Enabling long-read mRNA-seq for samples with limited input amounts

Lisa Welter, Gilma Sevilla, Yana Ryan, Kazuo Tori, Yue Yun, Sherry Wei, Tomoya Uchiyama, Mike Covington, Mohammad Fallahi, Saloni Pasta, Bryan Bell, and Andrew Farmer*

Takara Bio USA, Inc., San Jose, CA 95131, USA *Corresponding Author

Abstract

Due to its role in encoding proteins, the ultimate effectors of cell function, and the simplicity of its identification through sequencing, mRNA has become a key indicator in scientific fields such as cancer research, developmental biology, neurobiology, and immunology. Significantly, changes in mRNAs that effect protein function—for example, alternative splicing, SNVs and gene fusions—can be effective biomarkers in these settings.

Third-generation sequencing technologies, such as Oxford Nanopore Technologies (ONT), provide the opportunity to sequence full-length cDNA without the need for fragmentation and hence provide a more complete picture of isoforms present. However, a current limitation of long-read RNA sequencing is the requirement for high input amounts that can be unachievable for some primary sample types, like RNA isolated from tumor samples or sorted blood cancer cells.

SMART® technology (**S**witching **M**echanism **a**t the 5' end of **R**NA **T**emplate) has emerged as the best solution for detecting and analyzing RNA from a wide range of samples, especially from low input, due to its sensitivity, simplicity, and ability to capture the entire RNA transcript in a single step. An efficient and comprehensive analysis of the full-length cDNA generated by SMART technology has previously been hampered by short-read sequencing, which fragments the resulting output, making it challenging to obtain a full picture of isoform repertoire.

With the improved productivity and accuracy of the ONT platform, combining it with SMART technology in the SMART-Seq® mRNA Long Read (LR) workflow to capture full-length RNA transcripts provides a more complete picture of isoform-specific changes in critical sample types. It is now possible to interrogate up to 96 samples from as little as 10 pg–100 ng of input RNA, which highlights the new opportunities that this technology provides to the field of biomarker discovery.

