

Differential expression from single cells using the SMART-Seq v4 3' DE Kit

- [Overview of the SMART-Seq v4 end-capture method for generating sequencing libraries:](#)
SMART-based protocol and cell-barcoding enable up to 1,152 separate cell reads per sequencing lane.
- [Highly sensitive end-capture method:](#)
Twelve single K562 cells were processed, pooled, and sequenced in one pooled sample; we detected a large number of genes per cell with high correlations between cells.
- [Reads map to the 3' ends of transcripts:](#)
The end-capture protocol leads to a majority of reads mapping to the last 30% of transcript.

Differential expression (DE) analysis utilized for single-cell comparisons has become one of the key methods for studying transcriptome variability—especially when homogeneous cell populations are elusive, such as in cancer research, developmental biology, neurobiology, and immunology (Heaton et al. 2014; Henley et al. 2013; Saliba et al. 2014). SMART (**S**witching **M**echanism **a**t the 5' end of the **R**NA **T**emplate) technology has emerged as the most sensitive solution for processing the small amounts of mRNA present in single cells. Here we discuss the use of the SMART-Seq v4 3' DE Kit to enable differential expression (DE) analysis in a more efficient and cost-effective manner. SMART-Seq v4 kits incorporate LNA technology in order to produce high-quality, reproducible sequencing data with superior identification of genes, including those with low expression. By combining these features with cellular indexes, pooling, and 3' end-capture sequencing, we can obtain high-quality gene expression data without having to sequence the entire transcriptome.

Introduction ^

Differential expression (DE) analysis focuses on comparing the relative expression levels of different transcripts in the cell and is one of the primary analysis tools used to explore transcriptome variability. End-capture methods are appealing for DE analysis as they can decrease the number of reads necessary to determine differential expression between cells. By focusing the sequencing data on a portion of each transcript (in this case, the 3' end of each transcript), we can reduce the number of reads and hence the overall cost required to identify expressed genes. Additionally, samples can be pooled prior to sequencing, decreasing the work and resources required and increasing the multiplexing capabilities of each sequencing lane.

In this tech note, we demonstrate the use of SMART-Seq v4 chemistry combined with an end-capture method for low-input amounts down to the single-cell level. As we have previously reported, SMART-Seq v4 technology produces high-quality cDNA libraries from individual cells that closely represent the original *in vivo* mRNAs. It is extremely sensitive, works with transcripts of different lengths, and has excellent gene body coverage across a wide range of GC content (see References for details). In order to reduce labor and cost, we have adapted this robust technology for 3' end capture, and here we present this approach and validate it by sequencing single K562 cells.

This approach allows each pool of cDNA from 12 single-cell reactions to be tagged by one of the 96 Illumina HT barcode combinations, enabling up to 1,152 separate cell reads per run. Researchers using this kit can confidently determine differentially expressed transcripts while decreasing the cost and time required for discovery. This method utilizes a modified oligo(dT) primer including an in-line index which serves as a cell barcode and a portion of the Illumina read primer 2 sequence in order to accommodate a pooled library generation protocol (**Figure 1**). The in-line index is placed between the transcript and the Illumina read primer 2 (RP2), and enables pooled cell samples to be demultiplexed after sequencing.

