Sensitivity of gene fusion detection from high-throughput SMART-Seq chemistry on the ICELL8 system

Rachel Fish, Michael J. Young, Lyndon Liu, Sherry Wei, Karthikeyan Swaminathan, Magnolia Bostick, Nidhanjali Bansal, Necip Mehmet, Suvarna Gandlur, and Andrew Farmer* Takara Bio USA, Inc., Mountain View, CA, USA; *Corresponding author: andrew_farmer@takarabio.com

Abstract

The use of next-generation sequencing for transcriptome analysis in clinical and applied spaces requires accurate, parallel processing of large numbers of samples and the availability of chemistries that enable robust library preparation from the desired targets. The method chosen for library preparation—for example, end-capture versus full-length transcript capture—will yield different quality data to aid in answering specific questions about gene expression. Full-length capture, which provides more uniform coverage of the transcript, enables more thorough examination of gene fusions, SNP detection, and alternative splicing over other methods that capture only the 3′ or 5′ end of the transcript. The application of our full-length SMART-Seq® chemistry on the ICELL8® system provides a high-throughput solution to obtaining richer data on single-cell transcriptomics. We identified fusions in a solid tumor cell line as well as in a leukemia cell line using this application that could not be identified with a 3′ DE approach, with fusions discovered in more than three times as many single cells. This method can also be used to obtain a full understanding of the immune response in single immune cells by combining cell type identification with paired T-cell receptor and B-cell receptor information—all obtained using different analyses of a single whole transcriptome assay. This sensitive SMART-Seq method for automated, full-length RNA-seq offers benefits in increased gene body coverage, enabling improved detection of fusions, SNPs, and splice variants—applications that will aid in the advancement of biomarker identification and the development of novel therapeutics.

1 SMART-Seq ICELL8 application kit workflow

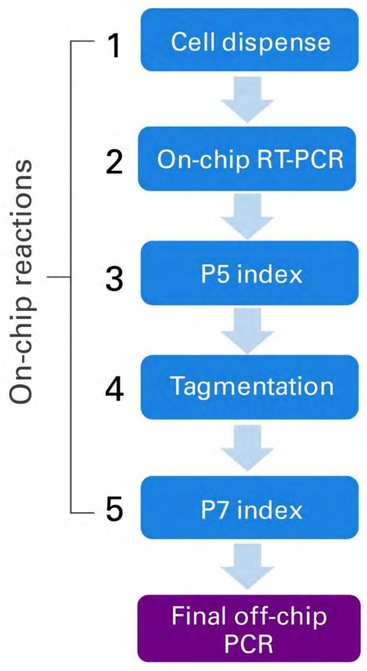

Figure 1. ICELL8 instrument and SMART-Seq ICELL8 application kit workflow: full-length scRNA-seq on the ICELL8 system. The ICELL8 system is an open-platform automated system that provides unparalleled sensitivity and flexibility. The full-length application kit's 8-hr protocol requires five ICELL8 dispensing steps. Cultured cells are dispensed into the wells of a blank, 5,184-well ICELL8 chip at an average of 1 cell/well (Step 1). Built-in cell-selection software is used to identify single cells. Cell lysis is followed by cDNA synthesis and amplification (Step 2). Full-length cDNA is tagmented with Illumina® Nextera® TDE1 and amplified with Illumina-specific indexed adapters added in a 72x72 grid-like fashion to generate uniquely indexed cDNA libraries in each well (Steps 3–5). The final libraries are pooled, further amplified, and purified prior to sequencing.

Methods and references

For experiments performed on the ICELL8 system: Cells (Figure 3: K-562 cells [ATCC CCL-243]; Figure 4: HCC2157 [ATCC CRL-2340] and control cells) were dispensed across an ICELL8 chip together with 48 negative-control wells (containing only PBS) and 48 positive-control wells (containing 10 pg K-562 RNA). The ICELL8 system's cell-selection software was used to identify candidate single-cell wells. For Figure 3, 1,000 single-cell wells of K-562 were selected, and for Figure 4, 1,122 single-cell wells of HCC2157 were selected. These single-cell wells, together with the positive and negative controls, were then processed according to the SMART-Seq ICELL8 application kit protocol using 12 cycles of PCR during the RT-PCR step. The resulting libraries were sequenced together on an Illumina NextSeq® instrument with paired-end 2 x 75-bp reads. HCC2157 cells had a mean of 284,000 reads per cell. Demultiplexing was performed using mappa™ software from Takara Bio.

