High-throughput single-cell transcriptomics with SMART-Seq technology

Survarna Gandlur*, Nathalie Bolduc, Sangwon Lee, Tommy Duong, Kazu Tori, Magnolia Bostick & Andrew Farmer

Takara Bio USA, Inc., Mountain View, CA 94043, USA *Corresponding Author: survarna_gandlur@takarabio.com

Abstract

Since the emergence of next-generation sequencing (NGS), the importance and demand for single-cell analysis have risen rapidly. Single-cell RNA-seq generates data for various analysis methods such as differential gene expression, alternative splicing, gene fusion identification, and so on, which allow for cell profiling in great detail. As a result, single-cell RNA-seq has been gaining prominence not only in basic research fields, but also in clinical fields. Extracting meaningful biological information from the small amount of mRNA present in a single cell requires a library preparation method with exceptional sensitivity and reproducibility. By providing the capability to obtain full-length mRNA sequence information (as opposed to merely capturing transcript 3' ends), the SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing (SMART-Seq v4) offers the most advanced single-cell RNA-seq method on the market. However, this method is relatively low throughput while researchers are interested in analyzing hundreds or thousands of individual cells for any given experiment. To address this need, our single-cell RNA-seq technology was further modified to create a simplified, high-throughput workflow with very little hands-on time. The reverse transcription (RT) and PCR amplification steps were combined into a single RT-PCR step, so that users can simply set up the RT-PCR and walk away. The updated workflow, available in the SMART-Seq HT Kit, is extremely fast, convenient, and generates a higher cDNA yield than its predecessor, all while providing the same unparalleled sensitivity and reproducibility.

1 Simplified workflow with the SMART-Seq HT Kit

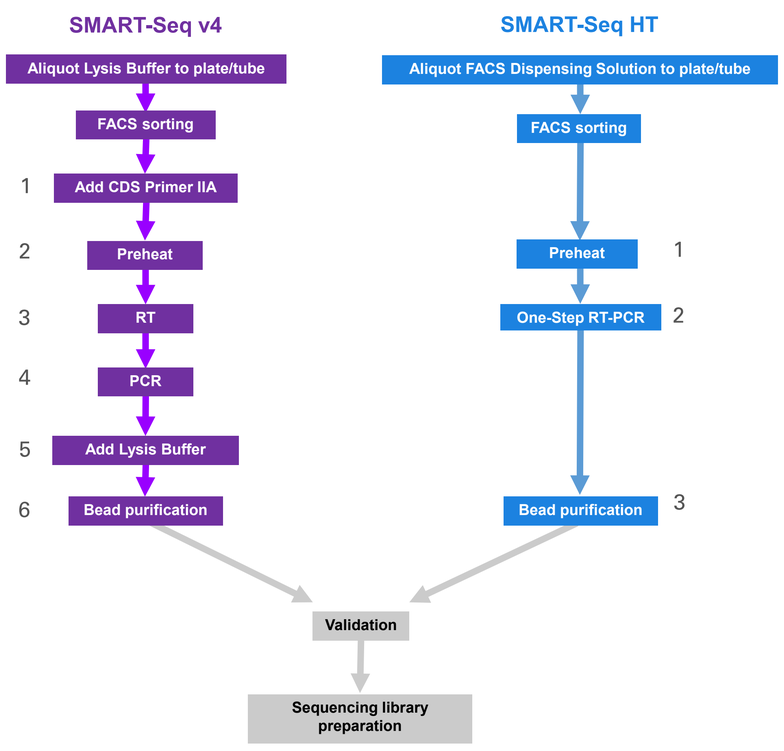

Figure 1. Comparison of the SMART-Seq v4 and SMART-Seq HT kit workflows. The SMART-Seq v4 method (left) was modified to generate a simplified, high-throughput workflow (SMART-Seq HT, right) with very little hands-on time. Once single cells have been obtained using FACS, the SMART-Seq HT Kit involves only three hands-on steps, while the original SMART-Seq v4 kit involves six hands-on steps. One key step in the SMART-Seq HT workflow is the One-Step RT-PCR, performed using the One-Step RT-PCR Buffer, formulated specifically for optimal reverse transcription followed by efficient PCR cDNA amplification. The One-Step RT-PCR Buffer is directly compatible with AMPure bead purification without the need for addition of Lysis Buffer. As with the original SMART-Seq v4 kit, the SMART-Seq HT Kit requires validation (quantification and assessment of high molecular weight, full-length cDNA) before cDNA is used for sequencing library preparation (Nextera® XT).