1 SMART-Seq mRNA LR workflow

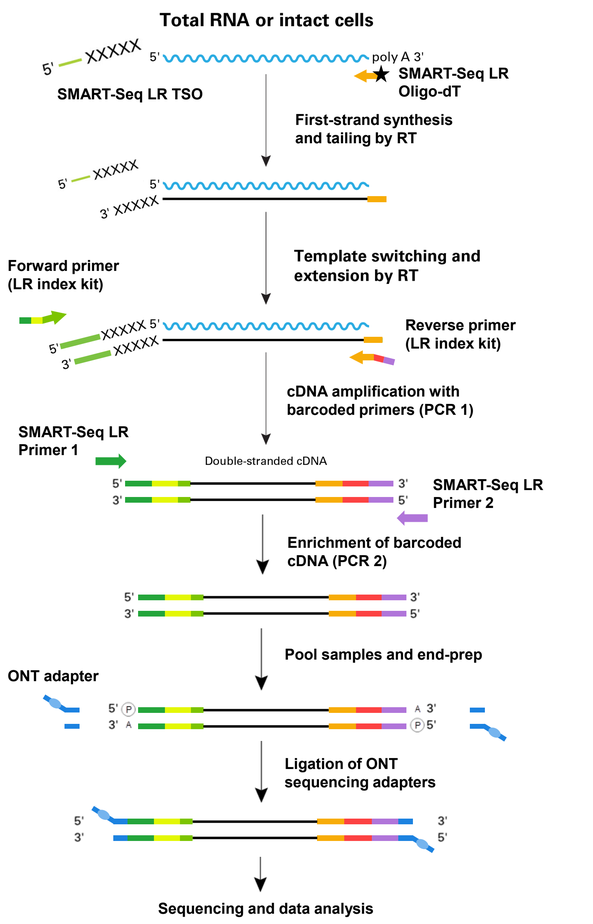


Figure 1. Library preparation workflow for the SMART-Seq mRNA Long Read kit. Starting with high-integrity RNA (RIN>8) or intact cells, first-strand cDNA synthesis is primed by the SMART-Seq LR Oligo-dT and performed by an MMLV-derived reverse transcriptase (RT). Upon reaching the 5' end of each mRNA molecule, the RT adds non-templated nucleotides to the first-strand cDNA. The SMART-Seq LR TSO contains a sequence that is complementary to the non-templated nucleotides added by the RT. This primer hybridizes to the first-strand cDNA. In the template-switching step, the RT uses the remainder of the SMART-Seq LR TSO as a template for the incorporation of an additional sequence on the end of the first-strand cDNA. The first-strand cDNA is then barcoded by forward and reverse primers from the SMART-Seq Long Read Index Kit (96 rxns) and amplified by the first PCR; a second PCR enriches for barcodes fragments. Samples are pooled and end-prepped, and sequencing adapters are ligated using the Ligation Sequencing Kit V14 (ONT). After sequencing, samples are basecalled and demultiplexed using guppy. Downstream data analysis was performed using cutadapt, minimap2, samtools, bedtools and salmon.

2 SMART-Seq mRNA LR read-length distribution

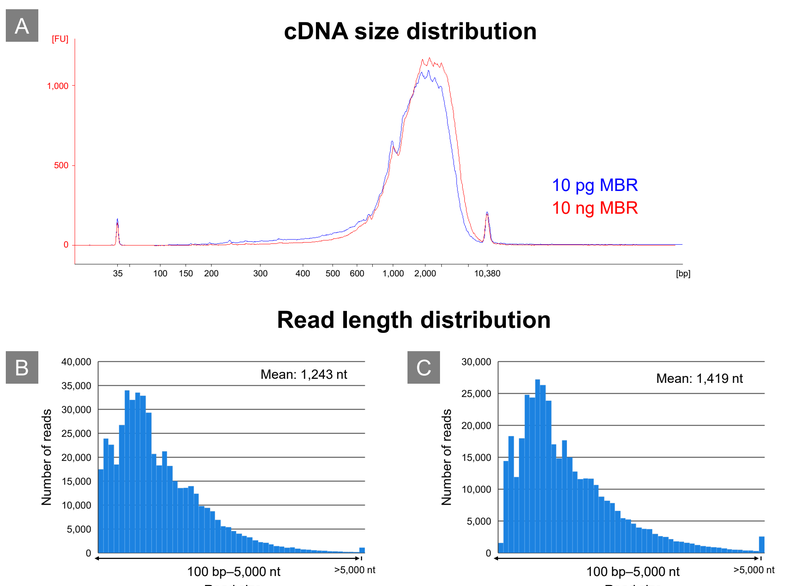


Figure 2. SMART-Seq mRNA LR generates barcoded cDNA with a wide range of fragment lengths. SMART-Seq mRNA LR workflow was used to create cDNA from 10 pg or 10 ng total mouse brain RNA (n=8). cDNA size distribution was measured on a 2100 Bioanalyzer (Agilent Technologies) using an Agilent High Sensitivity DNA Kit (**Panel A**). Barcoded cDNA was pooled per input, and libraries were generated using the Ligation Sequencing Kit V14. Libraries were sequenced on a MinION Flow Cell (ONT) for 72 hr. Samples were basecalled and demultiplexed using guppy, and read-length distribution was plotted using MS-Excel. Read length distributions are shown for a representative sample of 10 pg (**Panel B**) and 10 ng (**Panel C**) total mouse brain RNA.

