SMART-Seq Stranded Kit performance with ovarian cancer cells

Luke Sherlin*, Anastasia Potts, Nathalie Bolduc, Simon Lee, Magnolia Bostick, Andrew Farmer

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

Abstract

Single-cell RNA sequencing (scRNA-seq) approaches are increasingly being used to characterize the abundance and functional state of tumor-associated cell types, and have provided unprecedented detail into cellular heterogeneity. Extracting meaningful biological information from the small amount of RNA in single cells requires a library preparation method with exceptional sensitivity and reproducibility. The SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing (SMART-Seq v4) is an extremely sensitive scRNA-seq library preparation method in part due to its capability to retrieve information from full-length mRNA and not just the 3′ end. However, this method can only capture polyadenylated mRNA. To address this, we have modified our SMART® RNA-seq technology to create the SMART-Seq Stranded Kit, a single-cell RNA-seq library preparation method that relies on random priming instead of oligo dT priming. The SMART-Seq Stranded Kit captures any RNA regardless of polyadenylation status and preserves strand-of-origin information, making it more amenable for distinguishing overlapping genes and comprehensive annotation and quantification of long noncoding RNA (lncRNAs). To show the applicability of the SMART-Seq Stranded Kit in characterizing tumor heterogeneity, we analyzed single cells dissociated from a solid tumor in stage IV ovarian cancer (serous carcinoma). We sorted CD45+ leukocytes and EpCAM+ tumor cells in 96-wells plates. After library preparation, sequencing, and analysis, we detected an average of 4,717 genes in the CD45+ cells and 8,039 genes in the EpCAM+ tumor cells. This analysis enabled identification of well-accepted markers of tumor-infiltrating lymphocytes (TILs) and associated with ovarian carcinoma.

1 Simple workflow for generation of stranded libraries from cells

- SMART technology (Chenchik et al. 1998) is used in a ligation-free protocol to preserve strand-of-origin information. Random priming allows the generation of cDNA from all RNA fragments in the sample, including rRNA.
- When the SMARTScribe™ Reverse Transcriptase (RT) reaches the 5' end of the RNA fragment, the enzyme's terminal transferase activity adds nontemplated nucleotides to the 3' end of the cDNA (shown as Xs). The SMART-Seq Stranded Adapter base-pairs with the nontemplated nucleotide stretch, creating an extended template to enable the RT to continue replicating to the end of the oligonucleotide.
- In the next step, a first round of PCR amplification (PCR1) adds full-length Illumina® adapters, including barcodes.
- The ribosomal cDNA (originating from rRNA) is then cleaved by scZapR in the presence of the mammalian-specific scR-Probes. This process leaves the library fragments originating from non-rRNA molecules untouched, with priming sites available on both 5' and 3' ends for further PCR amplification.
- For inputs below 100 ng or 10 cells, an optional pooling of up to 12 samples after PCR1 allows for greater ease of use by minimizing the number of samples to be processed downstream.

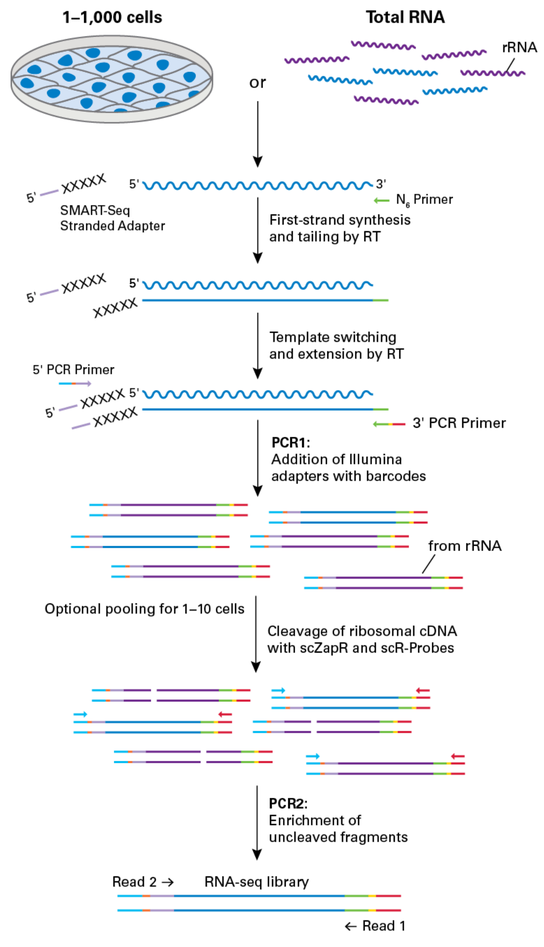

Figure 1. Schematic of technology in the SMART-Seq Stranded Kit.

