SMARTer® Stranded RNA-Seq Kit

Directional RNA-Seq Data from Small Samples

that's GOOD science!

- **Superior performance:** Get uniform transcriptome coverage, high reproducibility, and high correlation across a broad input range (100 ng–100 pg RNA)
- **99% accurate strand identification:** Determine every transcript's strand of origin
- **Simplified workflow:** Produce an Illumina®-ready library in less than 4 hours
- **Versatile kit:** Suitable for eukaryotic and prokaryotic samples, and for coding and noncoding RNA-Seq

Introduction

Although Next Generation Sequencing (NGS) is an extremely powerful set of technologies, its short reads can make it difficult to determine which gene or noncoding sequence a particular read comes from. Strand-specific sequencing maintains the orientation of each read, allowing overlapping reads to be distinguished from each other (Figure 1; compare Panels A and B).

There are three areas where strandedness plays a critical role in NGS: Identifying antisense transcripts, noncoding RNAs, and closely spaced or overlapping genes (Figure 1; 1). The latter case is particularly important for non-model organisms (with un-annotated transcriptomes or where annotation is incomplete) and for prokaryotic organisms, which have very compact genomes.

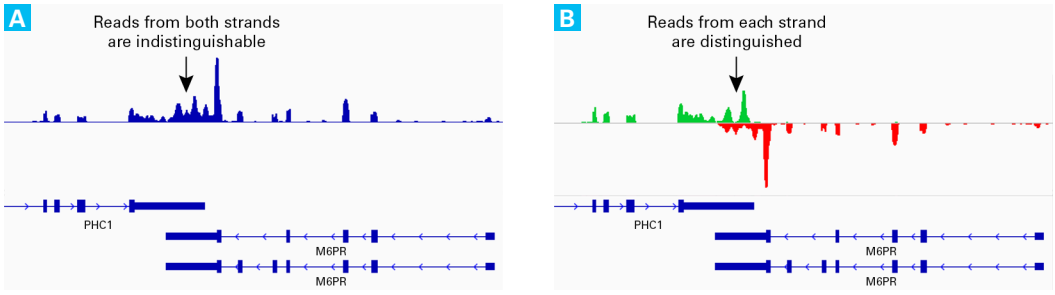

Figure 1. Capturing strand-specific information allows you to assign RNA-Seq reads to the correct strand of DNA. Panel A. Non-stranded RNA-seq cannot distinguish the reads from RNA transcribed from different strands of the same genomic sequence. **Panel B.** The SMARTer Stranded RNA-Seq Kit produces libraries that can be assigned to a strand with ≥99% accuracy. Short, overlapping sequences that originate from different DNA strands are distinguished from each other, enabling quantitative expression analysis and accurate genome annotation.

Clontech Laboratories, Inc. • A Takara Bio Company
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.543.7247

The SMARTer Stranded RNA-Seq Kit generates indexed cDNA libraries that are suitable for RNA-Seq on any Illumina platform. The protocol has been designed for ease of use and direct addition of adapters, so library construction can be completed in less than 4 hours (Figure 2). By comparison, other commercially-available methods require 1–2 full days of effort to prepare libraries for RNA-Seq.

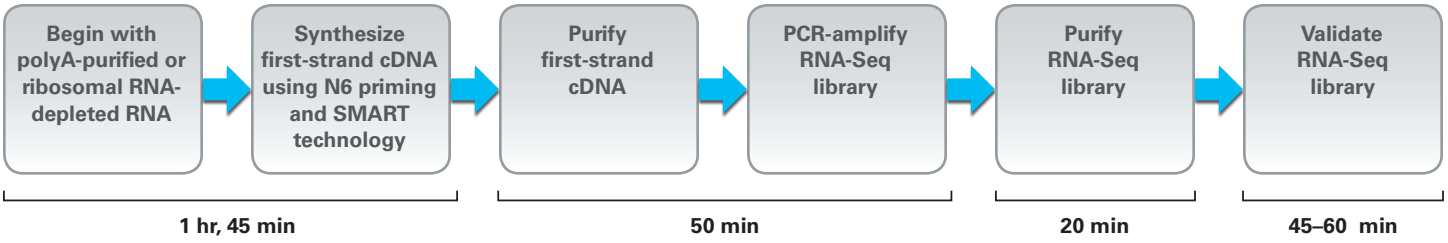

Figure 2. NGS Work flow for the SMARTer Stranded RNA-Seq Kit. The SMARTer Stranded RNA-Seq Kit synthesizes first-strand cDNA from as little as 100 pg of polyA-purified or ribosomal RNA-depleted RNA. cDNA is then PCR-amplified to generate Illumina-compatible libraries, without enzymatic clean-up or adapter ligation steps.

