Production of an unbiased, highly reproducible small RNA library for NGS using a novel circularization technology

Suvarna Gandlur[1]*, Amanda McNulty[1], Marta Gonzalez-Hernandez[1], Konstantinos Charizanis[1], Jing Ning[1], Edward Jan[1], Matthew Carroll[1], Emmanuel Kamberov[1] and Karl Hecker[2]

[1]Takara Bio USA, Inc., Mountain View, CA 94043, USA • [2]Rubicon Genomics, Inc., Ann Arbor, MI, 48108, USA • *Corresponding author: suvarna_gandlur@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 a miRNA-Seq technology that efficiently captures 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 80–86% of miRNAs captured fall within a +/- 2-fold variation of the expected read number they should receive. This means that the expression level of 80–86% of miRNAs in the equimolar mixture was truly and accurately represented. In contrast, frequency distribution analyses for technologies A, B, and C revealed that 47–56% of miRNAs are greatly under-represented (i.e., less than 2X fewer reads than expected), 29–38% of miRNAs are represented within a +/- 2-fold variation of the expected read number, and around 15% are over-represented 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, our product 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 a small RNA-seq technology that efficiently and accurately captures a broad representation of miRNA species with extremely low bias. Analysis of sequencing data from an equimolar mixture of 963 miRNAs shows that 70–75% of miRNAs are captured within a 2-fold range of their expected number of reads. 80–86% of miRNAs are captured within a 3-fold range. In contrast, state-of-the-art ligation-based technologies only capture 13–30% of the same miRNAs at around a 2-fold variation from expected read numbers, with the vast majority being grossly under-represented.

Our library preparation technology 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.

Library preparation workflow

Denature Samples (2 min)
↓
Adapter Ligation (1 hr 30 min)
↓
Dephosphorylation (35 min)
↓
Circularization (1 hr 40 min)
↓
ReverseTranscription (15 min)
↓
PCR (25–40 min)*
*Depending on input amount

~ 4.5 to 5 hours from RNA to final library (w/o hands-on time)

Over- and under-representation of miRNAs

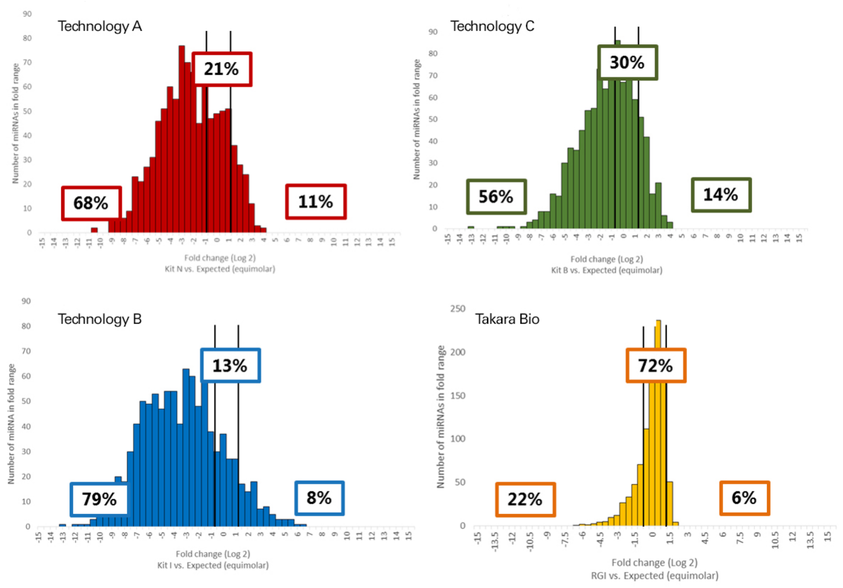

Figure 1. State-of-the-art ligation-based technologies have a significantly higher bias when compared to Takara Bio's technology. Small RNA libraries were prepared using 1 pmol (Takara Bio, Technology A, Technology B) or 5 pmol (Technology C) of a synthetic equimolar pool of 963 miRNAs (Miltenyi miRXplore Universal Reference, Cat. # 130-093-521) following each kit's manufacturer's protocol. The pool consists of HPLC-purified oligos identical to mature human, mouse, rat, and virus miRNAs (developed according to miRBase 9.2). After gel purification and library QC samples were sequenced on Illumina's NextSeq® 500 platform 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 after normalization of reads (reads per million). The bias of the reads was analyzed and shown as representative data of the percentage of miRNAs detected within a two-fold deviation from the expected read value (area between the vertical lines). Analyses also show the percentage of miRNAs detected with read numbers below the expected read value (area left of the two vertical lines) and the percentage of miRNAs detected with read numbers above the expected read value (right of the two vertical lines).

