An unbiased and highly reproducible method for constructing microRNA NGS libraries for accurate expression profiling

Marta González-Hernández[1]*, Suvarna Gandlur[1], Amanda McNulty[1], Konstantinos Charizanis[2], Jing Ning[3], Matthew Carroll[1], Emmanuel Kamberov[1], and Karl Hecker[1]

[1]Takara Bio USA, Inc., Mountain View, CA 94043, USA • [2] Gemphire Therapeutics Inc., Livonia, MI, 48152, USA • [3] QIAGEN Sciences, Inc., Germantown, MD 20874, USA • *Corresponding author: marta_gonzalez@takarabio.com

Abstract

miRNA sequencing (miRNA-seq) is a useful tool for aiding researchers in the examination of miRNA expression patterns, the characterization of novel miRNAs, and for uncovering miRNA disease associations. Since miRNAs are also unusually well-preserved in a range of specimens (e.g., urine, FFPE tissue, plasma), profiling their expression could become a powerful diagnostic tool. However, current methods for sequencing miRNA require large amounts of total RNA, are not very reproducible, and more importantly, have considerable systematic bias resulting in loss of many prospective biomarkers. This bias severely affects the trustworthiness of results as libraries are not a true representation of the biological state of the sample.

We have recently developed the SMARTer® microRNA-Seq Kit, which uses MAGIC (**M**ono-**A**dapter li**G**ation and **I**ntramolecular **C**ircularization) technology to efficiently capture miRNA species with extremely low bias. Libraries prepared using an equimolar mixture of 963 miRNAs, sequenced on Illumina® platforms and analyzed for read distribution reveal that >70% of miRNAs captured fall within a +/– 2-fold variation of the expected read number they should receive. This means that the expression level of ~70% of miRNAs in the equimolar mixture was truly and accurately represented. In contrast, frequency distribution analyses for kits from competitors (NEB, Illumina, Bioo, and QIAGEN) revealed that 49–79% of miRNAs are greatly underrepresented (i.e., less than 2X fewer reads than expected), 13–35% of miRNAs are represented within a +/– 2-fold variation of the expected read number, and around 8–16% are overrepresented by more than 2X. These findings highlight the importance of understanding the current technical state of miRNA sequencing technologies to better prepare for analyzing and validating miRNA expression data. Additionally, the SMARTer microRNA-Seq Kit is designed and developed to more accurately reflect the true biological state of a sample, which will be an important factor as miRNA research moves toward diagnostic tools specific for personalized medicine.

Introduction

miRNA-seq is a useful tool for the profiling of expression patterns, establishing disease associations, and detecting and identifying novel species in health and disease. Like many other miRNA profiling tools, currently practiced miRNA-seq protocols are not able to accurately represent miRNA expression levels, since these have been shown to be dependent on sequence, nature of miRNA modification, as well as library preparation conditions (Leshkowitz et al. 2013; Fuchs et al. 2015; Raabe et al. 2014; Van Dijk, Jaszczyszyn, and Thermes 2014; Shore et al. 2016). The inaccuracy of miRNA level representation by next-generation sequencing (NGS) is predominantly due to systematic biases introduced during the adapter ligation, but is also affected by factors such as cDNA synthesis and PCR amplification.

We have developed the SMARTer microRNA-Seq Kit, based on MAGIC technology, which affords a more accurate reflection of the true biological state of a sample as expressed in the miRNA or small RNA expression pattern, allowing for the development of more precise analytical tools for diagnostics and personalized medicine.

1 MAGIC technology

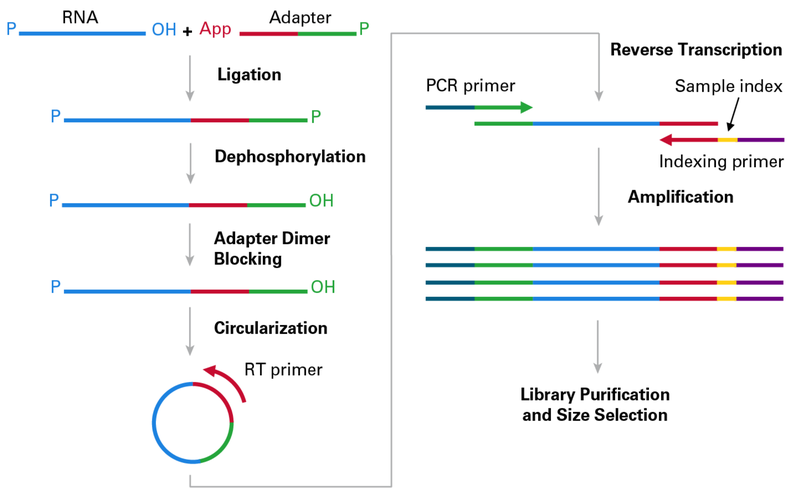

Figure 1. Schematic of MAGIC technology used by the SMARTer microRNA-Seq Kit. miRNA molecules containing 5'-phosphate and 3'-hydroxyl groups are first ligated to an adenylated adapter that is blocked at the 3'-end by a phosphate group. Ligated molecules are then dephosphorylated, circularized, reverse transcribed, and amplified in a PCR reaction. The amplified library is finally purified and size-selected, and the resulting library is ready for sequencing on an Illumina NGS platform.

