Pushing the limits of single-cell RNA-seq with SMART-Seq single cell technology

Nidhanjali Bansal, Tommy Duong, Magnolia Bostick, and Andrew Farmer*

Takara Bio USA, Inc., Mountain View, CA, USA *Corresponding author: andrew_farmer@takarabio.com

Introduction

Since the emergence of next-generation sequencing (NGS), the importance of and demand for single-cell analysis have risen rapidly. Extracting meaningful biological information from the small amount of mRNA present in each cell requires an RNA-seq preparation method with exceptional sensitivity and reproducibility. 3′ droplet-based sequencing has been the primary method used to date. However, using droplet sequencing and full-length mRNA information in parallel has become an emerging requirement to help generate and better understand rich single-cell datasets. To address this need, we developed the SMART-Seq® Single Cell Kit (SSsc) using new chemistry with unparalleled sensitivity and a highly scalable, easily automatable workflow.

The SMART-Seq Single Cell Kit outperforms all current commercial and noncommercial full-length methods, particularly with as little as 2 pg of total RNA. When validating with a B lymphocyte cell line or peripheral blood mononuclear cells (PBMC) from a healthy donor, we were able to detect 50–60% more genes with the new chemistry compared to current methods. The improvement in sensitivity was associated with a clear reduction of the dropout rate as well as an increase in reproducibility. This unparalleled sensitivity continues to be seen with automation and miniaturized workflows.

These features make SSsc chemistry extremely useful for difficult cells—e.g., clinical research samples that often have very low RNA content—making it ideal for highly detailed characterization of precious samples.

1 Improved chemistry, same easy workflow

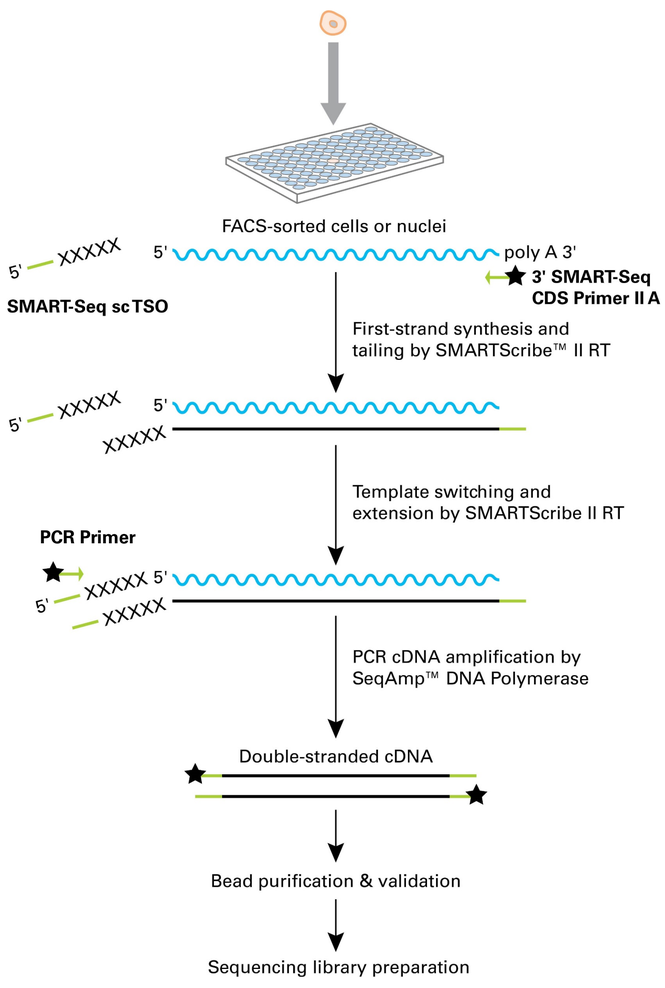

Figure 1. SMART-Seq Single Cell Kit technology and workflow. SMART® technology (Chenchik *et al.* 1998) is used in a ligation-free workflow to generate full-length cDNA. The reverse transcriptase (RT) adds nontemplated nucleotides (indicated by Xs) that hybridize to the SMART-Seq scTSO, providing a new template for the RT. Chemical modifications to block ligation during sequencing library preparation are present on some primers (indicated by the black stars). The SMART adapters, added by the oligo(dT) primer (3′ SMART-Seq CDS Primer II A) and SMART-Seq sc TSO, are indicated in green and used for amplification during PCR. The amplified cDNA is then purified, quantified, and used for sequencing library preparation (Illumina® Nextera® XT kit).

