Comparing the performance of RNA-seq kits for inputs of varying sample type and quality

Data kindly provided by Sri Ganesh Jammula, Ph.D., Karin Johanna Ferrari, Ph.D., Andrea Scelfo, Ph.D., Alessandra Rossi, Ph.D., and Diego Pasini, Ph.D., Department of Experimental Oncology, European Institute of Oncology

- In a customer-conducted performance analysis with standard RNA control samples, Takara Bio and Illumina RNA-seq kits yielded consistent and comparable sequencing metrics.
- Strong correlations were also observed between kits for data generated from varying input amounts of high-quality and partially degraded RNA.
- According to customer feedback, Takara Bio's SMARTer Stranded RNA-Seq Kit performed better than Illumina's TruSeq® RNA Sample Preparation Kit v2 for both intact and partially degraded mouse RNA inputs.
- The SMARTer Stranded RNA-Seq Kit features a fast and easy workflow of ~4 hours, which can quickly generate results from even the most challenging samples.

Introduction

RNA-seq has become an essential tool for analyzing transcriptional differences between cell populations and individual cells. To empower accurate transcriptome analyses in a range of biomedical research applications—including those in which RNA inputs are low or compromised—several protocols and advancements have been made for RNA-seq. This technical note features performance data generated for human and mouse RNA inputs of varying quality using four commercially available solutions for RNA-seq: the TruSeq Stranded Total RNA with Ribo-Zero™ Human/Mouse/Rat kit (Illumina), the TruSeq RNA Sample Preparation Kit v2 (Illumina), the SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian (Takara Bio), and the SMARTer Stranded RNA-Seq Kit (Takara Bio; RiboGone rRNA depletion technology is included in the high-input kit, and was used in conjunction with the SMARTer Stranded RNA-Seq Kit). The experiments were performed at the European Institute of Oncology and the Italian Institute of Technology Center for Genomic Science (both in Milan, Italy) under the direction of Dr. Diego Pasini, and by staff from the Takara Bio R&D department. The results indicate that the Takara Bio and Illumina kits yield consistent and comparable sequencing metrics for standard human RNA control samples, high-quality mouse embryonic stem cell RNA, and partially degraded mouse intestinal RNA. Additionally, analysis of partially degraded mouse intestinal RNA samples using the SMARTer Stranded RNA-Seq Kit revealed divergent expression levels for genes known to be differentially regulated between various intestinal cell types, evidence of the sensitivity afforded by this kit.

Results

Comparable sequencing metrics from experiments using standard RNA inputs

To compare the performance of the SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian and the TruSeq Stranded Total RNA with Ribo-Zero Human/Mouse/Rat, each kit was used to generate a sequencing library from 400 ng of Microarray Quality Control Consortium (MAQC) human total RNA-Human Universal Reference RNA (HURR) or Human Brain Reference RNA (HBRR)-spiked with External RNA Controls Consortium (ERCC) synthetic RNA, followed by sequencing and data analysis. The two methods yielded comparable sequencing metrics (**Table 1**) in terms of rRNA depletion, exon/intron mapping, and the number of genes identified at thresholds of 0.1 or 1 RPKMs (Reads Per Kilobase of Exon Per Million Fragments Mapped). Minor differences were observed between the data sets for various sequencing metrics, presumably because while both kits include similar cDNA-synthesis, rRNA-removal, and library-preparation steps, they employ different proprietary technologies.