For experiments performed on the 10X system: Single-cell capture and library preparation were performed by MedGenome using Chromium Single Cell 3′ gene expression v3 chemistry. For Figure 3, ~1,500 K-562 cells were captured and processed, and for Figure 4, HCC2157 cells were captured and processed. The resulting libraries were sequenced on an Illumina NovaSeq™ instrument with PE100 chemistry. The HCC2157 cells had a mean of 50,000 reads per cell, according to the manufacturer's recommendation. Demultiplexing was performed using Cell Ranger software.

References

Haas, B. J. *et al*. STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv* 120295 (2017). at <https://www.biorxiv.org/content/10.1101/120295v1>

Broad Institute. DepMap: The Cancer Dependency Map Project at Broad Institute. (2019). at <https://depmap.org/portal/>

Conclusions

Takara Bio's SMART-Seq ICELL8 application kit workflow:

- Captures both junction and spanning reads to confidently identify gene fusions
- Identifies fusions across the full transcript length
- Provides greater fusion identification at a lower sequencing depth
- Provides full-length sequence information, high sensitivity, and high reproducibility

2 Bioinformatics analysis

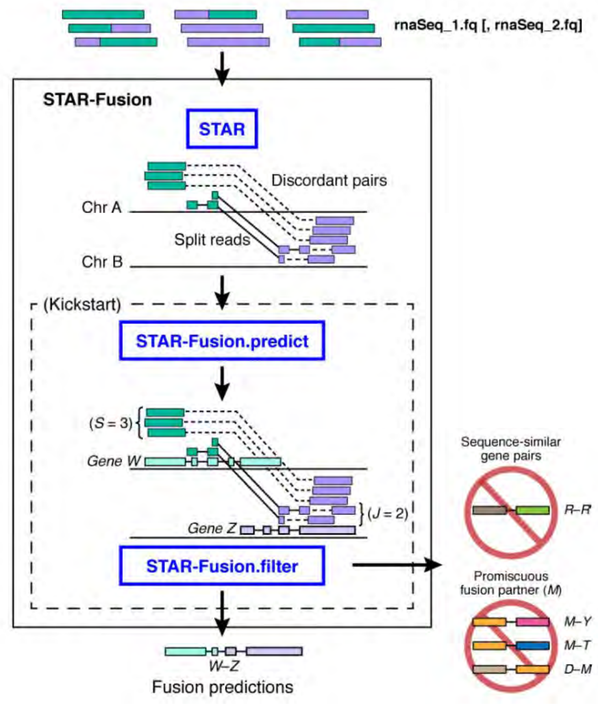

Figure 2. Overview of bioinformatics analysis. Figure 1 from Haas et al. 2017 was used unmodified and adopted under CC-BY-ND 4.0 (https://creativecommons.org/licenses/by-nd/4.0/). In the first step of the bioinformatics pipeline, labeled in the blue box as "STAR," the STAR algorithm aligns the reads to the reference genome (hg38, in our case). Two types of alignments can occur. One alignment type is "discordant" (dashed line), indicating that each read pair originates from two genes that have undergone chromosomal translocation. The second alignment type is "split" (solid line), indicating that a single read overlaps the junction between two genes. In the second step, labeled "STAR-Fusion.predict," STAR-Fusion interprets the reads: discordant reads become spanning reads (S) and split reads become junction reads (J). SMART-Seq full-length chemistry uses paired-end reads, so both junction and spanning reads are captured. 10X 3′ DE uses a single-end read approach, so *only* junction reads can be captured. The final step, labeled "STAR-Fusion.filter," was used in Haas *et al.* For our experiments detecting single-cell fusions, instead of applying the STAR-Fusion.filter, we used a relaxed setting that does not apply a filter. For bulk analysis, STAR-Fusion.filter may be more appropriate.

