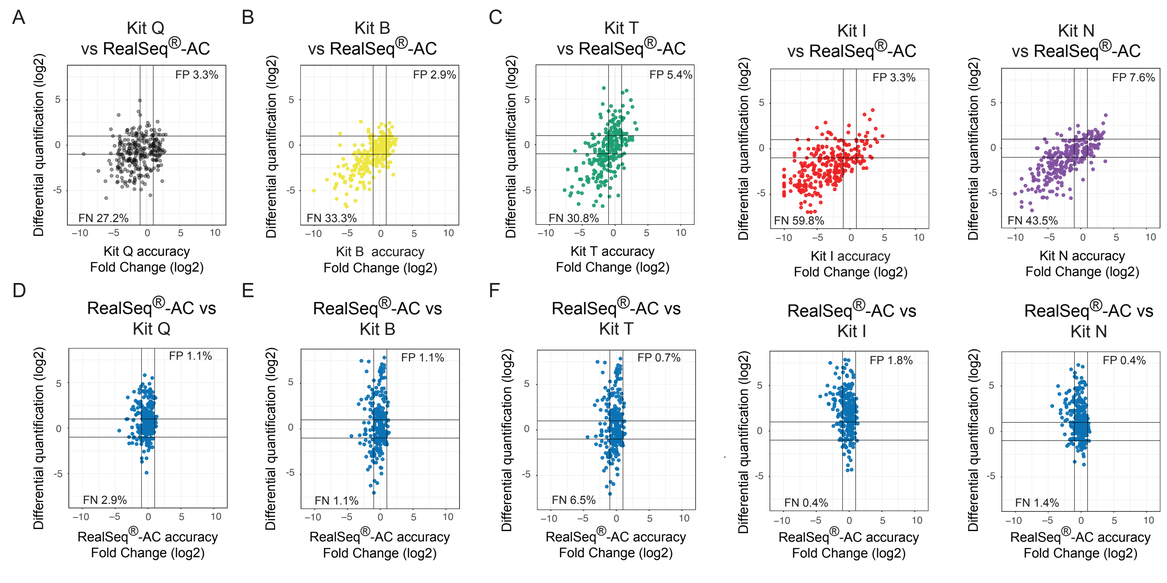


Figure 4. Quantification accuracy for brain miRNAs for each tested library preparation kit. A reference brain RNA sample was used to prepare libraries in triplicate for each kit. Purified libraries were sequenced on an Illumina MiSeq instrument. Sequencing reads were processed as described in Figure 3, 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 3. 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.