Ligation-free small RNA sequencing libraries from clinical samples using SMART® technology

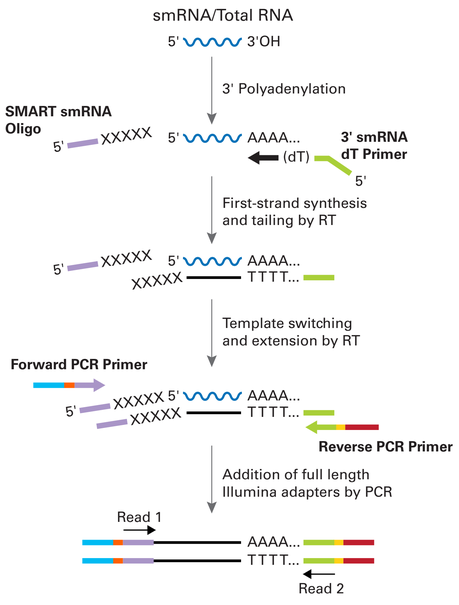
Nathalie Bolduc[1], Simon Lee[1], Rachel Fish[1], Kahraman Tanriverdi[2], Evangelia Malahias[2], Ekaterina Mikhalev[2], Jane E. Freedman[2] and Andrew Farmer[1]*

[1]Takara Bio USA, Inc., Mountain View, CA 94043; [2]University of Massachusetts Medical School, Department of Medicine, Division of Cardiovascular Medicine, Worcester, MA 01605, USA *Corresponding Author

Abstract

Small non-coding RNAs regulate gene expression via diverse mechanisms and facilitate fundamental cellular processes such as transcript splicing and protein translation. Moreover, small RNAs may act at sites different from their original production via incorporation and transport in bodily fluids. Obtaining an accurate portrait of small RNA expression levels from small sample inputs carries potential both for the fulfillment of basic research objectives and the development of novel therapeutics and clinical diagnostic solutions. Towards this end, we have developed a novel, ligation-free approach for the preparation of small RNA sequencing libraries that leverages 3′ RNA polyadenylation followed by cDNA synthesis and template switching. This approach minimizes sample representation bias and is sensitive enough to accommodate inputs of as little as 1 ng of total RNA. Here we present data demonstrating the accuracy, sensitivity, and reproducibility afforded by the SMARTer® smRNA-Seq Kit for Illumina®. Using a library of synthetic miRNAs, we show that the SMARTer method generates a more accurate representation than the typical adapter ligation method. Furthermore, we successfully created high-quality libraries using 400 pg to 1 ng of RNA extracted from plasma and serum samples, demonstrating the suitability of the SMARTer approach for applications such as biomarker discovery.

1 Workflow for the SMARTer smRNA-Seq Kit for Illumina



Schematic of technology used by the SMARTer smRNA-Seq Kit for Illumina. SMART technology is used in a ligation-free workflow to generate sequencing libraries for Illumina platforms. Input RNA is first polyadenylated in order to provide a priming sequence for an oligo(dT) primer (3′ smRNA dT Primer). When the MMLV-derived PrimeScript™ ReverseTranscriptase (RT) reaches the 5′ end of each RNA template, it adds non-templated nucleotides which are bound by the SMART smRNA Oligo—enhanced with locked nucleic acid (LNA) technology for greater sensitivity. In the template-switching step, PrimeScript RT uses the SMART smRNA Oligo as a template for the addition of a second adapter sequence (purple) to the 3′ end of each first-strand cDNA molecule. In the final step, full-length Illumina adapters (including indexes for sample multiplexing) are added during PCR amplification.

