A complete, ultra-low input RNA-seq solution for full-length transcriptome analysis and RNA counting

Ilona Holcomb*, Tommy Duong, Junya Seo, Karthikeyan Swaminathan, Bryan Bell, Yue Yun, and Andrew Farmer

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

Abstract

OBJECTIVE: RNA sequencing (RNA-seq) is a powerful way to investigate transcriptional highs and lows, allelic origins, and isoform preferences in the transcriptome that can underlie key biological states. One current limitation of single-cell RNA-seq methodologies is either the absence of unique molecular identifiers (UMIs), or the inability to maintain the yield, sensitivity, and reproducibility when UMIs are employed.

METHODS: To test the yield, sensitivity, and reproducibility, we benchmarked a new SMART-Seq® method, SMART-Seq mRNA (with UMIs), against the existing SMART-Seq v4 Ultra® Low Input RNA Kit for Sequencing (SSv4), and the Smart-seq 2 homebrew method (SS2). We included our novel library prep method in the testing to determine if a complete, end-to-end solution improved the data outcome. In addition, we tested the performance of this new method on common automation platforms.

RESULTS: Gene count and read distribution across major RNA-seq output components were comparable between SMART-Seq mRNA (with UMIs) and SSv4. However, the new method showed significantly increased sensitivity compared to the SS2 homebrew method. In addition, we demonstrate that SMART-Seq mRNA (with UMIs) can enable RNA counting, and while optimized for low RNA input, is compatible with single-cell RNA-seq analysis. Finally, we show that SMART-Seq mRNA (with UMIs) is compatible with common automation platforms.

CONCLUSION: Our data demonstrate that SMART-Seq mRNA (with UMIs) leveraging SMART® technology with UMIs for cDNA generation and our unique library preparation protocol, combined with our Cogent™ NGS analysis software (CogentAP), is a complete, robust, and sensitive solution for full-length transcriptome studies. The inclusion of UMIs allowed for RNA counting without compromising data quality, and lead to superior sensitivity compared to homebrew SS2 chemistries.

