Improved chemistry for single-cell transcriptome analysis

SMART-Seq v4 Ultra Low Input RNA Kit for the Fluidigm C1 System

- Improved mRNA-seq using SMART-Seq v4 chemistry on the Fluidigm C1
  A variety of improvements to the original chemistry and protocol
- Higher sensitivity, with more genes detected
  Significant increases in cDNA yield and number of genes detected
- Better representation of GC-rich genes
  More accurate expression data for genes with high GC content
- Reduced technical variability
  Stronger reproducibility between replicates

Introduction

In single-cell transcriptome studies, it is critical to obtain high-quality cDNA libraries from individual cells that represent the original *in vivo* mRNAs as closely as possible. To achieve this, we improved upon our existing commercially available mRNA-seq chemistry for single-cell transcriptome studies (the SMARTer Ultra Low Input RNA Kit for the Fluidigm C1 System; UL-v1). The SMART-Seq v4 Ultra Low Input RNA Kit for the Fluidigm C1 System (SS-v4) allows high-quality cDNA synthesis from up to 96 single cells that have been isolated and processed with the Fluidigm C1 Single-Cell Auto Prep System. We have developed the "SMART-Seq v4" C1 script for single-cell RNA-seq employing the Open App IFC and new chemistry. While the original UL-v1 kit runs only with the IFC for mRNA Seq, the SS-v4 kit is compatible with *both* the IFC for mRNA Seq and the IFC for Open App. No matter which IFC is used, the SS-v4 kit uses the SMART-Seq v4 script. The IFC for Open App was used for the SS-v4 experiments reported here.

The improved chemistry built into the SMART-Seq v4 Ultra Low Input RNA Kit for the Fluidigm C1 System:

- Incorporates locked nucleic acid (LNA) technology integrated with SMART technology. These improvements lead to higher sensitivity, higher cDNA yield, and more genes detected.
- Uses SeqAmp polymerase, which results in better representation of high-GC genes.
- Enables better reproducibility, which indicates reduced technical variation, increasing the likelihood of discovering true biological variation.

In addition to differences in chemistry, the newer SMART-Seq v4 script (SS-v4 chemistry) and the original mRNA-Seq script (UL-v1 chemistry) have different protocols. The lysis and RT thermal cycling conditions are the same in the SS-v4 and UL-v1 protocols; however, a variety of reagents and enzymes (for lysis, RT, and PCR), and PCR conditions vary between the two protocols. (For example, the SMART-Seq v4 script uses SeqAmp polymerase and conducts PCR at a higher temperature.) Key differences are highlighted in the table below.

	SS-v4 (SMART-Seq v4 script)		UL-v1 (mRNA Seq script)	
Lysis conditions	3 min 72°C 10 min 4°C 1 min 25°C			
Lysis reagents	10X lysis buffer		Dilution buffer	
RT conditions	90 min 42°C 10 min 70°C			
RT reagents	**SMART-seq v4 Oligonucleotide**		**SMARTer II A Oligonucleotide**	
PCR conditions	1 cycle	1 min 98°C	1 cycle	1 min 95°C
	5 cycles	20 sec 98°C 4 min 59°C 6 min 68°C	5 cycles	20 sec 95°C 4 min 58°C 6 min 68°C
	9 cycles	20 sec 95°C 30 sec 65°C 6 min 68°C	9 cycles	20 sec 95°C 30 sec 64°C 6 min 68°C
	7 cycles	30 sec 95°C 30 sec 65°C 7 min 68°C	7 cycles	30 sec 95°C 30 sec 64°C 7 min 68°C
	1 cycle	10 min 72°C	1 cycle	10 min 72°C
PCR reagents	**SeqAmp DNA Polymerase**		**Advantage 2 Polymerase**	

Comparison of sample preparation conditions. To employ the SS-v4 chemistry on C1 IFCs, a SMART-Seq v4 script was written using the C1 Script Builder (v.2.0.10). The existing Fluidigm mRNA Seq script that is built into the C1 system was used for the UL-v1 experiments (Protocol PN 100-7168, Fluidigm Corporation).