2 Experimental overview

- A solid tumor from a stage IV ovarian cancer (serous carcinoma) was dissociated to obtain single cells
- CD45+ tumor-infiltrating leukocytes were labeled with CD45-FITC and EpCAM+ tumor cells with EpCAM-PE
- FACS was used to deposit single cells into PCR strips
- SMART-Seq Stranded libraries were then prepared according to the user manual with the pooling workflow
- Libraries were sequenced on a NextSeq® 500 instrument (2 x 75 bp, about 1.5M paired-end reads per sample)
- Data was analyzed with the CLC Genomics Workbench

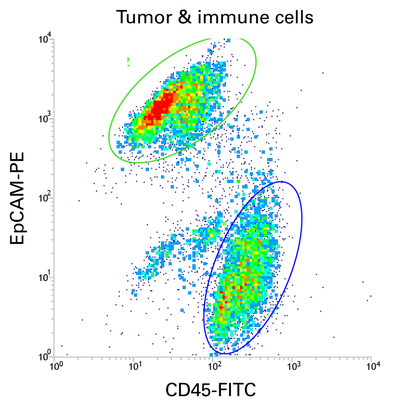

Figure 2. Separation of CD45+ and EpCAM+ cells by FACS.

3 Excellent mapping statistics even from CD45+ leukocytes

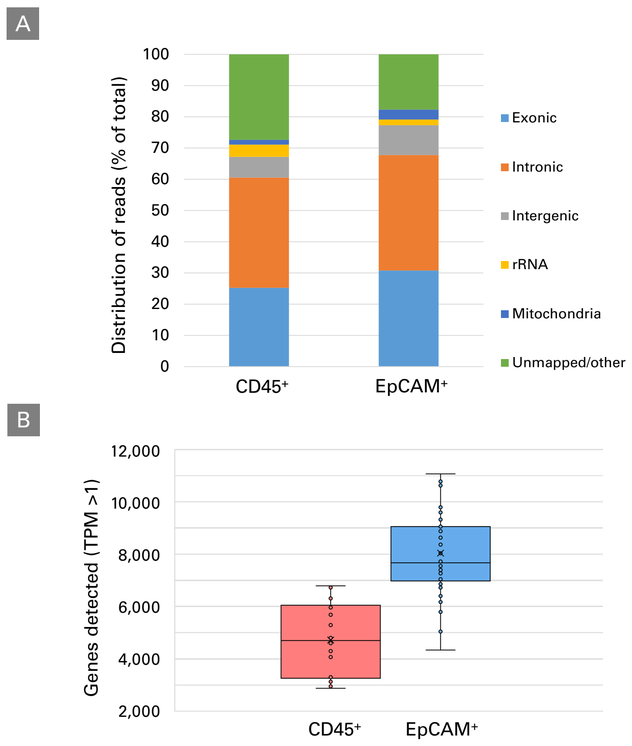

Figure 3. Distribution of reads and genes detected in the two populations. Panel A. Read distribution between the two populations (CD45+: n = 18; EpCAM+: n = 47) is fairly similar, with most of the reads mapping to exons and introns. **Panel B.** The number of genes detected using a TPM cutoff of 1 is higher in the EpCAM+ tumor cells than the CD45+ leukocytes, with a minimum of ~3,000 genes detected in any given cell. The gene detection in CD45+ leukocytes is of particular note given that a large proportion have very low RNA content (e.g., T cells which contain 1–5 pg of RNA).