1 SMART-Seq mRNA (with UMIs) sequencing workflow

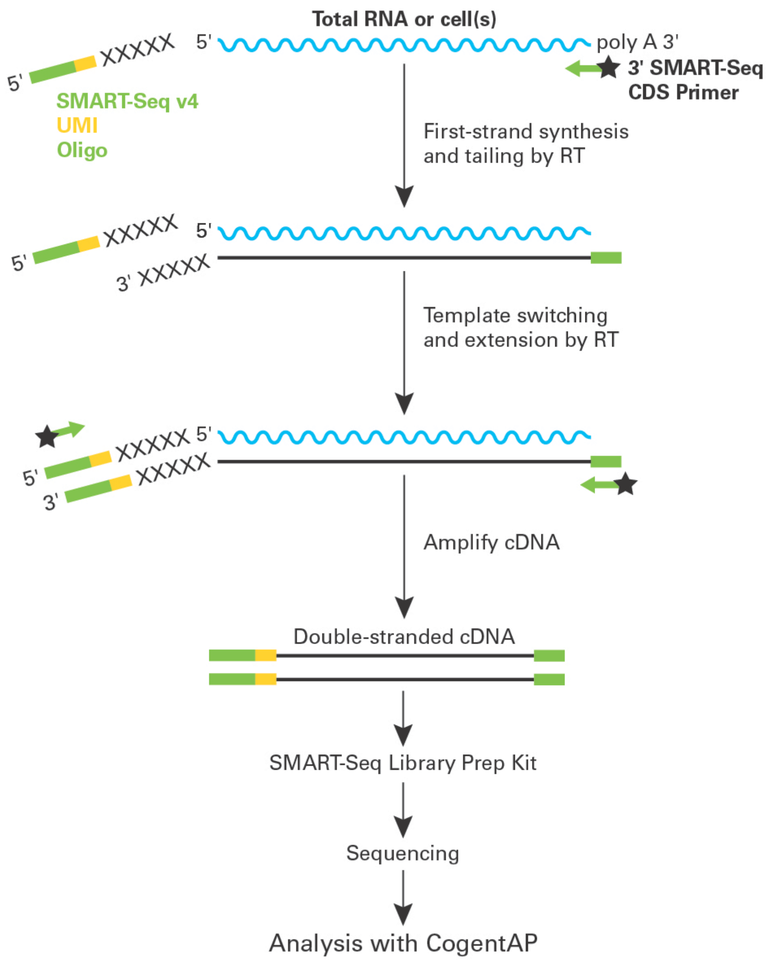


Figure 1. Library preparation and sequencing workflow for the SMART-Seq mRNA (with UMIs) kit. First-strand cDNA synthesis is primed by the SMART-Seq CDS Primer and performed by an MMLV-derived reverse transcriptase (RT). Upon reaching the 5' end of each mRNA molecule, the RT adds non-templated nucleotides to the first-strand cDNA. The SMART-Seq v4 Oligonucleotide contains a sequence that is complementary to the non-templated nucleotides added by the RT and the UMI sequence. This primer hybridizes to the first-strand cDNA. In the template-switching step, the RT uses the remainder of the SMART-Seq v4 Oligonucleotide as a template for the incorporation of an additional sequence on the end of the first-strand cDNA. Then, the first-strand cDNA is amplified by PCR. In the second part of the workflow, the SMART-Seq Library Prep Kit is used to generated sequencing-ready libraries. After sequencing, data is analyzed using CogentAP software.

2 SMART-Seq mRNA (with UMIs) performance: reproducibility

A

10 pg	Rep1	Rep2	Rep3
Rep1	1	0.9575	0.9574
Rep2	0.9575	1	0.9568
Rep3	0.9574	0.9568	1

B

	100 pg			1 ng			10 ng		
	Rep1	Rep2	Rep3	Rep1	Rep2	Rep3	Rep1	Rep2	Rep3
100 pg Rep1	1	0.989	0.985	0.989	0.989	0.961	0.969	0.966	0.962
100 pg Rep2	0.989	1	0.988	0.985	0.986	0.961	0.961	0.959	0.958
100 pg Rep3	0.985	0.988	1	0.983	0.984	0.974	0.96	0.958	0.962
1 ng Rep1	0.989	0.985	0.983	1	0.994	0.971	0.984	0.982	0.978
1 ng Rep2	0.989	0.986	0.984	0.994	1	0.973	0.984	0.982	0.98
1 ng Rep3	0.961	0.961	0.974	0.971	0.973	1	0.961	0.959	0.97
10 ng Rep1	0.969	0.961	0.96	0.984	0.984	0.961	1	0.994	0.991
10 ng Rep2	0.966	0.959	0.958	0.982	0.982	0.959	0.994	1	0.991
10 ng Rep3	0.962	0.958	0.962	0.978	0.98	0.97	0.991	0.991	1

0.5 — 1.0

Figure 2. SMART-Seq mRNA (with UMIs) is reproducible across replicates for a broad range of input amounts. The SSv4+UMI workflow was used to create sequencing libraries from 10 pg of K562 RNA (**Panel A**) or three different amounts (100 pg, 1 ng and 10 ng), of universal human reference (UHR) RNA (**Panel B**), in triplicate. Then, the cDNA libraries were sequenced on an Illumina® platform and the RPKM was calculated using CogentAP for each experimental condition. To determine method reproducibility, the R^2 was calculated between all possible replicate pairs.

