

Reduced-Bias Small RNA Library Preparation with Gel-Free or Low-Input Options

INTRODUCTION

Changes in microRNA (miRNA) expression have been shown to be associated with a variety of normal physiological processes, as well as diseases including cancer. Studies have already shown that miRNAs may provide useful markers for the development of disease diagnostic and prognostic assays. Next Generation Sequencing (NGS) of small RNAs has significantly improved our understanding of gene regulation, as small RNAs have been shown to fine tune or modulate the expression of genes important in tissue maintenance and disease. NGS brings sensitivity, specificity, and the ability to maximize data acquisition and minimize costs by using multiplex strategies to allow many samples to be sequenced simultaneously with small RNA analysis. This strategy has enabled researchers to perform genome-wide and high depth sequencing studies that would not be possible using other technologies. The ability to quickly and efficiently measure or profile these transcripts has great importance in research and clinical applications.

However, small RNA sequencing has typically suffered from three major drawbacks: 1) severe bias, such that sequencing data does not reflect original miRNA abundances, 2) the need to gel purify final libraries, and 3) lack of low-input protocols.

Bias is introduced into small RNA libraries during the ligation steps. Small RNA library preparation involves ligating adapters directly onto the 3' and 5' ends of the RNA molecule in two separate steps, and each of these ligation steps has been shown to introduce severe bias into library preparation. Recent studies have shown that inserting random bases at the ends of the adapters greatly reduces bias in comparison to using non-randomized adapters [1-4].

Another major drawback of small RNA sequencing is the need to purify final libraries by gel, typically by PAGE gel. This need is due to the small difference in the size of adapter-dimer molecules versus insert-containing molecules following the PCR step of library preparation. In typical DNA or RNA library prep, insert-containing molecules are at least 100 bp larger than adapter-dimer molecules, and thus can be removed using SPRI magnetic beads. However, since insert-containing molecules are only ~20 bp larger than adapter-dimer molecules in small RNA libraries, SPRI size selection is not feasible, and gel-based selection must be performed. The need for gel-based size selection greatly limits both throughput and automation potential of small RNA library preparation, as only a limited number of libraries can be run on a single gel and it is a labor-intensive process that is not amenable to automation.

The lack of low-input protocols for small RNA-Seq is also related to adapter-dimer formation. Small RNA sequencing is somewhat unique in that additional PCR cycles result in negligible bias; thus it should theoretically be possible to create low-input small RNA libraries by using a high number of PCR cycles. However, adapter-dimers present in the libraries will also be greatly amplified, which eventually leads to a library where adapter-dimer products are extremely abundant, making it difficult to isolate insert-containing products and leading to sequencing data where very few of the reads are useful. A number of methods have been developed to reduce adapter-dimer formation in small RNA library preparation, but unfortunately none are effective at reducing adapter-dimer formation to such an extent that gel-free or low-input small RNA library preparation is possible.

In response to the drawbacks associated with small RNA sequencing, Bioo Scientific has developed a small RNA library preparation method that utilizes randomized adapters for greatly reduced bias and features gel-free or low-input protocols, which are possible due to significant reduction in adapter-dimer formation. This method is currently available as the **NEXTflex™ Small RNA-Seq Kit v3**.

RESULTS

The NEXTflex Small RNA-Seq Kit v3 addresses major drawbacks of small RNA sequencing by using two strategies; randomized adapters to reduce ligation-associated bias and a dual approach to adapter-dimer reduction, thereby allowing gel-free or low-input small RNA library preparation. The dual approach to adapter-dimer reduction involves first depleting excess unligated 3' adapter through a magnetic-bead based method, and then inactivating any residual 3' unligated adapter with an enzymatic method. This combination of depletion and inactivation of excess unligated 3' adapter results in significant reduction of adapter-dimer formation, allowing gel-free or low-input library preparation.