Methods

Figures 2 and 3: All cells were labeled with CD81-FITC antibody and 7-AAD (to distinguish live from dead cells) prior to sorting using a BD FACSJazz cell sorter into a 96-well plate or PCR strips. After sorting, cells were flash-frozen on dry ice and then stored at –80°C until ready to use. Unless otherwise noted, all libraries created with the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing and the SMART-Seq Single Cell Kit were processed at full volume per the user manuals. For the comparison to the Smart-seq2 protocol, cells were sorted and processed as described in Picelli *et al.* 2014.

Figures 4 and 5: For processing the SMART-Seq Single Cell Kit at quarter volume, the lysis buffer containing the 3′ SMART-Seq CDS Primer II A oligo was prepared as in the full-volume protocol, but only 3.5 µl were aliquoted in each well prior to cell sorting. A quarter volume of the reverse transcription master mix was added to start the reverse transcription, bringing the volume to slightly over 5 µl. The PCR was started by adding 0.25 µl of PCR Primer, 0.4 µl of SeqAmp DNA Polymerase, 10 µl of SeqAmp CB Buffer, and Nuclease-Free Water up to a 20 µl total volume. PCR cycling and subsequent steps were performed as in the full-volume protocol.

Sequencing libraries were generated using 125 pg of cDNA and the Nextera XT DNA Library Preparation Kit (Illumina) with a quarter of the recommended volume, as described in the SMART-Seq Single Cell Kit User Manual. Libraries were sequenced on a NextSeq® 500 instrument using 2 x 75 bp paired-end reads, and analysis was performed using CLC Genomics Workbench (mapping to the human [hg38] genome with Ensembl annotation). All percentages shown—including the number of reads that map to introns, exons, or intergenic regions—are percentages of mapped reads in each library.

References

Chenchik, A. et al. in *Gene cloning analysis by RT-PCR* (Siebert, P. & Larrick, J.) 305–319 (Eaton Pub Co, 1998).

Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9,** 171–181 (2014).

2 Superior sensitivity and reproducibility with the SMART-Seq Single Cell Kit

Sequencing metrics comparing SMART-Seq v4 kit and SMART-Seq Single Cell Kit						
RNA source	2 pg of UHR total RNA					
cDNA synthesis method	**SSv4**			**SSsc**		
Replicate	A	B	C	A	B	C
cDNA yield (ng)	7.8	6.9	5.5	14.8	14.9	9.6
Number of genes with TPM >1	7,412	7,522	7,487	8,774	8,614	8,406
Number of genes with TPM >0.1	8,660	8,868	9,240	10,319	10,276	10,285
Average Pearson/Spearman	0.95/0.59			0.97/0.63		
Proportion of reads mapped (%):						
Genome	92.7	92.5	92.5	80.1	80.9	80.6
Exon	79.3	78.7	76.6	63.4	64.1	62.0
Intron	10.5	10.9	12.5	13.0	12.8	14.0
Intergenic regions	2.9	3.0	3.4	3.7	4.0	4.6
rRNA	0.8	0.7	0.6	6.1	6.0	4.3
Mitochondria	3.5	3.6	3.9	9.3	8.4	10.2

Figure 2. Increased sensitivity with the SMART-Seq Single Cell Kit. Replicate cDNA libraries were generated from 2 pg of Universal Human Reference (UHR) total RNA using the SMART-Seq v4 kit (SSv4) or the SMART-Seq Single Cell Kit (SSsc); all libraries were processed with 19 PCR cycles. As described in the methods, RNA-seq libraries were generated and sequences analyzed (after normalizing all samples to 1.6 million paired-end reads). SSsc identified about 15% more genes than SSv4, despite showing a higher number of undesirable reads mapping to rRNA and mitochondria.

