

Supplemental Note and Supplemental Figures

Quartz-Seq2: a high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads

Yohei Sasagawa^{1†}, Hiroki Danno^{1†}, Hitomi Takada^{2†}, Masashi Ebisawa¹, Kaori Tanaka¹, Tetsutaro Hayashi¹, Akira Kurisaki^{2*}, and Itoshi Nikaido^{1,3*}

[†]These authors contributed equally to this work.

*Corresponding author

¹Bioinformatics Research Unit, Advanced Center for Computing and Communication, RIKEN, Hirosawa 2-1, Wako, Saitama, Japan

²Laboratory of Stem Cell Technology, Graduate School of Biological Sciences, Nara Institute of Science and Technology, Takayama-cho 8916-5, Ikoma, Nara, Japan

³Single-cell Omics Research Unit, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Japan

Supplemental Note

RT primer and byproduct synthesis

We reported in our previous Quartz-Seq paper that the poly-A tagging strategy has great potential for increasing the number of genes detected [1]. However, the efficiency of poly-A tagging itself for single-cell RNA-seq has not been improved. To convert more cDNA molecules to amplifiable cDNA, we attempted to improve the poly-A tagging efficiency (Additional file 1: Fig. S1). However, we should first focus on suppressing the synthesis of byproducts because the byproducts from whole-transcript amplification cause severe problems that prevent quantitative performance via a low mapping ratio [1].

It has been reported that byproducts could be synthesized from reverse-transcription (RT) primers in poly-A tagging-based methods ([1-3], Additional file 1: Fig. S2a). Quartz-Seq required RT primer digestion by exonuclease I treatment to suppress byproduct synthesis. We thus investigated whether column purification could be used instead of exonuclease I treatment. The cDNA purified using columns was used for whole-transcript amplification based on conditions similar to those in Quartz-Seq. We found that cDNA purified with original Quartz-Seq RT primer (70-mer) was not associated with the synthesis of byproducts upon column purification for the pooling of cell-barcoded cDNA. On the other hand, we observed an enormous amount of byproduct synthesis when using long RT primer (126-mer). The length of DNA thus affects the tendency for byproduct synthesis [1]. Therefore, we designed a relatively short RT primer (v3.1: 73-mer), which has a cell-barcode sequence and a UMI sequence. This is designed based on the RT primer conjugated to barcoded beads of Drop-seq. This is because Drop-seq RT primer is relatively short, at the same length as the original Quartz-seq primer [1,4]. By using the v3.1 RT primer, we succeeded in amplifying cDNA without byproduct synthesis (Additional file 1: Fig. S2b). Thereafter, we used v3.1 RT primers for the technical development of Quartz-Seq2.

Streamline for pooling of cell-barcoded cDNA

After cell barcoding in reverse transcription, labeled cDNA is pooled and purified for subsequent whole-transcript amplification. A low volume of reverse-transcription solution for cell barcoding is efficient for cost reduction [4-6]. On the other hand, the pooling of labeled cDNA at a low volume from a PCR plate is complicated and time-consuming using pipette tips. Ideally, cell-barcoded cDNA should be pooled and purified with minimal loss for the subsequent whole-transcript amplification step in an expeditious manner. Therefore, we validated and streamlined the pooling step of cell-barcoded cDNA at a low volume. We collected 2 μ L of solution from every well of a 384-well PCR plate and estimated the collection efficiency. By using an eight-channel pipette, we collected $79.6 \pm 2.4\%$ ($n=3$) of solution in approximately 10 min from every well of a 384-well PCR plate. We also developed a spin-down collection system for the PCR plate (Figure 1 and Additional file 1: Fig. S4). Using this collection system (type A), we collected $88.7 \pm 0.99\%$ ($n=3$) of solution in 5 min from two to four plates. The collection efficiency of cDNA solution was improved by approximately 10% by constructing this spin-down collection system. In addition, the collection time per 384-well plate using this collection system was four- to eightfold faster than with a multi-channel pipette. We also validated the purification efficiency of cell-barcoded cDNA. We determined the qPCR score with eight genes using purified first-strand cDNA and unpurified first-strand cDNA. The purification efficiency was 93.77% ($n=4$) (Additional file 1: Fig S3a). By combining our approach with the spin-down collection system, we estimated that 80%–82% of cell-barcoded cDNA could theoretically be used for subsequent whole-transcript amplification in our system.

To import more cDNA molecules into subsequent whole-transcript amplification, we attempted to improve the efficiency of conversion from target RNA to first-strand cDNA in reverse transcription. First, we performed reverse transcription under conditions with various buffers and temperatures (Additional file 1: Fig. S3b). The Spearman's rank correlation coefficient (SCC) between the Quartz-Seq-like condition (PCR buffer, 45°C, 20 min) and positive control was 0.6957. In contrast, the SCC between the modified condition (T100, 50°C, 50 min) and the positive control was 0.8996. Moreover, reverse-transcription efficiency was slightly improved by approximately 20% on average (Additional file 1: Fig. S3b). These results show that

the modified RT condition reduced the bias of sequence preference in reverse transcription. In addition, we could import more cDNA molecules into subsequent whole-transcript amplification.

Relationship between sequence platform and Quartz-Seq2

We also focused on the relationship between sequencing platform and Quartz-Seq2. We did not observe clear differences between the NextSeq500 and HiSeq2500 sequencing platforms for determination of the UMI count and the gene count for Quartz-Seq2 (Additional file 1: Fig. S14a). On the other hand, the unique mapping ratio and UMI count clearly depended on the Read2 length for transcript mapping (Additional file 1: Fig. S14c, d). In this regard, the number of detected genes slightly depended on Read2 length (Additional file 1: Fig. S14d). Long Read2 length led to better UMI counts, but high sequencing cost (Additional file 1: Fig. S14b). Therefore, in this study, we mainly used the NextSeq500 sequence platform and a Read2 length of 62 due to the cost performance of this combination.

Library preparation method for Quartz-Seq2

Moreover, we mainly used the adaptor ligation method for sequence library preparation of Quartz-Seq2. We did not observe a clear difference between the two sequence library preparation methods (ligation and Nextera/Tn5 transposase) regarding determination of the UMI count and the gene count. However, we could easily identify specific familiar genes (such as *Pou5f1* and *Gapdh*) using the ligation method (Additional file 1: Fig. S14e). Therefore, we mainly used the ligation method for Quartz-Seq2.

References

1. Sasagawa Y, Nikaido I, Hayashi T, Danno H, Uno KD, Imai T, et al. Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* 2013;14:R31.

2. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods*. 2009;6:377–82.
3. Kurimoto K. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res*. 2006;34:e42–2.
4. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015;161:1202–14.
5. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*. 2015;161:1187–201.
6. Bose S, Wan Z, Carr A, Rizvi A, Vieira G, Pe'er D, et al. Scalable microfluidics for single-cell RNA printing and sequencing. *Genome Biol*. 2015;16:120.

Supplemental Figures

Fig. S1. Summary of improvements conferred by Quartz-Seq2

Target RNA was converted into amplifiable cDNA with a cell barcode sequence via several steps. We show a summary of the improvements in each step in the boxes. To convert more mRNA molecules to first-strand cDNA, we improved the efficiency of reverse transcription. Moreover, we used low-enzyme concentration for reverse-transcription. To collect more cDNA molecules for subsequent steps, we improved the collection efficiency in pooling steps. In addition, to convert more cDNA molecules to amplifiable cDNA, we improved the poly-A tagging efficiency. Finally, we reduced the PCR bias using an UMI sequence in the reverse-transcription primer.

Fig. S2. Suppression of byproduct synthesis for poly-A tagging strategy

a) A schematic representation of the suppression of byproduct synthesis. RT primer remained in purified cDNA solution. Survival RT primer could be targeted with terminal transferase to attach a poly-A tail, the same as in the first-strand cDNA. Both types of resulting DNA had complementary sequences at both ends. The total length of the byproduct was shorter than first-strand cDNA. Previously, we reported that the length of target DNA is important for the suppression of byproduct synthesis. Specifically, the byproduct tended to form a pan-like structure, which prevented binding between PCR primer and template DNA. Finally, byproduct amplification was suppressed.

b) The effect of RT primer on byproduct synthesis. We used three kinds of RT primer (Quartz-Seq RT primer, v2 RT primer, and v3.1 RT primer) for whole-transcript amplification, in accordance with the Quartz-Seq-like conditions. First-strand cDNA with respective RT primer was treated with column purification and/or exonuclease I digestion. After that, purified cDNA solution was used for whole-transcript amplification. The yellow arrowhead indicates the byproduct derived from RT primer.

c) Typical cDNA pattern of Quartz-Seq2 from a 384-well plate containing 10 pg of total RNA.

Fig. S3. The efficiency of first-strand cDNA synthesis and column purification

a) We validated the purification efficiency. We prepared first-strand cDNA with 1 ng of total RNA and 1 ng of total RNA with 10 ng of tRNA using oligo-dT primer (mixture of v3.1). We performed column purification for first-strand cDNA solution using DNA Clean & Concentrator™-5 kit.

Thereafter, we determined the qPCR scores for eight genes (*Eef1b2*, *Nanog*, *Pou5f1*, *Sox2*, *Utf1*, *Spp1*, *Ywhae*, and *Tbp*) using purified first-strand cDNA and unpurified first-strand cDNA. We estimated the purification efficiency using the slope score of linear regression analysis by comparison between unpurified and purified conditions. The purification efficiency was 93.77% (n=4). The 10 ng of tRNA as carrier RNA. We thus investigated whether column purification could be used instead of exonuclease I treatment, which reached 95.52% (n=4) (p-value, 0.987). SCC stands for Spearman's rank correlation coefficient, while PCC stands for Pearson's correlation coefficient.

b) Improvement of reverse-transcription efficiency. We prepared first-strand cDNA using 200 pg of total RNA and poly-dT primer (mixture of v3.1) under six conditions (n=4). We also prepared first-strand cDNA using 200 pg of total RNA and random hexamer as a control for comparison (n=4). For the reverse transcription with random hexamer, we used the SuperScript IV system in accordance with the manufacturer's instructions. We determined qPCR scores for eight genes (*Eef1b2*, *Nanog*, *Pou5f1*, *Sox2*, *Utf1*, *Spp1*, *Ywhae*, and *Tbp*) using the above cDNA. We describe the respective conditions (buffers and temperature conditions) for reverse transcription with poly-dT primer in the boxes. In Quartz-Seq, the reverse-transcription buffer and temperature conditions were as follows (PCR buffer and 35°C for 5 min, 45°C for 20 min). Finally, we adopted the following conditions for Quartz-Seq2 (T100 buffer and 35°C for 5 min, 50°C for 50 min). Details of the buffer compositions are presented in Additional file 5: Table S4. We estimated the relative efficiency of reverse transcription using the slope score of linear regression analysis by comparison between poly-dT conditions and random hexamer conditions. In the Quartz-Seq condition, the relative efficiency of reverse transcription was 48.48%. In the Quartz-Seq2 condition, the relative efficiency was 59.02%.

Fig. S4. How to assemble collector unit for pooling of cell-barcoded cDNA

a) Dimensional outline of the stainless-steel frame adaptor for collector type A. The metal frame was made by bending and welding one stainless-steel panel at a small factory (Yakkensha or AS ONE Corporation).

b) Assembly of collector type A. (1) First, we prepared a disposable one-well reservoir (Nalgene Microplate Robotic Reservoir with Flat Bottom 1200-1301, Thermo Scientific). (2) We put the metal frame around the one-well reservoir. (3) To prevent the spilling of liquid from the source plate to the reservoir plate, we sealed the foresides of the reservoir with paraffin film (Parafilm). (4) We then turned the reaction 384-well PCR plate upside down on the assembled collector type A. (5) Next, we set it on a centrifuge adaptor. (6) We then centrifuged the plate with the assemble collector at 3,010 g and 4°C for 3 min with swing-bucket rotors. By using spin-down collection (type A), we collected $88.7 \pm 0.99\%$ (n=3) of solution in 5 min from two to four plates. We mainly used type A for cDNA collection.

c) Assembly of collector type B. (0) By using type B, cDNA solution of 384 wells was collected and split into eight wells in a reservoir. (1) We prepared an eight-well disposable reservoir. (2) We put a joint-adaptor (384 Transfer Plate 1859-3845, Watson/Fukaekasei) around the eight-well reservoir. The joint-adaptor has 384 funnel-shaped wells, which enable liquid transfer. (3) We turned the reaction 384-well PCR plate upside down on the assembled collector type B. (4) We then fixed the assembled one using metal holders (Watson/Fukaekasei). (5) Next, we set it on a centrifuge adaptor. (6) We centrifuged the plate with the assemble collector at 3,010 g and 4°C for 3 min with swing-bucket rotors. By using spin-down collection (type B), we collected $86.9 \pm 0.4\%$ (n=2) of solution in 5 min from two to four plates.

Fig. S5. Reproducibility of improvement of poly-A tagging reaction

a) Bar plot represents the relative DNA yield in various poly-A tagging conditions using first-strand cDNA purified from 1 ng of total RNA (n=3 for each condition). We quantified cDNA yield (400–5,900 bp) and byproduct DNA yield (100–400 bp) by Fragment Analyzer (Advanced Analytical). A red color indicates the conditions for which Quartz-Seq2 was used.

- b) Another experiment for the improvement of poly-A tagging efficiency with T55 buffer and the “Increment” condition.
- c) The improvement of cDNA yield with the combination of T55 buffer and “Increment” (T55+Inc) was confirmed by qPCR assay. We detected eight genes (*Eef1b2*, *Nanog*, *Pou5f1*, *Sox2*, *Utf1*, *Spp1*, *Ywhae*, *Tbp*) with amplified cDNA of Figure 2a and nonamplified cDNA. We prepared nonamplified cDNA with 200 pg of total RNA and random hexamer. SCC stands for Spearman’s rank correlation coefficient.

Fig. S6. Cost comparison of single-cell RNA-seq methods including Quartz-Seq2

- a) Costs of sequence library DNA and sequence analysis with NextSeq500 for single-cell RNA-seq methods. Quartz-Seq refers to the method that we previously developed. We show the cost in Japanese currency [the monetary unit is the yen (¥)]. The exchange rates for one US dollar and one euro were 113 yen and 131 yen on average in November 2017. The blue bar represents the sequence cost, if we prepare 0.1 M or 0.2 M fastq reads on average for a single cell. The cost of several methods is derived from a previous study (asterisks) [1].
- b) The relative total cost including preparation cost and sequence cost for single-cell RNA-seq methods.

1. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell.* 2017;65:631–4.

Fig. S7. Cost reduction via a low concentration of enzymes in reverse transcription for Quartz-Seq2

- a) The step of reverse transcription has the highest cost in the single-cell experimental part of Quartz-Seq2. We present the different proportions of costs associated with each of the experimental steps. The step of reverse transcription accounts for approximately 65% of the total cost in the “RT100” condition. We thus established the “RT25” condition by reducing the enzyme concentration to 25% in reverse transcription (for details, see Methods). Moreover, the number

of columns for the concentration of first-strand cDNA affected the cost for the process downstream of purifying cDNA. The use of a small number of columns also decreased the cost.

b) Assay assessing the use of a low enzyme concentration in reverse transcription. We reverse-transcribed 200 pg of total RNA into first-strand cDNA with two buffers (PCR buffer and T100 buffer). T100 buffer was used for the reverse-transcription step of Quartz-Seq2. We reduced the levels of reverse transcriptase and RNase inhibitor in the reverse-transcription premix as follows: RT100 (20 units/ μ L SuperScript III, 2.2 units/ μ L RNasin plus), RT80 (16 units/ μ L SuperScript III, 1.76 units/ μ L RNasin plus), RT60 (12 units/ μ L SuperScript III, 1.32 units/ μ L RNasin plus), RT40 (8 units/ μ L SuperScript III, 0.88 units/ μ L RNasin plus), and RT20 (4 units/ μ L SuperScript III, 0.44 units/ μ L RNasin plus). We determined the expression of four genes by qPCR for the respective conditions (n=4). We show the relative qPCR score, which is normalized by RT100 with PCR buffer for the respective genes.

c) We prepared first-strand cDNA using 200 pg of total RNA and poly-dT primer (mixture of v3.1) under serially diluted enzyme conditions (technical replicates, n=5). The detail of enzyme concentration is presented in the figure. We determined qPCR scores for ten genes (*Eef1b2*, *Nanog*, *Pou5f1*, *Sox2*, *Fn1*, *Spp1*, *Utf1*, *Atp5a*, *Rex1*, and *Lefty2*) using the above cDNA. For Figure 2b, we calculated the average relative qPCR score with 10 genes per technical replicate.

d) Bar plots present the amount of amplified cDNA from three technical replicates of a 384-well plate with 10 pg of total RNA in all wells for the “RT100” and “RT25” conditions. The y-axis represents the yield of amplified cDNA. The cDNA yields from 10 pg of total RNA were 55.9 ± 5.5 ng for “RT100” and 65.7 ± 2.5 ng for “RT25.” The presented p-value was obtained using two-tailed Welch's t-test.