4 Higher correlations among the EpCAM+ tumor cells

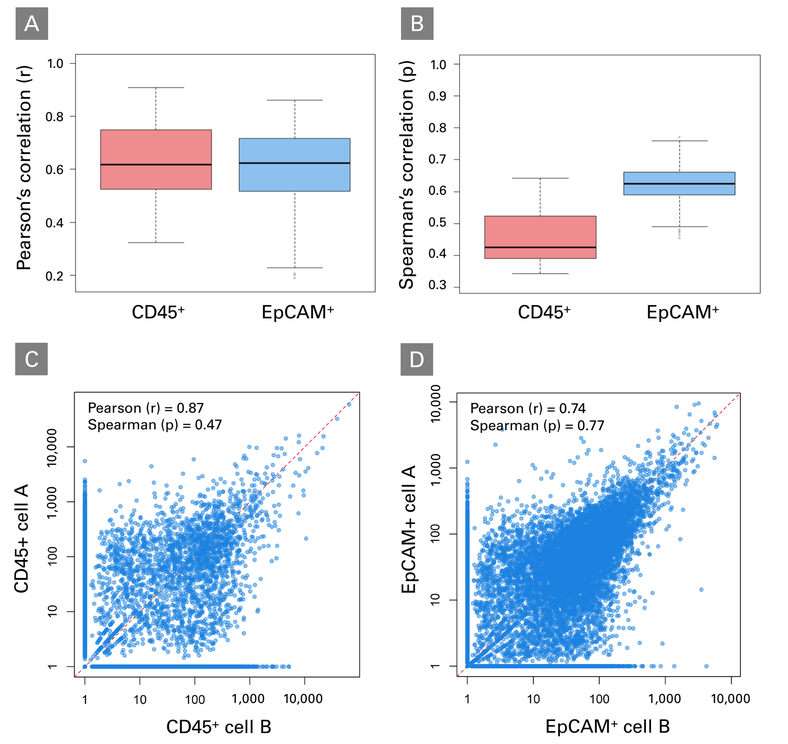

Figure 4. Higher intragroup correlations among the EpCAM+ cells. Correlation boxplots showing intragroup Pearson (**Panel A**) or Spearman (**Panel B**) correlation between all the cells labeled as CD45+ or EpCAM+. The tumor cells (EpCAM+) exhibit a stronger Spearman correlation than the leukocyte cells, possibly reflecting a higher heterogeneity across the leukocytes population. **Panels C and D**. Scatter plots between two cells of each group (CD45+ in Panel C, EpCAM+ in Panel D). The two CD45+ leukocytes in Panel C are CD3+CD8+ T cells.

5 Identification of markers for TILs and ovarian carcinoma

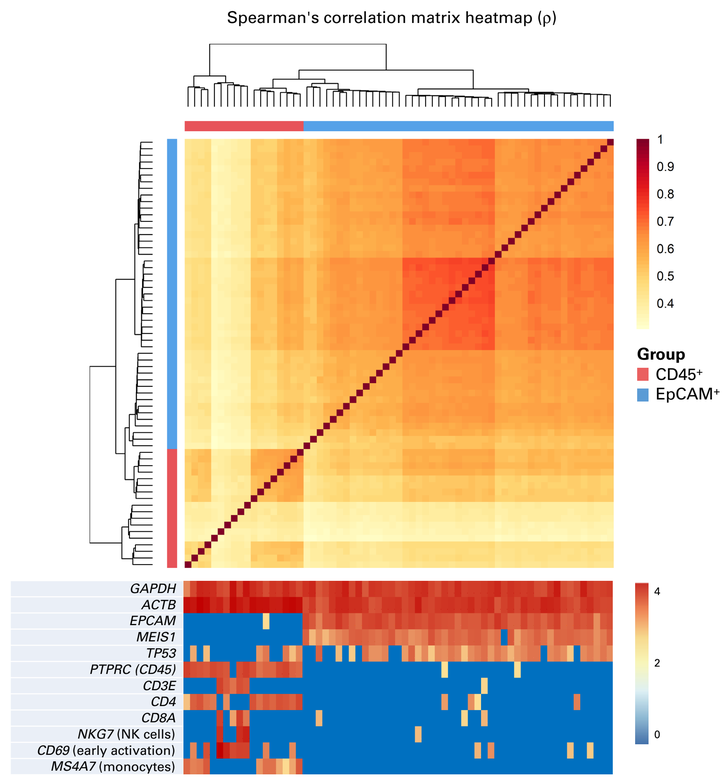

Figure 5. Cells cluster according to their surface marker. Top panel. Hierarchical clustering heatmap displaying the Euclidean distance between all the samples based on Spearman correlation coefficients. **Bottom panel.** Expression levels of a selection of housekeeping genes (*GAPDH, ACTB*); genes associated with cancer (*EpCAM, TP53*); genes particularly important in ovarian cancer (MEIS1 (Crijns et al. 2007)); and genes enriched in TILs (*CD45* and below). Values 0–4 are log10 of the expression level (TPM+1).