3 Greater sensitivity and reproducibility than the popular homebrew method

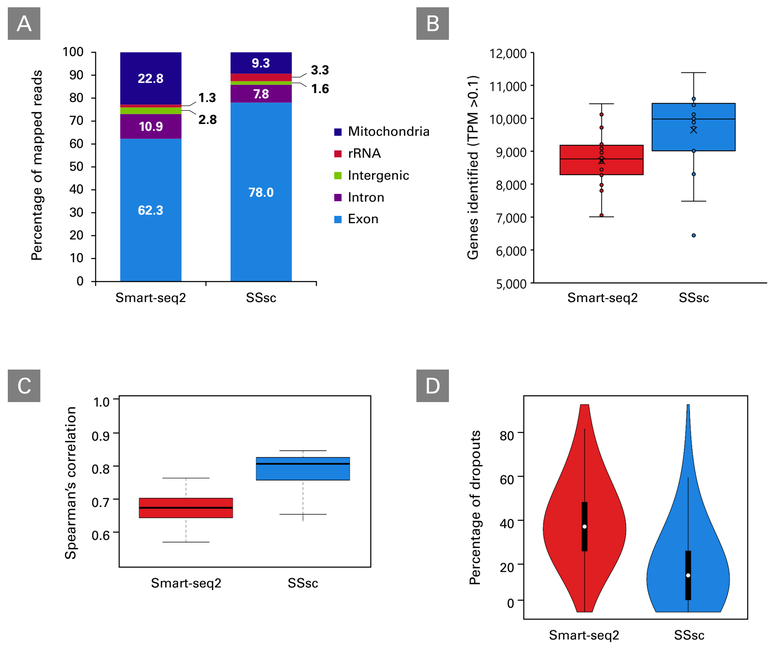

Figure 3. The SMART-Seq Single Cell Kit outperforms the Smart-seq2 protocol. Single cells from the lymphoblastoid cell line GM12878 were processed with SSsc (18 cells) or the Smart-seq2 protocol (20 cells; Picelli *et al.* 2014) using 19 cycles of PCR. As described in the methods, RNA-seq libraries were generated, and sequences were analyzed after normalizing all samples to 1.75 million paired-end reads. **Panel A.** The read distribution is different between the two chemistries, with a drastically greater number of reads mapping to the mitochondrial genome with Smart-seq2 chemistry. This indicates that fewer reads are available for gene identification with the Smart-seq2 method. **Panel B.** More genes are detected in the cells processed with SSsc. **Panel C.** Correlation boxplots showing the intragroup Spearman correlation between all cells processed with either method. The higher Spearman's correlation among the cells processed with SSsc indicates a higher reproducibility than the Smart-seq2 method. **Panel D.** The higher reproducibility of SSsc is also demonstrated by the lower dropout rate of the genes detected with a TPM >1.

4 Uncompromised performance for miniaturization

Sequencing metrics comparing full- and quarter-volume workflows of SMART-Seq Single Cell Kit		
RNA source	Single B cells (GM12878)	
Workflow	**Full volume (20 µl)**	**Quarter volume (5 µl)**
Replicates	n = 18	n = 20
cDNA yield average (ng)	79	61
Number of genes with TPM >1	8,660	8,325
Number of genes with TPM >0.1	9,642	9,270
Proportion of reads mapped (%):		
Genome	87.4	86.5
Exon	78.0	76.9
Intron	7.8	8.2
Intergenic regions	1.6	1.4
rRNA	3.3	3.9
Mitochondria	9.3	9.6

Figure 4. Similar yield and sensitivity between full-volume and quarter-volume workflows. Single cells from lymphoblastoid cell line GM12878 were sorted in a full volume (12.5 µl) or quarter volume (3.5 µl) of lysis buffer containing the 3′ SMART-Seq CDS Primer II A. cDNA libraries were generated using the regular SMART-Seq Single Cell Kit protocol (full-volume workflow; reverse transcription in 20 µl) or a quarter-volume workflow (reverse transcription in 5 µl). Pipetting for both workflows was performed using a MANTIS Liquid Handler (Formulatrix), and the PCR step was conducted with 19 cycles. The cDNA yield was similar between the two workflows, generating about 60–80 ng of cDNA per cell. As described in the methods, RNA-seq libraries were generated and sequences analyzed (after normalizing all samples to 1.75 million paired-end reads). The full-volume and quarter-volume workflows generated very similar mapping statistics, including the number of genes identified.