2 SMARTer microRNA-Seq Kit workflow

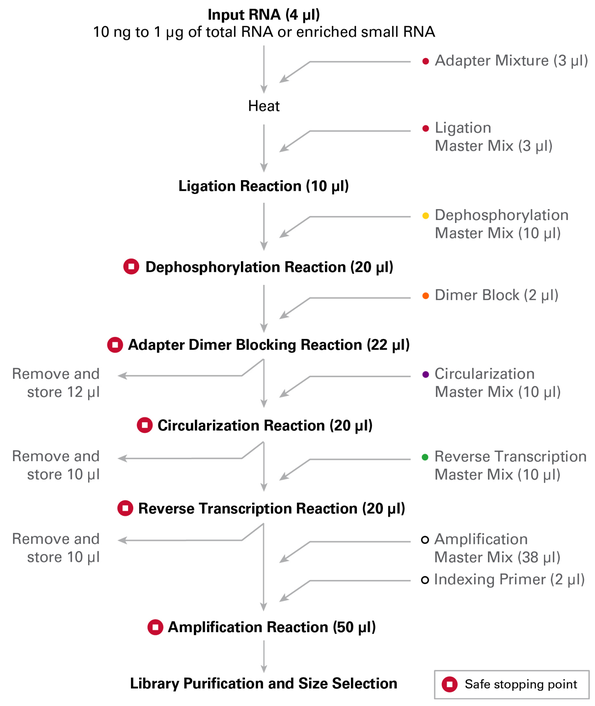

Figure 2. Library preparation workflow for the SMARTer microRNA-Seq Kit. This kit employs a single-tube workflow, which can be completed in a single day (about six hours) to generate amplified libraries from input RNA.

3 Accurate representation of microRNA expression

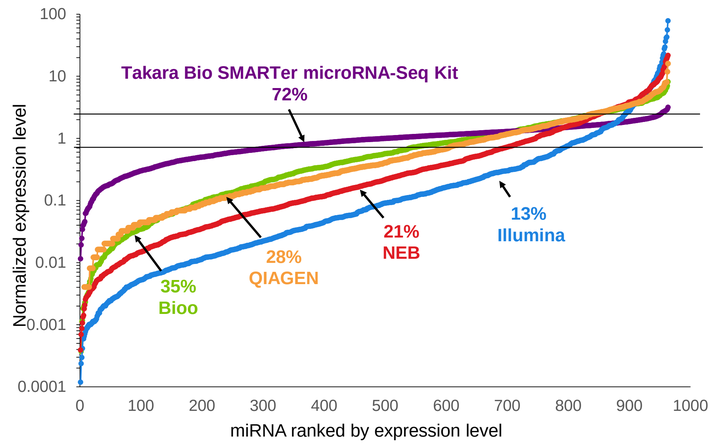

Figure 3. miRNA libraries generated with the SMARTer microRNA-Seq Kit are more accurate than other state-of-the-art ligation-based technologies. Libraries were generated from 10 ng of an equimolar pool of 963 synthetic miRNAs (Miltenyi miRXplore Universal Reference, Cat. No. 130-093-521) using our SMARTer library preparation protocol as well as different available technologies (Bioo, Illumina, NEB, and QIAGEN). Following sequencing, mapping (using a custom reference containing the 963 sequences in the pool), counting and normalization of reads (where equal expression level is assigned as 1), miRNA expression levels (Y axis, log scale) were plotted against miRNA species (ranked along the X-axis in order of expression level). A 2-fold cutoff was assigned both above and below the expected expression level (indicated by two horizontal lines). Of note, libraries prepared with 1 ng input amount show identical results. Additionally, libraries from the QIAGEN kit failed to capture 6 of the 963 miRNAs, even though it utilizes unique molecular identifiers (UMIs).

