TECH NOTE

Overcome ligation-induced bias and skewed miRNA representation in microRNA-seq

SMARTer microRNA-Seq Kit

- Even capture of microRNAs for an accurate microRNA expression profile
  Mono-adapter ligation and intramolecular circularization technology minimizes bias
- Higher sensitivity and accuracy than competitors' kits
  Illumina, NEB, Bioo, and QIAGEN approaches fall short of the SMARTer microRNA-seq method
- Reliably capture small RNA species across input sources and amounts
  Reproducible results provide a better reflection of the true biological state of a sample

Introduction

In cellular regulatory processes small, noncoding RNAs (between 15 to 200 nucleotides in length) have key roles in regulating the flow of genetic information (Phillips 2008). They are capable of regulating RNA splicing and protein translation, affect DNA replication through altering chromatin structure, mediate intercellular communication in a hormone-like fashion, and participate in genome defense (Bayraktar, Van Roosbroeck, and Calin 2017; Phillips 2008; Tao et al. 2017). While there are multiple types of small noncoding RNAs, the regulatory role of microRNAs—which act by binding to target mRNAs in a sequence-dependent fashion—is the most well known and studied.

MicroRNA sequencing is a useful tool for researchers to examine microRNA expression patterns, characterize novel microRNAs, and uncover disease-associated microRNAs. Since microRNAs are unusually well preserved in a range of specimens (e.g. urine, FFPE tissue), profiling their expression could serve as a powerful diagnostic tool (Stuopelyte et al. 2016; Kakimoto et al. 2016). However, most of the methods for sequencing microRNAs involve sequential intermolecular ligation events. These events introduce considerable systematic bias that results in the loss, or overrepresentation, of many prospective biomarkers and hence libraries that are not true reflections of the starting sample (Fuchs et al. 2015; Raabe et al. 2014).

We have recently developed the SMARTer microRNA-Seq Kit, which uses Mono-Adapter liGation and Intramolecular Circularization (MAGIC) technology to efficiently capture microRNA species with extremely low bias (Figure 1). Libraries are prepared via ligation but, instead of undergoing two sequential intermolecular ligation events, microRNAs receive a single combo adapter at their 3'-end. This adapter (after minimization of leftover, unligated adapter in the reaction) allows microRNAs to be flanked by two independent adapters through a circularization event with the 5'-end of the substrate molecules. Adapter-flanked circular microRNAs are reverse transcribed and the cDNA is PCR amplified using barcoded index primers for Illumina sequencing platforms.