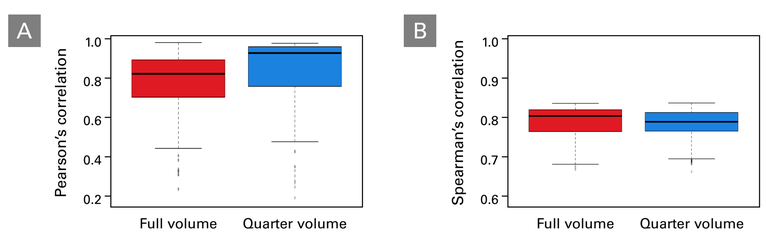

Figure 5. Similar reproducibility among the cells processed with the full-volume and quarter-volume workflows. Correlation boxplots showing intragroup Pearson (**Panel A**) or Spearman (**Panel B**) correlations between all of the cells processed at full volume or quarter volume. Processing the cells with the quarter-volume workflow neither introduced extra variability nor a significant decrease in sensitivity (see Figure 4), indicating that the SMART-Seq Single Cell Kit is perfectly suitable for miniaturization.

5 Customer data: unprecedented sensitivity with SSsc using miniaturization

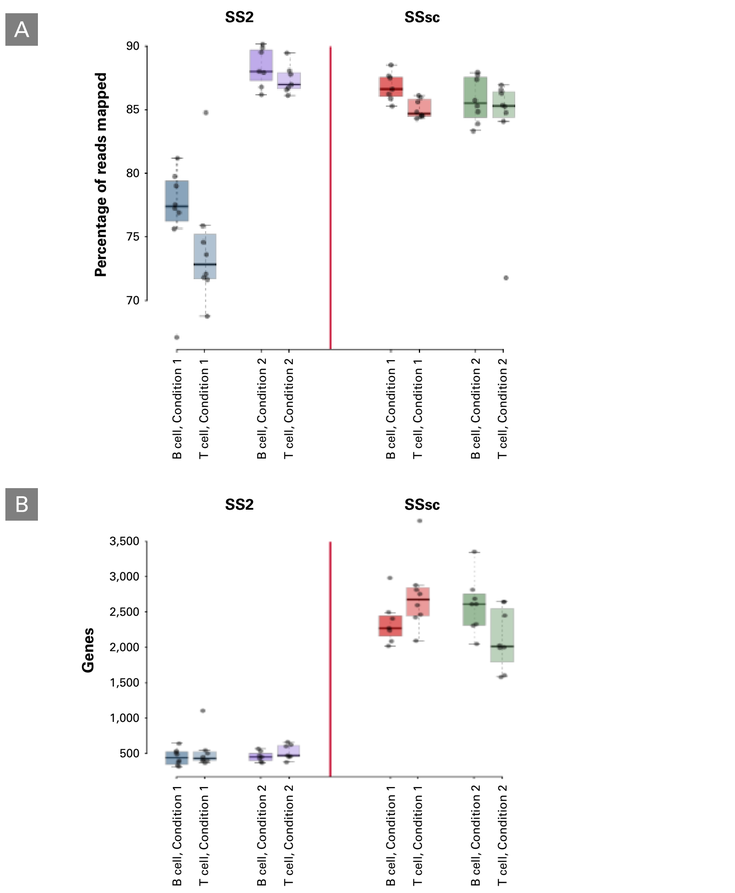

Figure 6. The SMART-Seq Single Cell Kit shatters sensitivity seen with the Smart-seq2 protocol. Libraries were prepared from both B and T cells with either 23 (Condition 1) or 25 (Condition 2) PCR cycles using quarter-volume on the Mantis. **Panel A**. SSsc has a similar percentage of reads mapped regardless of the PCR cycling, unlike Smart-seq2 (SS2). This indicates that small variations in the workflow have a significant impact on the reproducibility of SS2 libraries, a disadvantage of using a method that is not quality-controlled or manufactured under stringent quality standards. **Panel B**. In both the B and T cells, SSsc identified at least **five** times as many genes as SS2, exhibiting unprecedented sensitivity compared to SS2. *Data kindly provided by Dr. Holger Heyn, Team Leader at the National Centre for Genomic Analysis in Barcelona, Spain.*

