that's GOOD science!

TECH NOTE

High-throughput single-cell mRNA-seq

The SMART-Seq HT Kit generates full-length cDNA from cells sorted by FACS   >>

Simplified workflow for high-throughput transcriptome profiling from single cells, featuring a convenient one-step RT-PCR reaction   >>

Comparable sensitivity and reproducibility to the SMART-Seq v4 Ultra Low Input Kit for Sequencing (SMART-Seq v4)   >>

Introduction

The emergence of next-generation sequencing (NGS) technology has encouraged intense interest in single-cell transcriptome analysis. Single-cell RNA-seq has been gaining prominence in basic and clinical research since it can be used to examine differential gene expression, alternative splicing, gene fusions, differentiation processes, tumor composition, and much more (Strijp et al. 2017; Burns, Brooks and Spencer 2016; Cheng et al. 2016; Lau et al. 2016). Single-cell RNA-seq overcomes problems inherent to bulk RNA sequencing, which averages gene expression over a cell population and thereby obscures the diversity of single-cell expression levels, which are susceptible to intrinsic stochastic variation and extrinsic variation from the microenvironment. Therefore, single-cell transcriptome information is critical for understanding fundamental biological phenomena and complex diseases (Kanter and Kalisky 2015).

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 (Fish et al. 2016; Sugiyama et al, 2016; Paul et al. 2017). However, this method is relatively low throughput while researchers may be interested in analyzing hundreds or thousands of individual cells per 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 users can simply set up the one-step RT-PCR and walk away. The updated workflow, available in the SMART-Seq HT Kit, is extremely fast, convenient, and generates a comparable cDNA yield to its predecessor, all while providing the same unparalleled sensitivity and reproducibility.
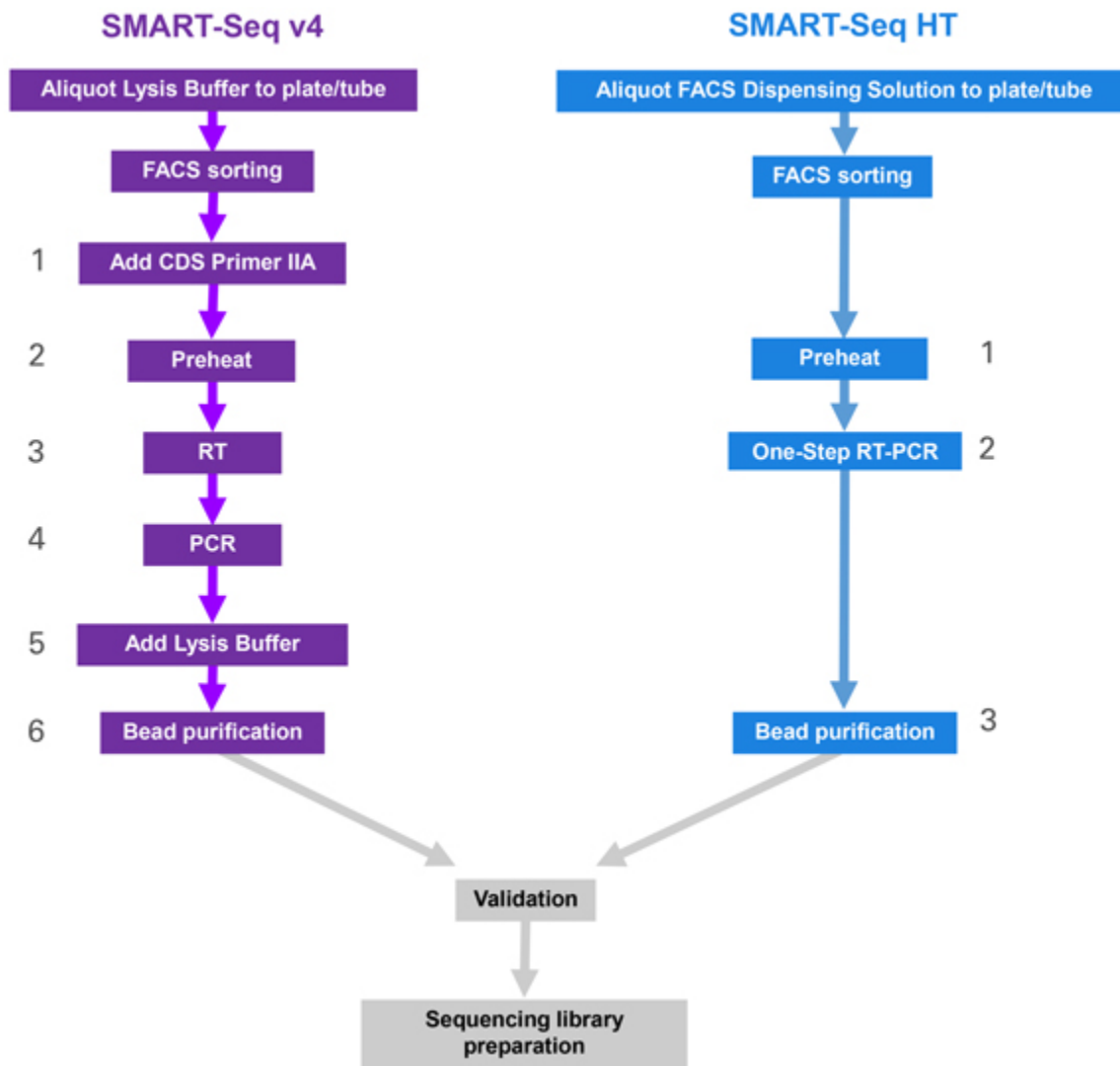
Figure 1. Comparison of the SMART-Seq v4 and SMART-Seq HT kit workflows. The SMART-Seq v4 method (left) was modified to generate a simplified, high-throughput workflow (SMART-Seq HT, right) with very little hands-on time. Once single cells have been obtained using FACS, the SMART-Seq HT Kit involves 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.

