BIOO SCIENTIFIC®

Constructing High Quality RNA-Seq Libraries from Limited Amounts of Total RNA

INTRODUCTION

An ever-present challenge for Next Generation Sequencing (NGS) applications is obtaining high-quality data from low amounts of starting material. The large number of processing steps required to generate RNA-Seq libraries presents additional challenges for efficiency and retention of material throughout each enzymatic processing and cleanup step. Here the robustness of the NEXTflex™ Poly(A) Beads and the NEXTflex™ Rapid Directional RNA-Seq Kit for mRNA isolation and library preparation was examined using a dilution series of total RNA isolated from a human cell line. Several metrics of library quality, including unique reads, ribosomal RNA contamination and the number of transcripts detected were analyzed to determine the quality of low-input total RNA samples. The data demonstrate that high-quality libraries are readily produced from as little as 10 ng total RNA using the NEXTflex Rapid Directional RNA-Seq Kit.

ROBUST LOW INPUT LIBRARIES

Messenger RNA makes up approximately 0.5%-3% of total RNA depending on the cell line or cell type; therefore 1000 ng of total RNA would produce at a maximum 30 ng of poly (A)+ RNA. The NEXTflex Rapid RNA-Seq Kits were developed for use with 10-100 ng of Poly(A) selected RNA; RNA-seq libraries were constructed from MCF7 Poly (A)+ RNA isolated from either 1000, 100, or 10 ng total RNA. RNA was poly (A)+ selected using NEXTflex Poly(A) Beads according to the protocol in the manual, except RNA was eluted in 15 µL Elution Buffer and 14 µL was used for NEXTflex Rapid RNA-Seq library preparation. Libraries were constructed according to the manual, except that after the final PCR, libraries were eluted in 13 µL Resuspension Buffer. After library construction, each sample was quantified, pooled into equimolar ratios, and sequenced on a HiSeq 2500 in Rapid mode. NEXTflex Poly(A) Beads paired with the NEXTflex Rapid RNA-Seq Kit reliably produces libraries with each amount of starting material, including 10 ng of high-quality total RNA (Figure 1). Several metrics of library quality were quantified across the range of starting material, including rRNA depletion, duplication rates and transcript detection. Excellent performance was observed across the entire input range.

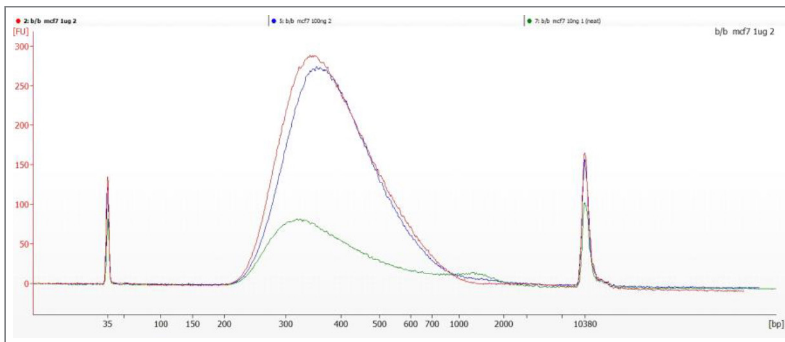

Figure 1. Libraries constructed from a range of total RNA input amounts. High Sensitivity DNA Bioanalyzer traces of RNA-Seq libraries constructed from Poly (A)+ selected total RNA ranging from 10 ng (green), 100 ng (blue), or 1000 ng (red) total MCF7 RNA. 1000 ng libraries were diluted 1:5 prior to loading.

EFFICIENT DEPLETION OF rRNA

For transcriptome studies, the degree to which ribosomal RNA (rRNA) is removed during poly (A) selection is a major factor in obtaining high-quality RNA-Seq data. In order to examine the quality of rRNA depletion, rRNA reads were quantified by mapping libraries to several human rRNA sequences. Methods were adapted from Adiconis et.al. Sequences examined included 18S rRNA, 28S rRNA, 5S rRNA, 5.8S rRNA, the 3' transcribed spacer region of 45S rRNA, and the human mitochondrial rRNA sequence. A significant reduction of rRNA reads was observed across the entire range of input amounts (Figure 2). All ranges of input amounts display very low levels of rRNA contamination, indicating the effectiveness of the NEXTflex Poly(A) Beads on any scale of input material.