In order to test the bias reduction achieved with randomized adapters, we compared the NEXTflex Small RNA-Seq Kit v3 to the NEB-Next® Small Library Prep Kit from New England Biolabs (NEB) and the TruSeq® Small RNA Library Preparation Kit from Illumina. Two miRNA standards were used for comparison. The first, miRNA Calibrator, is an equimolar mixture of 24 synthetic miRNAs with sequences matching known miRNAs from miRBase. Libraries were prepared in triplicate using 1 ng of RNA as starting material and using the respective kit manufacturer's instructions, with the NEXTflex and NEB libraries prepared using gel-free protocols. Libraries were sequenced to a depth of ~500,000 reads each on an Illumina MiSeq, and resulting data was mapped back to the mature miRNA sequences using Bowtie2. For miRNA calibrators, the proportion of reads mapped to each miRNA over total mapped reads was calculated. This value was then used to create an observed/expected value, where the observed value is the calculated proportion and the expected value is ~0.042 (1/24), as the expectation is for each miRNA to be represented equally. Figure 1 shows the results of this analysis, mapped on a log scale. On this scale, 1 is the x-axis and shows the ideal observed/expected value, values less than 1 show that the miRNA is underrepresented, and values greater than 1 show that the miRNA is overrepresented. From this analysis, the bias-reducing effects of the randomized adapters are clear, as almost every miRNA has a value closer to 1 in the NEXTflex libraries than in the NEB or Illumina libraries. Both NEB and Illumina libraries show many more underrepresented small RNAs than NEXTflex libraries, which is correlated with the fact that one or two small RNAs dominate the reads in those libraries. Another way to determine overall bias in these libraries is to calculate the coefficient of variation (CV) of the different miRNA proportions. The CV is defined as the standard deviation divided by the average, and a CV closer to 0 indicates less variation between values, and also less bias in this analysis. As demonstrated in the inset in Figure 1, the CV of the NEXTflex libraries is less than half of that of the NEB or Illumina libraries, demonstrating the reduced bias and more even coverage achieved with the NEXTflex Small RNA Seq Kit v3.

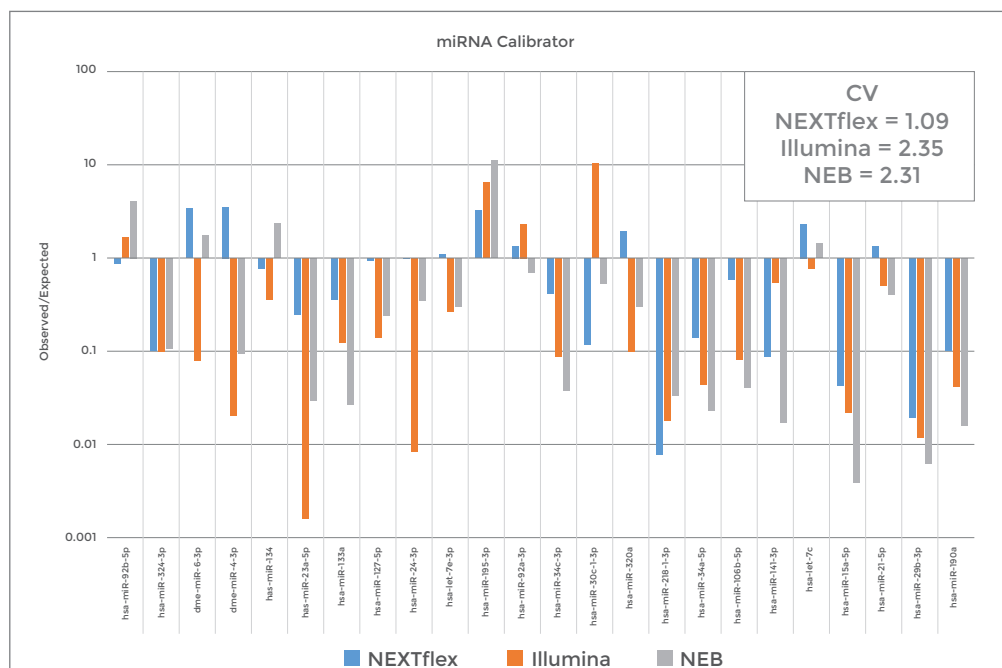


Figure 1. Sequencing results from small RNA libraries created in triplicate from 1 ng of miRNA Calibrator, an equimolar mixture of 24 miRNAs. The indicated number of reads were sampled from each library and the average number of miRNA groups with ≥ 20 reads determined. The inset shows the Coefficient of Variation of the 24 miRNAs in each sample.