Sequencing metrics for libraries generated from 400 ng total RNA inputs				
Human total RNA type	Universal (HURR)	Brain (HBRR)	Universal (HURR)	Brain (HBRR)
PCR cycles	12		15	
Library prep kit used	SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian with RiboGone - Mammalian		TruSeq Stranded Total RNA with Ribo-Zero Human/Mouse/Rat	
Number of reads (millions)	8.5 (paired-end)		7.0 (paired-end)	
Percentage of reads (%)				
After trimming	97.8	94.4	99.4	99.3
Mapped to rRNA	0.28	5.36	0.6	0.24
Mapped to mitochondria	13	15	22	14
Mapped to genome	87	87	95	95
Mapped to exons	45	50	54	41
Mapped to introns	45	41	40	49
Mapped to intergenic regions	11	9.5	6.4	9.2
Number of genes with at least 0.1 RPKM	25,465	24,026	24,578	24,587
Number of genes with at least 1.0 RPKM	15,672	14,960	15,975	15,624
Percentage of duplicates (%)	24	23	23	13
Biological strandedness (%)	98.4	96.4	99.4	98.9
ERCC strandedness (%)	99.3	98.8	99.9	100
Percentage of reads mapped to ERCC (%)	0.54	0.54	1	0.9

Table 1. Sequence alignment metrics. 400 ng of Human Universal Reference RNA (HURR) or Human Brain Reference RNA (HBRR) spiked with ERCC RNA was processed with each kit. Alignment data is displayed for each library, with the percentage of reads that mapped to rRNA, exons, introns, intergenic regions, and the correct strand, etc., as defined by Picard analysis.

In addition to yielding similar alignment metrics, RNA-seq data generated from HURR and HBRR samples with each kit correlated strongly (R^2 >0.8) with MAQC-generated data (**Figure 1**). This concordance speaks to the robustness and comparable performance of both kits.

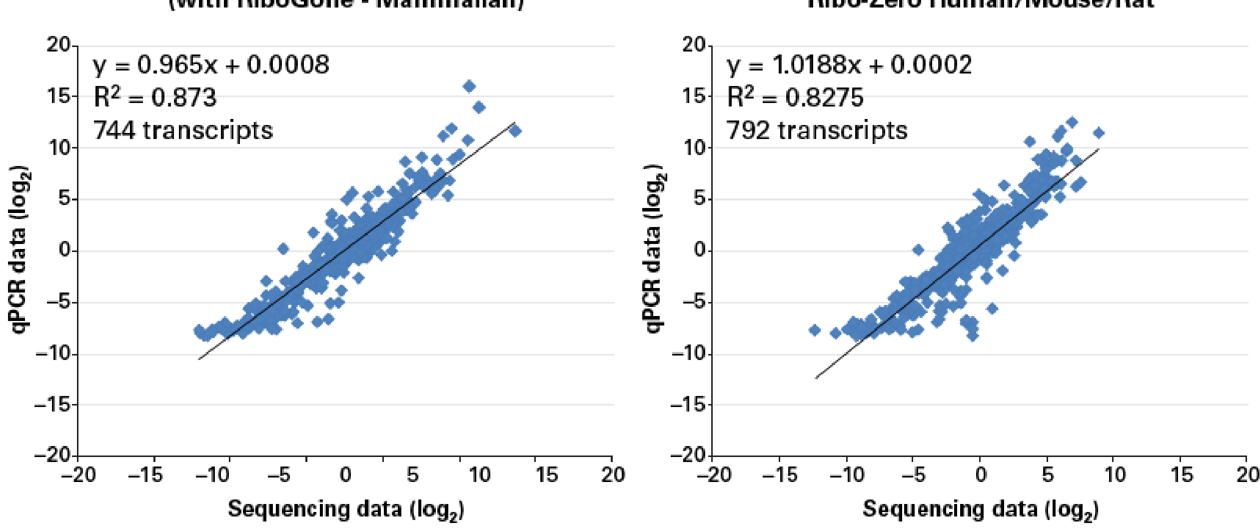

Figure 1. Correlation between RNA-seq data and MAQC data. RNA-seq libraries were independently generated from 400 ng of HURR or HBRR using each kit. Each scatter plot shows the $\log_2$ of the ratio of HURR/HBRR RPKMs graphed against the $\log_2$ of the ratio of HURR/HBRR expression data derived from qPCR Taqman probes. **Panel A.** Correlation for data generated with the SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian. **Panel B.** Correlation for data generated with the TruSeq Stranded Total RNA with Ribo-Zero Human/Mouse/Rat kit.