Results

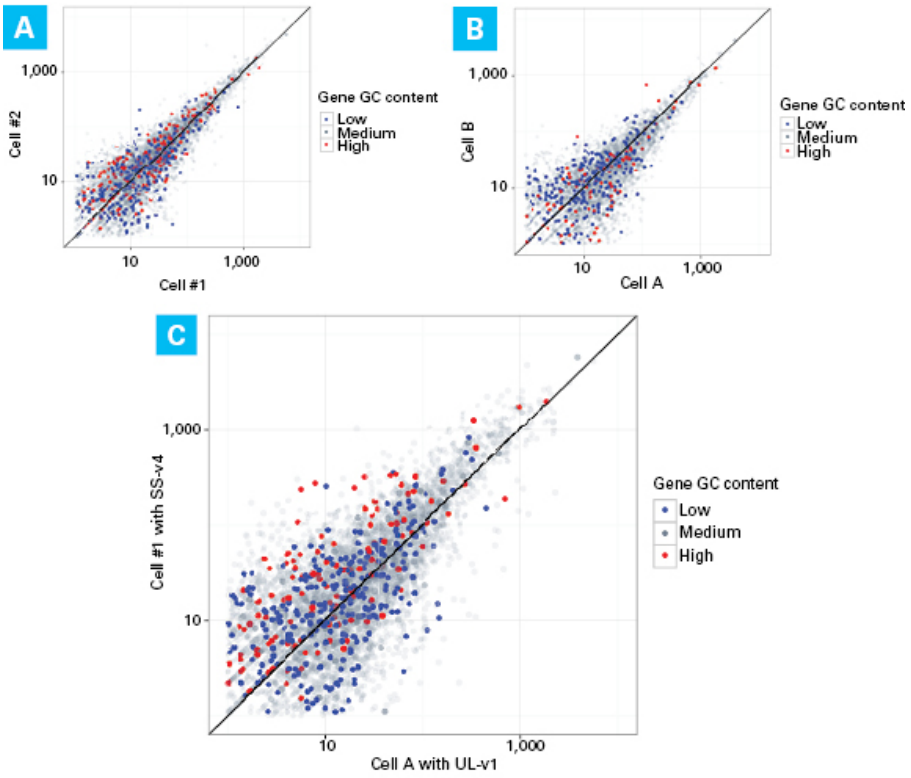
In a comparison, most single cells (96–100%) successfully produced good amounts (>6.5 ng) of cDNA by either method. However, the SS-v4 method significantly increased cDNA yield, with results 38–68% higher than the UL-v1 method. In addition, more genes were detected using the SS-v4 chemistry, and sequencing metrics showed better performance.

Sequencing alignment metrics			
Chemistry used	**SS-v4**		**UL-v1**
Replicate	T1	T2	
Number of single cells captured	65	75	81
Number of single cells that generated >6.5 ng cDNA	65	74	80
Average cDNA yield (ng per cell)	38 (±5.4)	31 (±5.6)	23 (±9.4)
Number of single cells sequenced	11		
Number of sequence reads per library	2M*		
Proportion of reads (%)			
Mapped to human genome	98 (±0.3)	98 (±0.2)	97 (±0.5)
Mapped to rRNA	1.6 (±0.4)	0.9 (±0.2)	0.5 (±0.1)
Mapped to mitochondria	11 (±1.5)	7 (±1.6)	6 (±0.9)
Mapped to NCBI-RefSeq	86 (±1.6)	90 (±1.7)	91 (±1.0)
Exonic	89 (±2.5)	90 (±2.2)	71 (±8.2)
Intronic	8 (±2.1)	7 (±1.8)	22 (±7.1)
Intergenic	3 (±0.5)	3 (±0.5)	7 (±2.5)
Number of genes identified with >0.1 RPKM	8,419 (±297)	7,669 (±538)	6,644 (±606)