Methods

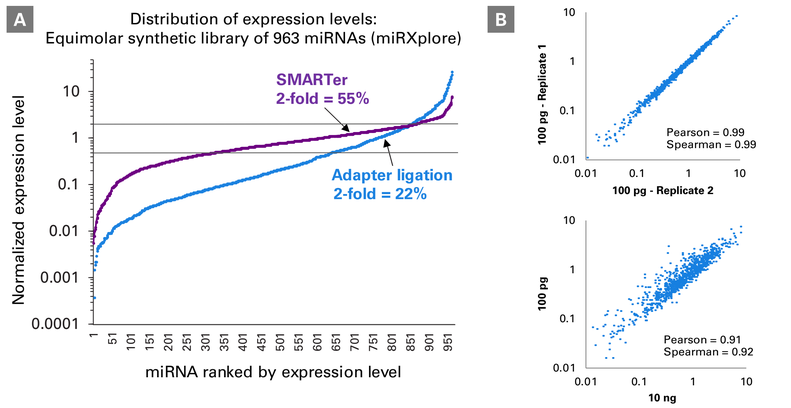
Sequencing libraries shown in Section 2 were generated from the miRXplore Universal Reference (Miltenyi Biotech Inc., Cat. No. 130-093-521) using a 1 ng input for the SMARTer method, and a 100 ng input for the adapter-ligation approach. The adapter-ligation approach was performed using a small RNA-seq kit from another vendor, according to its accompanying protocol. Human total RNA (from placenta, brain, and spleen, respectively) used to generate data shown in Section 3 was purchased from Thermo Fisher Scientific (Cat. Nos. AM7950, AM7962, and AM7970).

All blood samples were collected from a single individual in either EDTA or silica blood collection tubes. Standard centrifugation methods were used to isolate plasma and serum from whole blood. RNA was isolated from 200 µl of either plasma or serum (two technical replicates per sample type) using the miRNeasy Serum/Plasma Kit (Qiagen, Cat. No. 217184). RNA was eluted in 25 µl of PCR-certified water. Of the eluted RNA, 7 µl (containing either 400 pg or 1 ng of input RNA) was used for Illumina library preparation with the SMARTer smRNA-Seq Kit for Illumina as described in the user manual.

Most libraries were sequenced on an Illumina MiSeq® instrument. Libraries generated from RNA isolated from plasma or serum were sequenced on a NextSeq® instrument.

Sequencing reads for all libraries were trimmed and annotated using CLC Genomics Workbench 8.5.1 (Qiagen) and Small RNA Analysis tools, allowing no more than one mismatch during mapping. Overall mapping was performed against GENCODE (GRCH38). miRNA sequences were mapped to miRBase (release 21), while all other small RNA species were analyzed by mapping to the NONCODE v3.0 dataset. For analysis of the miRXplore libraries, reads were annotated using a reference file containing sequences present in the synthetic pool. Normalization was performed by determining, for each miRNA, the ratio between the observed number of reads and the predicted number of reads.

2 A ligation-free method providing unparalleled accuracy and reproducibility



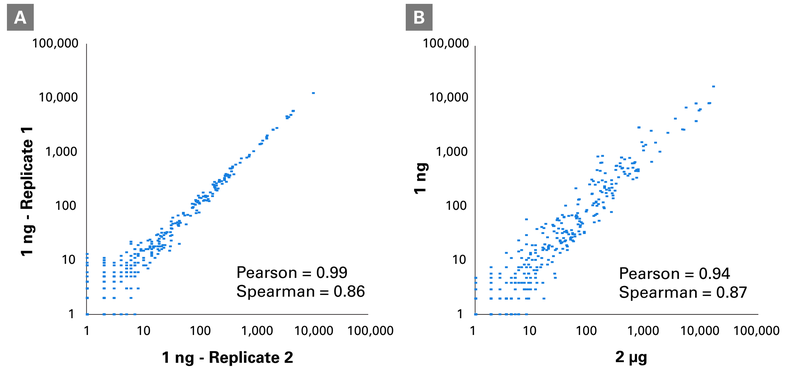
Accuracy and reproducibility of the SMARTer approach for small RNA-seq. Panel A. To gauge the accuracy of the SMARTer approach, sequencing libraries were generated from an equimolar pool of 963 synthetic miRNAs (miRXplore Universal Reference) using the SMARTer smRNA-Seq Kit for Illumina (1 ng input; purple), or a small RNA-seq kit from a different vendor (Competitor N) employing an adapter-ligation method (100 ng input; blue). miRNA expression levels (Y axis, log scale) were normalized, resulting in an expected expression level equal to 1 for each miRNA, and a 2-fold cutoff was assigned both above and below the expected expression level (indicated by two horizontal lines). For visualization purposes, miRNAs are ranked along the X axis in order of expression level. **Panel B.** To assess the reproducibility of data generated with the SMARTer approach, sequencing libraries were generated in parallel from the indicated input amounts of the miRXplore Universal Reference using the SMARTer smRNA-Seq Kit for Illumina. Expression levels of miRNAs identified for each library were quantified and plotted on correlation diagrams, and Pearson and Spearman correlation coefficients were calculated.