SMART® cDNA synthesis protocol for directional RNA-Seq

Clontech's patent-protected SMART technology (**S**witching **M**echanism **A**t the 5' end of the **R**NA **T**emplate) is an exceptionally efficient method for accurately synthesizing full-length cDNA in a single reaction. This method relies on the template switching activity of reverse transcriptase to enable the direct addition of a PCR adaptor to the 3' end of the first strand cDNA, thus avoiding inefficient ligation steps.

In the SMARTer Stranded RNA-Seq Kit, 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 (Figures 3 and 8). Importantly, the SMARTer Stranded RNA-Seq Kit produces whole transcriptome coverage without 5' or 3' bias, and yields excellent reproducibility and sensitivity, mappability, and ERCC correlation.

Directional cDNA library production from human brain RNA

A stock solution of Human Brain Poly A$^+$ RNA (Clontech®, Cat. No. 636102) spiked with ERCC (External RNA Controls Consortium) control RNA was serially diluted to prepare RNA samples containing between 100 ng–100 pg human brain polyA$^+$ RNA. cDNA libraries were successfully prepared in triplicate according to the SMARTer Stranded RNA-Seq Kit protocol, using twelve different Illumina indices. All libraries showed comparable yields and purity irrespective of input RNA concentrations (Figure 4). Libraries were sequenced on an Illumina HiSeq® 2000 instrument, with ~300M 2 x 100 bp paired end reads.