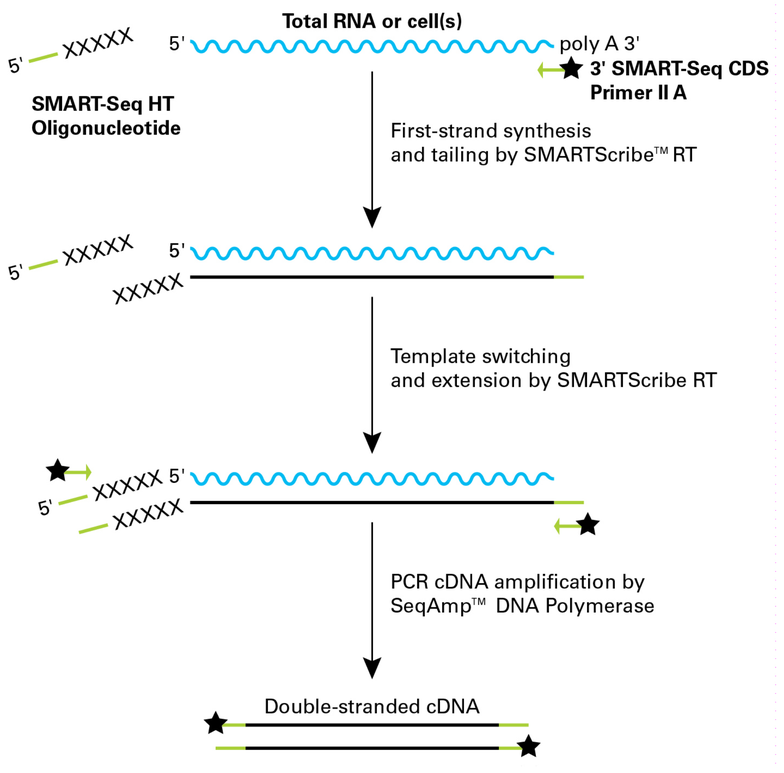

Figure 2. Schematic of technology in the SMART-Seq HT Kit. SMART® technology is used in a ligation-free workflow to generate full-length cDNA. The reverse transcriptase (RT) adds non-templated nucleotides (indicated by Xs) which hybridize to the SMART-Seq HT Oligonucleotide, providing a new template for the RT. Chemical modifications to block ligation during sequencing library preparation are present on some primers (indicated by black stars). The SMART adapters, added by the oligo(dT) primer (SMART-Seq CDS Primer IIA) and SMART-Seq HT Oligonucleotide and used for amplification during PCR, are indicated in green. The One-Step RT-PCR is set up as a single reaction so that all the reagents are mixed together but then used sequentially. SeqAmp is a hot-start DNA polymerase and is activated only after the reverse transcription/template-switching step is complete.

2 Comparable sensitivity and reproducibility with the SMART-Seq HT Kit

Sequencing metrics comparing SMART-Seq v4 and SMART-Seq HT kits

RNA source		10 pg Mouse Brain Total RNA					
cDNA synthesis		SMART-Seq v4			SMART-Seq HT		
Replicate		A	B	C	A	B	C
Yield (ng)		7.3	8.5	9.4	9.5	9.9	9.2
Number of transcripts	FPKM >0.1	14,168	13,934	14,147	14,510	14,650	14,553
	FPKM >1	11,455	11,267	11,349	11,582	11,670	11,454
Average Pearson/Spearman		0.97/0.67			0.97/0.68		
		0.96/0.66					
Proportion of reads mapped (%):							
rRNA		0.7	0.6	0.5	1.1	1.1	1.1
Mitochondria		2.8	4.2	4.1	3.7	3.8	4.2
Genome		92.3	90.8	86.6	89.1	89.0	87.8
Exons		74.8	72.5	69.0	67.3	67.1	65.3
Introns		13.3	14.1	13.3	16.9	16.9	17.4
Intergenic regions		4.2	4.2	4.4	4.9	5.0	5.2