Testing the efficiencies of two different library preparation methods using high-quality RNA inputs

High-quality RNA isolated from mouse embryonic stem cells was used to compare the efficiencies of the SMARTer Stranded RNA-Seq Kit, and the TruSeq RNA Sample Preparation Kit v2 (**Figure 2**). Data Set A was obtained from a sequencing library prepared using the Illumina kit with an input consisting of poly(A)-enriched mRNA derived from 1 µg of total RNA. Data Sets B and C were obtained from sequencing libraries generated from 10 ng and 100 ng, respectively, of total RNA using the Takara Bio kit. In summary, mRNA expression data for both libraries generated using the Takara Bio kit correlated strongly with data from the library generated using the Illumina kit ($R^2 > 0.9$ for each comparison), and the kits yielded comparable transcript mapping results. In particular, there was considerable overlap for the 25% most highly expressed transcripts identified in each data set, as indicated by the Venn diagram. These results suggest that the Takara Bio kit may offer greater efficiency than the Illumina kit, in that the former can generate comparable sequencing results from a lower input amount.

A

$R^2 = 0.907$

Data Set B (10 ng, Takara Bio)

Data Set A (1 µg, Illumina)

B

$R^2 = 0.942$

Data Set C (100 ng, Takara Bio)

Data Set A (1 µg, Illumina)

C

Overlap between data sets (top 25% most highly expressed genes)

Data Set B (10 ng, Takara Bio)

551

587

48

6559

221

760

378

Data Set A (1 µg, Illumina)

Data Set C (100 ng, Takara Bio)

Figure 2. Assessing the efficiency of the SMARTer Stranded RNA-Seq Kit for sequencing high-quality input RNA. RNA-seq libraries were generated from either total or poly(A)-enriched mouse embryonic stem cell RNA, using either the SMARTer kit (paired with RiboGone - Mammalian), or the Illumina TruSeq RNA Sample Preparation Kit v2. **Panel A.** Correlation of normalized mRNA expression data (log-transformed RPKMs) for the Illumina TruSeq library (Data Set A) vs. the Takara Bio SMARTer library (10 ng total RNA input, Data Set B). Data points are color-coded according to relative expression levels in the TruSeq dataset (red: 75th–100th percentile, green: 50th–75th percentile, blue: 25th–50th percentile, purple: 0–25th percentile). The correlation coefficient (R^2) for the two data sets is indicated in the top left corner of the graph. **Panel B.** Correlation of normalized mRNA expression data for the Illumina TruSeq library (Data Set A) vs. the Takara Bio SMARTer library (100 ng total RNA input, Data Set C). Data points are color coded as in Panel A. The correlation coefficient (R^2) for the two data sets is indicated in the top left corner of the graph. **Panel C.** Venn diagram demonstrating overlap for the 25% most highly expressed transcripts in each dataset.

Data reproducibility across varying input amounts

Expression data for the two SMARTer libraries generated from 10 ng or 100 ng inputs of high-quality mouse embryonic stem cell RNA were compared directly (**Figure 3**). The strong correlation observed ($R^2 = 0.971$) suggests that the SMARTer Stranded RNA-Seq Kit generates consistent results for varying input amounts of total RNA. In addition, the strand specificity of sequencing reads was analyzed by measuring the frequencies with which the reads indicated a gene's annotated strand of origin ("specific strand") vs. the complementary strand ("opposite strand"). The results confirmed that most reads mapped to the annotated strand regardless of the input amount (100 ng vs. 10 ng), suggesting that sequencing libraries produced with the SMARTer Stranded Total RNA-Seq Kit reliably capture strand-of-origin information.