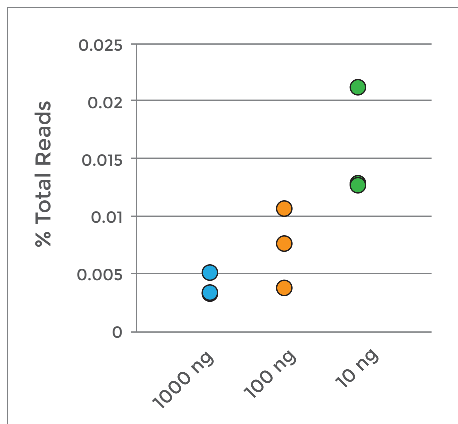

Figure 2. Percentage of rRNA contamination. The percentage of total reads of 28S rRNA, 18S rRNA, 5.8S rRNA, 5S rRNA, 3' spacer of 45S, and mitochondrial rRNA make up a very small fraction of the total reads. The amount of rRNA derived reads increased as the total amount of input decreased.

LIBRARY COMPLEXITY & TRANSCRIPT DETECTION IS MAINTAINED IN LOW INPUT SAMPLES

Library complexity was examined across each input level to determine whether decreased input affects transcript detection sensitivity or abundance measurements. The percentage of libraries that contained uniquely mapping reads as well as percentage of reads that were completely unique was analyzed. As expected, with lower input libraries, lower amounts of starting material result in higher duplication rates (Figure 2; Adiconis et. al., 2013). The duplication rate is increased in the 10 ng libraries compared to the similar 100 ng and 1000 ng libraries.

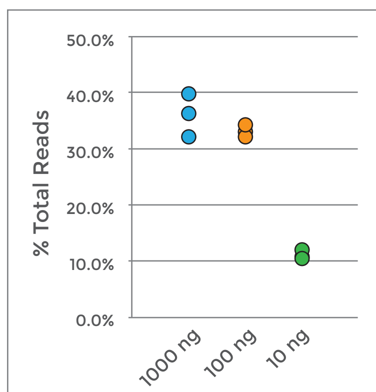

Figure 3. Percent unique reads. Duplicate reads are collapsed into a single read and the fraction of reads mapping to unique sites for each replicate is shown. Note, this analysis does not differentiate between duplicate reads due to PCR and those occurring as a result of high abundance transcripts.

Transcript detection was determined by cataloging the fraction of reads falling into each annotated Ensembl feature. While the detection profiles between the 1000, 100 and 10 ng inputs are distinct (Figure 4), the increase in unaligned reads in the 10 ng samples is likely a result of increased duplication rates. Despite this change in the percentage of raw reads, the 10 ng total RNA input library does not

significantly skew overall transcript representation compared to either 100 ng or 1000 ng input libraries (Figure 5). A decrease in the detection rate of low expression transcripts was observed; however, libraries across input amounts are highly correlated globally, indicating the ability to detect transcripts overall is not overly compromised as a result of low inputs. Importantly, the transcript detection, duplicate levels and global correlation are highly similar between 100 ng and 1000 ng total RNA starting material, indicating that even 100 ng of total RNA can produce libraries equivalent to the widely-accepted input amount of 1 μg total RNA.