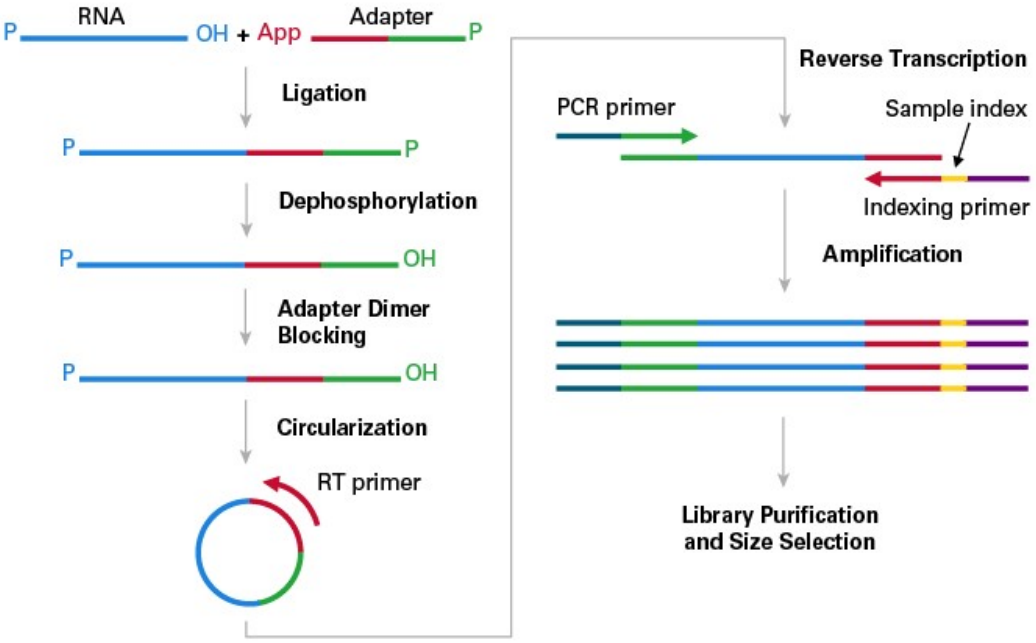

Figure 1. Schematic of Mono-Adapter liGation and Intramolecular Circularization (MAGIC) technology.

Following PCR amplification, libraries are purified using magnetic beads and their size profiles can be analyzed using an Agilent Bioanalyzer High Sensitivity chip (or similar instrument and consumables) to confirm that microRNA molecules were successfully captured and contain adapter molecules (Figure 2). As seen in the representative traces shown here, successful library preparations should show multiple peaks. Adapter dimers, while present, are not the predominant peaks and should be around 115–125 bp. They should be accompanied by slightly larger peaks in the range of 130–350 bp, with smears up to 800–1,000 bp. If only the single 115–125-bp peak is observed, the library preparation protocol did not work (i.e., the adapter dimer minimization step was faulty).

