Efficient Production of On-target Reads for Small RNA Sequencing of Single Cells Using Modified Adapters

Ruba Khnouf,[1,2†] Sabrina Shore,[3†] Crystal M. Han,[4,5] Jordana M. Henderson,[3] Sarah A. Munro,[4,6] Anton P. McCaffrey,[3] Hirofumi Shintaku,[7*] and Juan G. Santiago[1*]

[1] Department of Mechanical Engineering, Stanford University, Stanford, CA, United States

[2] Department of Biomedical Engineering, Jordan University of Science and Technology, Irbid, Jordan; [3]TriLink Biotechnologies LLC, San Diego, CA, United States; [4]Joint Initiative for Metrology in Biology, National Institute of Standards and Technology, Stanford, CA, [5]Department of Mechanical Engineering, San Jose State University, San Jose, CA, United States; [6]Minnesota Supercomputing Institute, University of Minnesota, MN, United States; [7]RIKEN Cluster for Pioneering Research, Japan

[†] Authors contributed equally to this work

[*]To whom correspondence should be addressed: hirofumi.shintaku@riken.jp; juan.santiago@stanford.edu

This document contains the following supplementary figures and information further describing our single-cell small RNA extraction methods, library prep, and sequencing results.

- Figure S-7. Reproducibility, as evaluated by the percentage of sRNAs detected in pairs of replicate samples out of the mean total number of sRNAs detected in this pair of samples.
- Figure S-8. Heatmap summarizing correlations among the sequenced samples and protocol types.
- Figure S-9. Violin plot showing total measured small RNA length distributions.

Table S-1. Modified conditions of individual samples in CleanTag small RNA library preparation.

	100 ng total RNA	**10 pg total RNA**	**Triton X-100**	**Microfluidics**
# Libraries Made	3	3	3	12
Adapter Dilution	1:2	1:16	1:16	1:16
RNA template (μL)	1.0*	1.0*	4.0	8.0
# PCR Cycles	18	27	27	27

*added 1 μL of nuclease free water to bring the 3' adaptor reaction volume to 10 μL

Table S-2. Summary of sequencing reads for the various sample types.

	Raw reads	Down-sampled	Average input reads	% too short	% unmapped	% rRNA	% unique	% multiple
Triton-1	20,198,484	10,501,685	9,577,403	19%	7%	59%	11%	4%
Triton-2	15,254,603	7,934,019		36%	15%	33%	12%	4%
Triton-3	19,803,727	10,296,504		28%	7%	43%	17%	5%
Microfluidics-1	11,770,967	11,770,967	9,577,831	78%	14%	5%	2%	1%
Microfluidics-2	11,003,997	11,003,997		49%	22%	17%	9%	3%
Microfluidics-3	11,911,275	11,911,275		45%	17%	33%	3%	1%
Microfluidics-4	9,363,265	9,363,265		53%	40%	6%	1%	0%
Microfluidics-5	8,145,223	8,145,223		66%	30%	4%	0%	0%
Microfluidics-6	9,458,328	9,458,328		58%	14%	22%	3%	3%
Microfluidics-7	10,192,448	10,192,448		51%	29%	11%	7%	3%
Microfluidics-8	8,325,487	8,325,487		61%	34%	4%	0%	0%
Microfluidics-9	9,411,707	9,411,707		46%	38%	11%	3%	2%
Microfluidics-10	7,640,640	7,640,640		62%	29%	9%	1%	0%
Microfluidics-11	9,944,730	9,944,730		73%	15%	10%	1%	1%
Microfluidics-12	7,765,900	7,765,900		51%	34%	10%	4%	2%
10pg-1	16,545,983	9,245,028	9,577,470	35%	27%	22%	13%	3%
10pg-2	17,120,837	9,565,861		28%	25%	26%	18%	3%
10pg-3	17,758,223	9,921,522		27%	29%	25%	16%	3%
100ng-1	16,727,988	10,229,697	9,578,133	29%	7%	32%	23%	10%
100ng-2	15,520,502	9,491,451		18%	5%	25%	43%	9%
100ng-3	14,737,788	9,013,252		30%	6%	24%	32%	10%

