Bioo Scientific®

Reducing Biases in Small RNA-Sequencing

Adam Morris Ph.D.*, Dawn Obermoeller, Masoud Toloue Ph.D.
Bioo Scientific Corp., 7050 Burleson Rd.
Austin, TX 78744 USA
*adamm@biooscientific.com

ABSTRACT

The past decade has seen an explosion of interest in cataloging the small RNA repertoires of animal and plant species, and in understanding the biological function of small RNAs. Distinguishing closely-related small RNAs is difficult using hybridization-based approaches, since imperfectly matched small RNAs may hybridize to primers or immobilized probes. These considerations have led to the realization that high throughput sequencing is the most practical method for large-scale small RNA studies that aim to identify and enumerate small RNAs in various species and tissues.

Unfortunately, NGS approaches for small RNA analysis are not without their own challenges. Several studies have now shown entire datasets, including those in miRBase to contain severe sequence bias, specifically, small RNA expression that is not accurately represented by sRNA-seq. Significant effort has gone into identifying the cause of this misrepresentation, and it is now generally accepted that bias in sRNA-seq libraries is primarily introduced during the ligation steps in library preparation. Specifically, RNA ligases show sequence-specific preferences toward flow cell adapters, resulting in preferential inclusion of some small RNAs in sRNA-seq libraries, at the expense of others. Simply using two different adapter sequences during ligation can result in up to 30-fold differential expression for some microRNAs. No single adapter sequence is able to efficiently ligate to all small RNAs, indicating that the target sequence, as well as adapter sequence, is a source of bias.

Our approach to overcoming ligation bias in sRNA-seq libraries involves using a pool of adapters having randomized sequences at the ligation site, where each adapter sequence is present in vast molar excess over any given small RNA in the sample. Experiments show that most of the bias in adapter ligation is due to the sequence of 2-4 adapter nucleotides adjacent to the target junction.

Using our randomized adapter strategy, small RNA libraries were prepared with both synthetic small RNAs and small RNA isolated from human brain and sequenced. Libraries utilizing randomized adapters demonstrated significantly more even coverage due to reductions in ligase bias. We will demonstrate why our new streamlined small RNA-seq protocol is critical for those needing to accurately assess small RNA abundance using high throughput sequencing.

INTRODUCTION

The study of small RNAs, including miRNAs, siRNAs, and pi-RNAs, is an ideal application of next-generation sequencing (NGS) technology. Although methods such as quantitative PCR and microarray analysis are useful for relative quantification of small RNAs, they suffer from two major drawbacks. The first is that these methods are hybridization-based, which presents problems when trying to discriminate two small RNAs whose sequence differs by only a nucleotide or two. The second drawback is that both of these methods are only able to interrogate an a priori determined set of small RNAs, which both limits the scope of studies and prevents discovery of new small RNAs.

Both of these drawbacks of hybridization based methods are addressed by using NGS for small RNA studies, as NGS can reliably distinguish small RNAs that differ by only a single base, and NGS is not limited to the study of a predetermined set of sequences. However, a major drawback of NGS methods for the study of small RNAs is the substantial bias that has been shown to exist in traditional library preparation protocols. This bias has been shown to be introduced during the two ligation steps, and the combined effect of the bias introduced in these steps results in some small RNAs being ligated to adapters much more efficiently than others.

Some of the studies that demonstrated the substantial bias introduced by RNA ligases showed that this bias resulted from the adapter sequence proximal to the ligation junction, and that adapters with 2-4 randomized bases at this junction could be used to substantially reduce ligation bias. Jayaprakash et al. first demonstrated that NGS libraries prepared with this strategy showed little evidence of ligase bias and that data generated from these libraries correlated well with microarray and qPCR data [1]. Jayaprakash et al. also showed that using randomized adapters in both the 5' and 3' adapters was more effective at reducing bias than randomization of the 5' adapter alone (Figure 1A). Bioo Scientific has obtained an exclusive license on this patent pending technology and has since developed a library preparation kit for high throughput sequencing of small RNAs which uses 5' and 3' adapters with 4 random bases at the ligation junctions to reduce bias (Figure 1B).