Fig. S8. Low enzyme concentration in reverse transcription improved the variability of UMI counts and gene counts

For three conditions [Quartz-Seq-like, Quartz-Seq (RT100), Quartz-Seq (RT25)], we performed three batch experiments using all wells of a 384-well plate, which contained 10 pg of total RNA. For Quartz-Seq2, we used different enzyme concentrations for reverse transcription. We used the following reverse-transcription premix: RT100 (2x Thermopol buffer, 20 units/ μ L SuperScript

III, 2.2 units/ μ L RNasin plus) and RT25 (2x Thermopol buffer, 5 units/ μ L SuperScript III, 0.55 units/ μ L RNasin plus) for Quartz-Seq2. The presented p-value was obtained using two-tailed Welch's t-test between the RT100 condition (three 384-well plates) and the RT25 condition (three 384-well plates). The UMI conversion efficiency levels at approximately 0.096 M fastq reads on average per well were as follows: Quartz-Seq-like ($14.83 \pm 1.3\%$, three 384-well plates), Quartz-Seq2 (RT100) ($27.76 \pm 1.05\%$, three 384-well plates), and Quartz-Seq2 (RT25) ($31.25 \pm 0.81\%$, three 384-well plates). We prepared library DNA with Tn5 transposase for the batch 3 experiment for the Quartz-Seq-like condition (asterisks). For accurate comparison of quantitative performance, we removed batch 3 of Quartz-Seq-like in Figure 2.

b) We analyzed the same sequence library DNA with NextSeq500 and HiSeq2500. Read2 lengths of NextSeq500 and HiSeq2500 were 62 and 98, respectively. Circles represent the data from NextSeq500. Triangles represent the data from HiSeq2500.

Fig. S9. Calculation of UMI conversion efficiency

We performed the conversion from fastq reads as initial reads to UMI counts via several steps. We define the formula for calculating the UMI conversion efficiency (boxed region). Each parameter is defined as follows: UMI_{sc} : the number of UMI count, assigned to a single-cell sample, $fastq_{sc}$: the number of fastq reads derived from each single-cell sample, $fastq_{non-sc}$: the number of fastq reads derived from non-single-cell samples, which include experimental byproducts such as WTA adaptors, WTA byproducts, and non-STAMPs. Initial fastq reads are composed of $fastq_{sc}$ and $fastq_{non-sc}$.

Fig. S10. Effect of lower enzyme concentration on quantitative performance of Quartz-Seq2

We prepared sequence library DNA of Quartz-Seq2 from 10 pg of total RNA of four 384-well plates by using two different reverse-transcription premixes as follows: RT25 (2x Thermopol buffer, 5 units/ μ L SuperScript III, 0.55 units/ μ L RNasin plus) and RT6.25 (2x Thermopol buffer, 1.25 units/ μ L SuperScript III, 0.1375 units/ μ L RNasin plus) for Quartz-Seq2. We performed four-batch experiments on different experimental days for the RT25 and RT6.25 conditions.

a) We present the quantitative performances of Quartz-Seq2. b) Technical variability of Quartz-Seq2 with RT25 and RT6.25. We present the distribution of gene expression variability from 1,536 wells. We show the mean and S.D of dispersion in a box. c) Quantitative performance of ERCC spike-in RNA for Quartz-Seq2 under two different conditions. The x-axis represents the average of the detection limit. The y-axis represents the ERCC capture efficiency.

Fig. S11. Technical introduction to Drop-seq

a) We observed a clear knee in the cumulative distribution plot of the UMI count, which corresponded to the cell barcodes derived from cells.

b) We validated our technical introduction to Drop-seq using a species mixing experiment. Specifically, we performed a Drop-seq experiment using a mixture of mouse ES cells and human iPS cells. We analyzed 16,000 barcode beads. We confirmed that sequence reads of each cell barcode were exclusively mapped to mouse and human transcripts.

c) Schematic representation of the experimental design for the validation of Drop-seq and Quartz-Seq2. ES cells and Dex-treated ES cells were dissociated from each other. Dex-treated ES cells differentiated into primitive endoderm-like cells. The cell suspensions for these were equally mixed. We used this mixture of cells for Drop-seq and Quartz-Seq2 in the “RT25” condition.

d) Knee plot of UMI count for Drop-seq using a mixture of ES cells and Dex-treated ES cells. We selected 500 STAMPs for downstream analysis of Drop-seq.

e) The number of UMI count and gene count per cell in the selected number of STAMPs.

f) Distribution of gene count for Quartz-Seq2 using ES cells and Dex-treated ES cells. We excluded low-quality cells that expressed fewer than 4,000 genes.

g) Drop-seq and Quartz-Seq2 could distinguish between ES cells and differentiated cells by principal component analysis. It is known that the *Dab2* gene is highly expressed in differentiated cells. Red dots represent differentiated cells that highly express the *Dab2* gene.

h) Schematic representation of filtering before the calculation of differentially expressed genes between ES and Dex-treated ES cells (PrE cells).

Fig. S12. Simulation-based power analysis for Figures 3 and 4

a) We performed simulation-based power analysis by using the data set of Quartz-Seq2 and Drop-seq in Figure 3 with G6GR ES cells. The true positive rate (TPR) and false discovery rate (FDR) were determined for 0.24 million reads on average per cell for the following sample sizes: $n = 16$, $n = 32$, $n = 64$, $n = 128$, $n = 256$, and $n = 512$ per group. Line plots represent the median of 25 simulations. We also show the absolute numbers of true-positive-simulated genes and false-discovery-simulated genes. Line plots represent the median and S.D. of 25 simulations.

b-c) Simulation-based power analysis by using the data set of Quartz-Seq2 and other methods presented in Figure 4 with J1 ES cells. True positive rate (TPR) and false discovery rate (FDR) were determined for 0.1 million reads on average per cell for the following sample sizes: $n = 16$, $n = 32$, $n = 64$, $n = 128$, $n = 256$, and $n = 512$ per group. Line plots represent the median of 25 simulations. We also show the absolute numbers of true-positive-simulated genes and false-discovery-simulated genes. Line plots represent the median and S.D. of 25 simulations.

Fig. S13. Assessment of quantitative performance for external control RNA using Quartz-Seq2 data and previously reported data

a) Quantitative data of ERCC spike-in RNA about Quartz-Seq2 data, including 25 batch experiments corresponding to 11,520-well samples. We plotted the ERCC capture efficiency and average detection limit at respective data points of initial fastq reads for 25 batch experiments. All batch experiments can ensure 30,000 fastq reads/well on average (red line). We calculated the ERCC capture efficiency and detection limit for CEL-seq2(C1), CEL-seq2, and MARS-seq by using a digital expression matrix from Svensson et al. [2]. We observed that these performances strongly correlated with the numbers of initial fastq reads and samples. Nevertheless, Quartz-Seq2 had good performance equivalent to or higher than the range of fewer initial fastq reads.

b) Further comparison of the detection limit between Quartz-Seq2 and CEL-seq(C1). We reanalyzed CEL-seq2(C1) ($n=96$) with SRA files, which were also used by Svensson et al. [2-3], using our pipeline. First, we confirmed that the detection limit of reanalyzed CEL-seq2(C1) was not an underestimate compared with that of previously analyzed CEL-seq2(C1). We observed

that the detection limit of all Quartz-Seq2 data from 25 batch experiments was lower than that of CEL-seq2 in the range of 30,000 initial fastq reads.

c) Strong correlation between the quantitative performance of ERCC spike-in RNA and the average amount of total RNA per single cell. The x-axis represents the average amount of total RNA per single cell. The y-axis represents the ERCC capture efficiency or average detection limit of ERCC spike-in RNA. We also show Spearman's rank correlation (SCC) between the average amount of total RNA per single cell and the quantitative performances of ERCC spike-in RNA from respective batch experiments.

d) We show the ERCC capture efficiency and detection limit from respective single-cell RNA methods using J1 ES cells.

2. Svensson V, Natarajan KN, Ly L-H, Miragaia RJ, Labalette C, Macaulay IC, et al. Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods*. 2017;6:150.

3. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol*. 2016;17:892.

Fig. S14. Sequence platform and sequence library preparation for Quartz-Seq2

a) There was no obvious difference between the NextSeq500 platform and the HiSeq2500 platform for Quartz-Seq2. We analyzed the same sequence library DNA with these different platforms. We did not observe any clear differences in UMI count and gene count.

b) We show the relative costs with the three sequence platforms for Quartz-Seq2. Read2 sequence of Quartz-Seq2 was used for the mapping to transcript. A Read2 length of 62 nt in NextSeq500 was the most cost-effective setting for Quartz-Seq2.

c) Read2 length affected the unique mapping ratio. We prepared the three kinds of sequence library DNA from the same amplified cDNA of Quartz-Seq2. We prepared the library DNA using the ligation method and the Nextera method. We also prepared data comprising a mixture of ligation data and Nextera data.

Fig. S15. We did not observe a clear batch effect among the twelve 384-well plates in Quartz-Seq2

We plotted the 4,484 cells on t-SNE space with color labeling for each 384-well PCR plates.

Fig. S16. Quality check of high proportion of mitochondrial RNA for a cell cluster

We show the UMI count, gene count, and the proportion of mitochondrial RNA for each cluster. We present the mean UMI count for each cluster. In cluster 2, both marker genes (ES and PrE marker genes) were expressed. In addition, the UMI count of cluster 2 was about twice those of clusters 1 and 4. Therefore, cluster 2 was judged to represent doublets of an ES cell and a PrE cell. Moreover, we determined the UMI count derived from mitochondrial genes. We also calculated the proportion of mitochondrial RNA for respective cell clusters. In cluster 6, the proportion was higher than for the other cell clusters.

Fig. S17. Magnified version of Figure 6c

Fig. S18. Quartz-Seq2 with ultra-shallow fastq reads enables identification of heterogeneity of cell cycle phase within a cell type.

a) Cells are plotted on t-SNE space with coloring representing the intensity of Hoechst 33342 staining. We used 0.01 M fastq reads on average per cell as initial reads. b) Identification of cell cycle-associated genes.

Fig. S19. Correlation between gene count and cell size in SVF populations

a) Distribution of gene counts from respective single cells. We excluded those single cells expressing 500 or less different genes. b) Correlation between gene count and cell size from flow cytometry data. Each dot represents a single cell. Different colors are used to identify the different cell clusters. c) Distribution of cell size for each cell cluster. d) Immunofluorescence images of stromal vascular fraction. Yellow arrow-heads represent stained cell with antibody. White scale bars represent 10 μ m.

Fig. S20. Full-sized image of Figure 7d

Fig. S21. Expression profile of ligand and receptor genes for stromal vascular fraction

We present the gene expression using Quartz-Seq2 data and previous reported list of ligand and receptor genes [4]. Listed gene expression is displayed as a blue color in a heatmap. Class labeling of single-cells in stromal vascular fraction is displayed as each color. a) Heatmap of gene expression for ligand gene. b) Heatmap of gene expression for receptor gene.

4. Ramilowski JA, Goldberg T, Harshbarger J, Kloppman E, Lizio M, Satagopam VP, et al. A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat Commun.* 2015;6:7866.

Fig. S22. Expression profile of previously reported MSC marker genes for stromal vascular fraction

We present the gene expression using Quartz-Seq2 data. We used a previously reported list of marker genes (*Thy1/Cd90*, *Eng/Cd105*, *Pdgfra*, *Lys6a/Sca1*, *Tnnt2*, *Nanog*, *Pou5f1*, *Sox2*, *Myog*, *Cd34*, *Cd31*, *Cd45*) in MSCs.

Fig. S23. Expression profile of previously reported ASC marker genes for clusters 1 and 8

We present the gene expression using Quartz-Seq2 data. We used a previously reported list of marker genes in CD34+/CD55+/Dpp4+ adipose-derived stem cells (ASC). Almost all marker genes were enriched in cluster 1.

Fig. S24. Definition of “variable genes the expression of which is less associated with the cell cycle phase”

a) In our strategy, cells were discretized into 40 equal-sized buckets based on the rank of Hoechst 33342 staining intensity. Then, the CV of averaged UMI counts for a gene in each bin was calculated (as mentioned in Methods). b) We represented the definition of “variable genes the expression of which is less associated with the cell cycle phase”.

Fig. S1

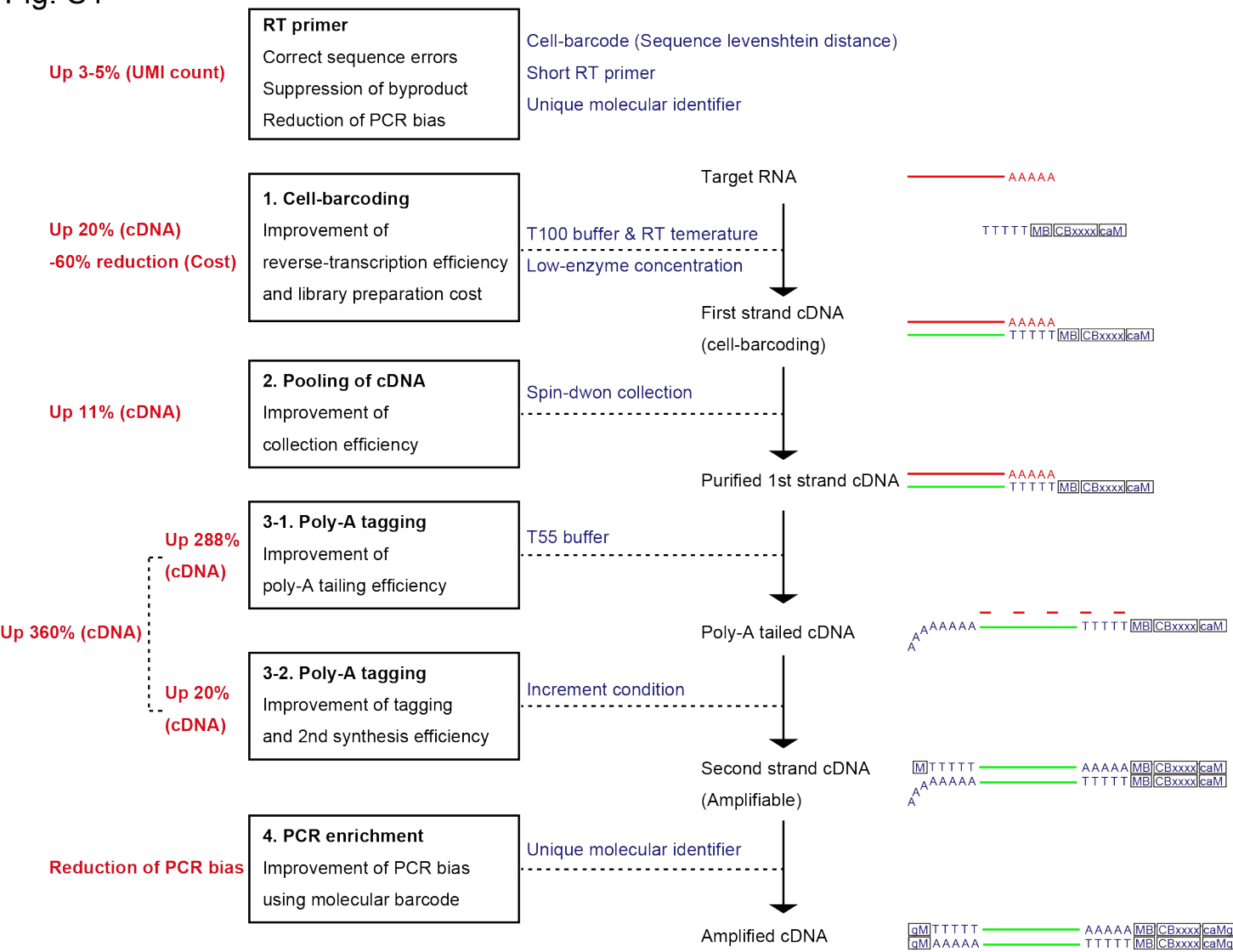
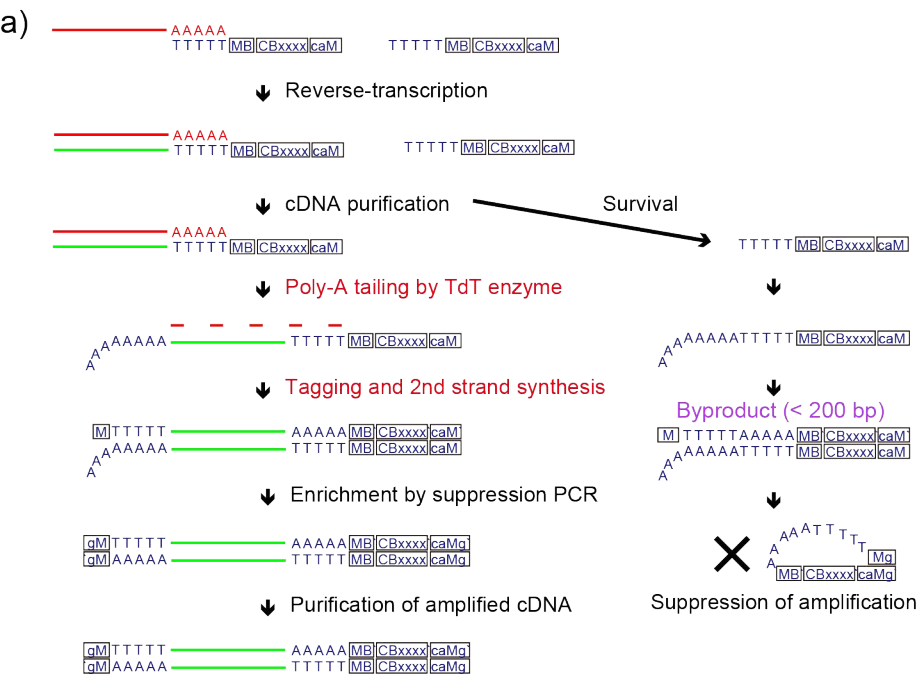