Table 1. Comparable sensitivity and mappability between SMART-Seq v4 and SMART-Seq HT kits. Replicate cDNA libraries were generated from 10 pg Mouse Brain Total RNA using the SMART-Seq v4 or the SMART-Seq HT kits. RNA-seq libraries were generated from output cDNA using the Nextera XT DNA Library Preparation Kit and sequenced on an Illumina® NextSeq® instrument (2 x 75 bp). Sequences were analyzed as described in the methods after normalizing all the samples to 13 million paired-end reads. The two kits generated similar sequencing metrics, with a high mapping rate and comparable number of transcripts identified, in addition to strong Pearson and Spearman correlations. These data indicate that the SMART-Seq HT Kit provides the same sensitivity and reproducibility as the SMART-Seq v4 kit.

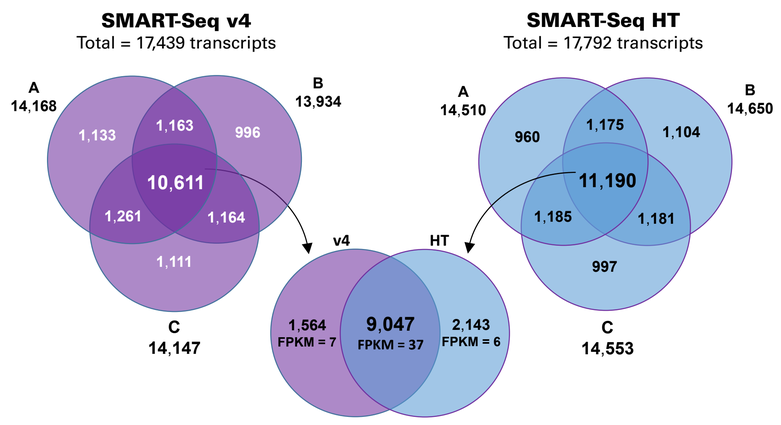

Figure 3. High overlap of transcripts identified for data generated with the SMART-Seq v4 and SMART-Seq HT kits. Libraries prepared from 10 pg of Mouse Brain Total RNA shown in Table 1 were further evaluated for the overlap in the number of transcripts identified (FPKM >0.1) between technical replicates within each kit, and found to be very similar (61–63% overlap). Transcripts identified by all three replicates for each kit were then compared against each other, indicating an overlap of 71%. The overlapping transcripts have an average expression level of 37 FPKM, while the transcripts uniquely identified with individual kits are less abundant, averaging between 6–7 FPKM, indicating that the transcripts more likely to not be identified are the ones expressed at a low level.

3 Even representation of low- and high-GC content genes with either approach

Gene GC content representation for SMART-Seq v4 and SMART-Seq HT kits

	Number of genes identified (% of total)			Total genes identified
	GC content ≤36%	GC content 37–54%	GC content ≥55%	GC content 0–100%
SMART-Seq HT Kit	272 (2.1)	11,788 (90.6)	954 (7.3)	13,014
SMART-Seq v4 kit	280 (2.2)	11,540 (90.4)	943 (7.4)	12,764
Reference genome	1,665 (4.7)	31,916 (89.9)	1,914 (5.4)	35,495

Table 2. Evaluation of gene GC content representation for the SMART-Seq v4 and SMART-Seq HT kits. The libraries made from 10 pg of Mouse Brain Total RNA shown in Table 1 were further analyzed for GC content representation. Genes were binned by GC content, and the number of genes identified is reported for each bin (numbers shown are the average of three technical replicates). The percentage of genes identified in each bin were identical for the two kits. For reference, there are 35,495 annotated RefSeq genes, of which 4.7% are arbitrarily classified as low CG content (≤36%), 89.9% are classified as medium CG content (37–54%), and 5.4% are classified as high GC content (≥55%).