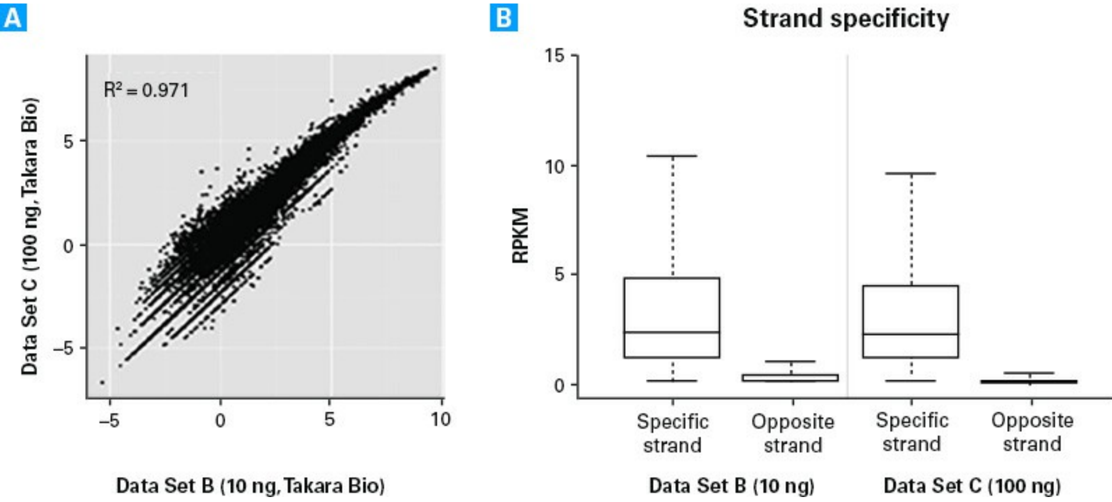

Figure 3. Data reproducibility across varying input amounts. Data sets for RNA-seq libraries generated from 10 ng or 100 ng of mouse embryonic stem cell total RNA using the SMARTer Stranded RNA-Seq Kit were compared. **Panel A.** Correlation of normalized mRNA expression data (log-transformed RPKMs) for 10 ng input (Data Set B) vs. 100 ng input (Data Set C). The correlation coefficient (R^2) for the two data sets is indicated in the top left corner of the graph. **Panel B.** Strand-specificity data for reads mapped to annotated genes in each data set. The boxplots summarize the normalized quantities of reads that identified the annotated strand of origin ("specific strand") vs. the corresponding complementary strand ("opposite strand") for each gene, for both data sets.

Testing the efficiencies of two different library preparation methods using partially degraded RNA inputs

RNA isolated from partially degraded mouse intestinal tissue was used to compare the efficiencies of the Illumina and Takara Bio kits in processing compromised samples (**Figure 4**). Data Set D was obtained from a sequencing library prepared using the Illumina TruSeq RNA Sample Preparation Kit v2. The input for this data set was poly(A)-enriched mRNA derived from 1 µg of total RNA. Data Set E was obtained from a sequencing library generated from 100 ng of total RNA using the SMARTer Stranded RNA-Seq Kit paired with RiboGone - Mammalian. Expression data from the two libraries were strongly correlated (R^2 = 0.948), and considerable overlap was observed for the 25% most highly expressed transcripts identified in each data set, as indicated by the Venn diagram. As with the above comparison involving high-quality mouse RNA, these results suggest that the Takara Bio kit may offer greater efficiency than the Illumina kit.