3 SMART-Seq mRNA (with UMIs) is as sensitive as SSv4

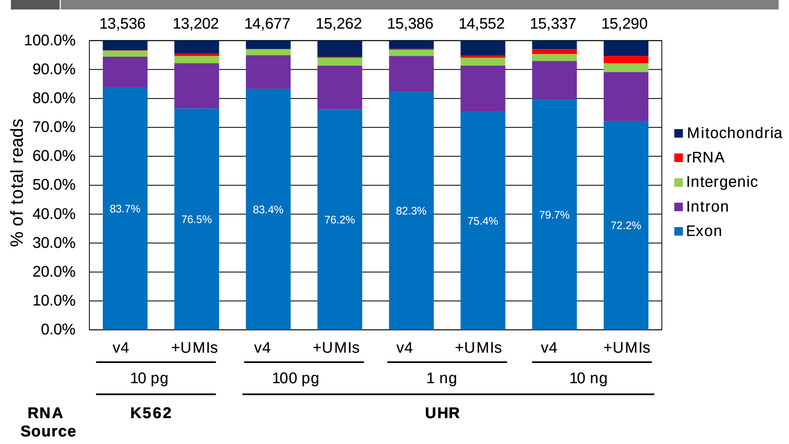
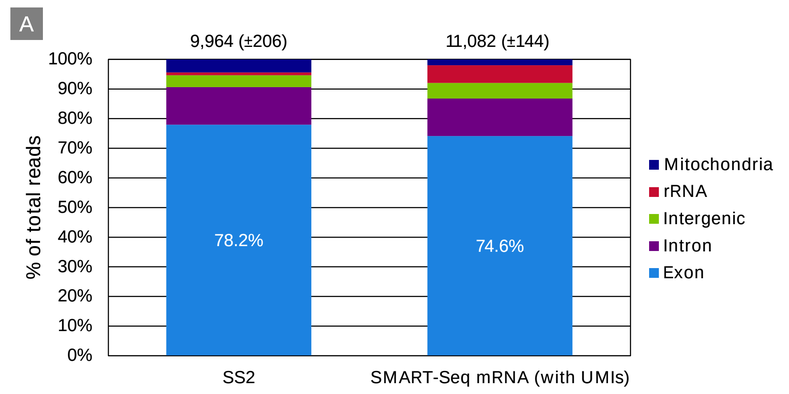


Figure 3. SMART-Seq mRNA (with UMIs) demonstrates sensitivity comparable to SSv4. Gene count and read distribution across major RNA-seq output components for SSv4 (v4) and the SMART-Seq mRNA (with UMIs) (+UMIs in figure) workflow. To evaluate the performance of the SMART-Seq mRNA (with UMIs) workflow compared to the SSv4 workflow, each method used to create sequencing libraries from K562 RNA (10 pg) or three different amounts of universal human reference (UHR) RNA (100 pg, 1 ng, and 10 ng). The resulting cDNA libraries were sequenced on an Illumina platform. After sequencing, data for the 10 pg input samples was downsampled to 1.8 million reads, while the data for all other input amounts was downsampled to 1.0 million reads. The average gene count for each condition is shown above the corresponding bar. n=3 for all conditions.

4 SMART-Seq mRNA (with UMIs) outperforms SS2 in terms of sensitivity and reproducibility

A



B

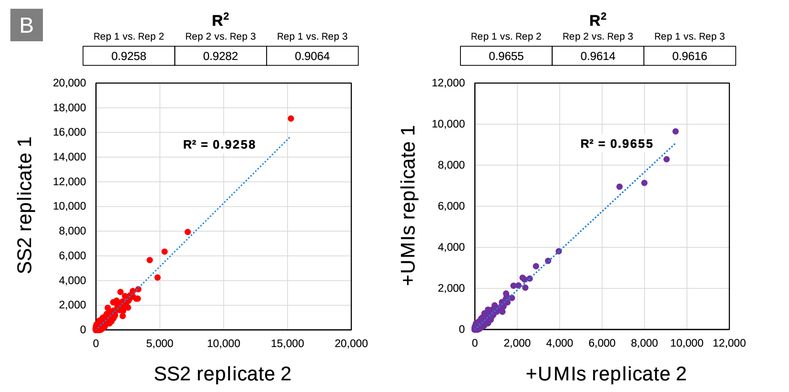


Figure 4. SMART-Seq mRNA (with UMIs) outperforms SS2 in terms of sensitivity and reproducibility. Panel A. Gene count and read distribution across major RNA-seq output components for the SMART-Seq mRNA (with UMIs) workflow and SS2. cDNA was generated from 10 pg UHR RNA using the SMART-Seq mRNA (with UMIs) method or the SS2 method. Then, the cDNA generated with the SMART-Seq mRNA (with UMIs) workflow was prepared for sequencing using the SMART-Seq Library Preparation Kit and The SS2 cDNA was prepared for sequencing using the Nextera® XT DNA Library Prep Kit (Illumina). Libraries from both were pooled, sequenced, and downsampled to 1.2 million reads. The average gene count for each condition is shown above the corresponding bar. n=3 for all conditions. **Panel B.** Absolute read counts for each gene matrix between each replicate pair for SS2 and SMART-Seq mRNA (with UMIs) (+UMIs in figure) were plotted, and regression analysis was performed. The R^2 values were consistently higher for the SMART-Seq mRNA (with UMIs) workflow using the SMART-Seq Library Preparation Kit. An example XY plot for both SS2 and SMART-Seq mRNA (with UMIs) (+UMIs in figure) are shown. All data analysis was performed with CogentAP.