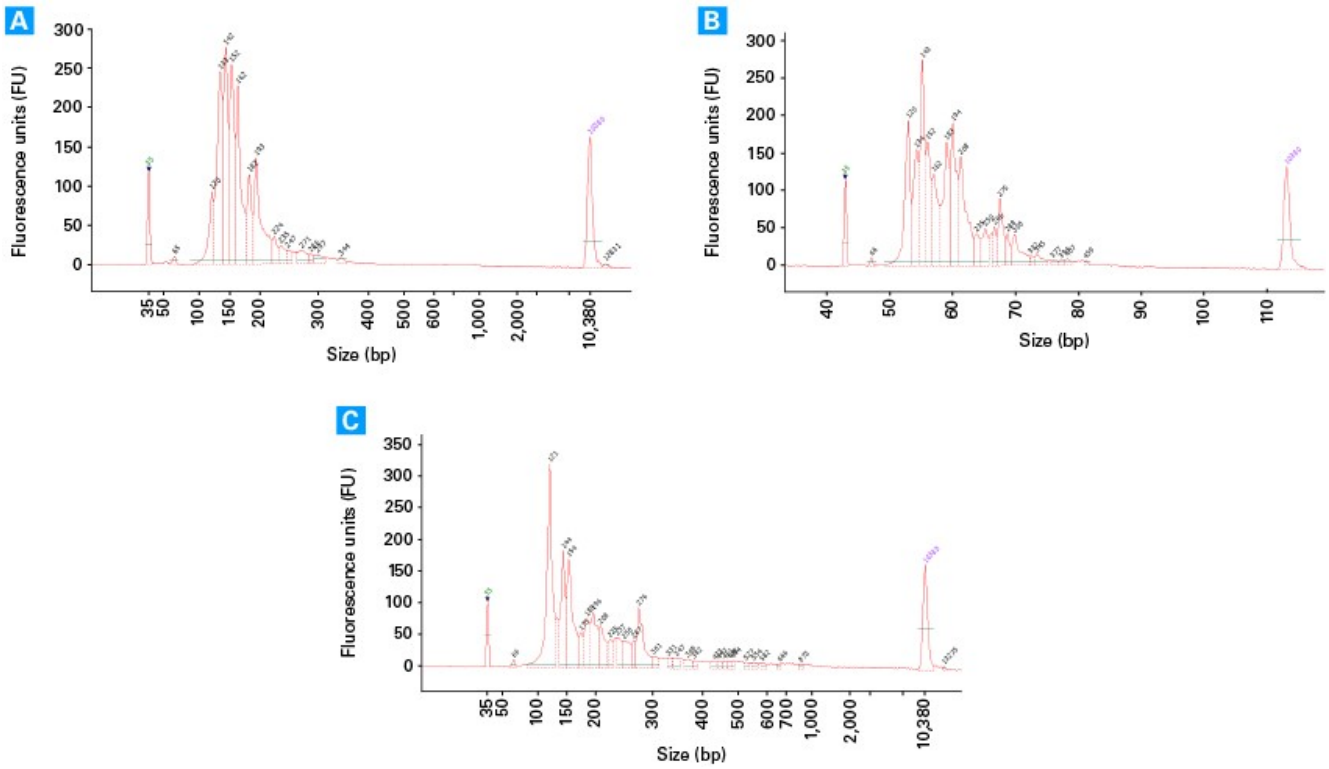

Figure 2. Example electropherogram traces of libraries generated with 100 ng human brain total RNA prior to size selection. The adapter dimer peak (115–125 bp) can vary in size compared to the library peak (140–150 bp), ranging from half the size (Panel A) to larger (Panels B and C). There may also be a smear of peaks up to 2,000 bp in size. All traces shown indicate effective ligation and adapter minimization due to the presence of peaks other than the adapter dimer peak.

Once samples have been confirmed to contain library molecules, a size-selection step is needed to remove adapter dimer contamination and to enrich for microRNA molecules. Two options are provided: size selection using the Pippin Prep (Figure 3, upper panel) or size selection using polyacrylamide gel electrophoresis (Figure 3, lower panel). The size selection will enrich for molecules around 150 bp. Following validation of this process, the libraries are ready for sequencing on Illumina platforms.