3 High performance across RNA sources and input amounts

Sequencing alignment metrics for small RNA from placenta, brain, spleen, and blood

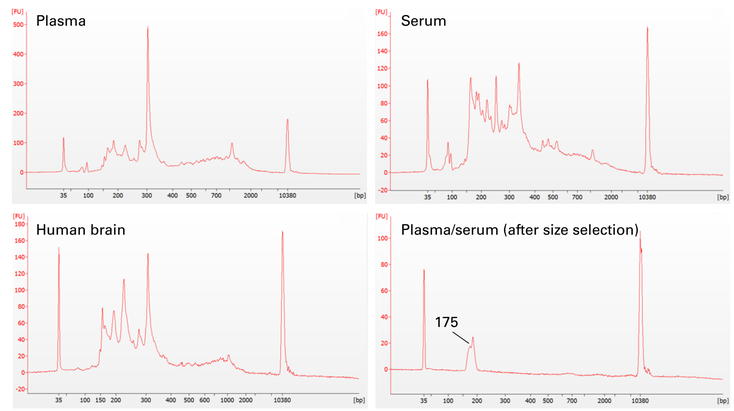
RNA source	Placenta		Brain		Spleen		Plasma-1	Plasma-2	Serum-1	Serum-2
smRNA <200 nt (% of total RNA)	13%		5%		2%		-	-	-	-
Input amount	2 µg	1 ng	2 µg	1 ng	2 µg	1 ng	400 pg	400 pg	1 ng	1 ng
Total number of reads	4,342,213	4,744,519	4,764,574	4,275,787	3,796,263	4,254,142	7,493,889	5,629,347	10,571,256	7,305,083
Proportion of reads discarded	15.1%	24.7%	23.2%	31.8%	38.6%	32.2%	19.6%	20.5%	11.1%	14.4%
Proportion of reads mapped to GENCODE	76.5%	65.4%	68.1%	56.9%	53.7%	56.1%	70.1%	68.8%	79.0%	74.1%
miRNA mapping (miRBase)										
Number of miRNA reads	485,331	509,129	657,282	340,814	326,870	353,553	256,128	197,537	204,627	144,229
Proportion of total reads	**11.2%**	**10.7%**	**13.8%**	**8.0%**	**8.6%**	**8.3%**	**3.4%**	**3.5%**	**1.9%**	**2.0%**
Unique miRNAs detected	260	263	286	253	198	221	186	167	151	125
Number of miRNAs in common	**247**		**243**		**187**		**158**		**120**	
Proportion of miRNAs in common	**89%**		**82%**		**81%**		**81%**		**77%**	
Other small RNA mapping (proportion of total reads, %)										
piRNA	3.5	3.7	8.9	5.1	4.1	3.1	2.9	2.8	3.2	1.7
snoRNA	1.0	0.7	0.8	0.5	1.1	1.3	0.2	0.2	0.4	0.4
snRNA	2.1	1.1	1.2	0.8	0.7	0.9	0.3	0.3	0.4	0.4
tRNA	2.1	3.4	4.0	2.9	0.9	0.9	1.0	1.0	4.5	2.3
Other RNAs (proportion of total reads, %)										
rRNA (5, 5.8, 18, and 28S)	19.2	12.8	12.1	10.1	14.9	13.8	9.6	10.0	34.5	31.7

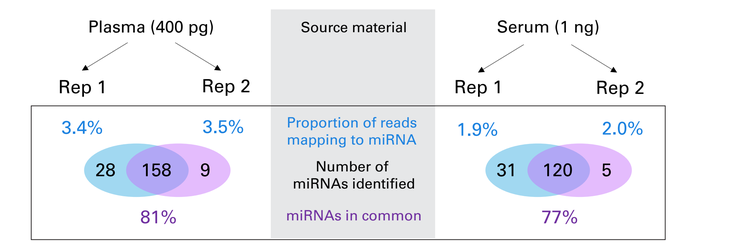
Evaluating the performance of the SMARTer method for small RNA-seq across RNA input types and amounts. Sequencing libraries were generated from 1 ng and 2 µg of human placenta, brain, and spleen total RNA, 400 pg of human plasma RNA, and 1 ng of human serum RNA. The miRNA fraction, corresponding to a final library size of about 175 bp (see 4A) was enriched prior to sequencing using a BluePippin instrument. Following trimming, reads were mapped either to the GENCODE dataset (for overall mapping), or to specific small RNA datasets, as indicated. Only miRNAs represented by at least five reads were included in count data for the number of miRNAs detected.