The second standard used was the Miltenyi miRXPlore Universal Reference, which contains an equimolar mixture of over 950 synthetic miRNAs matching known miRNAs from miRBase. Results from analysis of libraries created from the Miltenyi miRXPlore Universal Reference standard also demonstrated the reduced bias and more even coverage achieved using randomized adapter technology. The analysis in Figure 2 demonstrates how many of the miRNAs in the Miltenyi libraries were detected at various thresholds of detection. For example, at a threshold of 50 reads/100K total reads, 517 miRNAs were detected in the NEXTflex libraries, and only 365 and 213 were detected in the NEB and Illumina libraries, respectively. This represents a discovery rate in the NEXTflex libraries of 141.6% versus NEB and 242.7% versus Illumina. The CV was also calculated for these libraries, with results again demonstrating less bias and more even coverage in the NEXTflex libraries (Figure 2 inset).

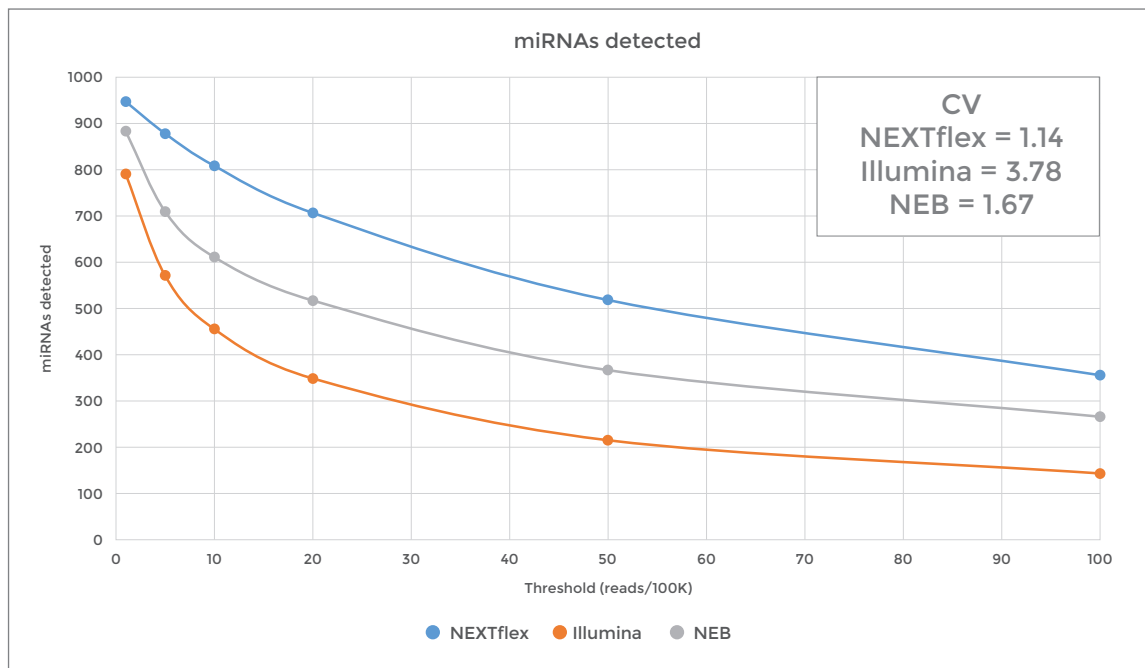


Figure 2. Sequencing results from small RNA libraries created in triplicate from 1 ng of Miltenyi miRXPlore Universal Reference, an equimolar mixture of 963 miRNAs. The number of miRNAs detected at various thresholds is shown. The inset shows the Coefficient of Variation of the 963 miRNAs in each sample.

To compare the performance of the three kits with an actual total RNA sample, libraries were prepared in duplicate with each kit from 100 ng of human brain total RNA. Following adapter trimming, reads were aligned to miRBase human miRNA hairpins using Bowtie2, then analyzed with miRUtils (<http://mirutils.sourceforge.net/>) in order to collapse individual miRNAs into groups. miRNA groups have identical mature miRNA sequences, thus for most small RNA analysis it is more appropriate to consider miRNA groups than individual miRNAs. A more sophisticated approach was taken to analyzing discovery rates in these libraries. The number of reads indicated on the x-axis in Figure 3 was sampled randomly from the adapter-trimmed reads 10 times and the results averaged, and then the number of miRNA groups discovered at a threshold of at least 20 reads was calculated. Figure 3 demonstrates that discovery rates were much higher in NEXTflex libraries than those provided by NEB or Illumina kits. For example, in order to detect 100 miRNAs, only 36,000 reads were required in the NEXTflex libraries, with 110,000 and 164,000 required to detect the same number of miRNAs in the NEB and Illumina libraries. Thus ~3x and ~4.5x more reads are required to detect 100 miRNAs in the NEB and Illumina libraries in comparison to the NEXTflex libraries.