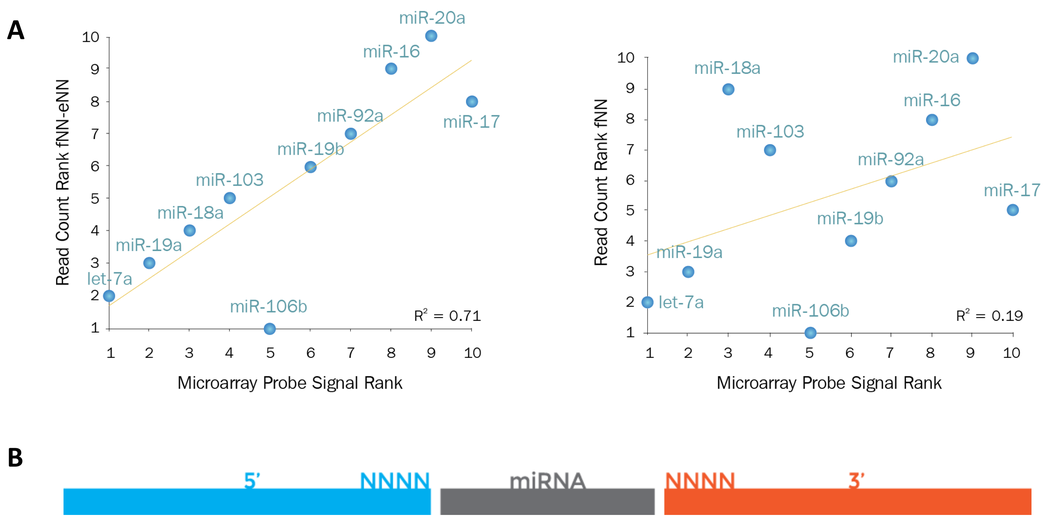

Figure 1. A. Small RNA sequencing libraries were constructed using a 5' adapter with 2 random bases at the ligation junction (fNN) and a non-randomized 3' adapter, or 5' and 3' adapters with 2 random bases at the ligation junctions (fNN_eNN) and the resulting sequencing data compared to microarray results from the same RNA [1]. B. The NEXTflex Small RNA Sequencing Kit v2 utilizes a 5' adapter (blue) and 3' adapter (orange) that contain a series of 4 degenerate bases (randomized adapters) to remedy adapter preferences shown by T4 RNA Ligase 1 and T4 RNA Ligase 2 during the ligation steps of small RNA-Seq library preparation.

REFERENCE

1. Jayaprakash, A.D., et al., Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic Acids Res, 2011. 39(21): p. e141.

METHODS

All libraries were prepared using the NEXTflex Small RNA Sequencing Kit v2 according to the included protocol (Figure 2). Briefly, a pre-adenylated 3' adapter was first ligated to small RNAs using AIR Ligase, a truncated T4 RNA ligase 2. Excess adapter was then removed using a bead-based purification. After removal of excess adapter, ligation to a 5' adapter was performed using T4 RNA ligase 1. This was followed by reverse transcription using a primer that anneals to the 3' adapter, then PCR with primers that anneal to the 5' and 3' adapter sequences and also contain sample barcodes and sequences necessary for cluster generation and sequencing. Following PCR this product was run on a TBE-PAGE gel and the library of the desired size was cut and eluted from the gel. Sequencing was performed on Illumina MiSeq and HiSeq instruments.

The "miRNA calibrator" library was created by mixing 24 synthetic small RNAs in equimolar amounts. Total RNA from human tissues was obtained from Ambion. Libraries from miRNA calibrator were analyzed with a word matching program, and total RNA libraries were aligned to human miRNA hairpins from miRBase (v20) and to the human genome (hg19)

