Development of a complete, end-to-end, automated single-cell system for biomarker discovery

Yue Yun*, Kazuo Tori, Sally Zhang, Ying Bao, Shuwen Chen, Bryan Bell & Andrew Farmer

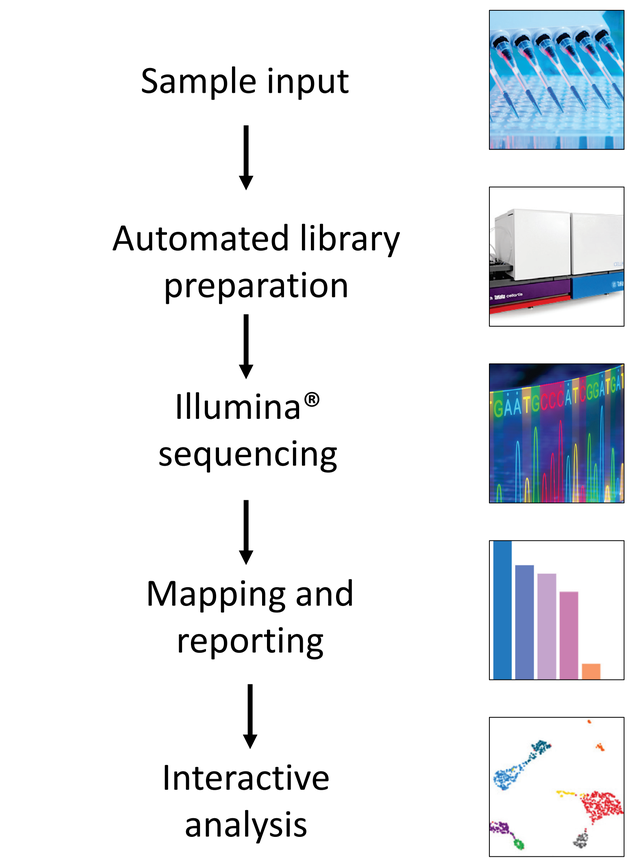
Takara Bio USA, Inc., San Jose, CA 95131, USA *Corresponding Author: yue_yun@takarabio.com

Abstract

Many RNA transcripts undergo post-transcriptional modification such as alternative splicing, with tumor-specific isoforms showing clear diagnostic value as biomarkers. Full-length single-cell mRNA sequencing has contributed to the discovery of these rare biological events; however, automated, high-throughput workflows are desired. Conventional plate-based methods cannot satisfy increasing research needs due to limited sample throughput and lengthy hands-on time. We developed the SMART-Seq® Pro Application Kit, which carries out full-length transcriptome analysis on the ICELL8® cx Single-Cell System. Combining the application kit with the included Cogent™ NGS bioinformatics tools creates an end-to-end, automated solution for biomarker discovery. We compared the sensitivity of the SMART-Seq Pro Application Kit against the Smart-seq2 (SS2)[1] homebrew method and assessed the ability of the SMART-Seq Pro kit to correctly identify clinically relevant splice variants of the *PTPRC* gene. We showed that SMART-Seq Pro kit detects more genes than SS2, highlighting the kit's ability to reveal rare biological events. The SMART-Seq Pro kit also successfully identified PTPRC isoforms of interest, which were missed by 3' end-counting methods. These data illustrate how the SMART-Seq Pro kit for the ICELL8 cx Single-Cell System can characterize clinically relevant isoforms and, together with Cogent NGS analysis tools, offer an efficient way to accelerate translational research by providing insight of important biomarkers in single-cell samples.

1. Picelli, Nat Protoc 9, 171–181 (2014).

SMART-Seq Pro Application Kit workflow



The SMART-Seq Pro kit workflow goes from sample isolation to cDNA synthesis to library prep of challenging samples. Input samples into the ICELL8 cx Single-Cell System to generate Illumina-ready sequencing libraries. Sequence on an Illumina instrument and use the resulting FASTQ files as input for our free analysis software, Cogent NGS Analysis Pipeline, for mapping and reporting, followed by Cogent NGS Discovery Software for interactive data visualization.