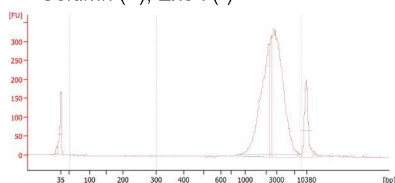
Fig. S2



b)

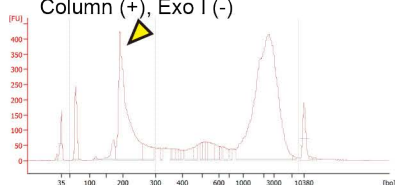
Quartz-Seq RT primer (70mer)

Column (+), Exo I (-)



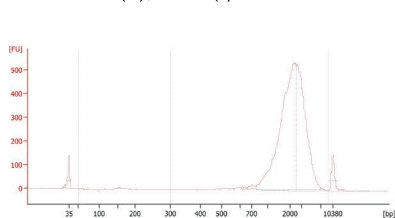
v2 RT primer (126mer)

Column (+), Exo I (-)



v3.1 RT primer (Quartz-Seq2) (73mer)

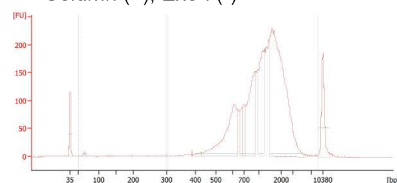
Column (+), Exo I (-)



c)

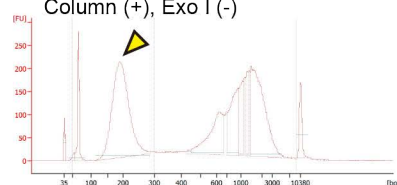
T55 (Quartz-Seq2)

Column (+), Exo I (-)



NBF40

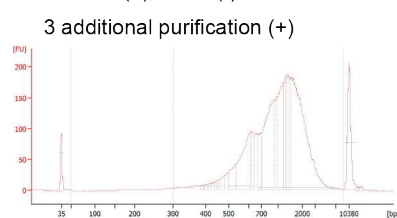
Column (+), Exo I (-)



NBF40

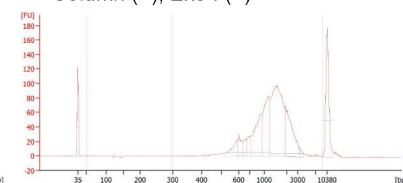
Column (+), Exo I (-)

3 additional purification (+)



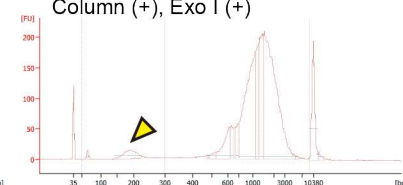
T55

Column (+), Exo I (+)



NBF40

Column (+), Exo I (+)



NBF40

Column (+), Exo I (+)

3 additional purification (+)

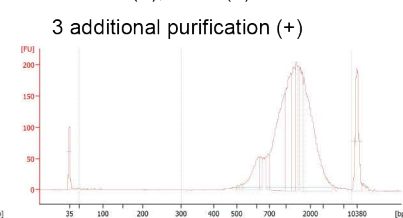
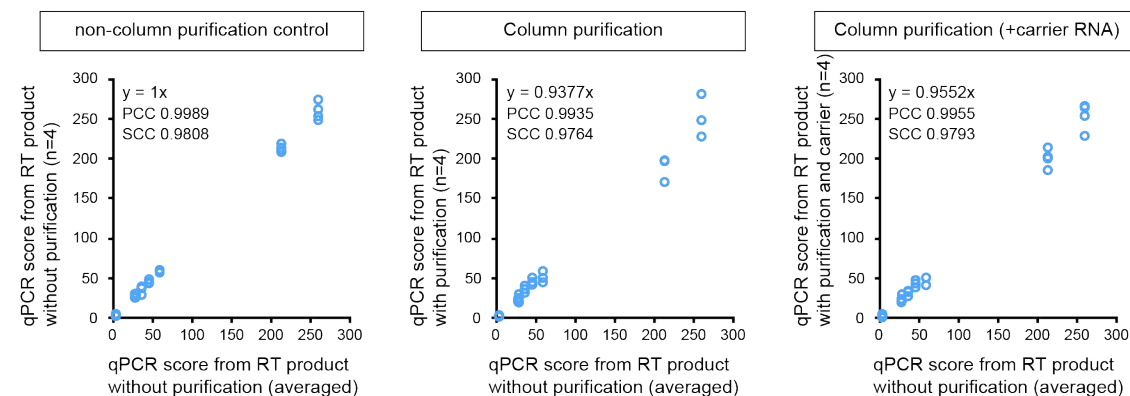
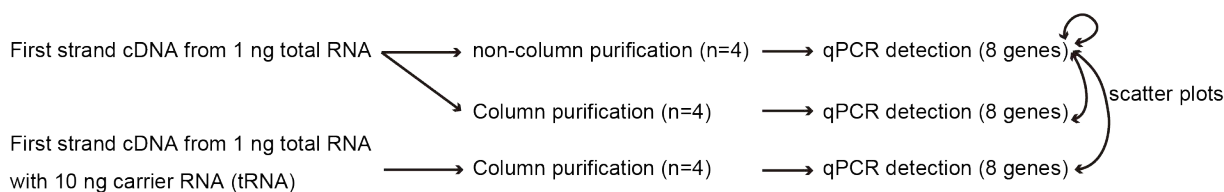


Fig. S3

a)



b)

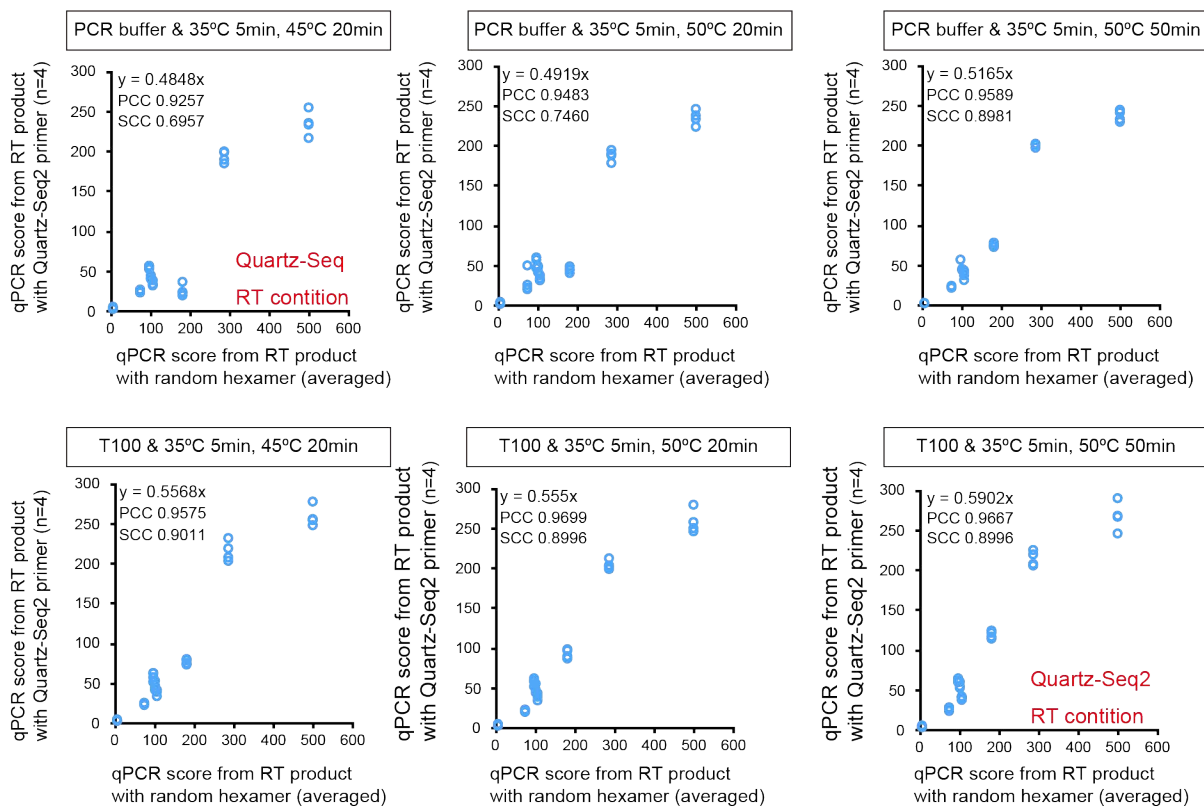
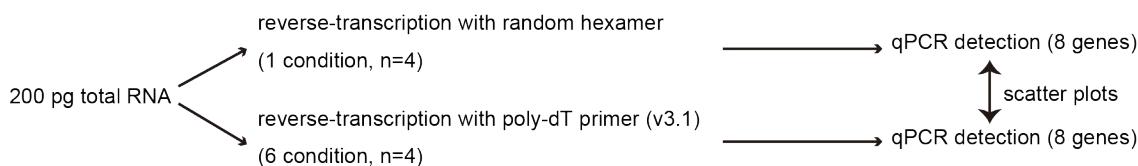
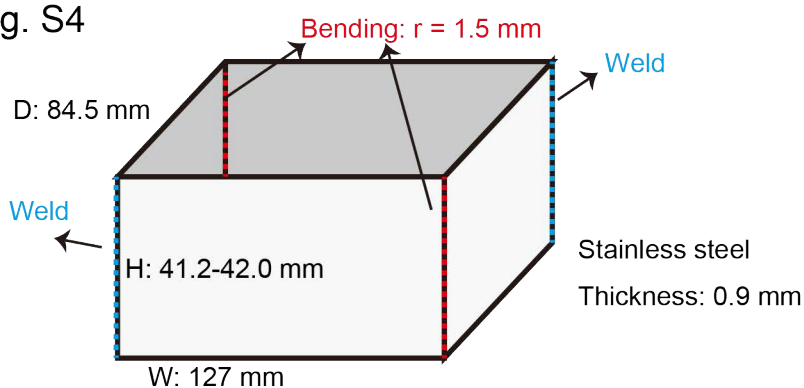


Fig. S4

a)

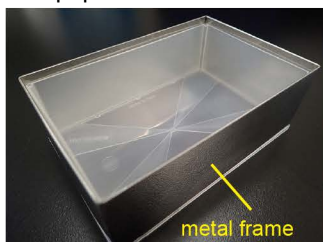


b)

1. One well reservoir



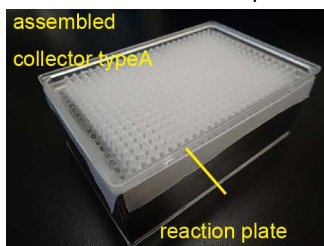
2. Equipment of metal frame



3. Equipment of paraffin films



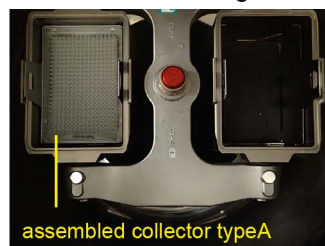
4. Place a reaction plate



5. Set on centrifuge adaptor

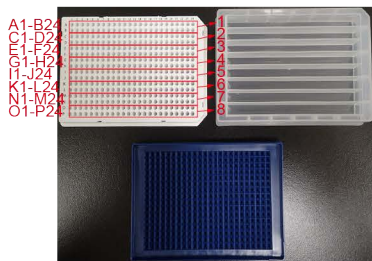


6. Set on centrifuge rotor

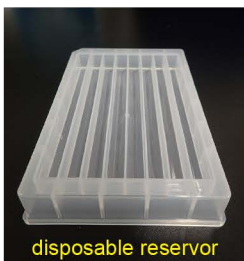


c)

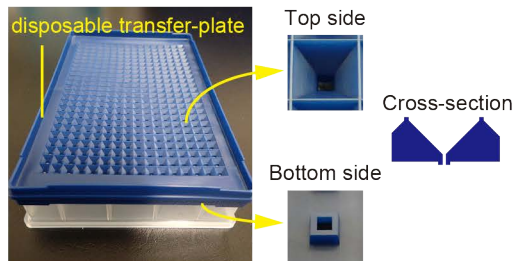
0. preparation for assembling



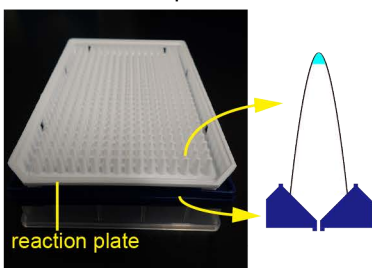
1. Eight well reservoir



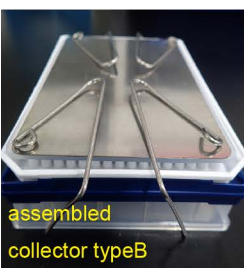
2. Equipment of transfer-plate



3. Place a reaction plate



4. Fixation



5. Set on centrifuge adaptor

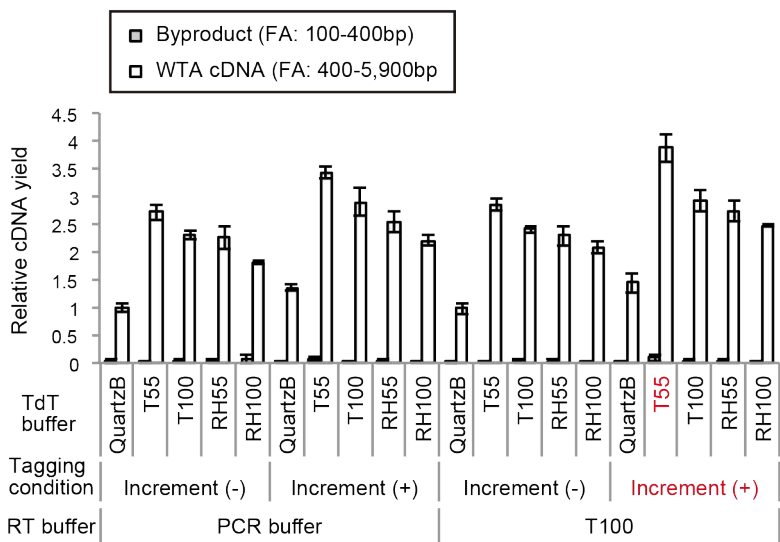


6. Set on centrifuge rotor

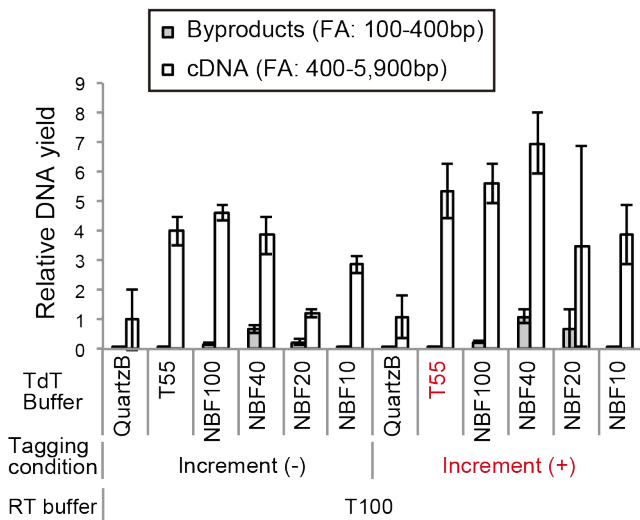


Fig. S5

a)



b)



c)