In order to test the performance of the NEXTflex Small RNA Sequencing Kit v3 with low inputs, libraries were created in duplicate with 10 ng of total RNA from human brain. Analysis of discovery rate was performed as described previously, with the low-input libraries showing much greater rates of discovery than NEB and Illumina libraries, despite the use of 10x less input material (Figure 3). miRNA abundance showed a strong correlation between libraries prepared with 100 ng and 10 ng of RNA, demonstrating that the low input protocol did not introduce bias into the library preparation (Figure 4). The high discovery rate and strong correlation seen with low input libraries demonstrates that the NEXTflex Small RNA Seq Kit v3 can be used to create small RNA libraries of superior quality compared to kits from NEB or Illumina, despite the use of considerably less RNA.

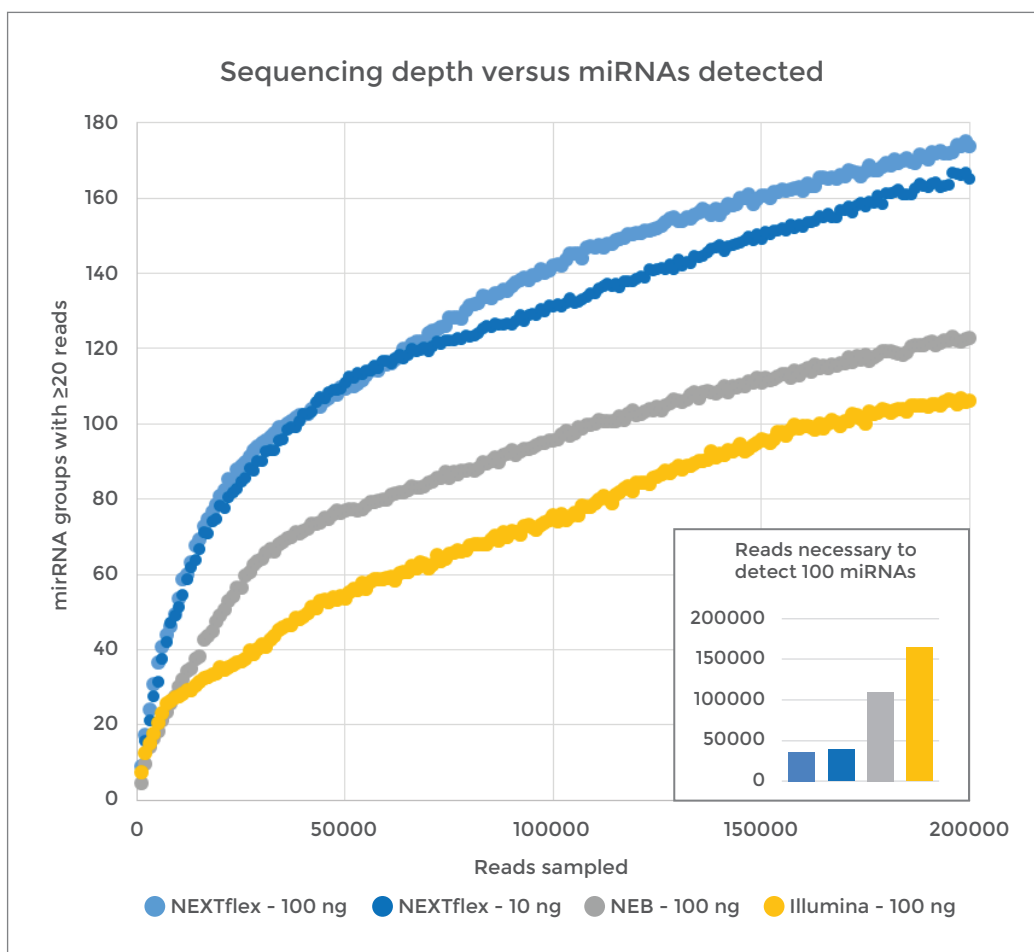


Figure 3. Small RNA libraries were created in duplicate from the human brain total RNA and sequenced on an Illumina MiSeq. The indicated number of reads were sampled from each library and the average number of miRNA groups with ≥ 20 reads determined. The inset shows the number of reads required to detect 100 miRNA groups at a threshold of ≥ 20 reads.