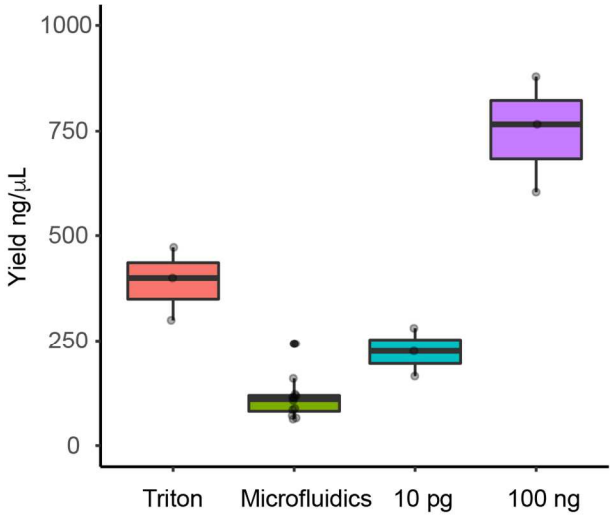

Figure S-1. Quantification of purified library yields observed for the various sample types. The mean abundance of the three are within a factor of about 5. Note the single cell ITP extracts only cytoplasmic fractions of single cells.

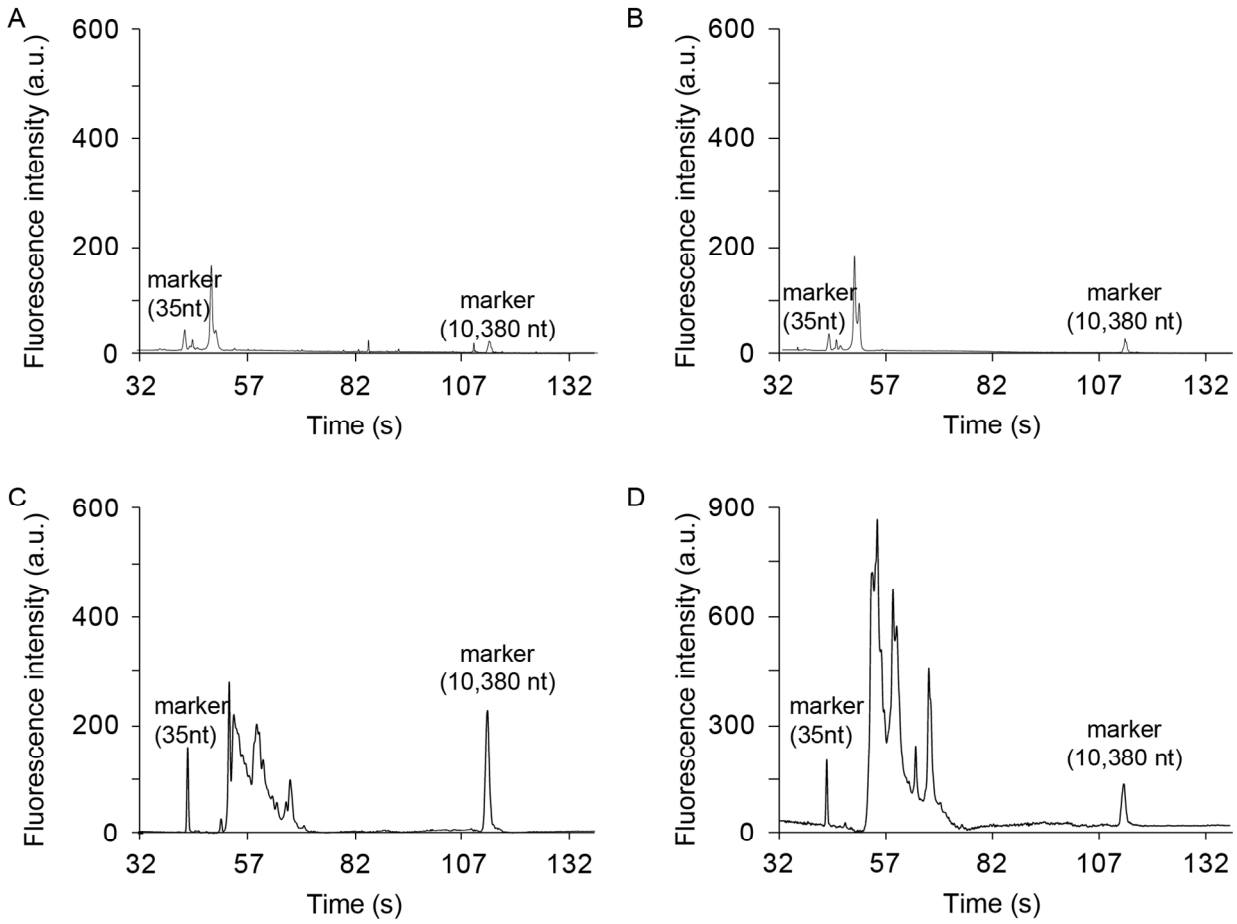

Figure S-2. Capillary gel electropherograms using the Agilent Bioanalyzer for products of the CleanTag library preparation. (A) Example electropherogram for library preparation with off-the-shelf lysis kit (Ambion's Single Cell Lysis Kit). Here the molecules associated with ligated sRNA is not present. (B) Example electropherogram for library preparation with single K562 cell lysed using Triton X-100 without overnight-freezing. Again, the library peak associated with ligated sRNA is not apparent. (C) Example electropherograms of 10 pg of bulk total RNA and (D) 100 ng of bulk total RNA.