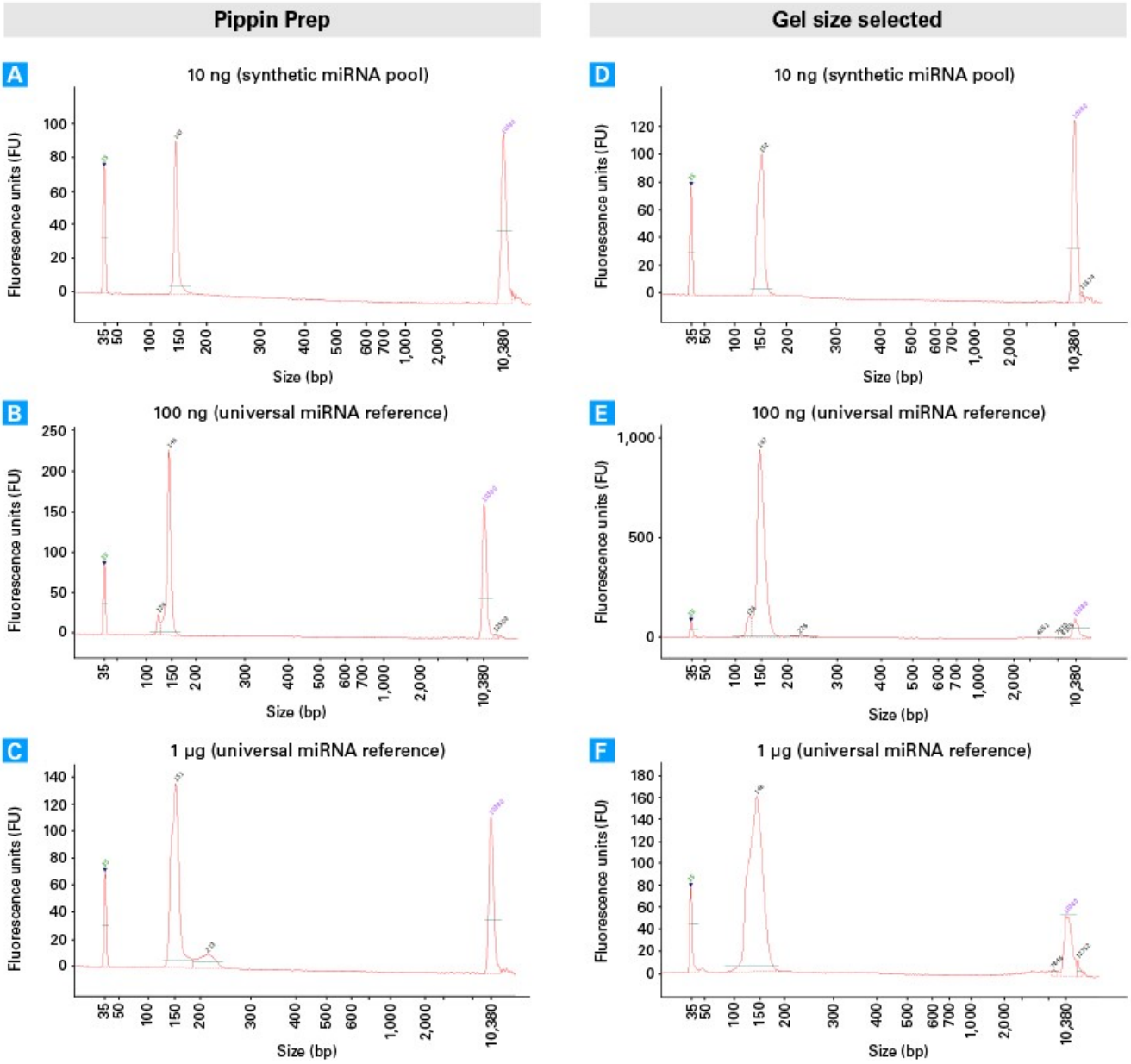

Figure 3. Example electropherogram traces for Pippin Prep or gel size-selected libraries. Libraries were size selected using either Pippin Prep (Panels A–C) or gel size-selection (Panels D–F). Libraries were generated using either 10 ng of a synthetic pool of miRNAs (Panels A and D), or 100 ng (Panels B and E) or 1 µg (Panels C and F) of Agilent Universal Human miRNA Reference RNA.

Results

SMARTer method avoids representation bias inherent to other approaches

To measure the accuracy of microRNA representation, libraries were prepared using an equimolar mixture of 963 microRNAs. After purification, size selection, and sequencing on an Illumina platform, the reads were mapped against a custom reference genome and analyzed for read distribution. In theory, assuming an unbiased library preparation, each microRNA in the pool should receive an equal number of reads.

To assess this, all reads corresponding to each microRNA were counted and normalized such that each microRNA in the library was expected to

have a value of 1. Plotting these values for each microRNA in libraries prepared with the SMARTer microRNA-seq Kit revealed that >70% of microRNAs captured fall within a +/– twofold variation of the expected read number (Figure 4, Panel A), indicating that the expression levels of >70% of microRNAs in the equimolar mixture were accurately represented. Libraries prepared from the same sample with competitor kits (Illumina, NEB, Bioo, or QIAGEN) had poorer microRNA representation: 49–79% of microRNAs were greatly underrepresented (i.e., less than twofold fewer reads than expected), 13–35% of microRNAs were represented within a +/– twofold variation of the expected read number, and 8–16% were overrepresented by more than twofold. Libraries from the QIAGEN kit failed to capture 6 of the 963 miRNAs. Furthermore, biases introduced by sequential linear ligation events could not be corrected by the presence of Unique Molecular Identifiers (UMIs, present in the QIAGEN technology), which can only correct for PCR-induced duplication.

These same libraries were also analyzed for the number of microRNAs detected at different counts per million (Figure 4, Panel B). The SMARTer microRNA-seq Kit detected more microRNAs at the indicated thresholds and with greater sensitivity. As compared to the competitors' kits, the SMARTer microRNA-seq Kit more accurately reflected the true biological state of a sample—an important consideration as microRNA research moves toward diagnostic tools.