3 Full-length chemistry detects more fusions

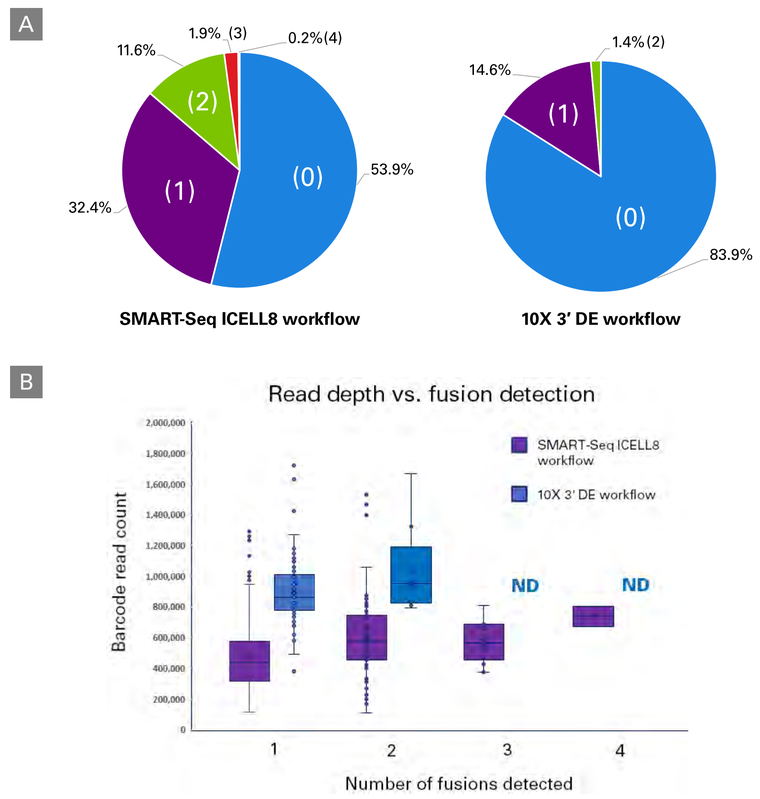

Figure 3. The SMART-Seq ICELL8 workflow detects a higher percentage of cells with fusions. Panel A. In this proof-of-concept experiment using K-562 cells, fusions were counted if they were reported by "DepMap: The Cancer Dependency Map Consortium" or supported in the literature. This led to a total of eight fusions identified across the two platforms. The SMART-Seq ICELL8 workflow detected fusions in 46% of the cell population, while the 10X 3′ DE workflow detected fusions in only 16% of the cell population. The ICELL8 workflow was able to identify cells with up to four unique fusions, while the 10X workflow could not identify three or four fusions in any cell. **Panel B.** The distribution of reads is shown for cells in which at least one fusion was identified. The SMART-Seq ICELL8 workflow detected up to four fusions per cell at a lower read depth than the 10X 3′ DE workflow, which was not able to detect more than two fusions per cell.