Results

Comparable sequencing metrics obtained from total RNA libraries generated with the SMART-Seq v4 and SMART-Seq HT kits

To evaluate the technical performance of the SMART-Seq HT Kit, replicate RNA-seq libraries were generated in triplicate from 10 pg of Mouse Brain Total RNA. 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 (Figure 2, Panel A). Between technical replicates within each kit, there was a 61–63% overlap in the number of transcripts identified (FPKM >0.1; Figure 2, Panel B). These data indicate that with 11,190 transcripts identified across three replicates, the

SMART-Seq HT Kit provides the same sensitivity and reproducibility as the SMART-Seq v4 kit (10,611 transcripts identified). Further analysis indicated that 9,047 transcripts are commonly identified by the two kits, representing an overlap of 71%. These 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 less likely to be identified are the ones expressed at a low level.

A

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

B

SMART-Seq v4
Total=17,439 transcripts

A 14,168   B 13,934

1,133   1,163   996

10,611

1,261   1,164

1,111

C 14,147

SMART-Seq HT
Total=17,792 transcripts

A 14,510   B 14,650

960   1,175   1,104

11,190

1,185   1,181

997

C 14,553

v4   HT

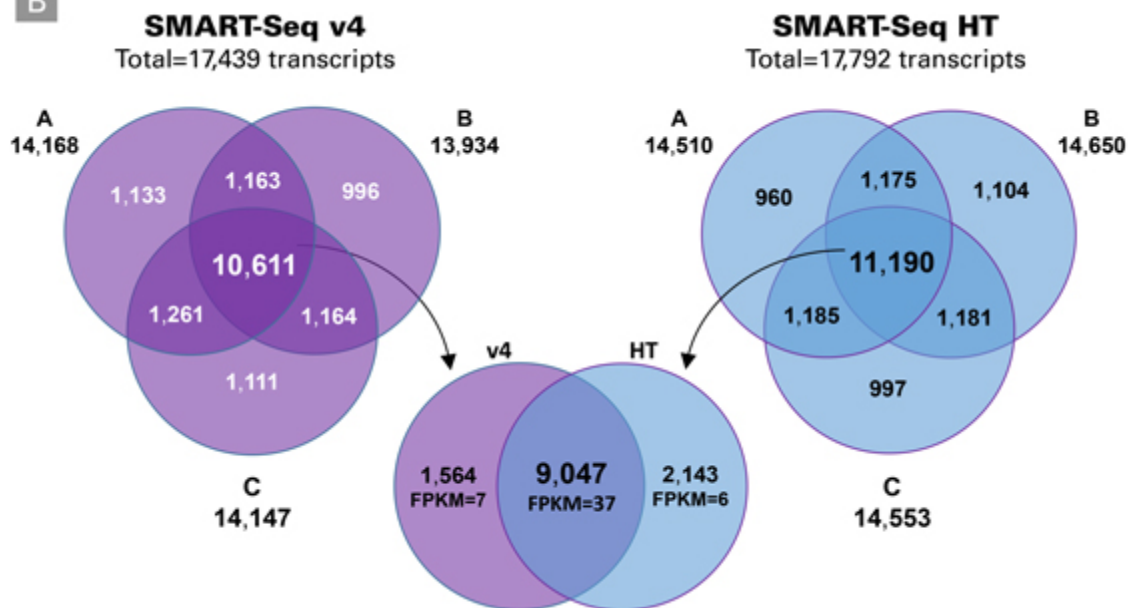1,564 FPKM=7   9,047 FPKM=37   2,143 FPKM=6