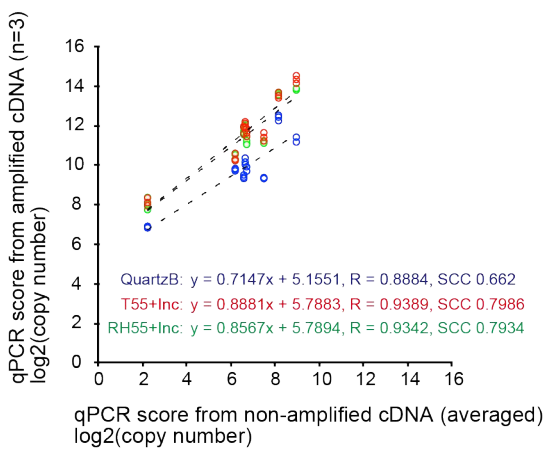
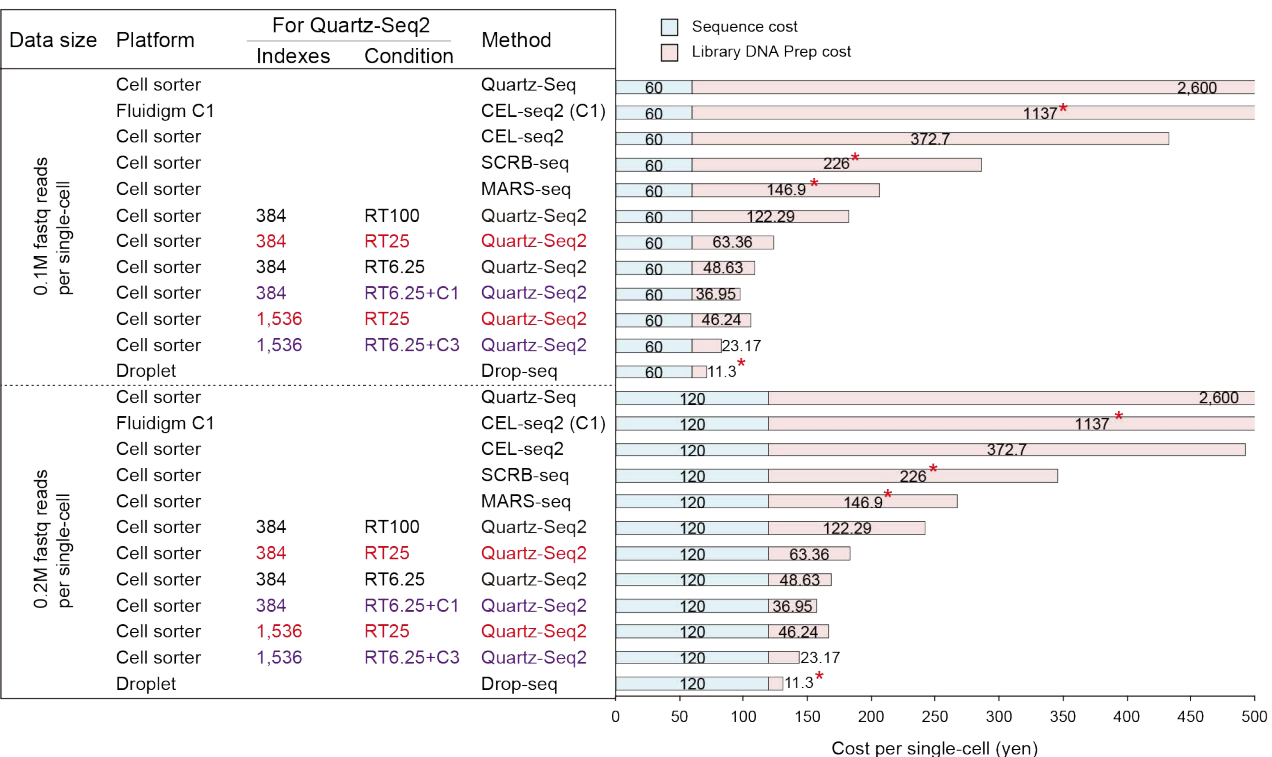


Fig. S6

a)



b)

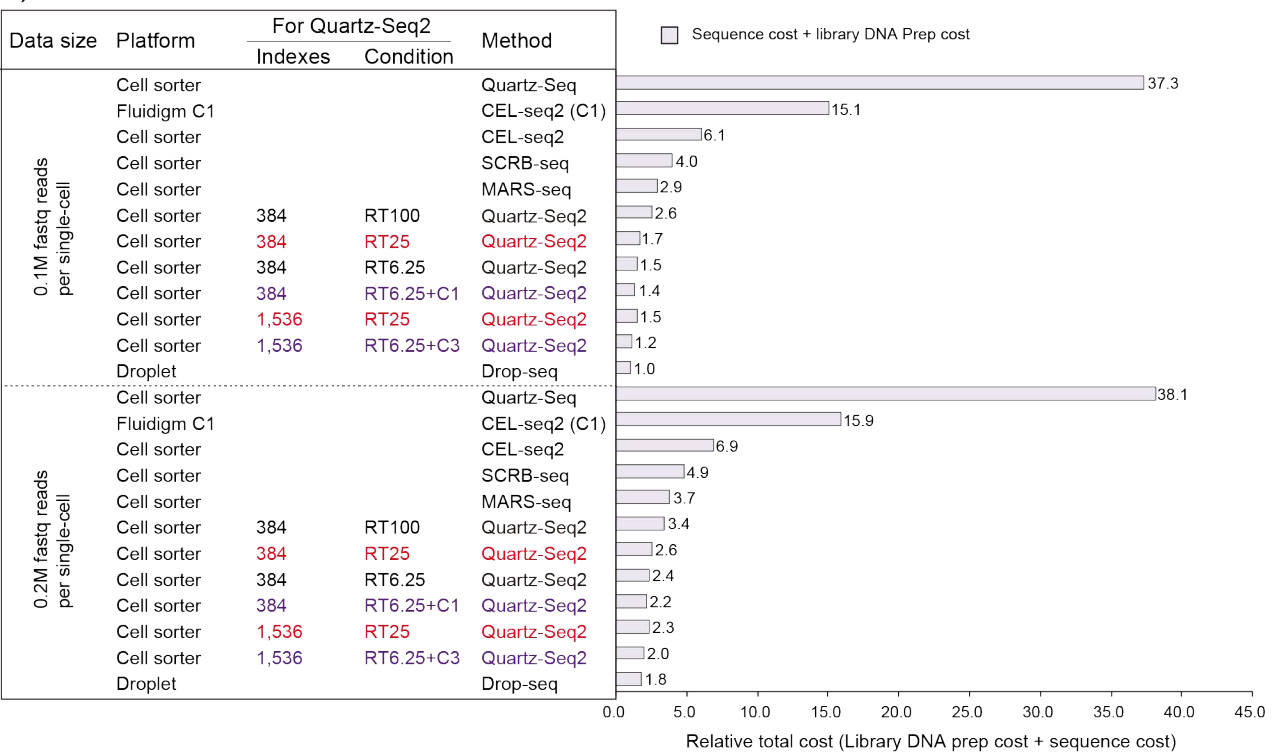
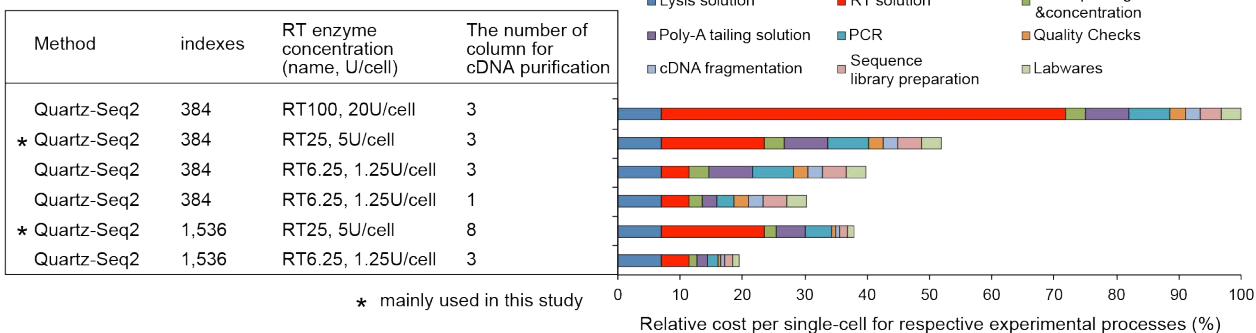
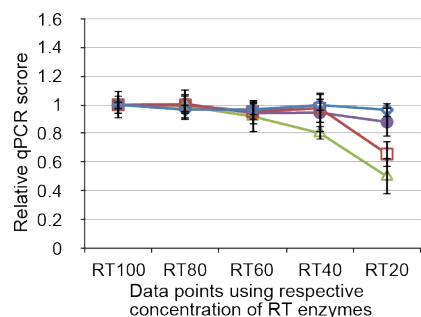
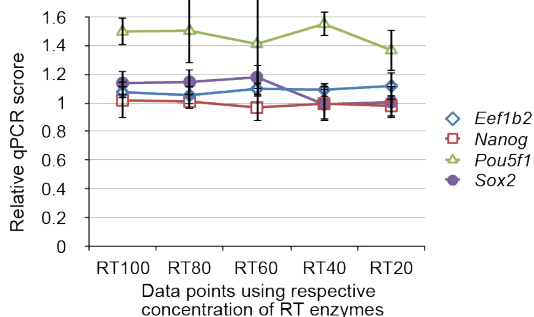


Fig. S7

a)

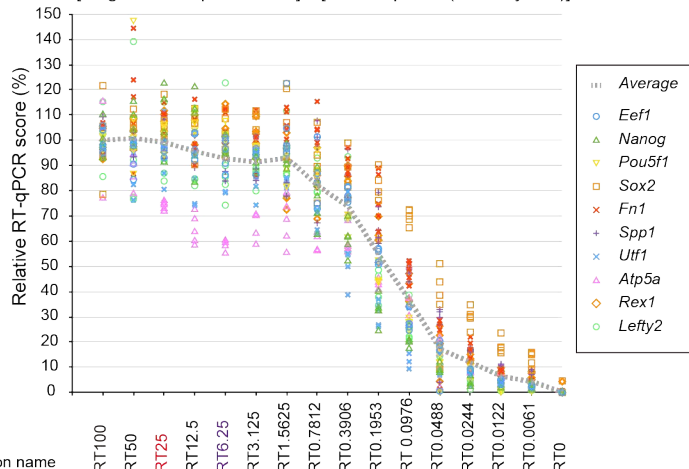


b)

Reverse-transcription buffer
PCR buffer (Quartz-Seq)Reverse-transcription buffer
T100 buffer (Quartz-Seq2)

c)

[10 genes RT-qPCR x n=5] x [16 data points (RT enzymes)]



Condition name

RT enzymes
concentrationInformation
about unit for
RT enzymes

100%	20U	5U	1.25U	0U
50%				
25%				
12.5%				
6.25%				
3.125%				
1.5625%				
0.7812%				
0.3906%				
0.1953%				
0.0976%				
0.0488%				
0.0244%				
0.0122%				
0.0061%				
0%				

SuperScript III reverse-transcriptase in RT solution(unit/uL)

2.2U 0.55U 0.135U 0U

RNase inhibitor (Unit/well) in RT solution (unit/uL)

Data points using respective concentration of RT enzymes

d)

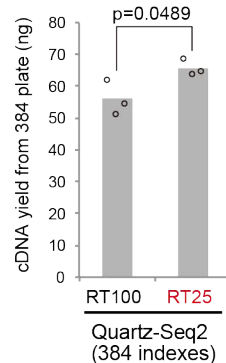
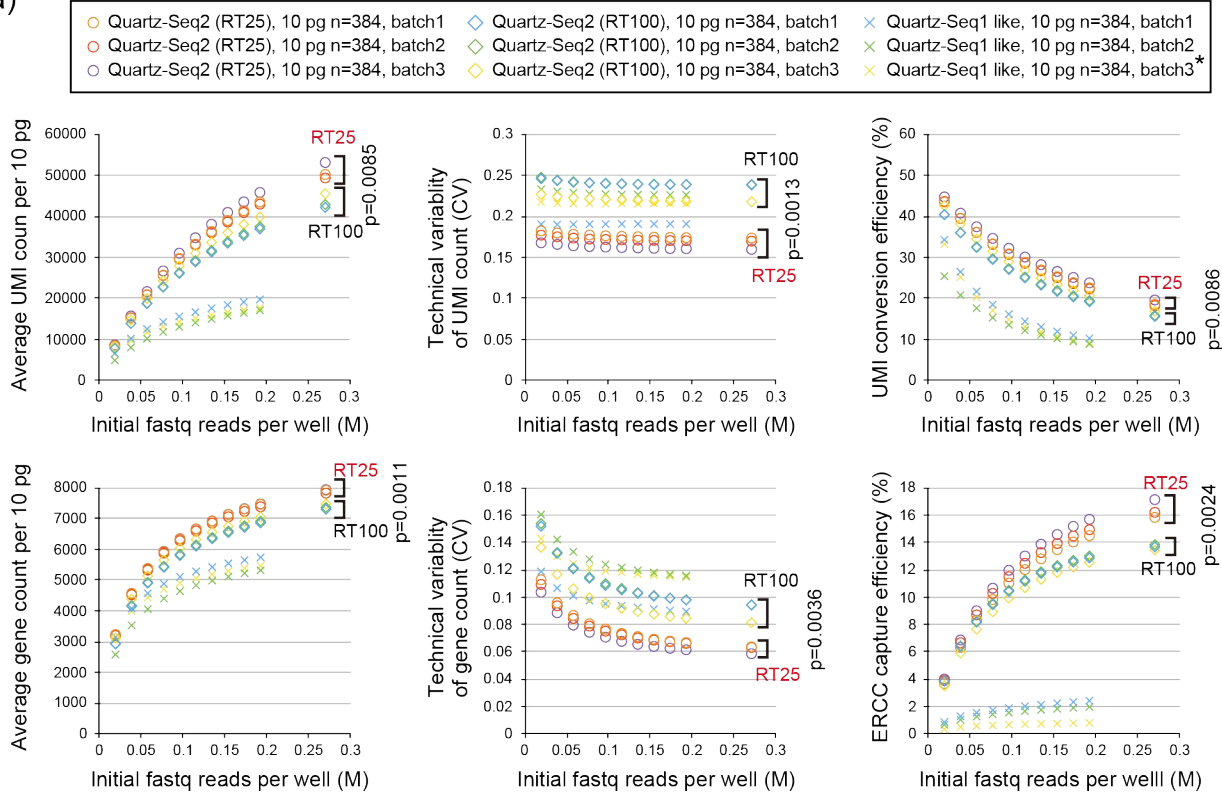


Fig. S8

a)



b)