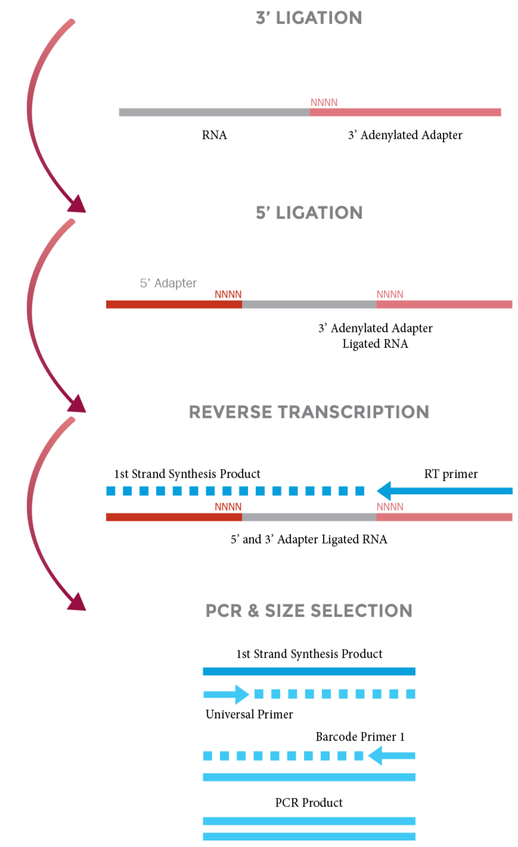

Figure 2. NEXflex Small RNA library preparation protocol. For details see Methods.

RESULTS

In order to demonstrate the reduction in bias achieved by the NEXTflex Small RNA Sequencing Kit v2, a "miRNA calibrator" sample of 24 synthetic miRNAs mixed in equimolar amounts was created. These miRNAs were chosen to represent a variety of sequence combinations at the 3' and 5' ends. 5 ng of this calibrator sample was used to prepare libraries in duplicate with the NEXTflex Small RNA Sequencing Kit v2, which uses adapters with randomized ends, and a traditional small RNA-Seq protocol, which uses standard (non-randomized) adapters. These libraries were then sequenced together, and the proportion of reads mapping to each miRNA present in the calibrator libraries was determined. The duplicate values were then averaged and plotted as a pie graph to demonstrate the relative proportion of reads that aligned to each small RNA in the sample (Figure 3). This analysis clearly demonstrates the substantial reduction in bias when using the latest improvements in small RNA-Seq technology.

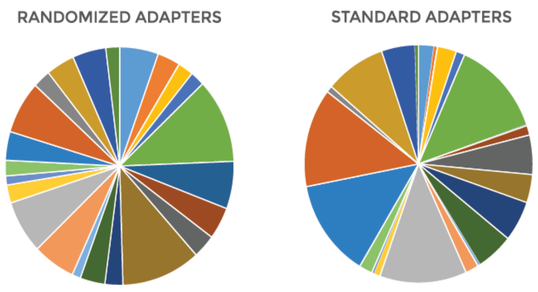

Figure 3. Standard small RNA sequencing vs sequencing using the NEXTflex Small RNA Sequencing Kit v2 with randomized adapters. Libraries were prepared from a "miRNA calibrator" sample consisting of an equimolar mixture of 24 synthetic miRNAs. Each slice in the pie graph represents one miRNA. Mean values from 3 duplicates are shown.

In order to demonstrate the ability to prepare small RNA libraries from various sample types with the NEXTflex Small RNA Sequencing Kit v2, libraries were prepared using total RNA from a variety of human tissues: brain, colon, lung, thymus, esophagus, and spleen. One microgram of total RNA starting material was sufficient to prepare libraries from all tissue types tested. Between 42% and 58% of reads in these libraries mapped to human miRNAs found in the miRBase database (Table 1).

RNA Source	miRBase	hg19
Esophagus	57.1%	64.9%
Brain	55.9%	64.7%
Thymus	52.4%	63.6%
Lung	51.9%	65.4%
Spleen	48.2%	51.3%
Colon	43.5%	63.3%

Table 1. Mapping rates of libraries prepared with the NEXTflex Small RNA Sequencing Kit v2 from 1 µg total RNA from the indicated human tissues. Mapping rates were determined for human miRNA hairpins from miRBase (v20) and for the human genome (hg19).

CONCLUSIONS

While much effort is being focused on understanding microRNA expression and the importance these small RNAs play in gene regulation, research is limited by skewed expression representations introduced during small RNA library preparation. NGS is an ideal technology for measuring small RNA expression, however results from deep sequencing, microarrays and qPCR often do not agree, making it difficult to extract conclusions. By combining adapters with randomized ends and a streamlined, user-friendly protocol, the NEXTflex Small RNA-Seq v2 kit allows creation of small RNA libraries with reduced bias from a variety of sample types. The reduced bias achieved by this kit will allow researchers studying small RNAs to make discoveries that may have been overlooked otherwise and allow for a greater understanding of many aspects of small RNA biology.