I MicroRNAs captured with a low bias

Average % miRNAs within 2-fold	Average % miRNAs within 3-fold	Average % miRNAs within 5-fold
72.1	84.7	91.3

Table I. More than 70% of miRNAs in a sample are captured within a small variation of their expected expression levels using the SMARTer microRNA-Seq Kit. Small RNA libraries prepared using 10 ng of a synthetic equimolar pool of 963 miRNAs (Miltenyi miRXplore Universal Reference) were sequenced, processed, and mapped using a custom reference containing all 963 sequences in the pool (allowing 1 mismatch). Fold changes were calculated assuming an equimolar representation of miRNAs in the pool using a custom Python script. Shown are the average percentages of miRNAs that fall within the stated fold range of the expected number of reads each miRNA in the equimolar pool should receive.

4 Synthetic microRNA as a functional control and for assessing bias

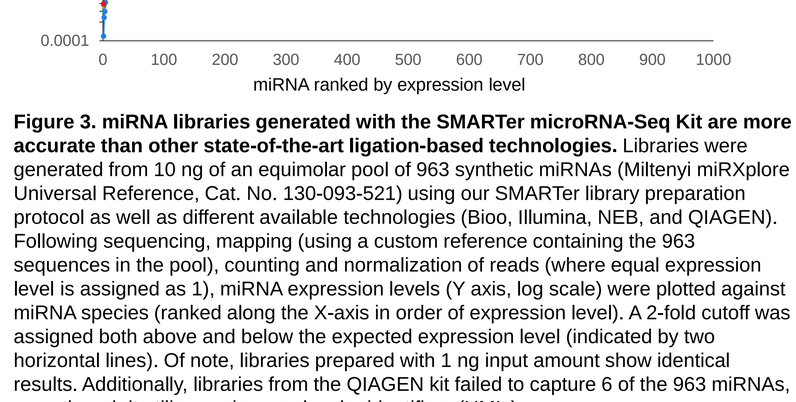

Figure 4. Using a high-diversity pool of miRNAs as a control for library preparation can allow the user to verify that reactions were successful as well as to determine their bias. The functional control is a pool of 4,096 synthetic miRNAs at equimolar quantity. The sequence is derived from human *let-7d-3p*, but with randomized ends (5' P-NNNUACGACCUGCUGCCUUNNN 3') (Shore et al. 2016). Libraries were prepared using 10 ng of input following our library preparation protocol. After gel purification (**Panel A**) and library QC, samples were sequenced on an Illumina platform, and sequences with an exact match were used to calculate fold change assuming an equimolar representation of miRNAs in the pool after normalization of the reads (reads per million). **Panel B.** The bias of the reads was analyzed and shown as representative data of the percentage of miRNAs detected within a 2-fold deviation from the expected read value (70%; area between the vertical lines). Analyses also show the percentage of miRNAs detected with read numbers below the expected read value (22%; area left of the two vertical lines) and above the expected read value (8%; area right of the two vertical lines).

5 More microRNA species detected

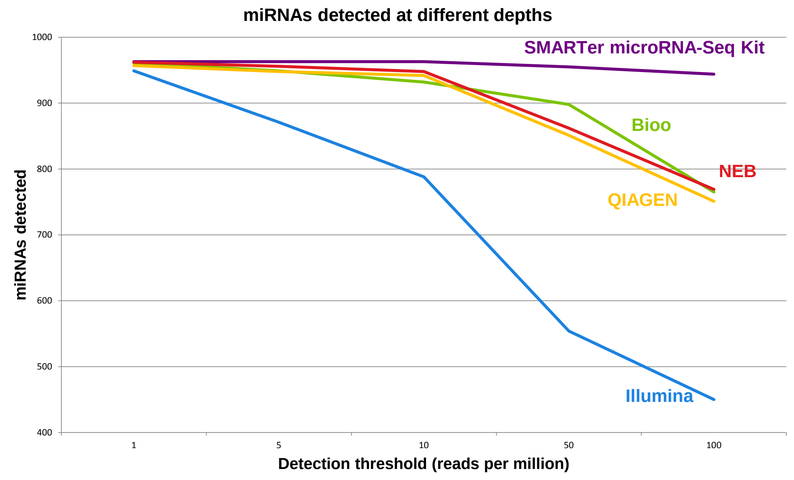

Figure 5. The SMARTer microRNA-Seq Kit detects more sequences in an equimolar pool of 963 synthetic miRNAs. Libraries were generated from 10 ng of an equimolar pool of 963 synthetic miRNAs (Miltenyi miRXplore Universal Reference, Cat. No. 130-093-521) using our SMARTer library preparation protocol as well as different available technologies (Bioo, Illumina, NEB, and QIAGEN). Following sequencing, mapping (using a custom reference containing the 963 sequences in the pool), counting, and normalization of reads, miRNAs detected were plotted against the detection threshold in reads per million.