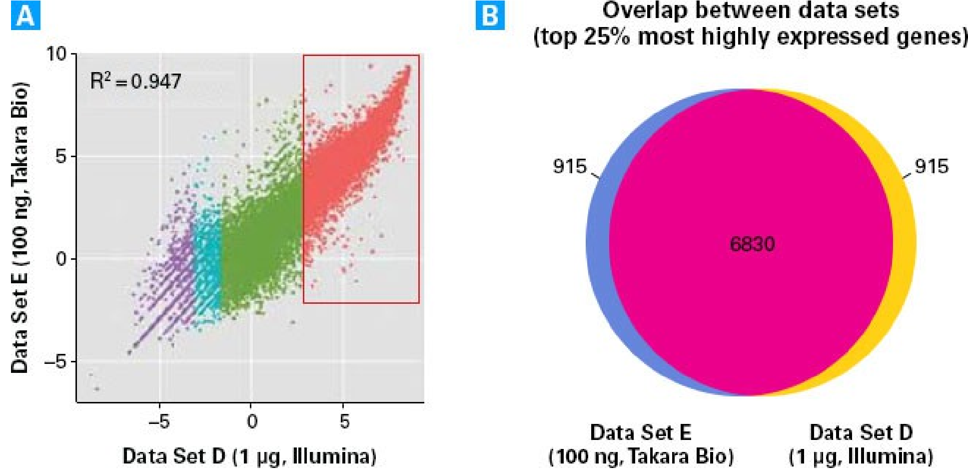

Figure 4. Assessing the efficiency of the SMARTer Stranded RNA-Seq Kit for sequencing of partially degraded input RNA. RNA-seq libraries were generated from partially degraded mouse intestinal RNA using either the SMARTer kit, or the Illumina TruSeq RNA Sample Preparation Kit v2. **Panel A.** Correlation of normalized expression data (log-transformed RPKMs) for the Illumina TruSeq library (Data Set D) vs. the Takara Bio SMARTer library (Data Set E). Data points are color-coded according to relative expression levels in the TruSeq dataset (red: 75th-100th percentile, green: 50th-75th percentile, blue: 25th-50th percentile, purple: 0-25th percentile). The correlation coefficient (R^2) for the two data sets is indicated in the top left corner of the graph. **Panel B.** Venn diagram demonstrating overlap for the 25% most highly expressed transcripts in each dataset.

Detecting tissue-specific differences in gene expression from partially degraded RNA inputs

To further test the capabilities of the SMARTer Stranded RNA-Seq Kit in processing partially degraded samples, sequencing libraries were generated from RNA inputs derived from the small intestine and colon (Data Sets E and F, respectively). While expression levels for many genes were correlated (**Figure 5**), the analysis also identified genes that were differentially expressed in the respective samples, and confirmed that much of this divergence was consistent with known transcriptomic differences between the small intestine and colon. As was the case for data generated from high-quality RNA inputs, analysis of strand specificity for the small intestine and colon data sets confirmed that the vast majority of sequencing reads correctly indicated the annotated strand of origin for the corresponding genes to which they mapped. These results demonstrate that the SMARTer Stranded RNA-Seq Kit successfully preserves strand-of-origin information for compromised RNA inputs.