6 Uniform gene-body coverage with the SMART-Seq Stranded Kit

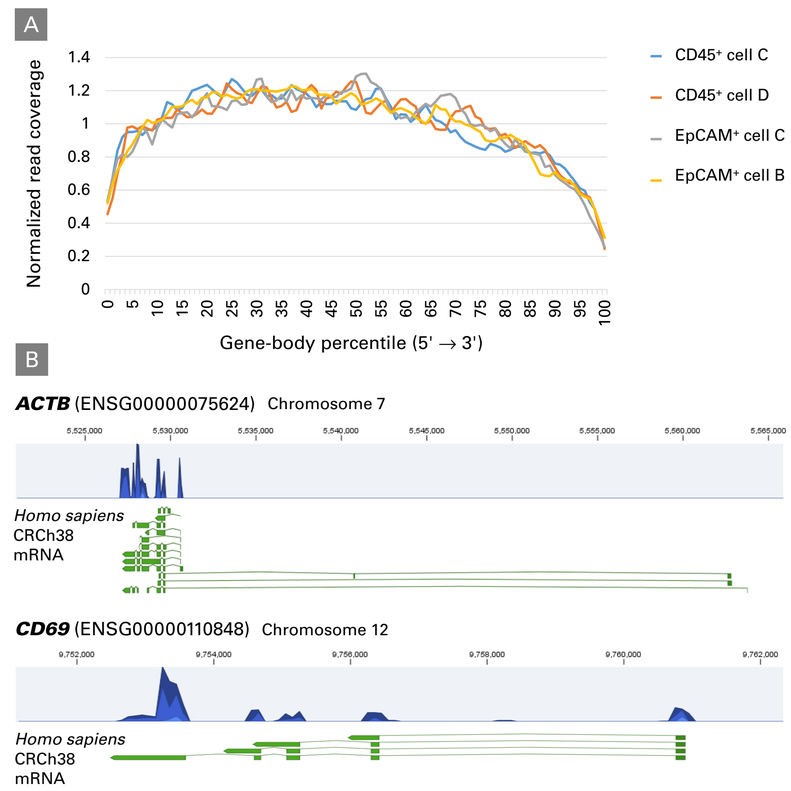

Figure 6. Uniform gene-body coverage with the SMART-Seq Stranded Kit. Panel A. The gene-body coverage is shown for two representative cells from each group (CD45+ or EpCAM+). Similar to what we observe for purified RNA (data not shown), the gene-body coverage from single cells exhibits a slight 5′ bias. **Panel B**. Read coverage at the *ACTB* (top) or *CD69* (bottom) gene locus for a CD3+CD8+ T cell (same as CD45+ cell B shown in Figure 4, Panel C).

Methods

Dissociated tumor cells were purchased from Conversant Bio (now Discovery Life Sciences). Frozen cells were thawed, washed in PBS, and labeled prior to sorting using a BD FACSJazz cell sorter. Cells were sorted in 7 µl of PBS buffer, flash frozen in dry ice, and stored at –80°C until ready to use.

Sequencing libraries were generated using the SMART-Seq Stranded Kit as specified in the 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 (hg19) genome with RefSeq annotation). All percentages shown—including the number of reads that map to introns, exons, or intergenic regions—are percentages of total reads in each library. Cells with high mitochondrial or intergenic read mapping were considered dying or necrotic cells and excluded from the analysis. The number of genes identified for each library was determined upon remapping to hg38 with Ensembl annotation. The gene-body coverage was determined using Picard tools after mapping with STAR. Clustering analysis and correlation boxplots were obtained using in-house software (Takara Bio USA, Inc.).

References

Chenchik, A., Zhu, Y. Y., Diatchenko, L., Li, R., Hill, J., Siebert, P. D. in *Gene Cloning Anal. by RT-PCR* (Siebert, P., Larrick, J.) 305–319 (BioTechniques Books, 1998).

Crijns, A. P. G. *et al.* MEIS and PBX homeobox proteins in ovarian cancer. *Eur. J. Cancer* **43,** 2495–2505 (2007).

Conclusions

- Simple workflow starting directly from 1–1,000 cells, including from cells with very low RNA content, to generate sequencing-ready Illumina libraries in 7 hours
- Unparalleled sensitivity for single-cell, full-length total RNA sequencing with strand-of-origin information
- Reproducible, sensitive detection of coding and noncoding transcripts from total RNA and single cells
- Convenient features make SMART-Seq Stranded Kit the smart choice for analyzing precious samples like tumor-infiltrating leukocytes and characterizing tumor heterogeneity

Scan to download your copy of this poster and view additional posters, or visit takarabio.com/ngs-posters

800.662.2566
Visit us at takarabio.com

Clontech TaKaRa cellartis