TECH NOTE

Improving the Sensitivity of Ultra-Low Input mRNA-Seq

SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing

Powered by SMART and LNA technologies:
Locked nucleic acid technology significantly improves template switching and sensitivity >>

Higher sensitivity and improved mapping to the genome:
Better sequencing metrics than SMARTer Ultra Low v3 and SMART-Seq2 >>

Reproducible results across input amounts:
Excellent sequencing libraries are produced across a 10 pg–10 ng input range >>

Overview

Good science is continuously looking for ways to improve the quality of data produced from precious samples. With this goal of constant improvement in mind, we have created the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing, the highest sensitivity single-cell mRNA-seq kit we have ever released. This kit has higher mapping to the genome, identifies more genes, and has better reproducibility than the SMART-Seq2 method or to any of the previous generations ofSMARTer Ultra Low kits.

For our previous ultra-low input mRNA-seq kit (the SMARTer Ultra Low Input RNA Kit for Sequencing - v3) we streamlined the protocol by removing a purification step, resulting in higher yield, and introduced a polymerase (SeqAmp DNA polymerase) better able to amplify GC-rich targets, increasing the representation of these transcripts. With the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing we further improved sensitivity by incorporating and improving upon the SMART-Seq2 method from Rickard Sandberg's lab at Ludwig Cancer Research (Picelli *et al.*, 2013). This new kit uses the SMARTScribe Reverse Transcriptase (RT), which is already optimized for template switching. The inclusion of locked nucleic acid (LNA) technology, from the SMART-Seq2 method, as part of the template-switching SMART-Seq v4 oligo likely stabilizes the interaction between the oligo and non-templated nucleotides added by the RT. Add to that our long history of optimizing template-switching technology, and you get the high yield and extreme sensitivity of this new kit. By building on our extensive experience with single-cell mRNA-seq, the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing allows researchers to capture transcriptome data with the highest confidence.

Significant Improvements in Data over Previous Methods

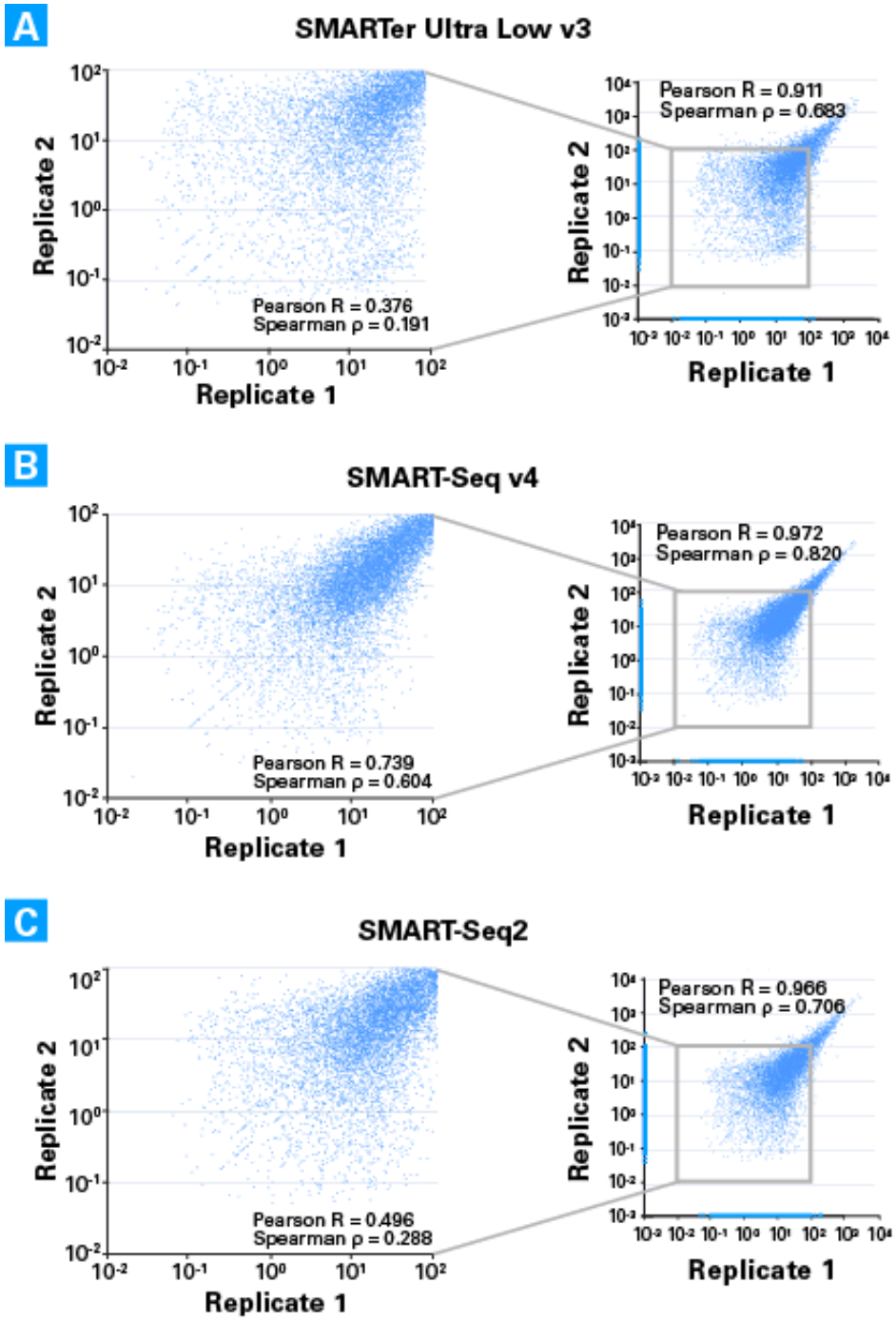
RNA-seq libraries generated using the SMART-Seq v4 kit outperform those made with either the SMARTer Ultra Low v3 kit or the SMART-Seq2 method. The cDNA library yield from the SMART-Seq v4 kit is consistently higher compared to the other two methods. A higher percentage of reads from the SMART-Seq v4 kit map to the genome compared to the SMART-Seq2 method, showing that improvements in the SMART-Seq v4 kit decrease background. Finally, the SMART-Seq v4 kit has extremely high sensitivity; the SMART-Seq v4 libraries identify more transcripts than either the SMARTer Ultra Low v3 kit or the SMART-Seq2 method.