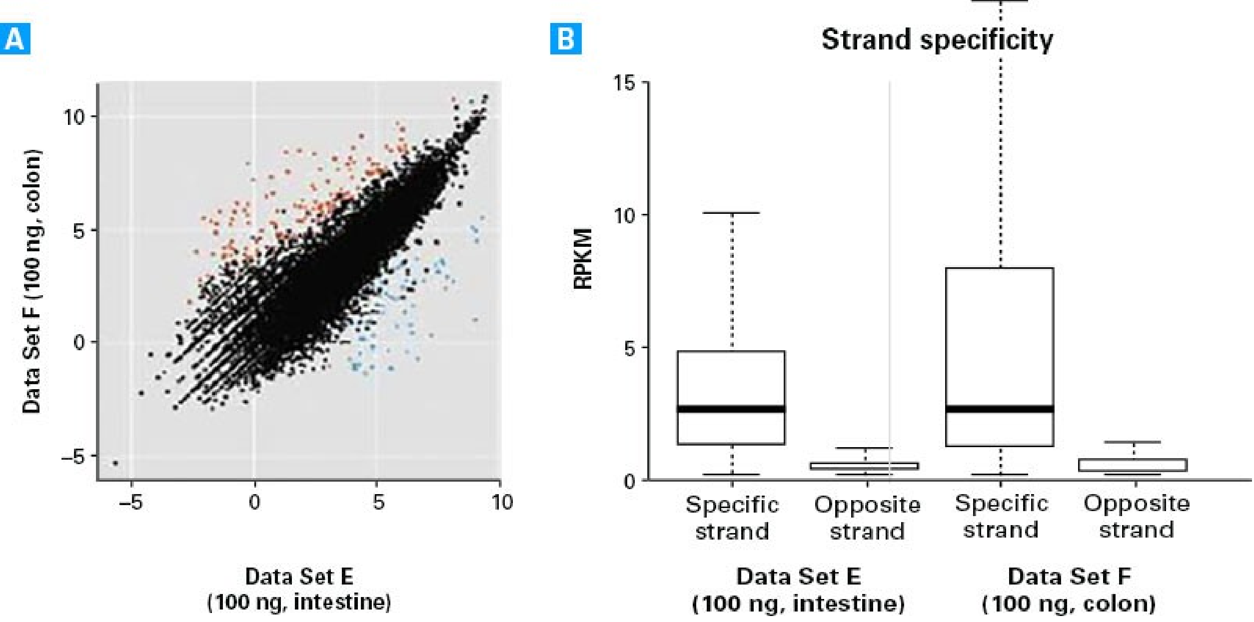

Figure 5. Observing tissue-specific differences in gene expression and obtaining strand-of-origin information from partially degraded RNA inputs. RNA-seq libraries were generated from 100 ng inputs of partially degraded RNA from mouse small intestine and colon (Data Sets E and F, respectively) using the SMARTer Stranded RNA-Seq Kit. **Panel A.** Correlation of normalized mRNA expression data (log-transformed RPKMs) for Data Set E vs. Data Set F. Differentially expressed genes attributed with elevated expression in the intestine and colon are indicated in red and blue, respectively. **Panel B.** Strand-specificity data for reads mapped to annotated genes in each data set. The boxplots summarize the normalized quantities of reads that identified the annotated strand of origin ("specific strand") vs. the corresponding complementary strand ("opposite strand") for each gene, for both data sets.

Summary

A customer comparison of Takara Bio and Illumina RNA-seq library preparation kits indicated that they yield comparable sequencing results for a range of RNA inputs, including control human RNA, high-quality mouse RNA, and degraded mouse RNA. Included in the comparisons were the SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian (Takara Bio), the SMARTer Stranded RNA-Seq Kit (Takara Bio), the TruSeq Stranded Total RNA with Ribo-Zero Human/Mouse/Rat kit (Illumina), and the TruSeq RNA Sample Preparation Kit v2 (Illumina). Whereas library preparation with the Takara Bio and Illumina stranded kits involved depletion of rRNA from total RNA inputs, mRNA processed with the Illumina RNA Sample Preparation Kit v2 was poly(A)-enriched. Correspondingly, for the comparisons involving mouse RNA inputs, the input amount used with the Illumina TruSeq RNA Sample Preparation Kit v2 (1 µg) was much greater than the input amounts used with the Takara Bio kit (10 ng or 100 ng). According to the customer who conducted the study, the Takara Bio SMARTer Stranded RNA-Seq Kit performed better overall than the Illumina TruSeq RNA Sample Preparation Kit v2 for both high-quality and partially degraded inputs of mouse RNA (data not shown). Importantly, known differences in gene expression between tissues of the mouse intestine and colon were represented in sequencing libraries generated from partially degraded RNA samples using the Takara Bio kit. For both mouse inputs tested, the Takara Bio kit captured strand-of-origin information with a high degree of accuracy. Taken together, these data suggest that the SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian and the SMARTer Stranded Total RNA-Seq Kit provide more efficient and robust methods for sequencing library preparation compared to available alternatives.