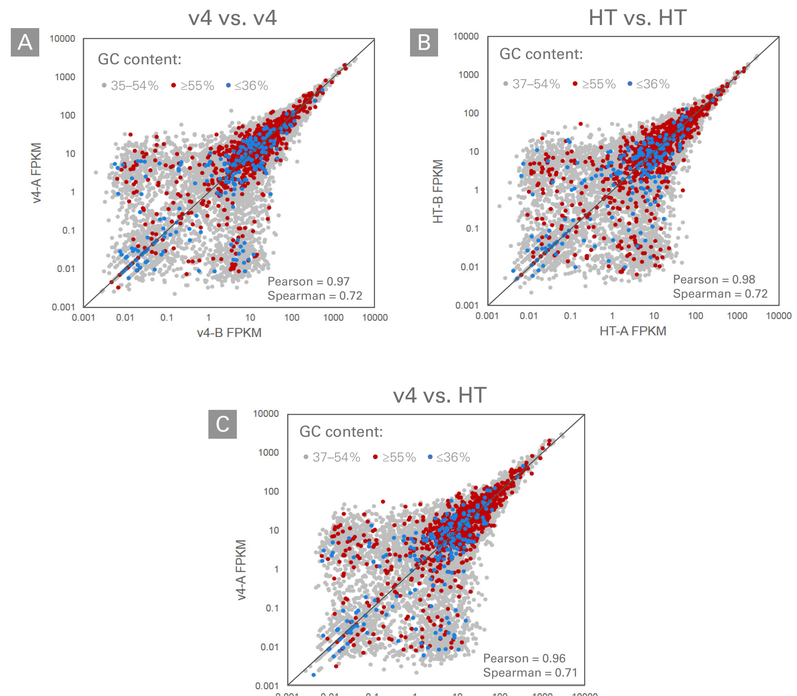

Figure 4. Comparison of expression level by gene GC content between the SMART-Seq v4 and SMART-Seq HT kits. The libraries made from 10 pg of Mouse Brain Total RNA shown in Table 1 were further analyzed for GC content representation (see Table 2). Genes were binned by GC content, and correlation plots were used to visualize the reproducibility of the expression levels (FPKM) of genes in each bin. The average gene counts are very reproducible for replicate samples analyzed using the SMART-Seq v4 (Panel A) or SMART-Seq HT kits (Panel B). Genes with high or low GC content (shown in red and blue, respectively) show similar expression levels in the SMART-Seq v4 and SMART-Seq HT kits (Panel C). Thus, the One-Step RT-PCR reaction introduced in the new SMART-seq HT Kit maintains the representation of the low- and high-GC content genes.

4 Comparison of data obtained by the v4 and HT kits using single cells isolated by FACS

Sequencing metrics comparing SMART-Seq v4 and SMART-Seq HT kits

RNA source		293T single cells			
cDNA synthesis		SMART-Seq v4 (n=12)		SMART-Seq HT (n=9)	
		Average	St. dev.	Average	St. dev.
Number of transcripts	FPKM >0.1	12,241	2,132	12,840	2,937
	FPKM >1	9,110	1,468	9,459	1,848
Proportion of reads mapped (%):					
rRNA		0.1	0.04	0.3	0.1
Mitochondria		2.7	0.6	4.2	1.3
Genome		95.1	0.6	93.1	1.4
Exons		87.8	1.3	83.1	4.0
Introns		5.4	0.9	7.6	2.7
Intergenic regions		1.9	0.2	2.5	0.5

Table 3. Comparison of the library data obtained from 293T cells using SMART-Seq v4 or SMART-Seq HT kits. Libraries were generated from individual 293T cells isolated by FACS using the SMART-Seq v4 or the SMART-Seq HT kits. RNA-seq libraries were generated using the Nextera XT DNA Library Preparation Kit and sequenced on an Illumina NextSeq instrument (2 x 75 bp). Sequences were analyzed as described in the Methods section after normalizing all the samples to 7 million paired-end reads. The two kits generated similar sequencing metrics, with a high mapping rate and around 600 additional transcripts identified in the SMART-Seq HT Kit. These data indicate that the SMART-Seq HT Kit provides the same or a slightly higher sensitivity than the SMART-Seq v4 kit.