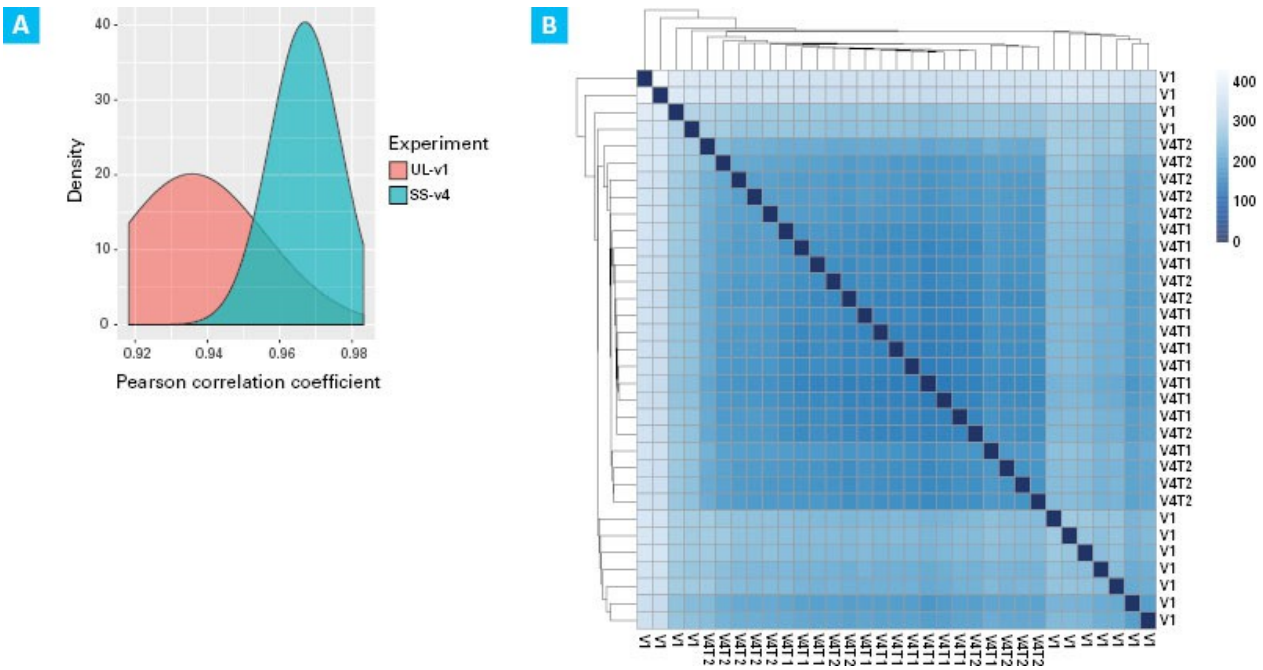
* Sequence reads were resampled to 2 million reads per library. Two out of eleven UL-v1 libraries had lower numbers of reads and were analyzed as is (1.12 million and 0.54 million reads, respectively).

More genes detected using the new SS-v4 chemistry. Eleven cDNA libraries per experiment were generated from K562 cells (human leukemia cell line) and sequenced using the Illumina Nextera® XT DNA Library Preparation Kit and a MiSeq® sequencer. The FASTQ files were analyzed using CLC Genomics Workbench (v.8.5.1, Qiagen Aarhus). The paired sequence reads were trimmed using quality score (<0.05) and primer and adapter sequences. In mapping, the paired-end sequences that mapped to rRNA were removed first. Next, the sequences that mapped to mitochondria were excluded from the data set. The remaining sequences were tested for mapping against the reference genome (GRCH37). The SS-v4 sequencing data showed a higher percentage of reads mapped to exons and a much lower percentage of reads mapped to introns compared to UL-v1. SS-v4 also detected a higher number of genes compared to UL-v1 (a 15–27% increase).

The components, chemistry, and protocol updates in the SS-v4 kit have resulted in higher overall performance. More genes with high GC content were also detected in the cDNA libraries constructed using the new SS-v4 chemistry relative to the cDNA libraries constructed with UL-v1 (data not shown). Expression levels were very reproducible for pairs of cells using either chemistry (compare Panels A and B). However, when the two chemistries were compared, genes with high GC content (shown in red) showed higher expression with SS-v4 than with UL-v1 (Panel C).