Figure 2. High overlap of transcripts identified with the SMART-Seq v4 and SMART-Seq HT kits. Libraries were prepared from 10 pg of Mouse Brain Total RNA. The output cDNA was converted into RNA-seq libraries using the Nextera® XT DNA Library

TaKaRa

Clontech TaKaRa cellartis
www.takarabio.com
© 2016

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 **(Panel A)**. Data were further evaluated for the overlap in the number of transcripts identified (FPKM >0.1) between technical replicates within each kit **(Panel B)**.

Even representation of genes with low and high GC content between the SMART-Seq HT and SMART-Seq v4 kits

One of the improvements in the SMART-Seq v4 kit compared to earlier generations was a better representation of genes with low or high GC content. In order to assess whether the changes introduced in the SMART-Seq HT Kit have affected the representation of genes with low or high GC content, the libraries made from 10 pg of Mouse Brain Total RNA shown in Figure 2 were further analyzed for GC content representation. We examined the number of genes identified in three GC content bins (≤36%, 37–54%, and ≥55%) (Figure 3, Panel A). 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 classified as low GC content (≤36%), 89.9% are classified as medium GC content (37–54%), and 5.4% are classified as high GC content (≥55%). 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. The average gene counts are highly reproducible for replicate samples analyzed using the SMART-Seq v4 or SMART-Seq HT kits. Genes with high or low GC content show similar expression levels in the SMART-Seq v4 and SMART-Seq HT kits (Figure 3).

A

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

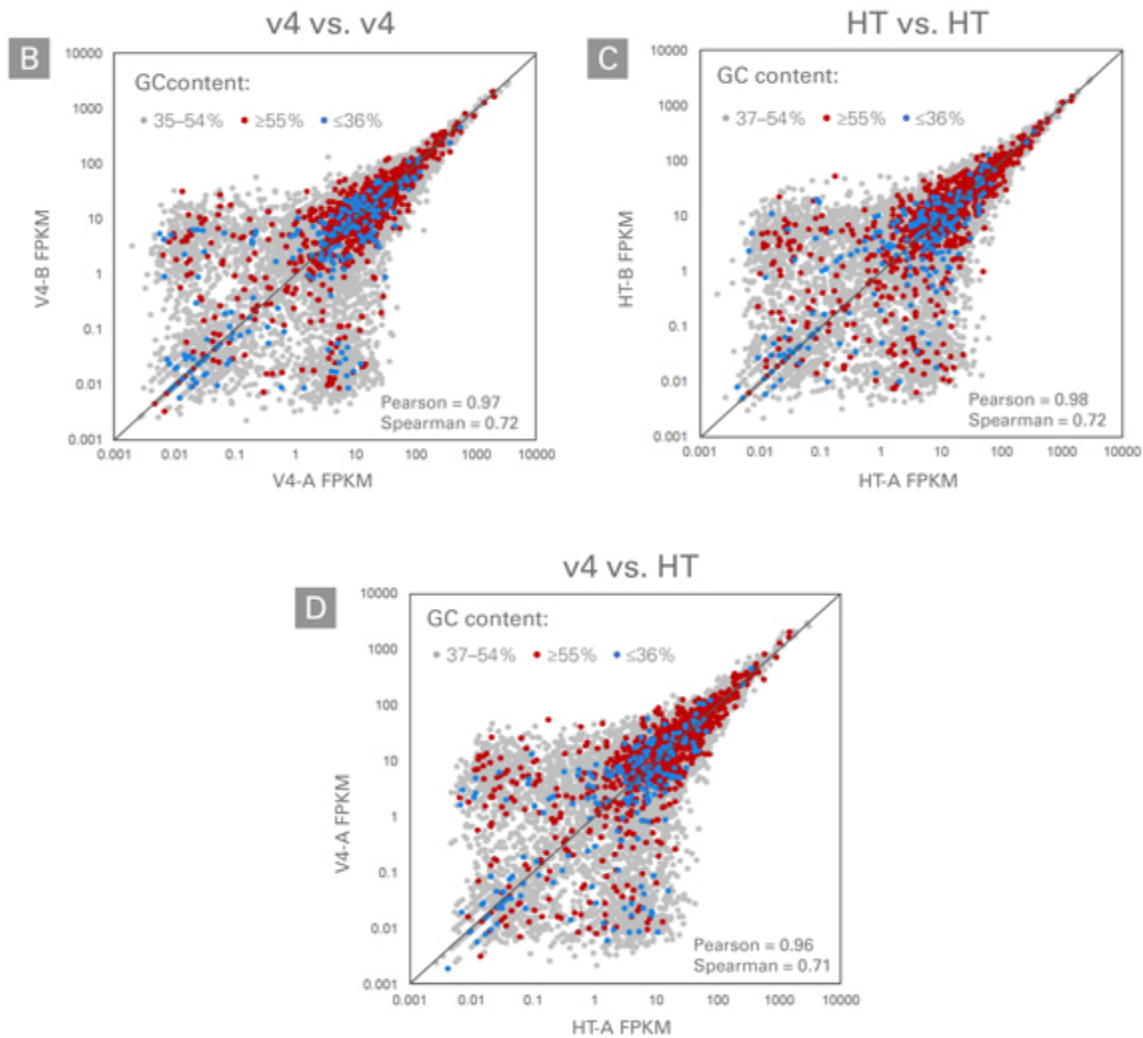|	Number of genes identified (% of total)			Total genes identified
	GC content ≤36%	GC content 37–54%	GC content ≥55%	GC content 0–100%
SMART-Seq HT Kit	272 (2.1)	11,788 (90.6)	954 (7.3)	13,014
SMART-Seq v4 kit	280 (2.2)	11,540 (90.4)	943 (7.4)	12,764
Reference genome	1,665 (4.7)	31,916 (89.9)	1,914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