3 SMART-Seq mRNA LR shows high sensitivity across broad input range

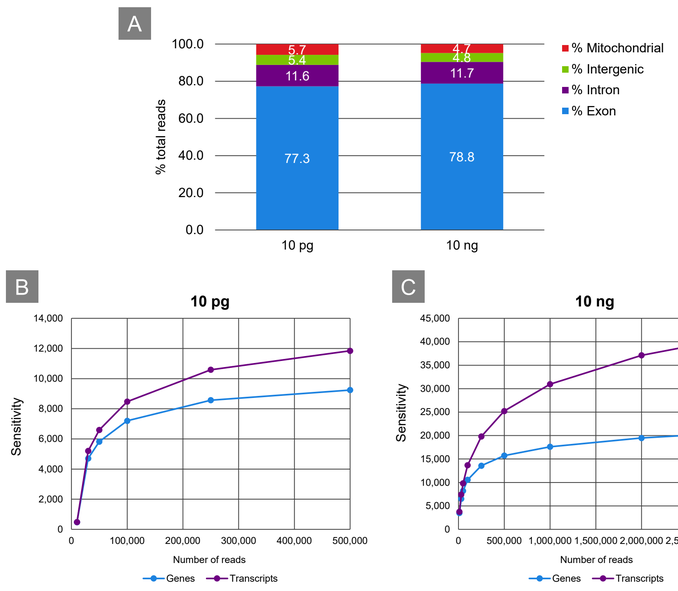


Figure 3. SMART-Seq mRNA Long Read demonstrates high sensitivity across a broad range of inputs. To evaluate the performance of the SMART-Seq mRNA LR workflow, cDNA was generated and sequencing libraries werefrom 10 pg and 10 ng total mouse brain RNA (MBR) using the workflow described in Figure 1. After sequencing, data was basecalled and demultiplexed using guppy, and reads per barcode were downsampled to 400,000 reads. The read distribution for 10 pg and 10 ng MBR input (**Panel A)** is similar, showing consistent performance across the input range (n=8 for 10 ng; n=7 for 10 pg). (**Panel B**) Reads of the 10 pg MBR dataset downsampled to assess gene and transcript sensitivity. (**Panel C**) 10 ng MBR dataset downsampled to assess gene and transcript sensitivity. While neither of the two conditions (10 pg and 10 ng) were sequenced until saturation, both highlight the high gene and transcript sensitivity across read depths.

4 SMART-Seq mRNA LR is compatible with single-cell inputs

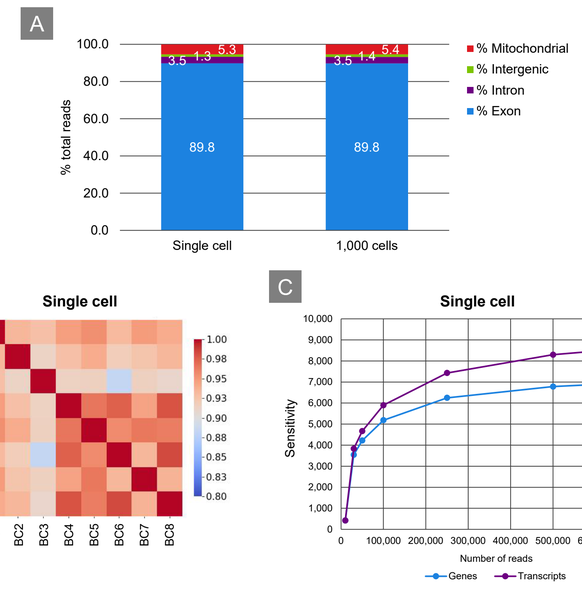


Figure 4. SMART-Seq mRNA LR is compatible with direct cell inputs. cDNA was generated with from single K562 cells or 1,000 K562 cells using the SMART-Seq mRNA LR workflow described in Figure 1. After sequencing, data was basecalled and demultiplexed using guppy, and reads per barcode were downsampled to 300,000 reads. (**Panel A**) The average gene count for each condition is shown above the corresponding bar (n=8 for single cell and n=2 for 1,000 cell). (**Panel B**) Pearson correlation was calculated from the gene matrix across 8 single K562 cells. (**Panel C**) Downsampling analysis of single K562 cells shows gene and transcript count per read depth.