5 SMART-Seq mRNA (with UMIs) allows for RNA counting

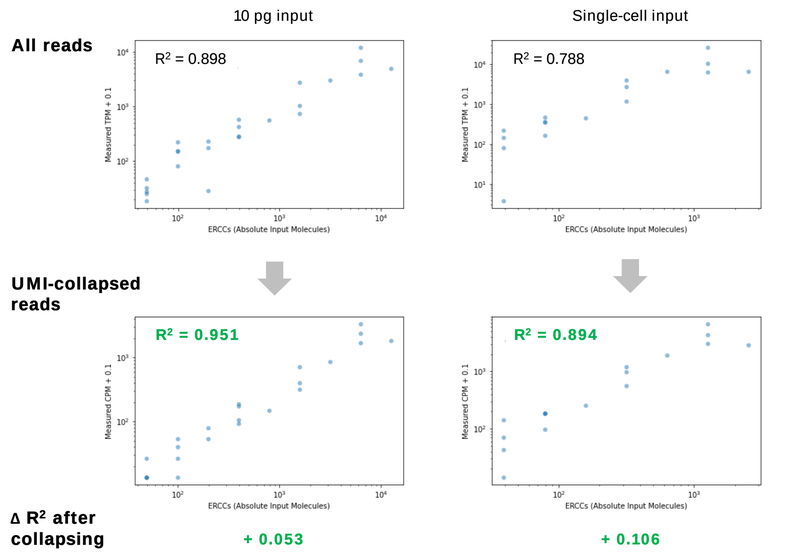
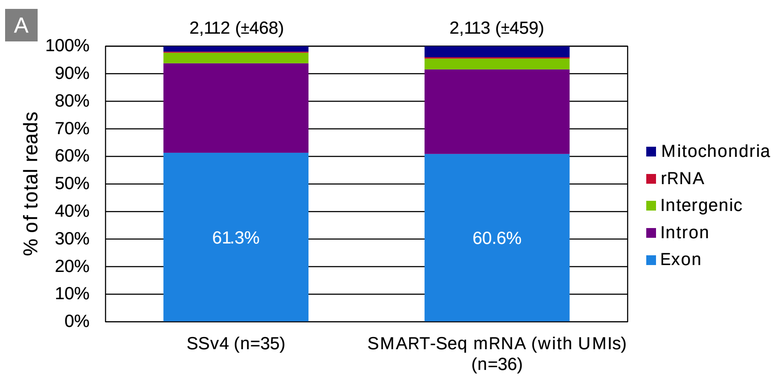


Figure 5. SMART-Seq mRNA (with UMIs) reduce PCR duplications and allows for RNA counting. 10 pg of input RNA or RNA from a single cell was spiked with ERCC control RNA, and libraries were prepared using the SMART-Seq mRNA (with UMIs) method. Libraries were sequenced and then analyzed using CogentAP. The measured TPM or CPM was plotted against the absolute number of input molecules from the ERCC spike-in and regression analysis was performed. There is a significant increase in the R^2 with the collapse of UMIs, indicating addition of UMIs to the SSv4 workflow leads to a reduction of PCR duplications and enables more accurate RNA counting.