Comparing expression levels by gene GC content for two cDNA synthesis protocols. Libraries made from single K562 cells using either the SS-v4 protocol or the UL-v1 protocol were compared. Genes were binned by GC content (Low: 0–36% GC; Medium: 37–62% GC; High: 63–100% GC) and correlation plots were used to evaluate the two protocols. The RPKM values were very reproducible for pairs of cells using the SS-v4 chemistry (**Panel A**) or the UL-v1 chemistry (**Panel B**), with even distribution of genes across all three GC-content categories. **Panel C.** When the two protocols were compared, genes with high GC content (shown in red) showed higher expression with the SS-v4 protocol, while genes with medium or low GC content (shown in gray and blue, respectively) showed an even distribution. A simple way to understand the reproducibility of the new method is to look at the Pearson correlation of sequencing data from replicates of libraries prepared with the UL-v1 and SS-v4 protocols. Transcriptomic data generated by SS-v4 were more consistent between replicates than the data generated by UL-v1. The SS-v4 protocol reduced technical variability, which may increase the opportunity to detect biologically relevant differences between samples.



Reduced variability among cDNA libraries synthesized using the SS-v4 protocol. Sequencing data were generated with the SS-v4 method (11 libraries x 2

replicate experiments) and the UL-v1 method (11 libraries), as described above. The FASTQ files were aligned against the human reference genome (Gencode release GRCh38) using STAR v. 2.5 (Dobin et al. 2013) with default options. Gene counts generated by STAR were subjected to regularized log transformation (DESeq2; v.1.10.1; Love et al. 2014). These transformed gene expression values were then used in the gene expression analyses. The analysis results were visualized using R. **Panel A**. The Pearson correlation coefficients were higher and tighter in the cDNA libraries generated with SS-v4. **Pane B**. The hierarchical clustering heat map demonstrates that the Euclidean distances between the SS-v4 cDNA libraries were smaller than the distances between the UL-v1 cDNA libraries. This indicates that the cDNA libraries generated with SS-v4 were more similar to each other than the cDNA libraries generated with UL-v1.

Conclusion

Good science always aims for the highest performance in biological research while maintaining the simplicity of the experimental strategy. With that in mind, Takara Bio scientists have developed this newly improved best-in-class solution for analyzing single-cell transcriptomics using a microfluidic-based approach. The SMART-Seq v4 Ultra Low Input RNA Kit for the Fluidigm C1 System (SS-v4) provides an improved single-cell, full-length mRNA-seq method. The new SMART-Seq v4 script with SS-v4 chemistry demonstrates great sensitivity and delivers high quality, robust, and reproducible transcriptomic data.

Methods

In the experiments described here, K562 cells (human leukemia cell line) were captured, lysed, and processed for cDNA library construction on the Fluidigm C1 platform. The C1 Single-Cell Auto Prep IFC for Open App (for medium-sized cells) was used as described here for the SMART-Seq v4 script (SS-v4 chemistry), while the C1 Single-Cell Auto Prep IFC (for medium-sized cells) was used according to the Fluidigm Protocol for experiments performed with the existing mRNA Seq script and UL-v1 chemistry. For all experiments, the cells were cultured to a log phase, mixed with suspension reagent at a 3:2 ratio for the best buoyancy, and 500–700 cells were loaded into an IFC. Every capture site was examined using microscopy to determine if a single cell was captured. Improved SMART (**S**witching **M**echanism **a**t 5' end of **R**NA **T**emplate) technology enables us to synthesize cDNA from total RNA using a poly dT primer and the SMART-Seq v4 Oligonucleotide, and eliminates ligation steps by using sequencing adapters in downstream steps (Ramsköld et al. 2012).

References

Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30,** 777–82 (2012).

Related Products

Cat. #	Product	Size	License	Quantity	Details
635026	SMART-Seq® v4 Ultra® Low Input RNA Kit for the Fluidigm® C1™ System, 10 IFCs	Each		*	

The SMART-Seq v4 Ultra Low Input RNA Kit for the Fluidigm C1 System, IFCs allows high-quality cDNA synthesis starting from 96 single cells that have been isolated and processed with the Fluidigm C1 Single-Cell Auto Prep System. Employing locked nucleic acid (LNA) technology integrated with SMART technology, this kit offers remarkable sensitivity, enriches for full-length transcripts, and maintains true representation of the original mRNA-critical factors for transcriptome analysis. Genes with high GC content show higher expression with this kit as compared to the previous generation. The kit includes SeqAmp DNA Polymerase for PCR amplification, and the libraries generated by this kit have been tested for compatibility with Illumina sequencing platforms.

Documents	Components	Image Data

635025	SMART-Seq® v4 Ultra® Low Input RNA Kit for the Fluidigm® C1™ System, 4 IFCs	Each		*	

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