1 SMART-Seq Pro uncovers the most biological information from single cells

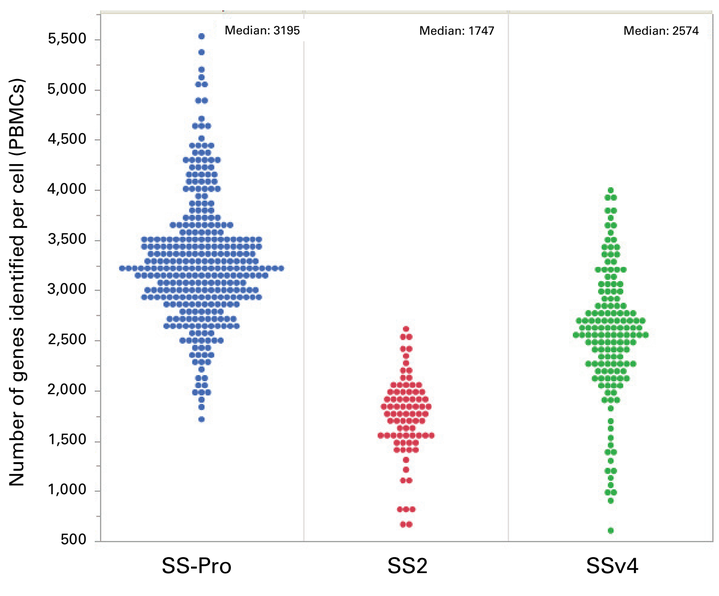


Figure 1. The SMART-Seq Pro Application Kit for the ICELL8 cx system offers robust detection of genes and transcripts. The SMART-Seq Pro Application Kit for the ICELL8 cx system (SS-Pro) identifies more genes than the popular plate-based Smart-seq2 (SS2) and SMART-Seq v4 (SSv4) methods in primary cells with low RNA content such as PBMCs.

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

2 Gene-level analysis shows high expression of *PTPRC* across all gene-based clusters

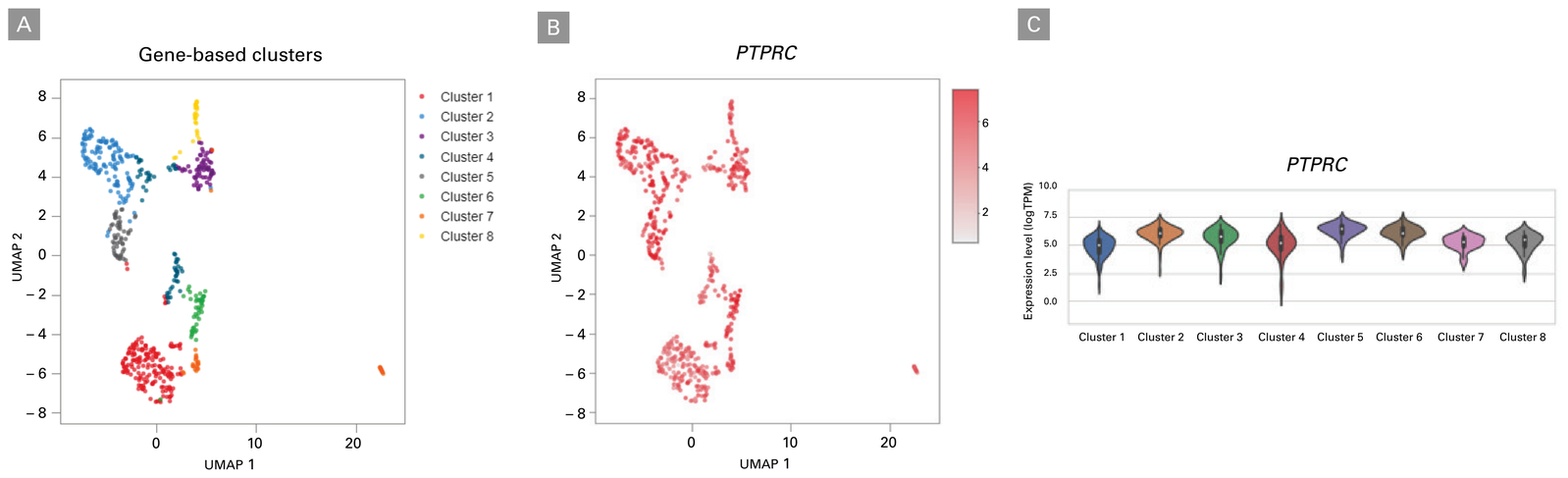


Figure 2. Gene-level analysis shows high expression of *PTPRC* across all clusters. Full-length, single-cell Human PBMC libraries were generated using the SMART-Seq Pro Application Kit for the ICELL8 cx Single-Cell System and sequenced using the Illumina NextSeq® 500 system. Cogent NGS Analysis Pipeline (CogentAP) v1.5 was then used to generate gene matrices that were then input into Cogent NGS Discovery Software (CogentDS) v1.5 to perform clustering analysis and generate UMAP plots. **Panel A.** UMAP plots showing Human PBMC cell clusters based on differentially expressed genes. **Panel B.** UMAP plot showing expression of the *PTPRC* gene, which encodes a hematopoietic-specific transmembrane protein tyrosine phosphatase, also known as CD45, across gene-based clusters. **Panel C.** Violin plot showing the overall distribution and density of the *PTPRC* gene by gene-based cluster. At the gene level, *PTPRC* had high expression across all clusters.