6 SMART-Seq mRNA (with UMIs) is compatible with single-cell inputs

A



B

Cell #

Cell type	Gene name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Leukocytes	PTPRC (CD45)	1139.9	882.5	1036.7	1330.6	1293.5	1083.0	309.8	1005.8	253.2	153.4	18.2	1031.4	336.5	574.2	1234.5	1019.1	1.6	855.3
T cells (general)	CD3E	873.9	0.0	508.1	0.0	179.0	0.0	67.1	868.4	0.0	94.5	0.0	0.0	1614.7	105.6	426.6	246.7	18.9	181.7
T cells (helper)	CD4	0.0	0.0	0.0	0.0	0.0	171.8	60.3	372.2	0.0	0.0	0.0	367.9	0.0	0.0	0.0	0.0	0.0	0.0
Cytotoxic T cells	CD8A	253.3	0.0	0.0	974.2	848.9	0.0	28.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30.3	111.9	0.0	0.0
B cells	CD19	0.0	660.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Natural killer cells	NCAM1 (CD56)	0.0	0.0	0.0	0.0	0.0	0.0	19.5	0.0	0.0	0.0	0.0	0.0	0.0	32.9	8.7	0.0	3.2	0.0
Dendritic cells	ITGAX (CD11c)	4.3	0.0	0.0	0.0	0.0	660.5	24.3	0.0	19.7	0.0	0.0	114.9	0.0	25.8	1.6	0.0	0.0	0.0
Monocytes/macrophages	CD14	0.0	0.0	0.0	0.0	0.0	13.5	177.0	0.0	0.0	0.0	0.0	0.0	0.0	119.5	0.0	0.0	0.0	0.0
Monocytes/macrophages	CD33	0.0	0.0	0.0	0.0	0.0	0.0	9.6	0.0	0.0	0.0	0.0	0.0	0.0	52.1	0.0	0.0	0.0	0.0

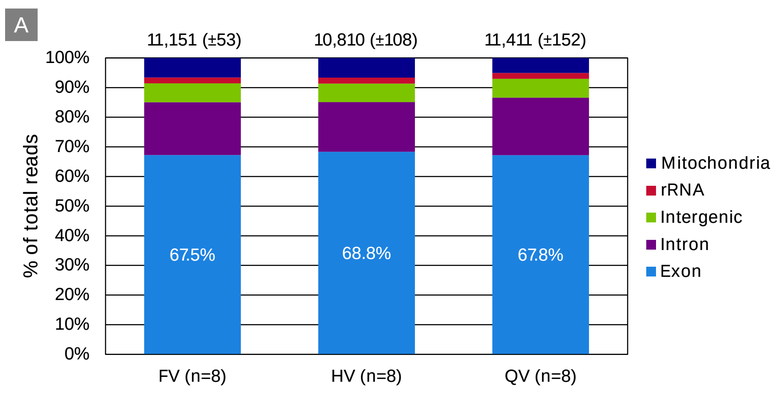
Cell type	Gene name	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Leukocytes	PTPRC (CD45)	958.8	2471.8	1749.3	988.1	764.3	2339.1	1057.3	1304.5	2376.9	796.1	1684.8	274.2	903.0	1380.3	218.0	1233.1	704.1	0.0
T cells (general)	CD3E	500.0	883.6	305.7	623.2	541.6	1662.7	69.5	367.3	93.1	331.4	0.0	203.4	0.0	521.4	0.0	0.0	201.4	10.5
T cells (helper)	CD4	0.0	0.0	60.4	0.0	0.0	0.0	5.4	0.0	0.0	0.0	0.0	166.9	0.0	0.0	0.0	0.0	0.0	0.0
Cytotoxic T cells	CD8A	0.0	0.0	32.6	690.9	195.6	1090.9	0.0	130.1	0.0	0.0	0.0	636.7	0.0	0.0	0.0	0.0	610.1	0.0
B cells	CD19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	281.7	0.0	0.0	0.0
Natural killer cells	NCAM1 (CD56)	0.0	0.0	0.0	5.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.2	0.0	0.0
Dendritic cells	ITGAX (CD11c)	0.0	6.2	7.0	10.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	470.4	0.0	239.9	0.0	0.0	0.0
Monocytes/macrophages	CD14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3788.5	0.0	0.0	0.0	0.0	0.0
Monocytes/macrophages	CD33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	21.6	0.0	0.0	0.0	0.0	0.0