Methods

- Kits used were the SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian (Cat. No. 634873), the SMARTer Stranded RNA-Seq Kit

(Takara Bio, Cat. No. 634836), RiboGone - Mammalian (Takara Bio, Cat. No. 634846), the TruSeq Stranded Total RNA with Ribo-Zero Human/Mouse/Rat (Illumina, Cat. No. RS-122-2201), and the TruSeq RNA Sample Preparation Kit v2 (Illumina, Cat. No. RS-122-9001).

- RNA inputs included Human Universal Reference RNA (Agilent, Cat. No. 740000), FirstChoice Human Brain Reference RNA (Thermo Fisher Scientific, Cat. No. AM6050), and ERCC Spike-In Mix (Thermo Fisher Scientific, Cat. No. 4456740).
- All libraries in this study were sequenced with paired-end reads (1 x 50 bp) on an Illumina HiSeq™ 2000 instrument.
- Sequencing data analysis was done using CLC Genomics Workbench.

Related Products

Cat. #	Product	Size	License	Quantity	Details
634836	SMARTer® Stranded RNA-Seq Kit	12 Rxns		*	

The SMARTer Stranded RNA-Seq Kit includes the components needed to generate indexed cDNA libraries suitable for next-generation sequencing (NGS) on any Illumina platform, starting from as little as 100 pg of polyA-purified or ribosomal RNA-depleted RNA. The kit consists of the SMARTer Stranded RNA-Seq Components, SeqAmp DNA Polymerase, and the llumina Indexing Primer Set (PCR primers for the amplification of indexed, paired-end Illumina-compatible sequencing libraries, which enable multiplexing of NGS library analysis).

The SMARTer Stranded RNA-Seq Kit utilizes our patented SMART (**S**witching **M**echanism **At** 5' end of **R**NA **T**emplate) technology, coupled with PCR amplification, to generate Illumina-compatible libraries without the need for enzymatic clean-up or adapter ligations. The directionality of the template-switching reaction preserves the strand orientation of the RNA, making it possible to obtain strand-specific sequencing data from the synthesized cDNA.

Documents	Components	You May Also Like	Image Data	Resources

Cat. #	Product	Size	License	Quantity	Details
634837	SMARTer® Stranded RNA-Seq Kit	24 Rxns		*	
634838	SMARTer® Stranded RNA-Seq Kit	48 Rxns		*	
634839	SMARTer® Stranded RNA-Seq Kit	96 Rxns		*	
634861	SMARTer® Stranded Total RNA Sample Prep Kit - Low Input Mammalian	24 Rxns		*	
634862	SMARTer® Stranded RNA-Seq Kit HT	96 Rxns		*	
634846	RiboGone™ - Mammalian	6 Rxns		*	
634847	RiboGone™ - Mammalian	24 Rxns		*	
634873	SMARTer® Stranded Total RNA Sample Prep Kit - HI Mammalian	12 Rxns		*	
634874	SMARTer® Stranded Total RNA Sample Prep Kit - HI Mammalian	24 Rxns		*	
634875	SMARTer® Stranded Total RNA Sample Prep Kit - HI Mammalian	48 Rxns		*	
634876	SMARTer® Stranded Total RNA Sample Prep Kit - HI Mammalian	96 Rxns		*	
634877	SMARTer® Stranded Total RNA Sample Prep Kit - HI Mammalian	192 Rxns		*	
634878	SMARTer® Stranded Total RNA Sample Prep Kit - HI Mammalian	480 Rxns		*	

Add to Cart

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999
FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES. © 2023 Takara Bio Inc. All Rights Reserved. All trademarks are the property of Takara Bio Inc. or its affiliate(s) in the U.S. and/or other countries or their respective owners. Certain trademarks may not be registered in all jurisdictions. Additional product, intellectual property, and restricted use information is available at takarabio.com.