A Complete Solution for Generating Stranded RNA-Seq Libraries from High-Input Total RNA

Magnolia Bostick, Tommy Duong, Suvarna Gandlur & Andrew Farmer[1]

[1]Corresponding Author: Andrew_Farmer@clontech.com

Abstract

Expression analysis of the entire transcriptome by RNA sequencing (RNA-seq) can benefit greatly from high sensitivity, a wide range of sample input amounts, and easy-to-use protocols. A streamlined workflow, combining time-saving techniques with high-performance reagents, provides the opportunity to push each experiment to its fullest potential, improving efficiency and accuracy. Traditionally, the high amounts of ribosomal RNA (rRNA) in the starting material, and lengthy protocols required to incorporate platform-specific adaptors via ligation are major challenges to the generation of RNA-seq libraries from total RNA. The SMARTer® Stranded Total RNA Sample Prep Kit - HI Mammalian is a unique and simplified solution for generating indexed cDNA libraries from mammalian total RNA suitable for next-generation sequencing (NGS) on any Illumina® platform.

In this poster, we illustrate the method for efficient rRNA removal and ligation-free library preparation used in this kit. We present data showing a high correlation between RNA-seq expression levels and qPCR data from the MAQC (Microarray Quality Control) analysis. Additionally, we present ERCC (External RNA Control Consortium) spike-in experiments demonstrating the dynamic range and accuracy of libraries prepared using this method. Finally, we probe the range of input RNA by comparing sequencing metrics from libraries generated using 100 ng–1 µg of total RNA with RNA integrity numbers (RINs) between 3 and 7. In total, these results demonstrate the high reproducibility, minimal bias, and high accuracy obtained using the SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian.

Introduction

The SMARTer Stranded Total RNA Samples Prep Kit - HI Mammalian uses inherently stranded SMART® (**S**witching **M**echanism **a**t 5′ End of **R**NA **T**emplate) technology (1) to both preserve strand information and add Illumina adaptors during cDNA synthesis. This total RNA-seq kit also seamlessly integrates RiboGone™ technology—which uses RNase H digestion to specifically remove rRNA—into the workflow prior to cDNA synthesis. The combination of these two technologies into a single kit decreases the total time needed for rRNA removal, cDNA synthesis, and Illumina library preparation down to approximately 5 hours, which is significantly shorter than other methods for generating total RNA-seq libraries that preserve strand of origin information.

Sequence Analysis

Libraries generated using the SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian from HURR (Universal Human Reference RNA; Agilent) and HBRR (Human Brain Reference RNA; Ambion) were aligned with STAR against hg19 with Ensembl annotation. Libraries generated from Mouse Liver Total RNA (Clontech) were aligned with STAR against mm10 with Ensembl annotation. The percentage of reads that mapped to rRNA, exonic regions, intronic regions, intergenic regions, and the correct strand were defined by Picard analysis.