0 — Highest TPM

Figure 6. SMART-Seq mRNA (with UMIs) is as sensitive as SSv4 at single-cell input levels. Panel A. Single peripheral blood mononuclear cells (PBMCs) were sorted into plates and sequencing libraries were produced using either SSv4 or the SMART-Seq mRNA (with UMIs) workflow. Libraries were then pooled, sequenced, and downsampled to 900,000 reads for analysis. The gene count and read distribution for SSv4 and SMART-Seq mRNA (with UMIs) were comparable. **Panel B.** Gene expression profiles of the 36 individual PMBCs were produced using the SMART-Seq mRNA (with UMIs) workflow**.** Transcripts per kilobase million (TPM**)** for genes corresponding to cell surface markers commonly used to do cell-type identification for PBMCs by FACs are shown. Gene expression profiles identified 70% (25/36) of PMBCs as T cells, which is the expected ratio for PBMCs from a normal, healthy donor. In addition, gene expression profiles identified one B cell, one NK cell, one monocyte, and two dendritic cells. Six cells could not be clearly classified.

7 SMART-Seq mRNA (with UMIs) is compatible with automation

A



B

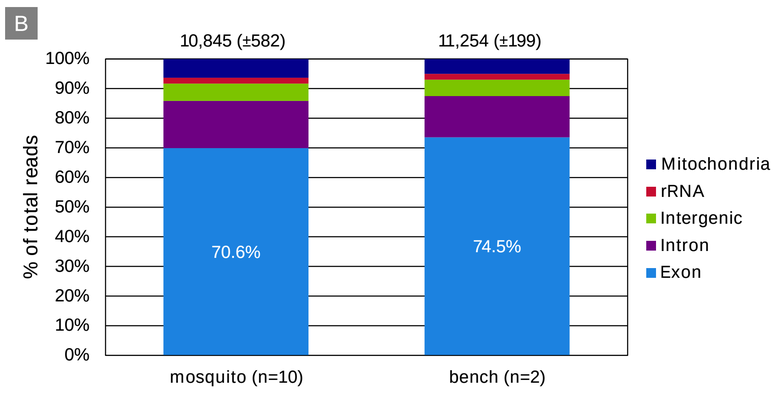


Figure 7. SMART-Seq mRNA (with UMIs) is compatible automation platforms. Panel A. Read count and distribution across major RNA-seq output components using the SMART-Seq mRNA (with UMIs) workflow on the Mantis Liquid Handler (Formulatrix). 10 pg of mouse brain control RNA was used as input for the SMART-Seq mRNA (with UMIs) workflow at full (FV), half (HV), and quarter (QV) volume reactions on the Mantis Liquid Handler. Libraries were pooled, sequenced, and downsampled to 1.3 million reads for analysis. Analysis was performed using CogentAP. The average gene count for each condition is shown above the corresponding bar. There is a high concordance in read distributions and gene counts across all volumes. **Panel B.** Gene count and read distribution across major RNA-seq output components using the SMART-Seq mRNA (with UMIs) workflow on the mosquito HV liquid handler (SPT Labs) or on the benchtop without the aid of automation. 10 pg of mouse control RNA was used as input for the SMART-Seq mRNA (with UMIs) workflow at 1/8 volume using the mosquito HV or manually on the benchtop. All libraries were prepared with the SMART-Seq Library Preparation Kit at full volume on the bench. The libraries were then pooled, sequenced, and downsampled to 620,000 reads for analysis. Analysis was performed using CogentAP. The average gene count for each condition is shown above the corresponding bar. There is a good concordance in read distributions and gene counts between the mosquito and bench preparations.

Conclusions

- Our new SMART-Seq mRNA (with UMIs) kit, leveraging SMART technology with UMIs for cDNA generation and our unique library preparation protocol, combined with our Cogent NGS analysis software (CogentAP), is a complete, robust, and sensitive solution for full-length transcriptome studies
- The inclusion of UMIs allows for RNA counting without compromising data quality
- While SMART-Seq mRNA (with UMIs) is optimized for ultra-low RNA inputs, it is compatible with single-cell inputs
- SMART-Seq mRNA (with UMIs) is compatible with automation platforms

Scan to download your copy of this poster and view additional posters, or visit
www.takarabio.com/abrf-2022

800.662.2566
www.takarabio.com