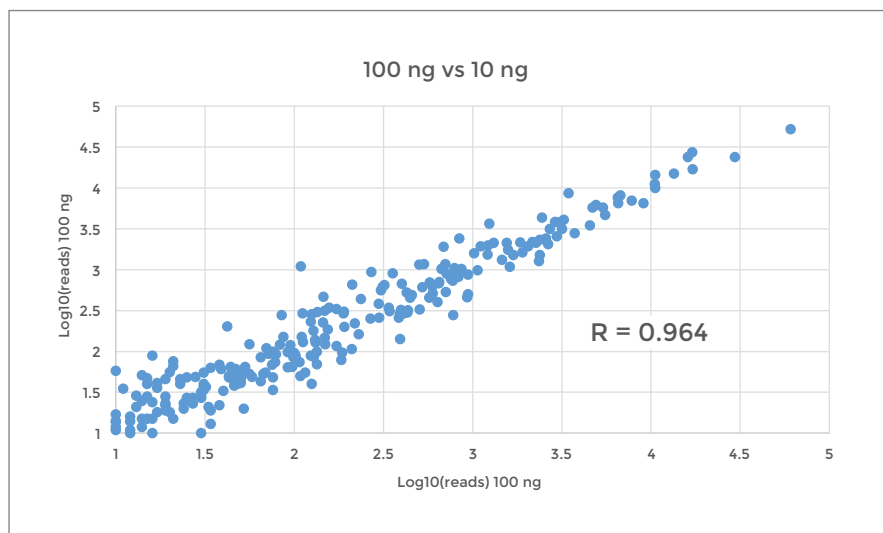


Figure 4. Correlation of miRNA expression between samples created with 100 ng and 10 ng total RNA from human brain with the NEXTflex Small RNA-Seq kit v3. The Pearson correlation coefficient is shown.

In order to generate high-quality small RNA libraries from low abundance starting material, up to 25 cycles of PCR may be necessary. Prevailing dogma in NGS library preparation is to perform the minimum number of PCR cycles needed to generate enough material for sequencing, as additional PCR cycles result in additional bias introduced into the library. While this concern is certainly valid for typical DNA-Seq and RNA-Seq library preparations, at least two studies have shown that PCR introduces negligible bias into small RNA-Seq libraries [1, 5]. In order to test bias introduced through additional PCR cycles, 4 library preparations were started with 2 µg of total RNA from human brain as starting material, and samples were pooled following reverse transcription. Serial 1/10 dilutions of this cDNA were prepared, and library preparation was continued in duplicate using 12-28 cycles of PCR, with 4 additional cycles added for each serial dilution. Thus, 12 cycles were performed on undiluted cDNA, 16 cycles on 1/10 diluted cDNA, 20 cycles on 1/100 diluted cDNA, and so on. Libraries were sequenced and aligned as described above, and the expression correlation of each condition calculated (Table 1). For this analysis, only miRNA groups with ≥10 reads in all samples were considered, resulting in 96 considered miRNA groups. Correlation of libraries created using various cDNA dilutions and corresponding increases in PCR cycles were extremely high, with all correlations from 12-24 cycles being greater than 0.995. A slight decrease in correlation to ~0.98 was seen at 28 PCR cycles, likely due in part to stochasticity involved with very small amounts of cDNA starting material (the equivalent of 0.2 pg of total RNA starting material). This data adds to previously published work demonstrating that additional PCR cycles introduce negligible bias into small RNA library preparation, showing that accurate data can be generated from low-input library preparation that requires relatively high PCR cycle numbers.

