Randomized Adapters for Reducing Bias in Small RNA-Seq Libraries

Authors:
R&D Laboratory, Bioo Scientific
7050 Burleson Rd, Austin, TX 78744 USA

Introduction

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. Small RNAs include not only microRNAs, but also piRNAs and other types of endogenous small RNAs.

Distinguishing closely related small RNAs is difficult using microarray- and qPCR-based approaches, since imperfectly matched small RNAs may still be able to hybridize to PCR primers or immobilized probes. These considerations have led to the realization that next generation sequencing (NGS) is the most practical method for large-scale small RNA studies that aim to identify and enumerate small RNAs in various species and tissues. NGS offers advantages of sensitivity, specificity, and the ability to maximize data acquisition and minimize costs by using multiplex strategies to allow many samples to be sequenced simultaneously.

NGS approaches for small RNA analysis are not without their own challenges. Small RNA libraries prepared for sRNA-Seq have been found to contain biases, resulting in libraries that inaccurately represent relative levels of the different small RNAs present in the starting RNA sample. Much effort has gone into identifying the cause of bias, and it is now generally accepted that bias in sRNA-seq libraries is primarily introduced by the T4-phage RNA ligases used during the ligation steps of small RNA library preparation (2, 3, 4). NGS libraries are made by ligation of adapter oligonucleotides to the 5' and 3' ends of the target nucleic acids. The adapters comprise sequences needed to amplify the library by PCR using generic Forward and Reverse primers, as well as sequences needed to associate the target nucleic acids with the NGS sequencing instrument (e.g. the flowcell in Illumina sequencers) and optionally, sequences comprising barcodes to allow multiplexing. For sequencing small RNAs, adapters are added directly to the population of small RNAs using RNA ligases derived from T4 bacteriophage (Rnl1 for 5' adapter ligation and Rnl2 for 3' adapter ligation). Ideally, the RNA ligases would show no preference for attaching adapters to small RNAs of different sequences, but the reality is that RNA ligases show sequence-specific bias, resulting in preferential inclusion of some small RNAs in the sRNA-seq libraries, at the expense of others. Bioo Scientific has developed a novel solution to this problem, which allows meaningful results to be obtained in comparing small RNA profiles between samples.

Materials and Methods

Bioo Scientific's proprietary 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. 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. The new NEXTflex™ Small RNA Sequencing Kit v2 uses a randomized adapter strategy to allow small RNAs of any sequence to "find" their corresponding optimal adapters, resulting in sRNA-seq libraries that show a dramatic reduction in bias.

To demonstrate the reduced bias when using a randomized adapter strategy, small RNA sequencing libraries were prepared in triplicate from an equimolar mixture of 23 synthetic miRNAs, using either standard or randomized adapters in the ligation steps.

Results and Discussion

To validate bias reduction when using the randomized adapter strategy employed in the NEXTflex Small RNA Sequencing Kit v2, libraries were constructed from an equimolar mixture of 23 synthetic miRNAs using the NEXTflex Small RNA Sequencing Kit v2 and using a traditional protocol without randomized adapters. These libraries were sequenced on an Illumina MiSeq, and the percentage of each of the 23 miRNAs was determined with a word-matching algorithm. Mean values from triplicate samples are shown in the pie graph in Figure 1. The results clearly demonstrate that randomized adapters show more equal coverage due to reduced ligase bias.

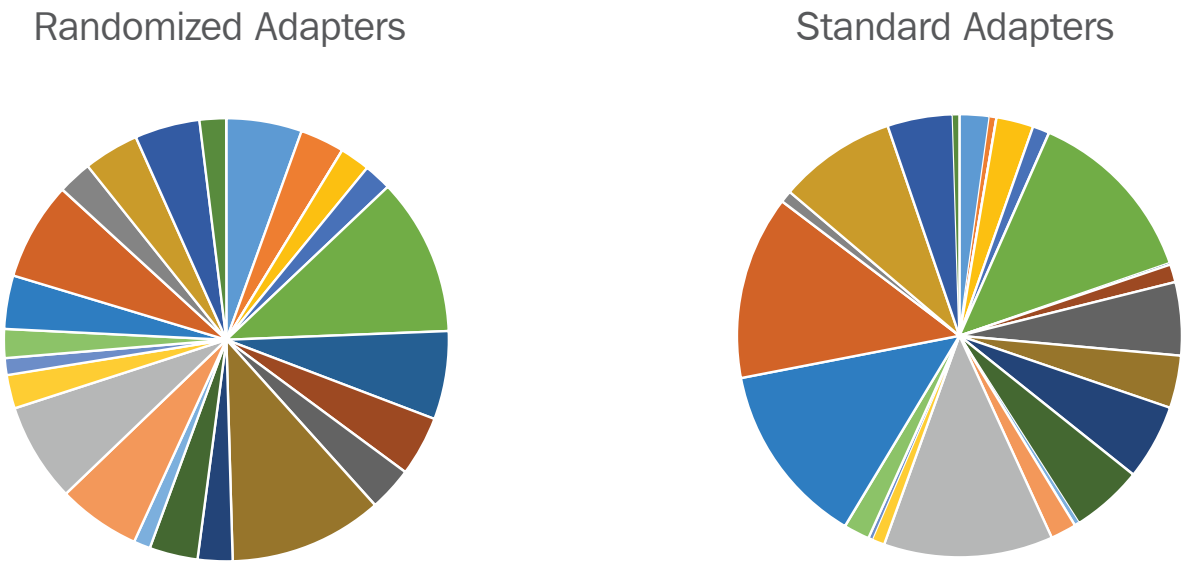

Figure 1. Standard small RNA sequencing vs sequencing using the NEXTflex Small RNA Sequencing Kit v2 with randomized adapters. Libraries were prepared from an equimolar mixture of 23 synthetic miRNAs. Each slice in the pie graph represents one miRNA.