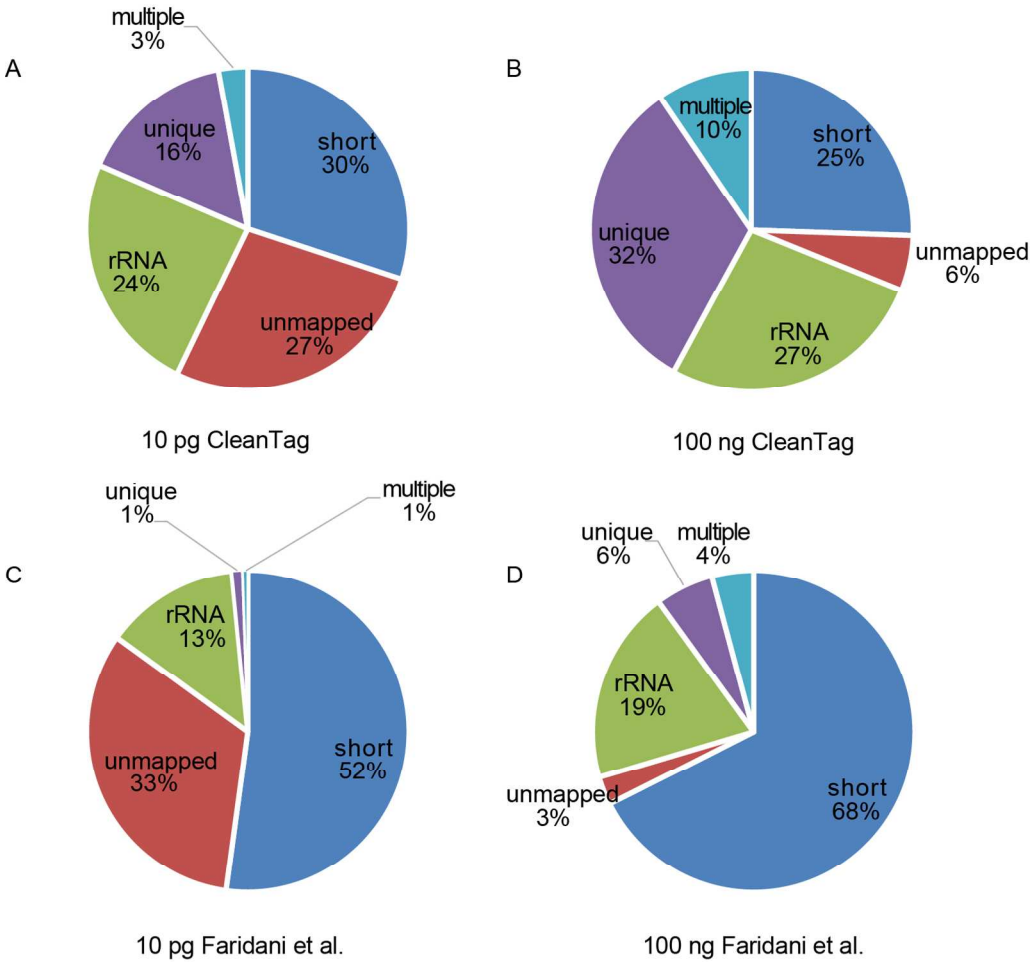

Figure S-3. Proportion of on-target reads (unique and multiple) and off-target reads (short, unmapped, and rRNA) from sRNA-seq library prepared with (A)10 pg and (B)100 ng total RNA of K562 cells. For reference, we compared our results to Faridani et al.[18] in which total RNA of HEK293FT cells was used (C and D).

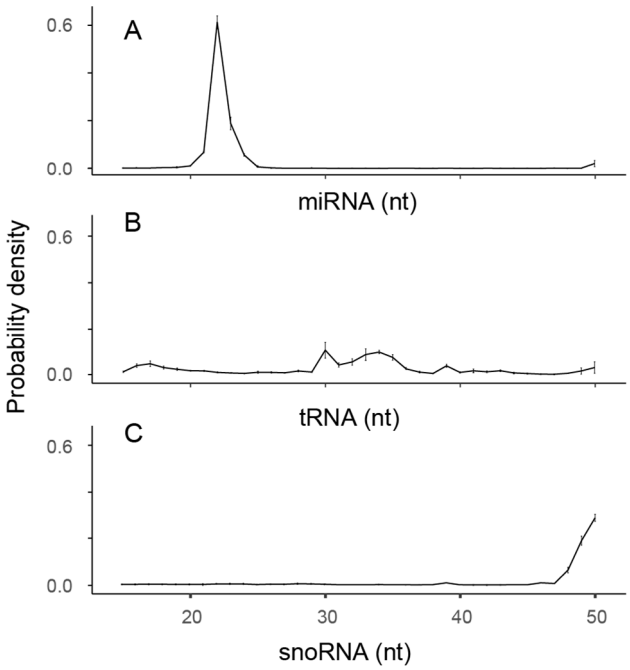

Figure S-4 Length distribution of aligned reads on (A) miRNA, (B) tRNA, and (C) snoRNA with 100 ng bulk sample