Table 1. Correlation of miRNA abundance in libraries created using serial 10x dilutions of cDNA and the indicated number of PCR cycles. The Pearson correlation coefficients calculated from the Log10(reads) values of miRNAs with ≥ 10 reads in all samples are shown.

| PCR Cycles | 12 | 16 | 20 | 24 | 28 |
|------------|----|--------|--------|--------|--------|
| 12 | 1 | 0.9999 | 0.9994 | 0.9966 | 0.9807 |
| 16 | | 1 | 0.9996 | 0.9970 | 0.9813 |
| 20 | | | 1 | 0.9975 | 0.9824 |
| 24 | | | | 1 | 0.9816 |
| 28 | | | | | 1 |

In order to achieve the best value for the cost of sequencing small RNA libraries, both overall alignment rates and discovery rates should be as high as possible. The NEXTflex Small RNA-Seq Kit v3 clearly shows higher discovery rates than NEB or Illumina kits, so we also wanted to calculate overall alignment rates to miRBase, which can be thought of as the number of “useful” reads. Library products that do not represent miRNAs, such as adapter-dimer products and products of longer cellular RNAs, reduce miRBase alignment rates, thus lower alignment rates typically indicate the presence of these other products. As seen in Figure 5, the NEXTflex Small RNA-Seq Kit v3 demonstrates higher overall alignment rates to miRBase, indicating less contamination of non-miRNA products. This result clearly shows that the greatest value in small RNA sequencing is achieved with the NEXTflex Small RNA-Seq Kit v3, as it achieves greater alignment and discovery rates compared with other kits.

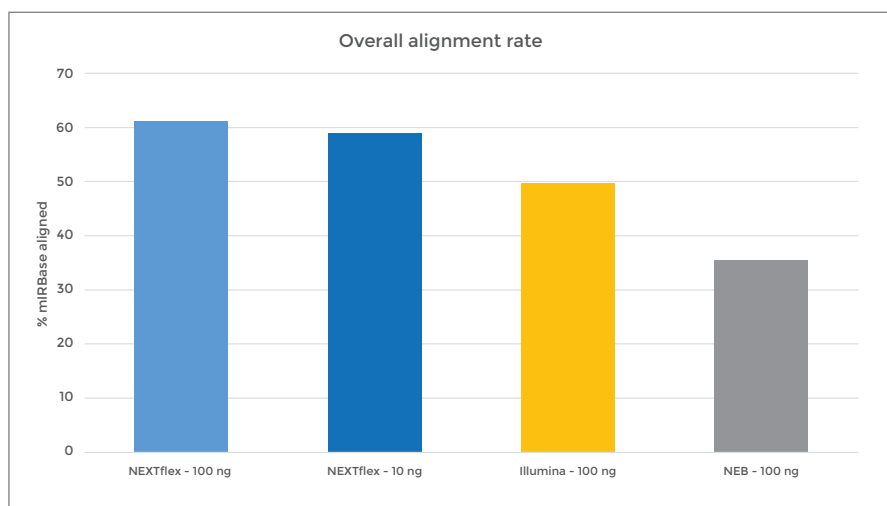


Figure 5. Overall miRBase alignment rates of libraries created with the indicated kit and amount of human brain total RNA input.

CONCLUSION

Small RNA-Seq library preparation has historically suffered from three major drawbacks: severe bias, the need for gel-based purification, and the lack of low-input protocols. The NEXTflex Small RNA Sequencing Kit v3 addresses all of these drawbacks with a library preparation kit that uses patent-pending technology to offer reduced bias and allow gel-free or low-input library preparation. Both gel-free and low-input libraries created with the NEXTflex Kit show greatly reduced bias in comparison to libraries created with kits from NEB and Illumina, resulting in greater depth of discovery and sequencing data that more accurately represents the starting material. The NEXTflex Small RNA-Seq Kit v3 is exceptional in allowing high quality, reduced bias small RNA libraries to be created with low-input or completely gel-free protocols.

This product and its use are the subject of one or more issued and/or pending U.S. and foreign patent applications.

REFERENCES

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.
2. Sun, G., et al., *A bias-reducing strategy in profiling small RNAs using Solexa*. *RNA*, 2011. 17(12): p. 2256-62.
3. Sorefan, K., et al., *Reducing ligation bias of small RNAs in libraries for next generation sequencing*. *Silence*, 2012. 3(1): p. 4.
4. Zhang, Z., et al., *High-efficiency RNA cloning enables accurate quantification of miRNA expression by deep sequencing*. *Genome Biol*, 2013. 14(10): p. R109.
5. Hafner, M., et al., *RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries*. *RNA*, 2011. 17(9): p. 1697-712.