Average miRNA fold levels for Takara Bio's technology

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

Figure 2. More than 70% of miRNAs in a sample are captured within a small variation of their expected expression levels. Small RNA libraries prepared using 1 pmol 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 above and below the expected number of reads each miRNA in the equimolar pool should receive.

Functional control and size marker

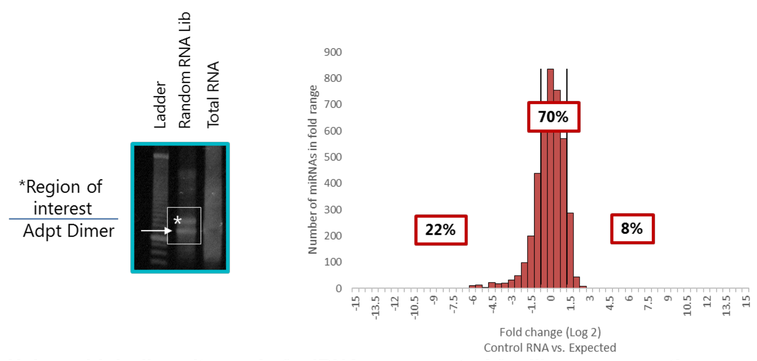

Figure 3. Using a high-diversity pool of miRNAs as a control for library preparation can allow the user to verify that reactions were successful as well as determine their bias. The functional control is a pool of 4096 synthetic miRNA at equimolar quantity. The sequence is derived from human Let7d-3p, but with randomized ends (5' P-NNNUACGACCUGCUGCCUUNNN 3') (Shore et al. 2016). Libraries were prepared using 1 pmol of input following our library preparation protocol. After gel purification and library QC, samples were sequenced on the Illumina MiSeq® platform. 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). The bias of the reads was analyzed and shown as representative data of the percentage of miRNAs detected within a two-fold deviation from the expected read value (area between the vertical lines). Analyses also show the percentage of miRNAs detected with read numbers below the expected read value (area left of the two vertical lines) and the percentage of miRNAs detected with read numbers above the expected read value (area right of the two vertical lines).

Total RNA mapping to miRBase

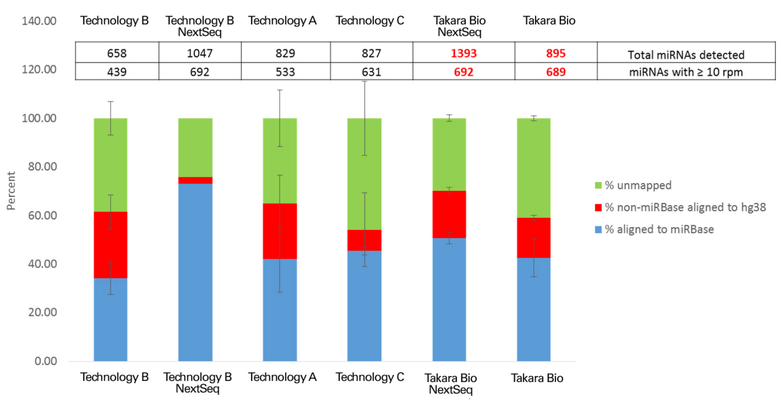

Figure 4. Takara Bio's library preparation technology exhibits similar mapping to the human genome as state-of-the-art technologies, but captures more of the under-represented miRNAs due to its lower bias. Barcoded libraries were prepared from 1 µg of total human RNA (Agilent Universal miRNA Reference Kit, Cat. # 750700) following each kit's manufacturer's protocol. Libraries were purified, size-selected, and checked for quality prior to sequencing. All runs were performed on the Illumina MiSeq platform (unless otherwise noted) using single-end reads (36 or 50 bp), generating at least 1 million reads per library (at least 15 million if sequenced on the NextSeq platform). Sequencing reads for all libraries were trimmed and mapped against GENCODE (GRCH38) and miRBase (release 21) using the CAP-miRSeq bioinformatics small RNA analysis tool set. Shown are the percentages of reads mapped to the human genome, the percentages of those mapped to miRBase and the percentages of reads that appear unannotated.