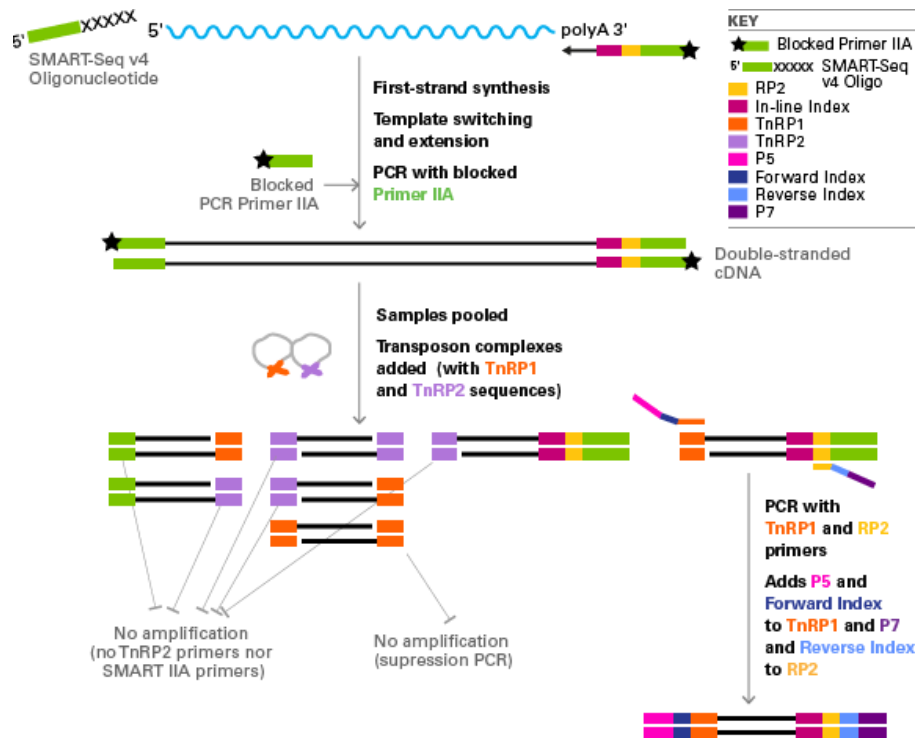


Figure 1. Overview of the end-capture method for analyzing differential expression. cDNA (black lines) is synthesized with a blocked (black star) and modified oligo(dT) primer that adds sequences for subsequent amplification and analysis—a cell barcode (magenta), part of the Illumina read primer 2 sequence (RP2, yellow), and the SMART IIA sequence (green). The SMART IIA sequence is used as a priming site during cDNA amplification, the Illumina RP2 sequence is used as a priming site during library amplification, and the cell barcode is used for demultiplexing pooled samples during analysis. The process works as follows: first, the template for SMARTScribe reverse transcriptase switches from the mRNA (blue wavy line) to the SMART-Seq v4 Oligonucleotide (green). After reverse transcription, the full-length cDNA is amplified by PCR with blocked Primer IIA oligonucleotides. After cDNA amplification, the presence of the barcode (magenta) allows for pooling up to 12 samples. The pooled samples are tagmented and Illumina Nextera® read primer 1 and 2 sequences are added by the Nextera Tn5 transposon (TnRP1 and TnRP2, orange and purple respectively). The 3' ends of the original mRNA are captured by selective PCR with primers for the TnRP1 and RP2 sequences. Other products of the transposon-based reaction are not amplified, either because there are no primer sites for amplification or because of suppression PCR. Cluster generation (pink and dark purple) and indexing sequences (light blue and dark blue) are added during this PCR stage to generate a library ready for sequencing on an Illumina platform.

Results

High sensitivity of the end capture method

We used this method to sequence 12 single K562 (human immortalized myelogenous leukemia) cells in one pooled sample. The total number of reads in the pool was 21.6 million, with reads from individual cells ranging from 0.2 to 5.0 million, as seen in **Table 1**. This variability is due to the different quantities of cDNA generated from each cell. These reads were then used to identify expressed genes in each cell and determine the correlation between expression profiles for each cell.

Sample ID	Pool	In-line Indexes											
		i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12
mtRNA (%)	3.3	3.1	3.6	1.7	3.9	4.1	1.8	2.5	2.8	5.5	4.1	2.7	4.8
rRNA (%)	0.6	1.2	0.7	1.1	0.4	0.4	0.3	0.3	0.7	1.2	0.5	0.1	0.6
Uniquely mapped reads (%)	69	68	68	70	68	71	71	72	68	69	69	74	69
Total mapped reads (%)	97	96	98	97	97	98	98	98	96	97	96	98	98
Number of reads (M)	21.6	2.3	5.0	1.3	2.2	2.0	1.9	0.8	1.3	0.6	0.2	1.5	1.6

Table 1. Mapping statistics for pooled libraries from K562 single cells. K562 (human immortalized myelogenous leukemia line) cells were diluted to one cell/ μ l in PBS buffer and twelve single cells were isolated, checked via optical microscopy, lysed, and subjected to cDNA synthesis. The pooled libraries were sequenced on an Illumina MiSeq® instrument with 47 bp for read 1 and 26 bp for read 2. The pooled libraries were demultiplexed based on the in-line barcode sequence from read 2. All libraries were mapped with STAR v.2.3.0.1 (Dobin et al., 2013) against the human genome (hg19). The reads map to the genome at a high rate (>96%) with a small proportion mapping to rRNA or mitochondrial regions.