5 SMART-Seq mRNA LR detects full-length isoforms and gene fusions

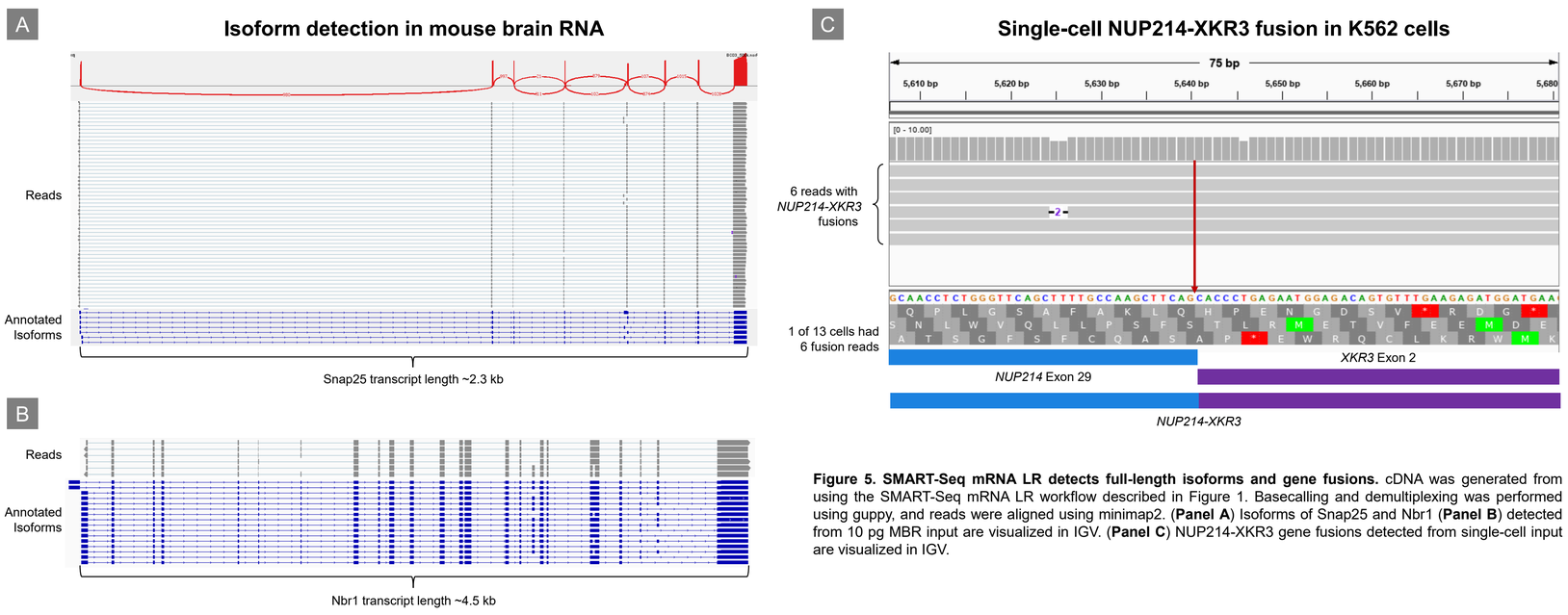


Figure 5. SMART-Seq mRNA LR detects full-length isoforms and gene fusions. cDNA was generated from using the SMART-Seq mRNA LR workflow described in Figure 1. Basecalling and demultiplexing was performed using guppy, and reads were aligned using minimap2. (**Panel A**) Isoforms of Snap25 and Nbr1 (**Panel B**) detected from 10 pg MBR input are visualized in IGV. (**Panel C**) NUP214-XKR3 gene fusions detected from single-cell input are visualized in IGV.

6 SMART-Seq mRNA LR generates even gene body coverage

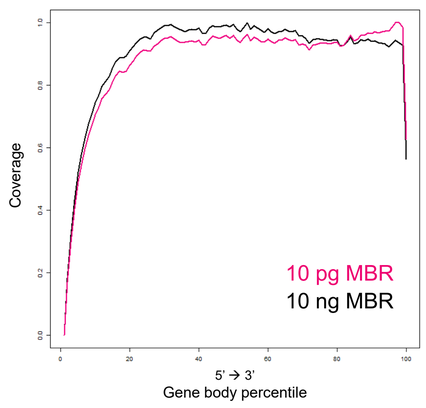


Figure 6. SMART-Seq mRNA LR yields even gene body coverage for ultra-low inputs. 10 pg and 10 ng MBR samples underwent the SMART-Seq mRNA LR workflow. Gene body coverage was assessed for the average of 8 replicates of 10 pg and 10 ng MBR samples.