Small RNA distribution patterns

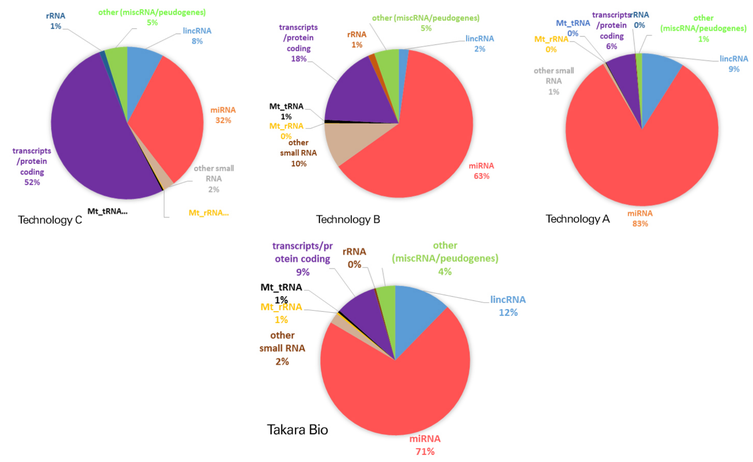

Figure 5. The majority of RNA species captured by Takara Bio's technology are small RNAs. Reads that mapped to the human genome were analyzed for alternative RNA species captured. The distribution of those reads are shown and comprise transcripts/protein coding fragments, rRNA, lincRNA, as well as other small RNAs and miRNA sequences included in the human reference genome (miRBase and non-miRBase).

Repeatability and reproducibility

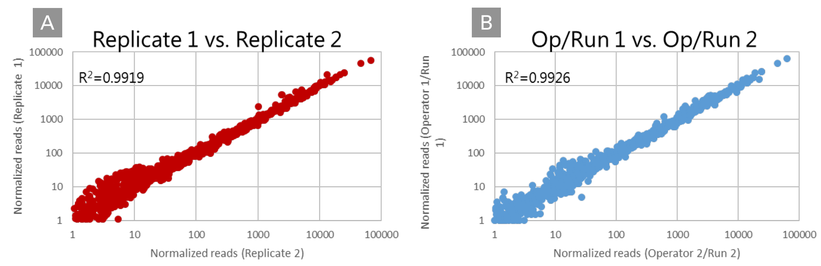

Figure 6. Replicate libraries and libraries prepared by multiple operators are extremely reproducible. Sequencing libraries were generated in parallel from 1 µg of total human RNA (Agilent Universal miRNA Reference Kit) using our small RNA library preparation protocol by two different operators. Libraries 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 comparing individual replicates, or individual operators/runs on correlation diagrams. **Panel A.** Correlation of miRNA expression levels for two technical replicates. **Panel B.** Correlation of miRNA expression levels for samples prepared by two different operators and from different sequencing runs.

Comparison of miRNA profiles captured by different library preparation technologies

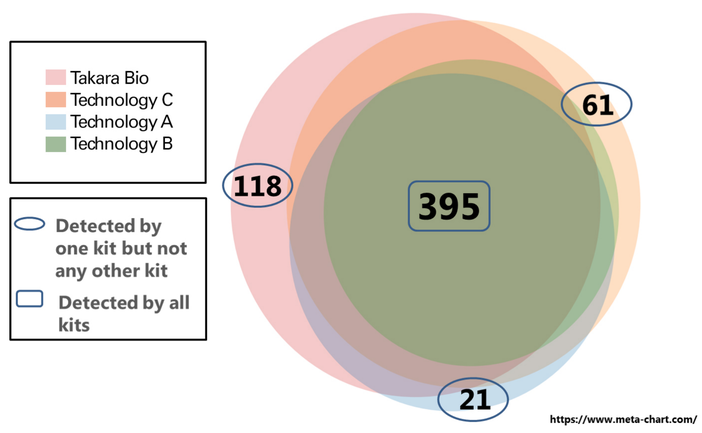

Figure 7. Overlaps exist between miRNAs captured by all state-of-the-art technologies and Takara Bio's technology, but ours captures more at a similar depth and with less bias. Venn diagrams showing the overlap between expressed miRNAs detected in human reference total RNA (Agilent Universal miRNA Reference Kit) using our small RNA-seq technology, as well as commercially available chemistries. miRNA mapping data from miSeq platform sequencing runs were normalized to reads per million (rpm), and the numbers of miRNAs identified with a cutoff of ≥10 rpm are shown.

Conclusions

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

Advantages of Takara Bio's Technology

- Captures more miRNAs
- Better representation of true miRNA expression profile
- Higher percentage of miRNA captured in pool of small RNA
- Random RNA oligo serves as a functional control to verify the success of library preparation and bias assessment
- Libraries prepared from the functional control serve as size markers for gel purification

References

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).

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).

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

Scan to download your copy of this poster or visit takarabio.com/ABRF-2017-smRNA

800.662.2566
Visit us at takarabio.com

Clontech TaKaRa cellartis