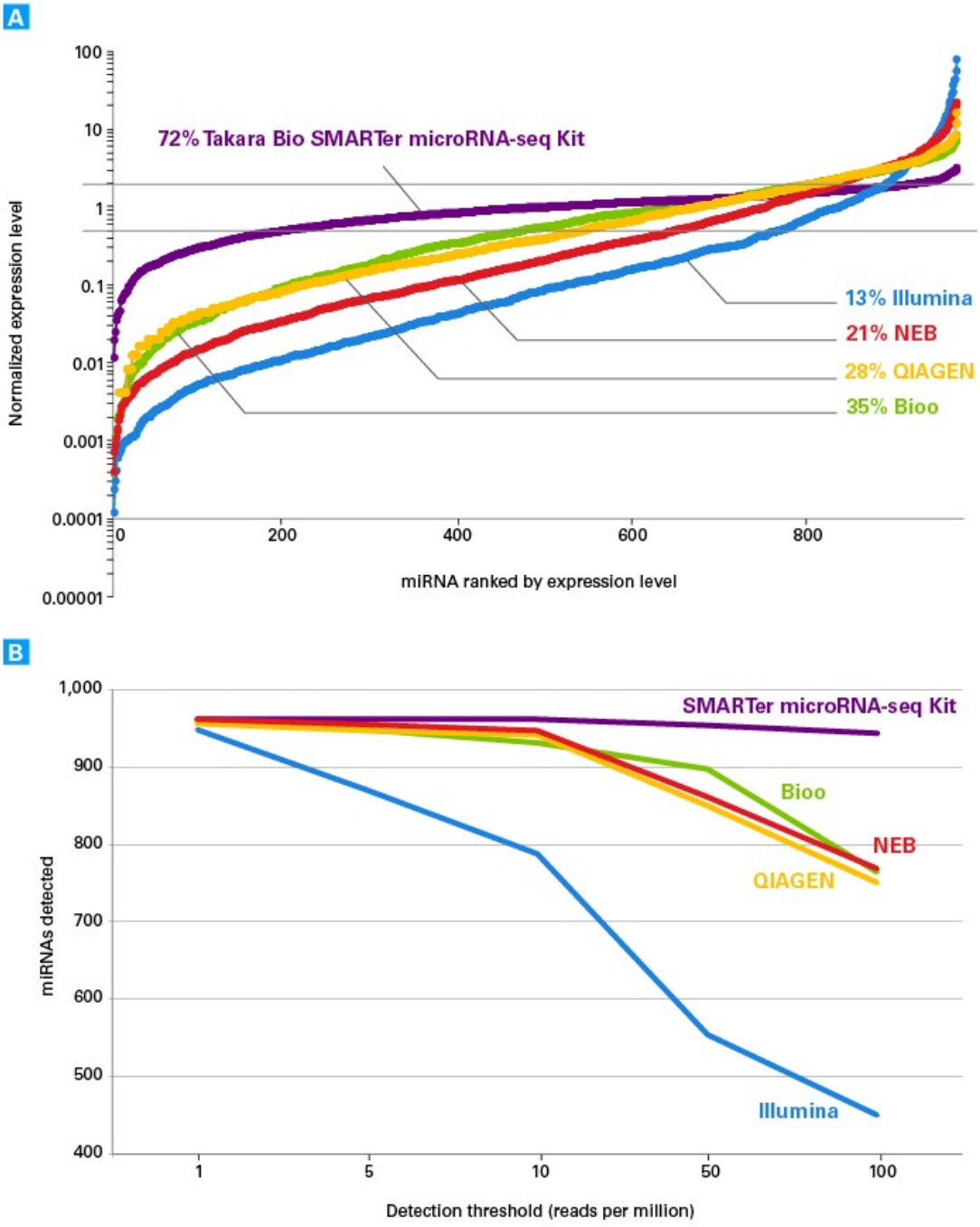

Figure 4. MicroRNA libraries generated with the SMARTer microRNA-seq Kit are more accurate and detect more reads than other state-of-the-art ligation-based technologies. Libraries were generated from 10 ng of an equimolar pool of 963 synthetic microRNAs using our SMARTer library preparation protocol as well as competitors' available technologies (Illumina, NEB, Bioo, or QIAGEN). Libraries were sequenced, mapped using a custom reference containing the 963 sequences in the pool, counted, and normalized (where equal expression was set as 1). Panel A. MicroRNA expression levels (Y axis, log scale) were plotted against microRNA species (X axis, ranked according to expression level). Horizontal lines above and below expected expression level indicate a twofold cutoff. Of note, libraries prepared from 1 ng input show identical results (data not shown). Panel B. Detected microRNAs (Y axis) were plotted against detection threshold (X axis).

SMARTer method yields highly reproducible results across input sources and amounts

To evaluate the performance of the SMARTer microRNA-seq Kit across multiple total RNA sources and input amounts, libraries were prepared with

100 ng or 1 µg of total RNA (RIN >8) from four sources that are known to contain different microRNA expression levels (between 2% and 13%, as described in Table I): human brain, human placenta, human spleen, and the universal reference RNA sample used in the miRQC study (Mestdagh et al. 2014). Sequencing metrics shown in Table I indicate consistent results across input sources and amounts.