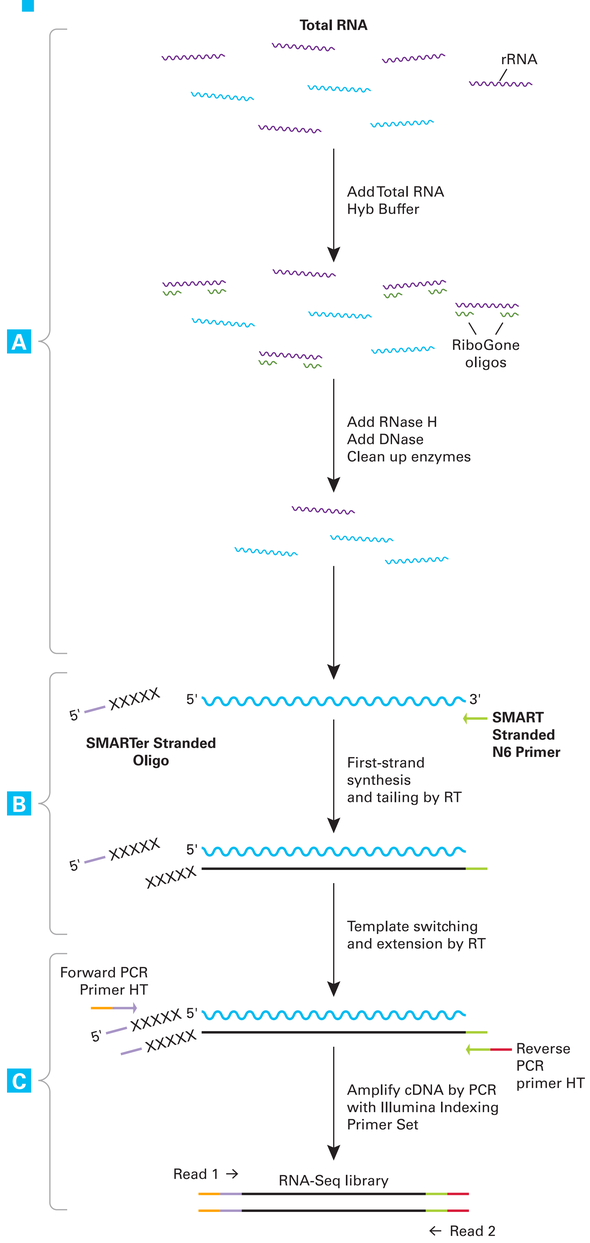
References

1. Chenchik, A., *et al.* (1998) *RT-PCR Methods for Gene Cloning and Analysis.* (BioTechniques Books, MA), pp. 305–319.

2. MAQC Consortium (2006) *Nat. Biotechnol.* **24**(98):1151–1161.

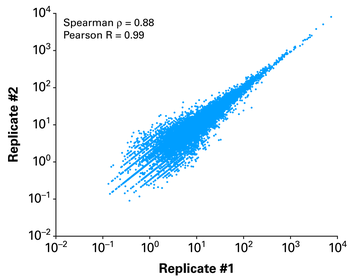
3. Mortazavi, A., *et al.* (2008) *Nature Methods* **5**(7):621–628.

1 Workflow overview



SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian workflow. Section A. Depletion of rRNA from total RNA samples with RiboGone technology. **Section B.** First-strand cDNA synthesis with SMART technology, incorporating Illumina Read Primers 1 and 2. **Section C.** Template switching and generation of sequencing libraries with Illumina cluster-generating sequences and indexes by PCR amplification.

2 Reproducible sequencing data



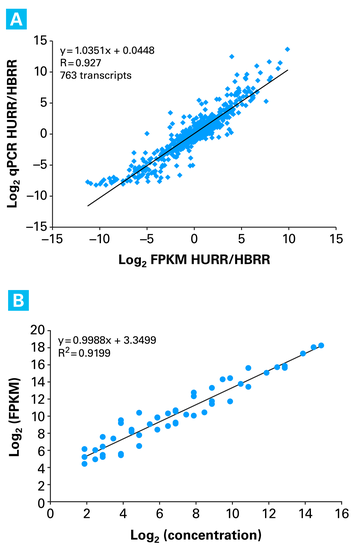
Reproducibility across replicates. RNA-seq libraries were generated from two samples of 100 ng of Universal Human Reference RNA (HURR). The two replicates were processed using the same protocol, except 13 PCR cycles were used for Replicate #1 and 14 PCR cycles were used for Replicate #2. Libraries were sequenced on an Illumina MiSeq® instrument generating 1.3 million single-end reads (1 x 50 bp). The scatterplot illustrates the high correlation between the FPKMs (Fragments Per Kilobase Of Exon Per Million Fragments Mapped) from the two libraries.