Gene expression analysis and cell-to-cell correlation

After obtaining the read data from the pooled cell samples, the data was analyzed to determine the number of identified genes using different cutoff values for the Counts Per Million mapped reads (CPM). As seen in **Figure 2**, a high number of genes were identified for all cells, with some differences due to the total number of reads varying between cells. We then determined the Pearson correlations of regularized log-transformed read-counts, shown below in **Figure 3**. As expected, we observed high correlations (>0.7) between all cells, with most cells demonstrating a correlation >0.9.

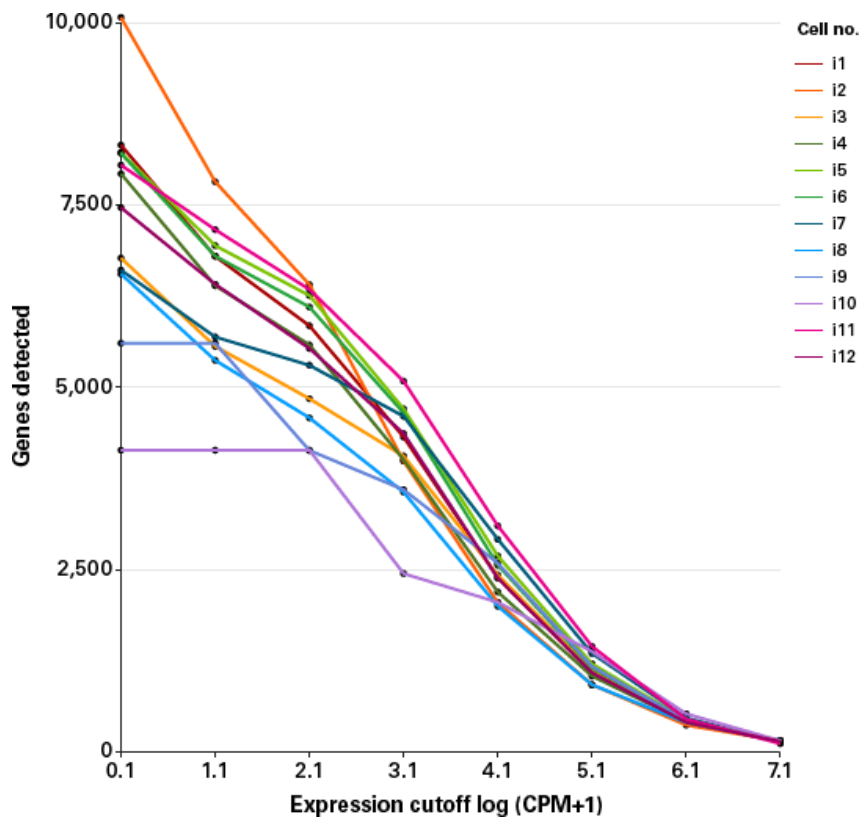


Figure 2. Number of genes identified from K562 single cells. Mapped libraries were analyzed with CPM) generated from STAR v.2.3.0.1 (Dobin et al. 2013). The number of genes identified with different cutoffs (0.1, 1.1, 2.1, 3.1, 4.1, 5.1, 6.1, and 7.1) for log-transformed CPM+1 are plotted. The amount of cDNA

produced from each cell varies, leading to different read depths per cell. This affects the number of genes identified, as seen most obviously for sample 10 at lower expression cutoff values.

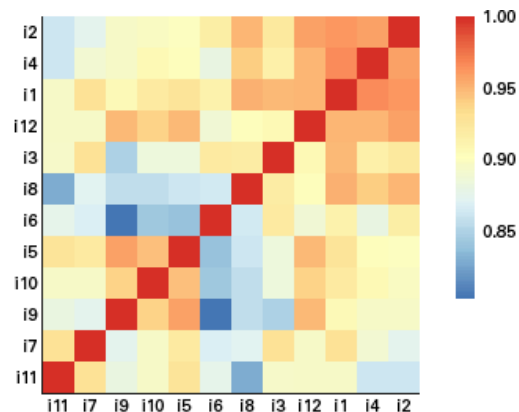


Figure 3. Pearson correlation heat map matrix of K562 single cells. The heat map represents the Pearson correlations of expression levels for the 12 single-cell libraries. For all comparisons, the correlation (R) was >0.8 , while the majority of single-cell libraries are highly correlated (>0.9).

Verification of the end-capture method

Finally, we verified that the end-capture method was functioning correctly by determining where the majority of reads were occurring along each transcript. As seen in **Figure 4**, the majority of reads mapped to the last 30% of the transcripts, as expected with our end-capture method.