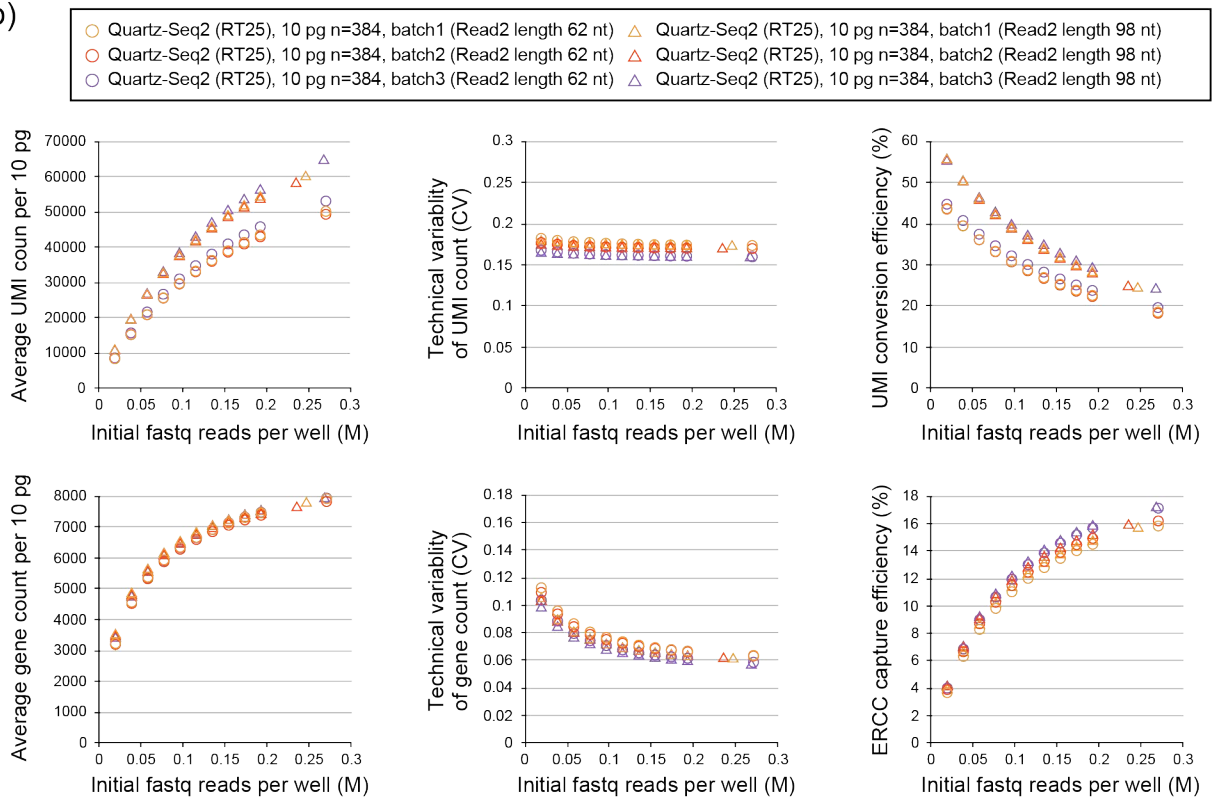


Fig. S9

$$\text{UMI conversion efficiency} = \frac{\sum UMI_{sc}}{\sum fastq_{sc} + fastq_{non-sc}}$$

UMI conversion
efficiency

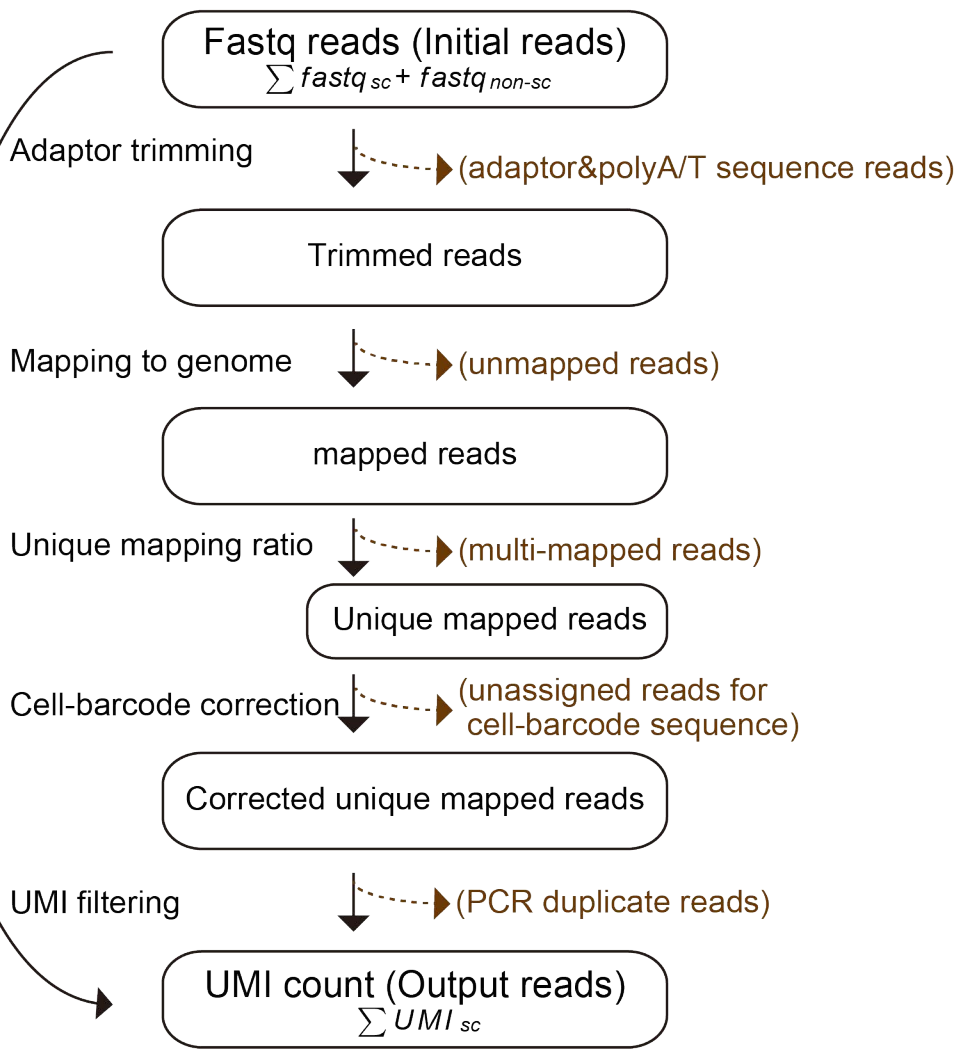
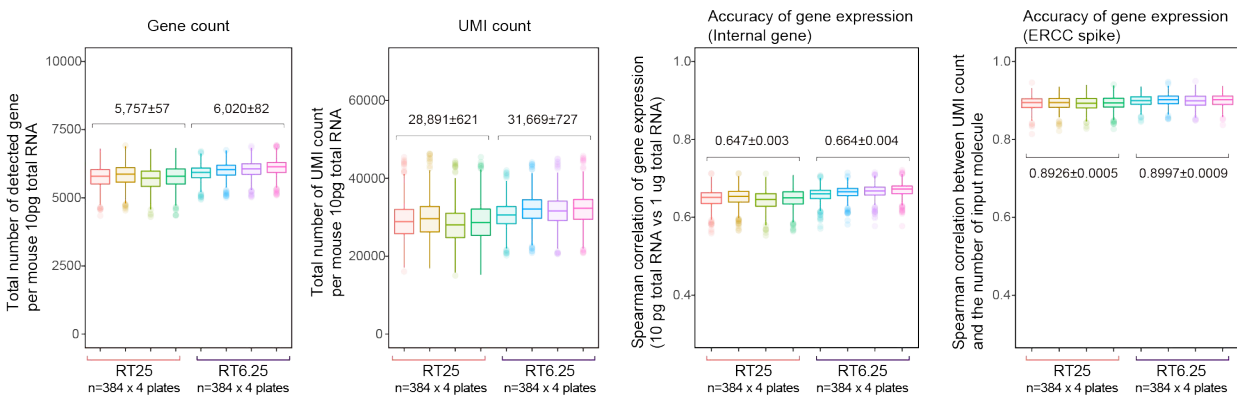


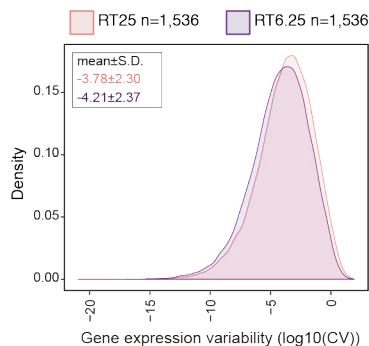
Fig. S10

a)

Sample name	Sample	Method	Type of indexes	Concentration of RT enzymes (%)	Experimental Day & batch number	The number of initial fastq reads per well (on average)
RT25_mES10pgV31_p24	mouse 10 pg total RNA (n=384)	Quartz-Seq2	384	25	Day 1 & batch 1	100,000
RT25_mES10pgV31_p25	mouse 10 pg total RNA (n=384)	Quartz-Seq2	384	25	Day 1 & batch 2	100,000
RT25_mES10pgV31_p28	mouse 10 pg total RNA (n=384)	Quartz-Seq2	384	25	Day 2 & batch 3	100,000
RT25_mES10pgV31_p29	mouse 10 pg total RNA (n=384)	Quartz-Seq2	384	25	Day 2 & batch 4	100,000
RT625_mES10pgV31_p22	mouse 10 pg total RNA (n=384)	Quartz-Seq2	384	6.25	Day 1 & batch 1	100,000
RT625_mES10pgV31_p23	mouse 10 pg total RNA (n=384)	Quartz-Seq2	384	6.25	Day 1 & batch 2	100,000
RT625_mES10pgV31_p26	mouse 10 pg total RNA (n=384)	Quartz-Seq2	384	6.25	Day 2 & batch 3	100,000
RT625_mES10pgV31_p27	mouse 10 pg total RNA (n=384)	Quartz-Seq2	384	6.25	Day 2 & batch 4	100,000



b)



c)

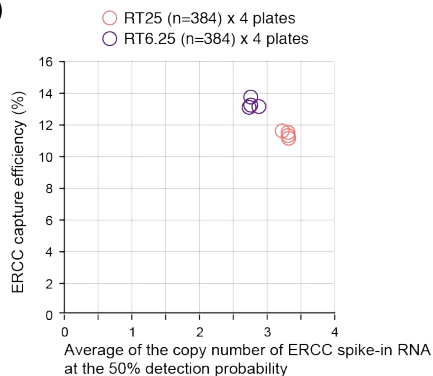


Fig. S11

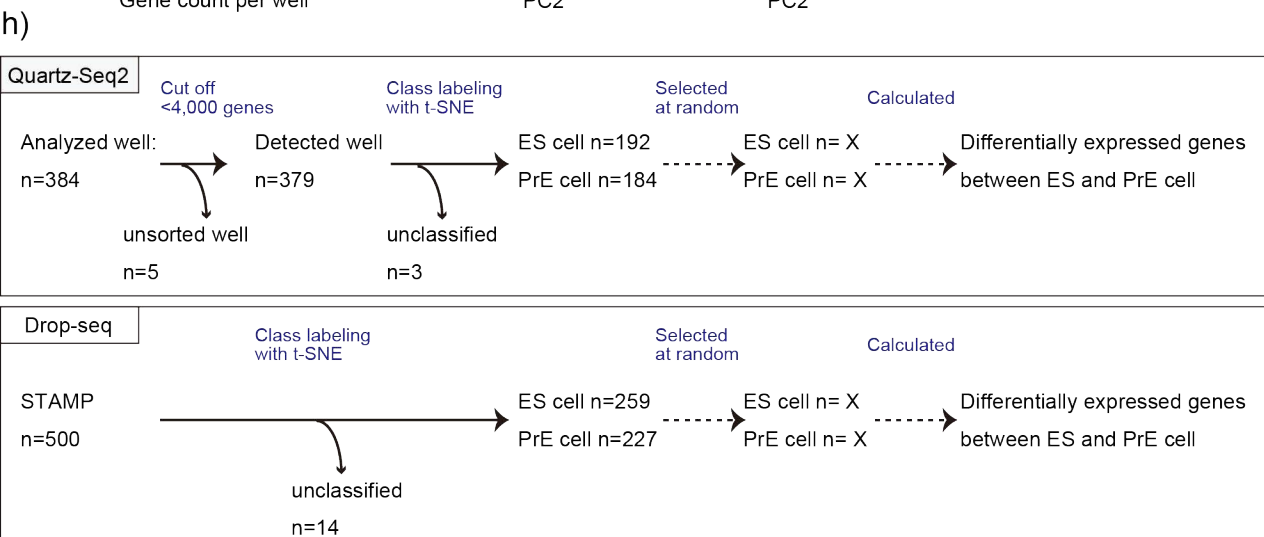
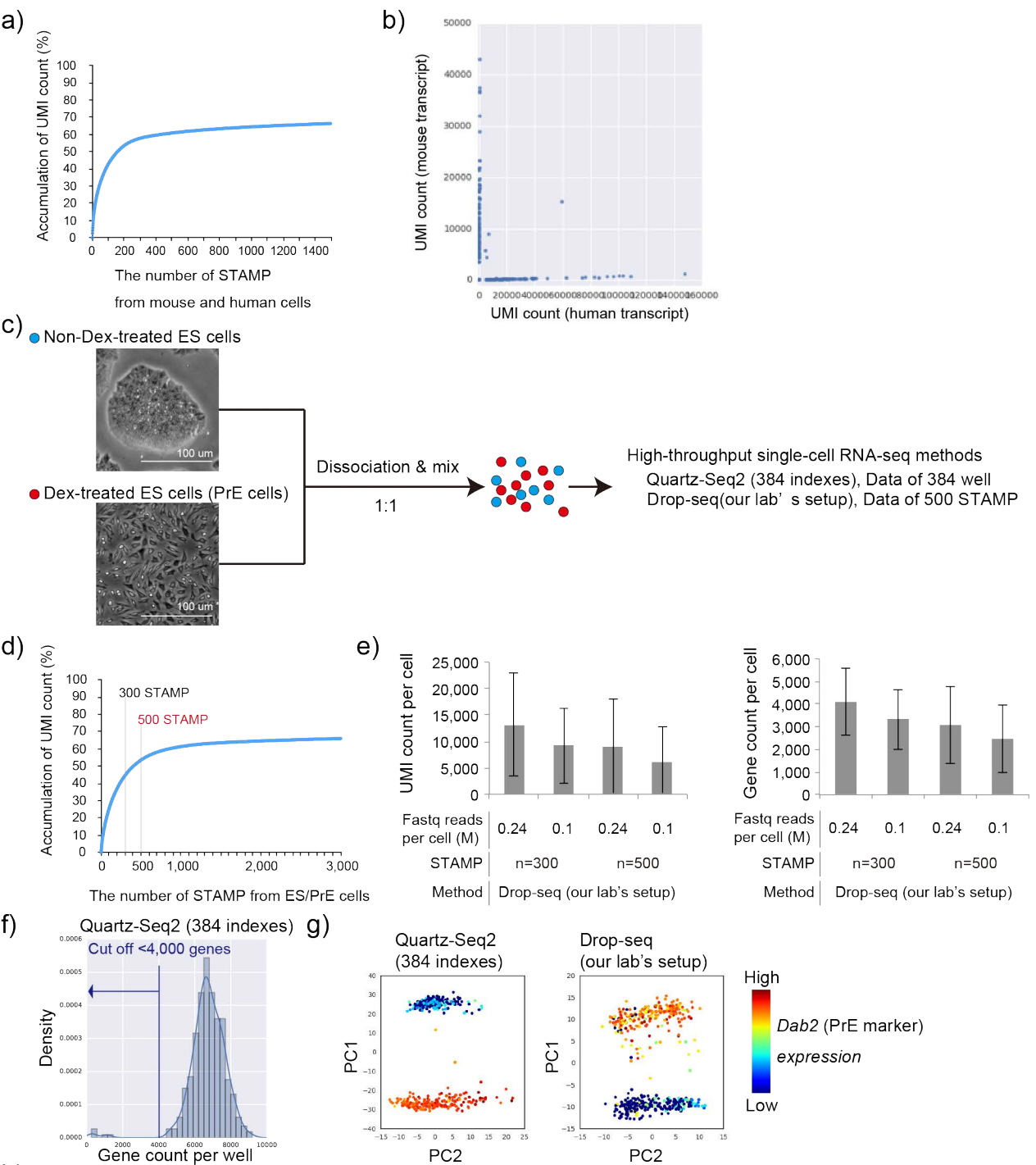
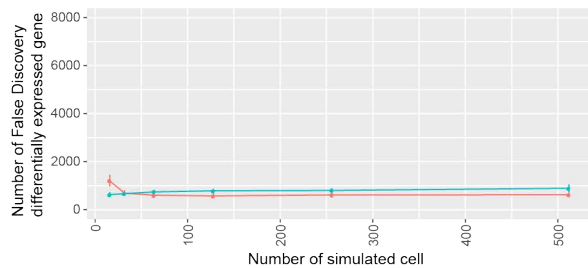
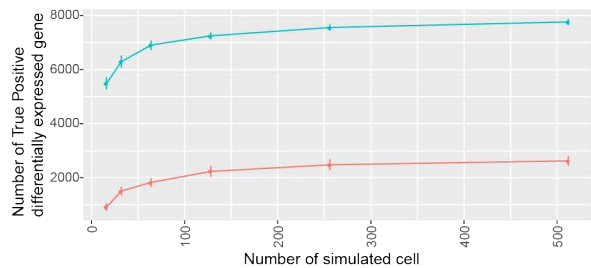
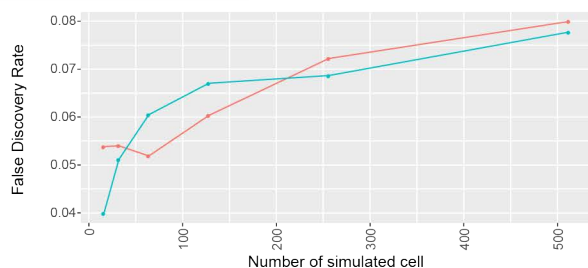
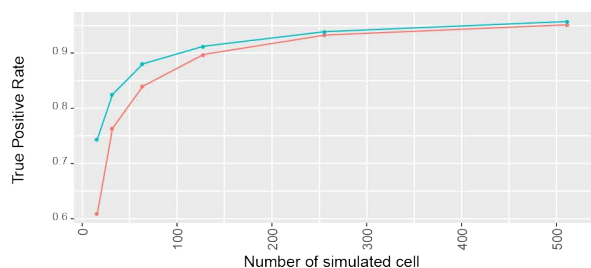