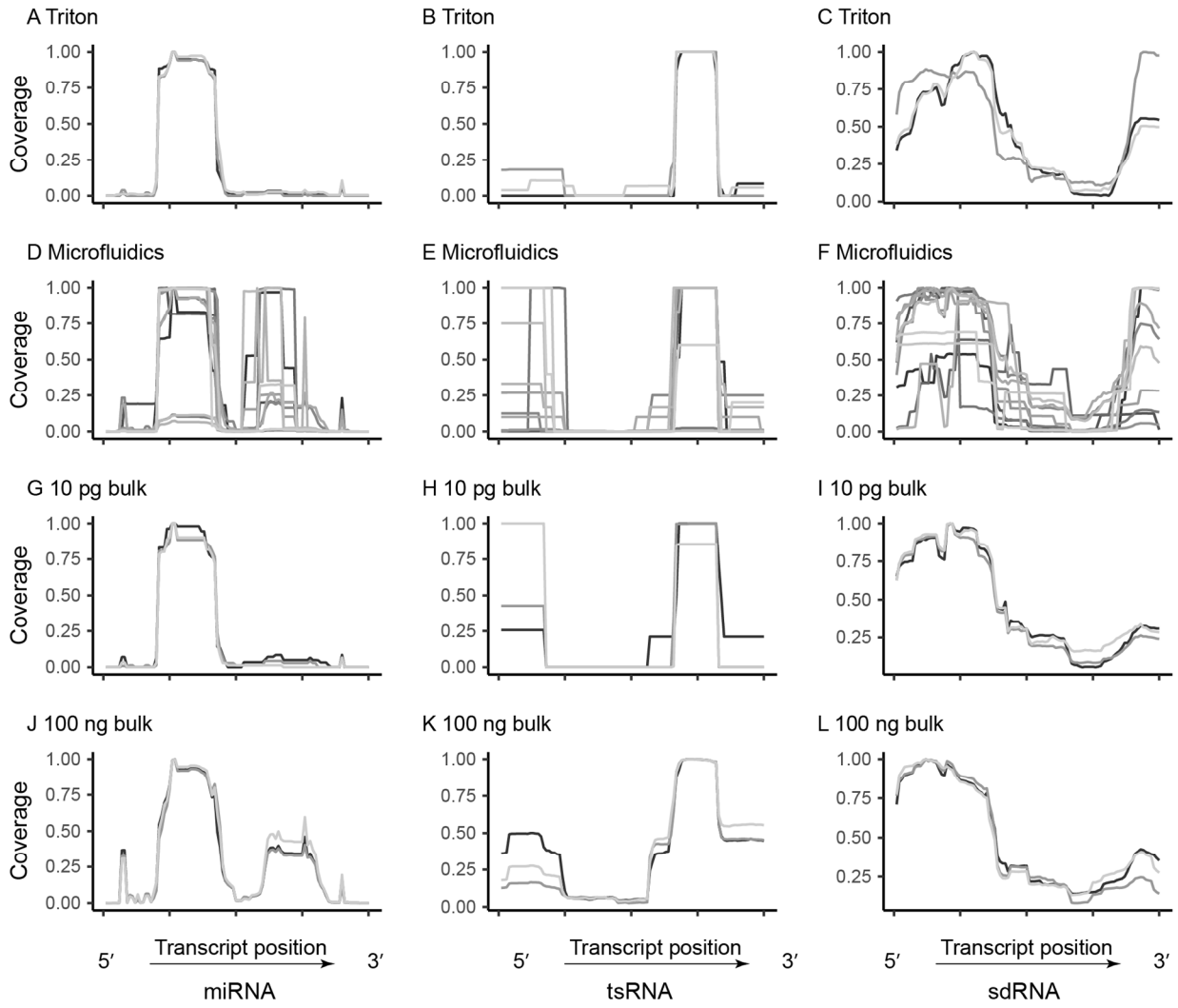

Figure S-5 Distribution of aligned reads on respective precursors. (A-C)Triton-based lysis, (D-F) Microfluidics-based lysis, (G-I) 10 pg of bulk sample and (J-L) 100 ng of bulk sample.