4 Full-length chemistry detects junction and spanning reads, improving fusion identification

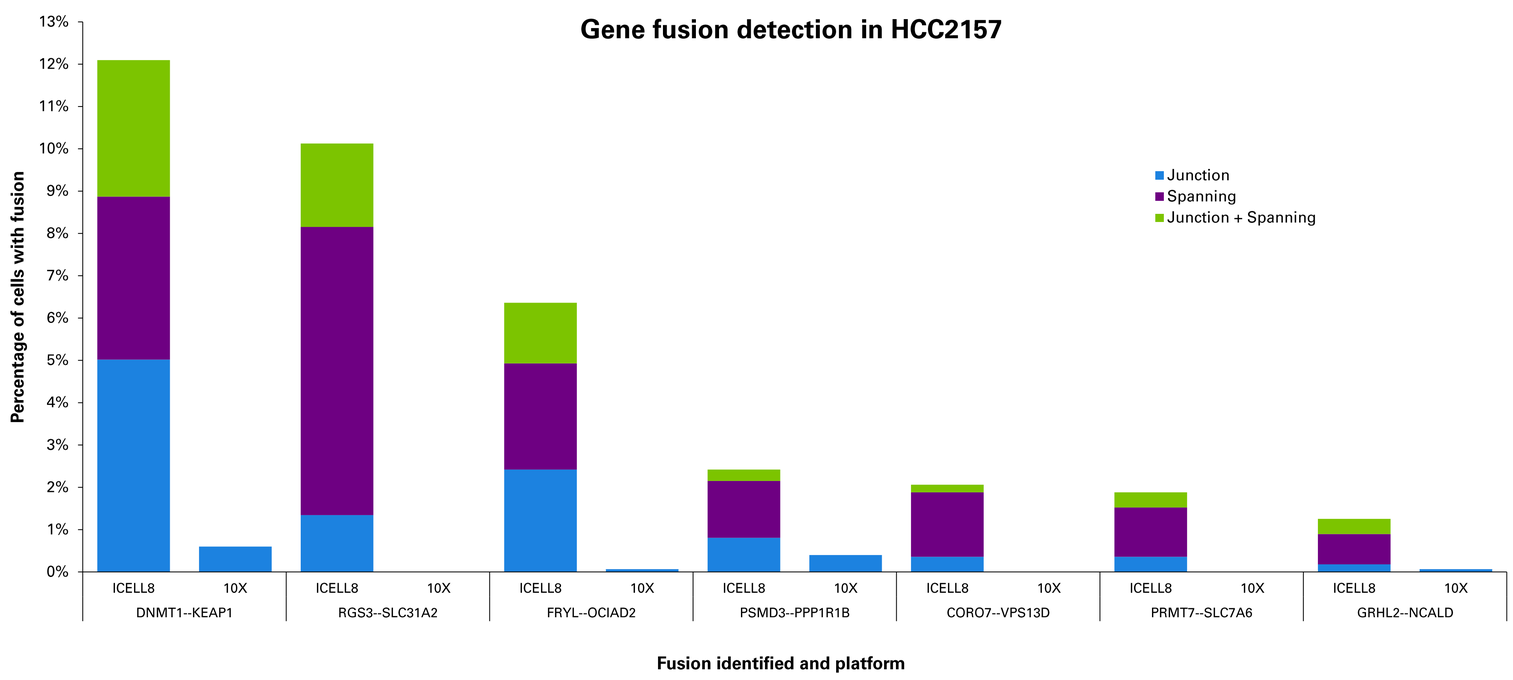

Figure 4. The SMART-Seq ICELL8 workflow uniquely detects both junction and spanning reads, improving confidence in fusion detection. HCC2157 is a breast cancer cell line. A total of 18 fusions were detected in the HCC2157 line and confirmed to be reported in "DepMap: The Cancer Dependency Map Consortium." In this figure, fusions identified in at least 1% of the cells on either platform are shown (with the exception of those examined in more detail in Figure 5). The 10X 3′ DE workflow ("10X") is only able to detect junction reads, whereas the SMART-Seq ICELL8 workflow ("ICELL8") detects a combination of junction and spanning reads. This ICELL8 advantage, made possible by full-length coverage, leads to a higher sensitivity of fusion identification, which is supported by the higher percentage of cells detected with fusion events.