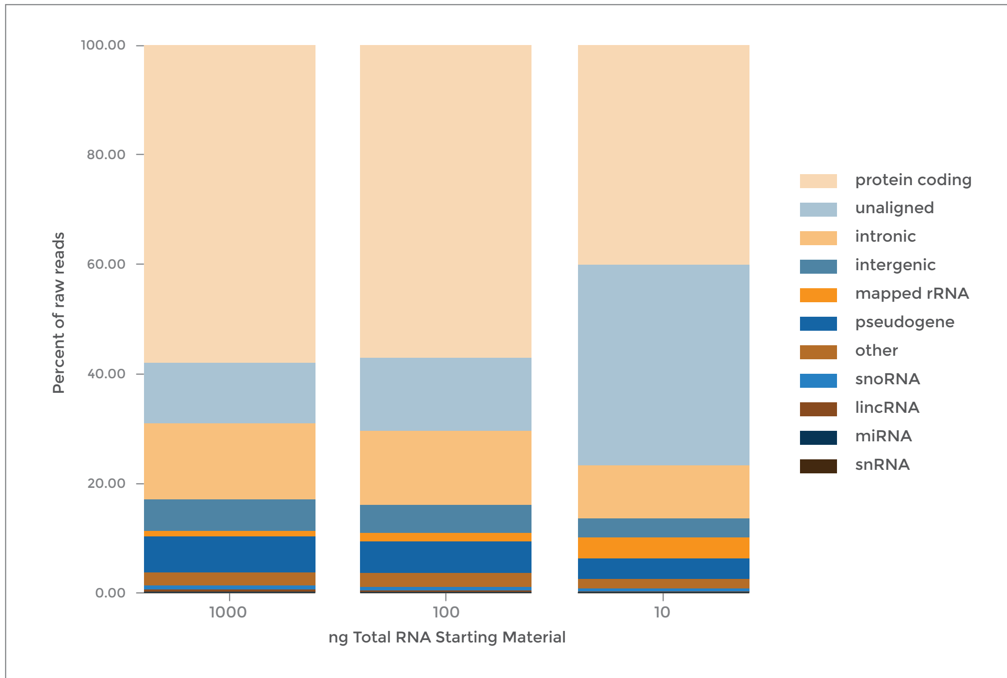

Figure 4. Percent of raw reads aligned to Ensembl feature classes. Reads were mapped to the hg19 assembly and reads aligning into each feature class were counted.

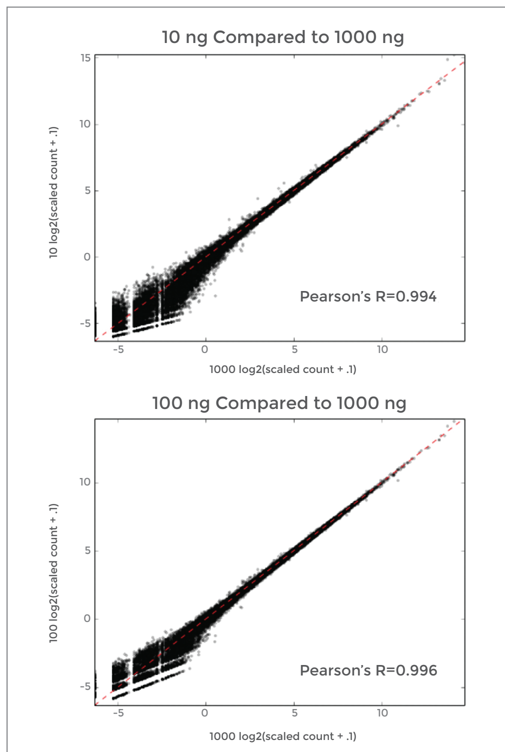

Figure 5. Correlation between libraries constructed with varying amounts of starting material. Scatterplot of the mean scaled counts per gene across 1000 ng input replicates (x-axis) and 10 ng replicates (top panel or 100 ng replicates (bottom panel; y-axis). Each point represents one annotated gene. Points along the edge of the axes are genes with zero reads.

CONCLUSIONS & PRACTICAL APPLICATION

The data presented here outline a method for generating stranded RNA-Seq libraries using low input amounts. Libraries generated from 10 ng total RNA using the NEXTflex Poly(A) Beads and NEXTflex Rapid Directional RNA-Seq Kits produced high quality libraries and sequencing data across several metrics. Importantly, no significant changes to the protocols were made, other than reducing the elution volume, to obtain successful library preparation. As expected, the main challenge of lower input range libraries is an increase in PCR duplicates. However, despite the increase in PCR duplicates, the ability to accurately detect transcript levels and relative expression information remains intact. For researchers unable to obtain large quantities of RNA for sequencing applications, the NEXTflex Poly(A) Beads and NEXTflex Rapid Directional RNA-Seq Kit provide a cost-effective and flexible library construction solution.

References

Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, Sivachenko A, Thompson DA, Wysoker A, Fennell T, Gnirke A, Pochet N, Regev A, Levin JZ. Comprehensive comparative analysis of RNA sequencing methods for degraded or low input samples. Nat Methods. 2013 Jul;10(7):623-9.

Li S, Tighe SW, Nicolet CM, Grove D, Levy S, Farmerie W, Viale A, Wright C, Schweitzer PA, Gao Y, Kim D, Boland J, Hicks B, Kim R, Chhangawala S, Jafari N, Raghavachari N, Gandara J, Garcia-Reyero N, Hendrickson C, Roberson D, Rosenfeld J, Smith T, Underwood JG, Wang M, Zumbo P, Baldwin DA, Grills GS, Mason CE. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. Nat Biotechnol. 2014 Sep;32(9):915-25.