Fig. S12

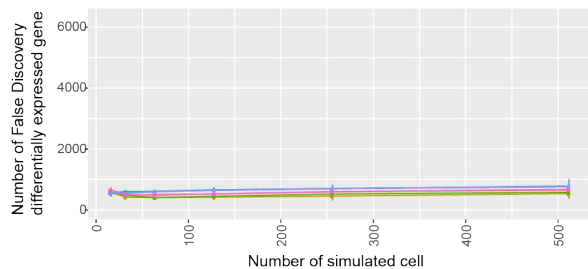
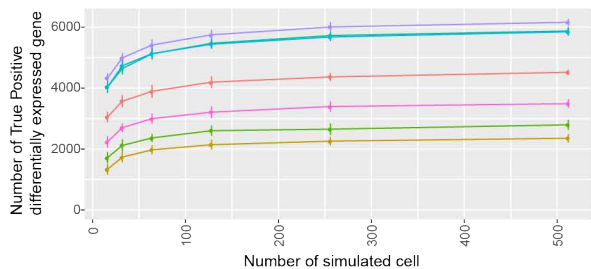
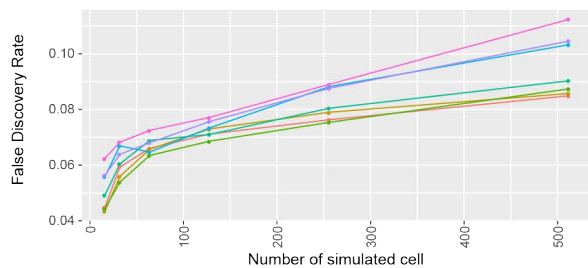
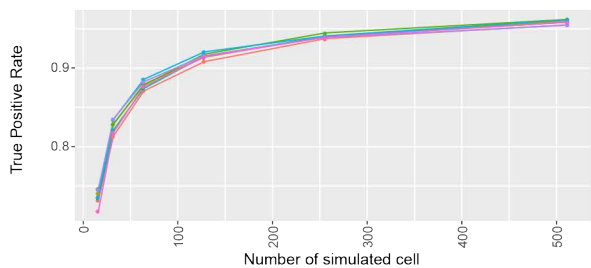
a)

Method	Sample	Data	Related figure
Drop-seq	G6GR ES cell (non-dex treated) n=192	Our data	Figure 3
Quartz-Seq2	G6GR ES cell (non-dex treated) n=259	Our data	Figure 3



b)

Method	Sample	Data	Related figure
Quartz-Seq2 (RT25) Day1	J1 ES cell (2i/LIF) n=1,152	Our data	Figure 4
Quartz-Seq2 (RT25) Day2	J1 ES cell (2i/LIF) n=768	Our data	Figure 4
Quartz-Seq2 (RT6.25, Column1)	J1 ES cell (2i/LIF) n=768	Our data	Figure 4
CEL-seq2	J1 ES cell (2i/LIF) n=71	<i>Ziegenhain et al.2017</i>	Figure 4
SCR-seq	J1 ES cell (2i/LIF) n=84	<i>Ziegenhain et al.2017</i>	Figure 4
MARS-seq	J1 ES cell (2i/LIF) n=65	<i>Ziegenhain et al.2017</i>	Figure 4
Drop-seq	J1 ES cell (2i/LIF) n=76	<i>Ziegenhain et al.2017</i>	Figure 4



c)

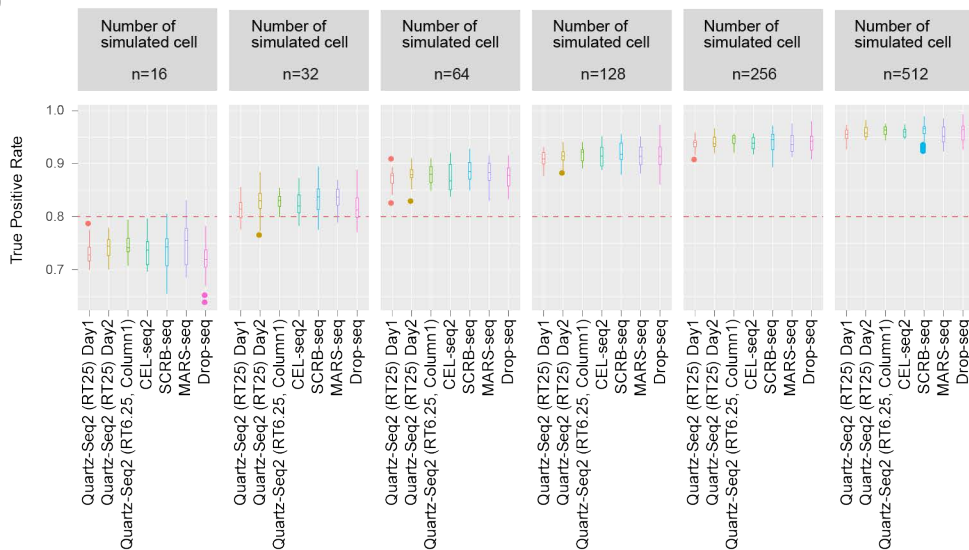
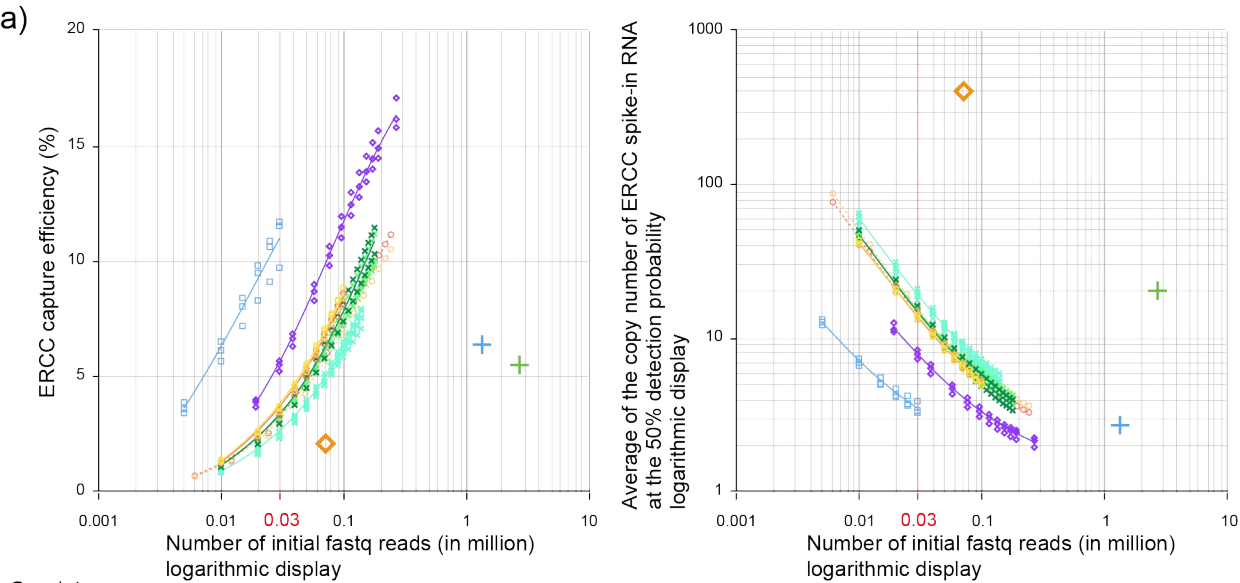


Fig. S13

a)



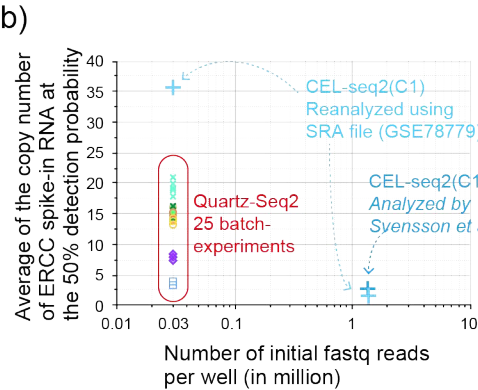
Our data

Platform	Cell-type	Method	RT condition	Type of Index	Cost of library DNA preparation
Cell-sorter	10 pg total RNA	Quartz-Seq2 n=1,152 3 batches (384 well x 3 plates)	RT25	384 indexes	¥61.3-63.3/well
Cell-sorter	Stromal vascular fraction	Quartz-Seq2 n=1,152 3 batches (384 well x 3 plates)	RT25	384 indexes	¥61.3-63.3/well
Cell-sorter	G6GR ES cell	Quartz-Seq2 n=192 1 batch (384 well x 0.5 x 1 plate)	RT25	384 indexes	¥61.3-63.3/well
Cell-sorter	Dex-treated G6GR ES cell (PrE)	Quartz-Seq2 n=192 1 batch (384 well x 0.5 x 1 plate)	RT25	384 indexes	¥61.3-63.3/well
Cell-sorter	G6GR ES cell	Quartz-Seq2 n=2,304 3 batches (384 well x 0.5 x 12 plates)	RT25	1,536 indexes	¥45.7-46.2/well
Cell-sorter	Dex-treated G6GR ES cell (PrE)	Quartz-Seq2 n=2,304 3 batches (384 well x 0.5 x 12 plates)	RT25	1,536 indexes	¥45.7-46.2/well
Cell-sorter	J1 ES cell (2i/LiF)	Quartz-Seq2 n=2,304 6 batches (384 well x 6 plates)	RT25	384 indexes	¥61.3-63.3/well
Cell-sorter	J1 ES cell (2i/LiF)	Quartz-Seq2 n=1,152 3 batches (384 well x 3 plates)	RT6.25	384 indexes	¥46.6-48.6/well
Cell-sorter	J1 ES cell (2i/LiF)	Quartz-Seq2 n=768 2 batches (384 well x 2 plates)	RT6.25 &Column1	384 indexes	¥34.9-36.9/well

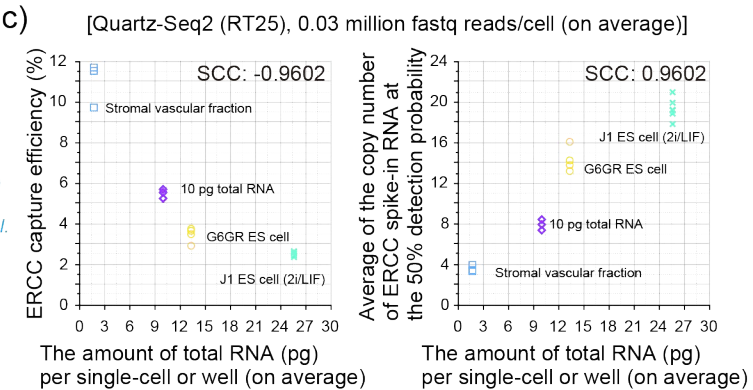
Data (digital expression matrix) from Svensson *et al. Nat. Methods. 2017;6:150.*

Platform	Cell-type	Method	Original data (Accession ID & Reference)
Fluidigm C1	Fibroblast	CEL-seq2(C1) n=96	GSE78779 SRP070989, Hashimshony <i>et al</i> 2016
Cell-sorter	Fibroblast	CEL-seq2 n=44	GSE78779 SRP070989, Hashimshony <i>et al</i> 2016
Cell-sorter	Dendritic cell	MARS-seq n=384	GSE54006 SRP035326, Jain <i>et al</i> 2014

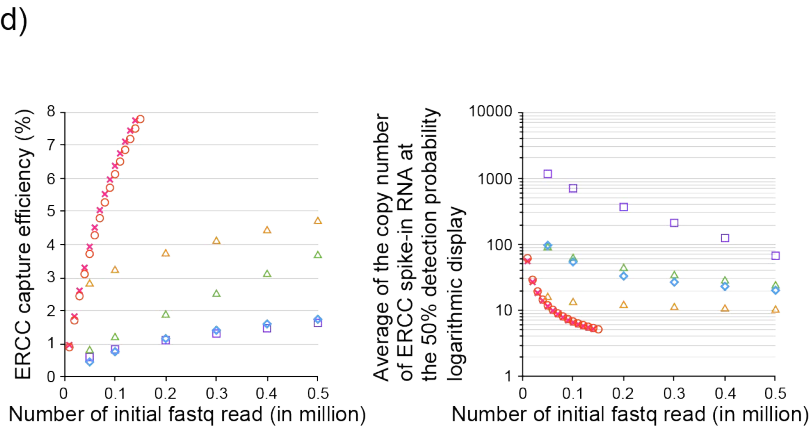
b)



c)



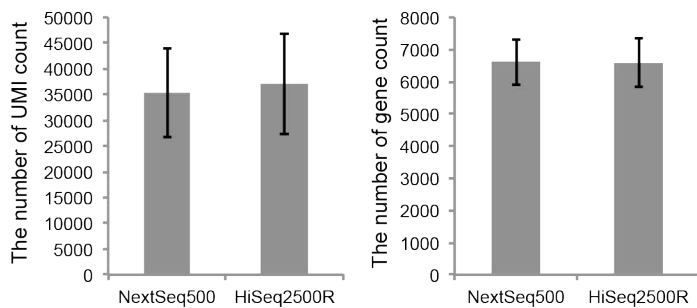
d)



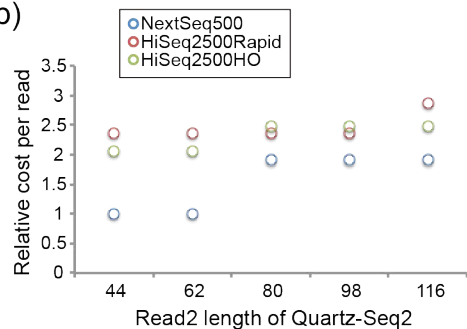
J1 ES cell 2i/LiF, HiSeq platform	
Our data	Quartz-Seq2 day1 (n=1,152) Quartz-Seq2 day2 (n=768) CEL-seq2(C1) (n=71)
Data from Ziegenhain <i>et al</i> 2017	SCRB-seq (n=84) MARS-seq (n=65) Drop-seq (w/o ERCC) (n=76)
without cell (ERCC data), NextSeq platform	
Data from Macosko <i>et al.</i> 2015	Drop-seq (w ERCC) (n=75)