3 SMART-Seq Pro adds transcript-level insights missed by basic gene expression profiling

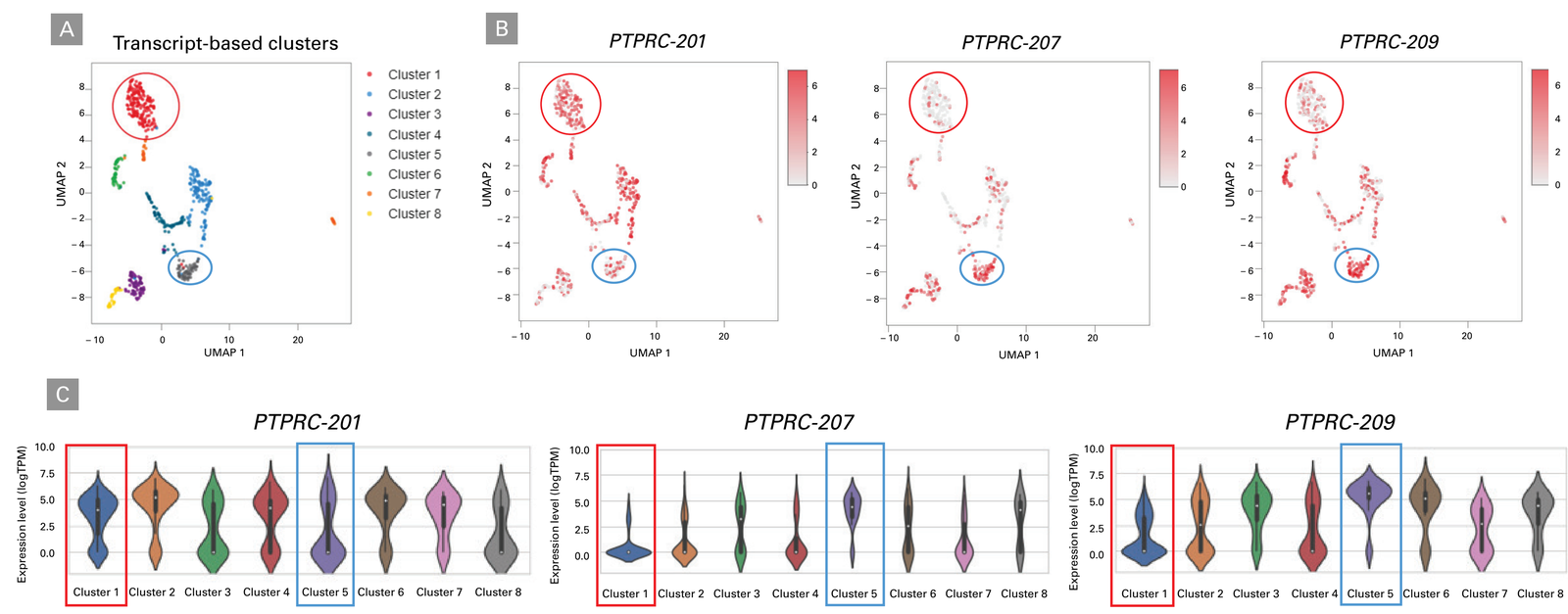


Figure 3. SMART-Seq Pro offers transcript-level insight for detection of different *PTPRC* isoforms in addition to gene expression profiling, providing extra information which can only be obtained using a full-length approach. Cogent NGS Analysis Pipeline (CogentAP) v1.5 was used to generate transcript matrices that were then input into Cogent NGS Discovery Software (CogentDS) v1.5 to perform clustering analysis and generate UMAP plots. **Panel A.** UMAP plots showing Human PBMC cell clusters based on transcript expression. Transcript-based Clusters 1 (red circle) and 5 (blue circle) were chosen for further analysis. **Panel B.** Expression of different *PTPRC* isoforms generated through alternative splicing depends on the state of activation and differentiation of hematopoietic cells. UMAP plots showing the different expression levels of distinct *PTPRC* isoforms, *PTPRC-201, PTPRC-207*, and *PTPRC-209* by transcript-based clustering analysis. Transcript-based Clusters 1 and 5 are highlighted by red circles and blue circles, respectively. **Panel C**. Violin plots showing the expression levels of *PTPRC* isoforms across each cell cluster. At the transcript level, different *PTPRC* transcripts, which result in different isoforms, displayed distinct patterns of expression within the same cluster, as highlighted by Cluster 1 (in red boxes) and Cluster 5 (in blue boxes).

4 SMART-Seq Pro successfully identified *PTPRC* isoform changes missed by 3′ end counting

Figure 4. The SMART-Seq Pro Application Kit on the fully-automated ICELL8 cx system could discern isoform expression data that was lost on the 10x Chromium system. The SMART-Seq Pro method identified junction reads across exons 3, 4, 5, 6, and 7 and revealed distinct profiles for *PTPRC-201* and *PTPRC-209* isoforms. Data generated from 10x Chromium end-counting methods could not identify junctions in these same exonic regions, with peaks representing noise from misaligned reads.

Conclusions

- SMART-Seq Pro is a fully automated workflow on the ICELL8 cx Single-Cell System with all reaction steps happening at nanoliter scale.
- SMART-Seq Pro detects more genes per single cell than plate-based, homebrew methods such as Smart-seq2, and other single-cell platforms.
- The SMART-Seq Pro kit, paired with Cogent NGS tools, detected differences in the expression of different *PTPRC* isoforms within a PBMC population that would have been missed with end counting methods like 5′DE or 3′DE.

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

takarabio.com