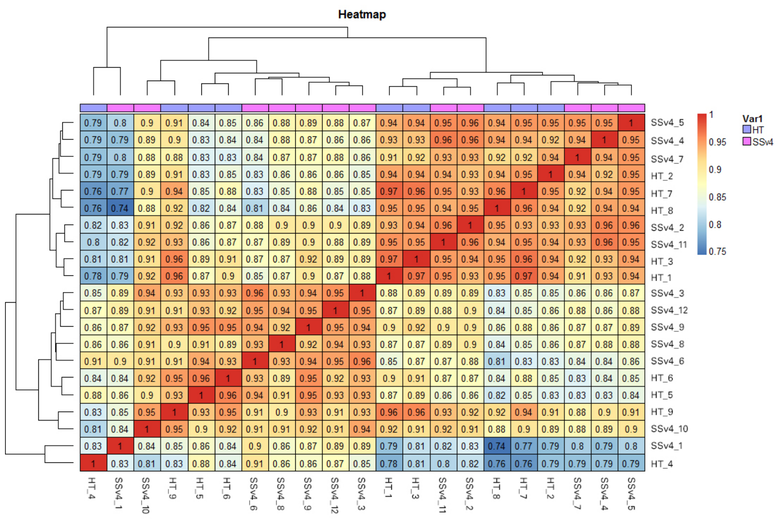

Figure 5. High reproducibility of gene expression data obtained from FACS-sorted 293T cells using the SMART-seq v4 and SMART-Seq HT kits. Libraries generated from twenty-one individual 293T cells shown in Table 3 were further analyzed to evaluate the reproducibility of gene expression measurements obtained for each cell with the SMART-Seq v4 kit (SSv4_1 to SSv4_12) and the SMART-Seq HT Kit (HT_1 to HT_9). The hierarchical clustering heat map shows Euclidean distances between all the cells and reports Pearson correlations ranging from 0.74 to 0.97. While the best correlations are observed between cells prepared with one or the other kit, the correlations are still very high between the two kits and the cells are not clustering based on the library preparation method. These data demonstrate that the modified workflow in the SMART-Seq HT Kit does not introduce major bias in measurement of gene expression levels.

Methods

For FACS sorting, 293T cells grown to near confluence were harvested by trypsinization, stained with FITC Mouse anti-Human CD47 (Clone B6H12; BD, Cat No. 556045), and resuspended in ice-cold BD FACS Pre-Sort Buffer (BD, Cat No. 563503). Sorting was done with a BD FACSJazz Cell Sorter in 12.5 µl of FACS Dispensing Solution. Cells were frozen at –80°C until ready for processing. The cDNA was synthesized and sequencing libraries prepared and sequenced as described in the figure legends. Reads from all libraries were trimmed and mapped to mammalian rRNA and the human or mouse mitochondrial genomes using CLC Genomics Workbench. The remaining reads were subsequently mapped using CLC to the human (hg19) or mouse (mm10) genome with RefSeq annotation. All percentages shown, including the number of reads that map to introns, exons, or intergenic regions, are percentages of the total reads in the library. The number of transcripts identified in each library was determined by the number of transcripts with an FPKM greater than or equal to 1 or 0.1.

Conclusions

- The new SMART-Seq HT Kit has a simplified workflow for high-throughput transcriptome profiling from single cells, featuring a convenient One-Step RT-PCR reaction.
- The SMART-Seq HT Kit generates full-length cDNA, yielding information about transcripts that would otherwise be missed, e.g., chimeric gene fusions, transcript isoforms, and splice variants.
- The SMART-Seq HT Kit provides the same, if not higher, sensitivity and reproducibility compared to the SMART-Seq v4 kit.
- The One-Step RT-PCR reaction does not generate any particular bias compared to the SMART-Seq v4 kit, and maintains the representation of low- and high-GC content genes.

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

Scan to download your copy of this poster visit www.takarabio.com/Transcriptomics_Poster

800.662.2566
Visit us at www.takarabio.com

Clontech TaKaRa cellartis