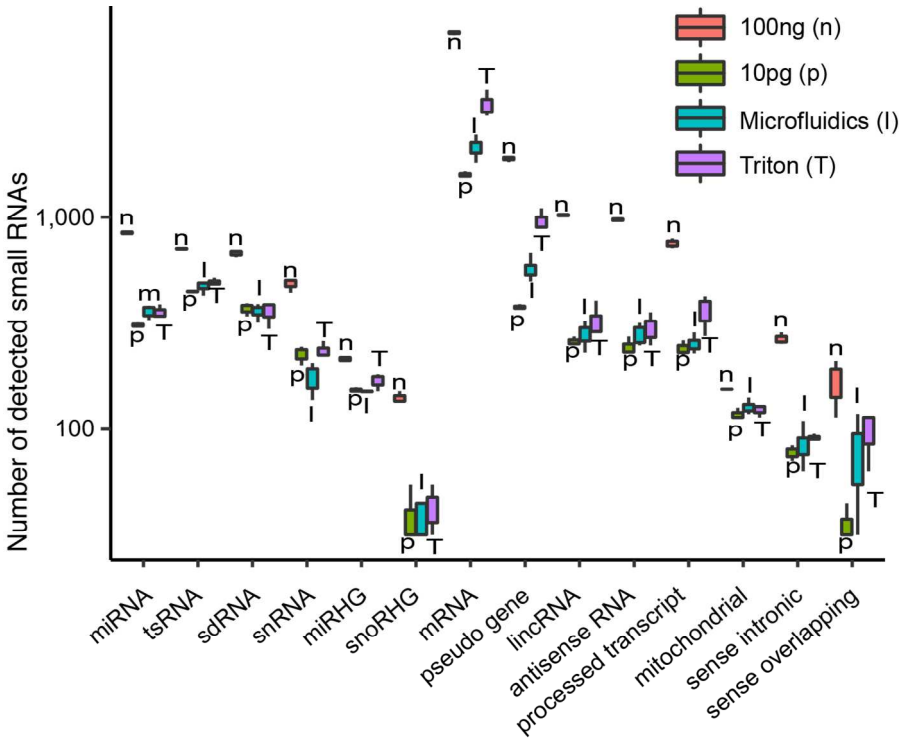

Figure S-6. Measured abundance of sRNAs (<40 nt).

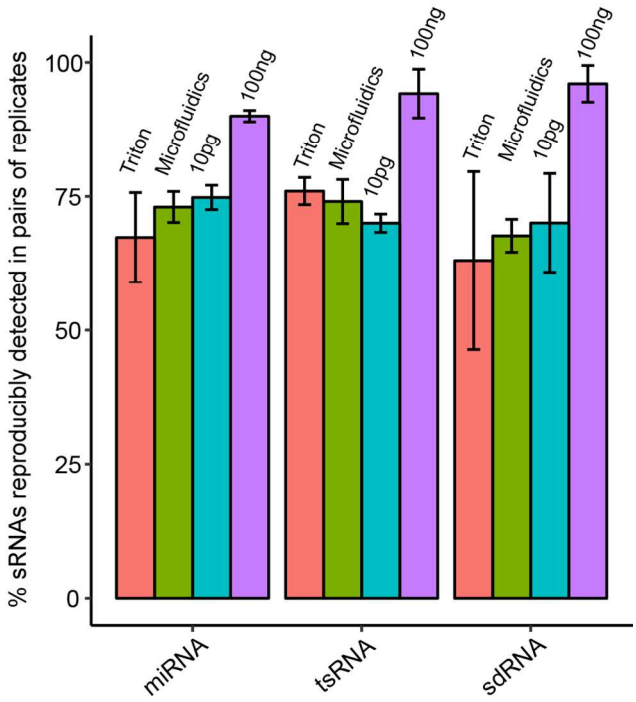

Figure S-7 Reproducibility, as evaluated by the percentage of sRNAs detected in pairs of replicate samples out of the mean total number of sRNAs detected in this pair of samples.