7 SMART-Seq mRNA LR evinces high reproducibility

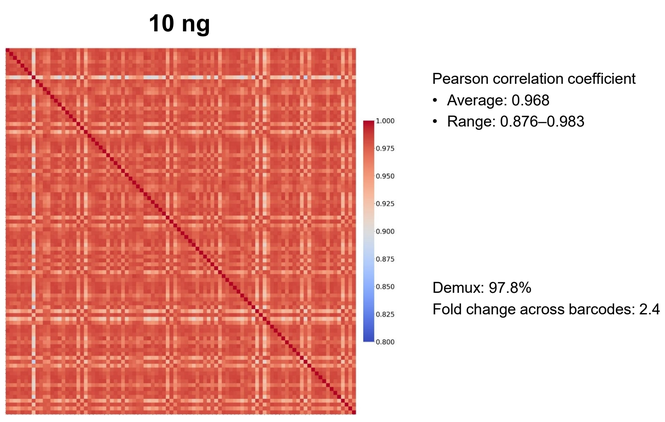
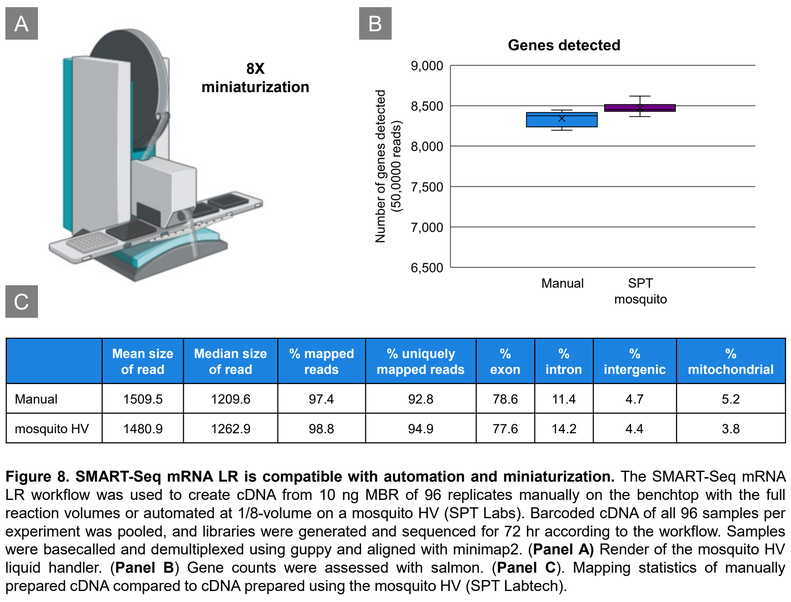


Figure 7. SMART-Seq mRNA LR is highly reproducible across replicates for a broad range of inputs. The SMART-Seq mRNA LR workflow was used to create cDNA from 10 ng MBR of 96 replicates. Barcoded cDNA of all 96 samples were pooled, and libraries were generated and sequenced according to the workflow.

8 SMART-Seq mRNA LR is compatible with automation and miniaturization



	Mean size of read	Median size of read	% mapped reads	% uniquely mapped reads	% exon	% intron	% intergenic	% mitochondrial
Manual	1509.5	1209.6	97.4	92.8	78.6	11.4	4.7	5.2
mosquito HV	1480.9	1262.9	98.8	94.9	77.6	14.2	4.4	3.8

Figure 8. SMART-Seq mRNA LR is compatible with automation and miniaturization. The SMART-Seq mRNA LR workflow was used to create cDNA from 10 ng MBR of 96 replicates manually on the benchtop with the full reaction volumes or automated at 1/8-volume on a mosquito HV (SPT Labs). Barcoded cDNA of all 96 samples per experiment was pooled, and libraries were generated and sequenced for 72 hr according to the workflow. Samples were basecalled and demultiplexed using guppy and aligned with minimap2. (**Panel A)** Render of the mosquito HV liquid handler. (**Panel B**) Gene counts were assessed with salmon. (**Panel C**). Mapping statistics of manually prepared cDNA compared to cDNA prepared using the mosquito HV (SPT Labtech).

Conclusions

- Our new SMART-Seq mRNA Long Read workflow generates high-quality barcoded cDNA from ultra-low inputs for sequencing on Oxford Nanopore Technologies (ONT) devices.
- High reproducibility is observed between samples.
- Although optimized for ultra-low inputs, SMART-Seq mRNA Long Read is also compatible with single-cell inputs.
- SMART-Seq mRNA Long Read is compatible with miniaturization and automation.
- SMART-Seq mRNA Long Read enables single-tube library preparation using ONT reagents.
- SMART-Seq mRNA Long Read is compatible with ONT analysis tools.

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. © 2024 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

For more information email
bryan_bell@takarabio.com

800.662.2566
www.takarabio.com