All libraries exhibited mapped to hg38 within a tight range (~48–71%) across input levels and sources. When microRNA reads were mapped to miRbase and expressed as a proportion of total reads, results showed consistent mapping across RNA sources for each input amount. Additionally, SMARTer library preparation can capture other small RNA species (collectively piRNA, snoRNA, and snRNA) and does not lose many reads to rRNA. This trend is also seen with lower input amounts (1–10 ng of total RNA; data not shown), although these samples suffer from increased adapter-dimer contamination that may affect sequencing quality. The library preparations show a similar number of microRNAs detected at 5X or higher reads regardless of microRNA proportion (data not shown), suggesting efficient and less-biased microRNA capture.

Sequencing metric	RNA source							
	Brain		Placenta		Spleen		Universal human reference RNA	
smRNA (<200 nt, % of total RNA)	5		10–13		2		3	
Input amount	1 µg	100 ng	1 µg	100 ng	1 µg	100 ng	1 µg	100 ng
Total number of reads	4,114,895	4,057,076	3,860,186	4,590,518	4,298,758	3,800,238	4,644,428	4,359,698
Reads trimmed (% of total reads)	77.91	83.07	61.18	57.38	69.62	66.6	89.09	84.29
Reads mapped to hg38	61.32	55.27	61.13	53.12	67.15	48.92	71.76	66.12
microRNA reads (% of total reads)	29.53	21.05	28.39	13.97	25.95	15.84	30.38	24.98
Number of microRNAs detected	676	497	520	515	442	410	708	603
piRNA, snoRNA, snRNA (% of total reads)	15.81	8.34	14.81	9.41	19.49	11.32	13.82	6.50
rRNA (% of total reads)	1.32	1.52	2.11	1.88	1.33	2.66	2.17	2.02

Table I. Sequencing metrics of RNA species detected from a variety of sources indicate consistent results using the SMARTer microRNA-seq approach.

Finally, there was a strong correlation between microRNA expression levels measured from technical replicates amongst all tissues analyzed (Figure 6, Panel A; brain total RNA shown as a representative example) and between input amounts across all tissues analyzed (Figure 6, Panels B and C). This indicates that data generated using the SMARTer microRNA-seq Kit was highly reproducible.