Fig. S14

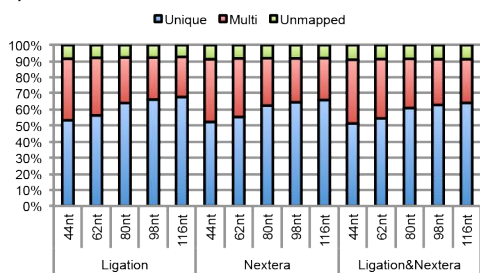
a)



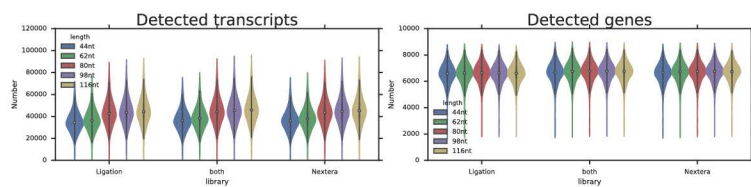
b)



c)



d)



e)

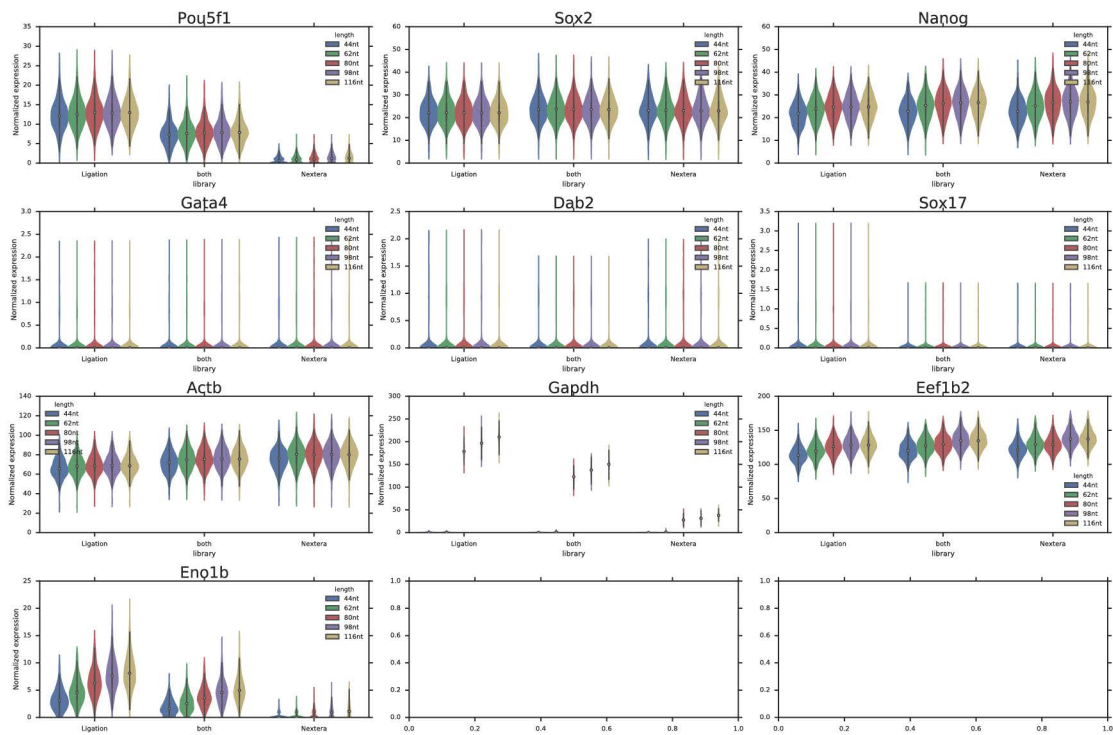


Fig. S15

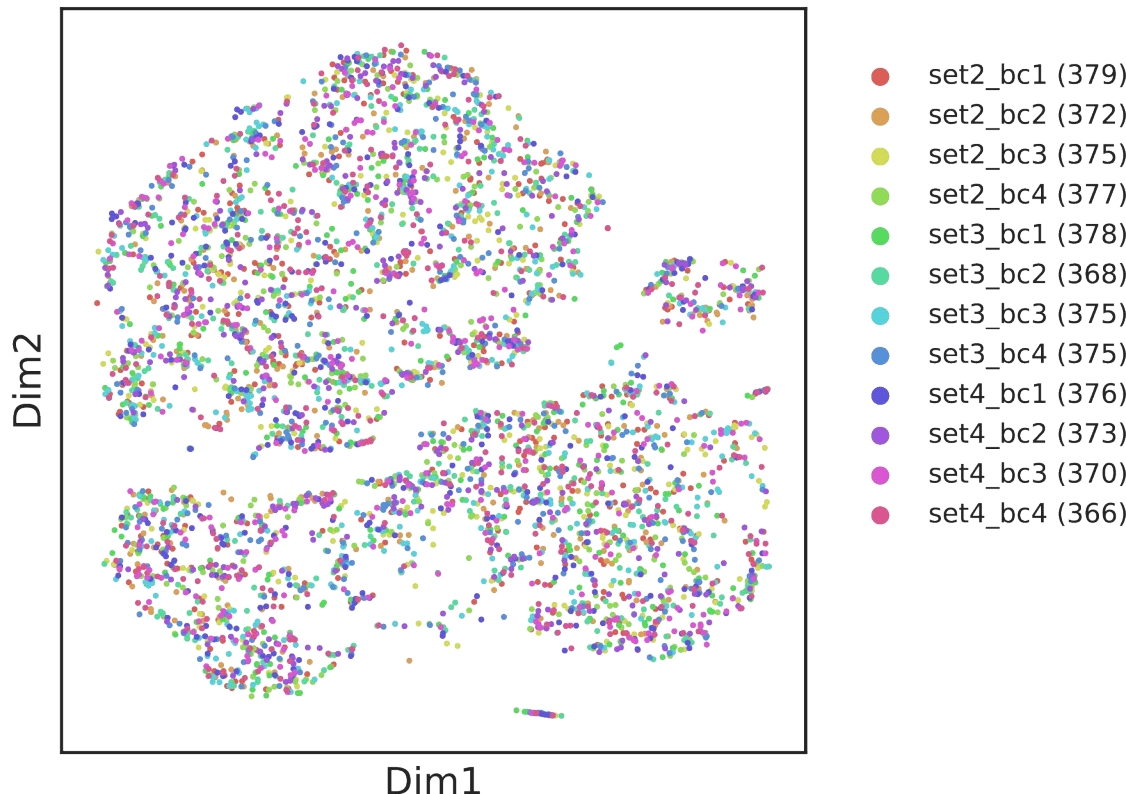
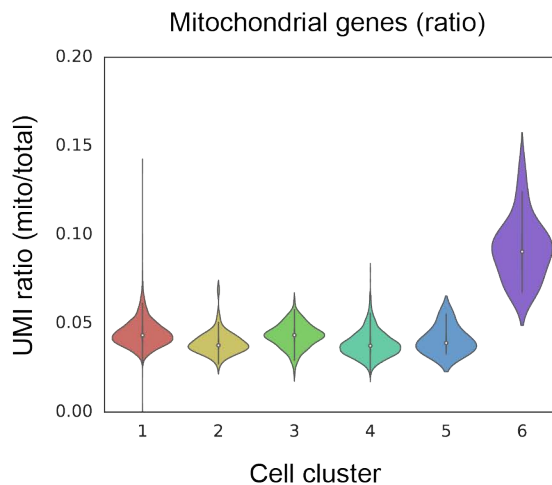
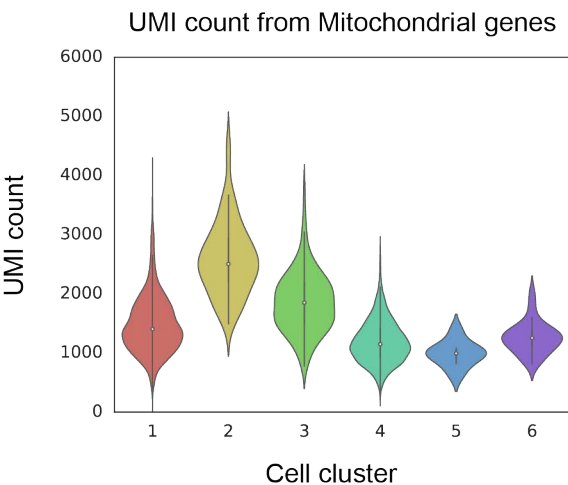
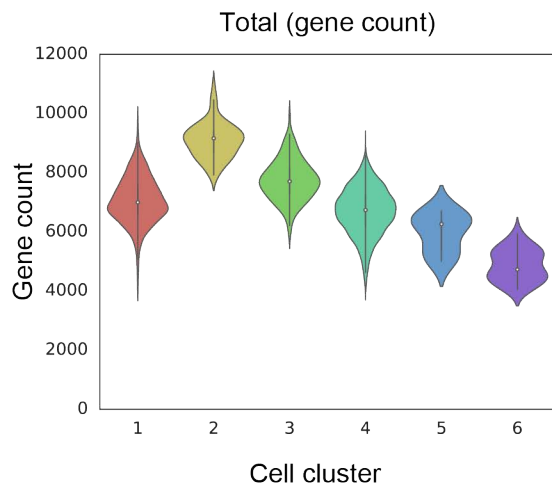
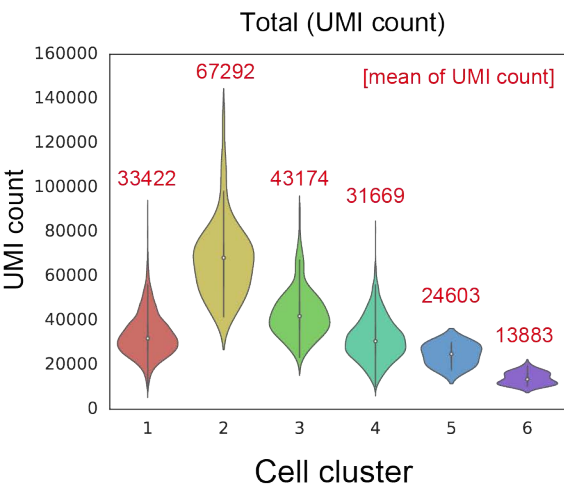


Fig. S16

- 1: ES (main), 44.6% (2,000 cells)
- 2: ES/PrE doublets, 1.38% (62 cells)
- 3: ES (*Stmn2* high), 3.3% (148 cells)
- 4: PrE (main), 49.6% (2,225 cells)
- 5: *Zscan4c/d* high, 0.17% (8 cells)
- 6: ES (low-quality cell), 0.57% (26 cells)



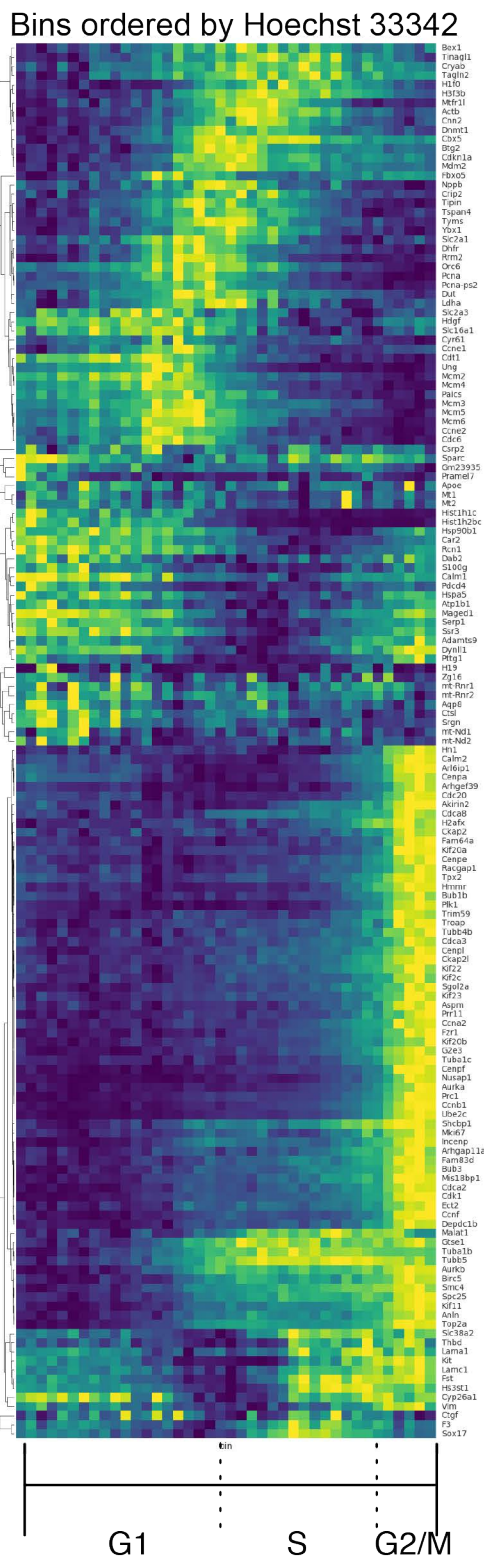
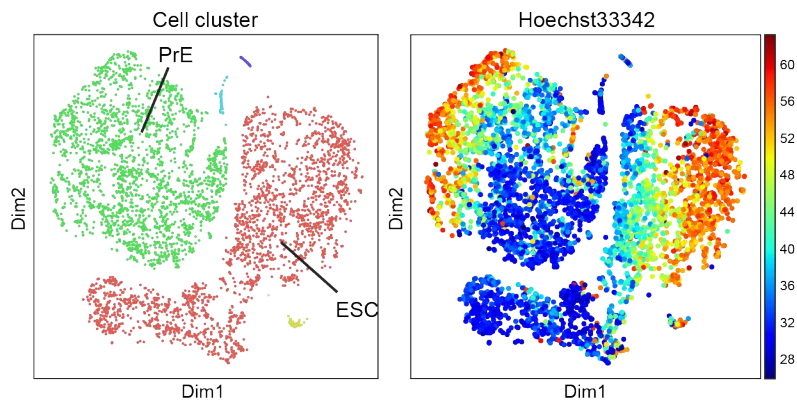
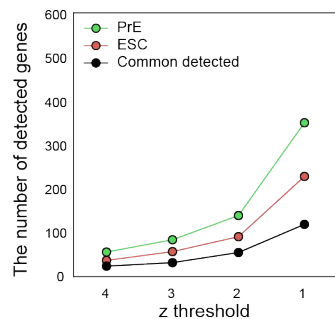


Fig. S18

a) 0.01M fastq reads/cell



b)



c)