5 Full-length chemistry provides higher confidence in fusion discovery

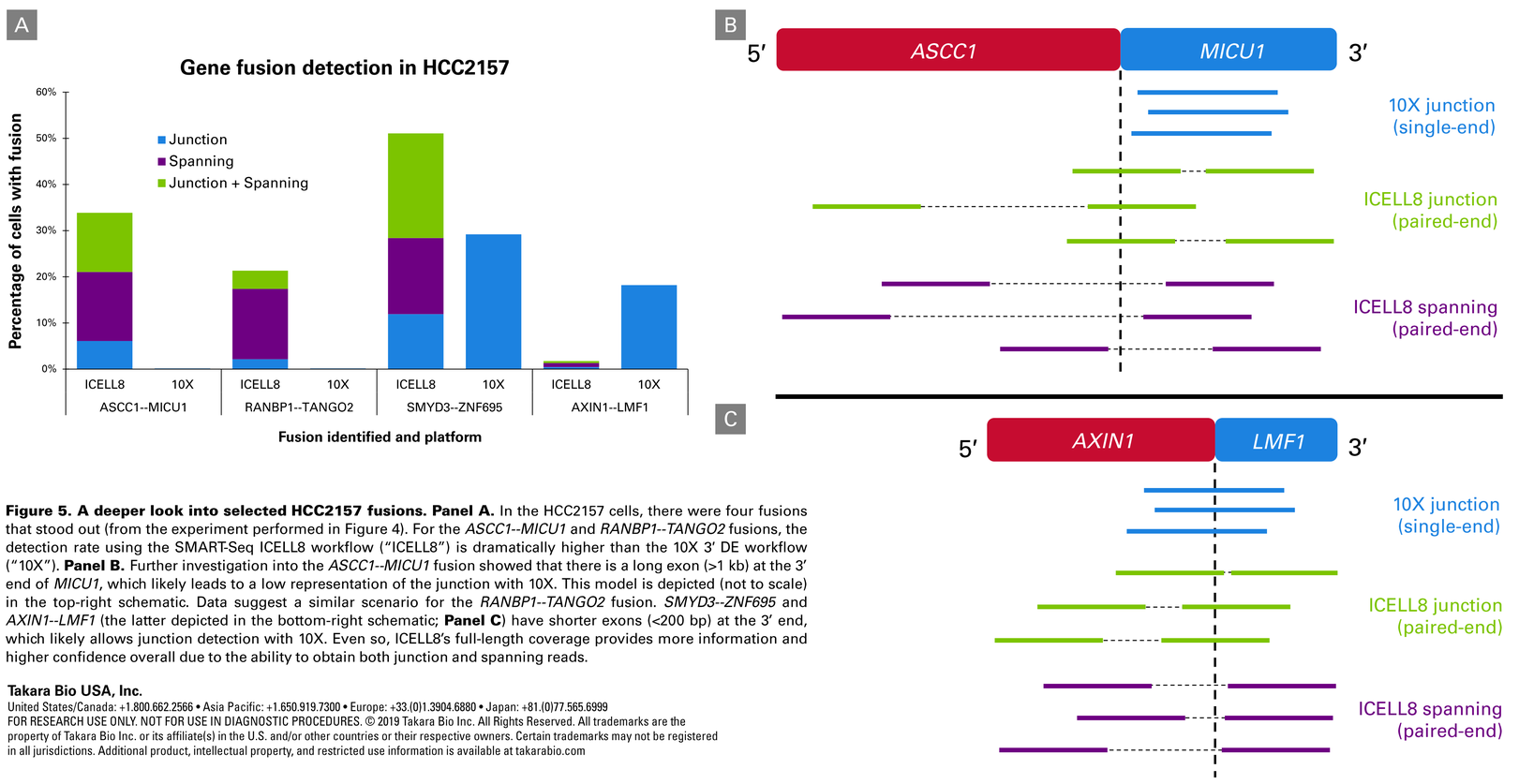

Figure 5. A deeper look into selected HCC2157 fusions. Panel A. In the HCC2157 cells, there were four fusions that stood out (from the experiment performed in Figure 4). For the *ASCC1--MICU1* and *RANBP1--TANGO2* fusions, the detection rate using the SMART-Seq ICELL8 workflow ("ICELL8") is dramatically higher than the 10X 3′ DE workflow ("10X"). **Panel B.** Further investigation into the *ASCC1--MICU1* fusion showed that there is a long exon (>1 kb) at the 3′ end of *MICU1*, which likely leads to a low representation of the junction with 10X. This model is depicted (not to scale) in the top-right schematic. Data suggest a similar scenario for the *RANBP1--TANGO2* fusion. *SMYD3--ZNF695* and *AXIN1--LMF1* (the latter depicted in the bottom-right schematic; **Panel C**) have shorter exons (<200 bp) at the 3′ end, which likely allows junction detection with 10X. Even so, ICELL8's full-length coverage provides more information and higher confidence overall due to the ability to obtain both junction and spanning reads.

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

800.662.2566
Visit us at takarabio.com

Clontech TaKaRa cellartis