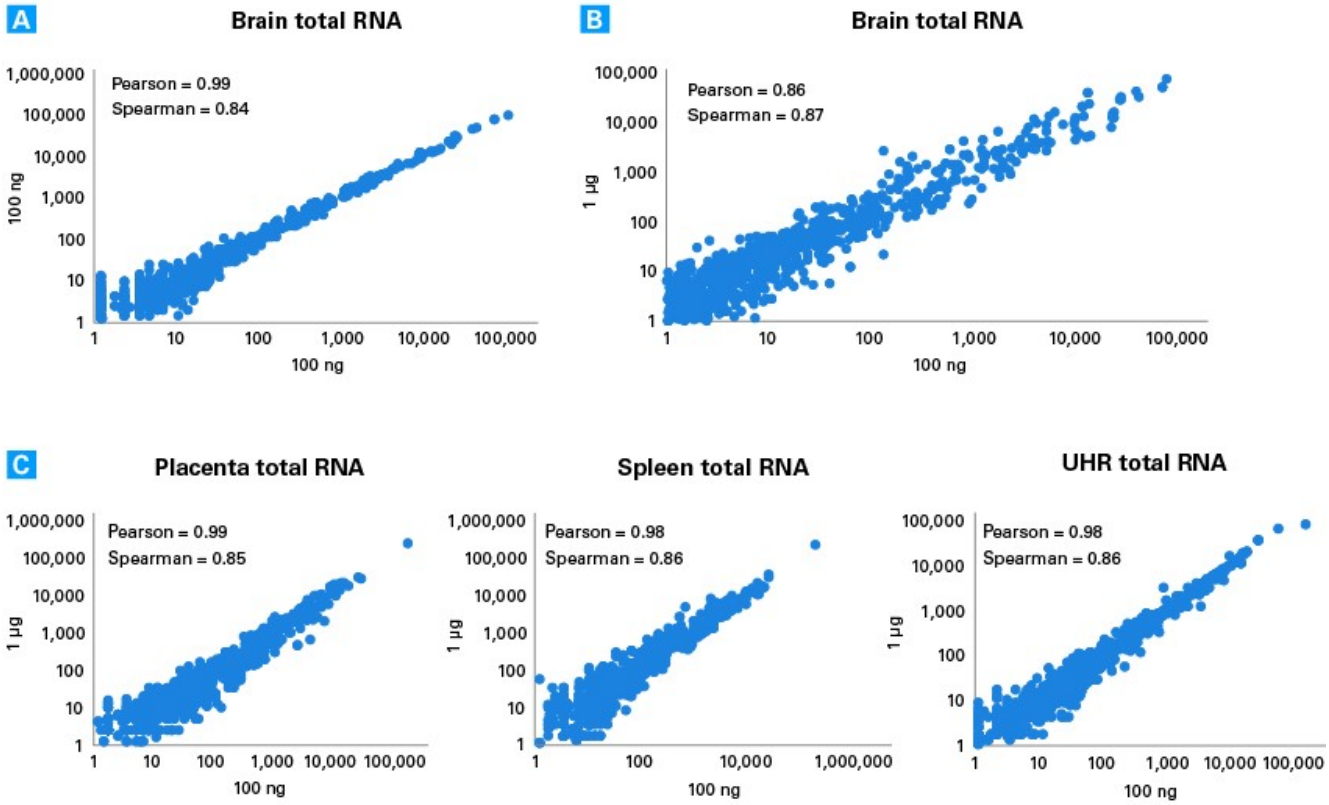

Figure 5. Replicate libraries and libraries prepared from different input amounts are highly reproducible. Libraries were generated in parallel using the SMARTer microRNA-seq Kit from 1 µg or 100 ng of total RNA from human brain, human placenta, human spleen, or universal human reference (UHR) total RNA. Libraries were size selected using the Pippin Prep system. Following sequencing and data processing, normalized expression levels of microRNAs (reads per million) were plotted on correlation diagrams. Panel A represents a correlation of microRNA levels for experimental replicates using 100 ng of human brain total RNA as input. Remaining panels represent correlations of microRNA levels using 1 µg versus 100 ng of input RNA derived from human brain total RNA (Panel B) or placenta, spleen, or universal human reference total RNA (Panel C).