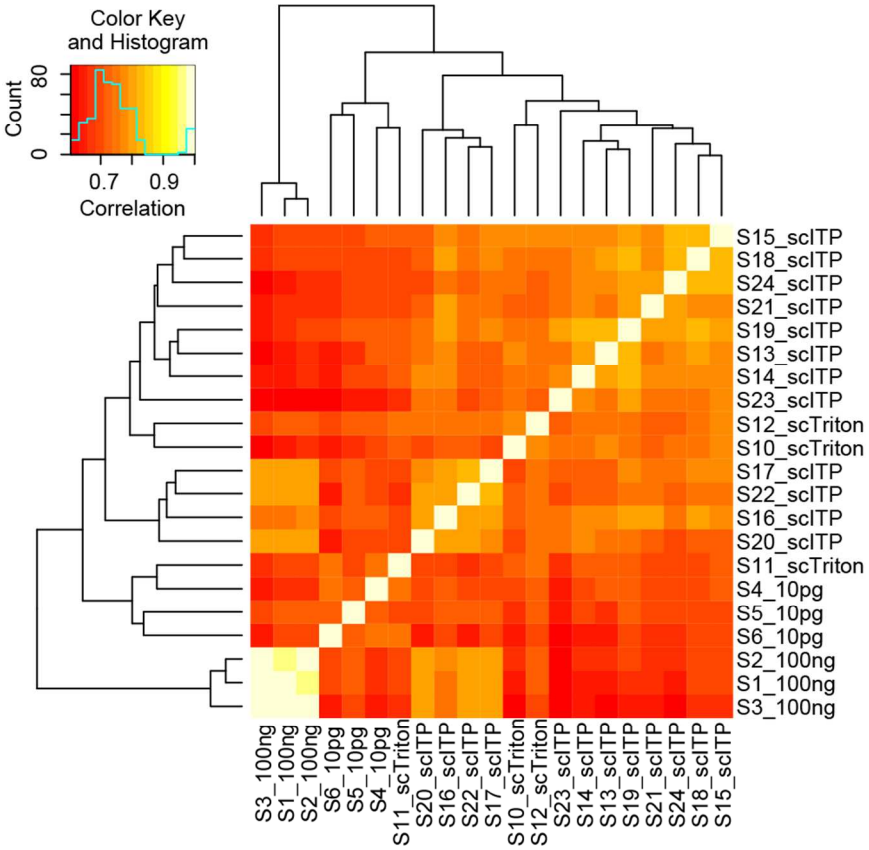

Figure S-8. Heatmap summarizing correlations among the sequenced samples and protocol types. The heatmap shows a higher correlation between the 10 pg commercial K562 RNA samples with the single cell samples than between the 100 ng samples and the single cell samples. We denote here Triton-based lysis and microfluidic-based lysis as scTriton and scITP, respectively, in this figure.

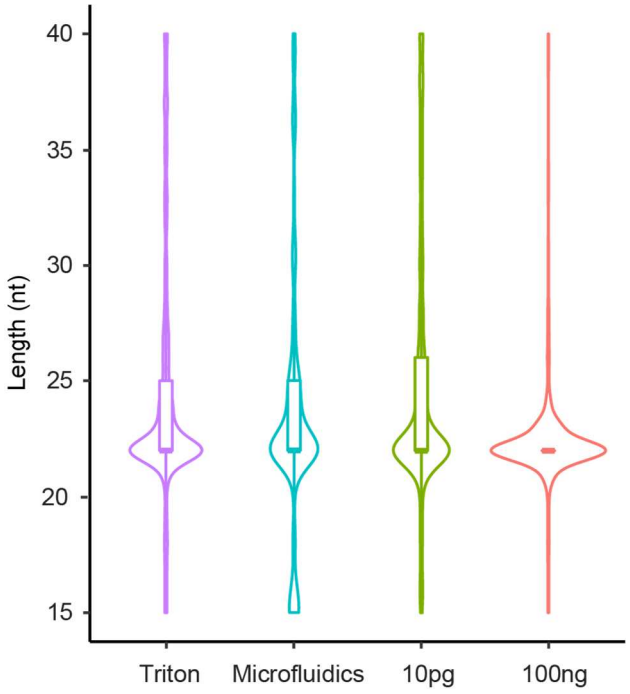

Figure S-9. Violin plot showing total measured small RNA length distributions. All four sample types and analyses show strong representation of small RNA sequences.