Figure 3. 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 Figure 2 were further analyzed for GC content representation. 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 **(Panel A)**. The average gene counts are very reproducible for replicate samples analyzed using the SMART-seq v4 **(Panel B)** or SMART-Seq HT kits **(Panel C)**. 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 D)**.

Sequencing metrics from libraries generated directly from cells using SMART-Seq HT and SMART-Seq v4 show similar library profiles

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). As observed for the libraries from Mouse Brain Total RNA, the two kits generated similar sequencing metrics, with a high mapping rate and around 600 additional transcripts identified in the SMART-Seq HT Kit (Figure 4, Panel A). These data indicate that the SMART-Seq HT Kit provides the same sensitivity as the SMART-Seq v4 kit. The hierarchical clustering heat map in Figure 4, Panel B, 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 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 the measurement of gene expression levels.

A

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

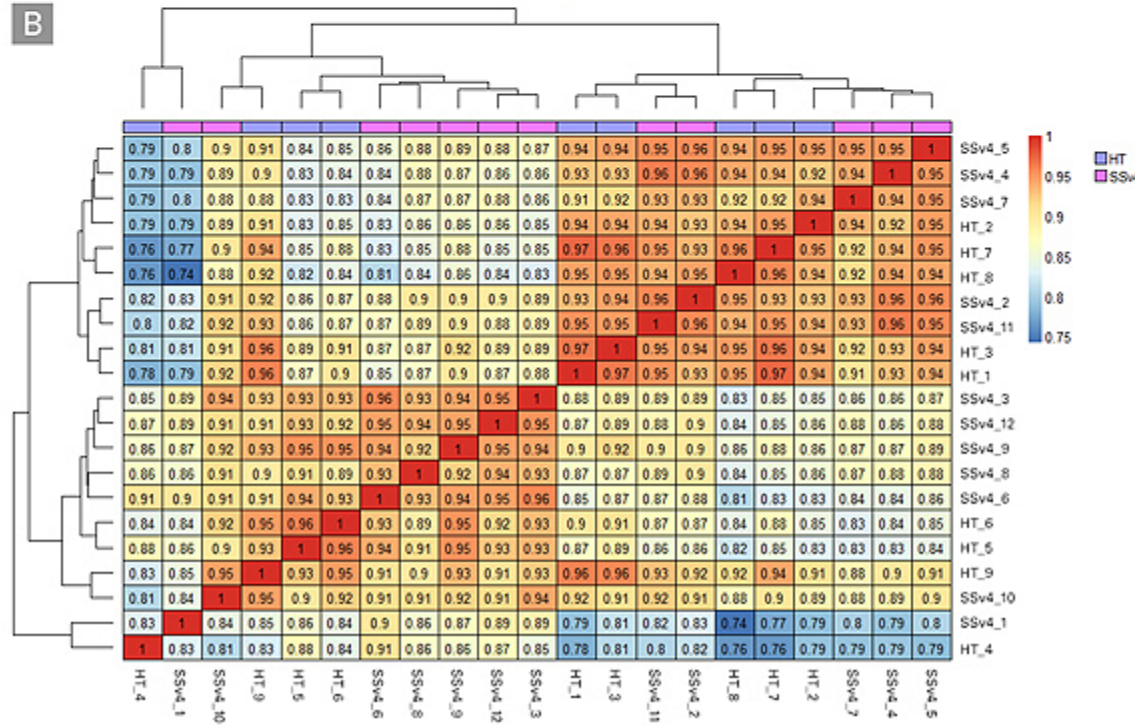| RNA source		293T single cells			
cDNA synthesis		SMART-Seq v4 (n=12)		SMART-Seq HT (n=9)	
		Average	St. dev.	Average	St. dev.
Number of transcripts	FPKM >0.1	12,241	2,132	12,840	2,937
	FPKM >1	9,110	1,468	9,459	1,848
Proportion of reads mapped (%):					
rRNA		0.1	0.04	0.3	0.1
Mitochondria		2.7	0.6	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