Sequencing Metrics Comparing Different cDNA Synthesis Protocols							
RNA source		10 pg Mouse Brain Total RNA					
cDNA synthesis		SMARTer Ultra Low v3		SMART-Seq v4		SMART-Seq2	
Yield (ng)		4.7	6.0	10.6	11.2	12.6	8.1
Percentage of reads (%):							
Mapped to rRNA		0.8	0.4	6.5	6.1	6.9	3.8
Mapped to mitochondria		6.0	5.4	3.4	3.4	5.1	7.2
Mapped to genome		96	97	96	95	72	93
Mapped to exons		73	73	76	76	66	67
Mapped to introns		21	21	19	20	28	27
Mapped to intergenic regions		6.0	6.2	4.7	4.7	5.8	5.8
Number of genes	**FPKM >0.1**	11,647	10,885	14,731	14,813	12,080	12,039
	FPKM >1	9,729	9,105	12,501	12,591	10,270	10,058

Higher sensitivity and better mappability with the SMART-Seq v4 kit. Replicate libraries were generated from 10 pg Mouse Brain Total RNA using the SMART-Seq v4 kit, the SMARTer Ultra Low v3 kit, or the SMART-Seq2 method. 18 PCR cycles were used to amplify cDNA libraries with the SMART-Seq2 method and SMARTer Ultra Low v3 kit; however, only 17 PCR cycles were needed for the SMART-Seq v4 libraries. RNA-seq libraries were generated using Nextera® XT DNA Library Preparation Kit and sequenced on an Illumina® MiSeq® instrument. Sequences were analyzed as described in the methods.See Methods >>