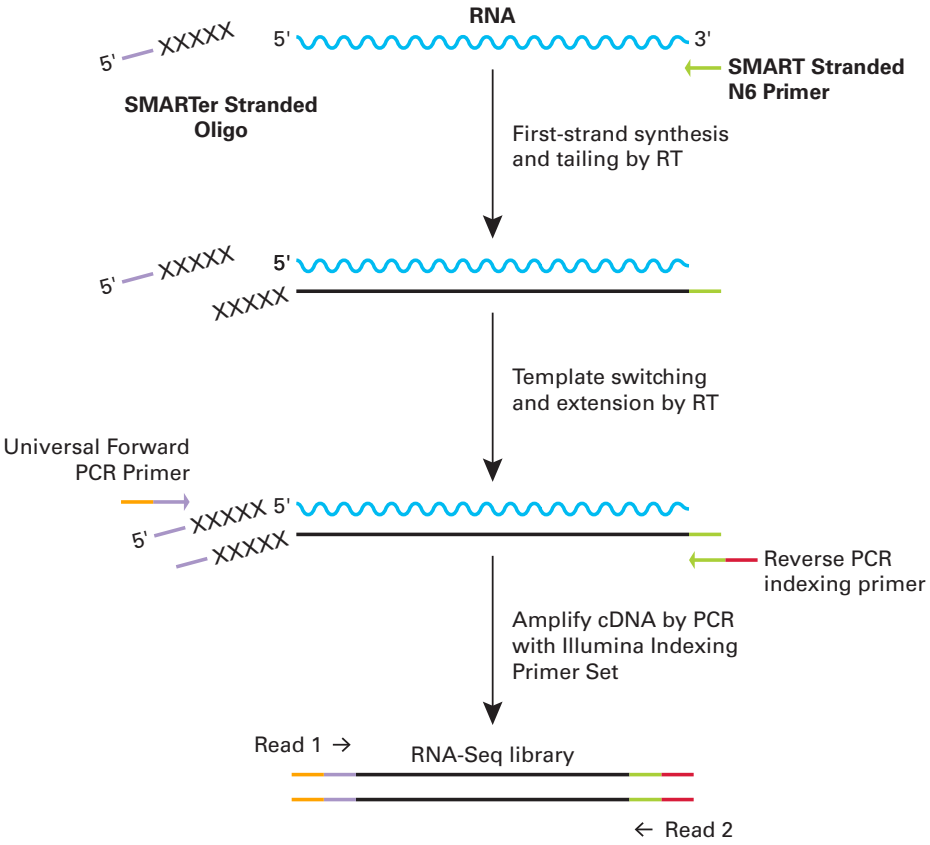

Figure 3. Flowchart of SMARTer Stranded RNA-Seq library generation. The SMARTer Stranded RNA-Seq Kit starts with as little as 100 pg of RNA and utilizes the proprietary SMART Stranded N6 Primer. SMARTScribe™ Reverse Transcriptase performs first strand synthesis and tailing. When SMARTScribe RT reaches the 5' end of the RNA fragment, its terminal transferase activity adds a few additional nucleotides to the 3' end of the cDNA. The SMARTer Stranded Oligo base-pairs with this non-template nucleotide stretch, creating an extended template to enable SMARTScribe RT to continue replicating to the end of the oligonucleotide (2). The resulting full-length, single-stranded (ss) cDNA contains the complete 5' end of the mRNA, as well as sequences that are complementary to the SMARTer Stranded Oligo.

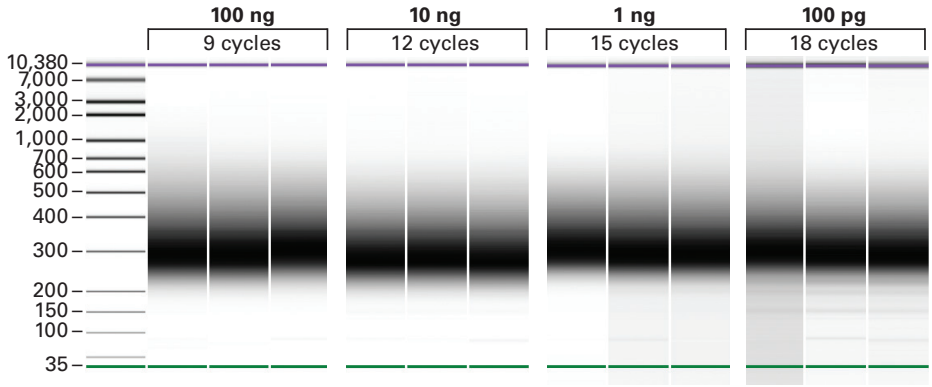

Figure 4. cDNA libraries produced using the SMARTer Stranded RNA-Seq Kit show comparable yields and purity irrespective of input RNA concentrations. Agilent 2100 Bioanalyzer gel-like image using a High Sensitivity DNA chip with cDNA libraries produced from between 100 ng–100 pg human brain polyA$^+$ RNA with the number of PCR cycles indicated. cDNA libraries were produced by amplification with SeqAmp DNA Polymerase (included in the SMARTer Stranded RNA-Seq Kit).

Highly reproducible and sensitive directional RNA-seq

Sequence data was compared for the libraries created from 100 ng and 100 pg of cDNA (Figure 5). The SMARTer Stranded RNA-Seq Kit yielded extremely linear and reproducible expression data over a thousand fold range of input RNA levels, down to 100 pg of polyA$^+$ RNA.

Data from technical replicates of each library indicated that the SMARTer Stranded RNA-Seq Kit delivers very consistent results across a wide range of input RNA concentrations. Pearson correlation data for expression levels (FPKM, fragments per kilobase of transcript per million mapped reads) shows high reproducibility for independent cDNA library replicates across a broad input range. Even at 100 pg input RNA, the results showed strong correlation (Figure 6).

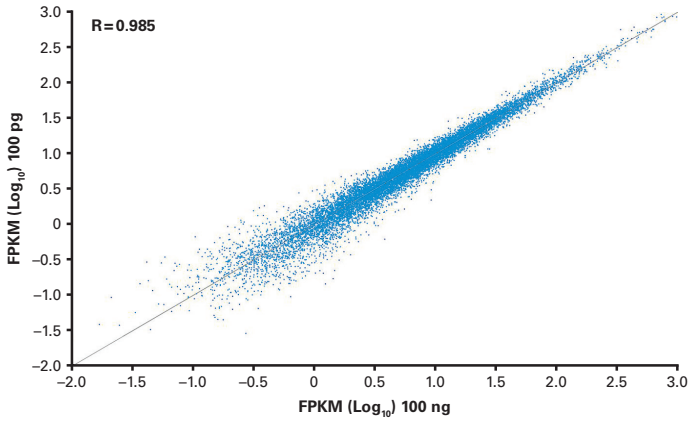

Figure 5. The SMARTer Stranded RNA-Seq Kit provides high reproducibility and sensitivity across a wide range of input RNA. Scatter plots of expression (FPKM) for cDNA libraries prepared from 100 ng and 100 pg of human brain polyA$^+$ RNA show a high correlation, suggesting consistency across input levels. Axes are plotted on a $\log_{10}$ scale. Insets indicate the Pearson coefficient of correlation between replicates (R).