Multiple strategies utilizing adapters with randomized ends have been proposed, with the main differences being the number of randomized bases and whether one or both adapters contain randomized bases. As demonstrated by Jayaprakash et al. better data is obtained when both 5' and 3' adapters contain random bases at the ligation junction. Small RNA libraries were constructed using adapters where only the 5' adapter contained 2 random bases (fNN) or where both 5' and 3' adapters contained 2 random bases (fNN-eNN), and the sequencing data from these libraries was compared to data obtained by microarray. The strategy using randomized bases in both adapters correlated much better with microarray data, demonstrating that randomization of both adapters is a superior strategy. Jayaprakash also showed that randomization of 4 bases at the ligation junctions resulted in more equal recovery of certain miRNAs, such as miR-106b, thus the adapters included in the NEXTflex Small RNA Sequencing Kit v2 contain 4 randomized bases at the ligation junctions of both the 5' and 3' adapters.

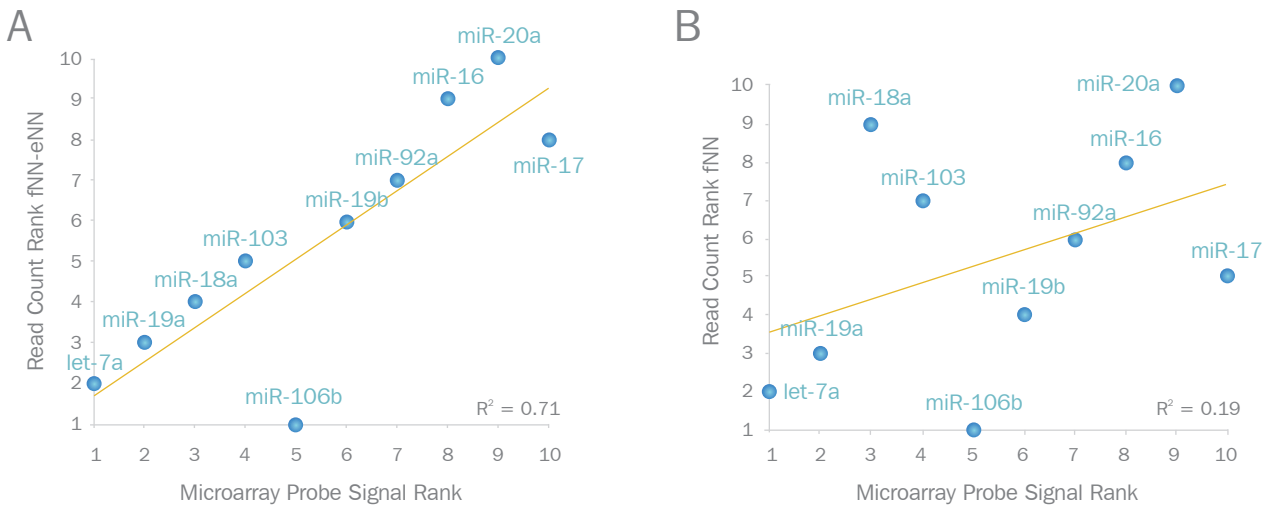

Figure 2. Small RNA sequencing libraries were constructed using 5' adapters with 2 random bases at the ligation junction (fNN) or 5' and 3' adapters with 2 random bases at the ligation junctions (fNN_eNN) and the resulting sequencing data compared to microarray results from the same RNA.

Conclusion

While NGS is the most practical method for studies that aim to identify and enumerate small RNAs, bias can skew the resulting data. Previous studies have shown that RNA ligases derived from T4-phage exhibit significant sequence specificity in their activity (2, 3, 4). As the profiles of small RNAs are strongly dependent on the adapters used for sample preparation, Bioo Scientific has developed the patent pending NEXTflex Small RNA Sequencing Kit v2 to significantly reduce this bias by incorporating a randomized adapter strategy that involves incorporating pools of adapters with randomized bases at critical positions adjacent to the ligation junctions. This technology substantially reduces the bias typically seen in small RNA sequencing data.

1. Betel, D., M. Wilson, A. Gabow, D.S. Marks, and C Sander. 2008. The microRNA.org resource: targets and expression. Nucl. Acids Res. Vol. 36, Database issue D149-D153.

2. Jayaprakash, A.D., O. Jabado, B.D. Brown, and R. Sachidanandam. 2011. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic Acids Res. 39(21):e141

3. Sorefan, K., H. Pais, A.E. Hall, A. Kozomara, S. Griffiths-Jones, V. Moulton, and T. Dalmay. 2012. Reducing ligation bias of small RNAs in libraries for next generation sequencing. Silence. 3:4-15

4. Zhang, Z., J.E. Lee, K. Riemondy, E.M. Anderson, and R. Yi. 2013. High-efficiency RNA cloning enables accurate quantification of miRNA expression by deep sequencing. 14:R109

Bioo Scientific Corporation
7050 Burleson Road
Austin, Texas 78744, USA

www.biooscientific.com
info@biooscientific.com
1.888.208.2246