II High performance across different DNA types

	RNA source							
	Placenta		Brain		Universal Human Reference		Spleen	
smRNA <200 nt (% of total RNA)	10–13		5		3		2	
Input amount	1 µg	100 ng	1 µg	100 ng	1 µg	100 ng	1 µg	100 ng
Total number of reads	3,860,186	4,590,518	4,114,895	4,057,076	4,644,428	4,359,698	4,298,758	3,800,238
% of reads trimmed	61.18	57.38	77.91	83.07	89.09	84.29	69.62	66.60
% of reads mapped to hg38	61.13	53.12	61.32	55.27	71.76	66.12	67.15	48.92

miRNA	Placenta		Brain		Universal Human Reference		Spleen	
% of total reads	28.39	13.97	29.53	21.05	30.38	24.98	25.95	15.84
Number of miRNAs detected	520	515	676	497	708	603	442	410

Other RNA (% of total reads)	Placenta		Brain		Universal Human Reference		Spleen	
smRNA (piRNA, snoRNA, snRNA)	14.81	9.41	15.81	8.34	13.82	6.50	19.49	11.32
rRNA	2.11	1.88	1.32	1.52	2.17	2.02	1.33	2.66

Table II. Evaluating the performance of the SMARTer approach across total RNA input types and amounts. Sequencing libraries were generated from 100 ng and 1 µg of human placenta, brain, or spleen total RNA and an Agilent Universal miRNA Human reference using the SMARTer microRNA-Seq Kit, and size-selected using the Pippin Prep. Following sequencing, trimming of adapter sequences, and removing reads shorter than 17 nt, the remaining reads were mapped to hg38 for overall mapping, and to miRBase.

6 Highly reproducible data

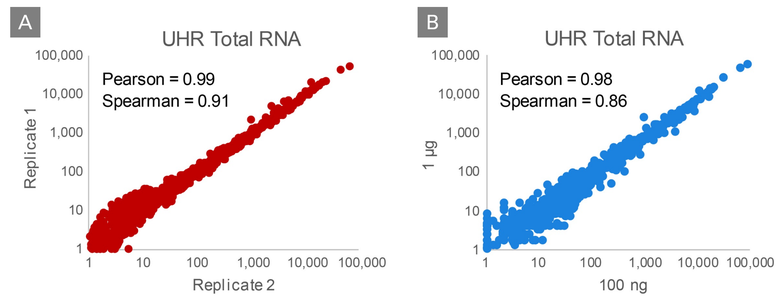

Figure 6. Replicate libraries and libraries prepared from different input amounts are extremely reproducible. Sequencing libraries were generated in parallel from 1 µg or 100 ng of total human RNA (Agilent Universal miRNA Human Reference Kit, UHR) using our library preparation protocol. Libraries were purified, size-selected, and checked for quality prior to sequencing on Illumina's NextSeq® 500 platform. Following sequencing, read processing, and mapping, expression levels of miRNAs identified for each library were quantified. Normalized reads were plotted on correlation diagrams, comparing either individual replicates or individual input amounts. **Panel A.** Correlation of miRNA expression levels for two technical replicates. **Panel B.** Correlation of miRNA expression levels for samples prepared from two different input amounts.

Conclusions

Takara Bio's SMARTer microRNA-Seq Kit has significantly less miRNA representation bias than state-of-the-art ligation-based technologies.

Advantages of Takara Bio's SMARTer microRNA-Seq Kit:

- Minimizes bias to provide better representation of the true miRNA expression profile
- Captures more miRNAs
- Generates highly reproducible results
- Includes miRNA oligos with randomized ends as a functional control to verify the success of library preparation and assess bias in representation. Libraries prepared from the functional control also serve as size markers for gel purification.

References

Fuchs, R. T., Sun, Z., Zhuang, F. & Robb, G. B. Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS One* **10**, (2015).

Leshkowitz, D., Horn-Saban, S., Parmet, Y. & Feldmesser, E. Differences in microRNA detection levels are technology and sequence dependent. *RNA* **19**, 527–538 (2013).

Raabe, C. A., Tang, T. H., Brosius, J. & Rozhdestvensky, T. S. Biases in small RNA deep sequencing data. *Nucleic Acids Res.* **42**, 1414–26 (2014).

Shore, S., Henderson, J. M., Lebedev, A., Salcedo, M. P., Zon, G., McCaffrey, A. P., Paul, N. & Hogrefe, R. I. Small RNA library preparation method for next-generation sequencing using chemical modifications to prevent adapter dimer formation. *PloS One* **11**, (2016).

Van Dijk, E. L., Jaszczyszyn, Y. & Thermes, C. Library preparation methods for next-generation sequencing: Tone down the bias. *Exp. Cell Res.* **322**, 12–20 (2014).

To download your copy of this poster, visit:
https://go.takarabio.com/ABRF-2018.html

800.662.2566
Visit us at takarabio.com

Clontech TaKaRa cellartis