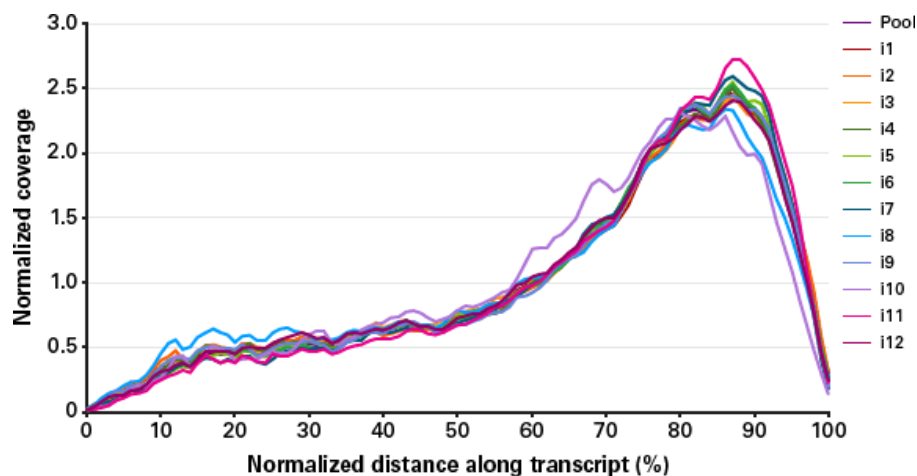


Figure 4. Gene body coverage analysis. Once the reads were mapped to the human genome, gene body coverage analysis was performed to assess the ability of the methods to capture the 3' ends of the cDNA. The majority of reads across all transcripts mapped to the last 30% of the transcripts (normalized in length to 100%).

Conclusion

The [SMART-Seq v4 3' DE Kit](#) provides a more cost-effective method for identifying differentially expressed genes in single cells. This will enable DE experiments with higher sensitivity for scientists interested in quickly identifying differences between cells without sequencing the entire transcriptome. Our method 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 isolated, evaluated via optical microscopy, lysed, and processed for cDNA library construction. This was accomplished by diluting the cells in PBS to 1 cell/ μ l, then spotting 1 μ l drops in each well of a 96-well plate and visually choosing single cells to pick and process. cDNA libraries were prepared from single cells according to the protocol provided with the SMART-Seq v4 3' DE Kit.

References

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

Heaton, N. S. *et al.* Long-term survival of influenza virus infected club cells drives immunopathology. *J. Exp. Med.* **211**, 1707–14 (2014).






Henley, B. M. *et al.* Transcriptional regulation by nicotine in dopaminergic neurons. *Biochem. Pharmacol.* **86**, 1074–83 (2013).

Saliba, A.-E., Westermann, A. J., Gorski, S. A. & Vogel, J. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* **42**, 8845–60 (2014).

Additional reading about SMART technology:

SMART-Seq v4 sensitivity is discussed in our [SMART-Seq v4 single-cell tech note](#) and additional data on sensitivity across a wide range of GC content is discussed in our [Fluidigm C1 single cell tech note](#).

Related Products

Cat. #	Product	Size	License	Quantity	Details
635040	SMART-Seq® v4 3' DE Kit	96 Rxns		*	
<p>The SMART-Seq v4 3' DE Kit includes the components to generate high-quality cDNA from 1-100 intact cells or ultralow amounts of total RNA. The cDNA generated from each reaction has a cell barcode that allows for pooling downstream. This kit uses the SMART-Seq v4 technology with locked nucleic acids (LNA) for excellent sensitivity and reproducibility. Additionally, this kit includes a complete set of index primers for downstream library preparation. By focusing the majority of cDNA fragments on the 3' end of mRNA transcripts, this kit enables differential expression (DE) analysis with fewer reads, reducing the cost of discovery. When combined with the Illumina® Nextera® XT DNA Library Preparation Kit, this kit results in a library ready for sequencing on Illumina platforms.</p> <p style="text-align: center;"></p> <div style="display: flex; justify-content: space-around; border: 1px solid #ccc; padding: 5px;"> Documents Components Image Data </div>					
635041	SMART-Seq® v4 3' DE Kit	192 Rxns		*	

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
 FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES. © 2023 Takara Bio Inc. All Rights Reserved. All trademarks are the property of Takara Bio Inc. or its affiliate(s) in the U.S. and/or other countries or their respective owners. Certain trademarks may not be registered in all jurisdictions. Additional product, intellectual property, and restricted use information is available at takarabio.com.