Conclusions

MicroRNAs are a family of small, noncoding RNAs that act as key posttranscriptional regulators of gene expression. Due to their key role in maintaining cellular function, dysregulated miRNA expression is implicated in many disease states, raising the possibility of using microRNAs as biomarkers (Wang, Chen, and Sen 2016). As such, the ability to accurately query microRNA expression is important for current and future scientific advancements, including therapeutics and clinical diagnostics. Towards this end, we have developed an approach that minimizes ligation-induced bias in microRNA library preparation by leveraging adapter addition to microRNAs via intramolecular circularization (MAGIC technology). This results in significantly lower ligation bias and provides an accurate profile of the true microRNA expression in a sample. As compared to microRNA-seq library preparation approaches available from Illumina, NEB, Bioo, and QIAGEN, the Takara Bio SMARTer microRNA-seq Kit significantly improves the accuracy of microRNA expression analyses.

Methods

Library preparation

NGS libraries were generated following the manufacturers' protocols.

MicroRNA bias analyses

Libraries were generated from the miRXplore Universal Reference (Miltenyi Biotec Inc., Cat. No. 130-093-521) using a 10-ng input for the SMARTer microRNA-seq Kit from Takara Bio, NEBNext Small RNA Library Prep Set for Illumina, QIAseq miRNA Library Kit from QIAGEN, and Illumina TruSeq® Small RNA Library Preparation Kit approaches, and a 15-ng input for the Bioo NEXTflex Small RNA-Seq Kit v3 approach.

Total RNA analyses

Libraries were generated using Human Brain Total RNA, Human Placenta Total RNA, Human Spleen Total RNA (Thermo Fisher, Cat. Nos. AM7962, AM7950, and AM7970), or Universal Human miRNA Reference RNA (Agilent, Cat. No. 750700).

Library purification

Libraries generated with the SMARTer microRNA-seq Kit were purified following PCR amplification using magnetic beads (AMPure XP, Beckman Coulter, Cat. No. A63881) following instructions outlined in the SMARTer microRNA-seq Kit User Manual. Libraries prepared using alternate technologies were purified using methods recommended in the manufacturers' protocols.

Library size selection

Post-PCR size selection was performed using the Pippin Prep Size Selection System (Sage Science, PIP0001) and 3% Agarose Gel Cassettes (Sage Science, CDP3010) specifying a size range of 128–160 bp, or via PAGE size selection as specified in the SMARTer microRNA-seq Kit User Manual.

Library quantification and validation

All libraries were quantified using a Qubit Fluorometer (Thermo Fisher Scientific) with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Cat. No. 32851) prior to sequencing. Libraries were analyzed for correct size distribution using an Agilent Bioanalyzer and the Agilent High Sensitivity DNA Kit (Agilent, Cat. No. 5067-4626).

Illumina sequencing

Libraries were sequenced on an Illumina MiSeq® platform using single-end reads (36 bp or 50 bp, depending on the technology), generating at least 1 million reads per library.

References

Bayraktar, R., Van Roosbroeck, K. & Calin, G. A. Cell-to-cell communication: microRNAs as hormones. *Mol. Oncol.* 11, 1673–1686 (2017).

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, e0126049 (2015).

Kakimoto, Y., Tanaka, M., Kamiguchi, H., Ochiai, E. & Osawa, M. MicroRNA Stability in FFPE Tissue Samples: Dependence on GC Content. *PLoS One* 11, e0163125 (2016).

Mestdagh, P. *et al.* Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat. Methods* 11, 809–815 (2014).

Phillips, T. Small non-coding RNA and gene expression. *Nat. Educ.* 1, 115 (2008).

Raabe, C. A., Tang, T.-H., Brosius, J. & Rozhdestvensky, T. S. Biases in small RNA deep sequencing data. *Nucleic Acids Res.* 42, 1414–1426 (2014).

Stuopelyte, K. *et al.* The utility of urine-circulating miRNAs for detection of prostate cancer. *Br. J. Cancer* 115, 707–15 (2016).

Tao, B.-B. *et al.* Evidence for the association of chromatin and microRNA regulation in the human genome. *Oncotarget* 8, 70958–70966 (2017).

Wang, J., Chen, J. & Sen, S. MicroRNA as Biomarkers and Diagnostics. *J. Cell. Physiol.* 231, 25–30 (2016).

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999