Improved Reproducibility for Genes with Low Expression

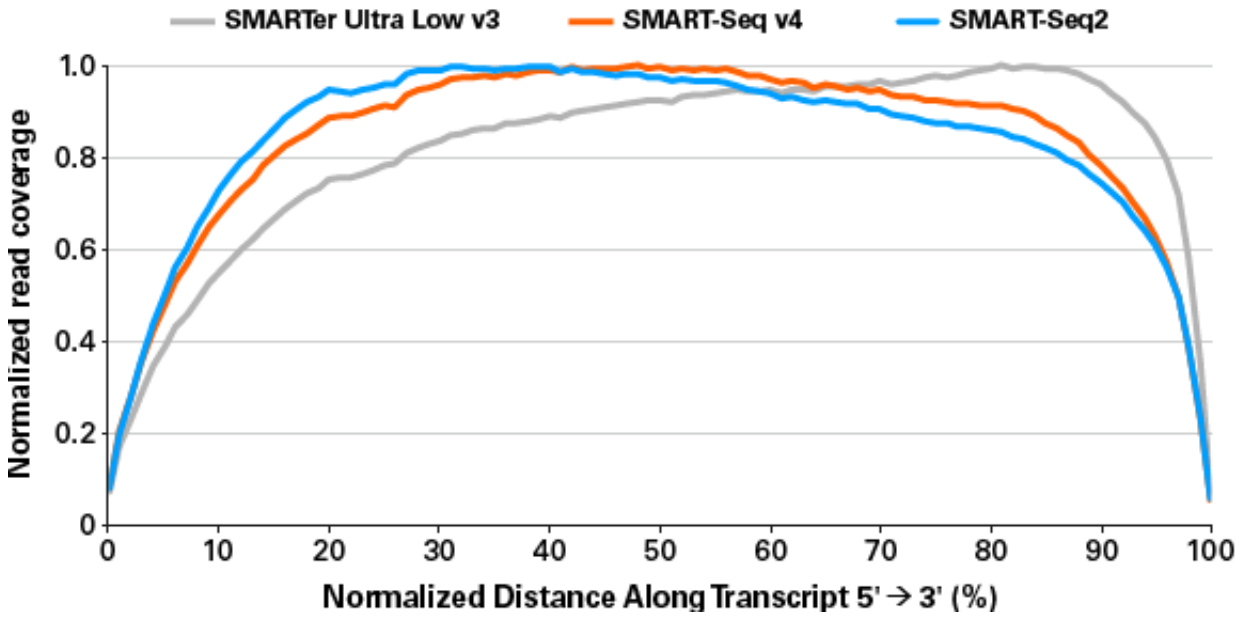
SMARTer Ultra Low kits and the SMART-Seq2 method are known to have good reproducibility, even at extremely low inputs, indicated by high Pearson correlations. The SMART-Seq v4 kit improves on this, in particular for transcripts with relatively lower expression levels. Across all expression levels (FPKMs; Fragments Per Kilobase Of Exon Per Million Fragments Mapped), correlation between replicates (Pearson R) ranges between 0.911–0.972 for the different cDNA synthesis methods. However, when high-expression transcripts (FPKM >100) are removed, the SMART-Seq v4 kit has much higher correlation between replicates (Pearson R = 0.739) compared to the SMARTer Ultra Low v3 kit or the SMART-Seq2 method. Additionally, fewer higher-expression transcripts are found in only one replicate of libraries generated with the SMART-Seq v4 kit. For the SMARTer Ultra Low v3 kit and SMART-Seq2 method, transcripts found only in one replicate have FPKMs of up to 364 or 236, respectively, while the highest FPKM of transcripts found in only one replicate is only 70 for the SMART-Seq v4 kit. This is consistent with the smaller variation across replicates and high reproducibility seen when using the SMART-Seq v4 kit.



Reproducibility is high for low-input samples. FPKMs from replicate libraries generated from 10 pg of Mouse Brain Total RNA using the SMART-Seq v4 kit, the SMARTer Ultra Low v3 kit, or the SMART-Seq2 method were compared. For transcripts with FPKM <100, the correlation between replicates was much higher for the SMART-Seq v4 kit (Pearson R = 0.739; **Panel B**) compared to the SMARTer Ultra Low v3 (Pearson R = 0.376; **Panel A**) or the SMART-Seq2 method (Pearson R = 0.496; **Panel C**). For all transcripts (shown in the scatterplots on the right) the correlation between replicates was high for each of the three methods (Pearson R between 0.911–0.972), though the SMART-Seq v4 kit did have the highest correlation. Transcripts represented in only one replicate can be seen along the X- and Y-axes of the scatter plots showing all transcripts.

Consistent Gene Body Coverage

One of the major advantages of using template switching and oligo(dT) priming for cDNA synthesis is that the final cDNA library is enriched for full-length transcripts. Ideally this results in even coverage across the gene body (i.e., a similar number of reads across the entire transcript). The SMART-Seq v4 kit has a slight 5′ bias, but generally even gene body coverage, comparable to the SMART Ultra Low v3 kit and SMART-Seq2 method.



Gene body coverage is good for all three library preparation methods. Gene body coverage shown is the average of two replicate libraries prepared from 10 pg Mouse Brain Total RNA using the three different cDNA synthesis methods. The SMARTer Ultra Low v3 kit produced a slight 3′ bias, and the SMART-Seq v4 kit produced a slight 5′ bias; however, the overall coverage was fairly even.

Excellent Metrics Across RNA Input Amount

The SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing produces libraries that identify a high number of transcripts, with very low percentages of reads mapping to rRNA, and high percentages mapping to the genome and to exons. It has been optimized for ultra-low-input total RNA samples, thus generating high-quality, consistent sequencing metrics across an input range of 10 pg–10 ng or from 1–1,000 intact cells (data not shown).