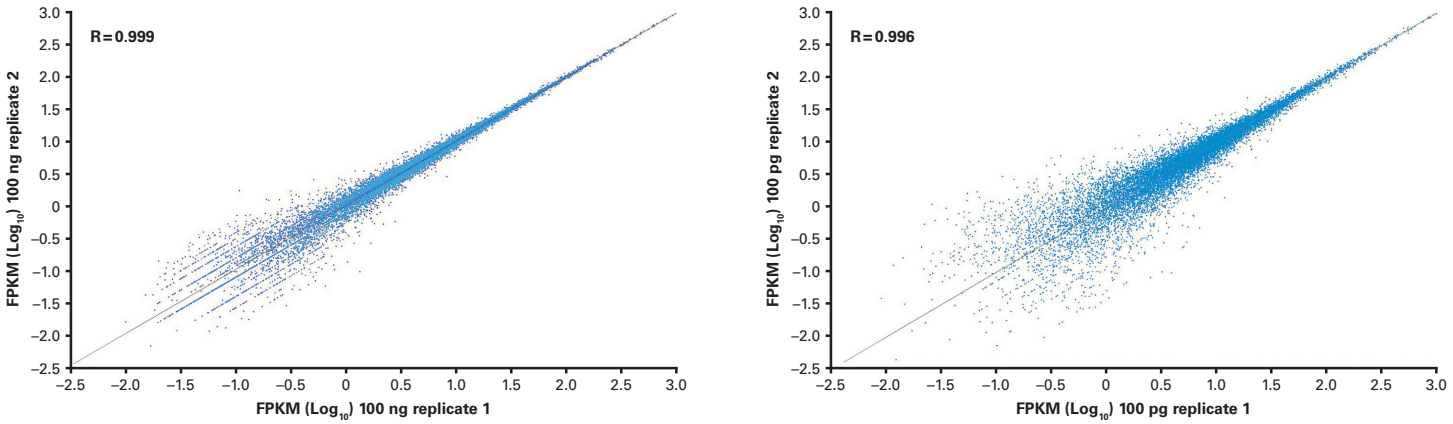

Figure 6. SMARTer Stranded RNA-Seq Kit provides high reproducibility across a wide range of RNA inputs. Comparisons of pairs of cDNA library replicas created from 100 ng and 100 pg of input RNA show high reproducibility across a wide range of input RNA. Axes are plotted on a $\log_{10}$ scale. Insets indicate the Pearson coefficient of correlation between replicates (R).

Reads from the previous experiments that mapped to the ERCC data set (External RNA Control Consortium; 3) were pooled together and FPKM was plotted against relative transcript abundance. The analysis showed a high degree of reproducibility between duplicate libraries (Figure 7). All 92 ERCC control transcripts were detected, and the expression levels determined by sequencing were consistent with the various levels of the control transcripts spiked in. Even at the lowest levels tested (100 pg polyA$^+$ input RNA), the SMARTer Stranded RNA-Seq Kit was able to maintain high accuracy (Figure 7) while detecting ~15,000 genes using strand-specific alignments (Table I).

Using the SMARTer Stranded RNA-Seq Kit to identify and analyze directional sequences

RNA-Seq libraries created using the SMARTer Stranded RNA-Seq Kit display over 99% strand accuracy, as indicated by analysis of the reads that mapped to the spiked-in ERCC control RNA. Figure 8 is an example of the improved discrimination that stranded data provides.

Table I: Sequence Alignment Metrics*

Input RNA	100 ng	10 ng	1 ng	100 pg
% pairs mapped to RefSeq	76.6%	76.4%	77.6%	76.2%
% pairs uniquely mapped to RefSeq	74.4%	74.1%	75.4%	74.0%
Total exons	64.9%	65.2%	64.4%	66.4%
Total introns	35.1%	34.8%	35.6%	33.6%
Genes identified (>0.1 FPKM)	15,632	15,642	15,568	14,993
Number of reads	23,775,433	26,176,325	27,720,909	22,280,680

* All data shown is based on strand-specific alignment to RefSeq genes.