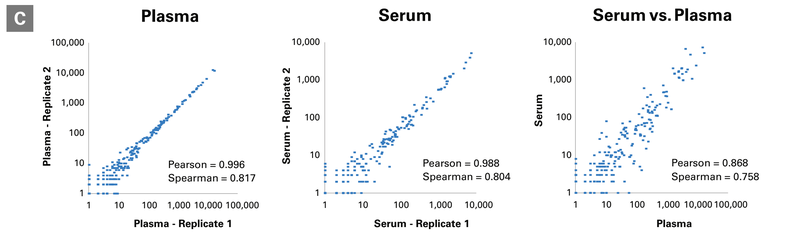


Reproducibility of SMARTer small RNA-seq data generated from human brain total RNA samples. Sequencing libraries were generated in parallel from the indicated input amounts of human brain total RNA using the SMARTer smRNA-Seq Kit for Illumina, and size selected using the BluePippin system. Following sequencing, data processing, and mapping, expression levels of miRNAs identified for each library were quantified and plotted on correlation diagrams, and Pearson and Spearman correlation coefficients were calculated. **Panel A.** Correlation of miRNA expression levels for experimental replicates involving 1 ng inputs. **Panel B.** Correlation of miRNA expression levels for 2 µg vs. 1 ng inputs.

4 High quality data from low input plasma and serum RNA samples







Evaluating the performance of the SMARTer smRNA-Seq method for plasma and serum RNA samples. Sequencing libraries were generated from RNA extracted from 200 µl of plasma or serum (two technical replicates per sample type, see methods). In each instance, the maximum input volume was used, representing either 400 pg or 1 ng of cfRNA. Libraries were amplified using only 16 cycles of PCR. **Panel A.** Bioanalyzer profiles of libraries before and after BluePippin size selection. For comparison, a profile generated from human brain total RNA is also shown. Libraries generated from plasma and serum RNA exhibit very different profiles, with the profile of the plasma library being similar to the profile of the brain total RNA library. Size selection was performed to specifically enrich for the miRNA fraction (insert + adapters = ~175 bp). **Panel B.** Mapping to miRBase revealed that more miRNAs could be identified from the plasma samples than from the serum samples. The percentage of miRNAs that were identified for both replicates is indicated below each Venn diagram (full mapping statistics are included in Section 3). **Panel C.** To assess reproducibility, expression levels of miRNAs identified in each library were quantified and plotted on correlation diagrams. While the best reproducibility was achieved using replicates of either plasma or serum, very good correlations in expression were also observed for miRNAs identified from both sample types.

Conclusions

- Using RNA 3′ polyadenylation and SMART template-switching technology, we developed a small RNA library preparation method that provides greater accuracy than approaches involving adapter ligation.
- In addition to capturing miRNAs, the SMARTer smRNA-Seq Kit for Illumina enables analysis of piRNAs, snoRNAs, snRNAs, etc.
- Comparable sequencing results were obtained across a 1 ng–2 µg input range of total RNA.
- Our approach successfully generated high-quality libraries from subnanogram amounts of RNA extracted from 200 µl of human plasma or serum.
- The number of unique miRNAs detected in plasma RNA was higher than the number detected in serum RNA, but strong correlations in miRNA expression were observed for comparisons involving either or both sample types.
- Data generated with the SMARTer smRNA-Seq Kit for Illumina exhibits the robustness necessary for biomarker discovery from clinical samples.

Scan to download your copy of this poster and view additional posters, or visit http://info.clontech.com/smRNA_poster_2016.html

800.662.2566
Visit us at www.clontech.com/ngs

University of Massachusetts Medical School

Clontech TaKaRa cellartis

09.16 US (633835)