Sequencing Metrics Comparing Input RNA Amounts									
RNA source		Human Brain Total RNA							
Input amount		10 pg		100 pg		1 ng		10 ng	
Yield (ng)		3.2	3.2	5.2	5.6	3.5	4.4	6.7	6.1
Percentage of reads (%):									
Mapped to rRNA		1.6	1.4	1.9	1.9	3.83	3.8	11.8	11.6
Mapped to mitochondria		8.7	9.0	8.7	9.0	9.5	9.4	9.6	9.3
Mapped to genome		93	94	95	96	95	95	94	95
Mapped to exons		76	77	76	76	75	75	73	73
Mapped to introns		19	19	19	19	20	20	22	22
Mapped to intergenic regions		4.7	4.4	4.6	4.7	4.8	4.8	5.2	5.1
Number of transcripts	**FPKM >0.1**	14,169	22,367	22,367	22,531	24,449	24,469	24,524	24,522
	FPKM >1	11,627	11,227	16,460	16,600	17,218	17,119	17,383	17,333
Pearson R		0.96		0.99		1.00		1.00	

Sequencing metrics are consistent across RNA input amounts. 10 pg–10 ng of Human Brain Total RNA were used to generate cDNA libraries in duplicate with the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing. cDNA libraries were amplified using 17, 14, 10, or 7 PCR cycles for the 10 pg, 100 pg, 1 ng, or 10 ng libraries, respectively. RNA-seq libraries were generated using the Nextera XT DNA Library Preparation Kit and sequenced on an Illumina MiSeq instrument. Sequences were analyzed as described in the methods. See Methods >>

Summary

Good science incorporates and builds on new technologies in exciting ways. The SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing builds on three previous generations of SMARTer Ultra Low kits and exclusively incorporates the SMART-Seq2 method developed by Rickard Sandberg's lab to achieve a new level of sensitivity in single-cell RNA-seq. The improved template-switching oligonucleotide incorporates LNA technology, as well as other optimizations developed in-house, to increase the sensitivity and yield of this kit compared to previous generations of SMARTer Ultra Low kits for mRNA-seq. This increase in sensitivity allows the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing to produce libraries that outperform both the SMART-Seq2 method and other SMARTer Ultra Low kits for mRNA-seq. The streamlined workflow, high sensitivity, and excellent reproducibility make the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing the most advanced kit yet in our suite of best-in-class tools for next-gen sequencing.

Methods and References

Comparison across methods:

cDNA libraries were prepared in duplicate from 10 pg Mouse Brain Total RNA using the SMARTer Ultra Low Input RNA Kit for Sequencing - v3, the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing, or the SMART-Seq2 method as described in the user manuals or Picelli *et al*. (2013). Illumina adapters and indexes were added using Nextera XT DNA Library Preparation Kit and 100 pg of cDNA input. Libraries were sequenced using an Illumina MiSeq instrument, generating 4.0 million paired-end reads (2 x 75 bp).

Comparison of different input amounts:

10 pg–10 ng of Human Brain Total RNA was used as input for the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing as described in the user manual. 100 pg of cDNA input was used as input for the Nextera XT DNA Library Preparation Kit to generate Illumina RNA-seq libraries. Libraries were sequenced using an Illumina MiSeq instrument, generating 3.1 million paired-end reads (2 x 75 bp).

Sequence analysis:

Reads from both Mouse and Human Brain Total RNA were trimmed by CLC Genomics Workbench and mapped to rRNA and the mitochondrial genome with CLC (% reads indicated). The unmapped reads were subsequently mapped with CLC to the human (hg19) or mouse (mm10) genome with RefSeq annotation. The number of reads that map to introns, exons, or intergenic regions is a percentage of the reads successfully mapped to RefSeq. The number of genes identified in each library was determined by the number of genes with an FPKM of at least 0.1.

Scatter plots were generated with FPKM values from CLC mapping to the transcriptome. In order to identify the transcripts found in only one replicate, 0.001 was added to each value prior to graphing. Gene body coverage was determined using the geneBody_coverage.py module of RSeQC. Read coverage was normalized using Excel.

Reference:

Picelli, S., *et al.*, (2013) *Nat. Methods* **10**(11)1096–1098.

View web page >>
http://www.clontech.com/US/Products/cDNA_Synthesis_and_Library_Construction/NGS_Learning_Resources/Technical_Notes/SMART-Seq-v4