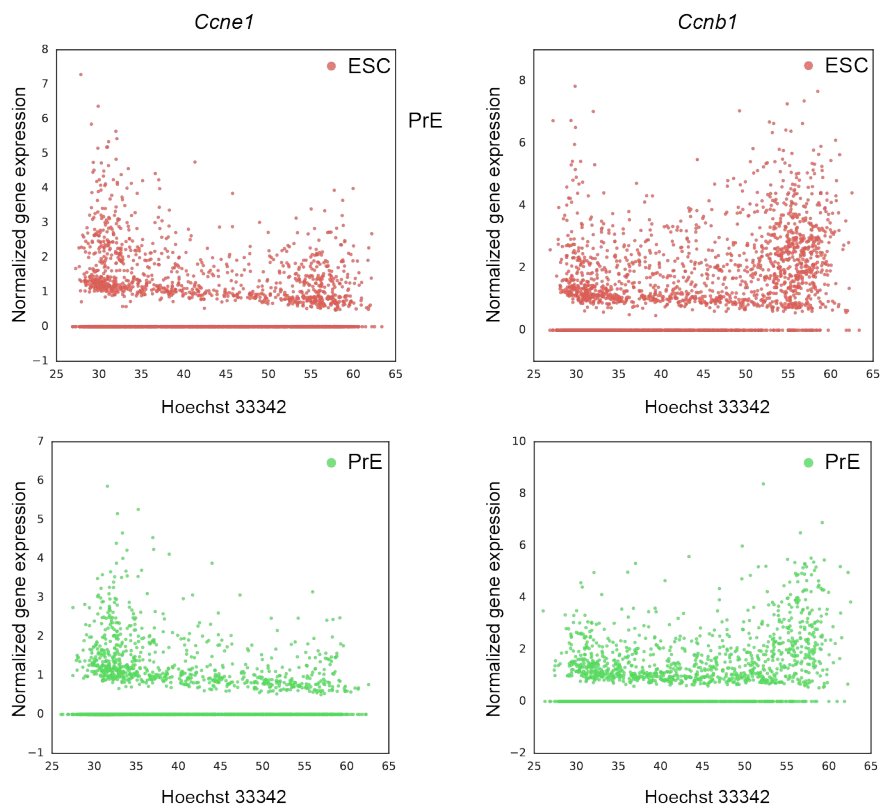
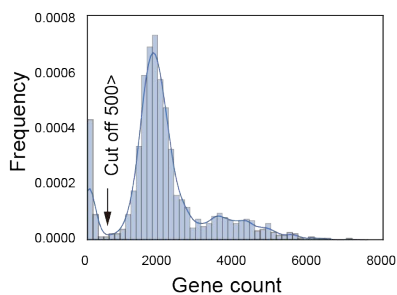
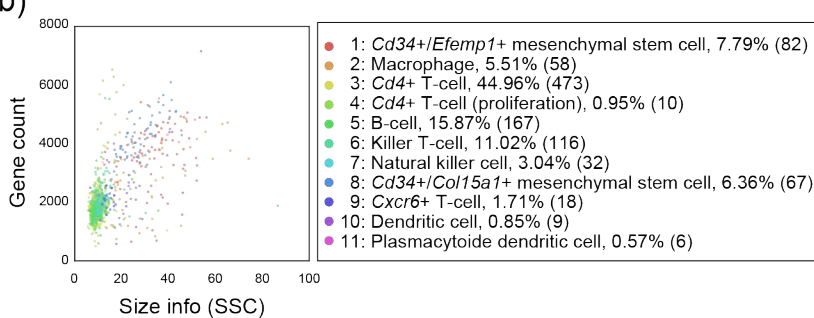


Fig. S19

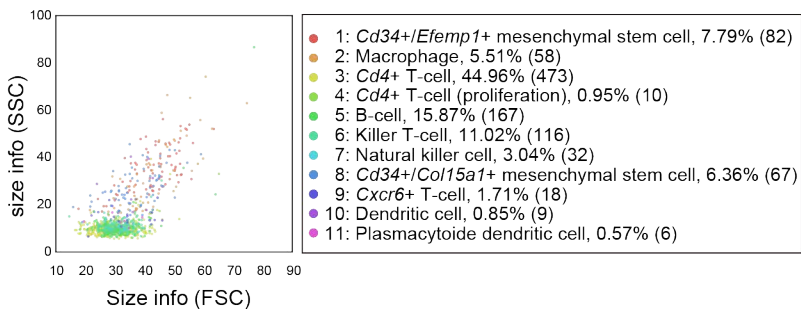
a)



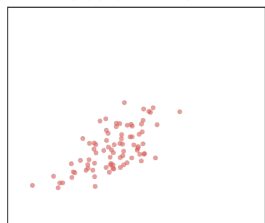
b)



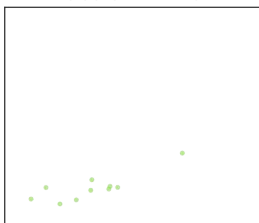
c)



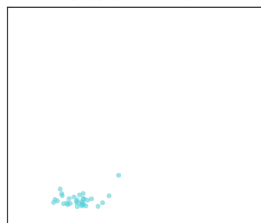
cluster1: n=82



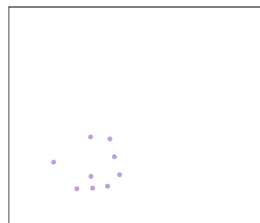
cluster4: n=10



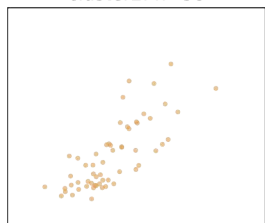
cluster7: n=32



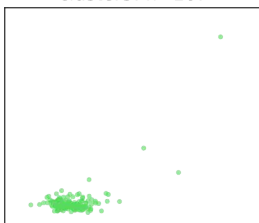
cluster10: n=9



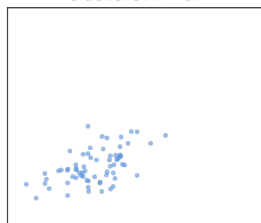
cluster2: n=58



cluster5: n=167



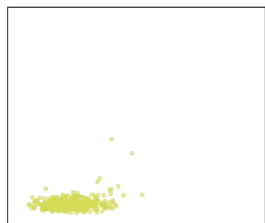
cluster8: n=67



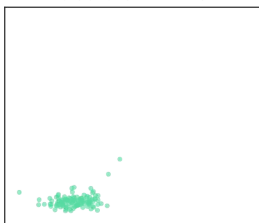
cluster11: n=6



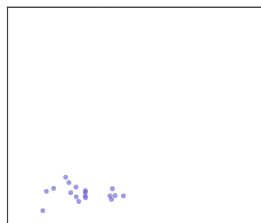
cluster3: n=473



cluster6: n=116



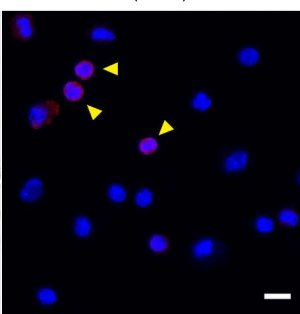
cluster9: n=18



d)

Bright field

Anti-CD4 (Red)
DAPI (Blue)



Bright field

Anti-CD79 (Red)
DAPI (Blue)

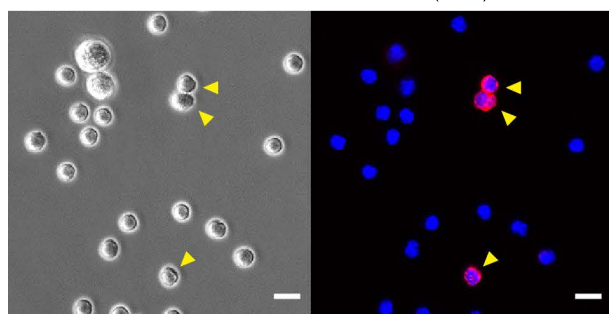


Fig. S20



Signatures

Fig. S22

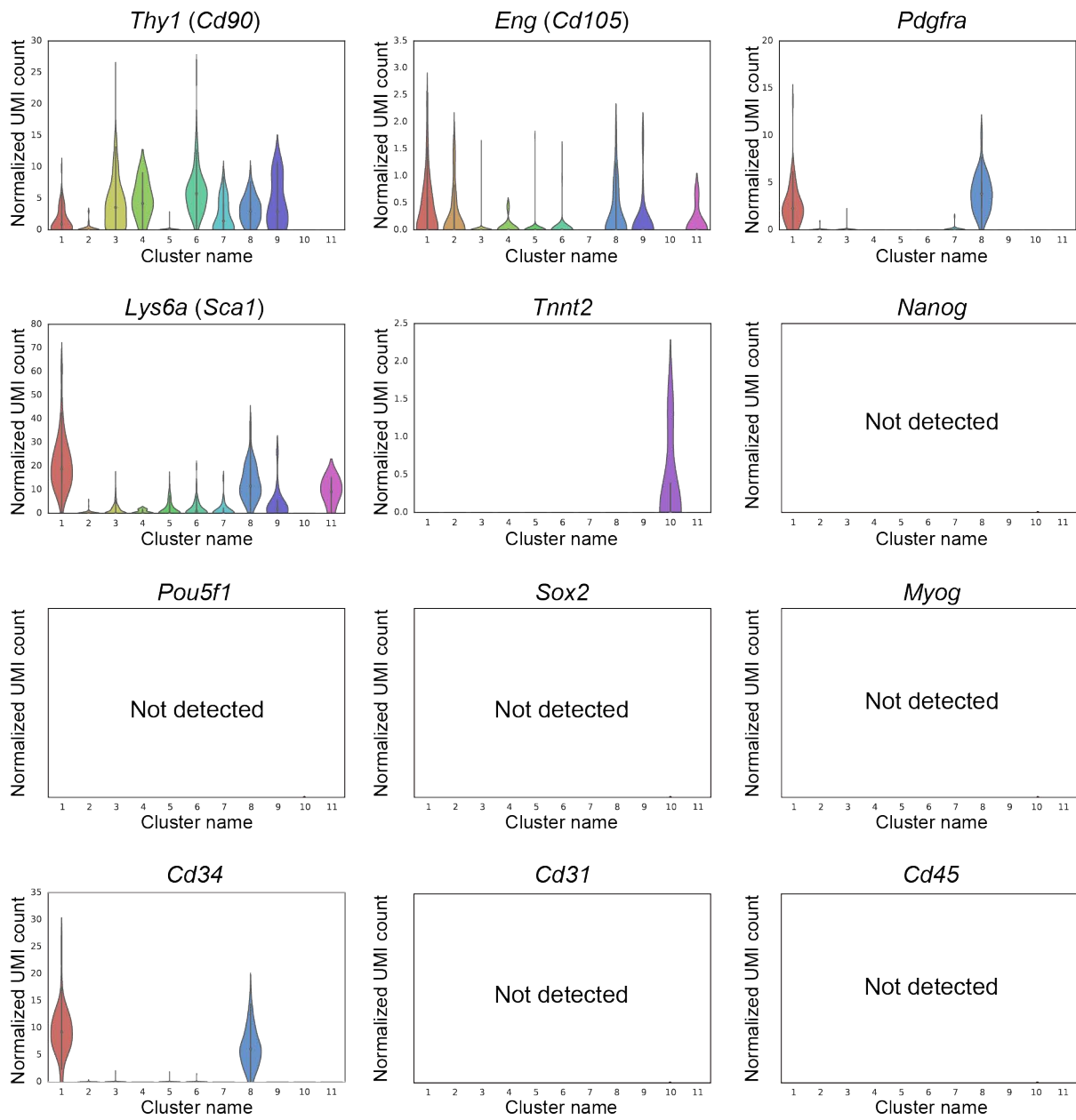


Fig. S23

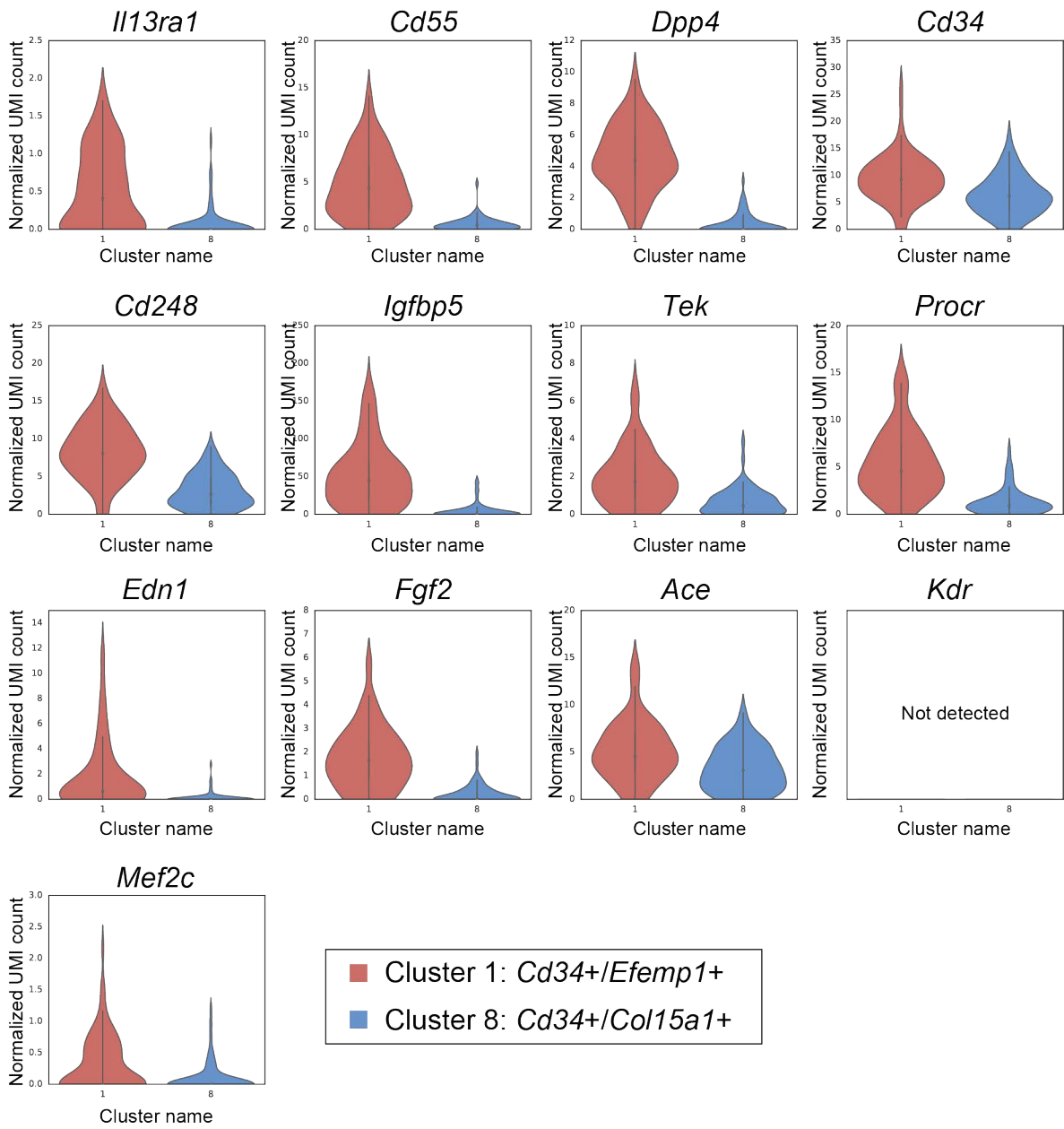
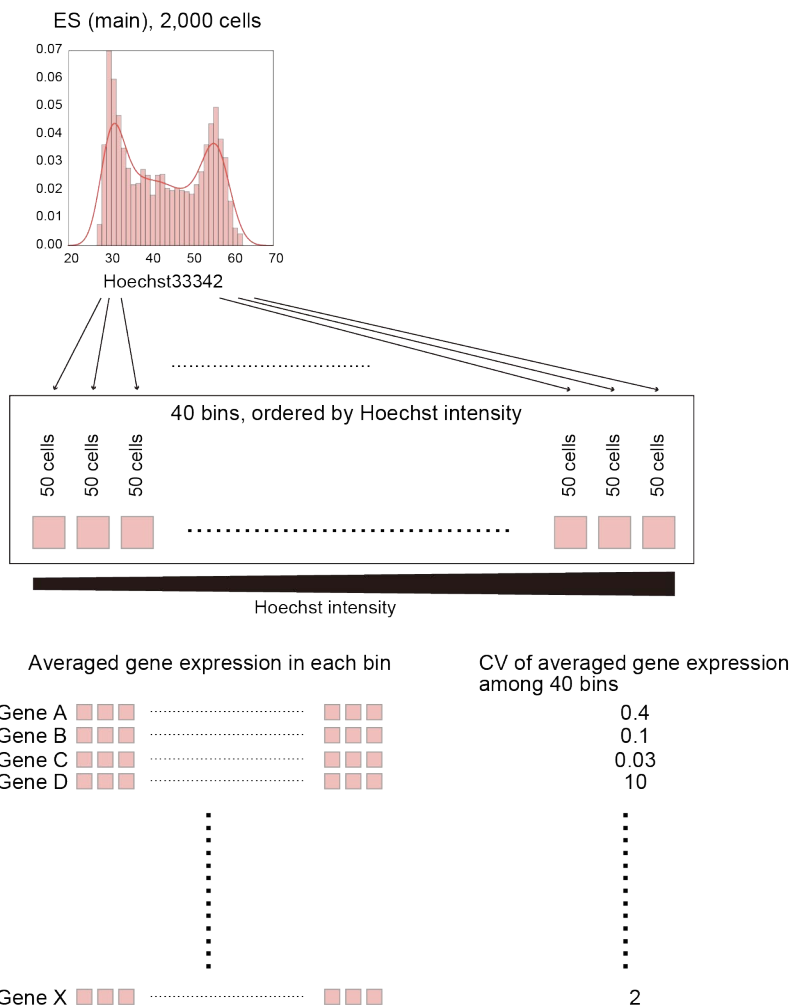


Fig. S24

a)



b)