6 Greater and more valuable sensitivity

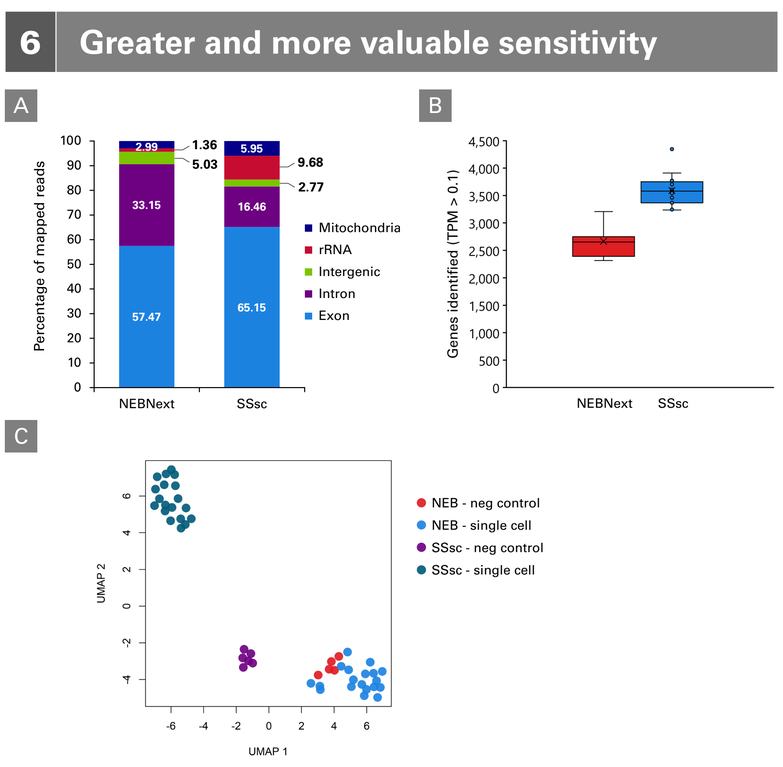

Figure 7. The SMART-Seq Single Cell Kit outperforms the NEBNext protocol. Libraries were prepared from T cells with either NEBNext Single Cell/Low Input RNA Library Prep Kit for Illumina (NEB) or SSsc according to manufacturer's instructions. **Panel A**. The read distribution is different between the two chemistries, with more reads mapping to exon regions for SSsc. **Panel B**. The number of genes identified with a TPM >0.1 is higher (~40%) for SSsc than for NEBNext. **Panel C.** The high number of cycles required to generate a library for the NEBNext protocol means that the negative controls for NEB (red) cluster strongly with the single-cell libraries (light blue), while the negative controls for SSsc (purple) are distinct from the single-cell libraries (gray-green).

Conclusions

- Full-length chemistries can complement 3′ DE data to provide a whole picture for single-cell data
- The SMART-Seq Single Cell Kit (SSsc) features an easy, plate-based workflow that starts directly from single cells isolated by FACS or other methods
- SSsc offers unparalleled sensitivity and reproducibility for single-cell, full-length RNA sequencing, particularly for cells with very low RNA content and for nuclei
- SSsc generated high cDNA yield and detected a greater number of genes compared to the Smart-seq2 method
- SSsc is compatible with miniaturization at quarter-reaction volume without compromising on performance (yield, sensitivity, or reproducibility)
- SSsc identifies a greater number of genes than NEBNext that clearly separate from the background

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

800.662.2566
Visit us at takarabio.com

Clontech TaKaRa cellartis