3 Accurate representation of input RNA

Table I: Sequence Alignment Metrics

RNA source	Human Universal	Human Brain
Input amount	400 µg	
Number of reads (millions)	8.5 (paired-end reads)	
Percentage of reads:		
rRNA (%)	0.3	5.3
Mapped to genome (%)	94	88
Mapped uniquely to genome (%)	91	84
Exonic (%)	43	50
Intronic (%)	43	33
Intergenic (%)	14	12
Number of genes identified	17,570	17,600
Percentage of ERCC transcripts with correct strand (%)	99.3	98.8

Sequence alignment metrics from different RNA sources. RNA-seq libraries were generated from 400 ng of HURR and Human Brain Reference RNA (HBRR) with ERCC RNA Spike-In Mix 1 (Life Technologies) added to HURR, and Mix 2 added to HBRR, at recommended concentrations. Libraries were sequenced on an Illumina MiSeq instrument generating 8.5 million paired-end reads (2 x 75 bp). Alignment data are displayed for both libraries, with the percentage of reads that mapped to rRNA, exonic regions, intronic regions, intergenic regions, and the correct strand.



Panel A. Comparison to MAQC analysis. To compare the HURR and HBRR libraries to the MAQC analysis (2), the Log_2 ratio of FPKMs from HURR/HBRR libraries was graphed against the Log_2 of the ratio of HURR/HBRR derived from qPCR. The scatter plot shows the good correlation between the expression levels derived from these two methods.

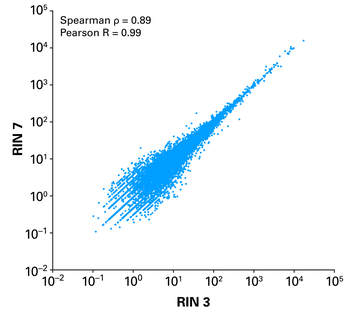
Panel B. Dynamic range and linearity of RNA-seq data. ERCC expression levels (as determined by FPKMs of libraries generated using HBRR with ERCC RNA Spike-In Mix 2) were compared to the ERCC input concentration. The graph shows strong correlation between the Log_2 of input concentrations of individual ERCC transcripts and the Log_2 of FPKMs for those transcripts.

4 High-quality data from a wide range of RNA

Table II: Sequence Alignment Metrics from RNA of Varying Quality

RNA source	Mouse Liver			
RNA quality (RIN)	RIN 3		RIN 7	
Input amount	100 ng	1 µg	100 ng	1µg
Number of reads (millions)	1.7 (paired-end reads)			
Percentage of reads:				
rRNA (%)	2	2	1	1
Mapped to genome (%)	82	86	81	88
Mapped uniquely to genome (%)	73	75	72	77
Exonic (%)	55	53	54	54
Intronic (%)	32	31	33	32
Intergenic (%)	12	14	12	13
Number of genes identified	12,079	12,172	12,099	12,212
Percent biological strandedness (%)	95.5	97.2	95.6	98.1

Sequence alignment metrics from RNA of varying quality. Libraries were generated from 100 ng or 1 µg of Mouse Liver Total RNA that was chemically sheared until it had a RNA Integrity Number (RIN) of 3 or 7 (3). Libraries were sequenced on an Illumina MiSeq instrument generating 1.7 million paired-end reads (2 x 25 bp). Sequencing data show the percentage of reads that mapped to rRNA, exonic regions, intronic regions, intergenic regions, and the correct strand.



Reproducibility across RNA quality. The scatterplot illustrates the good correlation between the FPKMs from two libraries generated from 1 µg of Mouse Liver RNA that was chemically sheared until it had a RIN of 3 or 7.

Conclusions

The SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian is a complete solution for preparing indexed Illumina sequencing libraries from 100 ng–1 µg of mammalian total RNA. This kit incorporates key RiboGone and SMART technologies to significantly reduce hands-on time and increase efficiency while maintaining the high-quality sequencing data of low-input SMARTer Stranded RNA-Seq Kits.

- **Simplified protocol**—A single kit for rRNA removal, cDNA synthesis, and indexed Illumina library preparation
- **High reproducibility**—Consistent data across replicates as well as between RNA of different quality
- **High accuracy and wide dynamic range**—High correlation with MAQC data and accurate detection of ERCC controls

To download this poster visit:
www.clontech.com/ABRF2015-Total-RNA-seq