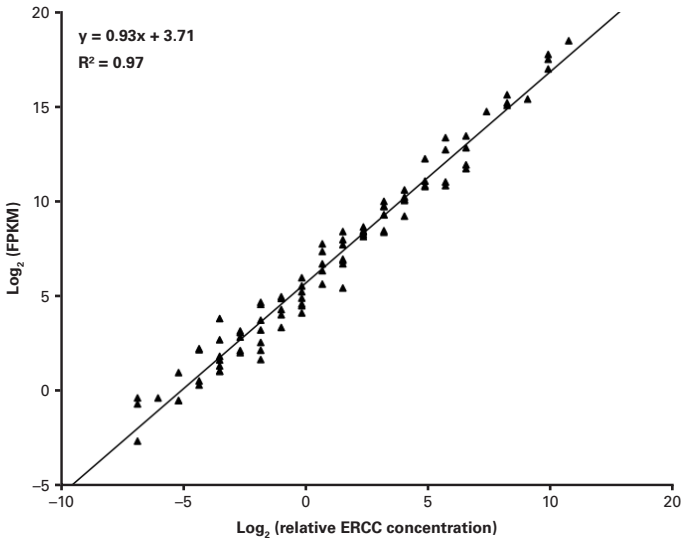

Figure 7. High reproducibility confirmed by ERCC analysis. Reads mapping to the ERCC data set from the previous experiments were pooled together and the FPKM was plotted against relative transcript abundance. The complete set of 92 ERCC transcripts was identified, with a slope of 0.934 and R^2 of 0.9725. Axes are plotted on a $\log_2$ scale.

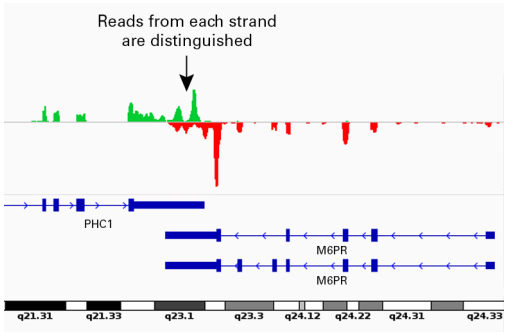

Figure 8. >99% of RNA-Seq reads map to the correct strand. Indexed cDNA libraries were prepared according to the SMARTer Stranded RNA-Seq Kit protocol using twelve indices, and sequenced on an Illumina HiSeq Platform with 2 x 100 bp paired end reads. Short, overlapping reads originating from different strands of the genomic DNA were distinguished from each other, enabling quantitative expression analysis and accurate genome annotation.

Summary

The SMARTer Stranded RNA-Seq Kit provides a simple and efficient solution for generating 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. Unlike other stranded RNA-Seq kits, the SMARTer Stranded RNA-Seq Kit integrates library preparation with cDNA amplification, so you can go from rRNA-cleared RNA to Illumina sequencing in less than four hours (Figures 2 and 3). Other methods typically require a separate, lengthy library prep protocol after cDNA synthesis and purification, for an overall two-day workflow.

The SMARTer Stranded RNA-Seq Kit uses a proprietary SMART Stranded N6 primer that provides uniform transcript coverage. In addition to a simplified workflow, the kit provides very high quality data and outperforms existing methods in reproducibility, mappability and ERCC correlation (Figures 4–7 and Table I). Furthermore, the SMARTer Stranded RNA-Seq Kit provides the same robust performance across a wide dynamic range, even when starting with low amounts of total RNA. The SMARTer Stranded RNA-Seq Kit is ideal for most RNA-Seq applications, including antisense transcript and noncoding RNA identification, and for identifying the source of overlapping reads (Figure 8), such as reads from closely spaced or overlapping genes, from closely related genes, or from prokaryotic organisms.

For more information about this and other kits for NGS, please visit: www.clontech.com/NGS

References

1. Levin, J., Yassour M., *et al.* (2010) *Nature Methods* **7**(9):709–715.
2. Chenchik, A. *et al.* (1998) In RT-PCR Methods for Gene Cloning and Analysis. Eds. Siebert, P. & Larrick, J. (BioTechniques Books, MA), pp. 305–319.
3. Jiang, L. *et al.* (2011) *Genome Res.* **21**(9):1543–1551

PRODUCTS

Cat. #	Product	Package Size
634836	SMARTer Stranded RNA-Seq Kit	12 Rxns
634837		24 Rxns
634838		48 Rxns
634839		96 Rxns
638504	SeqAmp DNA Polymerase	50 rxns

Learn more

www.clontech.com/NGS