B

Figure 4. High reproducibility of gene expression data for 293T cells using the SMART-seq v4 and SMART-Seq HT kits. cDNA libraries were generated from a total of twenty-one individual 293T cells isolated by FACS. The output cDNA was then converted into RNA-seq libraries 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 **(Panel A)**. Data 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) **(Panel B)**. The hierarchical clustering heat map shows Euclidean distances between all the cells and reports Pearson correlations ranging from 0.74 to 0.97.

Conclusions

The new SMART-Seq HT Kit has a simplified workflow for high-throughput transcriptome profiling from single cells or ultra-low input total RNA featuring a convenient one-step RT-PCR reaction. The kit generates full-length cDNA, yielding information about chimeric gene fusions, transcript isoforms, and splice variants that would otherwise be missed if relying solely on 3′ end capture. The SMART-Seq HT Kit provides the same 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 genes with low and high GC content.

Methods

cDNA synthesis was performed starting from single cells isolated by FACS or Mouse Brain Total RNA (Takara Bio, Cat. # 636601) as described in the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing User Manual or the SMART-Seq HT Kit User Manual, using 17 cycles of PCR.

For FACS sorting, 293T cells grown to near confluence were harvested by trypsinization, stained with FITC Mouse anti-Human CD47 (Clone B6H12; BD Biosciences), and resuspended in ice-cold BD FACS Pre-Sort Buffer (BD Biosciences). 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. cDNA was synthesized and 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). 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.

References

Burns, G. W., Brooks, K. E. & Spencer, T. E. Extracellular Vesicles Originate from the Conceptus and Uterus During Early Pregnancy in Sheep1. *Biol. Reprod.* **94,** (2016).

Cheng, L. *et al.* Blocking type I interferon signaling enhances T cell recovery and reduces HIV-1 reservoirs. *J. Clin. Invest.* **127,** 269–279 (2017).

Fish, R. N., Bostick, M., Lehman, A. & Farmer, A. Transcriptome Analysis at the Single-Cell Level Using SMART Technology. *Curr. Protoc. Mol. Biol.* **116,** 1–24 (2016).

Kanter, I. & Kalisky, T. Single Cell Transcriptomics: Methods and Applications. *Front. Oncol.* **5,** 53 (2015).

Lau, C. M. *et al.* Leukemia-associated activating mutation of Flt3 expands dendritic cells and alters T cell responses. *J. Exp. Med.* **213,** 415–31 (2016).

Paul, A.L. *et al.* Genetic dissection of the Arabidopsis spaceflight transcriptome: Are some responses dispensable for the physiological adaptation of plants to spaceflight? *PLoS One* **12,**e0180186 (2017).

van Strijp, D. *et al.* Complete sequence-based pathway analysis by differential on-chip DNA and RNA extraction from a single cell. *Sci. Rep.* **7,** 11030 (2017).

Sugiyama, H. *et al.* Nat1 promotes translation of specific proteins that induce differentiation of mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 340–345 (2017).

View web page >>

http://www.takarabio.com/US/Products/cDNA_Synthesis_and_Library_Construction/
Next_Gen_Sequencing_Kits/Single_cell_RNA_Seq_Kits_for_mRNA_seq/High-
throughput_single_cell_mRNA_seq